2013 IEEE XXXIII
International Scientific Conference
ELECTRONICS AND NANOTECHNOLOGY
(ELNANO)

CONFERENCE PROCEEDINGS

April 16-19, 2013
Kyiv, Ukraine

2013 IEEE XXXIII International Scientific Conference Electronics and Nanotechnology (ELNANO)

Copyright © 2013 by the Institute of Electrical and Electronics Engineers, Inc.
All rights reserved.

Copyright and Reprint Permission

Copyright and Reprint Permission: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Operations Center, 445 Hoes Lane, Piscataway, NJ 08854.

All rights reserved. Copyright @2013 by IEEE.

IEEE Catalog Number: CFP13O5U-CDR
ISBN: 978-1-4673-4671-9

Additional copies of this publication are available from
ELNANO 2013 ORGANIZING COMMITTEE
Faculty of Electronics,
National Technical University of Ukraine “KPI”
Polytekhnichna Str., 16, Block 12, off. 405,
03056, Kyiv, Ukraine

GENERAL ORGANIZERS

NATIONAL TECHNICAL UNIVERSITY OF UKRAINE «KYIV POLYTECHNIC INSTITUTE»

INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

ORGANIZERS

V.YE. LASHKARYOV INSTITUTE OF SEMICONDUCTOR PHYSICS OF NAS OF UKRAINE (ISP)

NATIONAL AVIATION UNIVERSITY

PUBLIC ORGANIZATION «LADY SCIENCE

GENERAL SPONSOR

Teleoptic PRA, Ltd.

IEEE SPONSORS

IEEE Region 8

IEEE Ukraine Section

IEEE Central Ukraine (Kyiv)
Joint Chapter ED/MTT/COM/CPMT/SSC

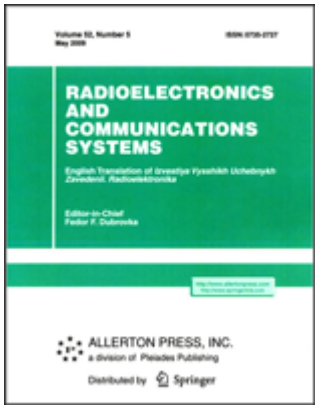
IEEE Central Joint Chapter SP/AES

IEEE East Joint Chapter AP/MTT/ED/AES/GRS/NPS

IEEE KPI Student Branch

MEDIA PARTNERS

International scientific journal
"RADIOELECTRONICS AND COMMUNICATIONS SYSTEMS"



International scientific journal
"ELECTRONICS AND COMMUNICATIONS"

COMMITTEES

STEERING COMMITTEE

Chair:

Prof. Yuriy Yakymenko	Academician, First Vice Rector of National Technical University of Ukraine "Kyiv Polytechnic Institute" (NTUU "KPI"), Head of Microelectronics Department, Faculty of Electronics, IEEE Senior Member, Ukraine

Co-Chairs:

Prof. Alexander Borisov	First Vice Dean of Faculty of Electronics, NTUU "KPI", Kyiv, Ukraine
Dr. Volodymyr Machulin	Academician, V. Ye. Lashkaryov Institute of Semiconductor Physics of NAS, Ukraine
Prof. Volodymyr Timofeyev	Head of Physical and Biomedical Electronics Department of NTUU "KPI", IEEE Senior Member, Ukraine
Prof. Klaus-Jurgen Wolter	Director of IAVT and ZμP, Technical University of Dresden, IEEE Senior Member, Germany
Prof. Felix Yanovsky	Head of Electronics Department of National Aviation University (NAU), IEEE Fellow, Ukraine
Prof. Valerii Zhuikov	Dean of Faculty of Electronics, NTUU "KPI", Chairperson of IEEE Ukraine Section, Kyiv, Ukraine

ORGANIZING COMMITTEE

General Chair:	**Dr. Anton Popov,** NTUU "KPI", Counselor of IEEE KPI SB, Ukraine
Technical Program Chair:	**Dr. Tetiana Khyzhniak,** NTUU "KPI", Ukraine
Technical Program Vice Chair:	**Ievgen Pichkalov,** NTUU "KPI", Secretary of IEEE Ukraine Section, Ukraine
Publication Committee Chair:	**Dr. Kateryna Ivanko,** NTUU "KPI", Ukraine
Publication Committee Member:	**Dr. Helen Semenovskaya,** NTUU "KPI", Ukraine
Publication Committee Member:	**Anna Poreva,** NTUU "KPI", Ukraine
Finance Chair:	**Prof. Julia Yamnenko,** NTUU "KPI", Director of Fund "Lady of Science", Vice Chair of IEEE Ukraine Section, Ukraine
Finance Vice Chair:	**Anna Kyselova,** NTUU "KPI", Treasurer of IEEE Ukraine Section, Ukraine
Electronic Communications Coordinator:	**Dr. Yuriy Vountesmery,** NTUU "KPI", Ukraine

TECHNICAL PROGRAM COMMITTEE

Prof. Ahmed Al-Shamma'a, Liverpool John Moores University, UK

Prof. Necmi Biyikli, Bilkent University, Ankara, Turkey

Prof. Maurizio Bozzi, University of Pavia, Italy

Prof. Hung T. Diep, University of Cergy-Pontoise, France

Prof. Wolf-Joachim Fischer, Technical University of Dresden, Germany

Dr. Patrick Henaff, University of Cergy-Pontoise, France

Dr. Dirk Van Hertem, KU Leuven, Belgium

Dr. Olga Korostynska, Liverpool John Moores University, UK

Prof. Igor Korotyeyev, University of Zielona Gora, Poland

Dr. Oleg Krasnov, Delft University of Technology, The Netherlands

Prof. Krzysztof Kulpa, Warsaw University of Technology, Poland

Prof. Sergey Edward Lyshevski, Rochester Institute of Technology, USA

Prof. Hagen Malberg, Technical University of Dresden, Germany

Dr. Alex Mason, Liverpool John Moores University, UK

Dr. Oleg Mironov, University of Warwick, Coventry, UK

Dr. Tamer M. Nassef, Misr University for Science and Technology, Egypt

Prof. Nenad Novkovski, University St. Cyril and Methodius in Skopje, Macedonia

Prof. Nikolay Potrakhov, Saint Petersburg Electrotechnical University "LETI", Russia

Dr. Anton Popov, NTUU “KPI”, Kyiv, Ukraine

Prof. Dr. Juergen Schreiber, Technical University of Dresden, Germany

Prof. Saurabh Sinha, University of Pretoria, South Africa

Prof. Volodymyr Timofeyev, NTUU “KPI”, Kyiv, Ukraine

Dr. Dmitri Vinnikov, Tallinn University of Technology, Estonia

Prof. Klaus-Jurgen Wolter, Technical University of Dresden, Germany

Prof. Felix Yanovsky, NAU, Kyiv, Ukraine

Dr. Janis Zakis, Tallinn University of Technology, Estonia

Dr. Sebastian Zaunseder, Technical University of Dresden, Germany

Prof. Valeriy Zhuikov, NTUU “KPI”, Kyiv, Ukraine

TECHNICAL PROGRAM SUBCOMMITTEE

Prof. Emir Aznakaev, NAU, Kyiv, Ukraine

Anuj Bhatia, Thermo Fisher Scientific, USA

Prof. Anatolii Bilous, Institute of general and inorganic chemistry, Ukraine

Dr. Artem Fediai, NTUU "KPI", Kyiv, Ukraine

Oleksandr Husev, NTUU "KPI", Kyiv, Ukraine

Dr. Kateryna Ivanko, NTUU "KPI", Kyiv, Ukraine

Dr. Natalia Ivanushkina, NTUU "KPI", Kyiv, Ukraine

Dr. Yevgeniy Karplyuk, NTUU "KPI", Kyiv, Ukraine

Dr. Victor Kazmirenko, NTUU "KPI", Kyiv, Ukraine

Prof. Oleksandr Kovalenko, Institute of cybernetics, Kyiv, Ukraine

Dr. Tetiana Khyzhniak, NTUU "KPI", Kyiv, Ukraine

Anna Kyselova, NTUU "KPI", Kyiv, Ukraine

Olga Kyselova, NTUU "KPI", Kyiv, Ukraine

Dr. Viktor Lopata, NTUU "KPI", Kyiv, Ukraine

Prof. Oleksandr Lysenko, NTUU "KPI", Kyiv, Ukraine

Prof. Igor Melnik, NTUU "KPI", Kyiv, Ukraine

Dr. Olexadr Melnik, NAU, Kyiv, Ukraine

Prof. Sergey Miroshnichenko, Teleoptic PRA, NAU, Kyiv, Ukraine

Stanislav Mudriievskyi, Technical University of Dresden, Germany

Dr. Nikolai Nikolov, NTUU "KPI", Kyiv, Ukraine

Dr. Anatolii Orlov, NTUU "KPI", Kyiv, Ukraine

Ievgen Pichkalov, NTUU "KPI", Kyiv, Ukraine

Dr. Yuriy Prokopenko, NTUU "KPI", Kyiv, Ukraine

Prof. Yuriy Poplavko, NTUU "KPI", Kyiv, Ukraine

Prof. Petro Stakhiv, Lviv Polytechnic National University, Lviv, Ukraine

Prof. Volodymyr Ulansky, NAU, Kyiv, Ukraine

Dr. Oleg Vityaz, NTUU "KPI", Kyiv, Ukraine

Dr. Yuriy Vountesmery, NTUU "KPI", Kyiv, Ukraine

Prof. Julia Yamnenko, NTUU "KPI", Kyiv, Ukraine

Jean van Niekerk, University of Pretoria, South Africa

Wynand Lambrechts, University of Pretoria, South Africa

Paul Ellinghaus, University of Pretoria, South Africa

Bongani Mabuza, University of Pretoria, South Africa

Tinus Stander, University of Pretoria, South Africa

Adriaan de Kamper, University of Pretoria, South Africa

Antonie Alberts, University of Pretoria, South Africa

C O N T E N T S

SECTION I. NANO- AND MICROELECTRONICS

Microwave Characterization of Dielectric Substrates for Thin Films Deposition *17*
Viktor Bovtun, Martin Kempa, Stanislav Kamba, Valeriy Pashkov, Vitaliy Molchanov Yuriy Poplavko, Yuriy Yakymenko

Broadband Dielectric and Conductivity Spectra of Dielectric – metal Nanocomposites for Microwave Applications *21*
A. Borisova, A. Machulyansky, Y. Yakimenko, V. Bovtun, M. Kempa, M. Savinov

Synthesis of ZnO Nanorods for Acoustic Wave Sensor *25*
Anatolii Orlov, Veronika Ulianova, Yuriy Yakimenko, Oleksandr Bogdan, Gennady Pashkevich

Electrochemical Design of Zinc Oxide Arrays for Dye-Sensitized Solar Cells *28*
E.I. Sokol, N.P. Klochko, G.S. Khrypunov, V.R. Kopach, K.S. Klepikova, V.M. Lyubov, A.V. Kopach, Y.O. Myagchenko, E.E. Melnychuk

Information-Theoretic Estimates of Communication and Processing in NanoScale and Quantum OptoElectronic Systems *33*
Sergey Edward Lyshevski, Leonid Reznik

Nano and Molecular Technologies in Microelectronics, MEMS and Electronic Systems *38*
Sergey Edward Lyshevski

Micromechanical Tuning of Microstrip Antenna in Frequency Domain *43*
N. Ruda, Y. Prokopenko, Y. Poplavko

Thermal Resistance Of Power Submicron Heterojunction Field-Effect Transistors *47*
Vladimir Timofeyev, Elena Semenovskaya

Structural and Electrical Characterization of SiGe Heterostructures Containing a Pure Ge Strained Quantum Well *51*
A.H.A. Hassan1, O.A. Mironov, A. Dobbie, .J.H. Morris, J.E. Halpin, V.A. Shah, M. Myronov, D.R. Leadley, S. Gabani, A. Feher, E. Cizmar, V.V. Andrievskii, and I.B. Berkutov

Ge/GaAs Thin Films for Thermometer and Bolometer Application *56*
V.F. Mitin, P.M. Lytvyn, V.V. Kholevchuk, V.V. Mitin, E.F. Venger, O.A. Mironov

Mathematical Simulation of Optical Nanoantenna Based on a Comb-Like Finite Nanostrip Grating *61*
Olga V. Shapoval, Jiří Čtyroký, Alexander I. Nosich

Modeling of 2-D Microcavity Lasers with Arbitrarily Shaped Active Regions *66*
Elena I. Smotrova, Melanie Lebental, Alexander I. Nosich

Electromagnetic Characterization of a Silver Nanowire Grating Embedded into a Dielectric Slab *70*
Volodymyr O. Byelobrov, Trevor M. Benson, Alexander I. Nosich

Accurate Numerical Study of Graphene Disk Surface Plasmon Resonances *73*
Mikhail V. Balaban, Ana Vukovic, Trevor. M. Benson, Alexander I. Nosich

Carbon-Rich Nanostructured a-SiC for Cold Emitters *76*
Gordienko S. A., Nazarov A.N., Lytvyn P.M., Stadnik A.A., Gomeniyuk Yu.Yu., Rusavsky A.V., Vasin A.V., Stepanov V.G., Lysenko V.S., Nazarova T.M.

Negative Differential Resistance Based Voltage-Controlled Oscillator for VHF Band *80*
V.V. Ulansky, S.F. Ben Suleiman

Optimization of LC Voltage-Controlled Oscillators in 90-nm CMOS Technology for 3G Transceivers *85*
V.V. Ulansky, H.M. Elsherif

Thermal Dependence of "Metal-Polymer" Type Composite Materials' Dielectric Properties in Ultra-high Frequency Band *90*
Didenko Yu.V., Patsora I.V., Tatarchuk D.D., Franchuk A.S., Kharabet Ie.I.

Novel Concept for Microstrip Stub Resonant Frequency Control *94*
Pavlo Sergienko, Yuriy Prokopenko, Guy A. E. Vandenbosch

The Three-Collector Magnetotransistor: Variable Sensitivity *99*
V.V. Amelichev, R.D. Tikhonov, A.A. Cheremisinov

Porous Silicon Formation on Solar Cells Surface *104*
Tetyana Obukhova

Modeling of Resonant-Tunneling Diodes with Uniform and Graded Emitter *107*
Artem Fediai, Volodymyr Moskaliuk

Functional Film Peculiarities of the "Metal-Insulator-Metal" Structure Devices *112*
N. Korobova, V. Vodopyanov, Yu. Stepanova, Yu. Cherkasova, S. Timoshenkov, A. Melnikov, O. Isaikina, N. Plis

Optimization of Phosphorous Diffusion in Silicon Photocells *115*
Maksym Klymov

Mode Matching Technique for Tunable Shielded Cylindrical Metal-Dielectric Resonator *118*
Kostiantyn Savin, Yuriy Prokopenko, Guy A. E. Vandenbosch

Modeling of Velocity "Overshoot" in the Multivalley Semiconductors *123*
Konstantin Bol, Vladimir Moskalyuk

Receiver/Transmitter Input/Output Termination Resistance Calibration Method *126*
Vazgen Melikyan, Abraham Balabanyan, Artak Hayrapetyan, Nazeli Melikyan

Transmission Spectra of the Graphene-based Fibonacci Superlattice *131*
A.M. Korol, V.M. Isai

Optical and Luminescence Effects in Cu, Au-doped Cadmium Iodide layered Nanostructures *134*
Mykola Rudka, Vladimir Antonyuk, Natalia Stetsyk

ZnO as a Conductive Layer for SolarCells Based on II-VI materials *137*
Semikina T.V., Mamykin S.V., Shmyryeva L.N., Godlewski M., Luka G., Pietruszka R., Kopalko K., Krajewski T.A., Gierałtowska S., Wachnicki Ł.

3D IC Cooling Mechanism by Using Signaling Vias *140*
Ara Gevorgyan

Improvement of Piezoceramic Scanners *144*
Sharapov V.M., Filimonov S.A., Sotula Zh.V., Bazilo K.V., Kunitskaya L.G., Zaika V.M.

Mathematical Modeling of Electromagnetic Wave Propagation in Inhomogeneous Medium *147*
I. Yu. Dmitrieva

Modern Printing Technologies in Micro- and Nanoelectronics *151*
Olha Sarapulova, Tetiana Kyrychok, Valentyn Sherstiuk, Anatolii Orlov

Modeling of the Dielectric Barrier RF Discharge at Atmospheric Pressure *156*
Chaplinskiy R.U.

Timing Yield and Reliability Improvement of Carbon Nano-tube FET Based Digital Circuits with Statistical Driven *160*
Amir Jalali, Hossein Pedram

Nonlinear Beams and Active Controllable Field Concentrator with Isotropic Metamaterials *165*
Yuriy Rapoport, Nikita Kalinich, Allan Boardman, Vladimir Grimalsky

Three-Level Approach to Graphene Metamaterials: Electron Density Waves and Linear and Nonlinear Electrodynamics *169*
Yuriy Rapoport ,Nikita Kalinich, Vladimir Grimalsky, Igor Nefedov, Igor Nefedov

Two-Channel Heterotransistors With Quantum Dots Systems *172*
Vladimir Timofeyev, Elena Faleyeva

The Ways of Silicon Carbide Usage in Field-emission Devices: The Technological Aspect *177*
N. Shelepin, I. Matyushkin, S. Orlov, A. Ermakov, K.Svechkarev, P. Bobovnikov, A. Mikhaylov, A. Belov

Photonic Crystal Filter with Fractal Structure 2D Modeling *181*
A.V. Vishnevsky, D.E. Aznakayeva

Controllable Growth of Copper Fractal Aggregates on Structurally Modified Silicon Surface *185*
Anton Boyko, Sergei Timoshenkov, Dmitry Litmanovich, Dahir Gaev

Periodicity Assisted Scattering and Absorption of Light by Finite Layered Gratings of Silver Nanowires *188*
Denys M. Natarov, Alexander I. Nosich, Marian Marciniak, Ronan Sauleau

Modeling of Electronic Radiation Influence to Transistors *192*
A.A.Tuyakbayev, M.K.Baizhumanov, D.A.Tuyakbaev

Excitation Processes Modeling in Two-Layer Graphene *195*
E.G. Aznakayev, D.E. Aznakayeva

PerformanceAnalysisofMultiwalledCarbonNanotubeBundles *200*
Manodipan Sahoo, Hafizu Rahaman

Nanocrystalline Cerium Oxide Films as a Functional Material for Photodetectors of Bioluminescent Signal *205*
N.V. Maksimchuk, A.V. Borisov

Investigation of Self-Shifted Solitons as Part of Femtosecond Supercontinuum *208*
Y. Bashkatov, O. Kutsenko, B. Tsyganok, V. Khomenko

Development of Technology of Multicomponent Instrument Compositions Formation on the Basis of Porous silicon *212*
Zubko Evgeniya Ivanovna, Shvets Evgenii Yakovlevich

Multiphysics Effects and Electronic Properties of Anisotropic Semiconductor Quantum Dots *217*
Sanjay Prabhakar, Roderick Melnik

SECTION II. BIOMEDICAL ELECTRONOCS AND SIGNAL PROCESSING

Evaluation of Two Methods Based on EMD for Elimination of Blink from EEG *223*
Mohammad Shahbakhti

QRS Pattern Recognition Using a Simple Clustering Approach for Continuous Data *228*
A. Noack, R. Poll, W.-J. Fischer, S. Zaunseder

Investigation of the Relative Sensitivity of Control of Impulse X-ray Television Systems *332*
Mikhailov S.R., Slobodian N.V., Shilo D.S.

Novel Self-folding Electrode for Neural Stimulation and Recording *337*
Melinda Varga, Marco Luniak, Klaus-Jürgen Wolter

Magnetic Nanotherapeutics of Guerin Carcinoma *241*
Orel V.E., Nikolov N.A., Rykhalskiy A.Y., Dzyatkovskya N.N., Romanov A.V., Yaroshenko O.Yu., Supruniuk D.A., Orel V.B., Khamrovskiy MM., Penievskyi M.I., Radaguz A.V.

Stochastic Properties of Nephroscintigraphic Images with ^{99m}Tc-DMSA *246*
Nikolov N.A., Supruniuk D.A., Smetaniuk A.N.

Saturation of Electroencephalogram Permutation Entropy for Large Time Lags *251*
Anton Popov, Oleksii Avilov, Oleksii Kanaykin

Noninvasive Method for Blood Glucose Measuring and Monitoring *255*
Loshitskiy Pavel Pavlovich, Dmitriy Yurievich Mynziak

Influence of Non-Thermal Intensity EHF Radiation on Properties of Water and NaCl Aqueous Solutions *258*
D.Y. Mynziak, P.P. Loshitskiy, O. Korostynska, A. Mason and A. I. Al-Shamma'a

The Analysis of Technical Solutions for Medical Ozonators *262*
Sokol E.I., Kipenskiy A.V., Kulichenko V.V., Tomashevskiy R.S., Barkhotkina T.M.

The Impact of Magnetic Field on Blood Vessels *266*
M.M. Baran, N.V. Kostevska, V.P. Virchenko, Y.S. Synekop, A.I. Mukhomor

Barrier Properties of Polymer Encapsulation Materials for Implantable Microsystems *269*
Sabine Kirsten, Jakob Wetterling, Jürgen Uhlemann, Klaus-Jürgen Wolter, Sergej Zigler

Experimental Investigation of Aqueous Solutions Superradiance *273*
Nikolov N.A., Solyar A.G., Yaroshenko O.Yu.

Computer-Aided Software for Early Diagnosis of Eerythemato-squamous Diseases *276*
Bekir Karlık, Güneş Harman

Nanoparticles of Ferromagnetic Materials and Possibilities of their Application in the Hyperthermia of Malignant Tumors *280*
A. Belous, S. Solopan, A. Yelenich, L. Bubnovskaya, S. Osinsky R. E. Kavetsky

Combined Method for Detection of Atrial Late Potentials *285*
Matveyeva N. A., Ivanushkina N. G., Ivanko K. O.

Heat Production, Heat Transfer and Heat Exchange in Man During Water Immersion: Mathematical Modeling *290*
IrenaYermakova, Yuliia Solopchuk, Lyudmila Khudyakova

Automatic Pattern Recognition of Epileptiform Discharges Using Morphological Descriptors and Linear Discriminant Analysis *293*
Christine Fredel Boos, Fernando Mendes de Azevedo

Adaptive Algorithm for Continuous Monitoring of Blood Pressure Using a Pulse Transit Time *297*
Ali Jadooei, Zaderykhin O., Shulgin V.I.

Stimulating Light System to Create a Controlled Effect Relaxation *302*
Barylo G.I., Gotra Z.Yu., Zazulyak A.M., Kozhuhar O.T., Kus N.I.

Differentiation of Moist Fine Rales and Crackles by Polyspectral Analysis *305*
Anna Poreva, Vladimir Fesechko

Automated Parallel Detection Based Analyzer for Integrated Bioassays in Digital Microfluidic Biochip *310*
Pranab Roy, Mahua Raha Patra, Hafizur Rahaman, Parthasarathi Dasgupta

On the Spirometry and Spirometers Standardization *316*
Lopata V., Sinekop Yu., ElShabbah M., Myasnyi I.

TheEnhancementoftheLinearX-RayTomographywithDigitalTomosynthesisAlgorithms *319*
SergiySenchurov, OlexanderMotolyga

Geometrical measuring at 3D stereoscopic radiography *322*
Miroshnychenko S.I., Nevhasymyy A.A., Volkov E.V.

The Possibilities of the Contrast Improvement in Fluorography Image with Digital Processing *325*
Aslamova L., Melenevska N., Miroshnichenko N.,Grabovska E., Miroshnichenko S., Radko D.A.

Efficiency and Pattern Noise Supression in Sensor Array X-Ray Digital Receptors *328*
Miroshnichenko S.I., Nevgasimy A.A.

SECTION III. POWER ELECTRONICS AND SMART GRID

Electronic System With Series Organization For Connection the Distributed Generators of Renewable Energy to Low Voltage Industrial Network *335*
Sokol E.I., Goncharov Y.P., Eresko A.V., Zamaruiev V.V., Krivosheev S.Y., Ivakhno V.V., Ilina O.V., Malyarenko E.A., Styslo B.A., Upyrenko K.G.

Method for Determining Energy Losses in Switched-Mode Converter Power Switches Depending on Storage Inductor Current Modes *339*
Oleksandr M. Afanasyev, Anatoliy A. Shcherba

Conversion of Shaded-Pole Induction Motor to Switched Reluctance Motor and Effects of Pole Shoe and Notch on SRM Noise *344*
A. Dadpour, K.Ansari, M. Divandari, M. R. Hosseinkhah

A Friendly Approach to Increasing the Frequency Response of Piezoelectric Generators *349*
Sam Ben-Yaakov, Gil Hadar, Amit Shainkopf and Natan Krihely

A Novel Fuzzy Dynamic Observer for High Speed BLDC Motor *354*
M. Divandari, F. Shahsavari, M. Rezaeimoghadam, A.Dadpour

Minimization of Cable Losses in Three-Phase Four-Wire Systems by means of Instantaneous Compensation with Shunt *359*
M.Yu. Artemenko, L.M. Batrak, S.Y. Polishchuk, V.M. Mykhalskyi, I.A. Shapoval

The use of the Dirichlet Kernel in the Control Systems of Active Filters for Industrial Power Line *363*
V.V. Zamaruiev

The Method of Determining the Duration of Transients in Semiconductor Devices for Current-Voltage Characteristics Measurement *367*
Ievgen O. Iermolenko, Oleksandr F. Bondarenko, Kristina I. Iermolenko

Multilevel Inverter as VAR-compensator *370*
Vladimir Bezhenar, Dmitry Mykolaets, Vitaly Mykytyuk, Tatiana Tereshchenko

Modeling of Electrical and Vibration Signals of Transformers with Different Magnetic Properties *373*
Valeriy Zhuikov, Nikolay Kuznietsov

Optimal Control of Uninterruptible Power System with Diesel Generator *378*
Ievgen Pichkalov, Anna Kyselova, Olga Kyselova, Gennadiy Kyselov

Design an Algorithm to Syntheses DC-DC Converters *381*
Murad Ahmed Ali Taher

Integration of Context-Aware Control System in Microgrid *386*
Valerii Zhuikov, Anna Kyselova

Modeling of Power-supply Subsystems of Microgrid Using Petri Nets *391*
Tetiana Khyzhniak, Viktor Kolesnyk

Simulation of Electromagnetic Environment of Class D Amplifier *396*
Y. Onikienko, V. Pilinsky, D. Probity, V. Shvaychenko

FPGA Based Capacitive Touch Pad/Interface *399*
Radovan Stojanović, Nedjeljko Lekić, Zoran Mijanović and Jovan Kovačević

SECTION IV. ELECTRONIC SYSTEMS

A 1.2V Wideband CMOS Mixer with High Conversion Gain and Low Flicker Noise *405*
Chao Chen, Jianhui Wu

A 1.2V Current Mirror Double Balanced Mixer with Current Reusing Technique *409*
Xiao Shi, Jianhui Wu, Zhilin Liu

Sustaining the Lower Bound of QoS in MIMO Ad-Hoc Network with Concurrent Virtual Streams *413*
Viktor V. Zaharov, Olexander B. Kokhanov, Angel Lambertt Lobaina

Peculiarities of Multichip Micro Module Frameless Design with Ball Contacts on the Flexible Board *417*
D.V. Vertyanov, K.S. Tikhonov, S.P. Timoshenkov, V.S. Petrov, G.A. Blinov

Investigation of Electron-Ion Optics of Pulse Technological Glow Discharge Electron Guns *420*
S.V. Denbnovetskiy, V.G. Melnyk, I.V. Melnyk, B.A. Tugay, S.B. Tugay

Waveguide Variable Attenuator Suitable for Electromechanical Control *425*
Victor Kazmirenko, Irina Golubeva, Yuriy Prokopenko

Simulation Parameters of the Microwave Synthesizer for Radar System *428*
Gramarchuk Y., Kobak N., Petrenko V., Slesarenko S.

Distributed Feedback Analysis *431*
Oleg Vityaz

Optimal Synthesis of Electronic System for Gyroscopic Nautical Compass Stabilisation *436*
O.A. Sushchenko

Real-time Implementation of the Optimal Two-Dimensional Filter for Highly Nonstationary Frequency-Modulated Signal Estimation *440*
Veselin N. Ivanović, Nevena Radović, Zdravko Uskoković, Srdjan Jovanovski

Electronic Hardware Design of a Low Cost Tactile Sensor Device for Physical Human-Robot Interactions *445*
Ganna Pugach, Viacheslav Khomenko, Artem Melnyk, Alexandre Pitti, Patrick Henaff, Philippe Gaussier

Wavelet Transform at Oriented Basis for Network Traffic Forecasting *450*
Tereshchenko T., Yamnenko Y., Veretiuk A., Veretiuk S.

Slot Allocation Model and Data Burst Scheduling in Downlink WiMAX Technology *455*
Lemeshko Olexandr, Garkusha Sergiy

Topological Analysis of Active Networks Containing Pathological Mirror Elements *460*
Vladimir Filaretov, Konstantin Gorshkov

Non-Destructive Volume and Thickness Measurements with Planar Microwave Sensors *465*
M. Al Jader, O. Korostynska, A. Mason and A. I. Al-Shamma'a

Fractal Spline Signals: Synthesis and Estimation *469*
Ihor Shelevytsky, Olga Novikova

Short-Range C-Band Noise Radar for Meteorological Application *473*
Lukasz Maslikowski, Krzysztof Kulpa, Dmitriy Glushko, Felix J. Yanovsky

SECTION I

NANO- AND MICROELECTRONICS

Microwave Characterization of Dielectric Substrates for Thin Films Deposition

Viktor Bovtun, Martin Kempa, Stanislav Kamba
Institute of Physics ASCR
Prague, Czech Republic
bovtun@fzu.cz

Valeriy Pashkov, Vitaliy Molchanov, Yuriy Poplavko, Yuriy Yakymenko
NTUU "Kyiv Polytechnic Institute"
Kyiv, Ukraine

Abstractð **A new electrode-free method is proposed for microwave characterization of low-loss single crystal and ceramic dielectric substrates used for thin films deposition. The substrate is considered and characterized as a thin dielectric resonator. Both $TE_{01ŭ}$ and $HE_{11ŭ}$ resonance modes were activated in the substrates with a dielectric permittivity above 10. The in-plane averaged dielectric permittivity and losses of a number of substrates were measured in a broad temperature range using the $TE_{01ŭ}$ mode. The $HE_{11ŭ}$ modes are proposed for the in-plane dielectric anisotropy characterization. In-plane components of anisotropic dielectric parameters of the (110) $DyScO_3$ substrate were measured.**

Keywordsð microwave characterization; dielectric resonator; dielectric substrate; anisotropy

I. Introduction: Substrates for Thin Films Deposition

Dielectric substrates with a relatively high permittivity ($\varepsilon \geq 10$) and low losses ($\tan\delta \leq 10^{-4}$) are widely used for a long time as the passive substrates of microwave (MW) hybrid and integrated circuits. They provide miniaturization of the planar waveguides between circuit elements and of the whole circuit as well. Beside the high permittivity and low losses, temperature stability of the MW dielectric parameters and mechanical stability are the main requirements applied to the MW substrates.

Thin films or multilayer heterostructures with special dielectric (or magnetic) properties deposited on dielectric substrates also belong to the basic elements of MW integrated circuits. For instance, ferroelectric films meet widespread applications as on-chip capacitors, voltage tunable capacitors for frequency agile resonant filters and MW phase shifters, etc [1,2]. Dielectric substrates for thin film deposition should provide conditions for growth of the structurally perfect, ordered films with desired properties. It requires crystal lattice parameters match of the film and substrate. Therefore single crystal substrates are mainly used to obtain the epitaxial thin films. On the other hand, the crystal lattice misfit can induce in a thin film the in-plane strains of a per cent level. Such strains are used to increase superconducting, ferromagnetic and ferroelectric transition temperatures, stabilize films in novel non-bulk phases (for example $SrTiO_3$ in the FE phase) or achieve simultaneously ferromagnetism and ferroelectricity in a single phase (see [3] and references there).

Consequently, a number of single crystals are proposed (and commercially available) for use as substrates for thin film deposition. Sometimes they do not satisfy the requirements applied to MW substrates and are characterized by higher losses, strong anisotropy and pronounced temperature dependence of their dielectric parameters. Microwave characterization of these substrates, as well as deposited films, is an actual task. Reliable knowledge of the substrate MW dielectric properties is necessary for the deposited film characterization. Taking into account that electrodes could essentially influence [4] or suppress the film properties, electrode-free MW characterization methods are preferable.

$TE_{01\delta}$ composite and split-post dielectric resonators (DRs) easy include a dielectric substrate or film on the substrate as a part of the resonance system and their parameters were shown to be sensitive to the substrate and film dielectric properties [5,6]. Recently we proposed a thin DR consisting only of the dielectric substrate (or substrate and deposited film) for the electrode-free microwave characterization of thin films and substrates [7,8]. Absence of the base DR component essentially increases the film volume fraction and enhances the method sensitivity comparatively to the composite [5] or split-post [6] DRs. Here we report new developments and applications of the thin DR method for the substrate characterization.

II. In-plane Averaged Dielectric Propertis of Substrates Measured by the $TE_{01\delta}$ Thin Dielectric Resonator

Thin (L=0.1-1 mm) rectangular (a=b=10 mm) dielectric substrates were studied as thin DRs. The $TE_{01\delta}$ resonance was activated in the shielding cavity using the transmission setup with a weak coupling [6,8,9] by Agilent E8364B network analyzer. Substrates were placed on the quartz support at the coupling loop level (Fig. 1a). The $TE_{01\delta}$ mode was certainly identified [8] by simulation of the electromagnetic field distribution (Fig. 1b, 1c) in our resonance system and by analysis of the continuous shift of the resonance frequency with the decrease of DR thickness (Fig. 2). The in-plane averaged dielectric permittivity and losses of our rectangular substrates were calculated using electrodynamic analysis [6,9] of the cylindrical DR with an effective diameter D = 10.82 mm. D was defined by the approximate formula [8] derived by comparison of cylindrical and rectangular $TE_{01\delta}$ DRs having equal thicknesses, resonance frequencies and transversal wave

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

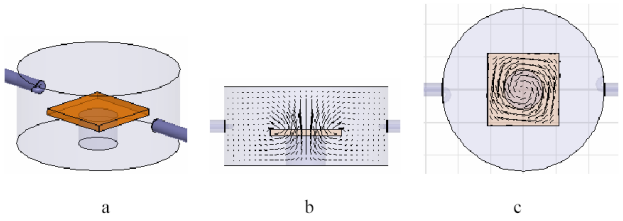


Fig. 1. Thin DR in the cylindrical shielding cavity: scheme of the measuring cell (a), magnetic (b) and electric (c) field distribution for the $TE_{01\delta}$ resonances.

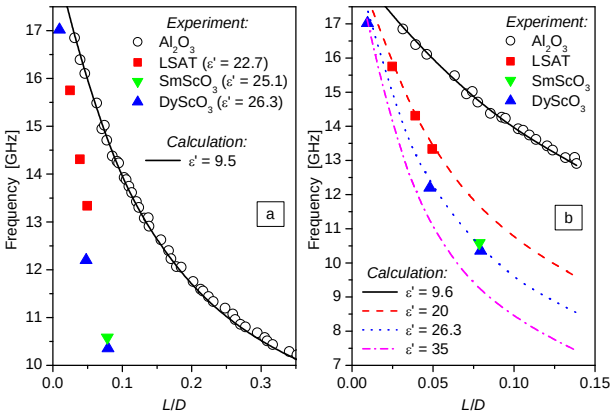


Fig. 2. $TE_{01\delta}$ resonance frequencies of substrates in the shielding cavity in dependence on the *L*/*D* ratio.

numbers:

$$D = \frac{1.53 \cdot a \cdot b}{\sqrt{a^2 + b^2}}, \qquad (1)$$

where *a* and *b* are side dimensions of the rectangular DR.

In-plane averaged dielectric parameters of a number of substrates were measured in the temperature range 10-400 K. (Fig. 3). Their values (Table I) and temperature evolution coincide well with those obtained by other MW methods [10].

The structural phase transition in LSAT [11] is clear evidenced by the small permittivity anomaly and dielectric loss maximum near 150 K. $DyScO_3$ substrates are characterized by relatively high dielectric losses (tanδ exceeds 10^{-3}) which could limit tentative MW applications of the ferroelectric or multiferroic films deposited on them [3]. Moreover, two pronounced tanδ maxima were revealed below 200 K. These anomalies could be related to the complicated phonon dynamics in $DyScO_3$ [3].

So, very thin dielectric substrates, down to 0.1 mm with ε > 10, could be successfully measured using $TE_{01\delta}$ thin DR. It is very important for the measurements of thin films deposited on the substrate: increase of the film/substrate thickness ratio results in the increase of the method sensitivity [8].

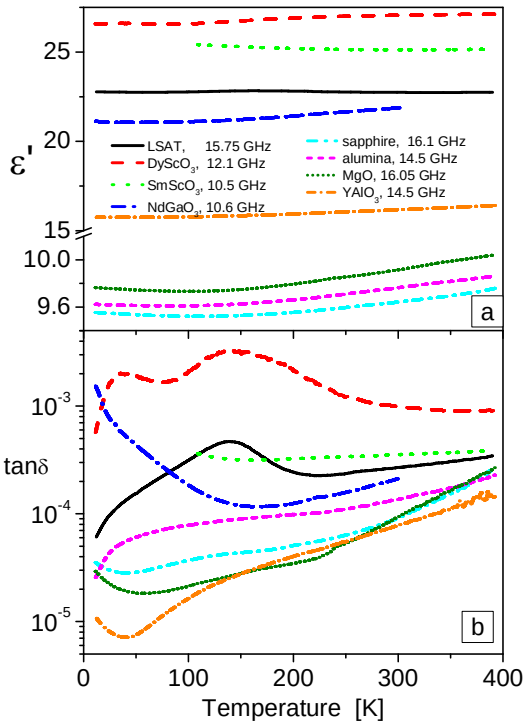


Fig. 3. In-plane averaged dielectric permittivity (a) and losses (b) of rectangular low-loss dielectric substrates 10*10*L mm^3 with a thickness L = 0.2-0.8 mm.

TABLE I. IN-PLANE AVARAGED MICROWAVE DIELECTRIC PARAMETERS OF SUBSTRATES AT ROOM TEMPERATURE

Substrate	**Frequency,** (GHz)	**ε̊**	**tanδ** (10^{-4})
(0001) sapphire[a]	15÷17	9.7÷9.9	0.5÷1
alumina[a]	14÷16	9.6÷9.8	~1
(100) MgO	16	9.7÷9.8	0.2÷1
(100) LSAT[b]	14÷16	22.7÷22.8	~2
(110) $DyScO_3$[a]	10÷17	26÷28	5÷15
(110) $TbScO_3$	12÷15	26÷27	5÷8
(110) $SmScO_3$	10.5	25.2	3
(110) $NdGaO_3$	10.5	21.8	1.5
(100) $YAlO_3$	14.3	16.2	0.8

[a] Multiple substrates with different thickness
[b] $(LaAlO_3)_{0.29}(SrAl_{1/2}Ta_{1/2}O_3)_{0.71}$

It is even more important in the case of single crystal substrates which are mainly used to obtain ferroelectric films with a biaxial tensile or compressive mechanical strain [12]. The strain is relaxed in the case of "thick" films (above 100 nm thickness). Therefore the very thin (20-100 nm) incipient ferroelectric films ($SrTiO_3$, $EuTiO_3$, etc.) are of interest [13]. Consequently, the thin substrate is preferable to ensure the method sensitivity.

III. In-plane Dielectric Anisotropy Measured by $HE_{11\delta}$ Thin Dielectric Resonator

While the $TE_{01\delta}$ resonance provides the in-plane averaged parameters, the in-plane dielectric anisotropy of the substrate can be characterized using the $HE_{11\delta}$ resonances of thin DRs. In the cylindrical shielding cavity, two $HE_{11\delta}$ resonances are activated in thin substrates together with the $TE_{01\delta}$ one (Fig. 4). In the case of thin enough substrates (L/D < 0.05), the E-field distribution at $HE_{11\delta}$ resonance frequencies corresponds to the permittivity components along $DyScO_3$ substrate edges ($\varepsilon_X = \varepsilon_c = 34.4$, $\varepsilon_Y = \varepsilon_{ab} = 20.3$, $\varepsilon_Z = \varepsilon_{ab} = 20.3$); at the $TE_{01\delta}$ resonance – to the in-plane averaged $\varepsilon_{XY} = \varepsilon_{abc} = 26.5$, $\varepsilon_Z = \varepsilon_{ab} = 20.3$. In the rectangular waveguide, only one $HE_{11\delta}$ mode is excited [14] and its resonance frequency corresponds to the in-plane permittivity component perpendicular to the waveguide axis.

Simulated dependence of the resonance frequencies on permittivity (Fig. 5) allows calculating the substrate in-plane permittivity components. Estimated microwave values of ε_X and ε_Y and their temperature dependences (Fig. 6) are well consistent with the THz parameters and with the permittivity components along crystallographic axes $\varepsilon_c = 34.4$, $\varepsilon_a = 22$, $\varepsilon_b = 19$ measured at low frequencies [15]. Essential anisotropy of microwave dielectric losses revealed at low temperatures is similar to that in THz range.

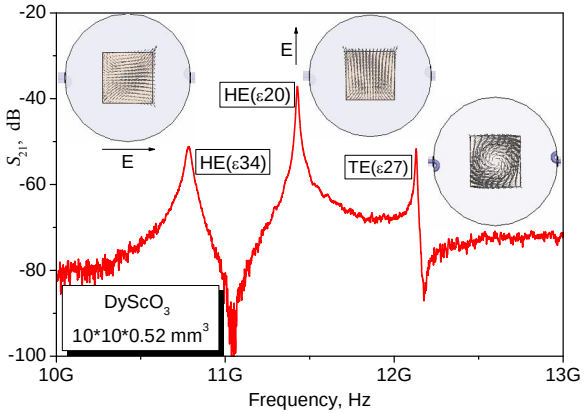


Fig. 4. Transmission coefficient spectrum indicating $HE_{11\delta}$ and $TE_{01\delta}$ resonances of the (110) $DyScO_3$ substrate oriented with $\varepsilon_X = 34.4$ along coupling loops in the cylindrical shielding cavity.

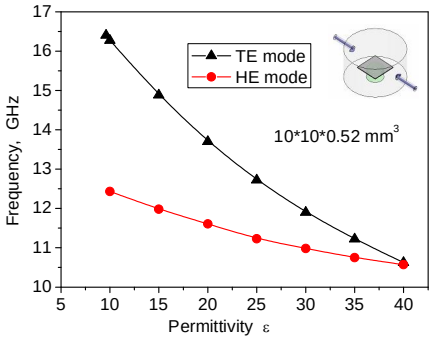


Fig. 5. Simulated dependences of the $HE_{11\delta}$ and $TE_{01\delta}$ resonance frequencies of a rectangular substrate on the substrate permittivity in the cylindrical shielding cavity.

High dielectric anisotropy of some single crystal substrates inserts influence on their $TE_{01\delta}$ resonance, which intensity and quality factor are dependent on the substrate orientation in relation to the coupling loops (Fig. 7). This effect is dependent on the substrate thickness and could be related to the anisotropic field distribution (Fig. 8), even in the cylindrical DR, which influence excitation and coupling conditions of the $TE_{01\delta}$ mode, as well as its interaction with other modes.

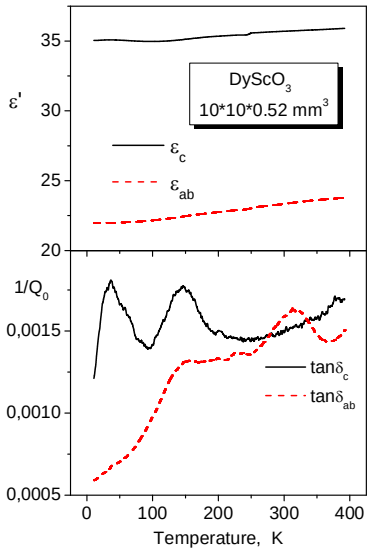


Fig. 6. In-plane components of dielectric permittivity ε and losses tanδ of the (110) $DyScO_3$ substrate at 11 GHz.

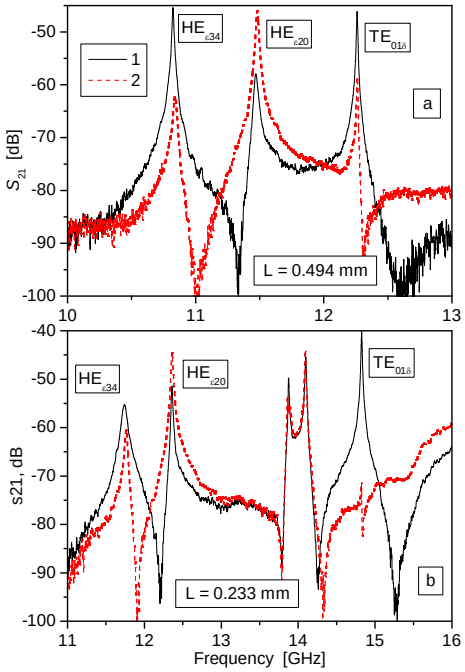


Fig. 7. Transmission coefficient spectra indicating $HE_{11\delta}$ and $TE_{01\delta}$ resonances of the (110) $DyScO_3$ substrates with different thickness in the cylindrical shielding cavity.Curves 1 and 2 correspond respectively to the substrate orientation with $\varepsilon_Y = 20.3$ or $\varepsilon_X = 34.4$ along the coupling loops.

Also a scattering range of the $TE_{01\delta}$ resonance frequencies of nominally identical substrates is dependent on the substrate orientation (Fig. 9). We should note the too low quality factors of the $HE_{11\delta}$ resonances in the cylindrical shielding cavity (Fig. 4, 6), which are lower than that measured in the rectangular waveguide [14]. It is also can be caused by peculiarities of the modes excitation and interaction in the anisotropic DR. Development of electrodynamical model of the anisotropic thin DR accounting for the modes interaction and coupling with the measuring cell is highly required for the correct estimation of MW parameters of substrates, as well as of the thin films which can be deposited on them.

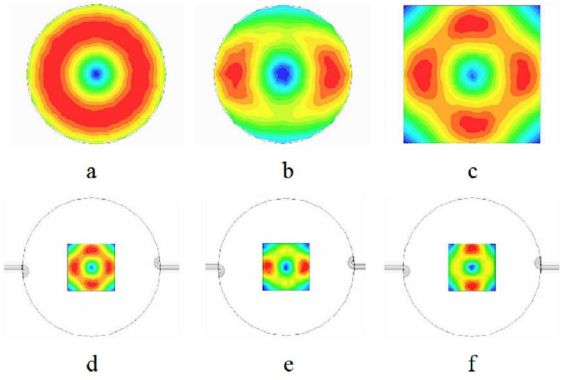


Fig. 8. Simulated E-field intensity distribution in the $TE_{01\delta}$ thin dielectric resonators with permittivity values similar to that of $DyScO_3$ substrates: a) cylindrical isotropic, $\varepsilon_{XY} = 26.5$; b) cylindrical anisotropic, $\varepsilon_X = 34.4$ along coupling loops; c,d) rectangular isotropic, $\varepsilon_{XY} = 26.5$; e) rectangular anisotropic, $\varepsilon_X = 34.4$ along coupling loops; f) rectangular anisotropic, $\varepsilon_Y = 20.3$ along coupling loops.

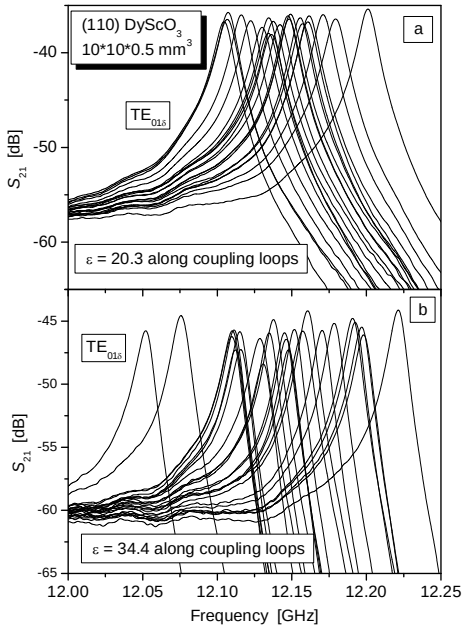


Fig. 9. Scattering of the $TE_{01\delta}$ resonance frequences of nominally identical (110) $DyScO_3$ substrates oriented in the cylindrical shielding cavity with $\varepsilon_Y = 20.3$ or $\varepsilon_X = 34.4$ along the coupling loops.

IV. Conclusions

Anisotropic dielectric substrates can be successfully characterized by the thin dielectric resonator method. The $TE_{01\delta}$ resonance mode allows measurements of their in-plane averaged dielectric parameters, while the $HE_{11\delta}$ modes provide characterization of the in-plane dielectric anisotropy.

Acknowledgment

The financial support by the Czech Science Foundation (Project P204/12/1163) is gratefully acknowledged.

References

[1] O.G. Vendik, E.K. Hollmann, A.B. Kozyrev, and A.M. Prudan, "Ferroelectric tuning of planar and bulk microwave devices", Journal of Superconductivity, vol. 12, pp. 325-338, 1999.

[2] A.K. Tagantsev, V.O. Sherman, K.F. Astafiev, J. Venkatesh, and N. Setter, "Ferroelectric materials for microwave tunable applications", Journal of Electroceramics, vol. 11, pp. 5-66, 2003.

[3] J.H. Lee, L. Fang, E. Vlahos, X. Ke, Y.W. Jung, et al., "A strong ferroelectric ferromagnet created by means of spin–lattice coupling", Nature, v. 466, pp. 954-958, 2010.

[4] A. K. Tagantsev, and G. Gerra, "Interface-induced phenomena in polarization response of ferroelectric thin films", J. Appl. Phys., vol. 100, 051607, pp.1-28, 2006.

[5] V. Bovtun, S. Veljko, A. Axelsson, S. Kamba, N. Alford, and J. Petzelt, "Microwave characterization of thin ferroelectric films without electrodes by composite dielectric resonator", Integrated Ferroelectrics, vol. 98, pp. 53-61, 2008.

[6] J. Krupka, "Frequency domain complex permittivity measurements at microwave frequencies", Meas. Sci. Technol., vol. 17, pp. R55-R70, 2006.

[7] V. Pashkov, V. Bovtun, Y. Prokopenko, M. Kempa, V. Molchanov, et al., "Measurement of dielectric films microwave parameters", Proc. 19th Int. Crimean Conf. "Microwave & Telecommunication Technology" (CriMiCo'2009), Sept. 2009, Sevastopol, Ukraine, pp. 769-770, 2009.

[8] V. Bovtun, V. Pashkov, M. Kempa, S. Kamba, et al., "An electrode-free method of characterizing the microwave dielectric properties of high-permittivity thin films", J. Appl. Phys., vol. 109, 024106, pp. 1-6, 2011.

[9] J. Krupka, K. Derzakowski, B. Riddle, and J. Baker-Jarvis, "A dielectric resonator for measurements of complex permittivity of low loss dielectric materials as a function of temperature", Meas. Sci. Technol., vol. 9, pp. 1751-1756, 1998.

[10] J. Krupka, R.O. Geyer, M. Kuhn, and H. Hinken, "Dielectric properties of single crystals of Al_2O_3, $LaAlO_3$, $NdGaO_3$, $SrTiO_3$, and MgO at cryogenic temperatures", IEEE Trans. MTT, vol. 42, pp. 1886-1890, 1994.

[11] B.C. Chakoumakos, D.G. Schlom, M. Urbanik, and J. Luine, "Thermal expansion of $LaAlO_3$ and $(La,Sr)(Al,Ta)O_3$, substrate materials for superconducting thin-film device applications", J.Appl. Phys., vol. 83, pp. 1979-1982, 1998.

[12] D.G. Schlom, L.-Q. Chen, Ch.-B. Eom, K.M. Rabe, S.K. Streifer, and J.M. Triscone, „Strain tuning of ferroelectric thin films", Annu. Rev. Mater. Res., vol. 37, pp. 589-626, 2007.

[13] D. Nuzhnyy, J. Petzelt, S. Kamba, P. Kužel, C. Kadlec, V. Bovtun, M. Kempa, J. Schubert, et al., „Soft mode behavior in $SrTiO_3$/$DyScO_3$ thin films: Evidence of ferroelectric and antiferrodistortive phase transitions", Appl. Phys. Lett., vol. 95, 232902, pp. 1-3, 2009.

[14] V. Bovtun, V. Pashkov, M. Kempa, et al., " Thin dielectric resonatots for microwave characterization of films and substrates", Proc. 21st Int. Crimean Conf. "Microwave & Telecommunication Technology" (CriMiCo'2011), Sept. 2011, Sevastopol, Ukraine, pp. 620-621, 2011.

[15] S. Coh, T. Heeg, J.H. Haeni, M.D. Biegalski, J. Lettieri, L. F. Edge, et al., "Si-compatible candidates for high-k dielectrics with the Pbnm perovskite structure", Phys. Rev. B, vol. 82, 064101, pp. 1-16, 2010.

Broadband Dielectric and Conductivity Spectra of Dielectric – metal Nanocomposites for Microwave Applications

A. Borisova, A. Machulyansky, Y. Yakimenko
NTUU "Kiev Polytechnic Institute"
Kiev, Ukraine
al-borisova@mail.ru

V. Bovtun, M. Kempa, M. Savinov
Institute of Physics ASCR
Prague, Czech Republic
bovtun@fzu.cz

***Abstract*—Three kinds of dielectric-metal nanocomposites were prepared and studied: a) epoxy resin-xFe (x = 0–30 vol.% is a content of the Fe nanoparticles); b) epoxy resin-xNi (x = 0–40 vol.% is a content of the Ni nanoparticles); c) paraffin-xFe (x = 10–60 vol.% is a content of the Fe nanoparticles). Their dielectric and ac conductivity spectra were measured in a broad frequency range from 10^{-2} Hz to 10^{10} Hz. Dielectric/conductivity dispersion was observed in the whole frequency range. Increase of the metal content results in the increase of the composite dielectric permittivity, loss and conductivity in the whole frequency range. The power-law dispersion of conductivity and permittivity characterizes response in the microwave range for all studied composites. For the compositions with high metal content, above the percolation threshold, the conductivity plateau screens the power-law dispersion at low frequencies. Three contributing polarization/conductivity mechanisms are considered, possibilities of the composites microwave applications are discussed.**

Keywords—dielectric-metal nanocomposites, dielectric and conductivity spectra, percolation, microwave properties

I. INTRODUCTION

Development of modern electronic and telecommunication systems related with an increase of their working frequency and sensitivity makes important development of the materials with a high absorption of electromagnetic radiation in a wide range of wavelengths. Such materials and coatings are widely used in various technical applications for improvement of the noise stability and reliability of electronic devices, radar countermeasure (radar-tracking masking), ecological and information safety, metrology, for design of the physical parameter converters, etc. [1, 2].

Analysis of the efficiency of electromagnetic field absorption by currently used materials allows to define their basic features and drawbacks and to conclude that creation of radio/microwave absorbing materials on the base of heterogeneous composite nanostructures is a perspective way of improving the shielding efficiency [2,3]. Dielectric-metal composites utilizing nanoparticles of the magnetic metals are of great importance for development of the broadband absorbing coatings [3]. However, formation of such metal-dielectric composites with a required set of their spectral electromagnetic parameters is a complicated problem, because of the strong influence of both composition and mesostructure of on their electromagnetic properties [4]. Consequently, theoretical and experimental studies of the composition-structure-properties dependences are incomplete and their results are sometimes discrepant. The electromagnetic properties are usually reported only for the narrow frequency ranges [3,4,5]. Therefore complex experimental study of the electromagnetic parameters of metal-dielectric composites in a broad frequency range is an actual task.

We extend our previous investigations [3-6] and report here on the dielectric and ac conductivity spectra of the metal-dielectric composites with various concentrations of nanodispersed particles of magnetic metals (Ni, Fe) in a broad frequency range from 10^{-2} to10^{10} Hz.

II. PROCESSING AND EXPERIMENTAL TECHNIQUES

Three series of dielectric-metal nanocomposites were prepared by electromechanical mixing of nanodispersed metal powders with a binding dielectric matrix at the room temperature [5]: a) epoxy resin-xFe (x = 0–30 vol.% is a content of the Fe nanoparticles); b) epoxy resin-xNi (x = 0–40 vol.% is a content of the Ni nanoparticles); c) paraffin-xFe (x = 10–60 vol.% is a content of the Fe nanoparticles). Structure of the composites was investigated by a raster electronic microscopy. The results evidence the nanodispersed metal phase built by the particles with a size from 50 nm to 100 nm.

Two kinds of samples were prepared from the fabricated compositions for dielectric measurements. Plate-shaped samples (with a diameter of 10-20 mm and thickness of 3-5 mm) were used in the low-frequency (LF, 10^{-2} – 10^{6} Hz) and microwave (MW, 10^{8} – 10^{10} Hz) measurements. For the high-frequency (HF, 10^{6} – 10^{9} Hz) measurements, cylindrical samples with a diameter of 2-3 mm and length of 5-7 mm were prepared. In LF and HF experiments, the pasted silver or evaporated gold electrodes were deposited on the flat surfaces of samples, while the MW experiment was performed on samples without electrodes.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

Dielectric and ac conductivity spectra were measured using impedance methods by the techniques described in [7,8]. Standard LF dielectric measurements were performed with the dielectric analyzer Novocontrol Alpha AN, HF measurements - with the Novocontrol BDS 2100 coaxial sample cell and Agilent 4291B impedance analyzer, MW measurements – with the open-end coaxial probe and Agilent E8364B vector network analyzer.

III. Results and Discussion

Experimental broadband spectra of the dielectric permittivity (ε′), dielectric loss (ε′′) and ac conductivity (σ′) of the epoxy resin-xFe, epoxy resin-xNi and paraffin-xFe composites with different concentrations of nanosized metallic particles are presented in Figures 1-3.

For all studied compositions, dielectric/conductivity dispersion was observed in the whole frequency range. The power-law dispersion of conductivity and permittivity characterizes response in the high-frequency and microwave ranges for all studied composites. For the compositions with high metal content, above the percolation threshold, the conductivity plateau screens the power-law dispersion at low frequencies. The spectra are similar to that of the composites with carbon nanoparticles [9] and nanotubes [10].

According to the different character of the ε′′ and σ′ frequency dependences at LF, HF and MW, we consider three mechanisms contributing to the dielectric / ac conductivity spectra. The low-frequency conductivity plateau can be attributed to the dc metal network conductivity (in percolated clusters), while the high-frequency power-law dispersion – to the ac conductivity in finite metal clusters (i.e., to the dynamics of the charge carriers in the fractal conducting network) and inter-cluster polarization of insulating matrix.

Increase of the metal content generally results in the increase of the composite dielectric permittivity, loss and conductivity at all frequencies. Nevertheless, some deviations from the monotonic concentration dependency are observed for LF conductivity and loss. Concentration dependence of the dc conductivity (estimated from the LF conductivity plateau) is monotonic for the paraffin-xFe compositions above percolation threshold (x = 20-60), the x = 10 composition is below the threshold. For the epoxy resin-xNi, compositions with x = 10, 30, 40 are already above percolation threshold, while the composition with x = 20 – below the threshold. Similar, the epoxy resin-xFe compositions x = 10, 30 are above percolation threshold, but x = 20 – below it.

These deviations from monotonic dependency are caused by inhomogeneous distribution of the metallic nanoparticles in the dielectric matrix. Therefore compositions differ one from others not only with their concentration, but also with the morphology and mesostructure. It makes possible creation of the percolated metal cluster in a composition with lower concentration rather than in that with higher concentration. Let us note that the inhomogeneous metal distribution influences mainly the dc conductivity and LF dielectric properties of composites. The HF and MW properties are mainly defined by the concentration and are less sensitive to the composite morphology, because both percolated and finite conductive clusters contribute to the ac conductivity.

MW ac conductivity (10^8 – 10^{10} Hz) of the studied metal-dielectric composites varies in a broad rage (from 10^{-7} S/cm to 1 S/cm) and can be controlled by changing of the metallic nanoparticles concentration. It is very important and promising for their tentative application as microwave absorbing or shielding materials. Additional useful contribution to the high-frequency absorbing or shielding efficiency of the studied composites could arise from magnetic properties of the Ni and Fe nanoparticles. Therefore HF study of the composite magnetic properties is under consideration.

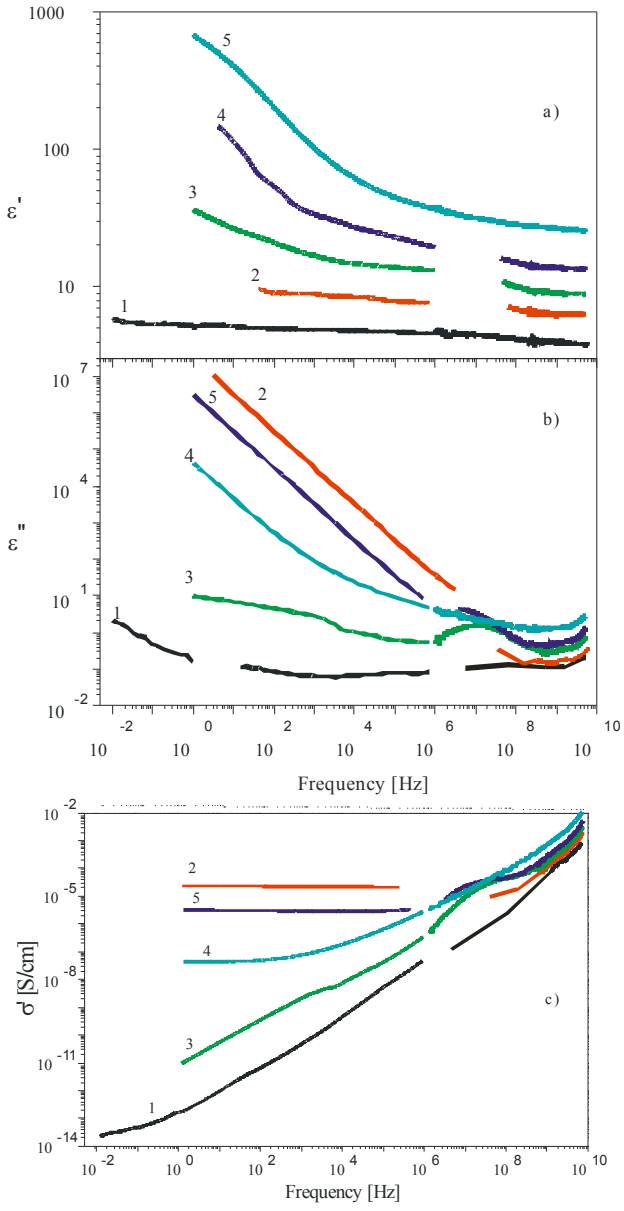


Fig. 1. Frequency dependence of dielectric permittivity (a), loss (b) and conductivity (c) of the epoxy - xNi composites with different concentrations of Ni (in vol.%) 1 – x = 0; 2 - 10, 3 - x =20, 4 - x =30, 5 – x =40.

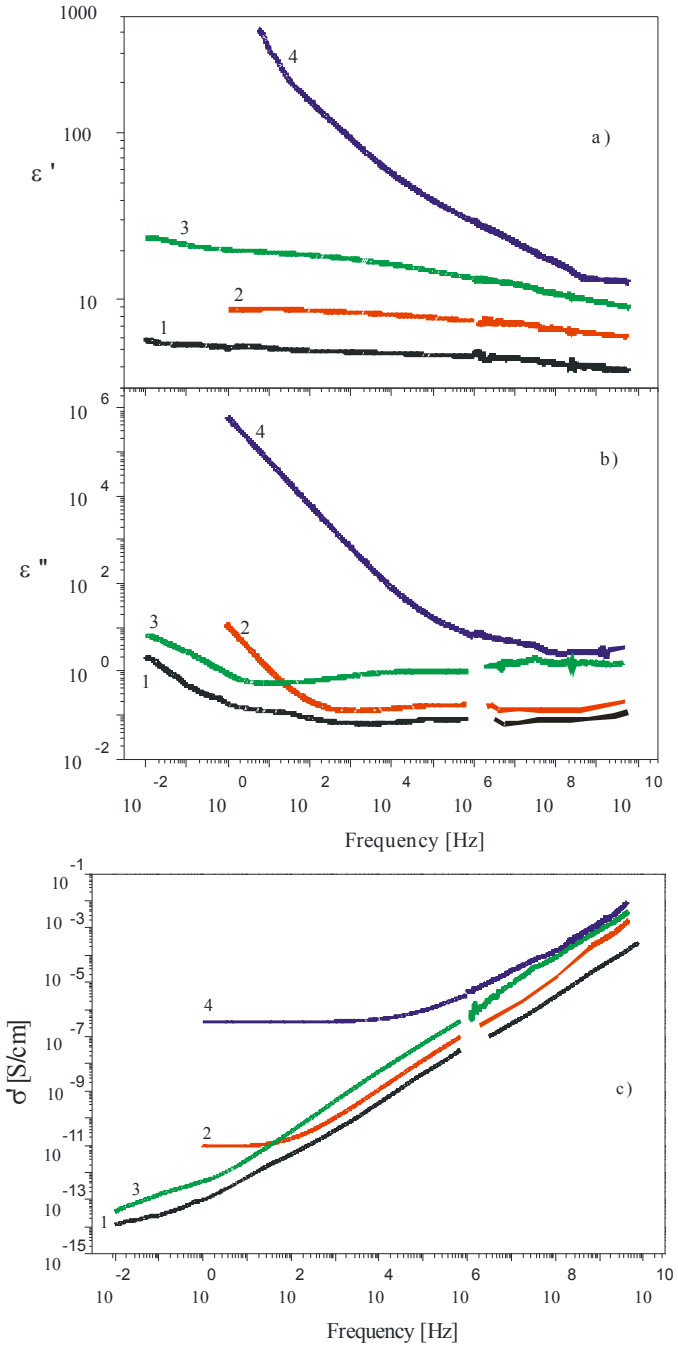


Fig 2. .Frequency dependence of dielectric permittivity (a), loss (b) and conductivity (c) of the epoxy-xFe composite with different concentrations of Fe (in vol.%) 1 - x = 0; 2 - x =10, 3 - x =20, 4 – x =30.

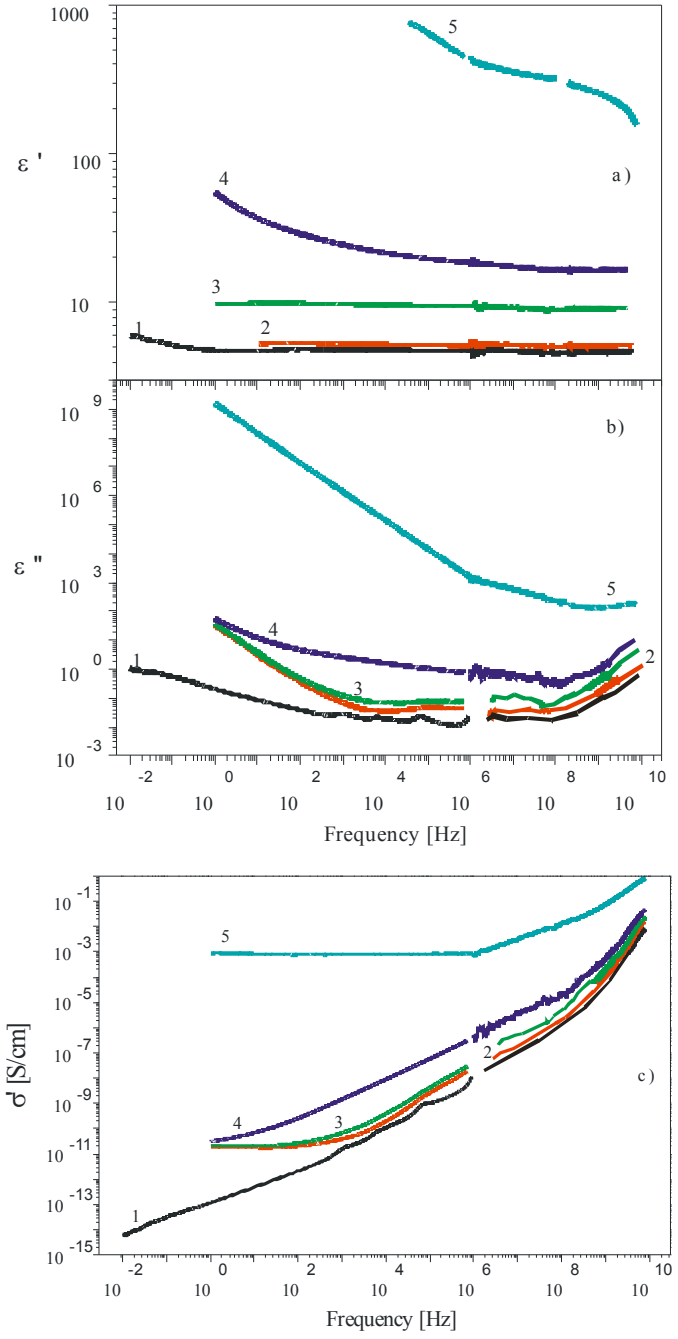


Fig 3. Frequency dependence of dielectric permittivity (a), loss (b) and conductivity (c) of the paraffin-xFe composite with different concentrations of Fe (in vol.%) 1 - x =10; 2 - x =20, 3 - x =30, 4 – x =40, 5 – x =60

IV. Conclusions

Dielectric and ac conductivity spectra of dielectric - metal nanocomposites were measured in a broad frequency range from 10^{-2} Hz to 10^{10} Hz. The power-law dispersion of conductivity and permittivity was shown to define their response in the high-frequency and microwave ranges. For the compositions with high metal content, above the percolation threshold, the conductivity plateau screens the power-law dispersion at low frequencies. The low-frequency conductivity plateau is attributed to the dc metal network conductivity (in percolated clusters), while the high-frequency power-law dispersion – to the ac conductivity in finite metal clusters. The HF and MW properties are mainly defined by the concentration and are less sensitive to the composite morphology. MW ac conductivity (10^{8} – 10^{10} Hz) of the studied metal-dielectric

composites varies in a broad rage (from 10^{-7} S/cm to 1 S/cm) and can be controlled by changing of the metallic nanoparticles concentration. Studied composites can be used as microwave absorbing or shielding materials.

ACKNOWLEDGMENTS

The work was supported by the Czech Science Foundation (project P204/12/0232).

REFERENCES

[1] L.N. Kechiev, P.V. Stepanov - *EMC and Information Security in Telecommunications*, M.: Technologies, 2005 – 320 p.

[2] H. W. Ott - *Electromagnetic Compatibility Engineering*, New York, I. Wiley & Sons, 2009 – 862 p.

[3] A.V. Machulyansky, V.V. Pylynsky, O.V. Telychkina, V.B Shvaychenko, M.V. Rodionova "Bases of application of nanostructural materials for providing electromagnetic compatibility of electronic devices" - *Electronics and Communications* №3, 83-85 (2010) – (in Russian).

[4] A.V. Machulyanskiy "High-frequency conductivity of nano-sized metal particles' - *Electronics & Telecommunication* Part 1, 41–45 (2007) – (in Russian.)

[5] A. Machulyansky, D. Tatarchuk, O. Telychkina, Y. Yakimenko "Electrophysical properties of nanocomposite materials on the base of metal-dielectric structures" - *Proceeedings of 33rd International Spring Seminar on Electronics Technology (ISSE 2010 Warsaw), Warsaw, Poland, May 12th - 16th*, 249-250 (2010).

[6] B.M. Bondar, A.V. Borisova, A.V. Machulyansky, Yu.I. Yakimenko "Characteristics of shields composite coating based on nanodisperse particles" - *Proceeedings of 22th Int. Crimean Conference "Microwave & Telecommunication Technology" (Sevastopol, September 12-14th, 2012)*, 683-684 (2012).

[7] D. Nuzhnyy., P. Vaněk, J. Petzelt, V. Bovtun, M. Kempa, I. Gregora, M. Savinov, R. Krupková, V. Studnička, J. Buršík, M. I. Samoylovich, W. Schranz "Properties of $BaTiO_3$ confined in nanoporous Vycor and artificial opal silica" - *Processing and Application of Ceramics* **4**, No.3, 215–223 (2010).

[8] D. Nuzhnyy, J. Petzelt, V. Bovtun, M. Kempa, M. Savinov, C. Elissalde, U.-C. Chung, D. Michau, C. Estournès, and M. Maglione "High-Frequency Dielectric Spectroscopy of $BaTiO_3$ Core - Silica Shell Nanocomposites: Problem of Interdiffusion" - *Journal of Advanced Dielectrics* **1**, No. 3, 309-317 (2011).

[9] M. Knite, A. Hill, V. Bovtun V. Teteris, A. Solovjovs, G Shakale, J. Zavickis, I. Aulika, B. Polakovs, S.J. Pas, S. Veljko, D. Noujni, I. Klemenoks, j. Zikans, A. Kiploka, D. Erts, J. Petzelt, A. Fuith "Polymer-nanostructured carbon composites as multifunctional sensor materials: design, processing and properties" - *Latvian J. of Physics and Technical Sciences*, N2, 15-29 (2006).

[10] D. Nuzhnyy, M. Savinov, V. Bovtun, M. Kempa, J. Petzelt, B. Mayoral, T. McNally „Broad-band conductivity and dielectric spectroscopy of composites of multiwalled carbon nanotubes and poly(ethylene terephthalate) around their low percolation threshold" - *Nanotechnology* 24, 055707, 1-9 (2013).

Synthesis of ZnO Nanorods for Acoustic Wave Sensor

Anatolii Orlov, Veronika Ulianova, Yuriy Yakimenko
Department of Microelectronics
National Technical University of Ukraine "KPI"
Kyiv, Ukraine
e-mail: atorlov@rambler.ru

Oleksandr Bogdan, Gennady Pashkevich
Research Institute "Applied Electronics"
National Technical University of Ukraine "KPI"
Kyiv, Ukraine
e-mail: bogdan@ee.ntu-kpi.kiev.ua

***Abstract*— We reported the electrochemical growth of zinc oxide nanorods in a zinc nitrate/hexamethylenetetramine solution at 60°C. High-density vertical nanorods were grown on Au-coated lithium niobate substrates and aluminum foil. Scanning electron microscopy results showed that hexagon-shaped ZnO nanorods with a diameter in the range of 150 nm to 300 nm and a length of about 3 μm were formed. Application of synthesized ZnO nanorods as sensing element of SAW devices will promote quality and accuracy enhancement, manufacturing simplification, and novel acoustic wave sensors simulation technique improvements due to the mathematical representation simplicity of the regular-shaped structures.**

Keywords— electrochemical growth; ZnO nanorods; SAW sensor

I. INTRODUCTION

Nowadays zinc oxide became a promising component in a wide range of nanoscale devices for different applications. It is an attractive material for electronics, photonics and sensing due to having exotic and versatile properties such as mechanical, piezoelectric, optical and electrical properties, biocompatibility, nontoxicity, chemical and photochemical stability, high specific surface area, optical transparency, electrochemical activities and so on [1]. On other hand, ZnO is attracting considerable attention due to its unique ability to form a variety of nanostructures such as nanowires, nanoribbons/nanobelts, nanocombs, nanorings, nanocages, nanocastle, nanofibers etc. ZnO nanorods/nanowires have been employed as bio- or gas sensitive element of acoustic wave sensors [2]. ZnO nanorods provide huge effective surface area and strong bonding sites and this way allow more precision control of their properties and characteristics.

Various fabrication techniques have been established for the growth of ordered ZnO nanorods, such as vapor liquid solid growth [3], vapor-phase synthesis [4], laser ablation [5] and metal organic vapor phase epitaxy [6]. However, those growth methods require relatively high synthesis temperatures. In recent years, low-temperature approaches as sol–gel techniques [7] and electrochemical growth [8] become more accepted. The electrochemical growth of ZnO has advanced to generate various kinds of nanostructures by using zinc nitrate [9] or zinc chloride [10] solutions, adding different precursors. This method doesn't require a preparation of a seed-textured ZnO thin layer before nanorods growing, it is more controllable than sol-gel method, and it's just as low-cost, low temperature, scalability and ease handling.

In this paper, we reported the new original results of the electrochemical growth of well-aligned ZnO nanorod arrays on the Au-coated lithium niobate substrate and Au-coated aluminum foil using a zinc nitrate solution with hexamethylenetetramine (HMT) as an additional precursor. Such ZnO nanorods arrays could be suitable as sensing element for acoustic wave based sensors and resonators.

II. EXPERIMENTAL DETAILS

The electrochemical growth method at low-temperature in aqueous solution was used to obtain zinc oxide nanorods on two different kinds of substrates. First, the lithium niobate substrate with Au layer is used. After a wafer cleaning (Fig.1, a) with a mixture of H_2O_2/H_2SO_4 during 30 min, 30-nm Au layer was deposited on the $LiNbO_3$ substrate by magnetron sputtering (Fig.1, b). For the second type of substrate, aluminum foil was Au-coated by the same technique.

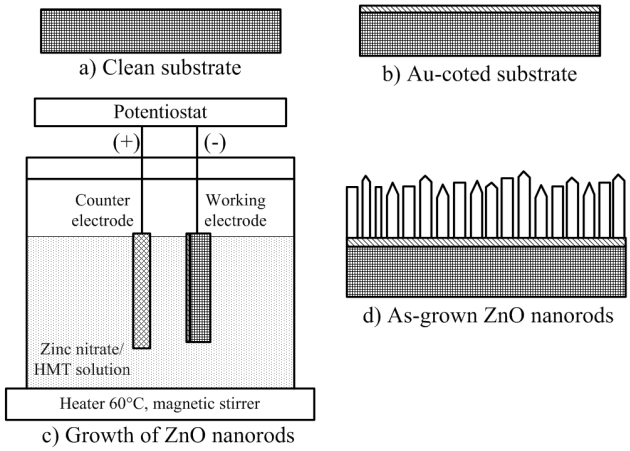


Fig. 1. Schematic diagram of growth of ZnO nanorods.

Generally, the ZnO growth was carried out in a temperature-controlled chemical batch reactor at 60°C under stirring with a magnetic agitator for 90 min (Fig. 1, c). The solution consists of analytically pure zinc nitrate hydrate

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

$Zn(NO_3)_2 \cdot 6H_2O$ and analytically pure hexamethylenetetramine $C_6H_{12}N_4$ which were mixed with an equal molar concentration. The chemicals were solved in 600 ml deionized water, resulting in a transparent solution. The concentration of both zinc nitrate and HMT was 0.025 mol/l. At a temperature of about 60°C the solution starts to become cloudy, indicating that the chemical reactions have started. During the growth process the solution is stirred in order to obtain a homogeneous concentration and temperature throughout the reactor volume. After the reaction the samples are thoroughly purged with deionized water and dried at 60°C.

In both experiments ZnO nanorods were electrodeposited on the Au-coated substrates at applying 2V between neighboring conducting lines for a deposition time of 90 min, a stainless steel sheet acts as a counter electrode. The spacing between both electrodes was about 2 cm.

The formation of ZnO takes place on the cathode side. Using a zinc nitrate based solution, the possible chemical reactions can be described as follows [9]:

$$Zn(NO_3)_2 \cdot \rightarrow Zn^{2+} + 2NO_3^-, \quad (1)$$

$$NO_3^- + H_2O + 2e^- \rightarrow NO_2^- + 2OH^-, \quad (2)$$

$$Zn^{2+} + OH^- \rightarrow Zn(OH)_2, \quad (3)$$

$$Zn(OH)_2 \rightarrow ZnO + H_2O. \quad (4)$$

Additional OH^- ions are obtained by the chemical reaction of HMT with water:

$$C_6H_{12}N_4 + 6H_2O \leftrightarrow COH_2 + 4NH_3, \quad (5)$$

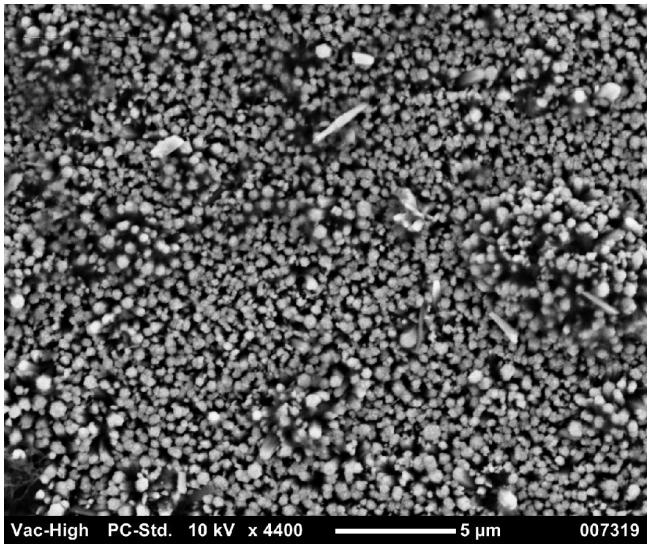
$$NH_3 + H_2O \leftrightarrow NH_4^- + OH^-. \quad (6)$$

Scanning electron microscopy (JEOL Ltd., NeoScope JCM-5000) was used to examine the morphology, especially the diameters and length of the ZnO nanorods across the substrates.

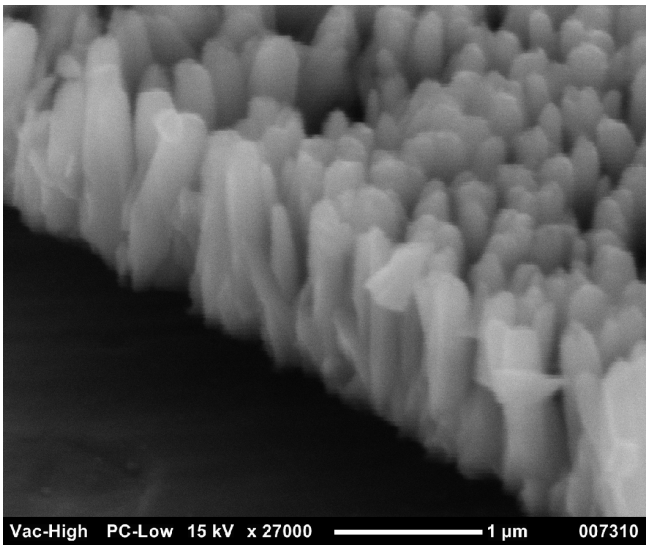
III. Results and Discussion

The top-view and cross-sectional SEM images of the ZnO nanorods on the Au-coated aluminum foil (Fig. 2) and the Au-coated lithium niobate (Fig. 3) substrate synthesized by electrochemical technique in the zinc nitrate/HTM aqueous solution at 60°C for 90 min were obtained.

The substrates is covered by the ZnO nanorods, which have a diameter in the range of 150 nm to 300 nm and a length of about 1 μm for ZnO nanorods on the Au-coated aluminum foil and 3 μm for ZnO nanorods on the Au-coated lithium niobate substrate. Standard deviation of length was 0.05 μm for ZnO nanorods on Au-coated aluminum foil and 0.22 μm for ZnO nanorods on Au-coated lithium niobate substrates. There were areas with longer rods on Au-coated lithium niobate substrates due to insufficient thickness of the Au layer and it caused higher std deviation of length. The density of the ZnO nanorods on Au-coated aluminum foil was $20 \cdot 10^8$ cm^{-2}, the density of the ZnO nanorods on Au-coated lithium niobate substrates was $30 \cdot 10^8$ cm^{-2}. Synthesized ZnO nanorods showed hexagonal structure almost along the full length.



a)

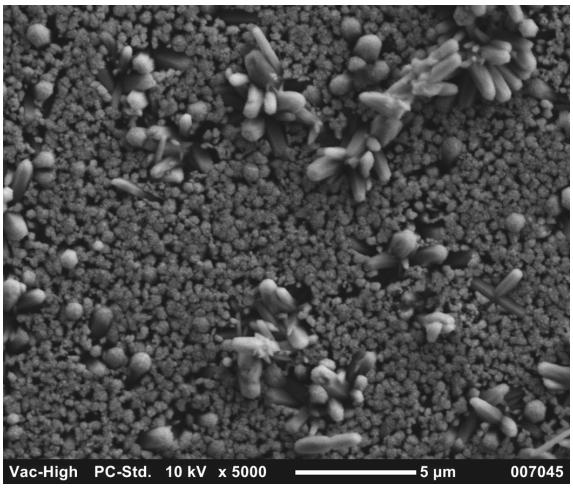


b)

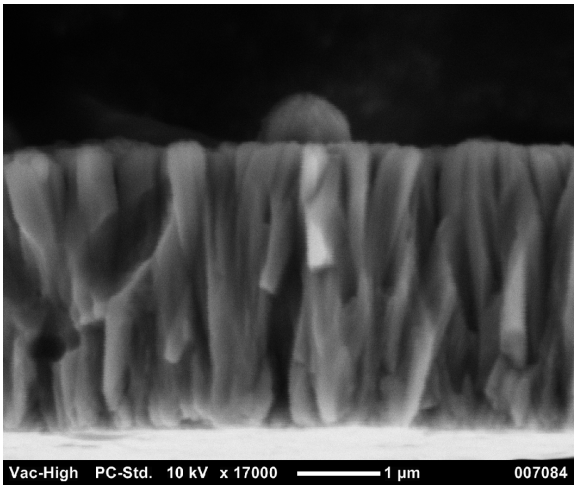
Fig. 2. SEM images of ZnO nanorods grown on Au-coated aluminum foil substrate: (a) top-view and (b) cross-sectional.

From observations of the interfaces (Fig. 2, b and Fig. 3, b), the ZnO nanowires had an excellent bonding interface with substrates and the growth direction was perpendicular to the surface.

The array of ZnO nanorods grown on the Au-coated aluminum foil is more regular than the array on the Au-coated lithium niobate substrate. The array of ZnO nanorods on the Au-coated lithium niobate substrate was denser than the array on the aluminum foil. On both substrates overgrowth at the roots occurred. The density of the incipient layer was significantly higher than density of the tops of the ZnO nanorods. It could be caused by defectiveness or insufficient thickness of the Au layer on dielectric lithium niobate substrate and better uniform distribution of electric field in the case of growth onto Au-coated aluminum foil.



a)



b)

Fig. 3. SEM images of ZnO nanorods grown on Au-coated lithium niobate substrate: (a) top-view and (b) cross-sectional.

We supposed that density and overgrowth could be controlled by matching of parameters of electrochemical growth such as concentrations of precursors, temperature and voltage. Effects of irregular growth could be overcome by the appropriate variation of the technological regimes during the growth process.

IV. Conclusion

In this paper we presented the array of ZnO nanorods grown on the Au-coated aluminum foil and on the Au-coated lithium niobate substrate by electrochemical technique in the zinc nitrate/HTM aqueous solution at 60°C for 90 min at applying 2VDC. Scanning electron microscopy results showed that ZnO nanorods are hexagon-shaped with a diameter in the range of 150 nm to 300 nm and a length of about 3 µm.

The obtained new results could be applied for the design of acoustic wave devices (specifically as sensing element of SAW sensors). Application of ZnO nanorods will promote quality and accuracy enhancement, manufacturing simplification, and improvements of simulation techniques of novel acoustic wave sensors due to mathematical representation simplicity of the regular-shaped structures. The proposed method will ensure high repeatability and consistency of production when all process parameters are satisfied.

Acknowledgment

The authors are grateful to Tokyo Boeki Technology Ltd. for providing the opportunity to carry out the research on Scanning electron microscope JCM-5000, NeoScope JEOL Ltd. at Laboratory of express-microscopy of National Technical University of Ukraine “KPI”.

References

[1] Ü. Özgür, D. Hofstetter, and H. Morkoç, “ZnO devices and applications: A review of current status and future prospects,” Proc. IEEE 98, 2010, pp. 1255-1268.

[2] Yicheng Lu, “Multifunctional Biosensor Based on Zno Nanostructures”, Acoust. Soc. Am. Vol. 131, Issue 2, 2012, pp. 1667-1667.

[3] D. Banerjee, J. Rybczynski, J. Y. Huang, D. Z. Wang, K. Kempa, and Z. F. Ren, "Large hexagonal arrays of aligned ZnO nanorods", Appl. Phys. A. 80, 2005, pp.749-752.

[4] M. H. Huang, Y. Y. Wu, H. Feick, N. Tran, E. Weber, P. D. Yang, “Catalytic growth of zinc oxide nanowires by vapor transport”, Adv. Mater. 13, 2001, pp. 113–116.

[5] Y. Sun, G. M. Fuge, M. N. R. Ashfold, “Growth of aligned ZnO nanorod arrays by catalyst-free pulsed laser deposition methods”, Chem. Phys. Lett. 396, 2004, pp. 21–26.

[6] W. I. Park, D. H. Kim, S. W. Jung, G. C. Yi, “Metalorganic vapor-phase epitaxial growth of vertically well-aligned ZnO nanorods”, Appl. Phys. Lett., 80, 2002, pp. 4232–4234.

[7] S. Ilican, “Preparation and characterization of ZnO thin filmsdeposited by sol-gel spin coating method” Journal of optoelectronics and advanced materials, Vol. 10, 2008, pp. 2578 – 2583.

[8] H. B. Zeng, J. B. Cui, B. Q. Cao, U. Gibson, Y. Bando, D. Golberg, “Electrochemical deposition of ZnO nanowire arrays: Organization, doping, and properties”, Sci. Adv. Mater., 2, 2010, pp. 336–358.

[9] B. Postels, A. Bakin, H.-H.Wehmann, M. Suleiman, T. Weimann, P. Hinze, A.Waag, “Electrodeposition of ZnO nanorods for device Application”, Appl. Phys. A 91, 2008, pp. 595–599.

[10] Hyunghoon Kim, Jin Young Moon, and Ho Seong Lee “Growth of ZnO Nanorods on Various Substrates by Electrodeposition”, Electronic Materials Letters, Vol. 5, No. 3, 2009, pp. 135-138.

Electrochemical Design of Zinc Oxide Arrays for Dye-Sensitized Solar Cells

E. I. Sokol, N. P. Klochko, G. S. Khrypunov,
V. R. Kopach, K. S. Klepikova, V. M. Lyubov,
A. V. Kopach
Materials Engineering for Solar Cells dept.
National Technical University (NTU "KhPI")
Kharkiv, Ukraine
klochko_np@mail.ru

Y. O. Myagchenko, E. E. Melnychuk
Experimental Physics dept.
National Taras Shevchenko University
Kiev, Ukraine
myagch@mail.ru

***Abstract*—Investigation of morphology, structural and optical characteristics of pulse plated zinc oxide arrays made possible the formation of the nanorods, size of which is varied by the electrolyte temperature. The surface morphology can be modified by duty cycle up to the growth of the mesoporous networks of zinc oxide. They together with nanorods will form hierarchical nanostructures with large specific surface that will ensure an efficiency of the created dye-sensitized solar cells.**

Keywords—pulse plating; zinc oxide; dye-sensitized solar cell.

I. INTRODUCTION

A dye-sensitized solar cell (DSSC) is a promising route toward solar energy conversion because it has advantages in both low-cost and high-efficiency, relative to the more established solar cell technologies [1-4]. Although standard DSSCs with current maximum efficiency of 11.1% are based on a mesoporous network of nanocrystallites it is believed [4] that the replacement of this network through one dimensional single-crystalline structures like ZnO nanowires could reduce the amount of grain boundaries and therefore the density of electron traps. On the other side, the drawback of ZnO nanowire arrays is the relatively small surface area compared to the one offered from a mesoporous network of nanocrystallites. Therefore, the growth of multiscale hierarchical structures has been shown to be an interesting mean to increase significantly the quantity of dye immobilized in the photoanode and to insure a fast charge collection. Many interesting hierarchical ZnO nanostructures with different geometries have been synthesized by a high-temperature thermal vapor evaporation method [5], however, the above process is poorly controlled. Electrochemical deposition (ECD) is known to be a simple, low-temperature, and cost-effective large-area deposition technique. High-quality ZnO films grown by ECD have been reported [1-3, 6-11], which can be easily scaled up for optoelectronic devices fabrications. According to [7-11], the pulsed electrochemical deposition with rectangular pulses (also called "pulse plating") exhibits a lot variables potentially affecting film growth. As shown in [7-11], the preferred orientation of films may be controlled by changing these pulse parameters. A detailed study on the influence of pulse parameters on the density and size of the ZnO crystals initially grown at short electrodeposition times presented in [9], but the researchers did not investigate ZnO arrays. So, the aim of this work is the study of possibilities of electrochemical design of zinc oxide arrays for dye sensitized solar cells, including research of modifying of ZnO arrays morphology, structural and optical characteristics by the adjustment of the pulse parameters.

II. EXPERIMENTAL

ZnO arrays were electrodeposited on transparent fluorine doped tin oxide covered glass cathodes (SnO_2:F, or FTO, Pilkington Corp., USA) in aqueous electrolytes containing 0.05 M $Zn(NO_3)_2$ and 0.1 M $NaNO_3$ in three-electrode cell with platinum counter-electrode and saturated Ag/AgCl reference electrode. Electrodeposition of each ZnO layer pattern was carried out during 1 hour at pulse plating mode with rectangular voltage impulses of cathode potential (with more negative cathode potential U_{on} = – 1.4 V and less negative one U_{off} = – 0.8 V versus saturated Ag/AgCl reference electrode). Pulse plating parameters were determined in accordance with [8]. A duty cycle (*Dc*) was given by:

$$Dc = \frac{T_{on}}{(T_{on} + T_{off})} = T_{on} \cdot f, \quad (1)$$

where T_{on} is time at potential U_{on}; T_{off} is time at potential U_{off}; f is frequency, defined as the reciprocal of the cycle time T.

Some ZnO layer patterns were dye-sensitized by immersion in a freshly prepared blackcurrant juice during 1 hour and by subsequent drying in air.

Study of the zinc oxide array surface morphology was conducted by using NTEGRA Prima Atomic Force Microscope (AFM) from NT-MDT in a tapping mode. Phase composition and structure of the deposited films were determined by XRD-method using an X-ray diffractometer DRON-4M in CoK_α radiation (the wavelength is λ = 1.7889 Å). Scanning was performed with Bragg–Brentano focusing (θ–2θ). The measured X-ray diffraction patterns were processed (background separation, $K_{\alpha1}$–$K_{\alpha2}$ doublet splitting, and others) and the parameters of the profile of the diffraction line were calculated using the New_Profile v.3.4 (486) and OriginPro v.7.5 software packages. The presence of crystal line phases was detected by comparing the data obtained from experimental X-ray diffraction patterns with the JCPDS database using the PCPDFWIN v.1.30 software package. The sizes of the grains as coherent scattering regions (*CSR*s) and

microstrains $\Delta d/d$ (where d is the crystal lattice parameter according to JCPCDS; Δd is the difference between the experimental and reference values for the crystal lattice parameter) were estimated by analyzing the broadening of the X-ray diffraction maxima, taking into account instrumental broadening by the Williamson–Hall approximation in accordance with [7]. The crystal lattice parameters were calculated by the position of the last two indexed lines of the X-ray diffraction patterns according to [1, 7, 10]. Preferable orientations of the films were researched by analytical treatment of the X-ray diffractions by means of obtaining of texture factor *P(hkl)* in accordance with [1, 6, 7].

The optical properties of the arrays of zinc oxide nanocrystallites were studied in the spectral range of 350–850 nm using an SF-46 double beam spectrophotometer. The transmittance spectra of ZnO layers and dye-sensitized ZnO arrays were measured, when FTO-glass substrates were placed in a reference channel. The optical band gap E_g of the zinc oxide layers was determined according to [7] by extrapolating the linear portion of the dependence $[(-\ln T)\cdot h\nu]^2$ on $h\nu$ to the energy axis. The structural disorder of the electrodeposited zinc oxide layers was estimated according to [7] by the Urbach energies E_o obtained from the slope of the linear portion of the dependence $\ln[-\ln(T)]$ on $h\nu$.

III. Results

The study of the crystal structure of the zinc oxide arrays electrodeposited in regimes presented in Table 1 has shown that all diffraction peaks match the hexagonal structure of wurtzite ZnO (with the exception of reflections assigned to FTO-glass substrates). Comparative analysis of XRD patterns of the ZnO layers electrodeposited at different pulse frequencies f has revealed (Fig. 1) that by changing the frequency one can affect the texture of the ZnO arrays. This observation is consistent with that described in [11]. Our calculations showed that all electrodeposited arrays were nanocrystalline (their coherent scattering regions do not exceed tens or hundreds of nanometers), characterized by a slightly increased lattice parameters compared to the reference ZnO (according to JCPDS 36-1451, a = 3,250 Å, c = 5,207 Å) and have minor compression microstrains. Unfortunately, for the samples electrodeposited at frequencies f = 3 Hz and f = 5 Hz, it was not possible to use the method of Williamson–Hall approximations for the determination of the contributions of crystal sizes and microstrains to the broadening of the diffraction peaks, since the calculated *CSR*s exceeded values allowed within this method, so data on microstrains for these patterns are missing in Table 1. The vertical preferential orientation of ZnO nanocrystals with the c axis perpendicular to the substrate (<001> texture) is manifested in the diffraction patterns (Fig. 1) as intense peaks corresponding to the planes (002) and (004). The increase in the pulse frequencies above 2 Hz decreased the <001>preferred orientation of the ZnO arrays. Fig. 2 shows how the f growth decreases the texture factor *P(002)* of ZnO arrays. Note, that in [7] we have shown a reduction in <001> texture perfection at frequencies below 2 Hz. Transmittance spectra and graphs for the determining the optical band gaps of ZnO layers are shown in Fig. 3. Analysis of dependence of the optical properties of zinc oxide arrays on

TABLE 1 Electrodeposition Parameters and Structure Characteristics of ZnO

ECD parameters			*Structure of the ZnO arrays*				
			P(002)	*CSR, nm*	*Δd/d, 10^{-3}*	*Lattice parameters*	
t, °C	*f, Hz*	*Dc*				*a*	*c*
60	1000	40	2.13	60-100	0.3-0.6	3.253	5.215
70	200		2.01	19-66	6.4-7.9	3.250	5.206
	20		3.12	73-163	0.3-0.6	3.253	5.220
	10		3.39	40-45	0.9-1.4	3.257	5.218
	6		4.91	73-150	0.1-0.8	3.258	5.215
80	5		4.88	> 200	–	3.257	5.217
70	3		5.25	> 200	–	3.259	5.217
70	2		5.43	60-80	0.5-0.8	3.256	5.215
50		60	–	20-30	0.9-1.9	3.258	5.211
30		60	2.14	20-30	1.3-2.4	3.259	5.217
60		60	–	40-70	0.6-1.1	3.255	5.218
85		40	4.00	60-100	0.2-0.4	3.256	5.213
70		20	3.21	65-105	0.3-0.5	3.256	5.214
		60	–	60-100	1.2-1.6	3.259	5.220
		80	3.21	40-65	0.3-1.1	3.257	5.212

the cathode potential pulse frequency has shown (Figs. 2, 3) that in the range of f from 2 to 200 Hz the transparency of ZnO arrays in the visible region decreases simultaneously with the deterioration of the texture <001> as the frequency grows. However, a further f increase up to 1 kHz makes arrays anomalously transparent, but the last is not due to ZnO perfect crystal structure, but is a result of the reduced thickness of the films grown at high frequencies. According to [12], at high-

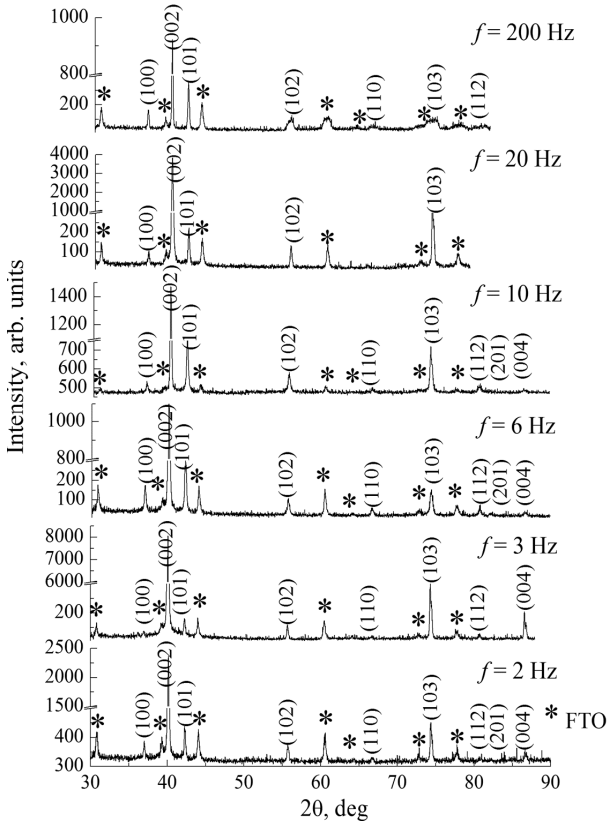


Fig.1 X-ray diffraction patterns of ZnO arrays deposited on FTO-glass substrates at t = 70 ° C, Dc = 40, and different frequencies f.

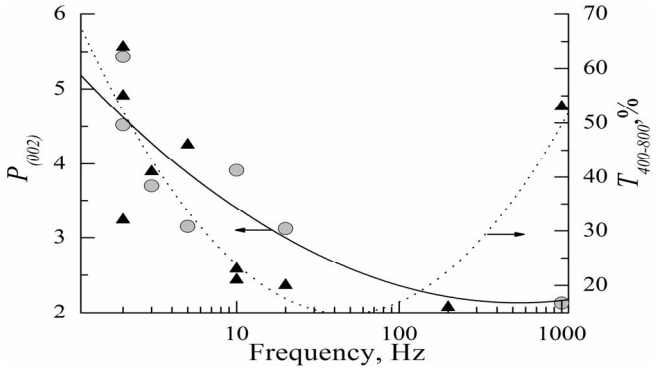


Fig. 2 Dependence of the texture factor *P(002)* and the average optical transmittance in the visible range $T_{400\text{-}800}$ on the pulse frequency *f* for ZnO arrays electrodeposited at *t* = 70 ° C, and *Dc* = 40.

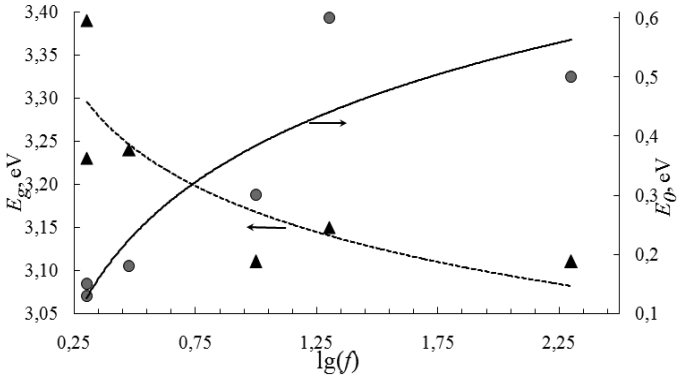


Fig.4 Dependence on the pulse frequency *f* of the band gap E_g and Urbach energy E_o of ZnO arrays, electrodeposited at *t* = 70 ° C, and *Dc* = 40.

frequency electrodeposition modes, most of the electric charge passed through the electrolyte is consumed not by Faraday oxidation at the anode and cathode, but he spent to recharge the electric double layers. The increase in the f above 2 Hz causes a reduction of the band gap E_g and an increase of Urbach energy E_o (Figs. 3, 4), that is, this semiconductor material becomes more disordered and has narrow band gap. As for using in the DSSC hybrid solar cells, the ZnO array can be as a more transparent in the visible range and well-textured in the <001>, so, here we consider more details the ZnO films electrodeposited with cathode impulse frequencies *f* = 2-20 Hz.

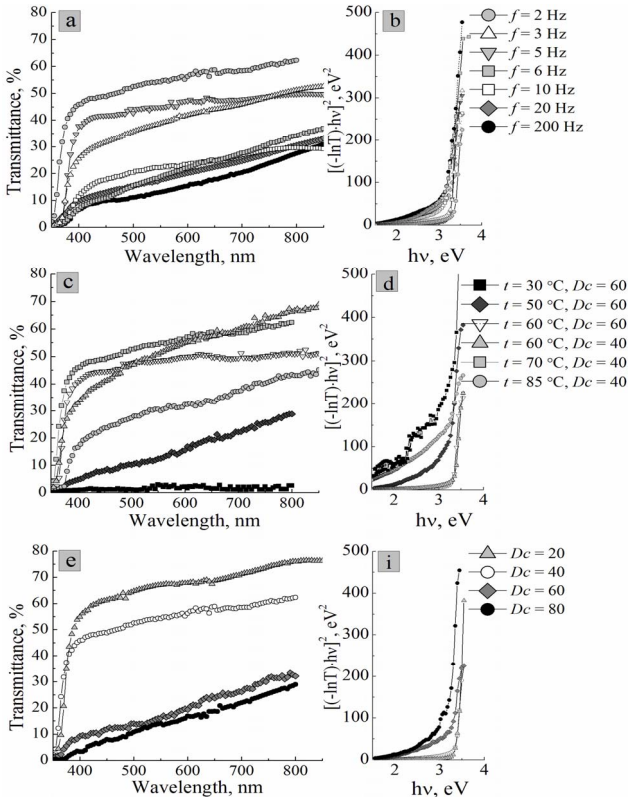


Fig. 3 Optical transmittance spectra (a, c, e) and dependences of $[(-\ln T)\cdot h\nu]^2$ on the photon energy (b, d, i) for pulse plated ZnO layers: a, b – at *t* = 70 ° C, Dc = 40 and different *f*; c, d – at *f* = 2 Hz and different *t*; e, i – at *f* = 2 Hz; *t* = 70 ° C, and different *Dc*.

As seen in Fig. 5, the surface morphologies of ZnO arrays are very diverse and depend not only on *f*, but also on other electrodeposition parameters, namely, duty cycle *Dc* and the electrolyte temperature *t*. Our investigations have shown that the pulse electrodeposition of zinc oxide at temperatures below 60 ° C did not provide the formation of well-structured arrays. Produced by atomic force microscopy image in Fig. 7 a shows spheroid formations of 200-600 nm in diameter grown at *t* = 50 °C. X-ray diffraction studies showed (Table 1) that the arrays obtained at *t* = 30 – 50 ° C are formed by very small ZnO nanocrystallites (*CSR* = 20 – 30 nm), and are in a disorderly manner blinded in a rather large "lumps". Optical transmittance in the visible range of the ZnO pattern electrodeposited at 50 ° C did not exceed 29%, the band gap was small E_g = 3.18 eV, and Urbach energy, which is a measure of disorder in the crystal structure, by contrast, was abnormally high E_o = 0.44 eV (Fig. 6). When the bath temperature was only 30 ° C, an array of ZnO was not only practically non-textured, but characterized by the presence of relatively large microstrains (Table 1), its transparency does not exceed 3%, E_g was 3.23 eV, and the E_o equaled 0.46 eV (Fig. 6).

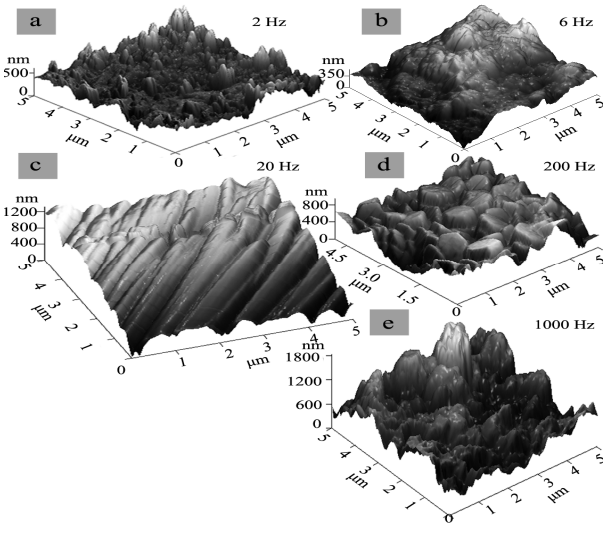


Fig.5 AFM images of ZnO arrays from Table 1 deposited at *Dc* = 40 and *t* = 70 ° C (at *f* range 2 – 200 Hz) (a, b, c, d) or *t* = 60 ° C (at *f* = 1000 Hz) (e).

Increasing of the temperature of the electrolyte in the range from 60 ° C to 85 ° C resulted in the growth of well textured zinc oxide arrays (Table 1). The sample grown at t = 60 ° C, was very poorly textured in the <102> and <103> ($P(102)$ = 1.90; $P(103)$ = 1.71), as reflected in its morphological features that can be seen in the AFM images, shown in Fig. 5 e and Fig. 7 b. The arrays of zinc oxide electrodeposited at t = 70 ° C and at t = 85 ° C shown in Fig. 5 a, c, Fig. 7 c and in Fig. 7 g, respectively, consisted of nanorods, oriented perpendicular to the substrate surface, since they had the perfect texture in the <001>. As can be seen in Fig. 7, nanorods became larger as the temperature of the electrolyte increased, the diameter of the hexagonal base increased from 100 nm (t = 70 ° C) to 1.5 μm (t = 85 ° C). Analysis of the optical properties of electrodeposited ZnO arrays (Fig. 3 c, d, and Fig. 6) in the electrolyte temperature range t = 30 – 70 ° C confirms an increase in the transparency and optical band gap and the shortening of Urbach energy of ZnO films with t growth. However, in the case of heating of the electrolyte to t = 85 ° C the material becomes more opaque, narrow band gap and disorder.

To vary the morphology of electrodeposited zinc oxide arrays we used not only the change in temperature of the electrolyte, but also such factor as the duty cycle. It turned out that at the same cathode pulse frequency and electrolyte temperature t = 70 ° C the preferred orientations, *CSR*s (Table 1), and overall morphology of electrodeposited zinc oxide arrays (Fig. 7 c, d, e, i) can greatly vary. From Fig. 3 e, i, it is clear that of view of the optical properties of zinc oxide arrays small quantities Dc = 20 – 40 are preferred. The AFM investigations presented in Figure 7 show the ability to change the morphological features of zinc oxide arrays in a wide range at the same t = 70 ° C. For example, at Dc = 20 "leaves" with a diameter of 1 μm and a thickness of 100 nm are formed (Fig. 7 e). The oriented perpendicular to the substrate zinc oxide nanorods of 100 nm in diameter have been obtained at Dc = 40 (Fig. 7 c). When Dc = 60, a stack of two-dimensional hexagonal plates with thickness of 50 nm and a diameter of 600 – 1000 nm have been formed (Fig. 7 d). If Dc = 80, three-dimensional networks of this material with a filament thickness of 100 nm and with hexagonal network cells of 200 – 600 nm in diameter have been deposited (Fig. 7 i). Our investigations have shown that a saturation of the electrodeposited ZnO layers with the natural dye from the freshly prepared blackcurrant juice provides a good absorption of visible light (Fig. 8) confirming suitability of these ZnO arrays for use in the hybrid solar cells.

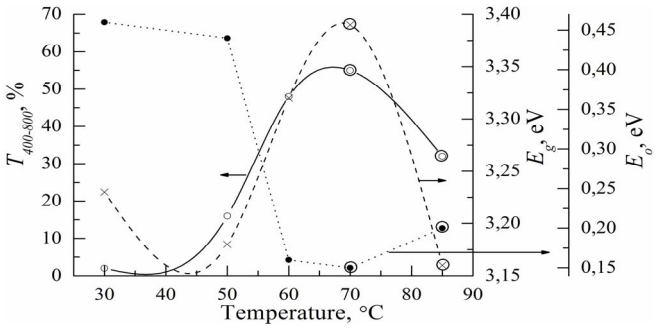


Fig. 6 Dependence on the electrolyte temperature t of the average optical transmittance in the visible range $T_{400\text{-}800}$, the band gap E_g and Urbach energy E_o of the ZnO arrays electrodeposited at f = 2 Hz, Dc = 40 (circle) and Dc = 60.

As we reported earlier in [7], during the pulse electrodeposition of the zinc oxide arrays, at T_{off}, the nitrate ions are catholically reduced to form hydroxyl groups, whose interaction with the zinc cations in the cathode region leads to the formation ZnO. At a negative potential of the cathode U_{on}, the above processes are more intense, and, in addition, a cathodic reduction of water releases hydrogen

$$2H_2O+2\bar{e}\rightarrow H_2\uparrow+2OH^- \quad (2)$$

and a reduction of zinc occurs

$$Zn^{2+}+2\bar{e}\rightarrow Zn \quad (3)$$

this is followed at U_{off} by heterogeneous chemical reaction leading to the formation of an additional amount of zinc oxide

$$Zn+NO_3^-+H_2O\rightarrow NO_2^-+Zn(OH)_2\rightarrow NO_2^-+ZnO+H_2O\ . \quad (4)$$

Previously [7, 8, 10] we proved that for the preferential growth of zinc oxide arrays of wurtzite modification in the direction perpendicular to the substrate surface, i.e., along with

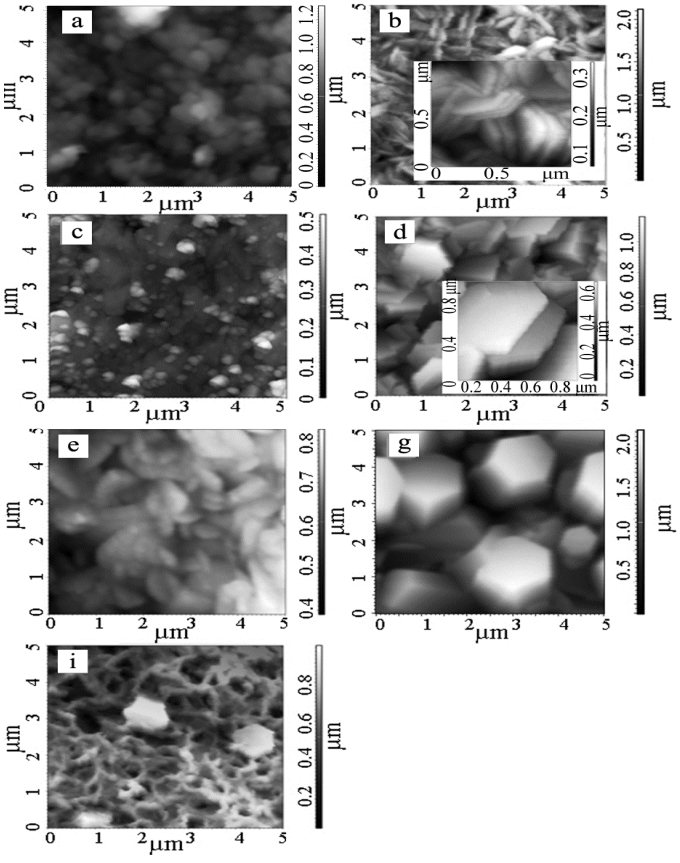


Fig.7 AFM images of ZnO arrays from Table 1 deposited at f = 2 Hz: a – t = 50 ° C, Dc = 60; b – t = 60 ° C, Dc = 60; c – t = 70 ° C, Dc = 40; d – t = 70 ° C, Dc = 40; e – t = 70 ° C, Dc = 20; g – t = 85 ° C, Dc = 40; i – t = 75 ° C, Dc = 80.

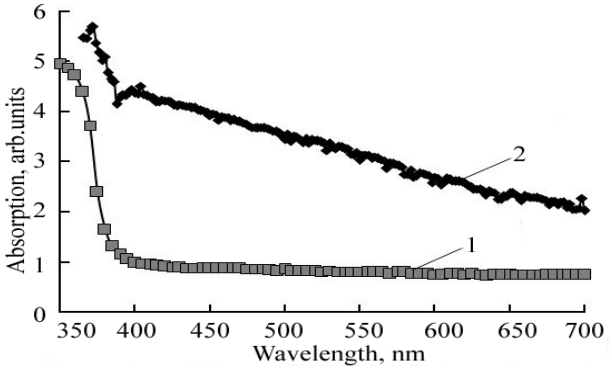


Fig.8 Absorption vs. wavelength curves for bare ZnO array from Table 1 electrodeposited at f = 5 Hz, t = 80 ° C, Dc = 60 (1) and for this ZnO layer sensitized by dye from the freshly prepared blackcurrant juice (2).

c axis, it is necessary to eliminate the adsorption of the hydrogen bubbles formed on the (002) plane during the electrochemical cathodic deposition of ZnO. In this paper, using the X-ray diffractometry and optical studies we demonstrated that above requirements are fulfilled by the pulse frequency f = 2 Hz, which favors the electrodeposition of the zinc oxide arrays with a perfect texture in the <001>, high transparency and acceptable optical quality ZnO in the visible spectrum, which is likely due to the rapid desorption of gaseous products from the cathode. The results of studies presented in this paper also showed that in order to ensure desorption of gases from the surface of the growing zinc oxide array to form ZnO nanorods textured in the <001>, the electrolyte should be heated to a temperature of at least 70 ° C, and, the more heat the electrolyte, the larger one-dimensional nanostructures are formed, due to the acceleration and intensification of diffusion and electrochemical processes with temperature. On the other hand, this means that by cooling the electrolyte during the electrodeposition from 85 ° C to 70 ° C, the fine zinc oxide nanorods on the surface of the large nanorods could be grown and thus hierarchical nanostructures will be created. With a greater reduction of the electrolyte temperature from 85 ° C to 60 ° C we can form thin (100 nm thick) layered hexagonal plates, such as those shown in Fig. 7 b, on the surfaces of the big nanorods. Another tool to control the surface morphology of the electrodeposited zinc oxide arrays for the creating of the hierarchical nanostructures was duty cycle, because by operating of T_{on} and T_{off} we can regulate the electrochemical and diffusion stages during ZnO growth. For example, our studies have shown, in particular, that ZnO nanorods are formed only if Dc = 40, and a structure with a large specific surface in the form of mesoporous networks, as shown in Fig. 7 i, is formed at Dc = 80. Therefore, by a simple and affordable admission, namely by changing the mode of electrodeposition with the programmer supplied with the potentiostat, the surface of the ZnO nanorods can be "encapsulated" by the mesoporous networks of the same material tightly bound with the nanorods both mechanically and electrically, and in this way hierarchical zinc oxide nanostructures are formed with a large surface area and a small amount of grain boundaries for the significant increasing of the quantity of the dye immobilized in the photoanode and to insure a fast charge collection in the produced DSSCs.

IV. CONCLUSIONS

Studies on the effect of pulse electrodeposition on the structural parameters, morphology and optical properties of zinc oxide arrays made possible to identify the modes that are optimal for the formation of ZnO nanorods oriented perpendicular to the substrate: electrolyte temperature t = 70 – 85 ° C, duty cycle Dc = 40; pulse frequency f = 2 Hz. It was found that the size of the nanorods can be varied by heating or cooling the electrolyte in the above-mentioned range, thus, in particular, small nanorods could be electrodeposited on the surface of the larger ones, thus forming a hierarchical nanostructure. In addition, by varying the duty cycle, the surface morphology can be modified up to the growth of the mesoporous zinc oxide networks. They together with ZnO nanorods will form hierarchical nanostructures with large specific surface and little grain boundaries, which will ensure an efficiency of the created dye-sensitized solar cells.

REFERENCES

[1] Lupan, O., Guérin, V.M., Tiginyanu, I.M., Ursaki, V.V., Chow, L., Heinrich, H., and T. Pauporté, "Well-aligned arrays of vertically oriented ZnO nanowires electrodeposited on ITO-coated glass and their integration in dye sensitized solar cells," *Journal of Photochemistry and Photobiology A: Chemistry*, vol. 211, pp. 65–73, 2010.

[2] Zhou, B., Han, X., Feng, Q., and M. Tao, "A greener electrodeposition recipe for ZnO films in terawatt photovoltaics," *Thin Solid Films*, vol. 520, pp. 6730–6733, 2012.

[3] Gue´rin, V.-M., Rathousky, J., and Th.Pauporte´, "Electrochemical design of ZnO hierarchical structures for dye-sensitized solar cells", *Solar Energy Materials & Solar Cells*, vol. 102, pp. 8–14, 2012.

[4] Wagner, A., Behrends, A., Waag, A., and A. Bakin, "Two step deposition method with a high growth rate for ZnO nanowire arrays and its application in photovoltaics", *Thin Solid Films*, vol. 520, pp. 4637–4641, 2012.

[5] Lao, J. Y., Huang, J. Y., Wang, D. Z., and Z. F. Ren, "Hierarchical oxide nanostructures," *Journal of Materials Chemistry*, vol. 14, pp. 770–773, 2004.

[6] Lupan, O., Pauporte´, T., Chow, L., Viana, B., Pelle´, F., Ono, L.K., Cuenya, B. Roldan, and H. Heinrich, 'Effects of annealing on properties of ZnO thin films prepared by electrochemical deposition in chloride medium," *Applied Surface Science*, vol. 256, pp. 1895–1907, 2010.

[7] Klochko, N. P., Khrypunov, G. S., Myagchenko, Yu. O., Melnychuk, E. E., Kopach, V. R., Klepikova, E. S., Lyubov, V. M., and A. V. Kopach**, "**Controlled growth of one-dimensional zinc oxide nanostructures in the pulsed electrodeposition mode", *Semiconductors*, vol. 46, no. 6, pp. 825–831, 2012.

[8] Klochko, N. P., Khrypunov, G. S., Myagchenko, Yu. O., Melnychuk, E. E., Kopach, V. R., Klepikova, E. S., Lyubov, V. M., and A. V. Kopach**,** "Electrodeposited zinc oxide nanostructures with adsorbed natural dyes for photovoltaic applications", *Functional Materials*, vol. 19, no. 2, pp. 276–281, 2012.

[9] Dunkel, C., Lüttich, F., Graaf, H., Oekermann, T., and M.Wark, "Investigation of the pulsed electrochemical deposition of ZnO", *Electrochimica Acta*, vol. 80, pp. 60–67, 2012.

[10] Khrypunov, G., Klochko, N., Volkova, N., Kopach, V., Lyubov, V., and K. Klepikova, "Pulse and direct current electrodeposition of zinc oxide layers for solar cells with extra thin absorbers," in World Renewable Energy Congress WREC-2011, Link□ping, Sweden, 8-13 May 2011, 1002PV.pp.1–8, 2011.

[11] Nomura, K., Shibata, N., and M. Maeda, "Orientation control of zinc oxide films by pulsed current electrolysis," *Journal of Crystal Growth*, vol. 235, pp. 224–228, 2002.

[12] Chandrasekar, M.S., and M. Pushpavanam, "Pulse and pulse reverse plating—Conceptual, advantages and applications," *Electrochimica Acta*, vol. 53, pp. 3313–3322, 2008.

Information-Theoretic Estimates of Communication and Processing in NanoScale and Quantum OptoElectronic Systems

Sergey Edward Lyshevski
Department of Electrical and Microelectronic Engineering
Rochester Institute of Technology
Rochester, NY 14623, USA
Sergey.Lyshevski@mail.rit.edu, http://people.rit.edu/seleee

Leonid Reznik
Department of Computer Science
Rochester Institute of Technology
Rochester, NY 14623, USA
lr@cs.rit.edu

***Abstract* – This paper applies information theory to examine communication and processing in nanoscale and quantum opto-electronic systems. Classical, semiclassical and quantum information-theoretic analyses and measures are proposed and examined. The reported results and new findings enable practical technological solutions supported by nanoscaled microelectronics and optoelectronics. Our results may advance communication and processing schemes. We contribute to areas of a critical importance and significance, such as communication, data processing, biotechnology, neurophotonics, photonics, etc.**

Keywords – electronics; information theory; processing; nanotechnology; quantum informatics

I. INTRODUCTION

Sensing, communication, data fusion, data exchange, computing and processing are accomplished applying different schemes and technologies. Quantitative analyses, complexity measures and communication estimates were proposed in [1-3]. Significant advances in microelectronics, photonics and optoelectronics led to exceptional communication and processing capabilities. However, fundamental and technological grand challenges were identified for nano- and quantum electronics, optoelectronics and photonics [4]. There are needs to further develop information science and technology.

Nanoscaled microelectronics identified and encountered fundamental and technological limits [4-8]. The majority of solid-state devices operate utilizing electrostatic phenomena, while quantum effects degrade device performance and system capabilities. Enabled microelectronics and optoelectronics guarantee high-performance interfacing, communication and computing. Further needs, practical advancements and technological inroads are outlined in [4]. Enabling quantum-centric communication, computing and processing schemes are the subjects of intensive research. Results in quantum-probabilistic and quantum-deterministic communication, computing and hardware solutions are reported in [4, 9-11].

Molecular and optoelectronics technologies, as well as new developments in physical sciences, enable emerging semi-classical and quantum communication and processing schemes. Quantum and compliant quantum↔classical processing, exhibited by nervous systems [9-13], is an open fundamental problem of a great significance. Sensing and processing, which are accomplished by living systems, substantiate quantum, classical→quantum and quantum→classical schemes [14, 15]. In medicine, as well as in life and physical sciences, advancements in a knowledge base and transformative discoveries are of a great importance.

We depart from theoretical quantum computing hypothesis which attempt to attest a possibility of "quantum computing" by means of narrative models of computing, hypothetical use of not observable mathematical operators, etc. We research quantum communication and processing on real-valued, detectable and measurable variables. The quantum transitions of these physical variables should ensure sensing, communication and processing in electronic, optoelectronic and photonic systems. This paradigm typifies asynchronous spatially- and time-distributed quantum sensing, communication and processing in living organisms.

Emerged nanoelectronics and photonics require new information processing and information theory paradigms [4]. High-speed, high-density and low-power nanoscaled logic, memory, photonic and other devices ensure enabling performance and capabilities. However, these systems lead to enormous challenges due to high noise, failures, error rates, distortions, etc. In fact, "The error rate of all nanoscale devices and circuits is a major concern" [4]. Advanced architectures, enabling organizations and fault tolerant schemes were demonstrated [4]. To achieve communication benchmarks and guarantee high-performance processing, a great number of open problems must be solved. Consistent information measures and practical quantum schemes are the grand challenges which will enable a knowledge base of information theory and technology, thereby contribution to the *beyond CMOS, big data* and other paradigms [4].

II. II.CLASSICAL, SEMICLASSICAL AND QUANTUM OPTOELECTRONICS

Optical photon absorption and emission, as well as electronic transductions, are utilized by natural biomolecular and engineered optoelectronic systems. Quantum phenomena are utilized by single-photon biomolecular detectors, transmitting and processing molecular fabrics in visual, photosynthesis and other systems. In contrast, engineered sensors, receivers, waveguides and processors operate on immense number of electrons and photons mainly utilizing

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

photoelectric and electrostatic effects. Classical, *hybrid* and quantum communication and processing on wave packets and a single photon are reported in [4-7, 9-11]. Utilizing nanotechnology and optoelectronics, prototypes of integrated laser – waveguide – sensor (detector) – post-processing electronics are designed [4-7, 9-11]. Communication, interfacing and computing by optical photons were extensively examined by IBM, Intel, HP, etc. Robust ~100 Gbit/sec interfacing and communication were achieved [4-7]. However, significant challenges remain. Nanoscale silicon, organic and polymer microelectronics and photonics for sensing, communication and processing are under developments in Rochester Institute of Technology as documented in Fig. 1.

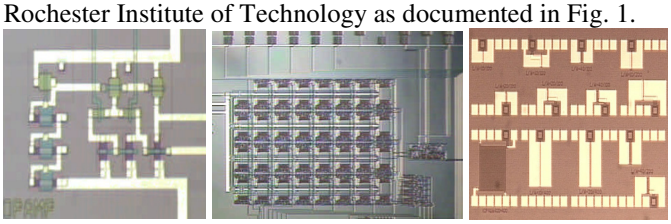


Fig. 1. Images of functional silicon and organic electronic, opto-electronic and photonic devices fabricated in RIT

III. Information-Theoretic Analyses

We examine communication and processing schemes in order to assess and approach the Shannon and quantum-mechanical performance limits in classical and quantum domains. Consistent quantitative information performance estimates and capability measures should be derived. Our goal is to derive coherent information-centric estimates, measures and metrics in order to evaluate communication and processing capabilities. The classical and quantum communication, information fusion, memory schemes and processing premises are different. We focus on conventional, *hybrid* and quantum deterministic processing.

For a random variable X_i with n outcomes $\{x_i: i=1,2,\ldots,n-1,n\}$, which has the probability distribution functions $p(x_i)$, the entropy is [1]

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)\cdot \qquad (1)$$

The classical mutual information $I(X;Y)$ between inputs X and outputs Y, as well as quantum $\mathcal{I}(\mathcal{X};\mathcal{Y})$, are predefined by the classical and quantum entropies H and $\mathcal{H}$ [9-11]. The mutual information between the stimulus and the response quantitatively defines the amount of information received on average. The mutual information provides a quantitative estimate on the dependence of two discrete random variables X and Y. One has [1, 3]

$I(X;Y)=I(Y;X)=H(X)-H(X|Y)=H(Y)-H(Y|X)=H(X)+H(Y)-H(X,Y)$,

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} p_{X,Y}(x,y)\log_2 \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} \text{ [bits/symbol]. (2)}$$

A positively-definite mutual information $I(X;Y)\geq 0$ provides the average amount of information received per symbol transmitted or processed. The conditional entropy $H(X|Y=y)$ of a random variable X, conditional on a particular realization y of Y, is the expected conditional information content, and $H(X|Y=y) = -\sum_x p_{X|Y}(x|y)\log_2 p_{X|Y}(x|y)$. The expected conditional information content with respect to both X and Y is given by the conditional entropy

$$H(X|Y) = -\sum_{x\in X}\sum_{y\in Y} p_{X,Y}(x,y)\log_2 p_{X|Y}(x|y),\ H(X|Y)\geq 0. \qquad (3)$$

One has $H(X)\geq H(X|Y)$, and, $H(X|Y)=H(X)$ if X and Y are independent, yielding $I(X;Y)=0$. The conditional entropy $H(X|Y)$ corresponds to the average loss of information. The joint entropy is the entropy of the joint random variable (X,Y)

$$H(X,Y) = -\sum_{x,y} p_{X,Y}(x,y)\log_2 p_{X,Y}(x,y), \qquad (4)$$

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y).$$

3. 1. Quantum-Mechanical Modeling and Entropy

Microscopic systems may be described by consistent models. Solving the Schrödinger equation, one finds the spatiotemporal wave function $\Psi(t,\mathbf{r})$. Mathematically, the system can be found in any given state with $|\psi_j\rangle$ with probability p_j. The density matrix is $\rho = \sum_{i=1}^{N} p_i\rho_i = \sum_{i=1}^{N} p_i|\psi_i\rangle\langle\psi_i|$. One may analyze the randomness of an ensemble $\mathcal{E}$. Using the eigenvalues λ_ρ of the source density matrix ρ, von Neumann's entropy $S(\rho_\mathcal{E})$ of a mixed state for non-orthogonal *pure states* is

$$S(\rho_\mathcal{E}) = -\sum_{i=1}^{N} \lambda_{\rho i}\ln\lambda_{\rho i} = -\mathrm{tr}(\rho\ln\rho)\cdot \qquad (5)$$

Quantum-mechanical analysis of *microscopic* systems may result in time-varying transition probability, yielding the time-varying quantum entropy [9-11]

$$\mathcal{S}(t,p) = -\sum_{i=1}^{n} p_i \ln p_i \qquad (6)$$

with the probability of i to be p_i. The entropy rate is $\mathcal{S}_r(t,p) = -\sum_i p_i\sum_j p_i(j)\ln p_i(j)$, where $p_i(j)$ is the probability of j given i as the possible state.

The classical entropy H and quantum entropies ($\mathcal{H}$ and S) are fundamentally distinct. Using the physical, detectable and measurable quantities, in *macroscopic* and *microscopic* systems one may measure (X,Y) or $(\mathcal{X},\mathcal{Y})$, obtaining $p(x_i)$ and $p(y_i)$, or, $p(\mathcal{x}_i)$ and $p(\mathcal{y}_i)$. This leads to H and $\mathcal{H}$. In contrast, the von Neumann's entropy $S(\rho_\mathcal{E})$ provides mathematical quantitative estimates using operators, and, the quantum quantities may not be detectable and observable.

3. 2. Quantum Processing

For many decades, there were obscure attempts to substantiate a practicality and effectiveness of stochastic-centric classical and quantum computing. In contrast, analog and digital deterministic computing and processing guarantee exceptional performance. These paradigms were implemented in all existing data processing platforms. The deterministic quantum processing on real-valued physical variables is emphasized. Quantum communication and processing on a single or few photons or electrons will result in significant uncertainties, distortions and errors [4].

We apply an inherent quantum determinism to ensure quantum-deterministic communication and processing on the *utilizable* initial (I) and final (F) *state* transductions $\mathcal{S}=[\mathcal{S}_I,\mathcal{S}_F]^T$ $\mathcal{S}_I:\boldsymbol{\nu}_I\Rightarrow\mathcal{S}_F:\boldsymbol{\nu}_F$ on real-valued physical variables $\boldsymbol{\nu}$ [9]. As $\mathcal{X}$ and $\mathcal{Y}$ can be measured, one yields $\mathcal{H}(\mathcal{X};\mathcal{Y})$. This concept is substantiated by *natural* systems, as well as by quantum-effect nanoelectronic, optoelectronic and photonic devices which are widely deployed and commercialized [4]. A quantum-effect processing primitive $\mathcal{P}_j$ exhibits transductions $\mathcal{S}_j(\boldsymbol{\nu})$ on *detectable*, *measurable* and *processable variables* $\boldsymbol{\nu}_j$ yielding *distinguishable* and *computable* transforms $\mathcal{T}_j(\mathcal{S},\boldsymbol{\nu})$.

Considering conventional and quantum-deterministic communication, the channel capacity of a stationary memoryless channel with finite input and output alphabets is

$$C=\max_{p_X(\cdot)} I(X;Y) \text{ and } \mathcal{C}=\max_{p_X(\cdot)} \mathcal{I}(\mathcal{X};\mathcal{Y}) \text{ [bits/symbol]}. \quad (7)$$

If the transition probabilities vary, for nonstationary memoryless channels, one obtains

$$C=\max_{p_{X_1}(\cdot),\ldots,p_{X_n}(\cdot)} \frac{1}{n}\sum_{i=1}^{n} I(X_i;Y_i)=\frac{1}{n}\sum_{i=1}^{n} C_i, \text{ [bits/symbol]}. \quad (8)$$

Significant errors and distortions are observed in low-power nanoscale microelectronics, nanoICs, nanowaveguides and ~100 Gbit/sec CMOS nanophotonic communication [4-7]. High-performance computing, communication and processing platforms must be designed despite the aforementioned errors and distortions. Using the probability of a bit error p_b, the maximum transmission rate is

$$r_{\max}(p_b)=\frac{C}{1-H_2(p_b)}, H_2(p_b)=-[p_b\log_2 p_b+(1-p_b)\log_2(1-p_b)], \quad (9)$$

where H_2 is the binary entropy function.

Consider a sequence $X_1,\ldots, X_n$ with $p(x)$ and a finite alphabet $\mathcal{A}$, $x\in\mathcal{A}$. Using the reproduced alphabet $\mathcal{A}_r$ with symbols $x_r\in\mathcal{A}_r$, the distortion measure $d:\mathcal{A}\times\mathcal{A}_r\to\mathbb{R}$ is finite, e.g., $d_{\max}=\max_{x\in A, x_r\in A_r} d(x,x_r)<\infty$. The distortion depends on the sequences, encoding and decoding functions, etc. The rate distortion function $(r,\mathcal{D})$ for a source X with $d(x,x_r)$ can be defined as $r_I(\mathcal{D})=\min I(X;X_r)$. This minimum must be found with respect all condition distributions $p(x_r|x)$ for which the joint distribution $p(x,x_r)=p(x)p(x_r|x)$ satisfies the imposed distortion constraints. Furthermore,

$$I(X;X_r)=H(X)-H(X|X_r)\ge H(p)-H(\mathcal{D}),\ r_I(\mathcal{D})\ge H(p)-H(\mathcal{D}). \quad (10)$$

IV. Information-Theoretic Measures

Communication and processing capabilities must be quantitatively evaluated for the emerged nanoelectronics, CMOS-integrated nanophotonics, as well as for molecular and optoelectronic nanotechnologies. Our objective is to derive coherent information quantities. We introduce the processing estimates, as well as data complexity and data quality measures, which consistently assess and evaluate classical, consistent quantum↔classical and quantum data processing.

4. 1. Processing Estimate

The data processing capability estimates are

$$D=BI(X;Y)H(X|Y)r(p_b),\ \mathcal{D}=\mathcal{R}\mathcal{I}(\mathcal{X};\mathcal{Y})\mathcal{H}(\mathcal{X}|\mathcal{Y})r(p_b), \quad (11)$$

$$I(X;Y)=H(X)-H(X|Y),\ \mathcal{I}(\mathcal{X};\mathcal{Y})=\mathcal{H}(\mathcal{X})-\mathcal{H}(\mathcal{X}|\mathcal{Y}),$$

where B is the bandwidth; $\mathcal{R}$ is the quantum transition rate in the *microscopic* processing system.

The proposed quantitative D and $\mathcal{D}$ allow one to evaluate the processing performance and processing platforms capacity.

4. 2. Data Complexity

The entropies H and $\mathcal{H}$ define the data set complexity. A finite length of the string $x\in\{0,1\}$ is denoted as $l(x)$. The Kolmogorov descriptive complexity $K_U(x)$ is defined as the minimal description length l of a string x with respect to a universal processor U within a processing realization p, e.g.,

$$K_U(x)=\min_{p:U(p)=x} l(p), \quad (12)$$

where U is the computable function of two arguments x and p.

Binary strings are the words in the alphabet $\mathcal{A}=\{0,1\}$. For any computable function $U:\mathcal{A}\to\mathcal{A}$. The complexity of $x\in\mathcal{A}$ can be defined with respect to U. For any processor $\mathcal{P}$, one has

$$K_U(x)\ge cK_{\mathcal{P}}(x),\ \forall x,\ c>0, \quad (13)$$

where c is the constant which depends on U and $\mathcal{P}$.

Recalling that $I(X;Y)=H(Y)-H(Y|X)$, we define the mutual complexity as

$$I_K(X;Y)=K(Y)-H(Y|X,K_U(X)). \quad (14)$$

The data processing complexity measures are

$$L=H(X)K_U(x)I_K(X;Y),\ \mathcal{L}=\mathcal{H}K_U(x)\mathcal{I}_K(\mathcal{X};\mathcal{Y}). \quad (15)$$

4. 3. Data Quality

A Markov information source is a pair (M,f) of stationary Markov chain M and function f of reachable states s_k, $f(s_k):S\to A$. The mapping $f(s_k)$ maps states S into the Markov chain to entities in the alphabet $\mathcal{A}$. To estimate the data quality of information sources Q and $\mathcal{Q}$, we use the positively-definite mutual information. The sequence of length n has a complexity $\sim O(n)$. The probability of an input p is $\sim 2^{-l(p)}$. The universal probability of a binary string x is

$$P_U(x)=\sum_{p:U(p)=x} 2^{-l(p)}=\Pr(U(p)=x),\ P_U(x)\approx 2^{-K(x)}. \quad (16)$$

We define the data quality measures as

$$Q=I(X;Y)P_U{}^{-1}(x),\ \mathcal{Q}=\mathcal{I}(\mathcal{X};\mathcal{Y})\mathcal{P}_U{}^{-1}(\mathcal{X}). \quad (17)$$

Here, for two discrete random variables X and Y

$$I(X;Y)=\sum_{y\in Y}\sum_{x\in X} p(x,y)\ln\frac{p(x,y)}{p(x)p(y)}, \quad (18)$$

where $p(x,y)$ is the joint probability distribution function of X and Y; $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y.

Illustrative Example 4. 1.

Let X and Y are independent. Then, $p(x,y)=p(x)p(y)$. Hence, $\ln\frac{p(x,y)}{p(x)p(y)}=\ln 1=0$, and $I(X;Y)=0$. ■

The proposed quantitative metrics yield consistent estimates and measures to evaluate robustness, effectiveness and capabilities of processing on extra-large-scale data as illustrated in Examples.

Example 4. 1.

Consider a two-state *microscopic* system with two distinct unperturbed states characterized by ψ_a and ψ_b, $\langle \psi_a | \psi_b \rangle = \delta_{ab}$ [9-11]. The time-independent Schrödinger equations are

$\hat{H}_a \psi_a = E_a \psi_a$ and $\hat{H}_b \psi_b = E_b \psi_b$

with Hamiltonian operators $\hat{H}_a$ and $\hat{H}_b$. The wave function is

$$\Psi(t) = c_a \psi_a e^{-i\frac{E_a}{\hbar}t} + c_b \psi_b e^{-i\frac{E_b}{\hbar}t}, \ |c_a|^2 + |c_b|^2 = 1.$$

We use the perturbation theory and apply the simplifying assumptions [9-11]. The evolutions of $c_a(t)$ and $c_b(t)$ are

$$\frac{dc_a}{dt} = -i\frac{1}{\hbar} H_{Eab} e^{-i\frac{E_b - E_a}{\hbar}t} c_b(t), \ \frac{dc_b}{dt} = -i\frac{1}{\hbar} H_{Eba} e^{i\frac{E_b - E_a}{\hbar}t} c_a(t).$$

The quantum transition probability $p_{a\to b}(t)$ defines the probability that the *microscopic* system, which started in ψ_a, will be found in ψ_b at time t. The probability is $p_{a\to b}(t) = |c_b(t)|^2$.

Then, $$p_{a\to b}(t,\omega) = |\boldsymbol{\mathcal{E}}_{ab}|^2 \left| \frac{1}{2\hbar} \left[\frac{e^{j(\omega_0+\omega)t} - 1}{\omega_0 + \omega} + \frac{e^{j(\omega_0-\omega)t} - 1}{\omega_0 - \omega} \right] \right|^2.$$

Consider a sinusoidal excitation. We have

$\hat{H}_E(t,\mathbf{r}) = \boldsymbol{\mathcal{E}}(\mathbf{r}) \cos \omega t$, or, $\hat{H}_{Eab} = \mathcal{E}_{ab} \cos \omega t$, $\Pi_{ab} = \langle \psi_a | \boldsymbol{\mathcal{E}} | \psi_b \rangle$.

The theoretical quantum transition rate is $r = (dp/dt)$.

This r is relevant to the device quantum transition rate $\mathcal{R}$, which can be viewed as an analog to the bandwidth B. One recalls that the channel capacity is $C_{\max} = B\log_2[1 + S/N]$, where S/N is the average signal-to-noise power ratio.

The *natural* transition frequency is $\omega_0 = (E_b - E_a)/\hbar$, $\omega_0 \geq 0$.

Considering quantum communication and processing by a *microscopic* system on optoelectronic transductions, we study two-state quantum-mechanical switching $p_{a\to b}(t,\omega)$ at $f = 1\times10^{15}$ Hz [9-11]. The three-dimensional plot for $p_{a\to b}(t,\omega)$ is illustrated in Fig. 2.

For a serial optical photon communication, the *gross bit rate* r depends on the transmission time T_t, $r = 1/T_t$. For the parallel communication, the *gross bit rate* is $r = \sum_{i=1}^{N} \frac{1}{T_i} \log_2 M_i$, where N is the number of parallel channels; M_i is the number of symbols or modulation levels in the ith channel; T_i is the symbol duration time for the ith channel.

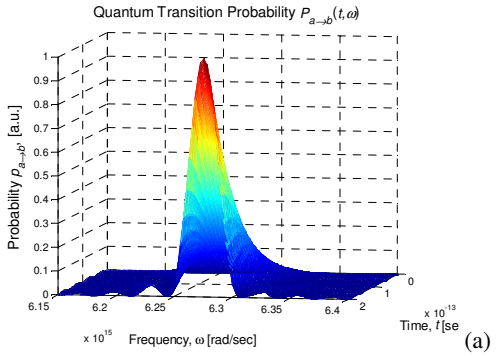
The quantum transductions occur within $\sim 1\times10^{-15}$ sec [9-11]. Let the maximum rate $r_{\max}$ varies from 1×10^{13} to 1×10^{14} pulse/sec, while the average rate r_0 changes from 1×10^{11} to 1×10^{12} pulse/sec. Assume that $r_{\min}$ is 1×10^{11} pulse/sec. The channel capacity $\mathcal{C}(r_0, r_{\max})$ is [9, 10]

$$\mathcal{C} = \frac{1}{\ln 2} \left((r_0 - r_{\min}) \ln \left(\frac{r_{\max}}{r_{\min}} \right)^{\frac{r_{\max}}{r_{\max} - r_{\min}}} - r_0 \ln \left(\frac{r_0}{r_{\min}} \right) \right) \text{ if } r_{\min} \leq r \leq r_{\max}.$$

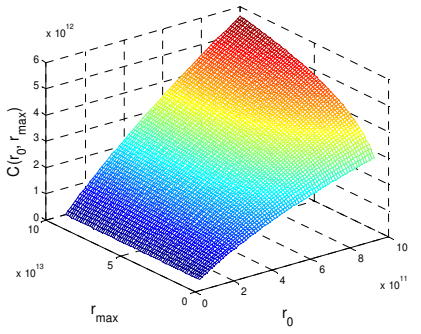
Fig. 2.b reports a three-dimensional plot for $\mathcal{C}(r_0, r_{\max})$. A very high channel capacity is achieved with $\mathcal{C}_{\max} = 5.66\times10^{12}$ bits

The probability of a bit error p_b, depends on the derived quantum transition probability $p_{a\to b}(t)$, The reported analysis also lead to the maximum transmission rate $r_{\max}(p_b)$.

This example reports quantum-mechanically consistent theoretical studies. The results must be substantiated by practical *microscopic* devices. Enormous challenges arise such as interfacing, compliance, measurability, etc.



(a)



(b)

Fig. 2. (a) Quantum transition probability $p_{a\to b}(t,\omega)$;
(b) Three-dimensional plot for channel capacity $\mathcal{C}(r_0, r_{\max})$. ■

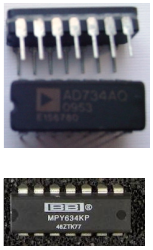
Example 4. 2.

Many data processing schemes require multiplication and division of variables. Different schemes and solutions exist. High-performance low-power four-quadrant multipliers are shown in Fig. 3.a. Consider low-power analog and digital multipliers with the measured inputs (random variables X_1 and X_2, $X_1 \in [-1\ 1]$ and $X_2 \in [-1\ 1]$) and output Y. The experimental results (realizations) of $X = X_1 X_2 \to Y$ are illustrated in Fig. 3.b.

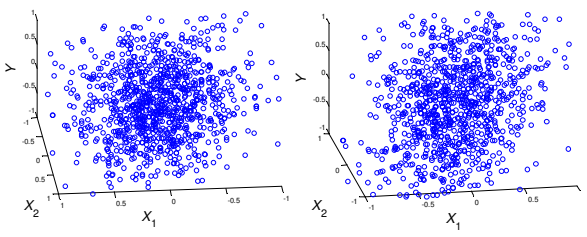
We find the marginal and joint probability density functions. For different functional ICs which realize the processing, one obtains $p_X(x)$, $p_Y(y)$, $p_{X,Y}(x,y)$, $p_{X|Y}(x|y)$ and $p_{Y|X}(y|x)$. The information measures, which depend on $I(X;Y)$

$$I(X;Y) = \int_Y \int_X p_{X,Y}(x,y) \log_2 \frac{p_{X,Y}(x,y)}{p_X(x) p_Y(y)} dxdy = \int_Y \int_X W(x,y) dxdy,$$

are proposed to assess the functionality and quantitatively evaluate performance and capabilities. The three-dimensional plots for the integrand $W(x,y)$, which defines of the information received, are reported in Fig. 3.c and 3.d. Functional multipliers lead to processing with the maximum information content. For faulty ICs, the information received is close to zero on average as documented in Fig. 3.d. We verified and substantiated the proposed concept.



(a)



(b)

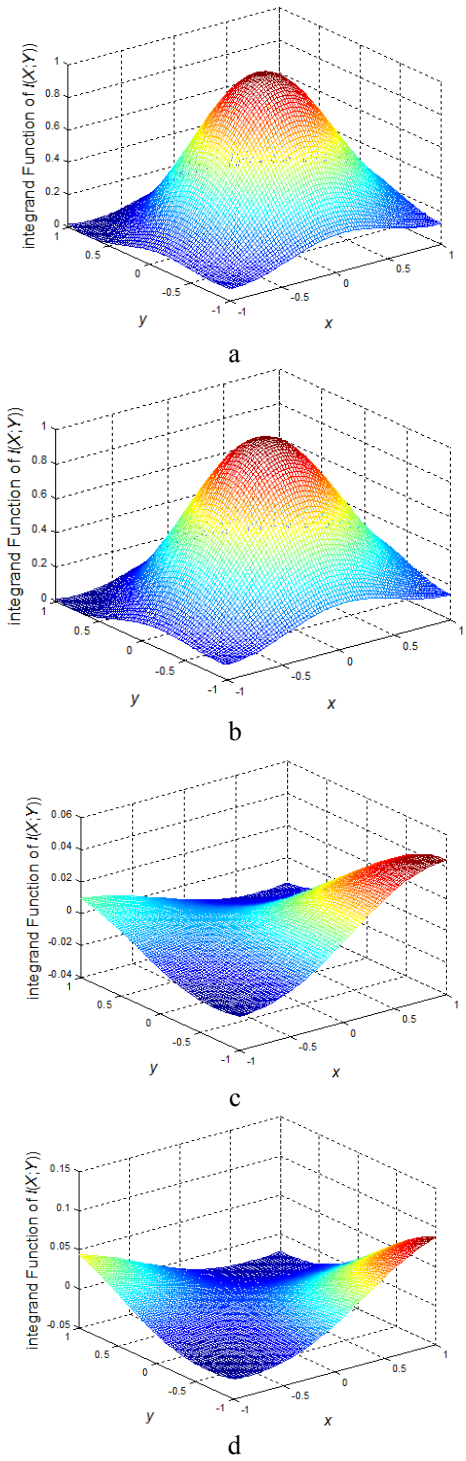


Fig. 3. (a) Processing on two random variables: AD734AQ and MPY634KP multipliers; (b) Experimental results: Multiplication X=X1X2 with the output Y; (c) The information content measure W(x,y) for functional multipliers; (d) The information content measure W(x,y) for faulty multipliers.

V. Conclusions

This paper advanced theoretical fundamentals of classical and quantum information technologies, computer science and computer engineering. We defined, analyzed and evaluated classical and quantum-centric information measures pertained to optoelectronic transductions as well as photon emission, propagation and absorption. Further developments of information theory were ensured by consistently examining information fusion, communication and processing. We proposed coherent estimates and measures to assess classical, quantum↔classical and quantum communication and processing. This paper contributed to a fundamental knowledge base in analysis and design of *microscopic–macroscopic* optoelectronic and photonic systems ensuring a quantum↔classical consistency. High-fidelity modeling, heterogeneous simulations and experimental studies were performed for lasers, waveguides, detectors and processing hardware. Our findings may enable revolutionary sensing, communication and processing capabilities beyond classical solutions and schemes. This paper contributed to devising, design and analysis of multi-physics quantum↔classical communication and processing, thereby enabling cognizant areas across disciplines. The proposed solutions enabled various existing and future communication, sensing, information fusion, computing and processing platforms. The reported results fostered technical readiness, technological capabilities and commercialization capacities by developing relevant tools and practical methods.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. Journal*, vol. 27, pp. 379-423, pp. 623-656, 1948.

[2] . N. Kolmogorov, "Logical basis for information theory and probability theory," *IEEE Trans. Information Theory*, vol. IT-14, no. 5, pp. 662-664, 1968.

[3] A. N. Kolmogorov, "Three approaches for defining the concept of information quantity," *Information Transmission*, vol. 1, pp. 3-11, 1965.

[4] *International Technology Roadmap for Semiconductors*, 2005, 2007, 2009 and 2011 Editions, Semiconductor Industry Association, Austin, Texas, USA, 2013.

[5] P. W. Coteus, J. U. Knickerbocker, C. H. Lam and Y. A. Vlasov, "Technologies for exasscale systems," *IEEE Communication Magazine*, pp. S67-S72, 2012.

[6] Y. A. Vlasov, "Silicon CMOS-integrated nano-photonics for computers and data communications beyond 100G," *IBM J. Res. Dev.*, vol. 55, no. 5, pp. 14:1-14:12, 2011.

[7] "The 50G silicon photonics pink: The Word's first silicon-based optical data connection with integrated lasers", *Intel Labs White Paper*, 2010.

[8] *National Nanotechnology Initiative Strategic Plan*, Executive Office of the President of the United States, Washington, DC, 2007.

[9] S. E. Lyshevski, *Molecular Electronics, Circuits, and Processing Platform*, CRC Press, Boca Raton, FL, 2007.

[10] S. E. Lyshevski, *Molecular and Biomolecular Processing: Solutions, Directions and Prospects*, *Handbook of Nanoscience, Engineering and Technology*, Ed. W. Goddard, D. Brenner, S. E. Lyshevski and G. Iafrate, CRC Press, Boca Raton, FL, pp. 125-177, 2012.

[11] S. E. Lyshevski, "High-performance computing and quantum processing," *Proc. IEEE Conf. High-Performance Computing*, Kiev, Ukraine, pp. 33-40, 2012.

[12] A. G. Gurwitsch, *Theory of the Biological Field*, Nauka, Moscow, 1944.

[13] J. J. Chang, J. Fisch and F. A. Popp, *Biophotons*, Kluwer Academic Publishers, Dordrecht-Boston-London, 1998.

[14] R. P. Feynman, *Lectures on Computation*, Addison-Wesley, Reading, MA, 1986.

[15] J. von Neumann, *The Computer and the Brain*, Princeton Press, 1957.

Nano and Molecular Technologies in Microelectronics, MEMS and Electronic Systems

Sergey Edward Lyshevski
Department of Electrical and Microelectronic Engineering
Rochester Institute of Technology, Rochester, NY 14623, USA
Sergey.Lyshevski@mail.rit.edu http://people.rit.edu/seleee

***Abstract* – For electronic, MEMS and electronic systems, we examine emerged technologies, research novel solutions and apply new transformative findings. To enable communication and processing schemes, nano and molecular technologies are applied. Transformative findings and developments are substantiated by analyzing and evaluating multi-physics *microscopic* systems. We advance physical foundations, engineering premises and information technology of classical, quantum and *mixed* sensing, communication and processing on quantum transductions. This paper examines fundamentals of quantum data processing on measurable and processable real-valued physical quantities. The developed paradigm: (1) Unifies and enables concepts of theoretical computer science, computer engineering and quantum mechanics; (2) Consistent with the *first principles* of quantum informatics, communication and processing; (3) Coherently examines device physics, switching algebra, processing arithmetics and calculus; (4) Enables practical *microscopic* hardware solutions for electronics, MEMS and electronic systems.**

Keywords – electronics; MEMS; microelectronics; nanotechnology; processing

I. Introduction

Analog, *hybrid* and digital ICs are used in different MEMS and electronic systems. Enabling designs, organizations, architectures and fabrication technologies are used. Fundamental, applied and technological developments in microelectronics led to mass production of high-performance MEMS, SoC and SiP with billions of transistors on a chip. Enabling materials, processes and tools led to a current lithography-defined 28-nm "DRAM Metal 1 Half-Pitch" [1]. Various fundamental and technology limits emerge within a foreseen scaling below ~20-nm [1]. The technology evaluation criteria are yield, scalability, energy efficiency, on/off current, operational reliability, operational temperature, architecture compatibility and technology compatibility. As the planar solid-state devices are scaled to hundreds of nanometer in dimensions, the undesirable quantum phenomena significantly degrade the device and ICs performance and functionality. The so-called "single electron transistor","single photon transistor", "quantum dots" and other devices are *macroscopic* microelectronic devices. The focused research activities are centered on photonic, optoelectronic and other quantum-effect devices. A significant progress is made in resonant-tunneling devices, solid-state, inorganic and organic lasers, etc. [1].

To fully utilize quantum phenomena and enable new features, engineering and technology must progress beyond current *macroscopic* microelectronics. We focus on quantum-centric solutions and nanotechnology. These imply new device physics of subatomic/atomic/molecular *microscopic* devices, novel processing principles, new interfacing, enabling networking schemes, innovative synthesis and fabrication, etc. There are enormous challenges and complexities which range from quantum-consistent design to synthesis, integration, interfacing, testing and characterization. Solutions of the aforementioned problem will enable processing with unprecedented performance and capabilities thereby advancing MEMS and electronic systems. The compliant quantum↔ classical processing and overall compatibility must be ensured.

II. Nanotechnology and Quantum Processing

In some applications, high-performance computing is expected to ensure sustainable performance reaching 1×10^{15} (petaflops) floating point operations per second. Vertebrates and invertebrates exhibit information and data processing. The data processing in living organisms greatly exceeds *exascale* performance significantly surpassing the *exaflops* range, e.g., one quintillion 1×10^{18} floating point operations per second. This performance and capabilities cannot be ensured by any envisioned super-computers due to fundamental and technological limits. The information processing in living organisms has not been comprehended. However, the exhibited performance can be accomplished assuming quantum-centric schemes which guarantee extraordinary information processing by biomolecules and molecular assemblies. It is very important to understand mechanisms and baseline phenomena of sensing, communication and processing by *microscopic* systems.

Optoelectronic and molecular nanotechnologies may enable one to design and synthesize functional *microscopic* systems which exhibit and utilize quantum phenomena, accomplishing quantum and compliant quantum↔classical processing. Our goals are to:

1. Study quantum data processing by *microscopic* systems;
2. Develop and substantiate quantum-deterministic and quantum-probabilistic communication and processing on optoelectronic transductions due to photon-electron interactions in classical and quantum domains;
3. Advance information theory and computing arithmetics;
4. Contribute to design of next generation of *mixed* compliant quantum↔classical photonic and optoelectronic communication and processing platforms which will maximize information content, capacities, information fusion capabilities and processing performance.

Recent advances in optoelectronic communication and processing, with applications in microelectronics, MEMS and electronic systems, were outlined in [1-4]. Further advances

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

are needed. Practical solutions are researched attempting to approach the performance exhibited by living organisms where biomolecules and macromolecules accomplish quantum sensing, information fusion and processing. Biomolecular *fabrics* operate by utilizing quantum phenomena, e.g., quantum processing on photons due emission, propagation and absorption of photons in near-infrared (heat sensors), visible (visual systems) and near-ultraviolet spectra [3-5].

Engineered photonic and optoelectronic systems are emerged, e.g., lasers, resonant-tunneling transistors, photon detectors, optical waveguides and others [1-4, 6]. The aforementioned assured evidences support and justify further developments. Technology advancements and recent findings foster practical quantum and *mixed* quantum↔classical processing by *microscopic* and *macroscopic* systems. Recalling the emphasized grand challenges [1], we intent to:

1. Research and substantiate multi-physics sensing, communication and processing by quantum-effect devices: Apply quantum mechanics, stochastic quantum thermodynamics and electrodynamics to examine complex inter-atomic, intra-atomic and molecular transitions due to strong electron and phonon–electron interactions;
2. Advance computer and information sciences: Derive quantitative information measures, quantum processing calculus and arithmetics;
3. Enable communication and processing schemes: Characterize and evaluate quantum-centric optoelectronic schemes which will ensure quantum and quantum↔ classical communication, data fusion and processing.

Publications [3-6] examine: (i) Photons emission and absorption, as well as electronic transductions in molecular and *microscopic* device; (ii) Photon propagation in organic and silicon optical waveguides; (iii) Information fusion and processing by photons and electrons; (iv) Photons detection and optoelectronic interface which may lead to computer–brain interfacing. An exploratory roadmap is reported in Fig. 1.

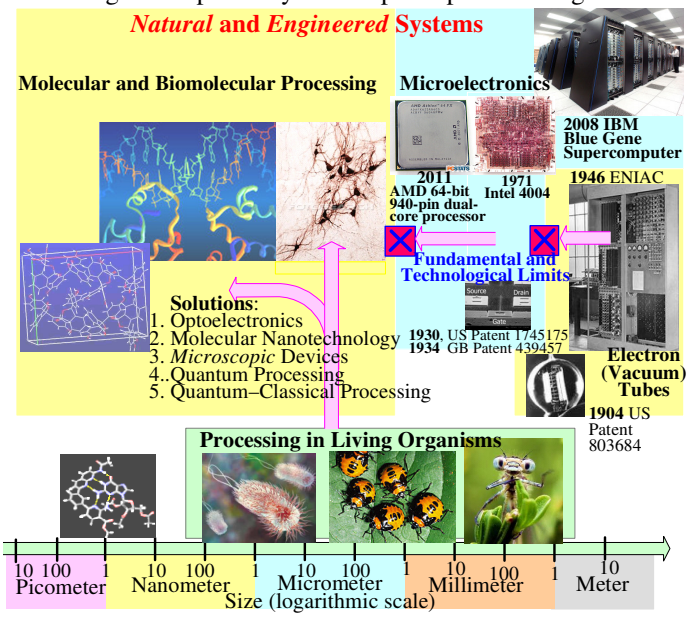


Fig. 1. Towards optoelectronic and molecular nanotechnologies [3, 4]

III. Quantum Mechanics and Informatics

Various concepts of information theory and computer engineering are applied to examine classical and quantum communication and processing [3, 4, 6, 7]. Conventional and quantum quantities, information measures and processing schemes are different. This section focuses on:

1. Analysis of electron- and photon-induced phenomena which lead to quantum *state* transitions and *utilizable* transductions on *detectable*, real-valued and *measurable* physical *variables* (*observables*). These *measurable variables* must be *controllable*, *algorithmically processable* and *hardware realizable*;
2. Quantum- and device-physics consistent fundamentals of communication and processing by *microscopic* systems;
3. Developments and substantiation of practical engineering paradigms and technologies of electronic and optoelectronic sensing, communication and processing.

It is important to progress from basic foundations to theory substantiation, practical engineering solutions and technologies by:

- Studying and applying *utilizable* quantum transductions on *detectable*, real-valued and *measurable* physical *variables*;
- Verifying coherent principles and mechanisms of energy-centric sensing and processing in *microscopic* systems as applied to practical solutions, schemes and technologies.

Under some hypotheses and conjectures, *microscopic* systems can be mathematically modeled and described by using wave functions in spatial, momentum and other spaces.

The spatiotemporal wave function $\Psi(\mathbf{r},t)=\psi(\mathbf{r})\varphi(t)$ is found by solving the time-dependent Schrödinger equation

$$\hat{H}\Psi(\mathbf{r},t)=i\hbar\frac{\partial\Psi(\mathbf{r},t)}{\partial t},\ \hat{H}\in\mathbb{C},\ \Psi\in\mathbb{C},\ c_n\in\mathbb{C}, \tag{1}$$

$$\Psi(\mathbf{r},t)=\sum_{n=1}^{\infty}c_n\Psi_n(\mathbf{r},t)=\psi(\mathbf{r})\varphi(t)=\sum_{n=1}^{\infty}c_n\psi_n(\mathbf{r})e^{-i\frac{E_n}{\hbar}t}=\sum_{n=1}^{\infty}c_n\psi_n(\mathbf{r})e^{-i\omega_n t},$$

where $\hat{H}$ is the *total* Hamiltonian operator, $\hat{H}=\hat{H}_0+\hat{H}_E+\hat{H}_P$; $\hat{H}_0$ is the unperturbed Hamiltonian in the absence of external excitations, and, for an unperturbed *microscopic* system $\hat{H}_0\psi_n=E_n\psi_n$, $\hat{H}_0=-\frac{1}{2m}\hbar^2\nabla^2+\Pi$; $\hat{H}_E$ and $\hat{H}_P$ are the excitation and perturbation Hamiltonians; Π is the potential function; c_n are the complex probability amplitudes, $\sum_{n=1}^{\infty}|c_n|^2=1$.

Using Ψ, one finds $|c_n|^2$ which gives the probability that a *microscopic* system at any given time t is in a state with E_n, or, the probability that a measurement of E at t would yield E_n.

We have

$$\Psi(\mathbf{r},t)=\psi(\mathbf{r})\varphi(t),\ \hat{H}\psi_n=E_n\psi_n,\ i\hbar\frac{1}{\varphi}\frac{\partial\varphi}{\partial t}=E,\ \varphi(t)=e^{-i\frac{E}{\hbar}t}. \tag{2}$$

Wave functions Ψ, derived using various spaces, may yield mathematically-consistent estimates on various quantities, such as probabilities, allowed *states*, E_n, expectation values, etc. The expectation value of a quantum canonical variable $C\in\mathbb{R}$, with an associated operator $\hat{C}\in\mathbb{C}$, is

$$\langle C\rangle=\int\Psi^*(\mathbf{r},t)\hat{C}\Psi(\mathbf{r},t)dV,\ C\in\mathbb{R},\ \hat{C}\in\mathbb{C}. \tag{3}$$

The governing equation for the operator $\hat{C}$ is

$$\frac{d}{dt}\langle\hat{C}\rangle=\left\langle\frac{\partial\hat{C}}{\partial t}\right\rangle+\frac{1}{i\hbar}\langle[\hat{C},\hat{H}]\rangle,\ [\hat{C},\hat{H}]=\hat{C}\hat{H}-\hat{H}\hat{C},\ \hat{H}\in\mathbb{C},\ \hat{C}\in\mathbb{C}. \tag{4}$$

Quantum-mechanically-consistent modeling and analysis result in a set of equations (1)-(4) with additional equations [6].

Example 3. 1.

A *microscopic* particle in an infinite potential well can be quantum-mechanically described by the wave function [8]

$\Psi(x,t)=\psi(x)\varphi(t)$, $\psi(x)=Ax(a-x)$, $0\le x\le a$, $A\neq 0$, $\varphi(t)=e^{-iEt/\hbar}$, and $\Psi=0$ if $x<0$ and $x>a$.

The normalization $\int \psi^*(x)\psi(x)dx=1$ gives $A=(30/a^5)^{1/2}$.

Fig. 2.a documents $\Psi(x,0)$.

From $c_n=\int_{-\infty}^{\infty}\psi_n^*(x)\Psi(x,0)dx$, one finds

$$c_n=\int_0^a\sqrt{\frac{2}{a}}\sin\left(\frac{\pi n}{a}x\right)\sqrt{\frac{30}{a^5}}x(a-x)dx=\frac{4\sqrt{15}}{\pi^3 n^3}\left[1-\cos(\pi n)\right]=\begin{cases}0, & n \text{ is even}\\ \dfrac{8\sqrt{15}}{\pi^3 n^3}, & n \text{ is odd}\end{cases}$$

We have $\Psi(x,t)=\sqrt{\frac{30}{a^5}}\frac{8}{\pi^3}\sum_{n=1,3,5}^{\infty}\frac{1}{n^3}\sin\left(\frac{\pi n}{a}x\right)e^{-i\frac{\pi^2\hbar n^2}{2ma^2}t}$.

There is the admixture of the exited states characterized by

$\Psi(x,t)=\sum_{n=1}^{\infty}c_n\psi_n(x)e^{-i\frac{E_n}{\hbar}t}$. From $\sum_{n=1}^{\infty}|c_n|^2=\left(\frac{8\sqrt{15}}{\pi^3}\right)^2\sum_{n=1,3,5,\dots}^{\infty}\frac{1}{n^6}=1$,

the probability of being in all exited states with $\forall n>1$ is 0.001445. In $n=1$, the system is mathematically described by

$$\Psi(x,t)_{n=1}=\sqrt{\frac{30}{a^5}}\frac{8}{\pi^3}\sin\left(\frac{\pi}{a}x\right)e^{-i\frac{\pi^2\hbar}{2ma^2}t}=\sqrt{\frac{30}{a^5}}\frac{8}{\pi^3}\sin\left(\frac{\pi}{a}x\right)e^{-i\omega_1 t}.$$

The probability that a system is in the *ground* state $n=1$ with $E_1=\frac{\pi^2\hbar^2}{2ma^2}$ is $|c_1|^2=\left(8\sqrt{15}/\pi^3\right)^2=0.998555$.

The expectation value of the energy is

$$\langle H\rangle=\sum_{n=1}^{\infty}|c_n|^2E_n=\sum_{n=1}^{\infty}|c_n|^2\frac{\pi^2\hbar^2}{2ma^2}n^2=\frac{480\hbar^2}{\pi^4ma^2}\sum_{n=1,3,5,\dots}^{\infty}\frac{1}{n^4}=\frac{5\hbar^2}{ma^2}.$$

Other important quantities can be derived.

The complex wave function $\Psi(x,t)$ is a mathematic operator. This $\Psi(x,t)\in\mathbb{C}$ is not *detectable*, not *measurable* and unlikely may be used to realize any processing tasks in real systems. The three-dimensional plots for the real and imaginary parts of $\Psi(x,t)$ are illustrated in Fig. 2.b letting $m=9.11\times10^{-31}$ kg, $a=0.154\times10^{-9}$ m, $n=1$ ($E_1=2.5424\times10^{-18}$ J).

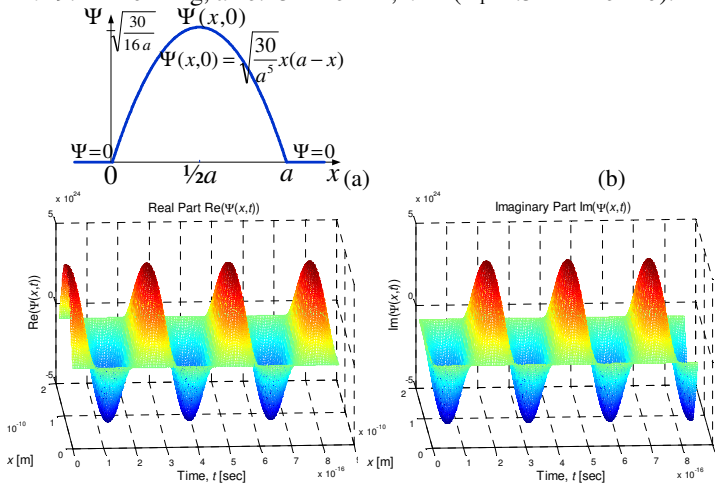


Fig. 2. (a) Wave function $\Psi(x,0)=\sqrt{\frac{30}{a^5}}x(a-x)$, $0\le x\le a$;

(b) Evolutions of real Re[$\Psi(x,t)$] and imaginary Im[$\Psi(x,t)$] parts of the wave function $\Psi(x,t)_{n=1}$. ■

Example 3. 2.

We examine photon absorption by a retinal molecule. Rhodopsin is a membrane-intrinsic protein. A photon is absorbed by the renal molecule $C_{20}H_{28}O$ which is bound to opsin. Retinal molecule is illustrated in Fig. 3.a. Paper [8] examines quantum interactions, photon absorption and transitions in a retinal molecule which is a *microscopic* target. In the spherical coordinates $\Psi(t,\mathbf{r})=\Psi(t,r)\,\Psi(t,\theta,\phi)$.

Consider quantum-probabilistic photon absorption for a realistic wavelength $\lambda\in[300\ 800]$ nm and pointing angle $\Omega_{\mathbf{k}}\in[-\frac{1}{2}\pi\ \frac{1}{2}\pi]$ rad. The maximum probability of photon absorption is observed at $\lambda=555$ nm when the photon energy is 3.6×10^{-19} J. The *microscopic* sensors and detectors ensure high selectivity for activation energy.

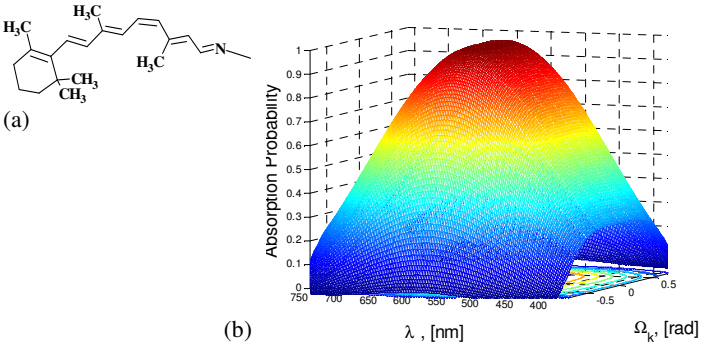


Fig. 3. (a) Retinal molecule $C_{20}H_{28}O$;
(b) Probability of absorption, $\lambda\in[400\ 800]$ nm, $\Omega_{\mathbf{k}}\in[-0.75\ \ 0.75]$ rad

The resulting model *mapping* of *microscopic* systems is

$$\mathcal{M}(\hat{k}(\Psi),\hat{C},C)\in\mathcal{K}\times\hat{\mathbf{C}}\times\mathbf{C},\ \hat{C}\in\mathbb{C},\ C\in\mathbb{R}. \qquad (5)$$

In (5), $\hat{k}(\Psi)$ gives the quantum-mechanically-consistent descriptive quantitative estimates on physical and auxiliary variables, mathematical quantities and operators. Using quantum mechanics with simplifying assumptions, the spatiotemporal evolutions of mathematical operators are modeled. One may describe the evolutions of *detectable*, real-valued and *measurable* physical *variables*. Modeling and analyses premises must be substantiated by examining measured evolutions of physical *variables* in *microscopic fabrics*. The biomolecular *fabrics*, as well as organic and inorganic devices, are illustrated in Fig. 4.

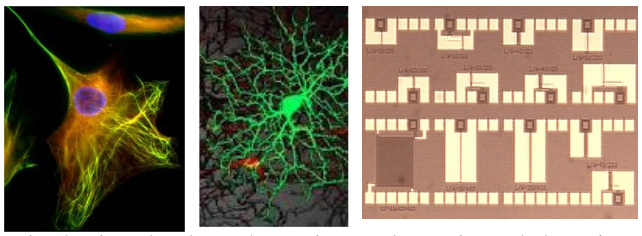


Fig. 4. Biomolecular and organic optoelectronics and photonic

The use of various mathematical operations, manipulations on operators, *superposition* and other premises of all-algorithmic quantum communication and computing were outlined in [9-14]. We depart from abstract quantum computing which assumes applicability and practicality of:

1. *Naive* all-algorithmic computing on not detectable, not observable and not measurable mathematical operators;
2. Algorithmic schemes of "quantum logic gates" networked within postulated computing structures by elusive quantum "interconnects", "channels" and "circuits"; etc.

Communication and processing must be accomplished by means of practical solutions, verifiable schemes and hardware. It is unlikely that any processing tasks can be established by abstractly using *quantum indeterminacy*, *incompleteness*, mathematical operators (wave functions, *kets*, *eigenkets*, probability density and others), *hidden variables*, "superposition" of *states*, etc.

We propose a solution which centers on three principles.

***Principle* 1.** *Engineering Science: Quantum Mechanics and Information Science* – Guarantee coherent quantum and informatics fundamentals as applied to communication, information fusion and data processing, thereby ensuring overall system functionality.

***Principle* 2.** *Engineering Design* – Guarantee processing on real-valued physical *variables* using *distinguishable*, *quantifiable* and *computable* transforms with corresponding data processing arithmetics (logics and calculi), organizations, architectures and schemes. Utilize quantum phenomena which lead to state transitions and utilizable transductions on measurable physical *variables* in molecular *fabrics*. These *variables* should be quantum-mechanically achievable, algorithmically processable and hardware realizable, thereby accomplishing processing. A compliant quantum↔classical interfacing, processing and compatibility must be ensured.

***Principle* 3.** *Physical Implementation and Validation* – Physical design, synthesis and implementation of molecular *fabrics*. Integration and characterization at device and system levels.

Example 3. 3. *Control of Electrons in a Processing Device* [4]

We study the propagation and control of *j*th *delocalized* electrons in an organic molecule which is illustrated in Fig. 5. The probability of finding the electron in a given volume is

$$P=\int_{r=0}^{\infty}\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}\Psi^{*}_{nlm_l}(r,\theta,\phi)\Psi_{nlm_l}(r,\theta,\phi)r^2dr\sin\theta d\theta d\phi. \quad (6)$$

Using equations (1)-(4), high-fidelity modeling and data-intensive analysis are performed, thereby obtaining a *mapping* $\mathcal{M}(\hat{k}(\Psi),\hat{C},C)$. The surfaces in the *arbitrary units* for $|\psi_{21\pm1}(r,\theta,\phi)|^2$ and $|\psi_{32\pm2}(r,\theta,\phi)|^2$ are depicted in Fig. 5. Applying the controlled field $u_{\mathcal{E}}(t,\mathbf{r})$ [3, 4], a quantum system is *governed* ensuring the controlled motion of the *delocalized* electrons. The valence electrons are most probable at $r_{\max}$ which corresponds to the 3*p* subshell of P atoms with the radial separation $\Delta r=r_{\max}-r_{\min}$, $r_0=½(r_{\max}-r_{\min})$. Fig. 5 reports a three-dimensional surface for $P_{nl}(r,u_{\mathcal{E}})$ if $u_{\mathcal{E}0}\in[u_{\mathcal{E}0\min}\ u_{\mathcal{E}0\max}]$ for $r\in[r_0\ r_{\max}]$.The controllable quantum transductions are ensured.

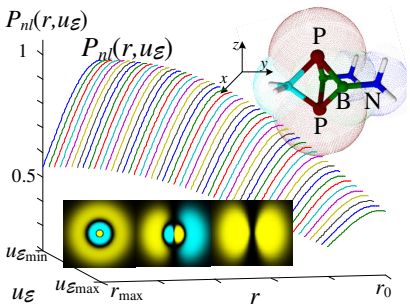


Fig. 5. Plot of $P_{nl}(r,u_{\mathcal{E}})$ for the valence electrons in the 3*p* subshell of P atoms in an organic molecule with B (green), P (maroon), N (blue) and C (cyan) atoms. ■

IV. Quantum Processing

Using three principles, we research quantum-mechanically consistent, algorithmically cohesive and hardware coherent quantum data processing. The *detectable*, real-valued and *measurable* physical *variables* $\boldsymbol{\nu}\in\mathcal{V}$ may lead to quantum-mechanically-, device- and algorithmically- (arithmetically) consistent processing if these *variables* $\boldsymbol{\nu}$ are:

1. Controlled and evolved during controlled quantum transductions;
2. *Algorithmically processable.*

For example, one may control measurable $\boldsymbol{\nu}=[E,\omega,\lambda,\ldots]^T$, and, perform processing *utilizable* initial (*I*), intermediate (*T*) and final (*F*) *state* transductions $\mathcal{S}=[\mathcal{S}_I,\mathcal{S}_T,\mathcal{S}_F]^T$ on $\boldsymbol{\nu}=[\nu_1,\ldots,\nu_k]^T$. The *microscopic* devices may accomplish the following irreversible and reversible *utilizable* transductions

$$\mathcal{S}_I:\boldsymbol{\nu}_I\Rightarrow\mathcal{S}_T:\boldsymbol{\nu}_T\Rightarrow\mathcal{S}_F:\boldsymbol{\nu}_F \text{ and } \mathcal{S}_I:\boldsymbol{\nu}_I\Leftrightarrow\mathcal{S}_T:\boldsymbol{\nu}_T\Leftrightarrow\mathcal{S}_F:\boldsymbol{\nu}_F. \quad (7)$$

Consider a physical *microscopic* processing *fabrics* with processing primitives $\mathcal{P}_1,\ldots,\mathcal{P}_k$. Each $\mathcal{P}_j$ exhibits transductions $\mathcal{S}_j(\boldsymbol{\nu})$ on *detectable*, *measurable* and *processable variables* $\boldsymbol{\nu}_j$ yielding *distinguishable* and *computable* transforms $\mathcal{T}_j(\mathcal{S},\boldsymbol{\nu})$. Using $\mathcal{T}_j(\mathcal{S},\boldsymbol{\nu})$, consistent with device physics and *admissible* arithmetic operand $\mathcal{A}_j$, one has

$$\mathcal{T}_\Sigma=\mathcal{T}_1\circ\ldots\circ\mathcal{T}_k. \quad (8)$$

The processing can be accomplished by using the infinite- and finite-valued logics [4]. Analog, digital and *hybrid* processing schemes may be supported by *microscopic* devices. Considering multiple-valued logic, the switching function on *r*-valued ν_j is $f:\{0,\ldots,r-1\}^n\to^{\mathcal{F}}\{0,\ldots,r-1\}^m$ with a truth vector $\boldsymbol{\mathcal{F}}$.

Any *f* can be represented as

$$f=\mathcal{A}(\boldsymbol{\mathcal{F}},\boldsymbol{\mathcal{T}}). \quad (9)$$

The evolutions of quantum transductions are mapped as

$$\boldsymbol{\nu}_{j,l}\to^{\mathfrak{Q}}\boldsymbol{\nu}_{j,l+1}=\mathcal{Q}_j(\boldsymbol{\nu}_{j,l}) \quad (10)$$

which defines the evolution of physically-realizable *computable* transforms $\mathcal{T}_j(\mathcal{S},\boldsymbol{\nu})$. The transductions $\mathcal{S}_j(\boldsymbol{\nu})$ on $\boldsymbol{\nu}$ in physical *microscopic* systems can be controlled by using device-specific control schemes by varying systems energy, potential or other quantities denoted as $\mathcal{H}$ [3, 4]. The *controllable* evolutions on *measurable* $\boldsymbol{\nu}_j$ are mapped as

$$\boldsymbol{\nu}_{j,l}\to^{\mathfrak{Q}}\boldsymbol{\nu}_{j,l+1}=\mathcal{Q}_j(\boldsymbol{\nu}_{j,l},\mathcal{H}_{j,l}), \quad (11)$$

where $\mathcal{Q}_j$ denotes transductions S_j on $\boldsymbol{\nu}_j$ consistent with $\mathcal{A}(\boldsymbol{\mathcal{F}},\boldsymbol{\mathcal{T}})$.

The binary switching function is $f:\{0,1\}^n\to^{\mathbf{F}}\{0,1\}^m$.

An *n*-variable *r*-valued function *f*, with r^{r^n} different

combinations, is defined as a mapping of a finite set $\{0,\ldots,r-1\}^n$ into a finite set $\{0,\ldots,r-1\}^m$. In particular,

$$f:\{0,\ldots,r-1\}^n \rightarrow^{\mathbf{F}} \{0,\ldots,r-1\}^m. \quad (12)$$

Truth vectors on n binary and r-valued variables $[x_1,\ldots,x_n]$ are

$$\mathcal{F}=[f(0),f(1),\ldots,f(2^n-1)]^T \text{ and } \mathcal{F}=[f(0),f(1),\ldots,f(r^n-1)]^T. \quad (13)$$

Digital computing and digital design use algebraic maps, Boolean algebra, digital (binary and multiple-valued) logics, decision diagrams, data structures, fundamental expansions, polynomial expressions, sequential networks, probabilistic concepts, stochastic schemes, and other approaches [15]. If quantum consistency, algorithmic cohesiveness and hardware coherency are satisfied, some aforementioned concepts may be applicable. Any arithmetically- and algorithmically-defined computable function must be: (1) *Definable*, *realizable* and *implementable* as derived by using the *distinguishable* and *computable* transforms; (2) Implementable using the *utilizable* quantum transductions on *measurable* and *algorithmically processable variables.* Some results on the calculi and arithmetics of quantum processing are reported in [3, 4].

V. Heisenberg Uncertainty Principle and Processing

To characterize processing devices and *fabrics*, as well as to accomplish processing, the physical quantities must be measured. The Heisenberg uncertainty principle provides the position-momentum and energy-time limits on the measurements of *observables*. The following inequalities

$$\sigma_x\sigma_p \geq \tfrac{1}{2}\hbar,\ \sigma_E\sigma_t \geq \tfrac{1}{2}\hbar, \quad (14)$$

must be guaranteed. The expression $\sigma_E\sigma_t \geq \tfrac{1}{2}\hbar$ specifies the uncertainties on the standard deviation on the energy E yielding limits on σ_t. The Heisenberg uncertainty principle does not define the device physics, device sizing, etc. The uncertainty principle defines the fundamental limits on the *measurability*, testing and characterization of *engineered* systems using the specific real-valued physical quantities which must be *observable*, *detectable*, *controllable* and *processable*.

Illustrative Example 5. 1.

The energy of a photon is $E=h\nu=hc/\lambda$. For $\lambda=500$ nm, $E=3.98\times10^{-19}$ J. The de Broglie relation is $\lambda=h/p$. Consider an electron which undergoes a transition from an exited state with E_2 to a ground state with E_1, $(E_2-E_1)>0$. From $\varphi(t)=\sum_{n=1}^{\infty} e^{-i(E_n/\hbar)t}=\sum_{n=1}^{\infty} e^{-i\omega_n t}$, the frequency of evolution is $\omega_{21}=(E_2-E_1)/\hbar$. The aforementioned equations provide relationships between some real-valued physical *variables* and transition estimates relevant to σ_E and σ_t. ■

VI. Conclusions

We reported nanotechnology solutions and examined core problems to enable MEMS and electronic systems. The *microscopic* processing devices must exhibit utilizable quantum-effect state transitions on the physical *variables* which result in computable transforms from viewpoints of device physics, processing arithmetics, calculus and design. Our consistent paradigm: (i) Eases enormous challenges in quantum and *hybrid* sensing, communication and processing; (ii) Overcomes foremost inconsistencies departing from all-algorithmic-centric computing; (iii) Enables new practical inroads, paradigms and solutions; (iv) Advances theory and practice of *natural* and *engineered* processing; (v) Enables unprecedented sensing and processing capabilities ensuring far-reaching benchmarks in MEMS and electronic systems.

Our findings support a broad spectrum of transformative research activities, engineering developments and technological advancements. Some core problems of physical and life sciences are studied and advanced. We enabled a knowledge base and proposed new solutions. Our findings are substantiated by means of basic, applied and numeric studies which are consistent with experiments, biophysics, quantum mechanics, information theory, computer science, computer engineering and emerging technologies.

References

1. *International Technology Roadmap for Semiconductors*, 2005, 2007, 2009 and 2011 Editions, Semiconductor Industry Association, Austin, Texas, USA, 2013.
2. P. W. Coteus, J. U. Knickerbocker, C. H. Lam and Y. A. Vlasov, "Technologies for exasscale systems," *IEEE Communication Magazine*, pp. S67-S72, 2012.
3. S. E. Lyshevski, *Molecular Electronics, Circuits, and Processing Platform*, CRC Press, Boca Raton, FL, 2007.
4. S. E. Lyshevski, *Molecular and Biomolecular Processing: Solutions, Directions and Prospects*, *Handbook of Nanoscience, Engineering and Technology*, Ed. W. Goddard, D. Brenner, S. E. Lyshevski and G. Iafrate, CRC Press, FL, pp. 125-177, 2012.
5. J. J. Chang, J. Fisch and F. A. Popp, *Biophotons*, Kluwer Academic Publishers, Dordrecht-Boston-London, 1998.
6. A. Yariv, *Quantum Electronics*, John Wiley and Sons, New York, 1988.
7. S. E. Lyshevski and L. Reznik, "Information-theoretic estimates of communication and processing by nanoscaled and quantum optoelectronic systems," *Proc. IEEE Conf. Electronics and Nanotechnology*, Kiev, Ukraine, 2013.
8. S. E. Lyshevski, "Quantum molecular sensing, communication and processing by photons," *Proc. IEEE Conf. Nanotechnology*, Birmingham, UK, pp. 476-481, 2012.
9. A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin and H. Weinfurter, "Elementary gates for quantum computation," *Phys.Rev. A*, vol. 52, no. 5. pp. 3457-3467, 1995.
10. C. H. Bennett, I. Devetak, P. W. Shor and J. A. Smolin, "Inequalities and separations among assisted capacities of quantum channels," *Phys. Rev. Let.*, vol. 96, no. 150502, 2006.
11. A. Imamoglu, D. D. Awschalom, G. Burkard, D. P. DiVincenzo, D. Loss, M. Sherwin and A. Small, "Quantum information processing using quantum dot spins and cavity QED," *Phys.Rev.Let.*,vol.83,no.4204, 1999.
12. D. C. Marinescu and G. M. Marinescu, "The Boole Lecture – Quantum information: A glimpse at the strange and intriguing future of information." *The Computer Journal*, vol. 50, pp. 505-521, 2007.
13. M. Mosca, R. Jozsa, A. Steane and A. Ekert, "Quantum-enhanced information processing," *Trans. R. Soc.*, vol. A 358, pp. 261-279, 2000.
14. A. Steane and E. Reiffel, "Beyond bits: The future of quantum information processing," *Computer*, vol. 33, no. 1, pp. 38-45, 2000.
15. S. Yanushkevich, V. Shmerko and S. E. Lyshevski, *Computer Arithmentics for Nanoelectronics*, CRC Press, FL, 2009.

Micromechanical Tuning of Microstrip Antenna in Frequency Domain

N. Ruda, Y. Prokopenko, Y. Poplavko
National Technical University of Ukraine "KPI"
Kiev, Ukraine
nruda219@gmail.com

***Abstract*—Microstrip antenna structure with micromechanical tuning in frequency domain is discussed. Frequency dependence of reflection coefficient for different sizes of air gap between metal patch and dielectric substrate is analyzed. It is shown that change of physical and geometrical parameters of microstrip antenna structure also effects on the antennas field distribution.**

Keywords—microstrip antenna, reflection coeffitient, directivity patter, air gap, micromechanical tuning

I. INTRODUCTION

As the general trend of modern communication systems development is the use of high operating frequencies, the manufacturing technology and design of antennas should also change according to the new requirements. Many antenna applications for satellite links, mobile and wireless LANs constrain the compactness, frequency stability, polarization and directivity pattern control and others.

Microstrip antennas (MSA) in which the line and antenna functions combined for manufacturing compact transceivers are best suited to meet these requirements. Since these antennas have a very low profile, mechanically strong and can be formed in accordance with the surface bending of the vehicle, they are often mounted on the external parts of aircraft and space vehicles, or included in the construction of mobile devices.

In most telecommunication applications it is desirable to have the opportunity to change the frequency characteristics of MSA. The tunable MSA is realized by changing the length of the small stub attached to the regularly shaped MSA, by changing the number of shorting posts used to make a compact configuration or by integrating active devices such as varactor or PIN diodes along with the MSA [1].

Ideally, the dielectric constant of substrate should be low ($\varepsilon < 2.5$) to enhance boundary fields that are important for radiation [2]. Increasing the dielectric permittivity of substrate reduces MSA bandwidth. Efficiency of antenna radiation and its bandwidth are regress as a result of fluctuations in the substrate. To eliminate spurious oscillations inside the substrate and to reduce its dielectric permittivity the method of selective substrate material removal from under antenna patch by etching or drilling holes is proposed in [3]. In [4] MSA construction without dielectric substrate using air fulfillment is analyzed.

Based on experience of microelectromechanical alteration of waveguide phase shifter parameters study [5] it was found that small air gap between dielectric substrate and metal patch causes certain change in measured antenna parameters. This change indicates sensitivity of electromagnetic field distribution on the air gaps size for such antenna construction shown in Fig. 1. Using microelectromechanical method of alteration during the device operation the parameters of MSA can be changed dynamically.

II. DEVICE SIMULATION

In this paper square and circular designs of metal antenna patch are considered. Metal patch is located under dielectric substrate which is laid on the metal ground plane. The size of air gap changes electromechanically using piezoelectric actuator. Antenna feeding is realized through the coaxial cable from the bottom of substrate. Such direct contacting feeding method has the advantage of simplicity, but it also has several disadvantages. It makes fabrication difficult and lowers reliability in an array having thousands of elements and the inherent asymmetry of these feeds generates some higher-order modes which produce cross-polarized radiation [6].

Frequency dependence of reflective coefficient of antenna without air gap was obtained using the finite-difference time-domain (FDTD) solver of the integral-equation based electromagnetic simulator CST Microwave Studio (CST, 2010). Results of FDTD simulation for antenna with square and circle patch is shown in Fig. 2. This figure shows that operating frequency and bandwidth of MSA depend on patch shape. Construction with circular patch shows better radiation characteristics. Radiation pattern of MSA with air gap is shown in Fig. 3.

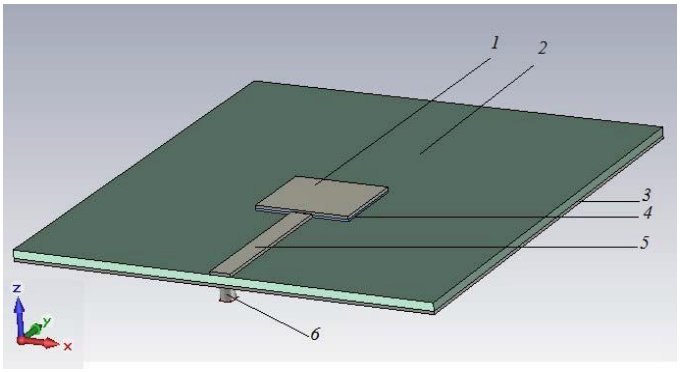


Fig. 1. Construction of microstrip antenna with square patch: 1 – square metal patch; 2 – dielectric substrate; 3 – ground plane; 4 – air gap; 5 – microstripe line; 6 – coaxial feed.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

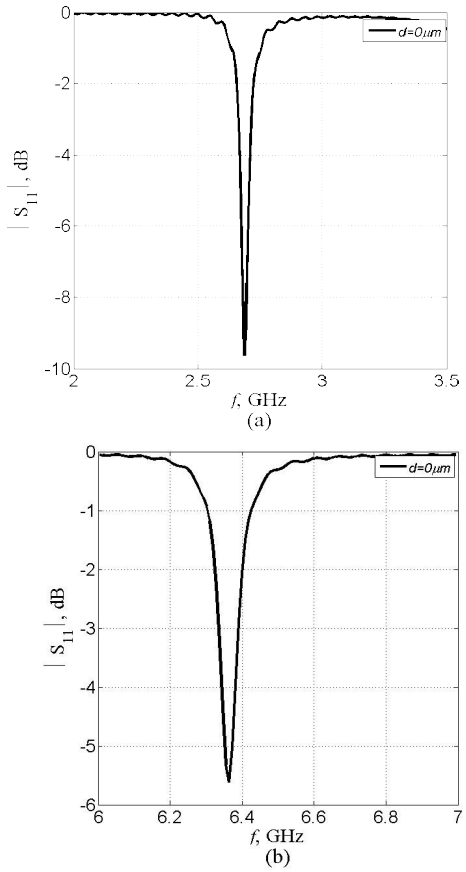


Fig. 2. FDTD results of reflective coeffitient of MSA with circular patch (a) and square patch (b) without air gap(thickness of air gap d = 0μm).

Fig. 2 shows that higher-order harmonics and spurious surface waves strongly affect the characteristics of such antenna construction resulting in asymmetry of the frequency dependence about operating frequency.

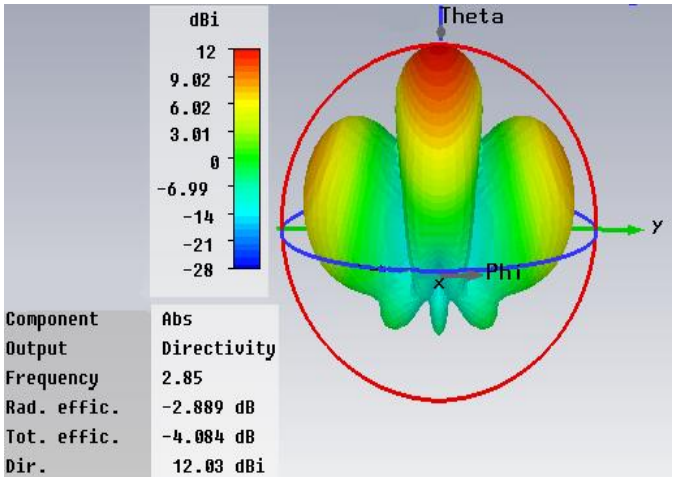


Fig. 3. Radiation pattern of MSA with square patch, thickness of air gap 0,2 mm.

Emission characteristics of microstrip antennas also depend on the design parameters of the radiator and relative dielectric permittivity and thickness of the substrate. Resonant frequency of MSA with circular resonator is obtained using the formula [7]:

$$f_r = \frac{\alpha_{nm} c}{2\pi R_{eff}\sqrt{\varepsilon_{eff}}}, \quad (1)$$

where α_{nm} is the m-th root of the derivative of the Bessel function of order n, c is velocity of propagation of electromagnetic waves in free space, R_{eff} is effective radius of the circular radiator, ε_{eff} - the effective dielectric permittivity of the substrate. For the fundamental TM_{11} mode, the value of α_{nm} is 1.84118. The fringing fields along the circumference of the circular MSA are taken into account by replacing the patch radius R and substrates permittivity ε_s by the effective radius R_{eff} and effective permittivity ε_{eff}, which values are obtained using

$$R_{eff} = R\left[1+\frac{2h}{\pi R\varepsilon_S}\left\{\ln\left(2\frac{R}{h}\right)+1,41\varepsilon_S+1,77+\frac{h}{R}(0,268\varepsilon_S+1,65)\right\}\right]^{\frac{1}{2}}, \quad (2)$$

$$\varepsilon_{eff} = \frac{C(R,h,\varepsilon_S)}{C(R,h,1)}, \quad (3)$$

where h is the thickness of dielectric substrate, $C(R, h, \varepsilon_s)$ and $C(R, h, 1)$ are the total capacitances of the dominant TM11 mode of CMSA with and without a dielectric substrate, respectively. For thin substrates, ε_s should be used instead of ε_{eff} in (1), and for thick substrates, ε_{eff} should be used.

In general, resonant frequency of MSA with rectangular resonator excited at any TM_{mn} mode is obtained using the following expression [8]:

$$f_r = \frac{c}{2\sqrt{\varepsilon_{eff}}}\left[\left(\frac{m}{L}\right)^2+\left(\frac{n}{W}\right)^2\right]^{\frac{1}{2}}, \quad (4)$$

where L and W are length (nonradiating edge) and width (radiating edge) of rectangular patch, m and n are the modes along the L and W, respectively. Effective permittivity ε_{eff} of substrate is obtained using

$$\varepsilon_{eff} = \frac{\varepsilon_S+1}{2}+\frac{\varepsilon_S-1}{2}\left[1+\frac{10h}{W}\right]^{-\frac{1}{2}}. \quad (5)$$

For the fundamental TM_{10} mode and square radiator with the side ength W, the value of f_r is obtained using the equation:

$$f_r = \frac{c}{2W\sqrt{\varepsilon_{eff}}}, \quad (6)$$

Equations (1) and (4) show increasing of the resonant frequency of MSA with the decrease of radiators size and relative dielectric permittivity of substrate.

The thickness of the dielectric substrate used for the manufacture of antennas, usually much smaller than the wavelength. Therefore, only small part of the waves excited by an

emitter inside the structure reaches its limits where radiation should occur. This explains the low efficiency of microstrip antennas. Expressions (2), (3) and (5) shows dependence of MSA resonant frequency on dielectric substrate thickness. Since the value of this frequency is inversely proportional to the dielectric permittivity of substrate it increases with increasing of the substrate thickness.

The expressions for approximately calculating the percentage bandwidth of the MSA in terms of patch dimensions and substrate parameters is given by [8]

$$\Delta f[\%] = \frac{A h f_r}{c\sqrt{\varepsilon_S}}\sqrt{\frac{W}{L}}, \qquad (7)$$

$$\text{where } A = \begin{cases} 180, \dfrac{hf_r}{c\sqrt{\varepsilon_S}} \le 0.045 \\ 200, 0.045 \le \dfrac{hf_r}{c\sqrt{\varepsilon_S}} \le 0.075 \\ 220, \dfrac{hf_r}{c\sqrt{\varepsilon_S}} \ge 0.075 \end{cases}.$$

Radiation from microstrip line structure similar to the microstrip antenna can be significantly reduced when using a thin substrate with a high value of relative dielectric permittivity. Radiation from microstrip antennas, on the other hand, maintained for higher efficiency. Therefore the thick substrate with a low dielectric permittivity is used in microstrip antennas. MSA fabricated on substrates with a high value of dielectric permittivity are prefered for easy integration with radio frequency input stage monolithic microwave integrated circuits. However, the use of substrates with high values of dielectric constant results in low efficiency and narrow bandwidth as it is shown in (7).

With the increase in h, the fringing fields from the edges increase, the W/h ratio reduces which decreases ε_{eff} and hence increases the resonance frequency and bandwidth.

Iincreasing the size of the air gap d between metal patch and dielectric substrat can increase the operating frequency of the antenna and reduce the return loss. Results of FDTD simulation of reflection coefficient magnitude are shown in Fig. 4. Results obtained from the given expressionss are in good agreement with the FDTD simulation in CST Microwave Studio.

Fig. 5 and Fig. 6 show the increasing of resonant wavenumber $k = 2\pi f_r/c$ and bandwidth of square MSA with the increase of relative air gap size d/h. Effective dielectric permittivity ε_{eff} for various ratios of d/h was computed by finite element method [9].

These dependences show good tunability of MSA parameters with the change of the distance between metal patch and dielectric substrate.

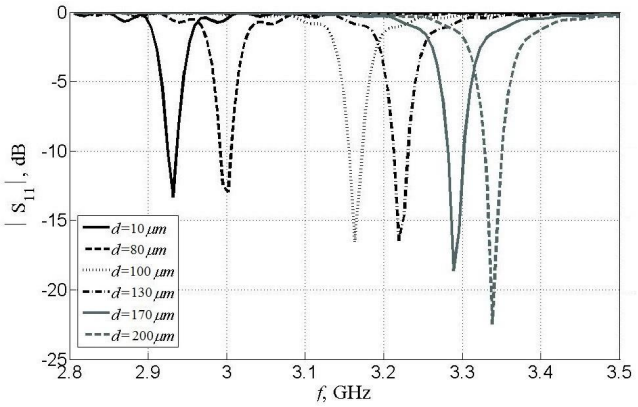


Fig. 4. FDTD results of reflection coeffitient magnitude of MSA with circular patch for different thickness d of air gap.

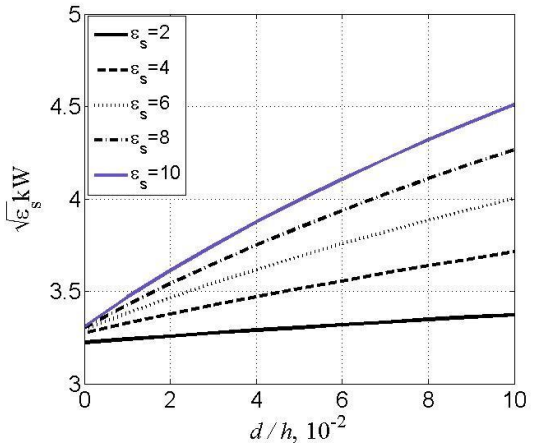


Fig. 5. Variations of resonant wavenumber of square MSA versus d/h for various values of ε_s, $W/h = 10$.

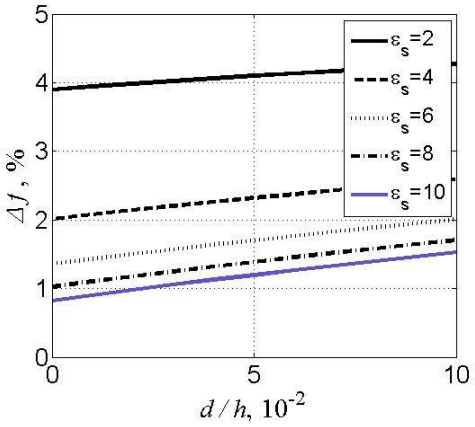


Fig. 6. Variations of bandwidth of square MSA versus d/h for various values of ε_s, $W/h = 10$.

III. Discussion

Microstrip antenna design with air gap between metal patch and dielectric substrate is characterized by tuning of its basic parameters such as magnitude and phase of S_{11} and radiation pattern in frequency domain. Increasing of the air

gaps size increases the operating frequency of the antenna and reduces the return loss.

The method of antenna feeding depends on many factors. The most important factor is the impedance matching between radiating and exciting structures. Coaxial excitation has the advantage of simplicity of design. However, it has several limitations. Firstly, coaxial excitation grating requires a large number of solder joints. Second, to increase the bandwidth thicker substrates are used and therefore the longer probe is required. This causes a rise in parasitic radiation from the probe, increasing the power of surface waves and rising input inductance. Coplanar microstrip exciters are easy to design and manufacture. However exciting microstrip lines contribute parasitic radiation. Therefore, they are used where performance requirements are not strict and exciter must be coplanar with the transmitter.

To build fully integrated systems is necessary for all system components to be compatible with each other. In the microwave frequency range it is difficult to find the optimal value of dielectric permittivity of substrate. It is better to use the substrate material with a high dielectric constant, which allows reducing the size of the device and radiation losses. Radiation characteristics of antennas and the antenna arrays worsen when using substrate materials with high values of dielectric constant. Consequently it is necessary to find a compromise when choosing a substrate material.

IV. CONCLUSION

Realization of active devices requires materials with high dielectric constant (typically order 12) that in the range of micro- and millimeter waves leads to the necessity of using substrates with high electrical thickness.

Large thickness leads to increased propagation of surface waves, which inevitably causes an increase in losses. Although the thickness of the substrate can then be reduced, this leads to a reducing the strength of the structure, which is a serious problem in manufacturing of large substrates required for circuits operating in the millimeter range. Therefore, the more promising are methods for reducing dielectric permittivity directly under the antenna. Results of simulation for MSA with different sizes of air gap under the patch show good ability of antenna parameters tuning.

Development antennas with frequency tuning is just started, and it is expected that further improvement in such antennas will be realized through the implementation of microelectromechanical part of their structures. Since modern communication systems operate at higher frequency, the size of antenna used for them should be decreased significantly. This makes the process of antennas manufacturing more complicated because smaller size is impossible to implement without the use of microfabrication. The last allows serial production of such devices organization that will have higher accuracy while the cost is reduced.

V. ACKNOWLEDGEMENT

Authors are indebted to Prof. F. Dubrovka for fruitful discussions.

REFERENCES

[1] R. Garg, P. Bhartia, I. Bahl, A. Ittipiboon, Microstrip Antenna Design Handbook. Artech House antennas and propagation library, 2001.

[2] D. G. Fang, Antenna Theory and Microstrip Antennas. CRC Press, Taylor & Francis Group, 2010.

[3] K.-F. Lee, “Microstrip patch antennas – basic properties and some recent advances,” Journal of Atmospheric and Terrestrial Physics, vol. 51, № 9/10, pp. 811 – 818, April 1989.

[4] F. F. Dubrovka, S. E. Martynyuk, “Stripline Antenna Array for Local Information-Telecommunication Radio Networks,” Radioelectronics and Communication Systems, vol. 52, № 8, pp. 445 – 448, August 2009.

[5] N. Ruda, Y. Prokopenko, Y. Poplavko, “Electromagnetic analysis of waveguide-dielectric phase shifter,” Electronics and Communications, № 2(61), pp. 46 – 51, 2011.

[6] D. M. Pozar, “Microstrip Antennas,” Proceedings of IEEE, vol. 80, № 1, pp. 79 – 92, January 1992.

[7] K. Guney, N. Sarikaya “A comparative study of models of adaptive systems with fuzzy logic, designed to calculate the resonant frequency of round microstrip antennas”, Radiotechnics and Electronics, vol. 54, № 4, pp. 389 – 400, 2009.

[8] G. Kumar, K. P. Ray, Broadband microstrip antennas. Artech House antennas and propagation library, 2003.

[9] I.Golubeva, V.Kazmirenko, P.Sergiyenko and Y.Prokopenko, “Effective permittivity in tunable microstrip and coplanar lines,” Proceedings of the XXXII International Scientific Conference ELNANO-2012, pp. 69-70, April 2012.

Thermal Resistance Of Power Submicron Heterojunction Field-Effect Transistors

Vladimir Timofeyev, Elena Semenovskaya
Department of physical and biomedical electronics
National Technical University of Ukraine «KPI»
Kyiv, Ukraine
v.timofeyev@kpi.ua, helensem@ukr.net

Abstractð **The paper presents a method for thermal resistance calculating of power submicron heterojunction field-effect transistor. Described method establishes dependence of the thermal resistance of the transistor on its thermo- and electrophysical parameters, geometrical size and shape. The results of thermal resistance calculation for submicron transistors are proved by experiment and show good agreement.**

Keywordsð thermal resistance, power submicron heterojunction field-effect transistor, thermal model

I. INTRODUCTION

Designing of integrated circuits and especially monolithic microwave integrated circuit (MMIC) requires estimation of the influence of electrothermal processes in the active region of power field-effect transistors (FETs). It is known, that the reliability and performance of power FETs and MMIC are limited by operation transistor channel temperature. In turn, the maximum operating temperature is determined by the input DC channel dissipation and the self-heating effects.

Generally, the temperature behavior is described by the three-dimensional Laplace equation: $\nabla^2 T(x, y, z) = 0$, where $T(x, y, z)$ – the temperature at any point in the space.

Only in special cases, for concentric spheroids, concentric cylinders and parallel plates, this equation can be solved analytically [1]. In practice numerical solutions have been based on finite element, finite difference and finite volume methods [2]. However, numerical solution of Laplace's equation is not usable for most circuit designers. First, it is due to the complexity of the boundary conditions definition for the multilayer structures. Second, frequently, the solution does not converge. Therefore, simplified method of transistor thermal resistance estimation, depending on the geometrical and thermal transistor parameters makes practical interest.

II. PROBLEM STATEMENT AND SOLUTION

Modeling power FET with constant highly localized heat sources (gates) is located on sapphire substrate with thickness h_2. Topology of power multiple finger FET with gate length L_g, gate width W_g and gate-gate spacing L_x, on sapphire substrate with thickness h_2 and GaN layer with thickness h_1 is presented on Fig. 1.

The frequently used method for analysis of thermal resistance of an FET, based on the Fourier's conduction law, allows to find the maximum channel temperature directly under the gate [2 – 4]. Expression thermal resistance calculation is as fallows

$$R_T = \frac{\Delta T}{Q} = \frac{h}{\lambda \cdot S}, \qquad (1)$$

where ΔT is temperature drop; Q is heat flux; h is thickness of structure; λ is thermal conductivity; S is cross-section area.

The FET case is more complicated than above mentioned. This approach uses simplifying assumption that the heat transfer is confined by $\alpha < 45°$, where α is spreading angle of the heat flow between the gate and base. The resulting equation is as follows:

$$R_T = \frac{h}{\lambda (W_g + h)(L_g + h)},$$

where W_g and L_g are width and length of the gate, respectively.

This method is well converged for special cases: for gates in the shape of a square and a circular disk. When the gate was made in the 'thin long line form", error was equal to ± 10% … 50% [3].

In [5] detailed studies of heat-flux behavior in FET structures were performed using finite-element simulation. Heat propagation can be divided into two regions. In first region (I) heat propagates in an ellipsoid manner, in second region (II) – to the elliptic cylinders case.

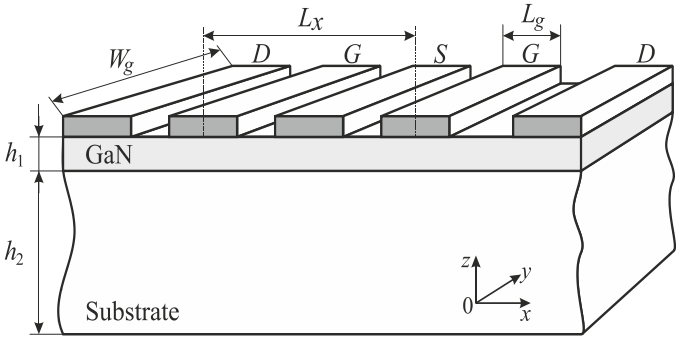


Fig. 1. Topology of power multiple finger FET.

Cross section of FET with typical isothermal surfaces for region I and II is shown on Fig. 2.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

Total thermal resistance equals to the sum of the thermal resistance of first and second regions:

$$R_{\text{T}} = \frac{1}{\pi W_g \lambda} \ln\left(\frac{V\left(f\left(g(\sqrt{2}s)+1\right)\right)}{V\left(f\left(g(L_g)\right)\right)}\right) + \frac{1}{2\pi s\lambda} \ln\left(\frac{h(2.3h)}{h(s)}\right), \quad (2)$$

where s is gate-gate spacing; $h(x) = \frac{\sqrt{1+g(\sqrt{2}x)}+1}{\sqrt{1+g(\sqrt{2}x)}-1}$;

$V(z) = \frac{z-1}{z+1}$; $f(w) = \sqrt{\frac{\sqrt{w}+1}{\sqrt{w}-1}}$; $g(y) = \left(\frac{W_g}{y}\right)^2$.

Equation (2) has an implicit form and can only solved numerically.

In [6] a method for calculating of thermal resistance of the power MOSFET according to thermal model and data on the form and size of the substrate and the heat source (the gate) is proposed. The method includes two stages: the development of the thermal model and definition of the cross section areas of the thermal model. In the first stage the heat source is placed on the surface of the chip, and then drawing line of heat flow spreading from the center and the edges of the heat source. Taking into account, that the heat flow is spreading from the edges of the heat source at the angle β, and from the center of the heat source at the angle α. Thus, the thermal model is the three-dimensional figure bounded from the top by the area of the heat source, and from the bottom by the area of the substrate. In the second stage: the thermal resistance is calculated by: $R_{\text{T}} = \sum_{i=1}^{n} R_{\text{T}i}$, K/W, where $R_{\text{T}i} = \frac{h_i}{\lambda \cdot \Delta S_i}$, K/W; n is number of cross-sections; h_i is height of i-th cross-section; ΔS_i is increase cross-sectional areas at the i-th height; λ is thermal conductivity.

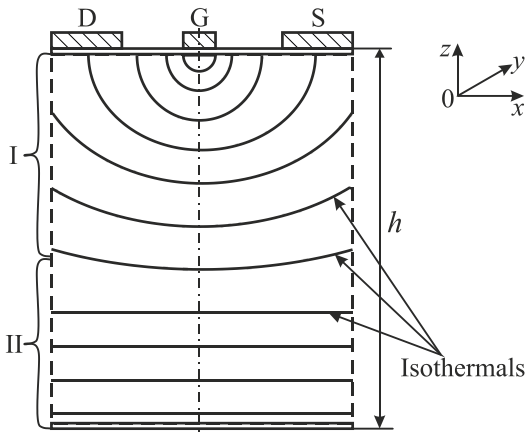


Fig. 2. Cross section of FET with isothermal surfaces shown for regions I and II.

In this paper, the dependence of spreading angles of the heat flux on the heat source shape (strip, circle, square, rectangle) and the operation modes of the transistor (average, maximum, minimum) are received. The advantage of this method is thermal resistance calculation of the transistor of any geometric and shape of the substrate and gate.

In [7] thermal resistance is calculated based on increment of heat flow spreading area, whereas in [8] is calculated from ratio of the thermal model volume elements. In [7] thermal model elements and thermal resistance in the growth area of the heat flow spreading where calculated by the following equation:

$$R_{\text{T}} = f^R \cdot f^S, \text{ K/W},$$

where $f^R = \frac{S_{chip}}{V_{chip} \cdot \lambda}$ – geometrical equivalent, where V_{chip} and S_{chip} – the volume and area of the chip, respectively; $f^S = \Delta S / S_{chip}$ – dimensionless coefficient. Tangents of the angles of the heat flow spreading tgα = 1,571 and tgβ = 0,819 are taken from [6] for strip-shaped gate.

The transistor thermal model at the top has the form of the gate with length L_g and width W_g, and at the bottom has the form of substrate with area $L \times W$ (Fig. 3 a). Lateral surface of the thermal model has two fractures: first at a height h_{f1}, where heat flows at the angle α and heat flow at the angle β are crossed; the second at a height h_{f2}, where the heat flows at the angle α crossed side edges of the substrate (Fig. 3 b).

The height h_{f1} is calculated as:

$$h_{f_1} = \Delta h'_{f1} + \Delta h''_{f1} + \Delta h'''_{f1} + \ldots + 0.01 =$$

$$= \frac{L_g}{2 \cdot \text{tg}\alpha} + \frac{L_g}{2 \cdot \text{tg}\alpha} \cdot \frac{\text{tg}\beta}{\text{tg}\alpha} + \frac{L_g}{2 \cdot \text{tg}\alpha} \cdot \left(\frac{\text{tg}\beta}{\text{tg}\alpha}\right)^2 + \ldots + 0.01,$$

where $\Delta h'_{f1} = \frac{L_g}{2 \cdot \text{tg}\alpha}$; $\Delta h''_{f1} = \frac{\Delta h'_{f1} \cdot \text{tg}\beta}{\text{tg}\alpha}$; $\Delta h'''_{f1} = \frac{\Delta h''_{f1} \cdot \text{tg}\beta}{\text{tg}\alpha}$.

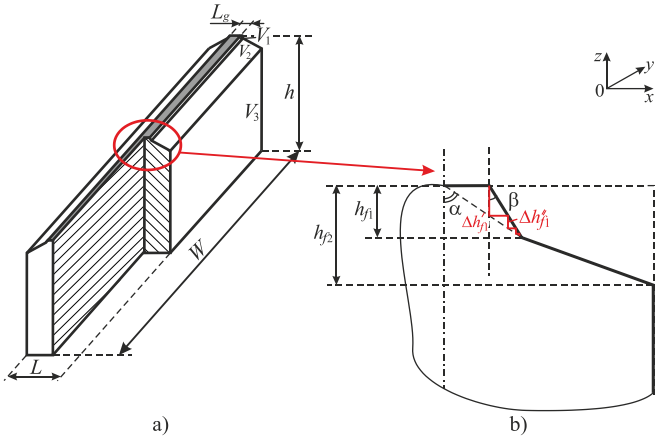


Fig. 3. The thermal model of submicron transistors (a), and its part (b).

The figure shows: h, W, L is dimensions of the substrate; L_g is length of the heat source; α, β is angles of the heat flow spreading; h_{f1} and h_{f2} is heights of the fracture of the lateral surface.

Height h_{f1} is composed from the elements of a decreasing series with decreasing rate $\mathrm{tg}\beta/\mathrm{tg}\alpha = 0{,}819/1{,}571 = 0{,}517$.

Height h_{f2} is calculated as $h_{f\,2} = (L/2)/\mathrm{tg}\alpha$.

The area S_{chip} and volume V_{chip} of chip are calculated as $S_{chip} = L \cdot W$ and $V_{chip} = h \cdot L \cdot W$.

Cross-sectional area of the thermal model at the height h_{f1} is calculated: $S_1 = L_1 \cdot W$, where $L_1 = L_g + 2 \cdot h_{f\,2} \cdot \mathrm{tg}\beta$ is a length of the cross-section of the thermal model at a height h_{f1}.

Cross-sectional area at a height h_{f2} is equal to the cross-sectional area of the chip $S_2 = S_{chip}$.

Coefficient f^S is calculated as $f^S = \Delta S / S_{chip}$, where ΔS is the increase in cross-sectional area of the thermal model.

In the method [8] the thermal resistance of the transistor is calculated by: $R_T = f^V \cdot f^R$, where f^V is dimensionless ratio of the volume of a chip have covered of heat flow, to the volume of chip.

III. THERMAL MODEL OF POWER FET

In this work we calculated the thermal resistance of GaN 4-finger power FET on sapphire substrate with thickness $h_2 = 100$ μm and constant thermal conductivity $\lambda_{sap} = 35$ W/(m·K), with gate length $L_g = 0.25$ μm, gate width $W_g = 250$ μm and gate-gate spacing $L_x = 30$ μm, thermal conductivity GaN layer $\lambda_{GaN} = 35$ W/(m·K).

The following simplifications are assumed:

1) All surface are adiabatic, no heat flux alowed.
2) The substrate has a constant thermal conductivoty λ.
3) The dissipated power generates a conctant heat flux directly under the gate.
4) All fingers have the same temperatures.
5) The heat source is long and thin, i.e. $W_g >> L_g$, $h_2 >> L_g$, $h_2 >> L_x$.

Thermal model of GaN 4-finger power FET on sapphire substrate consists of four regular tetrahedral prisms (V_1) with heights h_{f1}, four regular tetrahedral prisms (V_2) with heights ($h_{f2} - h_{f1}$) and the parallelepiped (V_3) with height ($h_1 - h_{f2}$) (see Fig. 4).In the top the thermal model of transistor is limited by gate area and in the bottom the area of the chip.

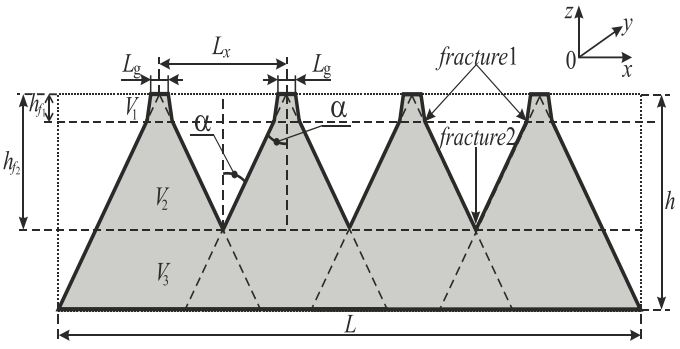


Fig. 4. Cross-section of the thermal model of the power multiple finger FET.

We define the value of the volume of the thermal model elements:

$$V_1 = h_{f\,1}/3 \cdot \left(S_1 + S_g + \sqrt{S_1 \cdot S_g}\right),$$

$$V_2 = \left(h_{f\,2} - h_{f1}\right)/3 \cdot \left(S_1 + S_{chip} + \sqrt{S_1 \cdot S_{chip}}\right);$$

$$V_3 = S_{chip}\left(h_1 - h_{f\,2}\right)$$

and the volume of the heat equivalent $V_{HE} = 4 \cdot V_1 + 4 \cdot V_2 + V_3$.

Thermal resistance is calculated as:

$$R_T = f^V \cdot f^R,$$

where $f^V = V_{HE}/V_{chip}$ and $f^R = \dfrac{S_{chip}}{V_{chip} \cdot \lambda}$.

Developed method shows that the geometric equivalent of the thermal resistance f^R is defined similarly to the method of the thermal resistance calculation [2]. The functions tgα and tgβ of heat flow spreading angles are identical to the same functions of the method [6]. Method [8] was further developed to calculate the thermal resistance of power multiple finger FETs.

IV. VERIFICATION MODEL

To verify our model, we used experimental data from [9] for GaN 4-finger power FET on sapphire substrate with thickness $h_2 = 100$ μm and constant thermal conductivity $\lambda_{sap} = 35$ W/(m·K), with gate length $L_g = 0.25$ μm, gate width $W_g = 250$ μm and gate-gate spacing $L_x = 30$ μm, $\lambda_{GaN} = 35$ W/(m·K). In the future, modeling results can be confirmed in the experiment.

In Figs. 5 – 8 we compare predicted thermal resistance as a function of substrate thickness (h_2), gate length (L_g), gate width (W_g) and gate-gate spacing (L_x) with experimental data. The parameters were varied across a wide range of practical situations.

V. CONCLUSION

The method for calculation the thermal resistance of GaN 4-finger power FET on sapphire substrate with the strip-shaped gate is presented. The method allowed to establish the dependence of the thermal resistance on geometrical dimensions of both gate and chip. Dependence of the thermal resistance on the substrate thickness and the gate width as well dependence on the gate length and gate-gate spacing are established.

The model was verified by comparising with experimental data. Difference between those does not exceed 10 %.

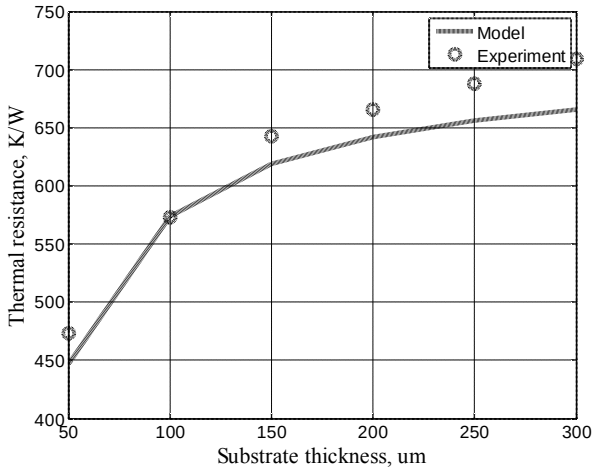


Fig. 5. Dependence of the thermal resistance on substrate thickness for several practical thicknesses, assuming L_g = 0.25 μm; L_x = 30 μm; W_g = 250 μm.

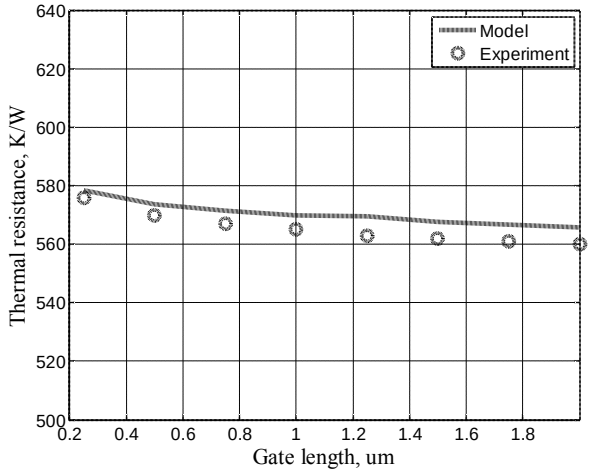


Fig. 6. Dependence of the thermal resistance on gate length for several practical gate lengths, assuming h_2 = 100 μm; L_x = 30 μm; W_g = 250 μm.

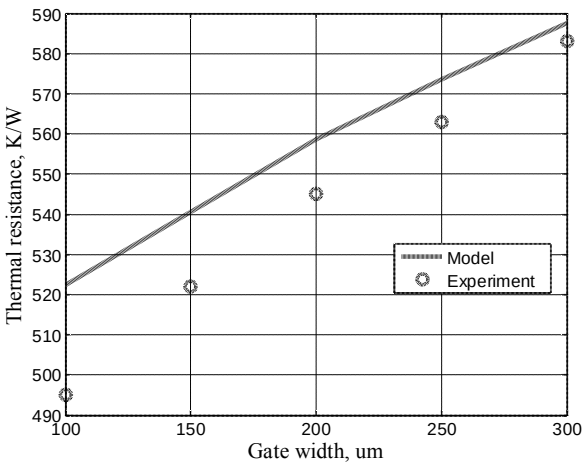


Fig. 7. Dependence of the thermal resistance on gate width for several practical gate widths, assuming h_2 = 100 μm; L_g = 0.25 μm; L_x = 30 μm.

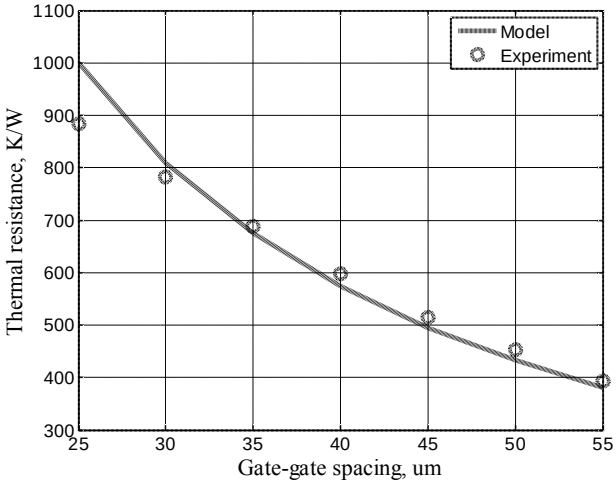


Fig. 8. Dependence of the thermal resistance on gate-gate spacing for several practical spacing between gates, assuming h_2 = 100 μm; L_g = 0.25 μm; W_g = 250 μm.

REFERENCES

[1] G Ellison, Thermal computations for electronics: Conductive, radiative and convective air cooling. Chicago, Illinois, USA: Jamnia & Associates, 1999.

[2] H. Yang, H. Wang, and C.L. Tan, "Thermal resistance of metamorphic InP-based HBTs on GaAs substrate using a linearly graded InxGa1-xP metamorphic buffer," IEEE Trans. Electron Devices, vol. 51, pp. 1221 – 1227, August 2004.

[3] A. Pacelli, P. Pacelli, and M. Mastrapasgua, "Compact modeling of thermal resistance in bipolar transistors on bulk and SOI substrate," IEEE Trans. Electron Devices, vol. 49, pp. 1027 – 1031, June 2002.

[4] R. Joy and E.S. Schlig "Thermal properties of very fast transistors," IEEE Trans. Electron Devices, vol. 17, pp. 586 – 594, March 1997.

[5] A.M. Darwish, A.J. Bayba, and H.A. Hung "Accurate determination of thermal resistance of FETs," IEEE Trans. Microwave Theory Tech., vol. 53, pp. 306 – 313, January 2005.

[6] A.L. Zakharov and E.I. Asvadurova, Calculation of the thermal parameters of semiconductor devices: Method of equivalents. Moscow, Russia: Radio and communication, 1983.

[7] Patent 89911 Ukraine, IPC G01R 31/26, H01L 21/66 O.V. Semenovska, V.I.Timofeyev "A method of the thermal resistance calculation of the submicron transistor," patented 31.10.08., published 10.03.10., Bulletin # 5.

[8] Patent 96061 Ukraine, IPC G01N 25/18 .V. Semenovska, V.I.Timofeyev "A method of the thermal resistance calculation of the submicron transistor," patented 19.02.10., published 26.09.11., Bulletin # 18.

[9] A.M. Darwish, A.J. Bayba, and H.A. Hung "Thermal resistance calculation of AlGaN-GaN devices," IEEE Trans. Microwave Theory Tech., vol. 52, pp. 2611 – 2620, November 2004.

Structural and Electrical Characterization of SiGe Heterostructures Containing a Pure Ge Strained Quantum Well

A.H.A. Hassan[1], O.A. Mironov[2], A. Dobbie, .J.H. Morris, J.E. Halpin, V.A. Shah, M. Myronov, D.R. Leadley
Department of Physics, University of Warwick Coventry, CV4 7AL. UK amnahassan53@yahoo.co.uk

A. Feher, E. Cizmar
Institute of Physics, Faculty of Science P.J. Safarik University Kosice, Slovakia
oamironov@yahoo.co.uk

S. Gabani
Centre of Low Temperature Physics
Institute of Experimental Physics SAS, Kosice, Slovakia

V.V. Andrievskii, I.B. Berkutov
B.I. Verkin Institute for Low Temperature Physics and Engineering of NAS of Ukraine, Kharkov, Ukraine

***Abstract* — In this paper we present structural characterization and magneto-transport properties of the two dimensional hole gas in strained germanium heterostructures. An extremely high hole mobility has observed, along with the lowest value of effective mass to date, using Schubnikov de Haas oscillations for both normal and inverted structures. The channel is confirmed to be pure germanium, with low background impurity scattering that improves the hole transport.**

Keywords — Ge/SiGe; 2DHG; Pure Ge; Shubnikov de Haas oscillations.

I. INTRODUCTION

Germanium is the most effective replacement for Si in terms of device technology, it has high hole mobility with low effective mass which helps to improve the performance of Si-based devices[1]. Introducing strain to nanometer thick Ge layers opens a new field for researchers to investigate the new transport properties of the carriers, as strain splits the heavy-light bands further reducing the effective mass [2] and the interband scattering. Selectively doping the heterostructure containing strained germanium (sGe) can lead to a two dimensional holes gas (2DHG) with improved transport properties of the carriers in the MODFET transistor channel. Many 2DHG results have been published regarding the mobility of holes in these structures [3-9]. The highest published mobility exceeded 10^6 cm^2/Vs at sheet density 3×10^{11} cm^{-2} for the "normal" heterostructure i.e. with the doping above the channel [10]. Two parameters that are significant in defining the scattering mechanisms that affected mobility are the effective mass and the Dingle ratio (α) of transport to quantum lifetime. The lowest value of effective mass measured previously in this system is 0.08 m_0 for strained Ge on $Si_{0.3}Ge_{0.7}$ [11]. Furthermore, the maximum Dingle ratio [2] obtained is 15 for an inverted structure i.e. doping below the Ge channel, at sheet density 2.4×10^{12} cm^{-2}.

In this paper we present the electrical and structural characterization of high purity sGe channel on Si0.2Ge0.8 using different techniques such as high resolution x-ray diffraction (HR-XRD), cross sectional transmission electron microscopy (XTEM) and ultra-low energy secondary ion mass spectrometry (SIMS). We also assess the electrical properties by analyzing the Schubnikov de Haas (SdH) oscillations in magnetotransport measurements to calculate mobility, effective mass and the ratio of transport to quantum lifetime (α). This study reveals the highest hole mobility and Dingle ratio in a fully strained Ge 2DHG, as well as the lowest effective mass.

II. EXPERIMENTAL RESULTS

A. Sample Structure

By comparing "normal" modulation doped heterostructures with otherwise similar "inverted" ones we can study the mobility of the carriers in the strained Ge channel with a minimum effect of background impurity scattering. For that reason, we study inverted and normal structures (Fig.1) grown in an ASM Epsilon 2000 reduced pressure chemical vapour deposition (RP-CVD) chamber. Both heterostructures contain a 20 nm fully strained germanium (sGe) channel, equivalent to 0.65% lattice mismatch with respect to the underlying $Si_{0.2}Ge_{0.8}$ buffer. This strain tuning buffer [12] is grown without any chemical mechanical polishing (CMP) and has a low threading dislocation density (TDD) of $\leq 4\times10^6$ cm^{-2} [13-14]. A 10 nm doping layer with B-density of 1×10^{18} cm^{-3} is placed above a 30 nm thick spacer for the normal structure and below a 20 nm thick spacer for the inverted one. The structure is capped by a 30 nm SiGe layer followed by 2 nm of Si grown at 500ºC to avoid strain relaxation [15].

[1]Also at: Department of Physics, Tripoli University, Tripoli, Libya
[2]Also at: International Laboratory of High Magnetic Fields and Low Temperatures, 53-421 Wroclaw, Poland.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

2 nm Si cap	2 nm Si cap
30 nm $Si_{0.2}Ge_{0.8}$ cap	30 nm $Si_{0.2}Ge_{0.8}$ cap
10 nm B:$Si_{0.2}Ge_{0.8}$ supply layer	20 nm Strained Ge channel
30 nm $Si_{0.2}Ge_{0.8}$ spacer	20 nm $Si_{0.2}Ge_{0.8}$ spacer
20 nm Strained Ge channel	10 nm B:$Si_{0.2}Ge_{0.8}$ supply layer
50 nm $Si_{0.2}Ge_{0.8}$	50 nm $Si_{0.2}Ge_{0.8}$
500 nm $Si_{0.2}Ge_{0.8}$	500 nm $Si_{0.2}Ge_{0.8}$
1000 nm $Si_{0-0.2}Ge_{1-0.8}$	1000 nm $Si_{0-0.2}Ge_{1-0.8}$
600 nm Ge	600 nm Ge
Si substrate (100)	Si substrate (100)
a	b

Fig. 1. Schematic diagram of the strained Ge QW heterostructure, (a) normal structure11-289,(b) inverted structure11-284.

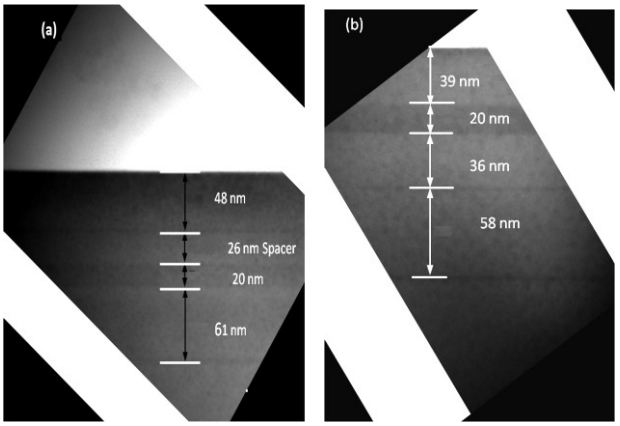


Fig. 2. Cross section XTEM for the top parts of (a) the normal structure 11-289 and (b) inverted structure 11-284.

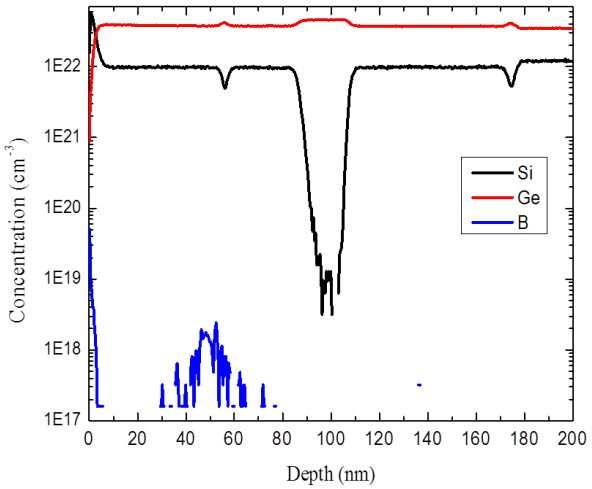


Fig. 3. Si, Ge and B SIMS profiles for the normal structure 11-289.

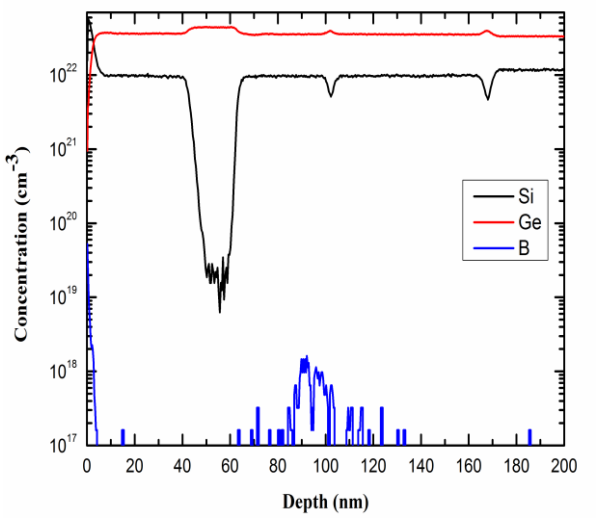


Fig. 4. Si, Ge and B SIMS profiles for the inverted structure 11-284.

B. Structural Characterization by XRD, SIMS and XTEM

The structures have a very low threading dislocation (TDD ~4×10^6 cm^{-2}) [17]. They were investigated as well using XTEM and SIMS depth profiles using a near normal incidence O_2^+ primary beam at 250 eV, which has previously been determined [16] to offer an accurate quantification of nanometre scale $Si_{1-x}Ge_x$/Ge heterostructures. XTEM shows that both structures have no defects and a smooth interface within the field of view. The techniques also each confirm the channel thickness of 20 nm for both the normal and inverted structures. However, the spacer thickness measured by XTEM (Fig.2) differs from its nominal value, with a thickness of 26±1 nm for the normal structure. The doping layer thickness was measured to be 13 nm for the normal structure from SIMS (Fig.3). Although the thickness of the doping layer in the inverted structure is not clear from XTEM, the SIMS measurement shows it is higher than the expected value (at 18±2 nm). SIMS shows (Fig.4) that the spacer thickness for the inverted structure is 20±2 nm. In general, the XTEM images in Fig 2 confirm these dimensions.

The average Boron concentration, measured from the SIMS profile in (Fig.3) for the normal structure, is 1.4×10^{18} cm^{-3}. Significant segregation of the B-doping in the spacer is seen, which has given a larger doped region than intended. For the inverted structure the Average doping concentration level is ~7×10^{17} cm^{-3}. These SIMS profiles also confirm that the purity of Ge channel for the normal structure, with a Si concentration in the centre of the channel < 0.01 at %. Slightly more Si is present in the channel centre of the inverted structure < 0.015 at %, but overall we can term both of these as pure Ge.

Strain and composition have been determined using HR-XRD. Out of plane and in plane lattice constant were obtained from (004) and (224) reciprocal space maps (RSM). Figure 5 and 6 show the five different layers in our structure, two peaks for $Si_{0.2}Ge_{0.8}$ appear here as result of the different of Ge composition between the buffer (x=0.79) and the cap (x=0.82). The buffer was found to be slightly over relaxed in respect of the Si substrate. The strained Ge layer was

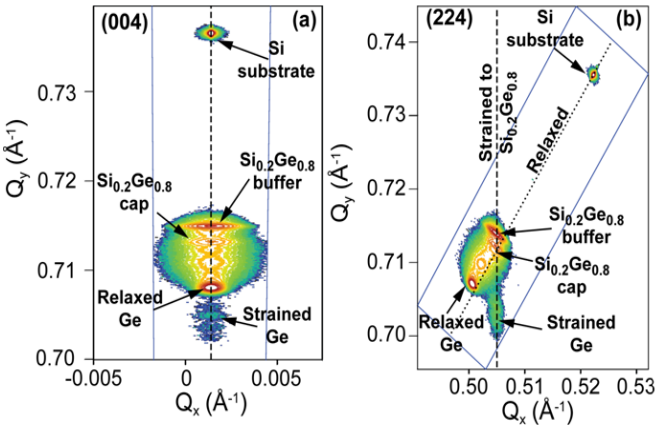


Fig. 5. XRD for normal structure (a) (004) symmetric reflection (b) (224) asymmetric reflection.

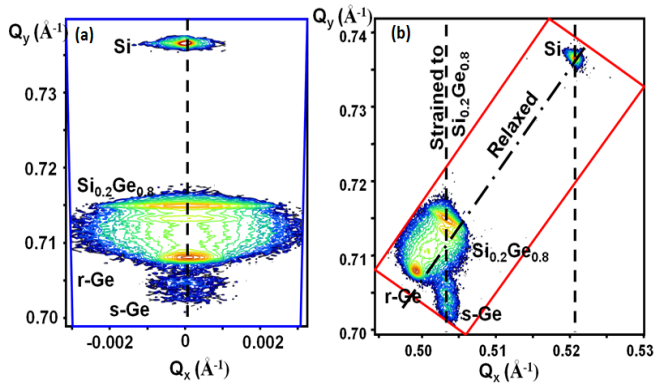


Fig. 6. XRD for Inverted structure (a) (004) symmetric reflection (b) (224) asymmetric reflection.

expected to be 0.84% fully strained with respect to $Si_{0.2}Ge_{0.8}$ buffer. However, from XRD it is clearly equal to a 0.65% lattice mismatch. XRD rocking curves also confirm that the s-Ge channel remains strained. Finally, from the (004) ω-2θ rocking curves the thickness of the channel was calculated to be 20±1 nm, in corroboration with the other measurement techniques for both normal and inverted structures.

C. *Galvanomagnetic Properties:*

Hall bars, of width 50 µm and probe arm separations either 250 µm or 500 µm, were fabricated using wet etching with Al contact deposited by thermal evaporation and annealed at 425ºC for 20 min under dry N_2, to enable Al-Si eutectic needle-like penetration happen to a depth of ~ 60-120 nm and establish an ohmic contact to the Ge-QW. Firstly, hole mobility and carrier sheet density were determined using standard resistivity and Hall measurements at temperatures down to 12 K. The result at 12 K for the normal structure was a mobility of 1.1×10^6 cm^2/Vs at a carrier sheet density of 3.0×10^{11}cm^{-2} (previously published [10, 18-19]). The inverted structure had a mobility of 0.26×10^6 cm^2/Vs at a sheet density of 6.3×10^{11}cm^{-2}. These significant results demonstrated that more study on these samples are essential.

Magnetotransport measurement of ρ_{xx} and ρ_{xy} were made at temperatures 1.5 K-3K for sample 11-289HB8b (normal structure) and 90mK–1.5K for 11-284HB4b (inverted) (Fig.7 and 8). An extremely high mobility of 1.34×10^6 cm^2/Vs at a sheet density of 2.9×10^{11} cm^{-2} has been calculated for the normal hetero-structure, while the mobility for the inverted structure was 0.509×10^6 cm^2/Vs at a sheet density of 5.14×10^{11} cm^{-2}. This result for holes in the normal structure is comparable with the highest reported electron mobility for a strained Si channel [20], where an electron mobility of 1.6×10^6 cm^2/Vs at a carrier sheet density of 1.5×10^{11} cm^{-2} was observed at 0.3 K.

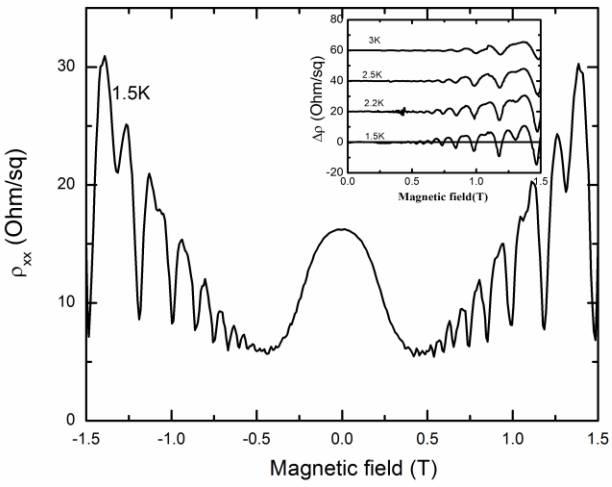


Fig. 7. Magnetoresistance ρ_{xx} versus magnetic field for the normal structure 11-289HB8b with 90 deg orientation of Hall Bar to <110> direction of (100) Si substrate.

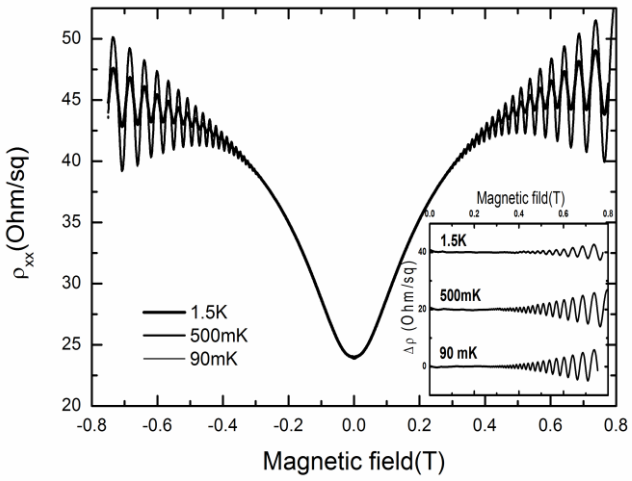


Fig. 8. Magnetoresistance ρ_{xx} versus magnetic field for the inverted structure 11-284HB4b with 90 deg orientation of Hall Bar to <110> direction of (100) Si substrate.

To calculate the effective mass, the amplitude of Schubnikov de Haas (SdH) oscillations were measured over different temperatures for range of magnetic fields [21]. The normal structure has the lowest value of hole effective mass published so far at 0.063 m_0 (Fig.9), which makes a significant contribution to the improved mobility for this structure. Previously measured effective mass values for strained "Ge" channels were significantly higher, with 0.083–0.095 m_0 and

0.087 m_0 being reported for LEPE-CVD [11] and MBE-grown [8] material with similar carrier sheet densities. The very low effective mass we observe may be a consequence of the low temperature RP-CVD growth leading to an extremely pure Ge channel, with less than 0.01 at.% of Si. We previously found that as the Si content in a strained Ge channel increases, due to annealing, the hole mobility decreases. The mass will also be low due to the channel retaining full strain at level of 0.65%.

The Dingle plot (Fig. 9) illustrates a straight line intercept with Y-axis at "4" that is compatible with the theory [22], and relates to homogeneous broadening of the Landau levels. The ratio of transport to quantum lifetime $\alpha = 78$ is also determined from the slope of the Dingle plot, and it is clear from the plot of Fig. 10, for various values of B at which the peak occurs, that the points are nearly in a line with a gradient equal to unity. This value of α indicates that the mobility in the normal sample is limited by small angle scattering and means that remote impurity scattering is dominant, with a negligible influence from background impurities. Moreover, it can be seen from the Dingle plot of the inverted structure that the curve is not a straight line as for the normal structure. Instead, two slopes can be seen for this data, which is because there are two subbands occupied by carriers in the quantum well of the inverted structure. On the other hand, the normal structure has just one subband occupied by holes in the quantum well. The slope of the line of the inverted sample, fixed to have the correct theoretical intercept, also produces a high transport to quantum lifetime ratio of $\alpha = 33$ (which is much higher than all previous results apart from the normal structure here). The effective mass of the inverted structure, at 0.070 m_0, is slightly higher than for the normal structure, which is expected as result of the higher sheet density. This result indicates that the mobility for the inverted structure is again limited by remote impurity scattering.

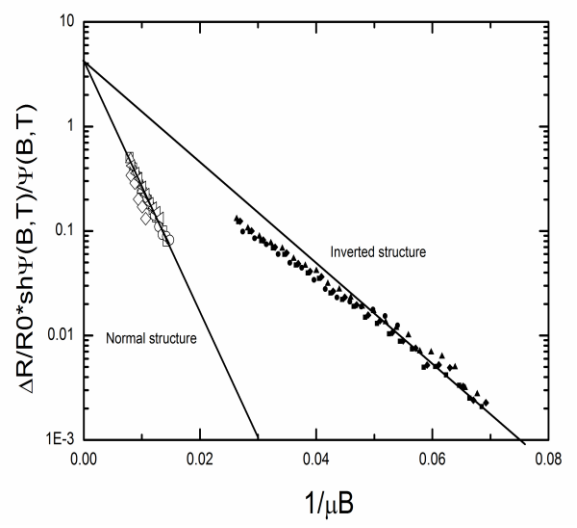


Fig. 9. Dingle plot at temperatures in the range 1.5K<T< 3K for the Hall bars samples of normal structure 11-289HB8b and 90 mK<T<1.5K for the inverted structure 11-284HB4b.

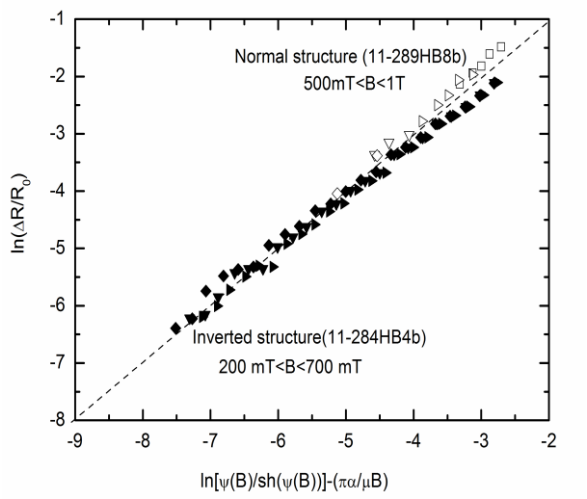


Fig. 10. Plot of $\ln \frac{\Delta R}{R}$ vs $\ln(\frac{\sinh \psi}{\psi}) - \left(\frac{\pi\alpha}{\mu B}\right)$ for the Hall bars of normal structure 11-289HB8b and inverted structure 11-284HB4b.

III. CONCLUSIONS

Our study of sGe heterostructures results in an extremely low effective mass for both the normal and inverted structure 0.063 m_0 and 0.070 m_0, respectively. These lower values represent one of the significant advantages of strain modulated band structure and it is a result of the lower sheet density. Furthermore, the low effective mass plays an important role for mobility improvement. The purity of the Ge channel (Si concentration < 0.010 at %) in the normal structure 11-289 and slightly higher than in the inverted one 11-284 (Si <0.015 at %) produces using RP-CVD, which could be another reason for its higher mobility than the literature, which they using another growth technique, the purity of the channel and the quality of the structure reduces both background impurity scattering and interface roughness which proved by the high Dingle ratio measure in this study. This result improve the hole mobility of the carriers and it is compatible with theoretical prediction indicated the mobility could be for bulk Ge mono crystals higher than 1×10^6 cm^2/Vs below 10 K for low background impurity level [23-25].

In summary, we have studied the structural and electrical properties of 2DHGs in sGe quantum wells grown on top of a $Si_{0.2}Ge_{0.8}$ relaxed buffer. Low temperature hole mobility exceeding one million has been observed, with the lowest effective mass of 0.063 m_0 and an extremely high transport to quantum lifetime $\alpha = 78$ for the normal structure. The inverted structure also has a higher mobility compared to previous literature reports, even though it is lower than the mobility of our normal structure. The purity of the fully strained Ge channel (in respect to Si penetration from the barriers during growth) together with low background B-impurities scattering are suggested as a significant reason for these new outstanding results.

ACKNOWLEDGMENT

This work has been partially supported by the EPSRC UK Renaissance Germanium project and the grant of the Slovak Research and Development Agency under the contract No. APVV-0132-11, VEGA-0106-13, and SAS Centre of Excellence CFNT MVEP. OAM acknowledged Dr T. Konstantinov for fruitful and stimulating discussions and National Scholarship Program of the Slovak Republic for the Mobility Support for Researchers in the Academic Year 2012/2013.

REFERENCES

[1] S. M. Sze, Physics of Semiconductor Devices, 2nd ed. New York: Wiley,1981.

[2] K. Sawano, K. Toyama, R. Masutomi, T. Okamoto, N. Usami, K. Arimoto, K. Nakagawa, and Y. Shiraki, "Strain dependence of hole effective mass and scattering mechanism in strained Ge channel structures," Applied Physics Letters, vol. 95, p. 3, Sep 2009.

[3] S. Madhavi, V. Venkataraman, and Y. H. Xie., "High room-temperature hole mobility in Ge0.7Si0.3/Ge/Ge0.7Si0.3modulation-doped heterostructures," Applied Physics Letters, vol. 89, p. 2497, 2001.

[4] M. Myronov, T. Irisawa, S. Koh, O. A. Mironov, T. E. Whall, E. H. C. Parker, and Y. Shiraki, "Temperature dependence of transport properties of high mobility holes in Ge quantum wells," Journal of Applied Physics, vol. 97, Apr 2005.

[5] B. Rossner, D. Chrastina, G. Isella, and H. von Kanel, "Scattering mechanisms in high-mobility strained Ge channels," Applied Physics Letters, vol. 84, pp. 3058-3060, Apr 2004.

[6] M. Myronov, K. Sawano, and Y. Shiraki., "Enhancement of hole mobility and carrier density in Ge quantum well of SiGe heterostructure via implementation of double-side modulation doping" Applid physics letters, vol. 88, 2006.

[7] T. Irisawa, H. Miura, T. Ueno, and Y. Shiraki, "Channel width dependence of mobility in Ge channel modulation-doped structures," Japanese Journal of Applied Physics Part 1-Regular Papers Short Notes & Review Papers, vol. 40, pp. 2694-2696, Apr 2001.

[8] T. Irisawa, M. Myronov, O. A. Mironov, E. H. C. Parker, K. Nakagawa, M. Murata, S. Koh, and Y. Shiraki, "Hole density dependence of effective mass, mobility and transport time in strained Ge channel modulation- doped heterostructures," Applied Physics Letters, vol. 82, pp. 1425-1427, Mar 2003.

[9] H. v. Kanel, "Very high hole mobilities in modulation-doped Ge quantum wells grownby low-energy plasma enhanced chemical vapor deposition," Applied Physics Letters,, vol. 80, p. 2922, 2002.

[10] A. Dobbie, M. Myronov, R. J. H. Morris, A. H. A. Hassan, M. J. Prest, V. A. Shah, E. H. C. Parker, T. E. Whall, and D. R. Leadley, "Ultra-high hole mobility exceeding one million in a strained germanium quantum well," Applied Physics Letters, vol. 101, p. 172108, 2012.

[11] Benjamin Rossner, Giovanni Isella, and H. v. K. nel, "Effective mass in remotely doped Ge quantum wells," Applied Physics Letters,, vol. 82, 2003.

[12] M. Myronov, A. Dobbie, V. A. Shah, X. C. Liu, V. H. Nguyen, and D. R. Leadley, "High Quality Strained Ge Epilayers on a Si0.2Ge0.8/Ge/Si(100) Global Strain-Tuning Platform," Electrochemical and Solid State Letters, vol. 13, pp. H388-H390, 2010.

[13] V. A. Shah, A. Dobbie, M. Myronov, D. J. F. Fulgoni, L. J. Nash, and D.R. Leadley, "Reverse graded relaxed buffers for high Ge content SiGe virtual substrates," Applied Physics Letters, vol. 93, Nov 2008.

[14] V.A.Shah, A. Dobbie, M. Myronov, and D. R. Leadley, "Reverse graded SiGe/Ge/Si buffers for high-composition virtual substrates," Journal of Applied Physics, vol. 107, Mar 2010.

[15] A. Dobbie, V. H. Nguyen, R. J. H. Morris, X.-C. Liu, M. Myronov, and D. R. Leadley, "Thermal Stability of Thin Compressively Strained Ge Surface Channels Grown on Relaxed Si0.2Ge 0.8 Reverse-Graded Buffers," Journal of the Electrochemical Society, vol. 159, pp. H490-H496, 2012.

[16] R. J. H. Morris, M. G. Dowsett, R. Beanland, A. Dobbie, M. Myronov, and D. R. Leadley, "Overcoming Low Ge Ionization and Erosion Rate Variation for Quantitative Ultralow Energy Secondary Ion Mass Spectrometry Depth Profiles of Si1−xGex/Ge Quantum Well Structures," Analytical Chemistry, vol. 84, pp. 2292-2298, 2012/03/06 2012.

[17] J. Mitard, B. De Jaeger, G. Eneman, A. Dobbie, M. Myronov, M. Kobayashi, J. Geypen, H. Bender, B. Vincent, R. Krom, J. Franco, G. Winderickx, E. Vrancken, W. Vanherle, W. E. Wang, J. Tseng, R. Loo, K. De Meyer, M. Caymax, L. Pantisano, D. R. Leadley, M. Meuris, P. P. Absil, S. Biesemans, and T. Hoffmann, "High Hole Mobility in 65 nm Strained Ge p-Channel Field Effect Transistors with HfO2 Gate Dielectric," Japanese Journal of Applied Physics, vol. 50, Apr 2011.

[18] Andy Dobbie, Maksym Myronov, Richard J. H. Morris, Martin J. Prest, James S. Richardson-Bullock, Amna H. A. H. Hassan, Vishal A. Shah, Evan H. C. Parker, T. E. Whall, and D. R. Leadley, "Ultra-High Hall Mobility (1 x 106 cm2V-1s-1) in a Two-Dimensional Hole Gas in a Strained Germanium Quantum Well Grown by Reduced Pressure CVD," ISTDM 2012,June 4-6,2012, Berkeley,CA,USA, 2012.

[19] Oleg A. Mironov, Maksym Myronov, Amna H. Hassan, Andy Dobbie, Richard J. H. Morris, Vishal Shah, Martin Prest, I.B. Berkutov, V. V. Andrievskii, and D. R. Leadley, "Effective mass and scattering in ultra-high hole mobility strained germanium," 6th international conference Materials Science and Condensed Matter Physics September 11-14, 2012, Chisinau, Moldova, 2012.

[20] R. R. Lieten, S. Degroote, M. Leys, N. E. Posthuma, and G. Borghs, "Solid phase epitaxy of amorphous Ge on Si in N-2 atmosphere," Applied Physics Letters, vol. 94, p. 3, Mar 2009.

[21] I.B. Berkutov, V.V. Andrievskii, Yu.F. Komnik, Yu.A. Kolesnichenko, R.J.H. Morris, D.R. Leadley, and O. A. Mironov., "Magnetotransport studies of SiGe-based p-type heterostructures: problems of the effective mass determination," Low Temperature Physics, vol. 38, 2012.

[22] A. Isihara and L. Smrcka, "Density and magnetic field dependences of the conductivity of two-dimensional electron systems," J. Phys. C: Solid State Phys, vol. 19, p. 6777, 1986.

[23] W. Shockley, Electrons and Holes in Semiconductors. Princeton, New Jersey: D. Van Nostrand Company, Inc, 1950.

[24] G. Ottaviani, C. Canali, F. Nava, and J. W. Mayer, "Hole drift velocity in high-purity Ge between 8 and 220 [degree]K," Journal of Applied Physics, vol. 44, pp. 2917-2918, 1973.

[25] F. Szmulowicz, "Calculation of optical- and acoustic-phonon—limited conductivity and Hall mobilities for p-type silicon and germanium," Physical Review B, vol. 28, pp. 5943-5963, 1983.

Ge/GaAs Thin Films for Thermometer and Bolometer Application

V.F. Mitin
V. Lashkaryov Institute of Semiconductor Physics
National Academy of Sciences of Ukraine
Kyiv 03028, Ukraine

P.M. Lytvyn
V. Lashkaryov Institute of Semiconductor Physics
National Academy of Sciences of Ukraine
Kyiv 03028, Ukraine

V.V. Kholevchuk
V. Lashkaryov Institute of Semiconductor Physics
National Academy of Sciences of Ukraine
Kyiv 03028, Ukraine

V.V. Mitin
V. Lashkaryov Institute of Semiconductor Physics
National Academy of Sciences of Ukraine
Kyiv 03028, Ukraine

E.F. Venger
V. Lashkaryov Institute of Semiconductor Physics
National Academy of Sciences of Ukraine
Kyiv 03028, Ukraine

O.A. Mironov
NANO-Silicon Group, Department of Physics,
University of Warwick, Coventry CV4 7AL, UK
International Laboratory of High Magnetic Fields
and Low Temperatures, 53-421 Wroclaw, Poland

***Abstract*—We demonstrate that the heavily doped and completely compensated Ge thin films on semi-insulating GaAs substrates are very promising for bolometer and thermoresistor application in the 250 K to 500 K temperature range. These films have single-crystal structure, the pronounced nano-relief surface and compositional nano-inhomogeneities. The conductivity of such films is two-dimensional percolation of charge carriers through a fluctuating electrostatic potential.**

Keywords—Ge/GaAs; heavily doped and strongly compensated semiconductors; percolation; thermometers; bolometers.

I. INTRODUCTION

The Ge/GaAs is promising for a number of practical applications and an excellent model system for IV/III-V heterostructures in general. This heterostructure is characterized by close values of lattice constants of both materials (Ge and GaAs) as well as practical coincidence of their coefficients of thermal expansion. These properties are favorable for epitaxial growth and fabrication of films and heterojunctions with almost perfect structure.

The defect formation in the Ge/GaAs heterostructures depends on which material serves as the substrate (Ge or GaAs). In this work, we consider Ge film epitaxy on the GaAs(100) substrate.

The Ge lattice constant exceeds that of GaAs. Therefore, the Ge films grown on GaAs substrates are compressed. In some conditions the elastically compressed Ge films on GaAs may exist up to film thickness of 5 μm. In that case, there are no mismatch dislocations which could impair the electrical properties of the heterojunction. Thus, it is possible to form a high-quality interface (without misfit dislocations) in the Ge/GaAs heterostructure.

The distinguishing feature of Ge films on GaAs is the strong dependence of their structural, electrical and optical properties, on the method of preparation. This is related to possible interdiffusion of the joining materials. For Ge films, diffused Ga and As atoms act as shallow acceptors and donors, respectively. Therefore, because of autodoping, the Ge films on GaAs may be *p*- or *n*-type with different degrees of compensation. Moreover the degree of structural perfection of the Ge films, as well as the concentration of structural defects in them, is determined by the film preparation condition. The Ge films can be amorphous, polycrystalline, or single crystalline with different degrees of structural perfection.

It is well known that for sensitive element of resistance thermometer and bolometer it is necessary to use the high heat sensitive materials. The heavily doped and strongly (better if completely) compensated Ge can be used for this purpose.

Here we present and generalize the results on development and study of completely compensated single-crystal Ge thin films [1–6]. We show that nature of high heat sensitivity of the developed completely compensated Ge/GaAs thin (~100 nm) films deals with two-dimensional (2D) percolation of charge carriers through a fluctuating electrostatic potential.

In this report we also discuss the preparation conditions of the films, as well as the profound effect of film growth rate and thickness on surface topography and electrical properties. This effect is essential and should be taken into account when developing and producing devices based on the Ge/GaAs heterostructure.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

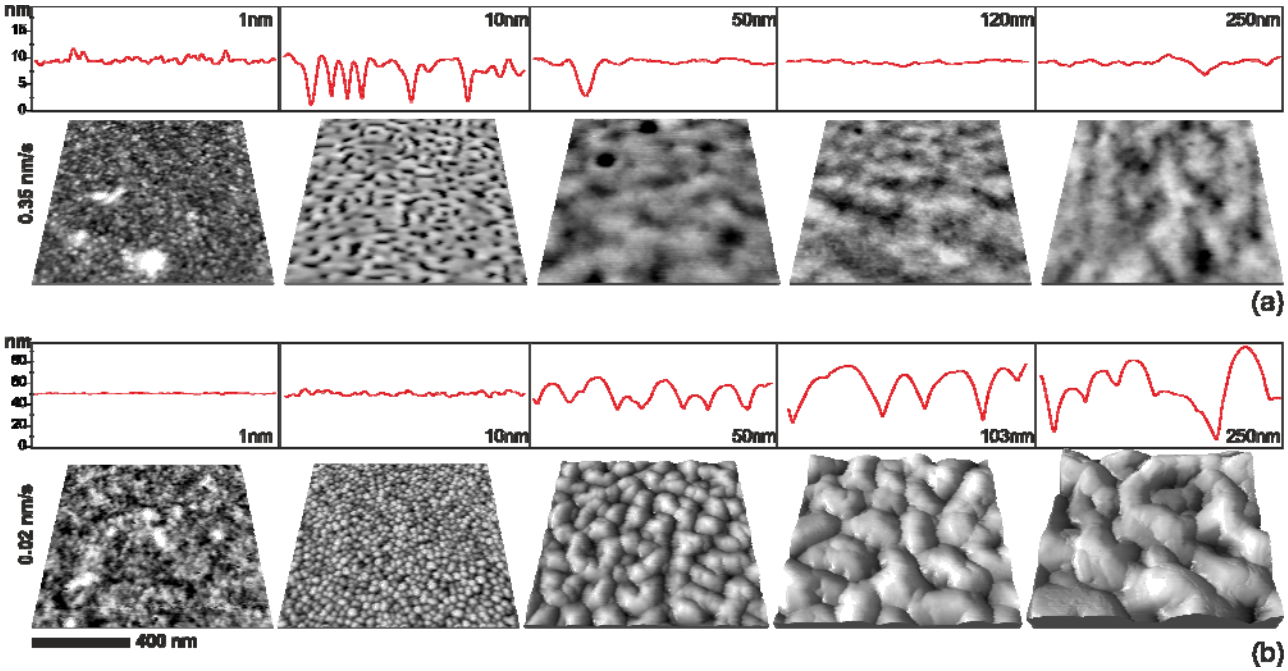


Fig. 1. 3D AFM images from (1×1) μm area of Ge films on GaAs with corresponding surface topography line scan for film thicknesses of 1, 10, 50, 120 and 250 nm. Ge growth deposition rates used are 0.35 nm/s (a) and 0.02 nm/s (b).

II. Experimental Results

A. Preparation Condition and Measurement Procedure

The Ge films were deposited using thermal evaporation of Ge in the vacuum onto semi-insulating (10^7 Ωcm) GaAs(100) substrates. The temperature of GaAs substrate during the film deposition was maintained at 500 °C. The film deposition rate was maintained constant over the course of each deposition; and varied (0.02 - 0.35) nm/s for different specimens. The film thicknesses varied from 1 nm to 250 nm. The morphology of Ge film surfaces was studied with atomic force microscope (AFM) NanoScope IIIa. The electrical properties of the Ge films were determined by the conductivity, Hall effect and magnetoresistance, which were measured in temperature range from 295 K to 380 K and magnetic field induction 2 Tesla.

B. Surface Morfology

Figure 1 show AFM images of films surfaces at different thickness obtained at low (0.02 nm/s) and high (0.35 nm/s) deposition rate. Figure 2 shows line-profile (height variations) of the corresponding film surfaces presented in figure 1 The morphology of film surfaces strongly depends on film thickness and deposition rate. The films prepared by low deposition rate have surfaces with pronounce morphology of grains of irregular shape and substructure. At a film thickness of about 1 nm, a nucleating layer with characteristic grain diameter of 12 nm is formed. The character of surface grain distribution changes from log-normal to normal and bimodal in a 1 nm, 10 nm and 20 nm thick layers, respectively. The root-mean-square (RMS) roughness and surface area increase (SAI) (defined as the difference of the surface area and its projection area divided by the projection area) of the films increase monotonically with thickness; when film thickness is 250 nm, their values are 27 nm and 25%, respectively (Fig. 2(b)). At film thicknesses over 50 nm, the grain size increases abruptly with broadening of the grain diameter histogram (Fig. 2(b)).

In contrast to low deposition rate film surfaces, the surfaces of high deposition rate films (0.35nm/s) are smoother, with the exception of low film thicknesses during the initial growth stages. The growth begins from a nano-grain nucleating layer (at a thickness of ~ 1 nm) with the same log-normal size distribution as low growth rate (not presented here). Typical grain diameter of the ~1 nm nucleating layer is 9 nm, these grains proceed the growth with pits. The density and depth of pits decrease as the film thickness increases; the pits are practically not observed at film thicknesses over 50 nm. RMS and SAI peaks at epitaxial layer thickness of d =20 nm (RMS and SAI values are 2.8 nm and 3%, respectively). At film thicknesses $d \geq 50$ nm, RMS and SAI are 0.7 nm and 0.15%, respectively, and show no further film thickness dependence (Fig. 2(a)).

C. Electrical Properties

The transport phenomena also differed drastically in films obtained at high and low deposition rate. Figure 3 shows typical temperature dependencies of resistivity ρ, Hall coefficient R_H and magnetoresistance $\Delta\rho/\rho$ for Ge films obtained at different deposition rates.

Our investigations showed that Ge films with thickness of about 100 nm obtained at low deposition rate are *p*-type and high-resistant (10–120 Ωcm), with low concentration (10^{15}–10^{16} cm^{-3}) and mobility (20 – 100 cm^{-2}/V s) of free charge

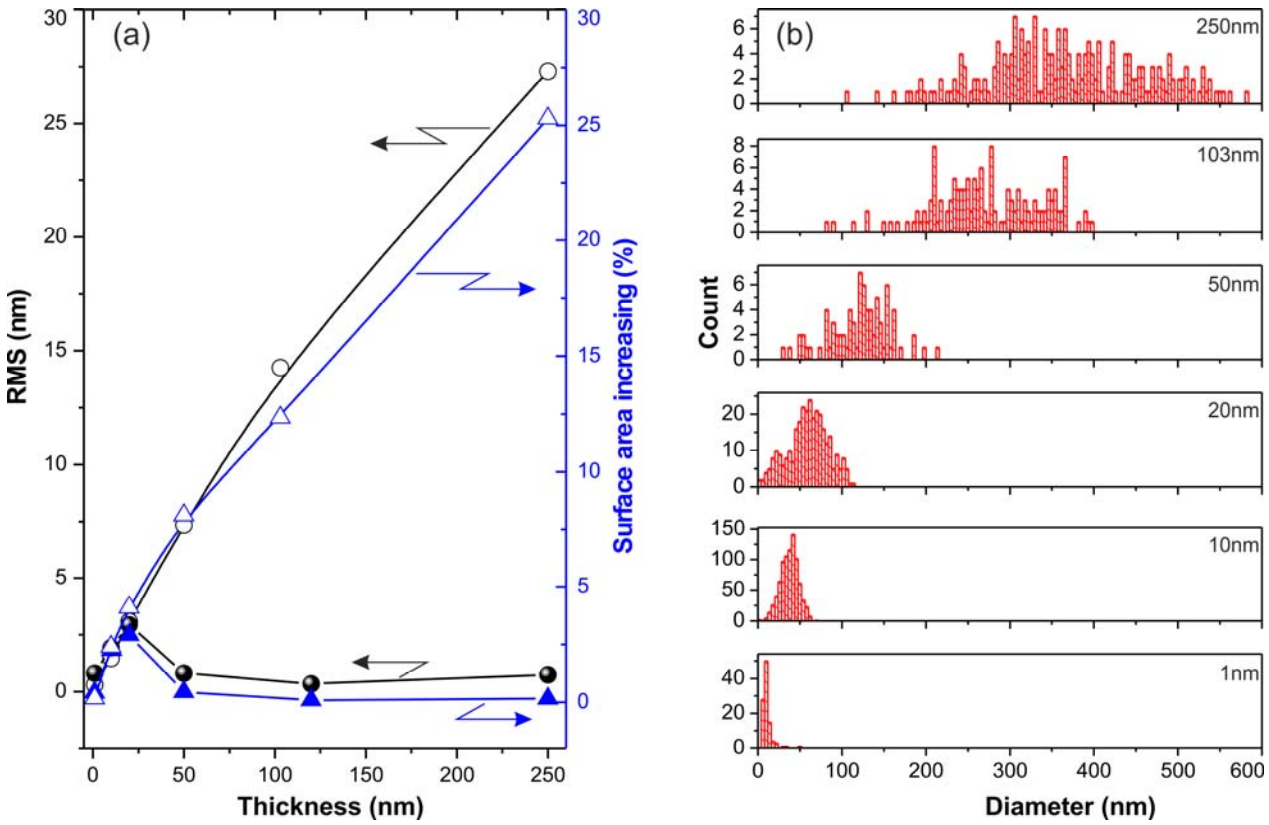


Fig. 2. Dependencies of RMS and surface area increasing on thickness (a) and grain diameter histograms (b) for the Ge films of different thicknesses grown at 0.02 nm/s (open symbols) and 0.35 nm/s (filled symbols) rates (V). The film thickness (d) is indicated in histograms.

carriers (determined from Hall effect measurements) at room temperature. Their conductivity is thermally activated and is described by the expression $\rho(T) = \rho_1 \exp(\varepsilon_1/kT)$ with activation energy ε_1 close or equal to $E_g/2$ for Ge ($\varepsilon_1 = 0.34$ eV for specimen presented on Fig. 3). In the most resistant (compensated) Ge films the Hall coefficient reverses sign at elevated temperatures. The transverse magnetoresistance greatly increases with temperature.

Contrary to the above, the Ge films obtained at low deposition rate are *n*-type and low-resistant (0.1 – 0.01 Ωcm), with high concentrations of free charge carriers (10^{18} – 10^{19} cm^{-3}). Thermal dependence of conductivity, Hall coefficient and magnetoresistance is weak (Fig. 3) characteristic of heavily doped degenerate semiconductors.

III. Discussion of Results

Our investigations showed [6, 7] that Ge films under consideration have single-crystal structure and coherent dislocation-free film–substrate interfaces. Deposition rate strongly effects the electrical properties and surface morphology of the Ge films. In addition the films surface morphology strongly depends on film thickness in addition to the deposition rate.

The early stages of film growth (up to thickness of 20 nm) appear similar for both low and high deposition rates. The growth begins from a nanograin nucleating layer. Then for high deposition rate pits are formed. As the film thickness increases pits heal and disappear at film thicknesses over 50 nm; in this case, surface roughness is below 1 nm (Fig. 1(a) and Fig. 2). Appearance of pits at high deposition rates can be explained by the effect of phase decomposition and formation of As atom clusters that prevent the acquisition of Ge adatoms, a phenomenon observed and discussed in [8, 9]. This correlates well with the electrical investigations that showed that such films were *n*-type.

At low deposition rate, there is no pits formation. Grains of irregular shape determine the film morphology. This can be clearly seen at film thicknesses over 50 nm. Our investigations showed [7] that both the continuous film and protruding grains have a single-crystal structure.

We suppose that for layers grown on mismatched substrates, there is a coupling between the local composition and the surface morphology, so that a rough film will decompose, while a flat layer will be homogeneous.

The presence of compositional inhomogeneities in the Ge films is confirmed by electrical transport measurements. Indeed the films grown at low deposition rate have higher resistivity, low concentrations and mobility of the free charge carriers. Their conductivity is thermally activated. To give consistent explanation for the electrical and optical properties of such films, we assume the presence of fluctuations of electrostatic potential in the films related to compositional

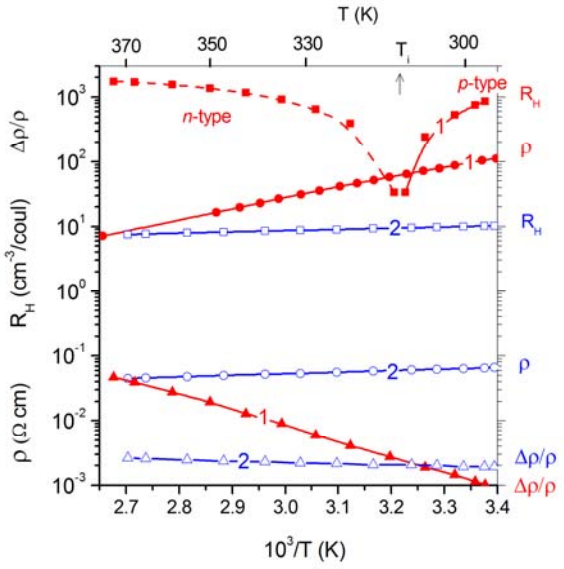


Fig. 3. Typical dependencies of resistivity (ρ), Hall coefficient (R_H) and magnetoresistance ($\Delta\rho/\rho$) on inverse temperature for 100 nm Ge films obtained at deposition rate of 0.02 nm/s (1) and 0.35 nm/s (2).

disorder. The model of potential fluctuations was developed [10] and successfully applied for interpretation of various properties in crystalline heavily doped and compensated, inhomogeneous, as well as an amorphous semiconductors. Such potential fluctuations modulate semiconductor energy bands leading to appearance of potential relief and deep tails of density of states in the semiconductor bandgap (Fig. 4). The main characteristics of potential relief are amplitude γ and spatial scale r that depend on the characteristics of random electrostatic field. In heavily doped and strongly compensated semiconductors γ and r depend on the type of spatial distribution of impurities (correlated or uncorrelated), doping level and degree of compensation [10]. In heavily doped and fully compensated semiconductors with uncorrelated distribution of impurities the amplitude γ of potential relief is equal $E_g/2$ (Fig. 4). Potential relief leads to appearance of regions with free charge carriers in the potential relief valleys and dielectric regions containing no charge carriers. In this case, conductivity is percolation-type and is characterized by percolation level ε_p [10].

At high temperatures ($T > 250$ K) conductivity of Ge films (that is realized by thermionic emission of charge carriers from the Fermi level ε_F to the percolation level ε_p) is characterized by activation energy $\varepsilon_1 = |\varepsilon_p - \varepsilon_F|$. It should also be taken into account that, in the presence of a symmetric random potential, the percolation threshold ε_p depends on the dimensionality of space [11–16]. This is a general property of all percolation systems. A three-dimensional (3D) potential

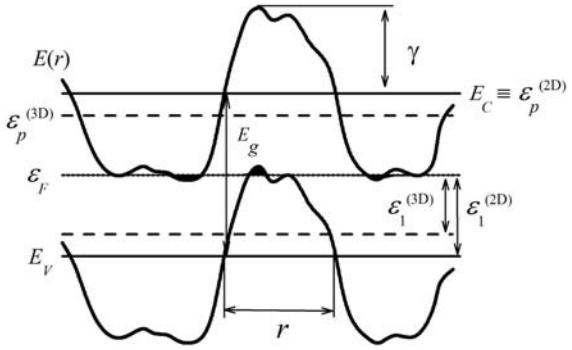


Fig. 4. Energy diagram of a heavily doped and completely compensated semiconductor: r is the size of a typical potential fluctuation; γ the LSFEP amplitude; ε_p the percolation level (dotted line); ε_F the Fermi level.

relief ($d > r$) reduces ε_p (Fig. 4). Only for a two-dimensional (2D) potential relief $d \le r$ the percolation threshold ε_p corresponds to the average value of $E(r)$ (i.e. E_C and E_V) and $\varepsilon_1 = |\varepsilon_p - \varepsilon_F|$ can reach up to half of the Ge band gap ($E_g/2$).

Indeed for thin Ge films (thickness of about 100 nm) potential relief characteristic size r can be equal or greater than the film thickness [4–6]. In this case the potential relief is two-dimensional and 2D percolation is observed. In such films, that are fully compensated, temperature dependence of conductivity is exponential with activation energy close or equal to $E_g/2$.

We observed $p \rightarrow n$ inversion of Hall coefficient R_H in the low deposition rate films as temperature increased (figure 3). The inversion temperature T_i is ~ 310 K. The sign of R_H corresponded to n-type (p-type) conductivity at $T > T_i$ ($T < T_i$). Such inversion indicates strong compensation and bipolar conductivity in the films. Indeed a high degree of compensation (the Fermi level is in the midgap) and fluctuations of the electrostatic potential lead to two independent conductivity channels. At relatively high temperatures, conductivity is realized by both electrons and holes thermally transferred to their percolation levels which are spaced by potential barriers, i.e. conductivity is bipolar. In this case, one should take into account contribution to kinetic effects from both electrons and holes. It is known that different charge carriers contribute differently to the kinetic coefficients. Inversion of the Hall coefficient in Ge films indicates predominance of electrons over holes as temperature increases, i.e., transition to intrinsic conductivity. Low inversion temperature ($T_i = 310$ K) indicates strong compensation of the films. A relatively high concentration of intrinsic charge carriers n_i (10^{15}–10^{16} cm^{-3}) as compared with that in pure intrinsic i-Ge ($n_i = 10^{14}$ cm^{-3} [17]) results from the presence of deep tails of the density of states in the

semiconductor bandgap, which leads to an increase in the number of intrinsic charge carriers in the semiconductor. This may cause effects uncommon to the viewpoint of the theory of homogeneous semiconductors.

It should be also noted that transverse magnetoresistance in the films increases with temperature (Fig. 3). The magnetoresistance effect should reflect the behavior of drift mobility of the free charge carriers. The increase of the drift mobility of free charge carriers with temperature could be explained by reduction of their scattering on fluctuations of random electrostatic potential as well as by reduction of probability of their capture to localized states near the conduction (valence) band edge. This also follows from the presence of a potential relief and compositional disorder in the single-crystal Ge films under investigation.

As to the Ge films grown at high deposition rate, there is practically no compositional disorder in them. The degree of compensation is reduced. The amplitude of electrostatic potential fluctuations is low ($\gamma < kT$). The potential relief ceases to affect the transport processes. The films are heavily doped with degenerate of charge carriers. Therefore, their conductivity is no longer thermally activated and does not depend on temperature.

Thereby, as the deposition rate decreases, the fluctuation of random electrostatic potential in the films is increased and a metal–insulator transition occurs; i.e., the metallic conductivity is replaced with that of an activation type.

IV. Conclusions

The large effect of film growth rate and thickness on properties of the Ge/GaAs thin films was found and studied. This effect is essential and should be taken into account when developing and producing devices based on the Ge/GaAs heterostructure.

The heavily doped and completely compensated Ge films were obtained and investigated. These films have single-crystal structure, the pronounced nano-relief surface and compositional nano-inhomogeneities. The high heat sensitivity of the films is conditioned with 2D percolation of charge carriers through a fluctuating electrostatic potential.

Due to high heat sensitivity, the strongly compensated films (considered in this work) can be used for resistance thermometer and bolometer application.

The manufacturing technology for Ge films on GaAs makes it possible to vary temperature sensitivity of Ge film resistance over wide limits in different temperature ranges (cover the 0.02 K to 500 K), as well as make different versions of resistance thermometers and bolometers with preset properties, i.e., the necessary resistance and heat sensitivity in the required temperature range. The films may demonstrate either metal conductivity (heavily doped films with degenerate charge carriers) or semiconductor (activation-type) dependence of conductivity on temperature with maximal activation energy of $E_g/2$ (completely compensated films).

Some versions of the thermometers are commercially available [18].

Acknowledgment

This work was partially funded under the Ukrainian State Program “Nanotechnologies and Nonmaterials,” Projects No. 2.2.6.15 and No. 3.6.1.10.

OAM thanks Prof. A. Feher (Institute of Physics, Faculty of Science, P.J. Safarik University, Kosice, Slovakia) and National Scholarship Programme of the Slovak Republic for the Mobility Support for Researchers in the academic year 2012/2013.

References

[1] V.F. Mitin, Yu.A. Tkhorik, Yu.M. Shvarts, Ukrainian Journal of Physics, 31, 1 (1986) 105.

[2] N. P. Garbar, L. A. Matveeva, V. F. Mitin, Yu. A. Tkhorik, R. Harman, Yu. M. Shvarts, Z. Stroubek, Sov. Phys. Semicond. 21 (1987) 245.

[3] V. F. Mitin, Yu. A. Tkhorik, E. F. Venger, Microelectron. J. 28 (1997) 617.

[4] V. F. Mitin, Appl. Phys. Lett. 92 (2008) 202111.

[5] V. F. Mitin, J. Appl. Phys. 107 (2010) 033720.

[6] V.F. Mitin, V.K. Lazarov, P.M. Lytvyn, P.J. Hasnip, V.V. Kholevchuk, L.A. Matveeva, E.Yu. Kolyadina, I.E. Kotenko, V.V. Mitin, E.F. Venger, Phys. Rev. B 84 (2011) 125316.

[7] V.K. Lazarov, L. Lari, P.M. Lytvyn, V.V. Kholevchuk, V.F. Mitin, J. Phys.: Conf. Ser. 371 (2012) 012040.

[8] Yu Bai, Kenneth E. Lee, Chengwei Cheng, Minjoo L. Lee, Eugene A. Fitzgerald, J. Appl. Phys. 104 (2008) 084518.

[9] Francois Leonarda, Rashmi Desai, Thin Solid Films 357 (1999) 46.

[10] B. I. Shklovskii and A. L. Efros, *Electronic Properties of Doped Semiconductors,* (Springer, Berlin, 1984).

[11] R. Zallen and H. Scher, Phys. Rev. B 4 (1971) 4471.

[12] A. S. Skal, B. I. Shklovskii, A. L. Efros, Sov. Phys. JETP Lett. 17 (1973) 377.

[13] B. I. Shklovskii, Sov. Phys. JETP Lett. 14 (1971) 269.

[14] A. S. Skal, B. I. Shklovskii, A. L. Efros, Sov. Phys. Solid State 15 (1973) 961.

[15] Brian Skinner, Tianran Chen, B. I. Shklovskii, Phys. Rev. Lett. 98 (2012) 106803.

[16] Tianran Chen, B. I. Shklovskii, arXiv:1212.4183v1 [cond-mat.str-el] 17 Dec 2012.

[17] W. C. Dunlap, Phys. Rev. 79 (1950) 286.

[18] MicroSensor Company: <http://www.microsensor.com.ua>.

Mathematical Simulation of Optical Nanoantenna Based on a Comb-Like Finite Nanostrip Grating

Olga V. Shapoval[1], Jiří Čtyroký[2] and Alexander I. Nosich[1],

[1] Laboratory of Micro and Nano-Optics, Institute of Radio-Physics and Electronics of National Academy of Sciences of Ukraine, Proskury st. 12, Kharkiv 61085, Ukraine

[2]Institute of Photonics and Electronics AS CR, v.v.i., Chaberská 57, 18251 Prague 8, Czech Republic

Abstractð **Mathematical simulation of the optical nanoantenna based on a comb-like finite nanostrip grating is presented. It is carried out using the generalized (effective) boundary conditions (GBCs) imposed on the strip median lines and the Nystrom-type discretization of the relevant singular and hyper-singular integral equations (IEs). Associated scattering and absorption spectra are computed and both surface plasmon resonances and periodicity-induced effects are analyzed.**

Keywordsð ***noble-metal strips, nanoantennas, generalized boundary conditions, singualr/hyper-singular integral equation, Nystrom-type discretization.***

INTRODUCTION

As active research into nanoscience and nanotechnology has been advancing rapidly in recent years [1-5], various methods have been developed to fabricate and theoretically model the optical properties of metals on nanometer scale.

The interest in these properties is associated with plasmon waves and plasmon resonances appearing on noble-metal nanostructures due to negative values of their dielectric functions [6]. They are promising for a variety of applications such as biosensing and efficient manipulation of light using the optical antennas in the visible and infrared ranges. Here, noble-metal strip-like nanostructures are attractive as fundamental objects, which are easily manufactured and serve as building blocks of nanoantennas and nanosensors with unique geometry-dependent optical properties.

Recently we have proposed a convergent, efficient, and accurate numerical algorithm to simulate electromagnetic wave scattering by stand-alone thin material strips and finite strip gratings [7]. Our mathematical simulation is carried out, first, on the basis of the special type of GBCs valid for a thin and high-contrast material layer [8]; they allow us to consider only the limit values of the field components and reduce integration contour to the corresponding strip median lines. Second, for the building of discrete model of the obtained singular and hyper-singular IEs we use a very efficient Nystrom-type numerical algorithm with weighted quadrature formulas (QFs) of interpolation type.

We have applied this technique to the analysis of finite coplanar arrays of silver nanostrips in the visible range [10,11] and graphene strips in the THz range [11]. Now we present a modified technique able to analyze comb-like gratings.

I. MATHEMATICAL SIMULATION

A. Problem Formulation

Consider the scattering of the H-polarized plane wave by a two-dimensional optical nanoantenna based on a comb-like finite nanostrip grating made of N identical silver strips. The corresponding freestanding geometry is shown in Fig. 1.

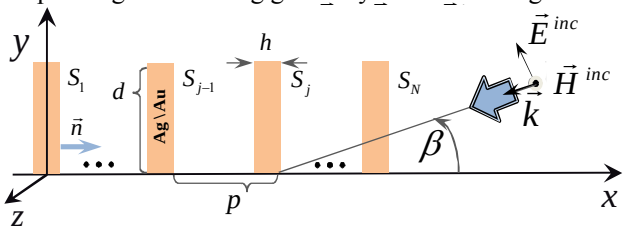


Fig. 1. Cross-section of a free-standing comb-like finite nanostrip grating of N infinitely long (along z-axis) noble-metal strips.

The strips are assumed to be infinite along the z-axis, have thickness h, width d and form a grating with the period p. The silver is characterized with complex-valued permittivity $\varepsilon_r(\lambda)$ obtained via experimental data of [6], where λ is the free-space wavelength.

In the 2-D scattering problem, one has to find a scalar function $H_z^{sc}(\vec{r})$ that is the scattered magnetic field z-component. Here, the total magnetic field $H_z(\vec{r})$ at is a sum of the incident and scattered fields,

$$H_z^{inc}(\vec{r}) = e^{-ik(x\cos\beta + y\sin\beta)} \quad (1)$$

$$H_z^{sc}(\vec{r}) = k\sum_{j=1}^{N}\left[\int_{S_j} v_j(\vec{r}\,')G(\vec{r},\vec{r}\,')d\vec{r}\,' + \int_{S_j} w_j(\vec{r}\,')\frac{\partial G(\vec{r},\vec{r}\,')}{\partial \vec{n}\,'}d\vec{r}\,'\right] \quad (2)$$

Here the scattered field is presented as a sum of single-layer and double-layer potentials, $k = 2\pi/\lambda$ is the free-space wavenumber, $G(\vec{r},\vec{r}\,') = (i/4)H_0^{(1)}(k\,|\vec{r}-\vec{r}\,'|)$ is the Green's function of the 2-D Helmholtz equation, and unknown functions $v_j(\vec{r})$ and $w_j(\vec{r})$, $j = 1,\ldots,N$ are the electric and the magnetic surface currents (jumps of the tangential components of H and E fields, respectively) induced on the strips for $S_j = \{(x,y): x \in [(j-1)p - h/2, (j-1)p + h/2],\ y \in [0,d]\}$.

Exploiting the small thickness of each the strips, $h << \lambda$, we assume it to be zero, i.e. we shrink their cross-sections to the set of median lines $L_j = \{(x,y): x = (j-1)p,\ y \in [0,d]\}$. This enables us to neglect the internal fields inside the strips and impose the two-side GBCs at $L = \mathrm{U}_{j=1}^{N} L_j$: (see [8] for more mathematical details),

$$\partial[H_z^+(\vec{r}) + H_z^-(\vec{r})]/\partial\vec{n} = -i2kR[H_z^+(\vec{r}) - H_z^-(\vec{r})], \quad (3)$$

$$[H_z^+(\vec{r}) + H_z^-(\vec{r})] = i2Qk^{-1}\partial[H_z^+(\vec{r}) - H_z^-(\vec{r})]/\partial\vec{n}. \quad (4)$$

Here, the coefficients *R* and *Q* are the so-called relative electric and magnetic resistivities (surface resistances),

$$R = i\cot(kh\sqrt{\varepsilon_r}/2)/(2\sqrt{\varepsilon_r}),\ Q = (i/2)\sqrt{\varepsilon_r}\cot(kh\sqrt{\varepsilon_r}/2) \quad (5)$$

which contain the strip characteristics such as electric thickness *kh* and relative dielectric function ε_r, $\vec{n}$ is the unit vector normal to the strip grating along the *x*-axis, and the superscripts $\pm$ denote the limit values of the field at the top and bottom faces of the strip, respectively. These GBCs are valid if $kh \ll 1$ and $|\varepsilon_r| \gg 1$ [7,8].

B. Singular and Hyper-Singular Integral Equations

By enforcing the GBC (3),(4) and using the properties of the limit values of potentials in (2), we obtain two dependent sets of IEs of the second kind for the unknown surface currents $w_j(\vec{r})$ and $v_j(\vec{r})$, where $\vec{r}_0, \vec{r} \in L$,

$$iQv_s(\vec{r}_0) + \sum_{j=1}^{N}\int_{L_j} w_j(\vec{r})\frac{\partial G(\vec{r},\vec{r}_0)}{\partial\vec{n}(\vec{r})}d\vec{r} + k\sum_{j=1}^{N}\int_{L_j} v_j(\vec{r})G(\vec{r},\vec{r}_0)d\vec{r} = -u_0(\vec{r}_0), \quad \vec{r}_0 \in L_s \quad (6)$$

$$ikRw_s(\vec{r}_0) + k\sum_{j=1}^{N}\int_{L_j} v_j(\vec{r})\frac{\partial G(\vec{r},\vec{r}_0)}{\partial\vec{n}(\vec{r}_0)}d\vec{r} + \sum_{j=1}^{N}\int_{L_j} w_j(\vec{r})\frac{\partial G(\vec{r},\vec{r}_0)}{\partial\vec{n}(\vec{r}_0)\partial\vec{n}(\vec{r})}d\vec{r} = -\partial u_0(\vec{r}_0)/\partial\vec{n}(\vec{r}_0), \quad \vec{r}_0 \in L_s \quad (7)$$

Here, the normal derivatives age given by expressions,

$$\frac{\partial G(\vec{r},\vec{r}_0)}{\partial\vec{n}(\vec{r}_0)} = k(x-x_0)\frac{i}{4}\frac{H_1^1(k\sqrt{(x_0-x)^2+(y_0-y)^2})}{\sqrt{(x_0-x)^2+(y_0-y)^2}}, \quad (8)$$

$$\frac{\partial G(\vec{r},\vec{r}_0)}{\partial\vec{n}(\vec{r})} = -\frac{\partial G(\vec{r},\vec{r}_0)}{\partial\vec{n}(\vec{r}_0)}, \quad (9)$$

$$\frac{\partial G(\vec{r},\vec{r}_0)}{\partial\vec{n}(\vec{r}_0)\partial\vec{n}(\vec{r})} = \frac{ik}{4}\cdot\left[k(x_0-x)^2\frac{H_0^{(1)}(k\sqrt{(x_0-x)^2+(y_0-y)^2})}{(x_0-x)^2+(y_0-y)^2} + \frac{(y_0-y)^2-(x_0-x)^2}{(x_0-x)^2+(y_0-y)^2}\frac{H_1^{(1)}(k\sqrt{(x_0-x)^2+(y_0-y)^2})}{\sqrt{(x_0-x)^2+(y_0-y)^2}}\right]. \quad (10)$$

On changing the variables, $y^j = d(t+1)/2$, $x^j = (j-1)p$ and $y_0^s = d(t_0+1)/2$, $x_o^s = (s-1)p$, $t, t_0 \in [-1,1]$, and introducing $\rho_{js}^2(t,t_0) = (x^j - x_0^s)^2 + [y^j(t) - y_0^s(t_0)]^2 = (d/2)^2[(t-t_0)^2 + \sigma_{js}^2]$, where $\kappa = kd/2$, $\sigma_{js} = 2p(j-s)/d$, the IEs can be rewritten in more detailed form,

$$4Q\kappa^{-1}v_s(t_0) + \sum_{j=1}^{N}\int_{-1}^{1} v_j(t)K_v^{js}(\kappa,t,t_0)dt + \sum_{j=1}^{N}\int_{-1}^{1}\tilde{w}_j(t)\sqrt{1-t^2}K_w^{js}(\kappa,t,t_0)dt = f_v^s(\kappa,t_0),\ s = 1,...,N, \quad (11)$$

$$4R\kappa^{-1}w_s(t_0)\sqrt{1-t_0^2} + \sum_{j=1}^{N}\int_{-1}^{1}\tilde{w}_j(t)\sqrt{1-t^2}M_w^{js}(\kappa,t,t_0)dt + \sum_{j=1}^{N}\int_{-1}^{1} v_j(t)M_v^{js}(\kappa,t,t_0)dt = f_w^s(\kappa,t_0),\ s = 1,...,N. \quad (12)$$

Note that the current functions $v_j(t)$ are finite and non-zero at the endpoints [-1,1]. The other currents $w_j(t)$ are the functions which behave at the edges as $w_j(t) = \tilde{w}_j(t)(1-t^2)^{1/2}$, where $\tilde{w}_j(t)$ are regular functions tending to finite limits at the endpoints. The right parts are known regular functions for all κ and $t \in [-1,1]$ given by the next formulas, while $n_p = d/p$.

$$f_v^s(\kappa,t_0) = 4i\kappa^{-1}e^{-i\kappa(t_0+1)\sin\beta}e^{-2i\kappa(s-1)\cos\beta/n_p}, \quad (13)$$

$$f_w^s(\kappa,t_0) = 4\kappa^{-1}\cos\beta e^{-i\kappa(t_0+1)\sin\beta}e^{-2i\kappa(s-1)\cos\beta/n_p}. \quad (14)$$

Here, the kernel functions have the form,

$$K_v^{js}(\kappa,t,t_0) = \begin{cases} H_0^1(\kappa|t-t_0|), & j = s \\ H_0^1(\kappa\rho_{js}(t,t_0)), & j \neq s \end{cases}$$

$$K_w^{js}(\kappa,t,t_0) = \begin{cases} 0, & j = s \\ \sigma_{sj}\dfrac{H_1^1(\kappa\rho_{js}(t,t_0))}{\rho_{js}(t,t_0)}, & j \neq s \end{cases} \quad (15)$$

$$M_v^{js}(\kappa,t,t_0) = \begin{cases} 0, & j = s \\ \sigma_{js}\dfrac{H_1^1(\kappa\rho_{js}(t,t_0))}{\rho_{js}(t,t_0)}, & j \neq s \end{cases}$$

$$M_w^{js}(\kappa,t,t_0) = \begin{cases} \dfrac{H_1^{(1)}(\kappa|t-t_0|)}{\kappa|t-t_0|}, & j = s \\ \sigma_{sj}^2\dfrac{H_0^{(1)}(\kappa\rho_{js}(t,t_0))}{\rho_{js}(t,t_0)^2} + \dfrac{H_1^{(1)}(\kappa\rho_{js}(t,t_0))}{\kappa\rho_{js}(t,t_0)}\dfrac{(t-t_0)^2-\sigma_{sj}^2}{\rho_{js}(t,t_0)^2}, & j \neq s \end{cases}.$$

The asymptotic behavior of the Hankel functions is $H_0^{(1)}(z) \approx 2i\ln(z/2)/\pi$ and $H_1^{(1)}(z)/z \approx (i/\pi)\ln z - 2i/(\pi z^2)$, $z \to 0$, respectively. Therefore it is only if the observation point coincides with the integration point (*j* = *s*), that the kernel functions are singular.

C. Nystrom-type Discretization of IEs

Following [7,9], we isolate the singularities and discretize the resulted sets of IEs using Nystrom-type method with two different quadrature rules (but the same discretization order, n) of interpolation type. For the integrals of IEs (11),(12) with unknown electric currents $v_j(t), j=1,...,N$ we use the Gauss-Legendre quadrature formulas with nodes in the nulls of Legendre polynomials $P_n(\tau_j)=0, \quad j=1,...,n$ for the singular and regular parts, respectively,

- $$\int_{-1}^{1} v_j(\tau)\ln|\tau-\tau_{0i}|d\tau \approx \sum_{k=1}^{n} A_k S(\tau_k,\tau_{0i}) v_j(\tau_k), \quad (16)$$
- $$\int_{-1}^{1} v_j(\tau) R_v(\tau,\tau_{0i})d\tau \approx \sum_{k=1}^{n} A_k v_j(\tau_k) R_v(\tau_k,\tau_{0i}), \quad (17)$$

where $\tau_{0i}\in(-1,1)$ and weighted coefficients are given as,

$$A_k = 2/[P_n'(\tau_k)^2(1-\tau_k^2)], \quad (18)$$

$$S(\tau,\tau_{0i}) = \frac{1}{2}\ln(1-\tau_{0i}^2) + L_1(\tau_{0i}) + \sum_{l=1}^{n-1} P_l(\tau)[L_{l+1}(\tau_{0i}) - L_{l-1}(\tau_{0i})] \quad (19)$$

and $L_l(t)$ is the Legendre function of the l-th order.

For the integrals of IEs (11),(12) with unknown magnetic currents $\tilde{w}_j(t), j=1,...,N$ we use the Chebyshev quadrature formulas (with the weight $(1-t^2)^{1/2}$) are more efficient, with nodes in the nulls of Chebyshev polynomials of the second kind, $t_j=\cos(\pi j/n)$, $j=1,...,n$. These formulas have the following form for the hyper-singular, log-singular and regular integrals meet in our analysis,

- $$\int_{-1}^{+1} \tilde{w}_j(t)\ln|t-t_{0i}|\sqrt{1-t^2}dt \approx \sum_{k=1}^{n}\alpha_{ik}\tilde{w}_j(t_k) \quad (20)$$
- $$\int_{-1}^{+1} \frac{\tilde{w}_j(t)}{(t-t_{0i})^2}\sqrt{1-t^2}dt \approx \sum_{k=1}^{n}\beta_{ik}\tilde{w}_j(t_k), \quad (21)$$
- $$\int_{-1}^{+1} \tilde{w}_j(t)R_w(t,t_{0i})\sqrt{1-t^2}dt \approx \frac{\pi}{n+1}\sum_{k=1}^{n}\tilde{w}_j(t_k)R_w(t_k,t_{0i})(1-t_k^2) \quad (22)$$

where coefficients α_{ik} and β_{ik} are given by formulas,

$$\alpha_{ik} = -\frac{\pi}{n+1}(1-t_k^2)\left[\ln 2 + 2\sum_{l=1}^{n}\frac{T_l(t_k)T_l(t_{0i})}{l} + \frac{(-1)^k}{n+1}T_{n+1}(t_{0i})\right] \quad (23)$$

$$\beta_{ik} = \begin{cases} \frac{\pi}{n+1}\frac{(1-t_k^2)(1-(-1)^{k+i})}{(t_k-t_{0i})^2}, & i=k \\ -\frac{\pi(n+1)}{2}, & i\neq k \end{cases} \quad (24)$$

Here, $R_v(t,t_0)$, $R_w(t,t_0)$ can be arbitrary regular functions on the interval [-1,1], and $T_l(t)$ is the Chebyshev polynomial of the second kind.

As the collocation nodes, we choose the corresponding discretization nodes $\{\tau_{0i}\}$ for the first set of IEs, and $\{t_{0j}\}$ for the second one, respectively.

Applying the above presented quadrature formulas, we arrive at a set of matrix equations of the orders $2nN$ for the values $v_j(\tau_k)$ and $\tilde{w}_j(t_k)$, where $j=1,...,N$ and $k=1,\ldots,n$,

$$\sum_{j=1}^{N}\sum_{k=1}^{n} v(\tau_k)A_{lk}^{js} + \tilde{w}(t_k)B_{lk}^{js} = f_v^s(\kappa,\tau_{0l}), \ l=\overline{1,n} \quad (25)$$

$$\sum_{j=1}^{N}\sum_{k=1}^{n} v(\tau_k)C_{lk}^{js} + \tilde{w}(t_k)D_{lk}^{js} = f_w^s(\kappa,t_{0l}), \ l=\overline{1,n}, \quad (26)$$

and the matrix coefficients are given by formulas,

$$A_{lk}^{js} = \begin{cases} (2i/\pi)A_k S(\tau_k,\tau_{0l}) + A_k N_v(\kappa,\tau_k,\tau_{0l}) + 4Q\kappa^{-1}\delta_{lk}, & j=s \\ A_k K_v^{js}(\kappa\rho(\tau_k,\tau_{0l})), & j\neq s \end{cases}$$

$$B_{lk}^{js} = \begin{cases} 0, & j=s \\ \pi(1-t_k^2)K_w^{js}(\kappa,t_k,\tau_{0l}))/(n+1), & j\neq s \end{cases} \quad (27)$$

$$C_{lk}^{js} = \begin{cases} 0, & j=s \\ A_k M_v^{js}(\kappa,\tau_k,t_{0l}), & j\neq s \end{cases}$$

$$D_{lk}^{js} = \begin{cases} \frac{i}{\pi}\alpha_{lk} - \frac{2i}{\pi\kappa^2}\beta_{lk} + \frac{\pi}{n+1}N_w(\kappa,t_k,t_{0l})(1-t_k^2) + 4R\kappa^{-1}\sqrt{1-t_k^2}\delta_{lk}, & j=s \\ \pi(1-t_k^2)M_w^{js}(\kappa,t_k,t_{0l})/(n+1), & j\neq s \end{cases}$$

while δ_{lk} is the Kronecker delta, and regular functions $N_{v(w)}(\kappa,\tau,\tau_0)$ are defined as,

$$N_v(\kappa,\tau,\tau_0) = H_0^{(1)}(\kappa|\tau-\tau_0|) - (2i/\pi)\ln|\tau-\tau_0| \quad (28)$$

$$N_w(\kappa,\tau,\tau_0) = \frac{H_1^{(1)}(\kappa|\tau-\tau_0|)}{\kappa|\tau-\tau_0|} - \frac{i\ln|\tau-\tau_0|}{\pi} + \frac{2i}{\pi\kappa^2|\tau-\tau_0|^2} \quad (29)$$

These finite-size matrix equations represent discrete model of our IEs (6) and (7). On solving them numerically we obtain approximate solution of IEs in the form of interpolation polynomials of the order n for each of the surface currents.

The chosen quadrature formulas ensure rapid convergence of numerical solutions to the accurate ones if discretization order $n\to\infty$. Conservative estimation gives the rates of convergence as $O(1/n)$, although the actual rate is always faster [7,9].

Thus, in order to build our numerical model we have

- imposed the two-side GBCs on the strip median-lines,
- neglected by the internal fields inside the strips,
- reduced boundary problem to a set of coupled 1D singular and hyper-singular IEs,

- discretized obtained IEs using Nystrom-type numerical algorithm with quadrature formulas of interpolation type.

D. Radiation Pattern and Scattering Characteristics

The scattered-field pattern is a function of the observation angle defined as,

$$\Phi(\varphi_0)=\Phi_v(\varphi_0)+\Phi_w(\varphi_0)=\lim_{r_o\to\infty}\sqrt{2/(i\pi kr_0)}^{-1}e^{-ikr_0}H_z^{sc}(r_0,\varphi_0)\,. \quad (30)$$

The use of the large-argument asymptotics for the Hankel functions in the scattered magnetic field in the far zone and above-mentioned corresponding quadrature formulas of interpolation type leads to expressions,

$$\Phi_v(\varphi_0)\approx(i\kappa/4)\sum_{j=1}^{N}\sum_{k=1}^{n}A_k v_j(\tau_k)e^{-2i\kappa(j-1)\cos\varphi_0/n_p}e^{-i\kappa(\tau_k+1)\sin\varphi_0} \quad (31)$$

$$\Phi_w(\varphi_0)\approx\frac{\kappa\pi}{4(n+1)}\cos\varphi_0\sum_{j=1}^{N}\sum_{k=1}^{n}w_j(t_k)(1-t_k^2)e^{-2i\kappa(j-1)\cos\varphi_0/n_p}e^{-i\kappa(t_k+1)\sin\varphi_0} \quad (32)$$

The total scattering cross section (TSCS) is obtained by the integration of the scattering pattern,

$$\sigma_{sc}=(2/\pi k)\int_0^{2\pi}|\Phi(\varphi)|^2 d\varphi, \quad (33)$$

while the absorption cross-section (ASC) is given by the formula that follows from the optical theorem,

$$\sigma_{abs}=-\sigma_{sc}-(4/k)\,\mathrm{Re}\,\Phi(\beta+\pi). \quad (34)$$

The total field at any point is obtained via the quadrature formulas of interpolation type as follows:

$$H_z(x,y)=H_z^{sc}(x,y)+e^{-ik(x\cos\beta+y\sin\beta)} \quad (35)$$

$$H_z^{sc}(x,y)\approx(i\kappa/4)[\sum_{j=1}^{N}\sum_{k=1}^{n}A_k v_j(\tau_k)H_0^{(1)}(k\tilde{\rho}_j(\tau_k,x,y))+$$

$$\frac{\pi}{n+1}\sum_{j=1}^{N}\sum_{k=1}^{n}\tilde{w}_j(t_k)(1-t_k^2)(x-(j-1)p)\frac{H_1^{(1)}(k\tilde{\rho}_j(t_k,x,y))}{\tilde{\rho}_j(t_k,x,y)}] \quad (36)$$

where $\tilde{\rho}_j^2(t,x,y)=(x-(j-1)p)^2+(y-d(t+1)/2)^2$.

II. Results and Discussion

In this paper we focus specifically on the details of the mathematical simulation of the comb-like nanostrip gratings. Besides of that, in this section we present some preliminary results on the plasmon-type and grating-type resonance effects of the associated comb-like nanostrip gratings and their dependence on the number of strips *N*.

To characterize the optical properties of considered gratings, we have computed the normalized TSCS and ACS as a function of the wavelength in the visible range.

In Fig. 2, the above-mentioned scattering and absorption spectra are presented for the single strip (N = 1) under the normal incidence (β = 0) and for the gratings of N = 2,5,10 and 20 identical silver strips of the width d = 300 nm, thickness h = 50 and period p = 500 nm. Note that for the gratings the angle β = 0 corresponds to the grazing incidence with respect to the plane of the grating.

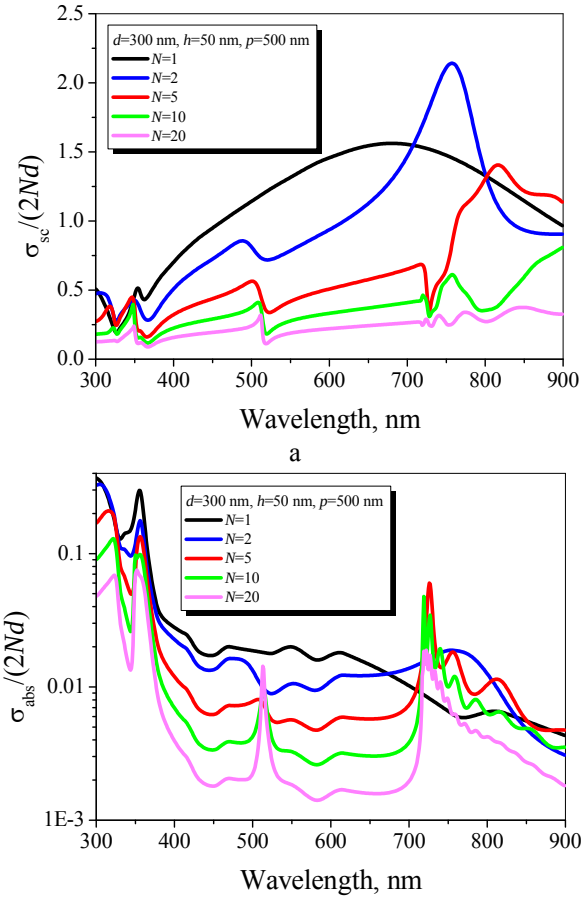


Fig. 2. Normalized TSCS (a) and ACS (b) as a function of the wavelength for N = 1,2,5,10 and 20 silver strips of the width d=300 nm, thickness h = 10 nm and period p = 500 nm; the discretization order is n = 20.

As one can see, if the number of the strips gets larger, the surface plasmon resonances deform and became less pronounced in contrast to the coplanar silver strip gratings [10,11], while the grating resonances near $\lambda = p$ become sharper.

In Fig. 3 and 4, respectively, the total near-field and the scattered far-field patterns for the single strip (N = 1) at λ = 680.01 nm and for two strips (N = 2) at λ = 488.5 nm and λ = 757.02 nm, are presented.

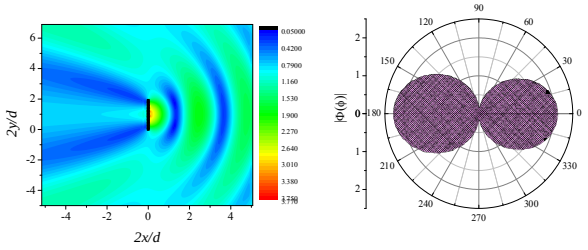


Fig. 3. Total magnetic near-field and scattered far-field pattern at the wavelength λ = 680.01 nm for stand-alone silver strip (black curve) of d=300nm, h = 50 nm

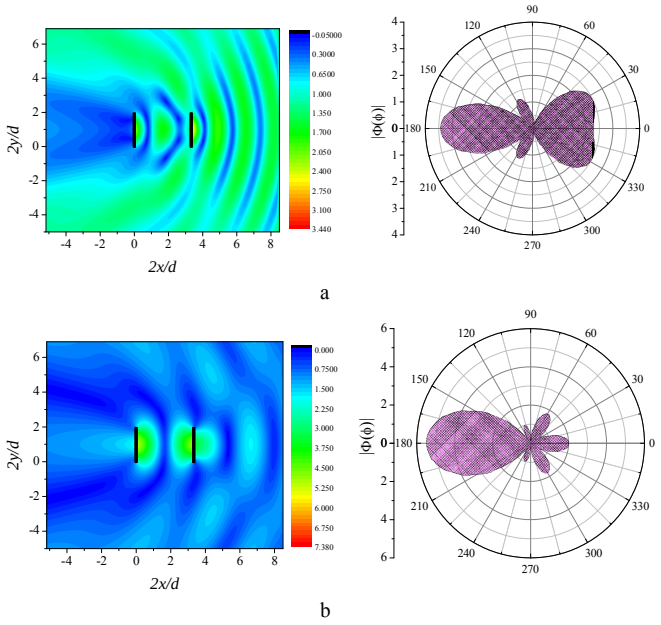


Fig. 4. Total magnetic near-field and scattered far-field pattern at the wavelengths λ= 488.5 nm (a) and λ= 757.02 nm (b) for a two-strip silver grating (N = 2) strip of d=300nm, h = 50 nm and p = 500 nm

In Fig. 5, the total near-field and the scattered field patterns are presented for the grating of N = 20 of the same parameters as in Fig. 4 at grating resonance λ = 511.02 nm.

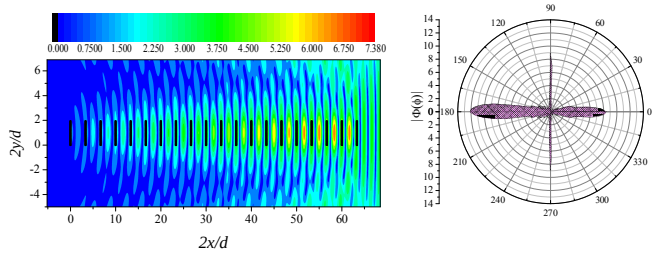


Fig. 5. The same as in Fig. 4 but for the grating of N = 20 silver strips at the grating resonance λ = 511.02 nm

ACKNOWLEDGMENT

Support of the National Academy of Sciences of Ukraine via the State Target Program "Nanotechnologies and Nanomaterials" and the European Science Foundation via RNP "Plasmon-Nanobiosense" is acknowledged. The first author is also grateful to the International Visegrad Fund for the support in the form of Ph.D. scholarship.

REFERENCES

[1] G. Schider, J.R. Krenn, W. Gotschy, B. Lambrecht, H. Ditlbacher, A. Leitner, and F.R. Aussenegg, "Optical properties of Ag and Au nanowire gratings," *J. Appl. Phys.*, vol. 90, no 8, pp. 3825-3830, 2001.

[2] V. Giannini and J. A. Sànchez-Gil, "Calculations of light scattering from isolated and interacting metallic nanowires of arbitrary cross section by means of Green's theorem surface integral equations in parametric form," *J. Opt. Soc. Am. A*, vol. 24, pp. 2822-2830, 2007.

[3] T. Søndergaard, "Modeling of plasmonic nanostructures: Green's function integral equation methods," *Phys. Stat. Sol. (b)*, vol. 244, pp. 3448-3462, 2007.

[4] G. Della Valle, T. Søndergaard, and S. I. Bozhevolnyi, "Plasmon-polariton nano-strip resonators: from visible to infra-red," *Opt. Exp.*, vol. 16, no 10, pp. 6867-6876, 2008.

[5] M.R. Gadsdon, I.R. Hooper, and J.R. Sambles, "Optical resonances on sub-wavelength lamellar gratings," Opt. Exp., vol. 16, no 26, pp. 22003-22028, 2008.

[6] P. B. Johnson and R. W. Christy, "Optical constants of the noble metals," *Physical Review B*, vol. 6, pp. 4370-4379, 1972.

[7] O. V. Shapoval, R. Sauleau, and A. I. Nosich. "Scattering and absorption of waves by flat material strips analyzed using generalized boundary conditions and Nystrom-type algorithm," *IEEE Trans. Antennas Propag.*, vol. 59, no 9, pp. 3339-3346, 2011.

[8] E. Bleszynski, M. Bleszynski, and T. Jaroszewicz, "Surface-integral equations for electromagnetic scattering from impenetrable and penetrable sheets," *IEEE Antennas Propag. Mag.*, vol. 35, pp. 14-25, 1993.

[9] O.V. Shapoval, R. Sauleau, A.I. Nosich, "Grating and plasmon resonances in the scattering of light by finite silver nanostrip gratings," *Int. J. Semicond. Physics, Quantum Electronics, Optoelectronics*, vol. 15, no 3, pp. 200-203, 2012.

[10] O.V. Shapoval, R. Sauleau, A.I. Nosich, "Modeling of plasmon resonances of multiple flat noble-metal nanostrips with a median-line integral equation technique," *IEEE Trans. Nanotechnology*, vol. 12, 2013, will be published.

[11] O.V. Shapoval, J. S. Gomez-Diaz, J. Perruisseau-Carrier, J. Mosig, A.I. Nosich, "Integral equation analysis of plane wave scattering by finite periodic gratings made of coplanar graphene strips in the THz frequency range," *IEEE Trans. Terahertz Science and Technology*, 2012, submitted.

Modeling of 2-D Microcavity Lasers with Arbitrarily Shaped Active Regions

Elena I. Smotrova,[1] Melanie Lebental,[2] and Alexander I. Nosich[1]

[1] Laboratory of Micro and Nano-Optics, Institute of Radio-Physics and Electronics NASU, Kharkiv 61085, Ukraine
[2] Laboratory of Quantum and Molecular Photonics, Ecole Normale Superieur de Cachan, Cachan, 94235, France

***Abstract*—We consider the mathematical model for the accurate and economic numerical study of an arbitrary-shape two-dimensional (2-D) microcavity laser with an active region of arbitrary shape. The model is based on the coupled Muller integral equations (IE) on the boundaries of passive and active parts of the cavity. The algorithm of their discretization can be based on the Nystrom interpolation of the unknown functions corresponding to the boundary electric and magnetic currents. The model enables one to extract the lasing-mode frequencies and thresholds and to use the geometry of the active region as a tool of manipulation of the lasing threshold.**

Keywords— active region, laser, microcavity, threshold gain, Muller boundary integral equation, Nystrom discretization

I. Introduction

Thin flat microcavity lasers are known as "2-D microcavity lasers" because, thanks to the effective refractive index concept, their modes can be considered in the cavity plane only [1,2]. The most important parameter of lasing is the threshold of pump power, or, in fact, of the material gain. Circular microdisks are famous for the ultra-low (exponentially low) thresholds of lasing for the whispering-gallery modes however the disk modes have low directionality of emission in the disk plane [3-6]. In applications, normally one seeks the sources having higher directionality that helps process the light signals with higher efficiency in a smaller volume. Still to keep the threshold reasonably low, the shapes of laser resonators that provide better directionality should not depart too far from the circle. This consideration has led to the investigation of such shapes as stadium, limacon and spiral [7-11]. In general, the shape of the 2-D cavity contour is widely used as an instrument of modification and control of the directionality of laser emission [1,2,12-15].

In analogous manner, the shape, size and location of the active region can serve as a tool for manipulation and control of the lasing threshold. This is because the threshold is inverse proportional to the overlap coefficient between the mode E-field and active region that provides gain [16]. In practice, the gain overlap can be varied using the focused pump beam (if the cavity is uniformly active and a photo-pumping is applied) or controlling the shape and location of electrodes (if the carrier injection is exploited) [17].

In this paper, we consider the mathematical and numerical model for the analysis of the lasing modes of an arbitrary 2-D dielectric resonator with an active region of arbitrary shape. Our model is based on the Muller boundary IEs [18,19]. As known, all natural modes of such a cavity split to two polarization classes, E and H, according to the field component directed along the *z*-axis. For either of polarizations, the 2-D eigenvalue problem for a time-harmonic electromagnetic field can be reduced to the Muller IE using the Green's formulas. Muller IE is actually a pair of coupled equations, which are together spectrally (in mathematical sense) equivalent to the original boundary-value problem for the Maxwell equations with boundary and radiation conditions. This means that, unlike many non-Muller IEs, they have no spurious eigenvalues.

I. Mathematical Simulation

A. Geometry Description

Fig. 1 shows generic geometry of a partially active 2-D dielectric resonator and corresponding notations.

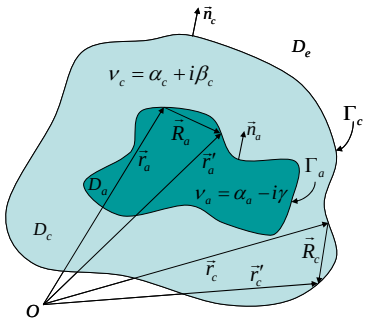


Fig.1. Generic geometry of the 2-D dielectric microcavity laser with arbitrary active region shown as a darker domain inside the cavity.

Denote the outer domain as D_e and the resonator domain as $D_i = D_c \cup D_a$, where domain D_a is the active region and D_c is the domain filled with a passive dielectric, so that $D_a \subseteq D_c$. Denote also the outer boundary of the passive dielectric as Γ_c, and the boundary of the active region as Γ_a. Assume that the unit vectors $\vec{n}_c$ and $\vec{n}_a$ are the outer normal vectors to the dielectric and active-region boundaries, respectively, and

$\nu_a = \alpha_a - i\gamma$, $\nu_c = \alpha_c + i\beta\ (\beta > 0)$ and $\nu_e = \alpha_e$ are the refractive indices of the materials of the corresponding domains. Here, $\gamma > 0$ is the threshold value of material gain needed to compensate for the modal losses that consist of the radiation and, if $\beta > 0$, the absorption in the domain D_c.

B. Derivation of Coupled Muller IEs

Suppose that $U_j = E_z$ or H_z in the domain D_j, $j = a,c,e$ is the solution to the problem that involves Helmholtz equations the corresponding domains and boundary conditions on their boundaries,

$$\left(\Delta + k_j^2\right)U_j(\vec{r}) = 0, \quad \vec{r} \in D_j, \tag{1}$$

$$U_a(\vec{r}_a) = U_c(\vec{r}_a), \quad \vec{r}_a \in \Gamma_a, \tag{2}$$

$$\eta_a \frac{\partial U_a(\vec{r}_a)}{\partial n_a} = \eta_c \frac{\partial U_c(\vec{r}_a)}{\partial n_a}, \quad \vec{r}_a \in \Gamma_a, \tag{3}$$

$$U_c(\vec{r}_c) = U_e(\vec{r}_c), \quad \vec{r}_c \in \Gamma_c, \tag{4}$$

$$\eta_c \frac{\partial U_c(\vec{r}_c)}{\partial n_c} = \eta_e \frac{\partial U_e(\vec{r}_c)}{\partial n_c}, \quad \vec{r}_c \in \Gamma_c, \tag{5}$$

and also the condition of the local finiteness of power and the Sommerfeld radiation condition for U_e at infinity in the outer domain D_e. The coefficients in (1) are $k_j = k\nu_j$. The conditions (2)–(5) are the demands that the tangential field components are continuous. The associated constants are as follows: $\eta_j = 1/\nu_j^2$ in the case of H-polarization and $\eta_j = 1$ in the case of E-polarization. Our goal is to find the pairs of real-valued parameters *k* (the frequency) and γ (the threshold material gain) that generate non-zero functions *U*, i.e. the eigenvalues and eignenfunctions of the formulated above Lasing Eigenvalue Problem (LEP). Details of the LEP approach and general properties of the LEP solutions have been discussed in [1,16].

Introduce three Green's functions of the homogeneous media with refractive indices ν_j that satisfy 2-D Helmholtz equations with Dirac deltas in the right-hand parts,

$$(\Delta + k_j^2)G_j(\vec{r},\vec{r}') = -\delta(\vec{r},\vec{r}') \tag{6}$$

As known, they are given by

$$G_j(\vec{r},\vec{r}') = (i/4)H_0^{(1)}(k_j R), \tag{7}$$

where $R = |\vec{r} - \vec{r}'|$ is the distance between the points $\vec{r}$ and $\vec{r}'$, and $H_0^{(1)}(\cdot)$ is the Hankel function of the first kind. Then, applying the first Green's formula to the functions $G_j(\vec{r},\vec{r}')$ and U_j and implying that $\vec{r} \in D_a$, we obtain (see [18,19]),

$$U_a(\vec{r}) = -\int_{\Gamma_a}\left(U_a(\vec{r}')\frac{\partial G_a(\vec{r},\vec{r}_a')}{\partial n_a'} - G_a(\vec{r},\vec{r}_a')\frac{\partial U_a(\vec{r}_a')}{\partial n_a'}\right)dl_a', \tag{8}$$

$$U_c(\vec{r}) = \int_{\Gamma_a}\left(U_c(\vec{r}_a')\frac{\partial G_c(\vec{r},\vec{r}_a')}{\partial n_a'} - G_c(\vec{r},\vec{r}_a')\frac{\partial U_c(\vec{r}_a')}{\partial n_a'}\right)dl_a' - \int_{\Gamma_c}\left(U_c(\vec{r}_c')\frac{\partial G_c(\vec{r},\vec{r}_c')}{\partial n_c'} - G_c(\vec{r},\vec{r}_c')\frac{\partial U_c(\vec{r}_c')}{\partial n_c'}\right)dl_c', \vec{r} \in D_c \tag{9}$$

$$U_e(\vec{r}) = \int_{\Gamma_c}\left(U_e(\vec{r}_c')\frac{\partial G_e(\vec{r},\vec{r}_c')}{\partial n_c'} - G_e(\vec{r},\vec{r}_c')\frac{\partial U_e(\vec{r}_c')}{\partial n_c'}\right)dl_c', \tag{10}$$

where $dl_{a,c}'$ is elementary arc of the corresponding contour.

Denote $\varphi_a(\vec{r}_a) = U_a(\vec{r}_a)$ and $\psi_a(\vec{r}_a) = \partial U_a(\vec{r}_a)/\partial n_a$, $\vec{r}_a \in \Gamma_a$, and also $\varphi_c(\vec{r}_c) = U_e(\vec{r}_c)$ and $\psi_c(\vec{r}_c) = \partial U_e(\vec{r}_c)/\partial n_c$, $\vec{r}_c \in \Gamma_c$. These functions have the meaning of equivalent electric and magnetic currents on the respective contours. Then from the boundary conditions (2) and (3) it follows that $U_c(\vec{r}_a) = \varphi_a(\vec{r}_a)$ and $\partial U_c(\vec{r}_a)/\partial n_a = (\eta_a/\eta_c)\psi_a(\vec{r}_a)$, $\vec{r}_a \in \Gamma_a$, while (4) and (5) yield $U_e(\vec{r}_c) = \varphi_c(\vec{r}_c)$ and $\partial U_e(\vec{r}_c)/\partial n_c = (\eta_c/\eta_e)\psi_c(\vec{r}_c)$, $\vec{r}_c \in \Gamma_c$.

Further, in (8) and (9) place the point $\vec{r}$ on the contour Γ_a. Using the continuity of the single-layer potential across the boundary [20], add the resulting expressions to each other and derive the first IE as

$$\varphi_a(\vec{r}_a) + \int_{\Gamma_a}\left[\varphi_a(\vec{r}_a')A_a(\vec{r}_a,\vec{r}_a') - \psi_a(\vec{r}_a')B_a(\vec{r}_a,\vec{r}_a')\right]dl_a' - \int_{\Gamma_c}\left[\varphi_c(\vec{r}_c')A_c(\vec{r}_a,\vec{r}_c') - \psi_c(\vec{r}_c')B_c(\vec{r}_a,\vec{r}_c')\right]dl_c' = 0. \tag{11}$$

On placing the point $\vec{r}$ on the contour Γ_a, differentiating (8) and (9) in $\vec{n}$, taking account of the properties of the normal derivatives of the single and double-layer potentials [20], add the results to each other and derive the second IE as

$$\frac{\eta_a + \eta_c}{2\eta_c}\psi_a(\vec{r}_a) + \int_{\Gamma_a}\left[\varphi_a(\vec{r}_a')C_a(\vec{r}_a,\vec{r}_a') - \psi_a(\vec{r}_a')D_a(\vec{r}_a,\vec{r}_a')\right]dl_a' - \int_{\Gamma_c}\left[\varphi_c(\vec{r}_c')C_c(\vec{r}_a,\vec{r}_c') - \psi_c(\vec{r}_c')D_c(\vec{r}_a,\vec{r}_c')\right]dl_c' = 0 \tag{12}$$

By analogy, considering the equations (9) and (10) with the point $\vec{r}$ placed on the contour Γ_c, we obtain a similar pair of the boundary IEs or the third and the fourth equations,

$$\varphi_c(\vec{r}_c) - \int_{\Gamma_a}\left[\varphi_a(\vec{r}_a')A_a(\vec{r}_c,\vec{r}_a') - \psi_a(\vec{r}_a')B_a(\vec{r}_c,\vec{r}_a')\right]dl_a' + \int_{\Gamma_c}\left[\varphi_c(\vec{r}_c')A_c(\vec{r}_c,\vec{r}_c') - \psi_c(\vec{r}_c')B_c(\vec{r}_c,\vec{r}_c')\right]dl_c' = 0. \tag{13}$$

$$\frac{\eta_c + \eta_e}{2\eta_e}\psi(\vec{r}_c) - \int_{\Gamma_a}\left[\varphi_a(\vec{r}_a')C_a(\vec{r}_c,\vec{r}_a') - \psi_a(\vec{r}_a')D_a(\vec{r}_c,\vec{r}_a')\right]dl_a' + \int_{\Gamma_c}\left[\varphi_c(\vec{r}_c')C_c(\vec{r}_c,\vec{r}_c') - \psi_c(\vec{r}_c')D_c(\vec{r}_c,\vec{r}_c')\right]dl_c' = 0. \tag{14}$$

Equations (11) to (14) form the set of the coupled Muller IEs for two contours. Their kernel functions are given by

$$A_a(\vec{r}_a,\vec{r}'_a)=\frac{\partial G_a(\vec{r}_a,\vec{r}'_a)}{\partial n'_a}-\frac{\partial G_c(\vec{r}_a,\vec{r}'_a)}{\partial n'_a},$$
$$A_c(\vec{r}_c,\vec{r}'_c)=\frac{\partial G_c(\vec{r}_c,\vec{r}'_c)}{\partial n'_c}-\frac{\partial G_e(\vec{r}_c,\vec{r}'_c)}{\partial n'_c} \quad (15)$$

$$B_a(\vec{r}_a,\vec{r}'_a)=G_a(\vec{r}_a,\vec{r}'_a)-\frac{\eta_a}{\eta_c}G_c(\vec{r}_a,\vec{r}'_a),$$
$$B_c(\vec{r}_c,\vec{r}'_c)=G_c(\vec{r}_c,\vec{r}'_c)-\frac{\eta_c}{\eta_e}G_e(\vec{r}_c,\vec{r}'_c), \quad (16)$$

$$C_a(\vec{r}_a,\vec{r}'_a)=\frac{\partial^2 G_a(\vec{r}_a,\vec{r}'_a)}{\partial n_a\partial n'_a}-\frac{\partial^2 G_c(\vec{r}_a,\vec{r}'_a)}{\partial n_a\partial n'_a},$$
$$C_c(\vec{r}_c,\vec{r}'_c)=\frac{\partial^2 G_c(\vec{r}_c,\vec{r}'_c)}{\partial n_c\partial n'_c}-\frac{\partial^2 G_e(\vec{r}_c,\vec{r}'_c)}{\partial n_c\partial n'_c}, \quad (17)$$

$$D_a(\vec{r}_a,\vec{r}'_a)=\frac{\partial G_a(\vec{r}_a,\vec{r}'_a)}{\partial n_a}-\frac{\eta_a}{\eta_c}\frac{\partial G_c(\vec{r}_a,\vec{r}'_a)}{\partial n_a},$$
$$D_c(\vec{r}_c,\vec{r}'_c)=\frac{\partial G_a(\vec{r}_c,\vec{r}'_c)}{\partial n_c}-\frac{\eta_c}{\eta_e}\frac{\partial G_e(\vec{r}_c,\vec{r}'_c)}{\partial n_c} \quad (18)$$

Here the Green's functions are given by (4) and their normal derivatives are found as $(j=a,c,e;\ i=a,c)$

$$\frac{\partial G_j(\vec{r}_i,\vec{r}'_i)}{\partial n'_i}=\frac{i}{4}k_jH_1^{(1)}(k_jR_i)\frac{(\vec{R}_i\cdot\vec{n}'_i)}{R_i}, \quad (19)$$

$$\frac{\partial G_j(\vec{r}_i,\vec{r}'_i)}{\partial n_i}=-\frac{i}{4}k_jH_1^{(1)}(k_jR_i)\frac{(\vec{R}_i\cdot\vec{n}_i)}{R_i}, \quad (20)$$

$$\frac{\partial^2 G_j(\vec{r}_i,\vec{r}'_i)}{\partial n_i\partial n'_i}=-\frac{ik_j}{4}\times$$
$$\left[k_jH_2^{(1)}(k_jR_i)(\vec{R}_i\cdot\vec{n}'_i)\frac{(\vec{R}_i\cdot\vec{n}_i)}{R_i^2}-H_1^{(1)}(k_jR_i)\frac{(\vec{n}'_i\cdot\vec{n}_i)}{R_i}\right], \quad (21)$$

In (19) to (21), $(\vec{R}_i\cdot\vec{n}_i)$, $(\vec{R}_i\cdot\vec{n}'_i)$ and $(\vec{n}'_i\cdot\vec{n}_i)$ are the scalar products. Suppose that the parameterizations of contour Γ_i is done using the regular analytic 2π -periodic functions $r_i(t)=\{x_i(t),y_i(t)\}$, $t\in[0,2\pi]$. Then the distance between two points on either curve is

$$R_i^2=\left[x_i(t)-x_i(\tau)\right]^2+\left[y_i(t)-y_i(\tau)\right]^2, \quad (22)$$

and the outer normal vector to the curve Γ_i and the Jacobian are given by the following expressions:

$$\vec{n}_i=\left(\frac{dy_i}{dt};-\frac{dx_i}{dt}\right)\frac{1}{L_i(t)}, \quad (23)$$

$$L_i(t)=\sqrt{(dx_i/dt)^2+(dy_i/dt)^2} \quad (24)$$

C. Regular and Singular Kernels

As one can see, the kernels of the Muller IEs are linear combinations of the Green's functions of homogeneous media with material parameters of the mentioned above three domains or their normal derivatives of the first and the second orders.

From the equations of (15) to (18) we derive that the kernels $A_i(t,\tau)$ and $D_i(t,\tau)$ are continuous in all points of a smooth contour and the kernels $B_i(t,\tau)$ and $C_i(t,\tau)$ have logarithmic singularities. The continuity of the functions $A_i(t,\tau)$ and $D_i(t,\tau)$ is conditioned by the limit expression (see [21,22]):

$$\frac{\partial G_i(t,t)}{\partial n_i}=\frac{1}{2}\chi_i(t),\quad i=a,c \quad (25)$$

where $\chi_i(t)$ is the curvature ([22], p. 564) of the contour Γ_i at the point $\vec{r}$.

The limit values of the singular kernels at $t\to\tau$ are given by the following formulas:

$$B_a(t,\tau)\underset{t\to\tau}{\cong}-\frac{1}{2\pi}\left(1-\frac{\eta_a}{\eta_c}\right)\ln R_a,$$
$$B_c(t,\tau)\underset{t\to\tau}{\cong}-\frac{1}{2\pi}\left(1-\frac{\eta_c}{\eta_e}\right)\ln R_c, \quad (26)$$

$$C_a(t,\tau)\underset{t\to\tau}{\cong}-\frac{1}{2\pi}\left(k_a^2-k_c^2\right)\ln R_a,$$
$$C_c(t,\tau)\underset{t\to\tau}{\cong}-\frac{1}{2\pi}\left(k_c^2-k_e^2\right)\ln R_c \quad (27)$$

Note that, according to (26), the kernels $B_i(t,\tau)$ of (16) have logarithmic singularities in the case of H-polarization only, while in the E-polarization case they are continuous. In either case the Muller IEs generate a Fredholm second kind operator equation. They are fully equivalent to the original boundary-value problem.

II. Remarks on Numerical Algorithm

There are several options for efficient discretization of the Muller IEs. Besides of simple collocations and traditional Galerkin techniques, today the most economic discretization technique is known as Nystrom method [23-26]. This method is based on the polynomial approximation of smooth unknown functions and the replacement of the integrals with sums using the certain quadrature formulas. These formulas have to account correctly for all singularities of integrand functions.

In our analysis, it is convenient to represent each of the kernel functions of IEs (even smooth ones) with extracted logarithmic singularity,

$$F(t,\tau)=F_1(t,\tau)\ln\left(4\sin^2\frac{t-\tau}{2}\right)+F_2(t,\tau) \quad (28)$$

Further, we introduce equidistant meshes of nods on the contours Γ_i at $t_p=\pi p/N$, $p=0,1,...,2N-1$. The integrals corresponding of the each part of (28) are replaced with the sums using the appropriate quadrature formulas. For the first parts we use the quadrature formula derived through the approximation of the integrand function with trigonometric

polynomials. For the second parts of kernels we use the trapezoidal rule.

On the replacement, in the Muller IEs, of the integrals with these quadratures, we arrive at the matrix equation of the total size $8N \times 8N$. In practical studies, it is convenient to introduce dimensionless frequency parameter $\kappa = ka$ where a is some characteristic dimension of the 2-D cavity. Then the lasing eigenvalues (κ, γ) are the roots of determinantal equation,

$$\mathrm{Det}\left[\mathbf{I} + \mathbf{A}(\kappa, \gamma)\right] = 0, \qquad (29)$$

where $\mathbf{I}$ is an identity operator and $\mathbf{A}$ is the operator corresponding to the integrals (11) – (14). These roots are found numerically using the iterative-type secant method. The number of iterations strongly depends on the nearness of initial guess to the true eigenvalue. Considering the accuracy of computations, we can establish that, according to [20] that if the integrand function is analytic and 2π-periodic, then the error of the interpolation algorithm has the order of $O\left[\exp(-\sigma N)\right]$, where $2N$ is the order of the interpolation polynomial and σ is a constant.

As a first step in the experimental verification of developed theory, we have performed measurements of the planar kite-shaped organic lasers with uniform optical pumping. This study has demonstrated good agreement between the emission patterns and in the relative values of the lasing thresholds for different types of modes.

ACKNOWLEDGMENT

Support of the National Academy of Sciences of Ukraine via the State Target Program “Nanotechnologies and Nano-materials” and the European Science Foundation via the Research and Networking Programmes “Newfocus” and “Plasmon-Nanobiosense” is acknowledged with gratitude.

REFERENCES

[1] A.I. Nosich, E.I. Smotrova, S.V. Boriskina, T.M. Benson, and P. Sewell, "Trends in microdisk laser research and linear optical modeling," *Optical and Quantum Electronics*, vol. 39, no 15, pp. 1253-1272, 2007.

[2] T. Harayama and S. Shinohara, “Two-dimensional microcavity lasers,” *Laser Photonics Rev*., vol. 5, no 2, pp. 247-281, 2011.

[3] S. L. McCall, A. F. J. Levi, R. E. Slusher, S. J. Pearson, and R. A. Logan, “Whispering-gallery mode microdisk lasers,” *Appl. Phys. Lett.*, vol. 60, no. 3, pp. 289–229, 1992.

[4] R. E. Slusher, A. F. J. Levi, U. Mohideen, S. L. McCall, S. J. Pearson, and R. A. Logan, “Threshold characteristics of semiconductor microdisk lasers,” *Appl. Phys. Lett.*, vol. 63, no. 10, pp. 1310–1312, 1993.

[5] E.I. Smotrova and A.I. Nosich, "Mathematical study of the two-dimensional lasing problem for the whispering-gallery modes in a circular dielectric microcavity," *Optical and Quantum Electronics*, vol. 36, no 1-3, pp. 213-221, 2004.

[6] E.I. Smotrova, A.I. Nosich, T.M. Benson, and P. Sewell, "Cold-cavity thesholds of microdisks with uniform and non-uniform gain: quasi-3D modeling with accurate 2D analysis," *IEEE J. Selected Topics in Quantum Electronics*, vol. 11, no 5, pp. 1135-1142, 2005.

[7] S.-K. Kim, S.-H. Kim, G.-H., Kim, H.-G. Park, D.-J. Shin, Y.-H. Lee, “Highly directional emission from few-micron-size elliptical microdisks,” *Appl. Phys. Lett*., vol. 84, no 6, pp. 861–863, 2004.

[8] Q.H. Song, L. Ge, A.D. Stone, H. Cao, J. Wiersig, J.-B. Shin, J. Unterhinninghofen, W. Fang, and G.S. Solomon, “Directional laser emission from a wavelength-scale chaotic microcavity,” *Phys. Rev. Letts*., vol 105, pp. 103902, 2010.

[9] M. Kneissl, M. Teepe, N. Miyashita, N.M. Johnson, G.D. Chern, and R.K. Chang, “Current-injection spiral shaped microcavity disk laser diodes with unidirectional emission,” *Appl. Phys. Lett*., vol. 84, no 14, pp. 2485–2487, 2004.

[10] T. Ben-Massaoud and J. Zyss, “Unidirectional laser emission from polymer-based spiral microdisks,” *Appl. Phys. Lett*., vol. 86, pp. 241110, 2005.

[11] M. Hentschel and T.-Y. Kwon, “Designing and understanding directional emission from spiral microlasers.” *Opt. Letts.*, vol. 34, pp. 163-165, 2009.

[12] H.G.L. Schwefel, H.E. Tureci, A.D. Stone, and R.K. Chang, “Progress in asymmetric resonant cavities; using shape as a design parameter in dielectric microcavity lasers,” in K. Vahala (Ed.), *Optical Microcavities*, Singapore, World Scientific, pp. 415-496, 2004.

[13] T.M. Benson, S.V. Boriskina, P. Sewell, A. Vukovic, S. Greedy, and A.I. Nosich, "Micro-optical resonators for microlasers and integrated optoelectronics," in S. Janz, J. Ctyroky and S. Tanev (Eds.), *Frontiers of Planar Lightwave Circuit Technology: Design, Simulation and Fabrication*, Berlin, Springer, pp. 40-70, 2005.

[14] R. Dubertrand, E. Bogomolny, N. Djellali, M. Lebental, and C. Schmit, “Circular dielectric cavity and its deformations,” *Phys. Rev. A*., vol. 77, no 1, pp. 013804(16), 2008.

[15] Q.J. Wang, C. Yan, N. Yu, J. Unterhinninghofen, J. Wiersig, C. Pflügl, L. Diehl, T. Edamura, M. Yamanishi, H. Kan, and F. Capasso, “Whispering-gallery mode resonators for highly unidirectional laser action,“ *Proc. Nat. Ac. Sci*., vol. 107, no 52, pp. 22407-22412, 2010.

[16] E.I. Smotrova, V.O. Byelobrov, T.M. Benson, J. Ctyroky, R. Sauleau, and A.I. Nosich, "Optical theorem helps understand thresholds of lasing in microcavities with active regions," *IEEE J. Quantum Electronics*, vol. 47, no 1, pp. 20-30, 2011.

[17] N. B. Rex, R. K. Chang, and L. J. Guido, “Threshold lowering in GaN micropillar lasers by means of spatially selective optical pumping,” *IEEE Photon. Technol. Lett.*, vol. 13, no. 1, pp. 1–3, 2001.

[18] C. Muller, *Foundations of the Mathematical Theory of Electromagnetic Waves*, Berlin, Springer, 1969.

[19] V. Rokhlin, “Rapid solution of integral equations of scattering theory in two dimensions,” *J. Comput. Phys.*, vol. 86, no.2, pp. 414–439, 1990.

[20] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, Berlin, Springer, 1998.

[21] A.N. Tikhonov and A.A. Samarskii, *Equations of Mathematical Physics*, New York, Dover Publ., 1990.

[22] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers: Eefinitions, Theorems, and Formulas for Reference and Review*, Dover Publ., 2000 (reprinted from the 1968 edition).

[23] Z.T. Nazarchuk, *Numerical Analysis of Wave Diffraction by Cylindrical Structures*, Kiev, Naukova Dumka Publ., 1989 (in Russian).

[24] L. Wang, J.A. Cox, and A. Friedman, “Modal analysis of homogeneous optical waveguides by the boundary integral formulation and Nystrom method,” *J. Opt. Soc. Am. A*., vol. 15, no 1, pp. 92–100, 1998.

[25] J.L. Tsalamengas, “Exponentially converging Nystrom methods applied to the integral-integrodifferential equations of oblique scattering / hybrid mode propagation in presence of composite dielectric cylinders of arbitrary cross-section,” *IEEE Trans. Antennas Propagat*., vol. 55, no11, pp. 3239-3250, 2007.

[26] M.V. Balaban, E.I. Smotrova, O.V. Shapoval, V.S. Bulygin, and A.I. Nosich, "Nystrom-type techniques for solving electromagnetics integral equations with smooth and singular kernels," *Int. J. Numerical Modeling: Electronic Networks, Devices and Fields*, vol. 25, no 5, pp. 490-511, 2012.

Electromagnetic Characterization of a Silver Nanowire Grating Embedded into a Dielectric Slab

Volodymyr O. Byelobrov,[1] Trevor M. Benson,[2] and Alexander I. Nosich[1]

[1] Laboratory of Micro and Nano-Optics, Institute of Radio-Physics and Electronics NASU, Kharkiv 61085, Ukraine
[2] The George Green Institute for Electromagnetics Research, The University of Nottingham, NG7 2RD, Nottingham, UK

***Abstract*—The electromagnetic properties of a silver nanowire grating embedded into a dielectric slab are investigated, using a Maxwellian formalism, to determine and distinguish effects caused by the structure periodicity and by the presence of silver wires which entail plasmon resonances. The interaction of these two effects gives an opportunity to control the qualitative characteristics of the plasmon resonance while the variation of the slab refractive index allows managing its wavelength.**

Keywords—grating; plasmon resonance; sub-wavelength wires

I. INTRODUCTION

A growing interest of today's nanooptics is concentrated on devices that use plasmonic effects on the nanometer scale, with a wide range of applications. Truly nanoscale lasers were recently demonstrated in the form of random ensembles of spherical gold particles, each of tens of nanometers in diameter and coated with a spherical layer of dye-doped silica [1]. Because the output power of such lasers generally is commensurable with device size and as the noble-metal cores or shells have considerable losses, one can foresee that elementary nanolasers should be assembled in groups or arrays. Such an arrangement brings new physical phenomena in the form of structure resonances caused by the periodicity.

A theoretical model of a grating of infinite dielectric cylinders placed periodically in free space and illuminated by a plane wave has been investigated as a fundamental electromagnetic problem since the 1950s – see, for instance, [2,3]. The high-Q resonances that occur near the "Rayleigh anomalies" were found numerically some time ago as can be concluded, for instance, from Figs. 2 and 3 in [3], however no attempt was done to explain them before analytical analysis was performed in [4].

In [5], we have found that a quantum-wire grating possesses specific grating modes in addition to the more conventional natural modes of a single dielectric cylinder perturbed by the presence of other cylinders. These modes display counter-intuitive properties: (i) the resonance Q-factor gets higher if the grating gets sparser, i.e. for the larger ratios of period to cylinder radius and (ii) mode near-field is a characteristic wave standing along the grating and stretching over many periods in the normal to the grating direction [6]. The assumed infinite extent of the structure, being an idealisation of real scatterers, leads to good agreement with the scattering by a finite grating of many wires, as was shown for a grating of silver cylinders [7]. Additionally the silver material brings the plasmon resonances, which are distinguished by their quality-factors and linkage to the particular frequency,

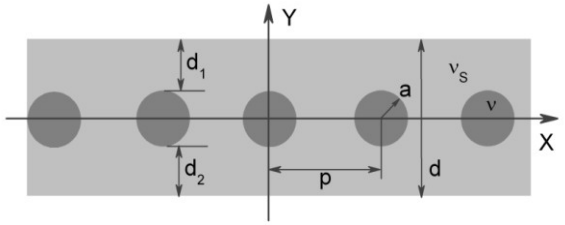


Fig.1 The cross-sectional geometry of a periodically structured dielectric slab.

In practical applications, gratings are frequently embedded in dielectric slabs. A slab with embedded nanowire grating of circular cylinders (Fig. 1) obviously possesses the above mentioned properties plus can support slab modes, now perturbed by the nanowires. We build an accurate model of such structure to reveal the impact of physical and geometrical parameters of the slab on the grating modes. A similar wave-scattering problem for a purely dielectric structure was investigated in [8], however here we will concentrate not on the scattering properties of the composite structure but rather on the interaction between the plasmon modes and the grating modes and its dependence on the slab refractive index. We will also emphasize the possibility of enhancement of the slab electromagnetic properties of the plasmon modes by changing geometrical parameters of the slab.

II. FORMULATION AND BASIC EQUATIONS

Consider a grating of circular nanowires lying parallel to the *z*-axis and periodic along the *x*-axis. The grating period is p, each silver nanowire is a cylinder of radius a, and its refractive index ν is taken from the Johnson and Christy table [9]. This grating is embedded in a dielectric slab, whose refractive index is ν_s and thickness is $d = d_1 + d_2 + 2a$, where $d_{1,2}$ are distances between the grating and the upper and lower slab surfaces. Suppose that the electromagnetic field is time-harmonic as $\sim \exp(-i\omega t)$ and does not vary along the *z*-axis, hence the field analysis problem is two-dimensional. Illumination by a plane wave incident from the upper half-space (in the negative direction of the *y*-axis) is investigated.

The function U, denoting the H_z component of the electromagnetic field in the *H*-polarization case and the E_z in the *E*-polarization case, must satisfy the Helmholtz equation with appropriate wavenumber inside and outside of the cylinders, the Sveshnikov radiation condition at infinity [5-7], the condition of local integrability of power, and the boundary conditions demanding continuity of the tangential field components at cylinder boundaries. The free space wavenumber is $k = \omega / c = 2\pi / \lambda$, where c is the free-space light velocity and λ is the wavelength, while inside the cylinders it is $k\nu$ and inside the slab $k\nu_s$. In this paper we will also use the normalized frequency as the ratio of period to wavelength, $\sigma = p / \lambda = k\nu_s p / 2\pi$.

The field inside the slab but outside the grating ($a < |y| < a + d_{1,2}$) and in the outer space ($|y| > d/2$) can be represented as Floquet series of the incoming harmonics $\{a_l^i e^{-i\tau_l \kappa^i y + i\pi_l k^i x}\}$ and the outgoing ones $\{b_l^i e^{i\tau_l \kappa^i y + i\pi_l k^i x}\}$, where $\pi_l = l/\sigma$, $\tau_l = i(\pi_l^2 - 1)^{1/2}$, and the index i corresponds to the domain so that $\kappa^1 = k$ in free space and $\kappa^2 = k\nu_s$ in the slab. Briefly the field and the incident wave can be written as a vector,

$$U^i = \begin{pmatrix} a^i \\ b^i \end{pmatrix}, \qquad U^{incident} = \begin{pmatrix} a^{incident} \\ 0 \end{pmatrix}. \tag{1}$$

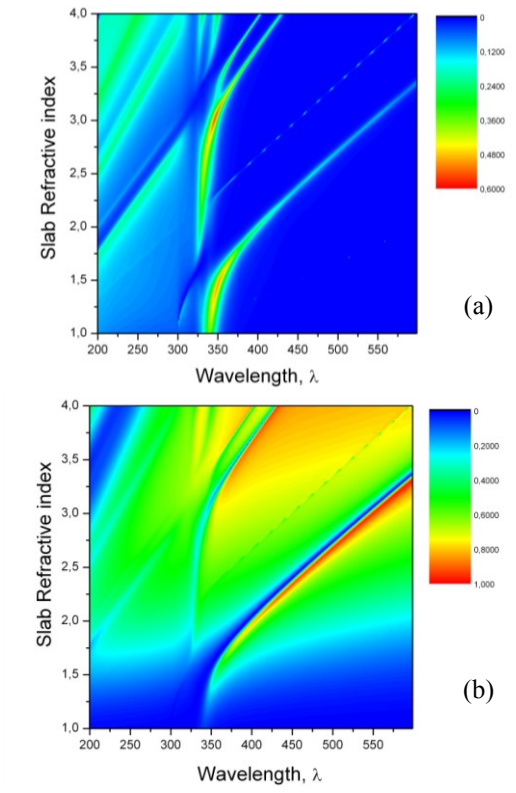


Fig.2 The reliefs of absorbance (a) and reflectance (b) from the dielectric slab of varying index and width equal to the grating period d=p=300 nm , in the middle of the slab there is a silver grating with a=15nm.

The total scattering matrix A that characterises a slab with a grating has a block-type nature and can be presented as

$$\begin{pmatrix} t \\ 0 \end{pmatrix} = A \begin{pmatrix} a^{incident} \\ r \end{pmatrix}, \qquad A = \begin{pmatrix} A^{II} & A^{I} \\ A^{III} & A^{IV} \end{pmatrix}. \tag{2}$$

Here $r = \{r_l\, e^{i\tau_l \kappa y + i\pi_l kx}\}$ and $t = \{t_l\, e^{-i\tau_l \kappa y + i\pi_l kx}\}$ are the vectors of the reflected and transmitted Floquet harmonics, respectively. Then the reflectance and transmittance are expressed as

$$R(\sigma) = \sum_{|i|<\sigma} \tau_i \,|\, r_i \,|^2, \qquad T(\sigma) = \sum_{|i|<\sigma} \tau_i \,|\, \delta_i^0 + t_i \,|^2, \tag{3}$$

where δ_i^j is the Delta function.

The matrix A is built by successive multiplication of the matrices that characterise the transmission of the incident field (1) throughout each elementary scatterer,

$$A = F^1 S^1 G S^2 F^2, \tag{4}$$

where F^i are the transfer matrices for the upper and lower interfaces of the slab. They can be expressed using well-known Fresnel matrices as

$$F^i = \begin{pmatrix} T^+ - R^- \left(T^-\right)^{-1} R^+ & R^- \left(T^-\right)^{-1} \\ -\left(T^-\right)^{-1} R^+ & \left(T^-\right)^{-1} \end{pmatrix}, \tag{5}$$

Further, S^i are the matrices of the phase shifts inside the slab,

$$S^i = \begin{pmatrix} S^+ & 0 \\ 0 & S^- \end{pmatrix}, \qquad S^{\pm} = \{e^{\pm ik\tau_s \nu_S d_i}\}, \tag{6}$$

and the matrix G is the transfer matrix of the field (1) through the grating. It can be obtained similarly to [8]. To build the vector of the reflected Floquet harmonics, we use equation (2),

$$r = -(A^{IV})^{-1} A^{III} a^{incident}, \tag{7}$$

$$t = A^{II} a^{incident} + A^{I} r. \tag{8}$$

Note that as our mathematical approach is similar to [7] then the resulting matrix equation (7) is the Fredholm second kind equation.

III. Numerical Results

Previous investigation of the silver-wire grating in free space has shown the presence of two different effects in its electromagnetic properties: the plasmon resonances aligned to the specific wavelength λ=348 nm and the grating resonances that appear close to the wavelength equal to the structure period. When the silver grating is sparse and the period does not coincide with plasmon resonance their interaction is

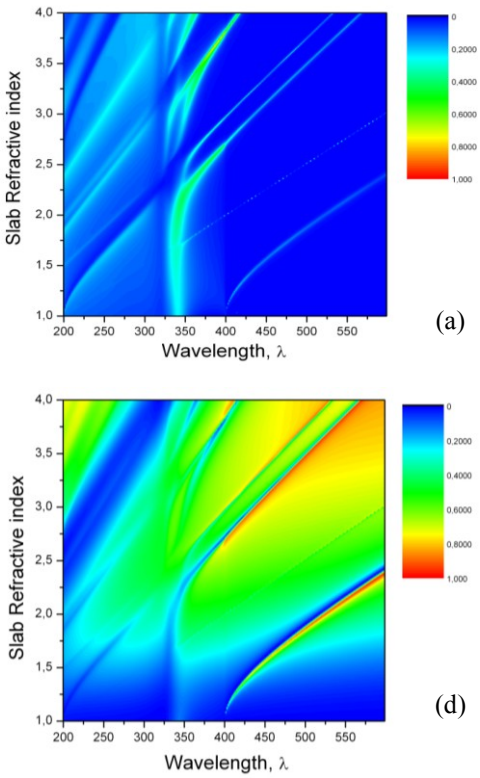


Fig.3 The same as in Fig.2 however for d=p=400 nm and a=20nm.

minimal. For such cases (e.g. radius is equal to 20 nm and period to 450 nm) the maximum reflectance might be achieved at the level of 0.15. Putting the silver grating in a dielectric medium gives opportunity to enhance its electromagnetic properties. Firstly the variation of the medium refractive index may enhance the plasmon resonance and shift it to the red side. Secondly changing the width of the slab and the position of the grating inside it can make visible higher-order grating resonances specific to any periodic structure.

The reliefs of absorbance and reflectance for the structure with parameters a=15 nm, d=p=300 mn, and d_1=d_2, and refractive index ν_s varying from the value 1, corresponding to the case of the silver grating suspended in free space, to 4 that represents a high-contrast dielectric, are shown in Fig.2. The bright ridges of high absorbance originate from the plasmon frequency and follow the inclined trajectories that are tracks of the grating modes. The high absorbance in this area for λ>400 nm where the silver has small losses is explained by the fact that the near field dramatically intensifies [6] and thus the absorbance increases. In this geometry the period selection is made in such a way that the grating resonances cross the plasmon wavelength under variation of the slab refractive index. If $\nu_s \to 1$, then the grating-resonance displays a Fano shape and its trajectory comes to λ=300 nm.

On the reflectance relief Fig.2 (b), with increase of the slab refractive index we see the formation of the wide flat hill that corresponds to the slab mode. It is split by much sharper ridges that correspond to the grating modes.

Absolutely the same in proportions geometry but with p=400 nm (Fig.3) and a=20 nm has been chosen to demonstrate more clearly the grating modes' behaviour. Here on both the absorbance and the reflectance reliefs we see two ridges of grating modes tending to λ=200 nm and λ=400 nm. There is, as previously, an area of high absorbance linked to the high bulk losses in silver for this frequency. On the reflectance relief, the slab-mode hill looks exactly as in Fig.2 (b) as it does not depend on the scale of periodic structure. However the plasmon resonances are brighter as the nanowires' radii are somewhat larger.

IV. Conclusion

We have built an accurate mathematical model for the dielectric slab with an embedded silver nanowire grating. The spector behavior of the reflectance and absorbance of the structure was investigated. The grating and the plasmon mode resonances have been visualised and explained. Dependences of the reflectance and absorbance on the slab refractive index have been studied.

Acknowledgment

This work was supported by the National Academy of Sciences of Ukraine via the State Target Program "Nanotechnologies and Nanomaterials" and by the European Science Foundation via the Research and Networking Programmes "Plasmon-Nanobiosense" and "Newfocus."

References

[1] M.A, Noginov, G. Zhu, *et al*. "Demonstration of a spaser-based nanolaser" *Nature,* vol. 460 pp 1110-1114, 2009.

[2] V. Twersky, "On scattering of waves by the infinite grating of circular cylinders," *IEEE Trans. Antennas Propagation*, vol. 10, no 3, pp. 737-765, 1962.

[3] K. Yasumoto, H. Toyama, and T. Kushta, "Accurate analysis of 2-D electromagnetic scattering from multilayered periodic arrays of circular cylinders using lattice sums technique," *IEEE Trans. Antennas Propagation*, vol. 52, no 10, pp. 2603-2611, 2004.

[4] R. Gomez-Medina, M. Laroche, and J.J. Saenz, "Extraordinary optical reflection from sub-wavelength cylinder arrays," *Optics Express*, vol. 14, no 9, pp. 3730-3737, 2006.

[5] V.O. Byelobrov, J. Ctyroky, T.M. Benson, A. Altintas, R. Sauleau, A.I. Nosich, "Low-threshold lasing modes of an infinite periodic chain of quantum wires," *Opt. Lett.*, vol. 35, no 21, pp. 3634-3636, 2010.

[6] V.O. Byelobrov, T.M. Benson, and A.I. Nosich, "Near and far fields of high-quality resonances of an infinite grating of sub-wavelength wires," *Proc. European Conf. Microwaves (EuMC-11)*, Manchester, 2011, pp. 858-861.

[7] D.M. Natarov, V.O. Byelobrov, R. Sauleau, T.M. Benson, and A.I. Nosich, "Periodicity-induced effects in the scattering and absorption of light by infinite and finite grating of circular silver nanowires", *Optics Express*, vol.19, no 22, pp. 22176-22190, 2011.

[8] K. Yasumoto, T. Kushta, and H. Toyama, "Reflection and transmission from periodic composite structure of circular cylinders," *Proc. Int. Symp. Physics Eng. Microwaves, MM and Sub-MM Waves (MSMW'2001),* Kharkiv, pp. 96-101, 2001.

[9] P.B. Johnson and R.W. Christy, "Optical constants of the noble metals," *Phys. Rev.*, vol. 6, pp. 4370-4379, 1972.

Accurate Numerical Study of Graphene Disk Surface Plasmon Resonances

Mikhail V. Balaban[1], Ana Vukovic[2], Trevor. M. Benson[2] and Alexander I. Nosich[1]

[1] Laboratory of Micro and Nano-Optics, Institute for Radiophysics and Electronics NASU, Kharkiv 61085, Ukraine
[2] George Green Institute for Electromagnetics Research, University of Nottingham, NG7 2RD, Nottingham, UK
mikhail.balaban@gmail.com

***Abstract*— The excitation of surface-plasmon resonances on a graphene disk is studied numerically as a 3D electromagnetic wave-scattering problem. The mathematical model for the scattering of a horizontal magnetic dipole field by a graphene disk is based on Maxwell's equations with electrically-resistive boundary conditions on the disk surface, where the graphene electron conductivity is included as a parameter and determined from the Kubo formalism. It is shown that plasmon resonance frequencies in the terahertz range shift with variation of the chemical magnetic potential of the graphene. Far field radiation patterns are plotted at two of the resonant frequencies.**

Keywords—graphene disk; surface plasmon resonance; wave scattering

I. INTRODUCTION

Graphene is a planar hexagonal structured layer of carbon atoms. Recently it has attracted the attention of the research community due to its amazing mechanical, electronic and optical properties. It is a zero bandgap semiconductor with conductivity tuned either by electrostatic or magnetostatic gating. This graphene property makes it a very promising material for the development of ultrathin fast nanoelectronic devices [1,2]. To study electromagnetic wave scattering by a graphene object it is necessary to couple the Maxwell boundary value problem with a phenomenological model of the graphene conductivity [3,4]. The main challenge of such an approach is to involve a zero thickness of the scatterer into the model. This can be done by modelling graphene as an equivalent resistive surface (also called an impedance surface because of the complex-valued resistivity). Then one can use various meshless codes for the numerical analysis of the wave scattering.

In this paper we use an analytical-numerical method for the study of electromagnetic wave scattering by a monolayer graphene (resistive) disk [5]. We suppose that the radius of the graphene disk is larger than 50 nm and thus disregard the edge effects on the graphene conductivity and use the electrical conductivity model developed for infinite graphene sheets. To solve the Maxwell boundary value problem we reduce it to a set of coupled dual integral equations (IEs) in the spectral domain and use the method of analytical regularization (MAR) with a Nystrom-type discretization to obtain the matrix equation.

The main goal of this paper is to develop a mathematically rigorous and numerically accurate method for the study of graphene disk surface plasmon resonances which are excited by a horizontal magnetic dipole located above the disk.

II. PROBLEM STATEMENT AND MATHEMATICAL MODEL

Consider the scattering of the time-harmonic ($e^{-i\omega t}$) electromagnetic field of a horizontal magnetic dipole by a graphene disk of radius a (Fig. 1). Suppose the dipole is located a distance h above the disk, and shifted from the disk axis by a distance r_0. In our analysis, we will use dimensionless cylindrical coordinates, (ρ,φ,ζ), normalized to the radius of the disk with the origin at the centre of the disk.

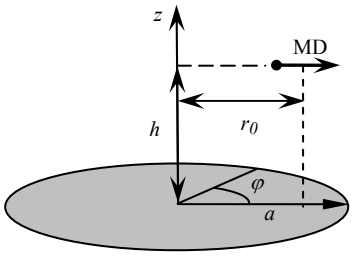


Fig. 1. Problem geometry

Decompose the total field as a sum of the electromagnetic fields incident and scattered by the disk. Assume that the scattered field satisfies homogeneous Maxwell equations outside the disk, while the total field satisfies the following resistive-type boundary conditions:

$$\left(E_{tg}^{+}+E_{rg}^{-}\right)=2Z_0\sigma^{-1}\vec{n}\times\left(H_{tg}^{+}-H_{tg}^{-}\right),\quad \left(E_{tg}^{+}-E_{rg}^{-}\right)=0\,. \quad (1)$$

Here Z_0 is the free-space impedance, $\vec{n}$ is the unit vector normal to the disk surface, and σ is the graphene surface conductivity, which can be determined from the Kubo formalism and expressed as a sum of intraband (σ_{intra}) and interband (σ_{inter}) contributions given by the following expressions [3]:

$$\sigma_{intra}=\frac{i e^2 k_B T}{\pi\hbar(\omega+i/\tau)}\left(\frac{\mu_c}{k_B T}+2\ln\left[\exp\left(-\frac{\mu_c}{k_B T}\right)+1\right]\right), \quad (2)$$

$$\sigma_{inter}=\frac{i e^2(\omega+i/\tau)}{\pi\hbar^2}\int_0^{\infty}\frac{f_d(-\varepsilon)-f_d(\varepsilon)}{(\omega+i/\tau)^2-4(\varepsilon/\hbar)^2}d\varepsilon\,, \quad (3)$$

where e is the charge of an electron, k_B is the Boltzmann constant, T is temperature, $\hbar$ is the reduced Planck constant, ω is the angular velocity, τ is the relaxation time of an electron, μ_c is the chemical potential and f_d is the Fermi-Dirac distribution function.

For completeness, the total electromagnetic field must satisfy the radiation condition and the condition of local integrability of power.

III. Solution Method

To solve the formulated problem, we use the method of dual IEs in the spectral domain together with the concept of analytical regularization [5]. We start with the following expressions for the scattered field components normal to the disk:

$$\begin{pmatrix} E_z^{sc,\pm}(\rho,\varphi,\zeta) \\ Z_0 H_z^{sc,\pm}(\rho,\varphi,\zeta) \end{pmatrix} = \sum_{m=-\infty}^{\infty} e^{im\varphi} \int_0^{\infty} e^{i\gamma(\kappa)|\zeta|} J_{|m|}(\kappa\rho) \begin{pmatrix} \kappa e_{m,z}^{sc,\pm}(\kappa) \\ \kappa h_{m,z}^{sc,\pm}(\kappa) \end{pmatrix} d\kappa \,, \quad (4)$$

where $\gamma(\kappa)=[(ka)^2-\kappa^2]^{1/2}$ is the complex-valued function of the branch $\mathrm{Im}(\gamma(\kappa)) \ge 0$, k is the free-space wavenumber, J_m is the first kind Bessel function of the order *m*, and $e_{m,z}^{sc,\pm}(\kappa)$, $h_{m,z}^{sc,\pm}(\kappa)$ are the images of the field components normal to the disk in the Fourier-Bessel spectral domain. By substituting expressions (4) into the boundary conditions (1) one can reduce the scattering problem to the following set of coupled dual IEs (CDIEs) for each azimuthal index:

$$\begin{cases} \int_0^{\infty} \bar{\bar{H}}_m(\kappa\rho) \begin{pmatrix} \gamma(\kappa)\left(u_m^{sc,-}(\kappa)+u_m^{in,-}(\kappa)\right)+ka\dfrac{2}{\sigma}u_m^{sc,-}(\kappa) \\ ika\left(v_m^{sc,+}(\kappa)+v_m^{in,+}(\kappa)\right)+\dfrac{2i}{\sigma}\gamma(\kappa)v_m^{sc,+}(\kappa) \end{pmatrix} d\kappa = \bar{0} \quad (\rho<1) \\ \int_0^{\infty} \bar{\bar{H}}_m(\kappa\rho) \begin{pmatrix} ika\,u_m^{sc,-}(\kappa) \\ -\gamma(\kappa)v_m^{sc,+}(\kappa) \end{pmatrix} d\kappa = \bar{0} \quad (\rho>1) \end{cases} . \quad (5)$$

Here $\bar{\bar{H}}_m(\kappa\rho)$ is the matrix kernel of the vector Hankel transform [6] given by:

$$\bar{\bar{H}}_m(\kappa\rho) = \begin{pmatrix} J'_{|m|}(\kappa\rho) & mJ_{|m|}(\kappa\rho)/(\kappa\rho) \\ mJ_{|m|}(\kappa\rho)/(\kappa\rho) & J'_{|m|}(\kappa\rho) \end{pmatrix}, \quad (6)$$

J'_m is the first order derivative of the Bessel function by the argument, and $u_m^{sc,\pm}(\kappa)$, $v_m^{sc,\pm}(\kappa)$ are the images of the jumps and the average values of the scattered field components normal to the disk respectively. These are the functions to be found. Functions $u_m^{in,\pm}(\kappa)$ and $v_m^{in,\pm}(\kappa)$ are known and determined by the incident field.

To solve the CDIEs obtained in (5), we first decouple them and reduce to a pair of DIEs. Then we invert the static part (the most singular part) of the integral operator of each DIE and obtain a set of the Fredholm second kind IEs (7)-(10) with kernel functions given by (11). Here $w(\kappa)=\gamma(\kappa)-i\kappa$ and $\Gamma(x)$ is Euler gamma function.

The features of Fredholm second kind IEs guarantee the existence and uniqueness of their exact solution, thus it is possible to use any reasonable analytical or numerical scheme to solve them. We use the numerical technique which is based on the truncation of the integration interval to the finite one combined with the descretization of the truncated equations using the Nystrom-type scheme. Finally we obtain a set of linear algebraic equations and solve them numerically using the Gauss inversion of the corresponding matrix.

IV. Numerical Results and Discussions

To study the excitation of the graphene-disk plasmon resonances by the time-harmonic electromagnetic field, we take a horizontal magnetic dipole as a source. Suppose that it is located a distance $0.0001a$ above the disk and oriented along the radial unit vector of the cylindrical coordinate system. We study the excitation of the resonances as the source point shifts from the disk axis (see Fig. 1). We start our consideration when the dipole is located on the disk axis ($r_0=0$), and then shift it towards the rim of the disk ($r_0=0.1a, 0.3a, 0.6a$).

Figures 2 and 3 show the powers (normalized to the radiated power of the dipole located in free space) radiated and lost by absorption by the magnetic dipole in the presence of the graphene disk. The material parameters taken for graphene are given in the boxes. One can clearly see the resonant nature of the curves. Resonant frequencies correspond to the excitation of surface plasmon resonances on the graphene disk. There are from one to several families of resonant peaks for different types of excitations. For example, in the case $r_0=0$ (black line)

$$u_m^{sc,-}(\lambda) = i\int_0^{\infty} \kappa^{-1} \left(\left(w(\kappa) + ka\frac{2}{\sigma} \right) u_m^{sc,-}(\kappa) + \gamma(\kappa) u_m^{in,-}(\kappa) \right) S_{|m|-1/2}(\kappa,\lambda) d\kappa - iA_m^l \sqrt{2} \frac{\Gamma(|m|+1)}{\Gamma(|m|+1/2)} \frac{J_{|m|+1/2}(\lambda)}{\lambda^{1/2}} \quad (7)$$

$$v_m^{sc,+}(\lambda) = -ka\frac{\sigma}{2}\frac{\lambda^{1/2}}{\gamma(\lambda)} \int_0^{\infty} \kappa^{-1/2} \left(v_m^{sc,+}(\kappa) + v_m^{in,+}(\kappa) \right) S_{|m|+1}(\kappa,\lambda) d\kappa - 2mD_m^r \frac{J_{|m|}(\lambda)}{\gamma(\lambda)} \quad (8)$$

$$A_m^l \frac{\Gamma(|m|+1)}{\Gamma(|m|+3/2)} - (ka)^{-1} 2D_m^r \frac{\Gamma(|m|+1/2)}{\Gamma(|m|)} = \sqrt{2} \int_0^{\infty} \kappa^{-3/2} \left(\left(w(\kappa) + ka\frac{2}{\sigma} \right) u_m^{sc,-}(\kappa) + \gamma(\kappa) u_m^{in,-}(\kappa) \right) J_{|m|+1/2}(\kappa) d\kappa \quad (9)$$

$$i\frac{\sigma}{2}\mathrm{sgn}(m) A_m^l + \mathrm{sgn}(m) D_m^r = ka\frac{\sigma}{2} \int_0^{\infty} \kappa^{-1} \left(v_m^{sc,+}(\kappa) + v_m^{in,+}(\kappa) \right) J_{|m|}(\kappa) d\kappa \quad (10)$$

$$S_{\mu}(\kappa,\lambda) = \kappa^{1/2}\lambda^{1/2} \int_0^1 J_{\mu}(\kappa v) J_{\mu}(\lambda v) v dv = \frac{\kappa^{1/2}\lambda^{1/2}}{\kappa^2-\lambda^2} \left(\lambda J_{\mu-1}(\lambda) J_{\mu}(\kappa) - \kappa J_{\mu-1}(\kappa) J_{\mu}(\lambda) \right) \quad (11)$$

$r_0 = 0$ (black line) one can see only the first family of peaks which correspond to the plasmon resonances with only one variation of the field along the azimuth and different numbers of variations along the radius. As we shift the source point to the disk rim, the number of the resonance families increases. For instance, in the case $r_0 = 0.6a$ (green line) one can see three families of plasmon resonances with different numbers of field variations along the azimuth and the radial coordinate.

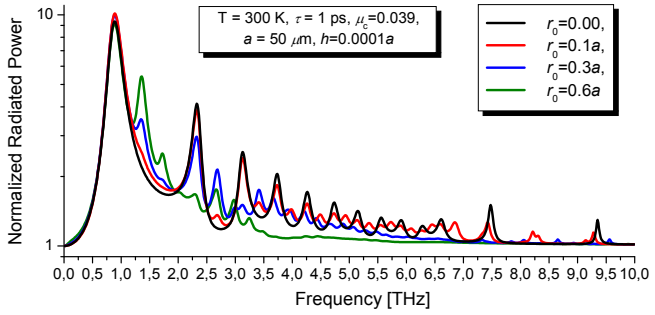


Fig. 2. Normalized radiated power as a function of frequency

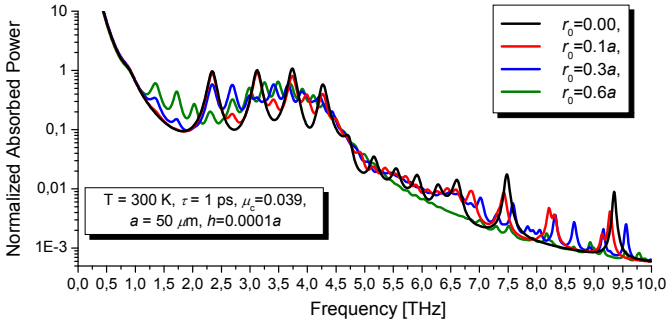


Fig. 3. Normalized absorbed power as a function of frequency

Figure 4 shows the normalized radiation patterns at two resonance frequencies, 0.8885 THz and 1.351 THz, for two different types of excitations. One can see the different number of azimuthal variations of the field that is due to the efficient excitation of different surface-plasmon modes. For the first frequency, the dominant mode has $|m|=1$ and for the second frequency this is the mode with $|m|=2$.

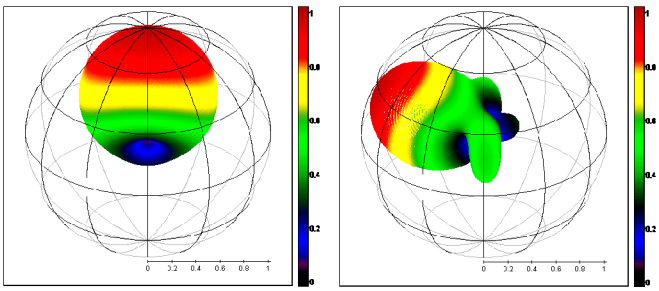


Fig. 4. Normalized radiation patterns of the magnetic dipole in the presence of the graphene disk at the frequencies 0.8885 THz (left) and 1.351 THz (right) and $r_0 = 0.1a$, $r_0 = 0.6a$, respectively

Figure 5 shows the shift of resonance frequencies in the case of the similar excitation but for the different types of graphene material (different values of chemical potentials).

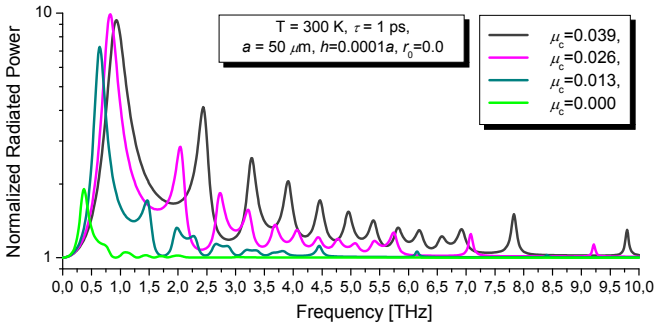


Fig. 5. The same as in Fig. 2 but calculated for different values of the chemical potential of the graphene material

V. Conclusions

We present a modelling method and fast convergent numerical algorithm for the analysis of the excitation of surface-plasmon modes on a graphene disk. The modelling is based on Maxwell's equations and resistive-type boundary conditions on the disk surface, where the graphene electron conductivity is included as a parameter. We reduce the problem to a set of Fredholm second kind integral equations which we solve numerically.

We study the different excitation of a graphene disk surface-plasmon modes by a horizontal magnetic dipole located above the disk. Also, we show how the resonance frequencies shift for different values of the chemical potential of the graphene disk material.

Acknowledgment

This work was partially supported by the National Academy of Sciences of Ukraine via the State Target Program "Nanotechnologies and Nanomaterials", and the European Science Foundation via the Research Networking Programme "Newfocus".

References

[1] H. Yan, X. Li, B. Chandra, G. Tulevski, Y. Wu, M. Freitag, W. Zhu, P. Avouris, and F. Xia, "Tunable infrared plasmonic devices using graphene/insulator stacks.," Nature Nanotechnology, vol. 7, no. 5, pp. 330–4, May 2012.

[2] I. Llatser, C. Kremers, A. Cabellos-Aparicio, J. M. Jornet, E. Alarcón, and D. N. Chigrin, "Graphene-based nano-patch antenna for terahertz radiation," Photonics and Nanostructures - Fundamentals and Applications, vol. 10, no. 4, pp. 353–358, Oct. 2012.

[3] G. W. Hanson, "Dyadic Green's functions for an anisotropic, non-local model of biased graphene," IEEE Trans. Antennas Propagation, vol. 56, no. 3, pp. 747–757, Mar. 2008.

[4] M. V. Balaban and A. I. Nosich, "Accurate modeling of the THz beam scattering by a graphene disk," in NATO Advanced Research Workshop Detection of Explosives and CBRN (Using Terahertz) TERA-MIR 2012, p. 51, 2012.

[5] M. V. Balaban, R. Sauleau, T. M. Benson, and A. I. Nosich, "Dual integral equations technique in electromagnetic wave scattering by a thin disk," Progress In Electromagnetics Research B, vol. 16, pp. 107–126, 2009.

[6] S. Ali and J. Kong, "Vector Hankel transform analysis of annular-ring microstrip antenna," IEEE Trans. Antennas Propagation, vol. 30, no. 4, pp. 637–644, Jul. 1982.

Carbon-Rich Nanostructured a-SiC for Cold Emitters

Gordienko S. A., Nazarov A. N., Lytvyn P. M., Stadnik A. A., Gomeniyuk Yu. Yu., Rusavsky A. V., Vasin A. V., Stepanov V. G., Lysenko V. S.
Lashkaryov Institute of Semiconductor Physics
National Academy of Sciences of Ukraine
Kyiv, Ukraine
nazarov51@yahoo.com

Nazarova T. M.
Department of Common and Inorganic Chemistry
National Technical University of Ukraine "KPI"
Kyiv, Ukraine
tmnazarova@gmail.com

***Abstract*—This work describes emission properties of a new nanostructured material which is carbon-rich amorphous (a) SiC deposited on silicon wafer. Even non-optimized technology demonstrates that the field enhanced factor of the electron emission can reach of 1000 with current density near $1x10^{-3}A/cm^2$ and efficiency of electron emission near 10%. A good correlation between charge transfer through the a-SiC layer and electron emission from the material in high vacuum is observed.**

Keywords—cold emitter; carbon-rich a-SiC; conductance AFM

I. INTRODUCTION

In papers [1], [2] it was shown that in the carbon-rich a-SiC after annealing at temperature higher 600°C the sp^2 coordinated graphitelike carbon nanoclusters precipitate in channels with size of 5-7 nm of the porous a-SiC material, which are oriented mainly perpendicular to surface. Additionally it was found [3] that such kind of the annealing leads to considerable (up to 10^3) conductance increase of the materials. In paper [3] it was suggested that such increased conductance is associated with conductivity either through the graphitelike carbon nanoclusters or even graphene flakes and could be used in effective field-effect electron emission from surface of the material.

In this study we focus our attention on using of carbon-rich a-SiC:H thin film on Si for cold electron emission which can be easy fabricated by reactive DC magnetron sputtering of a monocrystalline silicon target [1] and RF magnetron sputtering of SiC target [3] using a mixture of Ar/CH_4 as working gas.

II. SAMPLE PREPARATION AND EXPERIMENTAL APPROACH

A. Sample preparation

The carbon-rich hydrogenated a-SiC:H films with thickness near 300 nm were deposited by RF magnetron sputtering of crystalline SiC target using Ar (93 vol.%) + CH_4 (7 vol.%) mixture as a working gas on p- and n-type Si wafer. Following vacuum annealing in temperature range from 400°C to 1100°C for 60 minutes was used. Electron field emission was studied on the samples annealed at 900°C.

B. Measurement techniques

Chemical composition and local carbon structure were studied by scanning electron microscopy equipped with local high resolution Auger electron spectroscopy (JAMP-9500F) and by micro-Raman scattering spectroscopy (NR1800, JASCO) with excitation by a 514-nm-wavelength laser beam. The chemical bonds in the material were studied by Furrier transmission infrared spectroscopy (2000FT-IR, Perkin Elmer). The structural properties were examined by high-resolution scanning transmission electron microscopy (STEM, Jeol-JEM200CX) equipped with electron energy loss spectrometer (EELS, Gatan ENFINA 1000).

Conductance of the structures was studied by In probe station by semiconductor parameter analyzer Agilent 4156C and conductance AFM (C-AFM) mode in Quadrexed Dimensions 3000 Nanoscope IIIa scanning probe microscope. Current and topography maps were collected simultaneously by different measurement channels with a unique correspondence between points on topography and on the current map.

The emission current was studied in vacuum of 10^{-7} Torr; anode was located on distance of 105 μm from surface of the sample. Schematic design of the experimental equipment is presented in Fig. 1.

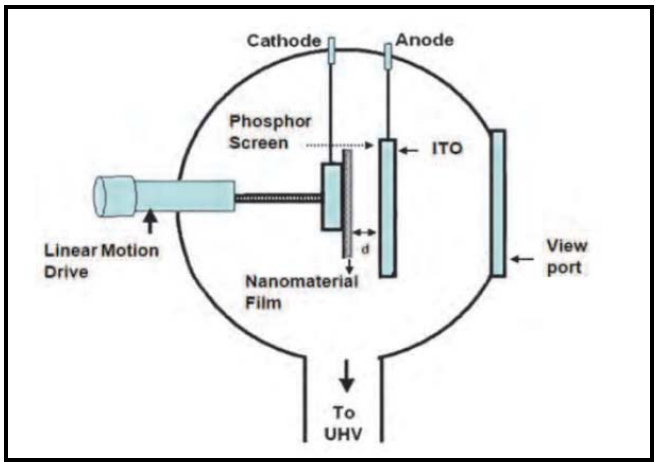


Fig. 1. Schematic design of chamber for UHV electron emission measurements.

This study was partially supported by project 1.1.4/15 in frame of Ukrainian State Program on Research and Development of Sensor High Technology Products.

III. Results and Discussion

A. Chemical composition of the samples

Fig. 2 demonstrates of Auger spectrum analysis a chemical composition of carbon-rich a-SiC layers deposited by RF magnetron sputtering of SiC:H target in Ar/CH_4 gas mixture with different power after vacuum annealing at 900°C. One can see that higher RF power (247 W; sample KO155) results in carbon concentration near 72 at %. At smaller RF power (195 W; sample KO156) carbon concentration in the sample reaches 77 at %. Both samples show small oxygen concentration near 7-8 at %. Additionally, an increase of RF power at a-SiC:H film deposition results in increase of the film density [4].

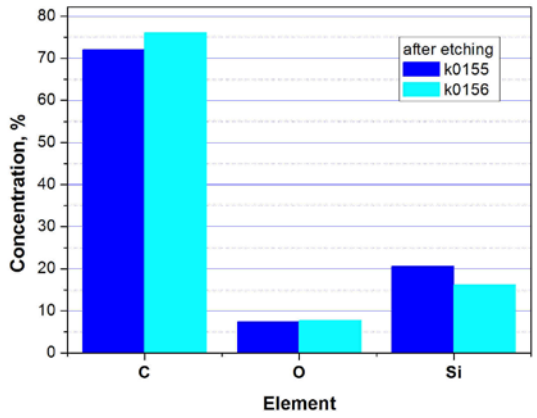


Fig. 2. Chemical composition of a-SiC films measured by Auger spectrum analysis.

Fig. 3 shows the Furrier-transform infra-red (FTIR) transmission spectra of the as-deposited and annealed a-SiC:H. The spectra of as-deposited samples in the range 400–3600 cm^{-1} consist of four main absorption bands at 790 cm^{-1}, 1000 cm^{-1}, 2100 cm^{-1}, and 2900 cm^{-1}. The absorption bands at 2100 cm^{-1} and 2900 cm^{-1} can be unambiguously attributed to the Si–H_n and C(sp^3)–H_n stretching vibration modes, respectively [5]. These bands decayed with increasing annealing temperature and disappeared completely after annealing at 600 °C. No detectable absorption in the range of 3000–3100 cm^{-1} inherent to C(sp^2)–H_n bonds in aromatic rings was observed. Thus the vacuum annealing (and annealing in argon [6]) at temperature above 600°C leaded to considerable decrease of hydrogen bonds concentration in a-SiC film.

A weak, broad, and featureless Raman scattering (RS) band in the range of 1200–1650 cm^{-1} was observed for the as-deposited sample (not shown here). After vacuum annealing at 900°C in Raman spectrum two pronounced bands centered at about 1352 and 1525 cm^{-1} appear (Fig. 4). These bands are attributed to *D*- and *G*-bands of a graphitelike local structure and formation of the graphitelike sp^2-coordinated nanoclusters. The higher RF power deposition of a-SiC:H layer with following high-temperature vacuum annealing resulted in decrease of defect component in the graphitelike structure (Fig. 4).

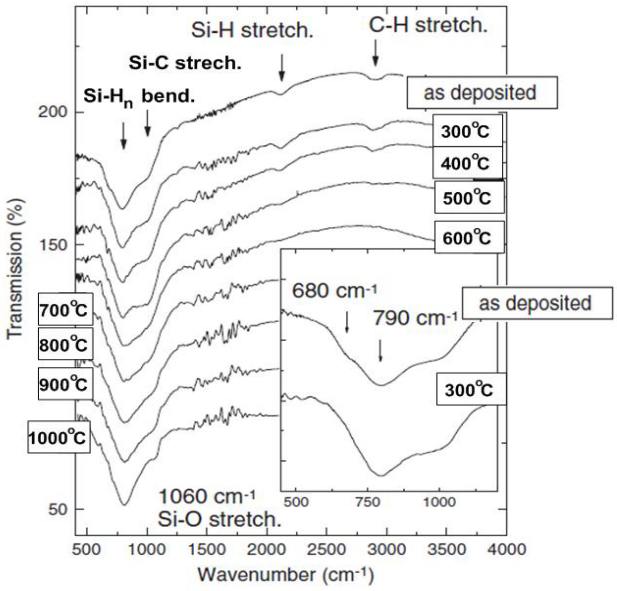


Fig. 3. FTIR spectra of the carbon-rich a-SiC layers after different annealing temperature in vacuum.

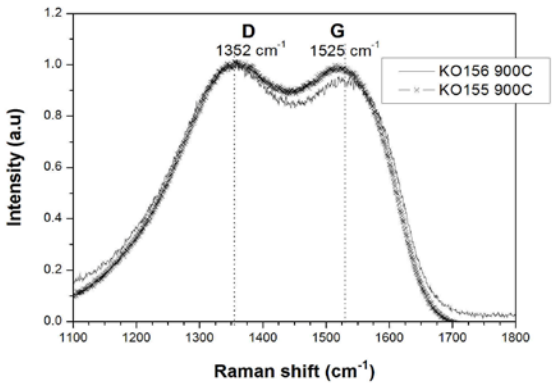


Fig. 4. Raman scattering spectrum of carbon-rich a-SiC layers annealed at 900°C.

B. Surface morphology and bulk structure of the a-SiC layers

The atomic structure of the carbon-rich *a*-SiC:H was investigated in as-deposited and annealed samples. Fig. 5 shows a typical AFM image in standard measurement (Fig. 5 (a)) and in phase contrast measurement (Fig. 5 (b)) of the as-deposited *a*-SiC:H surface. One can readily observe the nano-meter-size surface morphology of the films. The surface profile, shown in Fig. 5, confirms and quantifies the nanostructured columnar morphology that is inherent to thin films deposited at low temperatures. Large columns, 100–150 nm in diameter, are composed of thinner columns with diameters in the range of 20–40 nm.

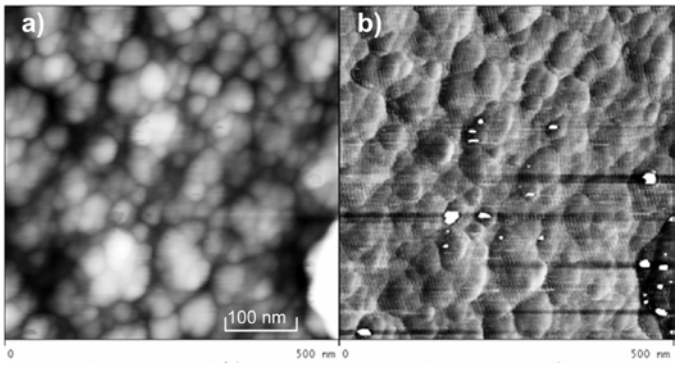


Fig. 5. AFM images of as-deposited a-SiC surface in standard (a) and phase contrast (b) regimes.

Elastic scattering STEM cross section image of *a*-SiC:H film annealed at 950 °C is shown in Fig. 6 (a, c). The structure of the film is not homogeneous and contains nanoscale low-density regions (voids) elongated roughly perpendicular to the film surface and having a typical size of less than 5 nm in width and up to 50 nm in length. It is suggested that the observation of the voids in Fig. 6 is also an indication of columnar growth morphology.

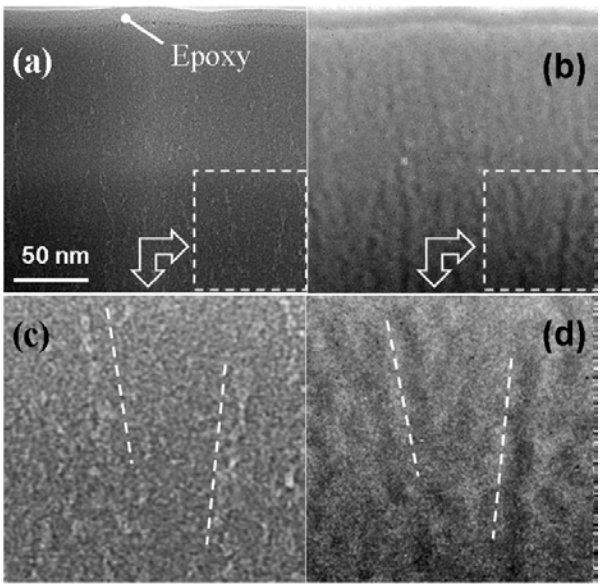


Fig. 6. STEM micrographs of an a-SiC:H sample annealed at 950 °C: (a) Elastic scattering STEM cross-sectional image; (b) STEM image based on EELS; (c) and (d) higher magnification of the regions marked by the dashed lines in (a) and (b) correspondingly.

The investigation of the spatial distribution of graphitelike carbon clusters was performed using energy filtered STEM measurements [7]. A narrow peak centered at 284 eV was observed in the electron energy loss spectrum of an carbon-rich *a*-SiC:H sample annealed at 950°C. This peak is a fingerprint of graphitelike carbon and corresponds to the electron transition from the 1*s* to $2p_z$ unoccupied states ($1s \rightarrow \pi^*$) in the sp^2 hybrid orbital of carbon atom [8]. Fig. 5 (b, d) shows the map of the distribution of sp^2-coordinated carbon atoms generated by energy filtering of the electrons with energy loss around 284 eV. The darker regions correspond to higher concentration of sp^2-hybridized carbon atoms. It is clearly seen that the spatial distribution of the graphitelike carbon is well correlated with the distribution of the nanovoids. Therefore, we can conclude that graphitic carbon that forms after the high-temperature annealing is predominantly localized in the nanovoids of the *a*-SiC:H film. It is well known that graphitelike clusters preferably condense on the surface in such a manner that graphene sheets form parallel to the surface. Thus, good conductance graphitelike clusters (or graphenelike sheets) incorporated in the dielectric a-SiC matrix can reveal effective electron emission from the surface.

C. Electrical conductance properties of the carbon-rich a-SiC/Si heterostructures

To understand local conductance in our material the conductance AFM measurements were employed. Surface topography map and conductance map are presented in Fig. 7 (a, b). It can be seen, that in main cases protuberances on the surface corresponds to enhanced value of electric current (Fig. 7 (c, d)). It should be noted, that in Fig. 7 (d) average current which equals 16 nA was shifted to zero for better comparison current and surface topographic data. Existence of average current in all areas of the surface indicates that conductive channels have dimensions considerable smaller than the tip diameter that is 50 nm. The data of previous chapter show that conductive channel dimension is near 5-10 nm that is considerable smaller than diameter of the used tip.

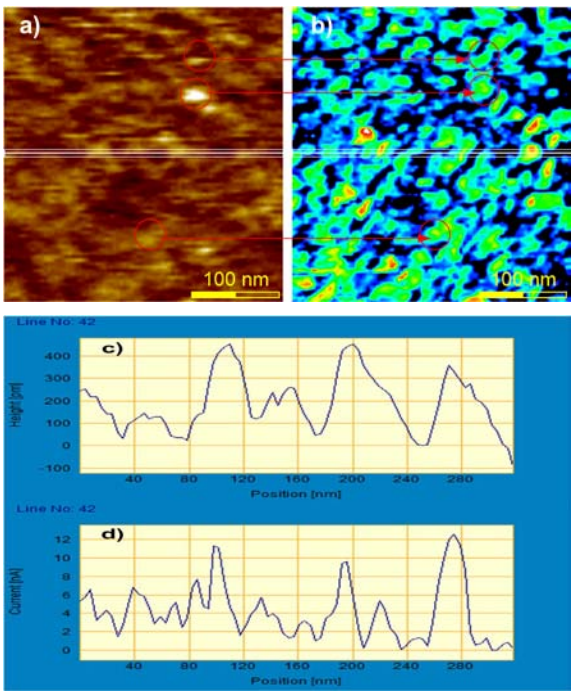


Fig. 7. AFM surface topography map (a) and conductance AFM map (b) of the sample KO155 measured at +1V by Pt tip. Height (c) and current (d) profiles along the line presented in Figures 3 (a) and (b) correspondingly. Width of the line corresponds to aria of average.

Dependence of average current on applied voltage was studied by In probe station with the probe diameter near 2 mm. All studied a-SiC/Si heterostructures demonstrated power low

of *I-V* characteristics ($I = aV^n$) where n was near 2 (see Fig. 8) that associated with space charge limited current (SCLC). Maximal current at electric field above 10 V/μm was observed for a-SiC/Si heterostructure fabricated on *n*-type Si wafer (sample A0094). In case of the p-type Si wafer the higher average current corresponded to higher carbon concentration in a-SiC layer.

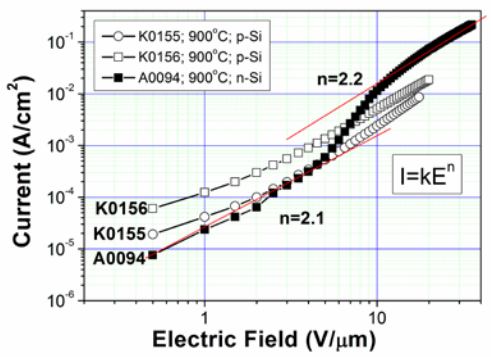


Fig. 8. I-V characteristics of the a-SiC/Si heterostructures measured by In probe station.

D. Electron emission of the carbon-rich a-SiC/Si heterostructures

Electron emission in vacuum from the a-SiC/p-Si heterostructures with current density 10^{-6} A/cm^2 was observed at turn-on electric field of 7 μV/cm for a-SiC film with high carbon concentration (77 at. %) and 9 μV/cm for n-type Si wafer. Current density of 10^{-3} A/cm was observed at electric field of 17 μV/cm for p-type Si wafer and 13 μV/cm for n-type Si wafer (Fig. 9).

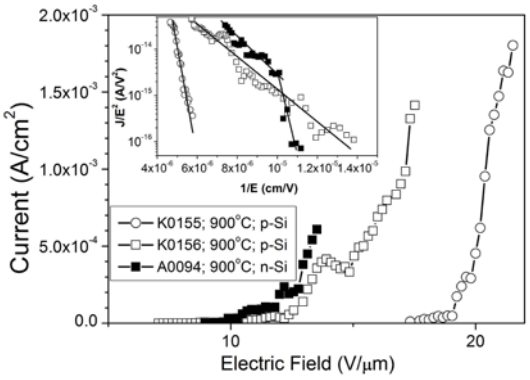


Fig. 9. Emission curves for different kind of studied heterostructures in linear coordinate and in coordinate of Fowler-Nordheim (inset).

The E-J characteristics show the Fowler-Nordheim law with field enhancement factor of 940 for p-type Si wafer and 800 for *n*-type Si wafer (see inset Fig. 9). The *E-J* characteristics were good reproducible and emission current was stable with changing near 15% during 60 minutes.

Comparison electron emission current with average current flowing through the structures allows us to extract efficiency of the electron emission in our structures. It was obtained that for higher carbon concentration in a-SiC layer and *p*-Si wafer the electron emission efficiency equals to 10% and for our structures with n-type Si wafer can reach of 5%.

IV. Conclusions

It was demonstrated that carbon-rich a-SiC films deposited on Si wafer by magnetron sputtering of SiC target in Ar/CH_4 gas mix posses porous structure saturated with graphitelike nanoclusters (or graphene flakes) and can be used as flat cold emitter with field enhanced factor near 1000. Minimal turn-on electric field is observed for the structure fabricated on n-type Si wafer and maximal electron emission efficiency for carbon concentration in a-SiC equaled to 77 at %.

Acknowledgment

The authors are grateful to Dr. Sci. V. Strelchuk and Dr. A. Nikolenko from Lashkaryov Institute of Semiconductor Physics NASU for micro-Raman scattering measurements and Prof. S. Muto from Nagoya University, Japan for possibility to perform measurements with scanning transmission electron microscopy.

References

[1] A. V. Vasin, S. P. Kolesnik, A. A. Konchits, A. V. Rusavsky, V. S. Lysenko, A. N. Nazarov, Y. Ishikawa, and Y. Koshka, "Structure, paramagnetic defects and light-emission of carbon-rich *a*-SiC:H films," J. Appl. Phys., vol. 103, pp. 123710-1 - 123710-7, June 2008

[2] E.N. Kalabukhova, S.N. Lukin, D.V. Savchenko, B. D. Shanina, A. V. Vasin, V. S. Lysenko, A. N. Nazarov, A. V. Rusavsky, J. Hoentsch, and Y. Koshka, "EPR study of carbon and silicon related defects in carbon-rich hydrogenated amorphous silicon-carbon films," Phys. Rev., vol. B81, pp. 155319-1 - 155319-9, August 2010.

[3] S.O. Gordienko, A.N. Nazarov, A.V. Vasin, A. V. Rusavsky, and V. S. Lysenko, "Correlation of nanostructure and charge transport properties of oxidized *a*-SiC:H films," Physica Status Solidi C, vol. 9, pp. 1477 – 1480, June 2012.

[4] A. V. Vasin, A. V. Rusavsky, A. N. Nazarov, V. S. Lysenko, P. M. Lytvyn, V. V. Strelchuk, K. I. Kholostov, V. P. Bondarenko, and S. P. Starik, "Identification of nanoscale structure and morphology reconstruction in oxidized a-SiC:H thin films," Appl. Surf. Science, vol. 260, pp. 73 - 76, January 2012.

[5] B. K. Ghosh, and B. K. Ageawal, "Vibrational structure of hydrogenated amorphous silicon carbide alloys," Physica Status Solidi B, vol.147, pp. 97- 102, January 1988.

[6] A. V. Vasin, Sh. Muto, Yu. Ishikawa, A.V. Rusavsky, T. Kimura, V. S. Lysenko, and A. N. Nazarov, "Comparative study of annealing and oxidation effects in a-SiC:H and a-SiC thin films deposited by radio-frequency magnetron sputtering," Thin Solid Films, vol. 519 , pp.2218–2224, 2011.

[7] R. F. Egerton, Electron Energy-Loss Spectroscopy in the Electron Microscope. Plenum Press, New York, 1996.

[8] "Transmission Electron Energy Loss Spectrometry in Materials Science and the EELS ATLAS," edited by C. C. Ahn,Wiley-VHC, New York, 2004.

Negative Differential Resistance Based Voltage-Controlled Oscillator for VHF Band

V.V. Ulansky
Department of Electronics
National Aviation University
Kyiv, Ukraine
vulanskyi@yahoo.com

S.F. Ben Suleiman
Department of Electronics
Carleton University
Ottawa, Canada
salibensuleiman@cmail.carleton.ca

***Abstract*—This paper presents a novel sinusoidal negative differential resistance (NDR) voltage-controlled oscillator (VCO) for VHF band. The VCO circuit is based on a junction field effect transistor (JFET) in combination with a current mirror (CM). Mathematical model of the proposed VCO is described and the start-up conditions are determined. The designed VCO covers the frequency band between 139.8 and 256.2 MHz. It has a phase noise of -111.1 dBc/Hz at 10 kHz offset from carrier, consumes only 6.2mW, suppresses harmonics to better than -31dBc, and achieves a very good figure of merit (FOM) performance of -190 dBc/Hz. The implemented VHF VCO generates low-noise low-distortion signal confirming simulation results.**

Keywords—voltage-controlled oscillator; negative differential resistance; phase noise; power consumption

I. INTRODUCTION

Voltage-controlled oscillators are critical building blocks in modern communications systems and phase-locked loop frequency synthesizers. Most of VHF VCOs have Colpitts or Clapp topology [1]. These VCOs are classified as negative-impedance oscillators because they have a negative real part in the input impedance. There is another class of VHF VCOs which is based on the Gun and tunnel diodes [2, 3]. Such oscillators have an N-type of *I-V* characteristic with negative-differential resistance (NDR) region. In the NDR oscillators the tank circuit losses are compensated by the induced negative resistance which is created at the operating point located in the NDR region. An N-type *I-V* characteristic can also be created by an electronic circuit where the NDR is found between two terminals like in the Gun and tunnel diodes. However, known NDR circuits usually have some properties that do not allow using them for VHF *LC* oscillator applications. For example, the NDR circuit proposed in [4] is based on CMOS NDR inverters with two power supplies, that is, three terminals are required for creating the N-type characteristic. Moreover, the proposed circuit is used for an oscillator similar to the CMOS ring oscillators that have poor phase noise performance compared to the class of *LC* oscillators. The effect of NDR can also be obtained by using a special connection of a BJT with a JFET [5, 6]. However, the *I-V* characteristics of these circuits suffer from BJT β_F-dependence and thermal runaway, and cannot be used in practical circuits. The proposed VCO has precisely controllable NDR region, low level of noise and harmonic distortions. The electronic circuit of the proposed VCO is a two terminal NDR device requiring only one power supply.

II. NEGATIVE DIFFERENTIAL RESISTANCE CIRCUIT

The proposed NDR circuit is shown in Fig. 1. Transistors T_1 - T_5 and resistors R_1 and R_2 provide an N-type characteristic between the voltage V_{DD} and the corresponding current I_t. Matched enhancement mode p-channel MOS transistors $T_1 - T_4$ form the improved MOS Wilson CM used for providing the feature that the current I_{D2} is a copy of the reference current I_{D1} and for any mode of the JFET T_5 operation $I_{D2}=I_{D1}$. A typical *I-V* characteristic of the proposed NDR circuit is shown in Fig. 2.

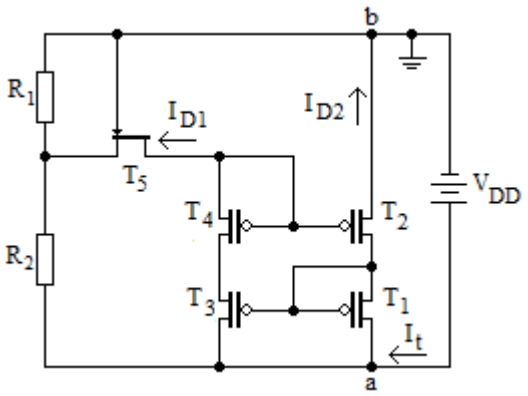


Fig. 1. Proposed NDR circuit.

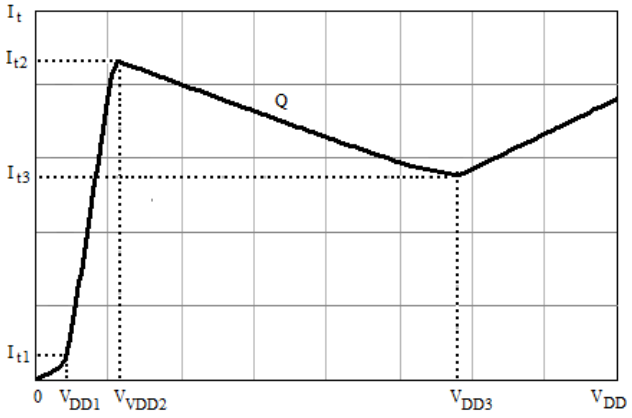


Fig. 2. Typical *N*-type *I-V* characteristic of the proposed NDR circuit.

In the first part of the characteristic, where

$$0 \le V_{DD} < V_{DD1} = 2V_{TH}\left(R_1 + R_2\right)/R_2$$

transistors T_1 - T_4 are in the cutoff mode of operation and no drain current flows (I_{D2}=0), the drain current of T_5 is also zero (I_{D1}=0) and the total current

$$I_t = V_{DD}/\left(R_1 + R_2\right) \quad (1)$$

where V_{DD} is the power supply voltage, V_{TH} is the threshold voltage of a p-channel MOS transistor in the CM. When voltage V_{DD} reaches V_{DD1}, transistors T_1 - T_4 and T_5 begin to operate in the ohmic region, so that

$$V_{DS} < V_{GS} - V_P \quad (2)$$

where V_{DS} is the drain-to-source voltage of T_5, V_{GS} is the voltage between gate and source of T_5, V_P is the pinch-off voltage (V_P <0) of T_5. Since

$$V_{GS} = -I_{D1}\left(R_1 IIR_2\right) - \frac{V_{DD}R_1}{R_1 + R_2} \quad (3)$$

$$V_{DS} = \frac{V_{DD}R_2}{R_1 + R_2} - I_{D1}\left(R_1 IIR_2\right) - 2V_{TH} \quad (4)$$

the inequality (2) is transformed into the form

$$V_{DD} < 2V_{TH} - V_P .$$

To derive the dependence of transistor T_5 drain current (I_{D1}) as a function of the circuit parameters in the ohmic region the following equation can be used [7]:

$$I_{D1} = I_{DSS}\left[\frac{3V_{DS}}{-V_P} - \frac{2\left(V_{DS} - V_{GS}\right)^{3/2}}{\left(-V_P\right)^{3/2}} + 2\left(\frac{V_{GS}}{V_P}\right)^{3/2}\right] \quad (5)$$

where I_{DSS} is the drain-to-source current in saturation for V_{GS}=0. Substituting (3) and (4) into (5), yields

$$2I_{DSS}\left(\frac{R_1 IIR_2}{-V_P} I_{D1} + \frac{R_1}{R_1 + R_2} \times \frac{V_{DD}}{-V_P}\right)^{3/2} - \left(\frac{3I_{DSS}\left(R_1 IIR_2\right)}{-V_P} + 1\right) I_{D1} +$$

$$\frac{3I_{DSS}R_2}{R_1 + R_2} \times \left(\frac{V_{DD}}{-V_P}\right) - \frac{6I_{DSS}V_{TH}}{-V_P} - 2I_{DSS}\left(\frac{V_{DD} - 2V_{TH}}{-V_P}\right)^{3/2} = 0 . \quad (6)$$

The drain current I_{D1} is found by solving the nonlinear equation (6).

When $V_{DS} \ge V_{GS} - V_P$, that is $V_{DD} \ge 2V_{GS,P} - V_P$, transistor T_5 begins to operate in the saturation region above pinch-off, where $V_{GS,P}$ is the voltage between gate and source of a p-channel MOS transistor in the saturation region. The voltage $V_{GS,P}$ is determined as

$$V_{GS,P} = V_{TH} + V_{OV}$$

where V_{OV} is the overdrive of a p-channel MOS transistor. In the saturation region, the drain current of T_5 can be approximated as [7]

$$I_{D1} = I_{DSS}\left[1 - 3V_{GS}/V_P + 2\left(V_{GS}/V_P\right)^{3/2}\right]. \quad (7)$$

Substituting V_{GS} from (3) into (7), gives the following nonlinear equation:

$$2I_{DSS}\left[\frac{V_{DD}R_1}{-V_P\left(R_1 + R_2\right)} + \frac{R_1 IIR_2}{-V_P} I_{D1}\right]^{3/2} + \left(\frac{3I_{DSS}\left(R_1 IIR_2\right)}{V_P} - 1\right) I_{D1} +$$

$$I_{DSS}\left[1 + \frac{3V_{DD}R_1}{V_P\left(R_1 + R_2\right)}\right] = 0 . \quad (8)$$

The current I_{D1} is found by solving (8).

Applying Kirchhoff's current law to the node *a* in the circuit of Fig. 1, we find

$$I_t = I_{D1} + I_{D2} - \frac{I_{D1}R_1}{R_1 + R_2} + \frac{V_{DD}}{R_1 + R_2} \quad (9)$$

where I_{D2} is the drain current of p-channel MOS transistor T_2. Since $I_{D2} \approx I_{D1}$, then (9) is reduced to

$$I_t = 2I_{D1} - \frac{I_{D1}R_1}{R_1 + R_2} + \frac{V_{DD}}{R_1 + R_2} . \quad (10)$$

In the third part of the *I-V* characteristic, the current I_t decreases because the term $2I_{D1}$ falls faster than the term $V_{DD}/\left(R_1 + R_2\right)$ rises. If $R_1 \le R_2/4$, the second term in (10) is much less than the first term and can be neglected. The current I_{D1} is found by solving (8).

Since the point V_{DD}=V_{DD2} is a local maximum of the *I-V* characteristic, the threshold voltage V_{DD2} can be found from the following condition:

$$I_{t2} = \max\left[I_t : V_{DD2} \le 2\left(V_{TH} + V_{OV}\right) - V_P\right]$$

where I_t is calculated from (10) by substituting I_{D1} from (6). The threshold voltage V_{DD2} is in the vicinity of the voltage

$$V_{DD} = 2\left(V_{TH} + V_{OV}\right) - V_P$$

at which the drain current of T_5 is saturated.

The threshold voltage V_{DD3} corresponds to the local minimum of the *I-V* characteristic. It is found from the following condition:

$$I_{t3} = \min\left(I_t : V_{DD3} < -V_P\left(1 + R_2/R_1\right)\right)$$

where I_t is calculated from (10) by substituting I_{D1} from (8). The voltage V_{DD3} is less than voltage

$$V_{DD} = -V_P\left(R_1 + R_2\right)/R_1$$

at which $V_{GS}=V_P$ and $I_{D1}=0$. Thus, when

$$V_{DD3} \le V_{DD} < -V_P\left(R_1 + R_2\right)/R_1$$

the differential resistance is positive, but the current I_t is still calculated from (10). And, finally, when

$$V_{DD} \ge -V_P\left(R_1 + R_2\right)/R_1$$

the current $I_{D1}=0$ and the total current I_t is calculated from (1).

In order to evaluate the validity of the VCO *I–V* characteristic analytical modeling, the SPICE simulations were performed for the NDR circuit of Fig. 1. All simulations were carried out using a MMBFJ310LT1 JFET with V_P=-3.5V and I_{DSS}=40.2mA and four FDV302P transistors with V_{TH}=-0.83V for the improved Wilson CM. The resistor values of R_1 and R_2 were selected to be 1kΩ and 4kΩ, respectively. Equations (6), (8) and (10) and special MATLAB program were used for plotting the theoretical *I-V* characteristic. The simulated and theoretical *I-V* characteristics are shown in Fig. 3. As seen, a good agreement exists between the simulated and theoretical results in the NDR region of the *I-V* curve, which confirms the validity of the derived analytical equations.

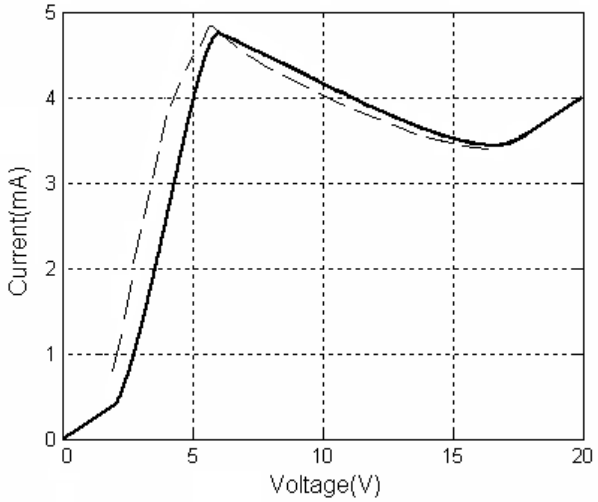


Fig. 3. The *I-V* characteristic of the NDR circuit: solid line corresponds to SPICE simulation; dashed line corresponds to analytical modeling.

III. VCO ARCHITECTURE

The VCO circuit is shown in Fig. 4, where the NDR electronic circuit of Fig. 1 is connected to the tank circuit. The VCO tank circuit consists of a coil L and two contrary connected varactor diodes VD_1 and VD_2. The varactor diodes are incorporated into the tank circuit to provide a voltage variable input for the VCO. The inductor L_s isolates the control voltage line, V_{ctr}, from the VCO tank. Capacitor C_1 is a small feedback capacitor, which facilitates the VCO self-excitation. Capacitor C_2 is used for reducing the level of phase noise and harmonic distortions.

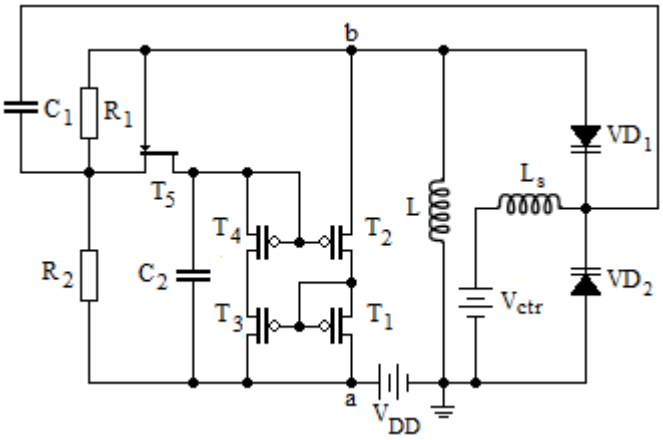


Fig. 4. VCO circuit.

IV. START-UP CONDITIONS

For finding the start-up conditions, the NDR is determined in an arbitrary point Q of the *I-V* characteristic (see Fig. 2):

$$R_D = \left(dI_t/dV_{DD}\right)^{-1}. \qquad (11)$$

Since $dI_t<0$, it follows that $R_D<0$. Using (8) and (10), R_D may be calculated from (11) and expressed as

$$R_D = \frac{R_1 + R_2}{1 - \dfrac{3I_{DSS}X_1X_3\left(R_1 + 2R_2\right)}{1 + 3I_{DSS}X_2X_3}} \qquad (12)$$

$$X_1 = R_1/\left[-V_P(R_1 + R_2)\right]$$

$$X_2 = R_1 \| R_2/\left(-V_P\right)$$

$$X_3 = 1 - \sqrt{X_1V_{DDQ} + X_2I_{DQ}}$$

where V_{DDQ} and I_{DQ} are, respectively, the power supply voltage and drain current of transistor T_5 at the selected operating point Q.

In order to guarantee the oscillation, the equivalent tank circuit resistance, R_{eqv}, must be greater than the absolute value of the NDR at the operating point, i.e.,

$$R_{eqv} > \left|R_D\right| \qquad (13)$$

where R_{eqv} is the equivalent unloaded tank circuit resistance at resonance. Assuming the VCO operates in the frequency range from f_1 to f_2 and expressing the equivalent tank circuit resistance at resonance as

$$R_{eqv} = 2\pi f Q_u L$$

inequality (13) can be represented as

$$|R_D| < 2\pi f L \min(f_1 Q_{u1}, f_2 Q_{u2}) \qquad (14)$$

where Q_{u1} and Q_{u2} are, respectively, the tank circuit unloaded quality factor values at f_1 and f_2. The quantity Q_{ui} (i=1, 2) is determined as follows

$$Q_{ui} = \sqrt{L/C_{ti}} \,/\, r(f_i)$$

where C_{t1}=$C_{t,\max}$ and C_{t2}=$C_{t,\min}$ are, respectively, the maximum and minimum total tank circuit capacitance in the tuning frequency range (f_1, f_2), $r(f_i)$ is the total series ac resistance of the tank circuit representing the tank circuit losses at frequency f_i. The total capacitance of the tank circuit is found as

$$C_t = 0.5C_V + C_{par} - C_{eff}$$

where C_V is the capacitance of one varactor, C_{par} is the parasitic capacitance including the terminal capacitances of transistors T_1-T_5, ($-C_{eff}$) is the effective negative capacitance induced into the tank circuit due to the action of capacitor C_1 (C_{eff}=0 when C_1=0). For the VCO circuit of Fig. 4, the resistance $r(f_i)$ is given by

$$r(f_i) = r_L(f_i) + 2r_V(f_i)$$

where $r_L(f_i)$ and $r_V(f_i)$ are the coil and varactor diode ac resistance at frequency f_i, respectively. If $r_L(f_1)\approx r_L(f_2)$ and $r_V(f_1)\approx r_V(f_2)$, the start-up condition (14) reduces to

$$r = r_L + 2r_V < L/\left[\left(0.5C_{V\max} + C_{par} - C_{eff}\right)|R_D|\right] = |r_{D\min}|$$

where $r_{D\min}$ is the minimum negative resistance induced into the tank circuit for compensating its losses.

V. Simulation Results

In order to demonstrate the efficiency of the proposed VCO, the circuit of Fig. 4 was simulated using SPICE parameters of real components. An *N*-channel JFET BF245C was selected as transistor T_5. As is well known, in the current mirrors are preferably used the long-channel MOS transistors, because the drain current in these transistors is less dependent on the drain-source voltage than in the short-channel devices. Therefore, four *p*-channel MOS transistors with channel length L_{ch}=1.2µm and channel width W_{ch}=10 µm (10 µm×20) were selected as transistors T_1-T_4. It can easily be calculated that the MOS transistors with L_{ch}=1.2 µm have a cutoff frequency of a few GHz, which allows them to be used over the entire frequency range of interest. The data used for SPICE models of p-channel MOS transistors corresponded to a standard 1.2µm CMOS technology. The high-Q varactors BB202 were selected to change the VCO frequency from f_1 to f_2. The selected inductor of the IC1210R10J type had a value of 100nH. The capacitances C_1=10pF and C_2=10nF were selected of the ECDG0E100C and C1005JB1E103K type, respectively.

The simulated I-V characteristic plotted with R_1=5.6kΩ (type ERJ1GEJ562) and R_2=47kΩ (type ERJ1GEJ473) is shown in Fig. 2, where V_{DD1}=1.75V, I_{t1}=36µA, V_{DD2}=5.75V, I_{t2}=0.88mA, V_{DD3}=29.25V, and I_{t3}=0.56mA. As seen, the dependence of current versus voltage in the NDR region is almost linear, which is not possible to reach in the known two-terminal NDR devices. Moreover, the NDR region is very long extending for more than 20V. The operating point Q was chosen at the beginning of the NDR region with V_{DDQ}=7.5V and I_{tQ}=0.83mA. Thus, the power consumption is only 6.2mW.

The VCO tuning characteristic is illustrated in Fig. 5. The VCO frequency is from 139.8 MHz to 256.2MHz when the dc voltage, V_{ctr}, applied to L_s=18µH varies from 1V to 6V resulting in the varactor capacitance change from $C_{V\max}$≈20pF to $C_{V\min}$≈5pF.

The simulated phase noise versus offset frequency at V_{ctr}=4V is shown in Fig. 6. As seen, the VCO shows an excellent phase noise of -111.1dBc/Hz at 10kHz offset.

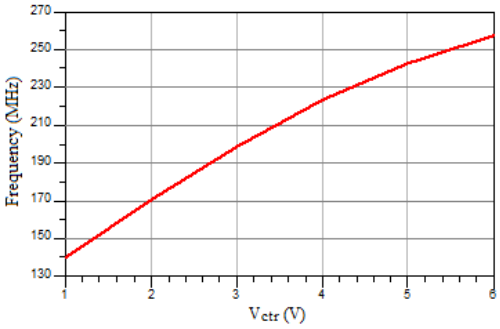


Figure 5. VCO tuning characteristic.

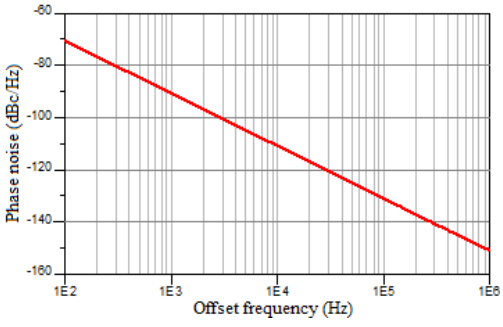


Fig. 6. Phase noise versus offset frequency.

For evaluating the overall performance of the proposed VCO, a common FOM is used [1]

$$FOM = L_N(\Delta f) - 20\log(f_0/\Delta f) + 10\log(P_{diss}/1\text{mW})$$

where $L_N(\Delta f)$ is the phase noise at a frequency offset Δf, f_0 is the central oscillation frequency, and P_{diss} is the power dissipation.

In Table 1, the phase noise, power consumption and *FOM* of the designed VCO are compared with the same characteristics of earlier published the best VCOs in VHF/UHF band.

TABLE I. COMPARISON OF DIFFERENT VCOS

Reference	Frequency MHz	Phase noise dBc/Hz	P_{diss} mW	*FOM* dBc/Hz
[8]	597	-92.5@10kHz	27	-173.7
[9]	275	-111@10kHz	75	-181.0
[10]	433	-99.7@100kHz	1.7	-171.3
[11]	903	-100.4@100kHz	4.5	-179.5
[12]	267	-87@10kHz	14	-164.0
This work	222	-111.1@10kHz	6.2	-190.0

VI. EXPERIMENTAL RESULTS

The VCO circuit was implemented using an *n*-channel JFET MMBFJ310LT1 and four p-channel MOS transistors FD302P as shown in the photograph of Fig. 7. The values of resistors R_1 and R_2 were selected as 1kΩ and 4kΩ, respectively. The measured and smoothed *I-V* characteristic is shown in Fig. 8. To isolate the VCO output from the oscilloscope input, a unity gain buffer was used. The VCO waveform at a frequency of 166.7MHz is shown in Fig. 9.

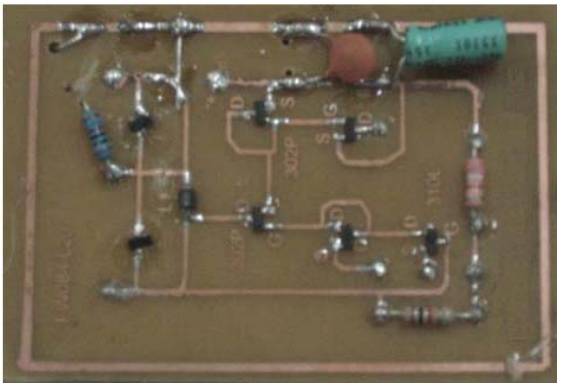


Fig. 7. Printed circuit board photograph.

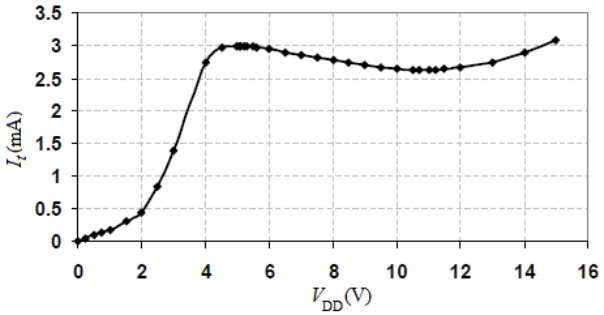


Fig. 8. Measured *I-V* characteristic of the implemented VCO.

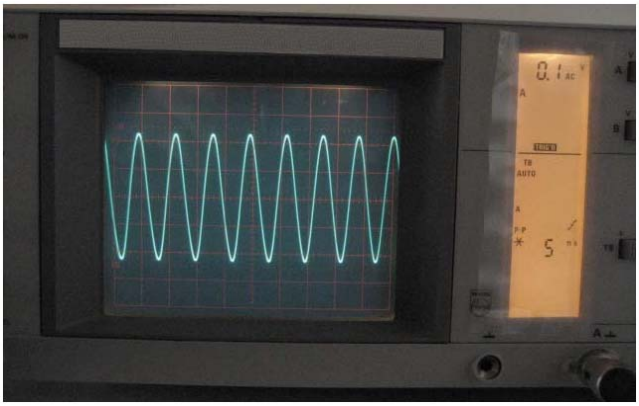


Fig. 9. The VCO output voltage waveform.

VII. CONCLUSION

A novel NDR VCO has been proposed for VHF range. It consists of a JFET and a *p*-channel MOS improved Wilson CM. The VCO has an N-type *I-V* characteristic with almost linear and precisely controllable NDR region. The VCO *I-V* characteristic has been mathematically modeled. The designed VCO has been simulated with Spice files of real components. It showed phase-noise of -111.1dBc/Hz at 10kHz offset from 222 MHz carrier and suppression of harmonics better than -31dBc. With power consumption of 6.2mW, the VCO achieves FOM of -190dBc/Hz, which is one of the best in the VHF/UHF band.

REFERENCES

[1] A. Grebennikov, "RF and microwave transistor oscillator design," John Willey&Sons, Inc., 2007, 441p.

[2] E. Alekseev and D. Pavlidis, "Large-signal microwave performance of GaN-based NDR diode oscillators," Solid-State Elect., № 44, pp. 941-947, 2000.

[3] L. Wang, "Reliable design of tunnel diode and resonant tunneling diode based microwave sources," PhD Thesis, Univ. of Glasgow, 2012, 222 p.

[4] K. J. Gan, C.C. Hsiao, and C. S. Tsai, "A novel voltage-controlled oscillator design by MOS-NDR devices and circuits," in Proc. of Fifth Int. Workshop on System-on Chip for Real Time Applications, pp. 372-375, 2005.

[5] I.W. Stanley and D.J. Ager, "Two-terminal negative dynamic resistance," Elect. Lett., vol. 6, is. 1, pp. 1-2, 1970.

[6] U. Kumar, "Simulation of a novel bipolar-FET type-S negative resistance circuit," Active and Passive Elec. Comp., vol. 26, pp. 129-132, Sept. 2003.

[7] S.M, Sze and Kwok K. Ng, "Physics of semiconductor devices," 3rd ed., Willey&Sons Inc., New Jersey, 2007, 815 p.

[8] R.T. Bosley, "A VHF/UHF voltage-controlled oscillator in 0.5μm BiCMOS," MSc Thesis, Virginia Polytech. Inst., 2003, 127 p.

[9] "V350ME12-LF VHF-band VCO", Microwave Journal, Jan. 17, 2011.

[10] Z. Ren, Y. Shi, Y. Wang, and Y. Tao, "A low phase noise VCO using enhanced LC resonator in a single-dip ASK receiver," in Proc. of IEEE 6th Int. Conf. on ASIC (ASICON 2005), 2005, pp. 611-615.

[11] K.H. Koo, S.W. Yoon, and C.W. Kim, "A 34% tuning range CMOS VCO with linear frequency tuning," in RWS`09 Proc. of the 4th Int. conf. on Radio and wireless symposium, IEEE Press, 2009, pp. 279-282.

[12] J.D. van der Tang and D. Kasperkovitz, "A low-phase-noise reference oscillator with integrated pMOS varactors for digital satellite receivers," IEEE J. of Sol.-State Circuits, vol. 35, no. 8, pp. 1238-1243, Aug. 2000.

Optimization of LC Voltage-Controlled Oscillators in 90-nm CMOS Technology for 3G Transceivers

V.V. Ulansky
Department of Electronics
National Aviation University
Kyiv, Ukraine
vulanskyi@yahoo.com

H.M. Elsherif
Department of Electrical and Electronic Engineering
Tripoli University
Tripoli, Libya
hassan_elsherif@hotmail.com

***Abstract*—This paper describes the design and optimization of *LC* voltage-controlled oscillators (VCOs) for 3G transceivers in 90-nm, 1.0-V CMOS technology. The proposed technique is illustrated by optimizing eight different MOS *LC* VCOs in the uplink band of 2500 to 2570 MHz. The best optimized VCO exhibits a maximum in-band phase noise of –90 dBc/Hz and –145 dBc/Hz, respectively, at 10 kHz and 10 MHz offset from carrier, 0.16 mW power consumption, and suppression of the second and third harmonics better than -40 dBc in the tuning range.**

Keywords—voltage-controlled oscillator; phase-noise; power consumption; harmonic distortion; spiral inductor

I. INTRODUCTION

3G technology is a high speed data transmission standard used by wireless operators in many countries. This technology has increased bandwidth and transfer rates for accommodating Web-based applications. Ukrainian 3G mobile network operators have deployed WCDMA technology in the 2100 MHz frequency band. Future developments in 3G technology in Europe are expected in the 2600 MHz frequency band, which is divided by uplink band (2500–2570 MHz) and downlink band (2620–2690 MHz). With a channel spacing of 5 MHz, the total bandwidth allocated for future 3G mobile services is 140 MHz, which will result in having a total number of 28 channels. WCDMA technology utilizes a bandwidth-efficient quadrature phase shift keying modulation scheme, which provides higher bit rates per unit of bandwidth. However, the modulation scheme efficiency can be degraded due to the phase noise of driving oscillator (DO) in transmitter and local oscillator (LO) in receiver. The phase noise in DO and LO can increase the error vector magnitude (EVM), which is one of the key performance characteristic of any 3G transceiver. For example, the specifications of transceiver ADF4602 (Analog Devices) guarantee that EVM $\leq$ 5% for transmitter, and EVM $\leq$ 8% for receiver [1]. The phase-locked loop (PLL) of a 3G transceiver has to be able to provide frequencies in the uplink and downlink bands performing both DO and LO functions. In respect to phase noise, the crucial PLL component is the VCO. The main concern for VCO design is a low phase noise, low power consumption and low energy level of harmonics of the fundamental frequency. For example, PLL synthesizer ADF4360–0 (Analog Devices), operating in the frequency range of 2400 to 2725 MHz, has VCO phase noise of –83 dBc/Hz and –145 dBc/Hz, respectively, at 10 kHz and 10 MHz offset from carrier, power consumption of 30 mW and suppression of the 2nd and 3rd harmonics better than –30 dBc and –39 dBc, respectively [2].

There are four commonly used MOS cross-coupled *LC* VCO topologies namely NMOS, PMOS, CMOS, and differential Colpitts. The differential Colpitts *LC* VCO topology can be based on NMOS or PMOS transistors [3], [4]. The CMOS *LC* VCOs have less performance because they require greater power supply voltage due to the series connection of the PMOS and NMOS transistors, which results in greater power consumption [5]. In published papers dealing with optimization of integrated MOS *LC* VCOs, for example [6]–[10], the oscillators were designed and optimized without taking into account the amount of distortion in the generated sinusoidal signals, which is important for 3G applications. Moreover, the list of design variables did not include some of the variables significantly affecting the characteristics of the generators. Besides that, the performance of the designed VCOs was not compared with the performance of VCOs having other topologies in the same CMOS technology. In this paper, the proposed optimization technique is illustrated by optimizing eight different MOS cross-coupled *LC* VCO topologies for perspective 3G applications.

II. VOLTAGE-CONTROLLED OSCILLATOR TOPOLOGIES

The differential cross-coupled *LC* VCO circuits, which are optimized in this paper, are shown in Fig. 1–4.

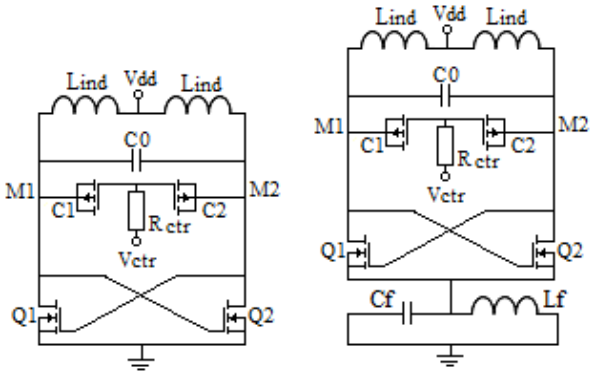


Fig. 1. NMOS differential cross-coupled VCO: (a) basic topology, (b) topology with a tail filter.

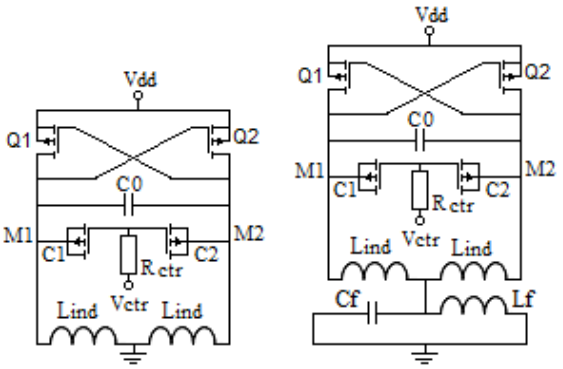


Fig. 2. PMOS differential cross-coupled VCO: (a) basic topology, (b) with a tail filter.

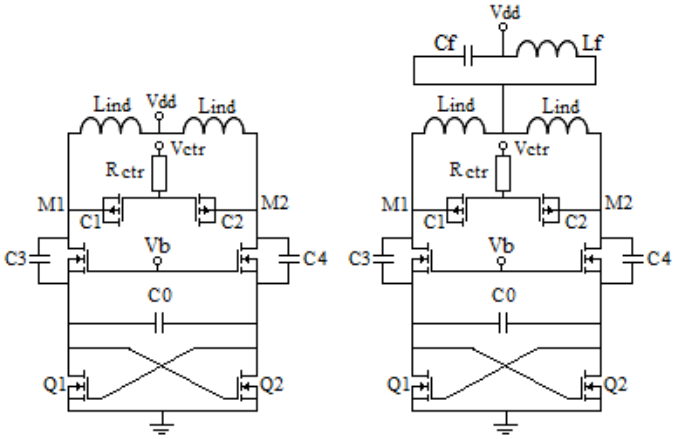


Fig. 3. NMOS differential Colpitts VCO: (a) basic topology, (b) with a tail filter.

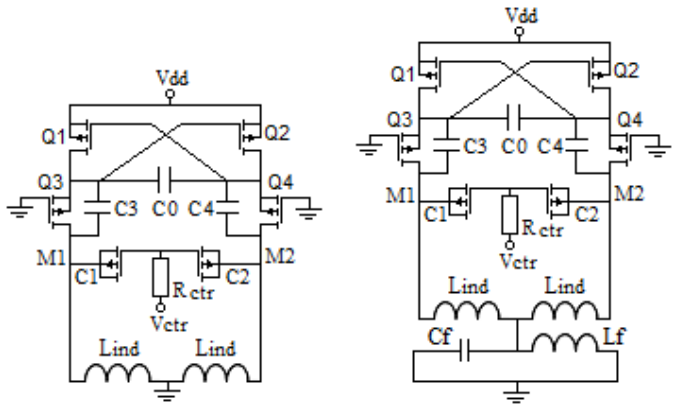


Fig. 4. PMOS differential Colpitts VCO: (a) basic topology, (b) with a tail filter.

Basic topologies of being compared MOS *LC* VCOs are shown in Fig. 1(a)–4(a). The MOS *LC* VCO topologies with a passive tail filter are shown in Fig. 1(b)–4(b). A tail filter technique was proposed in [11] for reducing phase noise and amplitude of the second harmonic in the NMOS *LC* VCO. In this paper, the technique is also applied to the PMOS and Colpitts *LC* VCOs.

In all topologies shown in Fig. 1 – 4, capacitor C_0 is added to reduce the VCO tuning band because the required band is only 70 MHz for uplink and down link.

In the VCO circuits of Fig. 1–4, transistors ($Q_1 - Q_4$) and (M_1 and M_2), are respectively, the core and varactor diode transistors. PMOS transistors M_1 and M_2 represent the contrary connected varactors controlled by a variable dc voltage V_{ctr} applied through the tuning source resistor R_{ctr}.

It should be noted that the NMOS differential Colpitts *LC* VCOs shown in Fig. 3 require additional supply voltage V_b for biasing transistors Q_3 and Q_4.

III. Figure of Merit

A common FOM for performance comparison of different *LC* VCOs is given by [6]

$$FOM_1 = L(\Delta f) - 20\log(f_0/\Delta f) + 10\log(P_{diss}/1\text{mW})$$

where $L(\Delta f)$ is the phase noise at a frequency offset Δf, f_0 is the central oscillation frequency in the VCO tuning band, and P_{diss} is the VCO power dissipation. Lower (more negative) values of FOM represent better VCO performance. The main drawback of FOM_1 is that it does not take into account the amount of distortion in the generated sinusoidal signal due to the higher harmonics presenting in the oscillator output signal spectrum, especially the second and third harmonics. The proposed FOM is expressed as

$$FOM_2 = L(\Delta f) - 20\log(f_0/\Delta f) + 10\log(P_{diss}/1\text{mW}) + 20\log\left(\sqrt{V_{m,2}^2 + V_{m,3}^2}/V_{m,1}\right) \quad (1)$$

where $V_{m,i}$ $(i = 1, 2, 3)$ is the amplitude of the i^{th} harmonic of the generated signal. The last term in (1) takes into account the amount of distortion introduced into the generated signal by the 2nd and 3rd harmonics. Thus, the VCOs with larger amplitudes of the 2nd and 3rd harmonics will have less performance than those having smaller distortions.

IV. Design Variables

For given CMOS technology with fixed transistor channel length, which is 90nm in our case, the VCO phase noise, power consumption and amount of distortions depend on the transistor channel width (W), tank circuit inductor (L_{ind}) and varactors (C_1 and C_2) performance, resistance of the tuning port (R_{ctr}), capacitances C_0, C_3 and C_4, and tail filter capacitance and inductance (C_f and L_f).

The standard way to design an on-chip spiral inductor is to lay out square spiral metal trace on the silicon substrate with one or more standard metal interconnection layers. The design parameters of a spiral inductor are usually given by the number of turns (n), metal width (w), metal-to-metal spacing (s), and outer radios (h). The inductor dimensions w, s, and h can be optimized by using computer aided tools such as ASITIC, MICROWIND, SONET, ADS Momentum, and etc.

The inductor frequency dependent series resistance $R_0(f)$ has the biggest impact on the quality factor Q_{ind}, which in turn seriously affects the VCO performance. For reducing $R_0(f)$ and increasing Q_{ind} the multilayer spiral inductors will be used in optimizing selected VCOs.

An analytical equation for frequency dependence of the resistance of a multi-turn spiral inductor with taking into account the skin and proximity effects is given by [12]

$$R_0(f) = R_{0dc}\left(1 + 0.1\,\omega^2 / \omega_{crit}^2\right) \quad (2)$$

$$\omega_{crit} = 3.1\frac{P \cdot R_{sheet}}{\mu_0 \cdot w^2}$$

where R_{0dc} is the dc resistance of the coil, w is the metal width, $P = w + s$ is the metal pitch, R_{sheet} is the metal sheet resistance, $\mu_0 \approx 1.257\times10^{-6}$ H/m is the magnetic permeability of the free space.

The dc resistance R_{0dc} of a multilayer inductor included into (2) can be simulated by CAD tools, for example, MICROWIND.

Equation (2) will be used when designing the spiral inductors and optimizing VCOs shown in Fig. 1 – 4.

As well known, a MOS transistor with drain, source and bulk connected together realizes a MOS capacitor with capacitance dependent on the voltage applied between source and gate. If the MOS transistor channel length is fixed, the varactor capacitance is proportional to the channel width W_{var}. However, the channel width of a MOS varactor can reach many hundreds of micrometers significantly increasing the parasitic resistance along the gate, which in turn reduces the varactor quality factor. Such transistors must be designed with big number of fingers (N_v), so that $W_{\text{var}} = N_v W_v$, where W_v is the channel width of a split varactor transistor. Thus, two new variables W_v and N_v have been introduced.

It should be pointed out that the core VCO transistors, labeled as Q_i (i=1, …, 4) in the VCO circuits of Fig. 1–4, must also be decomposed into multi-finger transistors with gates of minimum size such that $W = N_{tr}W_{tr}$ for the same reason as varactors, where N_{tr} and W_{tr} are, respectively, number of fingers and channel width of a split core VCO transistor.

The value of resistor R_{ctr} has certain influence on both VCO phase noise and amount of distortions. In general, when R_{ctr} increases the VCO generated signal distortions decrease, but the VCO phase noise increases due to the resistance thermal noise induced to the VCO through the tuning port. Thus, there should be a value of R_{ctr} which minimizes FOM_2. So, one more design variable has been introduced.

The last two variables are the capacitance and inductance of the tail filter C_f and L_f. It should be pointed out that commonly used FOM_1 is not sensitive to harmonic distortions. That is why C_f and L_f are not included into the set of design variables in [6-10]. On the contrary, FOM_2 is sensitive to the level of the 2nd and 3rd harmonics of the generated signal.

V. Optimization Criterion

The proposed VCO optimization criterion is presented as follows:

$$\min FOM_2\left(W_{tr}, N_{tr}, W_v, N_v, R_{ctr}, C_0, L_{ind}, C_f, L_f\right) \quad (3)$$

$$V_m \geq V_m^*, f_{osc,\min} \leq f_{\min}^*, f_{osc,\max} \geq f_{\max}^*$$

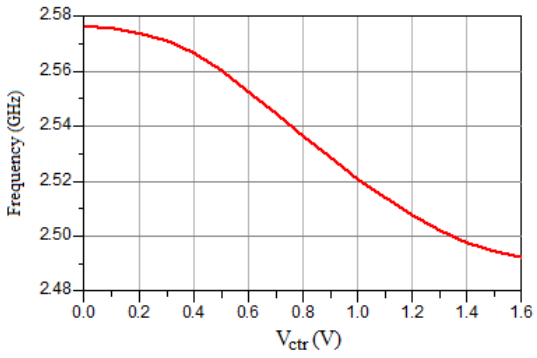
where V_m is the voltage amplitude across the tank, V_m^* is the minimum required output voltage amplitude, $f_{\min}^*$ and $f_{\max}^*$ represent the required frequency band of the VCO operation including additional bands, which must be used due to unavoidable process variations during the integrated circuit fabrication. The additional bands should be about 10-20% of the VCO frequency tuning band. In the case of WCDMA transceiver, we can choose $f_{\min}^* = 2488\,\text{MHz}$ and $f_{\max}^* = 2582\,\text{MHz}$. The first constraint ($V_m \geq V_m^*$) indicates that the oscillator must have an excess of the loop gain required to compensate all losses at resonance providing sufficient voltage amplitude across the tank. A VCO with small amount of excess gain results in $V_m < V_m^*$. The VCO with large amount of excess gain satisfies the first condition, that is $V_m > V_m^*$, but it generates a greater harmonic content in the output waveform. The VCO designer should balance the need for enough excess gain providing $V_m \geq V_m^*$ and the need to keep the harmonic content low. This feature is taken into account in the criterion (3), because FOM_2 is dependent on the amplitude of the 2nd and 3rd harmonics.

VI. Simulation Results

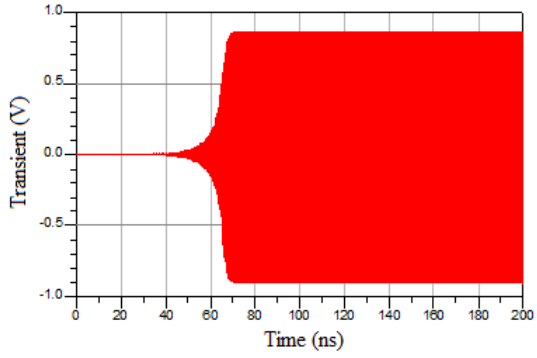
All MOS transistors were simulated using SPICE BSIM4v4.3 parameters of SCN009 90nm CMOS technology [13]. The designed three metal layers (metals 4, 5 and 6) spiral inductor of 6.5nH has the following geometric parameters: n=5, w=25μm, s=2μm and h=48μm. The value of V_m^* was chosen to be 80% of the supply voltage, that is 0.8 V. Since $V_m^* < V_{dd} = 1\,\text{V}$, the VCOs operate in the current-limited regime [3]. The results of simulation of the PMOS Colpitts VCO with LC tail filter, and W_{tr}=1μm (N_{tr}=10) and W_v=1μm (N_v=53) are shown in Fig. 6. The values of resistor R_{ctr} and capacitors C_0 and C_3 (=C_4) were selected to be 40 Ω, 0.43 pF, and 1.5 pF, respectively. The tail filter capacitance and inductance are, respectively, 0.5 pF and 6.5 nH. As shown in Fig. 5(a), the oscillation frequency of the VCO covers range from 2.493 to 2.577 GHz with tuning voltage variation from 1.6 to 0 V. The starting-up waveform of the VCO at V_{ctr}=0.8 V is shown in Fig. 5(b). The plots of the 1st, 2nd, and 3rd harmonic amplitudes versus tuning voltage are shown in Fig. 5(c). As seen, the designed VCO has very low level of 2nd and 3rd harmonics, which are suppressed to better than –40 dBc.

From Fig. 5(d) follows that the VCO phase noise is from –123 dBc/Hz to –125 dBc/Hz at 1 MHz offset from carrier in the tuning frequency band. It can also be seen that the maximum in-band phase noise is –90 dBc/Hz and –145 dBc/Hz, respectively, at 10 kHz and 10 MHz offset from carrier. As seen from Fig. 5(e), the dc current consumed from the 1 V power supply is 155.5μA. So, the power dissipation of the designed VCO is given by $1\text{V} \times 0.16\text{mA} = 0.16\text{mW}$.

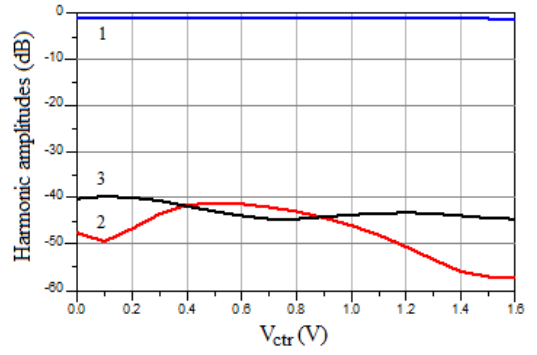
The results of simulation for the PMOS Colpitts VCO without a tail filter are shown in Fig. 6. Comparing Fig. 5(c) and 6(a), it can be noted that the PMOS Colpitts VCO without a tail filter has the amplitude of the 2nd harmonic greater at least by 18dB compared to that of the oscillator with the tail filter. Besides that, from comparison of Fig. 6(d) and 6(b), follows that at small offset frequencies from carrier the VCO without a tail filter has greater phase noise compared to the counterpart with the tail filter. At big frequency offsets from carrier, both VCOs have almost identical phase noise.



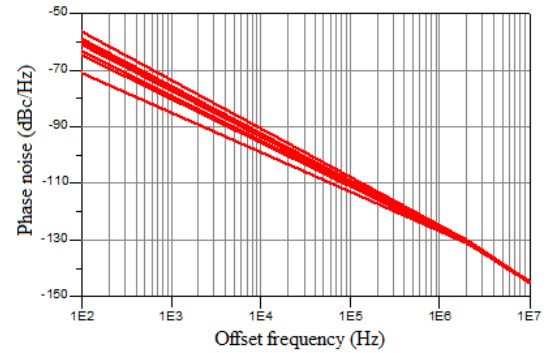
(a)



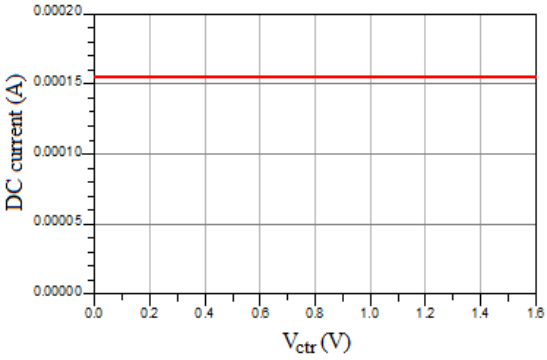
(b)



(c)

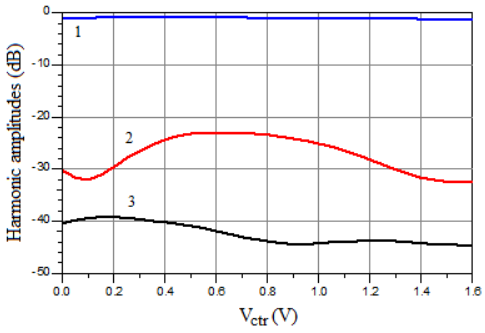


(d)

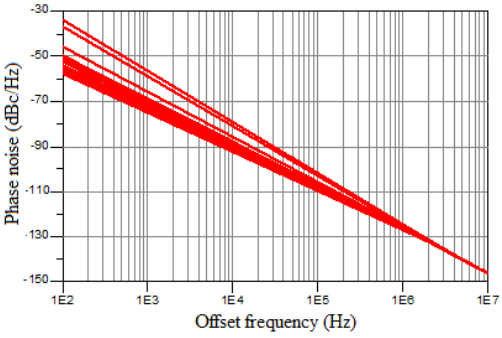


(e)

Fig. 5. Simulation results for the PMOS Colpitts VCO with tail filter: (a) VCO tuning characteristic, (b) transient response, (c) dependence of the 1st, 2nd and 3rd harmonic amplitude versus tuning voltage, (d) phase noise versus offset frequency at different voltage V_{ctr}, (e) consumed dc current versus tuning voltage.



(a)



(b)

Fig. 6. Simulation results for the PMOS Colpitts VCO without tail filter: (a) dependence of the 1st, 2nd and 3rd harmonic amplitude versus tuning voltage, (b) phase noise versus offset frequency at different voltage V_{ctr} .

VII. Optimization Results

Optimization of the VCOs was carried out using computer-aided design tools. The optimization results of the VCOs shown in Fig. 1 – 4 are given in Table 1. As seen, the best performance has the PMOS Colpitts VCO with the tail filter. It should be noted that the PMOS VCO with the tail filter has also good performance in terms of FOM_2 which is even 1dB higher than that of the PMOS Colpitts VCO without a tail filter. The NMOS VCOs have much less performance and cannot be proposed for realization in 3G transceivers.

TABLE I. VCO Optimization Results

VCO	W µm	W_{var} µm	P_{diss} mW	Phase noise dBc/Hz @1MHz	FOM_1 dBc	FOM_2 dBc
NMOS	1×1	69×1	0.83	-103	-172	-190
NMOS+F	1×1	69×1	0.83	-104	-173	-208
PMOS	3×1	64×1	1.19	-119	-186	-203
PMOS+F	4×1	62×1	1.55	-122	-188	-223
NCOLP.	2×1	60×1	1.18	-102	-169	-188
NCOLP.+F	2×1	60×1	1.18	-110	-177	-201
PCOLP.	9×1	53×1	0.14	-124	-201	-222
PCOLP.+F	10×1	53×1	0.16	-123	-199	-237

In Table 2, the phase noise, power consumption and FOM_1 of the designed PMOS Colpitts VCO with *LC* tail filter are compared with the same characteristics of earlier published the best CMOS VCOs. As seen in Table 2, the designed VCO has very good performance in terms of FOM_1.

TABLE II. Comparison of Different CMOS VCOs

Reference	Frequency GHz	Phase noise dBc/Hz	P_{diss} mW	FOM_1 dBc/Hz
[2]	2.563	-145@10MHz	30	-178.4
[3]	2.9	-142@3MHz	16	-189.5
[14]	3.7	-156@20MHz	22.8	-187.8
[15]	3.2	-117@1MHz	1.4	-185.5
[16]	25	-106@1MHz	6.6	-185.8
This work	2.535	-123@1MHz	0.16	-199.0

VIII. Conclusion

New criterion has been proposed for optimizing cross–coupled MOS VCOs. It is sensitive not only to the VCO phase noise and power consumption but also to the harmonic distortions of the generated signals. The design variables have been selected and analyzed. It has been shown that the PMOS Colpitts VCO with *LC* tail filter has the best performance among eight examined topologies for 3G applications.

References

[1] http://www.analog.com/en/rfif-components/rfif-transceivers/adf4602/products/product.html

[2] http://www.analog.com/en/rfif-components/pll-synthesizersvcos/adf4360-0/products/product.html

[3] P. Andreani, X. Wang, L. Vandi, and A. Fard, "A study of phase noise in Colpitts and *LC*-tank CMOS oscillators," IEEE J. of Sol.-St. Circ., vol. 40, no. 5, pp. 1107-1118, May 2005.

[4] C. Lee, C. Lu, and H. Chuang, "A 16-GHz 0.18-µm CMOS differential Colpitts VCO with a moderate output power for DS-UWB and 60-GHz applications," in Proc. of Asia-Pacific Microwave Conf., pp. 1132-1135, Dec. 2006.

[5] V. Ulansky, H. Elsherif, E. Aboadla, and I. Machalin, "The design and optimization of a 2.4-GHz integer-N frequency synthesizer in 0.18µm CMOS technology," in Proc. of 2010 IEEE Statistical Methods of Signal and Data Processing Conf., Kyiv, Ukraine, pp. 132-135, Sept. 2010.

[6] D. Ham, A. Hajimiri, "Concepts and methods in optimization of integrated *LC* VCOs," IEEE J. of Sol.-St. Circ., vol. 36, no. 6, pp. 896–909, June 2001.

[7] S. Ryu, "Design issues of CMOS VCO for RF transceivers," J. of the Korean institute of electromagnetic engineering and science", vol. 9, no. 1, pp. 25–31, Mar. 2009.

[8] H. J. Yao, "A comparative study of low phase noise VCOs in CMOS technology," MSc Thesis, Concordia University, 2006, 116 p.

[9] I. Krout, H. Muif, M. Fakhfakh, M. Loulu, "Optimising LC VCOs performances through a heuristic," WSEAS Transactions on Electronics, is. 6, vol. 5, pp. 275-281, June 2008.

[10] S. Arabshahi, M. Dousti, and H. Farahani, "Design of a CMOS VCO for RFID system," Int. J. of Eng. Research and Appl., vol. 2, is. 6, pp. 1386-1390, Dec. 2012.

[11] E. Hegazi, H. Sjoland, and A. Abidi, "A filtering teachnique to lower LC oscillator phase noise," IEEE J. of Sol.- St. Circ., vol. 36, no. 12, pp. 1921-1930, Dec. 2001.

[12] W. B. Kuhn and N. M. Ibrahim, "Analysis of current crowding effects in multiturn spiral inductors," IEEE Trans. Microwave Theory Tech., vol. 49, pp. 31–38, Jan. 2001.

[13] ftp://ftp.isi.edu/pub/mosis/vendors/ibm-90/t96u_9lprf_9m_lb-params.txt

[14] P. Andreani, K. Kozmin, P. Sandrup, M. Nilsson, and T. Mattsson, "A TX VCO for WCDMA/EDGE in 90nm RF CMOS," IEEE J. of Sol.-St. Cir., vol. 46, is. 7, pp. 1618-1626, July 2011.

[15] K. Siwiec, T. Boryko, and A. Pleskacz, "PVT tolerant *LC*-VCO in 90nm CMOS technology for GPS/Galileo applications," in Proc. of 2011 IEEE 14th Int. Sym. on DDECS, Cottbus, pp. 29-34, April 2011.

[16] M. Tormanen, "A 25-GHz differential *LC*-VCO in 90-nm CMOS," in Proc. of IEEE APCCAS 2008 Conf., Macao, pp. 554-557, Dec. 2008.

Thermal Dependence of "Metal-Polymer" Type Composite Materials' Dielectric Properties in Ultra-high Frequency Band

Didenko Yu.V., Patsora I.V., Tatarchuk D.D., Franchuk A.S., Kharabet Ie.I.
Department of Microelectronics
National Technical University of Ukraine "KPI"
Kiev, Ukraine
didenko@dp.ntu-kpi.kiev.ua

***Abstract* — A simple relation, which can calculate the temperature dependence of the dielectric permittivity and the dielectric loss tangent of composite metal-polymer is obtained. These materials in the frequency range of 8…12 GHz are investigated. It is shown that the thermal properties of these composites are mainly determined by the properties of the matrix material. The obtained results can be used to predict the temperature dependence of the dielectric parameters of composite metal-polymer with different volume part of the metal phase.**

Keywords — composite material, dielectric permittivity, dielectric loss tangent

I. INTRODUCTION

One of the most effective ways to protect against electromagnetic radiation to date is still shielding [1, 2]. Because with this promising direction is to develop the shielding coatings based on metal-polymer composite.

Composite materials based on dielectric materials containing metal particles are of interest for the experimental and theoretical studies [3, 4]. The physical properties of these materials is fundamentally different from the properties of continuous media, made of the same materials. The properties of these composites depend on:

- the properties of the matrix;

- the material and size of the metal particles;

- the ratio of components.

Selection of the optimal ratio between components provides reception of composite materials with the desired magnetic, dielectric, radio absorbing and other special properties [2]. However, the use of such material impeded by several difficulties associated with not a complete study of their properties [5]. It concerns the temperature dependences of the electrophysical properties of such materials. The aim of this study is to investigate the temperature dependence of the dielectric permittivity and dielectric loss tangent of composite metal-polymer in the microwave range.

II. THE TEMPERATURE DEPENDENCE OF THE DIELECTRIC PROPERTIES OF COMPOSITE METAL-POLYMER IN THE ULTRA-HIGH FREQUENCY BAND

Composite metal-polymer is a system consisting of metallic granules, randomly distributed in a dielectric matrix (Fig. 1).

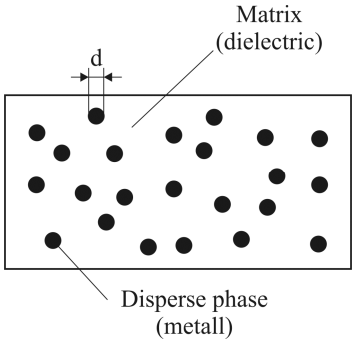


Fig. 1. The schematic representation of the composite metall-polymer (d is the size of metall particles)

It is known that the dielectric permittivity of these materials are determined by their composition and can be approximately calculated based on the effective medium theory of Maxwell-Garnett [6]. Assuming that the metal particles are spherical, the formula for determining the dielectric permittivity of the composite material is as follows:

$$\varepsilon_c = \varepsilon_m + \frac{3q\varepsilon_m\left(\varepsilon_d - \varepsilon_m\right)}{\left(\varepsilon_d + 2\varepsilon_m\right) - q\left(\varepsilon_d - \varepsilon_m\right)}, \qquad (1)$$

where ε_c is the complex dielectric permittivity of composite material, ε_m is the complex dielectric permittivity of matrix material (in this case polymer), ε_d is the complex dielectric permittivity of disperse phase material (in this case metal), q is the metal volume part in composite material.

On the other hand the temperature dependence of the dielectric permittivity of the material is determined by the

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

temperature coefficient of dielectric permittivity, which is given by:

$$\mathrm{TK}\varepsilon_c = \frac{1}{\varepsilon_c}\frac{d(\varepsilon_c)}{d(T)}, \quad (2)$$

where $\mathrm{TK}\varepsilon_c$ is the temperature coefficient of the composite's dielectric permittivity, T is the temperature.

Substituting (1) into (2) and differentiate. Consider that $|\varepsilon_d| >> |\varepsilon_m|$, $q < 1$, and after the cancellation of like terms get:

$$\mathrm{TK}\varepsilon_c \approx \left(\frac{1-2q}{1+2q}\right)\mathrm{TK}\varepsilon_m . \quad (3)$$

Similarly, the temperature dependence of the dielectric loss tangent of the material is determined by the temperature coefficient of dielectric loss tangent:

$$\mathrm{TKtg}(\delta_c) = \frac{1}{\mathrm{tg}(\delta_c)}\frac{d(\mathrm{tg}(\delta_c))}{d(T)}, \quad (4)$$

where $\mathrm{tg}(\delta_c)$ is the composite's dielectric loss tangent, and $\mathrm{TKtg}(\delta_c)$ is its temperature coefficient.

It is known, that dielectric loss tangent can be obtained as:

$$\mathrm{tg}(\delta_c) = \frac{\varepsilon_{c2}}{\varepsilon_{c1}}, \quad (5)$$

where ε_{c1} and ε_{c2} are real and imaginary parts of composite' dielectric permittivity respectively:

$$\varepsilon_c = \varepsilon_{c1} - j\varepsilon_{c2}, \quad (6)$$

Substituting (5) into (4), after transformations we obtain:

$$\mathrm{TKtg}(\delta_c) = \mathrm{TK}\varepsilon_{c2} - \mathrm{TK}\varepsilon_{c1}, \quad (7)$$

where $\mathrm{TK}\varepsilon_{c2}$ and $\mathrm{TK}\varepsilon_{c1}$ can be obtained from (3).

III. Results of Experimental Researches

To verify the theoretical results an experimental investigation of temperature dependence of the dielectric permittivity and dielectric loss tangent of composite materials in frequency range of 8...12 GHz is reported.

Composite materials were produced by electromechanical intermixing of disperse phase with matrix material at room temperature. Obtained material was formed into rectangular samples. Their size was chosen specifically so that the composite material filled section of the waveguide. The size of the waveguide in accordance with the frequency range was 23x10 mm (Fig. 2).

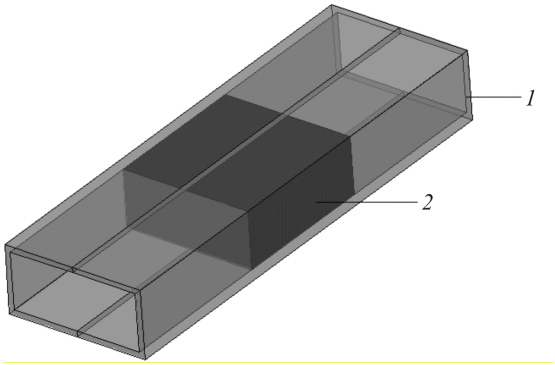


Fig. 2. The split waveguide with sample:
1 is the split waveguide; *2* is the test sample

The samples' thickness was 2 mm. Polymer on the base of epoxy with complex dielectric permittivity $\varepsilon = 2.73 - j\,0.2$ was used as matrix material. Aluminum nanopowder with particles' sizes d = 0.06…0.2 μm was used as disperse phase. Section with the sample was placed in a thermostat (Fig. 3).

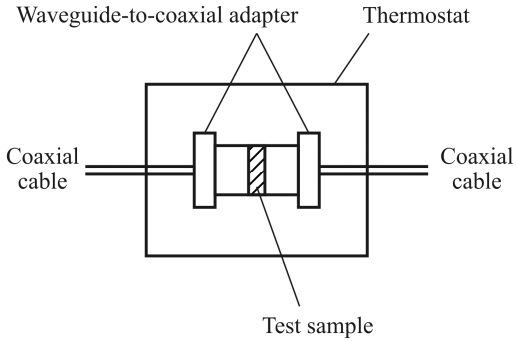


Fig. 3. The schematic representation of the thermostat with the test sample

Temperature control is carried out using a thermocouple TP-01, which provides an error of not more than 1%. Thermocouple is inserted through the hole in the narrow wall of the waveguide. Before taking the readings from the instrument sample was kept at a given temperature for 30 minutes. Then thermocouple was withdrawn from waveguide for a time of measurement. Through the coaxial inputs section with the sample was included in the circuit of panoramic meter, which consisted of a sweep generator (the frequency range 8…12 GHz, the frequency error no more than 0.1 %), an indicator of attenuation (the measurement error of attenuation no more than 0.01 %), two directional couplers with detectors, measuring unit and the matched termination. In accordance with the method of reflection-transmission coefficients [7] were measured transmission and reflection coefficients of the electromagnetic energy (the measurement scheme shown in Fig. 4a and 4b, respectively). From the measured values of these coefficients were calculated the dielectric permittivity and the dielectric loss tangent of the material at given temperatures. To minimize measurement errors used reference samples.

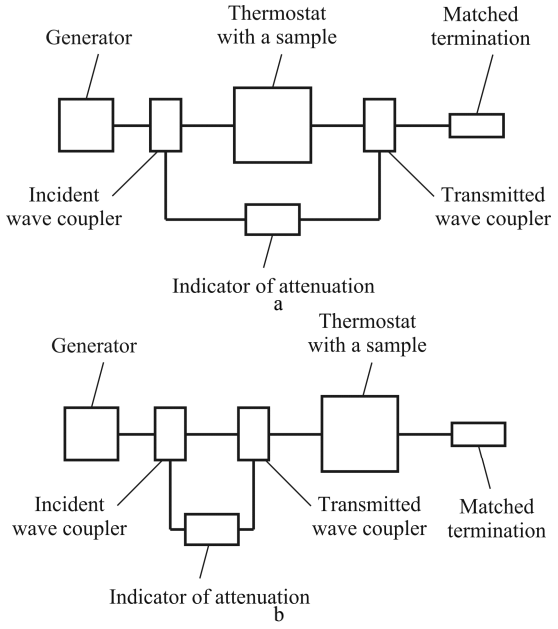


Fig. 4. Block diagram of measuring the transmission (a) and reflection (b) coefficients

Obtained theoretical and experimental results are represented on Fig. 5 and 6.

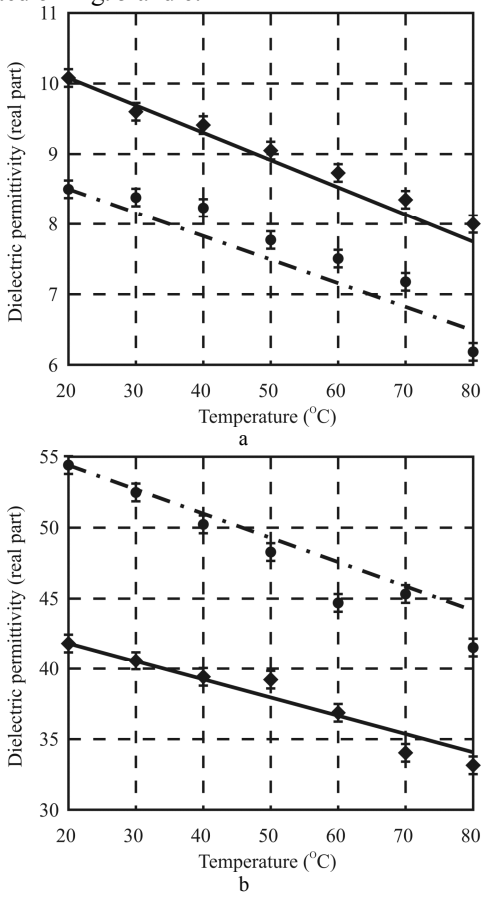


Fig. 5. Temperature dependence of dielectric permittivity's real part of composite materials with volume part $q = 0.1$ (a) and $q = 0.15$ (b):
——— – theory (Al); ◆ – experiment (Al);
– · – · – theory (Cu); ● – experiment (Cu)

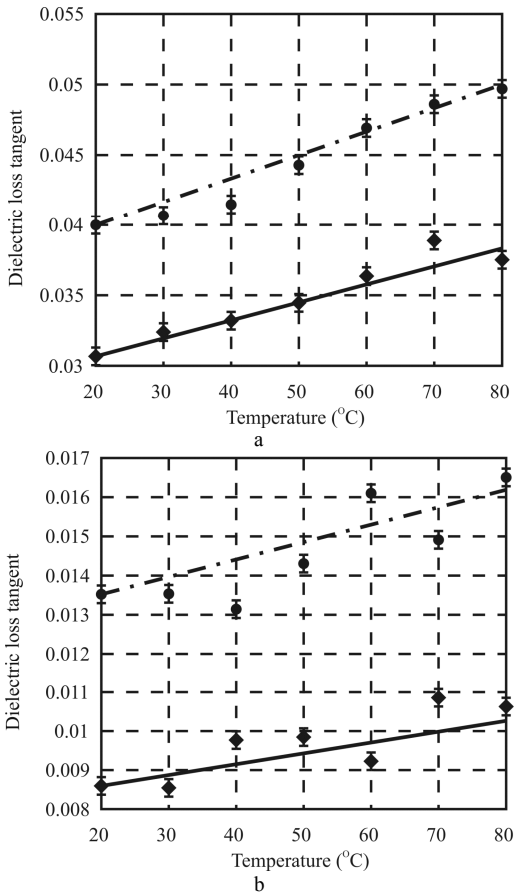


Fig. 6. Temperature dependence of dielectric loss tangent of composite materials with volume part $q = 0.1$ (a) and $q = 0.15$ (b):
——— – theory (Al); ◆ – experiment (Al);
– · – · – theory (Cu); ● – experiment (Cu)

Since the results of measurements at different frequencies of this range did not differ (within experimental error), then the figure shows the results only for the frequency of 10 GHz.

Analysis of the relations shows that the temperature coefficient of dielectric permittivity depends on the metal's volume part in the composite. So for samples with a volume part of aluminum $q = 0.1$ the temperature coefficient of the dielectric permittivity was $TK\varepsilon = -4.6 \cdot 10^{-3}$ °C^{-1}, and for $q = 0.15$ $TK\varepsilon = -4.5 \cdot 10^{-3}$ °C^{-1}. For composites containing copper, the increase of the temperature coefficient of the dielectric permittivity is expressed stronger: at $q = 0.1$ $TK\varepsilon = -6.6 \cdot 10^{-3}$ °C^{-1}, and for $q = 0.15$ $TK\varepsilon = -4.9 \cdot 10^{-3}$ °C^{-1}.

The dependence of the temperature coefficients of the dielectric loss tangent of the metal's volume part in the composite for the investigated material is weak. So for samples with a volume fraction of aluminum q = 0.1 temperature coefficient tan was $TKtg(\delta) = 2.1 \cdot 10^{-3}$ °C^{-1}, and for q = 0.15 $TKtg(\delta) = 3.3 \cdot 10^{-3}$ °C^{-1}. A sample containing copper, the temperature coefficient of dielectric loss tangent were: $TKtg(\delta) = 3.3 \cdot 10^{-3}$ °C^{-1} at q = 0.1, and $TKtg(\delta) = 3 \cdot 10^{-3}$ °C^{-1} at q = 0.15.

Thus, it is possible to provide the required values of material's temperature parameters, by selecting a metal's volume part in the composite.

As shown on Fig. 5 and 6, the difference between theoretical and experimental values of dielectric permittivity in whole measured specter is less than 3%, and for dielectric loss tangent this difference is about 7%.

IV. CONCLUSIONS

Theoretical equations for the temperature coefficients of dielectric permittivity and dielectric loss tangent of composite materials are obtained.

In the frequency range 8…12 GHz composite metal-polymer materials based on epoxy with the addition of aluminum and copper with different volume part of metal are experimentally investigated.

Using the method of transmission-reflection coefficients the temperature dependence of the dielectric permittivity and dielectric loss tangent of these materials are obtained.

By changing the proportion of the metal content can change the temperature coefficients of metal-polymer composites.

In the range 8…12 GHz dielectric permittivity and dielectric loss tangent of composite materials do not depend on the frequency.

In ultra-high frequencies band the temperature dependences of dielectric permittivity and dielectric loss tangent for metal-polymer composites, which include high conductivity metals as disperse phase, are determined mainly by matrix' properties.

The difference between theoretical and experimental values of dielectric permittivity in whole measured specter is less than 3%, and for dielectric loss tangent this difference is about 7%. Thus, expressions obtained in the theoretical part, can be used to predict the temperature dependence of the dielectric properties of the metal-polymer composites, for small values of the metal's volume part in the composite.

ACKNOWLEDGMENT

Authors are most grateful to Professor Yu.M. Poplavko for his help in analysis of experimental results.

REFERENCES

[1] A.V. Machulyansky, V.A. Popov, V.V. Pilinsky, D.D. Tatarchuk, V.B. Shvaychenko, "Electromagnetic shielding by coatings on the base of ultradisperse structures," Technical Electrodynamics, Special Issue "Power electronics and energy efficiency", vol. 5, pp. 23–25, Sept. 2009 (Russian).

[2] A.V. Machulyansky, V.V. Pilinsky, O.V. Telichkina, M.V. Rodionova, V.B. Shvaichenko, "Basics of application nanostructured materials for electromagnetic compatibility of electronic devices," Electronics and Communications, vol.3, pp. 9–15, 2010 (Russian).

[3] A.V. Machulyansky, D.D. Tatarchuk, V.A. Machulyansky, "Analysis of the microwave properties nanodispersed composite systems," Technical electrodynamics, vol. 1, pp. 303–304, 2010 (Russian).

[4] A. Machulyansky, D. Tatarchuk, O. Telychkina, Yu. Yakimenko, "Electrophysical properties of nanocomposite materials on the base of structures metal-dielectric," ISSE 2010: 33rd International Spring Seminar on Electronics Technology, May 12th–16th: thesis, Warsaw, Poland, pp. 249–250, 2010.

[5] N.G. Kokodiy, "Microwave radiation absorption very thin doublelayers cylinder," Radiotechnics and Electronics, vol. 51, №2, pp. 185–188, 2006 (Russian).

[6] A.P. Vinogradov, "Electrodynamics of composite materials," Edited by B.Z. Katsenelenbaum, Moscow: Editorial URSS, 2001 (Russian).

[7] Yu.M. Poplavko, V.I. Molchanov, V.A. Kazmirenko, "Microwave dielectric spectroscopy," Kyiv: NTUU "KPI", 2011 (Ukrainian).

Novel Concept for Microstrip Stub Resonant Frequency Control

Pavlo Sergienko,
Yuriy Prokopenko
College of Electronics
National Technical University of Ukraine "KPI"
Kiev, Ukraine
paulsergienko@gmail.com

Guy A. E. Vandenbosch
Department of Electrical Engineering (ESAT)
Katholieke Universiteit Leuven
Leuven, Belgium
guy.vandenbosch@esat.kuleuven.be

***Abstract*— A novel concept for controlling a low-cost tunable microstrip filter is proposed. The microstrip stub controlling is realized by detaching a part of the stub from the substrate. The air gap between the substrate and the signal electrode is the heterogeneity which controls the stub effective permittivity, in this way providing a resonance frequency shift. By moving the stub 50 microns the resonance frequency is shifted about 1 GHz. An equivalent circuit model of the resonator is developed. Calculation results are in good agreement with finite-difference time-domain method (FDTD) and finite elements method (FEM) simulations.**

Keywords—tunable; microstrip; stub; MEMS; electro-mechanical tuning

I. INTRODUCTION

Today, modern facilities of wireless communication and data transfer protocols put forward new device requirements for microwave technology. Multichannelness is one of these requirements. Alteration of the working frequency range is an enabling feature in this. The requirements of miniaturization and cost make microstrip lines one of the most important components in communication circuits. That is why development of microstrip controllable resonators is still under investigation.

There are several methods to control the resonant frequency of microstrip filters. This paper concerns the method of electromechanical tuning of microstrip filters.

Using varactors or other capacitors with variable capacitance is one of the most traditional ways for resonance frequency tuning [1-4]. This method is simple to implement, but provides a relatively low value of resonance frequency shift. Also, due to the presence of tunable semiconductor capacities it greatly increases the relative level of losses. As the frequency increases losses increase, which significantly limits the frequency range in which this method is efficient. A similar picture is observed when tuning is performed by p-i-n diodes. The tuning value is small, but the loss is less than with varactors [5-6].

In an attempt to avoid the use of varactors a number of different structural designs for tunable filters were proposed, like square wave shape elements and so on [7-9]. These methods eliminate the problem of the increasing losses with the increasing of the frequency. But often, such device is difficult to manufacture and for each band a new design has to be developed, which increases the cost of production. Also, because of the design complexity the loss increases due to radiation, which leads to a decrease of the quality factor. The publication [10] presents a method of cascade connecting several microstrip filters, which provides wide bandwidth and filter tuning opportunity by changing the working ranges.

Tunable filters based on Low Temperature Co-fired Ceramics (LTCC) technology are the most expensive and difficult to manufacture. They have a low insertion loss of 2.3 dB and very small sizes, e.g. 3.7 mm by 4.1 mm in [11], but provide lower resonance frequency shift than varactors and p-i-n diodes.

In contrast to expensive solutions, there are methods aimed at reducing the cost. For example magnetic tuning can be done with a simple cheap structure [12]. However, in comparison with other methods magnetic tuning has high insertion loss and low resonance frequency tuning. So, a compromise between cost and good amplitude frequency characteristic (AFC) has to be made.

Since their emergence tunable filters based on MEMS have come a long way, from devices with only several provisions to smoothly tunable devices [13-16]. Microelectromechanical systems (MEMS) have been upgraded significantly and now can cover a very wide frequency range, providing smooth tuning. An example of such implementation is given in [17].

In contrast to previous methods [18], tuning implemented by bending a part of the MEMS from the substrate, significantly simplifies the manufacturing of such device. However, since there is bending element the wear of the material from which MEMS are made is increased.

Varactors based on MEMS technology [19], with displacements up to 3 microns and providing about 0.5 GHz of center frequency tuning were developed in order to achieve greater integration. However the level of the insertion losses in such devices is close to the levels in devices based on varactors.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

This article investigates the characteristics of a tunable microstrip stub. Considering the advantages and disadvantages of all mentioned controlling methods, the control of the microstrip stub is performed by moving part of the stub from the substrate, thereby changing its electrical length. Creating an air gap between the substrate and the signal electrode will create a tunable heterogeneity. Moving the electrode up and down will result in a change of the stub's effective permittivity and resonance frequency shifting. This can be achieved by using MEMS.

II. TOPOLOGY AND PARAMETERS

Fig. 1 shows the proposed topology of the tunable microstrip filter with the stub as a resonant element.

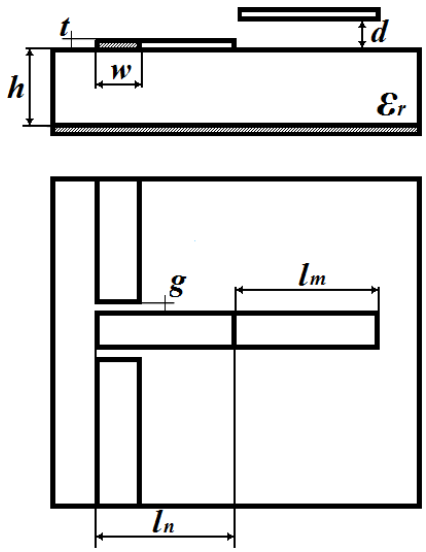


Fig. 1. Topology of the tunable microstrip filter with a stub as resonating element.

The filter consists of four parts: input and output microstrip transmission lines, the fixed part of the stub, and the tuning part of the stub. The electrode of the tuning stub part is moved up and down parallel to the substrate. ***d*** is the thickness of the air gap between the tuning part and the substrate. The height of the filter substrate is ***h*** and the substrate permittivity is ε_r. The microstrip tunable filter has the following parameters: ***w*** is the width and ***t*** is the thickness of the microstrip lines, ***g*** is the gap between the stub and the feeding input and output microstrip lines, l_n is the length of the fixed stub part, and l_m is the length of the tuning stub part.

The resonance frequency tuning concept lies in the effective permittivity alteration when the electrode of the tunable stub part is moving up and down. The air gap between the stub's moving electrode and the substrate is the tunable heterogeneity. The size of the heterogeneity affects the effective permittivity changes [20]. The dependency of the effective permittivity on the thickness of the air gap between the moving stub part and the substrate for different ratios of stub width and substrate height is shown in Fig. 2.

Reducing the effective permittivity while increasing the distance between electrode and substrate leads to a resonance frequency shift to higher frequencies. The dependency of the frequency on different thicknesses for the air gap for different ratios of stub width and substrate height is shown in Fig. 3. Since the effective permittivity varies in the range in between the effective permittivity value when there is no detachment to a permittivity of one, the tuning effectiveness will be limited. Also big effect on the resonance frequency value has capacitance C_s which appears in the tuning process between the movable and fixed stub parts. That is why the dependency of the frequency on the air gap thickness shows a region where the frequency varies slightly.

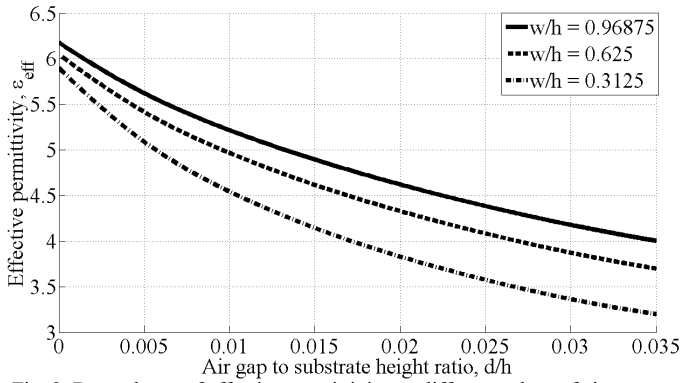


Fig. 2. Dependency of effective permittivity on different values of air gap between the moving stub part and the substrate for different ratios of stub width and substrate height. Substrate permittivity $\varepsilon_r = 9.9$.

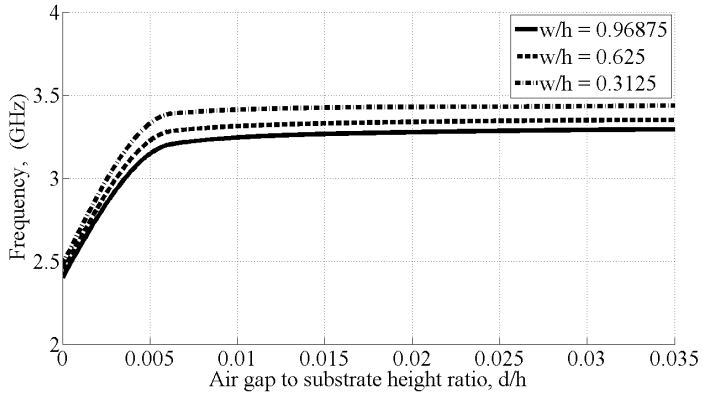


Fig. 3. Dependency of frequency versus different values of air gap between the moving stub part and the substrate for different ratios of stub width and substrate height. Substrate permittivity $\varepsilon_r = 9.9$.

Since the length of the tuning stub part affects the size of the controlled heterogeneity, changing its length can change the tuning range. The longer the tuning stub part, the greater the tuning range that can be achieved. So the ratio of the fixed and movable stub parts lengths should be chosen for the desired frequency tuning range.

III. EQUIVALENT CIRCUIT MODEL

The equivalent circuit of the resonator presented in Fig. 1 is shown in Fig. 4.

The characteristic impedances of input and output lines are equal to Z_1. The capacitors C_1 model the air gaps between the stub and the feeding lines. They depend on the air gap thickness. The C_2's represent the parallel capacitance between

the ground plate and feeding lines. They directly depend on substrate permittivity and height. These capacitors have a great influence on the filter frequency response.

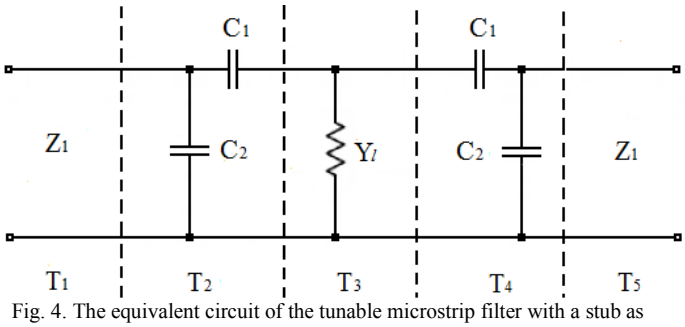


Fig. 4. The equivalent circuit of the tunable microstrip filter with a stub as resonating element.

The stub is represented by the complex conductivity Y_l. The whole circuit model is divided into five parts with transmission matrices that are simple to calculate. Matrices T_1 and T_5 can be calculated as transmission matrices of homogeneous transmission line section:

$$T_1 = T_5 = \begin{bmatrix} e^{-j\beta_{z1}l_1} & 0 \\ 0 & e^{j\beta_{z1}l_1} \end{bmatrix}, \quad (1)$$

where β_{z1} is the homogeneous transmission line wave propagation constant, and l_1 is the homogeneous transmission line length. Matrices T_2 and T_4 are given by:

$$T_2 = T_4 = \frac{1}{2Z_1Z_{C2}} \begin{bmatrix} -Z_1^2 + aZ_1 - b & -Z_1^2 - Z_{C1}Z_1 + b \\ Z_1^2 - Z_{C1}Z_1 - b & Z_1^2 + aZ_1 + b \end{bmatrix}, \quad (2)$$

where $a = (Z_{C1} + 2Z_{C2})$ and $b = Z_{C2}Z_{C1}$, Z_{C1} and Z_{C2} equivalent impedances of C_1 and C_2 respectively.

The capacitor values can be found with the closed form expressions of Gang and Bahl [21], modified by Yu and Chang [22]. Previously this approach was used for calculating air gaps between homogeneous transmission line sections, but it can be applied to this case too:

$$C_1 = \frac{2C_{odd} - C_{even}}{4}, \quad (4)$$

$$C_2 = \frac{C_{even}}{2}, \quad (5)$$

$$C_{even} = 1.167\left(\frac{\varepsilon_r}{9.6}\right)^{0.9}\left(12w\left(\frac{g}{w}\right)^{m_e} e^{k_e}\right), \quad (6)$$

$$C_{odd} = 1.1\left(\frac{\varepsilon_r}{9.6}\right)^{0.8}\left(w\left(\frac{g}{w}\right)^{m_0} e^{k_0}\right), \quad (7)$$

where

$$m_e = 0.8675,\ k_e = 2.043\left(\frac{w}{h}\right)^{0.12}, \quad (8)$$

$$0.1 \le g/w \le 0.5,$$

$$m_e = \frac{1.565}{\left(\frac{w}{h}\right)^{0.16}} - 1,\ k_e = 1.97 - \frac{0.03}{\left(\frac{w}{h}\right)}, \quad (9)$$

$$0.5 \le g/w \le 1.$$

$$m_0 = \frac{w}{h}\left(0.619\log_{10}\frac{w}{h} - 0.3853\right), \quad (10)$$

$$k_0 = 4.26 - 1.453\log_{10}\frac{w}{h}, \quad (11)$$

$$0.1 \le g/w \le 1.$$

Before calculating matrix T_3 it is necessary to calculate the overall complex impedance of the stub Z_l. This can be done in two steps. The first step is to calculate the input impedance of the movable stub part with the open circuit:

$$Z_m = \frac{Z_{m0}}{j\tan(\beta_{zm}l_m)} + \frac{1}{j\omega C_s(d)}, \quad (12)$$

where Z_{m0} is the characteristic impedance of the stub movable part, and β_{zm} is the stub's movable part wave propagation constant.

C_S is the capacitance between the fixed and movable stub parts, and depends on the air gap thickness between them. It is a very important model parameter, because its value significantly affects the resonant frequency. Then, the overall complex stub impedance can be calculated in the same way but loaded with the input movable part impedance:

$$Z_l = Z_{n0}\frac{Z_m + jZ_{n0}\tan(\beta_{zn}l_n)}{Z_{n0} + jZ_m\tan(\beta_{zn}l_n)}. \quad (13)$$

After the overall complex impedance of the stub Z_l is calculated, the T_3 matrix can be found as:

$$T_3 = \frac{1}{2Y_1}\begin{bmatrix} 2Y_1 - Y_l & -Y_l \\ Y_l & 2Y_1 + Y_l \end{bmatrix}, \quad (14)$$

where Yl is the stub's complex conductivity. Finally, the overall T-matrix can be calculated as:

$$T = \prod_{i=1}^{5} T_i. \quad (15)$$

Since all matrices are calculated the S-matrix can be expressed as:

$$S = \frac{1}{t_{22}}\begin{bmatrix} t_{12} & t_{11}t_{22} - t_{21}t_{12} \\ 1 & -t_{21} \end{bmatrix}, \quad (16)$$

where t_{11}, t_{12}, t_{21}, t_{22} correspond to the elements of the overall T-matrix.

IV. Results and Discussion

The frequency dependency of the transmission coefficient modulus calculated with (16), with FEM, and with FDTD is shown in Fig. 5-8.

As can be seen in Fig. 5, the detachment of the movable stub electrode from the substrate by 50 microns, results in a resonance frequency shift of more than 1 GHz. If the ratio of the movable stub electrode length and the fixed stub electrode length is changed, the resonance frequency shift increases up to 2 GHz (Fig. 7).

Fig. 5 illustrates the frequency dependence of the transmission coefficient for the following parameters: height of the filter substrate h = 1.6 mm, substrate permittivity ε_r = 9.9,

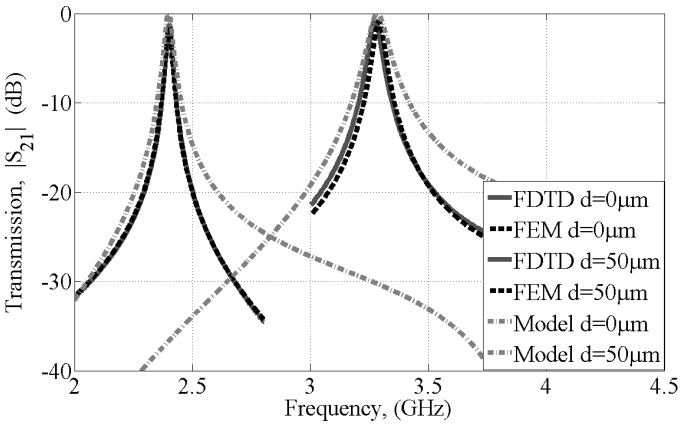


Fig.5. Frequency dependence of transmission coefficient for tunable microstrip stub for different values of the air gap between the moving electrode and the substrate, g = 0.2 mm.

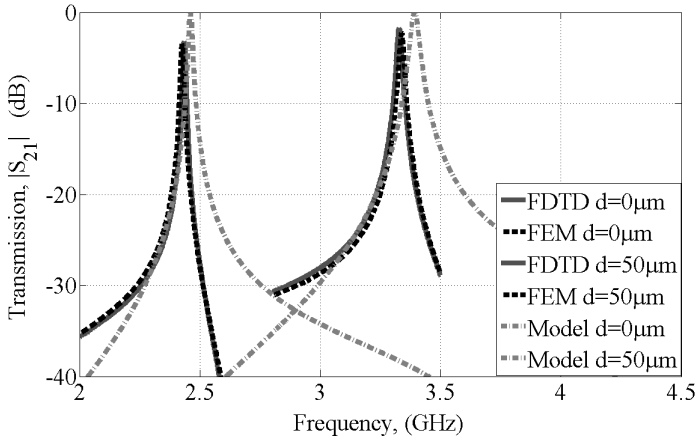


Fig. 6. Frequency dependence of transmission coefficient for tunable microstrip stub for different values of the air gap between the moving electrode and the substrate, g = 0.5 mm.

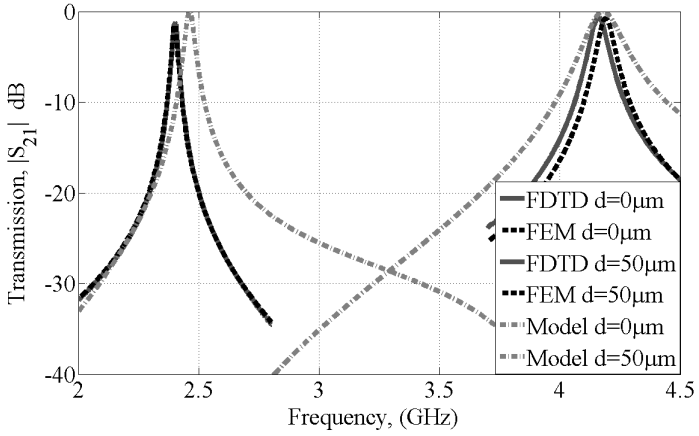


Fig. 7. Frequency dependence of transmission coefficient for tunable microstrip stub for different values of the air gap between the moving electrode and the substrate. ln = 10.6 mm, lm = 11 mm.

width of the microstrip parts w = 1.55 mm, thickness of the microstrip parts t = 0.01 mm, gap between the feeding In and Out microstrip lines g = 0.2 mm, length of the fixed stub part l_n = 14.6 mm, length of the tuning stub part l_m = 7 mm. Results calculated with the circuit model have a lower transmission loss compared to FDTD and FEM because the circuit model doesn't take into account the radiation losses. The results of the three methods are in good agreement.

The influence of the air gap g value on the resonance frequency tuning and the filter parameters is shown in Fig. 6. The gap between the stub and the feeding lines is increased from 0.2 mm to 0.5 mm. In this case a decrease of the bandwidth and increased transmission losses are observed. This can be explained by the fact that changing the g value affects the stub's excitation and the coupling coefficient. So we can obtain the necessary bandwidth and transmission loss level by choosing the correct air gap value between the feeding lines and stub. The circuit model results have a small resonance frequency shift compared to the FDTD and FEM results. This is caused by the stub excitation and the stub electrical length changing with the increasing of the air gap. The effective permittivity and the characteristic impedance calculation tolerance also affects the circuit model results.

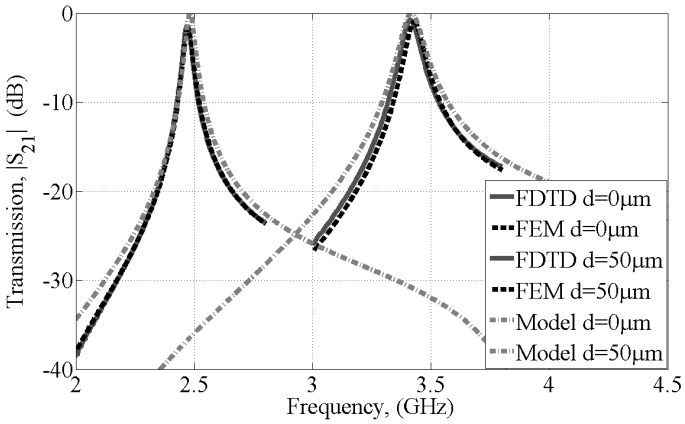


Fig. 8. Frequency dependence of transmission coefficient for tunable microstrip stub for different values of the air gap between the moving electrode and the substrate, w equal to 1 mm.

Changing the length of the movable and fixed stub parts allows to modify the tuning effectiveness, bandwidth, and transmission loss. In Fig. 7 the movable part length is increased from 7 mm to 11 mm, while the fixed part length is decreased from 14.6 mm to 10.6 mm. Due to the fact that the movable part is now longer, the heterogeneity between the moving stub part and the substrate is increased. The greater the heterogeneity, the greater its influence on the effective permittivity and filter resonant frequency, which allows to get more efficient resonant frequency tuning. This is due to the fact that of course the air gap is now longer and more energy is being transmitted through it, compared to the substrate, which decreases the dielectric loss, and thus also the overall transmission loss. The slight difference in resonance frequency values obtained using on the one hand the circuit model, and on the other hand the FDTD and FEM models, is related to the accuracy by which the effective dielectric constant and characteristic impedance are calculated in our circuit model.

To investigate the influence of the microstrip width to substrate height ratio on the filter parameters, the width of the

stub parts is changed to 1 mm; while the feeding line width remains at 1.55 mm (Fig. 8). Due to stub characteristic impedance variation the resonant frequency slightly shifts to higher frequencies. Decreasing the stub width results in a decreasing of the space between the feeding lines. This in turn increases the coupling between the feeding lines and therefore increases the bandwidth. Nonetheless, there are two competing processes that influence the bandwidth. From one side the stronger feeding line coupling increases the bandwidth, but from the other side increasing the stub characteristic impedance decreases the bandwidth. So the necessary bandwidth can be obtained with a correct choice of the microstrip stub width. Results calculated with the circuit model are in good agreement with FDTD and FEM.

Authors of the article [1] obtained results which are in good agreement with the experimental data. However, for the zero voltage there is a discrepancy in frequency between the simulated and experimental data. The tuning range was approximately 0.7 GHz. Presented in this paper methods have full alignment with each other, as well as predicting the filter tuning at range of about 2 GHz.

V. CONCLUSIONS

A novel concept of low-cost tunable microstrip filter tuning was proposed. A new equivalent circuit of the tunable microstrip stub was suggested. Comparing the proposed circuit model with FDTD and FEM shows a good agreement. The effect of the tunable microstrip stub's physical and topological parameters on the frequency range was investigated. The study shows that changing the air gap between the feeding lines and the stub affects the bandwidth and transmission loss. So the gap cannot be too large, because the coupling between stub and the feeding lines would not be enough to ensure good filter parameters. To increase the effectiveness of tuning the length of the tuning stub part should be increased. The air gap between the substrate and the signal electrode is the heterogeneity which controls the stub's effective permittivity and provides the resonance frequency shift. A detachment of the stub of 50 microns a resonance frequency shift of about 1 GHz is obtained.

REFERENCES

[1] Boutejdar, A.; Omar, A.; Senst, M.; Burte, E.P.; Batmanov, A.; Mikuta, R., "A new design of a tunable WLAN-band pass filter using a combination of varactor device, RF-choke and Hairpin-Defected Ground Structure ", Microwave Conference (EuMC), 2011 41st European, pp. 1067 – 1070, 2011.

[2] Xu-Guang Wang; Young-Ho Cho; Sang-Won Yun, "A Tunable Combline Bandpass Filter Loaded With Series Resonator", IEEE Transactions on Microwave Theory and Techniques, vol. 60, №. 6, pp. 1569-1576, 2012.

[3] Guyette, A.C., "Intrinsically Switched Varactor-Tuned Filters and Filter Banks", IEEE Transactions on Microwave Theory and Techniques, vol. 60, pp. 1044 – 1056, №. 4, 2012.

[4] Fei Cheng; Xianqi Lin; Yiyao Hu; Yong Fan, "Tunable Bandpass Filter using Varactor-Loaded Quarter-Wavelength Resonator", International Conference on Microwave and Millimeter Wave Technology (ICMMT), vol.5, pp. 1-3, 2012.

[5] Brito-Brito, Z.; Llamas-Garro, I.; Navarro-Muñoz, G.; Perruisseau-Carrier, J.; Pradell, L.; Giacomozzi, F.; Colpo, S., "Precise frequency and bandwidth control of switchable microstrip bandpass filters using diode and microelectro-mechanical system technologies",IET Microwaves, Antennas & Propagation, vol.6, pp. 713 – 719, 2012.

[6] Lan Yao; Wei Hong; Ke Wu, "Tunable Microstrip Ring Bandpass Filter"China-Japan Joint Microwave Conference Proceedings (CJMW), pp. 1-3, 2011.

[7] Ying Liu; Hongxing Zheng; Chengguang Sun, "A Tunable Compact Bandstop Filter Using Meander Spurline and Folding Open Stubs", International Conference on Consumer Electronics, Communications and Networks (CECNet), pp. 115 – 118, 2012.

[8] Habiba, H.U.; Malathi, K.; Masood, M.H.; Kunnath, R., "Tunable electromagnetic band gap-embedded multimode resonators for ultra-wideband dual band, lower-ultra-wideband and upper-ultra-wideband applications", IET Microwaves, Antennas & Propagation, vol.5, pp. 1182 – 1187, 2011.

[9] He-xiu Xu; Guang-ming Wang; Lu, K., "Tunable Low-Pass Filter Using Fractal Shaped Complementary Split Ring Resonator with Ultra-Wide Stop-Band and Excellent Selectivity", IEEE International Conference on Ultra-Wideband (ICUWB), vol. 2, pp. 1 – 3, 2010.

[10] Yun Zhu; Gang Qiu; Chi, K.H.; Wang, B.B.T.; Tsai, C.S., "A Compact X-Band Tunable Bandpass Filter Module Using a Pair of Microstrip Composite Bandpass Filters in Cascade", IEEE Transactions on Magnetics, vol. 46, pp. 1424 – 1427, №. 6, 2010.

[11] Mi, X.; Toyoda, O.; Ueda, S.; Nakazawa, F., "Miniaturized Microwave Tunable Bandpass Filters on High-k LTCC with Integrated Resistive Vias as Bias-T", International Microwave Symposium Digest (MTT), p.1, 2011.

[12] Pourush, R.; Tyagi, G.S., "A Low Cost Magnetically Tunable Microstrip Ring Resonator", IEEE Applied Electromagnetics Conference (AEMC), pp. 1-4, 2011.

[13] Pothier, A.; Orlianges, J.-C.; Guizhen Zheng; Champeaux, C.; Catherinot, A.; Cros, D.; Blondy, P.; Papapolymerou, J., "Low-Loss 2-Bit Tunable Bandpass Filters Using MEMS DC Contact Switches", IEEE Transactions on Microwave Theory and Techniques, vol. 53, pp. 354 – 360, №. 1, 2005.

[14] Pillans, B.; Malczewski, A.; Allison, R.; Brank, J., "6-15 GHz RF MEMS Tunable Filters", IEEE MTT-S International Microwave Digest Symposium, 2005.

[15] Abbaspour-Tamijani, A.; Dussopt, L.; Rebeiz, G.M., "A High Performance MEMS Miniature Tunable Bandpass Filter", IEEE MTT-S International Microwave Digest Symposium, vol. 3, pp. 1785 – 1788, 2003.

[16] Kraus, G.M.; Goldsmith, C.L.; Nordquist, C.D.; Dyck, C.W.; Finnegan, P.S.; Austin, F., IV; Muyshondt, A.; Sullivan, C.T., "A Widely Tunable RF MEMS End-coupled Filter", IEEE MTT-S International Microwave Digest Symposium, vol. 2, pp. 429 - 432, 2004.

[17] Pillans, B. W.; Malczewski, A.; Morris, F. J.; Newstrom, R. A., "A Family of MEMS Tunable Filters for Advanced RF Applications", IEEE MTT-S International Microwave Digest Symposium, p. 1, 2011.

[18] Bouyge, D.; Mardivirin, D.; Bonache, J.; Crunteanu, A.; Pothier, A.; Duran-Sindreu, M.; Blondy, P.; Martin, F., "Split Ring Resonators (SRRs) Based on Micro-Electro-Mechanical Deflectable Cantilever-Type Rings: Application to Tunable Stopband Filters", IEEE Microwave and Wireless Components Letters, vol. 21, pp. 243 – 245, №. 5, 2011.

[19] Mardivirin, D.; Barriere, F.; Pothier, A.; Crunteanu, A.; Blondy, P., "A 2 Pole 9.6-11.7 GHz Band Stop Filter Using Analog Tuning RF MEMS Varactors", Microwave Conference (EuMC), 2011 41st European, pp. 1091 – 1094, 2011.

[20] I. Golubeva, V. Kazmirenko, P. Sergienko, Y. Prokopenko, "Effective Permittivity in Tunable Microstrip and Coplanar Lines", XXXII International Conference "Electronics and Nanotechnology", Kiev, 10-12 April, pp. 1-4, 2012.

[21] R. Garg and I. J. Bahl, "Microstrip discontinuities," *Int'l J. Electronics*, vol. 45, pp. 81-87, 1978.

[22] C.-C. Yu and K. Chang, "Transmission-line analysis of a capacitively coupled microstrip-ring resonator," IEEE Trans. Microwave Theory and Techniques, vol. 45, №. 11, pp. 2018-2024, Nov. 1997.

The Three-Collector Magnetotransistor: Variable Sensitivity

V.V. Amelichev, R.D. Tikhonov and A.A. Cheremisinov
SMC "TECHNOLOGICAL CENTRE" MIET
124498, Zelenograd, Moscow, Russia
CheremisinovAA@gmail.com

***Abstract*—The sensitivity of bipolar magnetotransistor with the base in the well has been studied. An experiment is conducted to compare the performance of two designs of dual-collector lateral bipolar magnetotransistor, which are formed in a uniformly doped substrate or a diffused well. Their respective differential collector voltages are found to differ in sign under an applied magnetic field. It is noted that acquiring an adequate understanding of the mechanism underlying the operation of the device should help one improve its magnetic-field sensitivity. A low velocity of surface recombination and an extraction of the injected electrons by a base–well p-n junction determined operating mode with a deviation of two flows of charge carriers.**

Keywords—magnetotransistor; surface recombination

I. INTRODUCTION

Bipolar MagnetoTransistor (BMT) was studied since 1970. It has found reflection in many reviews and in the book, containing references to hundreds works [1]. Now BMT has no practical value. It is possible to note only two publications in which systems with magnetotransistor usage were described [2, 3]. BMT is very responsive to an applied magnetic field. It provides a linear output, operation at a high frequency, and also offers a good signal-to-noise ratio. All of this raises the perfectly legitimate question of why such a promising device still has not found any uses in magnetic-field-sensitive ICs that are being manufactured in millions. These issues suggest a more complicated mechanism of conversion lying at the heart of the device as compared with magnetoresistors and Hall probes. No one yet knows precisely what determines the magnitude and sign of BMT sensitivity.

II. SENSITIVITY MECHANISMS OF BIPOLAR MAGNETOTRANSISTORS

In an intrinsic semiconductor, electrons and holes drift toward one another under the action of the electric field, so that the magnetic field deflects them in the same direction. The result is that the density of the electron–hole plasma rises at one side and falls at the opposite one, with magnetoresistance and the Hall effect increasing and decreasing accordingly. The galvanomagnetic recombination effect also occurs [4]. In magnetodiodes, minority carriers are injected into the base region, whose size is comparable with the diffusion length [5]. This increases their magnetic-field sensitivity manifold, particularly when accompanied by accelerated recombination near a sidewall. The bipolar magnetotransistor of the vertical type employs the deflecting effect of a magnetic field applied parallel to the chip surface on carriers as they travel through the base, with the result that the carrier stream is divided between the two halves of a split collector. With the BMT of the planar type carriers are injected along the chip surface, whereas the magnetic field is applied perpendicular to the surface to deflect the carrier stream toward one of two collectors lying on the surface. The deflection alters the effective length of current paths and changes the ratio in which the carrier stream is divided between the collectors. With the BMT of the lateral type [6], the emitter is located between collectors on the chip surface, and a magnetic field is applied parallel to the surface. The carriers on one side of the emitter are deflected toward the surface; this shortens current paths, reduces bulk recombination, and encourages surface recombination. The magnetoconcentration effect associated with the former portion of carriers was discussed .When lateral BMT are fabricated as part of a complementary metal–oxide–semiconductor (CMOS) circuit, the emitter and collector regions of a BMT are formed in a diffused well, which also provides the base region. Injection along the surface is controlled by introducing the base doping level [7].

III. SIGN OF MAGNETIC FIELD SENSITIVITY

With J and B denoting the current density and the magnetic flux density, respectively, the Lorentz force F is written

$$\mathbf{F} = [\mathbf{J} \times \mathbf{B}]. \quad (1)$$

Their current sensitivity to a magnetic field is taken as the derived function of their current–flux density characteristic I(B). Accordingly, the sensitivity may be positive or negative depending on whether the current rises or falls with magnetic flux density. The resistance of rectangular or cylindrical magnetoresistors increases on applying a magnetic field [8]. The Lorentz force makes current paths curve, increasing the effective length of the device (Fig. 1a). For a given voltage across a magnetoresistor, the current density is related to the magnetic flux density by

$$j(B) = j(0)/(1+\mu^2 B^2) \quad (2)$$

where μ is the magnetoresistive mobility.

In a magnetodiode, electrons are injected from an N+ emitter into an intrinsic region before being collected by a P+ contact pad. The Lorentz force makes current paths curve, increasing the effective length of the intrinsic region (Fig. 1b). Compared with magnetoresistors, the device has the added

feature of injected carrier recombination in the intrinsic region and on its lateral sides. This is represented by a factor c that characterizes the ratio of the intrinsic-region length to the diffusion length, occurring in the forward-current relation

$$I=Ic(\exp(qU/ckT-1) \quad (3)$$

The factor c is related to the electron mobility μ_n, the hole mobility μ_p, the intrinsic-region length d, and the injected-carrier lifetime τ_n by the formula

$$c=2(b+ch\ d/L)/(b+1) \quad (4)$$

where $b=\mu_n/\mu_p$, $L=(D_n/\tau_n)^{1/2}[2b/(b+1)]^{1/2}$ is the effective length of diffusion displacement, and D_n is the diffusion coefficient. The factor c grows with the length of current paths. The resultant increase in recombination makes diode current decrease, with the sensitivity growing. Surface recombination contributes to this mechanism as the magnetic field drives carriers toward the surface. With the magnetic flux density perpendicular to the current flow, the current decreases with increasing magnetic flux density for any given voltage, implying a negative sensitivity.

In a dual-collector vertical BMT (Fig. 1c), electrons are injected from an N+ emitter into a P-base region before being collected by two N+ collectors. In addition to the three factors above, an applied magnetic field governs the ratio of the respective currents that flow into the collectors C1 and C2. As the field is increased, the current through C1 rises and that through C2 falls (Fig. 1f). The device is sensitive to the tangential direction of a magnetic field. In a dual-collector planar BMT (Fig. 1d), electrons are injected from an N+ emitter into a P-base region before being collected by two N+ collectors situated some distance to either side of the axis that joins the emitter and the base contact. In addition to the four factors above, an applied magnetic field alters the ratio of collector currents, as evidenced by changes in the pattern and effective length of current paths. As the field is increased, the current through C1 rises and that through C2 falls (Fig. 1f). In the basic dual-collector lateral BMT (Fig. 1e), electrons are injected from an N+ emitter into a P-substrate, which provides a base region. Two N+ collectors are situated on the substrate surface to either side of the emitter, and so are base contacts, which are farther from the emitter. In addition to the five factors above, an applied magnetic field alters the proportion in which injected electrons are divided between the collectors C1 and C2, as evidenced by changes in the pattern of current paths relative to the surface: those originating from C1 draw closer to the surface, and those from C2 move away from it. This changes the contribution of surface recombination to the two collector currents when the substrate is doped uniformly. If the contribution is insignificant, the current through C1 rises and that through C2 falls as the field is increased (Fig. 1f). Otherwise, the sensitivity changes sign with amount of injection; this occurs when the emitter and each collector are far apart. A more advanced version of dual-collector lateral BMT is formed in a nonuniformly doped diffused well in which doping concentration decreases with increasing depth from a high level at the surface to the substrate doping level (Fig. 1g). A depletion region is thus created in the form of a pn junction. When the substrate is doped nonuniformly, current paths close to C1 migrate toward higher doping levels due to the same factor that changes the position of current paths relative to the surface. An applied magnetic field drives current paths that are close to C2 toward lower doping levels. So that the sensitivity is positive. Thus, the collector currents vary with magnetic field in an opposite manner to what is shown by the basic version of the device. This fact was discovered by experiment, and was investigated carefully.

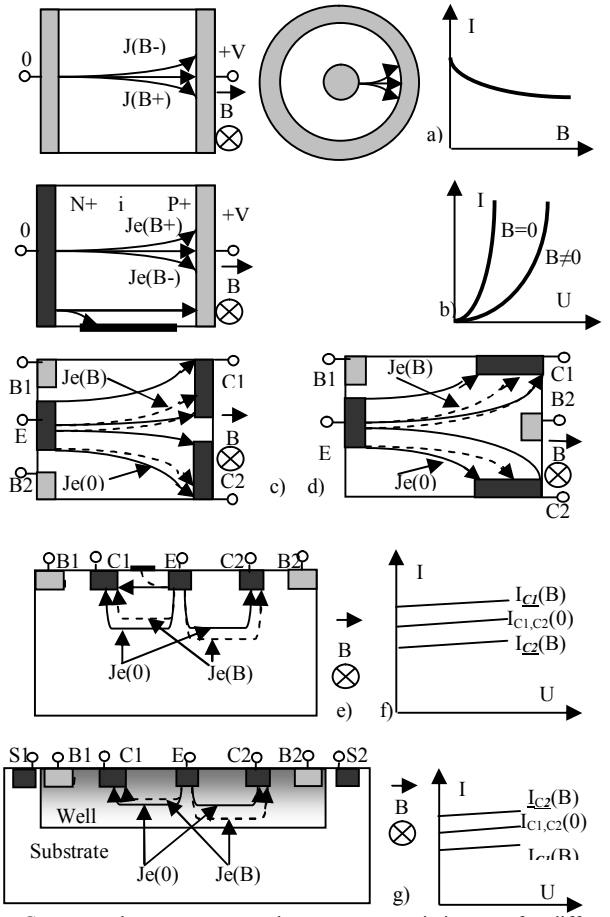


Fig. 1. Current-path patterns and current variations of different galvanomagnetic devices: (a) magnetoresistor, (b) magnetodiode, (c, f) dual-collector vertical BMT, (d, f) dual-collector planar BMT, (e, f) basic dual-collector lateral BMT, and (g) a dual-collector lateral BMT formed in a diffused well.

IV. Formulas to Calculate Magnetotransistor Sensitivity

Formulas to calculate magnetotransistor sensitivity are obtained on the basis that the device is effectively a controlled current source in which carrier transport is subject to an applied magnetic flux density B. Let $I_C(B)$ and $V_C(B)$ be the collector current and voltage, respectively, which are taken as the output, i.e., the response to the magnetic induction B. Let R_C and E_S be the collector load resistance and the supply voltage, respectively. Then, the collector current and voltage obey the relation

$$V_C(B) = E_S - R_C\, I_C(B) \quad (5)$$

As with any sensor, the sensitivity *S* is the most important parameter. Roumenin [3] has pointed out that the change in collector current, $\Delta I_C(B)$, as dependent on the magnetic field B

should be taken as the output signal of the dual-collector BMT the output signal is

$$\Delta I_C(B) = \Delta I_{C1}(B) - \Delta I_{C2}(B) \quad (6)$$

The differential absolute current sensitivity S_{DA}^I and the relative current sensitivity S_{DR}^I are defined for the dual-collector device as,

$$S_{DA}^I = [(I_{C2}(B) - I_{C1}(B)] / B, [A/T] \quad (7)$$

$$S_{DR}^I = [I_{C2}(B) - I_{C1}(B)] / [I_{C2}(0)+I_{C1}(0)] B, [1/T] \quad (8)$$

This is at odds with the reality, since the collector current may increase or decrease with magnetic flux density, making S_{DA}^I change its sign. Concerning integrated semiconductor magnetic field sensors, the output is provided by the collector voltage as a function of an applied magnetic field. Accordingly, the absolute and the relative voltage sensitivity are defined as

$$S_{DA}^V = [U_{C2}(B) - U_{C1}(B)] / B, [V/T] \quad (9)$$

$$S_{DR}^V = [U_{C2}(B) - U_{C1}(B)]/[U_{C1}(0)+U_{C2}(0)] B [1/T] \quad (10)$$

$$S_{DA}^V = - R_C S_{DA}^I \quad (11)$$

The voltage sensitivity is determined by the current sensitivity with the opposite sign to reflect the fact that an increase in collector current implies a decrease in collector voltage according to Eq. (5). The collector voltage varies in line with the voltage across R_C. Note that the relative sensitivity with respect to current as defined by Eq. (8) is not the same as that with respect to voltage as defined by Eq. (10):

$$S_{DR}^I = S_{DA}^V \cdot [\, U_{C2}(B) + U_{C1}(B) - 2E_S \,]^{-1} \quad (12)$$

The effect of the Lorentz force on a current is a defining factor for the emergence of magnetosensitivity. An increase in the collector current under the influence of the Lorentz force corresponds to a positive sensitivity, and a reduction of the collector current under the influence of the Lorentz force corresponds to a negative sensitivity. This definition corresponds to the effect of a deviation. The effect of the lifetime of minority carriers of a current can result in an identical sensitivity as the effect of a deviation and vice versa.

V. Experimental Study of Sensitivity

An experimental study was conducted on a dual-collector lateral BMT in a uniformly doped substrate (BMTS) and on one in a diffused well (BMTW) [8]. The BMTS and the BMTW were fabricated by the same set of photo masks. In fabricating the BMTW, the conduction type of the substrate was inverted and a diffused well was formed with a surface doping concentration two orders of magnitude higher than that of the BMTS substrate.

Measurements on BMTW specimens were made within the resistances as follows: Rb = 39.6 kΩ and Rc = 27.38 kΩ for both types of BMT with the substrate terminal disconnected from the base; Rb = 4.736 kΩ and Rc = 27.38 kΩ for the BMTW with the substrate terminal connected to the base. The supply voltage was gradually raised to 10 V. A permanent magnet provided B = 120 mT. Differential collector voltage was measured in zero or nonzero magnetic field. It is measured supply-voltage dependence of differential collector voltage,

$$\Delta U_C = [U_{C1}(B) - U_{C2}(B)] - [U_{C1}(0) + U_{C2}(0)] \quad (13)$$

For the BMTS, it rises monotonically with supply voltage. The maximum value of ΔU_C is 128 mV; it corresponds to the absolute voltage sensitivity $S^V_C = \Delta U_C/B = 0.9$ V/T. The BMTW with a floating substrate has a negative sensitivity; it attains its maximum magnitude at a 5 V supply voltage, above which the device goes into saturation. The maximum absolute value ΔU_C is 235 mV and $S^V_C = -2$ V/T. With the substrate terminal connected to the base, the differential voltage is again negative. It starts growing in magnitude once a 5 V threshold has been crossed, and continues to do so up to the end point of supply voltage. Its maximum magnitude is 524 mV and $S^V_C = -4.35$ V/T. We see that this configuration of BMTW offers a twofold increase in maximum sensitivity magnitude. This is accompanied by a twofold increase in emitter current. Diverting a proportion of injected electrons to the substrate reduces the collector currents, so that collector saturation occurs at a higher emitter current, which is subject to the Lorentz force. The net result is better sensitivity. In essence, the sensitivity of the device structure considered is determined by emitter current. For the BMTW, the two collector voltages change in the opposite direction to their counterparts in the BMTS.

VI. Magnetotransistor with Variable Sensitivity

Determining the sign of magnetic-field sensitivity is particularly important in cases where it can change. Consider the planar magnetotransistor studied by Vinal and Masnari [9]. It has two base contacts and two collector contacts arranged as in Fig. 2, with the substrate serving as the base region.

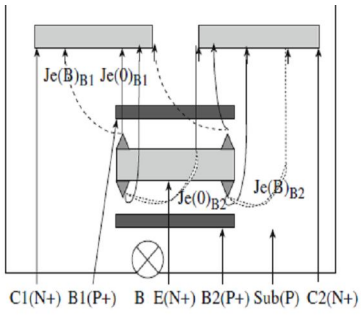


Fig.2. Dual-collector planar npn magnetotransistor studied by Vinal and Masnari [9]

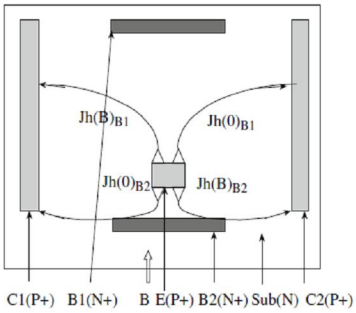


Fig.3. Dual-collector planar pnp magnetotransistor studied by Roumenin [1].

Let us now turn to the planar magnetotransistor studied by Roumenin [1]. It has two base contacts and two collector contacts arranged as in Fig. 3, with the substrate serving as the base region. The output signal, $\Delta I_{C1,2} = I_{C1} - I_{C2}$, is a linear function of magnetic flux density as long as this is below 1 T. The performance of the magnetotransistor was measured at a collector bias of –10 V. The effect was found to be independent of collector voltage. In the base region, the Hall

voltage was found to arise under drift current and to be small in magnitude.

VII. The Structure and Control Circuits 3CBMTBW

From these carried out researches the conclusion has been made that the main effect providing high sensitivity of a magnetotransistor surface recombination was the main reason of the main shortcomings of the device.

In reviewed paper such option of a design yields reproduced results and high sensitivity for deficiencies elimination of magnetotransistor with low speed of surface recombination has been investigated [10]. Research on magnetodiodes showed that instead of surface recombination, which is difficult to reproduce, it is possible to use extraction of carriers by a p-n junction.

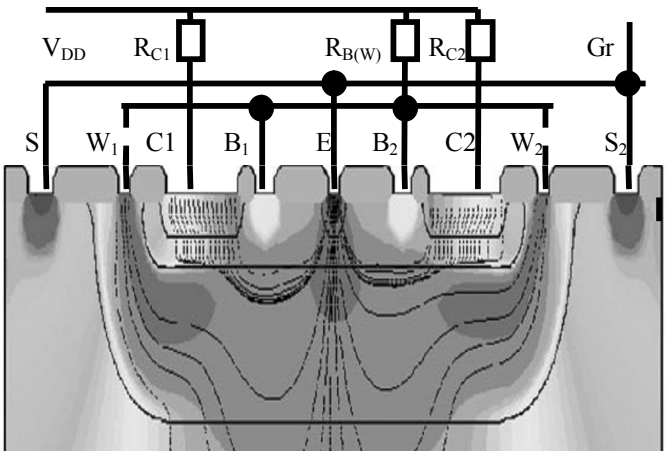


Fig.4. Structure and circuit of 3CBMTBW.

The structure of The Three-Collector Bipolar Magnetotransistor with Base into Well - 3CBMTBW is given in Fig. 4, where the distribution of a current of the injected electrons, obtained from the results of technological modelling of the device, is also shown. Generally electrons leave through the base and are extracted by the base–well p-n junction. Only an insignificant part of the stream of electrons along poorly doped areas near the base–well p-n junction reaches the collectors, where there is extraction of electrons the built in electric field of the collector junction.

VIII. The Influence of Control Circuits on Magnetotransistors Sensitivity

The measurement of the 3CBMTBW magnetotransistor is carried out fewer than two circuits of switching on. Sensitivity was measured in a constant magnetic field of the solenoid with a magnetic induction of B = 5 mT. The differential absolute sensitivity of the voltage was determined by the formula:

$$S_{DA}{}^{V} = [U_{C2}(B) - U_{C2}(0) - U_{C1}(B) + U_{C1}(0)] / B \qquad (14)$$

The voltage difference of the collectors was measured with a supply voltage of Vdd = 9 V. Without a magnetic field there was an initial imbalance, and with a magnetic field the working signal caused on an induction B. The sensitivity was calculated at the chosen voltage and with a bias current in the base I_B or in the base and well I_{BW}.

Eighteen samples of 3CBMTBW with a collector load resistance of 267.4 kiloohms at a connection of bias voltage only on the U_B base have an average value of sensitivity of 0.134 V/T when the voltage on the collectors is close to 1 V and the initial voltage imbalance of the collectors is –0.84 mV. With connection of the bias voltage on the base and well of U_{BW} the BMT parameters have an average value of sensitivity of 10.9 V/T with a voltage on the collectors close to 1.45 V and an initial voltage imbalance of the collectors of –50.5 mV. In the circuit with the general potential on the base and a well, the sensitivity increases 81 times and the initial voltage imbalance of the collectors increases 60 times. The sensitivity of different samples is highly dependent on the bias currents of the base or the base together with a well, which is testimony to a good reproducibility. The sensitivity of the voltage reaches a high value of 11 V/T.

IX. Comparation of Magnetotransistors

Sensitivity research at variable values of the base and well currents is carried out in the circuit including separate sources for the base and well shift. For a set base current it is necessary to choose the corresponding value of the well current to obtain high sensitivity. The sensitivity is at a maximum when the voltage on a well is close to the voltage of the base, when there is no base–well junction bias, and therefore the circuit with the general potential on the base and a well is best.

In Figs. 5 and 6 the dependencies of the voltage sensitivity versus of bias current I_{BW} and collector current $I_{C1,C2}$ of the magnetotransistor 3CBMTBW using sample number 832-01 are given. For comparison of the same schedules, data for the BMT magnetotransistor made on a coded as 83, which is similar to an SSIMT, are provided. A high speed of surface recombination was created in these transistors by surface doping with a boron dose of 10 μC/cm^2 and cultivation of a thick oxide of silicon of 0.9 μm thicknesses in water vapor.

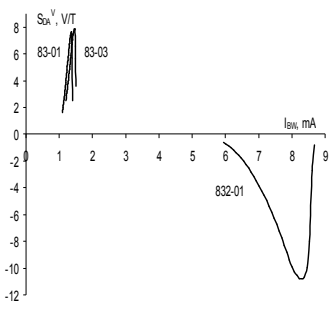


Fig.5. Sensetivity of 3CBMTBW versus base, well current I_{BW}.

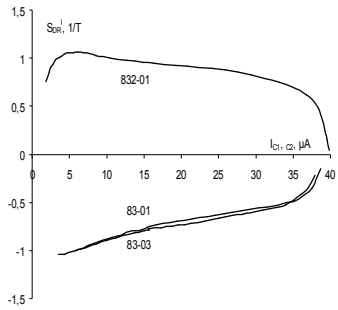


Fig.6. Sensitivity of 3CBMTBW versus collector current $I_{C1,C2}$..

As shown in Fig. 6, on samples 83-01 and 83-03 with a diffusive well and doped base the base current with a substrate in operating mode changes from 1.0 to 1.5 mA. On sample 832-01 the base current with a substrate in operating mode is 6.5–8.5 mA. Distance between the emitter and collectors is for

83 crystals 24 μm in topology and 50 μm in topology 832. Sensitivity increases as a result of increasing the base current and reaches its maximum values near the saturation mode of the collector junction: for samples 83-01 and 83-03, $S_{DA}^{V} = -8$ V/T; for samples 832-01, $S_{DA}^{V} = 11$ V/T. In this operating mode the sensitivity of the voltage of the magnetotransistor of sample 831-02 is higher than that of the other samples and has an inverse sign.

The relative sensitivity to the current is determined by (12). For the magnetotransistor of sample 83-01 the relative current sensitivity $S_{DR}^{I} = -1.04$ 1/T, and for sample 83-03, $S_{DR}^{I} = -1.0$ 1/T. For sample 832-01 the magnetotransistor has almost identical relative current sensitivity the $S_{DR}^{I} = 1.06$ 1/T, but an inverse sign in comparison with devices using samples 83-01 and 83-03.

X. Discussion

The explanation of inversion of the sign of sensitivity is given in the model of distribution of electron flows between electrodes of the 3CBMTBW, presented in Fig. 8 for cases of sensitivity of the magnetotransistor based on the following effects:

(1) the deviation of the injected flow of electrons proceeding in the base;

(2) a magnetoconcentration recombination on a surface of the injected flow of the electrons proceeding in the base;

(3) the deviation of the injected flow of electrons proceeding through a well.

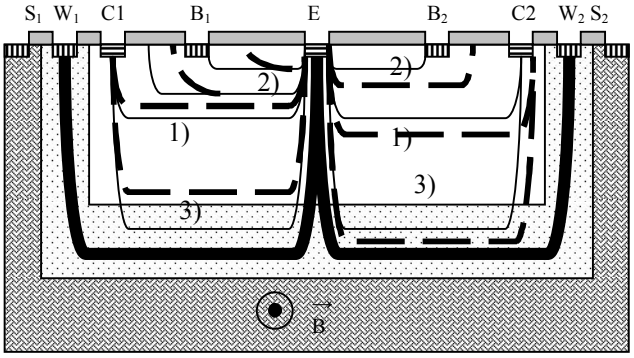


Fig.7. The sensitivity mechanism of 3CBMTBW.

In effects (1) and (2) the injected flow proceeds generally in the base. The high speed of the surface recombination provides high sensitivity of magnetotransistor due to the magnetoconcentration effect (2). In a magnetic field of the line of a current lies on a surface where there is their recombination and current reduction much more, than its growth due to the effect of deviation (1). At small speed of a surface recombination the deviation of the lines of the current passing through the base (1) and through the base–well junction (3) occurs in one direction and makes a joint contribution to the increase in sensitivity. Feature of characteristics is defined by current distribution in the structure of a three-collector magnetotransistor. The large current of the electrons injected from the emitter is extracted in pn-junction base- well. In a magnetic field it deviates in relation to measuring collectors in the same direction as the current, proceeding directly through base. That gives large total change of a current of measuring collectors.

XI. Conclusions

Magnetic field sensor is designed as a dual-collector lateral Bipolar Magnetotransistor with the Base area formed in the Well of BMTBW. The sensor operates with equal base and well potentials, which ensures a threshold of response of the device. An extraction of the injected electrons by a base–well p-n junction determined operating mode the Three-Collector Magnetotransistor – 3CBMTBW. The 3CBMTBW has been investigated with low speed of surface recombination. It was chosen as the best inclusion circuit of 3CBMTBW with the general base and a well. In such circuit the extraction of p-n junction “base –well” has been used instead of a surface recombination. It was established that such option of a design yields reproduced results and high sensitivity. High sensitivity defined by the deviation mechanism of two streams Lorentz's force of the carriers passing through base and through a well.

The establishment of mechanisms of sensitivity magnetotransistor allows to develop gauges of a magnetic field. The knowledge of mechanisms of sensitivity enables to create nanomagnetotransistor, for example, on heterojunction bipolar transistors with high mobility of carriers [11].

References

[1] . Ch. S. Roumenin, “Solid State magnetic Sensors”, ELSEVIER SCIENCE, Amsterdam-Lausanne-New York-Oxford, 1994.

[2] K. Maenaka, H. Okada and T. Nakamura, ”Universal Magneto-operational Amplifier (MOP)”, Sensor and Actuators, A21-A23, pp. 807-811, 1990.

[3] A. Häberli, M. Schneider, P. Malcovati, R. Castagnetti, F. Maloberti and H. Baltes, “2D Magnetic Microsensor with On-Chip Signal Processing for Contactless Angle Measurement”, IEEE Journal of Solid-State Circuits, vol.31, pp. 1902-1907, 1996.

[4] S. Cristoloveanu, “Magnetic Field and Surface Influences on Double Injection Phenomena in Semiconductors, Parts 1”, Phys. Status Solidi A, vol. 64, no. 2, pp. 683–695, 1981.

[5] I.M. Vikulin and V.I/ Stafeev, “Physics of Semiconductor Devices”, Moscow: Radio i Svyaz’, pp. 225–230, 1990.

[6] I.M. Mitnikova, T.V. Persiyanov, G.I. Rekalova and G.A. Shtyubner, “Characterization of Lateral Silicon Magnetotransistors with Two Sensing Collectors”, Fiz. Tekh. Poluprovodn., vol. 12, no. 1, pp. 48–50, 1978.

[7] L.J. Ristic, H.P. Baltes, T. Smy and I. Filanovsky, “Suppressed Sidewall Injection Magnetotransistor with Focused Emitter Injection and Carrier Double Deflection”, IEEE Electron Device Lett., vol. 8, no. 9, pp. 395–397, 1987.

[8] R.D. Tikhonov, “Sign of Bipolar Magnetotransistor Sensitivity as Dependent on Device Structure”, Russian Microelectronics, vol. 37, no. 5, 2008, pp. 334-345.

[9] A.W. Vinal and N.A. Masnari, “Operation Principle of Bipolar Transistor Magnetic Sensors”, IEEE Trans. Electron Devices, vol. 31, no. 10, pp. 1486–1494, 1984.

[10] A.V. Kozlov, M.A. Korolev, A.A. Zsukov, R.D. Tikhonov and A.A. Cheremisinov, “The characteristics of three-collector bipolar magnetotransistor” Russ. Univ. News., Electronics, no. 6, pp. 43-50, 2012.

[11] R.K. Oxland, A.R. Long and F. Rahman, “Magnetotransport characterization of surface-treated InP/InGaAs heterojunction bipolar transistors” Microelectronic Engineering, , vol. 86, no. 12, pp. 2432-2436, 2009.

Porous Silicon Formation on Solar Cells Surface

Tetyana Obukhova
Microelectronics Department
NTUU "KPI"
Kyiv, Ukraine
t.y.obukhova@gmail.com

***Abstract* — In the paper the process of formation of porous silicon on textured surface of solar cells is studied. It is shown that porous layers formed by the etchant HF:$NaNO_2$:H_2O, can be successfully applied in photovoltaic technology as antireflective coating . Technology of the obtaining of porous layer by chemical etching can be applied after contact grid formation. It is shown that as a result of a porous layer efficiency of solar cells increased by 20-30% on average.**

Keywords — porous silicon; solar cell; antireflection coating; chemical etching.

I. INTRODUCTION

Porous silicon has a wide range of properties that depend on its microstructure and etching method. Particular attention of scientists is attracted to photoluminescence properties of porous layers. However, wider study of porous silicon properties has shown possibility for many other applications of porous silicon. There are solar energy conversion, biotechnology and sensors among them.

One of the main problems of silicon solar cell performance is the light reflection. Under normal fall of the light on flat surface of silicon solar cells light losses will reach 30 %. To increase the absorption of the sunshine at fabrication of solar cells the texturing of the illuminated surfaces of silicon, with the help of its previous processing in anisotropic etchant, is used [1,2,3]. It is known that formation of a thin layer of porous silicon on the textured Si substrate allows considerably reduce the reflection of the light in comparison with usual textured substrates [4, 5]. However, methods and mechanisms of porous silicon formation are ambiguous and not studied enough. So the study of mechanisms of formation of porous silicon on textured substrate, as well as its characteristics requires detailed research.

II. SAMPLE PREPARATION

In the work substrates of single-crystal solar-grade silicon (100) orientation, p-type conductivity, resistivity 1 Ohm/cm^2 were used. Textured surface of silicon was formed by way of its previous processing in water solution of KOH. Then p-n junction was formed by phosphorous diffusion and Ti-Cu contact grid is obtained by magnetron deposition and photolithography.

Porous silicon was formed by stain etching in solution based on $NaNO_2$ and HF. Time of etching varied from 5 to 15 sec. The process of stain etching takes place under room temperature and day light illumination. Contact grid was protected by photoresist.

III. RESULTS AND DISCUSSION

A. Porous silicon formation on textured surfaces

During KOH texturing of initial surface of single-crystal silicon a pyramidal relief (Fig. 1) is formed. Formation of such relief occurs due to various velocity of etching of different crystallographic planes. Since the plane (111) is the most resistant to etching, the relief on the surfaces (100) constitute from polyhedrons in the form of tetragonal pyramids with side faces (111).

When porous silicon is formed it starts to appear along the bases of pyramids (Fig. 2 a). The height and sizes of the pores of structured surface do not exceed 50 nm [6].

With increasing of etching time porous layer extends to the pyramids sides moreover sizes of pyramids themselves do not change (Fig. 2 b).

In the case of further increase of the etching time pyramids are completely etched out [6].

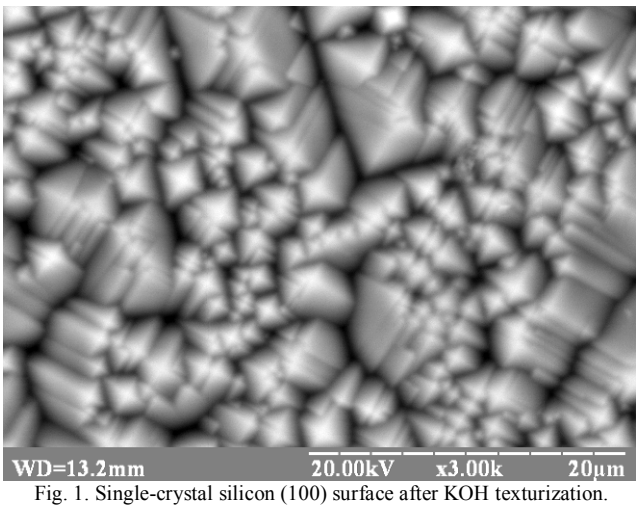
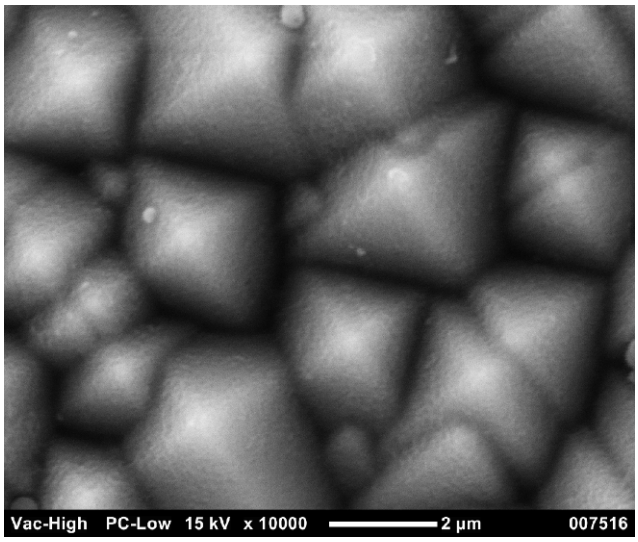
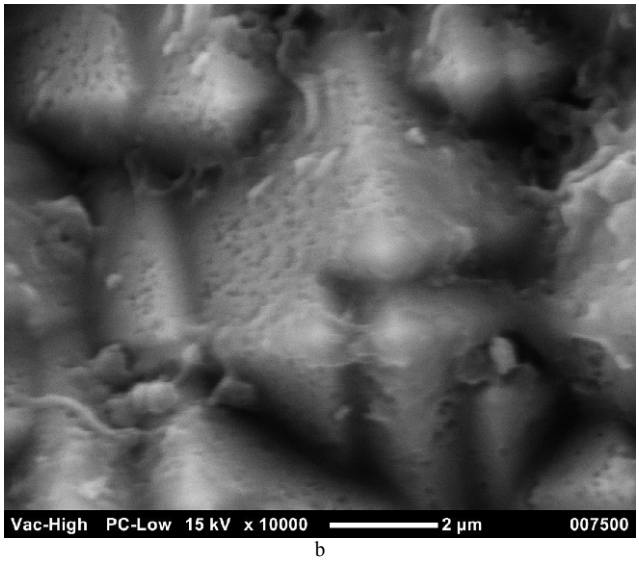


Fig. 1. Single-crystal silicon (100) surface after KOH texturization.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

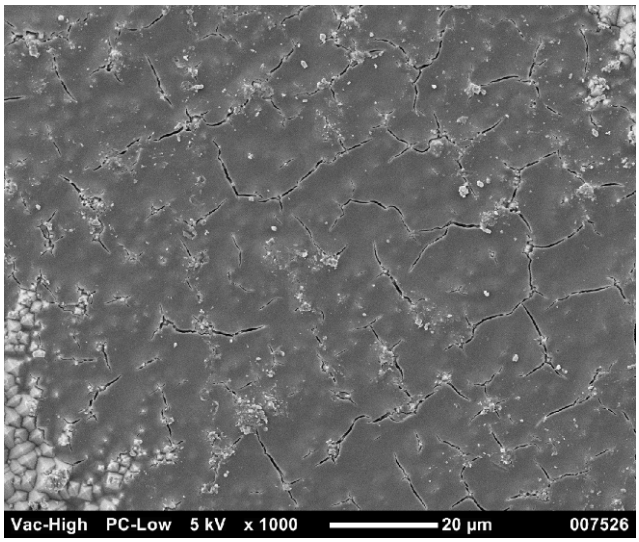


a

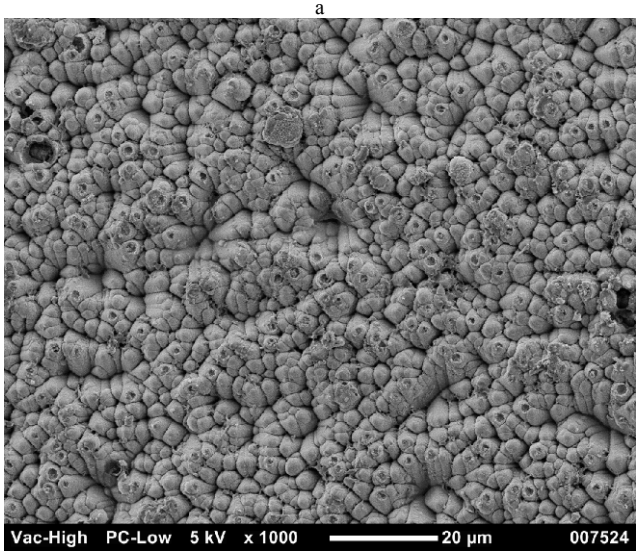


b

Fig. 2. Textured single-crystal silicon (100) surface after porous silicon formation during 5 sec (a) and 15 sec (b)



a



b

Fig. 3. Ti-Cu contact grid surface after porous silicon formation during 5 sec (a) and 15 sec (b)

B. Contact grid degradation

The main problem of porous silicon formation on solar cell surface remains contact grid protection.

Etchants used for porous silicon formation contains HF witch is very corrosive towards majority of metals and even after very little etching time contact grid can be damaged despite protection (Fig 3b). So it is necessary to use very "quick" etchants witch forms porous silicon in few seconds and more chemically resistant metals such as Cu instead of Ag and Al that a usually used.

Use of Au is impossible because of carrier lifetime declining.

C. Optimal etching time

In the case of porous silicon formation on solar cell surface follow needs to be considered:

- Porous silicon integral photosensitivity grows with increase of etching time, until saturation at approximately 90 sec [7].
- Solar cell p-n junction depth is about 100-200 nm and porous silicon etching rate is 3 nm/sec [8], so p-n junction will be damaged after 30-60secetching.

Experimental data show that the greatest increase of solar cell performance is between 7 and 10 sec (Fig. 4). These results are confirmed by SEM data showing that after 10 sec etching porous silicon is already extends to texture sides (Fig. 5a) and in the another hand contact grid still remain homogeneous (Fig. 5b)

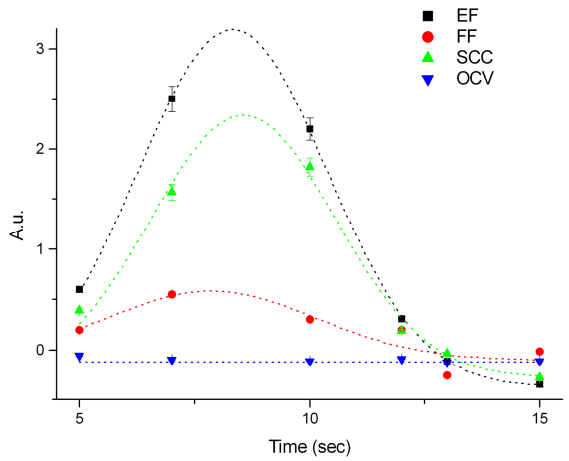


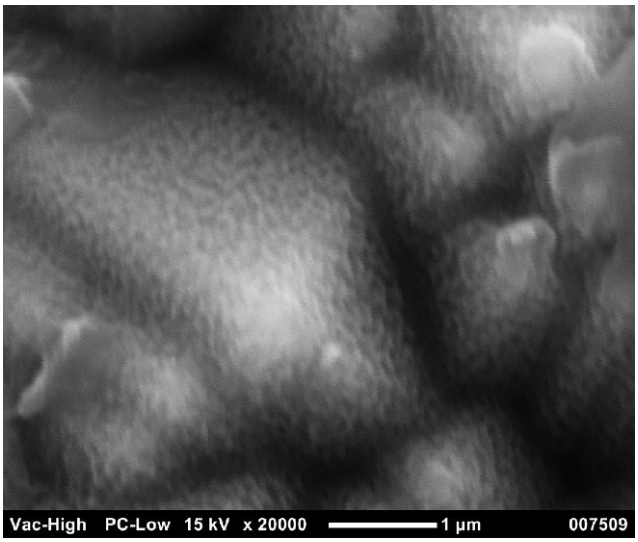
Fig. 4. Solar cell efficiency dependence from porous formation time

EF – efficiency, FF – fill factor, SCC – short circuit current, OCV – open circuit voltage

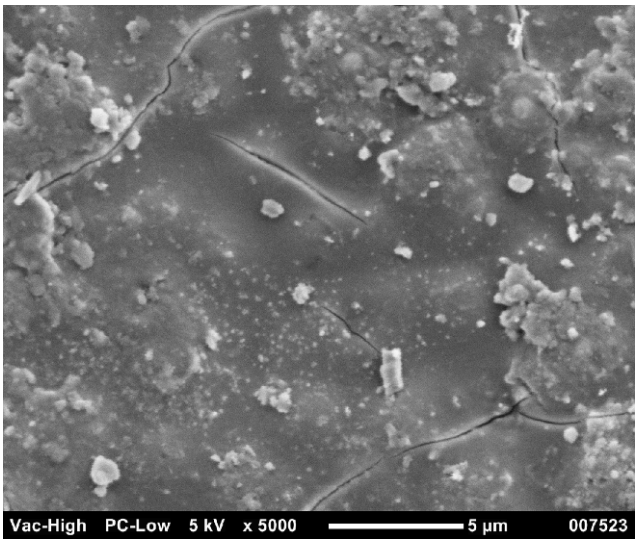
IV. CONCLUSION

Porous layers formed by using the etchant HF:$NaNO_2$:H_2O, can be successfully applied in photovoltaic technology as antireflective coating. Technology of the obtaining of porous layer by chemical etching can be adapted for industrial mass production. It can reduce its costs by replacing energy-intensive process of deposition of silicon nitride.

Studies of solar cells with porous layer revealed that the optimal etching time is 7 - 10 sec. It is shown that as a result of a porous layer formation main parameters of photovoltaic cells increased by 15 – 20 % on average



a



b

Fig. 5. Silicon surface (a) Ti-Cu contact grid surface (b) after porous silicon formation during 10 sec.

REFERENCES

[1] *T. Ya.Gorbach*, L. A. Matveeva, S. V. Svechnikov // Optoelectronics and semiconductor technology. – 1983. – 41. – P. 4.

[2] *T. Ya.Gorbach*, N.V. Kotova, E. V. Pidlisny, S. V. Svechnikov // Optoelectronics and semiconductor technology. – 1986. – 49 – P.10.

[3] *J. Zhao*, A. Wang, P. Altermalt, M.A. Green // Appl. Phys. Lett. – 1995. – 66. – P. 3636.

[4] *M. Saadoun*, H. Ezzaouia, D. Bessais // Solar Energy Materials and Solar Cells. – 1999. – 59. – P. 377

[5] *R.R. Belyalov*, R. Ludemann, W. Wettling, L. Stalmans, J. Poortmans, J. Nijs, L. Schirone, G. Sotgui, S. Strehlke and C. Levy-Clement // Solar Energy Materials and Solar Cells. – 2000. – 60. – P. 391.

[6] *T. Yu. Bilyk*, M. M. Melnichenko, O. M. Shmyryeva, K. V. Svezhentsova Improvement of silicon solar cells performance by using of nanostructured silicon layer // Electronics and Communication. – 2010. – 6, p. 2. – P. 101 – 105.

[7] *T. Yu. Bilyk* , S. S. Shkarnikov Photodensitive porous silicon structures // Electronics and Communication 2012 - 4. P.11-13

[8] *T. Yu. Bilyk* Stain etched porous silicon // Book of abstracts of The IX International Conference and the VIII School for Young Scientists "Silicon 2012" July 9-13, 2012, St. Petersburg, Russia

[9] *T. Yu. Bilyk,* M. G. Dusheiko Solar cells with porous silicon ARC// Book of abstracs of 14-th International Conference "Modern Information And Electronic Technologies", 27—31 May, 2013, Odessa, Ukraine

Modeling of Resonant-Tunneling Diodes with Uniform and Graded Emitter

Artem Fediai, Volodymyr Moskaliuk
Physical and Biomedical Electronics Department
College of Electronics, National Technical University of Ukraine "Kyiv Polytechnic Institute"
Kyiv, Ukraine
artem.fedyay@gmail.com

***Abstract*—Envelope function approach for resonant-tunneling diode (RTD) modeling is developed further. We have found a way of simultaneous treatment of scattering in the quantum well (QW) and incoherent transport, yielding better agreement with experimental I-V characteristics of RTDs with graded and uniform emitter. We also included all our previously developed improvements to the model to obtain maximum adequacy. Since I-V characteristic's shape is affected by numerous factors, we describe our vision on significance of each of them, based on the verification that was done.**

Keywords—resonant-tunneling diode; envelope-function method; current-voltage chatacteristic; modeling

I. INTRODUCTION

Resonant-tunneling diode (RTD) is the most high-frequency semiconductor electronic component; RTD-based oscillators' operation frequency reaches more than 1 THz [1]. RTD is semiconductor component based on a pure quantum effect – tunneling through double barrier, which dependent resonantly on transverse electron energy.

Many models of RTD have been developed by now. As the main feature of RTD is quantum tunneling, the main problem in its modeling is to describe quantum rather than semiclassical transport phenomena. Envelope-function method, Wigner's function method and nonequilibrium Green's function formulation are used for RTD modeling (see reviews [2, 3]). Many different models were developed within each approach having, however, one feature in common, which is a bad agreement with experiment. The best quantitative agreement with experiments was reached within Green's functions formalism and envelope function approach. However, envelope function method sometimes uses as many as 7 fitting parameters [4, 5]; Green's functions method, being powerful and universal, cannot, however, be properly reduced to computationally effective and good-predicting one at the same time [3]. Both of these approaches cannot inherently predict a correct shape of I-V characteristic and so-called plateau region in it in particular.

In this work we are intending to find unbalanced assumptions, responsible for low prediction abilities to replace those by more adequate. We try to improve prediction ability of the envelope-function models [4-7] and remove most of fitting parameters. At the end of a paper we compare modeling results with the experimental I-V characteristics to show particular significance of the work were done.

This paper is organized as follows. First, we introduce 1-band and 2-band coherent models. Then we discuss treating electrons' interactions within one-electron approach. Mentioned part of the theory can be viewed as a *core* of a model.

Secondly, we describe and discuss the ways of treating "secondary" factors, influencing electron transport. They are:

1) scattering in the quantum well and peculiarities of self-consistency procedure in presence of incoherent channel;

2) electron transport through emitter quantum well;

3) real shape of heterojunctions;

4) resistance of the passive regions.

We will conclude with discussion of the verification results for several RTDs.

II. MODEL (THE CORE)

A. Partitioning

Area of the RTD's mesastructure is always large enough to think of it as it was infinite and, thus, to skip it from explicit consideration.

In the direction of growth (hereafter, *z*-direction) resonant-tunneling diode is considering as consisted of the two regions, treating in a different way, because of the difference of those properties and dimensions (fig. 1). Spacers and double-barrier quantum system (DBQS) are referred to as "device"; envelope function approach and its extensions are applied for it. Long highly-doped regions, surrounding "device" are so-called "reservoirs". Electrons in the reservoirs are treating as free classical particles, having effective mass of reference (i.e. in a semiclassical way). It is assumed that reservoirs' electron gas is in equilibrium.

In contrast to [4] we fix a border between the device and reservoirs strictly between the spacers and highly-doped regions. Within 1-band model only Γ-valley electrons are taken into account.

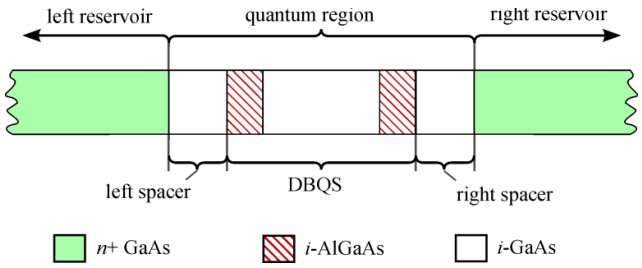


Fig. 1. Modeling region of AlGaAs/GaAs RTD and its partition.

B. Electric current formula

Principal point of the model is treating "device" as a "scatterer", characterized by *transmission function* $T(E_z)$. Landauer approach for 3D reservoirs and E_x, E_y-independent $T(E_z)$ yield famous *Tsu-Esaki formulation*

$$J = \frac{2m^* e k_B T}{(2\pi)^2 \hbar^3} \int_{eV}^{\infty} T(E_z) D(E_z) dE_z \qquad (1)$$

where e, k_B, $\hbar$ are known constants; T is temperature, V is applied voltage;

$$D(E_z) \equiv \ln\left(\left\{ 1 + \exp\left(-\frac{E_z - E_F^L}{k_B T} \right) \right\} \Big/ \left\{ 1 + \exp\left(-\frac{E_z - E_F^R}{k_B T} \right) \right\} \right)$$

is so-called *supply function*, in which $E_F^{L(R)}$ is a Fermi level in the left (right) reservoir.

From (1) it follows that the problem of finding electric current has reduced to that of defining transmission function $T(E_z)$.

C. Transmission function

Transmission function is defining from 1D effective mass Schrödinger equation (U and ψ are electron potential energy in the conduction band and envelope of z-direction wave function, accordingly)

$$\left[-\frac{\hbar^2}{2} \frac{d}{dz} \frac{1}{m^*(z)} \frac{d}{dz} + U(z) \right] \psi(z) = E_z \psi(z) \qquad (2)$$

as a ratio between transmitted ($j_R^{\rightarrow}$) and incident ($j_L^{\rightarrow}$) probability flux:

$$T = j_R^{\rightarrow} / j_L^{\rightarrow}. \qquad (3)$$

It follows from (2) that we need $U(z)$ to calculate $T(E_z)$. Within *Hartree approach*, we can write: $U(z) = E_c(z) + U_s(z)$, where $E_c(z)$ is a potential profile of conduction band due to band wrapping; $U_s(z) = -eV_s(z)$, where $V_s(z)$ is a Hartree self-consistent potential, which is to be found from Poisson equation:

$$\frac{d}{dz} \varepsilon(z) \frac{dV_s}{dz} = \frac{e}{\varepsilon_0} \left[n(z) - N_d^+(z) \right], \qquad (4)$$

where ε_0 (ε) is a vacuum (relative) permittivity; n and N_d^+ denote electrons concentration and ionized donor atoms concentration, respectively.

If we wish to perform quick and rough RTD simulation, we can simply assume that potential is constant within reservoirs and drops linearly in the device (just like in the plane capacitor):

$$V_s(z) = \begin{cases} 0, & z < z_L, \\ \dfrac{V}{L} z, & z \in [z_L, z_R], \\ V, & z > z_R, \end{cases} \qquad (5)$$

where V is a total applied voltage; $[z_1, z_5]$ is a device region; $L = z_R - z_L$.

We do not need to know $n(z)$, if we apply assumption (5); otherwise $n(z)$ should be specified. Models using Hartree approach are referred to as "*self-consistent models*"; "*nonselfconsistent*" models are those accepting condition (5) for potential distribution.

D. Concentration

Expressions for $n(z)$ differs for reservoirs and device region:

a) *reservoirs*. Electron gas is assumed to be in a local equilibrium with reservoir (voltage drop is equal to the difference of Fermi levels of reservoirs). For left (right) reservoir [defined as $z < z_L$ ($z > z_R$)] this yields an ordinary expression for 3D equilibrium electron gas in a low field:

$$n(z) = 4\pi (2m^*_{L(R)}/h^2)^{3/2} \int_{U(z)}^{\infty} \frac{\sqrt{E - U(z)}}{1 + \exp\left(\dfrac{E - E_F^{L(R)}}{k_B T} \right)} dE \; ; \qquad (6)$$

b) *device*. We suppose that distribution function of electrons within a device $N(E_z, z)$ is simply proportional to a sum $N_L |\psi_L(z)|^2 + N_R |\psi_R(z)|^2$, where subscripts "$L$" and "$R$" indicates left and right reservoir. Concentration $n(z)$ is to be found by integrating $N(E_z, z)$ by all relevant transverse energies E_z, yielding:

$$n_{L(R)} = C \int_{U_L}^{\infty} \frac{|\psi_{L(R)}|^2}{\sqrt{E_z - U(z_{L(R)})}} \ln\left(1 + e^{-(E_z - E_F^{L(R)})/k_B T} \right) dE_z \, , \qquad (7)$$

where $C \equiv \sqrt{2} m^*_{L(R)}{}^{3/2} k_B T / \left[(2\pi)^2 h^3 \right]$.

Schrödinger equation (2), subjected to the boundary conditions corresponding to left- and right-going electron must be solved to obtain $\psi_L(z)$ and $\psi_R(z)$. Amplitude of the correspondent incident waves must be normalized to 1.

E. Two-band model

For numerous of ternary alloys like $Al_xGa_{1-x}As$ ($x > 0{,}45$) elastic interface Γ-X intervalley scattering plays significant role. We have applied a method, derived in [8], within which

corresponding wave functions are solutions of the 2 coupled Schrödinger equations:

$$\begin{bmatrix} -\frac{\hbar^2}{2}\frac{d}{dz}\frac{1}{m_\Gamma^*}\frac{d}{dz} & 0 \\ 0 & -\frac{\hbar^2}{2}\frac{d}{dz}\frac{1}{m_X^*}\frac{d}{dz} \end{bmatrix}\begin{bmatrix}\psi_\Gamma \\ \psi_X\end{bmatrix} + \\ + \begin{bmatrix} U_\Gamma - E + \frac{\hbar^2 k_\parallel^2}{2m_\Gamma^*} & \alpha\delta(z_k) \\ \alpha^*\delta(z_k) & U_X - E + \frac{\hbar^2 k_\parallel^2}{2m_X^*} \end{bmatrix}\begin{bmatrix}\psi_\Gamma \\ \psi_X\end{bmatrix} = 0,$$

where subscripts X and Γ denotes the valley; α is so-called coupling parameter, characterized strength of intervalley scattering.

The rest of the two-band model is developed in the same manner as one-band model, except that new channel of transmission appears, and concentration distribution changes due to electrons redistribution between Γ and X–valleys [7].

III. SECONDRY FACTORS

A. Scattering in the quantum well

Non-resonant current is mainly caused by electron waves, which suffered scattering in the quantum well [9]. We apply "optical potential" *ad-hoc* method [10] to take this effect into account. Let $\hat{H}_0$ be Hamiltonian of the coherent model. Then $\hat{H} = \hat{H}_0 - iW$ is a Hamiltonian of a system, in which probability density decrease in the direction of propagation ($W \equiv \hbar/\tau$ where τ is chosen to be close to momentum relaxation time of a bulk semiconductor [4]). This decaying is attributed to exit of electrons from the *coherent channel*.

Loosen probability flux is then artificially returning to a system to preserve overall probability density conservation. An approach we use to take this into account is a model of "single scatterer" [9]. It is assumed that there is only one scatterer in the center of a quantum well. If electron suffers scattering, it tunnels to the left or to the right with probability proportional to T_L and T_R, which are the probabilities to reach left and right reservoirs being incident from the center of a well.

Total transmission probability T of a device is then a sum of the coherent term T_{coh} (which is calculating as usual), and incoherent term $T_{incoh} = (1 - R_{coh} - T_{coh}) \times T_R / (T_L + T_R)$; R_{coh} is a coherent reflection probability.

In contrast to [4, 5] we calculate T_L and T_R using strict expression for $U(z)$ rather than its approximation. There is a contradiction between distributed decaying of probability flux and locality of scattering centre (it is a matter for selfconsistent modeling only). We addressed this question and found that it is impossible to skip this contradiction in a traceable manner. However, if we are seeking for $V_s(z)$, assuming there is no scattering, and transmission coefficient is to be determined according to above mentioned procedure, modeling results seems to be reasonable. Such self-consistency can be referred to as a partial one; by our understanding, difference in the distribution of coherent and incoherent electrons mustn't be larger than that in case of using linear drop approach for potential.

It is worth to note that these questions have not been addressed in the works [4, 5] despite of their critical importance. Therefore, we cannot compare our approach with those.

B. Electron transport through emitter quantum well

Interaction between electrons states in emitter quantum well (EQW) and main quantum well (MQW) are proven to be a cause of plateau region in the I-V characteristics [11]. EQW can be populated, if and only if inelastic scattering take place in it. However, envelope function method inherently cannot predict any inelastic scattering. In [12] we have developed *ad-hoc* method taking inelastic scattering into account by introducing so-called "*virtual reservoir*", as depicted at fig. 2. Virtual reservoir models a source of inelastically scattered electrons. In [12] we have obtained parameters of it corresponding to a given inelastic scattering time.

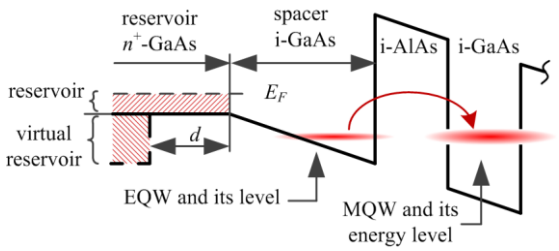


Fig. 2. Introduction of a virtual reservoir to take EQW-MQW transport into account (AlAs/GaAs material base shown for convenience).

C. Real shape of heterojunctions

Abrupt changing of electron's potential energy at the heterointerfaces is a rather rough assumption for RTD's layers having lattice's constant order thickness. In [4] smooth approximation was applied. It includes 2 parameters that regulate the shape of heterojunction; however, no physical reasons for such particular approximation were done (as a result, 2 more fitting parameters of a model appears). In [13] within pseudopotential method it was proven that one can reach better adequacy using two-step model of a barrier (as shown in fig. 3) – so-called "superlattice model of a smooth heterojunction". For AlAs/GaAs interfaces these parameters are as follows: $E_{\Gamma 1} = 0.3406$ eV, $E_{\Gamma 2} = 0.3864$ eV; $a = 0.5653$ nm (equal to the lattice constant of GaAs).

Using superlattice model of heterojunction adds no qualitative peculiarities to the shape of I-V characteristics. According to our numerical experiments, it predicts slightly enhanced current (5-10%) more or less uniformly along all voltage range.

D. Resistance of the passive regions

Passive layers are macroscopic and highly-doped, therefore they can be taken into account straightforwardly. In [4] resistance of the passive regions was used as a pure fitting parameter. Our calculation has shown, however, that this resistance can be neglected compared to resistance of active

region. Significant influence on I-V characteristic can cause nonlinear resistance of the contacts, but it is impossible to develop universal model to take this factor into account.

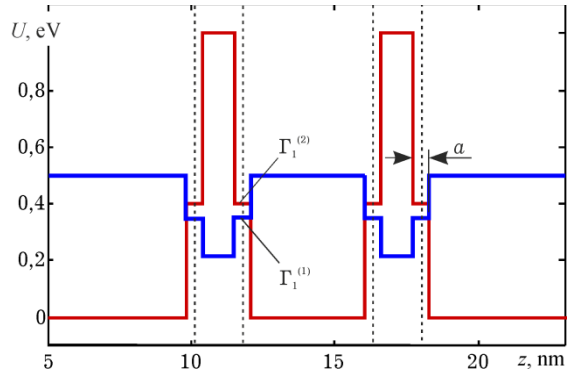


Fig. 3. Band structure of AlAs/GaAs RTD at Γ and X-valleys according to [13]. Double barrier structure is Γ-valley profile.

IV. VERIFICATION

Taking into account the mentioned factors, we have reached one of the best agreements with experimental data; but what is more important we have managed to keep only 2 fitting parameters: area of a mesa-structure (to fit experimental peak current value) and scattering rate, which appears during *W* calculation. In fig. 4 we compare experimental I-V characteristics of RTD1, described in [14] with those, calculated by our self-consistent 1-band model as well as with I-V curve, calculated by *Abramov et.al.* in [4], where authors use 7 fitting parameters. It is a typical verification result for RTD with ordinary emitter that justifies that we made appropriate reduction of the fitting parameters' list and appropriate changes to the model. We have used 2 fitting parameters: peak current density and scattering rate; we have also shifted obtained I-V curve to the right to take nonlinear resistance of the contacts into account. The same is done at fig. 5-7.

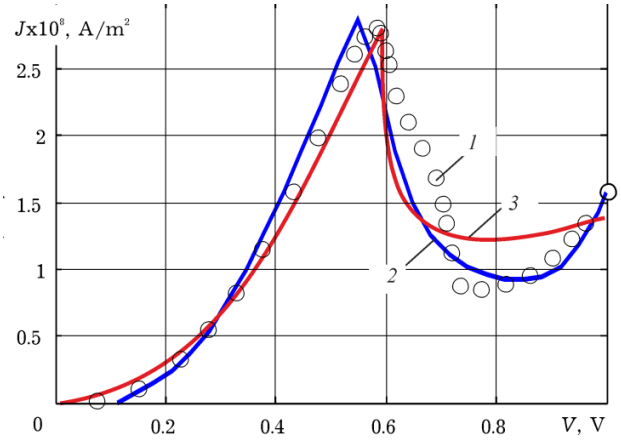


Fig. 4. Verification of RTD1: *1* – experimental I-V curve [14]; *2* – our model; *3* – model, developed by *Abramov et.al.* [3].

In [5] 2-band model of RTD was developed. Using a value of α as large as 1 eV Å and choosing appropriate values of 7 others fitting parameters, plateau region was obtained at I-V characteristic that resembles experimental one. But the value of parameter α is 1 order bigger then recommended value in [8]. That is why it is hard to trust in such model. According to our numerical experiments for most RTDs correction due to 2-band model weren't noticeable in the plateau region. However, for RTD2 from [15] current density for voltages exceeding those of a valley current, calculated by 2-band model is higher, thus, closer to experimental curve than that of 1-band model (fig. 5). But in most cases those region of voltages is not relevant, therefore 2-band model is usually have only fundamental importance. Moreover, one iteration of 2-band model is as time-consuming as 10 those of 1-band model. Therefore, we do not recommend using 2-band model for I-V characteristics prediction of ordinary double-barriers RTD.

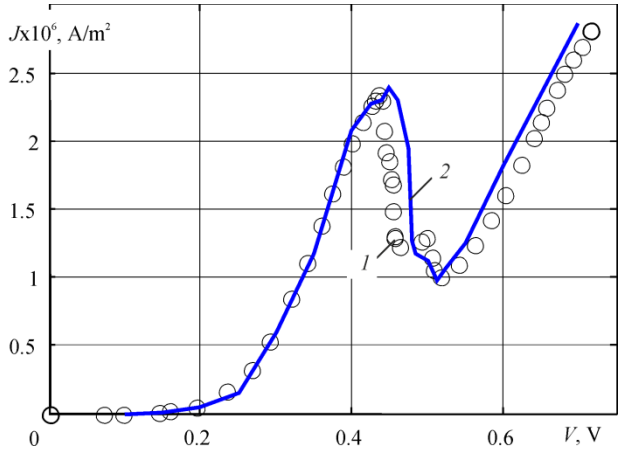


Fig. 5. Verification of RTD2: *1* – experimental I-V curve [15]; *2* – our 2-band model.

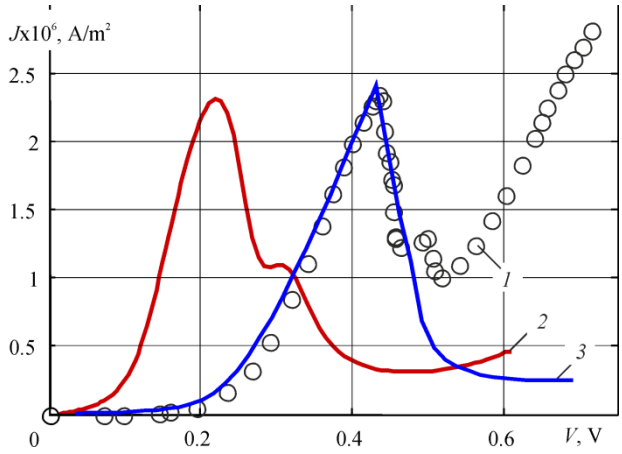


Fig. 6. Verification of RTD2: *1* – experimental I-V curve [15]; *2* – our 1-band nonselfconsistent model, taking EQW-MQW transport; *3* – our 1-band selfconsistent model.

Plateaus at I-V curve can be predicted only within the linear drop of approximation of potential (formula); it is true for RTDs with both ordinary (uniform) and graded emitters. In fig. 6 experimental data for RTD2 (it has uniform emitter) is shown at the same plot with the result of 1-band self-consistent and nonselfconsistent modeling. The same plot is presented in fig. 7 for RTD3 with graded emitter. It is obvious from those that Hartree self-consistence works fairly well at the raising part of I-V characteristic and fails at the falling one. In contrast,

to predict plateaus we should use linear drop approximation of potential, but peak voltage, of course, will be considerably lower. It is worth to note that for graded and uniform emitter physical mechanisms of plateau formation are different [16].

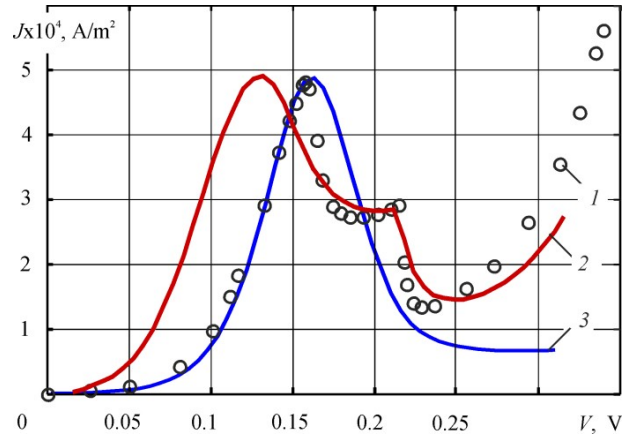


Fig. 7. Verification of RTD3: *1* – experimental I-V curve [17]; *2* – our 1-band nonselfconsistent model, taking EQW-MQW transport into account; *3* – our 1-band selfconsistent model.

CONCLUSION

In conclusion, we have further developed an envelope function model of RTD. Our strategy was to revise existent ways of modeling to exclude unnecessary factors and to correct existent assumptions appropriately. As a result we have reduced a number of fitting parameters from 7 to 2 by removing parameters as follows: a "flowing" border between device and reservoirs (now it coincides with the border between spacers and highly-doped regions); interface charge; energy of optical phonon; parameters of a smooth heterojunction (we use superlattice model with a fixed parameters), resistance of the passive regions (its value is always negligible). We have treated quantum charge in a whole modeling region by Hartree method as if electron transport were purely coherent, and we take the scattering into account within single scatterer model in more accurate way; we have introduced a method of treating EQW-MQW transport and disprove the statement [5] that plateau regions appears as a result of intervalley interface scattering.

Verification has shown that prediction abilities of the model have been improved. The problem left is prediction of plateaus: although we managed to obtain those within linear drop potential approximation, they are still unpredictable by selfconsistent model. Besides, nonlinear resistance of the contacts' junctions inherently cannot be taken into account within unified manner. Despite of these shortcomings, we made a significant step to understand physical processes in RTDs and to the realistic prediction of its static characteristics, having in hand only RTD's design parameters.

REFERENCES

[1] S. Suzuki, M. Asada, A. Teranishi, H. Sugiyama, and H. Yokoyama, "Fundamental oscillation of resonant tunneling diodes above 1 THz at room temperature," Appl. Phys. Lett. , vol.97, no.24, pp.242102-242102-3, Dec 2010.

[2] J.P. Sun; G.I. Haddad, P. Mazumder, and J.N. Schulman, "Resonant tunneling diodes: models and properties," *Proc. of the IEEE* , vol.86, no.4, pp.641-660, Apr 1998.

[3] I.I. Abramov, "Problems of physics and modeling of device structures of micro- and nanoelectronics. V. Resonant-tunneling structures," *Nano- and microsystem's technique*, vol. 3, pp.57-70, 2007 (in Russion).

[4] I.I. Abramov, I.A. Goncharenko, and N.V. Kolomeitseva, "Combined model of resonant-tunneling diode," *Fizika I tekhnika poluprovodnikov*, vol.39, no.9, pp.1138-1145, 2005 (in Russian).

[5] I.I. Abramov and I.A. Goncharenko, "Combined 2-band model of resonant-tunneling diode," *Fizika I tekhnika poluprovodnikov*, vol.41, no.11, pp.1395-1400, 2007 (in Russian).

[6] J.P. Sun and G.I. Haddad, "Self-consistent scattering calculation of Resonant Tunneling Diode Characteristics," *VLSI Design*, vol.6, no1-4, pp.83-86, 1998.

[7] J.P. Sun, R.K. Mains, K. Yang, and G.I. Haddad, "A self-consistent model of Γ-X mixing in GaAs/AlAs/GaAs quantum well using quantum transmitting boundary method," *J. Appl. Phys*., vol.74, no.8, pp.5053-5060, Oct 1993.

[8] T. Ando and H. Akera, "Connection of envelope functions at semiconductor heterointerfaces. II. Mixing of Γ and X valleys in GaAs/AlxGa1-xAs," *Phys. Rev. B*, vol.40, no.17, pp.11619-11633, 1989.

[9] M. Buttiker, "Coherent and sequantial tunneling in double barrier," *IBM. J. Res. Develop*, vol.32, no.1, pp.63-75, Jan. 1988.

[10] Y. Zohta and T. Tanamoto, "Improved optical model for resonant tunneling diode," *J. Appl. Phys*, vol.74, no.11, pp. 6996–6998, Dec 1993.

[11] Z.J. Qiu *et.al.* , "Experimental verification on the origin of plateau-like current–voltage characteristics of resonant tunneling diodes," *Applied Physics Letters* , vol.84, no.11, pp.1961-1963, Mar 2004.

[12] V.A. Moskaliuk, A.V. Fedyay, "Modeling of artifacts at the current-voltage characteristics of the resonant tunneling diode," *Microwave and Telecommunication Technology (CriMiCo), 2011 21th International Crimean Conference* , pp.787-788, 12-16 Sept. 2011.

[13] G.F. Karavaev and S.N. Griniaev, "Superlattice model of a smooth heterojunction GaAs/AlAs (001)," *Fizika tverdogo tela*, vol.48, no.5, pp.893–901, 2006 (in Russian).

[14] T. Boykin, R.C. Bowen, G. Klimeck, and K.L. Lear, "Resonant-tunneling diodes with emitter prewells," *Applied Physics Letters* , vol.75, no.9, pp.1302-1304, Aug 1999.

[15] S.V Evstigneev, A.L. Karuzskii, Yu.A. Mityagin, A.V. Perestoronin, D.S. Shipitan, "Multiple-barrier resonant tunneling structures for application in a microwave generator stabilized by microstrip resonator," *8th Int. Symp. "Nanostructures: Physics and Technology". St Peterburg, Russia*, pp. 494–497. 19–23 June 2000.

[16] A.V. Fediai and V.A. Moskaliuk, "Hartee approach and linear drop approximation of potential for modeling of resonant-tunneling diode," *5th International scientific conference "Functional base of nanoelectronics". Collected papers, Kharkov (Ukraine),. HNURE,* pp.42–45, 30 Sept – 5 Oct 2012 (in Russian).

[17] R. Yatskiv, J. Voves, "Analysis of the resonant tunneling diode with the stepped pre-barrier", *Journal of physics: conference series*, vol. 193, pp.1-4, 2009.

Functional Film Peculiarities of the "Metal- Iinsulator-Metal" Structure Devices

N. Korobova, V. Vodopyanov, Yu. Stepanova,
Yu. Cherkasova, S. Timoshenkov, A. Melnikov
Department of Microelectronics
National Research University of Electronic Technology
Moscow, Zelenograd, Russia,
korobova3@mail.ru

O. Isaikina
Topchiev Institute of Petrochemical Synthesis
Russian Academy of Sciences
Moscow, Russia

N. Plis
"Angstrom" Co, Moscow, Russia

***Abstract*— The important parts of material science in the film preparation fields and sol-gel technology are presented. Al_2O_3/SiO_2 and ZrO_2/Y_2O_3 films have been prepared by electrophoresis deposition of organometallic compounds as dielectric coatings on metal substrates and annealed at different temperatures. The films were characterized by X-ray diffraction (XRD), atomic force and scanning microscopy (AFM) and (SEM) respectively. The surface grains appeared to get larger in size and the root mean square roughness of the annealed films increased with the annealing temperature but was less than that of the as-deposited. The properties dependence of the film's chemical composition, impurities and microstructure, which are closely correlated with the deposition technique and conditions were investigated. A short analysis of sol-gel technology and thin film preparation methods of dielectric materials has been given.**

Keywords— thin films, functional dielectrics, microstructure

I. INTRODUCTION

Metal-insulator-metal structures are often utilized in a variety of different sensors. They are relatively inexpensive compared to other sensing technologies, robust, lightweight, long lasting and benefit from high material sensitivity and quick response times [1]. The need for miniaturization of electronic components will certainly continue. Advances in system technology in many respects have become materials limited. This recognized situation has resulted in research and development of imported electronic oxides ceramic materials. Many techniques such as electron beam evaporation, pulse laser deposition, sputtering and sol-gel [2] have been employed to deposit zirconia and alumina – silica films. Sol-gel processes are widely employed in the field of electronics to prepare ceramic powders and thin films, high-quality ceramics and insulators, offering the advantages of low cost, relative simplicity, and easy control of the composition of the layers formed [3].

In general, sol-gel processing requires a densification step at higher temperature to obtain dense glasses. Organic groupings can act as network modifier, reducing network connectivity and can affect a drastically change of properties compared to inorganic glasses, if organic groupings are cross linked, they can act as network formers.

A variety of unique microstructures (grades, multiphase, partially porous to fully dense, etc.) with unique properties can be investigated when processing is performed by electrophoresis. In addition, the precursor can be used to both increase the relative density and strengthen the films without shrinkage [3].

In this paper we consider preparation of oxide films like mullite composition (Al_2O_3-SiO_2) because of good dielectric properties for high-temperature sensors [4]. Zirconium oxide films have wide applications including high-reflectivity mirrors, broadband interference filters, waveguides with small losses and active electro-optical devices [5], due to their excellent properties such as high refractive index and high pulse laser damage threshold. We used stabilized ZrO_2 thin films as gate insulators with high dielectric constants in electronic devices.

II. EXPERIMENTAL PART

Purpose of the current work is to determine the kinetic preparation of electrophoretic films. For Seria A sol formed by mixing the aluminum alkoxide and silica sol (prepared using colloidal TEOS from 2.4 parts) was gelled at a pH of 3. Glacial acetic acid and 2-methoxy ethanol were used as the solvents. For Seria B sol formed with organic solvent with mixture of (Zr + Y) alkoxides which was pyrolyzed to ZrO_2/Y_2O_3 solid solution.

Seria A – preparation (Al_2O_3 + SiO_2) films

Sol was synthesized using aluminum iso-propoxide and stoichiometric amount of silicon tetraethoxide = 15g, and 3,7 g 95% ethanol, which were commercially obtained and used without further purification and 0,5-1 h under vigorous stirring to obtain translucent, homogeneous, stable sol. The stainless steel substrate was cleaned with ethanol-acetone solution in an ultrasonic bath for 20 min, washed with distilled water and then dried before used. Thin films of (Al_2O_3 + SiO_2) were manufactured on a stainless steel substrate by electrophoretic deposition process from prepared alcohol sol. We try to keep attention to the sol aging time- that is the time passed from its preparation to

the moment of utilization. In the work [4] we have considered the influence of aging sol on the morphology and structure of sol particles and also the sol-gel chemistry of $Al(iso\text{-}C_3H_7O)_3$ and $Si(C_2H_5OH)_4$ in various alcohols. The deposition always occurred at the cathode with a constant electric field of 200-300 V/cm for a time period of 0,5 - 2 min. The thickness of a deposited gel layer was usually in the range of 1-3 μm.

Seria B- preparation of stabilized ZrO_2 films

The stabilization of tetragonal zirconia at low temperatures in bulk materials is commonly achieved by decreasing the grain size or adding a dopant, such as Y_2O_3 or MgO, to lower the tetragonal-to-monoclinic transformation temperature. For the preparation ZrO_2/Y_2O_3 films zirconium and yttrium butoxides were used as the precursor materials and 2-methoxy ethanol as the solvent. With this precursor solution ethylene glycol has been added as an additive in order to avoid the film cracking which degrades the electrical properties of the films. The precursor solution of ZY was then deposited Pt substrates by spin coating at 2300 rpm for 60 sec and by electrophoretic deposition as in Seria A. The substrates were ultrasonically cleaned in acetone, methanol and rinsed in deionized water, followed by a drying process. After coating process, the films were kept in a preheated furnace at 300 ^oC for 10 minutes to remove the volatile components. The as-deposited amorphous films were subsequently annealed at different temperatures from 400^o to 700^{o}C in order to convert them into crystalline films.

The phase identification of the films was done by powder X-ray diffraction analysis using CuK_α radiation (λ=1.5418 Å). Surface morphology and microstructure of the films were observed using Scanning Electron Microscopy (SEM, Stereoscan 440model). X-ray Photoelectron Spectroscopy (XPS) for composition analysis was performed in a VG Microtech MT-500 Spectrometer with MgK_α radiation in the binding energy range of 0 to 1000 eV. Capacitance-Voltage properties were measured in a metal-insulator-metal configuration using Capacitance Bridge.

III. RESULTS AND DISCUSSIONS

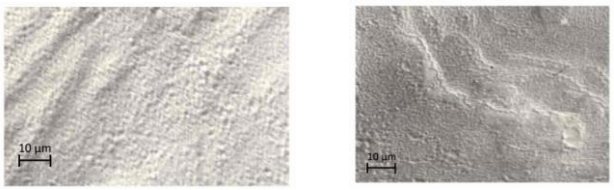
X-ray diffraction studies:

The structure of the Al_2O_3/SiO_2 films was carefully examined. The XRD patterns of the films prepared from fresh sols and annealed at various temperatures (up to 400^{o}C) show no diffraction peak of Al_2O_3 or SiO_2 films. This amorphous behavior remains unaltered up to an annealing temperature of 400^{o}C. With the annealing temperature increasing, a substrate diffraction peak became weaker, which may be due to the Al_2O_3/SiO_2 films got denser and weakened the intensity of the diffraction peak. While annealing at 500 ^{o} C, the Al_2O_3-SiO_2 film was found to be crystalline with (110), (111), (200) peaks arising at 2θ values. With further increase in the annealing temperature, peaks became increasingly sharp. After annealing the Y_2O_3-ZrO_2 film at 700 ^{o} C, the intensity of the diffracted peaks increased showing an increase in the crystallinity of the film. The structure was found to be tetragonal with grain size about 10-20 nm for 3,0 mol.% of Y_2O_3.

Surface morphologies

The morphology, microstructure and component of Al_2O_3/SiO_2 and ZrO_2/Y_2O_3 films were characterized, respectively by employing several experiment instruments. The surface morphologies of samples annealed at different temperatures were shown in Fig. 1(a)–(b). The grain size appeared to increase gradually with temperature, which may due to higher temperature annealing gathered more atoms together from available neighboring sites.

Fig. 1a shows the SEM image of the micro cracks observed on the Y_2O_3-ZrO_2 film annealed at 700 ^{o} C, which degrades the electrical properties of the film. The cause of the film cracking was due to the large volume change when the wet solution is fired to inorganic oxides during annealing. This problem was overcome by the addition of ethylene glycol in the precursor solution [3]. The high boiling point and latent heat of vaporization of ethylene glycol retains the atom mobility and significantly reduces the tendency to crack. Fig. 1b shows the dense film with fine grains and without any cracks annealed at 700 ^{o} C. It was also observed that with the increase in annealing temperature the grain size also increases.



a b

Fig. 1. SEM image of Y_2O_3-ZrO_2 films

Heat treatment temperature was raised stepwise at intervals of 100 ^{o}C for 10 min. The thickness of the coating films, as measured by surface profilometer, was in a range of 1300 to 1500 nm, depending on the mole ratio $Zr(OR)_4$ / $Y(OR)_3$ and on the degree of dilution of the coating solution by organic solution in $Al(OR)_3$ /$Si(OR)_4$. Many variables, such as composition of starting materials and low or big water content affect the gels produced during the sol-gel transition. The rheological property of a sol may provide information on sol and gel structures.

In-depth chemical composition of the film was analyzed by Auger Electron Spectroscopy with Ar+ ion beam of 2kV as the sputtering beam. From the depth profile, the film is believed to be divided into three regions: the surface, the bulk film and the interface. Although a carbon peak was detected on the surface of the films, as soon as the first monolayers are removed by the Ar ion bombardment, the carbon peak intensity decreases to the level which cannot be distinguished from noise. The peak is assumed to be from contamination during the sample handling and preparation.

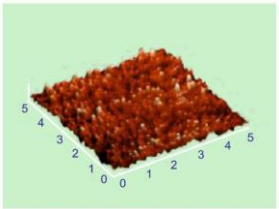
Electrical properties. The dielectric and leakage current density measurements were made on the films fabricated on

platinum or steel foils. A metal-insulator-metal (M-I-M) structure was fabricated by evaporating circular platinum dot electrodes of 0.5mm diameter on to Al_2O_3-SiO_2 film by shadow masking. Platinum has been widely used as the electrode because of its excellent oxidation resistance. Also the use of platinum for top and bottom electrode resulted in good electrical properties. A ramp bias voltage with a superimposed a.c signal of amplitude 100 mV and frequency (100 kHz) was applied to the Al_2O_3-SiO_2 capacitor, and the applied bias voltage was swept at a ramp rate of 0.5 V/s for both polarities. Dissipation factor was 0.0236. This value is comparable with other sol-gel deposited films, but lower than that of epitaxial films. This may be due to the variation in the microstructure. But in the present case the films were polycrystalline with random orientation.

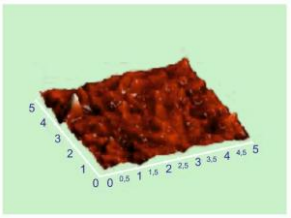
The major distinction of the present work lies in the method of accomplishing the hydrolysis reaction. In our case, water is not added to the system. Instead the metal alkoxide complex/alcohol solution is exploited by the electrophoretic deposition in order to prepare thin solid films. Gels prepared by hydrolysis of metal alkoxides during electrophoretic deposition at room temperature and standard atmospheric pressure are generally porous and contain water and alcohol. When the gels are heated, they liberate —water + alcohol" and convert into the nonporous impervious glasses.

In all methods of producing monolithic films, the chemical polymerization of Si and Al is accomplished by use of metal - organic compounds. One alkyl in the alkoxide molecule is converted to a hydroxyl group by partial hydrolysis and reacts with the alkyl groups of the other alkoxide.

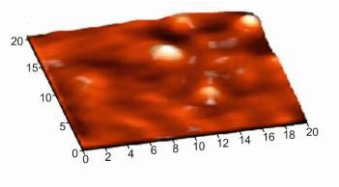
The resultant material contains Si and Al or Zr/Y bound via oxygen. Excess water in alcohol must be accompanied by the addition of acid to act as a peptizing agent, preventing the agglomeration of large particles. Once the clear liquid precursor is obtained, it can be applied to a substrate by any liquid application method, e.g., dipping, spraying, electrophoretic deposition, etc. Almost any substrate can be used, so long as the thickness of the coating is kept below a critical value, e.g., 1 μm, which varies slightly from one substrate to another. Porous α-Al_2O_3 tubes that are available commercially were coated with continuous films successfully. Detailed discussion of suitable substrates is not presented herein, since this paper is intended to report primarily on the newly developed Al_2O_3-SiO_2 formation technique (Fig. 2).



a



b



c

Fig. 2. ASM image of composite Al_2O_3-SiO_2 films; a) heat treatment at 350^0C; b) heat treatment at 450^0C; c) heat treatment at 550^0C

Heat treatment at 300°C converts the film to an essentially organic -free amorphous complex oxide; upon further heating above 400 °C crystals phase takes place [4].

CONCLUSION

Method of forming thin, continuous Al_2O_3-SiO_2 and ZrO_2-Y_2O_3 films from metal – organic - derived precursor solutions has been developed. The precursor is produced in a clear liquid state that contains oxide forms of Si and Al and Zr-Y in a chemically polymerized form. When applied to substrates, this liquid leaves a transparent, continuous film that converts to crystalline Al_2O_3-SiO_2 and ZrO_2-Y_2O_3 upon heating to 400 °C. A significant change of the film density takes place in the crystallization process, thus leading to strict requirements as to the thickness of the film which can survive crystallization. In this paper authors compared the crystallization behavior of oxide functional films derived from the same precursors, stressing the influence of experiment conditions and where it was possible to obtain the uniform amorphous or crystalline dielectric films.

REFERENCES

[1] Societal Implications of Nanoscience and Nanotechnology. Eds. M.C. Roco, W.S. Bainbridge. Dordrecht: Kluwer Acad. Publ., 2011, 384p.

[2] D. Vincenzi et al."Gas-sensing device implemented on a Micromac-hined membrane: A combination of thick-film and very large scale integrated technologies. J. Vac. Sci. Technol. B 2000, 18, 2441–2445.

[3] M.G.Blanchin et.al. –Structure and dielectric properties of films prepared by a sol-gel route", Journal of Sol-Gel Science Technology, 2008, 47, 165-172.

[4] N. Korobova "Thin Oxide Amorphous Functional Films from Metal Alkoxides", Proceeding of IX Intern. Conf. on the Structure of Non-crystalline Materials, July 11-15 2004, NY USA. O-45.

[5] N. Korobova, S.Timoshenkov, V. Vodopyanov –Design and fabrication of piezoelectric MEMS" //Proc. Int. Conf. –Micro- and Nanoelectronics – 2012". 1-5 Oct., 2012, Moscow- Zvenigorod, Russia. O3-28.

Optimization of Phosphorous Diffusion in Silicon Photocells

Maksym Klymov
Faculty of Electronics
NTUU KPI
Kyiv, Ukraine
maksym.klymov@iss.org.ua

***Abstract*— Today the main goals for photovoltaic industry are improvement of solar cell parameters and decreasing its cost. Author describes an experiment in manufacturing conditions that proves satisfactory results and achieves the both mentioned goals of photovoltaic industry.**

Keywords — low-temperature diffusion, photocell, solar cell, phosphorous diffusion, photovoltaics

I. INTRODUCTION

Since the first practical photovoltaic cell was developed in 1954 at Bell Laboratories [1] there have always been taken efforts to increase the efficiency of solar cell using minimum costs. One of the most obvious ways of increasing the solar cell efficiency is improving of the p-n junction parameters.

The experiment described in this article was dedicated to get better solar cell maximum generating power and efficiency than by the standard process. For reasons described in section III, decreasing of the diffusion process temperature was chosen as the most perspective way of getting the desired results.

Usually, the diffusion process held at temperatures in the range from 900 to 950°C during five to fifteen minutes can be considered representative [2]. However, use of lower temperatures also proved good results. It has been investigated [3] that satisfactory solar cell parameters can be obtained after pre-gettering at temperatures lower than 900°C.

Another target in the experiment was decreasing the solar cell cost (this is also one of today`s goals for photovoltaic industry [3]).

II. TECHNOLOGICAL PROCESS

A. Technology of photocells manufacturing

The basis for experiments was the serial technology of photocells manufacturing using monocrystalline high-resistivity (ρv = 3.8÷4 Ohm/cm) p-type silicon with diffusion n-type layer. Wafers (surface orientation <100>) are textured by anisotropic etching at temperature of 70-80°C in a weak, usually 2%, solution of NaOH or KOH with addition of isopropanol. This etching produces randomly distributed upside pyramids with height of 7-10 μm. Such surface texturing can reduce the optical reflection from more than 35 to less than 10%. In addition to the reduced reflection, front surface texturing ensures that light rays are coupled into the solar cell under an oblique angle, making it less probable to escape from the front surface after reflection at the rear. This light trapping effect results in an improvement in internal quantum efficiency in the wavelength range of 750-1000 nm. Fig. 1 shows the SEM micrograph of a randomly textured <100> oriented silicon surface.

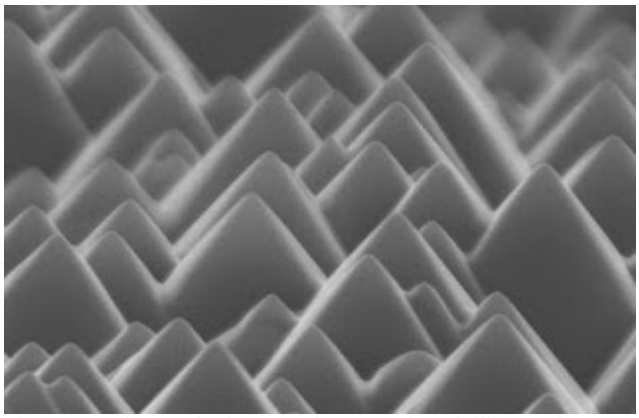


Fig. 1 SEM micrograph of a textured < 100> oriented silicon surface.

Diffusion is performed in a Tempress diffusion system (The Netherlands) by the open tube method. Gaseous diffusant ($POCl_3$) is supplied from the feeder and before entering the reactor is mixed with nitrogen and oxygen. In the reaction zone alloying element oxide is formed and on the surface of silicon wafers elementary impurity is allocated.

Diffusion is held in a single-band diffusion furnace with resistive heaters. The method of diffusion in an open tube makes it easy to manage the composition of the gas-vapor mixture, gas flow rate, and provides the required profile of the impurity distribution. The reproducibility of the diffusion parameters from wafer to wafer depends on the distribution of temperature in the working zone of the furnace, the number of wafers, their location relative to the gas flow, diffusant type and purity of the experiment [4]. Diffusion layer is 400-430 nm deep as shown in fig. 2.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

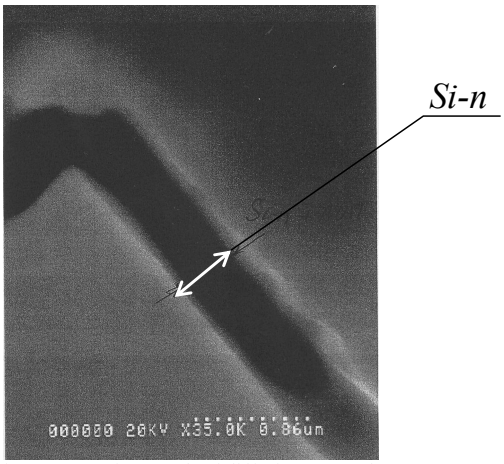


Fig.2. SEM micrograph of diffusion n-type layer

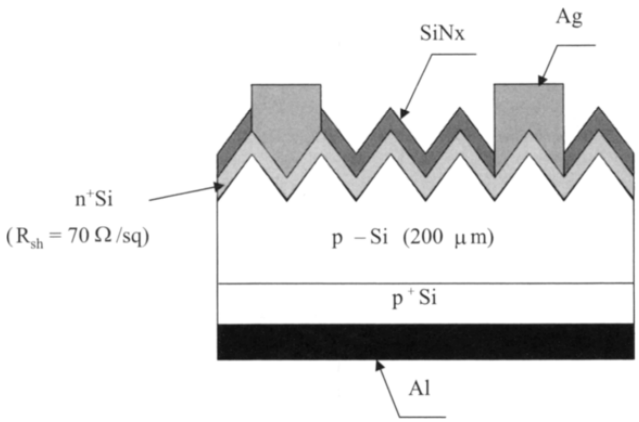


Fig. 3. Cross section of an advanced thin screen printed solar cell with shallow homogeneous emitter

The wafer surface is passivated with PECVD hydrogenated silicon nitride, due to its possibility of achieving optimal refraction index and excellent surface and bulk passivation properties at the same time It is impossible to talk about a unique SixNy, since a real variety of different silicon nitride layers with different physical properties can be deposited with the same machine. Finally, contact metallization is formed by buried front silver contacts and aluminum back surface field (BSF). Fig. 3 shows the structure of the photocell due to technology described.

B. Diffusion temperature profile

Temperature mode of diffusion system at regular technological process is shown in fig. 4.

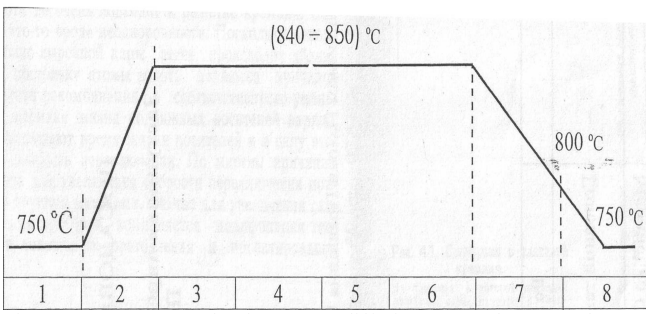


Fig. 4. Temperature mode of Tempress diffusion system due to regular process

The description of the fig. 4 and cycles duration are given in table I.

TABLE I. DIFFUSION SYSTEM TEMPERATURE PARAMETERS

Interval #	1	2	3	4	5	6	7	8
Temperature °C	750	-	840-850				800	750
Description	Loading	Heating	Stabilization	Gas inject	Phosphorous diffusion		Cooling down	Unloading
Duration, min	15	15	5	5	45±15		15	15

III. REASONS FOR DIFFUSION TEMPERATURE DECREASE

Whereas in IC manufacturing process emitter diffusion is held with temperatures over 1000°C, it is desirable to minimize the temperature profile of photocells manufacturing process. It can be explained by the number of factors that appear in photocell structure at high temperatures:

A. Silicon temperature resource

During the silicon treatment at the temperatures close up to its melting point (1414°C) some degradation processes appear in its crystalline gate; parasitic dopants in silicon begin to activate at such temperatures [5]; and there also are forming internal tensions that can cause the wafer destruction.

B. Dissimilarity of p-n junction

Silicon substrates used in commercial solar cell processes contain a nearsurface saw-damaged layer which has to be removed at the beginning of the process. The thickness of the damaged layer depends on the technique used in wafering of the ingot but in average it is from 7 to 40 μm. At the high diffusion temperature the dopant can penetrate into the mass of silicon especially through the cracks of the damaged layer, thus forming dissimilations of p-n junction.

C. Deep p-n junction

There is an important feature in photocells manufacturing to create a shallow p-n junction, aiming that charge carriers could recombinate in the junction. Nonetheless, high diffusion temperature leads to the deep p-n junction formation, which is a negative factor in photocells processing. It is investigated that a deep emitter and poor surface passivation lead to a voltage loss and collection losses at the short wavelengths of light [3].

D. 'Dead' layer in silicon

At the diffusion of a huge amount of phosphorous in silicon (about 5×10^{19} cm^{-3}) a small part of it remains on the wafer surface, forming so called 'dead' layer in silicon. There is a very small probability of charge carriers allocation in this layer. Because of the very huge concentration of electrons in this diffusion layer they start to recombinate with each other thus leading to charge carriers lifetime decrease.

Considering the fact that the 'dead' layer in silicon appears at the surface concentrations of $\approx10^{19}$ cm^{-3}, the process temperature should be chosen so that the maximal solubility of phosphorous in silicon was $\leq 10^{19}$ cm^{-3}.

IV. EXPERIMENTAL RESULTS

Here are given tables comprising experimental results due to those described above. All measurements have been held at automatic Pasan Cs-1200 tester.

A. Regular process

Due to make the understanding of experimental results importance here are given the measurement results of solar cells made by the regular process.

TABLE II. PARAMETERS OF REGULAR CELLS

#	Isc, A	Uoc, V	Eff, %	FF, %	Rser, Ohm	Rsht, Ohm	Pmax, W	Ir, mA
1	5.313	0.5949	15.55	76.12	0.0109	27	2.4	260
2	5.413	0.6191	16.34	76.16	0.0101	55	2.53	260
3	5.414	0.6105	16.12	75.45	0.0103	28	2.49	190
4	5.41	0.6098	16.26	76.24	0.0101	36	2.52	700

Note: I_{sc} – solar cell current
U_{oc} – generating voltage
Eff – solar cell efficiency
FF – fill factor of Volt-Ampere characteristic
R_{ser} – solar cell series resistance
R_{sht} – shunt resistance
P_{max} – maximal generating power
I_r – reverse current after metallization

B. Decreased diffusion temperature

The diffusion process temperature has been decreased to 840°C (instead of regular 845-850°C). Surface resistance of wafers used in the experiment was 3.8-4 Ohm/cm.

Regular process covers the use of wafers with surface resistance ρv = 1-1.5 Ohm/cm. Use of wafers with higher resistance ρv = 3.8÷4 Ohm/cm proved better results as such wafers are less doped, thus, have fewer parasitical dopants than the wafers with lower surface resistance.

TABLE III. RESULTS OF DECREASED DIFFUSION TEMPERATURE

Wafer #	1	2	3	4
Isc, A	5.523	5.527	5.526	5.566
Uoc, V	0.62	0.62	0.6206	0.62
Eff, %	17.01	17.06	17.03	17.09
FF, %	76.8	76.94	76.83	76.6
Rser, Ohm	0.0088	0.009	0.0089	0.009
Rsht, Ohm	100	100	51	100
Pmax, W	2.63	2.64	2.64	2.64
Ir, mA	140	130	50	170

Usually, solar cells on high-resistant silicon have lower reverse current and potentially should be more reliable because of lesser defectivity of initial silicon. Because of the same reason it is easier to get a “violet” cell [6] with a wide spectral band on a high-resistant silicon.

Results in table III are obviously better in comparison with parameters of regular cells (see table II). The best improvement can be seen in efficiency, power and shunt resistance. Better stability of parameters can be noticed in reverse current and generating voltage.

V. CONCLUSION

Obviously, the process technology used in the experiment is one of the most simple and cheap production methods developed up to date. Of course, this makes it uncompetitive in relation to the elements, which use multijunctions, thin-film metallization, innovative materials (based on III–V technologies like GaAs and GaInP, amorphous or thin-film silicon), etc.

However, the key advantage of a simpler process technology is its lower cost. In its price range, increasing the efficiency of the solar cell from 15.55-16.2% (see table II) to 17.0-17.1% (see table III) is a major competitive factor. Reference [3] mentions that the resulting cell efficiency (the process technology is equal to the one used in the experiment) is typically 12-15% and application of a antireflection coating gives up to 4% relative improvement. However, ref. [7] gives a value of 16,7% for the equal process technology laboratory cells efficiency. It seems clear that despite the possible variety in process implementation our experimental results of efficiency values (at least) are significant and competitive.

The use of high-resistivity silicon (which is more easily processed and is more convenient in the presence of the human factor), though more expensive compared to cheap impure low-resistivity silicon is justified, as it reduces the number of losses and greatly improves the quality of the finished product, thus, leading to cost lowering. The silicon used in the experiment with surface resistivity ρv = 3.8÷4 Ohm/cm is rather high-resistivity than low-resistivity, and shows good results corresponding with forecasts.

Assuming these results, decreasing of diffusion process temperature seems to be a perspective way. For more reliable results there should be held further experiments, stepwise reducing the temperature of the diffusion of up to 830°C, 820°C and so on, taking into account the gettering process and minority charge carriers lifetime.

REFERENCES

[1] K. Tsokos, Physics for the IB Diploma, Fifth edition, Cambridge: Cambridge University Press, 2008, p.423.

[2] A. Luque, S. Hedegus, Handbook of Photovoltaic Science and Engineering. Chichester: Wiley, 2003, pp.19, 273

[3] T. Markvart, L. Castaner, Solar cells: materials, manufacture and operation. Oxford: Elsevier, 2006, pp.92–105, 200-204.

[4] Z. Gotra. Tekhnologiya mikroelectronnykh ustroystv. Spravotchnik [Technology of microelectronic devices. Handbook]. Moscow: Radio and communication, 2001. (rus)

[5] V. Glazov, V. Zemskov, Fiziko-khimitcheskie osnovy legirovaniya poluprovodnikov [Physico-chemical basis of semiconductors doping]. Moscow: Science, 1967, p.142. (rus)

[6] S. Sze. Physics of semiconductor devices. New York: John Wiley & Sons. 1981

[7] J. Zhao, M. Green, Optimized Antireflection Coatings for High Efficiency Silicon Solar Cells. IEEE Trans. Electron Devices, Vol. 38, 1991, pp. 1925-1934

Mode Matching Technique for Tunable Shielded Cylindrical Metal-Dielectric Resonator

Kostiantyn Savin, Yuriy Prokopenko
Department of Physical and Biomedical Electronics
National Technical University of Ukraine "KPI"
Kyiv, Ukraine
savinkg@gmail.com, prok@phbme.ntu-kpi.kiev.ua

Guy A. E. Vandenbosch
TELEMIC group, ESAT
KU Leuven
Leuven, Belgium
guy.vandenbosch@esat.kuleuven.be

***Abstract*— A novel tunable cylindrical metal-dielectric resonator topology is described. It is designed with a dedicated Mode Matching technique. A Finite Element Method analysis of the resonator is also carried out. The computational results of both methods are compared and in good agreement.**

Keywords—mode matching technique; shielded metal-dielectric resonator; computational efficiency

I. INTRODUCTION

Wireless technologies have made great progress during the last 10 years. Since compact components are required, tunability has become one of the main trends in their design. Dielectric resonators (DRs) are widely used in modern radiofrequency equipment because of their small size and high quality (Q) factor [1–2]. The tunable DR with high-Q performance within the whole tuning range is highly demanded for different types of communication devices like filters, transceivers, receivers, phase shifters, etc.

Tunability and high-Q performance can be achieved by insertion a tunable perturber to the DR's design. There are tuning methods [3–4], in which MEMS switches are used to change the electromagnetic coupling between the DR and the perturber. Such designs are compact and maintain high Q-factor. The drawback is that they provide only step tuning. Using varactors instead of MEMS solves the problem [3] but leads to increased losses. Another way is moving the perturber. There are also composite DRs (CDRs). CDRs allow tuning by changing the distance(s) between its parts [5]. However, devices based on CDRs need very accurate positioning of these parts. This makes the CDR's performance vulnerable to vibration.

It was demonstrated in [6] that the electric field in the tunable air gap between a DR and a perturber has to be perpendicular to both components. Otherwise, large movements are needed for rather small frequency shifts.

A tunable DR with metal perturber, often called metal-dielectric resonator (MDR), was described in [7–8]. It matches the criterion demonstrated in [6]. These structures provide wide tuning range for small movements and high-Q performance. The disadvantage is that their performance can be calculated only with volume or surface discretization methods which demand a lot of computational resources.

The new MDR topology introduced in this paper does not show this disadvantage. It is such that a dedicated mode matching technique (MMT) can be used in order to solve the characteristic eigen problem describing the topology. Volume or surface discretization is not necessary. It is well-known that a MMT is computationally much more efficient than for example FEM.

II. THE SHIELDED MDR TOPOLOGY

The shielded MDR model is shown in Fig. 1. The model consists of the cylindrical-shaped DR radius R and height h between two parallel metal plates and has shielding metal wall of radius R_s. The bottom side of the DR is fixed to the lower plate and there is a tunable air gap of width d between the DR and the upper plate, see Fig. 1. Note that the upper plate can be made movable.

A cylindrical coordinate system is used. The origin is placed in the center of the DR bottom side. The longitudinal axis matches the cylinder axis and is directed from the lower to the upper plate.

The model volume is divided on two regions. The DR and the volume exactly above it form region 2. The rest is region 1.

III. THE SHIELDED MDR MMT PROBLEM FORMULATION

Electric and magnetic fields in both regions can be described in terms of electric and magnetic Hertz vectors:

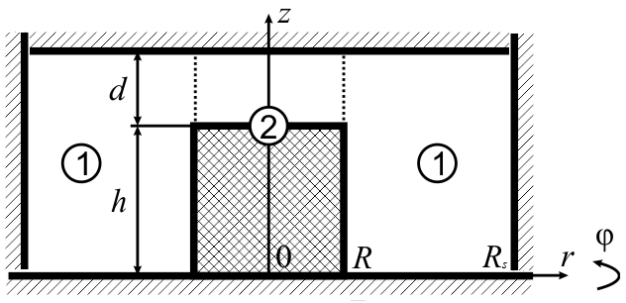


Fig. 1. The tunable shielded cylindrical MDR topology.

$$\mathbf{\Gamma}^{e(m)} = \left[0, 0, \Gamma^{e(m)}\right]^T.$$

Region 1 fields can be described with:

$$\Gamma_1^e = \sum_{i=0}^{\infty} a_{1i}^e \Re_{1in}(r)\Phi_n^e(\varphi) Z_{1i}^e(z),$$
$$\Gamma_1^m = \sum_{i=1}^{\infty} a_{1i}^m \Re_{1in}(r)\Phi_n^m(\varphi) Z_{1i}^m(z),$$

where $a_{1i}^{e(m)}$ are the eigenwaves amplitudes. $\Phi_n^{e(m)}(\varphi)$ are the functions which satisfy the conditions:

$$\frac{d\Phi_n^e(\varphi)}{d\varphi} = \pm n\Phi_n^m(\varphi),$$
$$\frac{d\Phi_n^m(\varphi)}{d\varphi} = \mp n\Phi_n^e(\varphi),$$
$$n \in N^0$$

$Z_{1i}^{e(m)}(z)$ are the MDR eigenfunctions which satisfy the boundary conditions in planes $z = 0, z = h + d$:

$$Z_{1i}^e(z) = \begin{cases} \sqrt{\frac{1}{h+d}}, i = 0, \\ \sqrt{\frac{2}{h+d}} \cos(\beta_{z1i} z), i \neq 0, \end{cases}$$
$$Z_{1i}^m(z) = \sqrt{\frac{2}{h+d}} \sin(\beta_{z1i} z),$$

where the $\beta_{z1i} = \frac{i\pi}{h+d}$ are the longitudinal wavenumbers. The functions $\Re_{1in}(r)$ are the solutions of the Bessel equations in region 1:

$$\Re_{1in}(r) = \begin{cases} \begin{cases} Y_n(\beta_{1i} R_s) J_n(\beta_{1i} r) - J_n(\beta_{1i} R_s) Y_n(\beta_{1i} r), \\ \mathrm{Re}\left(k^2 - \beta_{z1i}^2\right) \geq 0, \end{cases} \\ \begin{cases} K_n(\beta_{1i} R_s) I_n(\beta_{1i} r) - I_n(\beta_{1i} R_s) K_n(\beta_{1i} r), \\ \mathrm{Re}\left(k^2 - \beta_{z1i}^2\right) < 0. \end{cases} \end{cases}$$

$J_n(x)$ and $Y_n(x)$ are the Bessel functions of the first and the second kind of the n[th] order, but $I_n(x)$ and $K_n(x)$ are the modified Bessel functions of the first and the second kind of the n[th] order. β_{1i} is the transverse wavenumber in region 1:

$$\beta_{1i} = \begin{cases} \sqrt{k^2 - \beta_{z1i}^2}, \mathrm{Re}\left(k^2 - \beta_{z1i}^2\right) \geq 0, \\ \sqrt{\beta_{z1i}^2 - k^2}, \mathrm{Re}\left(k^2 - \beta_{z1i}^2\right) < 0. \end{cases}$$

Region 2 fields can be described with:

$$\Gamma_2^e = \sum_{i=0}^{\infty} a_{2i}^e \Re_{2in}^e(r)\Phi_n^e(\varphi) Z_{2i}^e(z),$$
$$\Gamma_2^m = \sum_{i=0}^{\infty} a_{2i}^m \Re_{2in}^m(r)\Phi_n^m(\varphi) Z_{2i}^m(z),$$

where $a_{2i}^{e(m)}$ are the eigenwaves amplitudes, $Z_{2i}^{e(m)}(z)$ are the MDR eigenfunctions which satisfy the boundary conditions in planes $z = 0, z = h, z = h + d$:

$$Z_{2i}^e(z) = \begin{cases} \frac{\cos(\beta_{z1i}^e z)}{N_i^e \varepsilon \cos(\beta_{z1i}^e h)}, 0 \leq z \leq h, \\ \frac{\cos(\beta_{z2i}^e (z - h - d))}{N_i^e \cos(\beta_{z2i}^e d)}, h \leq z \leq h + d, \end{cases}$$
$$Z_{2i}^m(z) = \begin{cases} \frac{\sin(\beta_{z1i}^m z)}{N_i^m \sin(\beta_{z1i}^e h)}, 0 \leq z \leq h, \\ \frac{\sin(\beta_{z2i}^m (h + d - z))}{N_i^m \sin(\beta_{z2i}^m d)}, h \leq z \leq h + d, \end{cases}$$
$$N_i^e = \sqrt{\frac{\frac{h}{2} + \frac{\sin(2\beta_{z1i}^e)}{4\beta_{z1i}^e}}{\varepsilon \cos^2(\beta_{z1i}^e h)} + \frac{\frac{d}{2} + \frac{\sin(2\beta_{z2i}^e)}{4\beta_{z2i}^e}}{\cos^2(\beta_{z2i}^e d)}},$$
$$N_i^m = \sqrt{\frac{\frac{h}{2} - \frac{\sin(2\beta_{z1i}^m)}{4\beta_{z1i}^m}}{\varepsilon \sin^2(\beta_{z1i}^m h)} + \frac{\frac{d}{2} - \frac{\sin(2\beta_{z2i}^m)}{4\beta_{z2i}^m}}{\sin^2(\beta_{z2i}^m d)}},$$

where ε is the DR permittivity, $\beta_{z1(2)i}^{e(m)}$ is the i[th] solution of the equations:

$$\frac{\beta_{z1i}^e}{\varepsilon} \tan\left(\beta_{z1i}^e h\right) + \beta_{z2i}^e \tan\left(\beta_{z2i}^e d\right) = 0, (\varepsilon - 1) k^2 = {\beta_{z1i}^e}^2 - {\beta_{z2i}^e}^2,$$
$$\beta_{z1i}^m \cot\left(\beta_{z1i}^m h\right) + \beta_{z2i}^m \cot\left(\beta_{z2i}^m d\right) = 0, (\varepsilon - 1) k^2 = {\beta_{z1i}^m}^2 - {\beta_{z2i}^m}^2.$$

The functions $\Re_{2in}^{e(m)}(r)$ are the solutions of the Bessel equations in region 2:

$$\Re_{2ni}^{e(m)}(r) = \begin{cases} J_n\left(\beta_{2i}^{e(m)} r\right), \mathrm{Re}\left(\varepsilon k^2 - \left(\beta_{z1i}^{e(m)}\right)^2\right) \geq 0, \\ I_n\left(\beta_{2i}^e r\right), \mathrm{Re}\left(\varepsilon k^2 - \left(\beta_{z1i}^{e(m)}\right)^2\right) < 0. \end{cases}$$

$\beta_{2i}^{e(m)}$ are the transverse wavenumbers in region 2:

$$\beta_{2i}^{e(m)} = \begin{cases} \sqrt{\varepsilon k^2 - \left(\beta_{z1i}^{e(m)}\right)^2}, \mathrm{Re}\left(\varepsilon k^2 - \left(\beta_{z1i}^{e(m)}\right)^2\right) \geq 0, \\ \sqrt{\left(\beta_{z1i}^{e(m)}\right)^2 - \varepsilon k^2}, \mathrm{Re}\left(\varepsilon k^2 - \left(\beta_{z1i}^{e(m)}\right)^2\right) < 0. \end{cases}$$

The boundary conditions at the interface between areas 1 and 2 require that tangential components of both electric and magnetic fields in both areas are equal.

Two new auxiliary functions are introduced in order to achieve the equality of z-components:

$$f^m(z) = j\frac{Z_0 H_z}{\Phi^m(\varphi)} = jZ_0\frac{\frac{\partial^2\Gamma_i^m}{\partial z^2} + \varepsilon_i(z)k^2\Gamma_i^m}{\Phi^m(\varphi)},$$
$$f^e(z) = \frac{E_z}{\Phi^e(\varphi)} = \frac{\frac{\partial^2\Gamma_i^e}{\partial z^2} + \varepsilon_i(z)k^2\Gamma_i^e}{\Phi^e(\varphi)}, i = 1,2, \quad (1)$$

where

$$\varepsilon_i(z) = \begin{cases} 1, & i = 1, \\ \varepsilon, & 0 \le z \le h \\ 1, & h \le z \le h+d \end{cases}, i = 2.$$

H_z and E_z are z-components of magnetic and electric field, and Z_0 is the characteristic impedance of free space. Taking the dot product of both parts of (1) and the MDR eigenfunctions provides the eigenwaves amplitudes expression in integral form of the functions $f^e(z)$ and $f^m(z)$.

The field φ-components equality at the surface $r = R$ provides the opportunity to form the set of Fredholm integral equations of the first kind:

$$\int_0^{h+d}\left(G_i^e(z,z')f^e(z') + G_i^m(z,z')f^m(z')\right)dz' = 0, i = 1,2,$$

where $G_i^{e(m)}(z,z')$ are the kernels of the equations (Green's functions of the problem) and expressed via eigenfunctions:

$$G_1^e(z,z') = \pm\frac{n}{R}\sum_{i=0}^{\infty}\left(\frac{Z_{1i}^e(z')\frac{dZ_{1i}^e(z)}{dz}}{\beta_{1i}^2} - \frac{\varepsilon_2(z')Z_{2i}^e(z')\frac{dZ_{2i}^e(z)}{dz}}{{\beta_{2i}^e}^2}\right),$$

$$G_1^m(z,z') = k\sum_{i=0}^{\infty}\left(\frac{Z_{1i}^m(z')Z_{1i}^m(z)\frac{d}{dR}\ln\left|\Re_{1ni}(R)\right|}{\beta_{1i}^2} - \frac{Z_{2i}^m(z')Z_{2i}^m(z)\frac{d}{dR}\ln\left|\Re_{2ni}^m(R)\right|}{{\beta_{2i}^m}^2}\right),$$

$$G_2^e(z,z') = \frac{k}{Z_0}\sum_{i=0}^{\infty}\left(\frac{Z_{1i}^e(z')Z_{1i}^e(z)\frac{d}{dR}\ln\left|\Re_{1ni}(R)\right|}{\beta_{1i}^2} - \frac{\varepsilon_2(z)\varepsilon_2(z')Z_{2i}^e(z')Z_{2i}^e(z)\frac{d}{dR}\ln\left|\Re_{2ni}^e(R)\right|}{{\beta_{2i}^e}^2}\right),$$

$$G_2^m(z,z') = \pm\frac{n}{RZ_0}\sum_{i=0}^{\infty}\left(\frac{Z_{1i}^m(z')\frac{dZ_{1i}^m(z)}{dz}}{\beta_{1i}^2} - \frac{Z_{2i}^m(z')\frac{dZ_{2i}^m(z)}{dz}}{{\beta_{2i}^m}^2}\right).$$

The Galerkin method is used to solve the system. The auxiliary functions are expanded in a complete set of orthogonal basis functions (coordinate basis) with weight functions $\rho_{\varphi^{e(m)}}(z)$ and unknown coefficients $\alpha_l^{e(m)}$:

$$f^{e(m)}(z) = \sum_{l=0}^{\infty}\alpha_l^{e(m)}\rho_{\varphi^{e(m)}}(z)\varphi_l^{e(m)}(z).$$

Then another complete set of orthogonal basis functions (projection basis) with weight functions $\rho_{\psi^{e(m)}}(z)$ is introduced:

$$\left\{\psi_n^{e(m)}(z)\right\}, n = \overline{0,\infty}.$$

The orthogonality of the system solution residual to all projection functions is demanded.

Thus, the problem is transformed in the homogenous system of linear equations:

$$\mathbf{AX} = 0, \quad (2)$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}^{e1} & \mathbf{A}^{m1} \\ \mathbf{A}^{e2} & \mathbf{A}^{m2} \end{bmatrix},$$

$$A_{nl}^{e1} = \int_0^{h+d}\int_0^{h+d} G_1^e(z,z')\rho_{\varphi^e}(z')\varphi_l^e(z')\rho_{\psi^m}(z)\psi_n^m(z)dz'dz,$$
$$A_{nl}^{m1} = \int_0^{h+d}\int_0^{h+d} G_1^m(z,z')\rho_{\varphi^m}(z')\varphi_l^m(z')\rho_{\psi^m}(z)\psi_n^m(z)dz'dz,$$
$$A_{nl}^{e2} = \int_0^{h+d}\int_0^{h+d} G_2^e(z,z')\rho_{\varphi^e}(z')\varphi_l^e(z')\rho_{\psi^e}(z)\psi_n^e(z)dz'dz,$$
$$A_{nl}^{m2} = \int_0^{h+d}\int_0^{h+d} G_2^m(z,z')\rho_{\varphi^e}(z')\varphi_l^m(z')\rho_{\psi^e}(z)\psi_n^e(z)dz'dz,$$
$$n = \overline{0,\infty}, l = \overline{0,\infty},$$

$$\mathbf{X} = \left[\alpha_0^e, \alpha_1^e, ..., \alpha_0^m, \alpha_1^m, ...\right]^T.$$

The homogenous linear system has nontrivial solution only if the determinant of the coefficient matrix is equal to zero:

$$\det(\mathbf{A}) = 0. \quad (3)$$

From this requirement it is possible to obtain the MDR eigenfrequencies. Then, the eigenfields can be achieved by solving the homogenous system of linear equations.

One of the main problems of solving the system (2) is the choice of coordinate and projection basis functions. Choosing the right basis will provide a rapid convergence of the system solution which will allow to reduce the order of the A matrix in (2). In general, the basis functions should be similar to the real distribution of the electromagnetic field in order to obtain the A matrix with high convergence and conditionality. There is a dielectric edge which is formed by boundaries of three dielectric mediums. According to [9] the field distribution near the edge has the singularity described by

$$E \sim r^{\nu-\frac{1}{2}},$$

where r is the distance to the edge, and

$$\nu = \frac{\arccos\left(\dfrac{(\varepsilon-1)^2}{2\varepsilon^2+4\varepsilon+1}-1\right)}{\pi}-\frac{1}{2}.$$

The basis functions should have a similar singularity and satisfy the boundary conditions at $z=0$ and $z=h+d$. The ultraspherical (Gegenbauer) polynomials $C_{2l}^{\nu}(\xi), C_{2l+1}^{\nu}(\xi)$ which are orthogonal with weights $\left(1-\xi^2\right)^{\nu-\frac{1}{2}}$ and complete on the interval [0, 1] match these conditions. That is why it is appropriate to represent functions $f^{e(m)}(z)$ as

$$f^{e(m)}(z)=\sum_{l=0}^{\infty}\begin{pmatrix}\alpha_{l1}^{e(m)}\varphi_{l1}^{e(m)}(z)\Theta(0,h)+\\ \alpha_{l2}^{e(m)}\varphi_{l2}^{e(m)}(z)\Theta(h,h+d)\end{pmatrix},$$

where $\Theta(\alpha,\gamma)=\Theta(z-\alpha)-\Theta(z-\gamma)$, $\Theta(z)$ is the Heaviside step function,

$$\varphi_{l1}^{e}(z)=\left(1-\left(\frac{z}{h}\right)^2\right)^{\nu-\frac{1}{2}}C_{2l+1}^{\nu}\left(\frac{z}{h}\right),$$

$$\varphi_{l2}^{e}(z)=\left(1-\left(\frac{h+d-z}{d}\right)^2\right)^{\nu-\frac{1}{2}}C_{2l+1}^{\nu}\left(\frac{h+d-z}{d}\right),$$

$$\varphi_{l1}^{m}(z)=\left(1-\left(\frac{z}{h}\right)^2\right)^{\nu-\frac{1}{2}}C_{2l}^{\nu}\left(\frac{z}{h}\right),$$

$$\varphi_{l2}^{e}(z)=\left(1-\left(\frac{h+d-z}{d}\right)^2\right)^{\nu-\frac{1}{2}}C_{2l}^{\nu}\left(\frac{h+d-z}{d}\right).$$

It is also appropriate to use Gegenbauer polynomials for the projection basis functions:

$$\psi_{n1}^{e}(z)=\varphi_{n1}^{e}(z)\Theta(0,h),$$
$$\psi_{n2}^{e}(z)=\varphi_{n2}^{e}(z)\Theta(h,h+d),$$
$$\psi_{n1}^{m}(z)=\varphi_{n1}^{m}(z)\Theta(0,h),$$
$$\psi_{n2}^{m}(z)=\varphi_{n2}^{m}(z)\Theta(h,h+d).$$

Using such basis functions transforms the A matrix and X vector in the form:

$$\mathbf{A}=\begin{bmatrix}\mathbf{A}_1^{e1} & 0 & \mathbf{A}_1^{m1} & 0\\ 0 & \mathbf{A}_2^{e1} & 0 & \mathbf{A}_2^{m1}\\ \mathbf{A}_1^{e2} & 0 & \mathbf{A}_1^{m2} & 0\\ 0 & \mathbf{A}_2^{e2} & 0 & \mathbf{A}_2^{m2}\end{bmatrix},$$

$$\mathbf{X}=\left[\alpha_{10}^{e},\alpha_{11}^{e},...,\alpha_{20}^{e},\alpha_{21}^{e},...,\alpha_{10}^{m},\alpha_{11}^{m},...,\alpha_{20}^{m},\alpha_{21}^{m},...\right]^{T}.$$

In general, (3) is the characteristic equation for the hybrid resonant modes HE_{nm} and EH_{mn}, where m is the root order. However in the case of $n=0$ $G_1^e(z,z')\equiv 0$, $G_2^m(z,z')\equiv 0$, and (3) reduces to two equations: for the TM_{0m} modes

$$\det\left(\mathbf{A}^{e2}\right)=0 \tag{4}$$

and for the TE_{0m} modes

$$\det\left(\mathbf{A}^{m1}\right)=0. \tag{5}$$

For a uniform resonator ($d=0$) the notation of resonant modes include the additional index l: HE_{nml}, EH_{mnl}, TM_{0ml} and TE_{0ml} [10], where l corresponds to the longitudinal wavenumber $\beta_z=\dfrac{l\pi}{h}$. For the uniform resonator the l index is integer. However for tunable MDR with $d\neq 0$ the index is not integer and is denote as δ. Consequently resonant modes of tunable MDR are: $HE_{nm\delta}$, $EH_{mn\delta}$, $TM_{0m\delta}$ and $TE_{0m\delta}$.

In the case of uniform resonator there is the additional TM mode: TM_{nm0} [10]. Let us denote analogue mode of tunable MDR as $TM_{nm\delta}$ regardless of the fact that the mode is actually $HE_{nm\delta}$ where $\delta<1$.

IV. Numerical Results and Discussion

The lowest resonant frequency of the shielded MDR is of the type $TM_{01\delta}$. This is fundamental mode of the tunable shielded MDR. Computed from (4) the normalized resonant wavenumber $\sqrt{\varepsilon}kR$ of this mode in the case of $R_s=2R$ and $R/h=2$ versus normalized air gap thickness d/h for various values of dielectric permittivity is presented in Fig. 2.

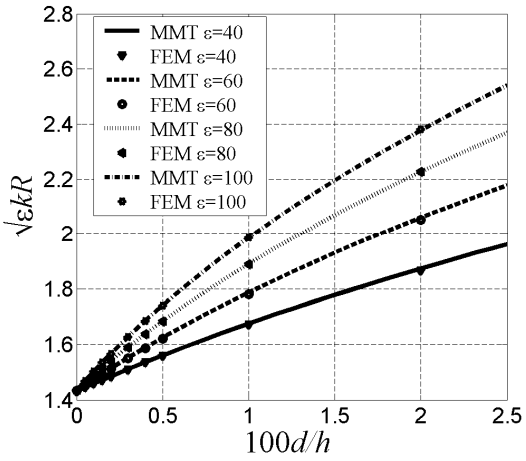


Fig. 2. Normalized resonant wavenumber of the $TM_{01\delta}$ mode versus normalized air gap thickness for an MDR with $R_s=2R$, $R/h=2$.

For comparison results computed by the finite elements method (FEM) are presented as well. There is a good agreement between the MMT and FEM results.

As it seen in Fig. 2 the resonant frequency of $TM_{01\delta}$ mode is very sensitive to variations of the air gap width d. Alterations of d in the order of only a few percents leads to resonant frequency shifts in the order of tens of percents. Such sensitivity gives an opportunity to apply MEMS or piezoelectric actuators for resonant frequency tuning. Analysis shows that the tuning efficiency increases in case R/h ratio and/or permittivity increase.

The $TM_{0n\delta}$, $TM_{nm\delta}$ and $HE_{nm\delta}$ modes demonstrate a similar behavior of the resonant frequencies in terms of the air gap thickness. However, the resonant frequencies of the $TE_{0m\delta}$ and $EH_{mn\delta}$ modes are less sensitive to a variation of the air gap width. As an example computed from (5) the dependence of the normalized resonant wavenumber of the $TE_{01\delta}$ mode in the case of $R_s = 2R$ and $\varepsilon = 80$ versus normalized air gap thickness d/h for various values of the R/h ratio is demonstrated in Fig. 3.

The fundamental reason of this different behavior is that for the $TM_{0n\delta}$, $TM_{nm\delta}$ and $HE_{nm\delta}$ modes the E_z component is dominant, while for the $TE_{0m\delta}$ and $EH_{mn\delta}$ modes this component is either absent or negligible in comparison with the H_z component. Thus, the $TM_{0n\delta}$, $TM_{nm\delta}$ and $HE_{nm\delta}$ modes are most suited for electromechanical tuning of the shielded MDR.

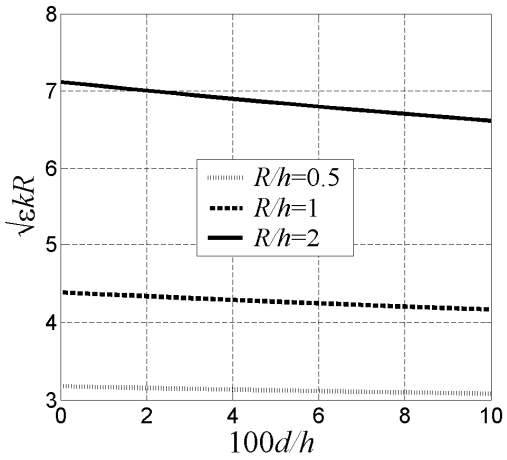


Fig. 3. Normalized resonant wavenumber of the $TE_{01\delta}$ mode versus normalized air gap thickness for MDR with $\varepsilon = 80$, $R_s = 2R$.

V. Conclusion

The mode matching technique in combination with a boundary integral equation was used to compute resonant frequencies of a tunable MDR. The combined technique reduces the electromagnetic field problem to a set of homogenous Fredholm integral equations of the first kind. This set was solved by the Galerkin method. It is illustrated that the developed method provides very similar result compared to a general purpose FEM solver, while requiring considerably reduced computational resources.

It is also shown that in order to achieve efficient electromechanical tuning of the shielded MDR resonant frequency, the resonant modes with a dominant electrical field component E_z have to be selected. These modes are the $TM_{0n\delta}$, $TM_{nm\delta}$ and $HE_{nm\delta}$ modes. The tuning efficiency grows with the DR radius to thickness ratio and increasing permittivity.

References

[1] L. Z. Thamae and Z. Wu, "Dielectric resonator-based multiple-input multipleoutput antennas and channel characteristic analysis**",** IET Microwaves, Antennas & Propagation, vol. 6, no. 9, pp. 1084–1089, June 2012.

[2] A. Petosa and S. Thirakoune, "Rectangular dielectric resonator antennas with enhanced gain", IEEE Transactions on Antennas and Propagation, vol. 59, no. 4, pp. 1385–1389, April 2011.

[3] F. Huang, S. Fouladi, and R. Mansour, "High-Q tunable dielectric resonator filters using MEMS technology", IEEE Transactions on Microwave Theory and Techniques, vol. 59, no. 12, pp. 3401–3409, December 2011.

[4] N. Klein, G. Panaitov and K. S. Daya, "Dielectric resonator with discrete electromechanical frequency tuning", IEEE Transactions on Microwave Theory and Techniques, vol. 53, no. 11, pp. 3371–3377, November 2005.

[5] T. Shen, K. A. Zaki and C. Wang, "Tunable dielectric resonators with dielectric tuning disks", IEEE Transactions on Microwave Theory and Techniques, vol. 48, no. 12, pp. 2439–2445, December 2000.

[6] Y. M. Poplavko, Y. V. Prokopenko, V. I. Molchanov and A. Dogan, "Frequency-tunable microwave dielectric resonator", IEEE Transactions on Microwave Theory and Techniques, vol. 49, no. 6, pp. 1020–1026, June 2001.

[7] F. Huang and R. R. Mansour, "Tunable compact dielectric resonator filters", Proceedings of the 39th European Microwave Conference (EuMC 2009), 2009, pp. 559–562.

[8] B. Pratsiuk, Y. Prokopenko and Y. Poplavko, "Tunable filters based on metal-dielectric resonators", Proceedings of the 18th International Conference on Microwave Radar and Wireless Communications (MIKON-2010), 2010, pp. 309–313.

[9] G. N. Brooke and M. Z. Kharadly, "Field behaviour near anisotropic and multidielectric edges", IEEE Transactions on Antennas and Propagation, vol. 25, no. 4, pp. 571–575, July 1977.

[10] Y. Kobayashi and S. Tanaka, "Resonant modes of a dielectric rod resonator short-circuited at both ends by parallel conducting plates", IEEE Transactions on Microwave Theory and Techniques, vol. 289, no. 10, pp. 1077–1085, October 1980.

Modeling of Velocity "Overshoot" in the Multivalley S emiconductors

Konstantin Bol
The National Technical University of Ukraine "Kyiv Polytechnic Institute"
Physical and Biomedical Electronics Department
Kyiv, Ukraine
airwave@mail.ua

Vladimir Moskalyuk
The National Technical University of Ukraine "Kyiv Polytechnic Institute"
Physical and Biomedical Electronics Department
Kyiv, Ukraine
mosk@phbme.ntu-kpi.kiev.ua

***Abstract* — Using the relaxation equations of energy balance, momentum and occupation of valleys, the possibility of modeling of the effect of drift velocity „overshoot" is shown in multivalley semiconductors on the example of aluminium nitride.**

Keywords — multivalley semiconductors; aluminium nitride; „overshoot" of drift velocity; times of momentum, energy relaxation, intervalley relaxation

I. INTRODUCTION

Developing of new semiconductor materials, such as threenitrids, having wide band gap and hence large intervalley distances, requires more detailed analysis of relaxation parameters especially for high-frequency and pulse-mode applications.

Basic relaxation parameters of electrons in semiconductors and hence their kinetic and dynamic parameters, such as mobility, pulse and high-frequency field response, etc., are mainly realized by numerical Monte Carlo method, which provides information redundance but does not allows to obtain analytical relation. It concerns to such known dynamic phenomenon as „overshoot" of drift velocity [1]. An attempt to apply relaxation equations for its modeling can not be recognized as fully successful [2]. In our opinion the reason of it consists in incorrect use of these equations, this article clarifies some of them.

II. RELAXATION EQUATIONS

Standard form of relaxation equations for momentum and energy, obtaining by averaging of Boltzmann kinetic equation, is

$$d\vec{p}/dt - e\vec{\mathcal{E}} = -\vec{p}/\tau_p, \qquad (1)$$

$$dE/dt - e\vec{v}\vec{\mathcal{E}} = -(E - E_0)/\tau_E . \qquad (2)$$

Here τ_p and τ_E are times of momentum and energy relaxation, which are determined from static field-velocity and field-temperature data, as it is done in [2], or from the analytical relation obtained from averaging of velocities of different dispersion mechanisms [3],[4].

For the multivalley semiconductors it is necessary to solve (1) and (2) for each valley, results must be averaged taking into account the electron distribution in valleys. The dynamics of occupation of valleys is determined by equation for a concentration in *i*-th valley:

$$dn_i/dt = -n_i/\tau_{nij} + n_j/\tau_{nji} \qquad (3)$$

and by analogical equation for *j*-th valley:

$$dn_j/dt = -n_j/\tau_{nji} + n_i/\tau_{nij}., \qquad (4)$$

where τ_{nij} and n_j/τ_{nji} are concentration times of intervalley relaxation that can be found through times of momentum relaxation with intervalley dispersion [4].

The second circumstance that must be taken into account when use the equation (2) and not taken into account in [2] is that the energy E includes both energy of chaotic heat motion E_t, and energy of the directed drift motion $E_d = m^*v^2/2$, which in the strong fields are commensurable quantities.

The drift motion energy is determined by electric field and relaxation rate, and is given by momentum balance equation (1), using multiplication by the drift velocity it is easy to obtain drift energy balance equation

$$dE_d/dt - e\vec{v}\vec{\mathcal{E}} = -2E_d/\tau_p . \qquad (5)$$

Subtracting (5) from (2), we will obtain heat energy balance equation

$$\frac{dE_t}{dt} = -\frac{E_t - E_0}{\tau_E} + \frac{E_d}{\tau_E}\left(\frac{2\tau_E}{\tau_p} - 1\right), \qquad (6)$$

that describes heating intensity of electrons. Here $E_0 = 3kT/2$ is equilibrium thermal energy, $2\tau_E/\tau_p \equiv \gamma$ is a parameter that determines efficiency of from drift energy to conversion thermal and depends on the rate of inelasticity of the collisions. For an electron temperature being a measure of energy of chaotic motion $E_t = 3kT_e/2$ and taking into account

that drift velocity $v = \mu\mathcal{E} = (e\tau_p/m_c^*)\mathcal{E}$, equation (6) takes the form

$$\frac{dT_e}{dt} = -\frac{T_e - T}{\tau_E} + \frac{m^*\mu^2\mathcal{E}^2}{3k\tau_E}(\gamma - 1) \qquad (7)$$

In a static field equation (7) turns into the equality

$$\frac{T_e}{T} - 1 = \frac{m^*\mu^2\mathcal{E}^2}{3k\,T}(\gamma - 1) \qquad (8)$$

that determines field-temperature data, i.e. the connection of electron temperature and electric field strength. It is difficult to express temperature as a function of intensity in explicit form due to complicated dependence of the momentum relaxation times on electron temperature.

It is necessary to notice that the reverse relaxation times in each valley are calculated by summation of reverse times for different dispersion mechanisms. Average valley mobility and ohmic effective mass are determined for a two-valley model from the relattions as follows:

$$\mu = b_i\mu_i + b_j\mu_j, \qquad (9)$$

$$1/m^* = b_i/m^*_i + b_j/m^*_j, \qquad (10)$$

where $b_i \equiv n_i/(n_i + n_j)$, $b_j = 1 - b_i$ are normalized concentrations (occupation of valleys). It is necessary to use (3) and (4) for their calculation.

Average momentum relaxation time in the valleys, is as usual determined from field-velocity data or from resultant mobility

$$\tau_p = \mu m^*/e, \qquad (11)$$

and average energy relaxation time, as opposed to [2], is determined from field-temperature data

$$\tau_E = \tau_p/2 + 3k(T_e - T)/2e\mu\mathcal{E}^2. \qquad (12)$$

The above equations were used for the modelling of „overshoot" effect in aluminium nitride the initial parameters of which were the same as in [2]. For time-dependent electrical field, changing from zero to a fixed value, temporal dependences of velocity, electron temperature, relaxation times, occupation of valleys were calculated. Temporal dependence of the velocity was converted in the spatial one by integration

$$x(t') = \int_0^{t'} v(t)\,dt\,. \qquad (13)$$

calculation results depicted in Fig. 1 show a good matching with the data from [5]. Fig. 2 shows energy and momentum relaxation times.

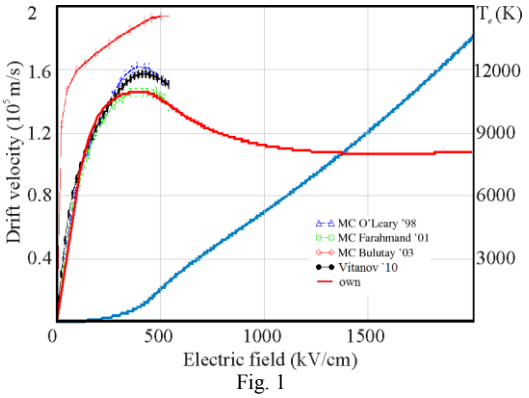


Fig. 1

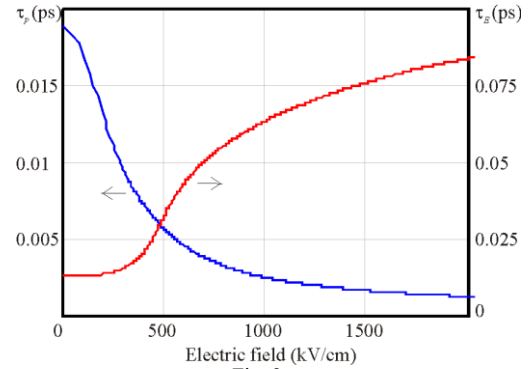


Fig. 2

Fig. 3 and Fig. 4 show relaxation times of momentum and energy as a function of time.

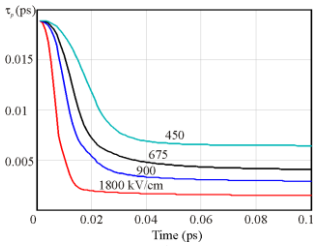


Fig. 3

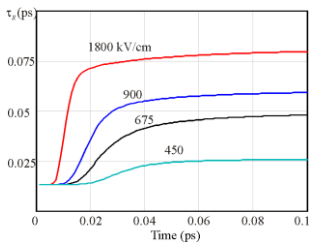


Fig. 4

Fig. 5 and Fig. 6 show an occupation of Γ and M-L valleys and electrons drift velocity as a function of time.

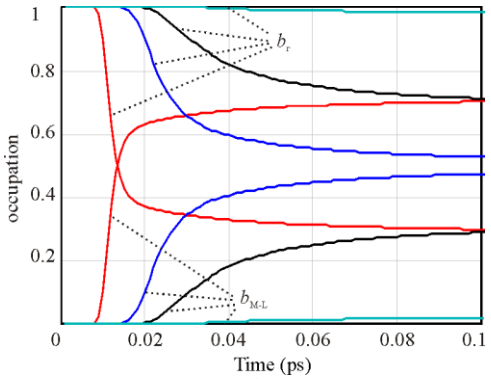


Fig. 5

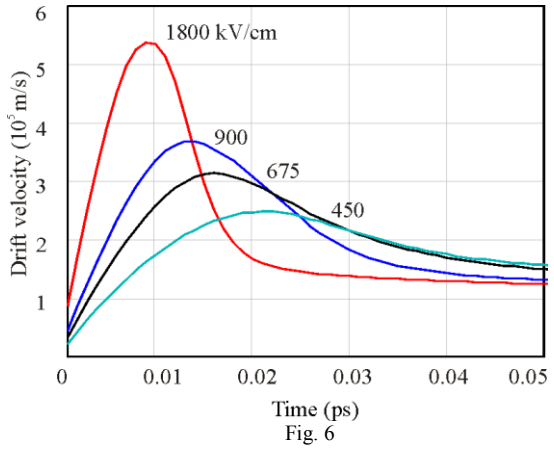


Fig. 6

Fig. 7 its a result of calculation electrons drift velocity as a function of distance, one can compare it with a Fig. 8, in which results are taken from [2] (Monte Carlo simulation).

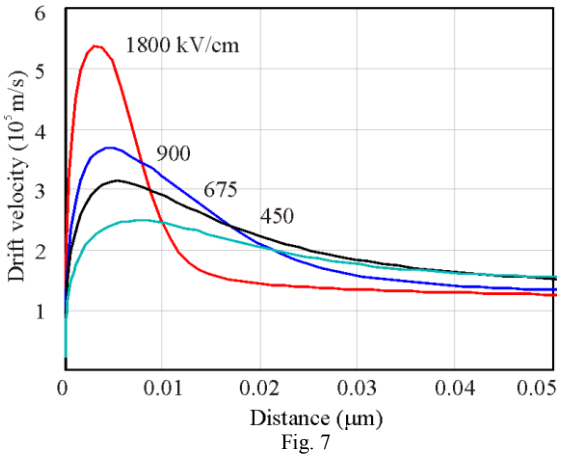


Fig. 7

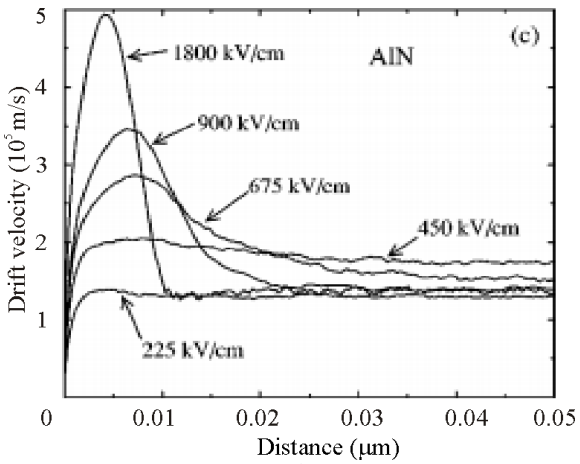


Fig. 8

The calculation results depicted in the figures show that on the initial section under the influence of power $e\mathcal{E}$ the electrons are accelerated and ballistic motion with linearly increasing velocity $(e\mathcal{E}/m^*)t$ and a parabolically increasing distance is observed. Such motion continues with time close to τ_p for a low field, and then as a result of the scattering a relaxation of the motion parameters occurs to the values typical for the drift motion in a given high field. A comparison of Fig. 7 and Fig. 8 shows much lesser difference of the results obtained by the Monte Carlo method and a method of relaxation equations than depicted in Fig. 5 in [2]. Obviously the reason of such difference is in the incorrect use by the authors of the equation (1) for the full energy [2].

CONCLUSIONS

Obtained results are in good accordance with the results of Monte Carlo method and indicate that the method of relaxation equations is suitable for the modeling of both static and dynamic properties of semiconductor materials, obtains greater temporal economy and sufficient information capability in comparison with the Monte Carlo method.

REFERENCES

[1]. Michael Shur, GaAs Devices and Circuits, Plenum Press, New York and London, 1987, p. 632.
[2]. Michael Shur, Transient electron transport in wurtzite GaN, InN, and AlN, American Institute of Physics, 1999.
[3]. K. Seeger, Semiconductor Physics, Wein : Springer - Verlag, 1973, p. 616.
[4]. V. A. Moskaliuk, Physics of Electron Processes. Dynamic Processes : Textbook, Kyiv: Politehnika, 2004, p. 180. (in Russian)
[5]. http://www.iue.tuwien.ac.at/phd/vitanov/node40.html

Receiver/Transmitter Input/Output Termination Resistance Calibration Method

Vazgen Melikyan, Abraham Balabanyan, Artak Hayrapetyan, Nazeli Melikyan
Department of Microelectronics
State Engineering University of Armenia
Yerevan, Armenia
vazgenm@synopsys.com, abrahamb@synopsys.com, artakh@synopsys.com, nazeli@synopsys.com

***Abstract*—A method of receiver and transmitter input/output resistance calibration, using external reference resistor is presented in this paper. The proposed architecture produces a calibration code corresponding to 50-Ohm PVT compensated termination impedance, which is needed to avoid reflections in transmission lines. The presented calibration mechanism can be used in the special input/output circuits of several standards such as Peripheral Component Interconnect (PCI), Universal Serial Bus (USB), Double Data Rate (DDR) etc.**

Keywords—receiver; transmitter; resistance calibration; PVT compensation; reflection

I. INTRODUCTION

In high speed systems and interfaces signal delays in PCBs (Printed Circuit Board) [1] long lines are an important factor and their effects cannot be ignored. When delays in interconnects are commensurable with signal transition times, they are considered transmission lines [1]. One of the main factors limiting the operating frequency of high speed devices are the I/Os and their transmission lines. Signal termination [2] is required to reduce voltage reflections. Excessive transmission line reflections can cause random logic false-triggering. As a result of the mentioned phenomena, the system may fail to function under some operating conditions such as high temperatures or over-voltages.

In general, the amplitude of the wave reflected at the end of a transmission line is determined by the reflection coefficient, ρ.

$$\rho = \frac{Z_{term} - Z_0}{Z_{term} + Z_0}, \quad (1)$$

where Z_0 is the characteristic impedance of transmission line, Z_{term} is the termination resistor in the far end of line.

When a voltage wave with amplitude V_{wave} hits the end of transition line, a wave with amplitude ρV_{wave} is reflected.

As it is seen from (1), in order to have no reflections (ρ=0), termination resistance and Z_0 must be matched (i.e., the same). If source impedance does not equal Z_0, then reflections occur at the near end of the line as well. Each end of the line has its own value of ρ.

In proposed configuration, the transmission line impedance is assumed 50-Ohm (Z_0=50 Ohm). To avoid reflections, 50-Ohm terminations in the near and far ends of the line are required.

The proposed calibration method provides relatively constant 50-Ohm termination impedance in the driver output and the receiver input for all PVT [3] conditions.

II. CALIBARATION CIRCUIT ARCHITECTURE

The structure of proposed Input/Output termination resistance calibration circuit is presented in Fig. 1.

Compensation circuit contains both analog and digital blocks. An external 200 Ohm high precision resistor (4 times the design target 50 Ohm impedance) and an integrated reference current generator [4] (which provides 3.25 mA DC current) are used for compensation. Current switch block consists of current mirrors with three branches (Fig. 2). Each of them is controlled by switches implemented with transmission gates (TG) [5]. The same reference current flows through all current mirror branches.

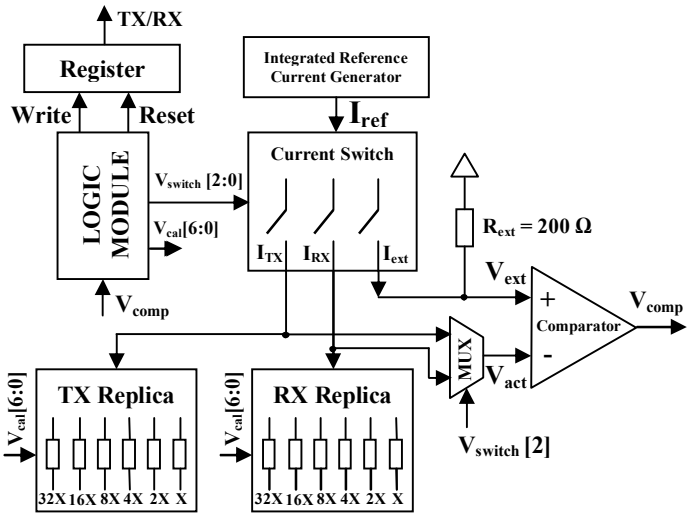


Fig. 1. I/O termination resistor calibration circuit structure.

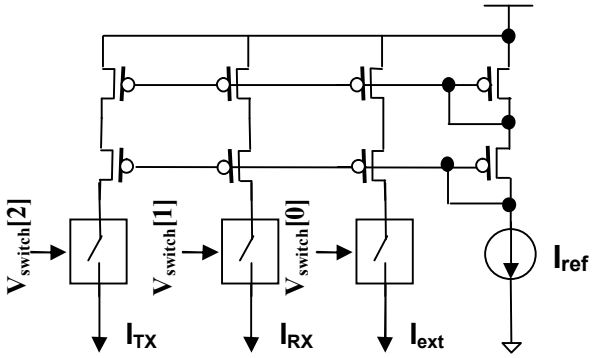


Fig. 2. Current switch circuit.

TX and RX replica blocks duplicate the actual transmitter and receiver input/output termination resistor circuits with one difference – total resistance of the replica blocks is four times higher than that of the actual ones.

In Fig. 3 a block diagram illustrating the structure of high speed links is presented. It consists of transmitter (TX) output resistance and output buffer, transmission line (which is typically 50 Ohm for PCBs) and receiver (RX) with its input termination resistance block. The ordinary TX is a buffer and RX is a differential amplifier.

In the process of calibration a binary code is obtained from replicas, which sets 50 Ohm on the TX output and the RX input. The replica resistances (200 Ohm) are four times higher than termination ones in real high speed links, so if the same binary code is applied to the pull up/pull down resistance section of TX and input resistance of the RX, 50 Ohm resulting impedance will be achieved. In this case there will not be reflection in transmission lines (Fig. 3).

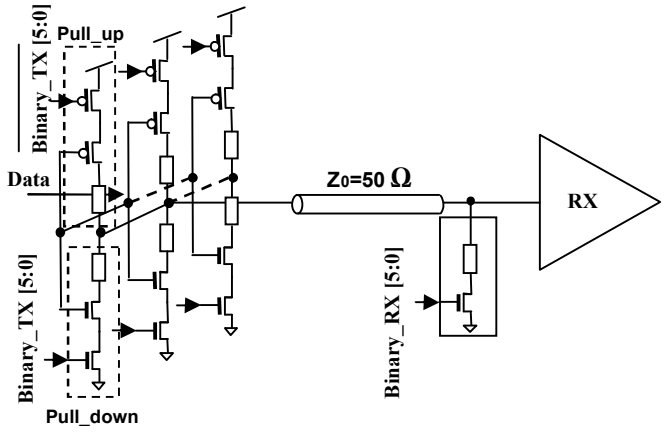


Fig. 3. Structure of high speed link.

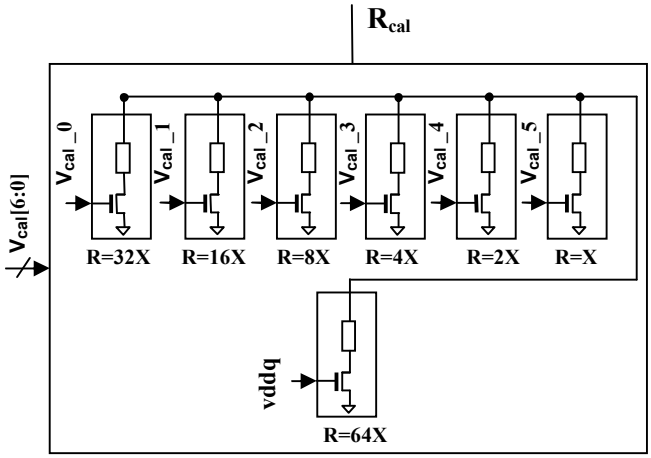


Fig. 4. RX replica circuit.

Each replica block contains parallel connected resistors and control transistors which can operate in cut off or triode regions. All resistors which are active in any situation define the cumulative resistance of RX or TX replica. Since these resistors are controlled by binary calibration signals (which come from logic block), each block has a resistance two times smaller than the next one. In proposed solution replica circuit contains seven blocks with resistors: for example, the first block has the smallest resistance and contains 32 *X* resistors connected in parallel, the sixth one contains only one resistor with the same *X* value and has the highest resistance. As it is mentioned above RX replica consists of resistances groups connected in parallel. The MSB [5] binary code connected to the smallest resistance group (with 32 resistances connected in parallel) in replica (R=32X) and LSB [5] to the highest resistance group of replica (R=X). To avoid excessively large resistances of R_{cal}, there is also a block of resistance which is always on and has the smallest resistance (R=64X). RX replica block is presented in Fig. 4.

Calibration circuit provides about 200 Ohm constant resistance in replica circuits for all PVT conditions. Usually resistance variations can reach up to ±50 Ohm so in typical case it is required to have at least 150-250 Ohm resistance range during calibration. This fact is used to choose a resistor value in each block of replica. In proposed solution the block which is always open has 300 Ohm resistance (64X=300 Ohm). So when binary code is zero (V_{cal}=000000) the total resistance is about 300 Ohm. When all blocks are open (V_{cal}=000000) the total resistance is about 150 Ohm. It is important to have about 200 Ohm replica resistance value in typical case for the middle input binary code. This condition is also provided in proposed architecture (when V_{cal}=000000, R_{cal}=200 Ohm).

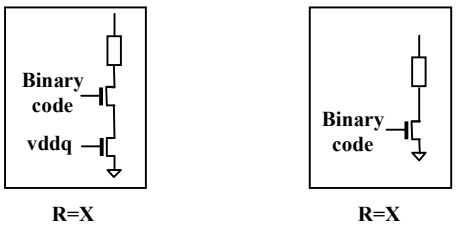


Fig. 5. a) TX replica single resistor block,
b) RX replica single resistor block.

The principle of calibration used for TX and RX replica circuits is the same, with a little difference: as in TX the output termination resistance block is a part of driver circuit, there should be two transistors on both pull up and pull down paths. The first transistor serves for data input and the second one is controlled by a binary code obtained from TX replica. For TX only pull down circuit is used in replica block. After calibration the same binary code will be given to TX pull up circuit. Actually each resistor in TX replica block is presented as two NMOS transistors and a resistor connected in series (Fig. 5a). The first transistor gets binary code while the second gets the highest potential and is always open (this transistor is added to have similar conditions with real TX output resistor). The single resistor block of RX replica is presented as NMOS transistor (with binary input) and a resistor connected in series (Fig. 5b).

In I/O termination resistor calibration circuit block diagram a logic module block, register, multiplexer and comparator are also used. Logic module provides all digital input signals such as calibration binary code (V_{cal}), switch signal (V_{switch}) controlling current switch block and multiplexer, write and reset signals to write data or reset register (where calibration final code is saved). Comparator compares calibration voltage of replica with external reference voltage. Output voltage of comparator (V_{comp}) is applied to logic module as an input.

III. Operation Principle and Compensation

For RX and TX input/output termination resistance calibration RX replica and TX replica circuits are used. As it is mentioned above replicas are added for simple and fast simulation. They have the same structure but four times bigger cumulative resistance than RX and TX termination resistance blocks.

The proposed principle of termination calibration method is presented as a block diagram in Fig. 6. The calibration starts from RX: A 3-bit signal, generated in logic module, is used to control the current switch block, turning it into RX mode. This means that no current flows through TX Replica (V_{switch}=011). So reference current generated by integrated reference current generator flows through RX replica and the external resistor. The V_{switch} signal controls multiplexer and enables path for actual voltage of RX replica to comparator. In the next phase logic module starts increasing 6-bit calibration code (V_{cal}) and changes RX replica resistance. Calibration precision depends on number of calibration bits (the higher is number of bits, the more accurate is calibration). During calibration replica resistance starts from the highest value (V_{cal}=000000). In these conditions in the replica resistance is about 300 Ohm. At the end of calibration the resistance of replica is minimum because all resistor blocks are open and connected in parallel (V_{cal}=111111). For such combination the replica resistance falls to about 150 Ohm. Number of calibration steps in proposed solution is 63.

$$N_{step}=2^n-1, \quad (2)$$

where N_{step} is the number of calibration steps and n is the bit number of input calibration signal.

The positive input of comparator has constant value V_{ext} which is the voltage drop of external 200 Ohm resistor (V_{ext}=200I_{ref}). Current in RX replica is constant as well and equal to I_{ref} but its actual resistance decreases during calibration. So the voltage drop of RX replica decreases which, in the meantime, is the voltage on negative comparator input. When the actual voltage of replica reaches the external reference voltage (V_{act}=V_{ext}) comparator output switches from logic "0" to logic "1" (V_{comp}=1). This is the case when the actual resistance of RX replica becomes equal to resistance of reference external resistor ($I_{ref}V_{act}$ = $I_{ref}V_{ref}$ => $R_{replica}$=200Ω). By the positive value of V_{comp} the logic module stops decreasing calibration code (V_{cal}) and stores fixed code in the register. This code is transferred from register to the actual input termination resistance block of receiver and provides its fixed 50 Ohm resistance (one quarter of RX replica). To fix the right calibration code it is important that the delay between calibration code bit changes be bigger than the summary of comparator and multiplexer delays.

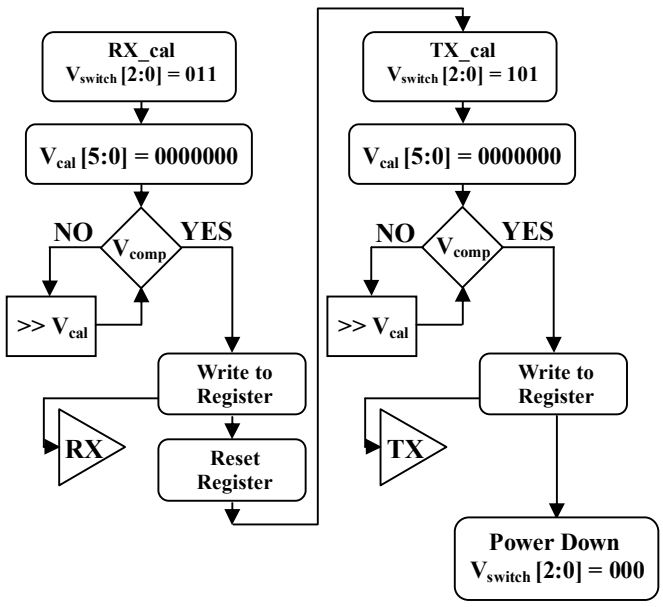


Fig. 6. Algorithm of calibration method.

After RX calibration the logic module resets register and starts calibration of transmitter. Logic module changes V_{switch} signal (V_{switch}=101) and switches current switch block to TX mode which means that no current flows through RX replica. In TX replica the same principle is implemented as it was in RX replica calibration. When the output of comparator switches the calibration binary code is being fixed, written to register and transferred to actual transmitter pull up/pull down sections providing their fixed 50 Ohm termination resistance value.

After receiver and transmitter termination resistance calibration is completed, the function of this circuit is finished and it does not affect the RX and TX operation. It should be noted that the compensation circuit uses large currents: two branches of current switch block are always open so they consume approximately 6.5mA current. To reduce power consumption the compensation circuit is switched to power down mode by logic module. It is performed by V_{switch} signal, which deactivates all TGs of three branches in current switch block (V_{switch}=000).

IV. Simulation Results

Simulations have been performed using circuit level simulator Hspice[6] for 65 PVT corners, including SS (slow-slow), TT (typical-typical), FF (fast-fast), SF (slow-fast), FS (fast-slow) with supply voltage and temperature variations.

Fig. 7 shows RX replica calibration results for TT(55^O) typical corner. It is seen that comparator switches to logic "1" when the actual voltage of RX replica equals to reference voltage and actual resistance of replica becomes about 200 Ohm (R_{act}=200.83 Ohm). Calibration is finished approximately in the middle of binary code (V_{cal}=100000 and N_i=32 for ideal case) at 33rd step of binary code, so in this case V_{cal}=100001 and N_i=33 (N_i is the step index of calibration binary code).

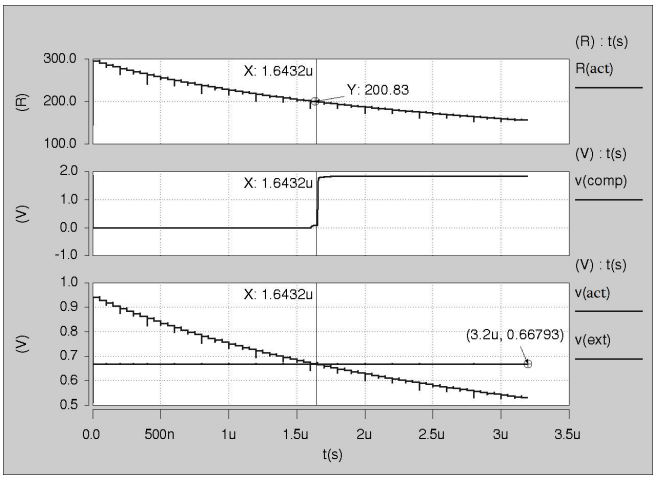


Fig. 7. RX replica calibration results for typical corner.

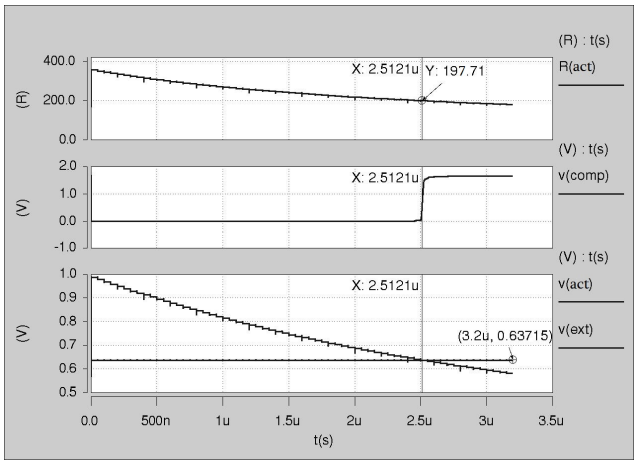


Fig. 8. RX replica calibration results for slow corner.

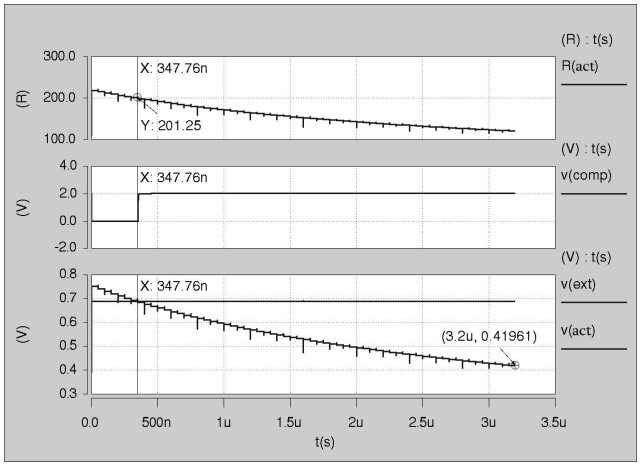


Fig. 9. RX replica calibration results for fast corner.

Fig. 8 and Fig. 9 show calibration simulation results for, respectively, SS(125^O) and FF(-40^O) main PVT corners. It is seen from waveforms that in slow case RX replica resistance reaches 200 Ohm for higher V_{cal} code values. In this case V_{cal}=110010, N_i=50 and R_{act}=197.71 Ohm (Fig. 8). In fast corner the calibration is finished more quickly and results for this case are the following: V_{cal}=000111, N_i=7 and R_{act}=201.25 Ohm (Fig. 9).

TX replica calibration waveforms are similar.

RX replica actual resistance waveforms (presented in Fig7-9) are obtained by dividing replica's actual voltage (V_{act}) to the current flowing through it (I_{RX}).

As it is seen from waveforms the voltage step during calibration is not ideally constant. Step changes in a range (4-10)mV. The reason of this, is not ideally constant behavior of the reference current.

An important parameter for calibration is the comparator sensitivity [5]. For calibration a comparator with sensitivity smaller than 1mV is designed. So calibration (4-10)mV voltage steps can easily be detected by comparator.

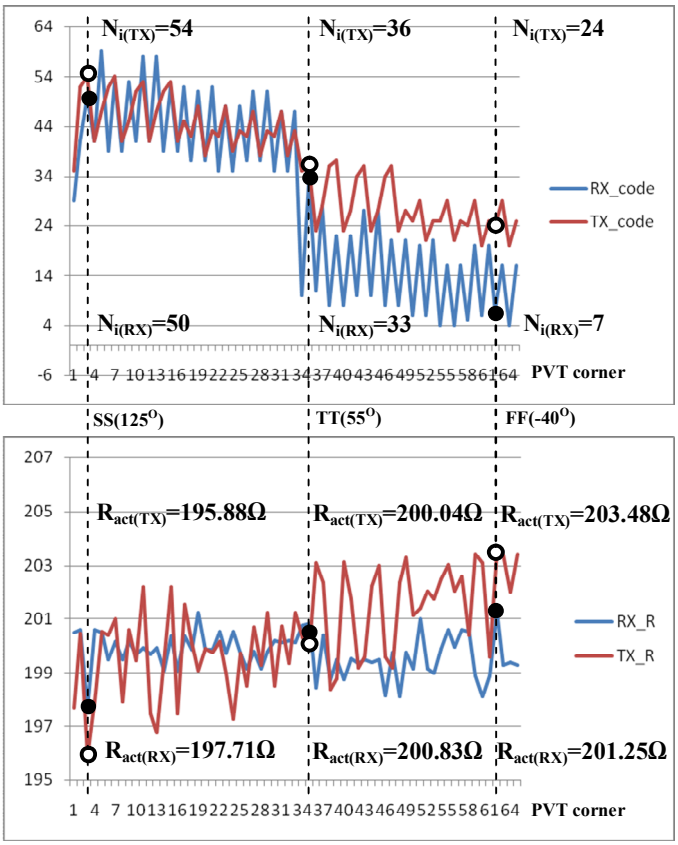


Fig. 10. Dependence of RX/TX replica calibration step index (N_i) and after calibration actual resistance (R_{act}) from 65 simulation PVT corners.

Fig. 10 shows the dependence of RX and TX replica calibration step index (N_i) and after calibration actual resistance (R_{act}) from 65 simulation PVT corners. Three main cases for RX replica discussed above are highlighted in the waveforms. From waveforms shown in Fig. 10 it is seen that the worst conditions of termination calibration both for TX and RX are two extreme cases: variation of resistance is maximum in SS(125^O) and FF(-40^O) PVT corners. Three main cases for TX and RX replica calibration results are circled in the waveforms (the circles showing RX results are filled, other ones are empty).

Simulation results of the three main corners for termination actual value using proposed calibration method are presented in Table I. The results show that termination resistance variation of transmitter is a little higher than variation of receiver (up to 2.1%). The result of this is the fact that TX replica resistor block is a little more complicated and contains one extra transistor which is always on during calibration.

The blocks which define main power consumption of the compensation circuit are RX/TX replicas, the external resistor (each of them uses approximately 3.25mA current and their supply voltage is 1.8V) and comparator (with average current approximately 400uA and supply voltage 1.8V). Logic block has small currents and 1V nominal supply voltage. Average power of compensation circuit is approximately 12.4mW and obtained from the following equation:

$$W_{avg}=V_{vddq}I_{vddq}+V_{vdd}I_{vdd}, \quad (3)$$

where W_{avg} is the average power of compensation circuit, V_{vddq} is the supply voltage of analog blocks, V_{vdd} is the supply voltage of digital blocks, I_{vddq} and I_{vdd}, respectively, are the currents flowing through analog and digital power sources.

TABLE I. SIMULATION RESULTS OF THE THREE MAIN CORNERS FOR TERMINATION ACTUAL VALUE.

Corner	RX Replica		TX Replica	
Unit	R_{act} (Ω)	*Var. (%)*	R_{act} (Ω)	*Var. (%)*
TT(55)	200.83	0.4	200.04	0.02
SS(125)	197.71	1.16	195.88	2.1
FF(-40)	201.25	0.63	203.48	1.74

V. CONCLUSIONS

A calibration method for receiver/transmitter input/output termination resistance is proposed and the respective circuit designed. Calibration system generates a digital code obtained from replica circuit calibration, both for driver and receiver. The proposed method provides 200 Ohm PVT compensated termination resistances with 2.1% maximal variation from the reference value. Without compensation variation of CMOS resistors can reach about ±20% due to PVT variations. Calibration is performed prior to RX/TX activation and has 3.2µs duration. Maximal power consumption of the designed circuit is approximately 12.4mW. Main disadvantage of the proposed solution is the necessity of an external 200 Ohm (±0.01%) reference resistor.

REFERENCES

[1] Reydezel Torres-Torres, Víctor H. Vega-González "An Approach for Quantifying the Conductor and Dielectric Losses in PCB Transmission Lines" IEEE 18th Conference on Electrical Performance of Electronic Packaging and Systems, Portland, OR, pp.235-238, 2009.

[2] Monica Zolog, Dan Pitică "Analysis of the Effect of Nonlinear Input/Output Characteristics of Digital Integrated Circuits on Signal Integrity" IEEE 31st International Spring Seminar on Electronics Technology, Budapest, pp.604-609, 2008.

[3] Gerald Esch Jr., Tom Chen "Design of CMOS IO Drivers with Less Sensitivity to Process, Voltage, and Temperature Variations" Second IEEE International Workshop on Electronic Design, Test and Applications, Perth, Australia, pp.312-317, 2004.

[4] Dong-Ok Han, Jeong-Hoon Kim, Nam-Heung Kim "Design of Bandgap Reference and Current Reference Generator with low Supply Voltage" IEEE 9th International Conference on Solid-State and Integrated-Circuit Technology, Beijing, pp.1733-1736, 2008.

[5] R. Jacob Baker "CMOS Circuit Design, Layout, and Simulation", Second Edition -2008 –1045p.

[6] Hspice Application Manual, Synopsys Inc. -2010., -196p.

Transmission Spectra of the Graphene-based Fibonacci Superlattice

A.M. Korol
National University for Food Technologies,
Kiev, Ukraine;
Laboratory on Quantum Theory
Linköping, Sweden.
korolam@ukr.net

V.M. Isai
National University for Food Technologies,
Kiev, Ukraine;
ivn@edu.nuft.ua.

***Abstract* — We consider the gapped graphene superlattice (SL) constructed in accordance with the Fibonacci rule. We propose to create the quasi-periodic modulation due to the difference in the values of the energy gap in different SL elements. It is shown that the effective splitting of the allowed bands and thereby forming a series of gaps is realized under the normal incidence of electrons on the SL as well as under oblique incidence. Energy spectra reveal periodical character on the whole energy scale.The splitting of allowed bands is subjected to the inflation Fibonacci rule.The gap associated with the new Dirac point is formed in every Fibonacci generation. The location of this gap is robust against the change in the SL period but at the same time it is sensitive to the ratio of barrier and well widths; also it is weakly dependent on values of the mass term in the Hamiltonian. Results otained allows for controlling the energetic spectra of the graphene-based superlattices and may be useful for operating of the nanoelectronic devices.**

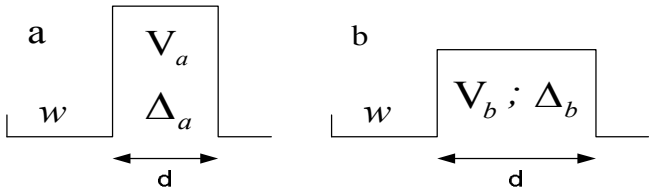
Keywords - graphene-based superlattice; Fibonacci rule; transmission spectra

I. INTRODUCTION

In recent years both graphene and graphene structures attracted much attention which is naturally explained by their non-trivial properties[1-3]. On the other hand, notable among the semiconductor structures including graphene ones are superlattices (SL) intermediate between periodic and disordered – quasi-periodic structures, e.g. Fibonacci, Thue-Morse and others. This is due to their unusual properties, such as self-similarity, fractal-like electronic spectrum etc. Attempts has already been started to study some kinds of superlattices based on graphene (Fibonacci and Thue-Morse [4-5]) which shed light on the behavior of Dirac chiral fermions in the quasi-periodic chains. Motivated by these circumstances, in this report we study the energetical spectra of the Fibonacci SL based on the monolayer gapped graphene. We propose to create the quasi-periodic modulation due to the difference in the values of the mass term Δ in the Hamiltonian in different elements of the superlattice. Because of the substantial progress in technology of graphene structures, in particular concerning the fabrication of the gapped graphene [6-9] we vary the value of Δ in wide range regardless of its nature which may be different [6-9].

II. METHOD OF EVALUATION

Consider the SL built of two elements "a" and "b".



Both elements contain the quantum well of width "w" and a potential barrier of width "d". The term Δ_a corresponds to the "a" element and the barrier height is denoted as V_a, for an element "b" we have Δ_b and V_b respectively. The SL is constructed according to the Fibonacci inflation rule. Thus for the fifth Fibonacci generation one can write : S_5=abaababa. The value of the transmission coefficient T(E), E being the electron energy, through the SL built for a certain generation is determined by the period of this generation (sequence). Energy intervals in which the condition T(E)=1 holds form the allowed bands in energy spectrum, gaps are associated with values T<<1. The transmission coefficient can be evaluated by the transfer matrix method, by expressing T either in terms of the Green functions or the eigenfunctions. The latter are the solutions of the Dirac-like equation

$$\left[v_F(\vec{\sigma},\vec{p}) + m^2 \cdot {v_F}^2 \cdot \sigma_z + V(x) \cdot \hat{I} \right] \cdot \Psi = E \cdot \Psi \qquad (1)$$

where $v_F=10^6$ $^m/_s$, $p=(p_x,p_y)$ the momentum operator $\sigma=(\sigma_x,\sigma_y),\sigma_x,\sigma_y,\sigma_z$ - Pauli matrices for the pseudospin, V(x) the external potential, which depends only on coordinate x, I – two-dimensional unit matrix, the mass term is denoted by the symbol Δ as adopted in the literature (units $c=\hbar=e=v_F=1$ are adopted). The function Ψ is a two-component pseudospinor $\Psi=[\Psi_A, \Psi_B]^T$, Ψ_A, Ψ_B are the envelope functions for the graphene sublattices A and B. Suppose that the potential consists of repetitive rectangular barriers along the x-axis and $V_j(x)$= const within each j-th barrier. If we present solutions for the eigenfunctions $\Psi_{A,B}$ as a sum of plane waves traveling in the forward and backward directions of the x-axis then we obtain:

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

$$\Psi(x)=\left[a_j\cdot e^{i\cdot q_j\cdot x}\begin{pmatrix}1\\ g_j^{+}\end{pmatrix}+b_j\cdot e^{-i\cdot q_j\cdot x}\begin{pmatrix}1\\ g_j^{-}\end{pmatrix}\right] \quad (2)$$

Transfer matrix connecting the wave functions at the points x and x + Δx is found in several studies (see e.g.[6]) and has the form:

$$M_j=\frac{1}{\cos\theta_j}\begin{pmatrix}\cos(q_j\cdot\Delta x-\theta_j) & i\cdot z_j^{-1}\sin(q_j\cdot\Delta x)\\ i\cdot z_j\cdot\sin(q_j\cdot\Delta x) & \cos(q_j\cdot\Delta x+\theta_j)\end{pmatrix} \quad (3)$$

where $z_j=\frac{s_{j-}}{k_j}$, $\theta_j=\arcsin\left(\frac{k_y}{k_j}\right)$,

k_y - is y-component of the wave number,

$k_j=\mathrm{sign}(s_{j+})[(E-V_j)^2-\Delta^2]^{1/2}$, $s_{\pm}\equiv E-V(x)\pm\Delta$,

$q_j=\mathrm{sign}(s_{j+})\sqrt{k_j^2-k_y^2}$ if $k_j^2>k_y^2$

and $q_j=i\cdot\sqrt{k_y^2-k_j^2}$ otherwise,

$g_j^{\pm}=\frac{\pm q_j+i\cdot k_y}{k_y}$,

the top line in (2) pertains to the sublattice A, the lower one – to the sublattice B.

The coefficient of the transmission of quasi–particles through the lattice $T=|t|^2$,

$$t=\frac{2\cos\theta_0}{R_{22}e^{-i\cdot\theta_0}+R_{11}e^{i\cdot\theta_0}-R_{12}-R_{21}} \quad (4)$$

where θ_o is the angle of incidence, and the matrix R is expressed as a product of matrices M_j: $R=\prod_{j=1}^{N}M_j$,

N - the number of elements in the SL.

III. Results and Discussion

The trace map for 6 initial Fibonacci sequences is depicted in Fig.1 for the case where we have the gapped graphene in elements "a" and the gapless one in elements "b".

First of all it is noteworthy that the quasi-periodic modulation by setting different values of Δ in different SL elements used in this study results in the efficient splitting of the allowed energy bands and thus in the formation of a number of gaps. And this is realized at normal incidence of the electrons on the lattice.

Fig.2 exhibits the tunneling spectrum, i.e. the dependence of the transmission coefficient T on energy E for the 4-th Fibonacci generation under normal incidence of the electron wave on the lattice.

We see that the structure of some parts of the spectra is repeated periodically throughout the whole energy scale; one of these fragments of the spectrum can be considered as its period. The characteristic features of the period – the number of allowed (forbidden) bands and their widths change so that with E increasing the gap's widths on average decrease. The natural result of this reduction is that the transmission coefficient asymptotically approaches to unity in the sufficiently far over-barrier region.

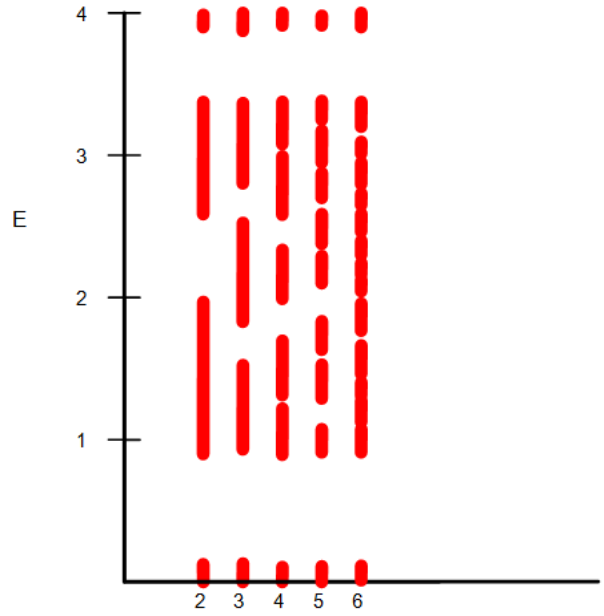


Fig. 1. Trace map for the Fibonacci SL, k_y =0, d=w=0,5, $V_a=V_b=1$, $\Delta_a=1$, $\Delta_b=0$.

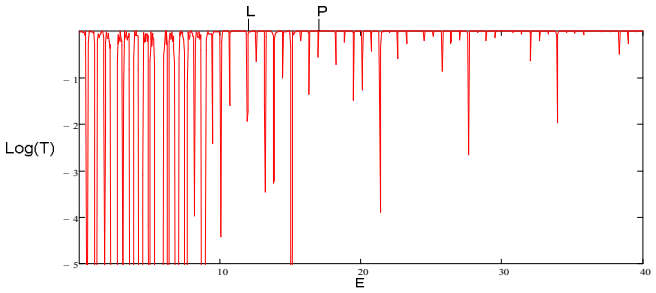


Fig. 2. Dependence of the transmission coefficient T on electrons energy E for the 4-th Fibonacci generation; values of the parameters: k_y =0, d=w=0,5, $V_a=V_b=5$, $\Delta_a=1$, $\Delta_b=0$.

The same spectrum in the energy range [0, 6] is shown in Fig3.

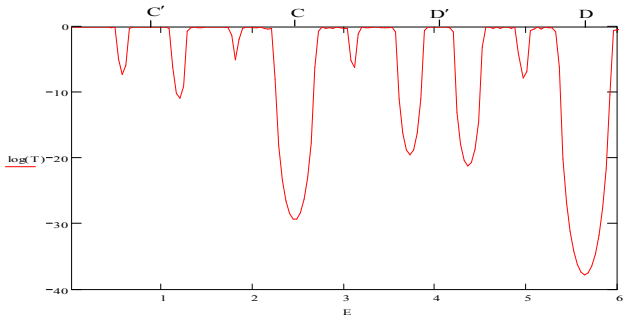


Fig. 3. Tunneling spectrum for the 4-th Fibonacci generation in the energy range [0,6]; values of the parameters are the same as those in Figure 2.

Spectra similar to that shown in Fig.2 for the 4-th generation are observed also for other Fibonacci sequences.

The number of bands and the width of each of them depend on one hand on the SL parameters and on the other hand on the number of the generation.

As follows from the calculations, the number of allowed (forbidden) bands contained in one period, in particular in the interval CD in Fig.3, corresponds to the Fibonacci sequence and this number is subjected to the Fibonacci inflation rule: $Z_n=Z_{n-1}+Z_{n-2}$, where n is the number of the Fibonacci generation (see Fig.1). Note that this rule is applied not only to the CD period but to larger periods as well (e.g. LP in Fig.2); every new super-period has its own number of bands.

As we can see in Fig.1,2, 3 at a certain energy all Fibonacci generations form the gap associated with the new Dirac point – "new Dirac point gap" [6]. Its location is almost unchanged in different Fibonacci sequences. A typical feature of the new Dirac point gap is that it is independent of the lattice period (d + w) and at the same time is sensitive to the ratio d/w. This is proved in particular by Fig.4 which shows the dependence of T on E for the third Fibonacci generation for different values of w and d.

Once again we draw attention to the fact that the band's splitting (thereby forming a series of gaps) is realized even in the case of the normal incidence of electrons on the lattice surface.

IV. Conclusions

In summary, we have demonstrated that an effective quasiperiodic modulation can be achieved due to difference in mass term in different elements of the superlattice based on the monolayer graphene. The number of gaps within a period of the energy spectrum is subjected to the Fibonacci inflation rule. The superlattice Dirac gap is observed in the spectrum and its features are analyzed.

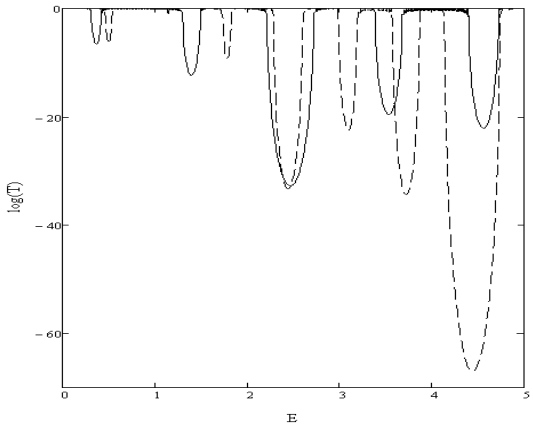


Fig. 4. Spectrum of the third Fibonacci sequence for different SL periods, d=w=0,5 (solid line); d=w=0,8 (dashed line); values of other parameters are the same as those in Figure 2.

References

[1] A.K.Geim, K.S.Novoselov, Nat. Materials 6,183 (2007).

[2] A.N.Castro Neto, F.Guinea, N.M.R.Peres et al., Rev. Mod. Phys. 81,109 (2009).

[3] J.M.Pereira, F.M.Peeters, A.Chaves et al., Semicond. Science Technology 25,033002 (2010).

[4] P.Zhao, X.Chen, Appl.Phys. lett. 99, 182108 (2011).

[5] T.Ma, C.Liang,L.Wang et al., Appl.Phys. lett. 100, 252402 (2012).

[6] J.C.Meyer, C.O.Girit, M.F.Crommie et al., Appl. Phys. Lett. 92,123110, (2008).

[7] P.W.Sutter, J.Flege, E.A.Sutter, Nat. Materials. 7, 406 (2008).

[8] J.Coraux, A.T.N'Diaye, C.Busse et al., Nano Lett. 8,565 (2008).

[9] M.Y.Han, B.Ozyilmaz, Y.Zhang et al., Phys. Rev. Lett. 98, 206805 (2007).

Optical and Luminescence Effects in Cu, Au-doped Cadmium Iodide layered Nanostructures

Mykola Rudka
Lviv Polytechnic National University Department of Physics
Lviv, Ukraine
mrud@polynet.lviv.ua

Vladimir Antonyuk, Natalia Stetsyk
Ivan Franko National University of Lviv Department of Physical and Biomedical Electronics
Lviv, Ukraine
stetsyck_natali@ukr.net

Abstract — **The crystals CdJ_2 activated by gold and copper were investigated. It was shown that activated by gold and copper samples CdJ_2:Au,Cu are optically resistant to high-energy electromagnetic radiation. Also the Cu-doping effects in CdJ_2 layered nanocrystals were studied experimentally. We used a photostimulated second harmonic generation (PSHG) in order to investigate the nanoconfined effects. It was found that the PSHG increased with increasing Cu-doping within a low-density limit.**

Keywords: cadmium iodide; donor-acceptor luminescence; photochromic impurities; photoinduced second harmonic generation (PSHG); donor-acceptor pair (DAP).

I. INTRODUCTION

The layered nanocrystals of cadmium iodide belong to the class $A^{II}B^{VII}_2$ and have a strong anisotropy of the internal crystal interactions. Within the limits of the structural sandwich package J-Cd-J there are strong covalently-ion interactions between the components, with dominate of covalent component, and there are weak van der Waals forces between layers. The fundamental absorption edge of the CdJ_2 crystals at room temperature is 3.2 eV and from decrease in temperature to 77 K shifts to 3.48 eV [1], [2].

It is known [3] that the dominant luminescence centers in pure crystals CdJ_2 are strongly oriented donor-acceptor (DA) complexes of point structure defects in the cation vacancy and interstitial cadmium. The spectral position and duration of donor-acceptor luminescence are determined using of energetical parameters of such luminosity centers and the rate of localization upon them the nonequilibrium genetic electron-hole pairs.

The copper-activated cadmium iodide crystals are efficient photochromic materials. The photoelecrtical properties of CdJ_2:Cu were studied before in [1], [2], [4]. The photochemical transformation in those systems have been shown to be related mainly to changes in the charge state of activator centers.

Earlier investigations performed in the pure CdJ_2 single crystals demonstrated higher-order optical nonlinearities. The photoinduced second harmonic generation (PSHG) investigations of magnetic field-stimulated ferroelectricity in the CdJ_2–Cu systems have also been reported [6].

However, these measurements were rather preliminary and require additional exploration. More recent reports on these systems are also available [7], [8]. In this work the optical luminescent properties of CdJ_2:Cu, CdJ_2:Cu,Au layered nanocrystals have been studied.

II. EXPERIMENTAL TECHNIQUE

Band structure calculations for the CdJ_2 crystals have also shown large anisotropy of the space charge density distribution causing high anisotropy of the corresponding optical spectra. The anisotropic behavior of the CdJ_2 crystals favors the local noncentrosymmetry that allows the PSHG investigations due to increased local hyperpolarizability [7], [8].

In the present work we experimentally studied the Cu-doping effects on nonlinear optical effects in the CdJ_2 nanocrystals. The principles of the effects consist in the formation under the photoexcited filed of the photoinduced local polar states due to optically stimulated anharmonic electronphonon interactions.

Investigated samples were the undoped as well as the Cu-doped and Au-doped CdJ_2 crystals of thickness varied within 0.8–10 nm. These nanocrystals were grown using the standard Bridgman–Stockbarger technique in sealed quartz ampoules (CdJ_2:Cu – 0,5 % by mass, CdJ_2:Cu,Au – 0,4 % by mass, CdJ_2 – 0,15 % by mass). Their crystalline structures were monitored using an x-ray diffractometer. The investigations were performed at liquid nitrogen temperature because at this temperature the effect should be the enhanced. We used a neodymium-doped yttrium aluminum garnet laser. The laser generates picosecond pulses having a repetition rate of 80 Hz. The output PSHG (λ=530 nm) and the fundamental (λ=1064 nm) signals were spectrally separated by a grating monochromator.

Detection of the doubled-frequency output PSHG signal was performed by a photomultiplier equipped with an electronic boxcar integrator (EBI) for registration of the output. A schema of the experimental setup is shown in Fig.1

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

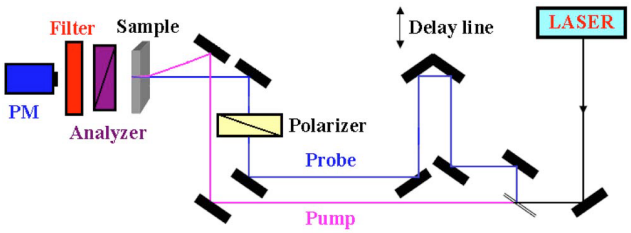


Fig. 1. A schema of the PSHG experimental setup

III. RESULTS

Activation CdJ_2 with using by impurities allows considerably to increase the intensity, and to change the spectrum of their fluorescent glow. The presence of impurities causes the increasing concentration both their own complexes, and the generate own-impurity or pure-impurity defects. It is shown in the absorption spectra, especially in photoluminescence excitation spectra of crystals [4], [5]. Impurities of copper and gold in cadmium iodide crystals are photochromic (Fig. 2). On the one hand, they are increasing the intensity of luminescence of cadmium iodide, on the other hand they make material CdJ_2:Me more sensitive to the high-energy electromagnetic radiation. It depending on the dose of radiation causing varying degrees of reduction of the afterglow.

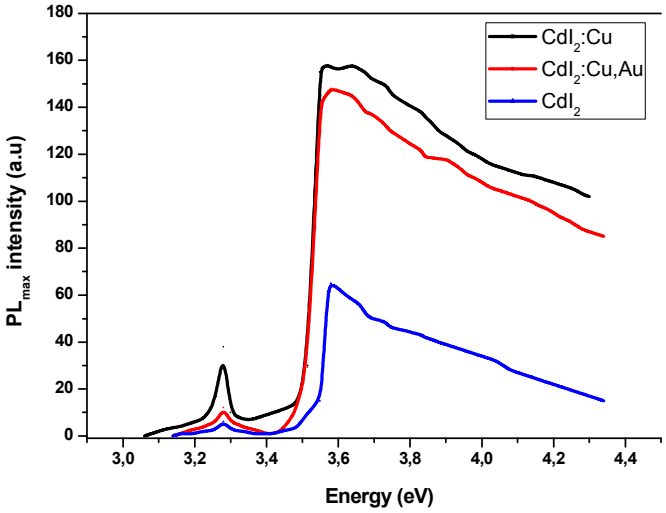


Fig 2. Excitation and luminescence spectra of cadmium iodide crystals doped by Cu and Au impurities.

Thus, much better in terms of improving the luminescent material output must be activation of crystals with using by simultaneously two activators. On one hand, they would have properties which are suitable well for the selected crystal lattice, and on the other hand, one of them performs functions of donor in this matrix and the other one performs the functions of acceptor there.

This behavior of the crystals with two activators confirms the hypothesis about recombination of nonequilibrium charge carriers which are created due to high-energetical irradiation, through donor-acceptor pair (DAP) defects. This, in turn, gives us reason to confirm that a purposeful creation of complexes defects type of DAP in the crystal lattice allows not randomly but predictably to look for prospective luminophores.

The dependence of the PSHG signal on pumping power density was measured. Results for the crystal thickness of 0.8 nm are shown in Fig. 3(a). The PSHG signal increases with increasing power density reaching a maximum and then in decreases to the background. However, a significant enhancement of the PSHG response occurs in the nanocrystals.

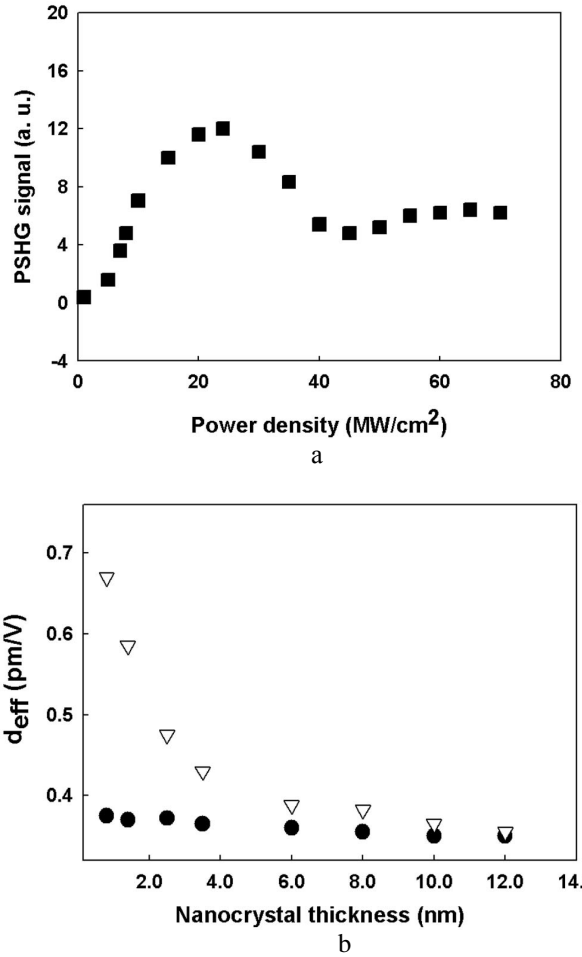


Fig. 3. Pumping power density dependence of the PSHG signal for a sample (a) and Second-order susceptibility as a function of sample thickness for the undoped (circles) and doped with 0.6% Cu (triangles) sample (b).

The second-order susceptibility determining the PSHG response as a function of sample thickness for the undoped, as well as the doped sample, is shown in Fig. 3(b). It can be noticed that the PSHG signal decreases with increasing thickness of the crystal. For thickness larger than 10 nm, the PSHG signal reduces to value observed for the undoped crystal.

The second-order susceptibility as a function of Cu-doping density is presented in Fig. 3 for nanocrystals of different thicknesses. Based on results presented in Fig.3

can conclude that the PSHG response significantly increases with increasing Cu content up to 0.6%. Insertion of the Cu impurities favors the stronger local electron-phonon interaction, particularly its anharmonic part, through the alignment of the local anharmonic dipole moments by the pumping light [7]. The local disordering of the Cu aggregators plays an additional role in the nanosize confined effects [8].

IV. CONCLUSIONS

It was shown that CdJ_2:Cu and CdJ_2:Au,Cu crystals are optically stable to the action of high electromagnetic radiation. Besides, this crystal has significant fluorescent yield and wide spectral range in luminosity as green-yellow and yellow region of the spectrum. The excitation of such crystals in selective absorption band at 3.23 eV at temperatures lower than 200 K causes also the donor-acceptor luminescence in the green and red spectral regions. Thus according to the time of afterglow the green luminescence can be attributed to the category of fluorescence, and the red luminescence can be attributed to the phosphorescence.

Creation of the Cu aggregates by Cu-impurities favors a reduction in the active electron-phonon centers, effectively contributing to noncentrosymmetry of the output charge density and leads to occurrence of the metallic clusters. In the result, it suppresses the effect at higher Cu content because of limitation of the enhancement of the local hyperpolarizability for the Cu aggregates, as well as the corresponding nonlinear dielectric susceptibility. The crucial role of the photoinduced anharmonic electron-phonon interactions was therefore confirmed in the processes observed.

REFERENCES

[1] Wide-layered crystals and their physical properties. Ed. A.B.Lyskovich. (Lviv, High School, 1982) 148 p. (In Russian)

[2] M.S. Brodin, I.V. Blonsky. Exciton processes in layered crystals. (Kiev Sciences, Dumka, 1986) 256 p. (In Russian)

[3] M. Rudka. Visnyk of Lviv Polytechnic National University. "Physics and mathematics science". - № 566, Vol. 97 (2006). (In Ukrainian)

[4] M. Rudka, Lahotskyy T.V., Harambura S.B., Kravchuk I.M. ELEKTROinform. - № 3, Vol.12 (2009). (In Ukrainian)

[5] M. Rudka Visnyk of Lviv Polytechnic National University. "Physics and mathematics science." – № 687, № 687, Vol. 181 (2010). (In Ukrainian)

[6] M. Catalano, A. Cingolani, R. Ferrara, and M. Lepore, Helv. Phys. Acta 58, 329 (1985).

[7] V. Kityk, S. A. Pyroha, T. Mydlarz, J. Kasperczyk, and M. Czerwinski, Ferroelectrics 205, 107 (1998).

[8] Bondar, Mater. Sci. Eng., B 71, 258 (2000).

ZnO as a Conductive Layer for Solar Cells Based on II-VI materials

Semikina T.V., Mamykin S.V., Shmyryeva L.N.
V.Lashkarev Institute of semiconductor physics
Kiev, Ukraine
tanyasemikina@gmail.com

Godlewski M., Luka G., Pietruszka R., Kopalko K., Krajewski T.A., Gierałtowska S., Wachnicki Ł.
Institute of Physics Polish Academy of Sciences
Warsaw, Poland
godlew@ifpan.edu.pl

***Abstract*—ZnO films with high conductivity are obtained by atomic layer deposition for the application in solar cells based on CdS/CdTe/Cu_2S films. The parameters of solar cells with ZnO electrode are compared with the ones with Mo. The advantages of ZnO electrode are discussed.**

Keywords-conductive ZnO films, atomic layer deposition, solar cells

I. INTRODUCTION

ZnO is intensively studied a wide band gap semiconductor material for applications in electronics and photovoltaics, for use as a transparent conductive material ([1] and references given there). Among the numerous technological methods used to deposit ZnO (pulse laser deposition, radio-frequency magnetron sputtering and etc.) atomic layer deposition method (ALD) demonstrates several advantages (see e.g. [1-3]). For example, the ALD allows deposition of ZnO films very homogeneous in thickness and with good structural properties on very big substrates. The films are highly transparent and conductive. Moreover, the ALD allows to load into the reaction chamber the tens of substrates in one run reducing process costs considerably. It also allows to deposit the ZnO films at low temperature (100-200 ^{0}C) using cheap precursors, resulting in costs reduction, a crucial factor for further spread of solar cell industry.

In this work we use the ALD to get ZnO films for application as conductive electrodes in solar cells of the second generation based on heterojunctions of II-VI materials. It replaces molybdenum often considered as the best electrode for structures based on II-VI materials. Mo is deposited by either magnetron sputtering or ion beam deposition. The main disadvantage of these methods is a low productivity. It is impossible to deposit a film on a big substrate, which is why we test in the present study an alternative approach. We compare the photoelectrical characteristics of solar cells structures CdS/CdTe/Cu_2S deposited on pyroceram glass/Mo or quartz glass covered by ZnO.

II. TECHNOLOGY AND PROPERTIES OF ZNO FILMS

The ALD is a variant of chemical vapor deposition (CVD), but contrary to CVD reaction precursors are introduced sequentially into the growth chamber and their pulses are separated by purging of a chamber with an inert gas (nitrogen). Therefore, precursors meet only at a surface of a growing film, where they react forming the required compound. ZnO films studied were grown using de-ionized water and diethyl zinc [$Zn(C_2H_5)_2$] in a commercial ALD reactor (Cambridge Nanotech Savannah 100, USA). Diethylzinc (DEZn) and de-ionized water serve as zinc and oxygen precursors, respectively. Each cycle of precursor's doze is separated by purging of a chamber with nitrogen.

A. Influence of technological regimes on ZnO films properties

The different applications require suitable and reproducible properties of the materials like surface roughness, optical and electrical properties, suitable crystallographic orientation. For example, for transparent conductive electrode applications in solar cells it is necessary to get ZnO films with a high conductivity, due to a high concentration of charge carriers and high mobility. Our previous investigations proved that the ALD technology allows control of films properties in a wide range via altering process parameters, in particular temperature [1-3]. We determined technological regimes for deposition of the films with good structural properties [1-3]. For example, X-ray diffraction (XRD) measurements showed that the growth mode of the ZnO films (e.g. the c-axis perpendicular or parallel to the surface) and grain diameters depend on purging time [1]. In the case of samples deposited with short purging times the main diffraction peak corresponds to the [10.0] direction, which indicates the growth mode with c-axis parallel to the substrate. In turn, for the layers prepared with an increased purging time, the c-axis is perpendicular to the surface and the [00.2] XRD peak dominates the patterns, especially at higher growth temperature (200 ^{0}C). The lattice parameters (*c* and *a*) approach the value of lattice parameters of the bulk ZnO crystal with an increase of a growth temperature [1]. The surface morphology was investigated by atomic force microscopy (AFM). We found that roughness clearly depends on a growth temperature (100 to 200 ^{0}C), thickness of the film and purging time. The roughness values range from 0.9 to 3.9 nm.

The electrical measurements (Hall effect studies) showed that the higher growth temperature leads to increasing of free electron concentration (n) [3]. The obtained results are presented in Table 1. The electron concentration varied between 10^{20} cm^{-3} (for samples grown at 200 ^{0}C and above)

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

TABLE I. RESULTS OF HALL EFFECT MEASUREMENTS FOR THE FILMS DEPOSITED ON A GLASS SUBSTRATE WITH FOLLOWING GROWTH PARAMETERS: TIME (H_2O)=15 MS, PURGING TIME=20 S, TIME (DEZN)=60 MS AND PURGING TIME=8 S (AFTER THE REFERENCE [3])

As-grown layer	Growt. temper. (^{0}C)	Thick-ness (nm)	Concen-tration (cm^{-3})	Mobility (cm^2/Vs)	Conduc-tivity ($\Omega^{-1}cm^{-1}$)
ZnO/glass	200	66.5	1.67×10^{20}	16.1	431.03
ZnO/glass	140	93.0	7.00×10^{18}	38.0	42.60
ZnO/glass	130	110.0	1.10×10^{18}	2.25	0.39
ZnO/glass	100	188.0	4.50×10^{17}	8.0	0.58

down to 10^{16} cm^{-3} (films obtained at 100 ^{0}C and below). We notice that a lower electrons concentration was obtained with a longer purging time applied after providing H_2O precursor to the growth chamber [3]. The effect was tentatively explained as follows - at longer purging process the growing layer is better protected from build-up of hydroxyl groups, which are responsible for increasing n-type concentration as they contain hydrogen atoms [3]. The values of electron mobility were usually between 10 and 50 cm^2/V·s, though for ZnO layers grown on crystalline substrates, mobility exceeds 100 cm^2/V·s. We observe the important correlation between the carriers concentration and mobility (fig.1). An increase of a growth temperature results not only in a higher *n* but also in an improved crystalline structure thus a higher mobility. Thus, to get highly conductive layers of ZnO for solar cell electrode we choose the temperature of 200 ^{0}C.

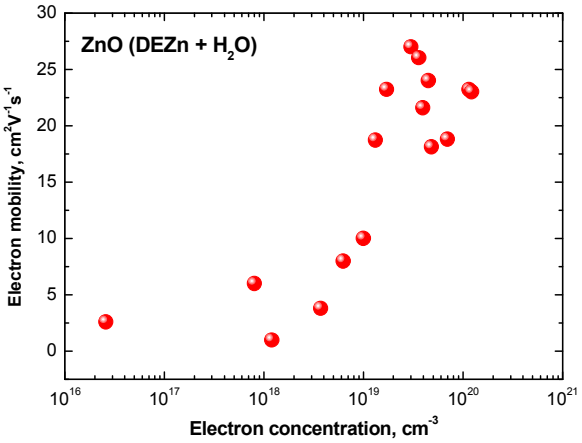


Fig. 1. Dependence between carriers concentration and electron mobility.

B. Properties of ZnO films for electrode applications

Based on the results of mentioned investigations, we selected the technological parameters for deposition of ZnO films used in the present study. ZnO films were deposited at 200 ^{0}C with thickness of 200 nm. All ZnO films have atomically flat surfaces as is derived from the AFM studies. The root mean square (RMS) of surface roughness is below 2 nm. The XRD spectra show the polycrystalline structure of ZnO films with privilege (00.2) crystallographic orientation, i.e., the situation when the c-axis is perpendicular to the layer surface. Relatively good structural quality of our ZnO films is unique feature. It was generally assumed that films grown at low temperature should be rather amorphous. In contrast, we found that ZnO films grown at low temperature (200 ^{0}C) are still ordered and are polycrystalline. The good structural quality of the films is additionally confirmed by the results of photoluminescence (PL) measurements at room temperature. It is observed a bright edge excitonic PL at room temperature. Furthermore, we do not observe defect related bands in the PL spectra. The Hall effect measurements of the prepared ZnO films show high concentration of carriers (1.42×10^{20} cm^{-3}), mobility 26.5 cm^2/V·s and small value of specific resistance ($1.66\times10^{-3} \div 1.00\times10^{-4}$ Ω·cm).

III. SOLAR CELLS ON ZNO AND MO

CdS/CdTe based solar cells are prepared on two different bottom electrodes. We first deposit the bottom electrode (Mo or ZnO), then thin layer of CdS, followed by an absorbing layer of CdTe and finally a top electrode. Cells are illuminated from the side of top electrode. Solar cells are prepared by a quasi-closed sublimation method from CdS and CdTe powders. The temperature of sublimation for CdS powder is 750 ^{0}C and durability of the process varied from 10 s to 1 min. Chamber's temperature is kept constant at 550 ^{0}C during the whole technological procedure. After the finishing of CdS deposition the heating of CdTe powder starts. The powder of CdTe is sublimated at temperature of 500 ^{0}C. Two types of substrates are installed in a holder at the same time. This allows to carry out more reliable comparison of properties of the obtained structure's. We use pyroceram covered by Mo and quartz glass with ZnO film deposited by ALD, at the conditions described above.

The goal of n-CdS layer in our case is formation of the Ohmic contact with a bottom electrode (Mo or ZnO). CdS also serves as a template for a following growth of CdTe that is an absorbing layer of the cell. The thickness of CdS layer is varied from 60 to 800 nm. The next step is deposition of a thick film of a lightly n-type CdTe (resistive film). Time of deposition varied from 3 to 6 min to get layers with a thickness 1÷3 μm. Since both CdS and CdTe layers have the n-type conductivity it is necessary to deposit the layer of p-type for p-n junction formation. For this purpose we use a layer of p-Cu_xS. Degenerated semiconductor Cu_xS is also used as a top electrode to CdTe. The size of solar cell working area is 0.196 cm^2.

The measurements of solar cells characteristics are performed at dark and under two illumination powers of 0.013 and 0.136 W/cm^2. The dark and illuminated current-voltage (I-V) characteristics of solar cells are presented in fig.2. It is evidently seen that better results are obtained for structures deposited on ZnO films in comparison with ones on Mo. The calculated efficiency of the best prepared solar cells is 1.89 % for samples on ZnO (short circuit current I_{sc}=0.249 mA, open circuit voltage U_{oc}=0.3695 V, fill factor FF=0.267). It should be noted that we do not observe any correlation between device performance and thickness of CdS and CdTe layers.

In fig.3, fig.4 we present the energy band gap diagrams of examined structures. The working barrier in our solar cells is formed between n-type CdTe and p-Cu_xS. On energy band diagram we assume that Fermi level in CdTe approximately is

in the middle of a band gap. The open circuit voltage (U_{oc}) is determined by value of band bending. The obtained small value of U_{oc} relates likely to high resistivity of CdTe layer. This value can be increased theoretically to U_{oc} = 0.7÷0.8 eV by electron doping of CdTe.

The short circuit current I_{sc} depends on recombination losses in heterojunction, at interface and in the volume of each layer. Density of surface states is here crucial. Seemingly, the interface between CdS and ZnO has a lower density of surface states as a contact of CdS and Mo. Therefore the recombination current in the case of ZnO is smaller and value of I_{sc} increases.

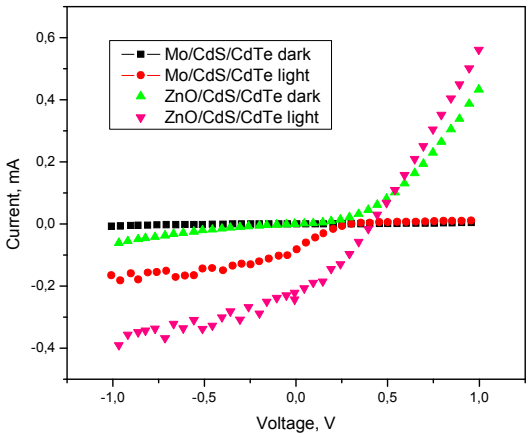


Fig. 2. I-V characteristics of CdS/CdTe solar cells deposited on metallized with Mo pyroceram substrate and on quartz glass covered by ZnO.

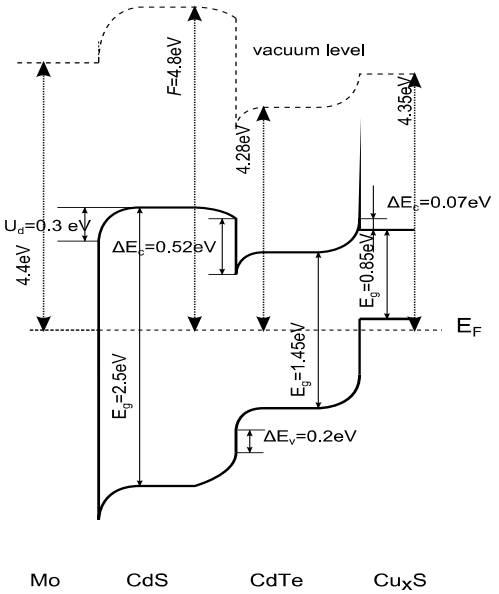


Fig. 3. Energy band diagram of structure CdS/CdTe/Cu_2S on Mo.

Moreover, in the case of contact with ZnO the diagram indicates a formation of the bigger barrier for holes and consequently the stronger value of built-in electric field.

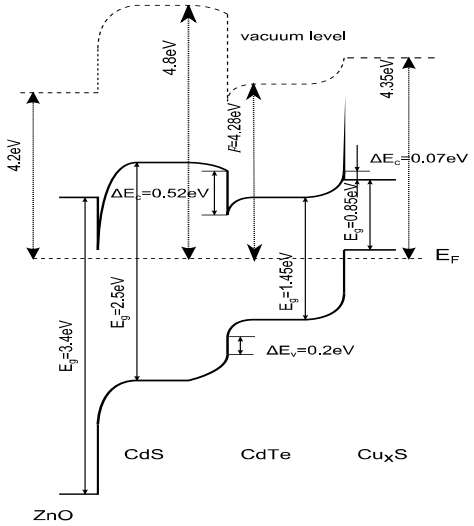


Fig. 4. Energy band diagram of solar cell CdS/CdTe/Cu_2S on ZnO.

Due to this field the holes move away from the ZnO/CdS contact and take part in photocurrent generation. It can be the reason of the better solar cell characteristics of cells deposited on ZnO. The additional argument to explanation of the improved solar cells characteristics is relative good polycrystalline structure of ZnO that promotes the more ordered growth of CdS layer. In contrary the XRD results show amorphous structure of the molybdenum contact that possibly leads to amorphous growth of thin CdS layer.

CONCLUSION

The ALD growth allows varying the structural and electrical properties of ZnO in a wide range. In particular, it allows obtaining the ZnO suitable for electrode application in solar cells. We get the ZnO films with good polycrystalline structure and high conductivity. The solar cells CdS-CdTe-Cu_xS deposited on ZnO show the improved photoelectrical characteristics as the ones deposited on Mo.

This work was partially supported by the European Union within the European Regional Development Fund, through the Innovative Economy grant (POIG.01.01.02-00-108/09).

REFERENCES

[1] M.I. Łukasiewicz, A. Wójcik-Głodowska, E. Guziewicz, A. Wolska, M. T. Klepka, P. Dłużewski, R. Jakieła, E. Łusakowska, K. Kopalko, W. Paszkowicz, Ł. Wachnicki, B.S. Witkowski, W. Lisowski, M. Krawczyk, J.W. Sobczak, A. Jablonski, M. Godlewski, "ZnO, ZnMnO and ZnCoO films grown by Atomic Layer Deposition" in Semicond Science and Technology, IOP Science, vol.27, N. 7, 2012, pp. 074009.

[2] S. Gieraltowska, L. Wachinski, B.S. Witkowski, M. Godlewski, and E. Guziewicz, "Atomic layer deposition grown composite dielectric oxides and ZnO for transparent electronic applications," in Thin Solid Films, ScienceDirect, Elsevier, vol.520, 2012, pp. 4694-4697.

[3] T. Krajewski, E. Guzievicz, M. Godlewski, L. Wachicki, I.A. Kowalik, A. Wojcik-Glodowska, M. Lukasiewicz, K. Kopalko, V. Osinniy, and M. Guziewicz, "The influence of growth temperature and precursors' doses on electrical parameters of ZnO thin films grown by atomic layer deposition technique," in Microelectronics Journal, ScienceDirect, Elsevier, vol.40, 2009, pp. 293-295.

3D IC Cooling Mechanism by Using Signaling Vias

Ara Gevorgyan
SEUA
Microelectronics chair
Yerevan, Armenia
aragev88@gmail.com

***Abstract*—In this paper is discussed the problem of efficient thermal energy removal in 3D integrated circuits (IC) based on thermal via insertion in parallel with using existing signaling vias. This method allows to minimize the number of additional thermal vias and by this save useful area of chip.**

Keywords—3D IC; thermal via; locus; convex hull

I. INTRODUCTION

The growing number of electronic components in IC requires more complex solutions from chip designers. One of the most actual problems is the problem of area of die. Growing number of elements requires more area for their placement, but the sizes of die cannot be increased without any side effects. One of such effects is that by increasing the size of die the count of defects increases too. An effective solution can be to have several dies connected with each another with signaling wires. This will allow keeping small sizes for dies, but in total have more area. This approach was realized in 3D ICs. In 3D structures several dies are stacked vertically one on another and are packed into one package. This method has several advantages. A chips are stacked vertically their interconnects are made through silicon, this allows to minimize wire length in opposite to classical traditional chips where a wire lies on horizontal surface and actually can play major role in parasitic effects like signal propagation delay, signal's edges sharpness. Also crosstalk effects can become more evident as wire length increases. Another advantage is that much more functionality can be included into one package and by this reduce the price for packaging. But with all this advantages there are several problems that this technology brings up. One of the most vital problems is the thermal dissipation problem. Due to stacking dies vertically the inner chip's layers do not have direct connection to package and the thermal energy cannot be removed effectively. To overcome this problem there are several approaches that are listed below.

- Thermal aware floorplan
- Thermal aware placement
- Thermal via insertion

In this paper is discussed usage of thermal vias that can carry thermal energy to upper layers. Thermal vias should be placed in areas with high thermal activity to effectively remove energy. The disadvantage of placing thermal via is that useful are should be used for placement. In this article is described a method to use signaling and power vias in parallel with thermal vias, and by this minimize thermal vias quantity and useful area used by them. Will be described a general flow to identify hot zones and thermal via's quantity.

II. METHOD FORMULATION

A. Problem formulation

By migrating to 3D IC technology the problem of thermal dissipation becomes more vital. Thermal via placement is one of solutions of this problem.

There are three main possible types of thermal vias. In Fig. 1. is shown a general structure of thermal via stack.

The first type of vias is intended to connect layers in the same chip. It is mostly used for cooling local points in the same chip. The second type of vias is designed to connect nearby chips. This method's usage is mostly preferable in case if difference of temperatures at the ends of via is significant. Due to low thermal conductance value of Si the thermal energy cannot dissipate to another layer and this second type of thermal vias is intended to resolve that problem. The third approach assumes to have thermal vias that go through several chips and layers. This method allows to save more area, but is difficult in implementation due to need of availability of whitespaces that are located vertically one on another in several layers. The problem of whitespace can be resolved by performing replacement of elements and blocks in some layers. This problem becomes more important in case if design is done with non standard library cells and sizes of blocks can vary. Also elements should be located on some "safe" distance from vias.

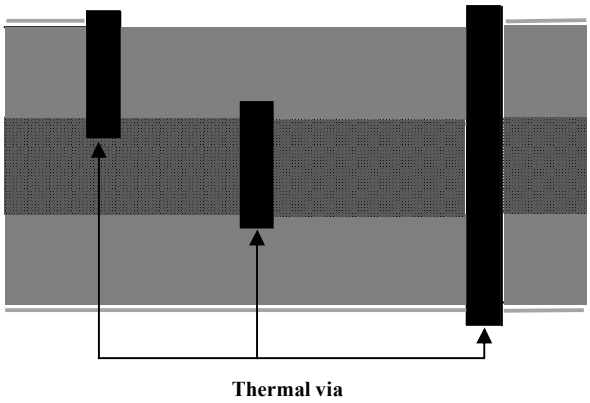


Fig. 1. Thermal via types

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

All this arguments show that thermal via insertion requires certain additional surface area. To minimize the area used by vias it is needed to minimize their number. One of the approaches can be to use signaling and power vias in parallel with thermal vias and by this minimize the area used for cooling purposes. In paper [1] is shown that signaling and power vias can be used very effectively to remove thermal energy from hot zones. In the following section will be described the general flow of such procedure.

B. Method formulation

The procedure of determining quantity of thermal vias consists of several steps. The main operations are done during physical design procedure. The general flow is shown in Fig. 2. Some steps are not included into that diagram as they don't have impact on thermal via calculation process.

As could be seen from diagram in 3D IC hot zones can be determined after Placement procedure is completed. Also to be able to use signaling vias for cooling purpose it is needed to have set of that signaling vias that will be included in cooling procedure before routing process.

Thermal via insertion itself is represented as a post placement procedure [2]. This can require some replacement actions before routing can be done. There are several parameters that are important during thermal via places determination.

Those parameters are shown below:

- coordinates of blocks
- geometrical sizes of blocks: W_B, L_B
- thermal via's sizes: W_{TV}, L_{TV}
- impact area of thermal via
- signaling via's sizes: W_{SV}, L_{SV}
- impact area of signaling via

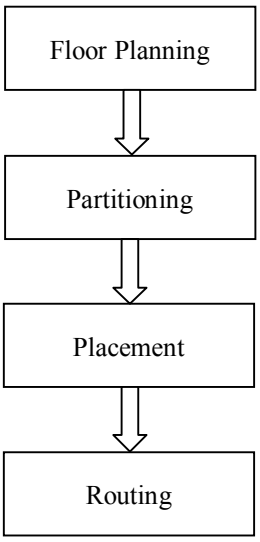


Fig. 2. Physical design flow

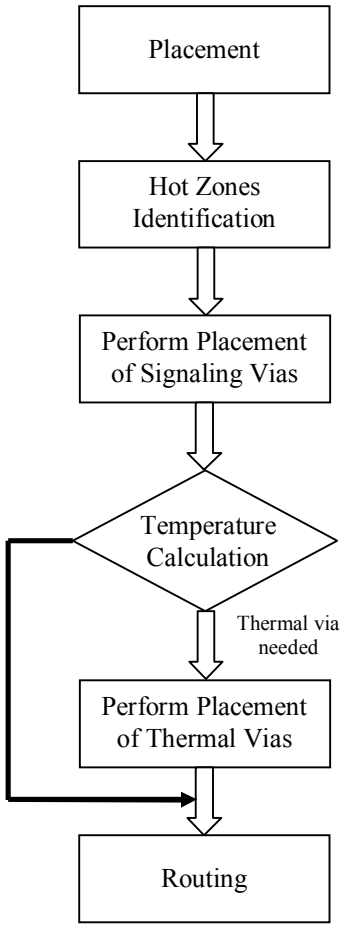


Fig. 3. Updated Physical design flow

The impact areas are the zones near vias that should be free to avoid any parasitic effects that can be produced by vias.

So after modifying of physical design flow the steps from Placement to Routing can be as is shown in Fig. 3.

The modified flow represents several steps that should be done before routing process. As Placement is completed hot zones identification should be done. After that thermal directed placement of signaling vias is performed. Then hot zones' temperature recalculation should be done and based on result is determined is there still a need for thermal vias. In case if hot zones are not cooled down enough thermal via insertion should be done. After all this steps the main routing of circuit could be done.

C. Method steps description

In this chapter are discussed the main parts of the algorithm. So the first step that should be done at post placement stage is to identify hot zones and compute the boundaries of accepted areas where in the next steps vias can be placed. This could be done by using differential equations that describes energy dissipation:

$$K_x \frac{\partial^2 T}{\partial x^2} + K_y \frac{\partial^2 T}{\partial y^2} + K_z \frac{\partial^2 T}{\partial z^2} + Q(x,y,z) = 0 \qquad (1)$$

where T is the temperature, K_x, K_y and K_z are the thermal conductivities, and Q is the heat generated per unit volume. A unique solution exists when convective, isothermal and insulating boundary conditions are appropriately applied. Boundary conditions can be applied to this equation to include packaging effects.

As hot zone identification is complete the next task that should be done is to determine whitespaces where signaling and thermal vias can be placed.

To construct whitespace the following procedure is applied. At first need to be identify the vertexes of schema blocks that lie in thermally active hot zone. This could be done by using vector multiplication. In Fig. 4 is shown this step for one point.

The set of vectors should be traversed in one direction and if all vectors multiplication has the same sign than the specified point is located insight a hull, else it is located outside of hull [3]. This operation for one vertex is done in liner time and represents an operation with complexity of O(N), where N is a number of vertexes of thermal zone. For all k points this operation will have O(N*k).

$$\begin{cases} \overline{m_1} = |\overline{h_1}| \times |\overline{v_1}| \times \cos \alpha_1 \\ \overline{m_1} = |\overline{h_2}| \times |\overline{v_2}| \times \cos \alpha_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \overline{m_n} = |\overline{h_n}| \times |\overline{v_n}| \times \cos \alpha_n \end{cases} \qquad (2)$$

This method has an advantage compared to square calculation method, because it allows avoiding problems connected to corner cases like if the point is located in one of the edges or floating calculation precision problem.

Another advantage of this method that it allows parallelization of computational process. If the computation is done on grid then each edge can be traversed separately.

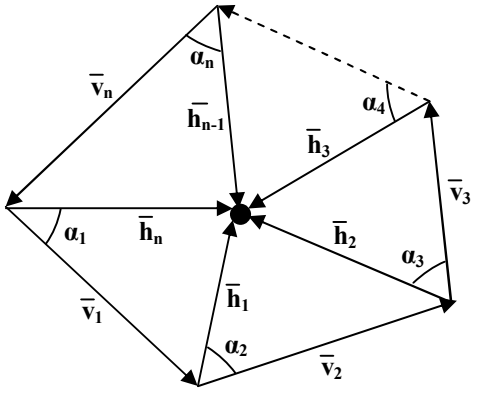


Fig. 4. Vertex's position calculation

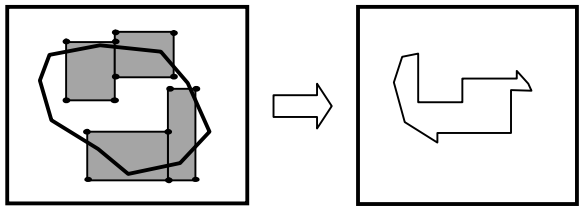


Fig. 5. Whitespace calculation

After identification of block's vertexes that are located in hot zone the next task is to calculate hot zone's edges' and block's edges' intersection points. In general this operation could be done in time of $O(n^2)$, where n is the total number of hot zone's and block's edges. But as the blocks have rectangular structure the algorithm can be improved and have complexity of O(n*log(n)).

The next step is to construct whitespace shape based on point's set from previous operation Fig. 5. There are several methods to calculate vertexes of whitespace. In this paper that task is solved by "gift wrapping" method. By starting from one the vertexes of blocks the line goes in clockwise direction until meets another vertex that belongs to block, hot zone's or hot zone and block intersection point. The complexity of this step is O(n*h), where n is the number of points and h is the number of points on the whitespace hull. If there are island blocks that are located in whitespace than their edges are considered as whitespace inner edges.

The distribution of signaling vias is preferable to do based on the strategy to fill small areas first. This is done to keep more place for thermal via insertion as thermal vias require more are for placement. To do this Manhattan grid is used Fig. 6. The signaling vias are located at the vertexes of grid and grid is calibrated on whitespace so that by this could be determined the parts of whitespace with lowest area. The cells of grid are calculated based on sizes of signaling vias and the safe zones around vias.

$$a = W_{SV}/2 + 2*dx_{SV} \qquad (3)$$

$$b = L_{SV}/2 + 2*dy_{SV} \qquad (4)$$

The same calculation is done for thermal vias.

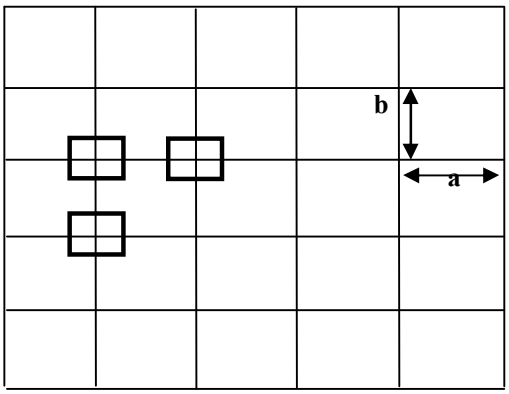


Fig. 6. Grid construction

$$a_T = W_{TV}/2 + 2*dx_{TV} \quad (5)$$

$$b_T = L_{TV}/2 + 2*dy_{TV} \quad (6)$$

To apply Manhattan grid to whitespace the calibration is done by fitting grid's vertex with whitespace's vertex.
After signaling via placement thermal simulation should be done again to check whether there is still need to insert thermal vias or not.

III. Results Discussion

The algorithm was implemented in C++. For effective data processing STL and Boost libraries are used. Computational geometry related algorithms are implemented using SGAL library. For hot zones simulation was used specially designed software module. By using computational geometry structures like inner convex hull was able to determine the whitespace boundaries. The usage of gift wrapping algorithm is preferable as the amount of memory is less and it has acceptable performance. The following are presented performance results of algorithms for signaling vias placement zone and whitespace identification (Fig. 7).

As could be seen from graphics for the first task working time is O(n). This is achieved by usage of grids.

For whitespace identification was achieved O(NlogN) speed, which is optimal in its class (Fig. 8).

After signaling via placement thermal analysis should be done to check whether there is still need to insert thermal vias or not. In case if signaling vias are not enough to cool down the hot region then for the hot zone should be re-identified the whitespaces available for thermal via placement [4-7]. As thermal vias require more area so the signaling vias should be inserted as compacted as possible. This was achieved by using Manhattan grid with the cell required for signaling via's sizes and safe zones.

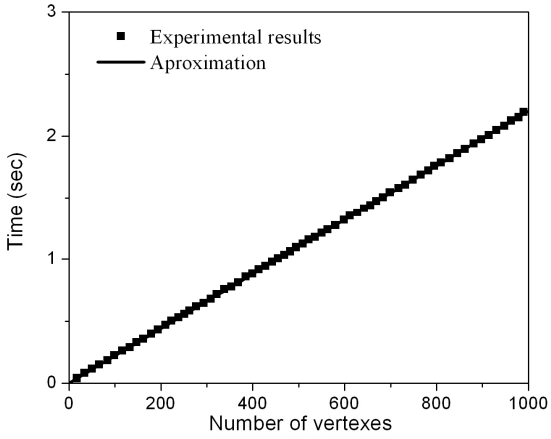


Fig. 7. Inner hull calculation performance

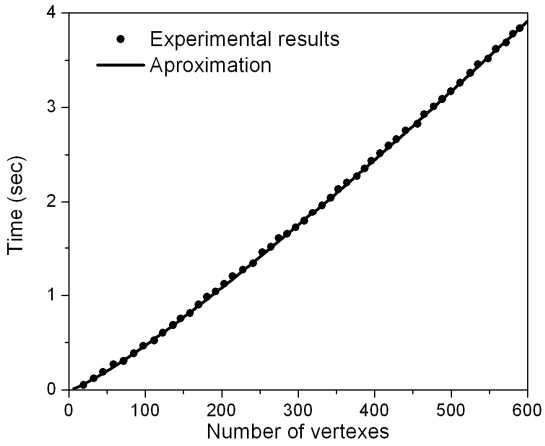


Fig. 8. Whitespace calculation performance

Conclusion

In this paper was discussed a general method that can be used during 3D IC design to minimize thermal dissipation effects. Was discussed the possible approach to use signaling vias as a cooling mechanism. Was presented a method to perform signaling via placement. In case if whitespace for thermal via placement is not available it is possible to perform replacement of blocks. This option could be chosen during design. By using triangulation the whitespace identification algorithm's time is O(NlogN), which is optimal in this class of algorithms. Experimental results were presented as well.

References

[1] Min NiQing SuZongwuTangJamil Kawa. An Analytical Study on the Role of Thermal TSVs in a 3DIC Chip Stack. DATE 2010.

[2] J. Cong and Y. Zhang, "Thermal via planning for 3-d ics," in *Proc. IEEE Int. Conf. on Computer-Aided Design*, 2005.

[3] M. D. Breg, M. V. Kreveld, M. Overmars and O. Schwarzkopf. Computational Geometry, Algorithms and Applications, Springer, 1997.

[4] X. Li, Y. Ma, X. Hong, S. Dong and J. Cong, "LP based white space redistribution for thermal via planning and performance optimization in 3D ICs," Asia and South Pacific Design Automation Conference, pp.209-212, 2008.

[5] J. Cong and Y. Zhang, "Thermal via planning for 3D ICs," International Conference on Computer-Aided Design, pp.745-752, 2005.

[6] S. Majumder, S. Sur-Kolay, S. C. Nandy,B. B. Bhattacharya and B. Chakraborty, "Hot Spots and Zones in a Chip: A Geometrician's View", Proc. Int. Conf. VLSI Design, pp. 691-696, 2005.

[7] Ming-Che Hsieh, Chih-Kuang Yu, Sheng-Tsai Wu "Thermo-Mechanical Simulative Study for 3D Vertical Stacked IC Packages with Spacer Structures", Semiconductor Thermal Measurement and Management Symposium, 2010. SEMI-THERM 2010. 26th Annual IEEE, pp. 47-54, 2010.

Improvement of Piezoceramic Scanners

Sharapov V.M., Filimonov S.A., Sotula Zh.V., Bazilo K.V., Kunitskaya L.G., Zaika V.M.
Cherkasy State Technological University
Cherkasy, Ukraine
e-mail: v_sharapov@rambler.ru

***Abstract* — The article is devoted to the problems of piezoscanners. The analysis of existent domestic and foreign constructions of piezoscanners is conducted. The design of the piezoceramic scanner on the basis of flat bimorph piezoelements of a trapezial form for scanning probe microscopes is offered and investigated. Built and probed model of this scanners through the package of the program Femlab 3.5.**

Keywords — nanotechnology; scanning probe microscopy; piezoceramic scanner

I. INTRODUCTION

The scanning probe microscopy allows to analyze at nuclear level of structure of the most different materials – glasses, ceramics, plastic, metals, semiconductors. Measurements can be carried out not only in vacuum, but also on air, in the environment of any gas and even in a liquid drop. This method is irreplaceable and for research of biological objects. Except the research functions, the scanning probe microscopy can carry out also the active – to provide capture of separate atoms, their transfer in a new position, atomic assembly of conductors one wide atom, local chemical reactions, a manipulation with separate molecules, etc. [1]-[3].

Basic element of scanning probe microscopy which provides possibility of operation of the device in modes of nuclear permissions and which significantly influences quality of the received image, is the piezoceramic scanner [1].

II. PROBLEM AND APPROACH

In scanning probe microscopy the wide circulation was gained by scanners on the basis of tubular piezoelements [1]-[4]. Lack of such scanners is that when scanning the next sections of the scanner have impact at each other that worsens the accuracy of measurements, reduces their reliability, reduces the limit resolution of the device [3]. Besides, the sections of the tubular scanner which aren't participating in movement, interfere with a bend of the piezoceramic scanner in the necessary direction that reduces scanning range [4].

Performance of this scanner in the form of the cylindrical bimorph element divided into sections (actuators) (Fig. 1), allows to increase sensitivity and to reduce communication between movements on X and Y coordinates, however the manufacturing techniques of the scanner become more difficult [4], [5].

The wide circulation was gained also by scanners on the basis of flat bimorph piezoelements. The bimorph piezoelement (BPE), as it is known [6]-[10], represents two piezoelements or a piezoelement and the metal plate, connected among themselves by means of epoxy glue or fusible solder [8], [9]. At influence of electric tension the bimorph piezoelement makes flexural fluctuations that is used for probe movement concerning object of researches or object of researches concerning a probe.

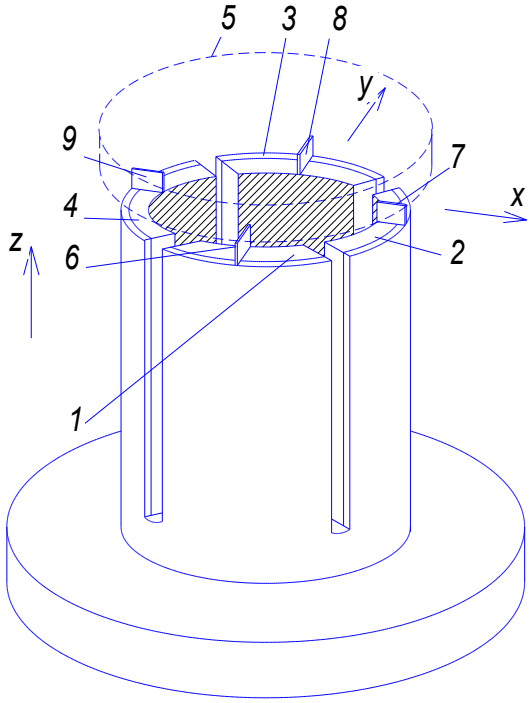


Fig. 1. Cylindrical bimorph scanner: 1-4 – actuators on X and Y coordinates; 5 – the actuator on Z coordinate; 6-9 – planar elastic plates.

In literature [3] the three-coordinate scanner on the basis of three round bimorph piezoelements (tripod) is described. Lack of this scanner is strong asymmetry of a design, and also communication between movements on coordinates.

The work's purpose is improvement of piezoceramic scanner on the basis of planar bimorph piezoelements.

For elimination of the specified shortcomings it is offered to use on two asymmetric bimorph squared piezoelements (1,3 and 2,4), fixed on the one hand on the basis of 14 (fig. 2) [11]. These bimorph piezoelements (actuators) are used for creation of movements on X and Y coordinates. Bimorph piezoelements consist of metal plates (6,7 and 8,9) and plates from piezoceramics (10,11 and 12,13), stuck together with the help of epoksid glue.

Except a design presented in Fig. 2 the design on the basis of a bimorph planar tuning fork [12] is developed. Distinctive

feature of this design from offered above scanners is the reduced number of actuators which became possible thanks to change of their relative positioning. Besides, this difference gives the chance of scanner turn round its pivot-center that allows to use it in case of use of the massif of probes in the form of a ruler or a matrix, and respectively, the massif of nanoelements, for example, in memory.

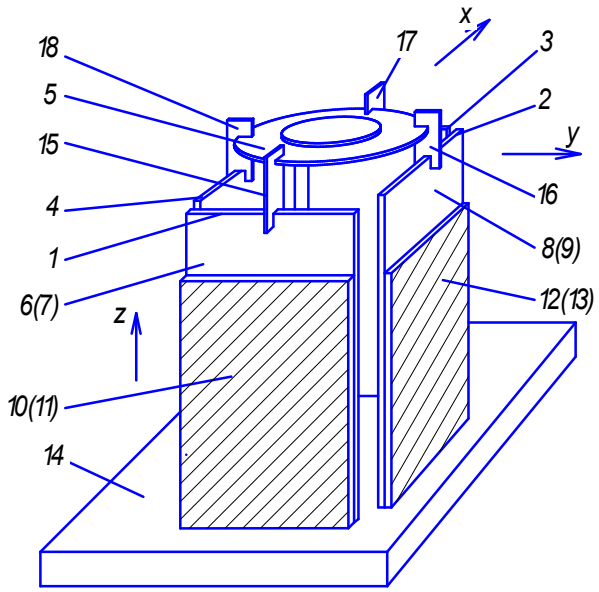


Fig. 2. The piezoceramic scanner on the basis of planar bimorph piezoelements: 1-5 - bimorph elements; 6-9 - metal plates; 10-13 - piezoelements; 14 - the basis; 15-18 - elastic plates.

Influence of fluctuations of one of bimorph piezoelements on other bimorph piezoelements was defined also. On an entrance of this piezoelement give electric tension at a resonant frequency of this bimorph piezoelement (886 Hz for BPE on coordinates of XY and 1340 Hz for BPE on Z coordinate) and measure movement on it and other bimorph piezoelements.

In all the offered designs (developed) by authors the metal plate on which the piezoelement is fixed represents a beam of equal section [13], [14].

Having carried out the analysis of a beam of equal section it is possible to draw a conclusion. By means of the section of a beam it is possible to influence its own frequency significantly. Reduction of thickness of a beam for the purpose of increase in sensitivity leads to equivalent reduction of own frequency of fluctuations of an elastic element. Characteristics of a material influence sensitivity also as well as in other cases.

It is possible and other option of a form of an elastic element is a beam of equal resistance to a bend, Fig. 3.

The elastic element of such form differs from the previous subjects that the size of tension as for the top layer of metal in a beam, and for the bottom don't depend on a section choice, are constant on length of a beam and are equal to tension in seal. Therefore the constructive coefficient of sensitivity of such elastic element doesn't depend neither on section coordinates, nor on completeness of use of a zone of deformations. Its size for all cases is constant, only its sign depending on that what deformations changes are used: top or bottom layers of a material.

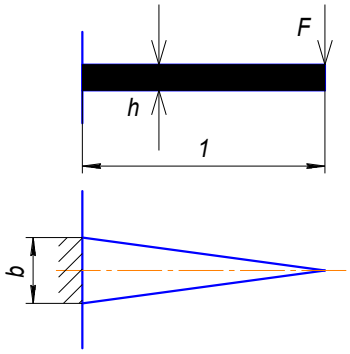


Fig. 3. Beam of equal resistance to a bend.

The deflection of the end of a beam to be defined by expression $\delta = \frac{6l^3}{b\,h^3\,E} F$. For determination of own frequency of a console beam of equal resistance to a bend we will use a ratio

$$f_0 = 0.159 \frac{B}{l^2} \sqrt{\frac{E\,J_0}{m_0}}, \qquad (1)$$

where J_0 – the inertia moment at seal; m_0 – the mass of unit of length of a beam at seal;

$$B = \sqrt{\frac{(1-\eta)(2-\eta)}{2(C_1 - C_2)}};$$

$$C_1 = \frac{1}{4-\eta} - \frac{1+\eta}{6};$$

$$C_2 = \frac{1}{3-\eta}\left[0.5 - \frac{1}{(4-\eta)(5-\eta)}\right] - \frac{3\eta+2}{24};$$

η – the indicator characterizing change of rigidity depending on the current coordinate x, a beam changing in the direction of length (the beginning of coordinates is located in seal). From a formula it is visible that η =1 size B=0 that doesn't correspond to physical representations. Really, if to construct the dependence schedule B(η), near unit of a branch of this function will significantly disperse. Thus, the offered mathematical model gives the chance of calculation B only at η<1 or η>1. For obtaining values η=1 build both curves and graphic interpolation which η=1 gives B=0,88 value is carried out. Substituting value B in a formula and making similar previous the transformations we will receive

$$f_0 = \frac{0.136\,h}{l^2} \sqrt{\frac{E}{\rho}} \qquad (2)$$

Thus, the considered designs of elastic elements have function of transformation and sensitivity which are defined by the constructive sizes, the module of elasticity and constructive coefficient of sensitivity. The constructive coefficient of sensitivity defines not only the sensitivity size,

but also its sign which depend on coordinates of section, nature of deformation and extent of use of zones of deformations.

Analyzing expressions for own frequencies of all considered designs, it is possible to draw the following main conclusions:

- Own frequency of an elastic element is defined by some numerical coefficient, one design data (the generalized parameter of length of an elastic element) and two characteristics of a material (the module of elasticity and density).
- Own frequency of subjects is more, than it is more relation of the module of elasticity of a material to density less generalized parameter of length of an elastic element.

In the software package Femlab 3.5 authors carried out modeling of one of bimorph elements of the developed piezoceramic scanner on the basis of flat bimorph piezoelements. In one option the metal plate of equal section, and in other of the equal resistance to a bend is used. Results of modeling are presented in fig. 4, 5.

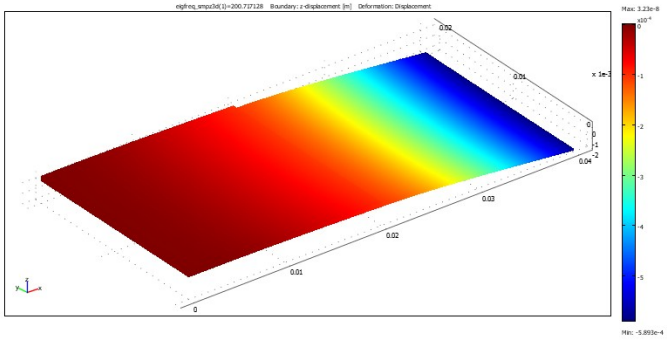


Fig. 4. Bimorph piezoelement at the heart of which the metal plate of equal section is put.

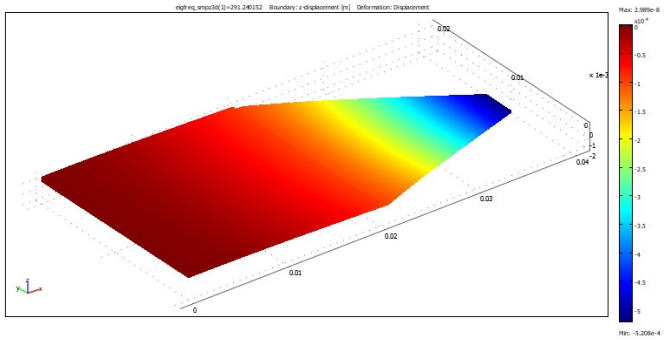


Fig. 5. Bimorph piezoelement at the heart of which the metal plate of equal resistance to a bend is put.

From fig. 4,5 it is possible to draw a conclusion that when using a metal plate of equal resistance to a bend resonant frequency increased by 45%, amplitude of fluctuations slightly decreased (by 7%).

These results correspond to the theory of resistance of materials according to which, as we know, at a beam bend with the jammed end the maximum movements arise on the free end of a beam whereas the maximum tension in a jamming place. Respectively for a bimorph piezoelement it is rational to use a metal plate such at which on the end the minimum area of section is used, and in a place of its jamming – maximum.

III. CONCLUSIONS

1. The design of the piezoceramic scanner on the basis of planar bimorph piezoelements of a trapezial form for scanning probe microscopes is offered and investigated, in which resonant frequency increased by 45%, amplitude of fluctuations slightly decreased (by 7%).

2. The constructed equivalent model of this scanner by means of software packages of FemLab 3.5 allows to predict work the piezoceramic scanner on the basis of planar bimorph piezoelements.

3. It is established that communication between bimorph elements which create fluctuations on X, Y, Z coordinates, practically is absent in the developed construction.

REFERENCES:

[1] Binnig G., Rohrer H. – Scanning tunneling microscopy. // Helv. Phys. Acta, v. 55, №6, pp. 726-735 (1982).

[2] Nanotehnology in electronics. Under rel. U.A. Chapligina. – M.: Technosfera, 2005. – 448 p. (in Russian).

[3] Mironov V.L. Fundamentals of scanning probe microscopy Moscow: Technosfera, 2004. – 144 p. (in Russian).

[4] Bimorph cylindrical piezoceramic scanner for scanning probe nanomicroscopes / V. Sharapov, A. Vladishauskas, S. Filimonov // Ultrasound. Kaunas: Technologija. 2009. vol.64, №4.

[5] Patent of Ukraine №22600. Piezoscanner / Sharapov V.M., Gyrgii A.N., Alpatov A.P., Filimonov S.A. – Cl. G12B 21/20 H01L 41/00; Bull. № 5. – 2007 (in Ukraine).

[6] Domarkas V., Kazys R. Piezoelectric transducers for measuring devices. – Vilnus: Mintis, 1975. – 258 p. (in Russian).

[7] Kazys R. Ultrasonic information-measuring systems. Vilnius: Moklas, 1986. – 216 p. (in Russian).

[8] Sharapov V.M., Mysienko M.P., Sharapova E.V. Piezoceramic transducer of physical sizes. Moscow: Technosfera, 2006. – 632 p. (in Russian).

[9] Sharapov V. Piezoceramic sensors. – Springer Verlag, 2011. – 498 p.

[10] Sharapov V.M. and other. Gauges / Under edit. V.M. Sharapov and E.S Polishuk. – Cherkassy: Brama, 2008. – 1072 p. (in Russian).

[11] Patent of Ukraine №22601. Piezoscanner / Sharapov V.M., Gyrgii A.N., Alpatov A.P., Filimonov S.A. – Cl. G12B 21/20 H01L 41/00; Bull. № 5. – 2007 (in Ukraine).

[12] Improvement of piezoceramic scanners for probe microscopes / V. Sharapov, S. Filimonov // Collection of reports IX International conference "Methodological Aspects of Scanning Probe Microscopy". Minsk: 2010. – pp. 172 – 176. (in Russian).

[13] Landau L.D. Theoretical physics. In 10 t. T.VII. Elasticity theory: Studies. grant. 4 prod. / L.D.Landau, E.M.Livshits – M: Science, hl. edition physical. - a mat. litas. 1987 . –248 p. (in Russian).

[14] Smirnov A.B. Mekhatronics and robototekhnic. Systems of microsupply with piezoelectric actuator: Teaching aid. SPb.: Publishing SPbGPU, 2003. (in Russian).

Mathematical Modeling of Electromagnetic Wave Propagation in Inhomogeneous Medium

I. Yu. Dmitrieva
Higher Mathematics Department
Odessa National Academy of Telecommunications, ONAT
Odessa, Ukraine
irina.dm@mail.ru

***Abstract*—Analytical study of electromagnetic wave propagation is done in the case of excited inhomogeneous isotropic linear medium. Mathematical modeling is based on the respective boundary problems whose partial differential equation (PDE) is the general wave one regarding the unknown electromagnetic field intensities. In its turn, this PDE is generated by the specific statement of differential Maxwell system. Solvability criterion of the latter is proved in terms of equivalence to the general wave equation in the class of non generalized functions.**

Keywords—mathematical modeling; electromagnetic wave propagation; inhomogeneous medium; differential Maxwell system; general wave equation

I. INTRODUCTION

It is known that the physical essence of signals and waves is electromagnetic field. It can be described mathematically by the system of PDEs (partial differential equations) with respect to the unknown vector field intensities [1], [2]. Such systems represent analytical simulation of the appropriate industrial / physical phenomena in the various kinds of media, and their explicit solution remains traditionally required even nowadays when computer technique develops so fast. In particular, the mentioned scientific trend concerns the modern problems of radio engineering and telecommunications.

It explains the main reason why the present results deal not only with the theoretical modeling of electromagnetic field behavior and/or wave propagation. This current research also tentatively suggests mathematical study of the relevant applied problems describing processes of technical electrodynamics.

II. PRELIMINARIES

The given paper is governed by [3] that is evoked, in its turn, by [4], [5]. Thus it was shown in [4] that the first two differential Maxwell equations were not equipollent with respect to the third and fourth ones considering them as the additional restrictions in the original system. Moreover, only immobile media were taken into account. Further, it was pointed out in [5] that the third and fourth differential Maxwell equations were the corollary of those first two from the same system. This case dealt with electromagnetic fields being harmonic in the time meaning. Similar analogy with expofunctional fields was noticed in [6] where the kernels of expofunctions did not depend on time or turned out to be harmonic regarding the temporal argument.

All mentioned conclusions concerned only linear homogeneous isotropic immobile media. Hence, it was quite natural that the next research step ran into the case of excitation. Such problem was raised in [3] basing on the experimental data from [6]. It was proved there in [3] that the third and fourth differential Maxwell equations were the corollary of two other while the medium was expofunctional and excited. Thus the so called "symmetrical" differential Maxwell system has appeared in [7] generalizing the original system from [3]:

$$\mathbf{rot}\left[\vec{H},-\vec{E}\right]^T=\left[\sigma+\varepsilon_a\partial_0^*,\ r+\mu_a\partial_0^*,\right]^T\left[\vec{E},\vec{H}\right]^T+\left[\vec{j}^{OS},\vec{e}^{OS}\right]^T. \quad (1)$$

At last medium was excited in (1) remaining isotropic homogeneous and linear.

Explaining all symbols in (1) it should be noted at first that $[\ ,\ ]^T$ is the two-dimensional transposed matrix column and $\vec{E},\vec{H}=\vec{E},\vec{H}(x,y,z,t)$ are the unknown electromagnetic field vector intensities with scalar components $E_k, H_k = E_k, H_k(x,y,z,t)\ \left(k=\overline{1,3}\right)$. Further, $\sigma,\mu_a,\varepsilon_a=\text{const}>0$ determine specific conductivity, absolute magnetic and dielectric permeability of the medium respectively. Partial differential operators look like $\partial_0^*=\partial_0\pm\lambda,\ \partial_0=\partial/\partial t$; $\lambda=\text{const}>0$ is the parameter of a signal exciting the medium. The sign in front of λ means the medium's reaction. Absorption of signal corresponds to the "+", and activity of the medium connects with the "-". Theoretical constant $r>0$ is responsible for the symmetry of the right parts in (1), and is accepted here only for the future mathematical computation. At the last stage, r can be deleted simplifying calculating procedure and not infringing the original physical problem statement. The given functions $\vec{j}^{OS},\vec{e}^{OS}=\vec{j}^{OS},\vec{e}^{OS}(x,y,z,t)$ whose scalar components are $j_k^{OS},e_k^{OS}=j_k^{OS},e_k^{OS}(x,y,z,t)\ \left(k=\overline{1,3}\right)$, describe the outside current sources and intensities. Vectors

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

$\vec{E}, \vec{H}, \vec{j}^{OS}, \vec{e}^{OS}$ belong to those functional classes that are required later by the particular boundary problem statement. In comparison with [3], system (1) has the additional constant $r > 0$ and the known vector $\vec{e}^{OS}$. Besides, analytical study of the original Maxwell system was not proposed in [3].

System (1) represents mathematical model of electromagnetic field behavior / wave propagation in the various kinds of media, complicated structures in the multidimensional analogous circuits and filters with the distributed parameters as well [3], [5], [8], [9].

Using the operator analogy [7] of the algebraic Gauss method [10], (1) was reduced to the general wave equation with respect to the unknown electromagnetic vector field intensities [7]. As far as it is known, such procedure did not appear earlier. The same result was obtained for an inhomogeneous medium when $\sigma, \mu_a, \varepsilon_a$ were continuous functions over the space (x, y, z) [11]. In both cases, the corresponding general wave PDE was solved explicitly [11].

Nevertheless, though the proposed mathematical methods were more or less innovatory, and all mentioned results looked somehow useful practically [11], they did not permit simulating directly the particular problems of technical electrodynamics. It was completely explicable because of restrictions related with the aforesaid operator analogy of Gauss method [7]. It claimed invertibility and commutativity in pairs of all matrix operator elements. Actually, the first demand implied investigation of all operator kernels and their further analysis. This approach was rather tiresome and finally vague.

Only the inverse matrix operator construction [11] jointly with the solvability theorem has allowed formulating correctly boundary problems mathematically describing electromagnetic field features.

The specific statement of (1) was considered briefly in [12], and a medium was homogeneous there. This case had to be studied separately since could not be got as the corollary of general solvability conditions [11]. However, more interesting inhomogeneous special statement of (1) was not even affected yet. It meant that a lot of current phenomena in radio engineering and telecommunications remained out of the effective analytic research.

Hence, *the goal of the given paper is mathematical modeling and explicit investigation of the specific wave propagation governed by (1), while a medium is excited and inhomogeneous.*

III. The Problem Statement

Let the specific inhomogeneous case of (1) be given

$$\partial_1 [H, -E]^T = [\sigma + \varepsilon_a \partial_0^*, r + \mu_a \partial_0^*]^T [E, H]^T + [j^{OS}, e^{OS}]^T . \quad (2)$$

In (2), all designations remain the same as in (1), only instead of **rot**, operator $\partial_1 = \partial / \partial x$ appears. Since the medium is inhomogeneous now, its parameters become continuous functions $\sigma, \mu_a, \varepsilon_a(x)$. The original vector functions from (1) are scalar here, i.e. $E, H, j^{OS}, e^{OS} = E, H, j^{OS}, e^{OS}(x,t)$, though their physical meaning does not change. The restrictions concerning these functions' belonging to the particular class are not as rigorous as in (1). It is quite explicable by the simpler structure of (2) in comparison with (1). Besides, inhomogeneity of a medium does not influence either on this functional class or on the further investigation procedure [12]. As it appears later, only the solvability criterion of (2) must be considered here more carefully than the similar homogeneous case from [12].

Taking into account all recent information, it should be noted that the explicit research of (2) consists of three stages. The first is the reduction of (2) to the specific equivalent wave PDE. It should be done using the inverse matrix operator construction. The second step is the solvability criterion statement. This theorem is proved basing on the interrelations between (2) and the general wave equation. The last, third approach deals with the formulation of the particular boundary problems in terms of the aforesaid criterion. These problems simulate mathematically respective phenomena of technical electrodynamics.

IV. The Main Results

Study of (2) at the first step leads to the following matrix form

$$MF = f , \quad (3)$$

where

$$M = \{m_{ij}\}_{i,j=1}^{2} =$$
$$\{m_{11} = -m_{22} = \partial_1;\ m_{12} = -C;\ m_{21} = -D\};$$
$$C = \sigma + \varepsilon_a \partial_0^*;\ D = r + \mu_a \partial_0^*;\ F = [H, E]^T ; \quad (4)$$
$$f = [j^{OS}, e^{OS}]^T .$$

Then we construct the inverse matrix operator M^{-1}

$$M^{-1} = -(\det M)^{-1} M , \quad (5)$$

and

$$\det M = -(\partial_1^2 + \tilde{\partial}_0^2) \quad (6)$$

is the determinant of matrix M. Partial differential operator $\tilde{\partial}_0^2$ from (6) is written below

$$\tilde{\partial}_0^2 = CD = \mu_a \varepsilon_a (\partial_0^*)^2 + (\sigma \mu_a + r \varepsilon_a) \partial_0^* + \sigma r . \quad (7)$$

All designations in (6), (7) are introduced in (4), and each "operator power" means the usual consecutive operator application. E.g., M^2 means the double application of M .

Influence of (5) – (7) upon (3), (4) leads to the general scalar wave PDE with respect to the unknown electromagnetic field intensities

$$-\left(\partial_1^2+\widetilde{\partial}_0^2\right)F=f^*. \tag{8}$$

In (8):

$$f^*=\left[f_1^*,\,f_2^*\right]^T,$$

$$f_1^*=-\partial_1 j^{OS}+Ce^{OS},\; f_2^*=Dj^{OS}+\partial_1 e^{OS}. \tag{9}$$

Investigation of $\det M$ from (6) as the square operator polynomial regarding ∂_0^*, with the additional restriction of $\mu_a,\,\varepsilon_a(x)\neq 0$, gives the real opportunity getting those conditions when inverse matrix operator M^{-1} exists. Either

$$\partial_1^2>\left(\sigma\sqrt{\mu_a/\varepsilon_a}-r\sqrt{\varepsilon_a/\mu_a}\right)^2\Big/4, \tag{10}$$

or

$$\partial_1^2\le\left(\sigma\sqrt{\mu_a/\varepsilon_a}-r\sqrt{\varepsilon_a/\mu_a}\right)^2\Big/4\text{, and}$$

$$\partial_0\neq\mp\lambda- \tag{11}$$

$$\left(\left(\sigma/\varepsilon_a+r/\mu_a\right)\pm\sqrt{\left(\sigma/\varepsilon_a-r/\mu_a\right)^2-4\partial_1^2/\left(\mu_a\varepsilon_a\right)}\right).$$

Formulas (10), (11) are based on the classical requirements of inverse operator M^{-1} existence. It means that $\operatorname{Ker}M=\varnothing$ [13], and $\operatorname{Ker}M$ is the kernel of original operator M . Numerical values of inequalities (10), (11) are understood in terms of their acting with the appropriate class of functions.

In addition, it should be stressed here that the sign interchange in front of λ in (11) is independent of "$\pm$" near the square root "$\sqrt{\ }$" from the same expression.

Further, it is necessary to assume the numerical zero values of either μ_a or ε_a [14], [15]. Then the conditions of the inverse matrix operator M^{-1} existence follow directly from the inequality $\det M\neq 0$ (look and compare with (5) – (7)):

$$\mu_a=0\Rightarrow\partial_1^2\neq-r\left(\varepsilon_a\partial_0^*+\sigma\right), \tag{12}$$

$$\varepsilon_a=0\Rightarrow\partial_1^2\neq-\sigma\left(\mu_a\partial_0^*+r\right). \tag{13}$$

Comparing solvability conditions (10), (11) with those from [12], one can notice their identity. However, in the present case of inhomogeneous medium, $\sigma,\,\mu_a,\,\varepsilon_a(x)$ are continuous functions. It means that their numerical values can be as positive as negative. In terms of (10), (11), the last fact implies that $\operatorname{sgn}\mu_a=\operatorname{sgn}\varepsilon_a$. This equality is in conformity with the necessary condition following from the electrodynamics equations. It concerns the possibility of electromagnetic wave propagation only in those media whose signs of magnetic and dielectric permeability are equal [14] – [16].

In such cases, the backward waves appear and they can extend to the whole volume of an infinite medium. Moreover, the parameters of a medium become controlled because of the magnetic field modification [16]. Additionally, these specific media whose both electromagnetic field permeabilities are negative form the bases of the metamaterials construction [14] – [16]. Importance of its application to modern radio engineering and telecommunications is incomparable. It includes creation of various absolutely new types of the waveguides, filters, antennae, etc. [14] – [19].

Turning back to solvability conditions (10), (11) we propose the following theorem

Solvability criterion of (2): *the specific case of symmetrical Maxwell system – (2) is solved explicitly in the meaning of its equivalence to the general scalar wave PDE (8) iff conditions (10), (11) (in particular, (12), (13)) are true. Only ordinary classical, non generalized functions are taken into account.*

Necessity of this theorem was already confirmed by existence of (10), (11). Sufficiency is proved easily applying the prototype operator M from (4) to the general wave equation (8). As the result, we obtain the original system (2). Thus the second step of the initially declared investigating scheme is fulfilled. Afterwards, we suggest several typical boundary problems simulating mathematically corresponding electromagnetic field behavior / relevant wave propagation.

Namely,

$$-\left(\partial_1^2+\widetilde{\partial}_0^2\right)F=f^*, \tag{14}$$

$$\begin{cases}x,t\in[0,+\infty);\,F(x,0)=g_1(x),\,F(x,t)|_{t=+\infty}=g_2(x);\\ F(0,t)=g_3(t);\,\partial^k F(x,t)/\partial x^k\,|_{x=+\infty}=0\;(k=0,1).\end{cases} \tag{15}$$

$$x\in[0,+\infty),\;t\in[0,T];$$

$$\begin{cases}F(x,0)=g_1(x),\,F(x,T)=g_2(x);\,F(0,t)=g_3(t);\\ \partial^k F(x,t)/\partial x^k\,|_{x=+\infty}=0\;(k=0,1).\end{cases} \tag{16}$$

$$\begin{cases}x\in[0,\pi],\,t\in[0,T];\,F(x,0)=g_1(x),\\ F(x,T)=g_2(x);\,F(0,t)=g_3(t),\,F(\pi,t)=g_4(t).\end{cases} \tag{17}$$

Everywhere in (15) – (17), the general wave PDE is from (14), and was determined completely in (8), (9). The known functions $g_k(x)\,(k=1,\,2),\,g_l(t)\,(l=3,\,4)$ are continuous in the respective intervals, and each of boundary conditions (15) – (17) has its own fixed functions.

Problems (14), (15) – (17) are solved explicitly using the integral transform method [20]. It is applied to the spatial variable x taking temporal argument t as the main one. So, the last step of study is done and the purpose of present paper is completely attained.

V. Concluding Remarks

Solvability criterion of (2) for excited inhomogeneous medium is proved in the meaning of equivalence between specific differential Maxwell system (2) and general wave PDE (8). Suggested result allows formulating the required boundary problems for (2) in terms of (8). It is clear, that such approach is easier than dealing with (2) directly. Thus (14), (15) – (17) represent mathematical simulation of the respective wave propagation in technical electrodynamics.

Given results are confirmed by the experiment from [21] in the homogeneous statement. Experimental study of the inhomogeneous case is in its preparation process.

Moreover, the modest advantage of the suggested criterion is application to those very important media whose both electromagnetic field permeabilities are negative.

As far as it is known, the proposed method of inverse matrix operator construction was not used for solution of PDEs systems. It concerns the differential Maxwell system too. Author tentatively hopes that this method can be useful for engineers because of its simplicity. Really, the given procedure is based on the ordinary algebraic inverse matrix construction. Analysis of its determinant zero value gives the solvability conditions. Substitution of the particular numerical values allows finding the correct boundary problem statement. The latter, in its turn, generates the right mathematical modeling of the respective physical / engineering process.

Acknowledgment

The author would like to thank all reviewers for their useful remarks.

References

[1] V.V. Nikolskiy and T.I. Nikolskaya, Electrodynamics and Radio Wave Propagation. The Textbook for Technical Universities, 3rd ed. Moscow: Nauka (Science), 1989. (Russian).

[2] J. Clerk Maxwell, Selected Works on the Electromagnetic Field Theory, 1st ed. Moscow: GITTL (the State Technical Theoretical Literature Edition), 1954. (Russian).

[3] A.M. Ivanitckiy, "Dependence of the third and fourth Maxwell equations upon the first two equations with an arbitrary excitation of the electromagnetic field", in Scientific Works of ONAT, no. 2, 2004, pp. 3-7. (Russian).

[4] A. Zommerfeld, Electrodynamics, 1st ed. Moscow: IL (FL), 1958. (Russian).

[5] Yu.V. Pimenov, V.I. Volman and A.D. Muravtsov, Technical Electrodynamics, 6th ed. Moscow: Radio i Svyaz (Radio and Communications), 2002 (Russian).

[6] A.M. Ivanitckiy, "Electric charge and magnetic flux of the expofunctional field", in Scientific Works of ONAT, no. 1, 2004, pp. 3-8. (Russian).

[7] A.M. Ivanitcliy and I.Yu. Dmitrieva, "Diagonalization of the "symmetrical" system of the differential Maxwell equations", in Scientific Works of ONAT, no. 2, 2007, pp. 15 -24. (Russian).

[8] A.M. Ivanitckiy, Principles of the Multidimensional Analogous and Discrete Circuit Theory. Educational Textbook. Odessa: ONAT, 2003. (Russian).

[9] V.P. Bakalov, P.P. Vorobiyenko and B.A. Kruk, Circuit Theory. Moscow: Radio i Svyaz (Radio and Communications), 1998. (Russian).

[10] A.G. Kurosh, General Algebra Course, 11th ed. Moscow: Science, 1975. (Russian).

[11] I.Yu. Dmitrieva, Technical Scientific Report. Odessa: ONAT, 2008, the state registered no. 0108u010946; ONAT, 2010, the state registered no. 0109u008009; ONAT, 2012, the state registered no. 0111U009013. (Russian).

[12] I. Dmitrieva, "Specific boundary problems as an analytic investigation of signal transmissions", in Proc. of the 14th Intern. Conf. on Mathematical Methods in Electromagnetic Theory (MMET 2012), August 28 – 30, 2012 , Kharkiv, Ukraine. IEEE Catalog Number CFP 12761 – CDR; ISBN: 978-1-4673-4480-7112 (print), pp. 146-149.

[13] A.N. Kolmogorov and S.V. Fomin, Elements of Functional Analysis and Function Theory, 4th ed. Moscow: Science, 1976. (Russian).

[14] I.B. Vendik, O.G. Vendik and M.A. Odit, "Isotropic metamaterial on the basis of segnetoceramic spherical inclusions", in Solid-State Physics, vol. 51, issue 8, 2009, pp. 1499-1503. (Russian).

[15] D.S. Ryzhenko, Application of Metamaterials in Design of the Waveguide Arrangements with the Microwave Frequency, Abstract of the Ph.D. Thesis. Moscow: Moscow State Technical N.E. Bauman Univ., 2011. (Russian).

[16] V.I. Scheglov, "Calculation of the dynamical permeability of a medium containig magnetic and electric components", in. J. Radioelectronics, no. 7, 2001, pp. 1–7. (Electronic version). (Russian).

[17] C. Caloz and T. Itoh, Electromagnetic Metamaterials:: Transmission Line Theory and Microwave Applications. New York: Wiley, 2005.

[18] G.V. Eleftheriades and K.G. Balmain, Negative Refraction Metamaterials: Fundamental Principles and Applications. New Jersey: Wiley & Sons Inc., 2005.

[19] R. Marques, F. Martin and M. Sorolla, Metamaterials with Negative Parameters: Theory, Design and Microwave Applications. New Jersey: John Wiley & Sons Inc., 2008.

[20] C.J. Tranter, Integral Transforms in Mathematical Physics, 1st ed. London: Methuen and Co. Ltd., New York: John Wiley & Sons Inc., 1951.

[21] A.M. Ivanitckiy and M.V. Rozhnovskiy, "The telegraph equations of homogeneous lines under expofunctional signals", in Scientific Works of ONAT, no. 1, 2009, pp. 56-63. (Russian).

Modern Printing Technologies in Micro- and Nanoelectronics

Olha Sarapulova, Tetiana Kyrychok,
Valentyn Sherstiuk
Institute of Publishing and Printing
National Technical University of Ukraine "Kyiv Polytechnic Institute"
Kyiv, Ukraine
olga-s@voliacable.com

Anatolii Orlov
Faculty of Electronics
National Technical University of Ukraine "Kyiv Polytechnic Institute"
Kyiv, Ukraine
atorlov@rambler.ru

***Abstract*—The trends in microminiaturization in printing are considered and some features of dispersed systems in nano sized scale as well as their applications discussed. Nanoparticles, nanostructures, nanocomposites are regarded as dispersed formations with unique properties due to quantum-sized and surface effects. The summary of photolithography in macro- and microelectronics is given, underlining the disadvantages of it in fabrication of electronic devices due to the light wavelength limits. The detailed analysis of nanotechnology in photography and printing is made, including micro- and nanocontact printing, or imprinting lithography, for what the short history of Graphic Arts development is overviewed as for applications of dispersed materials. Some perspectives of digital micro- and nanoprinting in nanoelectronics, nanophotonics, security printing, active and "intelligent" packaging are outlined, pointing out the importance of nanocomposites of metal oxides and halides for printed luminescent tags for smart food packaging. Nanosized ZnO was considered as a promising substance for printed application to active and "intelligent" packaging as well as nanophotonic devices. The experiments were carried out in order to determine the influence of the state of a packaged product on the properties of ZnO nanofilms. The results open the possibility to combine the properties of the films in development of devices for nanophotonics and smart packaging.**

***Keywords*—printing, electronics, nanoelectronics, nanophotonics, photolithography, nanoprinting, security printing, active packaging, intelligent packaging, smart packaging**

I. Introduction

The importance of printing methods for producing electronic devices can hardly be overestimated. Printing have been employed for fabricating electronics for decades, and with the development of electronic devices in terms of miniaturization the role of printing in producing them increases significantly [1]. After the beginning of printing processes in China and Korea at the end of the first millennium, they have been subjected to many changes and improvements, which resulted in development of classic printing techniques [2]. Other printing techniques, such as ink-jet printing, emerged much later, and found their place in the processes of producing electronics (nanoelectronics, nanophotonics in due course) because of their technological features (they allow low-cost and rapid deposition of nanowires [3]) and the development of materials, including nanomaterials, with possess unique characteristics, for example, conductive inks with metal nanoparticles with lower melting points [4]. This article aims to analyze printing techniques employed in the electronics industry, is devoted mainly to the printing technology in the production of nanoelectronic and nanophotonic devices.

II. Photography in Printing and Printing in Photography

Since 19th century, when photography and photomechanical printing processes were introduced, photography and printing has been connected not only with the necessity to reproduce photographic pictures, but also technologically, as non-digital printing methods required photoprocesses in order to produce printing plates. Modern technology allows us to avoid using these analogue methods in producing printing plates, replacing conventional pre-press and press processes with more convenient digital technologies. However, in some micro- and nanoprinting methods, it is of great value.

If we look at the link between photography and printing from another point of view, we will see that the process of reproducing digital photographs with printing techniques triggers the development of both technologies. A color fidelity issue in printing reproduction of photographs is indeed the most important mover of the progress of both technologies [5].

III. Analogue and Digital Photography and Printing

The invention of photographic processes for the information registration occurred in the middle of the 19th century, and photography was discovered in the first half of the 19th century [2]. Niciphor Nieps (1765-1833), Jack Dagguere (1787-1851), Fox Talbot (1800-1877) may be accounted as the first inventors of photography [6].

Graphic Artists were among the first areas that employed photoregistration in producing illustrative printing plates, using nitrocellulose and then gelatin layers which contained light sensitive argentums halide dispersions. The important consecutive step was the invention of light sensitive material dichromate gelatin layers of water soluble denatured protein collagen with dichromate – the salts of chromic acid (Talbot,

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

1840) [7]. This material had been used in most classic printing techniques; nowadays it is widely used in holography. The fact that almost all widespread photographic layers are dispersed systems limited the resolution of the registering media and registrant images. For electronics high resolution ability of copying layers or photo resists is critical [2]. Processes of photography are in fact of sol-gel technology, widely used for the production of nanomaterials.

IV. Photolithography in Macro- and Microelectronics

Lithography, a technique used to produce nanopatterns, is divided into optical lithography (photolithography) and non-optical [8]. Optical lithography is divided into four main types: electron-beam lithography (EBL), or electron projection lithography (EPL), which uses a beam of electrons as an exposure source; ion-projection lithography (IPL), which uses an ion beam; proximity lithography, where the image is produced on exposure by the shadow of a mask, held in close proximity to the surface of a semiconductor (for instance X-Ray lithography); and extreme ultraviolet lithography (EUVL), which uses an exposure source of wavelength of 11-13 nm [8].

V. The Photo-, X-Ray and Electron-Beam Lithography Limits in Electron Devices Production

The resolution of conventional projection optical lithography, as well as other types of lithography (EBL, IPL, EUVL) is limited by wavelength of energy, produced by the exposure source. The diameter of the minimum element or the minimum feature size that can be obtained is practically equal to the wavelength of light of other energy employed [9]. There are other limits of optical lithography, such as effects of wave diffraction, scattering and interference in a resist, and backscattering from a substrate. In the modern microelectronics production the MOS transistors with sizes 22 and 14 nm are obtained due to application of technological tricks using an ultraviolet laser with a wavelength of 193 nm. Difficulties arise when addressing of such conflicting claims, as nanometer resolution and accuracy of the geometric dimensions of a large area [10].

VI. Nanotechnology in Photography and Printing

Nanotechnology involves fabrication and use of nanosized (less than 100 nm) materials and devices. It is an interdisciplinary area [11]. Printed electronics and photography have also been influenced, changed, and improved by this fast-growing area.

Regarding nanoelectronic devices fabrication, as nanotechnology develops at enormous pace, it has become important for photography to ensure the existence of appropriate resolution. The technique of electron speckle photography was reported, which is claimed to have the spatial resolution in micrometer/nanometer domain [12]. There are also mentions of achieving incredible resolutions of silver halide images by using silver halide grains, J-aggregates of sensitizing dyes, dispersions of dyes and couplers. The combination of silver clusters and nanoparticles with silver halide grains makes considerable contribution to high performance of silver halide imaging materials [13].

The term “nanoprinting” includes methods where patterns are applied to substrates by physical contact or exposure through stamps, nozzles (ink-jet printing in nanoscale), masks (protecting areas from exposure to a printed material) etc. in nanoscale [14]. Apart from the necessity to engineer the required properties of all components, used in nanoprinting, it is critical to develop the methods of large-area, low-cost, rapid deposition of material layers onto the surface of an electronic device.

VII. Micro- and Nanocontact Printing, or Imprinting Lithography. Nanoprinting as Eventually the Best Way for Further Development of Electronics and Photonics

Techniques based on self-assembly and molding are thought to be a convenient and low-cost approach to micro- and nano-fabrication [15]. Soft lithography can produce patterns in micro- and nanoscale, employing contact printing techniques [16].

Nanoimprint lithography is used to fabricate devices for a wide range of applications. As imprint lithography does not use any energetic beams, its resolution is not limited by the optical effects mentioned in section V [17]. Nanotransfer printing technique uses high-resolution stamps coated with thin solid layers of ‘inks’, which are efficiently transferred from the stamp to the substrate by contact printing [18]. This nanofabrication technique is reported to be used for printing electrical contacts directly onto molecular layers at ambient conditions [19]. Usually transfer printing is a dry process that uses a poly(dimethylsiloxane) (PDMS) stamp as a mediator to transfer nanomaterials from one substrate to another [16]. Besides, another printing technique is mentioned – a poly(methyl methacrylate) (PMMA)-mediated nanotransfer printing technique. It is based on utilization of PMMA film as a mediator, on which nanomaterials are loaded [20]. In microtransfer molding a drop of liquid prepolymer is applied to the PDMS mold, which is then contacts with a substrate and irradiation or heat is applied in order to cure the liquid precursor to a solid [16]. Micromolding in capillaries (MIMIC), where the relief structure in the PDMS mold forms a network of empty channels, is claimed to be capable of patterning a broader range of materials than is photolithography [16]. It is reported that the stamps for nanoimprint lithography can have integrated heater [21]. There are also mentions of roll stamps for so called RNIL (roller nanoimprint lithography) [22].

New types of nanolithography are being developed in order to achieve nanofabrication in less complicated, cheaper and quicker ways. «Dip-pen» nanolithography (DPN) transfers ink on a sharp object (atomic force microscope probe tip) onto a solid “paper” substrate (for example, Au) via capillary forces. DPN does not exploit any resists or stamps, which sets this technique aside from conventional nanolithography methods [23]. Polymer pen lithography was developed from DPN as a low-cost, high-throughput scanning probe lithography method which employs a soft elastomeric tip array to deliver inks to a

surface [24]. Acceptable for mass production performance can be achieved with the use of multi-probe heads.

Step and flash imprint lithography (SFIL) is reported to be a high throughput, low cost technique which is based on a low viscosity, photocurable liquid and a UV transparent rigid template rather than flexible PDMS mold [25]. There is a method called «Particle Replication in Nonwetting Templates», which uses low-surface energy, chemically resistant fluoropolymers as molding materials [26].

There is an alternative micro-fabrication process called LIGA, which combines deep X-ray lithography, plating-through-mask and molding, and enables precise manufacture of high-aspect-ratio microstructures from hundreds to thousands of micrometers in thickness [27]. The most important application is the creation of MEMS devices [28].

Nanoimprint lithography (NIL) is successfully used for manufacturing nanoscale photodetectors, silicon quantum-dots, quantum-wires, and ring transistors [17]. Microelectronics and biotechnology applications such as microelectrode arrays, well-defined polymer patterns, memory devices, and biosensors can be fabricated with mentioned above techniques [29].

Ink-jet technology can also be considered for electronic manufacturing, using metal nanoparticles for direct patterning of micro- and nanoscale conductive metal lines [30]. In ink-jet deposited System-in-Package the package contains components encapsulated with resin mold [31].

VIII. Perspectives of Digital Micro- and Nanoprinting in Nanoelectronics, Nanophotonics, Security Printing, Active and Intelligent Packaging

Nanotechnology can be used in printing technology, for example, in optical lithography (EBL, IPL, proximity lithography, EUVL), non-optical lithography methods (such as contact nanoprinting, NIL, SFIL etc.), and ink-jet printing for producing microelectronic and photonic devices, security elements in forgery-proof printed products etc.

A. Perspectives of Digital Micro- and Nanoprinting in Security Printing

Nanotechnology is widely used in printing materials. For example, inks for ink-jet printing containing silver, gold, and copper nanoparticles are used to fabricate printed electronics, devices for photonics etc. Zinc oxide is a potential pigment material for paper coating to impart brightness and better printing properties [32]. These possibilities can be employed to bestow special properties onto paper for security printing, or nanosized functional materials can be delivered onto the surface of the substrate by nanoprinting techniques; they can possess conductive properties, which can be exploited as a level of safety measures in security printing.

B. Perspectives of Digital Micro- and Nanoprinting in Active and Intelligent Packaging

The use of multifunctional nanosized materials has a great potential in active or intelligent packaging [33]. Nanoscale materials play a significant role from packaging of electronic and photonic devices to packaging of foods [34]. Active and "intelligent" systems, applied to packaging materials via printing techniques, can track and alert to the processes that occur in packaged food during its storage by changing the intensity of luminescence while being exposed to light or the electric potential or change their magnetic properties in case of usage of magnetic-luminescent nanocomposites, therefore serving as "intelligent" systems for packaging. The usage of ZnO in "smart" packaging is justified by its ability to serve as UV light and ionizing radiation barrier and a biosensor to identify proteins [35]. There are examples of the use of nano-ZnO in the antimicrobial nanocomposite films [36]. The significant advantage of ZnO is the fact that it is considered to be a safe food additive, which enables the usage of it for food packaging [37]. Use of photovoltaic and piezoelectric properties of ZnO structures provides additional opportunities for the creation of printed sensors and energy harvesting devices. The investigation and modeling of interaction of luminescent hybrid nanocomposites with packaged products and compounds indicating food spoilage or decay allowed developing multifunctional nanocomposites for active and "intelligent" packaging with anti-microbial, forgery-proof, and product quality detection properties.

The research was carried out with ZnO nanosized films as model systems for nanophotonic devices and printed sensors, which can serve as spoilage indicators at the same time. Amines such as dopamine, tyramine, tryptamine, and histamine were used as model substances which emerge in protein products during storage. The interaction between ZnO and the amines was studied by recording the luminescence and absorption spectra of the films. The concentration of ZnO nanoparticles in solutions applied to substrates was $2 \cdot 10^{-3}$ mol/l, the concentration of the amines varied.

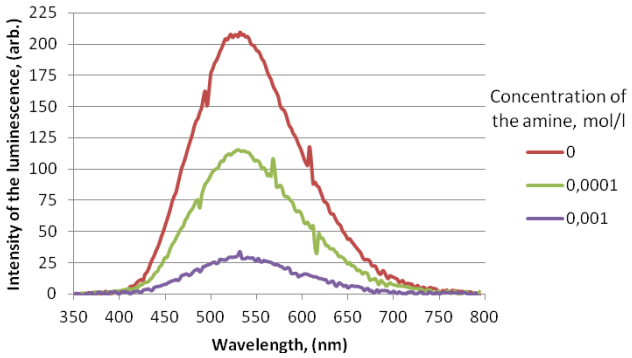


Fig. 1. The luminescence spectra of tryptamine with ZnO nanoparticles.

As can be seen from fig. 1, the quenching of the luminescence of ZnO nanofilm by tryptamine is quite significant. Similar results were obtained for the other model amines. The fact of quenching of the luminescence can be confirmed and explained with the analysis of the absorption spectra of the amines. The absorption spectra of tryptamine with ZnO nanoparticles (fig. 2), as well as the ones of dopamine and tyramine, suggest that with the increase of the concentration of tryptamine in the system, it absorbs light in UV region of the spectrum much more distinctively, therefore blocking light from excitation of the luminescence of ZnO nanoparticles.

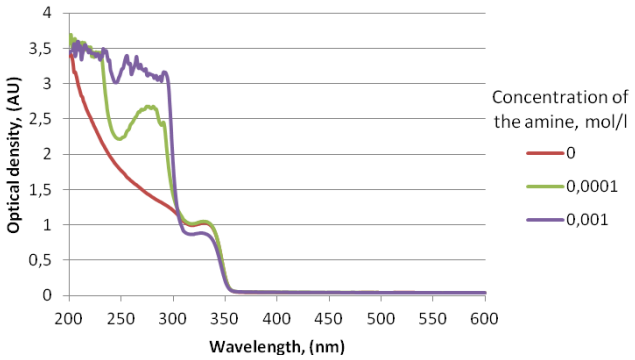


Fig. 2. Absorption spectra of tryptamine with ZnO nanoparticles.

Similar situation can be observed from experimental data with luminescence spectra of histamine with ZnO nanoparticles. However, distinctive quenching of the luminescence at the concentration of histamine 10^{-3} mol/l can not be explained by intensive absorption of the system in the UV region of the spectrum (fig. 3) as the intensity of absorption decreases at concentration of the amine 10^{-3} mol/l. Formation of non-luminescent complexes of histamine with ZnO nanoparticles can be accounted for the decrease of the intensity of the luminescence.

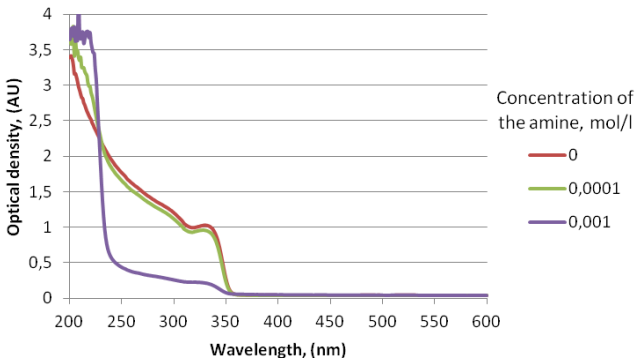


Fig. 3. Absorption spectra of histamine with ZnO nanoparticles.

Fig. 4 shows the combined data of changes in luminescence of ZnO nanofilms in the presence of the model amines, where the quenching of the luminescence of ZnO nanofilm by dopamine appears to be the most significant.

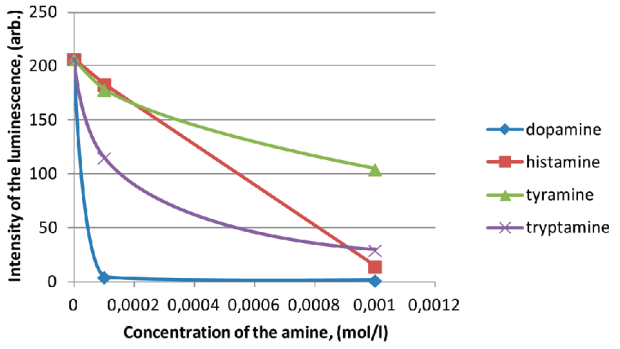


Fig. 4. The change of the luminescence intensity of nano-ZnO films depending on concentration of the amines.

The results show that the intensity of the luminescence of printed films with ZnO nanoparticles and amines decreases with the increase of concentration of the amines. This fact enables the possibility of using such films for detection of freshness of a packaged product alongside with exploiting antimicrobial, photovoltaic and piezoelectric properties of ZnO nanostructures.

Nanocomposites with CdS, CdSe, and similar ones may also be accounted as the components for materials with electroluminescent properties for applications in active informational packaging [38].

It is important to examine the possibility of usage of the components in the form of nanophotonic systems with electro-, magnetic and catalytic properties to account for the interaction of such systems with the components of inks, as well as determining whether these components and substances of packaged products interfere with the performance of nanophotonic and photocatalytic systems. If the interference occurs, there will be the need to find some ways to overcome such problem by controlled changing of the contents of the systems or the inks or the possible products to be packaged.

C. *Perspectives of Digital Micro- and Nanoprinting in Nanoelectronics, Nanophotonics*

Micro- and nanoprinting are relatively new and fast-growing areas. They develop according to the demands of modern nanoelectronics and nanophotonics, which means that they are required to achieve appropriate resolution and quality of fabricated patterns, being at the same time low-cost and rapid for large-area application. There are mentions that these challenging requirements might be achieved with microcontact printing or imprint lithography or by the combined use of such approaches with ink-jet printing [11]. The combination of different techniques can possess advantages of several techniques, which makes such an approach worth consideration. It is reported, however, that there are no fully developed viable manufacturing techniques for organic electronics, which leaves a wide gap in nanoprinting methods to be filled in due course. The situation is complicated by the fact that organic semiconductors now have unsatisfactory performance and reliability [14]. Therefore, optimized nanopatterning methods and materials for them are yet to be developed, investigated and improved.

CONCLUSIONS

In this paper, printing methods employed in fabrication of electronics were analyzed. The development of printing technology techniques for manufacturing nanoelectronic and nanophotonic devices is happening at a very high pace; however, certain difficulties and problems still remain unsolved. As printing in micro- and nanoscale is directly connected to photography, it is critical to achieve appropriate resolution in photographic technology. Another important limitation of minimum feature size is the fact that photolithographic methods are limited by the wavelength of an exposure source. This limitation resulted in invention of EBL, IPL, and EUVL. However, the reduction of wavelength was not enough to obtain nanosized printed elements for electronic

devices, so it triggered the development of NIL and similar nanolithographic techniques, the number of which is constantly increasing. New materials are being developed too. The combinations of different techniques are considered perspective in enabling low-cost, large-area, and rapid fabrication of nanoelectronic and nanophotonic devices. The research was conducted with ZnO nanofilms and model amines which emerge in packaged products during storage and indicate spoilage. It was determined that with the changes in intensity of the luminescence of such systems it is possible to exploit luminescent properties of ZnO nanostructures alongside with its antimicrobial, photovoltaic and piezoelectric properties for printed sensors and nanophotonic devices.

ACKNOWLEDGEMENTS

The authors thank Dr. Vlasova and Dr. Markitan from O.O. Chuiko Institute of Surface Chemistry of the National Academy of Sciences of Ukraine for bioamines to use in the experiments and Dr. Shvalagin from L.V. Pisarzhevsky Institute of Physical Chemistry of the National Academy of Sciences of Ukraine for the help in the spectra registration.

REFERENCES

[1] J. Perelaer, P. J. Smith, D. Mager, et al. "Printed electronics: the challenges involved in printing devices, interconnects, and contacts based on inorganic materials", J. Mater. Chem., 2010, vol. 20, pp. 8446-8453.

[2] O. V. Gumeniuk, V. P. Sherstyuk, "Nanotechnologies in Printing and Packaging", IC, 2010, vol. 3, pp. 32-41.

[3] J. U. Park, M. Hardy, S. J. Kang, et al., "High-resolution electrohydrodynamic jet printing", Nature Mater., 2007, vol. 6, pp. 782-789.

[4] H. Ishibashi, "Preparation of metal nano-particles and their application to circuit pattern formation by ink-jet printing", Journal of the Imaging Society of Japan, 2008, vol. 47, no. 6, pp. 138-146.

[5] H. Wilhel, "A 15-year history of digital printing technology and print permanence in the evolution of digital fine art photography – from 1991 to 2006", Final Program and Proceedings: NIP, pp. 308-315, September 2006.

[6] K. V. Chibisov, "The nature of photosensitivity", Moscow, 1980.

[7] Kosar J., "Light-sensitive systems", John Wiley and Sons, Inc., N.Y.-L.-S, 1965.

[8] T. Ito, S. Okazaki, "Pushing the limits of lithography", Nature, 2000, vol. 406, pp. 1027-1031.

[9] M. Wautelet, "Scaling laws in the macro-, micro- and nanoworlds", Eur. J. Phys., 2001, vol. 22, pp. 601–611.

[10] G.Pashkevich, A. Orlov, O.Bogdan "Use of "Electromask" pattern generator and image repeater for large-area subwavelenght nanolitography", Electrotechnica & Elektronica E+E, vol. 47, No 5-6, Union of Electronics, Electrical Engineering and Telecommunications (CEEC), Sofia, Bulgaria, ISSN 0861-4717 2012, pp. 18-20

[11] B. D. Gates, Q. Xu, M. Stewart, et al., "New approaches to nanofabrication: molding, printing, and other techniques", Chem. Rev., 2005, vol. 105, pp. 1171-1196.

[12] F. Chiang, "Some applications of electron speckle photography", unpublished

[13] T. Tani, "Nanoparticles and nanotechnology in silver halide imaging", J. Disper. Sci. Technol., 2004, vol. 25, no. 4, pp. 375-388.

[14] E. Menard, M. A. Meitl, Yugang Sun, et al., "Micro- and nanopatterning techniques for organic electronic and optoelectronic systems", Chem. Rev., 2007, vol. 107, no. 4, pp. 1117-1160.

[15] X. Zhao, Y. Xia, and G. M. Whitesides, "Soft lithographic methods for nano-fabrication", J. Mater. Chem., 1997, vol. 7, pp. 1069-1074.

[16] Y. Xia, G. M. Whitesides, "Soft lithography", Annu. Rev. Mater. Sci., 1998, vol. 28, pp. 153-184.

[17] S. Y. Chou, P. R. Krauss, and P. J. Renstrom "Nanoimprint lithography", J. Vac. Sci. Technol., 1996, vol. 14, no 6, pp. 4129-4133.

[18] J. Zaumseil, M. A. Meitl, J. Hsu, et al., "Three-dimensional and multilayer nanostructures formed by nanotransfer printing", Nanoletters, 2003, vol. 3, no. 9, pp. 1223-1227.

[19] Y.L. Loo, D. V. Lang, J. A. Rogers, et al., "Electrical contacts to molecular layers by nanotransfer printing", Nanoletters, 2003, vol. 3, no. 7, pp. 913-917.

[20] L. Jiao, B. Fan, X., et al., "Creation of Nanostructures with Poly(methyl methacrylate)-Mediated Nanotransfer Printing", J. Am. Chem. Soc., 2008, vol. 130 (38), pp 12612-12613.

[21] M. Tormena, R. Malureanua, R. H. Pedersenb, et al., "Fast thermal nanoimprint lithography by a stamp with integrated heater", Microelectron. Eng., 2008, vol. 85, no. 5-6, pp. 1229–1232.

[22] H. J. Lim, K. B. Choia, G. H. Kima, et al., "Roller nanoimprint lithography for flexible electronic devices of a sub-micron scale", Microelectron. Eng., 2011, vol. 88, no. 8, pp. 2017–2020.

[23] R. D. Piner, J. Zhu, F. Xu, et al., ""Dip-Pen" Nanolithography", Science, 1999, vol. 283, no 661, pp. 661-663.

[24] F. Huo, Z. Zheng, G. Zheng, et al., "Polymer Pen Lithography", Science, 2008, vol. 321, no. 5896, pp. 1658-1660.

[25] B. K. Long, B. K. Keitz, and C. G. Willson, "Materials for step and flash imprint lithography", J. Mater. Chem., 2007, vol. 17, pp. 3575-3580.

[26] J. P. Rolland, B. W. Maynor, L. E. Euliss, et al., "Direct fabrication and harvesting of monodisperse, shape- specific nanobiomaterials", 2005, J. Am. Chem. Soc., vol. 127 (28), pp. 10096-10100.

[27] N. H. Loh, S. B. Tor, B. Y. Tay, et al., "Micro powder injection molding of metal microstructures", Mater. Sci. Forum, 2003, v. 426-432, pp. 4289-4294.

[28] Y. Qin, A. Brockett, Y. Ma, et al., "Micro-manufacturing: research, technology outcomes and development issues", Int. J. Adv. Manuf. Tech., 2009, vol. 47 no. 9-12, pp. 821-837.

[29] D. G. Choi, H. K. Yu, and S. M. Yang, "2D nano/micro hybrid patterning using soft/block copolymer lithography", Mater. Sci. Eng., 2004, vol. 24, pp. 213-216.

[30] K. Murata, J. Matsumoto, A. Tezuka, et al., "Super-fine ink-jet printing: toward the minimal manufacturing system", Microsyst. Technol., 2005, vol. 12, no. 1-2, pp. 2-7.

[31] V. Pekkanen, M. Mantysalo, K. Kaija, et al., "Utilizing inkjet printing to fabricate electrical interconnections in a System-in-Package", Microelectron. Eng., 2010, vol. 87, no. 11, pp. 2382-2390.

[32] V. Prasada, A. J. Shaikha, A. A. Kathea, et al., "Functional behaviour of paper coated with zinc oxide-soluble starch nanocomposites", J. Mater. Process. Tech., 2010, vol. 210, no. 14, pp. 1962-1967.

[33] B. S. Sekhon, "Food nanotechnology – an overview", Nanotechnology, Science and Applications, 2010, vol. 3, p. 1-15.

[34] S. Otles, B. Yalcin, "Smart food packaging", LogForum, 2008, vol. 4, no. 3, pp. 1-7.

[35] J. Seol, G. Jeon, S. Jang, et al., "Preparation and properties of poly(propylene carbonate) and naosized ZnO composite films for packaging applications", J. Appl. Polym. Sci., 2011, vol. 122, no. 2, pp. 1101-1108.

[36] A. Emamifar, M. Kadivar, M. Shahedi, et al., "Preparation and evaluation of nanocomposite LDPE films containing Ag and ZnO for food-packaging applications, Adv. Mater. Res., 2010, vol. 129-131, pp. 1228-1232.

[37] X. Li, Y. G. Xing, L. W. Li, et al., "Antibacterial and physical properties of poly(vinyl chloride)-based film coated with ZnO nanoparticles", Int. J. Food Sci. Tech., 2010, vol. 16, pp. 3225-3232.

[38] V. P. Sherstyuk, O. O. Sarapulova, V. V. Shvalagin, "Luminescent hybrid nanocomposites and prospects of molecular and nanophotonic systems in modern packaging and printing", Book of Abstracts of the 3-International Symposium "Molecular Photonics" (Repino, St. Peterburg, Russia), 2012, p. 79.

Modeling of the Dielectric Barrier RF Discharge at Atmospheric Pressure

Chaplinskiy R.U.
Department of electronic devices
Faculty of electronics
NTUU"KPI"
Kyiv, Ukraine
chapok86@gmail.com

***Abstract* — one dimensional hybrid modeling of the dielectric barrier radio frequency (DBRF) discharge at atmospheric pressure in argon is performed. Electrons in this model are described in kinetic way and heavy species in hydrodynamic way. Electron density, capacitive power deposition and electrons temperature distributions in the discharge gap for low and high current modes of the DBRF discharge as well as volt-ampere characteristics of DBRF discharge are calculated. The differences between radio-frequency discharge with bare electrodes and DBRF discharge are established.**

Keywords — dielectric barrier RF discharge, non-thermal plasma, atmospheric pressure.

I. INTRODUCTION

Non-thermal plasma is widely used in different technologies like manufacturing of micro- and nano- electronic devices, coating and etching technologies, surface cleaning and activation [1-3]. In industrial application, non-thermal plasma is generated mostly in the low pressure discharges. However, this method has such disadvantages as using the expensive vacuum technique, low rates of coating, etching and reactions. Besides, in this method the size of the treated surface is limited by a volume of the vacuum chamber.

The most promising alternative for low pressure plasma is generation of large volume non-thermal plasma in atmospheric pressure discharges. Atmospheric pressure non-thermal plasma can be generated in such discharges as corona discharge, dielectric barrier discharge and etc. Using atmospheric pressure plasma would even expand the scope of material processing. Nevertheless, discharges with non-thermal plasma at atmospheric pressure are not stable. They are prone to shrinking and transformation into discharges with thermal plasma. Temperature of heavy species in thermal plasmas can be greater than 5000 K. Because of this, the thermal plasma cannot provide low temperature technological processes. From all atmospheric pressure discharges, the dielectric barrier radio frequency discharge (DBRF) is more suitable for non-thermal plasma generation. This is due to low temperature of neutral species homogeneous and large volume of plasma, and ability to treat large surfaces.

The experimental study of the RF discharge at atmospheric pressure was performed by different groups [4, 5]. However, many parameters of the RF discharge at atmospheric pressure (AP) are difficult to be measured. These parameters are the following: mass fraction and density of heavy plasma species, concentration, energy, and temperature of electrons. Even measuring of electrical characteristics of the RF discharge at atmospheric pressure is quite difficult because of electrical disturbances around the RF discharge device. These electrical disturbances can be much greater than the measured signal. That is why the modeling of this discharge is necessary.

A few models of the capacitively coupled plasma (CCP) generated in the RF discharges at atmospheric pressure are developed [6]. Most of them deal separately with describing plasma dynamic, dominant physical and chemical mechanisms of CCP. The most general one-dimensional hybrid model was developed for studying an atmospheric pressure RF discharge with bare electrodes in helium in [7]. The modeling [7] allows to establish that low current α- and high current γ-modes of the AP RF discharge are different not only in the intensity of gas ionization, but in its localization: the ionization take place mostly at the middle of the discharge gap in α-mode and in two places near electrodes in the γ-mode.

In this work, the hybrid model of the AP DBRF discharge has been developed. It is used to determine electron density, capacitive power deposition and electron temperature in the discharge gap during the period of the applied voltage for the two modes of the DBRF discharge in argon. In presented model each electrode covered with dielectric plate.

Presented information can be useful for the development of material processing devices, in which AP DBRF discharge is used for non-thermal plasma generation.

II. MATHEMETICAL MODEL

The presented model can be characterized as a hybrid one. It means that in this model electrons are considered in the kinetic way and heavy particles in the hydrodynamic way. Electrons are treated in the kinetic way by two continuity equations: one of them – for electron density and another for electron energy. Heavy particles are modeled only by the mass conservation equation.

Joule heating, fluid motion of heavy species, influence of a matching box are neglected. Only the simple argon plasma chemistry (reactions with molecular argon are neglected) is assumed, gas temperature is a constant input value and the Maxwelian electron energy distribution function (EEDF) is assumed.

A. Fundamental equations

The electron density (n_e) and mean electron energy (n_ε) are computed by solving a pair of drift-diffusion equations for

the electron density and mean electron energy. Convection of electrons due to fluid motion is neglected:

$$\frac{\partial n_e}{\partial t} + \nabla(-n_e \mu_e E - D_e \nabla n_e) = R_e$$

$$\frac{\partial n_\varepsilon}{\partial t} + \nabla(-n_\varepsilon \mu_\varepsilon E - D_\varepsilon \nabla n_\varepsilon) + E\Gamma_\varepsilon = R_\varepsilon,$$

where E is the electric field and Γ_ε is electron energy flux. The electron source (R_e) and the energy loss due to inelastic collisions (R_ε) will be given below. The electron diffusivity (D_e), energy mobility (μ_ε) and energy diffusivity (D_ε) are computed from the electron mobility (μ_e) using:

$$D_e = \mu_e T_e, \ \mu_\varepsilon = \frac{5}{3}\mu_e, \ D_\varepsilon = \mu_\varepsilon T_e,$$

where N is gas density, ε is electrons energy (eV) and $f(\varepsilon)$ is the electron energy distribution function, T_e is electron temperature. The electron mobility can be calculated by the next equations:

$$\mu_e N = -\frac{\gamma}{3}\int_0^\infty \frac{\varepsilon}{\tilde{\sigma}_m}\frac{\partial f(\varepsilon($$

$$\tilde{\sigma}_m = \sigma_m + \frac{\bar{\nu}_i}{N\gamma\varepsilon^{1/2}},$$

where σ_m is momentum transfer cross section, $\bar{\nu}_i$ is electron net production frequency.

The source coefficients in the above equations are determined by the plasma chemistry using reaction rate coefficients. Let us assume that there are *M* numbers of reactions, which contribute to the growth or decay of the electron density, and *P* number inelastic electron-neutral collisions. In general *P>>M*. In the case of rate coefficients the electron source term is given by:

$$R_e = \sum_{j=1}^{M} x_j k_j N_n n_e,$$

where x_j is the mole fraction of the target species for reaction j, k_j is the rate coefficient for reaction (m^3/s), and N_n is the total density of the neutral species ($1/m^3$). The electron energy loss is obtained by summing the collisional energy loss over all reactions:

$$R_\varepsilon = \sum_{j=1}^{P} x_j k_j N_n n_e \Delta\varepsilon_j,$$

where ε_j is the energy loss from reaction j (eV). The rate coefficients may be computed from cross section data by the following integral:

$$k_j = \gamma\int_0^\infty \varepsilon\sigma_k(\varepsilon) f(\varepsilon) d\varepsilon,$$

where $\gamma = \sqrt{\frac{2q}{m_e}}$ ($C^{1/2}/kg^{1/2}$), q is electron charge (C), m_e is the electron mass (kg), and σ_k is the collision cross section (m^2).

For non-electron species, the following equation is solved for the mass fraction of each species.

$$\rho\frac{\partial \omega_k}{\partial t} + \rho\nabla(u\nabla)\omega_k = \nabla j_k + R_k,$$

where ρ is gas density, ω and j are the mass fraction and the current density of the k species, u is a flow velocity, R is a source term.

The electrostatic field is computed by the Poisson equation:

$$-\nabla\varepsilon_0\varepsilon\nabla V = \rho,$$

where ε_0 and ε are dielectric permittivity of vacuum and considered material, V is sinusoidal voltage applied to the exterior boundary of the dielectric plates.

The space charge density ρ is automatically computed based on the plasma chemistry specified in the model using the formula:

$$\rho = q(\sum_{k=1}^{N} z_k n_k - n_e),$$

where n_k is concentration of the charged particles, z_k is a tuning parameter.

B. Bondary conditions

Electrons are lost on the walls due to random motion within a few mean free paths to the wall and generated due to secondary emission, which results in the following boundary condition for the electron density flux:

$$-n\Gamma_e = \left(\frac{1}{2}v_e\right)n_e - \sum_p \gamma_p(\Gamma_p n) \quad (1)$$

and the electron energy flux:

$$-n\Gamma_\varepsilon = \left(\frac{1}{2}v_e\right)n_e - \sum_P \varepsilon_p\gamma_p(\Gamma_p n) \quad (2)$$

The second term on the right hand side of equation (1) is the generation of electrons due to secondary emission effects, γ_p being the secondary emission coefficient. The second term in equation (2) is the secondary emission energy flux, ε_p being the mean energy of the secondary electrons.

Ions are lost on the wall due to surface reactions. The fact that the electric field is directed towards the wall is described by the following equation:

$$-nj_k = M_\omega R_k + M_\omega c_k Z\mu_k(En)[Z_k\mu_k(En) > 0],$$

where $n\,j_k$ is a normal component of flux of the species *k*, M_ω is molecular weight, Z is a charge, R_k is a source term, c_k is molar concentration, μ_k is mobility.

Surface charge accumulation is added to the dielectric surface by the way of the following boundary condition:

$$n(D_1 + D_2) = \rho_s,$$

where ρ_s is the surface charge density, which is computed by solving the following distributed ordinary differential equation on the surfaces:

$$\frac{d\rho_s}{dt} = nJ_i + nJ_e,$$

where nJ_i is the normal component of the total ion current density to the wall, and nJ_e is the normal component of the total electron current density to the wall.

The discharge is driven by a sinusoidal voltage (V) applied to the exterior boundary of the dielectric plates:

$$V = V_0 \sin(2\pi f t),$$

where V_0 is the amplitude of applied voltage (V), f is excitation frequency, t is time.

C. Plasma chemistry

The list of plasma chemistry reactions considered in the presented model is shown in Tab. I.

In the CCP reactor the electron density and density of excited species is relatively low, so stepwise ionization is not as important as in the high density discharges. In addition to volumetric reactions, the surface reactions which are shown in the Tab. II are considered.

When a metastable argon atom makes contact with the wall, it turns back to its ground state with some probability (the sticking coefficient).

D. Numerical scheme

The CCP is sustained by applying a sinusoidal voltage across a small gap (1 mm in this case) filled with AP argon. Because the mechanism of the power dissipation in a CCP

TABLE I. TABLE OF MODELED REACTIONS

REACTION	FORMULA	TYPE	Δε(eV)[a]
1	e+Ar=>e+Ar[b]	Elastic	0
2	e+Ar=>e+Ars[c]	Excitation	11.5
3	e+Ars=>e+Ar	Superelastic	-11.5
4	e+Ar=>2e+Ar+[d]	Ionization	15.8
5	e+Ars=>2e+Ar+	Ionization	4.24
6	Ars+Ars=>e+Ar+Ar+	Penning ionization	
7	Ars+Ar=>Ar+Ar	Metastable quenching	

[a] Δε is the energy lost in the concrete reaction;
[b] Ar is argon atom in the ground state;
[c] Ars is an excited argon atom;
[d] Ar+ is a positive ion of argon.

TABLE II. TABLE OF SURFACE REACTIONS

REACTION	FORMULA	STICKING COEFFICIENT
1	Ars=>Ar	1
2	Ar+=>Ar	1

reactor is highly nonlinear and occurs at multiple different frequencies, electrostatic potential cannot be solved in the frequency domain. The model must be transient and must allow to reach a periodic steady state solution. This condition is reached after a suitable number of RF cycles (1000 cycles in this case).

The presented mathematical model was solved by the finite element method in COMSOL Multyphysics software [8]. The electron mobility was calculated in BOLSIG+ software for each amplitude of applied voltages [9].

E. Input data

In presented model taking into account that the both metal electrodes covered with a dielectric plate which has 1 mm thick, and they are made with material which has dielectric permittivity (ε) equal to 4.7. Whole input parameters are shown in the Tab. 3.

III. RESULTS AND DISCUTION

In the gas discharges electrical energy of the power source transforms into the kinetic energy of the free electrons which transfers it to the gas atoms by elastic and inelastic collisions. That is why free electrons are the most important particles in the gas discharges. Most of the discharge parameters can be determined from the distribution of the electron density, energy and temperature in the discharge gap. That is why these parameters were modeled in this work.

It is well known that the RF discharge has two operation modes (α- and γ-) at low pressure as well as at atmospheric pressure [10]. The relatively low current mode is called α-mode, as the first Townsend coefficient of bulk ionization. This is because of the ionization mechanism which takes place mostly in the middle of the discharge gap. This mode can be characterized as oscillation of the free electrons cloud from one electrode to another with the conformity to the electric field applied to the electrodes. This means that averaged electron density must be higher at the middle of the discharge gap. It is shown on Fig. 1 as a result of the modeling for input voltage equal to 100 V.

TABLE III. INPUT PARAMETERS

Input parameters	Values of parameters
V_0	150-350 [V]
f	13.56 [MHz]
Pressure	750 [Torr]
T	400 [K]
μ_e	1.266E+25 – 8.593E24 [1/(V*m*s)]

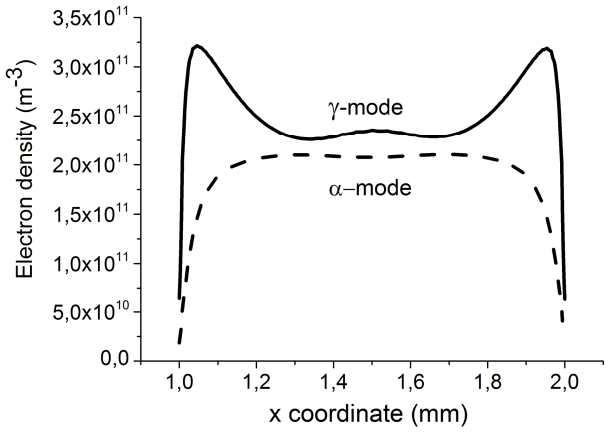


Fig. 1. Electron density averaged for the period of RF cycle, in the discharge gap at α and γ-modes of the dielectric barrier RF discharge.

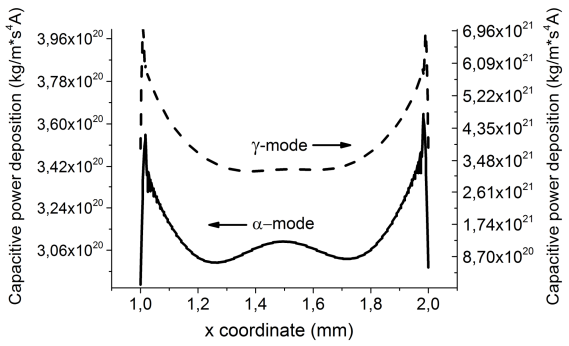


Fig. 2. Capacitive power deposition, averaged for the period of RF cycle, in the discharge gap for α- and γ-modes of the dielectric barrier RF discharge. Input voltage in α – mode is equal to 100 V and in γ-mode it is equal to 350 V.

The second mode of the RF discharge is called γ-mode also in accordance with the third Townsend coefficient (γ).
The γ describes secondary emission from the electrode surface. Mechanism of γ-mode of the RF discharge is similar to the classic glow discharge at low pressure. Positive ions hit the surface of the dielectric barrier on the electrode, which has more negative potential at this moment, and push electrons from this surface to the discharge gap. These electrons are accelerated in the cathode fall region and ionize neutral atoms, when their energy becomes enough for this. This means that ionization process in γ-mode takes place near the electrodes. Distribution of the electron density along the discharge gap in the γ-mode is shown in Fig. 1. This mode is characterized by the higher power dissipation in the discharge gap which can be seen in Fig. 2 and Fig. 3. They show that electron energy and electron temperature are higher in the γ-mode. In Fig. 2 distribution of the capacitive power deposition has its maximum near dielectric surface due to fact that active species lost their energy when they touch dielectric. Experimental study of the RF discharge with bare electrodes [10] shows that volt-ampere characteristics of the discharge are changed when α-γ transition takes place. That happens because impedance of the RF discharge in γ-mode is more resistive than the one in α-mode.

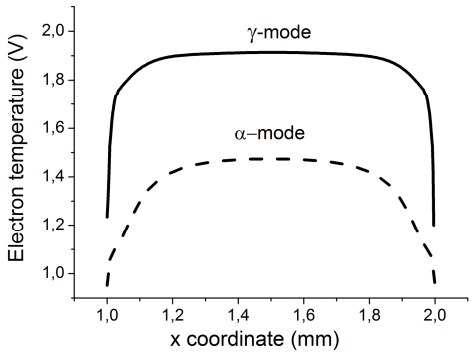


Fig. 3. Electron temperature averaged for the period of RF cycle in the discharge gap in α- and γ-modes of a dielectric barrier RF discharge. Voltage applied to the discharge gap.in α- mode is equal to 100 V and in γ-mode to 350 V.

The modeled volt-ampere characteristic (VACh) of the DBRF discharge is presented on Fig. 4. Fig. 4 shows that there is no signal distinction in VACh of the RF discharge in α- and γ-modes. This happens because a RF discharge capacitance is connected in series with the dielectric barriers having much greater capacitance. This data well agrees with experimental investigation [11].

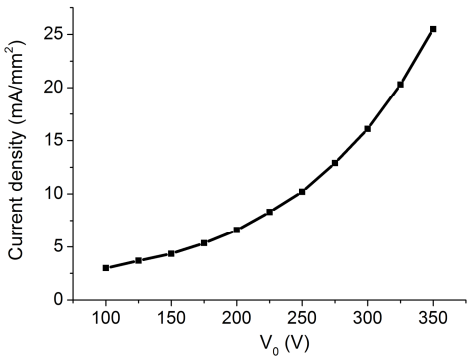


Fig. 4. Volt-ampere characteristic of the AP DBRF discharge.

IV. CONCLUSION

Modeling of the argon CCP of the dielectric barrier RF discharge by a hybrid model is performed. Distribution of the electron density, capacitive power deposition and electron temperature in the discharge gap for two modes of the DBRF discharge and volt-ampere characteristic of the DBRF are calculated. These results show main difference between the RF discharge with bare electrodes and the DBRF discharge at atmospheric pressure. This difference is almost linear volt-ampere characteristic of the DBRF discharge.

The presented model can be improved by calculation of the heavy species transport, matching box influence and considering flow dynamic. It will be a topic of the future works.

REFERENCES

[1] J. R. Roth, *Industrial Plasma Engineering*. Bristol: Institute of Physics Publishing, vol. 1, 1997.

[2] A. Fridman, *Plasma Chemistry*. New York: Cambridge university press, 2008.

[3] A. Bogaerts, E. Neyts, R. Gijbels, J. V. Mullen, "Gas discharge plasma and their applications," *Spectrochemica Acta Part B,* vol. 57, pp. 609-658, 2002.

[4] J. Laimer, S. Haslinger, W. Meissl, J. Hell and H. Stori, "Investigation of an atmospheric pressure radio frequency capacitively coupled plasma jet," *Vacuum*, vol. 79, pp 209-214, 2005.

[5] W. Zhu, B. Wang, Z. Yao and Y. Pu, "Discharge characteristics of an atmospheric pressure radio-frequency plasma jet," *J. Phys. D: Appl. Phys.,* vol. 38, p. 1396, 2005.

[6] X. Yuan and L. L. Raja "Computational study of capacitively coupled high-pressure glow discharges in helium," *IEEE Trans. Plasma Sci.*,. vol. 31, p 495, 2003.

[7] J. J. Shi and M. G. Kong, "Mode characteristics of radio-frequency atmospheric pressure glow discharges," *IEEE Trans. Plasma Sci.,* vol. 33, p. 624, 2005.

[8] Comsol Multiphysics 4.2., Herfordshire, AL10 9AB: COMSOL Ltd, 2007.

[9] G. J. M. Hagelaar, L. C. Pitchford, "Solving the Boltzmann equation to obtain electron transport coefficients and rate coefficients for fluid models," *Plasma Sources Sci. Technol.,* vol. 14, no. 4, pp. 722–733, 2005.

[10] Y. P. Raizer, M. N. Shneider, and N. A. Yatsenko, *Radio-Frequency Capacitive Discharges*. Boca Raton, FL: CRC Press, 1995.

[11] J. J. Shi, D. W. Liu, M. G. Kong, "Effects of dielectric barriers in radio-frequency atmospheric glow discharges," *IEEE Trans. Plasma Sci.,* vol. 35, pp. 137-142, 2005.

Timing Yield and Reliability Improvement of Carbon Nano-tube FET Based Digital Circuits with Statistical Driven Correlation-aware Placement

Amir Jalali
Department of Computer Engineering and Information Technology
Amirkabir University of Technology
Tehran, Iran
amirjalali@aut.ac.ir

Hossein Pedram
Department of Computer Engineering and Information Technology
Amirkabir University of Technology
Tehran, Iran
Pedram@aut.ac.ir

***Abstract*—Carbon Nano-tube Field Effect Transistor (CNFET) is one of the most promising successors of CMOS technology because of its superb electrical features. Although these features are proper for implementing in various practical circuits, CNFET-based circuits will encounter enormous fabrication problems due to their size. Two of the most challenging problems are timing yield and reliability reduction. Consequently methods for improving robustness of CNFET-based circuits should be conducted. Considering these problems, in this paper, the statistical model of reliability and timing yield of CNFET-based circuits is presented and then we propose a statistical driven correlation-aware placement for CNFET-based gates. We illustrate that our placement engine improves the reliability and timing yield of various circuits. Following these observations, in our method, a statistical approach is conducted to get optimum timing yield of register-to-register path in a sequential circuit. Subsequently, experimental results show improvement of about 19% in timing yield and 17% in reliability of some ISCAS89 circuits.**

Keywords—Carbon Nano-tube, Carbon Nano-tube Field Effect Transistor (CNFET), Timing yield, Reliability, Placement.

I. INTRODUCTION

Shrinking device scale of conventional CMOS technology has got a contemporary concern nowadays. This attitude translates into strict limit in physical and electrical properties which impel researchers to think of another alternative as substitution. Among many proposed candidates, carbon nano-tube (CNT) seems to be the most promising fundamental replacement for conventional CMOS devices due to its fascinating properties [1- 3].

As the result, Carbon Nano-tube Field Effect Transistor (CNFET) which is consisted of single-walled carbon nano-tubes are likely to be thriving candidates as a replacement of MOSFET technology so far [1-2]. Furthermore, the channel length of a CNFET can be scaled theoretically down to 5-10 nm without reducing the electrical characteristics [4]. Consequently, it is unanimous for many researchers that CNFETs have a potential for performing far superior to FETs build in silicon technology [2].

Albeit CNFET technology demonstrates many attractive characterizations, its disadvantages should not be neglected. To be more precise, we can refer to imperfections and fluctuations in the fabrication of CNFET technology which is caused by its small scaled size. To put into details, miss-positioned CNTs and CNT density fluctuations are some of the imperfections coming from the small size of CNFETs [5]. Simply put, CNFET technology is expected to deal with enormously higher fluctuations rather than conventional silicon CMOS.

Since the variations in CNFET-based digital circuits are high, some characteristics of CNT devices are not deterministic and should be described with statistical analysis and models. One of the most significant variations in CNFET technology is CNT density fluctuations among single transistors.

Recent studies have developed parameterized model for CNT density variations based on experimental data [6-7]. Moreover, rigorous derivation of the CNT density distribution using a binomial probability density function has been derived [8]. Based on these investigations the distribution of CNT in a single transistor is predicted to be consonant with a Gaussian approximation. That is, the number of CNT in a CNFET is not deterministic and should be modeled with a probability density function with relevant mean and variance [6].

In this paper, we first introduce the statistical timing yield model in sequential CNFET-based digital circuits. Subsequently, based on this model and the statistical reliability model which will be mentioned in next sections, we introduce our correlation-aware placement method. Finally, we conduct series of simulations in order to analyze and discuss about results obtained from our method.

II. TIMING YIELD MODEL

Timing yield is the probability that circuit delay is within specific timing constraints. These timing constraints are setup time and hold time which are meant in register-to-register paths.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

The statistical estimation of timing yield totally depends on gate delays. In order to estimate the timing yield of a digital circuit, we use the statistical estimation of CNFET delay which considers CNT fabrication imperfections [7]. As a result, CNFET delay can be estimated with statistical approach based on Gaussian distribution as it is mentioned below:

$$D_{CNFET} = \frac{\left(1+ g_{CNT}L_{s,CNT}\rho_{s,CNT}\right)\left(nC_{g-CNT,1}L_{g,CNT}+ C_{g-p}W_{g,CNT}\right)\Delta V}{mg_{CNT}\left(V_{supply}- V_{th,CNT}\right)} \quad (1)$$

where g_{CNT} is transconductance per CNT, $L_{s,CNT}$ is the length of the doped CNT source region, $\rho_{s,CNT}$ is the source resistance per unit length of a doped CNT, C_{g-CNT} stands for capacitance CNT per unit gate length ($L_{g,CNT}$), C_{g-p} stands for capacitance per unit gate width, *n* represents the total number of CNTs, $W_{g,CNT}$ is the width of lithographically defined gate, $V_{th,CNT}$ is the threshold voltage, V_{supply} is supply voltage and *m* is the number of semiconducting CNTs.

Regarding this equation, the timing yield of CNFET-based circuits, is the probability that the circuit path delays satisfy the setup time and hold time restrictions both which can be shown as:

$$Y_{T-CNT} = P\left\{\max_{i=1:N_{path}} D_{path}(i) < STC\right\} \times P\left\{\min_{i=1:N_{path}} D_{path}(i) > HTC\right\} \quad (2)$$

where N_{path}, is the number of paths in the circuit, $D_{path}(i)$ is the path delay of each path, STC is the setup time constraint and HTC is the hold time constraint.

Based on these equations, we can estimate the timing yield of CNFET-based sequential circuits by viable approximation. As we show later, we use this approach for estimating timing yield analysis in our correlation-aware placement which it will be described comprehensively in the next sections.

III. CNFET Reliability Model

According to International Technology Roadmap for Semiconductor (ITRS), the relative failure rate per transistor will be improved by shrinking the device size [9]. Consequently, the probability of failure in CNFET should be considered as one of the most sophisticated problems in designing circuits based on CNFETs. With regard to this fact, proposing a method which can reduce the transistor failure rate seems to be essential.

Recent studies developed some innovative methods for improving the reliability of CNFET-based circuits. Transistor level redundancy [8], and independent N^2-transistor structures [10] are two proposed methods for reducing the probability of failure in the CNFET-based circuits. Although these methods show reliable results in reducing the probability of failure, most of them depend on redundancy in different design layers.

The reliability of a CNFET is the probability of the transistor is being functional due to defined defect probability in manufacture. Regarding to this fact, the defect model in CNFET is categorized namely into two groups, *open* and *short*. We use this assumption in estimating the reliability in our method. In other words, the probability of a circuit is being functional depends on the CNFETs reliability which compound the digital circuit as shown in Fig. 1. In this figure, each CNFET is consisted of drain, gate and source. CNFETs are randomly placed on aligned CNTs.

According to this Figure, if the probability of failure for each CNFET is assumed to be $P_{f-CNFET}$, the reliability of the whole circuits in worst case can be determined by:

$$R_{cicuit} = (1 - P_{f-CNFET})^n \quad (3)$$

where *n*, is the number of CNFET [8].

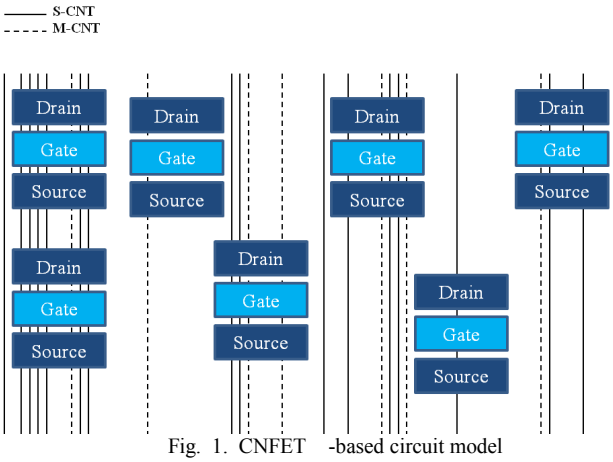


Fig. 1. CNFET -based circuit model

Regarding to the models which are discussed above, we can estimate the reliability and timing yield of a CNFET-based circuit with viable approximation, followed by developing our correlation-aware placement method to improve the functionality of the circuit.

IV. Correlation Model

As we discussed, the carbon nano-tubes can be regarded as the most promising candidates for replacing CMOS technology so far. Although applying this technology is limited by many constraints, several applicable electrical features of this technology facilitates designers' job to deal with these constraints. One of these plateful features is correlation characterization. Zhang *et al*. [11], investigate that large degree of spatial correlation observed in directional CNT growth provides a unique opportunity to significantly enhance the functional yield.

Based on these aligned-active layout restrictions, it is reasonable to measure this phenomenon and its effect on timing yield and reliability in this article. To do so, we aim to adopt some restrictions during CNT growth, and especially during the placement level, which are improving the timing yield and functional yield simultaneously.

Consider the CNFET-based logical circuit which is target [12], the problem is to place the components at gate level as it shows the best timing behavior and consequently, this will lead to higher rate of timing yield. The correlation model is described as a correlation matrix (COV_{Gate}) which its entries are "0" which it means that two junction nodes are not correlated and "1" which shows the fully correlated gates.

$$COV_{Gate} = \begin{bmatrix} C_{1,1} & \cdots & C_{1,gate} \\ \vdots & \ddots & \vdots \\ C_{Ngate,1} & \cdots & C_{Ngate,Ngate} \end{bmatrix} \quad (4)$$

where $C_{i,j}$ is represented for correlation between $CNFET_i$ and $CNFET_j$. That is, in our timing yield estimation, the correlated CNFETs which are existed in correlated gates are the same in the number of CNTs and semiconductor CNTs respectively, and this will lead to identical behavior in delay deviation and probability of defect. In other words, if two CNFETs are totally correlated, the deviation in delay of both transistors are varying the same and the probability of defect in both transistors are assumed to be identical. In Fig. 2, two correlated CNFETs are shown, as shown in this Figure if one of the CNTs which is aligned between transistors is metallic, both transistors are not functional due to short defect and the probability of failure for both transistors are the same and equal to the probability of a CNT is metallic.

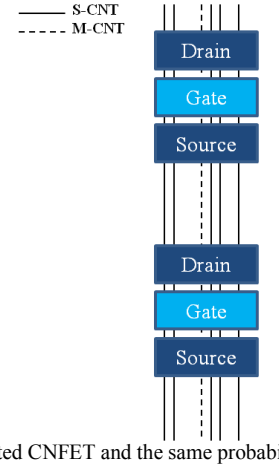


Fig. 2. Correlated CNFET and the same probabi lity of short defect

In order to consider the timing yield and reliability both in our method, we employ the Simulated Annealing (SA) algorithm for optimizing the placement strategy [13]. That is, the reliability, timing yield and wiring cost are the essential factors to be considered in our placement method and we intend to represent a solution which can improve these parameters simultaneously.

V. PLACEMENT METHOD

A. Movement Strategy

In our placement algorithm, each cell movement is defined as placing a cell from uncorrelated area to correlated areas which are in the nearest position from the movable gate. The uncorrelated gate will be moved to the nearest free space which can be correlated with at least one gate. This will not lead to better result for every movements due to wiring cost, area overhead, deviation in delay and CNTs defects, but the intention of our algorithm is to correlate as many gate as possible with other gates in order to enhance essential parameters of the design. Furthermore, under some circumstances, as it is shown in figure 3, where there is no uncorrelated gate in the design, random correlated gates will be relocated to another correlated area if that movement leads to better functionality. With this approach, not only many optimized placements will be explored, but also the placement method works more precisely.

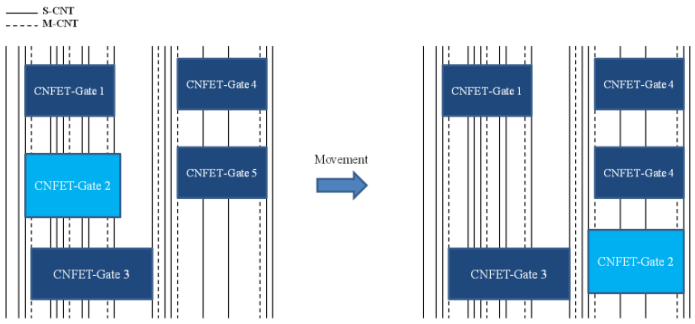


Fig. 3. Cell movement str rategy in placement algorithm

B. Metrics

We calculate three placement metrics in our method: first the reliability, which means the probability of the circuit, is being functional. Second, Timing yield of the circuit which means the probability of the circuit satisfies the timing constraints and third wiring cost and area overhead which is absolutely depended on the position of the circuit gates and logic.

In practice, we define the reliability of a circuit with equation (3), and then we use equation (2) to figure out the timing yield of the circuit. For approximating the wiring cost and area, we use half parameter method which is the simplest and widely used method to approximate the wire length in digital circuits.

In order to estimate the chip area, since there is no valid guideline for CNFET-based gate size, we had to refer to CMOS-based gate size with a constant ratio for all gates. Regarding to the gate positions, the total area for the circuits have been figured out, which will give us realistic overview of the circuit area.

We also limit our movement in order to remove overlaps in placement, We define x_i and y_i variables for new x and y of the new cell i for every movable cell and also we define $leftx_j$, $rightx_j$, upy_j and $lowy_j$ for the four boundaries of bounding box of net j. Now, for every cell i connected to net j:

$$leftx_j = \min_i(x_i), \qquad rightx_j = \max_i(x_i)$$
$$lowy_j = \min_i(y_j), \qquad upy_j = \max_i(y_j) \quad (5)$$

These min and max functions are converted to the linear constraints below:

$$\begin{aligned} \text{left}x_j \le (x_i), \qquad & \text{right}x_j \ge (x_i) \\ \text{low}y_j \le (y_j), \qquad & \text{up}y_j \ge (y_j) \end{aligned} \tag{6}$$

Additionally, cell overlaps are needed to be removed. It can be done with four limitations as below:

$$\begin{aligned} x_{G_{right}} \ge x_{G_{left}} + l_{G_{left,}} \qquad & x_{G_{left}} \le x_{G_{right}} + l_{G_{left}} \\ x_{G_{right}} \ge x_{G_{left}} + l_{G_{left,}} \qquad & x_{G_{left}} \le x_{G_{right}} + l_{G_{left}} \end{aligned} \tag{7}$$

where x_G and y_G show the position of gate G, with l_G stands for the length and w_G is representative of the width of the gate respectively.

C. Cost Function

As mentioned above, we employed SA algorithm in order to optimize our parameters. The Simulated Annealing is a generic probabilistic algorithm for global optimization. It randomly moves gates, and either accepts or rejects the move based on certain cost function. In [14], for instance, the Simulated Annealing is used for timing driven placement by augmenting the cost function to include path based timing information. Same approach is adapted in our method to increase the cost function value in each step.

During incremental placement, in each step, placement metrics are calculated and regarding to their value in previous step, next move will be defined. In order to declare the cost function, we should consider that every parameter has the same impact on the cost function with individual design coefficient. Here is our innovative cost function declaration which involves our placement metrics and intends to be augmented in each step of algorithm:

$$F_{cost} = \alpha(R) + \beta(TY) + \gamma\left(\frac{1}{S+W}\right) \tag{8}$$

where R is the reliability of the whole circuit, TY is timing yield of the circuit, S is the estimated area of the chip and W is the wiring cost which is approximated as mentioned in the previous section. α, β and γ are design coefficient which are defined based on the necessities and requirements of the design.

In our placement method, we ignore some placement characteristics like routability and congestion map for the sake of simplicity.

D. Experimental Results

We have implemented our algorithm for simulating the incremental statistical driven correlation-aware placement as a linear programming problem in C++ on Linux. In order to evaluate the effectiveness of our method in various sets of circuits, we tested our proposed placement method on some of ISCAS89 circuits. Although α, β and γ in our simulations are found to be 0.3 with equal impact on the cost function, they can be easily set to another quantity based on design requirements. Each step of the SA algorithm attempts to replace the current placement with a random placement which is chosen according to candidate gates as mentioned in the previous section. The initial and the final temperatures of the SA algorithm are set to be optimized for each circuit.

As illustrated in the Fig. 4, the best improvement in timing yield and reliability are respectively about 19% and 17% in ISCAS89 circuits. Table 1 and the corresponding chart show timing yield and reliability improvement by placing a small set of gates based on our method. The average number of gates moved by our placer is about 4.7%; however we were able to improve the timing yield and reliability of circuits on average by 11% and 10.5% respectively (Note that the initial and final timing yield and reliability are estimated by statistical approach).

Fig.4 shows the percentage of improvements in timing yield and reliability of carbon nano-tube based circuits. Moreover, it illustrates the area and wiring overheads percentage per circuits. As depicted in this figure, area and wiring overheads have significant impact on our method, but in order to achieve more reliable and robust implementation, we have to reduce their impact by reducing their design coefficient.

TABLE I. TIMING YIELD AND RELIABILITY IMPROVEMENT FOR CNFET BASED DIGITAL CIRCUITS

Circuit	Total cells	Cell moved	Initial		Final	
			Timing Yield (%)	Reliability	Timing Yield (%)	Reliability
S27	19	5	46.4	8.32E-2	54.6	9.14E-2
S1196	575	48	32.3	4.43E-3	38.8	4.79E-3
S5378	2958	36	39.6	6.40E-4	58.9	7.23E-4
S9234	5825	278	42.2	1.96E-4	53.1	2.29E-4
S13207	8584	196	54.2	2.78E-4	61	2.91E-4
S15850	10369	255	57.8	8.98E-5	65.1	9.67E-5
S35932	17829	432	45.6	3.21E-5	57.5	3.71E-5
S38417	23815	406	41.2	9.82E-6	48.3	1.12E-5
S38584	20705	263	44.8	8.71E-6	59.6	9.23E-6

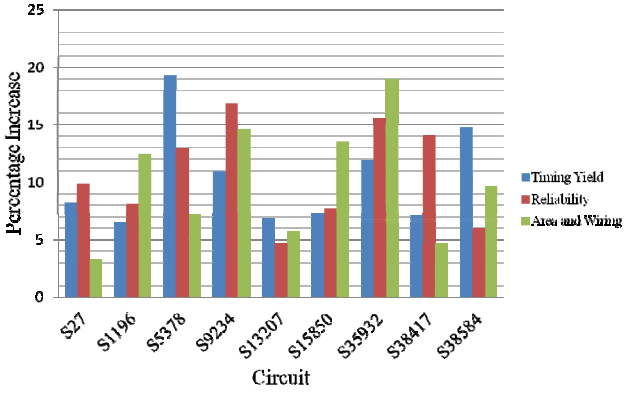


Fig. 4. Percentage increase of circu uit parameters from using the placement method

As illustrated in Fig. 4, there is a trade-off between improvement in timing yield and reliability, and design overheads which are depended on design necessities and requirements.

CONCLUSION

Since there are enormous variations existing in fabrication and implementation of CNFET-based digital circuits, timing yield and reliability of these circuits are exposed to significant reduction. Consequently, development of methods which are capable of improving these circuit characterizations is vitally needed. In this work, we have proposed a statistical driven correlation-aware placement which can improve the robustness of CNFET-based circuits significantly. Area and wiring overheads are two obstacles which are considered in our method as design parameters. As a result of study, we have accomplished remarkable improvements in timing yield and reliability on average of 11% and 10.5% respectively for ISCAS89 circuits by statistical driven placing 4.7% on average, even when the initial placement is generated by random engine. The proposed technique can be extended to consider transistor level placement for more accurate and detailed results.

REFERENCES

[1] Z. Chen, J. Appenzeller, Y. Lin, J. Sippel-Oakley, A. G. Rinzler, J. Tang, S. J. Wind, P. M. Solomon, P. Avouris, "An Integrated Logic Circuit Assembled on a Single Carbon Nanotube," *Science*, vol. 311, no. 5768, pp. 1735, 2006.

[2] A. Raychowdhury, A. Keshavarzi, J. Kurtin, V. De, K. Roy, "Carbon Nanotube Field-Effect Transistors for High-Performance Digital Circuits-DC Analysis and Modeling Toward Optimum Transistor Structure," *IEEE Trans. Electron Devices,* vol. 53, no. 11, pp. 2711-2717, 2006.

[3] A. Javey, J.Guo, D. B. Farmer, Q. Wang, R. G. Gordon, M. Lundstrom, H. Dai, "Carbon Nanotube Field-Effect Transistors with Integrated Ohmic Contacts and High-K Gate Dielectrics," *Nano Letters*, vol. 4, no. 3, pp. 447-450, 2004.

[4] P. McEuen, M. Fuhrer, P. Hongkun, "Single-Walled Carbon Nanotube Electronics," *IEEE Trans. Nanotechnology*, vol. 53, pp. 78-85, 2002.

[5] H. S. P. Wong, J. Appenzeller, V. Derycke, R. Martel, S. Wind, Ph. Avouris, "Carbon Nanotube Field Effect Transistor – Fabrication, Device Physics, and Circuit Implementations" , *IEEE International Solid State Circuits Conference (ISSCC)*, pp. 370-371, 2003.

[6] J. Zhang, N. Patil, A. Hazeghi, S. Mitra, "Carbon Nanotube Circuits in the Presence of Carbon Nanotube Density Variation," *IEEE/ACM Design Automation Conference (DAC)*, pp. 71-76, 2009.

[7] R. Ashraf, M. Chrzanowska-Jeske, S. G. Narendra, "Functional Yield Estimation of Carbon Nanotube based Logic Gates in the Presence of Defects", *IEEE Trans. Nanotechnology*, vol. 29, no.6, pp. 687-700, 2010.

[8] P. Zarkesh-Ha, A. A. Shahi, "Stochastic Analysis and Design Guidelines for CNFETs in Gigascale Integrated Systems', *IEEE Trans. Electron Devices*, vol. 58, no.2, pp. 2284-2295, 2011.

[9] ITRS, International Technology Roadmap for Semiconductors Emerging Research Devices, 2012.

[10] B. Ghavami, M. Raji, H. Pedram, "Metallic-CNT and Non-uniform CNTs Tolerant Design of CNFET-based Circuits using Independent N^2-Transistor Structures", *IEEE Computer Society Annual Symposium on VLSI*, pp. 242-247, 2011.

[11] J. Zhang, S. Bobba, N. Patil, A. Lin, H. S. P. Wong, G. De Micheli, S. Mitra, "Carbon Nanotube Correlation: Promising Opportunity for CNFET Circuit Yield Enhancement," *IEEE/ACM Design Automation Conference (DATE)*, pp. 889-892, 2010.

[12] A. Bachtold, P.Hadley, T.Nakanishi, C. Dekker, "Logic Circuits with Carbon Nanotube Transistors", *Science*, vol. 294, pp. 1317-1320, 2001.

[13] S. Lim, R. Ravichandaran, M. Niemier, "Partitioning and Placement for Buildable QCA Circuits", *ACM Journal on Emerging Technologies in Computing Systems*, vol. 1, no. 1, pp. 50-72, 2005.

[14] J. A. Chandy, P. Banerjee, "A Parallel Circuit-Partitioned Algorithm for Timing Driven Cell Placement", *IEEE International Conference on Computer Design (ICCD)*, pp. 621-627, 1997.

Nonlinear Beams and Active Controllable Field Concentrator with Isotropic Metamaterials

Yuriy Rapoport and
Nikita Kalinich
Physics Faculty
Taras Shevchenko National University of Kyiv
Kyiv, Ukraine
yuriy.rapoport@gmail.com

Allan Boardman
Joule Physics Lab., Inst. For Material Res., University of Salford
Salford, Greater Manchester, UK
A.D.Boardman@salford.ac.uk

Vladimir Grimalsky
Autonomus University of Morelos
Cuernavaca, Mor., Mexico
v_grim@yahoo.com

Abstract — **The idea of nonlinear field concentrator, inspired by papers on "linear black hall" and transformational optics [1-3] was originally proposed in the series of our conference papers [5-8]. We emphasize an importance of the strong nonlinearity, not considered in the resent papers [9, 10], where, in fact, only weak (Kerr-type) nonlinearity was included without any saturation. Namely, inclusion of strong (non-Kerr) saturable nonlinearity becomes very important in the "regime of superfocusing". Such a regime occurs during consequent linear and nonlinear focusing in the layered cylindrical structure, respectively in the external cylinder, where dielectric constant increases towards the center of the system and in the internal cylindrical region with nonlinear permittivity, linear part of which remains constant in space (and permittivity depends only on field intensity distribution). When amplitude exceeds some threshold value, the tendency to strongly nonlinear focusing, actually, "hot spot" formation, is observed. The new step done in the present paper is the demonstration of effective nonlinear controllability of the proposed [5-8] nonlinear effect. Namely, we have shown that the state of isotropic nonlinear field concentrator can be pushed into a regime of strongly nonlinear focusing, when two incident beams are used and an amplitude of one of these beams is gradually increased. Therefore we propose a possibility of a control of focusing of one ("signal") beam by means of amplitude of another (controlling") beam in a nonlinear field concentrator. In spite of a fact that of our evaluations are done for optic regime, a possibility of similar nonlinear signal processing is expected in wide range of frequencies, "from acoustic to optics". We demonstrate only a general idea of "beam-beam control", while more detail description of this new proposition will be a subject of another paper. Systems under consideration could be used for creation of antenna with strongly nonlinear focusing, achieving strongly nonlinear effects such as higher harmonic generation etc.**

Keywords—nonlinear supefocusing; concentrator; saturation; amplification; active

I. INTRODUCTION AND FORMULATION OF THE PROBLEM

The necessity to develop energy concentration is connected in part with their role in solar cells, but they are need also to ensure a high intensity electromagnetic field [1,2,3]. To start with, it is natural to involve metamaterials and use linear, classic, transformation optics. Going this way, some special treatment is necessary [4], since the practical use of metamaterial resonators can lead directly to large losses and narrow frequency band operation. Then the use of classical linear transformation optics, including form-invariance of Maxwell's equations leads to the necessity to engage both dielectric and magnetic properties [1]. Following this way, one may ask about the possibility to develop a new form of nonlinear transformation optics with form-invariance Maxwell equations changes that might provide an access to exciting and new high-intensity behavior. In this connection, a preliminary reports [5-8] were devoted to developing our new method included the new version of complex geometrical optics (CGO) and full-wave nonlinear electromagnetic solution (FWNES) in the different (external and internal) regions of nonlinear electromagnetic concentrator, and their matching on the proper boundary. This method is used in the present paper. An attempt to create the complete transformation optics approach appeared [9], but it is limited to a weak nonlinearity. We have found a possibility, however, to construct highly nonlinear concentrator, without using classical transformation optics, using gradient and non-magnetic material [3], which can be developed as an analogue of Kepler problem concerning gravitational capturing by massive bodies. Such an approach with graded index dielectric materials with cylindrical symmetry is used herein. At the same time, we make a very important addition of nonlinear electromagnetic capturing effect to the core region of concentrator. Thus, the high intensities, in particular in the core region of a concentrator, are put under very realistic control and therefore nonlinear controllability of a concentrating devices is achievable. One particular way of such control will be discussed in the present paper. As will be shown here, superconcentrator can be created, and a new kind of superfocusing conditions provide a tendency to the formation of "hot spots." New methods of analysis are developed to model this phenomenon. It does not follow from the ordinary transformation optics theory [1-3, 9]. However, it points to the need to try a new approach, which would be some type of regulated invariant form of Maxwell's equations. New methods used now are not to refer to the latter. And the emphasis is on target to achieve ultra-high field concentration or at least the tendency to such a concentration. Specifically, new methods were developed to search for strongly nonlinear effects in the system, which has a small size. We remain for a future studies, the question, whether it will be possible to detect highly

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

nonlinear transformations generalizations (which also would include as a particular case the method used in the present article), which are necessary for classical metamaterial route. In the present paper, heterogeneous, external, linear medium is used for the initial stage of the capture of the electromagnetic field, and then controllable nonlinear homogeneous inner core is used for the final capture. This system is sketched in Fig. 1, which contains a cross-section of the cylindrical field concentrator used in this study.

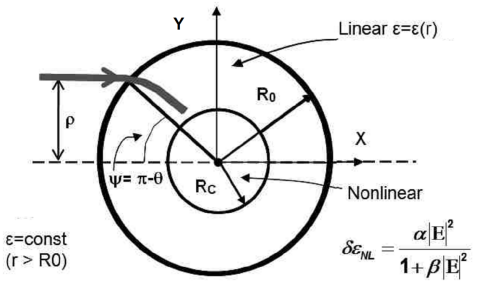


Fig. 1. Geometry of the problem: Nonlinear Field Concentrator. Core region $r \le R_c$ includes a medium with saturable (non-Kerr) nonlinear dielectric permittivity $\delta\varepsilon_{NL}$, shown in Fig. 1, where α, β are constants (complex and real, respecrively), and imaginary part of α describes nonlinear (saturable) losses.

The coordinates are (r, θ, z), and the cylinder axis coincides with the z-axis. The linear external region $(r > R_0)$ includes a homogeneous medium with relative permittivity ε_0. The inner core $(r \le R_c)$ is a non-linear and homogeneous. The latter also has an additional saturable nonlinear dielectric permittivity $\delta\varepsilon_{NL}$, as it is shown in Fig. 1. Outer cylindrical shell $(R_c \le r \le R_0)$ is both heterogeneous and linear. Heterogeneous permeability increases with decreasing radius, thus providing light capturing through graded index means [3]. For the treatment of wave propagation in such media, geometrical optics is used [11]. In particular, we developed the new version of complex geometrical optics [12, 13]. The complex eikonal describes how the amplitude and phase of the electric field change, while a ray propagates through weakly inhomogeneous medium for the beam. Moreover, even the propagation of beams can be included in our theoretical description.

The new exciting possibility of a nonlinear controllability considered in the present paper is connected with adding another (controlling) beam, incident on the system at some other impact distance (ρ) than the signal beam shown in Fig. 1. Consider the case, when signal and controlling beam propagate in external medium in parallel directions, and this direction is also parallel to the axis X of the chosen coordinate system (Fig. 1).

Our model provides a very accurate description of the intensity profiles, forming in a non-linear core. For internal non-linear region the appropriate technique, used in our model, is full-wave nonlinear electrodynamics. This model (CGO, FWNES and their matching with proper boundary conditions), in general, is the most adequate for modeling the nonlinear field concentration and a tendency to "hot spot" formation in the system (nonlinear concentrator) depicted in Fig. 1.

II. The Numerical Method and Algorithm for Modeling Nonlinear Field Concentration

In the external region of concentrator ($R_c \le r \le R_0$, Fig. 1), for the given incident ray, we solve the system of ordinary differential equations, based on the method of CGO, described in [4-8]. Then, a considered ray is supposed to be a central ray of a narrow beam with properly prescribed angle field distribution. In the internal nonlinear cylinder, where saturable nonlinearity is included and strongly nonlinear regime can develop (such as superfocusing and formation of hot spots), we use a full-wave nonlinear electromagnetic solution (FWNES) [4-8]. Using derived special boundary conditions, matching CGO and FWNES we become able to get, finally boundary conditions at the surface of internal cylinder ($r = R_c$) for the system of few hundred nonlinear coupled equations for the amplitudes of cylindrical harmonics. Such boundary conditions for each of the harmonics include a linear combination of an amplitude and its derivative by radius, and "external source" coming from the beam, incident on the concentrator from eternal medium. At $r = 0$, the boundary conditions, natural for corresponding cylindrical functions (harmonics) are used. Electric field E_z is presented as a series of angular (θ) Fourier harmonics B_m (m is the number of harmonic): $E_z = \sum_{m=-N/2+1}^{N/2} B_m(r)e^{-im\theta}$. This nonlinear system of coupled harmonics has a form:

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dB_m}{dr}\right) - \frac{m^2}{r^2}B_m + k_0^2[\varepsilon(r) + \sum_l \delta\varepsilon_l(r)]B_m$$
$$+k_0^2 \sum_{\substack{l=-N/2+1 \\ (l\neq 0)}}^{N/2} \delta\varepsilon_l(r)(B_{m-l} - B_m) = 0$$

Here $k_0 = \omega / c$, $\delta\varepsilon_l(r)$ includes nonlinearity, N is the number of the Fourier modes accounted for. To solve the system of coupled nonlinear equations, a special implicit nonlinear iterative grid-spectral method is used, while method of 3-diagonal running is applied along radius for each harmonic, with additional nonlinear iterations. This new method is a development of our methods [14], proposed before for modeling solitons and bullets in nonlinear gyrotropic [15, 16] and metamaterial [17] structures. Effective convergence of the numerical procedure is obtained.

III. The Results of Numeical Modeling: Nonlinear Beam Control

To illustrate a possibility of nonlinear control over nonlinear concentration, superconcentration and "hot spot" formation, we consider how nonlinear structure in the core region ($r \le R_c$, see Fig. 1) change, if we add another (controlling) beam with gradually changed input amplitude, to the signal beam with constant amplitude, shown in Fig. 1. Some parameters, namely input normalized amplitudes and impact distances of signal and controlling beams are shown in the captions to Fig. 2.

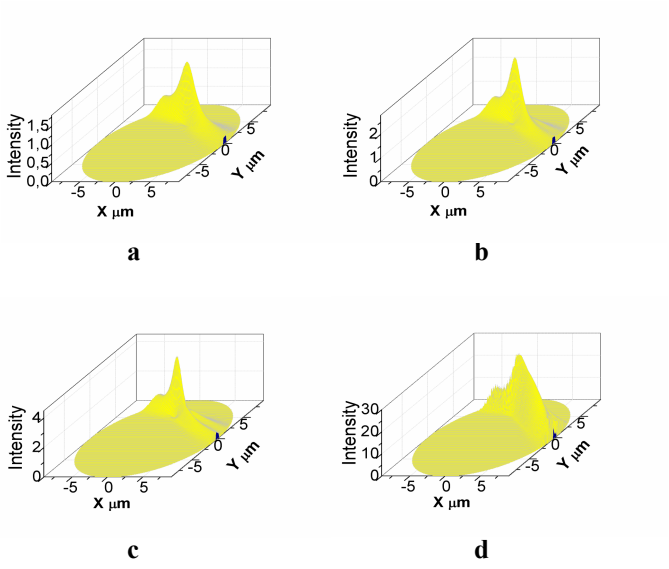


Fig. 2. Influence of controlling beam with incident normalised amplitude A_2 on amplitude and shape of nonlinear structure in the core of nonlinear concentrator (region $r \le R_c$, Fig. 1); impact distances $\rho_{1,2}$ of incident signal and controlling beams are equal to 10 μm and 10.5 μm, respectively; a- A_2=0; b- A_2=0.1; c- A_2=0.2; d- A_2=0.28.

The incident amplitudes are normalized in such a way, that corresponding nonlinear coefficients for a normalizing amplitude is equal to 1 [14-17], namely square of normalizing amplitude is equal to α^{-1}. Estimations show that the intensities of the input electromagnetic waves necessary for observations of nonlinear phenomena for a concentrator based upon doped n-Si should be of the order of (10-100) MW/cm^2. As it was reported in [5-8], due to superfocusing, a tendency to a formation of strongly nonlinear "hot spot" starts to reveal itself, when input amplitude reaches some "threshold" value. An input amplitude of signal beam is chosen to be below a "threshold value". Then, while incident amplitude of controlling beam increases gradually, shape and amplitude of the nonlinear structure, forming in the core region also change gradually (Fig. 2 a-c). At the same time, maximum of the nonlinear structure is placed somewhere inside a nonlinear core (region $r \le R_c$). Then, when incident controlling beam amplitude becomes close to definite value (around 0.28) in the case under consideration, an amplitude of the nonlinear structure starts to increase in the vicinity of definite point ($x = -R_c, y = 0$) placed on the axis X and on the boundary between nonlinear core and external linear cylinder (Fig. 2d). As it was shown in [5-8], grows of field at the point ($x = -R_c, y = 0$) is a symptom of the developing tendency to the formation of "hot spot" at the above mentioned point.

IV. Possibility of Controllable Hot Spot Formation

With farther increase of input amplitude of controlling beam, a tendency to a formation of "hot spot" reveals itself in the form of strongly localized and highly concentrated maximum near the point ($x = -R_c, y = 0$). This means that input normalized amplitude of controlling beam, in fact, exceeds corresponding "threshold value" (which is in the range between 0.28 and 0.30). Such a tendency is entirely controlled by external ("controlling") beam. This shows an exciting possibility of effective control over such strongly nonlinear phenomena as "superfocusing" and "hot spot" formation. Let us emphasize that numerical convergence in the "above threshold" region of parameters, to which Fig. 3 does correspond, is problematic, because in such a region "quasichaotic" dependence of "output" on "input" amplitude is observed.

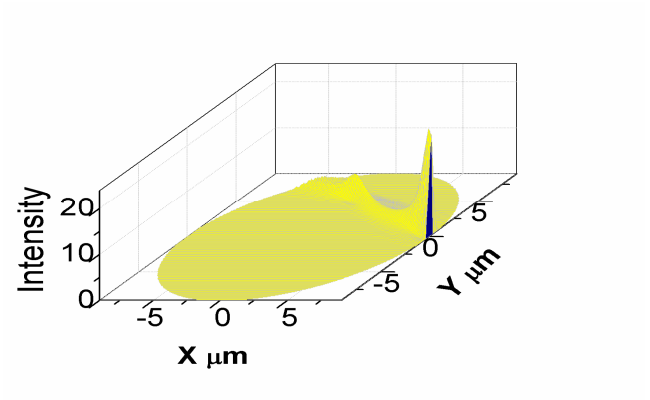


Fig. 3. Tendency to the formation of "hot spot" in the vicinity of the point ($x = -R_c, y = 0$) when incident normalised amplitude of the controlling beam reaches definite value (lying in the range between 0.28 and 0.30; compare Fig. 2d and 3). Numerical convergence in this range of parameters is problematic due to "quasichaotic" dependence of "output" on "input" amplitude.

To clarify this question, now a new, non-phenomenological model of an active medium accounting for non-locality and non-small nonlinearity is under development. Note also that the new results on a concentration with a nonlinear core medium possessing linear gain, nonlinear losses and saturation will be reported elsewhere.

V. Conclusions

We have demonstrated an exciting possibility of nonlinear control over such strongly nonlinear phenomena, as superconcentration and tendency to "hot spot" formation in a field concentrator, by means of applying an additional (respectively to the signal) "controlling" beam. Systems under consideration could be used for creation of antenna with strongly nonlinear focusing and providing strongly nonlinear effects such as higher harmonic generation etc.

References

[1] M. Rahm, D. Schurig, D. A. Roberts, S. A. Gummer, D. R. Smith and J. B. Pendry, "Design of electromagnetic cloaks and concentrators using form-invariant coordinate transformations of Maxwell's equations," Photonics and Nanostructures – Fundamentals and Applications, vol. 6, pp. 87–95, Apr. 2008.

[2] A. V. Kildishev and V. M. Shalaev, "Transformation optics and metamaterials," Uspekhi Fizicheskikh Nauk vol. 181 (1), pp. 59-70, Jan. 2011 (English Translation: Physics- Uspekhi, vol. 54 (1), pp. 53-63, Jan. 2011).

[3] E. E. Narimanov and A. V. Kildishev, “Optical black hole: broadband omnidirectional light absorber,” Appl. Phys. Lett,. vol. 95, 041106, Apr. 2009.

[4] A. D. Boardman, V. V. Grimalsky, Y. S. Kivshar, S. V. Koshevaya, M. Lapine, N. M. Litchinitser, V. N. Malnev, M. Noginov, Yu. G. Rapoport and V. M. Shalaev, “Active and tunable metamaterials,” Laser and Photonics Rev., vol. 5, pp. 287–307, March 2011.

[5] A. D. Boardman, V. V. Grimalsky, Yu. G. Rapoport, "The hybrid method of nonlinear transformational and complex geometrical optics for energy concentration in metamaterials", Metamaterials’2011: The Fifth International Congress on Advanced Electromagnetic Materials in Microwaves and Optics Barselona, pp. 763-765, Oct. 2011.

[6] Yu. G. Rapoport, A. D. Boardman, V. V. Grimalsky, Yu. Selivanov and N. A. Kalinich, "Metamaterials for space physics and the new method for modeling isotropic and hyperbolic nonlinear concentrators,” Proceedings of the International Conference on Mathematical Methods in Electromagnetic Theory (MMET) Kharkiv, pp. 76 – 79, Aug. 2012.

[7] Yu. G. Rapoport, A. D. Boardman , V. V. Grimalsky, R. Mitchell-Thomas and T. P. Orekhov, “Metamaterial-based electromagnetic and acoustic field concentrators and new physical phenomena: nonlinear focusing switching,” XXXII International Scientific Conference ELNANO Kyiv, , pp. 84-85, Apr. 2012.

[8] A. D. Boardman, V. V. Grimalsky and Yu. G. Rapoport, “Nonlinear transformational optics and electromagnetic and acoustic fields concentrators,” The Fourth International Workshop on Theoretical and Computational Nanophotonics, AIP Conf. Proc. 1398, 120-122 , 2011.

[9] L.Bergman, P.Alitalo and S. A. Tretyakov, “Nonlinear transformation optics and engineering of the Kerr effect,“ Phys. Rev. B, 84, 205103, Nov. 2011.

[10] A. Pandey and N. M. Litchinitser, “Nonlinear light concentrators,” Optics Letters, vol.37, no.24 , pp. 5238-5240, Dec. 2012.

[11] B. N. Gershman, L. M. Erykhimov and Yu. Ya. Yashin, Wave Phenomena In The Ionosphere and Space Plasma, Moscow: Nauka, 1984 (in Russian)

[12] A. Bravo-Ortega and A.H. Glassor, “Theory and application of complex geometrical optics in inhomogeneous magnetized plasma,” Phys. Fluids B, vol. 3, pp. 528-538, March 1991.

[13] A. V. Timofeev, “Geometrical optics and diffraction phenomena,” Uspekni fizicheskih nauk, vol. 175, pp. 637-641, June 2005 (in Russian) (English Translation: Physics-Uspekni, vol. 48(5), pp. 609-613, June 2005)

[14] A.N. Slavin, O. Büttner, M. Bauer, S.O. Demokritov, B. Hillebrands, , M.P. Kostylev, B.A. Kalinikos, V.V. Grimalsky and Yu. Rapoport, “Collision properties of quasi-one-dimensional spin wave solitons and two-dimensional spin wave bullets”, Chaos, vol.13, pp. 693–701, June 2003.

[15] Yu.G. Rapoport, , C.E. Zaspel, J.H. Mantha and V.V. Grimalsky, “Multisoliton formation in magnetic thin films” Phys. Rev. B , Vol. 65, 024423, pp. 244231-244234, Febr. 2002.

[16] C.E. Zaspel, J.H. Mantha, Yu.G. Rapoport and V.V. Grimalsky, , “Evolution of solitons in magnetic thin films” Phys. Rev. B, Vol. 64, 064416, June 2001.

[17] A.D. Boardman, O. Hess,Mitchell- R.C. Thomas, Y.G. Rapoport and L. Velasco, “Temporal solitons in magnetooptic and metamaterial waveguides” , Photonics and Nanostructures - Fundamentals and Applications, vol. 8, pp. 228-243, Sept. 2010.

Three-Level Approach to Graphene Metamaterials: Electron Density Waves and Linear and Nonlinear Electrodynamics

Yuriy Rapoport and Nikita Kalinich
Physics Faculty
Taras Shevchenko National University of Kyiv
Kyiv, Ukraine
yuriy.rapoport@gmail.com

Vladimir Grimalsky
Autonomus University of Morelos
Cuernavaca, Mor., Mexico
v_grim@yahoo.com2-

Igor Nefedov
Dept. of Radio Science and Engineering
Aalto University
Aalto, Finland
igor.nefedov@aalto.fi

Vadim Malnev
Dept. of Physics
Addis Abeba University
Addis Abeba 1176, Efiopia
vadimmalnev@yahoo.com -

Abstract — **We propose "three-level" approach to graphene metamaterials including "metamaterial treatment" of electron density waves in graphene. First, we determined the characteristics of electron wave transport in graphene layer, accounting for applied external electric and magnetic fields, exchange and spin-orbit interactions. The new system of equations for 2D electron beam in graphene layer with a 2D inhomogeneity is derived and the special method for a solution of such a system is developed. W search a possibility of "capturing" by the "magnetic ring resonator" of an electron beam. Then we propose the paradigm of "solid state graphene electron optics". This approach includes different ways of control over characteristics of a propagation of electron density waves, passive components, typical for electron optics and nanoelectronics, such as waveguides, lenses, diffraction gratings. Finally, the new equations are derived, to model linear (with great slowdown) and nonlinear (with electronic nonlinearity) electromagnetic waves in graphene layered lossy structures.**

Keywords—graphene; metamaterials; metatronics; electron optics; electronic nonlinearity; electromagnetic waves

I. INTRODUCTION

The new concept, proposed recently, namely, metatronics [1] is merging with graphene metamaterials. In this connection, at most electromagnetic (plasmonic) aspects of graphene metamaterials had been considered. In [2], the generation and manipulation of a flow of separate electrons by the periodically driven mesoscopic systems was considered. Such a system includes contacts (boundaries) which serve as a source/sink of particles which are scattered by the mesoscopic sample. Few quantum devices, such as particle emitter and electron collider, had been proposed. We propose an approach to "graphene metamaterials" including three levels of consideration (see also [3]). This approach includes "metamaterial treatment" of electron density waves in graphene (known example of which are graphene electron "perfect lenses" [4]), electron wave control upon electromagnetic characteristics (such as conductivity etc.) and, finally, electromagnetic/plasmonic effects (with proper periodical structures, electronically controllable waveguides and resonators etc.). Moreover, we propose that controllable electron waves in graphene could be considered not only to form the desirable and tunable macroscopic electromagnetic conductivity [5], but they are very interesting as such. In particular we propose to use controllable electron wave processes in graphene as the basis of 2D solid-state graphene electron optics. This approach includes different ways of control over characteristics of a propagation of electron density waves, passive components, typical for electron optics and nanoelectronics, such as "positive-negative phase" waveguides, lenses, diffraction gratings [6] etc. To form and control beams, magnetic and electric fields, as in electron optics can be used, as well as superlattices with periodical chemical potential [6], exchange and spin-orbit interaction coefficients. Such electron devices as spin filters, electron collimators, modulators, beam splitters etc. could be used as elements of solid-sate graphene electron optics, as a part of emerging radio-frequency nanoelectronics, a new paradigm in electron systems design [7]. In this paper, we consider three problems in the context of formulated above "three-level approach". First, we determine the characteristics of the electron wave transport in graphene layer, where the electric or magnetic field, the coefficient characterizing exchange or spin-orbit interaction, or magnetization, or these coefficients altogether could have proper inhomogeneity/periodicity. The second problem is modelling guiding electromagnetic waves in the layered structure "carbon nanotubes- graphene- dielectric (CNTGD)", and investigation of controllable slowing down and energy flow propagation in THz range. These two problems include direct development of

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

our previous works [3], [8], while the new results of 2D electron function capturing and slowing down of electromagnetic waves in lossy CNTGD structures will be presented. Finally, we will present an idea of searching nonlinear electrodynamics phenomena in the layered graphene-dielectric structures, where, in fact, quantum theory of 2D electron gas and nonlinear electromagnetic phenomena are synthesized.

II. The Method for Modeling Linear Electron Density and Electromagnetic Waves

A. 2D electron beams in graphene with 1D and 2D inhomogeneities

The method involves single-particle Hamiltonian and 4-component wave function $\vec{\psi} = (\psi_{A\uparrow}, \psi_{A\downarrow}, \psi_{B\uparrow}, \psi_{B\downarrow})^T$ where index "T" means transposition, "A", "B" correspond to one of two non-equivalent graphene sublattices ("pseudospin"). Arrows "$\uparrow$" and "$\downarrow$" correspond to "spin up" and "down". Extrinsic (or Rashba), intrinsic spin-orbit coupling, and an effects of magnetization (of an inhomogeneous or periodical ferromagnetic structure), and of scalar electrostatic potential describing (inhomogeneous/periodical) bias electric field are accounted for [3], [8]. Stationary Dirac-Weyl equations yields the new 2-component system of equations. If electron wave propagates in x direction and coefficients in obtained system of two second-order equations depend only on x (see, for example, [3]), 2D wave packet is modeled numerically, using Fourier transform in transverse (y) direction. It is shown that for the system, finite in x direction (with length L_x), the boundary conditions at $x = 0$ and $x = L_x$ are reduced to the linear combination of vector wave function and it's x derivative. For the case of the general 2D dependence of coefficients on coordinates (x, y), the new system of equations for pseudospin functions is derived. Method of numerical solution of such a system is developed and the results on trapping of 2D wave functions by means of "2D magnetic resonator" are obtained. Note that, in fact, spatial dispersion is accounted for by means of direct solution of the system of Dirac-Weyl equations, and no any supposition concerning a smallness of the effect of spatial dispersion, such as in [9], are used. The developed method is a combination of split-operator method [10-12] and pseudospectral method and includes non-stationary problem and especially derived boundary conditions (both at $x = 0$ and $x = L_x$, and at lateral sides of a graphene strip with the finite width), accounting for a propagation of 2D beam/pulse in a media with 2D inhomogeneity.

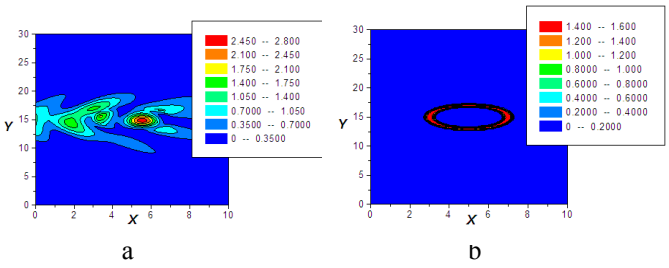
B. Linear electromagnetic slow waves in the layered CNTGD structures

This method includes the effective medium model of CNT array [13], surface conductivity model of graphene [5] and transfer matrix method. Then, basing on functional minimization technique with proper initial approximation, control of slowing down electromagnetic waves in CNTGD system and an effect of losses in the regime of very small group velocity is modeled.

III. Results of Numerical Computations for Linear Wave Phenomena

A. 2D electron beams in inhomogeneous graphene: capturing in 2D magnetic resonators and filtration

Transmission of 2D electron beam through the system with quasiperiodic 1D structure (in particular, "magnetic barrier") had been investigated [3], [8]. The modeling had shown high effectiveness of control over transfer characteristics of 2D electron waves. In particular, for different quasiperiodical "magnetic barriers" (function), dependence on structure period of pseudospin transmission function can be of "resonant" (for oscillating with constant sign) or "switching" type (for oscillating with alternating sign) [3], [8]. The new results of trapping of graphene electron waves in 2D "magnetic resonator" are obtained. A possibility of capturing electron distribution function is illustrated in Fig. 1 a, b.



a b

Fig. 1. Capturing of electron probability function using "magnetic ring resonator": a- spatial distribution of $|\psi_{A\downarrow}|^2$ component in graphene plane (XY) ; b- corresponding (to Fig. 1a) 2D distribution of function $F(x, y)$, to which stationary magnetic vector-potential is proportional.

A scattering of 2D electron beam on a 2D "diffraction grating" in graphene layer is under a consideration. We propose that controllable electron density waves in graphene can be a basis of solid-sate graphene electron optics. An example of such devices could be electron multichannel splitter, based on a system of graphene electron waveguides, similar to corresponding microstrip or integral optics/optic fiber structure.

B. Characteristics of slow electromagnetic waves in layered CNTGD structure

Electromagnetic guiding wave characteristics in CNTGD structure are modeled. An effective control of electromagnetic wave characteristics in CNTGD structures is provided by an external bias electric field which tunes the chemical potential in graphene layer. Distributions of the longitudinal and the transversal components of the local Poynting vector and energy velocity over the waveguide cross-section are calculated. We have found [3], [8] very interesting regime with large slowing down of waves, even with zero group velocity V_g at frequency 16.6 THz (and, at the same time, with increased energy concentration with given energy flow) in lossless CNTGD structure. In the present paper, we account losses for and found that V_g does not reach zero, but can have relatively very small value, controllable by means of chemical potential (for example, through external electric field).

IV. THE IDEA OF NONLINEAR ELECTRODYNAMICS APPLICATIONS IN LAYERED GRAPHENE STRUCTURES

We are interested now in the possibility of formation of electromagnetic solitons [10-12], [14-16] and bullets [11] in plane of graphene layer and effect of bistability for the electromagnetic waves propagating through a (multi)layered graphene-dielectric structures. Electrodynamics nonlinearity comes from electron nonlinearity. Namely nonlinear current i_s in the plane of graphene is expressed in the form $i_s \approx i_o F(\psi)$, where $i_o = eV_F n_2$, $F(\psi) = \psi(1+\psi^2)^{-1/2}$, $\psi = eA/(\hbar k_F)$, V_F, n_2 and A are Fermi velocity, effective surface concentration and amplitude of magnetic vector-potential, respectively, k_F is characteristic wavenumber. Nonlinear equation respectively to amplitude $C = i\omega A$ of electric field (ω is wave frequency) with first derivative by time and second derivative by coordinate z includes a new form of nonlinearity, proportional to the multiplier of the form $NL = 1 - Q^{-1/2}[1 - Q^{-1}(1 - \delta Q^{-1} \mid C \mid^2)]$, $Q = 1 + \alpha \mid C \mid^2$ (coefficients δ, α are not presented here). Corresponding equation and preliminary computations show a possibility of bistability for electromagnetic wave penetrating through multilayered graphene-dielecric structures.

V. CONCLUSIONS

Three-level approach to the layered graphene-dielectric structures is developed, including investigation of electron waves, electromagnetic waves and electromagnetic characteristics based on electron waves. Transmission of 2D electron beam through the system with quasiperiodic structure (in particular, “magnetic barrier”) and electromagnetic guiding wave characteristics in CNTGD structure are modeled. High controllability of transmission electron wave function and electromagnetic characteristics can be useful for sensors, modulators and nonlinear effects in graphene, such as a very effective harmonic generation etc. The new method is developed for modeling 2D electron waves in graphene with 2D inhomogeneity of bias electric and magnetic fields, exchange and spin-orbital interactions. The possibility of trapping of graphene electron waves in 2D “magnetic resonator” is shown. Controllable electron density waves in graphene can be a basis of solid-sate graphene electron optics. Large slowing down of waves in CNTGD structure are possible. Accounting for losses, group velocity does not reach zero, but still can have relatively very small value. Therefore highly controllable graphene-based delay lines and filters can be built. Electronic-based nonlinearity (of new form) can provide nonlinear effects and applications for electromagnetic waves in layered graphene-dielectric structures.

REFERENCES

[1] N. Engheta, “Taming light at the nanoscale”, Physics World, pp. 32–34, Sept. 2010.

[2] M. V. Moskalets, Scattering Matrix Approach to Non-Stationary Quantum Transport, Imperial College Press, London, 2012.

[3] Yu.G. Rapoport, V.V. Grimalsky and I.S. Nefedov, “Graphene as Electron Wave Density Metamarerial and Modeling 2D Electron Dynamics”, XXXII Intern. Sci. Conf. Electronics and Nanotechnology, ELNANO 2012, Kyiv, Ukraine, 10-12 Apr. 2012, Proceedings, pp. 86-87, 2012.

[4] V.M. Cheianov, V. Falko and B.L. Altshuler, “The focusing of electron flow and a Veselago lens ”, Science, vol. 315, pp. 1252–1255, March 2007.

[5] G.W. Hanson, “Dyadic Green’s functions and guided surface waves for a surface conductivity model of graphene”, J. of Appl. Phys., vol. 103, 064302, June 2008.

[6] M. Begliarbekov, O. Sul, J. Santanello, N. Ai, X. Zhang, E.-H. Yang, S. Strauf, “Localized States and Resultant Band Bending in Graphene Antidot Superlattices”, Nano Lett., vol. 11, pp. 1254-1258, 2011.

[7] L. Pierantoni and F. Coccetti, “Radio-frequency nanoelectronics: A new paradigm in electronic systems design”, Proc. of Asia-Pacific Micr. Conf. 2010, pp. 1007-1014, 2010.

[8] Yu.G. Rapoport, V.V. Grimalsky and I.S. Nefedov , “2D Electron Dynamics in Single Layer ``Graphene Metamaterial", TaCoNa-Photonics 2011. AIP Conf. Proc., Vol. 1398, pp. 138-140, 2011.

[9] M. G. Silveirinha and N. Engheta, “An Effective Medium Approach to Electron Waves: Graphene Superlattices”, http://arxiv.org/pdf/1203.1902.pdf

[10] Yu.G. Rapoport, , C.E. Zaspel, J.H. Mantha and V.V. Grimalsky, “Multisoliton formation in magnetic thin films” Phys. Rev. B , Vol. 65, 024423, pp. 244231-244234, 2002.

[11] C.E. Zaspel, J.H. Mantha, Yu.G. Rapoport and V.V. Grimalsky, , “Evolution of solitons in magnetic thin films” Phys. Rev. B, Vol. 64, 064416, 2001.

[12] A.N. Slavin, O. Büttner, M. Bauer, S.O. Demokritov, B. Hillebrands, , M.P. Kostylev, B.A. Kalinikos, V.V. Grimalsky and Yu. Rapoport, “Collision properties of quasi-one-dimensional spin wave solitons and two-dimensional spin wave bullets”, Chaos, vol.13, pp. 693–701, 2003.

[13] I.S. Nefedov and S.A. Tretyakov, “Ultabroadband electromagnetically indefinite medium formed by aligned carbon nanotubes”, Phys. Rev. B, vol. 84, 113410, 2011.

[14] A.D. Boardman, O. Hess,Mitchell- R.C. Thomas, Y.G. Rapoport and L. Velasco, “Temporal solitons in magnetooptic and metamaterial waveguides” , Photonics and Nanostructures - Fundamentals and Applications, vol. 8, pp. 228-243, Sept. 2010.

[15] A.D. Boardman, L. Velasco, N. King and Y.Rapoport, “Ultra-narrow bright spatial solitons interacting with left-handed surfaces”, J. Opt. Soc. of America B: Opt. Phys., vol. 22, pp. 1443-1452, July 2005.

[16] H. Dong, C. Conti and F. Biancalana “Terahertz relativistic spatial solitons in doped graphene metamaterials”, arXiv:1107.5803 [physics.optics]

Two-Channel Heterotransistors With Quantum Dots Systems

Vladimir Timofeyev
Physical and Biomedical Electronics Lab.
National Technical University of Ukraine "Kyiv Polytechnic Institute"
Kyiv, Ukraine
v.timofeyev@kpi.ua

Elena Faleyeva
Physical and Biomedical Electronics Lab.
National Technical University of Ukraine "Kyiv Polytechnic Institute"
Kyiv, Ukraine
elena.faleeva@gmail.com

***Abstract* — Two-channel field-effect heterotransistor characteristics with embedded systems of quantum dots (QDs) are presented. A physical interpretation of processes and analysis of field-velocity dependences in these structures is given. The interaction mechanism of QDs in quantum well with two-dimensional electron gas in high electric fields related to the polar optical phonon influence and charge carriers' emission from QDs is described.**

Keywords — quantum dot, quantum well, field-effect heterotransistor, two-channel heterotransistor, polar optical phonons, size quantization.

I. INTRODUCTION

Quantum effects appear in heterojunctions with two-dimensional electron gas and embedded systems of quantum dots when the size of at least one direction is comparable with de Broglie wavelength. The carrier transport in the whole structure is impacted by these areas due to changes in both the electron and phonon spectra.

Previous researches showed that quantum dots (QDs) integration to a transistor's heterojunction increases the output current [1, 2]. This occurs due to the properties of quantum dots in high electric fields and quantization of the polar optical phonons in channel region. These processes lead to increasing of drift velocity of charge carriers in the channel and to transistor output current-voltage characteristics saturation at higher drain voltages.

The longitudinal and transverse transport of charge carriers in these structures has features associated with short-channel effects, high potential, and concentration gradients, quantum effects, etc. The electron mean free path l_e at T = 300 K has the characteristic scale of 10^{-5} cm for scattering by acoustic phonons and 10^{-6} cm for strong optical scattering. Charge carriers in a channel are accelerated in the electric field to the optical phonon energy and then the velocity is lost due to the inelastic scattering with the optical phonon emission.

QDs influence

For GaAs with polar optical phonon energy $h\omega_0$ = 36 meV electron's wave vector is limited by the value $k_{opt} = \pi / L_{opt}$, where L_{opt} = 12,5 nm [1], and then further velocity increase impossible because of polar optical scattering. There were several attempts to solve this problem by spatial confinement of polar optical phonons. And the most successful way was offered by authors of work [1]. They proposed to build in QDs near a heterojunction. InAs QDs layer formation in GaAs quantum well leads to the momentum quantization of optical phonons in GaAs between InAs QDs in both the longitudinal and transverse directions of the electron motion. Each lens-like QD is a reflector of the optical phonons in GaAs [1].

The minimal wave vector of the optical phonon in the x direction is given by:

$$q_x = \pi / l_x$$

where l_x is GaAs layer thickness between InAs QDs and heterojunction [1].

If l_x < 12,5 nm, the electrons moved by the field reaches the optical phonon energy and do not stuck on phonons that could dissipate, so their drift velocity increases with increasing of electric field, until the inelastic scattering appears on the quantized optical phonons [1].

The necessary condition for significant increasing of the maximum drift velocity is the small size of the quantization length l_x. This condition can be performed at the sufficient density of QDs which form a continuous barrier along the channel structure and the small thickness of the GaAs layers between the layers of QDs.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

II. NUMERICAL SIMULATION OF HETEROTRANSISTORS WITH QUANTUM DOTS

The description of physical processes in high electric fields for submicron heterostructures with quantum wells and quantum dots based on the set of relaxation equations for electrons' velocity (1-2) and electron temperature (3), that was obtained from Boltzmann's kinetic equation in approximation of relaxation time [3], complemented by the continuity (4), the Poisson (5), and the current density (6) equations is following:

$$v_x \frac{\partial v_x}{\partial x} + v_y \frac{\partial v_x}{\partial y} = \frac{e\varepsilon_x}{m^*} - \frac{v_x}{\tau_p} \quad (1)$$

$$v_x \frac{\partial v_y}{\partial x} + v_y \frac{\partial v_y}{\partial y} = \frac{e\varepsilon_y}{m^*} - \frac{v_y}{\tau_p} \quad (2)$$

$$v_x \frac{\partial T_e}{\partial x} + v_y \frac{\partial T_e}{\partial y} = -\frac{T_e - T_0}{\tau_E} + \frac{2E_D}{3k_B}(\gamma - 1) \quad (3)$$

$$v_x \frac{\partial n}{\partial x} + v_y \frac{\partial n}{\partial y} = 0 \quad (4)$$

$$\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} = -\rho / \varepsilon\varepsilon_0 \quad (5)$$

$$\vec{j} = -en\vec{v} \quad (6)$$

where v_x, v_y – x and y projection of the charge carriers' drift velocity, n – charge carriers concentration, m^* – effective mass, e – electron charge, ε_x, ε_y – x and y projection of the electric field; τ_p, τ_E relaxation time of momentum and energy, T_0 – equilibrium temperature (the temperature of the crystal lattice), φ – potential, ρ – charge density, ε – permittivity, ε_0 – electric constant.

The drift energy is determined as:

$$E_D = m^* v^2 / 2 = \frac{m^*}{2}(v_x^2 + v_y^2)$$

The γ coefficient is determined as: $\gamma = 2\tau_E/\tau_p$.

Submicron multilayered structure modeling with complex topology is related with nonlinear function in the second member of Poisson's equation, specifying distribution of charge density, and with application of numerical methods with specified discretization pitch and inhomogeneous boundary conditions.

On receiving of solution for potentials (Poisson`s equation) and concentration (continuity equation) with prescribed validity, equations of the set above are solved conjointly. The solution algorithm of system of Poisson and continuity equations, taking into account approximations and electric field intensity equation $\varepsilon = -\text{grad}(\varphi)$, is a several nested loops. To solve the systems sweep method in combination with Newton's iterative multi step method is used. After calculating of the electric field intensity distribution in a structure with given values of concentration and doping profiles (Poisson's two dimensional equation), system of relaxation equations and also equation for current in common was solved.

To solve the equations of the electron temperature and the projections of the drift velocity the method of successive approximations was used. Since in these equations is some function of the coordinates $z(x, y)$ and its first derivative, we can write $z = f(z)$. Then z_n may be expressed through the z_{n-1} by the following recursive formula: $z_n = f(z_{n-1})$.

The problem becomes a self-consistent after adding to the set of equations the expressions for the relaxation times. For the calculation of the relaxation times used described in [3] procedure.

Then drain current was obtained from the expression:

$$I_d = -envS_d,$$

where S_d is a drain area.

The size quantization effects must be also considered. Methodology of quantization effect accounting is presented in the following part.

A schematic image of topology of a two-channel heterotransistor with two quantum dot systems is shown in Fig. 1.

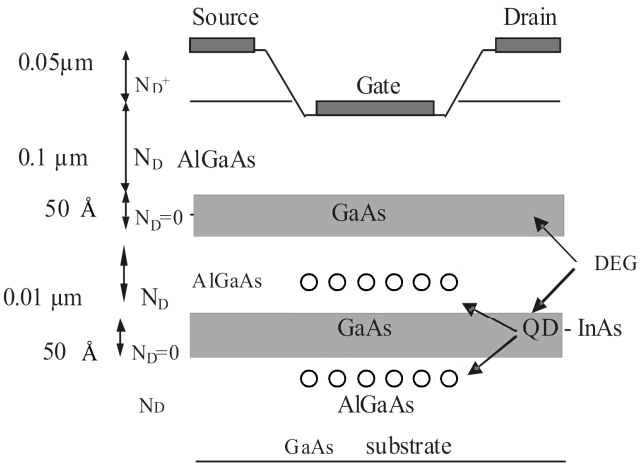


Fig. 1. The structure of a two-channel heterotransistor with quantum dot topology.

III. SIZE QUANTIZATION EFFECTS IN HETEROSTRUCTURES WITH QUANTUM DOTS

Consider the energy spectrum discretization and Fermi electron statistics in corresponding two-dimensional quantization sub-bands in high electric fields in structures with thin functional layers.

In doped heterostructure with carrier concentrations of about 10^{16} cm^{-3} and higher a significant electrostatic potential modification takes place near the quantum well which in turn causes considerable energy quantum level shifts and charge carrier concentration redistribution. To escape the rectangular quantum well approach for the two-dimensional electron gas, the self-consistent solution of Poisson and Schrödinger equations is needed:

$$-\frac{\hbar^2}{2m^*}\Delta\psi(x,y)+(\varphi(x,y)-E_0)\psi(x,y)=0,$$

where $\hbar$ – Planck constant, m^* – the effective electron mass, $\psi(x,y)$ – electron wave function, $\varphi(x,y)$ – quantum well potential profile, E_0 – energy of the main quantum level.

To calculate the charge carriers' concentration in the quantum well we should take into account that in two other directions electrons move freely and form a size quantization sub-bands. Mobile charge carriers' concentration (not taking into account the band bending) can be calculated as integration by sub-band states considering the Fermi-Dirac distribution [4]:

$$n(x,y)=\frac{m^*k_{\mathrm{B}}T}{\pi\hbar^2}\ln\left[1+\exp(q\frac{E_{\mathrm{F}}-E_i}{k_{\mathrm{B}}T})\right]\left|\psi(x,y)\right|^2,$$

where q is a unit charge.

To evaluate the location of the main energy level, the assumption of a "shallow" quantum well is used, for which $U_0 << \hbar^2/ma^2$ and the main quantum level location is evaluated as:

$$E_0 \approx U_0-\frac{ma^2}{2\hbar^2}U_0^2,$$

where U_0 is the depth of the quantum well, a is the width of quantum well, m is the electron mass .

Quantum dots obtained by the Stranski-Krastanov growth mechanism have a spherical shape at the bound of heterojunction. In the high field, as shown by numerical calculations, the difference of the energy level localizations in quantum dots in rectangular and spherical approximations is not significant.

QDs ionization

The emission of electrons from quantum dots is possible due to impact ionization, tunnelling of carriers from the QD to quantum well, and due to the thermal emission of carriers. The probability of thermal emission due to the carriers' interaction with lattice vibrations can be estimated by the Boltzmann factor: $\delta=\exp(-\Delta E_{\mathrm{a}}/k_{\mathrm{B}}T)$, where $E_{\mathrm{a}}=U_0-E_i$. Thus, if the depth of the potential well $U_0=0{,}8$ eV and the energy of the upper level about $E_n=0{,}5$ eV, the probability of thermal emission is negligible.

Acceleration of electrons to the ionization threshold depends on the ratio of two factors – acceleration in the electric field and scattering collisions with phonons, mainly with polar optical phonons and intervalley phonons. The distribution of electron temperature which was obtained by solving the relaxation equations set considers these factors. So, the ionization event occurs when the temperature is high enough for ionization in QDs location. It is assumed that the energy transferred to the charge carriers localized in a QD is spent to overcome the potential barrier and for ejection of carrier from the QD. And the electron in a transistor channel between the QDs again has the initial energy equal to the QD main energy level which it was emitted from.

When the negative voltage is applied to the gate then the width of the triangular barrier between the QW and QDs is growing up, hence, fewer electrons can tunnel through it, and the impact ionization becomes a decisive factor for the QD ionization If quantum wells are not tunnel-coupled, the carriers redistributed between the wells lose some energy to overcome the potential barrier, then the velocity in the bottom well increases [5-7].

The integration of the second channel in transistor leads to the additional carrier confinement, so the carriers cannot "spread" over the substrate [8]. Thus, even when carrier drift velocity is the same in both channels, the current through a device will increase.

IV. NUMERICAL RESULTS

The simulated structure represents a quantum system "heterojunction – QDs' layer – heterojunction – QDs' layer" in a plane with a normal being the direction of drain field applied. In this case, layers of GaAs and AlGaAs form heterojunctions and quantum dots are formed of InAs. In the researched structure the gate length is 0.2 μm, width of the wells of the two-channel transistors is 7.5 nm. The simulation results for the potential energy surface of the described structure are shown in Fig. 2.

It is interesting to compare two-dimensional distribution of the average electron drift velocity for three versions of the transistor structures: heterotransistor with two channels, heterotransistor with two channels with the quantum dots embedded between them, and heterotransistor with two channels and two systems of quantum dots. Fig. 3-5 shows the distribution of average drift velocity of the three structures.

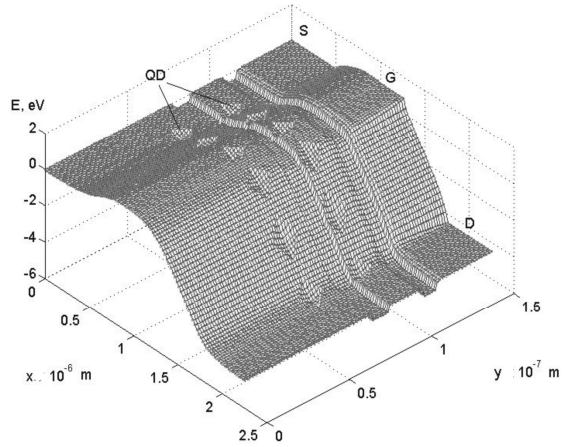


Fig. 2. Two-channel heterotransistor with two systems of QDs potential energy

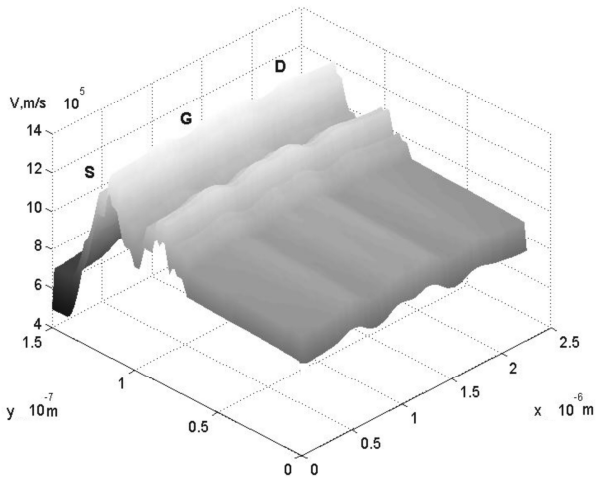


Fig. 3. The drift velocity of electrons in a two-channel heterotransistor.

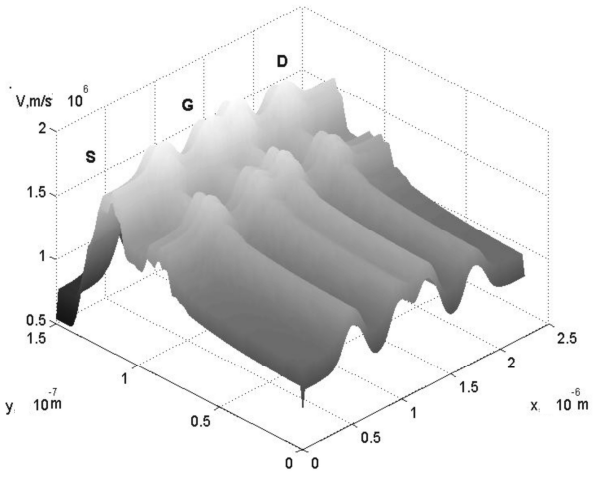


Fig. 4. The drift velocity of electrons in a two-channel heterotransistor with the system of QDs .

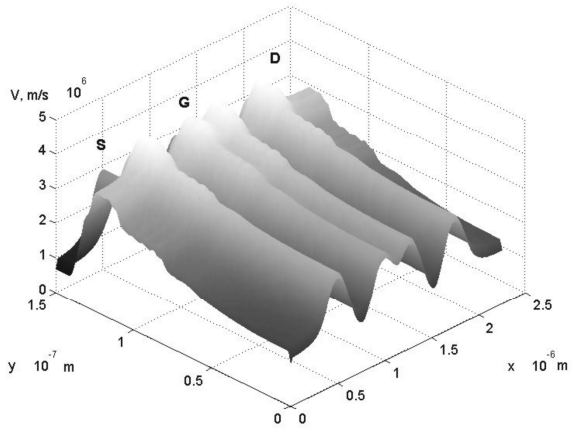


Fig. 5. The drift velocity of electrons in a two-channel heterotransistor with two systems of QDs

These results are calculated for the following electrical bias to the transistor electrodes: 0V on source; -0.4V and 4.0V the gate and the drain respectively. Systematic data for the drain current and maximum values of the drift velocity of the analyzed structures are presented in Table 1.

TABLE I.

	Researched structure	Drain current, mA	Maximum value of the drift velocity, m/s
1	two-channel heterotransistor	4.7	$1.2 \cdot 10^6$
2	two-channel heterotransistor with system of QDs	6.6	$1.7 \cdot 10^6$
3	two-channel heterotransistor with two systems of QDs	6.7	$4.2 \cdot 10^6$

These results show the increase in both the drain current and maximum values of the carriers' drift velocity.

CONCLUSIONS

The second channel and the systems of embedded quantum dots cause a "cooling" effect on electrons by reducing the probability of scattering by polar optical phonons and the processes of capture-emission in QDs, thus increasing injection into the channel of electrons with high velocity and low energy values. The probability of intervalley scattering is also reduced. The technique considering the quantum dots influence on the longitudinal charge carriers' transport in heterotransistor has been developed. The integration of QDs in heterojunction increases the output current and transc

onductance of a transistor. The drift velocity in a channel of heterotransistor with QDs is increasing because of the injection of "cold" electrons from quantum dots due to the impact ionization and tunnelling, and when the distance between the QD layer and the heterocontact is less than the wavelength of optical phonon – due to the phonon spectrum quantization.

REFERENCES

[1] J.K. Požela, V.G. Mokerov, "A large enhancement of the maximum drift velocity of electrons in the channel of a field-effect heterotransistor," Semiconductors, StPtb. : Ioffe Institute, vol. 40, pp. 357-361, 2006.

[2] V. Mokerov, J. Pozhela, K. Pozhela, V. Juciene, "Quantum-Dot Heterostructure Transistor With Enhanced Maximum Drift Velocity of Electrons," Semiconductors, StPtb.: Ioffe Institute, vol. 40, pp. 367-371, 2006.

[3] V. Moskalyuk, Physics of Electron Processes. Part II: Dynamic Processes. Kiyv: Avers, 2004.

[4] A.J. Schick, L.G. Bakueva, S.F. Musikhin, S.A .Rykov, ed. A. J. Schick. Physics of Low-Dimensional Systems. St. Petersburg.: Science, 2001.

[5] V. I. Timofeyev, E. M. Faleeva, "The Relaxation Processes Analysis in the Submicron Heterojunction Transistor with Quantum Dots," ISSE 2009 32nd International Spring Seminar on Electronics Technology, Brno, Czech Republic, pp. 262-263, May 13-17, 2009.

[6] V. I. Timofeyev, E. M. Faleeva, "Model of Heterotransistor with Quantum Dots," Semiconductor Physics, Quantum Electronics and Optoelectronics, vol. 13, №2, pp. 186-188, 2010.

[7] V. I. Timofeyev, M. Amini, E. M. Faleeva, "Non-stationary Drift of Electrons in Submicron High Electron Mobility Transistor with two Heterojunctions," Electronics and Electrical Engineering, Kaunas, Technologija, №. 4(76), pp. 33-36, 2007.

[8] Jie Liu, Yugang Zhou, Jia Zhu, Yong Cai, Kei May Lau, Kevin J. Chen, "DC and RF Characteristics of AlGaN/GaN/InGaN/GaN Double-Heterojunction HEMTs," IEEE Transactions on Electron Devices, vol. 54, №.1, pp. 2-10, 2007.

The Ways of Silicon Carbide Usage in Field-emission Devices: The Technological Aspect

N. Shelepin, I. Matyushkin, S. Orlov, A. Ermakov,
K.Svechkarev, P. Bobovnikov
Lab. of Perspective Technologies,
JSC "Molecular Electronics Research Institute", MERI
Moscow, Russia
imatyushkin@sitronics.com

A. Mikhaylov, A. Belov
NIFTI,
Nizhegorodsky State University
Nizhnyi Novgorod, Russia
mian@nifti.unn.ru

***Abstract* — In general cold field emission of SiC and other cathodes are compared. Possible applications of SiC for field emission cathodes and some of technological methods are considered. Also our results on the CVD-manufactured P-doped SiC-films are presented. Lastly our results on the ion-synthesized SiC-nanoinclusions in the silicon surface layer are discussed.**

Keywords—field emission; SiC; carbon; CVD; ion implantation

I. INTRODUCTION

Currently there is an expansion of carbon materials (carbon nanotubes, fullerenes, organic semiconductor materials, etc.) in conventional silicon electronics. Accordingly, silicon carbide (SiC) can be considered from a "naive chemical" point of view as a compromise between the two materials, Si and C. High binding energy in the SiC results in a high temperature of the crystal growth (>2000°C) and its chemical inertness and mechanical strength results in the complexity of processing. SiC could be etched very hard, only of KOH and a mixture HNO_3+HF. However, that major drawback in terms of traditional CMOS technology became the advantage in terms of longevity of field-emission cathodes. Moreover, this advantage became even more important for SiC-diodes or transistors, built by the technology of surface micromechanics, where distance "cathode - anode" is about L = 0,5−1,5 μm, and operating voltage, at which the threshold of field emission achieved, is equivalently 10−30 V, and it is and under conditions of poor vacuum (in that case the electron mean free path is comparable to L). Field emission of materials associated with SiC is relatively poorly studied. We note that, most authors came to that subject initially as to extended branch of the extensive studies of field emission from carbon materials.

During recent research technological methods of obtaining nanosized SiC are investigated. According to information contained in the database SCOPUS, about 100 articles are focused on nanocrystalline and amorphous objects based on SiC (review [1]).

The aim of this work is to systematize (but does not claim to be exhaustive) the literature in this fields and to present our own results of technological aspects and structural properties of SiC-nanointrusions and layers. In the first part we will compare the field emission of SiC and other cathodes. In the second part we will consider possible applications of SiC materials for field emission and some technological methods for SiC-structures producing. The third part contains our results to obtain SiC layers by CVD-method and doping of these films. The fourth part contains our results of the ion-synthesized nanointrusions of SiC into the surface layer of silicon.

II. EMISSION PROPERTIES OF SIC COMPARED WITH THE TRADITIONAL MATERIALS

In the last 15 years field emission from carbon materials has been actively studied, in particular, from structural carbon nanotubes (CNT). As a whole, despite considerable variation of parameters, metal cathodes have a large threshold voltage (turn-on) and large emission current, compared to the CNT-cathodes. Numerical parameters of the field emission from inorganic materials are also given in the review [2] as well as SiC-materials. Sample characteristics of these cathodes are given in Table.1.

TABLE I. THE CHARACTERISTICS OF SOME NANOSTRUCTURED CATHODES

Material of emitter	**The basic characteristics of field emitters: A threshold field V/um; threshold current, mA/cm^2; maximum field V/ μm; maximum current, mA/cm^2**	**Ref.**
ZnO-film	3,5; - ; 7,5; 2	[3]
ZnO nanowires ZnO nanopins Ga- alloyed ZnO- nanowires	4,3; 10^{-5}; 7,5; 10^{-4} 1,92; 10^{-5}; 6; $9*10^{-4}$ 2,4; $5*10^{-5}$; 7; $22,5*10^{-4}$	[4]
AlGaN/GaN nanojets	0,3; 0,001; >1,8; >5,2	[5]
CuO-nanowire	6; - ; 12; 0,18	[6]
Nanostructured Mo: 1)on the SIC substrate 2)on a metal substrate	1) 11,2; - ; 37,6; 57,5 2) 15,5; - ; 46,6; 19,6	[7]
α-Fe_2O_3 nanoprotrusions: 1) before anneal; 2) after 200°C	 1) 10; 0,06; >20; - 2) 10; 0,5; >12,2; -	[8]
$W_{18}O_{49}$ nanowire : 1) un-annealed; 2) annealed in Ar and H_2;	1) 2,5; - ; 6; 7,5 ; 2) 3; - ; 8; 3	[9]
TiO_2-polyaniline nanotubes	0,95; 0,001; >2,28; >1	[10]
C-graphite film	<1,5; - ; 4; 1	[11]
1) CNT 2) The CNT with diamond C-film before anneal 3) The CNT with diamond C-film after anneal	1) 2; - ; 4,75; 6,5; 2) 2,5; - ;4,5; 7,5 3) 3,5; - ; 5,75; 7,5	[12]
C-structure: 1)whiskers 2)film	1) 8; - ; 12; 0,05 2) 12; - ; 20; 0,1	[13]
CNT	2; 1nA; 15; 600 μA	[14]
CNT	1; - ; 2,5; 120	[15]

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

The adsorption by metal of electronegative atoms, such as oxygen, halogens, dioxide carbon and etc., leads to an increase work function In this case, the electric field of the double layer of charges prevents the exit of electrons. The work function of the adsorption, for example, of oxygen is increased by 1,5−2 eV. On the other hand, the formation of a thin layer of the oxide on metal does not, at least worsen the emission properties, as is commonly considered, in oxide cathodes, for example, a specifically coating of oxides of alkaline-earth metals is special used.

The main disadvantage of metal cathodes is the instability of current and short service life in high vacuum. Massive ion bombardment of residual gases causes cathode sputtering of the tip, which leads to a change in its microgeometry and, therefore, field intensity at the surface. These processes cause the instability at the emission current and short service life of the cathode.

The critical factors of CNT are: reproducible CNT height and form for irreversible transformation in the geometry of "forest" of nanotubes, compatibility with the technology of microelectronics. For micro-cathode which based on diamond-graphite layers the main obstacle is seen as high deposition temperature of about 1000°C (at low temperatures less than 400°C suffers reproducibility by form-factor inhomogeneities). In general, for carbon materials a distinctive disadvantage is degradation of material in low vacuum caused by reaction like

$$C + O^- \rightarrow CO\uparrow + e^-.$$

III. FIELD EMISSION FROM NANOSIZED SIC

Apparently, the first observation of field emission from nanosized SiC should be considered a note by Baker [16] in 1970. Since emission depends on the morphology of SiC and technology of SiC-forming nanostructures, we list the options, available now in the silicon microelectronics:

- Planar heterostructure SiC/Si-structure [17].
- Porous SiC-layer [18].
- Horizontal SiC-nanowires.
- Vertical SiC- whiskers and/or twigs (for the first time obtained in 1995 - by reaction forests CNT with SiO-vapor, [19]).
- SiC-powder particles on the Si-surface [20].
- Ion-synthesized SiC-intrusions on the Si-surface.
- Graphene after graphitization of SiC-layer.

Some data are collected in Table 2.

Detected metallization and/or hydrogen passivation [33] of SiC-surface is of particular interest, so it can be used for field emission (and the hydrogen bonds are stable with respect to oxygen).

Achievement of sufficient surface conductivity of SiC-structures directly influences on pre-exponential factor in Fowler-Nordheim equation. One of potentially possible ways of emission improvement consists in doping by phosphorus (or others electrically active impurity). Other mechanism is suggested by vacuum electronics and by experiments with SiC/SiO_x-nanocable. The effect of 10-nm oxide films is due to appearance of trapped levels in band gap of SiC and SiO_x (the last, obviously, lowers work function). The third way consists in gap structure refinement for account of quantum dimensional effects. Diminution of the lateral size of SiC-nanowhiskers, SiC-particles (in the form of intrusions or protrusions) or thickness of SiC-coating over, for example, silicon moves conduction band bottom to vacuum level. At the same time this diminution brings to increase in field-enhancement form-factor β. The basic problem of field emission from SiC-whiskers is disorderliness of their spatial arrangement that can affect stability of emission as well as in case with CNT. Therefore, choosing from variety of constructive decisions, we prefer SiC-intrusions (SiC-protrusions) and thin SiC-coverings of edge polycrystalline Si-cathode.

IV. CVD SIC-FILM: PROCEDURE AND PROPERTIES

Stability and field emission of microcathode, based on SiC, require strength Si−C bonds and fixing of SiC crystalline type. Acceptable CVD-process is characterized by temperatures higher than 1000 °C, in particular, for the evaporation of powders. There are studies of the polyenergetic ion synthesis [33] of SiC-layers with certain stoichiometry which show that the minimum temperature of Si-C bonds formation is about 700°C.

Associated with plasma activation CVD-production of SiC-films was carried out on an industrial plant “Izoplaz – 150” with silicon wafer doped boron (10 Ohm/sq). Monosilane (SiH_4) 5% in argon and 100% methane (CH_4) were used at working pressure 70 Pa and temperature of gas mix 350 °C. For the doping of the films 1,5% phosphine in argon was also used. Gases flow was respectively (liter per hour): SiH_4 - 10, CH_4 – 45, PH_3 – 6; under these conditions average rate of deposition was 115 nm. Following thermal anneal in N_2 and H_2 ambient with temperature about 1150 °C and duration about 1hour was performed. Table 3 summarizes data on resistivity (via Prometrix OmniMap RS35e) and typical SEM-image of SiC-samples (Fig.1).

V. ION SYNTHESIZED SIC-PROTRUSIONS

The samples of boron doped Si(111)-wafers were irradiated with C^+ ions in the pulse mode (without special separation of different charge states) at the accelerating voltage of 80 kV with a doses 10^{17} cm^{-2}, $2\cdot10^{17}$ cm^{-2}, $5\cdot10^{17}$ cm^{-2} and 10^{18} cm^{-2}. Following annealing was carried out in N_2 atmosphere at 1150−1200 °C.

At the dose of $1\cdot10^{17}$ cm^{-2} on the sample surface there are a lot of heterogeneous distributed protrusions (mostly linear groups, with distances of about 0,07 μm within the group and of 0,5 μm between groups). The protrusion heights and diameters are 10−30 nm of 20−40 nm respectively. At doses of $(2-5)\cdot10^{17}cm^{-2}$ its distribution is more uniform size approximately the same, and the density is slightly less. At the 10^{18} cm^{-2} dose its density sharply decreases as their size increases to about 100 nm

SEM-images indirectly show (Fig.3) that both species (C- and SiC-clusters) may be present simultaneously at the same sample. Below FN-curve is presented (Fig.4) for the sample according $2-10^{17}$ cm^{-2} dosage. The measurements were carried out in vacuum at about 1000V voltage stress through 40μm mica piece providing certain distance between cathode and anode. IV-curve consists of a linear plot with weak slope that is interpreted as noise, and second quite linear plot (88% covariation) in Fowler-Nordheim’s coordinates; it confirms the non-breakdown nature of measured current. Obtained maximal current density is 0,17mA/cm^2 and turn-on field is 20V/μm that is worse than average in literature which are accordingly 1 and 10.

TABLE II. FIELD EMISSION CHARACTERISTICS OF SiC NANOSTRUCTURES IN LITERATURE

Morphology	Fabrication method	Turn-on field and current density, V/μm and mA/sm^2; cathode-anode distance, μm	Achieved maximal field and current density, V/μm and mA/sm^2	Ref.
β-SiC nanorods (5÷10/1000 nm), non-aligned forest	LPCVD, H_2+CH_4+SiO, 1200°C, total 10 Torr, Fe catalist, VLS- mechanism of growth	13-17 and 0,01; 35-135	30 and about 0,32	[21]
6H-SiC, bulk, nitrogen doped: surface 3×3 и 1×1	Cree Research wafer of 0,005'' thickness, before and after H-plasma passivation (electron affinity – 3,0 и 3,9eV)	33 and 0,2nA (3×3), 50 и 0,2nA (hysteresis, 1×1); 0,1 – 12	-	[22]
SiC-intrusions and thin layers (150nm) on Si-surface	Implantation of C ions with energy 35keV and dosage about 10^{18}cm^{-2} и following anneal in N_2 ambient at 1200°C (2 hour)	1÷18 and 0,001; 20	>4 и >0,03	[23]
SiC-nanowhiskers of 50±300nm diameter in SiO_x shell of 10÷20nm thickness	Evaporation of C and Si powders at 1000°C during 2h in air (5×10^3 Pa). Catalyst – Fe_3O_4.	3,2÷5,3 and 0,001÷10; 200 1,1÷2,3 and 0,001÷10; 1000	>7,2 и >50; -	[24]
β-SiC/C-intrusions on Si-surface	Implantation of C ions with energy 35keV and dosage about 10^{18}sm^{-2} и following anneal in Ar ambient at 1200°C (1-2 hour)	12 and 0,001; 25 (without anneal) 25÷17,5 and 0,001; 25 (annealed, 1÷2 hours)	24 и 0,2 (without anneal) 33 и 0,23 (annealed, 2 hours)	[25]
Graphitized SiC-nanotip cathode (0,2μm^2 – emission area)	SiC-wafer and anneal	60 and -; 0,5 computation	100 and 1500; computation	[26]
SiC-nanowhiskers of 15÷25nm core and SiO_2-shell of 10÷20nm	Catalyst – NiO-particle на Si surface. Heating in Ar ambient at 1100°C during 3h at presence of graphite and WO_3 powders	4 and 0,01; 200 (bare SiC) 3,3÷4,5 and 0,01;	5,2 and 0,07; (bare SiC) 4,6÷5,7	[27]
β-SiC-nanowires (8÷20nm) on Pt-plain (non-aligned, horisontal)	Evaporation of SiO powder in graphite crucible (inductive heating with cylinder temperature of 1450°C)	3,5÷3,1 and 0,01; 100÷140	>6,0÷4,2 and >0,08÷0,06 (2h, 15% current deviation)	[28]
β-SiC-nanowire on graphite plain (non-aligned, horisontal)	Thermal evaporation of Si-powder	2,1 and 0,01	>3,8 and >0,68	[29]
Nanosized SiC-cap on Si-cathode tip (matrix, 10^{10} cm^2, 1nm/1μm)	Electron cyclotronic resonance associated with silicon etching in H_2+Ar+CH_4+SiH_4 plasma, 200°C	0,35 and 0,01	1 and 3	[30]
SiC-nanowhiskers (50nm/100μm) formed inside etch pits of SiO_2 layer on Si	Pyrolysis of CH_3CN on catalyst Ni at 800°C	1 and –; 500÷1000	>2,1÷1,2 and 6÷3 little hysteresis	[31]

TABLE III. RESISTIVITY OF SiC-SAMPLES

№	Mode of anneal	Thickness before anneal, mkm	RS Min Ohm/sq	RS Max Ohm/sq
1 (SiC)	1200°C H_2 60 min	0,27	210	336.2
2 (SiC)	1150°C N_2 40 min	0,43-0,44	670	849
3 (SiC)	1150°C N_2 40 min	0,43-0,44	674	834
3a (SiC)	1150°C N_2 40 min 1200°C H_2 60 min	0,43-0,44	413	521
4(SiC-P)	1150°C N_2 40 min	0,43-0,44	2.115	2.505
5(SiC-P)	1150°C N_2 40 min	0,43-0,44	2,14	2,5779

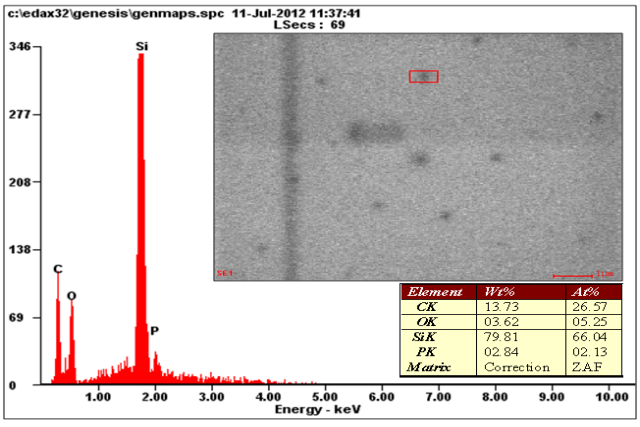


Fig. 1. Typical SEM-image of SiC-film via EDAX TSL system.

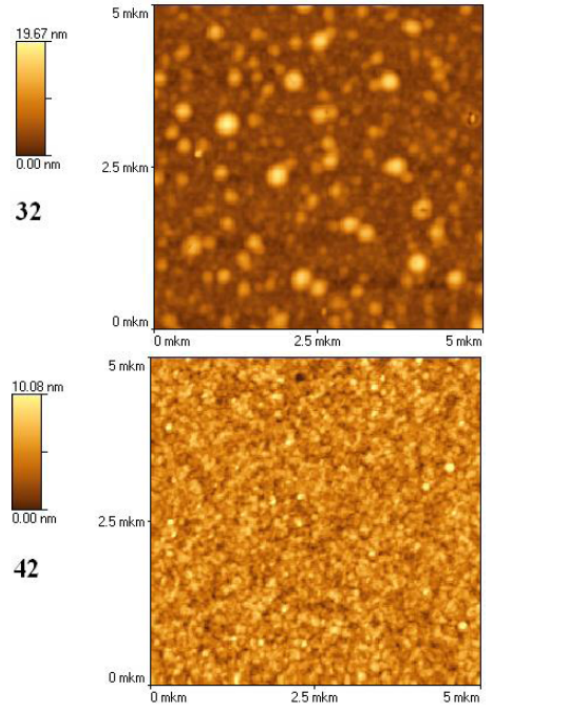


Fig. 2. Typical photography of the surface of samples according dosage 5·1017 cm-2 (№32) and 1018 cm-2 (№42) under anneal 1150°C condition.

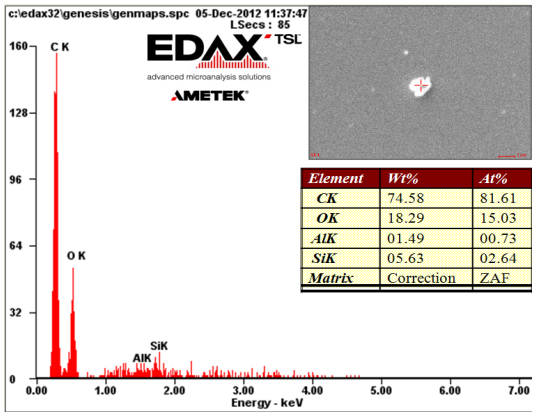


Fig. 3. SEM-image at point of a sample that proves C-protrusion presence.

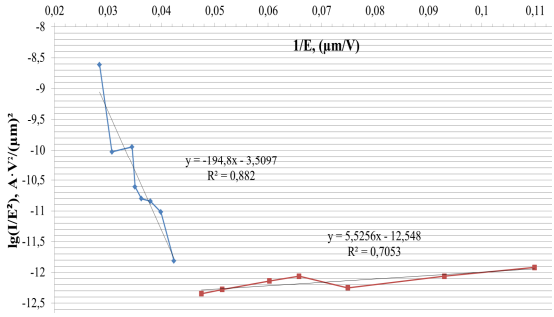


Fig. 4. IV-curve at FN-coordinates: right plot means a noise, left plot obeys classical FN-relationship.

VI. Conclusion

Colligating the previous the better topography is observed on un-annealed samples with a dose of $5 \cdot 10^{17}$ cm^{-2} (but we didn't study lower temperatures of annealing). At higher doses and at higher temperatures, in our opinion, two processes occur: i) separate C-protrusions merge into monolayer; ii) silicium atoms react with carbon and C-protrusion transforms into complex Si_xC_y or even $Si_xC_yO_z$-clusters. Our studies of entire layer surface conductivity indirectly prove this hypothesis. Our CVD SiC-film shows the absence of any field emission, perhaps, because of small form-factor or poor quality and layer heterogeneity. On the other hand, the phosphorous doping has decreased film resistivity in about 200 times.

The key strategies of field emission enhancement remain actual: the band gap engineering (to improve conductance) and the relief formation (to improve form-factor). In case of SiC-protrusions/intrusions the first is mostly an issue, and in case of sheeted SiC-layer the second is still useful.

References

[1] R.A. Andrievski, "Nano-sized silicon carbide: synthesis, structure and properties", Russ. Chem. Rev., vol. 78(9), pp. 821–831, 2009 (in Russian).

[2] Xiaosheng Fang et al. "Inorganic semiconductor nanostructures and their field-emission applications", Journal of Materials Chemistry, vol.18, pp.509–522, 2008.

[3] Xiuxia Zhang et al, "Nano- ZnO Film for field emission lighting tube", Proc. of 24th International Vacuum Nanoelectronics Conference (IVNC-2011), Wuppertal, Germany, July 18-22, 2011

[4] The usage of zinc oxide nanorods in nanoelectronics and nanophotonics, http://nanotechnology.sfedu.ru/mod/page/view.php?id=100

[5] Ru-Zhi Wang el al, "Field emission from AlGaN/GaN quantum-well nanotip arrays", IVNC-2011.

[6] G.X. Liu et al, "Preparation of CuO nanowire field emitter arrays by thermal oxidation in water vapor containing environments and their field emission properties", IVNC-2011.

[7] Y. Shen et al, "Synthesis and field emission properties of molybdenum nanowall arrays", IVNC-2011.

[8] J.Q. Wu et al, "Enhancement of field emission of iron oxide nanoemitters by thermal treatment", IVNC-2011.

[9] F. Mo et al, "Improvement of field emission uniformity of tungsten oxide nanowire arrays via in-situ gas plasma treatment", IVNC-2011.

[10] Sandip S. Patil, "Synthesis of titanium oxide-polyaniline nanotube composite and its superior field emission characteristics", IVNC-2011.

[11] A.N. Obraztsov, I.Yu. Pavlovsky, A.P. Volkov, "Field emission from graphite-like films", Zhurnal Technicheskoy Phisiki, vol.71 (11), pp.89-96, 2008 (in Russian).

[12] Yingzhi Liang et al, "The improvement of environmental stability of carbon nanotube cold cathode with a diamand-like carbon coating", IVNC-2011.

[13] Hidenori Gamo et al, "Liquid phase deposition and field emission measurements of the carbon nanomaterials grown in 1-octanethiol", IVNC-2011.

[14] Pavel Serbun et al, "Field emission from carbon nanotube (CNT) arrays and triode test of single CNT columns", IVNC-2011.

[15] Jae Woo Kim et al, "Development of high-temperature endurable CNT emitter for a cold cathode X-ray tube", IVNC-2011.

[16] Francis S. Baker, "Field Emission from Silicon Carbide Whiskers" // Nature, vol. 225, pp. 539-540, 07 February 1970.

[17] A.A. Evtukh, N.I. Klyui, V.G. Litovchenko, Yu.M. Litvin, O.B. Korneta, V.M. Puzikov, A.V. Semenov, "Peculiarities of field emission from silicon carbide films", Applied Surface Science, vol. 215 (1-4), pp. 237-241, 15 June 2003.

[18] Myung-Gyu Kang et al, "Stable field emission from nanoporous silicon carbide", Nanotechnology, vol. 24, Article 065201, 2013.

[19] H.J. Dai, E.W. Wong, Y.Z. Lu, S.S. Fan, C.M. Lieber, Nature (London), vol. 375, p.769, 1995.

[20] Tomomi Yoshimoto, Naohiro Yokogawa and Tatsuo Iwata, "Field Emission from Silicon Carbide (SiC) Micropowders", Jpn. J. Appl. Phys., vol.45, L482, 2006.

[21] X.T. Zhou, H.L. Lai, H.Y. Peng et al, "Thin β-SiC nanorods and their field emission properties", Chemical Physics Letters, vol. 318, pp. 58–62, 2000.

[22] Brandon Lea Ward. "Correlation of surface properties with electron emission characteristics for wide bandgap semiconductors", PhD thesis, North Carolina State University, Raleigh, 1999, 160p.

[23] Dihu Chen et al, "Electron field emission from SiC/Si heterostructures synthesized by carbon implantation using a metal vapor vacuum arc ion source", Appl. Phys. Lett., vol. 72(15), pp.1926-1928, 13 April 1998,.

[24] X.J.Wang, "Large scale SiC/SiOx nanocables: Synthesis, photoluminescence, and field emission properties", J. of Applied Physics, vol. 102, Article 014309, 2007

[25] W.M. Tsang, S.P. Wong, J.K.N. Lindner, "Origins of field enhancement in electron field emission from ion beam synthesized SiC layers", Applied Physics Letters, vol. 81 (21), pp. 3942-3944, November 2002.

[26] A.M. Svetlichnyi, O.B. Spiridonov, E.Y. Volkov, L.G. Linets, M.N. Grigoriev, "The evaluation of the characteristics of field emission nanostructures based on Si and SiC", Izvestia Yuzhnogo Fed. Univ. vol. 117 (4), pp. 27-35, 2011 (in Russian)

[27] Yonghwan Ryu, Youngjo Tak and Kijung Yong, "Direct growth of core–shell SiC–SiO2 nanowires and field emission characteristics", Nanotechnology, vol.16, S370–S374, 2005

[28] W.M. Zhou et al. "Field emission from nonaligned SiC nanowires", Applied Surface Science, vol.253, pp. 2056–2058, 2006.

[29] Jianjun Chen, Qiang Shi, Weihua Tang, "Field emission performance of SiC nanowires directly grown on graphite substrate", Materials Chemistry and Physics, vol. 126 (3), pp. 655-659, 15 April 2011.

[30] H.C. Lo, D. Das, J. S. Hwang, and K. H. Chen. "SiC-capped nanotip arrays for field emission with ultralow turn-on field", Applied Physics Letters, vol. 83(7), pp. 1420 – 1422, 18 August 2003.

[31] A.V.Okotrub, A.G.Kurenya, A.V.Gusel'nikov, Russian nanotechnologies, vol.4 (9-10), pp.72-76, 2009 (in Russian)

[32] P. Soukiassian, "Engineering Cubic Silicon Carbide Surfaces Properties Using Hydrogen: Metallization versus Passivation", Appl. Phys. A, vol. 82, pp. 421–430, 2006.

[33] K.H. Nusupov, N.B. Beisenhanov, I.V. Valitova, E.A. Dmitrieva, D. Zhumagaliuli, E.A. Shilenko, "Structural studies of thin layers of silicon implanted with multiple carbon ions", Fizika tverdogo tela, vol.48 (7), pp.1187-1199, 2006 (in Russian).

Photonic Crystal Filter with Fractal Structure 2D Modeling

A.V. Vishnevsky, D.E. Aznakayeva
Dept. of Electronics,
National Aviation University,
Kiev, Ukraine
aznakayev@nau.edu.ua

***Abstract*— In the paper a planar model (30-by-30 μm in size) of a band-pass filter, based on 2D photonic crystal with fractal inner structure has been considered. The frequency selectivity of this filter is due to its complex geometry, built on the base of "Sierpinski carpet" plane three-order fractal. The filter has been PC-simulated on TE electromagnetic waves in a band of wavelengths between 5 and 200 μm.**

Keywords— fractal; photonic crystal; filter; Sierpinski carpet

I. INTRODUCTION

Nowadays photonic crystals (PhC) [1], [2] are used in different areas of applied physics, including thin-film optics, holey fibers, LEDs, filters (a very promising branch) etc. PhCs are comparatively new optical materials for controlling and molding the flow of light. One dimensional PhCs are in widespread use today in form of thin-film optics (with applications from low and high reflection coatings on lenses and mirrors to and color changing inks and paints) . 2D and 3D photonic crystals are of great interest for both fundamental and applied scientific research, and the two dimensional ones are beginning to find some commercial applications already. The first commercial products involving two-dimensionally periodic photonic crystals are already available (in form of PhC-based fibers). The 3D PhCs are still off from commercialization but they offer additional features possibly leading to totally new devices (e.g. optical computers), when some technological aspects such as manufacturability and principal difficulties such as disorder are held under control. In this paper an idea of using a fractal 2D GaAs/Air PhC for the purpose of narrow-band filtering has been submitted.

II. GENERAL, TE AND TM PARTIAL DIFFERENTIAL EQUATIONS FORMULATION

A. General Partial Differential Equations Formulation

Particular partial differential equations formulations for high-frequency electromagnetic waves can be derived from classical Maxwell-Ampère's and Faraday's laws [3]:

$$\nabla\times\mathbf{H}=\mathbf{J}+\frac{\partial\mathbf{D}}{\partial t},\ \nabla\times\mathbf{E}=-\frac{\partial\mathbf{B}}{\partial t}$$

When using the constitutive relations for linear medias $\mathbf{B}=\mu_0\mu_r\mathbf{H}+\mathbf{B}_r$ and $\mathbf{D}=\varepsilon_0\varepsilon_r\mathbf{E}+\mathbf{D}_r$ and current $\mathbf{J}=\sigma\mathbf{E}$, these equations transform to

$$\nabla\times\mathbf{H}=\sigma\mathbf{E}+\frac{\partial\varepsilon\mathbf{E}}{\partial t},\ \nabla\times\mathbf{E}=-\mu\frac{\partial\mathbf{H}}{\partial t}$$

With help of relations $\mu\mathbf{H}=\nabla\times\mathbf{A}$ and $\mathbf{E}=-\partial\mathbf{A}/\partial t$ we can write down:

$$\mu_0\sigma\frac{\partial\mathbf{A}}{\partial t}+\mu_0\frac{\partial}{\partial t}\varepsilon\frac{\partial\mathbf{A}}{\partial t}+\nabla\times(\mu_r^{-1}\nabla\times\mathbf{A})=0$$

Having written now the fields in a time-harmonic form, it's possible to obtain (for 3D research)

$$\mathbf{E}(x,y,z,t)=\mathbf{E}_z(x,y,z)e^{j\omega t}$$

$$\mathbf{H}(x,y,z,t)=\mathbf{H}_z(x,y,z)e^{j\omega t}$$

From above field equations for either electric or magnetic component are derived

$$\nabla\times(\mu_r^{-1}\nabla\times\mathbf{E})-\omega^2\varepsilon_c\mathbf{E}=0$$

$$\nabla\times(\varepsilon_r^{-1}\nabla\times\mathbf{H})-\omega^2\mu\mathbf{H}=0$$

where

$$\varepsilon_c=\varepsilon-j\frac{\sigma}{\omega}$$

is the complex permittivity of the material in which propagates electromagnetic wave.

Because the filter will be investigated for TE as well as for TM polarization, for both cases we show appropriate differential equations formulation.

B. Partial Differential Equations Formulation for TE Waves

When the field propagates through the modeling *xy*-plane, a TE wave has only one electric field component in z direction (normal to observer), the magnetic field lying in the modeling plane. This is why one can write

$$\mathbf{E}(x,y,t)=E_z(x,y,t)=E_z(x,y)\mathbf{e}_z e^{j\omega t}$$

$$\mathbf{H}(x,y,t)=H_x(x,y,t)\mathbf{e}_x+H_y(x,y,t)\mathbf{e}_y=$$

$$=(H_x(x,y)\mathbf{e}_x+H_y(x,y)\mathbf{e}_y)e^{j\omega t}$$

μ_r denoting a 2-by-2 tensor, and ε_{rzz} and σ_{zz} being the relative permittivity and conductivity in *z* direction, the equation (which should be our main working equation)

$$\nabla \times (\mu_r^{-1} \nabla \times \mathbf{E}) - k_0^2 \varepsilon_{rc} \mathbf{E} = 0 \quad (1)$$

where

$$\varepsilon_{rc} = \varepsilon_r - j\frac{\sigma}{\omega \varepsilon_0}$$

can be simplified to a scalar equation for E_z:

$$-\nabla \cdot (\tilde{\mu}_r \nabla E_z) - \varepsilon_{rczz} k_0^2 E_z = 0$$

where

$$\tilde{\mu}_r = \frac{\mu_r^{\mathrm{T}}}{\det(\mu_r)}$$

If relation $\varepsilon_r = n^2$ is being used, where *n* is the index of refraction, one can rewrite the last equation in the following way (assuming that $\mu_r = 1$ and $\sigma = 1$)

$$-\nabla \cdot \nabla E_z - n_{zz}^2 k_0^2 E_z = 0$$

The wave number in vacuum k_0 is defined by the relation:

$$k_0 = \omega\sqrt{\varepsilon_0 \mu_0} = \omega / c_0$$

where c_0 is the speed of light in vacuum.

When we move to time domain, the working equation is

$$\mu_0 \sigma \frac{\partial \mathbf{A}}{\partial t} + \mu_0 \varepsilon_0 \frac{\partial}{\partial t}(\varepsilon_r \frac{\partial \mathbf{A}}{\partial t} - \mathbf{D}_r) + \nabla \times (\mu_r^{-1}(\nabla \times \mathbf{A} - \mathbf{B}_r)) = 0$$

and it can be simplified to

$$\mu_0 \sigma \frac{\partial A_z}{\partial t} + \mu_0 \varepsilon_0 \frac{\partial}{\partial t}(\varepsilon_r \frac{\partial A_z}{\partial t} - \mathbf{D}_{rz}) + \nabla \times (\mu_r^{-1}(\nabla A_z - \mathbf{B}_r)) = 0$$

The constitutive relations used here are $\mathbf{B} = \mu_0 \mu_r \mathbf{H} + \mathbf{B}_r$ and $\mathbf{D} = \varepsilon_0 \varepsilon_r \mathbf{E} + \mathbf{D}_r$

Using the relation $\varepsilon_r = n^2$, where *n* is the refractive index, the equation could alternatively be rewritten

$$\mu_0 \varepsilon_0 \frac{\partial}{\partial t}(n^2 \frac{\partial A_z}{\partial t}) + \nabla \cdot (\nabla A_z - \mathbf{B}_r) = 0$$

When using the refractive index, the assumption that $\mu_r = 1$ and $\sigma = 1$ is being made and only the constitutive relations for linear materials can be used.

C. Partial Differential Equations Formulation for TM Waves

The fields for the case of TM waves can be written as

$$\mathbf{H}(x, y, t) = H_z(x, y, t) = H_z(x, y)\mathbf{e}_z e^{j\omega t},$$

$$\mathbf{E}(x, y, t) = (E_x(x, y)\mathbf{e}_x + E_y(x, y)\mathbf{e}_y) e^{j\omega t}$$

It is required that ε_r and σ denote 2-by-2 tensors, and μ_{rzz} is the relative permeability in *z* direction.

The time-harmonic equation for TM waves looks as follows

$$\nabla \times (\varepsilon_{rc}^{-1} \nabla \times \mathbf{H}) - k_0^2 \mu_r \mathbf{H} = 0 \quad (2)$$

and can be simplified to a scalar one for H_z:

$$-\nabla \cdot (\tilde{\varepsilon}_{rc} \nabla H_z) - \mu_{rzz} k_0^2 H_z = 0$$

where

$$\tilde{\varepsilon}_{rc} = \frac{\varepsilon_{rc}^{\mathrm{T}}}{\det(\varepsilon_{rc})}$$

Or alternatively through $\varepsilon_r = n^2$

$$-\nabla \cdot (\tilde{n}^2 \nabla H_z) - k_0^2 H_z = 0$$

If moving to time domain the equation for TM waves is simplified to

$$\mu_0 \sigma \frac{\partial \mathbf{A}}{\partial t} + \mu_0 \varepsilon_0 \frac{\partial}{\partial t}(\varepsilon_r \frac{\partial \mathbf{A}}{\partial t} - \mathbf{D}_r) +$$

$$+ \nabla \times (\mu_r^{-1}(\nabla \times \mathbf{A} - \mathbf{B}_{rz})) = 0$$

where $\mathbf{B} = \mu_0 \mu_r \mathbf{H} + \mathbf{B}_r$ and $\mathbf{D} = \varepsilon_0 \varepsilon_r \mathbf{E} + \mathbf{D}_r$. Using refraction index it's possible to write down:

$$\mu_0 \varepsilon_0 \frac{\partial}{\partial t}(n^2 \frac{\partial \mathbf{A}}{\partial t}) + \nabla \times (\nabla \times \mathbf{A}) = 0$$

Here the constitutive relations for linear media can be used as well.

III. PROXY MODEL

The 2D finite elements (FEM) model of a PhC with a fractal "Sierpinski carpet" inner structure [4], that were treated in this research, is depicted in Fig. 1. It contains 14871 triangular elements, but in the research made there were 59484 such elements. Number of elements in the figure is reduced for the reason of better wire-grid mesh visibility. The numerical model was realized with help of Comsol Multiphysics ® software, as for example in [5] and [6]. Although experimental test hasn't been implemented, the validity of the research results is true because the main condition for numerical FEM modeling of electrodynamic structures is satisfied for all frequencies dealt with: the length of a proxy model element is always kept at least 10 times less than λ.

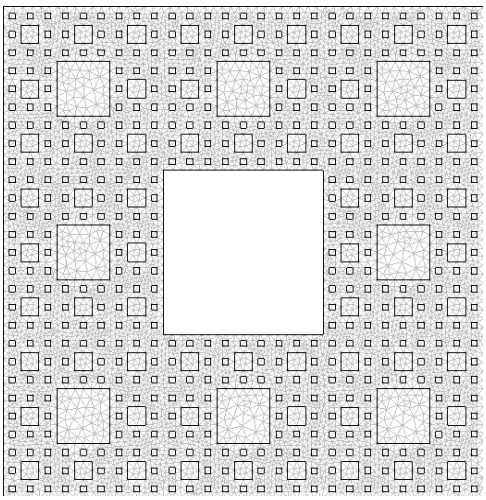


Fig. 1 The proxy FEM model of a fractal GaAs/Air 2D PhC

IV. BOUNDARY CONDITIONS

Continuity boundary conditions were used on all the edges joining GaAs (square pillars) and air (surrounding the last) subdomains. At the edges where the plane TE or TM wave is incoming into the structure of PhC, scattering boundary conditions were used with magnitude of electric field $E_m = 1$ V/m for TE and $H_m = 0,01$ A/m for TM waves respectively.

V. Investigation of Electromagnetic Field Propagation Through the Fractal Photonic Crystal at Different Wavelengths

The power flow patterns obtained in the research for TE waves on different frequencies using (1), are shown in Fig. 2. Here the red regions correspond to the highest levels of power flow (through the PhC body), yellow ones – middle levels, and blue ones – minimal power flows. The exciting planar EM wave is coming to the PhC through its inner four edges. Near each picture (at the right side) there is a level-bar, showing connection between colors and power flow levels. The Fig. 2,a shows case of weak transmission (15μm) – power flows through crystal in a perfect cross-like pattern.

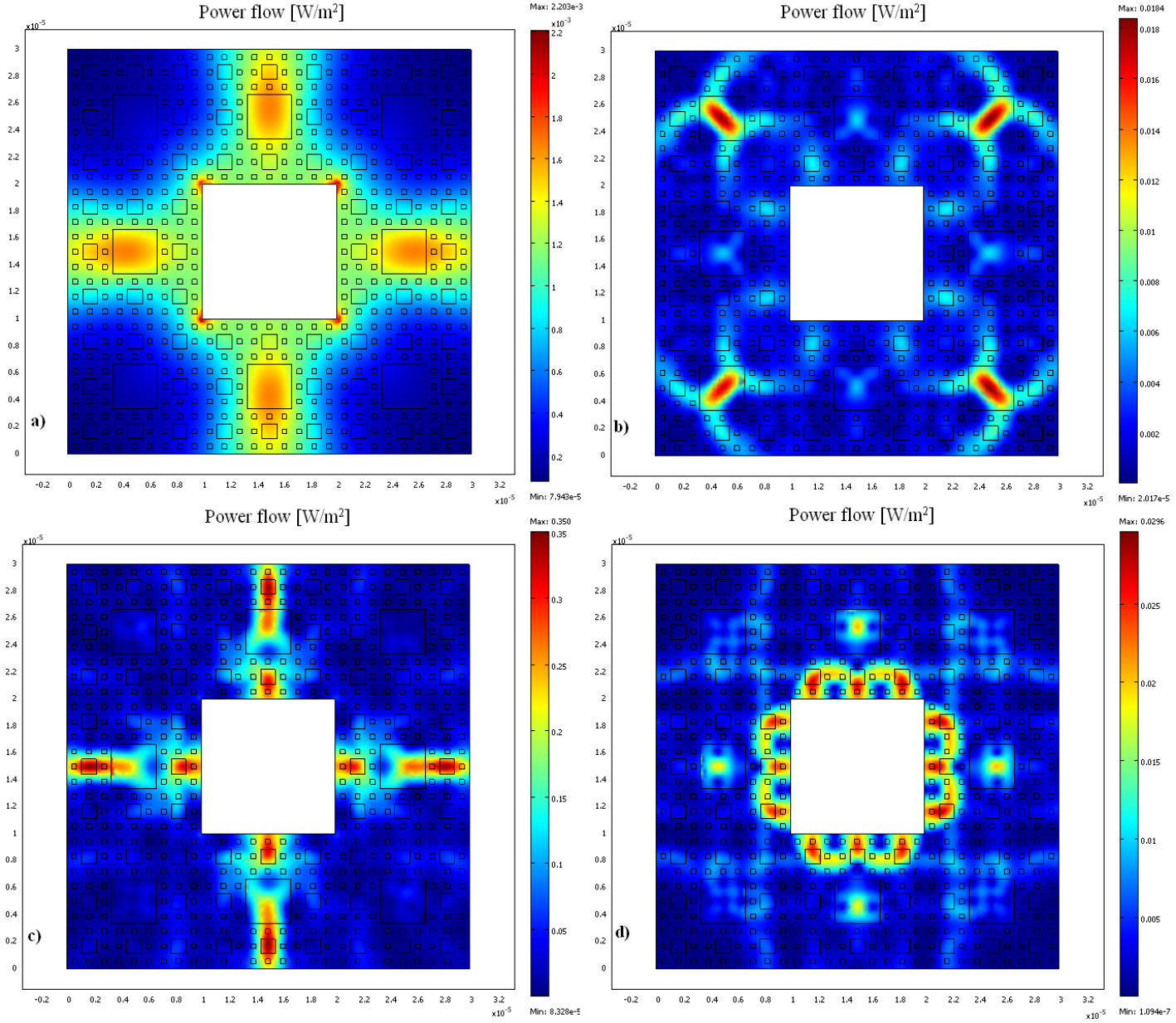


Fig. 2 Power flow patterns on PhC, wavelengths 15μm (a), 55μm (b), 65μm (c), 90μm (d)

Fig. 2,b exhibits case of bad power flow pattern for any practical use in bandpass filtering – power is dispersed on the crystal's body (55μm) because the wavelength is unsuited for given filter size and lattice constant. The Fig. 2,c shows case of the best transmission (65μm) – power again flows in a cross-like pattern, but now with much bigger amplitudes.

Finally, Fig. 2,d illustrates power flows relating mainly to stopband wavelengths $\lambda = 90...200$μm. Wavelength is becoming too big for given filter size and lattice constant. In this picture an interesting phenomenon of "binding" electromagnetic field to the inner edges of PhC can be seen, too. Such a model can be classified as a partial dielectric mirror and

could be investigated separately for creating a novel optical frequencies energy storage device.

For case of TM polarization the same research has been done using (2), but filtering properties of such a system were judged as not significant in comparison to TE polarization.

Also, similar research has been performed for 2D Air/GaAs PhC with same fractal geometry and size – it can be a bandpass filter, too. Low-contrast 2D fractal PhC deserves a special study and it's possible to try different dielectrics and geometries for its implementation.

In Fig. 3 transmission through the center of upper edge of PhC versus wavelengths is shown. From this figure it can be seen that the filter has the best transmission in the narrow range of wavelengths 62 – 69 μm, at the same time suppressing the signals with wavelengths below and above.

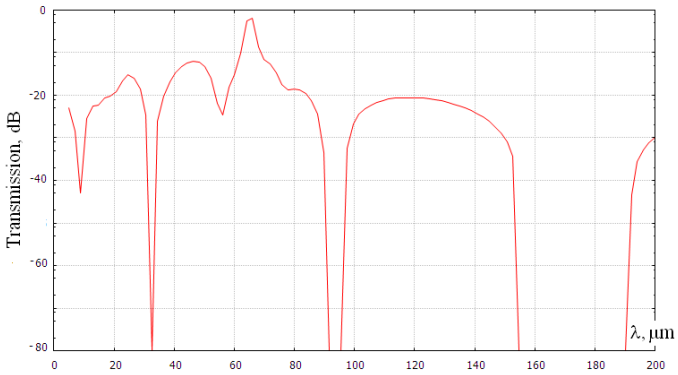


Fig. 3 Filter's transmission at wavelengths 5 - 200 μm

CONCLUSION

A mathematical model of a band-pass filter, based on three-order "Sierpinski carpet" fractal 2D GaAs/Air PhC has been proposed. This model was investigated in the range of wavelengths between 5μm and 200μm. The filter has the best power transmission in a narrow range between 62μm and 69μm. In comparison to [7], where one-order Sierpinski carpet based photonic bandgap structure filter had passband width (at level -10dB) equal to 14,76% of its working frequency range, the filter proposed gives 3,59% relative passband width, which means more than 4 times better filter's frequency selectivity.

It's important, that due to scaling properties of Maxwell's equations [1], [2] it is possible to change size of the filter and its working frequencies in a wide range. This allows a way to stretch greatly the possible area for the filter's application.

REFERENCES

[1] J. D. Joannopoulos, R. D. Meade, J. N. Winn, Photonic crystals: molding flow of light. Princeton University Press, 1995.

[2] M. Skorobogatiy, Jianke Yang, Fundamentals of photonic crystal guiding. Cambridge University Press, 2009.

[3] M. Born, E. Wolf, Principles of Optics: Electromagnetic Theory of Propagation, Interference, and Diffraction of Light (Cambridge University Press, Cambridge, 2003).

[4] B. B. Mandelbrot Les objets fractals: forme, hasard et dimension. – Paris : Flammarion, 2010. [in French].

[5] A. Shahmoon , M. Birenboim , A. Frydman, Z. Zalevsky, Design and Fabrication of 1 × 2 Nanophotonic Switch, Journal of Nanotechnology, vol. 2010, Article ID 953212, 5 pages, 2010. doi:10.1155/2010/953212

[6] A. V. Vishnevsky, Radiation Of Complex Formed Current-Conductive 2D Object : Proceedings of the 4rd World Congress 'Aviation in the XXI-st century. Safety in aviation and space technology', IEEE Microwaves, radar and remote sensing symposium 'MRRS–2011'. – Kiev: NAU, 2011, pp.279-282.

[7] H.W. Liu, X.W.Sun, Z.F.Li, A low-pass filter of wide stopband with a novel multilayer fractal photonic bandgap structure. Active and Passive Elec. Comp., June 2004, Vol. 27, pp. 107–110

Controllable Growth of Copper Fractal Aggregates on Structurally Modified Silicon Surface

Anton Boyko, Sergei Timoshenkov, Dmitry Litmanovich
Dept. of Microelectronics
National Research University of Electronic Technology
Moscow, Russia
ant_nico@mail.ru, spt@miee.ru, msiet21@gmail.com

Dahir Gaev
Dept. of Microelectronics
Kabardino-Balkarian State University
Nalchik, Russia
dahir@mail.ru

Abstract — **Copper layers of highly developed surface have been formed by electrodeposition technique on structurally modified silicon substrate. Silicon patterns were obtained by electrochemical anodizing as well as by wet etching and acted the part of solidification centers at the growth process. The effect of the current density on the morphology of copper fractal aggregates has been investigated. It was found that a change of current density has a significant impact on texture and fractal dimension of the aggregates. Increasing of the current density from 0.25 to 1.5 mA/cm^2 changes form of the fractals from spherical to symmetrically branched (dendritic) and fractal dimension from 2.55 up to 2.8 respectively. All the samples are of uniform quasi-periodical distribution of the aggregates over the surface that is substantially conditioned by properties of the modified surface.**

Keywords—fractals; nanostructuring; advanced materials; porous silicon.

I. INTRODUCTION

Nanostructuring with the use of nonequilibrium growth processes can change the properties of materials in a wide range including via formation of materials are of fractal texture. Fractal structures have a number of unique properties and are promising for creating advanced materials [1], [2]; metal layers with developed surface and high fractal dimension are widely used in different sensors, getters, catalysts, supercapacitors, etc. [3]-[6].

Fractal growth using electrodeposition is elaborated sufficiently [7]-[9], but formation of fractal layers of predictable and reproducible properties stays a quite complex task because the process preferably taking place under nonequilibrium conditions. So the approaches providing controllable growth of fractals are of great interest especially for industrial scale.

In the work we have solved two main tasks to form fractal aggregates. The former was to choose and realize a technique for formation of structural modifiers on silicon surface; the latter was to study an influence of electrochemical deposition conditions on the aggregates texture. The creation of structure modifiers on silicon substrate consists in formation of porous silicon layer and next removing of it. Resulting remains of pore layer act the part of solidification centers or patterns during metal deposition. Preferential growth of nuclei on the tops of irregularities can be explained by the fact that reduction potential is higher on the edge of silicon cathode [10]. Electrochemical deposition of copper has been used to study the growth process; in further researches it is planned to apply various materials providing specific functional properties.

The study was supported by the Ministry of science and education of Russian Federation, project №14.B37.21.1619

II. STRUCTURAL MODIFICATION OF THE SURFACE

A. Formation of Porous Silicon Layer

Formation of porous layer by electrochemical anodizing technique was the first step of patterns creation. The process was carried out in one-chamber electrochemical cell of vertical type using reticular platinum electrode. Porous silicon samples were prepared from n-type (10 ohm·cm) single-crystal (100) oriented silicon wafers with 1.0 cm^2 exposed area. After the wafers were held in 10% aqueous HF solution to remove natural surface oxide, they were rinsed thoroughly with deionized water and placed in the electrochemical cell. Anodizing solution was a mixture of HF (40%) and C_2H_5OH (96%). For the process of porous silicon formation the anodizing at direct current was used, the magnitude of one was maintained within 0.5%. Variation of both the ion current density from 15 to 40 mA/cm^2 and the anodization time in the range of 10-30 min allowed obtaining porous layers with different morphology. The samples texture has been studied by scanning electron microscope Helios Nanolab 650. SEM image of a porous layer formed by 15 minutes anodizing with 30 mA/cm^2 current density is shown in Fig. 1.

B. Removing of Porous Layer

The last step of surface modification has been done by etching of pSi with water solution of caustic soda at room temperature. This incomplete removal of porous layer results in desired nucleation centers on the surface (see Fig. 2).

Placement of nucleation centers on the surface depends on the properties of the original porous structure; patterns are distributed over the surface quasi-periodically with distance between the tops of a few micrometers (see Fig. 2).

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

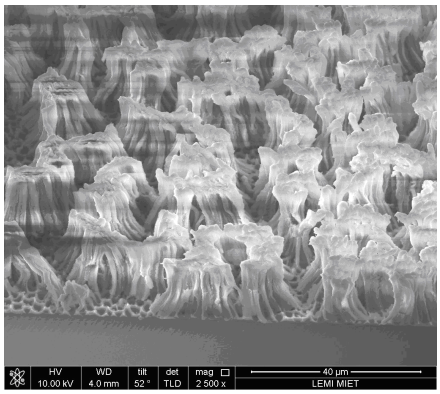


Fig. 1. Porous silicon layer formed by electrochemical anodization.

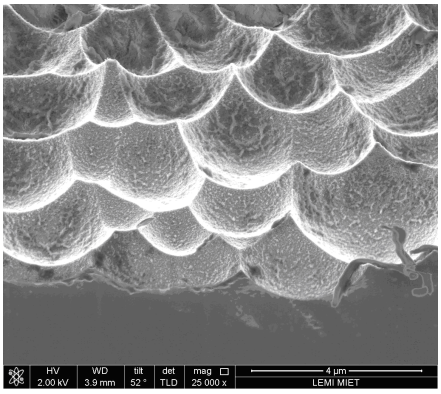


Fig. 2. SEM image of structurally modificated silicon surface

III. Metal Layer Deposition

Electrolytic deposition of copper was carried out from water solution of copper sulfate with concentration of 0.06 mol/l with addition of 2 vol. % of ethanol. The current density took on three values: 0.25, 1 and 1.5 mA/cm^2; those modes will be called like “slow”, “middle” and “rapid” respectively. SEM images of copper fractal aggregates grown on the modificated silicon surface with varying deposition modes are shown on the Fig. 3-5.

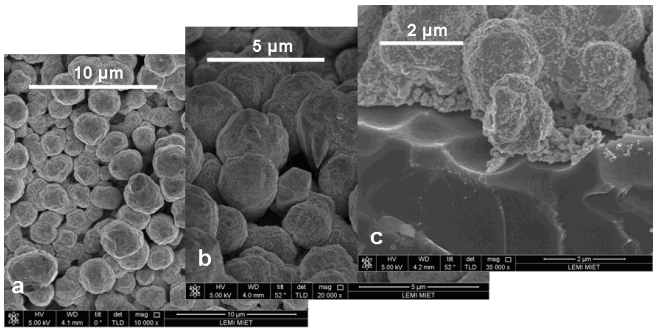


Fig. 3. SEM images of samples obtained at "slow" deposition mode (deposition time is 180 s): (a) – 10000^X; (b) –20000^X; (c) – 35000^X.

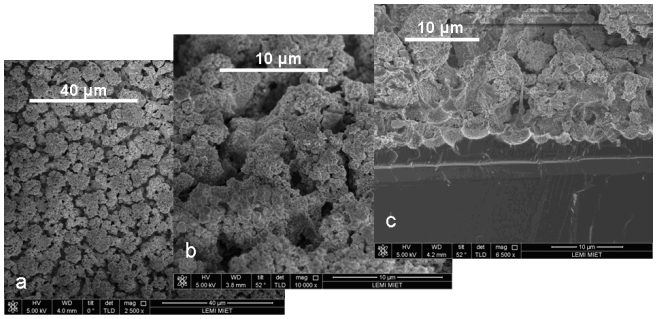


Fig. 4. SEM images of samples obtained at "midlle" deposition mode: (a) – 2500^X, 100 s; (b) – 10000^X, 100 s; (c) – 6500^X, 160 s.

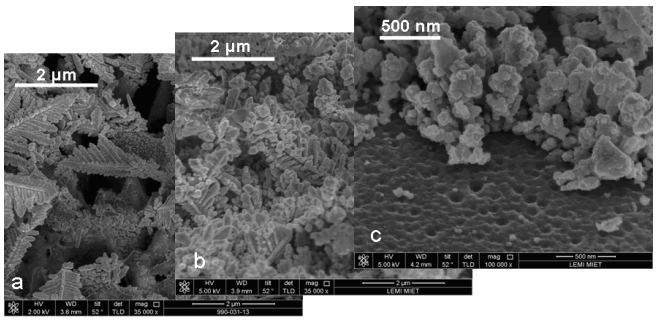


Fig. 5. SEM images of samples obtained at "rapid" deposition mode: (a) – 35000^X, 80 s; (b) – 35000^X, 140 s; (c) – 100000^X, 60 s.

IV. Analysis of Experimental Structures Using Fractal Geometry

The tools of fractal geometry were applied for morphological analysis of the aggregates under investigation; computation of surface fractal dimension allows to make quantitative assessment of the texture features in connection with deposition regimes. Fractal dimension can be calculated in different ways defined by explore tasks as well as by studied objects [11], [12]. We have used an “area-perimeter” method proposed by Mandelbrot is both easy-to-implement and accurate [11]. In line with one the surface imagery were cut by horizontal planes at different levels, in the issue in the plane of the section we get “islands” described by total area S and total length of “coastline” or perimeter P. Logarithmic area-perimeter curve for set of level sections is of linear type in ideal case. Angle of inclination α of the line to abscissa axis is related to fractal dimension D_F of surface by equation:

$$D_F = 2tg\alpha + 1 \qquad (1)$$

Dependence *lnS(lnP)* built using results of processing of 90 sections is shown in Fig. 5,a; the value of the fractal dimension in this case is 2.783. Expressed linear region on the graph testifies marked fractal properties of the surface, the same feature holds for other samples. The value of fractal dimension of the surface in Fig. 3 is 2.576, for the surface in Fig. 4 one is 2.654 respectively.

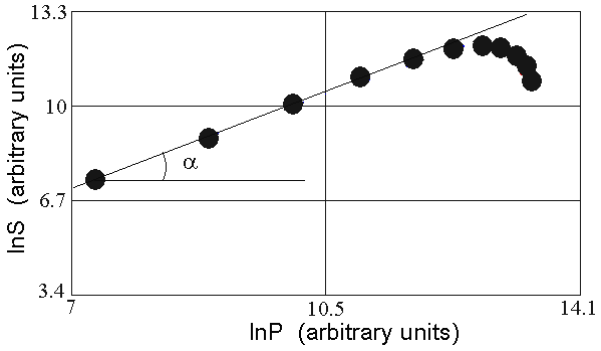


Fig. 6. Dependence *lnP (lnS)* of the sample formed in "rapid" mode.

The current density dependence of fractal dimension of the aggregates is shown in Fig. 7.

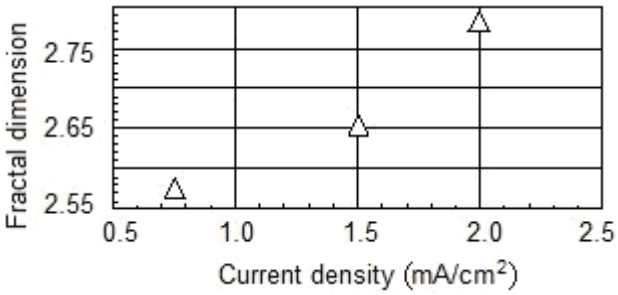


Fig. 7. Current density dependence of fractal dimension of the aggregates.

V. Results and Discussion

The proposed method of structural modification based on the use of porous silicon for making of patterns is promising in the field of advanced materials. Distribution of nucleation centers on the surface is determined by properties of the original porous layer. The exploration of SEM images presented in Fig. 2, 3, and 4,a showed that distribution and size of fractal aggregates correlate with quasi-periodic distribution of the patterns.

Variation of mode of copper deposition enables to get structures of different morphology. At low current density (0.25 mA/cm^2) spherical fractal aggregates, in turn consisting of smaller units, are growing on structured surfaces; thus marked attributes of self-similarity can be observed (see Fig. 3,c). Fractal dimension of surface of the aggregates equal to 2.576 indicates an expressed fractality of the texture. Increasing of current density to 1 mA/cm^2 leads to growth of "coral" structures are of more complex texture and pronounced fractality. Further rise of ion current (1.5 mA/cm^2) gives the growth of "fern" dendritic structures described by fractal dimension of 2.783.

Quantitative estimation of morphological features using fractal analysis showed dependence of fractal dimension of deposition rate of copper (see Fig. 6). The results confirm the assumption that increasing of non-equilibrium of deposition process leads to highly probable growth of fractals.

It can be confidently assumed that decreasing of deposition rate (less than 0.25 mA/cm^2) gives a dense non-porous film with fractal dimension of surface close to the topological one. The fractality will increase and then discontinue to depend on the rate of deposition at high deposition rates (over 1.5 mA/cm^2), the growth process may become uncontrollable in this occasion.

The graph of current density dependence of fractal dimension shown in Fig. 7 is more likely qualitative than shows unique dependence. Exploration of strict relationships goes beyond the purpose of this work and is the subject of further researches because it requires a detailed consideration of stages of growth process, demands study of dependence images scale on fractal dimension and expanding the range of metal deposition rates at last.

In summary, the results have showed that controllable growth of fractal aggregates can be achieved using structural surface modification. The variation of deposition regimes allowed to get structures characterized by different texture, the study exhibits marked dependence between the deposition rate and morphological features of the formed aggregates. The investigation results are useful to produce advanced materials with specific properties.

References

[1] E. Hornbogen, "Fractals in microstructure of metals", Intern. Mat. Rev., 1989, vol. 34, pp. 277–296.

[2] B. Mandelbrot, D. Passoja, A. Paullay, "Fractal character of fracture surfaces of metals", Nature, 1984, vol. 308, pp. 721–722.

[3] Y.-P. Zhao, G.-C. Wang, T.-M. Lu, G. Palasantzas, J. Hosson, "Surface-roughness effect on capacitance and leakage current of an insulating film", Phys. Rev. B.,1999, vol. 60, No. 12, pp. 9157–9164.

[4] O. Hilders, D. Pilo, "On the development of a relation between fractal dimension and impact toughness", Materials characterization, 1997, vol. 38, No. 3, pp. 121–127.

[5] P. Pfeifer, D. Avnir, D. Farin, "Scaling behavior of surface irregularity in the molecular domain: From adsorption studies to fractal catalyst", J.Stat. Phys., 1984, vol. 36, pp. 399–716.

[6] N. Yu, L. Gao, Sh. Zhao, Z.Wang, "Electrodeposited PbO_2 thin film as positive electrode in PbO_2/AC hybrid capacitor", Electrochimica Acta, 54, 2009, pp. 3835–3841.

[7] J.M. Costa, F. Sagués, M. Vilarrasa, "Growth rate of fractal copper electrodeposits: Potential and concentration effects.", Phys Rev A, 43(12), 1991, pp. 7057–7060.

[8] H. Brune, C. Romainczyk, H. Röder, K. Kern, "Mechanism of the transition from fractal to dendritic growth of surface aggregates", Nature 369, 1994, pp. 469–471

[9] L. Bicelli, B. Bozzini, C. Mele, L. D'Urzo, "A review of nanostructural aspects of metal electrodeposition", Int. J. Electrochem. Sci., 3, 2008, pp. 356–408.

[10] M. Belenky, A. Ivanov, "Electrodeposition of metal coatings", Metallurgy, Moscow, 1985, 287.

[11] B. B. Mandelbrot, "The Fractal Geometry of Nature", W. H. Freeman and Company, 1982, 468.

[12] J. Feder, "Fractals", Springer, 1988, 310.

Periodicity Assisted Scattering and Absorption of Light by Finite Layered Gratings of Silver Nanowires

Denys M. Natarov,
Alexander I. Nosich
Laboratory of Micro and Nano-Optics,
Institute of Radio-Physics and
Electronics NASU,
Kharkiv 61085, Ukraine
den.natarov@gmail.com

Marian Marciniak
National Institute of
Telecommunications,
Warsaw 04-894, Poland

Ronan Sauleau
IETR, Universite de Rennes 1,
Rennes Cedex 35042, France

***Abstract* — We consider the two-dimensional (2-D) problem of the H-polarized plane wave scattering by discrete arrays made of finite number of periodically arranged circular cylindrical nanowires. Using the Fourier expansions of the field in local coordinates and addition theorems for cylindrical functions, we perform analytical regularization and obtain a block-type matrix equation. This equation is cast to the Fredholm second-kind form that guarantees, after truncation, the convergence of numerical solution. In computations, we consider the gratings with shifted layers made of silver nanowires in the visible-light range of wavelengths. The scattering and absorption cross-sections and near-field patterns are calculated and the interplay of plasmon and periodicity-induced resonances is studied.**

Keywords—absorption; grating; nanowire; periodicity; plasmon; resonance; silver; scattering

I. INTRODUCTION

In today's nanophotonics, various periodically structured scatterers, such as finite-periodic gratings, arrays or chains of particles and holes in metallic screens (in 3-D) or wires and slots (in 2-D), attract increasingly large attention of researchers [1-7]. This is caused by the amazing effects of extraordinarily large reflection, transmission, emission, and near-field enhancement that have been found in the scattering of light by periodic scatterers. Recently it has been discovered that these phenomena are explained by the existence of so- called grating resonances [4-7] (a.k.a. geometrical, lattice and Bragg resonances). Their wavelengths are just above the Rayleigh wavelengths [8] (period being a multiple of the wavelength) if all elementary scatterers of a grating are excited in the same phase and their size is a fraction of the period. In the wave scattering by infinite gratings, they lead to almost total reflection of the incident field by a thin-dielectric-wire grating in a narrow wavelength band in the either of two alternative polarizations [5]

The other type of resonances is observed only for sub-wavelength noble-metal particles and wires in the mid-infrared and optical bands [9-11]. It is known that small material objects can exhibit resonance behavior at certain frequencies for which the object permittivity is negative and the free-space wavelength is large in comparison to object dimensions.

The latter condition clearly suggests that these resonances are electrostatic in nature. Excitation of plasmons results in powerful enhancement of scattered and absorbed light that is used in the design of optical antennas and biosensors for advanced applications. Plasmon resonances have unique physical property: in the leading terms, their wavelengths depend on the object shape but not on its dimensions. The goal of this paper is extension of this study to 2-D periodically structured silver wire configurations.

II. SCATTERING PROBLEM

Consider a finite grating of M parallel wires illuminated by an H-polarized plane wave as shown in Fig. 1.

The wires are assumed to be identical infinite circular cylinders, each having radius a and complex relative dielectric permittivity ε. In a 2-D problem, one has to find a scalar function $H_z(\vec{r})$ that is the scattered-field *z*-component. It must satisfy the Helmholtz equation with corresponding wavenumbers inside and outside the cylinders, the tangential field components continuity conditions, the radiation condition, and the condition of the local power finiteness.

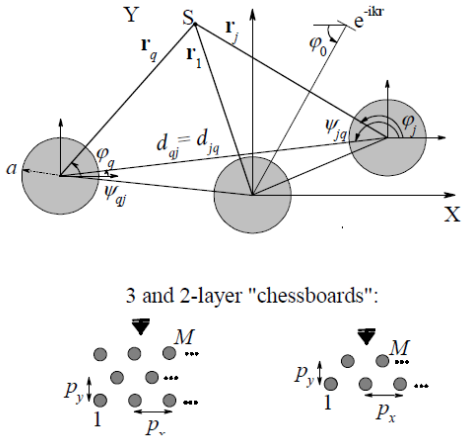


Fig. 1. Generic problem geometry and corresponding notations: a plane wave is incident on a finite number of identical parallel circular cylinders or wires (upper panel). Two and three layer collections of finite periodic gratings of wires (lower panel).

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

The full-wave solution can be obtained similarly to [12-15] by expanding the field function in terms of Fourier series in the azimuth exponents in local coordinates (Fig. 1), using addition theorems for cylindrical functions, and applying the boundary conditions on the surface of all M wires. This leads to an infinite $M \times M$ block-type matrix equation where each block is infinite (see equations (13)-(15) in [6]). When doing this we have found that the matrix equations used in the papers [12-15] did not provide the convergence of numerical solutions. Here, we imply the convergence in mathematical sense, as a possibility of minimizing the error of computations by solving progressively larger matrices. It should be noted that the convergence is frequently mixed up with accuracy. Many divergent numerical schemes are able to provide a few first digits correctly however fail when a better accuracy is requested. Example of such behavior can be seen in Figs. 3 and 4 of [15] where the ultimate achievable accuracy is only 10^{-2}. Fortunately, this defect of the earlier papers can be corrected by re-scaling the unknown coefficients – for details see equations (1)-(6) in [6]. The obtained in such a way matrix equation is a block-type Fredholm second kind equation. It is only in this case that the solution of equation with each block truncated to finite order N converges to exact solution if $N \to \infty$. The results presented below have been computed with $N = 4$ to provide 3 correct digits in the far-field characteristics of the sparse gratings of silver wires with radii $a \leq 60$ nm and periods $p \geq 200$ nm. Note that denser gratings need larger values of N to achieve the same accuracy because of stronger interaction between wires.

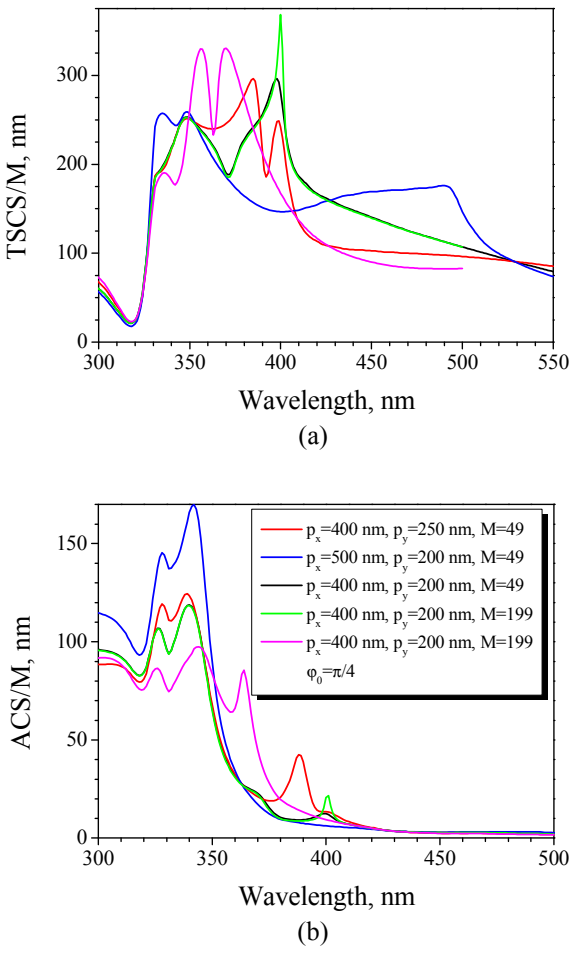


Fig. 2. TSCS (a) and ACS (b) per wire as a function of wavelength for the H-wave scattering by the different 2-layer gratings of silver nanowires with radii a = 50 nm.

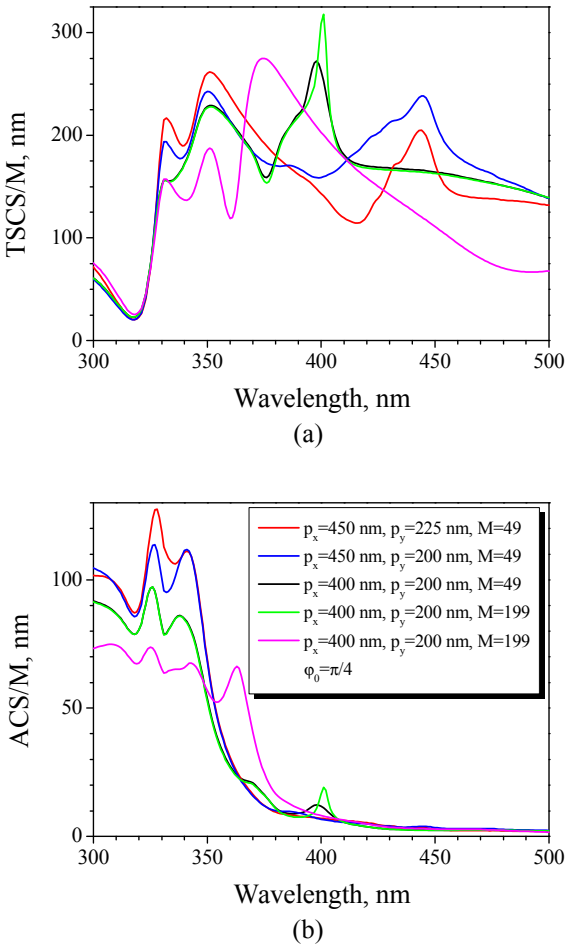


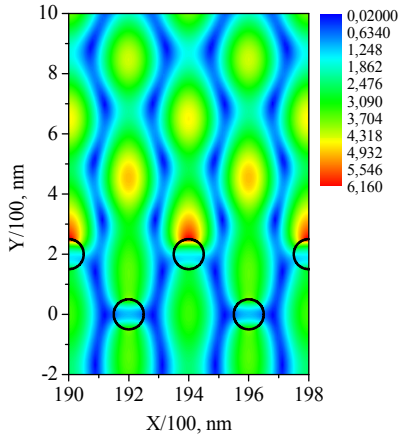
Fig. 3. The same as in Fig. 2 for the different 3-layer gratings.

III. Results of Numerical Simulation

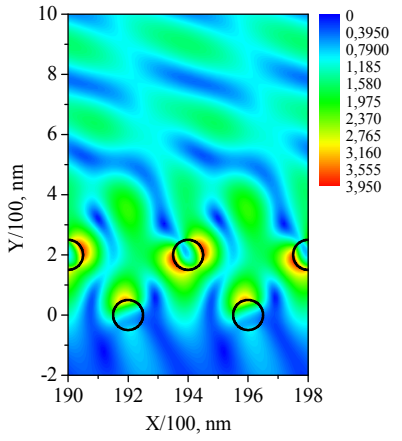
In this study, we have considered sparse ($\rho - 2a > a$) finite gratings of sub-wavelength silver nanowires: 2 and 3 stacked linear chains, where the second chain is shifted by the half of period p_x along the x-axis with respect to the first (and the third) chain and also reduced by one period. Such gratings can be equally considered as finite-thickness photonic-crystal slabs with triangular cells. To characterize the far-field scattering and the absorption properties of the considered gratings we have used the frequency dependences of the total scattering (TSCS) and absorption (ACS) cross sections and have also calculated the field patterns in the near zone. In Figs. 2 and 3, the wavelength varies from 300 nm to 550 nm and the

complex-valued dielectric function of silver has been borrowed from [14].

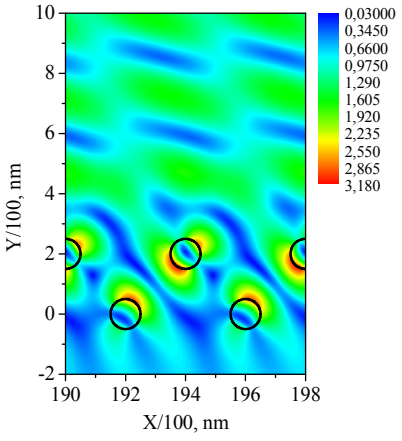
In Figs. 2 and 3, the wavelength dependences are presented of the per-wire values of TSCS and ACS for 2 and 3-layer gratings shown in Fig. 1 made of wires with radii of 50 nm, for different values of the periods p_x and p_y, the numbers N, and the angles of incidence. On each plot one can see 2 or 3 resonances whose Q-factors vary. They belong to two types:



(a)



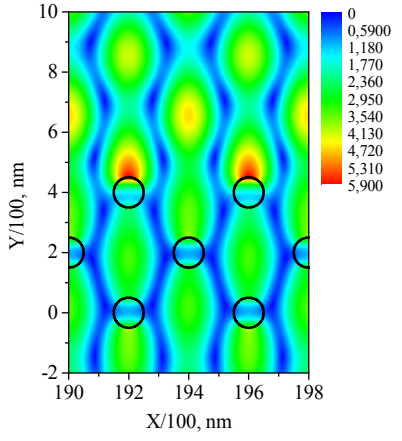
(b)



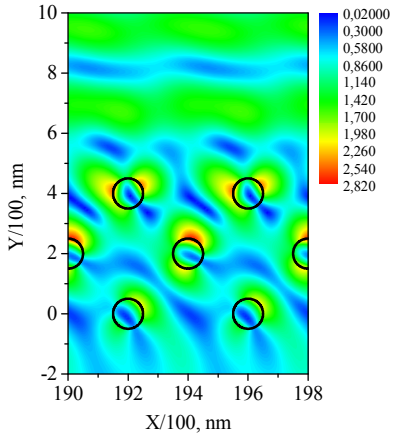
(c)

Fig. 4. Near-field amplitude patterns (central part of gratings) at the corresponding grating-type resonance wavelengths for the 2-layer gratings from 199 silver nanowires with parameters p_x = 400 nm, p_y = 200 nm, a = 50 nm; (a) λ = 400.1 nm, broadside incidence, (b) λ = 363 nm, minimum in TSCS, $\varphi_0 = \pi/4$, (c) λ = 370 nm, $\varphi_0 = \pi/4$.

localized surface plasmon resonances (P–resonances) whose



(a)



(b)

Fig. 5. Near-field amplitude patterns (central part of gratings) at the corresponding grating-type resonance wavelengths for the 3-layer gratings from 299 silver nanowires with parameters p_x = 400 nm, p_y = 200 nm, a = 50 nm; (a) λ = 400.8 nm, broadside incidence, (b) λ = 374.6 nm, $\varphi_0 = \pi/4$.

main resonance wavelength is always near to 350 nm for circular silver wires, and grating resonances (G–resonances), whose resonance wavelengths depend on the grating period and the angle of incidence. For the investigated gratings one can observe several P-resonances because of both dipole-type and higher-order plasmon modes excitation on each wire. Similar behavior becomes better observed in the scattering and absorption characteristics of denser gratings. Identification of the G-resonances is best on the plots for the gratings with identical parameters but with different number of wires: see, for instance, black and blue curves in Figs. 2 and 3. One can see that at the broadside (normal) incidence TSCS and ACS values dramatically increase in the vicinity of the wavelength equal to p_x if the number of periods in grating gets greater. However, if the wavelength of the G-resonance coincides with the P-resonance wavelength of 350 nm, like for the magenta curve in Fig. 2 (a) that corresponds to the inclined incidence under $\varphi_0 = \pi/4$, then the G-resonance splits the broader P-resonance and a narrow-band effect of suppressed scattering

(the optically-induced transparency) appears.

In Figs. 4 and 5, we present the in-resonance near-field patterns of 2 and 3-layer gratings, respectively. In Fig. 4 (a) presented is near-field pattern for the several central periods of the grating of 199 wires with periods p_x = 400 nm and p_y=200 nm under the broadside incidence at the G-resonance wavelength of 400.1 nm. The field pattern demonstrates typical behavior for the G-resonances in the H-polarization: the standing wave with strip-like bright spots along the y-axis. In Figs. 4 (b) and 4 (c), presented are the near-field patterns for the same gratings, but the incident H-polarized plane wave comes under the angle $\pi/4$. Fig. 4 (b) demonstrates the field at the minimum of the corresponding TSCS curve (λ = 363 nm), while Fig. 4 (c) – at the maximum (λ =370 nm). As mentioned above, the minimum of TSCS in this case corresponds to the G-resonance that is confirmed by the field pattern: the field amplitude is higher in Fig. 4 (b). In Figs. 5 (a) and 5 (b), we show the near-field patterns for the 3-layer gratings under the broadside and the $\pi/4$ incidence angles, respectively, at the G-resonance wavelengths of 400.8 nm and 374.6 nm, respectively. Note that addition of the third layer slightly red-shifts the G-resonance (to a higher wavelength) and the maximum field amplitudes in the case of the broadside incidence are approximately 2 times larger than in the case of the $\pi/4$ incidence.

IV. Conclusion

We have presented results of accurate calculations with a convergent numerical algorithm of the scattering and absorption characteristics for the 2 and 3-layer periodically structured ensembles of sparsely placed silver nanowires. We have investigated two types of resonances: plasmon and grating-type and computed the components of the extinction cross-section and near-field patterns of such layered gratings. An interesting observation is a minimum on the TSCS spectral dependence due to the G-resonance, if the wavelengths of the broader P-resonance and the sharper G-resonance coincide.

Acknowledgment

This work was supported, in part, by the National Academy of Sciences of Ukraine via the State Target Program "Nanotechnologies and Nanomaterials." The first author is also grateful to the International Visegrad Fund for a Ph.D. scholarship and to the Rennes-Metropole Programme for the researcher mobility grant.

References

[1] S. Zou and G. C. Schatz, "Narrow plasmonic/photonic extinction and scattering lineshapes for one and two-dimensional silver nanoparticle arrays," J. Chem. Phys., vol. 121, pp. 12606-7, 2004.

[2] B. Auguie and W. L. Barnes, "Collective resonances in gold nanoparticle arrays," Phys. Rev. Lett., vol. 101, pp. 143902-4, 2008.

[3] M. Laroche, S. Albaladejo, R. Gomez-Medina, and J. J. Saenz, "Tuning the optical response of nanocylinder arrays: an analytical study," Phys. Rev. B, vol. 74, no. 9, pp. 245422-10, 2006.

[4] F. J. G. Garcia de Abajo, "Colloquium: Light scattering by particle and hole arrays," Rev. Mod. Phys., vol. 79, no. 4, pp. 1267-1289, 2007.

[5] V. O. Byelobrov, J. Ctyroky, T. M. Benson, R. Sauleau, A. Altintas, and A. I. Nosich, "Low-threshold lasing modes of infinite periodic chain of quantum wires," Optics Lett., vol. 35, no 21, pp. 3634-3636, 2010.

[6] D. M. Natarov, V. O. Byelobrov, R. Sauleau, T. M. Benson, and A. I. Nosich, "Periodicity-induced effects in the scattering and absorption of light by infinite and finite gratings of circular silver nanowires," Optics Exp., vol. 19, no 22, pp. 22176-22190, 2011.

[7] D. M. Natarov, R. Sauleau, and A. I. Nosich, "Periodicity-enhanced plasmon resonances in the scattering of light by sparse finite gratings of circular silver nanowires," IEEE Photonics Technology Lett., vol. 24, no1, pp. 43-45, 2012

[8] Lord Rayleigh, "On the dynamical theory of gratings," Proc. Royal Soc. London A, vol. 79, pp. 399-416, 1907.

[9] J. P. Kottman and O. J. F. Martin, "Plasmon resonant coupling in metallic nanowires," Opt. Exp., vol. 8, pp. 655–663, 2001.

[10] O. J. F. Martin, "Plasmon resonances in nanowires with non-regular cross-section," in J. Tominaga and D. P. Tsai (Eds.), Optical Nano-technologies, Topics Appl. Phys., Springer, vol. 88, pp. 183–210, 2003.

[11] D. R. Fredkin and I. Mayergoyz, "Resonant behavior of dielectric objects (electrostatic resonances)," Phys. Rev. Lett., vol. 91, pp. 3902-3905, 2003.

[12] V. Twersky, "On a multiple scattering theory of the finite grating and the Wood anomalies," J. Appl. Phys., vol. 23, no 10, pp. 1099-1118, 1952.

[13] H. A. Ragheb and M. Hamid, "Scattering by N parallel conducting circular cylinders," Int. J. Electronics, vol. 59, pp. 407-421, 1985.

[14] D. Felbacq, G. Tayeb, and D. Maystre, "Scattering by a random set of parallel cylinders," J. Opt. Soc. Am. A, vol. 11, pp. 2526–2538, 1994.

[15] X. Antoine, C. Geuzaine, and K. Ramdani, "Computational methods for multiple scattering at high frequency with applications to periodic structure calculations," in M. Ehrhardt (Ed.), Wave Propagation in Periodic Media, Progress in Computational Physics, Bentham Science Publ., pp. 73-107, 2010.

[16] P. B. Johnson and R. W. Christy, "Optical constants of the noble metals," Phys. Rev. B, vol. 6, pp. 4370–4378, 1972.

Modeling of Electronic Radiation Influence to Transistors

A.A.Tuyakbayev, M.K.Baizhumanov
Academy of the Civil Aviation
Almaty, Republic Kazakhstan
altai_aga@mail.ru

D.A.Tuyakbaev
"GAZI" University
Ankara, Turkey
t.dak@mail.ru

***Abstract* − The paper presents a single-level model to identify the constant of the radiation changes in the lifetime of minority carriers in the base of n-p-n transistors, DLTS spectra base of n-p-n transistors and describes the results.**

Keywords* − *radiation effects, bipolar transistors, lifetime, recombi-nation component, electron irradiation, DLTS−spectra, vacancy, divacancy, radiation center.

I. INTRODUCTION

The irradiation of radiation effects in semiconductor devices lead to the degradation of almost all parameters of bipolar transistors. Of these the most highly variable static gain of the base current. These changes are caused by the fact that the irradiation leads to a change of the lifetime, the concentration and mobility of charge carriers in different areas of semiconductor devices [1, 2].

II. MODELING

In the silicon doping levels corresponding to the base region of the transistors, the most highly variable lifetime of minority carriers. For this you can use the formula

$$\frac{1}{\tau_{\phi}} = \frac{1}{\tau_o} + K_{\tau} * \Phi \ , \qquad (1)$$

where Kτ - proportionality coefficient determining the change rate in the lifetime of minority carriers in irradiation (the constant of radiation change in the lifetime of minority carriers);

Φ - integrated flux of radiation;

τ_o - the initial value of the lifetime of the charge carriers in any given area of the structure.

In this formula it is difficult to determine the value of the constant of radiative changes in the charge carriers' lifetime during irradiation. For this purpose, during the neutron irradiation, Messenger's two-level model can be used [1]. However, to simulate the effect of irradiation by fast electrons, this model cannot be used. It should be noted that there is currently no well-known model to simulate the effect of electron irradiation on the transistors.

Experimental studies of DLTS spectra [2] have been taken for the base region irradiated NPN transistors (Fig. 1), led to the conclusion that the recombination in the base region occur mainly through the center of H (0.45) (vacancy-boron), which is closest to the middle of the band gap of silicon and, consequently, increasing the concentration of the center leads to a decrease in the lifetime of minority carriers in the base, which is the reason for the increase of the recombination component of the base current of transistors. This, in turn, makes a significant contribution to reducing the gain of the transistor under irradiation. From all this it can be concluded that in order to model the effect of electron irradiation on the transistors is enough to create a mathematical model in which all the calculations would be based on the determination of changes in the concentration, namely, the center of H (0.45). Analysis Messenger's two-level model for the neutron irradiation and well-known model of Shockley-Read, allowing to determine the rate of recombination processes in the base NPN transistors, depending on the doping level and the excess charge carriers, allowing a single-level model to simulate the effect of electron irradiation on the NPN transistors [3]:

$$K_{\tau} = \frac{1}{\frac{1}{C R_p}\left(\frac{n_0 + n_e + \Delta n}{n_0 + p_0 + \Delta n}\right) + \frac{1}{C R_n}\left(\frac{p_0 + p_e + \Delta n}{n_0 + p_0 + \Delta n}\right)} \qquad (2)$$

where

$$n_e = N_c * \exp\left(-\frac{\varphi_c - \varphi_e}{\varphi_T}\right) , \qquad (3)$$

$$p_e = N_v * \exp\left(-\frac{\varphi_e - \varphi_V}{\varphi_T}\right) , \qquad (4)$$

$$R_p = \sigma_p * \sqrt{3 * K * T / m_p} \ , \qquad (5)$$

$$R_n = \sigma_n * \sqrt{3 * K * T / m_n} \ , \qquad (6)$$

N_c , N_v - the density of quantum states;

φ_c – the width of the restricted zone of semiconductor material;

m_n , m_p - effective mass of electrons and the hole;

K - Boltzmann constant;

σ_n, σ_p - effective cross section capture of the electron and the hole;

C - the rate of introduction of the recombination center.

Parameter φ_e in this expression corresponds in our case, to the level of H (0.45), ie φ_e = 0,45 eV. The latter is due to the fact that the experimental studies of DLTS base of n-p-n

transistors show, that center of H (0.45) (vacancy-boron) is nearest to the middle of the restricted zone and therefore an increase in the concentration, of this center particularly, leads to a decrease in the lifetime of minority carriers in the base.

In the expression (2) it is difficult to model the dependence of the rate of introduction of the radiation center C, on the doping level of the material, on the energy of the fast electrons, and consider repeated exposure.

The experimental data for the specified parameter p-type silicon suggests using the following formula:

$$C=K_1 * C_o * \lg(N_A/N_{Ao}) , \quad (7)$$

$$C_o=K_2 * \ln(E_e - 1) , \quad (8)$$

where

N_{Ao} - reference concentration of boron impurities in silicon, such as 10^{14} см$^{-3}$;

K_1 , K_2 - aspect ratios, determined from experimental data for the p-Si, $K_1 = 3$, $K_2 = 3{,}4 * 10^{-3}$;

N_A - impurity concentration of boron in silicon in the tested area of the device.

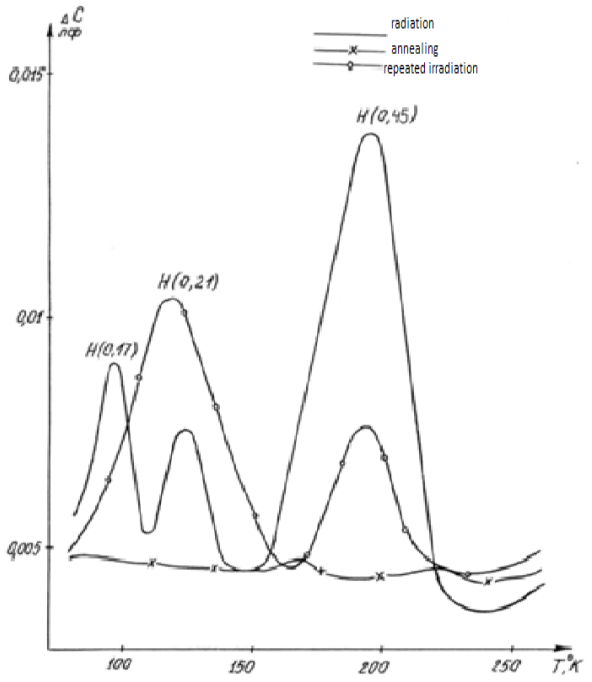


Fig. 1. DLTS - spectra in the base of transistors, irradiated by electrons with an energy of 5 MeV and flow 10^{15}e/sm^2.
Center H (0,45) – is a complex vacancy + atom boron
Center H (0,21) – is a complex divacancy + atom boron.

It should be noted that the above mentioned formulas are valid for irradiation of devices with fast electrons in normal conditions, ie, at room temperature. Figure 2 shows a family of curves obtained with these formulas, depending on the rate of introduction of the radiation center of H (0.45) on the electron energy E , and the boron concentration N_A . The figure shows a satisfactory agreement between the calculated and experimental data.

Using equation (2) the values of the constants of the radiation changes of the carrier lifetime, depending on the energy of fast electrons and boron concentrations NA were calculated. These data are shown in Figure 3.

The values of constants K_τ were used into the formula

$$h_{21\Theta\Phi} = \frac{1}{\frac{1}{h_{21\Theta}} + \frac{1}{2\pi f_T} K_\tau \Phi}$$

The values of the coefficients of the current amplification of base transistors $h_{21э\phi}$ were calculated. These are shown in Figure 3 and are in good agreement with experimental data.

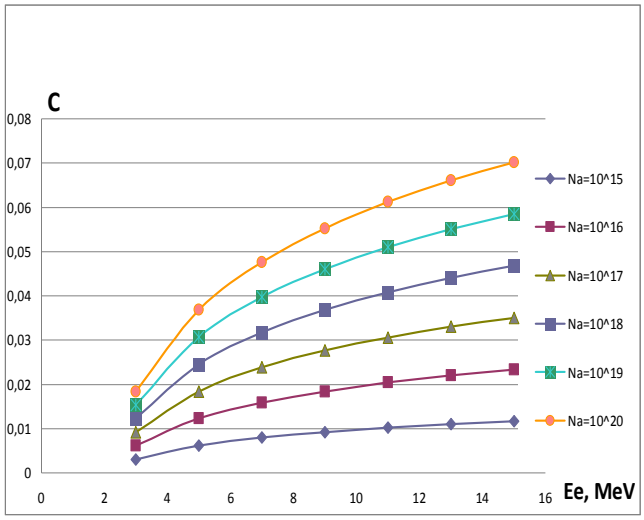


Fig. 2. Calculated dependencies of the rate of introduction of the radiation center of H(0.45) on the energy of fast electrons and the degree of doping of the material.

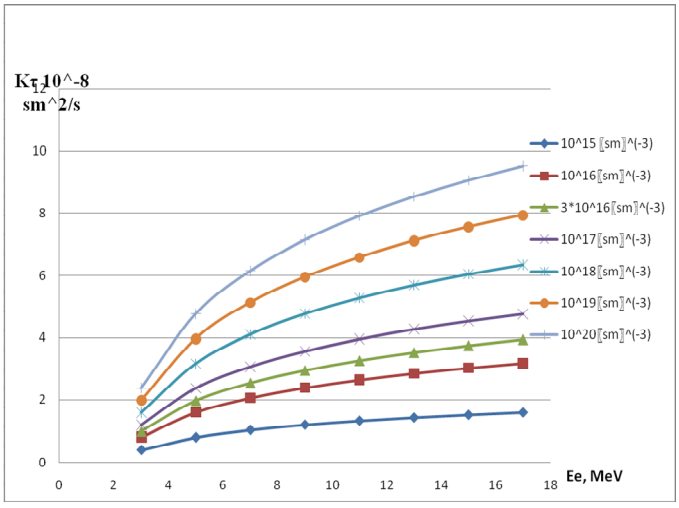


Fig. 3. Calculated dependencies of the radiation changes of the carrier lifetime on the energy of fast electrons and the degree of doping of the material.

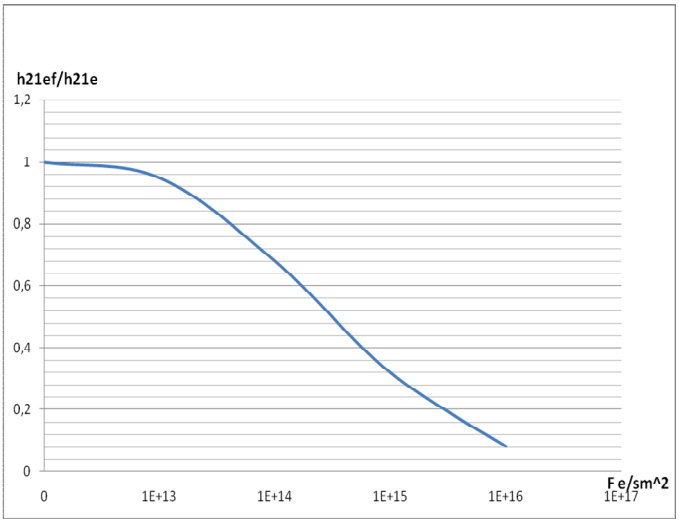


Fig. 4. Calculated dependence of the relative decrease in the gain of silicon bipolar transistor KT 315 under electron irradiation (E_e = 5 MeV).

In calculating the radiation resistance of transistors produced using radiation-thermal processes, the need to obtain the velocity of the centers of H (0.45) reduce to three times, as is evident from the data on DLTS [2] (Figure 4), and the rest calculation method is the same. The results of calculation of radiation resistance of these transistors are shown in Figure 5, and they are in good agreement with experimental data.

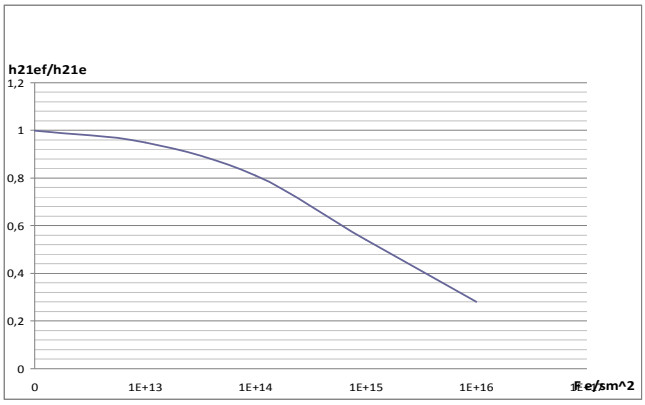


Fig. 5. Calculated dependence of the relative decrease in the gain of silicon bipolar transistors with dimensions KT 315, manufactured using radiation-thermal processes under electron irradiation (E_e = 5 meV).

Radiation centers in semiconductors can act not only as recombination centers, but also as traps and scattering centers [4], which effectively sucked the carriers out of the process of electromigration. As a result of this effect in silicon irradiated with increasing electron flux increases the electrical resistivity ρ. As the electron beam size ρ tends to the value ρi to own material.

For small streams of radiation in silicon use most simple form of dependence ρ (Φ) [4]:

$$\rho\phi = \rho o \exp(K\rho * \Phi) \qquad (9)$$

or

$$\rho\phi = \rho o (1 + K\rho * \Phi) \qquad K\rho * \Phi < 1 , \qquad (10)$$

where Kρ - constant characterizing the growth rate ρ during irradiation.

The physical meaning of the constant Kρ is given by:

$$K\rho = 1/no * \Delta n/\Phi , \qquad (11)$$

where Δn/Φ - the rate of removal of the majority carriers during irradiation.

Constant K_ρ , as well as K_τ , determined by the properties of the initial samples and the irradiation conditions. However, with the "deep" and the introduction of irradiated large number of defects begins to show decreased mobility of the majority carriers.

Thus, in silicon irradiated along with changes of the lifetime of minority carriers there is a change in the conductivity of the active regions of semiconductor devices.

III. CONCLUSION

The single-level model for definition of a constant of radiating change of lifetime of non-basic carriers in base n-p-n transistors and DLTS spectra of base n-p-n transistors have been resulted.

Experimental data show that loss in mobility plays a smaller role in growth ρ at a silicon irradiation, compare to capture of charge carriers by radiation centers. However, at a "deep" irradiation and introduction of a considerable quantity of defects the loss in mobility of the main carriers starts to be significant.

Thus, in silicon irradiated along with changes of the lifetime of minority carriers there is a change in the conductivity of the active regions of semiconductor devices.

REFERENCES

[1] V.S. Vavilov, N.A. Ukhin, "Radiation effects in semiconductors and semi-conductor instruments." Atomizdat, Moscow, 1969. (in Russian)

[2] Gorjunov N.N., Ladygin E.A., Tujakbaev A.A."Research of kinetics of accumulation and annealing of radiation defects in bipolar transistor structures." News of AN KazSSR. A series physical and mathematical, Alma-Ata, 1987, №2

[3] The Fifth World Congress "Aviation in the XXI-st Century", "Safety in Aviation and Space Technologies". A.A.Tuyakbayev, M.K.Baizhumanov, D.A.Tuyakbaev. "Single-level Model for Simulation of the Effect of Electron Irradiation on Transistors". Kyiv, Ukraine, September 25-27, 2012, Volume 2, page 3.7.76-3.7.79.

[4] "Questions of radiation technology of semiconductors. Under the editorship of Smirnova of h.p." – Novosibirsk: the Science, 1980.

Excitation Processes Modeling in Two-Layer Graphene

E.G. Aznakayev, D.E. Aznakayeva
Dept. of Electronics,
National Aviation University,
Kiev, Ukraine
aznakayev@nau.edu.ua

***Abstract* - Absorption vibronic spectra are investigated for molecular crystals with two molecules in an elementary cell that corresponds to two-layer graphene. The absorption vibronic spectra, their intensities, frequencies and polarizations are calculated for molecular crystals containing two molecules in an elementary cell or two-layer graphene.**

Keywords: two-layer graphene; modeling; excitation processes

I. INTRODUCTION

Exciting achievements for creation a new perspective nanomaterial graphene for nanoelectronics were made by outstanding scientists Novoselov, K. S. and Geim, A. K. from the Manchester University (Nobel Prize in Physics 2010) [1].

Graphene is a perfectly two-dimensional single-atom thin membrane. It has excellent transport properties, novel magnetic characteristics, and low spin-orbit coupling. Graphene and its nanostructures may have potential applications in spintronics, photonics, plasmonics and electronics. Electrical characteristics of graphene are linked to its structure. It is reported n-type graphene-based field-effect transistor fabricated at room temperature. Researching graphene nanoelectronics for a post-silicon world is presented in many papers [2-7]. For introduction to practical electronics technology the developing device-oriented theory for graphene device design is required.

The overwelming majority of the bands in the absorption spectrum correspond to simultaneous emergence of the light-induced electron and intramolecular vibrational excitation (vibronic bands).

The vibronic state theory for molecular crystals was firstly developed by A.S. Davydov on the assumption of connection of electronic and intramolecular oscillatory excitations. In this model the vibronic states are described by energy spectrum of a quasi particle - vibron corresponding to the molecular vibronic excitation transportation along the crystal.

Vibronic spectra of molecular crystals are investigated for a case of a combination of electronic excitation with intramolecular oscillation in the case of one molecule per unit cell [8]. Such model allows to describe as one-phonon, as many-phonon vibronic excitations.

The Born-Oppenheimer approximation is the assumption that the electronic motion and the nuclear motion in molecules can be separated. This adiabatic principle allows us to separate the nuclear and electronic motion, leaving an electron-phonon interaction. Size of the nucleus displacement from equilibrium state is small. So, it is possible to spread out the adiabatic potential in series under this amount. Optical vibronic spectra of complex molecules are well described in harmonic approximation, and adiabatic potential dependence from value of nucleus displacement in *f*-th electronic state looks like potential:

$$W^f(R) = W^f\left(R_0^f\right) + \frac{1}{2}\left(\frac{\partial^2 W^f}{\partial R^2}\right)_{R=R_0^f} \cdot (R - R_0^f)^2,$$

where R_0^f is an equilibrium position of nuclear configuration in the *f*-th electronic state.

Interaction between electronic and oscillation motions in the ground and *f-th* electronic states is defined by the operator of an electron-oscillation interaction:

$$\Delta W(q) = \sqrt{2 \cdot \nu^f \cdot \varepsilon_{FC} \cdot q} + \frac{1}{2} \cdot \Delta\nu q^2, \quad (1)$$

where $q = R \cdot \sqrt{M \cdot \nu}$ is dimensionless displacement of the nucleus from equilibrium position; $\varepsilon_{FC} = \frac{1}{2} \cdot \nu^f \cdot (q_0^f)^2$ is a minimum value of potential energy displacement as a result of equilibrium position displacement under electronic excitation; ν^f is frequency of oscillations; q_0^f is a coordinate of molecule's equilibrium state in *f*-th electronic state; equilibrium state of molecule in the ground state is a reference mark ($q_0^f = 0$), $\Delta\nu = \nu^f - \nu^0$ is the defect of oscillation frequency.

The first term in the right part of (1) characterizes the energy distortion of a molecule as result of nucleus equilibrium position changing at electronic excitation. This energy sometimes is called a molecule's dispersion energy and designated as ε_d. Second term in right part of (1) characterizes the changing of frequency oscillation.

There is no the equilibrium position displacement at excitation of vibronic transition for symmetric oscillations and interaction between electronic and oscillation motions is defined only by its frequency changing. The displacement of equilibrium position is characterized mainly for non-symmetric oscillation, but its frequency also changes at electronic excitation also.

II. STATEMENT OF THE PROBLEM

In this paper, based on the method proposed in [8], the vibronic spectra of crystals for a case of a combination of electronic excitation with asymmetrical intramolecular oscillation for a case of two molecules in the unit cell are investigated, which corresponds to the model of two-layer graphene.

Model: molecular crystal with two strongly fixed molecules in each unit cell; only ground electronic state $f = 0$ and one electronic excitation state $f = 1$ of a molecule and only one oscillation mode are considered.

In this paper we investigate the resonant interaction of the molecules associated with the electronic excitation exchanges, and the resonant interaction of molecules associated with the phonon exchanges at localization of electronic excitations.

Let's write down the Hamiltonian for our model in Heitler-London approximation:

$$H_0 = \Omega_0 \cdot \sum_{\vec{n},\alpha} b^+_{\vec{n},\alpha} \cdot b_{\vec{n},\alpha} + \sum_{\vec{n},\alpha}[\mathcal{E} + \Delta\Omega \cdot (b^+_{\vec{n},\alpha} \cdot b_{\vec{n},\alpha} + \tfrac{1}{2})] \cdot B^+_{\vec{n},\alpha} B_{\vec{n},\alpha} + H_{res.} \quad (2)$$

where

$$H_{res} = \sum\nolimits'_{\substack{\vec{n},\alpha;\\ \vec{m},\beta}} M^{(1)}_{\substack{\vec{n},\alpha;\\ \vec{m},\beta}} \cdot \left(b^+_{\vec{n}\alpha} b_{\vec{m}\beta} + b^+_{\vec{m}\beta} b_{\vec{n}\alpha}\right) \cdot B^+_{\vec{n}\alpha} B_{\vec{m}\beta} + \sum\nolimits'_{\substack{\vec{n},\alpha;\\ \vec{m},\beta}} D^{(1)}_{\substack{\vec{n},\alpha;\\ \vec{m},\beta}} \cdot (b^+_{\vec{n}\alpha} b_{\vec{m}\beta} + b^+_{\vec{m}\beta} b_{\vec{n}\alpha}) \cdot B^+_{\vec{n}\alpha} B^+_{\vec{n}\alpha};$$

$$\mathcal{E} = \Delta E + D^{(0)}, \quad .$$

We consider that all molecules are of one sort. In this Hamiltonian the deviation from Condon approximation minus the constant term is considered:

$$\mathcal{E}^0 = \mathcal{E}^{0,0} + W^{(0)};$$

where $W^{(0)}$ is a total interaction energy of crystal's molecules corresponding to the basic state of molecules.

$$W^{(0)} \equiv \sum\nolimits'_{\vec{n},\alpha;\vec{m},\beta} W^{(0)}_{\vec{n}\alpha,\vec{m}\beta} \equiv \tfrac{1}{2} \cdot \sum\nolimits'_{\vec{n},\alpha;\vec{m},\beta} \int \left|\Psi^0_{\vec{m}\beta}\right|^2 \cdot W_{\vec{n}\alpha,\vec{m}\beta} \cdot \left|\Psi^0_{\vec{n}\alpha}\right|^2 \cdot dr_{\vec{m}\beta} \cdot dr_{\vec{n}\alpha}$$

$\mathcal{E}^{0,0}$ is the total energy of molecule in vacuum state; $\Delta E = E_1 - E_0$ is a difference of minimal values of the excited and ground electronic terms of molecule in crystal; Ω_0 – oscillation frequency of molecule's nucleus in ground electronic state of molecule. $\Delta\Omega \equiv \Omega_1 - \Omega_0$ is the frequency difference of the nuclear vibrations in the ground Ω_0 and excited Ω_1 electronic states of molecule.

$$D^{(0)} \equiv \sum\nolimits'_{\vec{n},\alpha;\vec{m}\beta} D^{(0)}_{\vec{n}\alpha,\vec{m}\beta} \equiv \sum\nolimits'_{\vec{n},\alpha;\vec{m}\beta} \left\{ \int \left|\Psi^{(1)}_{\vec{n}\alpha}\right|^2 \cdot W_{\vec{n}\alpha,\vec{m}} \cdot \left|\Psi^{(0)}_{\vec{m}\beta}\right|^2 \cdot dr_{\vec{m}\beta} \cdot dr_{\vec{n}\alpha} \right.$$

is a change in the interaction energy of a molecule with the rest of the crystal during its transition to the excited electronic state.

$$M^{(1)}_{\vec{n}\alpha,\vec{m}\beta} \equiv \frac{1}{2} \cdot \left(\frac{\partial^2 M_{\vec{n}\alpha,\vec{m}\beta}}{\partial\xi_{0,\vec{m}\beta} \cdot \partial\xi_{0,\vec{n}\alpha}}\right)_{\xi_{0,\vec{m}\beta}=\xi_{0,\vec{n}\alpha}=0};$$

$$D^{(1)}_{\vec{n}\alpha,\vec{m}\beta} \equiv \frac{1}{2} \cdot \left(\frac{\partial^2 D_{\vec{n}\alpha,\vec{m}\beta}}{\partial\xi_{0,\vec{m}\beta} \cdot \partial\xi_{0,\vec{n}\alpha}}\right)_{\xi_{0,\vec{m}\beta}=\xi_{0,\vec{n}\alpha}=0}$$

are the matrix elements of molecules of the resonance interaction (the amendment to Condon approximation),

$$\xi_{0,\vec{n}\alpha} \equiv \sqrt{\frac{\Omega_0 \cdot M}{\hbar}} \cdot q_{\vec{n}\alpha}$$

is a dimensionless nuclear coordinate corresponding to the ground electronic term; M is the effective mass corresponding to the normal nuclear coordinate $q_{\vec{n}\alpha}$ (the minimal value of an electronic term is accepted as readout zero for $q_{\vec{n}\alpha}$); ħ is the Planck constant.

$W_{\vec{n}\alpha,\vec{m}\beta}$ is the molecules $\vec{n}\alpha$ and $\vec{m}\beta$ interaction operator; n and m are the numbers of cells, α and β - sorts of molecules.

$\Psi^{f_{\vec{n}\alpha}}_{\vec{n}\alpha} \equiv \Psi^{f_{\vec{n}\alpha}}_{\vec{n}\alpha}(r_{\vec{n}\alpha}, 0_{\vec{n}\alpha})$ is the electronic wave functions in adiabatic approximation and dependence neglecting from nuclei coordinates; $B_{\vec{n}\alpha}, b_{\vec{n}\alpha}$ are the annihilation operators of electronic excitation and intramolecular phonon; $B^+_{\vec{n}\alpha}, b^+_{\vec{n}\alpha}$ are the Hermitian conjunction operators for same excitations creation.

III. PROBLEM SOLUTION

True absorption curve can be calculated if the molecule's electric dipole moment average value is known. Let's calculate the average specific electric dipole moment $\langle P_{\vec{n}\alpha} \rangle$ arising in crystal under electromagnetic wave influence at the moment of time *t*.

a) The molecule's electric dipole moment

Let's define the molecule's electric dipole moment arising from the interaction with an electromagnetic wave

$$\vec{E} = 2 \cdot \vec{E}_0 \cdot \cos((\vec{Q} \cdot \vec{r}) - \omega t)), (\vec{E}_0 \cdot \vec{Q}) = 0 \ .$$

$\vec{Q}$ is the wave vector, $\vec{r}$ is the space coordinate, $\omega = 2\pi\nu$.

Operator of interaction with electromagnetic wave in coordinate representation looks like [9]:

$$H_{int} = U_\alpha e^{-i\omega t} + U^*_\alpha \cdot e^{i\omega t}, \quad (3)$$

where

$$U_\alpha = \frac{ie\vec{E}_0}{m\omega} \cdot e^{i\vec{Q}\cdot\vec{r}_\alpha} \cdot \vec{P}_\alpha,$$

$$\vec{r}_\alpha = \sum_i \vec{r}_{i\alpha}, \ \hat{\vec{P}}_\alpha = \sum_i \hat{\vec{P}}_{i\alpha} \quad .$$

In representation with basic functions

$$\Phi^{0\alpha}_{f\mu}(r, q) = \Psi^\alpha_f(r, q_\alpha) \cdot \varphi^\alpha_{f\mu}(q_\alpha); \quad f = 0,1$$

the energy operator for molecule's electronic-vibronic excitations has a diagonal form. Wave functions are normalized and form the full system:

$$\int \Phi^{0\alpha`*}_{f`\mu`}(r, q_\alpha) \int \Phi^{0\alpha}_{f`\mu}(r, q_\alpha) dr dq_\alpha = \delta_{ff`} \cdot \delta_{\mu\mu`} \cdot \delta_{\alpha\alpha`}$$

The matrix element of interaction on these functions looks like:

$$\langle \Phi^{0\alpha`*}_{f\mu`} | \hat{U}_\alpha | \Phi^{0\alpha}_{f\mu} \rangle = -\vec{E}_0 \cdot \langle \Phi^{0\alpha`*}_{f\mu`} \left| \hat{\vec{P}}_\alpha \right| \Phi^{0\alpha}_{f\mu} \rangle = -\vec{E}_0 \cdot \langle \varphi^{\alpha`*}_{f\mu`} \left| \hat{\vec{P}}_{ff`\alpha}(q_\alpha) \right| \varphi^\alpha_{f\mu} \rangle$$

where $\hat{\vec{P}}_\alpha = e \cdot \hat{\vec{r}}_\alpha$ is the electric dipole moment operator; e is the electrical charge of electron.

$\hat{\vec{P}}_{ff`\alpha}(q_\alpha) \equiv \int \Psi^\alpha_{f'}(r, q_\alpha) \cdot \hat{\vec{P}}_\alpha \cdot \Psi^\alpha_f(r, q_\alpha) \cdot dr$ is the dipole moment for quantum transition $f \to f'$.

The diagonal matrix elements are equal to zero under absence of stationary dipole moment of quantum states of a free molecule. Since $\vec{P}_{ff`\alpha}$ is a slowly changing function of q_α

$$\vec{P}_{ff`\alpha}(q) = [\vec{d}_{0\alpha} + \vec{d}_{1\alpha}(q_{0\alpha})] \cdot (1 - \delta_{ff`}),$$

where

$$\vec{d}_{0\alpha} \equiv \vec{P}_{0,1\alpha}(q^0_{0\alpha}) = P_{1,0\alpha}(q^0_{0\alpha}) \equiv \int \Psi^\alpha_1(r, q^0_{0\alpha}) \cdot \hat{\vec{P}}_\alpha * \Psi^\alpha_0(r, q^0_{0\alpha}) \cdot dr_\alpha$$

$$\vec{d}_{1\alpha}(q_{0\alpha}) = q_{0\alpha} \cdot \left(\frac{\partial}{\partial q_{0\alpha}} \vec{P}_{1,0\alpha}\right)_{q_{0\alpha}=0} \equiv q_{0\alpha} \cdot \vec{d}_\alpha,$$

and $\hat{q}_{0\alpha} = \left(\frac{\hat{b}^+_\alpha + \hat{b}_\alpha}{\sqrt{2}}\right)$ is a dimensionless nuclear coordinate, $q^0_{0\alpha}$ is coordinate of the equilibrium state of nucleus which is accepted as reference point.

Since $\vec{d}_{0\alpha}$ is responsible for the absorption in the purely electronic transition, and $\vec{d}_{\boldsymbol{\alpha}}$ - in the one-phonon

vibronic transitions which do not appear in the Condon approximation, and because we are interested in absorption at one-phonon electronic-vibronic transitions, let's consider only $\vec{d}_{\alpha}$.

In long-wave approximation relation $\langle\Psi_1|\vec{P}|\Psi_0\rangle = im\omega_{10}\cdot\langle\Psi_1|\vec{r}|\Psi_0\rangle;\ \hbar\omega_{10} = (E_1 - E_{0)}$ and for intramolecular transition frequencies ($\omega_\alpha \sim \omega_{10}^{\alpha}$) these matrix elements look like:

$$\langle\Phi_{1\mu}^{0\alpha}|\hat{U}_\alpha|\Phi_{0\gamma}^{0\alpha}\rangle = \frac{ie}{m\omega}\cdot\vec{E}_0\cdot e^{i\vec{Q}\vec{\rho}_\alpha^0}$$
$$\langle\Phi_{1\mu}^{0\alpha}\left|\hat{\vec{P}}_\alpha\right|\Phi_{0\gamma}^{0\alpha}\rangle = \frac{ie}{m\omega}\cdot\vec{E}_0\cdot e^{i\vec{Q}\vec{\rho}_\alpha^0}\cdot im\omega_\alpha \cdot\langle\Psi_1^{0\alpha}|\hat{\vec{x}}_\alpha|\Psi_0^{0\alpha}\rangle\cdot$$
$$\cdot\langle\varphi_{1\mu}^{0\alpha}|\hat{q}_{0\alpha}|\varphi_{0\gamma}^{0\alpha}\rangle = -(\vec{E}_0\cdot\vec{d}_\alpha)\cdot e^{i\vec{Q}\vec{\rho}_\alpha^0}\cdot\langle\Phi_{0\mu}^{0\alpha}|D_{0\alpha}^{+}\hat{q}_{0\alpha}|\Phi_{0\nu}^{0\alpha}\rangle \quad (4)$$

and

$$\langle\Phi_{0\mu}^{0\alpha}|D_{0\alpha}^{+}\hat{q}_{0\alpha}|\Phi_{0\nu}^{0\alpha}\rangle = \langle\Phi_{1\mu}^{0\alpha}|B_\alpha^{+}D_{0\alpha}^{+}\hat{q}_{0\alpha}|\Phi_{0\nu}^{0\alpha}\rangle = $$
$$= \langle\Phi_{1\mu}^{0\alpha}|V_\alpha^{+}\hat{q}_{0\alpha}|\Phi_{0\nu}^{0\alpha}\rangle = \langle\Phi_{1\mu}^{0\alpha}\left|V_\alpha^{+}\cdot\frac{(\hat{b}_\alpha+\hat{b}_\alpha^{+})}{\sqrt{2}}\right|\Phi_{0\nu}^{0\alpha}\rangle \quad (5)$$

From this expression follows:

$$\langle\varphi_{1\mu}|\varphi_{0\nu}\rangle = \langle\varphi_{0\mu}|\hat{D}_0^{+}|\varphi_{0\nu}\rangle = \langle\Phi_{0\mu}^{0}|\hat{D}_0^{+}|\Phi_{0\nu}^{0}\rangle,$$
$$\langle\varphi_{0\nu}|\varphi_{1\mu}\rangle = \langle\varphi_{0\nu}|\hat{D}_0|\varphi_{0\mu}\rangle = \langle\Phi_{0\nu}^{0}|\hat{D}_0|\Phi_{0\mu}^{0}\rangle, \quad (6)$$

where

$$\hat{D}_0 = exp\{\xi(\hat{b}_0^{+}-\hat{b}_0\}\cdot exp\{\beta\cdot(\hat{b}_0^{+}\hat{b}_0^{+}-\hat{b}_0\hat{b}_0)\}$$
$$\hat{D}_0^{+} = exp\{-\xi(\hat{b}_0^{+}-\hat{b}_0)\}\cdot exp\{\beta\cdot(\hat{b}_0^{+}\hat{b}_0^{+}-\hat{b}_0\hat{b}_0)\}, \quad (7)$$

value $\beta \equiv \frac{1}{4}\ln\left(\frac{\Omega_0}{\Omega_1}\right)$ characterizes the frequency of intramolecular oscillations change at electronic excitation; $\xi \equiv q_0$ characterizes the displacement of equilibrium position of an oscillation configuration; $\hat{V}_\alpha^{+} = \hat{B}_\alpha^{+}\cdot\hat{D}_{0\alpha}$; $\hat{V}_\alpha = \hat{B}_\alpha\cdot\hat{D}_{0\alpha}$ are vibronic operators.

Correlation between electronic and phonon excitations occurs at the oscillation frequency change only, and displacement of equilibrium position of an oscillation configuration is absent at electronic excitation (ξ =0). Since the molecule with non-degenerate states has independent normal oscillation coordinates, we will consider only one normal oscillation coordinate.

By using (4) – (7) we obtain the next equation:

$$\langle\Phi_{1\mu}^{0\alpha}|\hat{U}_\alpha|\Phi_{0\nu}^{0\alpha}\rangle = -(\vec{E}_0\cdot\vec{d}_\alpha)\cdot e^{i\vec{Q}\cdot\vec{\rho}_\alpha^0}\langle\Phi_{1\mu}^{0\alpha}\left|V_\alpha^{+}\cdot\frac{(\hat{b}_\alpha+\hat{b}_\alpha^{+})}{\sqrt{2}}\right|\Phi_{0\nu}^{0\alpha}\rangle; \quad (8)$$
$$\langle\Phi_{0\nu}^{0\alpha}|\hat{U}_\alpha|\Phi_{1\mu}^{0\alpha}\rangle = -(\vec{E}_0\cdot\vec{d}_\alpha^{*})\cdot e^{i\vec{Q}\cdot\vec{\rho}_\alpha^0}\langle\Phi_{0\nu}^{0\alpha}\left|\frac{(\hat{b}_\alpha+\hat{b}_\alpha^{+})}{\sqrt{2}}\cdot\hat{V}_\alpha\right|\Phi_{1\mu}^{0\alpha}\rangle$$

Hence

$$\hat{U}_\alpha = -\left\{\frac{\vec{d}_\alpha^{*}}{\sqrt{2}}(\hat{b}_\alpha+\hat{b}_\alpha^{+})\cdot\hat{V}_\alpha + \frac{\vec{d}_\alpha}{\sqrt{2}}\cdot\hat{V}_\alpha^{+}\cdot(\hat{b}_\alpha+\hat{b}_\alpha^{+})\right\}\cdot\vec{E}_0\cdot e^{i\vec{Q}\cdot\vec{\rho}_\alpha^0}; \quad (9)$$

$$\hat{\vec{P}}_\alpha = \frac{1}{v}\cdot e\hat{\vec{x}}_\alpha = \frac{1}{v}\left\{\frac{\vec{d}_\alpha^{*}}{\sqrt{2}}(\hat{b}_\alpha+\hat{b}_\alpha^{+})\cdot\hat{V}_\alpha + \frac{\vec{d}_\alpha}{\sqrt{2}}\cdot\hat{V}_\alpha^{+}(\hat{b}_\alpha+\hat{b}_\alpha^{+})\right\} \quad (10)$$

is an operator of electric dipole moment in a new basis at the non-Condon approximation.

b) Dielectric permeability of crystal

Expression for operator of an electromagnetic wave

$$\vec{E} = 2\cdot\overrightarrow{E_0}\cdot e^{-\eta t}\cdot\cos(\vec{Q}\vec{r}-\omega t),$$
$$(\vec{Q}\cdot\vec{E}) = 0$$

interaction H_{int} with crystal at the frequencies near to frequency of electronic transition ($\omega - \omega_{o1}^{\alpha}$) according to (3) and (9) it is possible to write down as

$$H_{\text{int}} = -E_0\cdot\sum_{m,\beta}\left\{U_{m\beta}(t)\cdot e^{-i\omega t+\eta t} + \text{herm.conj.}\right\},$$
$$V_{m\beta}(t) = \exp\left\{\frac{i}{\hbar}H_0 t\right\}\cdot V_{m\beta}\cdot\exp\left\{-\frac{i}{\hbar}H_0 t\right\}.$$

H_0 is an operator for crystal without interaction with electromagnetic wave (2). The small positive parameter η specifies that interaction takes on in the infinite past time. The parameter η describes the attenuation which always is available in any real system.

Let's calculate the average specific electric dipole moment $\left\langle\overrightarrow{P_n}\right\rangle$ arising at the moment of time *t* in a crystal under influence of H_{int}. This operator is defined in coordinate representation as $\vec{P}_{\vec{n}\alpha} = e\cdot\vec{x}_{n\alpha}$. Under (10) this operator can be written in a new basis as

$$\vec{P}_{\vec{n}\alpha} = \frac{1}{v}\left\{\frac{\overrightarrow{d_{\vec{n}\alpha}^{*}}}{\sqrt{2}}(\hat{b}_{\vec{n}\alpha}+\hat{b}_{\vec{n}\alpha}^{+})\cdot\hat{V}_{\vec{n}\alpha} + \frac{\overrightarrow{d_{\vec{n}\alpha}^{*}}}{\sqrt{2}}\cdot\cdot\hat{V}_{\vec{n}\alpha}^{+}(\hat{b}_{\vec{n}\alpha}+\hat{b}_{\vec{n}\alpha}^{+})\right\}$$

where *v* is a volume of a crystal's unit cell.

The average values of a physical values *L* are defined by a density matrix ρ by the expression $\left\langle\hat{L}\right\rangle = Sp(\rho\cdot\hat{L})$. If the density matrix equals to the equilibrium density matrix ρ_0 before inclusion of interaction, then at the interaction presence at the moment of time *t* it will be defined in linear approximation on the interaction operator by equation

$$\rho(t) = \rho_0 - i\int_{-\infty}^{t}\left[\tilde{H}_{\text{int.}}(\tau),\rho_0\right]\cdot d\tau \quad ,$$

where

$$H_{\text{int.}}(t) = e^{iH_0 t}\cdot H_{\text{int.}}\cdot e^{-iH_0 t} \quad .$$

So,

$$\langle\vec{P}_n(t)\rangle = \sum_\alpha Sp(\rho(t)\vec{P}_{\vec{n}\alpha}(t)) =$$
$$\sum_\alpha Sp(-\frac{i}{\hbar}\int_{-\infty}^{t}\left[\tilde{H}_{int}(\tau),\rho_0\right]d\tau\cdot\vec{P}_{\vec{n}\alpha}(t)) =$$
$$\sum_\alpha Sp\left\{-\frac{i}{\hbar}\int_{-\infty}^{t}\left[\sum_{\vec{m},\beta}(U_{\vec{m}\beta}(\tau)e^{-i\omega\tau+\eta\tau} + herm.conj.),\rho_0\right]d\tau\left\{\frac{\overrightarrow{d_{\vec{n}\alpha}^{*}}}{\sqrt{2}}(\hat{b}_{\vec{n}\alpha}+\hat{b}_{\vec{n}\alpha})\cdot\hat{V}_{\vec{n}\alpha} + \frac{\overrightarrow{d_{\vec{n}\alpha}}}{\sqrt{2}}\cdot\right.\right.$$
$$\left.\left.\hat{V}_{\vec{n}\alpha}^{+}\ (\hat{b}_{\vec{n}\alpha}+\hat{b}_{\vec{n}\alpha})\right\}\right.$$

We assume that the specific electric moment of a crystal without external influence is equal to zero $\left(Sp\left(\rho_0\cdot\vec{P}_{\vec{n}}(t)\right)=0\right)$. The decomposition terms of electronic and one-phonon Green's electronic-vibronic functions following to unit in expressions for vibronic operators (8) give zero contributions. So further , we can write the relation

$$\hat{V}_{\vec{n}\alpha} = (1-X^2)^{1/4}\ \hat{B}_{\vec{n}\alpha}$$

where $X = \frac{\Omega_0-\Omega_1}{\Omega_0+\Omega_1}$.

Going over to the Fourier component of the retarted Green's function:

$$G_r^{\alpha\beta}(\vec{Q},\omega) = \sum_{\vec{n},\alpha} e^{-i\vec{Q}\cdot\overrightarrow{\rho_{\vec{n}\alpha}^0}} \cdot G_r^{\alpha\beta}(\vec{n},\omega);$$

$$G_r^{\alpha\beta}(\vec{n},\omega) = \int_{-\infty}^{\infty} e^{-i\omega t-\eta t} \cdot G_r^{\alpha\beta}(\vec{n},t)dt,$$

where

$$G_r^{\alpha\beta}(\vec{n},t) = \\ = -i\theta(t-\tau)Sp\{\rho_0[\hat{B}_{\vec{n}\alpha}(t)\hat{b}_{\vec{n}\alpha}(t), \hat{B}^+_{\vec{o}\beta}(0)\hat{b}^+_{\vec{o}\beta}(0)]\} \quad (11)$$

So, the expression for the average specific electric dipole moment is

$$\left\langle \vec{P}_{\vec{n}}(t)\right\rangle = -\frac{\sqrt{1-X^2}}{2v}\cdot\sum_{\alpha,\beta}\left\{\overrightarrow{d_\alpha'}\left(\overrightarrow{d_\beta'^*}\cdot\overrightarrow{E_0}\right)\cdot\right. \\ \cdot\left(G_r^{\alpha\beta}(\vec{Q},\omega) - G_r^{+}(-\vec{Q},-\omega)\right)\cdot e^{i\left(\vec{Q}\overrightarrow{\rho_{\vec{n}\alpha}^o}-\omega t\right)+\eta t} + \\ \left. + \text{herm.conj.}\right\} \quad (12)$$

The average specific electric dipole moment is expressed through the cross-section dielectric permeability tensor characterizing the response of a crystal on the external perturbation [1]:

$$\left\langle \vec{P}_{\vec{n}}(t)\right\rangle = \left\{\frac{\varepsilon^{\perp}(\vec{Q},\omega)-1}{8\pi}\cdot\vec{\mathrm{E}}_o\cdot e^{i\left(\vec{Q}\cdot\overrightarrow{\rho_{\vec{n}}^o}-\omega t\right)+\eta t} + \text{herm.conj.}\right\} \quad (13)$$

Comparing formulas (12) and (13) we can write the expression for components of this tensor:

$$\varepsilon_{xy}^{\perp} = \delta_{xy} + \frac{4\pi}{v}\cdot\sqrt{1-X^2}\sum_{\alpha,\beta}\left\{G_r^{\alpha\beta+}(-Q,-\omega) - \right. \\ \left. - G_r^{\alpha\beta}(\vec{Q},\omega)\right\}\cdot\left|\overrightarrow{d_\alpha'}\right|^2$$

because $(\vec{Q}\cdot\vec{\rho}_\beta)\to 0$, ($\vec{\rho}_\beta$ is a radius-vector of molecule in the unit cell).

Since the energy of intramolecular excitations are well above the thermal energy even at room temperature ($\Delta E, \Omega_o >> kT$), it is possible to replace the averaging with the help of density matrix on averaging by the vacuum state which is characterized by function $\left|\{0,0\}\right\rangle$. k is the Boltzmann constant, T is the temperature.

In this case, expressions for Green's functions are simplified. For example, the function (11) is presented as

$$G_r^{\alpha\beta}(\vec{n},t) = \left\langle\left\langle B_{\vec{n}\alpha}\cdot b_{\vec{n}\alpha}; B^+_{o\beta}\cdot b^+_{o\beta}\right\rangle\right\rangle_t \equiv \\ \equiv i\theta(t)\cdot\left\langle\{0,0\}\left|B_{\vec{n}\alpha}(t)\cdot b_{\vec{n}\alpha}(t), B^+_{o\beta}(0)\cdot b^+_{o\beta}(0)\right|\{0,0\}\right\rangle.$$

In this case the main contribution to the interaction of the crystal with an electromagnetic wave is provided by processes of the photons absorption with vibronic excitation creation, which is described by function $G_r^{\alpha\beta}(\vec{Q},\omega)$. Absorption of light by a crystal for these frequencies is characterized by an imaginary part of dielectric permeability.

Let's find a vibronic absorption spectrum of crystal. For this purpose we calculate the Fourier transformation by time of retarded Green's functions. Such calculation for any operators A and B can be made by means of the equation

$$\omega\cdot\left\langle\left\langle A;B\right\rangle\right\rangle_\omega = \left\langle[A,B]\right\rangle + \left\langle\left\langle[A,H_o];B\right\rangle\right\rangle_\omega.$$

Here

$$\left\langle\left\langle A;B\right\rangle\right\rangle_\omega = \int_{-\infty}^{\infty} e^{i\omega t-\eta t}\cdot\left\langle\left\langle A;B\right\rangle\right\rangle dt,$$

where averaging of commutator $[A,B]$ also is provided under vacuum state $\left|\{0,0\}\right\rangle$.

Let's find $\left\langle\left\langle A;B\right\rangle\right\rangle_\omega$. That is the same value $\left\langle\left\langle B_{\vec{n}\alpha}b_{\vec{n}\alpha}; B^+_{o\beta}b^+_{o\beta}\right\rangle\right\rangle_\omega$.

Let's consider the crystal's states with one-electronic excitation only $\left(\sum_{\vec{n},\alpha} B^+_{n\alpha}B_{n\alpha} = 1;\ \ f = 0,1\right)$ and members keeping a full number of intramolecular oscillations.

For two identical molecules in an elementary cell (two identical sub-lattices in crystal) we have relations:

$$G_r^{11} = g_{0(\omega)}^{1111} = g_{0(\omega)}^{2222} = G_r^{22}$$

$$G_r^{12} = g_{0(\omega)}^{1122} = g_{0(\omega)}^{2211} = G_r^{21}.$$

Let's decompose these functions on poles. For example, for next functions $g_{0(\omega)}^{1111} = g_{0(\omega)}^{2222}$:

$$\frac{(\omega-\omega_0+L)\cdot(2(\omega-\omega_0)+L)}{2(\omega-\omega_1)\cdot(\omega-\omega_2)\cdot(\omega-\omega_3)} = \frac{A}{(\omega-\omega_1)} + \frac{B}{(\omega-\omega_2)} + \frac{C}{(\omega-\omega_3)},$$

where ω_1, ω_2, ω_3 are the Green's function poles for $g_{0(\omega)}^{1111} = g_{0(\omega)}^{2222}$ minus $\omega_0 = 2\pi\nu_0$, L is the resonant interaction between molecules of one sub-lattice; and L_3 is the resonant interaction between molecules of different sub-lattices within one unit cell. Similarly h is a resonant dipole interaction between molecules of one sub-lattice for one-phonon process at electron-oscillation transition (which don't appear in Condon approximation).

For A, B and C values we find that:

$$A = \frac{(3ly_1 + 2l^2 + y_1^2)}{(y_1-y_3)\cdot(y_1-y_2)},$$

$$B = \frac{(3ly_2 + 2l^2 + y_2^2)}{(y_2-y_3)\cdot(y_2-y_1)},$$

$$C = \frac{(3ly_3 + 2l^2 + y_3^2)}{(y_3-y_1)\cdot(y_3-y_2)}, \qquad y_i = \frac{\omega_i}{|h|}.$$

Similar expressions may be written for the functions $g_{0(\omega)}^{1122} = g_{0(\omega)}^{2211}$:

$$g_0^{1122} = g_0^{2211} = \frac{K}{(\omega-\omega_1)} + \frac{D}{(\omega-\omega_2)} + \frac{F}{(\omega-\omega_3)},$$

$$K = \frac{l_3(l+y_1)}{(y_1-y_3)\cdot(y_1-y_2)}, \qquad D = \frac{l_3(l+y_2)}{(y_2-y_3)\cdot(y_2-y_1)},$$
$$F = \frac{l_3(l+y_3)}{(y_3-y_1)\cdot(y_3-y_2)}.$$

The equation for finding Green's function poles is given by:

$$y^3 + 5ly^2 + y(4l^2 - l_3^2 - 0{,}5) + l(4l^2 - l_3^2 - 1) = 0,$$

$$y = \omega_{(i)}(|h|)^{-1} = \frac{\omega - \omega_0}{|h|}; \quad l = \frac{L}{|h|}; \quad l_3 = \frac{L_3}{2|h|}.$$

Hence we define the polarization of strip with frequency $\omega = \omega_0 + \omega_i$:

$$\varepsilon_{II}^{\perp} = 1 + \frac{a(1+\varphi)}{(y-y_i)}\left\{\frac{3ly_i + 2l^2 + y_i^2 + l_3(l+y_i)}{(y_i-y_j)\cdot(y_i-y_k)}\right\};$$

$$\varepsilon_{II}{}^{\perp}{}_{II} = 1 + \frac{a(1-\varphi)}{(y-y_i)}\left\{\frac{3ly_i + 2l^2 + y_i^2 - l_3(l+y_i)}{(y_i-y_j)\cdot(y_i-y_k)}\right\},$$

where *i, j, k* are indexes equal to 1, 2, 3.

Results for the different interaction parameters are shown in Tab. I.

TABLE I

l	l_3	y	I_{II}	I_I
0.1	0.1	0.58	0	0.49
		-0.16	0	0.01
		-0.92	0.5	0
	0.3	0.61	0	0.498
		-0.16	0	0.002
		-0.96	0.5	0
	0.5	0.72	0	0.499
		-0.13	0	0.001
		-1.08	0.5	0
0.3	0.1	0.36	0	0.462
		-0.33	0	0.038
		-1.53	0.5	0
	0.3	0.41	0	0.4994
		-0.32	0	0.006
		-1.57	0.5	0
	0.5	0.47	0	0.4953
		-0.31	0	0.0047
		-1.65	0.5	0
0.5	0.1	0.025	0	0.444
		-0.025	0	0.056
		-2.30	0.5	0
	0.3	-0.85	0	0.46
		-0.265	0	0.04
		-2.33	0.5	0
	0.5	0.175	0	0.47
		-0.31	0	0.03
		-2.375	0.5	0

The results of this paper generalize the results of paper [8]. Only electronic excitation of a crystal is realized at h=0. Expression for poles in this case looks like

$$\omega_{1,2} = \mathcal{E} - 2L \pm L_3 .$$

These poles correspond to the different polarizations. One of them is directed along the sum, and another is directed along a difference of dipole moments of transitions in molecules. Intensities of these peaks are identical.

The result for variant h=0 and L_3=0 corresponding to only electronic excitation of crystal with one molecule in the unit cell follows from the previous formula. Absorption line is not polarized in this variant, $\omega = \mathcal{E} - 2L$ .For variant with L_3=0 (there is no interaction between sub-lattices corresponding to one molecule in the unit cell) we have received two non-polarized poles $\omega_{1,2} = \omega_0 - \frac{3L}{2} \pm \frac{\sqrt{L^2+2h^2}}{2}$.

Including the resonant interaction characterized by parameter h removes degeneration and splits a spectral line. So, for variant with L_3=0 and $h \gg L$, the poles (non-polarized) are as follows:

$$\omega_{1,2} = \omega_0 - \frac{3L}{2} \pm \frac{h}{\sqrt{2}}.$$

For variant with L=0 there are the two poles with frequencies: $\omega_{1,2} = \omega_0 \pm \sqrt{L_3^2 + \frac{h^2}{2}}$. If $h \gg L_3$, then $\omega_{1,2} = \omega_0 \pm \frac{h}{\sqrt{2}}$ and we have two poles in both polarizations with intensities I_I and I_{II}. If $L_3 \gg h$, then $\omega_{1,2} = \omega_0 \pm L_3$ and we have two poles in different polarizations.

CONCLUSIONS

In this paper the absorption vibronic spectra have been investigated for molecular crystals with two molecules in the unit cell that corresponds to two-layer graphene. The non-Condon amendments have been take into account for absorption vibronic spectra defining. The absorption vibronic spectra, their intensities, frequencies and polarizations are calculated for molecular crystals containing two molecules in the unit cell or two-layer graphene. Numerical calculations show that, for certain values of the parameters of the intermolecular interactions either continuous or line absorption spectra can be observed in a sample.

ACKNOWLEDGMENTS

E. Aznakayev is grateful to Dr. Alexander Serikov for previous cooperation, attraction attention to this problem.

REFERENCES

[1] K.S. Novoselov, A.K. Geim, "Two-Dimensional Gas of Massless Dirac Fermions in Graphene", *Nature*, vol. 438**,** pp**.** 197-200, 2005.

[2] H.Raza. *Graphene Nanoelectronics*. Berlin-Heidelberg. Springer-Verlag Publ. 2012.

[3] P. Avouris, "Graphene: Electronic and Photonic Properties and Devices," *Nano Letters*, vol. 10, pp. 4285-4294, 2010.

[4] P. Sutter, "Epitaxial graphene: How silicon leaves the scene", *Nature Materials*, vol. 8, pp.,171–172, 2009.

[5] G. Gu, S. Nie, R. M. Feenstra, R. P. Devaty, W. J. Choyke,W. K. Chan, M. G. Kane, "Field effect in epitaxial graphene on a silicon carbide substrate", *Applied Physics Letters*, vol. 90, p.253507, 2007.

[6] G. Fiori, G. Iannaccone, "Simulation of Graphene Nanoribbon Field-Effect Transistor", *IEEE Electron Devices Letters*, vol. 28, No. 8, 2007.

[7] Kim, K. L. Shepard, J. Hone, "Boron Nitride Substrates for High-Quality Graphene Electronics," *Nature Nanotechnology*, vol. 5, p. 722, 2010.

[8] A.A. Serikov, "Vibronic Spectra of Molecular Crystals with Participation of Non-Totally Symmetrical Vibrations", *Phys. Stat. Sol.*, vol. 44, pp. 733-746, 1971.

[9] A.S. Davydov. *Quantum Mechanics*. Moscow: Nauka Publ.,1973.

2013 IEEE XXXIII International Scientific Conference Electronics and Nanotechnology (ELNANO)

Performance Analysis of Multiwalled Carbon Nanotube Bundles

Manodipan Sahoo
School of VLSI Technology
Bengal Engineering & Science University
Shibpur, Howrah-711103, India
Email: manodipansahoo@gmail.com

Hafizur Rahaman
School of VLSI Technology
Bengal Engineering & Science University
Shibpur, Howrah-711103, India
Email: rahaman_h@it.becs.ac.in

***Abstract*—In this work, we have presented a comprehensive analysis of the performances of Copper wire and MWCNT (Multi-walled Carbon Nano Tube) bundles across deep submicron technology nodes like 45nm, 32nm, 22nm and 16nm. The analysis has been done for local, intermediate and global level interconnects. The analytical closed form delay expressions for both the Copper wire and MWCNT bundles have been found out. It is observed that MWCNT bundles show better performance over Copper wire at all level of interconnects and technology nodes. In 16 nm technology node, the performance advantage numbers of MWCNT bundle over Copper wire are 50%, 90% and 85% for 20-μm long local interconnect, 200-μm long intermediate interconnect and 10000-μm long global interconnect respectively. It is also observed that performance numbers improve with scaling for all level of interconnects and the ratio of delay of MWCNT bundles and Copper wire for various level of interconnects agree well with the existing work.**

***Keywords*—MWCNT; delay; performance; scaling.**

I. Introduction

Carbon Nano Tubes(CNTs) have immense potential in dictating the future VLSI interconnect technology[1]. Because of their great electrical[2] and thermal properties[3], high mechanical stability[4], current carrying capacity[5], they are the most promising one to replace the existing Copper interconnect technology. As CNTs have long mean free paths(MFP) of the order of several micrometers (compared to Copper having mean free path of 40 nm at room temperature), it can provide ballistic transport at short interconnect lengths[6]. CNTs have excellent reliability compared to Copper due to later's electromigration concerns at high temperature[7]. CNTs are classified as SWCNT (Singlewalled Carbon Nano Tube), DWCNT (Doublewalled Carbon Nano Tube) and MWCNT (Multiwalled Carbon Nano Tube). SWCNT consists of single graphene sheet rolled up, it can have diameters varying from 0.4nm to 4nm. DWCNT and MWCNTs are two and multiple concentric shells of graphene sheets. Though SWCNT can be modeled easily due to it's simple structure [8-11], DWCNT and MWCNT's are far easier to fabricate[12]. In [13], performance comparison of MWCNT bundle and Copper wire is reported, but the analysis has been carried out based on the SPICE simulations of the complex MWCNT model. In [14], closed form expressions for 50% time delay of MWCNT bundles have been reported, but no comparison has been shown with the SPICE simulated data in the existing literatures. In this work a compact model of MWCNT bundle is considered which can significantly improve the simulation time without any loss of accuracy[15]. Contribution of this work, Firstly, Elmore delay estimation methodology is used to find out the analytical closed form 50% time delay expressions for the MWCNT bundles[16]. Secondly, the performance of MWCNT bundles is analysed at local, intermediate and global level interconnects across technology nodes like 45nm, 32nm, 22nm and 16nm. Thirdly, the ratio of delay of MWCNT bundles and Copper wire for various level of interconnects is compared with the existing work and shown that they agree well. The rest of the work is organised as follows. Section II introduces the equivalent RLC model of MWCNT bundles. Section III discusses the simulation methodology for Copper wire and MWCNT bundles. In Section IV, we analyze the performances of Copper wire and the MWCNT bundles. Finally, we conclude the discussions in Section V.

This work is partially supported by the DIT, Government of West Bengal, India under VLSI Design Project.

II. Equivalent RLC Parameters of MWCNT bundle

A MWCNT consists of multiple concentric shells. Shell diameters vary from D_{min} to D_{max} as shown in fig.1. The two adjacent shells are separated by a minimum distance δ called Van Der Waal's gap which is equal to 0.34 nm. The number of conducting channels for a particular shell can be approximated as [17],

$$N_{shell}(D) \approx (aD + b), \quad D > 3nm$$

where, D is the diameter of the shell, a is 0.0612 nm^{-1} and b is 0.425. According to [17], we have considered the ratio D_{min}/D_{max} to be equal to 0.5. Number of shells of MWCNT is denoted as p and it is given by,

$$p = 1 + \lfloor \frac{(D_{max} - D_{min})}{2\delta} \rfloor$$

So, the diameter of the i^{th} shell is given by,

$$D_i = D_{max} - 2\delta(i-1), \quad 1 \leq i \leq p$$

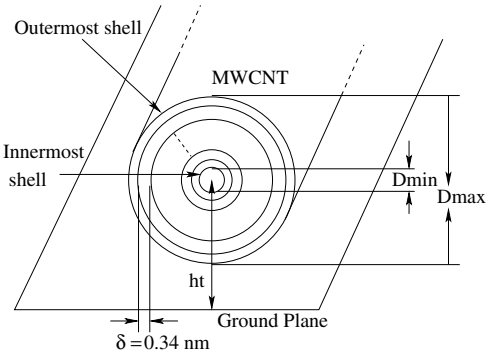


Fig. 1. MWCNT above a ground plane.

The intrinsic resistance R_{CNT} is determined by,

$$R_{CNT} = R_Q + R_O l_{CNT} = (\frac{h}{2q^2N})(1 + \frac{l_{CNT}}{\lambda})$$

R_Q and R_O are the quantum and ohmic resistances[8]. Here, l_{CNT}, λ and N are the length, MFP and number of conducting channels of MWCNT respectively. The metal-nanotube contact resistance(R_{mc}) can vary from zero to hundreds of Kilo Ohms. But it is mentioned in [18] that R_{mc} can be very small of the order of few Kilo Ohms. The MFP also can be approximated as $\lambda \approx$ 1000D [17]. The inductance of MWCNT has two components: a)Magnetic inductance and b)Kinetic inductance [13]. The per unit length magnetic and kinetic inductance per shell is given in (1),

$$L_M(p.u.l) = \left(\frac{\mu}{2\pi}\right)\ln\left(\frac{ht}{d}\right) \tag{1a}$$

$$L_{K/channel}(p.u.l) = \frac{h}{2q^2\upsilon_F} \times \frac{1}{2} \approx 8nH/\mu m \tag{1b}$$

$$L_{K/shell}(p.u.l) = \frac{L_{K/channel}(p.u.l)}{N_{shell}(D)} \tag{1c}$$

Where, ht is the height of the MWCNT from the ground plane, d is the diameter of the MWCNT, υ_F is the Fermi velocity($\sim 8\times10^5$m/s for CNTs). The capacitance of MWCNT also has two components: a)Quantum capacitance(C_Q) and b)Electrostatic capacitance(C_E)[13]. They are expressed as shown in (2),

$$C_{Q/channel} = 2 \times \frac{2q^2}{h\upsilon_F} \tag{2a}$$

$$C_{Q/shell} = C_{Q/channel} N_{shell}(D) \tag{2b}$$

The per unit length shell to shell capacitance is given in (3),

$$C_S(p.u.l) = \frac{2\pi\varepsilon}{\ln(D_{max}/D_{min})} = \frac{2\pi\varepsilon}{\ln\left[D_{max}/\left(D_{max}-2\delta\right)\right])} \tag{3}$$

The per unit length mutual inductance between shells for a MWCNT of length l_{CNT} is given by,

$$M_{shell}(p.u.l) = \frac{\mu}{2\pi}\left[\ln\left(\frac{4l_{CNT}}{D_{max}}\right) - 1 + \frac{\left(D_{max}+D_{min}\right)}{\pi l_{CNT}}\right]$$

This can be estimated to be $\sim$ 2pH/μm, which is far smaller than other components of inductance.

The per unit length electrostatic coupling capacitance between parallel CNT's is given in (4)[19].

$$C_c = \frac{\pi\varepsilon}{\ln\left(s/d + \sqrt{\left(s/d\right)^2+1}\right)} = \frac{\pi\varepsilon}{\ln\left(2s/d\right)}\ if\ diameter\ d \ll s\ (spacing) \tag{4}$$

The schematic of MWCNT bundle is shown in fig.2. The number of MWCNTs along x and z axes are denoted as N_w and N_h and they are shown in (5)[13],

$$N_w = \frac{(w - D_{max})}{\left(D_{max}+\delta\right)} \tag{5a}$$

$$N_h = \frac{2(t - D_{max})}{\sqrt{3}\left(D_{max}+\delta\right)} + 1 \tag{5b}$$

where, D_{max} is the outer shell diameter of an isolated MWCNT and (D_{max}+δ) is the inter-MWCNT distance. All the MWCNTs are metallic in nature due to their high conductance. So all the MWCNTs will take part in signal transmission. Total number of MWCNTs in a bundle is shown in (6)[13],

$$N_{MWCNT} = N_w N_h - \frac{N_h}{2}, \quad if\ N_h\ is\ even \tag{6a}$$

$$N_{MWCNT} = N_w N_h - \frac{(N_h - 1)}{2}, \quad if\ N_h\ is\ odd \tag{6b}$$

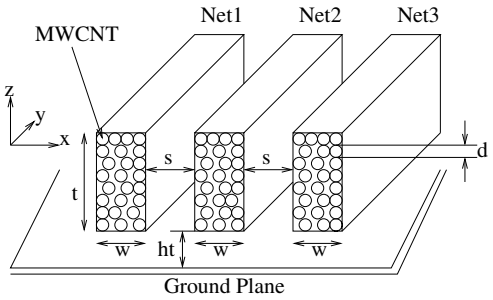


Fig. 2. MWCNT Bundle structure.

III. Simulation Methodology

Typical driver-interconnect-load structure is shown in fig.3. The inverting buffers are designed for driver and load as shown in the figure. The inverters can be modeled using simple R-C circuits with a high degree of accuracy[13]. In the figure, R_{d0} is the unit sized equivalent switching resistance per unit width, C_{d0} is per unit width equivalent diffusion capacitance of gate and C_{g0} is per unit width equivalent gate capacitance of the inverting buffer respectively. The estimated values of R_{d0}, C_{d0} and C_{g0} are given in Table(I). The simulations are performed at 45, 32, 22 and 16 nm technology nodes using Spectre simulator of Cadence. The technology nodes and parameters are derived from the ITRS 2006 roadmap as illustrated in Table(II)[20]. The SPICE models are used from Predictive Technology Model(PTM) source[21]. Most of the simulations are performed in MATLAB, as the CNT bundle delay can be analytically estimated. The delay estimation is done for Copper(Cu) and MWCNT bundles for local, intermediate and global level interconnects.

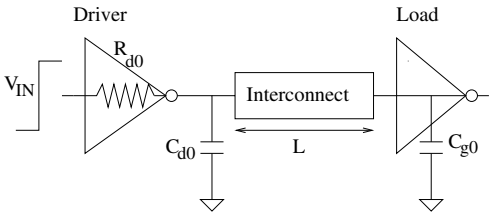


Fig. 3. Schematic of a typical interconnect structure.

TABLE I: Equivalent Minimum Sized Gate Parameters

Technology Node	**45 nm**	**32 nm**	**22 nm**	**16 nm**
R_{d0} (in KΩ)	17.22	23.68	34.45	47.37
C_{d0} (in fF)	0.105	0.073	0.05	0.036
C_{g0} (in fF)	0.215	0.156	0.137	0.07

TABLE II: Interconnect Parameters(ITRS 2006)

Year	2010	2013	2016	2019
DRAM1/2 pitch(nm)	**45**	**32**	**22**	**16**
MPU M1 1/2 pitch(nm)	45	32	22	16
MPU gate length(nm)	18	13	9	6
No. of Metal level(nm)	12	13	13	14
M1 wiring pitch	90 90 135	64 64 96	44 44 66	32 32 48
M1 A/R	1.8 1.8/1.6 2.4/2.2	1.9 1.9/1.7 2.5/2.3	2 2/1.8 2.6/2.4	2 2/1.8 2.8/2.5
Effective Resistivity (μ Ω-cm)	4.08 4.08 3.1	4.83 4.83 3.52	6.01 6.01 4.2	7.34 7.34 4.93
VDD(Volt)	1	0.9	0.8	0.7
Tox(nm)	1	0.9	0.8	0.7
Dielectric constant	2.5-2.8	2.1-2.4	1.9-2.2	1.6-1.9

A. Simulation Model for Copper Wire

Electrical equivalent model of a typical Copper interconnect structure is shown in fig.4. S is the actual gate size in μm. The interconnect parameters are denoted by R_{eqpul}, C_{eqpul} and l_{Cu} where, l_{Cu} is the length of the Copper interconnect, R_{eqpul} is the per unit length Copper wire resistance and C_{eqpul} is the per unit length Copper wire capacitance. We have considered R-C based models instead of R-L-C based models, because for different interconnect lengths, dimensions and driver sizes conidered here, R-C circuits will be sufficient to model MWCNT bundle based interconnect delay[9]. The analytical equations used to calculate RLC values of Cu wire are summarized below [21],

$$R = \frac{\rho}{\omega.t}$$

$$C_g = \varepsilon\left[\frac{\omega}{h} + 2.22\left(\frac{s}{s+0.70h}\right)^{3.19} + 1.17\left(\frac{s}{s+1.51h}\right)^{0.76}\left(\frac{t}{t+4.53h}\right)^{0.12}\right]$$

$$C_c = \varepsilon\left[1.14\frac{t}{s}\left(\frac{h}{h+2.06s}\right)^{0.09} + 0.74\left(\frac{\omega}{\omega+1.59s}\right)^{1.14} + 1.16\left(\frac{\omega}{\omega+1.87s}\right)^{0.16}\left(\frac{h}{h+0.98s}\right)^{1.18}\right]$$

where, R is the per unit length resistance, C_g is the per unit length capacitance with respect to ground, C_c is the per unit length coupling capacitance between two adjacent interconnects, l denotes length, ω is width, t is thickness, ρ is resistivity, ε is dielectric constant, s is the spacing between adjacent interconnects and h is the height of the interconnect above ground plane. The delay referred in this work is considered to be 50% propagation delay from input to output. For a local interconnect(1μm$\leq l_{Cu} \leq$20μm), the delay expression($t_{50\%}^{Cu}$) is given as below:

$$t_{50\%}^{Cu} = 0.69R_{d0}\left(C_{g0}+C_{d0}\right) + 0.69\frac{R_{d0}}{S_{opt}}C_{eqpul}l_{Cu} + 0.38R_{eqpul}C_{eqpul}l_{Cu}^2 + 0.69R_{eqpul}S_{opt}l_{Cu}C_{g0} \quad (7)$$

where,

$$S_{opt} = \sqrt{\frac{R_{d0}C_{eqpul}}{C_{g0}R_{eqpul}}} \quad (8)$$

R_{eqpul} and C_{eqpul} can be calculated using the analytical equations described above. For intermediate level interconnects (10μm$\leq l_{Cu} \leq$200μm) and global level interconnects (100μm$\leq l_{Cu} \leq$10000μm), repeaters are inserted to linearize the delay with respect to interconnect length. The expressions for the optimal interconnect length(l_{opt}), optimal driver and load sizes(S_{opt})(i.e. both driver and load sizes are considered the same for simplicity), optimal number of repeaters(N_{opt}) and optimal delay(D_{opt}) are given by,

$$l_{opt} = \frac{l_{Cu}}{N_{opt}} \quad (9)$$

$$S_{opt} = \sqrt{\frac{R_{d0}C_{eqpul}}{C_{g0}R_{eqpul}}} \quad (10)$$

$$N_{opt} = l_{Cu}\sqrt{\frac{R_{eqpul}C_{eqpul}}{2R_{d0}(C_{d0}+C_{g0})}} \quad (11)$$

$$\frac{D_{opt}}{l_{Cu}} = \sqrt{R_{d0}C_{g0}R_{eqpul}C_{eqpul}}\left(2+\sqrt{2\left(1+\frac{C_{d0}}{C_{g0}}\right)}\right) \quad (12)$$

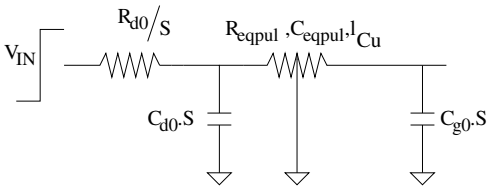


Fig. 4. Schematic of a typical Copper wire.

B. Simulation Model for MWCNT bundle

Equivalent electrical model(i.e. RC model) of MWCNT bundles can be represented like SWCNT bundles as shown in fig.5[15]. We have utilized the compact model developed in[15] for analytical delay estimations. SPICE solves KCL and KVL equations, so the simulation time in SPICE would increase with the complexity of the circuit. The analytical delay estimation is done in MATLAB which computes based on some analytical expressions, so it would be definitely faster than SPICE. So the compact model considered here not only improves the SPICE simulation time by easing the complexity of the original MWCNT model[13] but also allows us to derive the analytical closed form delay expressions by significantly improving the simulation time without any significant loss of accuracy. In fig.5, R_{lumped} is the summation of imperfect metal-nanotube contact Resistance(R_{mc}) and equivalent quantum resistance(R_Q) of an isolated MWCNT. R_{eqpul} is the per unit length ohmic resistance(i.e. scattering induced resistance) of an isolated MWCNT, C_{eqpul} is the per unit length equivalent capacitance(i.e. series combination of quantum and electrostatic capacitance) and l_{CNT} is the length of the MWCNT bundle. Equivalent resistance and capacitance can be calculated considering all of the

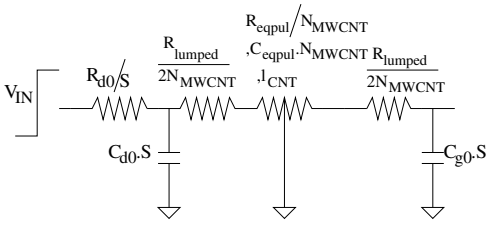


Fig. 5. Equivalent R-C model of a typical MWCNT bundle.

MWCNT's in parallel. We use online field solver CNIA(Carbon Nanotube Interconnect Analyzer) for estimating the electrostatic capacitance values of a MWCNT bundle interconnect[22]. Estimated electrostatic capacitances are shown in Table(III). Now, we denote equivalent bundle resistance and capacitance as follows:

$$Req_{bundle} = \frac{R_{eqpul}}{N_{MWCNT}} \tag{13a}$$

$$Ceq_{bundle} = C_{eqpul} \times N_{MWCNT} \tag{13b}$$

Here, N_{MWCNT} is the number of MWCNTs in a bundle as defined in Section(II). R_{eqpul} and C_{eqpul} have to be evaluated for an isolated MWCNT using the expressions given in Section(II). Then Req_{bundle} and Ceq_{bundle} for a MWCNT bundle are evaluated using (13). For the simulation, we have chosen maximum imperfect contact resistance(R_{mc}) of 20 KΩ. Outer shell diameter (D_{max}) is decided by the technology and inner shell to outer shell diameter ratio is considered to be 0.5[13]. For a local level interconnect, the delay expression(D_{opt}) is given below:

$$\begin{aligned} D_{opt} &= 0.69 R_{d0}(C_{g0} + C_{d0}) \\ &+ 0.69(\frac{R_{d0}}{S_{opt}} + \frac{R_{lumped}}{2N_{MWCNT}})Ceq_{bundle} l_{CNT} \\ &+ 0.38. Req_{bundle} Ceq_{bundle} l_{CNT}^2 \\ &+ 0.69(Req_{bundle} l_{CNT} + \frac{R_{lumped}}{N_{MWCNT}})C_{g0} S_{opt} \end{aligned} \tag{14}$$

where, S_{opt} is given by,

$$S_{opt} = \sqrt{\frac{R_{d0} Ceq_{bundle} l_{CNT}}{Req_{bundle} l_{CNT} C_{g0} + \frac{R_{lumped} C_{g0}}{N_{MWCNT}}}} \tag{15}$$

For intermediate and global level interconnects, the repeaters are inserted just like Copper wire. The expressions for the optimal interconnect length (l_{opt}), optimal number of repeaters (N_{opt}), optimal driver and load sizes (S_{opt}) (i.e. both driver and load sizes are considered the same for simplicity),and optimal delay (D_{opt}) are given by,

$$l_{opt} = \frac{l_{CNT}}{N_{opt}} \tag{16}$$

$$\frac{N_{opt}}{l_{CNT}} = \sqrt{\frac{Req_{bundle} Ceq_{bundle}}{2R_{d0}(C_{d0} + C_{g0})}} \tag{17}$$

$$S_{opt} = \sqrt{\frac{R_{d0} Ceq_{bundle} l_{CNT}}{Req_{bundle} l_{CNT} C_{g0} + \frac{R_{lumped} C_{g0}}{N_{MWCNT}}}} \tag{18}$$

$$\begin{aligned} D_{opt} &= 0.69 R_{d0}(C_{g0} + C_{d0}) N_{opt} \\ &+ 0.69(\frac{R_{d0}}{S_{opt}})Ceq_{bundle} l_{CNT} \\ &+ 0.38. \frac{Req_{bundle} Ceq_{bundle} l_{CNT}^2}{N_{opt}} \\ &+ 0.69(Req_{bundle} l_{CNT} + \frac{R_{lumped}}{N_{MWCNT}})C_{g0} S_{opt} \end{aligned} \tag{19}$$

TABLE III: Electrostatic capacitance for MWCNT bundle

Bundle Capacitance (in fF/μm)	**45 nm**	**32 nm**	**22 nm**	**16 nm**
Local	0.0103	0.0103	0.0101	0.0099
Intermediate	0.0103	0.0103	0.0101	0.0099
Global	0.0121	0.0116	0.0114	0.0093

IV. Results

A. Local level Interconnects

At the local interconnect level, the delay is decided by the capacitance of the bundle, as the ohmic resistance of such a small wire will be insignificant. Resistance will be decided largely by the fixed contact and quantum resistance(R_{lumped}). Fig.6 shows the comparison of MWCNT bundle with Copper wire. We observe that the performance improves with length of interconnect and technology scaling. N_{opt} and r for various level of interconnects are shown in Table(IV). Here Optimum Driver size ratio(r) is the ratio of Optimum driver size(S_{opt}) to minimum driver size. Further it has been observed that the optimum driver size ratio is 4 times in the case of Copper wire compared to MWCNT bundles in 16nm node.

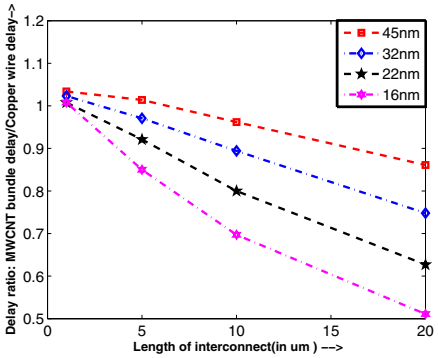


Fig. 6. Delay comparison of local level interconnects.

TABLE IV: N_{opt} and r for various interconnects in $16nm$ node

Interconnect level	Length	Copper		MWCNT bundle	
		N_{opt}	r	N_{opt}	r
Local	20μm	-	28	-	7
Intermediate	200μm	8	28	2	20
Global	10000μm	182	66	35	21

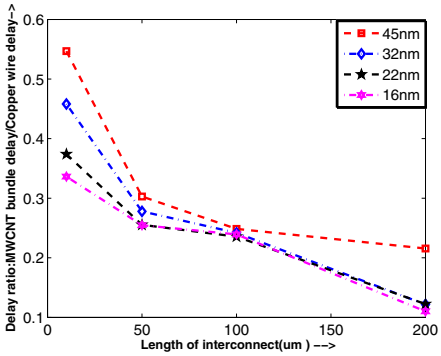


Fig. 7. Delay comparison of intermediate level interconnects.

B. Intermediate level Interconnects

At the intermediate level slowly the ohmic resistance plays a role as it becomes equivalent to it's lumped counterpart as shown in fig.7. As observed from Table(IV), in 16nm node, number of repeaters(N_{opt}) required for MWCNT bundles is approximately $(1/4)^{th}$ as that of Copper wire and the optimum driver size requirement(r) for MWCNT bundle is approximately 70% to that of Copper.

C. Global level Interconnects

Like intermediate level, the ohmic resistance plays a major role in the case of global level interconnects as shown in fig.8. It is also observed from the figure that the delay of MWCNT bundle is approximately 15% compared to Copper wire delay for 16nm node thus acheiving a 85% performance advantage of MWCNT bundle over Copper wire. As observed from Table(IV) requirement of number of repeaters(N_{opt}) for MWCNT bundles are approximately 20% to that of Copper wire, whereas for MWCNT bundle the optimum driver size requirement is $\sim(1/3)^{rd}$ to that of Copper wire in 16nm node. We have compared the ratio of delay of MWCNT bundles and Copper wire with literatures in Table(V). It is observed that our data matches well with data given in[13]. To summarize, the MWCNT bundles prove really competitive compared to Copper wire at all local, intermediate and global level interconnects. Further the performance of MWCNT bundles improve with scaling compared to Copper wire as observed in figs. 6, 7 and 8. So it can be one promising alternative to Copper interconnects in the future generation technology nodes.

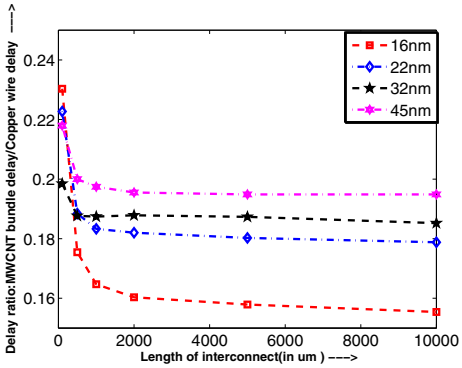


Fig. 8. Delay comparison of global level interconnects.

TABLE V: Ratio of delay of MWCNT bundle and delay of Copper for various technology nodes and interconnect lengths

Interconnect level	Length	Comparison with Literatures	**32 nm**	**22 nm**	**16 nm**
Local	1 μm	SPICE results [13]	1.01	1.02	1.04
		This paper	1.02	1.01	1
Intermediate	0.5mm	SPICE results [13]	0.15	0.15	0.15
		This paper	0.13	0.12	0.12
Global	1mm	SPICE results [13]	0.2	0.18	0.15
		This paper	0.19	0.18	0.16

V. Conclusions

We have explored the feasibility of MWCNT bundles as a replacement of existing copper interconnects. Due to their large diameter, MWCNTs can have higher conductivity compared to SWCNTs. In this work, analytical delay expression of MWCNT bundles have been found out considering a compact and efficient electrical model of MWCNT bundles. It is observed that MWCNT bundles show better performance over copper wire at all level of interconnects and technology nodes. In 16 nm technology node, the performance advantage numbers are 50%, 90% and 85% for 20-μm long local interconnect, 200-μm long intermediate interconnect and 10000-μm long global interconnect respectively. It is also observed that performance numbers improve with scaling for all level of interconnects. Since MWCNT's are easier to fabricate compared to SWCNT's, due to lesser dependence on chirality, they are attractive options for all local, intermediate and global level interconnects in deep submicron technology nodes.

References

[1] S. Ijima,"Helical microtubules of graphitic carbon," *Nature*,vol. 354,pp.56-58, 1991.

[2] Avouris et. al.,"Carbon nanotube electronics," *Proc. IEEE*, vol. 91, no. 11, pp. 1772-1784, Nov. 2003.

[3] Berber et. al.,"Unusually high thermal conductivity of carbon nanotubes,"*Phys. Rev. Lett.*, vol. 84, no. 20, pp. 4613-4616, 2000.

[4] Thess et. al.,"Crystalline ropes of metallic carbon nanotubes," *Science*,vol.273,no.5274,pp.483-487,1996.

[5] Wei et. al.,"Reliability and current carrying capacity of carbon nanotubes," *Appl. Phys. Lett.*, vol. 79, no. 8, pp. 1172-1174, 2001.

[6] P.L.McEuen et. al.,"Single-walled carbon nanotube electronics ",*IEEE Trans. Nanotechnology*,vol. 1 ,no. 1 ,pp. 78-85,Mar. 2002.

[7] Im et. al.,"Scaling analysis of multilevel interconnect temperatures for high-performance ICs," *IEEE Trans. Electron Device*, vol. 52, no. 12, pp. 2710-2719, Dec. 2005.

[8] P. J. Burke, "Luttinger Liquid Theory as a Model of the Gigahertz Electrical Properties of Carbon Nanotubes", *IEEE Trans. Nanotechnology*, Vol. 1, No. 3, pp. 129-144, 2002.

[9] Kaustav Banerjee et. al.,"Current Status and Future Perspectives of Carbon Nanotube Interconnects",*IEEE Conference on Nanotechnology*,pp. 432-436, 2008.

[10] A. Naeemi and J. D. Meindl,"Design and performance modeling for single-walled carbon nanotubes as local, semiglobal, and global interconnects in gigascale integrated systems," *IEEE Trans. Electron Devices*, vol. 54, no. 1, pp. 26-37, Jan. 2007.

[11] A. Naeemi and J. D. Meindl,"Physical modeling of temperature coefficient of resistance for single- and multi-wall carbon nanotubes interconnects," *IEEE Electron Device Lett.*, vol. 28, no. 2, pp. 135-138, Feb. 2007.

[12] Kreupl et. al.,"Carbon nanotubes for interconnect applications", *IEDM Tech. Dig.*, 2004, pp. 683-686.

[13] H. Li et. al.,"Circuit modeling and performance analysis of multi-walled carbon nanotube interconnects," *IEEE Trans. Electron Devices*, vol. 55, no. 6, pp. 1328-1337, Jun. 2008.

[14] Liang et. al.,"Estimation of Time Delay and Repeater Insertion in Multiwall Carbon Nanotube Interconnects",*IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2712-2720, Aug. 2011.

[15] Manodipan Sahoo et. al.,"Efficient and Compact Electrical Modeling of Multi Walled Carbon Nanotube Interconnects", 3^{rd} *IEEE International Symposium on Electronic System Design(ISED 2012)*, pp. 236-240.

[16] W. C. Elmore,"The Transient response of damped linear networks with particular regard to wideband amplifiers," *Journal of applied Physics*, vol. 19, no. 1, pp. 55-63, Jan,1948.

[17] A. Naeemi and J. D. Meindl,"Compact physical models for multiwall carbon-nanotube interconnects," *IEEE Electron Device Lett.*, vol. 27, no. 5, pp. 338-340, May 2006.

[18] H. J. Li et. al., "Multichannel ballistic transport in multiwall carbon nanotubes," *Phys. Rev. Lett.*, vol. 95, pp. 086601-086604, 2005.

[19] A. Raychowdhury and K. Roy,"Modeling of metallic carbon-nanotube interconnects for circuit simulations and a comparison with Cu interconnects for scaled technologies,"*IEEE Trans. Comput. Aided Design Integr. Circuits Syst.*, vol. 25, no. 1, pp. 58-65, Jan. 2006.

[20] International Technology Roadmap for Semiconductors (ITRS) Reports, [Online]. Available:http://www.itrs.net/reports.html

[21] Predictive Technology Model [Online]. Available: http://ptm.asu.edu/

[22] Carbon Nanotube Interconnect Analyzer tool[Online]. Available:http://www.nanohub.org/tools/

Nanocrystalline Cerium Oxide Films as a Functional Material for Photodetectors of Bioluminescent Signal

N.V. Maksimchuk, A.V. Borisov
National Technical University of Ukraine "Kiev Polytechnic Institute"
Kiev, Ukraine
lejanel@ukr.net

Abstract — **Photosensitive nanocrystalline cerium oxide films (CeO_x-films) have been obtained. Physical properties of the thin CeO_x-films have been studied for the purpose of their application as a functional material of new type of photodetectors for bioluminescence registration. Through changing microstructure, morphology and composition of the films we have determined the main principles of their physical properties control. The influence of technological factors on the photoelectrical, optical, temperature and electrophysical properties of the cerium oxide films has been studied. On the basis of the synthesized nanocrystalline CeO_x-films obtained by the "explosive evaporation" method we have developed new types of photodetectors for registration of bioluminescent signal (photoresistors and photodiodes) with enhanced photosensitivity (310-330 μA/lm·V) in the visible spectrum. Photoresistors and photodiodes with CeO_x-films allowed us to create a new type of bioluminescent sensor based on the effective semiconductor photodetectors. Such photodetectors haven't disadvantages typical for photomultiplier tubes and avalanche photodiodes.**

Keywords—nanocrystalline cerium oxide films; CeO_x; CeO_2; Ce_2O_3; thin-film photoresistors; photodiodes; bioluminescent radiation; microelectronic transducers for biosensors

I. INTRODUCTION

Bioluminescent methods are typically based on the measuring of changes in intensity of bioluminescence with the help of photomultiplier tubes (PMT), avalanche photodiodes and CCD detectors. The problems of using the photoelectron multipliers and avalanche photodiodes in the bioluminescent sensors are: limited dynamic range, high cost, complicated technology and high-voltage power supply of the PMT. Other problems are: need of power supply voltage stabilization, existence of uncontrolled breakdown, low sensitivity of avalanche photodiodes to the short-wave bioluminescent radiation. Use of means for luminescent signal registration equipped with photo- or video cameras based on the refrigerated CCD and computer. It is characterized by high cost and cumbersome, which also limits its use. Thus finding and investigation of new sensitive materials as well as obtaining of high-effective films are necessary for development of photodetectors of bioluminescent signals with increased sensitivity.

To eliminate the drawbacks of luminescent biosensors we have developed a new type of photodetectors of the luminescent signal – photoresistors and photodiodes based on cerium oxide films. This photodetectors are characterized by high sensitivity, low dimensions, stability of operations, low cost and rguire low-voltage supply. Using of cerium oxide films as a functional material of microelectronic sensor transducers (thin-film photoresistors, photodiodes) is proposed because of favorable combination of its optical and electrical properties.

Cerium dioxide is not the classical semiconductor and it was not considered in literature as an active material for biosensor transducers. Only in [1] information that mesoporous cerium dioxide films which consist of crystallites with a 5 nm diameter can be used as a functional material of solar cells is presented. Such films are characterized by the increased photoinduced charge carriers lifetime which is greater in comparison with carrier's lifetime for nanocrystalline titanium dioxide. The advantages of solar cells based on CeO_{2-x} are more stable characteristics and 3 times higher operation term compared to the solar cells with TiO_2 [2].

On the basis of publications analysis devoted to the study of cerium oxide films it was determined, that properties of the cerium oxide films and its quantitative relationships with the technological parameters (especially for its use as the active material for sensor transducers) are studied insufficiently. Moreover, the use of CeO_x-films as an active material for biosensor transducers never was considered. On this basis it has been tasked to carry out a comprehensive study of formation of cerium oxide films with the desired functional properties for the creation of bioluminescent signal photodetectors.

II. FABRICATION AND EXPERIMENTAL TECHNIQUES

Thin cerium oxide films were obtained by the vacuum flash evaporation method (the "flash" method) [3]. CeO_x films with thickness of 400–500 nm were precipitated by the method of vacuum flash evaporation from CeO_2 powder. Evaporation was realised in a vacuum from 10^{-3} to 10^{-2} Pa, and the temperature range was 175–300°C. For preparation of photoresistors, two types of monocrystalline silicon substrates with films made of oxidized silicon and silicon nitride were used.

Microstructure of the obtained CeO_x films was studied by scanning electron microscope JEOL JSM 6490LV (accelerating voltage is 20 kV), transmission electron microscope ПЕМ У-М (resolution is 0.21 nm, accelerating voltage is 125 kV) and X-diffractometer DMAX - B Japan [4].

The structural properties of cerium oxide films have been studied by using X-ray photoelectron spectroscopy (SERIES 800 XPS, Kratos Analytical). An X-ray tube with a magnesium anode (U = 12 kV, I = 30 mA) with nonmonochromatic MgK.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

CeO_x-films surface morphology was studied by atomic force microscope NanoScope IIIa (DI, USA) using the Periodic contact silicon probe with the nominal tip radius of 5-10 nm.

Electrophysical properties of the structures based on the cerium oxide films were studied by the I-V characteristics [5], the study of optical and photoelectric properties - with the help of absorption and transmission spectra, spectral photosensitivity and lux-ampere characteristics. The method of high-frequency C-V characteristics was used to study the cerium oxide film/singlecrystal silicon interface [4].

III. Properties of Cerium Oxide Thin Films and Its Application

Structural research of cerium oxide films has shown that cerium oxide films with nanocrystals in amorphous matrix are formed at low substrate temperatures (T_s=175^{0}C). The amorphous phase is represented by Ce_2O_3, while nanocrystalline – by CeO_2 with face-centered cubic crystal lattice. At higher substrate temperatures T_s the "explosive evaporation" method gave nanocrystalline films of CeO_2 with the CaF_2 structural type and amorphous phase was absent. Besides, such CeO_x films obtained at 200^{0}C and 300^{0}C had different grain size (close to 24 nm and 27 nm respectively), different shape of grains, and different preferred orientation.

Information about a significant photosensitivity of the cerium oxide films is absent in literature until recently. We revealed that the polycrystalline CeO_x films were sensitive to the weak bioluminescence signal at the illumination level of 600 quanta/s·cm^2 in the short-wave visible spectrum, moreover the nanocrystalline films had a higher photosensitivity compared to the amorphous ones. As the photosensitivity of films is strongly connected with its optical properties, investigation of optical transmission and absorption spectra of the CeO_x films obtained at different substrate temperatures had been carried out. As a result it was revealed that the CeO_x-films with nanocrystalline structure and CeO_2 phase composition obtained at 200^{0}C characterized by the largest absorption coefficient in the short-wave visible spectrum.

As the bioluminescent methods are usually used at 450-600 nm of wavelengths, the research of the received photodetectors photosensitivity, based on cerium oxide films, in this wavelength range was carried out.

Spectral photosensitivity research of the CeO_x film photoresistors in the spectral wavelengths range 200-800 nm (Fig. 1) has shown that maximum photosensitivity is observed in the range of 450-550 nm, which is typical for bioluminescent radiation.

Transfer characteristics of the optron pairs (semiconductor LED/CeO_x-photoresistor) were investigated. In this case bioluminescent radiation modeled by the LEDs with corresponding spectral characteristics. To calibrate of received photoresistors the range of LEDs currents (0.1-1.2 mA) was set with a standard lumenmeter on the basis of PMT. The photosensitivity was evaluated as the change of the optron output current at low levels of semiconductor LED illumination. Thus for the photoresistors with CeO_x films deposited on the SiO_2/Si substrate a linear dependence of photocurrent density from luminous flux was observed, unlike to this dependence for Si_3N_4/Si substrate.

Photodiode and photoresistor structures were exposed by the calibrated white light source and lux-ampere characteristics measured to determine the integrated photosensitivity.

Finally it vas experimentally determined that maximal photosensitivity coefficient (310-330 μA/lm·V) was registered both for photoresistors and for photodiode structures in which CeO_x-films were obtained at 200°C (Fig. 2). It follows from this that an ordered structure with a minimal amount of recombination centers forms at substrate temperature T_s=200°C and yields higher photosensitivity for both photoresistors and photodiodes.

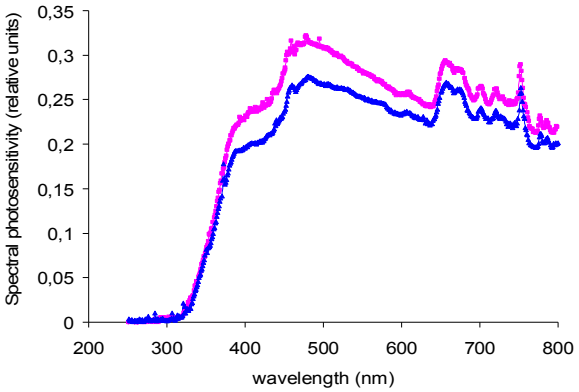


Fig. 1. Spectral photosensitivity of photoresistor samples based on the cerium oxide films obtained at substrate temperature T_s=200°C.

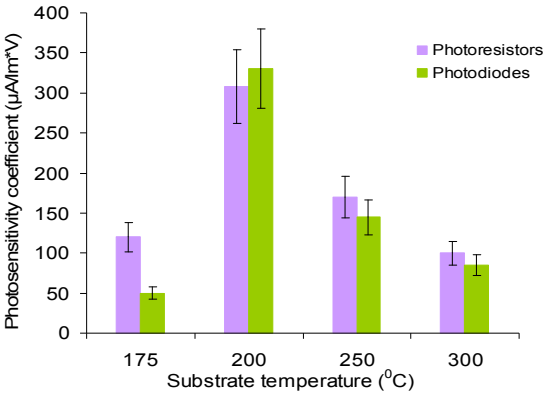


Fig. 2. Photosensitivity coefficients of photodetectors samples based on the CeO_x-films obtained by the explosive evaporation method at different substrate temperatures.

Increased photosensitivity of the cerium oxide films deposited at T_s=200°C can also be explained by the fact that CeO_x-films received at such substrate temperature by explosive evaporation method have prevailing CeO_2 phase which is characterized by several times higher absorption coefficient compared with the Ce_2O_3 phase.

According to researches of the structural, optical, photoelectrical, electrophysical, thermal and C-V

characteristics of CeO_x-films it was revealed that increasing of the substrate temperature from 175 to 200^0C leads to the increasing of photosensitivity, electrical conductivity, coefficient of absorption in the visible range, reduction of temperature resistance coefficient and decreasing of charge states density at the film-substrate interface.

This is caused by the transition from amorphous to nanocrystalline phase and also by changes in film composition (from two-phase Ce_2O_3 and CeO_2 to single-phase CeO_2) of the cerium oxide films produced by the "explosive evaporation" technology. Further increase of the substrate temperature up to 250-300^0C leads to deterioration of the above properties due to the changes in the morphology and phase composition of the CeO_x films.

Thus it was experimentally determined that the best characteristics of bioluminescent signal photodetectors creation had the cerium oxide films deposited by explosive evaporation at substrate temperature of 200^0C. These technological conditions promote formation of nanocrystalline CeO_2 films with a minimum number of recombination centers (the value of the surface states density at the film-substrate interface N_{ss} on a level of $7 \cdot 10^{10}$ $sm^{-2} \cdot eV^{-1}$) and a maximum value of the absorption coefficient ($\sim 10^4$ sm^{-1}), which also explains the increased photosensitivity (3-3.3 mA /lm).

Based on the synthesized thin nanocrystalline cerium oxide films we developed new types of photodetectors for bioluminescent signal registration (photoresistors and photodiodes) which have an enhanced photosensitivity value of 310-330 μA/lm·V and have no typical deficiencies of the photoelectron multipliers and avalanche photodiodes.

On the basis of the efficient photodetectors elaborated in this work a new type of bioluminescent sensor – a portable electronic biolumenometric device for the detection of toxic substances – was developed and studied [6]. The application of semiconductor photoresistors based on cerium oxide films in biolumenometric device instead of photoelectron multipliers or avalanche photodiodes allows to reduce the biolumenometers cost and raise its sensitivity due to the extended sensitive surface area and highly sensitive bridge circuits using alternating current.

IV. Conclusions

It was revealed that increasing of the substrate temperature from 175 to 200^0C is accompanied by the growth in photosensitivity, electrical conductivity, absorption coefficient in the visible range, temperature coefficient of resistance reduction and decreasing of charge states density at the film-substrate interface. This is caused by the transition from amorphous to nanocrystalline phase and also by changes in film composition (from two-phase Ce_2O_3 and CeO_2 to single-phase CeO_2) of the cerium oxide films produced by the "explosive evaporation" technology.

It was experimentally determined, that the best characteristics for creation of bioluminescent signal photodetectors had the cerium oxide films deposited by explosive evaporation method at substrate temperature of 200^0C. These technological conditions promote formation of nanocrystalline CeO_2 films with a minimum number of recombination centers (the value of the surface states density at the film-substrate interface N_{ss} on the level of $7 \cdot 10^{10}$ $sm^{-2} \cdot eV^{-1}$) and a maximum value of the absorption coefficient ($\sim 10^4$ sm^{-1}), which also explains the increased photosensitivity (3-3.3 mA/lm).

Based on the synthesized thin nanocrystalline cerium oxide films we have developed new types of photodetectors for the bioluminescent signal registration (photoresistors and photodiodes) which have an enhanced photosensitivity value of 310-330 μA/lm·V and have no typical deficiencies of the photoelectron multipliers and avalanche photodiodes. Application of cerium oxide based photoresistors in the bioluminometers instead of photomultiplier tubes and avalanche photodiodes allows significantly reducing a cost of bioluminometer and increasing its sensitivity at alternating signal measuring.

Microelectronic transducers based on cerium oxide films developed in this work might be a basis for creation of affordable, portable, highly sensitive and reliable biosensors for environmental monitoring, investigation of the effect of toxic agents on various chemical and biological processes and for other purposes, – instead of a costly, cumbersome and sophisticated analytical equipment used today.

Acknowledgment

Authors are grateful to the M.G. Dusheiko and Yu.V. Yasievich for the providing samples of vacuum condensates.

References

[1] A. Corma, P. Atienzar, H. García and J.-Y. Chane-Ching, "Hierarchically mesostructured doped CeO_2 with potential for solar-cell use," J. Nature Material, vol. 3, pp. 394-397.

[2] V.K. Ivanov, O.S. Polezhajeva, J.D. Tretijakov, "Nanocrystalline cerium dioxide: synthesis, structure-sensitive properties and application", Russian Chemical Journal, vol.53, № 2, pp. 56-67(in Russian).

[3] A.N. Shmyryeva, A.V. Borisov, N.V. Maksimchuk, "Electronic sensors Built on Nanostructured Cerium Oxide Films," J. Nanotechnologies in Russia, vol.5, № 5-6, pp. 382-389.

[4] N.V. Maksimchuk, A.N. Shmyryeva, A.V. Borisov, "Properties and Application of Nanocrystalline Cerium Oxide Films," J. Technology and Designing of Electronic Devices, № 5-6, pp. 54-59 (in Russian).

[5] A.V. Borisov, A.N. Shmyryeva, N.V. Maksimchuk, " Nanocrystalline Cerium Oxide Films for Bioluminescent Sensory Systems," J. Nanosystems, Nanomaterials, Nanotechnology, vol.7, pp. 245-254 (in Russian).

[6] V.G. Melnyk, V.I. Nazarenko, N.F. Starodub, N.V. Maksimchuk, A.N. Shmyryeva, "Electronic Bioluminescent Device for Determination of Toxic Substances," J. Electronics and Communications, vol. 2, pp. 110-114 (in Ukrainian).

Investigation of Self-Shifted Solitons as Part of Femtosecond Supercontinuum

Y. Bashkatov, O. Kutsenko, B. Tsyganok
Faculty of Electronics
NTUU "KPI"
Kyiv, Ukraine
yuriy.bashkatov@gmail.com

V. Khomenko
Department of Quantum and Coherent Optics
IOP NASU
Kyiv, Ukraine
khomenko.vadim@gmail.com

***Abstract* — This article is devoted to description of red shifted solitons appearing during supercontinuum generation in photonic crystal fibers by femtosecond pump. Numerical method for modeling supercontinuum generation is presented and it is shown that experimental results have agreement with theory. Simple relations for description of soliton behavior and Raman influence included.**

Keywords—soliton, photonic crystal fiber, supercontinuum, modeling, femtosecond, Raman shift.

I. INTRODUCTION

Supercontinuum (SC) generation [1] is creation of a broadband laser light source due to the propagation of intense laser pulses in a nonlinear optical medium. SC is useful applications in such diverse fields as spectroscopy, data transition and others. SC generation in photonic crystal fiber (PCF) (Fig. 1) by ultrashort-pulse has become a subject of intense study [2].

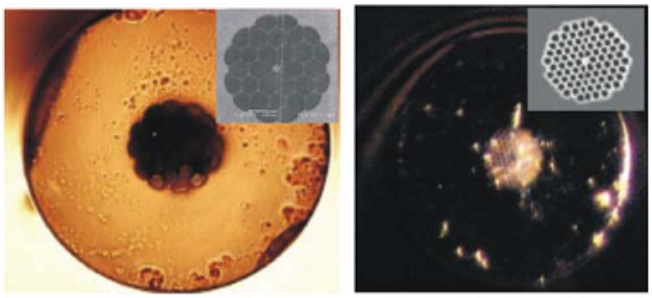


Fig. 1. Photonic crystal fiber

This is caused by ability of PCF to decrease the pulse energies required to generate a supercontinuum due to long distances of effective nonlinear interactions and high nonlinear coefficient [3].

SC cannot be considered as single nonlinear process. It is summary of a couple nonlinear processes Raman scattering, self-phase modulation, four-wave mixing, parametric amplification and soliton generation [4].

Remember mentioned above it can noted that describing each process behavior of SC is predictable.

It have been shown previously [5] that in the anomalous dispersion regime SC became flatter and wider due to increasing Kerr nonlinearities causing self-phase broadening and generated soliton shifted to infrared region due to Raman scattering. This phenomenon known as soliton self-frequency shift (SSFS). The shifting can be controlled by varying the pump power, fiber length and linear chirp of the initial pulse [6].

Solitons are essential to implement nonlinear optical microscopy [7] techniques and reflectance spectroscopy [8]. Practical interest to these scientific techniques gave rise necessity to develop methods for SSFS study, allowing numerically analyze the properties of them.

In this article is reported experimental results related with the generation of a supercontinuum in a PCF. This paper presents a complete characterization of the solitons that are tunable in the 850-1000 nm range. Simple relation for not complicated analysis of soliton characteristics and SSFS is obtained. Also fast method for numerical modeling solitons as part of SC is developed. Theoretical results compared with the experimental ones and determine conditions under which it is possible to vary the soliton pulse characteristics.

II. THEORETICAL INVESTIGATION

A. Soliton characterisation

Remember legend above, we can notice that the effects of group velocity dispersion (GVD) to compensate for the nonlinear effect when the GVD is negative. Under appropriate conditions an optical pulse can propagate through a dispersive, nonlinear medium with an invariant shape - soliton. Their spectral and temporal characteristics can be obtained as:

$$u(z,t) = \sqrt{\frac{|\beta_2|}{\gamma T_0^2}} S \sec h(\frac{t}{T_0}) \exp(j\frac{\beta_2 z}{2T_0^2})$$

$$u(z,w) = \sqrt{\frac{|\beta_2|}{4\gamma}} S \sec h(\frac{\pi\omega T_0}{2}) \exp(j\frac{\beta_2 z}{2T_0^2})$$

where is T_0 pulse width, P_0 – pulse energy, γ – nonlinearity coefficient, β_2 – propagation constant. And pulse energy is related

$$P_0 = \frac{\beta_2}{\gamma T_0}$$

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

B. Raman response

In these formulas the response of the medium was assumed to be instantaneous. However, the contributions to the nonlinearity due to optical phonons are non-instantaneous and give rise to the Raman Effect [9]. This results in a red-shift (mentioned above SSFS) and it is strongest for the highest order soliton (corresponding to the shortest pulse). SSFS can be written as

$$\Delta\omega = \frac{8\beta_2 t_r L}{15 T_0^4}$$

where t_r is a characteristic response Raman effect.

Relations shown above are good for simple description of soliton behavior during supercontinuum generation. For numerical simulation of SSFS solitons, solving expanded nonlinear Schrödinger equation still is necessary.

C. Numerical simulation

Theoretical analysis of propagation of laser pulses in PCFs is based on the numerical solution of the generalized nonlinear Schrödinger equation (GNSE):

$$\frac{\partial A(z,t)}{\partial z} = \sum_{k=2} \frac{i^{k+1}}{k!}\beta_k \frac{\partial^k A(z,t)}{\partial T^k} + i\gamma \times (1 + \frac{\partial}{\partial t}\frac{i}{\omega_0})(A(z,t)\int_{-\infty}^{+\infty} R(t') \times \left|A(z,t-t')\right|^2 dt')$$

where is A is amplitude, β_z –Tailor expansion of propagation constant, $R(t)$ - Raman response function.
Usage of fast Fourier methods for solving GNSE [10] provide easily obtained results (Fig.2).

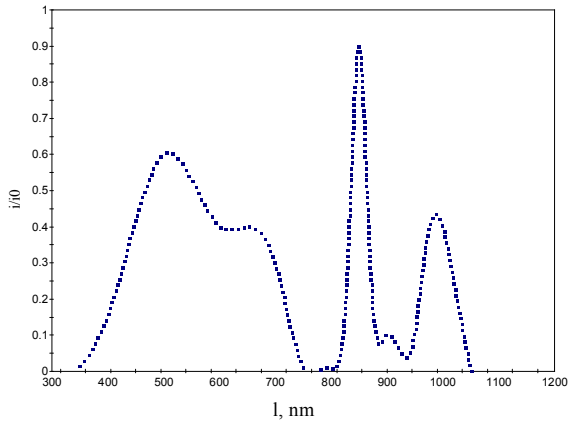


Fig. 2. Average SC simulated numerously with step of 15 ns.

III. EXPERIMNTAL RESULTS AND DISCUSSION

A. Exeperimental setup

The laser radiation was supplied by a femtosecond lasercomplex (Coherent Mira Optima900-F) at the Institute of Physics of the NAS of Ukraine. A typical scheme of the experimental setup is shown in Fig.3.

Basic functional parameters of the complex:

–minimum pulse duration T =70fs;

–maximum pulse power W =2:5 MJ;

–maximum pulse peak power P =3 * 1011 W;

–maximum radiation intensity I0 =1015 W/cm2;

–76-MHz pulse repetition rate.

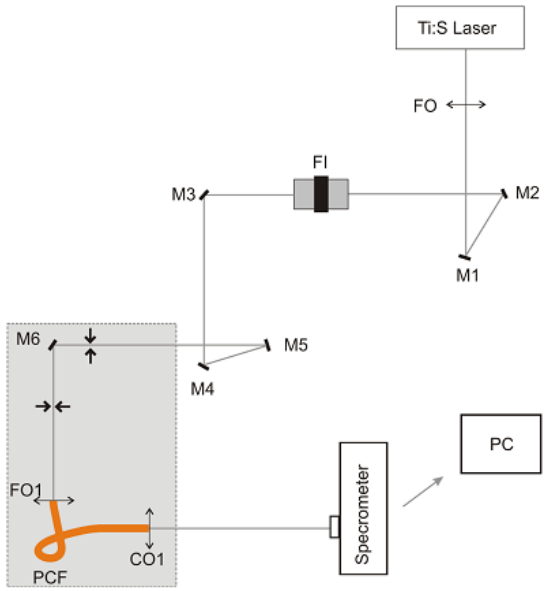


Fig. 3. SC generation scheme: a femtosecond Ti:S laser900Mira Optima-F, FI–Faraday optical isolator, M – mirrors, PCF – photonic-crystal fiber, FO–focusing lenses, CO–collimating lens, PC–personal computer.

B. Results and discussion

SC spectrum obtained at this experimental setup is shown on Fig. 4

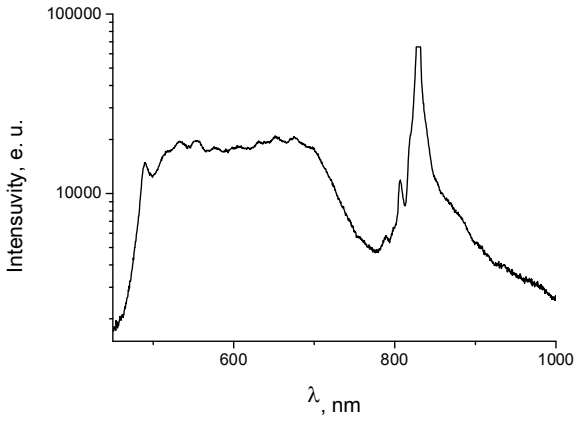


Fig.4. Emission spectra of a supercontinuum generated by femtosecond pulses of a Ti:Sa laser at a wavelength of 820 nm, a power of 500 mW in microstructured photonic-crystal fibers: Thorlabs NL–2.4–800.

In SC spectra analytically filtered SSFS in near IR region (850-950 nm) on Fig. 5 and using Fourier transform obtained temporal characteristics (Fig. 6).

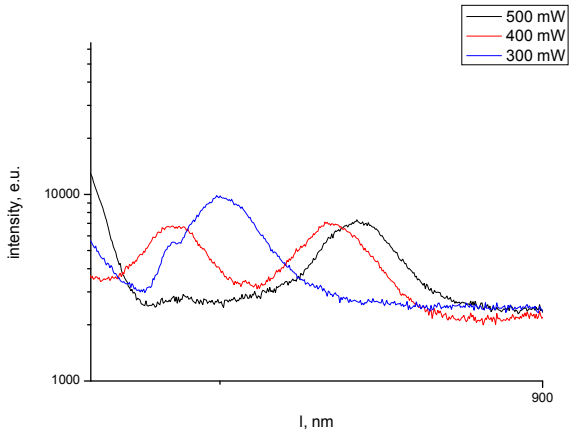


Fig. 5. Obtained SSFS spectra as a part of SC for different pump power

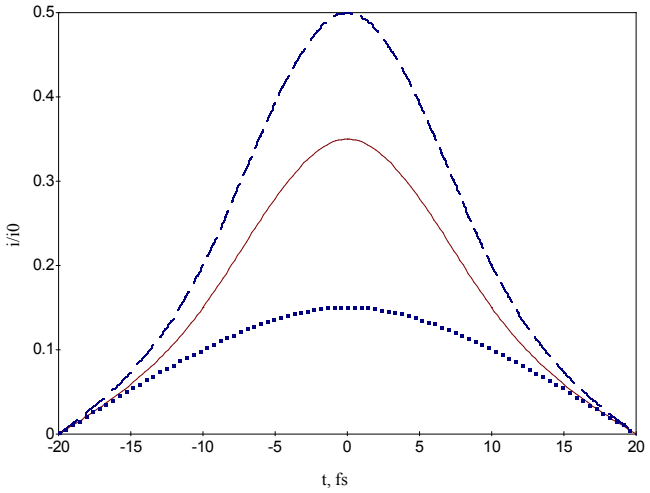


Fig. 6. Temporal characteristics of soliton in SC. Dashed line for 300 mW pump power, doted for 500 mW and solid – 400 nW.

Dissimilarity between the experimental and the numerical results indicate that modeling implemented in this paper not included influence of environment parameters and do not taking into account nonlinearity of high birefringence in PCF core. But it should be mentioned that as can also be seen from the results of experiments and simulations presented above are useful for investigation of SSFS and further filtering for soliton laser sources or spectroscopy. And difference between them is not bigger than 0.01 and is caused by complication of description of external parameters and nonlinear birefringence, but not by choosing and implementation of experimental and theoretical methods. Using fibers with more predictable nonlinearity response and effective compensation of environment influence systems can provide exact concurrency of results.

In Fig. 7 we present the input power dependence of the frequency shift of the red shifted soliton generated in the PCF.

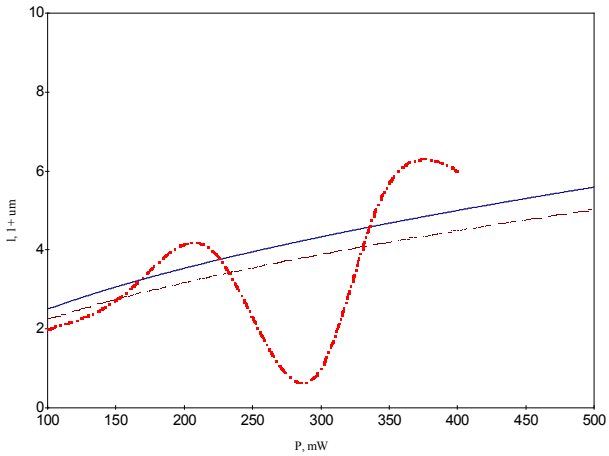


Fig. 7. Raman shifted as a function of input power. Solid line for theoretical results, dashed for adopted experimental results and dashed-doted – experimental results.

It can be seen that the numerically obtained results have an agreement with experimental results due to usage of approximation and decreasing of birefringence and external influence. Both the numerical and experimental data show:

- the Raman-shift monotonically increases with the input power up,
- the spectral width of solitons stays almost constant despite the input power variations.

This behavior is consistent with previously reported results [6, 11]. The discrepancy between the experimental and theoretical results can be attributed to a lack of accurate knowledge of the input pulse chirp and fiber dispersion, in particular of the higher order dispersion [6].

IV. CONCLUSIONS

This article presents complete characterization of SSFS. Existing formulas describing accurate and simple calculating of soliton temporal and spectral characteristics can provide only simple analysis of Raman shifted solitons. For exact and full analysis of SSFS numerical methods of solving GNSE should be built.

And numerical simulations presented in this paper provide adequate description for the central wavelength of the red-shifted soliton in SC spectra and the bandwidth of this component. Also accurate and effective relations for prediction of soliton behavior and evaluation of Raman shift is shown. Using Fourier transform temporal characteristics of soliton can be obtained.

Numerical and experimental data show:

- the Raman-shift monotonically increases with the input power up,
- the spectral width of solitons stays almost constant despite the input power variations.

Dissimilarity between the experimental and the numerical results indicate that modeling implemented in this paper not included influence of environment parameters and do not

taking into account nonlinearity of high birefringence in PCF core. Using fibers with more predictable nonlinearity response and effective compensation of environment influence systems can provide exact concurrency of results

ACKNOWLEDGMENT

Authors acknowledge help researchers of Institution of Physics NASU Department Quantum and Coherent Optics and Department of Radiation Processes (especially Natalia Kachalova, Valerii Voitsekhovich and Boris Danilchenko). Work was done under support of project NASU № 42-02-12.

REFERENCES

[1] R.R Alfano, The Supercontinuum Laser Source. New York: Springer-Verlag, 1989.

[2] J.M . Dudley, G. Genty, S. Coen, "Supercontinuum generation in photonic crystal fiber," Rev. Mod. Phys., Vol. 78, n. 11, pp. 1135–1184 2006.

[3] P.St. J. Russel, "Photonic Crystal Fibers" , Science, Vol.299, pp.358-362, 2003.

[4] A. Zheltikov, A. L'Huillier, F. Krausz, "Nonlinear optics" Springer Handbook Of Lasers And Optics New York: Springer, 2007.

[5] N.M. Kachalova, V.S. Voitsekhovich, A.M. Borodin, L.P. Yatsenko "Investigation of possibilities to control femtosecond supercontinuum characteristics," Ukrainian Journal of Physics, Vol. 55, n. 9, pp. 961–965, 2010.

[6] Licea Rodríguez, J., K. Garay Palmett, R. Rangel Rojo, "Femtosecond pulse source based on soliton filtering from a supercontinuum generated in a microstructured fiber," Revista Mexicana de Física,Vol. 56, n. 4, pp. 311-316, 2010.

[7] R. Hellwarth, P. Christensen, "Nonlinear optical microscopic examination of structure in polycrystalline ZnSe," Opt. Commun., Vol.12, pp.318–22, 1974.

[8] G. Kortum, "Regular and Diffuse Reflection in Reflectance Spectroscopy" Berlin Springer-Verlag, pp. 5–71, 1969

[9] G. P. Agrawal, Nonlinear Fiber Optics. Optics and Photonics, Academic Press, 2001,

[10] Yu. Bashkatov, B. Tsyganok, V. Khomenko, N. Kachalova, V. Voitsekhovich. I. Uvarova, "Modeling supercontinuum generation in microstructured fibers by femtosecond pump pulses using Split-Step Fourier method," International Spring seminar on Electronics technology ISSE, IEEEXplore, pp.309-311, 2012

[11] J. C. Travers, "Blue solitary waves from infrared continuous wave pumping of optical fibers," Opt. Express 17, 1502–1507 (2009).

Development of Technology of Multicomponent Instrument Compositions Formation on the Basis of Porous silicon

Zubko Evgeniya Ivanovna, Shvets Evgenii Yakovlevich
Department of Physical and Biomedical Electronics
Zaporozhye State Engineering Academy, ZSEA
Zaporozhye, Ukraine
evgeniya-zubko@mail.ru

Abstract — **Technological scheme for a formation of electron instrument compositions PcCu /por-Si/n-Si and PcAl/por-Si/n-Si has been developed. The data on the technological adaptation of the deposition and annealing of samples have been given. The effect of annealing for morphological and electrical characteristicks of the obtained compositions has been investigated.**

Keywords—technology; compositions; phthalocyanine; porous silicon; pulverization; structuring; substrates

I. INTRODUCTION

Trends in the development of semiconductor electronics are connected with the search for new materials [1-4]. Among many works of researchers multicomponent electronic instrument compositions based on porous silicon is of the utmost interest. At present known structures made on pore filling technology with magnetic materials CoNi (P), Fe (P), FeNi (P) [5]; deposited by electrolytic anodization Ag [6], Fe [7], Pt, Pd [8]; formed with the help of electrophoretic method from suspension of detonation synthesis ultra-dispersed diamond are known [9]. Recently, the structure development modified with organic substances — the liquid crystals 4-n-octyl-4-cyanobiphenyl (8CB) [10], polymers [11], fullerenes [12] rhodamine, 6G [13] etc. are being conducted. But among the great diversity of organic materials it should be distinguished abound with electrons phthalocyanine (Pc), which have a number of functional characteistics: high degree of purification, environmental safety, low cost and availability of technological methods of synthesis [14]. Therefore, they were chosen to form the multi—electronic compositions in this work.

Porous silicon (por—Si), used as a solid matrix, the pores of which are filled with functional materials is a part of the multi—component electronic instrument compositions. Porous silicon is formed by electrolytic anodization (EA) of crystalline silicon. Wide range of control parameters of EA allows to get both Bragg mirrors and structures used in photovoltaic and nanoelectronics. Thus the formation of structures based on por—Si is simple, predictable transaction, that has good reproducibility. All this makes the por—Si an attractive material for the development of new composite structures. Therefore the aim of this work was to develop the technology of forming of multi—electronic instrument compositions PcCu/por—Si/n—Si and PcAl/por—Si/n—Si by pulverization method with annealing of a sample. In this paper, the following problems have been solved: the creation of hardware and technological scheme of forming electronic instrument compositions PcCu/por—Si/n—Si and PcAl/ por—Si/n—Si; establishing of technological conditions of deposition and annealing of samples; studying of annealing effects on morphological, electrical and electrophysical characteristics from the formed layers.

II. EXPERIMENTAL PART

Wafers of solar quality doped with boron monocrystalline silicon of resistivity 4,5 Om·cm brand with a focus of (100) n—type conductivity of 300 μm thickness have been used in the work. The concentration of boron dopant was $1 \cdot 10^{16}$ cm^{-3}. Anodic etching of the sample was conducted at room light for 10 min. Layers of porous silicon of 7 μm thickness and porosity of 55% have been got. Porous silicon thickness was determined with the help of optical microscopy at "NEOFOT—32", the porosity was determined with gravimetric method. Surface resistivity (ρ_v) of the films was measured with four—probe method in accordance with GOST (All—Union State Standard) 20214-74.

To study the layers morphology of the por—Si raster electron microscope investigation on REM—106H has been used. The phase composition researches have been conducted by X—ray diffraction methods on diffractometer 3M were used. Radiographic studies of samples were carried out immediately after getting them, inseparable from monocrystalline silicon wafer. $Co_{K\alpha}$ — radiation (λ = 1,79021 Å) in the mode of 2θ from 15 to 100º in 0,1 step, and the exposure 3 sec and the time constant of $0{,}5 \cdot 10^3$ have been used.

Drift mobility studies were performed by time—of—flight measurement of the charge carriers through the electric field strength of $F = 2 \cdot 10^3$ — $3 \cdot 10^4$ V/cm. Excess density of charge carriers in the sample was being made by radiation pulse with a wavelength of 0.337 m and duration of ~ 8 ns, obtained from a nitrogen laser LGI-21.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

Impulse voltage of duration ~ 1 ms in the presence of delay ~ 100 ms between the moments of the supply voltage and the radiation pulse was applied to the samples. The measurements were performed in the strong injection mode, when it wasn't possible to neglect the influence of the electric field drift package carriers. The time constant at the amplifier input was much less than the time—of—flight.

As a filler copper phthalocyanine PcCu ($C_{32}H_{16}N_8Cu$) and aluminum phthalocyanine PcAl ($C_{32}H_{16}N_8Al$) were used, which previously were purified by recrystallization from a solution in H_2SO_4 by the following procedure: 100 g of phthalocyanine was dissolved in 1 liter of concentrated H_2SO_4, and then was crystallized by pouring a solution into 3 liters of hot H_2O. The suspension that was formed was heated with stirring till boiling, desilted and filtered. The precipitate was washed with H_2O until neutral reaction of filtrate, and then was dried in the oven at a temperature of 140 °C. Solutions preparation was carried out in four sets: $C_{32}H_{16}N_8Cu$:C_3H_6O:H_2O = 2:39:59 weight. % (1 set)—2:59:39 weight. % (2 set); $C_{32}H_{16}N_8Cu$:C_2H_5OH:H_2O=2:39:59 weight. % (3 set)—2:59:39 weight. % (4 set). Then, on the por—Si, which has not been isolated from the substrate of monocrystalline silicon, was precipitated by spraying solutions methods of the four sets. Thus the usage of solution concentrations 1 and 2, 3 and 4 sets is grounded the formation of grid structure of copper with dimensions of 200—1000 nm, the grid structure of aluminum with dimensions of 500—1200 nm [15]. The given concentration of the solvent is determined the fact that the concentration of sample $C_{32}H_{16}N_8Cu$:C_3H_6O:H_2O= 2:9,8:88,2 weight. %, in the micrographs of the samples there was observed the formation of island—type copper phthalocyanine, and at the concentration of $C_{32}H_{16}N_8Cu$: C_3H_6O:H_2O=2:68,6:29,4 weight. %, spraying is not possible because the hot acetone evaporates, the spray nozzle becomes clogged and the spray operation becomes impossible. Then the samples were being saturated and dried for 2 hours in a desiccator. After that they were annealed at 250—300 ° C in the air for 5—10 min. Hardware and technological scheme of composite layer formation based on porous silicon is shown in Fig. 1.

Films processing PcCu and PcAl was carried out on the experimental equipment, which consists of a pulverization system, heater, ventilation system. The pulverization system has a compressor, pressure gauges, pipes: an induction system of solution, reaction chamber of quartz, pulverizer, pulverization mobility system. A thermostat, heating element, chromel—alumel thermocouple, temperature controller belong to the heater. Ventilation of the reaction chamber is carried out by means of a piping system and a fan.

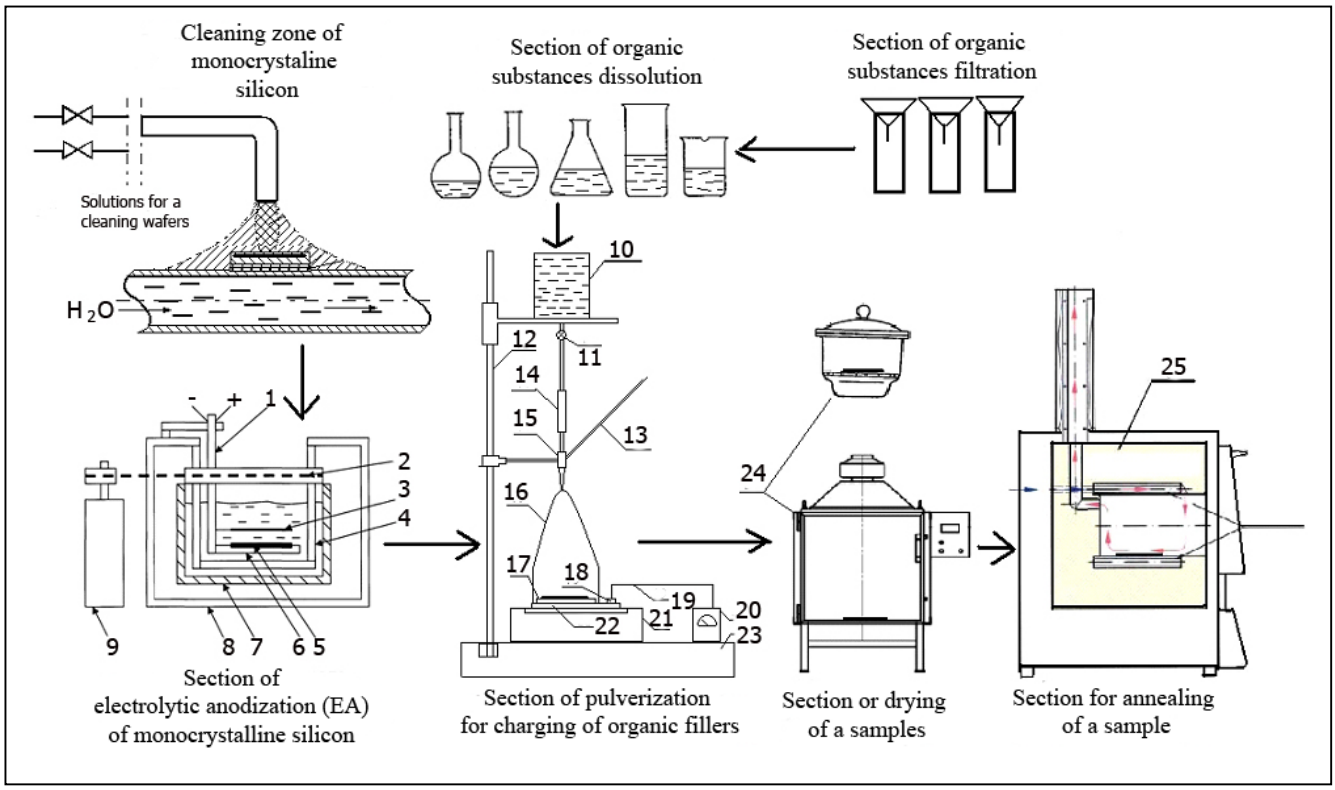


Fig. 1. Technological scheme for lines of manufacturing process for a composite layer: 1—rod, 2—fold, 3— wire of a platinum, 4—rotating electrolytic bath, 5—silicon of a sample, 6—holder, 7—thermostat, 8—frame, 9— electric motor, 10—container with solution, 11—controller, 12—stand 13—pipe to the compressor, 14—rotameter, 15—aerograph, 16—camera 17—substrate, 18—plate satellite, 19— thermocouple; 20—temperature meter, 21—heater, 22—holder, 23— footing, 24—oven or desiccator, 25—muffle furnace

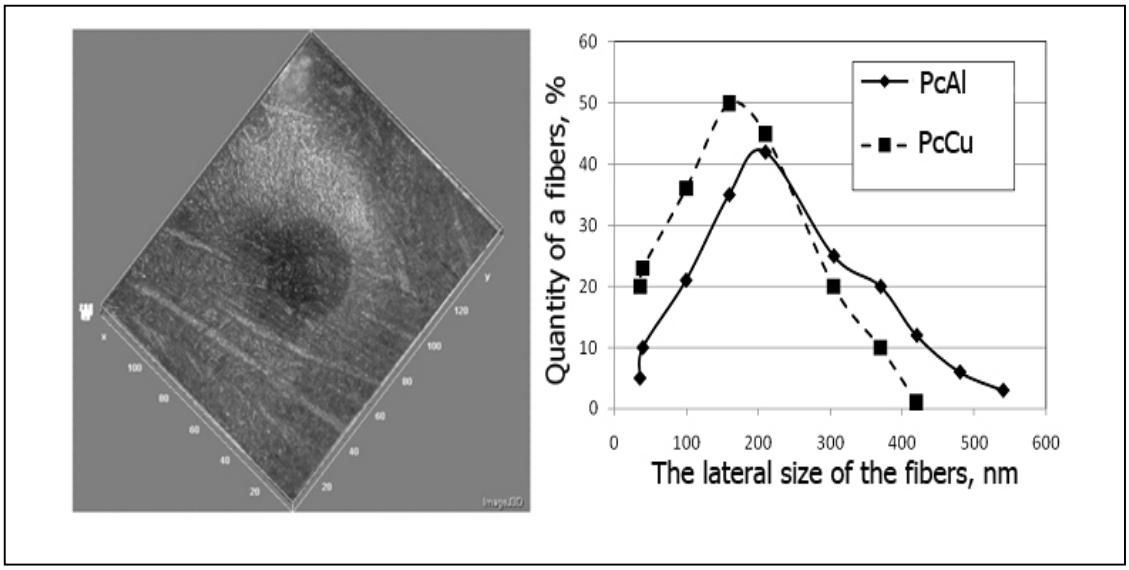


Fig. 2. Fracture photomicrograph structuring of PcAl on the PS surface (a) distribution histogram of fibers Cu and Al according to the sizes on the surface of 2 × 2 μm^2 for electronic instrument compositions (b)

The main parameters of film deposition are as follows: temperature measurement accuracy wasn't worse than 5 %; oxygen pressure at the outlet of the cylinder was 0,1 MPa, the flow rate of the solution during pulverization was maintained at 7 — 10 ml/min. The distance from the pulverizer nozzle to the surface plate was established experimentally and was 28 cm; a nozzle bore diameter from which the solution was sprayed did not exceed 0.5 ml. The oxygen was used as a gas carrier and was supplied intermittently in 3 — 4 sec to prevent surface plate cooling. The spray process lasted for 25 — 40 sec.

III. Basic Material and Results

Ordering in the form of a nanostructured grid on the surface of the porous silicon is observed in the formed electronic instrument compositions PcCu/por—Si/n—Si and PcAl/por—Si/n—Si (Fig. 2, a). The film PcAl/por—Si/n—Si, incorporated into the matrix por—Si is bound, randomly arranged fibers (in the form of "thrown straw") by the diameter of 150—300 nm and by the length of 1 μm, which can be used as a system "contact layer — antireflective coating". Because in the process of sample structuring (saturation and annealing) crosslinked structure formed by cross—linked structure coatings with high adhesion to porous silicon are formed. Electronic instrument composition PcCu/por—Si/n—Si have slightly different locations on the surface of the por—Si than PcAl/por—Si/n—Si. While these are bound fibers, but with short-range order and branch circuit formation in the form even smaller branches. The size of metal fibers in PcCu/por—Si/n—Si differs 1.5 the amount of fibers PcAl/por—Si/n—Si (Fig. 2, b). Analyzing distribution histograms of lateral size for Cu and Al fibers on their number implies that the typical size for structures PcAl/por—Si/n—Si is 160 nm in diameter, and for PcCu/por—Si/n—Si — 220 nm in diameter. Reflexes corresponding to monocrystalline silicon with the corresponding h, k, l have been set in X—ray diffraction patterns in the angular range of 30º and 34º (Fig. 3). Wide "halo" (the extension of diffraction reflections) in the angular range of 17º—19,75º and 27º—29º was also recorded, indicating about the porous silicon, and it is related not only with the existence of the amorphous phase, but also with the size decreasing of silicon microcrystallites. Registered reflex angles of 43,5º, 51º, 73º, 91º indicate that the surface of porous silicon is coated with Cu crystallites.

During the instrument electronic compositions research the dependence of physical effects (annealing of samples) on electrical and electrophysical characteristics of compositions was found (Tabl. 1). In the work [16] on substrates p—Si it is shown that intermittent annealing in the inert environment at 500—550ºC for 5—7 min leads to the sharp decrease (for 4—5 orders) in specific resistance of the porous layer and it is largely due to destruction of complexes which have atoms of hydrogen and boron.

In this article electric characteristics of the received structures have been studied. The experimental results showed that a solution with temperature $t_{solution}$ = 40 ºC sheet resistance of annealed samples was 12 Om/cm and 9,8 Om/cm for 1 and 3 sets respectively. The absence of annealing significantly increases the surface resistance, due to the formation of organic complexes on the surface. In composite films we can observe internal stresses correlation which appear in the film due to the solvent evaporation and structuring of compositions. High values of the internal stresses of compositions cause cracking along the border "monocrystalline silicon—porous silicon" because of annealing. Therefore, to avoid cracking of the structure it is necessary to use a solution with temperature $t_{solution}$ = 40ºC and wafers monocrystalline silicon temperature $t_{mono\text{-}Si}$ = 45ºC.

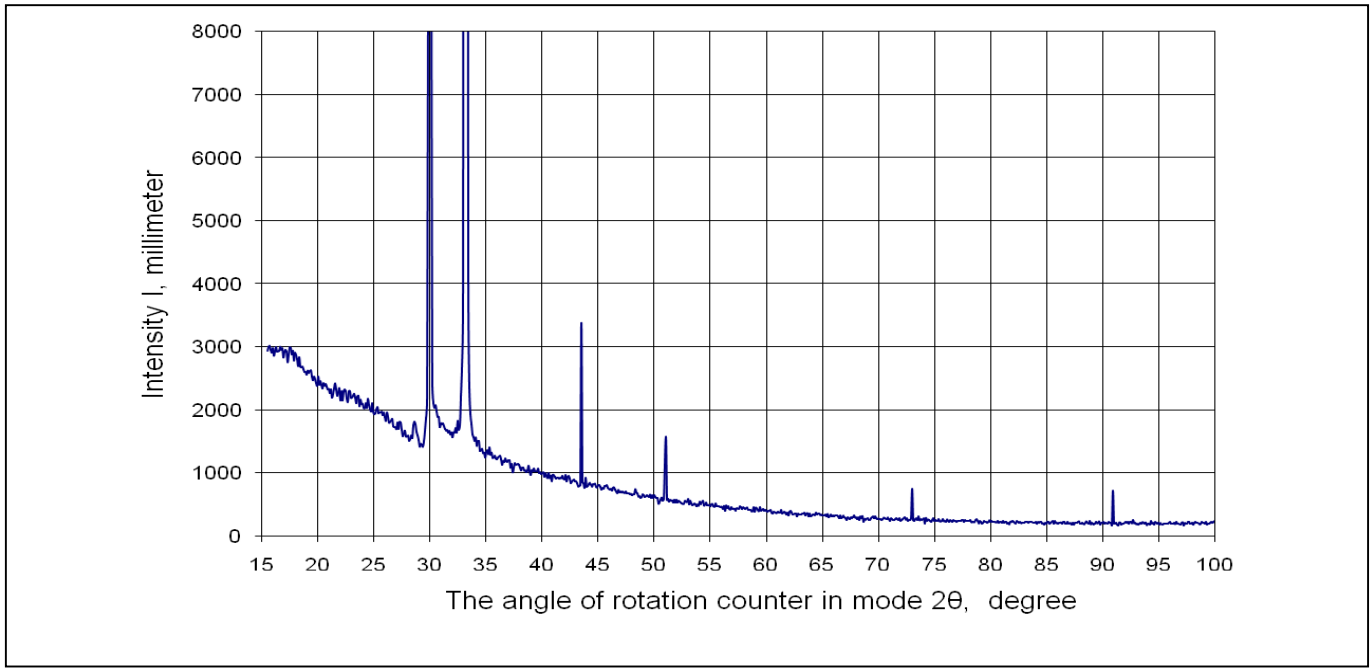


Fig. 3. X-ray analysis of a multicomponent composition based on por—Si

At research of electrophysical characteristics such results are received. At increase in temperature of a substrate in structures PcCu/por—Si/n—Si increase of mobility of carriers of a charge is observed. At solution rise in temperature decrease in mobility of carriers of a charge is observed. In structures PcAl/por—Si/n—Si at increase in temperatures of a substrate and a solution decrease in mobility of carriers of a charge is observed. At solution rise in temperature decrease in mobility of carriers of a charge is observed.

The annealing samples in a current of 5 minutes is of great importance by preparation of such structures. The annealing samples leads to structurization phthalocyanine in the branched out metal network. It is necessary for the greatest sensitivity of porous silicon and the greatest collecting electrons from a surface of porous silicon. Last quality can have the greatest value at creation of a contact and antireflective layer of solar elements of new generation. As such structures PcCu/por—Si/n—Si have straightening characteristics of contact. In structures PcAl/por—Si/n—Si ohmic contacts. At the expense of formation of complexes.

As such samples have stable characteristics in application as touch structures. Also sensitivity researches have been carried out in work [17] to gases of such structures for check of touch properties of porous silicon. Chlorine and bromine steams concern them.

TABLE I. ELECTRICAL AND ELECTROPHYSICAL CHARACTERISTICS OF COMPOSITE LAYERS PCCU/POR—SI/N—SI AND PCAL/POR—SI/N—SI

Composition	Substrate temperature ($t_{mono\text{-}Si}$), ºC	Temperature of the solution, ($t_{solution}$), ºC	Sample annealing	Surface resistivity (ρ_v), Om/cm	Drift mobility of charge carriers (μ), $cm^2/V \cdot c$
PcCu/por—Si/n—Si	50	37	-	25	$6{,}5 \cdot 10^{-3}$
PcCu/por—Si/n—Si	45	40	+	12	$5{,}5 \cdot 10^{-2}$
PcCu/por—Si/n—Si	20	20	-	38	$4{,}2 \cdot 10^{-3}$
PcCu/por—Si/n—Si	20	20	+	10,5	$4{,}1 \cdot 10^{-2}$
PcAl/por—Si/n—Si	50	35	-	21	$5{,}2 \cdot 10^{-3}$
PcAl/por—Si/n—Si	42	40	+	9,8	$5{,}3 \cdot 10^{-2}$
PcAl/por—Si/n—Si	20	20	-	45	$5{,}5 \cdot 10^{-3}$
PcAl/por—Si/n—Si	20	20	+	8,5	$5{,}0 \cdot 10^{-2}$

Thus it is possible to name the main advantage of such structures that they can be created on surfaces of the big area. Thus it is necessary to underline, that the layer phthalocyanine gets structure similar to structure of porous silicon.

IV. CONCLUSIONS

For the first time the technology of multi-electronic instrument compositions PcCu/por—Si/n—Si and PcAl/por—Si/n—Si has been developed. By this technology porous silicon is formed on the plate of monocrystaline silicon n-conductivity and the on top of which a wafers monocrystalline silicon temperature $t_{mono\text{-}Si}$ = 40 ºC is deposited by spraying a solution of $C_{32}H_{16}N_8Cu$:C_3H_6O:H_2O=2:39:59 weight. % (1 set) — 2:59:39 weight. % (2 set); $C_{32}H_{16}N_8Cu$:C_2H_5OH:H_2O = 2:39:59 weight. % (3 set) — 2:59:39 weight. % (4 set) for $t_{solution}$ = 50 º C. The samples were saturated and dried for 2 hours in a desiccator and subjected to annealing at $t_{an.}$ = 250—300 º C in air for 5—10 min. To prevent cracking of the structure along the border, "monocrystalline silicon—porous silicon", on annealing the sample, it is recommended to use from a solution with temperature $t_{solution}$ = 40 ºC and wafers monocrystalline silicon with temperature $t_{mono\text{-}Si}$ = 45 ºC. These data allow to get coatings of crosslinked structures, which are bound randomly arranged fibers (in diameter 150—300 nm and a length of one μm). It should be noted that fiber compositions sizes of PcAl/por—Si/n—Si differ 1,5 times from fibers PcCu/por—Si/n—Si. This should be considered when applying contacts and using in multilayer structures. As formed electronic compositions PcCu/por—Si/n—Si and PcAl/por—Si/n—Si are formed of densely arranged films on the developed surface of por—Si, repeating its structure (relief). Therefore, a porous silicon layer and the composition should have thickness greater than 5 μm.

REFERENCES

[1] A. M. Antropov.G. B. Demidovich, S. N. Kozlov, "Adsorption sensitivity nanocomposite «porous silicon — nickel» to methane", Letters in journal of theoretical physics, vol. 37, № 5, 2011, pp. 43—48.

[2] V. V. Bolotov and others, "Research of layers nanocomposite porous silicon — tin oxide by means of a method spectral ellipsometry", Journal of technical physics,vol. 81, № 11, 2011, pp. 52—57.

[3] A. N. Vinogradov and others, "Magnetooptical and magnetic properties nanocomposite the granulated cobalt—porous silicon", Letters in journal of theoretical physics, vol. 27, № 13, 2001, pp.84—89.

[4] V. M. Kashkarov and others,"Reception nanocomposites porous silicon with iron and cobalt and research of their electronic structure by methods of x-ray spectroscopy", Letters in journal of theoretical physics, vol. 35, № 17, 2009, pp. 89—96.

[5] R. S. Iskhakov and others, "Magnetostructural study ferromagnetic alloy filaments CoNi (P) in the matrix is porous silicon", Technical Physics Letters, vol. 29, № 7, 2003, pp. 1—9.

[6] K. Kleps and others, "Investigation of Silver—, Meso— and Nanoporous Silicon Composite Layers for Biomedical Applications", Romanian journal of information science and technology, vol. 10, № 1, 2007, pp. 97 — 111.

[7] P. Granitzer and others, "Investigation of a Mesoporous Silicon Based Ferromagnetic Nanocomposite", Nanoscale Res. Lett, vol. 5, 2010, pp. 374 — 378.

[8] N. A. Yashultov, "Electron-platinum and palladium nanoparticles on porous silicon", Journal of Fine Chemical Technology, vol. 6, № 3, 2011, pp. 87 — 92.

[9] V. A. Barashkov, A. A. Zakharov and V. A. Uzova, "Production and investigation of diamond dies for regular structures and membranes", Vestn. Krasnoyarsk. State. Tech. Univ.: Heat transfer and hydrodynamics, vol. 14, 1998. pp. 223 — 232.

[10] N. A. Drokin, V. A. Timashov and V. A. Uzova, "Measurements of microwave - permittivity of liquid crystals in porous silicon", Sat tr. 18 Intern. Crimean Conference "Microwave and Telecommunication Technology" (CriMiCo 2008), (in 2 volumes), Sevastopol, Weber, 8—12 September, 2008, pp. 745 — 747.

[11] Yang Yang Li, Vijay S. Kollengode and Michael J. Sailor, "Porous silicon / polymer nanocomposite photionic crystals formed by microdroplet patterning", Advance materials, vol. 17, 2005, pp. 1249—1251.

[12] V. B. Pikulev and others, "Energy transfer in the nanocomposite fotovozbudzheniya poristyvy silicon—fullerene in an oxygen atmosphere", Technical Physics Letters, vol. 32, № 3, 2006, pp. 75—80.

[13] H. Elhouichet, M. Oueslati, "Photoluminescence properties of porous silicon nanocomposites", Materials Science and Engineering, vol. 79, 2001, pp. 27—30

[14] Z. Lukasiak, "Photoluminescence of electrochemically etched porous silicon coated with small—molecule based thin organic films", 13th International Conference "Transparent Optical Networks" (ICTON), Torun, Poland, 26—30 June 2011, pp. 1—3

[15] E. I. Zubko, Zaporozhye State Engineering Academy, Zaporozhye, UK, "A method for manufacturing a contact layer on the surface anti-reflective coating of the solar cell", Ukraine, patent for utility model 67830 Ukraine: IPC (2012.01) N 01 L 43/00, owner. — U 2011 08780, appl. 07/12/2011, publ. 12.03.2012, Bull. № 5, pp. 4.

[16] S. P. Zimin, A. I. Bragin, "Conductivity relaxation in an enclosed porous silicon after heat treatment", Semiconductors, vol. 33, №. 4, 1999, pp. 476 — 479.

[17] E. I. Zubko, E. Y. Shvets, Zaporozhye State Engineering Academy, Zaporozhye, UK, "The gas-sensitive sensor", Ukraine, patent for utility model 74835 Ukraine: IPC (2012.11) N 01 L 43/00, owner. — U 2012 05463, appl. 03/05/2012, publ. 12.11.2012, Bull. № 21, pp. 3.

Multiphysics Effects and Electronic Properties of Anisotropic Semiconductor Quantum Dots

Sanjay Prabhakar[1]
[1] M[2]NeT Laboratory,
Faculty of Science,
Wilfrid Laurier University,
75 University Avenue West,
Waterloo, ON, Canada N2L 3C5
http://www.m2netlab.wlu.ca

Roderick Melnik[1,2]
[2] Gregorio Millan Institute for
Fluid Dynamics, Nanoscience and Industrial Mathematics,
Universidad Carlos III de Madrid,
Avenida de la Universidad 30
E28911, Leganes, Spain
Email: rmelnik@wlu.ca

***Abstract*—In this paper we analyze a number of important coupled effects in semiconductor quantum dots with multiband models, including magneto-electromechanical coupling. We also demonstrate that the Rashba spin-orbit coupling may provide appreciable contributions to electronic properties of quantum dots by focusing on gate-controlled electron spins in quantum dots.In particular, we analyze the properties of anisotropic semiconductor quantum dots formed in the conduction band in the presence of the magnetic field. For this case, we formulate the Kane-type model and based on it we study the properties of dots by using both analytical and finite element techniques. It is shown that that in semiconductor quantum dots, the electron spin states in the phonon-induced spin-flip rate can be manipulated with the application of externally applied anisotropic gate potentials. The spin flip rates can be enhanced by such potentials, which can also reduce the level crossing points to lower quantum dot radii. We provide numerical examples providing further insight into these new findings where it is evident that these observed effects are due to the suppression of the g-factor towards bulk crystal. Based on these findings, the phonon induced spin-flip rate can be controlled through the application of spin-orbit coupling. Other coupled effects that affect the electronic properties of quantum dots are also discussed.**

Keywords*—Low dimensional nanostructures, quantum dots, coupled multiphysics effects, electron-phonon coupling, finite element method, electromechanics, band structure calculations, quantum-continuum coupling.

I. Introduction

A number of coupled multiphysics effects, such as electromechanical, can be used as tuning parameters for the optical response of low dimensional nanostructures, examples of which include quantum wires, rods, and dots. The latter objects have applications in diverse areas from photonics, optoelectronics and band gap engineering to biotechnological applications and medical engineering.

Although in the recent years some coupled effects, e.g. piezoelectric, have often been included in the studies of properties of low dimensional nanostructures, the influence of coupled magneto-thermo-electromechanical effects has barely been included in such studies. This often needs to be supplemented by accounting for the Rashba spin-orbit coupling that can be used as a key parameter in controlling the electron spin in low dimensional nanostructures and, as a result, can profoundly influence their properties. Both of these contributions in the analysis of electronic properties of low dimensional nanostructures are considered in this paper (see details of the related models also in [1], [2]).

In our studies we focus on the so-called 0-dimensional nanostructures or quantum dots (QDs) where we consider several important examples (including, e.g., ferroelectric, piezomagnetic, and magnetoelectric nanostructures [2]). Based on these examples, we have analyzed the influence of magneto-thermo-electromechanical effects by using the fully coupled quantum-continuum models implemented via finite element methodology. In particular, we have shown that the influence of the mentioned coupled effects can be significant within the general fully coupled framework we developed. We have quantified this influence in each of the above three cases.

We note that a special consideration deserves the influence of thermal effects which are commonly neglected in similar studies in the literature [3]. There is a reason for that. Indeed, if we take magnetic quantum dots (in particular, GaN/BaTiO3), we have confirmed that the influence of electromechanical effects on the band structure calculations (and hence on the electronic properties of these nanostructures) and the spin splitting energy are practically independent of temperature. Nevertheless, we have also demonstrated that in piezoelectric quantum dots (in particular, AlN/GaN), the influence of temperature on the electromechanical effects, electronic properties and spin splitting energy is significant. One of the observations is that in such piezoelectric quantum dots, the energy difference between ground and first excited state decreases with the increase of temperature.

Finally, we would like also to highlight our results on the spin splitting energy of ground and first excited states through the application of magnetic fields for several types of quantum dots with and without wetting layers. An interesting novel finding is that in the presence of wetting layers in the quantum dots, Kramer's degenerate eigenvalues and spin splitting energy occur at lower levels. Moreover, the divergence in the Zeeman spin splitting energy increases with the increase in magnetic field due to the Rashba spin-orbit coupling in both cases: quantum dots with wetting layers and without wetting layers. The calculated g-factor provides further insight into the issues discussed here.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

A. Electron Spin States in Gated Quantum Dots

Recently, electron spin states in gated QDs in presence of magnetic fields along arbitrary direction have been measured experimentally [4]. The authors in Refs. [5], [6] by applying both computational and theoretical techniques, have shown that the spin-orbit coupling can be used as a control parameter in the electric/magnetic field tunability of the electron g-factor tensor. Long spin relaxation times have also been measured experimentally in both GaAs and InGaAs QDs by these authors.

These spin-flip rate measurements in QDs confirm theoretical predictions of the suppression of the phonon induced spin-flip rate by spin-orbit coupling with respect to the environment [7], [8], [9].

In the reminder of this paper, we focus on the anisotropic orbital effect on the phonon induced spin-flip rate for InAs quantum dots. Based on both theoretical and finite element numerical simulation methods, we find that the anisotropic potential enhances the spin flip rate and reduces the level crossing point to a lower quantum dot radius. Also we find that the anisotropic potential reduces the variation in the g-factor towards bulk crystal.

II. Mathematical Model

We consider 2D anisotropic semiconductor QDs formed in the conduction band in the presence of magnetic field B along z-direction. The total Hamiltonian $H = H_{xy} + H_{so}$ of an electron in the conduction band under the Kane model [10], [11] can be written as

$$H_{xy} = \frac{\vec{P}^2}{2m} + \frac{1}{2}m\omega_o^2(ax^2 + by^2) + \frac{\hbar}{2}\sigma_z\omega_z, \qquad (1)$$

$$H_{so} = \frac{\alpha_R}{\hbar}(\sigma_x P_y - \sigma_y P_x) + \frac{\alpha_D}{\hbar}(-\sigma_x P_x + \sigma_y P_y), \qquad (2)$$

$$\alpha_R = \gamma_R eE, \qquad \alpha_D = 0.78\gamma_D\left(\frac{2me}{\hbar^2}\right)^{2/3} E^{2/3}, \qquad (3)$$

where $\vec{P} = -i\hbar\nabla + e\vec{A}$ is the 2D electron momentum operator in the asymmetric gauge $\vec{A} = \frac{B}{\sqrt{a}+\sqrt{b}}(-y\sqrt{b}, x\sqrt{a}, 0)$ and $\omega_z = g_0\mu_B B/\hbar$ is the Zeeman frequency. Also, m is the effective mass, μ_B is the Bohr magneton, $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is the Pauli spin matrices, $\omega_0 = \frac{\hbar}{m\ell_0^2}$ is the strength of the parabolic confining potential with quantum dot radius ℓ_0. From Eq. 3, it can be calculated that the strength of the Rashba and Dresselhaus spin-orbit coupling becomes equal at the electric field, $E \approx 3 \times 10^3$ V/cm for InAs quantum dots. However, at this value of electric field, the rotational symmetry is not broken and the spin splitting energy mainly corresponds to the Zeeman spin splitting energy. Above this value of the electric field, only Rashba spin-orbit coupling has an appreciable contribution to the manipulation of spin states in quantum dots via the g-factor tensor and phonon induced spin flip rate.

We now turn to the calculation of the phonon induced spin relaxation rate in between two lowest energy states in QDs. The interaction between electron and piezo-phonon can

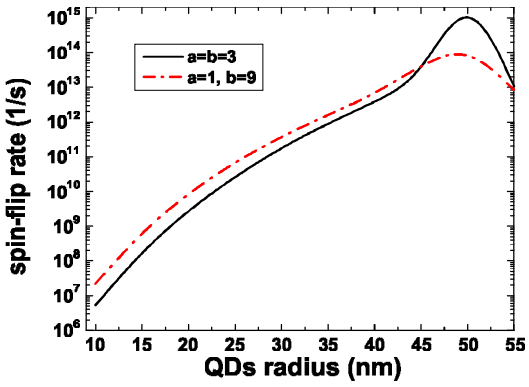


Fig. 1. (Color online) Phonon induced spin-flip rate due to spin-orbit admixture mechanism as a function of QDs radius. We choose the potentials characterized by $a = b = 3$ for symmetric QDs (solid line) and $a = 1$ & $b = 9$ for asymmetric QDs (dashed-dotted line). Also we choose $B = 1\ T$ and $E = 7\times10^5 V/cm$. The material constants for InAs QDs are chosen from Refs. [13] as $g_0 = -15$, $m = 0.0239$, $\gamma_R = 110$ Å^2, $\gamma_D = 130$ eVÅ^3, $eh_{14} = 0.54\times10^{-5}$ erg/cm, $s_l = 4.2\times10^5$ cm/s, $s_t = 2.35\times10^5$ cm/s and $\rho = 5.6670$ g/cm^3.

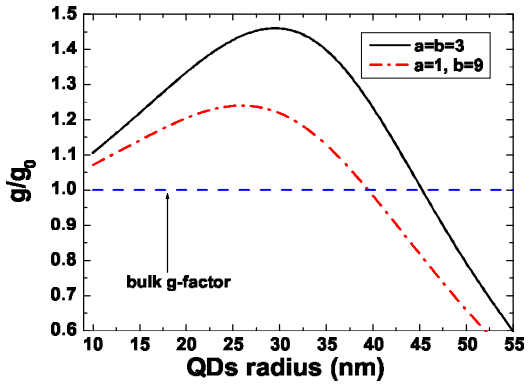


Fig. 2. (Color online) g-factor vs. QDs radius at the potentials characterized by $a = b = 3$ (solid line) for isotropic QDs and $a = 1, b = 9$ (dashed-dotted line) for anisotropic QDs. We choose $E = 7 \times 10^5\ V/cm$ and $B = 1\ T$.

be written as [8], [11], [12]

$$u_{ph}^{\mathbf{q}\alpha}(\mathbf{r}, t) = \sqrt{\frac{\hbar}{2\rho V\omega_{\mathbf{q}\alpha}}}e^{i(\mathbf{q}\cdot\mathbf{r}-\omega_{q\alpha}t)eA_{\mathbf{q}\alpha}b_{\mathbf{q}\alpha}^\dagger} + H.c. \qquad (4)$$

Here, ρ is the crystal mass density, V is the volume of the QDs, $b_{\mathbf{q}\alpha}^\dagger$ creates an acoustic phonon with wave vector $\mathbf{q}$ and polarization $\hat{e}_\alpha$, where $\alpha = l, t_1, t_2$ are chosen as one longitudinal and two transverse modes of the induced phonon in the dots. Also, $A_{\mathbf{q}\alpha} = \hat{q}_i\hat{q}_k e\beta_{ijk}e_{\mathbf{q}\alpha}^j$ is the amplitude of the electric field created by phonon strain, where $\hat{\mathbf{q}} = \mathbf{q}/q$ and $e\beta_{ijk} = eh_{14}$ for $i \neq k, i \neq j, j \neq k$. The polarization directions of the induced phonon are $\hat{e}_l = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$, $\hat{e}_{t_1} = (\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta)$ and $\hat{e}_{t_2} = (-\sin\phi, \cos\phi, 0)$. Based on the Fermi Golden Rule, the phonon induced spin

ransition rate in the QDs is given by [13], [11]

$$\frac{1}{T_1} = \frac{2\pi}{\hbar} \int \frac{d^3\mathbf{q}}{(2\pi)^3} \sum_{\alpha=l,t} |M(\mathbf{q}\alpha)|^2 \delta(\hbar s_\alpha \mathbf{q} - \Delta), \quad (5)$$

where sl ,st are the longitudinal and transverse acoustic phonon velocities in QDs. The matrix elements M (qα) for the spin-flip rate between the Zeeman sublevels with the emission of phonon qα have been calculated perturbatively [11], [14]. Also $\Delta = E_2 - E_1$ is the energy difference between the two lowest states in the quantum dots including spin-orbit coupling.

III. Result s and Discussions

In Fig. 1, we quantify the influence of anisotropic gate potentials on the phonon induced spin-flip rate as a function of quantum dot radius in both symmetric and asymmetric InAs QDs. It can be seen that the anisotropic potential (a = 1, b =9) enhances the spin-flip rate compared to that of symmetric potentials (a = b = 3). Note that we chose the above confining potentials in such a way that the areas of the symmetric and asymmetric QDs are held constant.

The level crossing takes into account, due to mixing, the Zeeman spin states |0, 0,–> and |0, 1,+> in QDs. The level crossing point, determined by the condition $\hbar(\omega_+ - \omega_-) = |g_0|\mu_B B$, for anisotropic QDs is smaller than for the case of isotropic QDs (given that the area of the QDs in both cases is constant). By applying both theoretical and numerical methods in Fig. 1, we report that the anisotropic potential in the spin-flip rate reduces the level crossing point to a smaller QD radius.

In Fig. 2, we study the g-factor vs quantum dot radius in both isotropic and anisotropic quantum dots. It can be seen that the anisotropic potential reduces the variation in the g- factor towards bulk crystal (see dashed-dotted line) compared to isotropic quantum dots (see solid line). This tells us that the anisotropic potential leads to the quenching effect in the orbital angular momentum [15] that pushes the g-factor of an electron towards the bulk crystal.

IV. Concl usions

To conclude, we have provided details on some of the most important coupled effects that affect electronic properties of anisotropic semiconductor quantum dots. This includes magneto-electromechanical effects and spin-orbit coupling. We have also shown that the electron spin states in the phonon induced spin-flip rate can be manipulated with the applica- tion of externally applied anisotropic gate potentials in QDs. The anisotropic potential causes the reduction of the g-factor towards bulk crystal that causes the enhancement of the spin-flip rate and reduces the level crossing point to a lower QD radius. The phonon induced spin-flip rate can be tuned with the application of the Rashba spin-orbit coupling in QDs.

Acknowledgment

This work has been supported by NSERC and CRC programs (Canada).

References

[1] S.Prabhakar, R.V.N.Melnik, P.Neittaanmaki, and T.Tiihonen, Coupled electromechanical effects in wurtzite quantum dots with wetting layers in gate controlled electricfelds: The multiband case, Physica E: Low-dimensional Systems and Nanostructures 46, 97104 (2012).

[2] S. Prabhakar, R.V.N. Melnik, P. Neittaanmaki, and T. Tiihonen, Coupled magneto-thermo-electromechanical effects and electronic properties of quantum dots, Journal of Computational and Theoretical Nanoscience 10, 534547 (2013).

[3] S. Patil and R. V.N. Melnik, Thermoelectromechanical effects in quantum dots, Nanotechnology, 20, 125402, 2009.

[4] S. Takahashi, R. S. Deacon, K. Yoshida, A. Oiwa, K. Shibata, K. Hi-rakawa, Y. Tokura, and S. Tarucha, Large anisotropy of the spin-orbit interaction in a single InAs self-assembled quantum dot, Phys. Rev. Lett. 104, 246801 (2010).

[5] J. M. Elzerman, R. Hanson, L. H. Willems van Beveren, B. Witkamp, L. M. K. Vandersypen, and L. P. Kouwenhoven, Single-shot read-out of an individual electron spin in a quantum dot, Nature 430, 431 (2004).

[6] M. Kroutvar, Y. Ducommun, D. Heiss, M. Bichler, D. Schuh, G. Ab-streiter, and J. J. Finley, Optically programmable electron spin memory using semiconductor quantum dots, Nature 432, 81 (2004).

[7] V. N. Golovach, A. Khaetskii, and D. Loss, Phonon-induced decay of the electron spin in quantum dots, Phys. Rev. Lett. 93, 016601 (2004).

[8] A. V. Khaetskii, and Y. V. Nazarov, Spin relaxation in semiconductor quantum dots, Phys. Rev. B 61, 12639 (2000).

[9] S. Prabhakar, J. Raynolds, A. Inomata, and R. Melnik, Manupulation of single electron spin in a GaAs quantum dot through the application of geometric phases: The Feynman dientangling technique, Phys. Rev. B 82, 195306 (2010).

[10] F. G. Pikus, and G. E. Pikus, Conduction-band spin splitting and negative magnetoresistance in A3 B5 heterostructures, Phys. Rev. B 51, 16928 (1995).

[11] A. V. Khaetskii, and Y. V. Nazarov, Spin-flip transitions between Zeeman sublevels in semiconductor quantum dots, Phys. Rev. B 64, 125316 (2001).

[12] S. Prabhakar, R. Melnik, and L. Bonilla, The influence of anisotropic gate potentials on the phonon induced spin-flip rate in GaAs quantum dots, Appl. Phys. Lett. 100, 023108 (2012).

[13] R. de Sousa, and S. Das Sarma, Gate control of spin dynamics in III-V semiconductor quantum dots, Phys. Rev. B 68, 155330 (2003).

[14] P. Stano, and J. Fabian, Orbital and spin relaxation in single and coupled quantum dots, Phys. Rev. B 74, 045320 (2006).

[15] C. E. Pryor, and M. E. Flatte´, Lande g-factors and orbital momentum quenching in semiconductor quantum dots, Phys. Rev. Lett. 96, 026804 (2006).

SECTION II

BIOMEDICAL ELECTRONOCS AND SIGNAL PROCESSING

Evaluation of Two Methods Based on EMD for Elimination of Blink from EEG

Mohammad Shahbakhti
Department of Biomedical Engineering Islamic Azad University
Dezful, Iran
Mohammad_shahbakhti@yahoo.com

***Abstract* — EEG signals collected from cortex are often contaminated with the eye electrical activity, which increase the difficulty in analyzing the EEG and to obtaining clinical information. In this paper, two different methods based on Empirical Mode Decomposition, for elimination of blink in EEG is compared. The performance index of this experiment is the correlation coefficient between bands of pure EEG and filtered EEG, which indicate the method based on combination of EMD and adaptive filtering outperforms the EMD approach for elimination of blink contamination from EEG signal.**

Keywords—EEG; Blink; EMD; Adaptive filtering; Correlation coefficient

I. INTRODUCTION

An EEG signal is a measurement of currents that flow during synaptic excitations of the dendrites of many pyramidal neurons in the cerebral cortex. EEG signals are often affected with different artifacts that have biological sources similar to EOG, ECG and EMG or external sources like line interference and/ or leads. EOG creates significant artifacts in the electroencephalogram [1]. The eye is similar to an electrical dipole which the cornea acts the negative and the retina acts as positive part. Each movements of eye result in changes in electrical field around this dipole (eye) and causes EOG signal. Two kinds of ocular artifacts can be observed in EEG records, eye blinks and eye movements. Eye blinks are represented by a low frequency signal (< 4 Hz) with high amplitude. It is a symmetrical activity mainly located on the front electrodes (FP1, FP2) with low propagation. Eye movements, are also represented by a low frequency signal (< 4 Hz) but with higher propagation [2].To correct or remove ocular artifacts in EEG, many methods have been proposed, including ICA [3], Regression-based methods, High pass-filtering [4] and Adaptive filtering [5]. ICA can not be applied online and it requires storing the data and off-line processing [6]. All regression methods, whether in time or frequency domain depend on having one or more regressing (EOG) channels [7]. Since EOG artifacts generally consist of low-frequency components, using High-pass filter will remove most of the EOG's artifacts [4]. This method, however, fails when EEG of interest and EOG artifact overlaps or lies in the same frequency band. As a result, a simple filtering approach cannot remove EOG without removing a portion of EEG [4], [8]. Adaptive filtering is useful for on-line removal of EOG [5], however, methods based on adaptive filtering require extra recording channels. This paper compares two new methods for elimination of eye blinks in EEG, which are based on the Empirical Mode Decomposition (EMD) [9] and combination of adaptive filtering and EMD. The EMD was introduced by Huang [10] and has become an established tool in analyzing real world data such as nonlinear and non-stationary signals with a number of important applications in signal processing. The basis of this technique consists of decomposing any signal into a finite number of intrinsic mode functions, which is called IMF [11]. The next sections of this paper describe the two proposed method and the strategy for elimination of blink artifacts from EEG signal. Furthermore, we compare the performance of our last method, which was based on EMD and entropy, to the method based on combination of adaptive filtering and EMD.

II. METHODS

A. The Empirical Mode Decomposition

The EMD step builds on the assumption that any data set consists of different, simple, intrinsic modes of oscillation that need not be sinusoidal, with the non sinusoidal character of each mode of oscillation derived from the data. At any given time, the recorded data may have many different coexisting modes of oscillation, which may or may not relate to different seismological phases. Each of these oscillatory modes, called an intrinsic mode function (IMF), is defined by the following conditions:

1-In the whole dataset, the number of extrema and the number of zero crossings should be equal or differ at most by one.

2-At any point of IMF the mean value of the envelope defined by the local maxima and the envelope defined by the local minima should be zero. A process that leads to IMF's extraction is called, "sifting". This process is done by identify all the local extrema. Connect all local maxima by a cubic spline to produce the upper envelope, and repeat the procedure for the local minima to produce the lower envelope. The mean of these two envelopes is designated as m_1, and the

difference between the data x(t) and m_1 is the first component h_1

$$x(t) - m_1 = h_1 \quad (1)$$

If h_1 had 2 condition of a IMF, it will be consider as first extracted component from the data, otherwise h_1 Will be a signal and the process will be done again.

$$h_1 - m_{11} = h_{11} \quad (2)$$

This flow will be continued k times until h_{1k} is chosen as the first component of signal (IMF).

$$h_{1(k-1)} - m_{1k} = h_{1k} \quad (3)$$

$$c_1 = h_{1k} \quad (4)$$

In next step, c_1 will be subtracted from the data, r_1, and the process would be done again for the extraction of second IMF.

$$x(t) - c_1 = r_1 \quad (5)$$

The residue r_1, which contains longer-period components, is treated as the new data and subjected to the same sifting process as described above. This procedure can be repeated to obtain all the subsequent r_j functions as follows

$$r_1 - c_2 = r_2 \quad (6)$$

The original data are thus the sum of the IMF components plus the final residue

$$x(t) = \sum_{j=1}^{n} c_j + r_n \quad (7)$$

Thus, the data are decomposed into n IMF components and a residue r_n that can be either the mean trend or a constant [12].

B. *The algorithm for elimination of eye blink contamination based on the EMD*

Since, the contaminating blinks increase the power of the frequencies in the lower end of EEG spectrum, the artifact components lie in the last several IMFs. Hence, while sifting process does not reach the artifact components the entropy between two consecutive IMFs decreases toward the decomposition levels. That is, when suddenly the entropy starts to go up, the decomposition of artifact contamination starts. In our last study [9], we proposed the following procedure for EEG signal filtering:

1. Decompose the signal into IMFs.
2. Calculate entropy between two consecutive IMFs.
3. Find the decomposition level, M, at which the entropy starts to go up.
4. Reconstruct the filtered signal by summing up the first M IMFs.

There are 2 defects with the algorithm. Our study in 5 blink contaminated EEG displayed that entropy could always not be applied as prominent index for stopping condition in EMD. In some cases, entropy between two consecutive IMF increase while the process did not reach the artifact components. In addition, this method could not separate artifact and original signal when they overlap in the same frequency band. As a result, the algorithm could not work well in delta band.

C. *Adaptive Filtering*

An adaptive filter is defined as a self-designing system that relies for its operation on a recursive algorithm. Figure 1 illustrates the block diagram of a blink canceller. The primary input is contaminated EEG which is picked up by a particular electrode (e.g Fz) s(n). It is a combination of pure EEG, x(n) and effect of blink artifact, z(n). Reference input to the system, r(n), is the blink. Reference input and the noise components of primary input are correlated in some unknown way. h(m) represents a finite impulse response (FIR) filter of length M. Adjusting the coefficients of the filter, the noise canceller produces an output signal e(n) which is an estimation of background EEG, x(n) [6].

$$e(n) = s(n) - r_1(n) = x(n) - [z(n) - r_1(n)] \quad (8)$$

where

$$r_1(n) = \sum_{i=m}^{M} h(m)\, r(n+1-m) \quad (9)$$

is the filtered reference signals. Among the various algorithms of adaptive filtering, the recursive least-squares (RLS) algorithm is chosen since its superior stability and fast convergence [5].

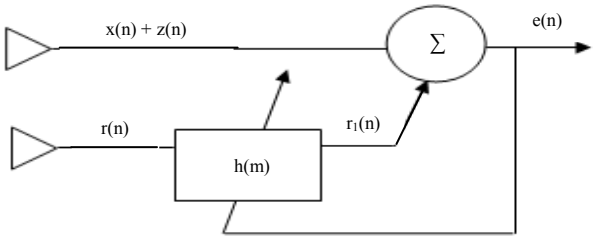


Fig. 1 blink artifact canceller system using adaptive filtering

The significant problem of adaptive method is requiring to an extra channel recording. In this paper, we explain a method to extract blink from contaminated EEG by EMD approach and apply extracted artifact as artifact reference in adaptive filtering.

D. *The proposed algorithm for elimination of eye blink contamination based on combination of adaptive filtering and EMD*

The following procedure is proposed for EEG signal filtering in this experiment:

1. Decompose the signal into n IMFs.
2. Find the decomposition level, M, at which the correlation

coefficient between the summing up the first M IMFs and true EEG be more than 0.90.
3. The last residual will be used as the artifact input in adaptive filtering.
In practice, which the pure EEG does not exist to compare, MSE between two consecutive IMFs is calculated. While, sifting process does not reach the artifact components the MSE between two consecutive IMFs decreases toward the decomposition levels. That is, when the error starts to go up, the decomposition of artifact contamination starts [10]. Therefore, we suggest following steps to reduce blink in EEG:
1. Decompose the signal into n IMFs.
2. Calculate MSE between two consecutive IMFs.
3. Find the decomposition level, M, at which the MSE starts to go up.
4. The residual, $r_{m-1,}$ will be used as the artifact input in adaptive filtering.

The method is implemented by RLS algorithm that includes a forgetting factor 0.9978 and the length of filter is 6.

E. Data Acquisition for the known contaminated EEG

EEG data were collected from F_z electrode and eye blink data were collected from a pair of electrodes which were placed above (negative terminal) and below (positive terminal) the eye. Each record was 3.5 seconds and sampled at 200 Hz [13]. Both signals were bandpassed filtered at 0.1–40 Hz. Using linear combination of the EOG-contaminated EEG signal and the EOG signal is the most common technique for removing ocular artifacts from EEG signals [14].Thus, to generate blink contaminated EEG, 700 samples from pure EEG and blink, under assumption linear combination, have been summed.

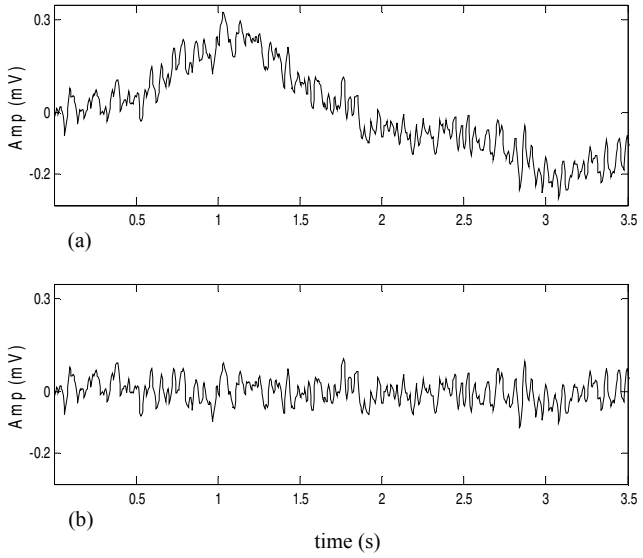


Fig. 2. (a): blink-contaminated EEG signal and (b): uncontaminated EEG signal, respectively.

III. RESULTS

3.5 seconds EEG from Fz electrode with sampling rate of 200 Hz, recorded for 15th ICBME conference, has been used in this study. Here, we compare obtained results from those methods.

A. Empirical Mode Decomposition

In this experiment, after third IMF, entropy between 3rd and 4th IMF is started to go up .Hence, third IMF, is proposed as the end level of the decomposition.

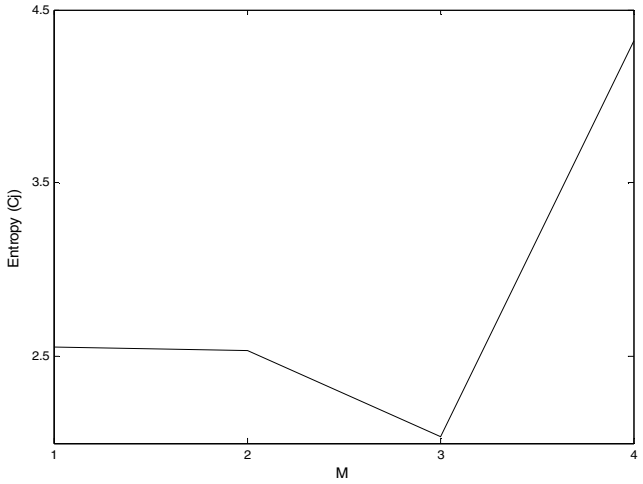


Fig. 3. The Entropy plot shows that the reconstructed signal is obtained by summing up the first three IMFs. (after third IMF, entropy starts to go up, therefore, third IMF should be considered as end level of decomposition)

The filtered signal is formed by summing up first three IMFs. Figure 4 illustrates the corrected EEG by EMD technique and pure EEG.

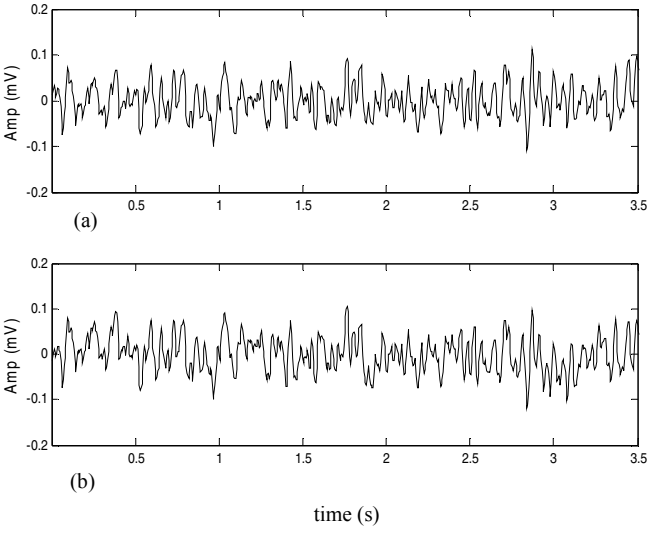


Fig. 4. (a): corrected EEG by EMD method and (b): uncontaminated EEG, respectively

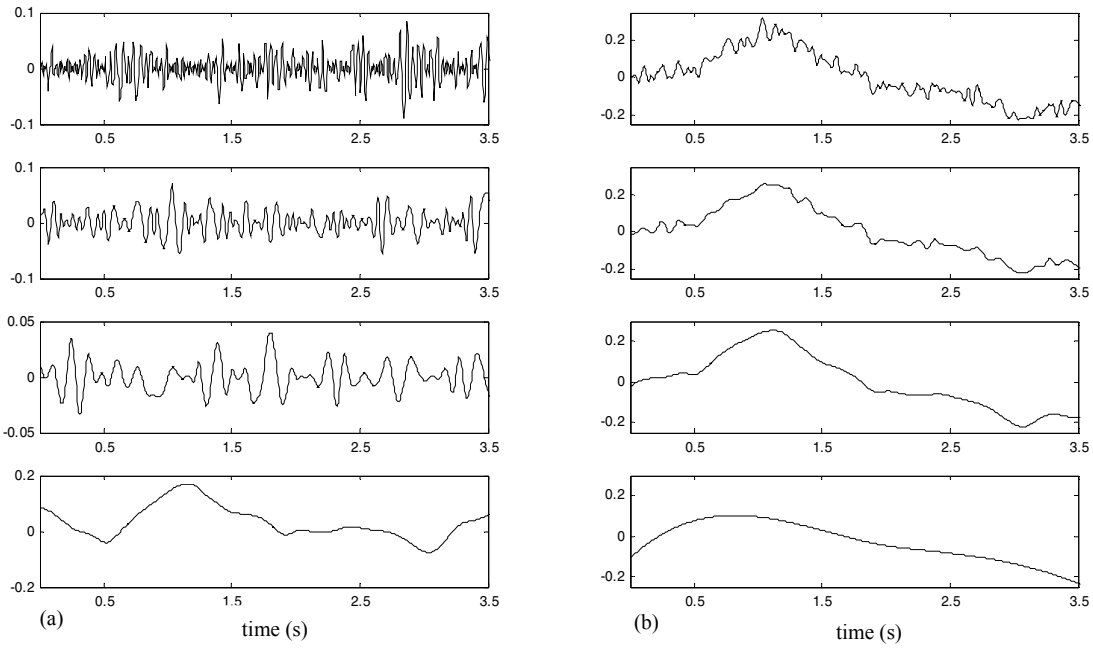


Fig. 5. (a): The IMFs and (b) the residues for known contaminated signal (Fig. 2(a)).

B. Combination of Adaptive filtering and EMD

In this study, after third IMF, the correlation coefficient between true EEG and summing up first three IMFs is raised to 0.9194. Hence, the residual, r_3, is proposed as the blink artifact and has been used as artifact input in adaptive filtering. Figure 6 depicts pure EEG and denoised signal with the novel method and pure EEG, respectively.

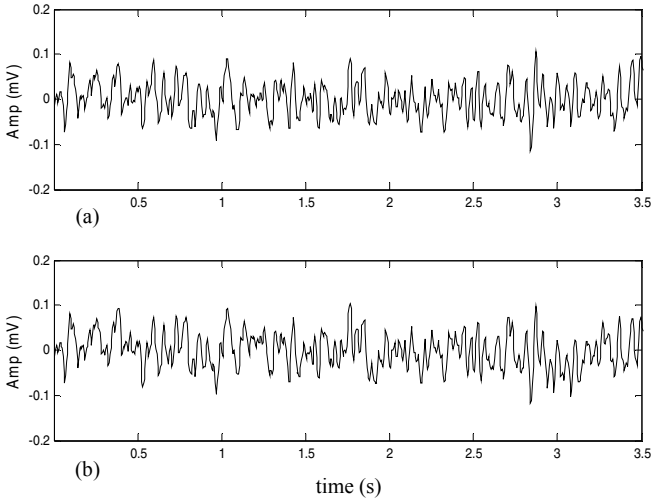


Fig. 6. (a): corrected EEG by EMD & Adaptive filtering and (b): uncontaminated EEG, respectively

Filter coefficients at the last sample (n=700) for M=6, are listed in Table 1.

TABLE I

FILTER COEFFICIENTS AT LAST SAMPLE (n=700) FOR FILTER LENGTH M=6

Coefficients	h_1	h_2	h_3	h_4	h_5	h_6
h(M=6)	0.207	0.187	0.170	0.156	0.141	0.123

Elimination quality is assessed by calculating correlation coefficient between bands of uncontaminated signal and denoised signal. Table II shows the values of correlation coefficient between bands of uncontaminated and denoised signal, for both denoising techniques.

TABLE II

CORRELATION COEFFICIENTS BETWEEN BANDS OF THE TRUE EEG AND THE CORRECTED EEG BY DIFFERENT METHODS

Band	beta	alpha	theta	delta
EMD	0.9996	0.9990	0.9920	0.3717
EMD & adaptive filtering	0.9993	0.9979	0.9889	0.5287

IV. CONCLUSION

In this paper, two different methods based on EMD for elimination of blink from EEG, were compared. Both methods worked almost equal in beta, alpha and theta bands. However, the algorithm based on EMD and adaptive filtering, worked more efficient in delta band.

REFERENCES

[1] Fisch, B.J,. Artifacts. In: Fisch, B.J. (Ed.), Spehlmann's EEG Primer, 2nd edition. Elsevier, Amsterdam, The Netherlands,pp. 108–124, 1991

[2] Crespel, A.; Gélisse, P.; Bureau, M. & Genton, P. Atlas of Electroencephalography (1st ed), Vol. 1 & 2, J. Libbey Eurotext, ISBN 2-7420-0600-1, Paris, 2005

[3] T. JUNG, S. MAKEIG, M. WESTERFIELD, J. TOWNSEND, E. COURCHESNE, nd T. SEJNOWSKI, " Removal of eyeactivity artifacts from visual event-related potential in normal and clinical subjects, " Clin. Neurophysiol., pp. 1745–1758, 2000

[4] G. Geetha, S.N. Geethalakshmi, "Scrutinizing different techniques for artifact removal from EEG signals," International Journal of Engineering Science and Technology (IJEST), Vol. 3 No. 2, pp. 1167-1172, 2011

[5] [5] P. He, G. Wilson, and C. Russell, "Removal Of Ocular Artifacts From Electro-Encephalogram By Adaptive Filtering, " Med. Biol. Eng. Comput., vol. 42, pp. 407–412, 2004

[6] [6] P. Shooshtari, G. Mohamadi, B. Ardekani, and M. Shamsollahi, " Removing Ocular Artifacts from EEG Signals using Adaptive Filtering and ARMAX Modeling, " World Academy of Science, Engineering and Technology., vol. 1, pp 457-460, 2005

[7] V. Krishnaveni, S. Jayaraman, S. Aravind, V. Hariharasudhan, K. Ramadoss, , " Automatic Identification and Removal of Ocular Artifacts from EEG using Wavelet Transform, " MEASUREMENT SCIENCE REVIEW, Volume 6, Section 2, No. 4, pp.45-57, 2006

[8] J. Gao, H. Sultan, J. Hu, and W, Tung, " Denoising Nonlinear Time Series by Adaptive filtering and Wavelet Shrinkage: A Comparison, " IEEE SIGNAL PROCESSING LETTERS, VOL. 17, pp 237-240, 2010

[9] M. Shahbakhti, V. Khalili, G. Kamaee "Removal of blink from EEG by empirical mode decomposition (EMD)," Unpublished.

[10] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis," Proc. R. Soc. Lond, vol. 454, pp.903-995, 1998.

[11] M. Naji, M. Firoozabadi, and S. Kahrizi, " The Application of Empirical Mode Decomposition in Elimination of ECG contamination from EMG signals, 18th Iranian Conference on BioMedical Engineering, Tehran, Iran, pp 143-146, 2011

[12] Ch. Hemanth, A. Sivasa ram, N. Rama krishna, and P. S. Barahmandam, " Non linear and non-Stationary data analysis using Hilbert-huang transform , " Journal of Theoretical and Applied Information Technology, Vol. 29, No.2, pp. 74-84, 2011

[13] N. Hafezimotlagh, M. Khalilzadeh, A. Moghimi, "Removal oof blinks from EEG by PCA and Adaptive filtering, " 15th Iranian Conference on BioMedical Engineering, Mashahd, Iran, pp 308-315, 2008

[14] [14] R. Croft, R. Barry, "Issues relating to the subtraction phase in EOG artefact correction of the EEG, " International Journal of Psychophysiology, vol. 44, pp. 187-195, 2002

QRS Pattern Recognition Using a Simple Clustering Approach for Continuous Data

A. Noack, R. Poll, W.-J. Fischer
Fraunhofer Institute for Photonic Microsystems (IPMS),
Department of Wireless Microsystems,
Dresden, Germany
alexander.noack@ipms.fraunhofer.de

S. Zaunseder
TU Dresden, Faculty of Electrical and Computer Engineering, Institute of Biomedical Engineering,
Dresden, Germany
sebastian.zaunseder@tu-dresden.de

Abstract — **This Paper describes a clustering approach to be used for incoming data under computational constraints at an early stage of the signal processing chain. The algorithm is evaluated on the MIT-BIH Arrhythmia Database (MIT) and the European ST-T-Database (EDB) using a pseudo classification method to estimate the beat identification rates. The algorithm allows an extensive computational simplification, still providing reliable pattern recognition results for normal QRS beat types (Se=96.18 %; +P=99.61 % on MIT and Se=98.26 %; +P=99.95 % on EDB) as well as for ventricular ectopic QRS types (Se=97.61 %; +P=99.64 % on MIT and Se=99.07 %; +P=98.93 % on EDB). Besides its performance in terms of pseudo classification, the computational simplicity and few restrictions regarding its applicability render the proposed clustering method an interesting choice for online-clustering applications even apart from ECG processing.**

Keywords—Electrocardiography, Pattern recognition, Clustering algorithms

I. INTRODUCTION

Holter recording systems are used in a wide range since their first practical introduction by N. J. Holter in 1961 [1], constantly being improved in terms of size, weight, recording quality and functionality. They even found their way into med- ical guidelines. Yet, they mainly aim at arrhythmia detection with focus on ventricular rhythm distortions. However, there are several devices available nowadays, not only providing the usual recoding task but also atrial arrhythmia detection, heart rate estimation, P- and T-wave annotation or ST-segment evaluation [2]–[5]. These developments show that ambulant long time recording tends towards new applications such as

- Identification of atrial fibrillation (AF) in terms of stroke risk stratification [6],
- Estimation of ST-T-Segment elevations as indicator for myocardial ischemia [7],
- Evaluation of QT-Interval prolongation and dynamics for drug effect studies [8], [9],
- Estimation of heart rate variability (HRV) to draw conclusions on the autonomous nervous system [10],
- Calculation of heart rate turbulence (HRT) as a powerful myocardial infarct (MI) risk predictor [11]
- or to provide post-surgery surveillance and therapy monitoring [6].

Also the Fraunhofer IPMS is developing an ambulant 24/7 monitoring system which includes automated electrocardio gram (ECG) analysis algorithms, focusing on ST-T-episode evaluation [12], [13].

The ST-T-episode algorithms as well as most of the above mentioned analysis methods need accurate identification of physiologically excited sinus rhythm heart beats to ensure a proper application of such further processing methods. Therefore, an accurate detection of clinical relevant arrhythmic events does not suffice the needs of those devices but a high accuracy normal beat detection is required additionally. Further development to reach this high accuracy is directed at the exploration of a full cardiac cycle. For now our work focuses on a morphological characterization of the most prominent part of the cardiac cycle, the QRS-complex.

Within this paper the task of morphological characterization is transferred to an online clustering problem, applied to a potentially infinite data input stream. Because the following evaluation algorithms in our mobile device are quite costly in terms of computational effort a simple and fast way to do this evaluation is desired.

The first section of this paper will give a brief outline on data clustering followed by a description of the used data material. In the next section the proposed algorithm will be elucidated. Thereafter its results are presented and discussed. Finally some conclusions will be drawn.

II. METHODS AND MATERIALS

A. Data Clustering

Data clustering describes a structure discovering multivariate analysis method, which assigns objects depending on their individual feature representation into a number of groups. These groups internally ought to be homogeneous, meaning objects of the same group should resemble to each other. Objects associated with different groups should have clearly heterogeneous characteristics to allow a good separation of the groups. [14]

In contrast to a classification task, clustering does not rely on predefined classes but objects are arranged based on the data immanent structures itself. [15], [16]

There are basically three components necessary to describe a clustering algorithm:

- A cluster method,

- A cluster representation approach,
- And a proximity measure.

The *cluster method* specifies the rules, if and how new clusters are created or old ones are deleted, how clusters fuse or how to split them and which connection rules may exist between clusters. Typically cluster methods are separated into so called hard- and soft-clustering methods, partitioning or hierarchically working approaches and disjunctive or overlapping object assignment. [14], [17] However, there may be methods not fitting in one of these categories.

The *cluster representation* describes the way how a cluster is referenced by other clusters or new input objects. Possible representations are nearest-neighbors-approaches (Single-Linkage, Complete-Linkage or K-Nearest-Neighbors), averaging models (Average-Linkage, Weighted-Average-Linkage), input objects as representatives or artificially constructed cluster centers. [14]

The *proximity measure* quantifies the relation between two elements, either the object-cluster-relation or the cluster1-cluster2-relation. Upon this calculation it is determined whether two objects are alike and therefore are assigned to the same cluster or not. An often used measure is the Euclidean Distance given by

$$d(\vec{x},\vec{y}) = \sqrt{\sum_{n=1}^{N}(x_n - y_n)^2}, \tag{1}$$

where N is the number of features in the elements $\vec{x}$ and $\vec{y}$.

B. Databases

To evaluate the proposed algorithm the MIT-BIH Arrhythmia Database [18] (MIT) and the European ST-T Database [7] (EDB) by Physionet [19] are selected. Both provide complete reference annotations (beat times, beat types, rhythm annotations) over a record duration of 30 minutes (MIT) and 2 hours (EDB), respectively. This is a total of more than 850 000 beats to be evaluated with a wide variation of different QRS morphologies. For both databases all records have been evaluated without any exclusion.

C. Quantitative performance assessment by Pseudo Classification

As described in the previous section data clustering discovers data structures and is capable of ordering the given data based on these structures. However, in the context of our work it is important to know if a proposed cluster algorithm is able to distinguish between diagnostically relevant morphologies. Therefore we apply a procedure which we call pseudo classification to allow an assessment, using the provided reference annotations.

If a new cluster is generated it receives a class identifier, based on the annotation of the respective beat in the reference annotation file. For each following beat assigned to this particular cluster, the annotated beat class is compared with the corresponding cluster class identifier. In this way it is possible to estimate if the discovered cluster structures are related to groups of diagnostic value and if the proposed clustering algorithm is capable of distinguishing between beat morphologies of medical relevance.

III. Algorithm Description

A. Boundary Conditions

The algorithm is intended to be used on an ambulatory device for continuous incoming data. This implies heavy computational constraints and requires a clustering approach capable of adapting to the input at the moment of its occurrence, not having the possibility to present the same input vector once again to the algorithm. Additionally the memory space available on a microcontroller is very limited. To define a maximum of iterations through existing clusters and to preliminary determine a memory heap the number of available clusters is limited to ten.

B. Cluster Method

The proposed cluster method is based on a procedure described by Schwenker et al. [20] as *adaptive cluster analysis*. This algorithm is able to create a new cluster if the distance $d(\vec{x},\vec{c}_{j^*})$ between the presented object $\vec{x}$ and the nearest cluster $\vec{c}_{j^*}$ is higher than the threshold θ_{new}, or to merge two cluster if their distance $d(\vec{c}_i,\vec{c}_j)$ (where $i,j = 1\ldots10 \wedge i \neq j$) is smaller than the threshold θ_{merge}. Additionally we added the possibility to delete a cluster if the maximum number of clusters available is reached and a reduction by merging two similar clusters is not possible. In such a case the cluster, not being activated for the longest time, is removed. This keeps the cluster representation adapted to the current input data stream, always having a good representation of the actual QRS complexes.

C. Cluster Representation

Initially there are no clusters defined and the first QRS complex will be handled like having a higher threshold than θ_{new}. A new cluster will be created from it. Following complexes, meeting the merge distance, will be assigned to this cluster and will adapt the cluster center using a similarity depended moving average approach as follows.

$$\vec{c}(n)_{j^*} = (1 - a \cdot s) \cdot \vec{c}(n-1)_{j^*} + a \cdot s \cdot \vec{x}(n) \tag{2}$$

In this equation $\vec{c}(n)_{j^*}$ is the actual winning cluster, $\vec{x}(n)$ is the actual input object at time n, a is a fix adaption constant (a=0.1) and s is the similarity score of $\vec{x}(n)$ and $\vec{c}(n-1)_{j^*}$, which is described in the next passage. This slight modification of the usual moving average adaption diminishes the influence of outliers, preventing the cluster centers from adapting too fast to input vectors, which are only being presented rarely.

D. Proximity Measure

The similarity score s describes the alikeness of two compared objects and is based on the Euclidean Distance in (1). It is defined as

$$s = 1 - \frac{d(\vec{x},\vec{y})}{d_{max}(\vec{x},\vec{y})}, \tag{3}$$

with $d_{max}(\vec{x},\vec{y})$ describing the maximum Euclidean Distance of $\vec{x}$ and $\vec{y}$, by summing the absolute values of their respective

feature components. This maximum distance results if each feature in $\vec{x} = \{x_1, x_2, \ldots, x_N\}$ has the opposing sign to the corresponding feature in $\vec{y} = \{y_1, y_2, \ldots, y_N\}$.

$$d_{max}(\vec{x}, \vec{y}) = \sqrt{\sum_{n=1}^{N} (|x_n| + |y_n|)^2} \tag{4}$$

In that way distance normalization is attained, penalizing a certain difference between two large scale feature vectors less than between two low scale ones. Using this similarity score the merge and creation thresholds θ_{merge} and θ_{new} do not have to be adjusted to the input range but can be chosen as a value between 0 and 1, i.e. independent from signal gain or individual QRS morphologies.

E. Feature Representation

As feature representation a well-known strategy is pursued, which has proven its worth in former applications [21]–[23]. For each QRS complex the feature vector $\vec{x}$ is extracted from a time window of 100 ms before and after the corresponding fiducial point (FP). This generates 10 subsequent 20 ms sum integral values of the area below the curve as shown in Fig. 1.

To allow a separation of normal beats and supraventricular ectopic beats the algorithm is using the beats preceding RR-interval as an additional feature with veto-characteristics. If the difference between the RR-interval of the current object and a cluster center is higher than 15 % the similarity score for this pairing is set to zero regardless of the morphological pattern. In any other case this $10 + 1^{\text{th}}$ feature is not contributing to any similarity estimations.

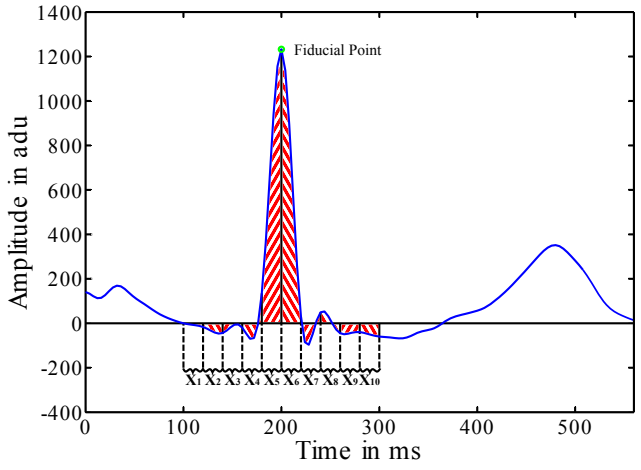


Fig. 1. Feature extraction starting 100 ms before FP and ending 100 ms after it. Each value of $\vec{x}$ is holding the sum integral of 20 ms.

IV. Results

Using the pseudo classification approach with the proposed algorithm a new annotation is created for each record of the database. To compare the new annotation with the provided reference the program `bxb` is used, which is included in the Physionet software package. As stated in [24] the annotations are distinguished into 5 class labels: normal beats (N), supraventricular ectopic beats (S), ventricular ectopic beats (V), fusion beats of a ventricular and a normal beat (F) and paced beats, fusion of a paced and a normal beat or beats that cannot be classified (Q).

TABLE I. Pseudo Classification Results of MIT and EDB

	MIT		EDB	
Class	Se [%]	+P [%]	Se [%]	+P [%]
N	95.74g 96.18a	99.59g 99.61a	98.30g 98.26a	99.95g 99.95a
S	92.23g 75.10a	42.01g 67.08a	77.00g 69.16a	6.36g 53.19a
V	98.12g 97.61a	99.66g 99.64a	97.20g 99.07a	83.40g 98.93a

Abbreviations: g...gross, a...average

TABLE II. Beat to Cluster Ratio in MIT and EDB

Database	Gross [bpc]	Average [bpc]
MIT	27.14	43.87
EDB	22.74	56.93

Using the comparison statistics of `bxb` we calculated the sensitivity (Se) and the positive predictive value (+P) for N, S and V beat types. Table I gives a summary for both databases. Note that the calculations for normal beats and ventricular ectopic beats do not include beats which have been annotated as fusion, since the algorithm shall not be penalized nor rewarded for treatment of this beat type, as stated in [24].

Additionally the number of created cluster centers during a complete evaluation of one record is logged. Combined with the total number of beats present in this record it is now possible to estimate the beat to cluster ratio (BCR) for each data file. This value is showing the average number of beats being represented by one cluster center, and hence, being classified by only one full feature classification if no pseudo classification would have been used. A BCR of 10 is therefore equal to a reduction of the full classification effort by 90 % ($1/10 = 0.1$). Table II is showing the average number of beats per cluster (bpc) for MIT and EDB.

V. Discussion

It has been demonstrated that the proposed clustering approach allows to separate normal beats from ectopic beat forms with a sensitivity of at least 95 % and a positive predictive value above of 99,5 %. This is a quite valuable result considering the vast mass of applications which rely on a secure identification of normal beats to be working correctly.

ST-T-segment analysis for example, is only possible if applied to normal beat signal morphologies, since it is necessary to estimate an isoelectric reference level within the PQ-intervall to calculate a reasonable ST-T-level deviation. Moreover, repolarization behavior will of course alternate depending on the focus of the myocardial excitation spot. [13], [29]

Heart rate variability estimations in are also heavily relying on accurate identification of sinus excited beats. Any ectopic beat will disturb the assessed beat interval series, possibly leading to wrong conclusions about the vegetative nervous system.

TABLE III. COMPARISON OF THE PROPOSED APPROACH TO METHODS DESCRIBED IN LITERATURE

Reference	Used Methods		N	S	V
this algorithm	10 sum values across QRS, dynamic clustering	Se [%]	96.18	75.10	97.61
		+P [%]	99.61	67.08	99.64
Yeh et al. [25]	9 features across PQRST-Segment, fuzzy C-neans	Se [%]	98.28	94.86	92.19
		+P [%]	97.38	93.87	86.82
Jekova et al. [26]	26 features across 2 channels				
	K-Nearest-Neighbors	Se [%]	99.91	-	97.76
		+P [%]	99.82	-	99.92
	fuzzy fogic	Se [%]	99.85	-	97.40
		+P [%]	99.75	-	99.47
	linear discriminant analysis	Se [%]	96.64	-	96.94
		+P [%]	99.37	-	72.29
	neural networks	Se [%]	99.88	-	97.99
		+P [%]	99.83	-	98.65
Korürek et al. [27]	2 RR-intervalls and QRS-amplitude, ACO	Se [%]	95.49	86.78	93.33
		+P [%]	-	-	-
Kamath [28]	TEO in time and frequency domain, NN	Se [%]	100	-	100
		+P [%]	-	-	-

Abbreviations: ACO...ant colony optimization, TEO...Teager energy operator, NN... neural networks

There are smart algorithms available dealing with those disturbances [30], but they are not usable for every case as far as they are relying on RR-intervals only. A morphological estimation would introduce additional physiological information, which may help to avoid drawing wrong conclusions.

Correct estimation of normal and ventricular escape beats is crucial for calculations of the heart rate turbulence (HRT) as well. According to the International Society for Holter and Noninvasive Electrophysiology [11] the calculation of the HRT needs certain recognition of a single ventricular ectopic beat (VES) preceded by at least 2 and succeeded by at least 15 sinus rhythm beats to be used correctly. Moreover, this estimation has to be made for at least 5 isolated VES, since the HRT has to be calculated on averaged tachograms.

With an average sensitivity of 96.18 % and a positive predictive value of 99.61 % for normal beats and a Se of 97.61 % with a +P of 99.64 % for ventricular ectopic beat morphologies, the estimated results on the MIT for the proposed clustering algorithm are reaching similar values as several approaches described in literature. Of course, one has to keep mind, that the presented results are obtained using pseudo classification as a verification method. They are therefore not directly comparable to results of usual classification methods. However, the results of the pseudo classification are well suited to document the usability of the proposed algorithm. Here, a high performance is obtained with a comparatively simple approach. The above mentioned classification methods which include fuzzy hybrid neural networks [25], [26], third order cumulants [31], wavelet entrophy [31], autoregressive coefficients [31], linear discriminant analysis [26] or Teager energy functions [28] usually imply a higher computational complexity of the preliminary feature extraction procedure and their evaluation algorithms. The higher calculation effort in comparison to the proposed method becomes even more evident if the used features are extracted form the whole PQRST-sequence. Jenkova et al. [26] for example, is using a total of 26 features including several slopes, amplitudes, durations, ratios and energy-/area- estimations collected from two channels. This will most likely not be reconcilable with the needs to be implemented on a microcontroller and parallel processing requirements. Table III summarizes some of the mentioned methods and their statistics.

Though its simplicity, the proposed morphological clustering approach reaches considerable identification results with a comparably low computational effort. Particularly the high positive predictive value has to be pointed out, meaning a normal beat labeling is true with an average error of less than 1 %, providing a very good basis for the above explained subsequent estimations.

Identification of SVES though, is not reaching this high level of quality (Se=75.10 %; +P=67.08 % on MIT and Se=69.16 %; +P=53.19 % on EDB). These results are not surprising. Since supraventricular beats are normally following the same electrical excitation path through the heart as normal beats do, their morphologies are typically quite similar. However, the aim of introducing the RR-interval-veto was not to achieve an ideal identification of SVES, but to provide a tool preventing SVES to be considered as normal beats. In this way the positive predictive value of normal beat identification is raised significantly.

As the presented algorithm is only identifying similar signal morphologies on an early stage of the signal processing chain a diagnostic classification step has to follow, if necessary. The idea is to prelocate the proposed approach to a high level classification method, which will speed up classification procedures, only needing to do a full classification on every 50^{th} beat on average. This reduces the full feature classification effort by 98 %. In one case it was even possible to describe more than 2400 beats with only 6 cluster centers, resulting in a BCR of more than 400 bpc, which equals a reduction of 99.75 %. Several other records reached at least values of more than 100 bpc. But there is also the same number of records having a lower ratio than 10 bpc. Considering the cluster center limit of maximum 10 clusters, such low beat per cluster ratios had to be expected in some cases. In records yielding more than 10 different reoccurring beat morphologies this may result in recreation of former known cluster centers, which is of course raising the cluster creation count.

Note, that we did not use any filtering at this early processing stage. Thus, of course more clusters will appear in disturbed signal periods than in such without noise.

VI. Conclusion

Using pseudo classification the capability to separate diagnostically relevant signal morphologies was estimated for the proposed cluster algorithm as individual processing step. By only using a simple morphologic representation of the QRS complex the reliability of the approach was demonstrated. Beat identification quality for normal beats and ventricular ectopic morphologies exhibit similar results as existing methods described in literature. However, the computational effort in terms of algorithmic complexity and preliminary feature extraction procedures is considerably lower in the proposed approach.

The described cluster algorithm is introduced as a preceding evaluation step in an early stage of the signal processing chain. Used in this manner the method proved to be utilized to estimate signal quality and to even reduce the need for a full feature beat classification to 2 % on average.

References

[1] N. J. Holter, "New Method for Heart Studies," *Science*, vol. 134, pp. 1214–20, Oct 1961.

[2] "Vitaphone." [Online]. Available: www.vitaphone.de/

[3] "Schiller Medilog." [Online]. Available: www.medilogholter.com.au/

[4] "Davita cardio." [Online]. Available: www.davita.de/

[5] "Instromedix." [Online]. Available: www.instromedixindia.com/

[6] A. Müller *et al.*, "VDE /DGK Thesenpapier TeleMonitoring- Systeme in der Kardiologie," Nov 2009.

[7] A. Taddei *et al.*, "The European ST-T database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography." *European Heart Journal*, vol. 13, no. 9, pp. 1164–1172, Sep 1992.

[8] B. Darpo, T. Nebout, and P. T. Sager, "Clinical evaluation of QT/QTc prolongation and proarrhythmic potential for nonantiarrhythmic drugs: the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use E14 guideline." *J Clin Pharmacol*, vol. 46, no. 5, pp. 498–507, May 2006.

[9] H. Bonnemeier, "Clinical significance of dynamic QT-interval-analyses," *Herzschrittmacherther Elektrophysiol*, vol. 18, no. 1, pp. 30–38, Mar 2007.

[10] M. Malik, "Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology." *European Heart Journal*, vol. 17, no. 3, pp. 354–381, Mar 1996.

[11] A. Bauer *et al.*, "Heart Rate Turbulence: Standards of Measurement, Physiological Interpretation, and Clinical Use: International Society for Holter and Noninvasive Electrophysiology Consensus," *Journal of the American College of Cardiology*, vol. 52, no. 17, pp. 1353 – 1365, 2008.

[12] S. Zaunseder, W.-J. Fischer, R. Poll, and M. Rabenau, "Wavelet-Based Real-Time ECG Processing for a Wearable Monitoring System." in *International Joint Conference on Biomedical Engineering Systems and Technologies*, 2008, pp. 255–260.

[13] S. Zaunseder, "Anwendungs und prozessorientierte verarbeitung elektrokardiografischer langzeitaufnahmen zur ischämiedetektion," Ph.D. dissertation, Technische Universität Dresden, 2011.

[14] J. Bacher, *Clusteranalyse: Anwendungsorientierte Einführung in Klassifikationsverfahren*. Oldenbourg Wissenschaftsverlag, 2010, vol. 3.

[15] B. Everitt, S. Landau, M. Leese, and S. Daniel, *Cluster Analysis*, 5th ed., ser. Wiley series in probability and statistics. Wiley, 2011.

[16] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data Clustering: A Review." *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.

[17] G. Gan, C. Ma, and J. Wu, *Data clustering - theory, algorithms, and applications*, M. T. Wells *et al.*, Eds. SIAM - Society for Industrial and Applied Mathematics, 2007.

[18] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database." *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[19] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals." *Circulation*, vol. 101, no. 23, pp. E215–E220, Jun 2000.

[20] F. Schwenker, H. A. Kestler, and G. Palm, "Visualization and Analysis of Signal Averaged High Resolution Electrocardiograms Employing Cluster Analysis and Multidimensional Scaling," in *Proc. Computers in Cardiology 1996*, 8–11 Sept. 1996, pp. 453–456.

[21] K. von Olshausen, C. Leitl, and W. Kübler, "Rechnergestützte Datenanalyse für das Pathfinder-Langzeit- EKG-System." *Biomedizinische Technik/Biomedical Engineering.*, vol. 27, pp. 272–279, 1982.

[22] U. Tietze, E.-R. v. Leitner, D. Andresen, and R. Schröder, "A Long term ECG-Analysing System for Quantitative Evaluation of Cardiac Rhythm Disturbances," *Biomedizinische Technik*, vol. 24, no. 11, pp. 275–280, 1979.

[23] J. Millet, M. Perez, G. Joseph, A. Mocholi, and J. Chorro, "Previous identification of QRS onset and offset is not essential for classifying QRS complexes in a single lead," in *Computers in Cardiology 1997*, Lund, Sep. 1997, pp. 299–302.

[24] *ANSI/AAMI EC57:1998 - Testing and reporting performance results of cardiac rhythm and ST-segment measurement algorithms*, Std.

[25] Y.-C. Yeh, W.-J. Wang, and C. W. Chiou, "A novel fuzzy c-means method for classifying heartbeat cases from ECG signals," *Measurement*, vol. 43, no. 10, pp. 1542–1555, 2010.

[26] I. Jekova, G. Bortolan, and I. Christov, "Assessment and comparison of different methods for heartbeat classification." *Medical Engineering and Physics*, vol. 30, no. 2, pp. 248–257, Mar 2008.

[27] M. Korürek and A. Nizam, "A new arrhythmia clustering technique based on Ant Colony Optimization." *J Biomed Inform*, vol. 41, no. 6, pp. 874–881, Dec 2008.

[28] C. Kamath, "ECG beat classification using features extracted from Teager energy functions in time and frequency domains," *IET Signal Processing*, vol. 5, no. 6, pp. 575 –581, sep 2011.

[29] F. Jager, *Ch 9: Introduction to Feature Extraction*, 1st ed., ser. Engineering in Medicine and Biology. Norwood, MA, USA: Artech House, October 2006, vol. 1, pp. 245–264, iSBN 1-58053-966-1.

[30] N. Wessel, H. Malberg, R. Bauernschmitt, and J. Kurths, "Nonlinear Methods of Cardiovascular Physics and their Clinical Applicability," *I. J. Bifurcation and Chaos*, vol. 17, no. 10, pp. 3325–3371, 2007.

[31] M. Engin, M. Fedakar, E. Z. Engin, and M. Korürek, "Feature measurements of ECG beats based on statistical classifiers," *Measurement*, vol. 40, no. 9-10, pp. 904–912, 2007.

Investigation of the Relative Sensitivity of Control of Impulse X-ray Television Systems

Mikhailov S.R.
Faculty of electronics
National Technical University of Ukraine "KPI"
Kiev, Ukraine
smihajlov@ukr.net

Slobodian N.V.
Faculty of electronics
National Technical University of Ukraine "KPI"
Kiev, Ukraine
kafedra.epp@gmail.com

Shilo D.S.
Faculty of electronics
National Technical University of Ukraine "KPI"
Kiev, Ukraine
Dolphin@inet.ua

***Abstract* — method of calculating the relative sensitivity for X-ray television systems based on X-ray screens, CCD cameras and impulse x-ray devices was proposed. To visualize the images formed by the pulsed X-ray radiation, an adjustable duration of accumulation on the CCD camera was used. The relative sensitivity of the control is determined by the intersection of dependences of the luminance and threshold contrast on the size of the defect. The comparison of calculated and experimental values of the relative sensitivity of control is made in this paper.**

Keywords—adjustable accumulation duration; threshold contrast; luminance contrast; penetrameter; monocystalline screen.

I. INTRODUCTION

X-ray system television systems (XRTS) are effectively used for nondestructive testing of materials and products in the industry. Such systems have advantages in performance and cost of control over the systems of film radiography and systems which use storage phosphor plates. XRTS can also monitor objects in statics and dynamics. The most promising are systems based on X-ray screens and CCD cameras, with a number of advantages over other XRTS: the ability to change the X-ray screen, which allows you to change the size of the working area and other parameters XRTS, simple construction, low cost, high performance. Such XRTS effectively used for non-destructive testing of objects by stationary X-ray devices, providing the relative sensitivity of control 1,5 ÷ 2% over a wide range of thickness of steel. Further improvement of the relative sensitivity of control and range of controlled thickness of the object is possible using CCD cameras in an adjustable duration of accumulation [1, 2].

However, the use of impulse devices in non-destructive testing systems is also of great interest [3]. Due to its small overall sizes such systems are mobile and can be used not only in stationary specially equipped laboratories but in the "field" conditions. These systems can also be constructed from a CCD camera with adjustable duration of accumulation.

II. CONSTRUCTION OF A MODEL

The investigated XRTS is made on the basis of an impulse X-ray device MIRA-2D with the X-ray tube IMA2-150D [4] and a CCD camera designed for the problems of industrial nondestructive testing, which allows changing the number of accumulated frames in the wide range (up to 2048). The CCD camera is designed on supersensitive CCD ICX 429ALL [5] manufactured by the "Sony" company with the microlens surface array (EXview HAD). The diagonal size of the matrix ICX-429ALL - 8 mm (format 1/2"), the number of supersensitive elements 752×582, horizontal and vertical sizes of the unit cell of the matrix equal $L_{py} = 8{,}6$ μm and $L_{pz} = 8{,}3$ μm respectively. The camera is equipped with the Computar lens with a relative aperture 1:1,4 and focal length 12 mm.

Calculation of the relative sensitivity was made using groove and wired penetrameters with the parameters, presented in [6]. The works of modeling the X-ray shadow image radiographic systems for NDT with continuous [7] and impulse operation [8 – 10] was added to develop a method of calculating the relative sensitivity.

As shown in [1], the identified flaw in the thickness of controlled object can be detected in the direction of X-ray radiation with some probability if the luminance contrast k_B exceeds (in the extreme case, equal to) the threshold contrast k_{por}.

In order to determine the minimum size of the defect that can be detected by the system, it is necessary to plot the dependencies of luminance and threshold contrast on the size of flaw and to find their point of intersection.

To determine k_B and k_{por} the following formulas were used:

$$k_B = K_S M(\nu), \tag{1}$$

$$k_{por} = \sqrt{\frac{\Psi_{por}^2}{(T_i / T_{EX})(\Omega / A_{CCD})} M^2(\nu) K_N^2 + k_\nu^2}, \tag{2}$$

where $M(\nu)$ – special frequency characteristic (SFC) of XRTS; ν – the spatial frequency, which corresponds to cross sizes of the flaw; Ψ_{por} – the threshold signal to noise ratio; T_i – integration time of visual analyser; T_{EX} – the duration of image accumulation; Ω – the area of flaw reduced to the surface of CCD-array, A_{CCD} – the area of light-sensitive element of CCD-array, k_ν – the threshold contrast of visual analyser; K_S – the image charge contrast defined as the ratio of accumulated

electrons in the unit cell of CCD-matrix that corresponding to the difference signal from the flaw in the object that enlightened $Q_S(H_1,\Delta H_1)/q$ (i.e. useful signal) and the signal from the flawless area of the object $Q_\Phi(H_1)/q$ (background signal):

$$K_S=\frac{Q_S/q}{Q_\Phi/q}; \quad (3)$$

K_N – the relative average fluctuation in calculation per area of supersensitive element of CCD-array during the time of accumulation:

$$K_N=\frac{1}{\sqrt{Q_\Phi/q}}. \quad (4)$$

The expressions $Q_S(H_1,\Delta H_1)/q$ and $Q_\Phi(H_1)/q$ are defined as follows:

$$\frac{Q_S(H_1,\Delta H_1)}{q}=K_{opt}N_kT_{EX}\mathrm{A}_{CCD}\frac{C_{ef}}{hc}\frac{\int_{\lambda_1}^{\lambda_2}\varphi_{nS}(\lambda)F(\lambda)\lambda d\lambda}{\int_{-\infty}^{+\infty}\varphi_{nS}(\lambda)d\lambda}\times$$

$$\times\int_0^{E_{\max}}N^{\sim}(E_{\max},E)\exp\left[-\mu_1(E)H_1\right]\left[1-\exp\left(-\mu_1(E)\Delta H_1\right)\right]\times$$

$$\times\left[1-\exp\left(-\mu_S(E)H_S\right)\right]EdE, \quad (5)$$

$$\frac{Q_\Phi(H_1)}{q}=K_{opt}N_kT_{EX}\mathrm{A}_{CCD}\frac{C_{ef}}{hc}B(E_{\max})\times$$

$$\times\frac{\int_{\lambda_1}^{\lambda_2}\varphi_{nS}(\lambda)F(\lambda)\lambda d\lambda}{\int_{-\infty}^{+\infty}\varphi_{nS}(\lambda)d\lambda}\times\int_0^{E_{\max}}N^{\sim}(E_{\max},E)\times$$

$$\times\exp\left(-\mu_1(E)H_1\right)\times\left[1-\exp\left(-\mu_S(E)H_S\right)\right]EdE, \quad (6)$$

where $N^{\sim}(E_{\max},E)$ – the energy spectrum of flux density for impulse X-ray device; H_1 – sample thickness, ΔH_1 – the local flaw's depth (height); E – the energy of X-rays, keV; $E_{\max}$ – maximum energy of the quanta in keV (numerically equal to the voltage at the anode of the X-ray tube U_A, kV); H_S – monocrystalline screen thickness, cm; $\mu_1(E)$ – the linear attenuation coefficient of X-rays of the controlled sample material, cm^{-1}; $\mu_S(E)$ – the linear coefficient of true absorption of the monocrystalline screen's material, cm^{-1}; $B(E_{\max})$ – accumulation factor, which depends on the accelerating voltage of the X-ray tube, material of the object and its thickness; K_{opt} – the transfer coefficient of the optical system; N_k - the quantity of accumulated frames; T_{EX} – time of the accumulated frame, s; C_{ef} – conversion efficiency of the X-ray monocrystalline scintillation screen; λ – wavelength of the optical radiation, cm; $F(\lambda)$ – dependence of the quantum efficiency of CCD-matrix on the wavelength of optical radiation; $\varphi_{nS}(\lambda)$ – normalized spectral emission characteristic of the monocrystalline screen; λ_1 and λ_2 – limit the range of wavelengths in which product characteristic of $\varphi_{nS}(\lambda)$ and $F(\lambda)$ not zero; h – Planck's constant; c – the speed of light.

To determine the energy spectrum of flux density of the impulse X-ray device $N^{\sim}(E_{\max},E)$, a model of characteristics of the impulse X-ray devices was used, developed in [8 - 10].

As a factor of pulsed operation of such X-ray sources the model takes into account altering voltage at the anode, the altering anode current during the impulse duration τ and the characteristics of their design (impulse tube with transmissive and reflective anodes). For the calculations a simplified form of the time dependences of voltage and current was used, which reflects the appropriate real dependences of characteristics of the tubes with explosive emission, namely in the time interval from 0 to τ_1 accelerating voltage is constant, and the current increases linearly, from τ_1 to τ_2 - the voltage and the current linearly decrease to zero. According to the model $N^{\sim}(E_{\max},E)$ - the distribution of quantum radiation energies of the impulse X-ray device in average during the period of pulses repetition τ is determined by the equations:

$$N^{\sim}(E_{\max},E)=[D_1(E_{\max},E)+D_2(E_{\max},E)]/\tau, \quad (7)$$

where:

$$D_1(E_{\max},E)=\int_0^{\tau_1}N_E(E_{\max},E)I_{\max}(E_{\max})\frac{t}{\tau_1}dt, \quad (8)$$

$$D_2(E_{\max},E)=\int_{E(\tau_2-\tau_1)/E_{\max}}^{\tau_2-\tau_1}N_E\left(E_{\max}\frac{t}{\tau_2-\tau_1},E\right)\times$$

$$\times I_{\max}(E_{\max})\frac{t}{\tau_2-\tau_1}dt, \quad (9)$$

$E_{\max}$ – maximum energy of X-ray radiation, keV; E – energy of X-ray quants, keV; t – time, s; τ_2 – duration of the pulses; τ – repetition period of the pulses; $I_{\max}$ – current amplitude of the pulse, A;

$$I_{\max}(E_{\max})=1.12\cdot10^6(C/d_{AK})^{1/2}E_{\max}^{3/4}, \quad (10)$$

$$N_E(E_{\max},E)=\frac{k_1Z\cdot I_{\max}\cdot(E_{\max}-E)}{r_s^2E}\times$$

$$\times\exp[-\mu_m(E)y(E,E_{\max})-\mu_0(E)H_0], \quad (11)$$

For the tubes with transmissive anode:

$$y(E,E_{\max})=\Delta-\frac{E_{\max}^2-E^2}{\rho_mC_{TB}(E_{\max})}, \quad (12)$$

where C – capacitance which discharges on the impulse tube during the pulse duration, F; d_{AK} – distance between the anode and the cathode of the tube, cm; $k_1=1.1\cdot10^{12}$ quanta/ /A·s·keV; Z – atomic number of the anode material; r_s – distance from the X-ray tube, cm; H_0 – thickness of the output window of the X-ray tube, cm; $\mu_m(E)$ and $\mu_0(E)$ –

linear coefficients of weakening of the X-rays in the anode material and in the window of X-ray tube respectively, cm^{-1}; ρ_m – density of the anode material, g/cm^3; $C_{TB}(E_{\max})$ – Thomson-Whiddington constant; Δ – transmissive anode thickness, cm.

Equation (10) expresses the ratio of amplitudes of the current pulses and the anode voltage (corresponding to the value $E_{\max}$), typical for the tubes with explosive emission. Equation (11) defines $N_E(E_{\max}, E)$ - instant distribution of photons of the impulse X-ray device radiation by the energies per unit current. Equation (12) takes into account the geometrical factor: the path length $y(E, E_{\max})$ which photons with different energies pass in the anode volume from their place of origin to the exit, defines self-absorption of X-rays in the anode matter.

III. CALCULATIONS

This paper deals with the calculation performed for the impulse X-ray tube IMA2-150D with transmissive tungsten anode, anode thickness is 0.02 mm. Distribution of photons of the impulse X-ray device radiation was averaged over the period of pulses repetition, which for the machine MIRA-2D is 0.1 s [4].

For the optical system used, transfer coefficient $K_{opt} \approx 0.075$; $A_{CCD} = L_{py} \times L_{pz} = 0.71 \cdot 10^{-10}$ m^2; conversion efficiency of the scintillator CsI(Tl) $C_{ef} = 0.06$ [11]. Normalized spectral dependence of scintillator radiation CsI(Tl) $\varphi_{nS}(\lambda)$ is given in [11]. Reference data on the energy dependence of the mass coefficients of weakening for W (X-ray tube anode material), Be (material of the window of the X-ray tube), Fe (material of the inspected object) and mass coefficients of true absorption for Cs and I (elements of which an X-ray monocrystalline screen material consists) was taken from [12]. The mass coefficient of true absorption CsI(Tl) was calculated using the formula for complex compounds by adding the products of the mass fractions of each element to its mass coefficient of true absorption. Linear coefficients of attenuation and true absorption $\mu_m(E)$, $\mu_0(E)$, $\mu_1(E)$ and $\mu_S(E)$ were obtained by multiplying the mass coefficient and density for the respective materials. Factor of accumulation $B(E_{\max})$, which is taken into account when calculating the background signal, according to [7] defined by the formula: $B(H_1, E_{\max}) = 1 + 0.013(E_{\max} H_1)^{0.8}$. The thickness of the object for the case of wired penetrameters equals to the controlled sample thickness: $H_1 = H_{kz}$, and the size of the flaw - a penetrameter's wire diameter: $\Delta H_1 = d$. For the case of a groove penetrameter thickness of the object is equal to the sum of controlled sample and penetrameter thickness: $H_1 = H_{kz} + h_e$; the size of the flaw – is the depth of the groove, the value of which is taken with the opposite sign: $\Delta H_1 = -h_k$.

The area of flaw reduced to the surface of CCD-array $\Omega = d^2 / \delta^2$, where d – diameter of the wired penetrameter. Duration of one accumulated frame $T_{EX} = 40$ ms. Integration time using visual analyser $T_i = 0{,}1$ s, threshold contrast $k_v = 2\%$, threshold signal to noise ratio $\psi_{por} = 5$ [13, 14].

IV. RESULTS AND DISCUSSION

Fig. 1 - 2 present obtained, using this model, dependences of threshold contrast and contrast of the output signal of the matrix on the defect size for accelerating voltage 150 kV.

The thickness scale is marked with dimensions corresponding penetrameters GOST 7512-82 [6]. Value of the relative sensitivity is calculated by the smallest groove depths of the groove penetrameter or by the smallest wire diameter of the wired penetrameter, for which the criterion (1) (i.e., the nearest from the side of the larger values of depth or diameter, relative to the reference point of intersection of the corresponding curves in Fig. 1 and Fig. 2).

Table 1 show the values of the relative sensitivity of XRTS control by the groove K_k and wired K_d penetrameters, which was obtained experimentally. Calculated from the model and obtained from the graphs values of the relative sensitivity of control are also given here. Accelerating voltage of the X-ray tube (U_A), the thickness of the controlled steel object H_{kz} and number of accumulated frames of CCD camera are shown in the table. The numbers of the lowest groove depth and the smallest wire diameter, which still can be revealed by XRTS and designation of the appropriate types of penetrameters (Fe-1 or Fe-3) are also given here.

Comparison of calculated and experimental values of the sensitivity shows satisfactory agreement of values obtained using the proposed model, with experimental data. The diagrams show that increasing the number of accumulation frames improves sensitivity. In addition, the proposed model is universal and can be used for calculation of the relative sensitivity of systems with the impulse devices of other models, with other options and features.

TABLE I. EXPERIMENTAL AND CALCULATED VALUES OF THE RELATIVE CONTROL SENSITIVITY OF XRTS

№	U_A, kV	H_{kz}, mm	N_k	Groove penetrameter			Wired penetrameter		
				№ of the groove and type of the penetrameter	K_k, % exp.	K_k, % calc.	№ of the wire and type of the penetrameter	K_d, % exp.	K_d, % cacl.
1	150	10	375	4Fe-1	2,5	1,7	6Fe-3	4	3,2
2	150	10	450	4Fe-1(5 Fe-1)	2,5(1,7)	1,7	6Fe-3	4	3,2
3	150	14	450	4 Fe-1	1,8	1,8	6Fe-3	2,9	2,3

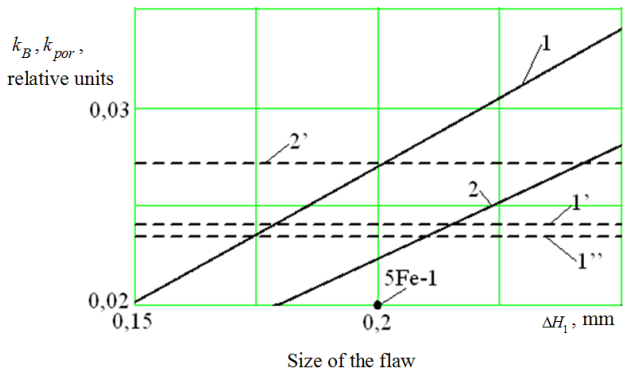


Fig. 1. Dependencies of CCD output contrast k_B (solid line) and threshold contrast k_{por} (dashed line) on the size of the defect ΔH_1 for the groove penerameter Fe-1 at a different duration of accumulation of frames: curve 1 – k_B with $H_{kz}=10$ mm; curve 2 – k_B with $H_{kz}=14$ mm. Dashed curves marked with strokes, indicate k_{por} that corresponds to k_B with the same number 1' – $N_k = 375$, 1" – $N_k = 450$, 2' – $N_k = 450$.

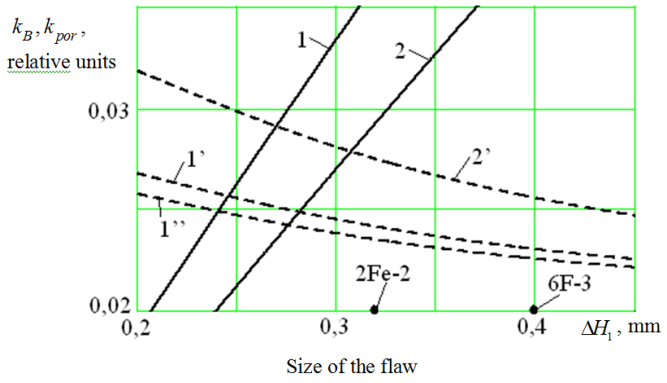


Fig. 2. Dependencies of CCD output contrast k_B (solid line) and threshold contrast k_{por} (dashed line) on the size of the defect ΔH_1 for the wired penerameters Fe-2, Fe-3 at a different duration of accumulation of frames: curve 1 – k_B with $H_{kz}=10$ mm curve 2 – k_B with $H_{kz}=14$ mm. Dashed curves marked with strokes, indicate k_{por} that corresponds to k_B with the same number 1' – $N_k = 375$; 1" – $N_k = 450$; 2' – $N_k = 450$.

V. CONCLUSIONS

Using the variable duration of the accumulation of images in CCD cameras can eliminate low-frequency flicker when using X-ray impulse device with small overall sizes for the control of objects in the "field" and remote places.

The proposed mathematical model takes into account all the main stages of the signal conversion in X-ray systems, and allows the calculation of the relative sensitivity of RTVS with the impulse X-ray devices when using the regulated duration of accumulation. Calculated values of relative sensitivity using the proposed model are in a good agreement with the experimental data.

X-ray television systems based on a CCD cameras, operated in the regime of regulated duration of accumulation of the images and impulse X-ray machines with the X-ray tube IMA2-150D, provide sensitivity of control of steel objects with thickness up to 14 mm equal to 1.8% and 2.9%, using groove, and wired penetrameter respectively.

Increasing the number of accumulated frames in CCD cameras of impulse XRTS allows to achieve sensitivity control that is not inferior to the sensitivity of XRTS based on continuous operation X-ray device.

REFERENCES

[1] Belyi N.G., Denbnovetsky S.V., Leshchishin A.V., Mikhailov S.R., Slobodian N.V., Troitsky V.A. Relative sensitivity of X-ray TV systems based on highly-sensitive CCD-cameras and X-ray monocrystalline screens // Technical diagnostics and non-destructive testing. – 2007. - № 2. – pp. 34 – 40 (in Russian).

[2] Troitsky V.A, Belyi N.G, Buhensky V.N., Leschishin A.V., Mikhailov S.R., Slobodian N.V. Highly sensitive X-ray television systems based on X-ray screens and a CCD cameras with the accumulation of images // Technical diagnostics and non-destructive testing. – 2009. – № 3. – pp.41 – 46 (in Russian).

[3] Vavilov S.P. Impul'snaya rentgenovskaya tekhnika (Impulse X-Ray Equipment), Moscow, Energoatomizdat 1981. – p. 80. (in Russian)

[4] Komyak N.I., Peliks E.A. Portable impulse X-ray soursces of the series MIRA // Electronic engineering. Series 4. Electro-vacuum and gas discharge devices. – 1979. – pp. 9 – 11 (in Russian).

[5] Diagonal 8 mm (Type 1/2) CCD Image Sensor for CCIR B/W Video Cameras ICX429ALL. // www.sony.net.

[6] State Standard of the USSR. Nondestructive testing. Welded joints. Radiographic method. GOST 7512-82. – Moscow, Publishing House of Standards, 1983. – p. 30 (in Russian).

[7] Mikhailov S.R. Modeling of shadow X-ray image of object controlled in fluoroscopic systems of non-destructive testing // Electronics and communication. – 2002. - № 16. – p. 59 – 70 (in Russian).

[8] Slobodian N.V. Features of modeling the characteristics of impulse X-ray devices // Electronics and communication. - 2004 - № 21. - pp. 80 – 82 (in Russian).

[9] Slobodian N.V. Simulation and comparison of the characteristics of impulse X-ray devices of various types // Electronics and communication. - 2006 - "Problems of Electronics", part 2 - p. 33 – 36 (in Russian).

[10] Denbnovetsky S.V., Slobodyan N.V. Simulation of radiation characteristics of pulse X-ray devices for non-destructive testing the semiconductor materials // Semiconductor Physics, Quantum Electronics and Optoelectronics. – 2006. – Vol. 9. – № 1. – pp. 68 – 72 (in Russian).

[11] Abramov A.I., Kazanski Y.A, Matusevich E.S. Fundamentals of the experimental methods of nuclear physics, Moscow, Energoatomizdat, 1985. – p. 488 (in Russian).

[12] Storm E., Israel H. Gamma-radiation interaction cross sections (for energies of 0.001–100 MeV and the elements from 1 to 100) [Russian translation], Moscow, Atomizdat, 1973. – p. 256 (in Russian).

[13] Gurvich A.M. Physical basis of X-ray monitoring and diagnostics. – Moscow, Energoatomizdat, 1989. – p. 168 (in Russian).

[14] Rose A. Human sight and electronic sight / Translated from English.- Moscow, Mir, 1977. – p. 216 (in Russian).

Novel Self-folding Electrode for Neural Stimulation and Recording

Melinda Varga, Marco Luniak, Klaus-Jürgen Wolter
Department of Electrical Engineering and Information Technology
Technische Universität Dresden
Dresden, Germany
varga@avt.et.tu-dresden.de

***Abstract*—The paper presents a novel self-folding, bipolar electrode structure for continuous evaluation and identification of neural dysfunction. The self-folding design realized with a careful choice of biocompatible materials, enhances the easy of implantation and positioning of the electrode during the medical procedure. In this regard, the proposed electrode offers a promising alternative for applications in neural electrical stimulation and recording.**

Keywords— functional electrical stimulation; electrode; self-folding, sterilization

I. INTRODUCTION

Functional electrical stimulation (FES) of nerves is currently one of the most important medical technologies used to control and treat numerous physiological conditions e.g. rehabilitation after stroke, spinal cord injuries, epilepsy, migraine, vestibular dysfunction, occipital neuralgia and cerebral palsy [1-7]. Its development and implementation largely depend on successful engineering of implantable electrodes and on effective transmission of power and signals to and from the electrodes to other in vivo or external devices.

Medical electrodes dedicated to functional neurostimulation need to ensure neural stimulation accuracy and on the same time, functional biocompatibility. They have to operate the injection of large amount of electrical charges through a reversible faradaic process. The electrode material must provide charge transfer under low interface impedance in order to avoid tissue trauma. Furthermore, the mechanical design i.e. geometry of the electrode structures has to ensure the preservation of the surrounding tissues by avoiding damage due to material rigidity. Implantable electrode configurations developed so far include e.g. the tubular and the cuff designs [8, 9]. Although stimulating efficiently and selectively, cuffs for example, require intraoperative procedures and may damage the nerves upon enclosure. An attractive alternative to the already available geometries would constitute a self-folding design that would allow an easy and straightforward application around the nerve upon exposure to specific stimuli.

Self-folding devices and materials for biomedical applications are extensively described in literature. For instance, self-folding bilayers as drug delivery platform in the digestive tract, self-folding biodegradable polymer capsules capable of controlled capture and release of yeast cell as well as reconfigurable, self-folding microgrippers serving as miniaturized surgical tools were reported [10-12]. The forces that drive the self-folding process can be varied and result from material heterogeneities that are either stimulated or engineered *in situ* [13]. In this regard, one common approach is based on thin (polymer) bilayers that are able to roll due to the internal stress produced as a result of thermal expansion, lattice mismatch or swelling (Fig.1).

Here we utilize this approach to fabricate a self-folding, bipolar electrode for neural stimulation and recording. The electrode consists of crosslinked gelatin, a polypeptide commonly used as gelling agent in food, and a paper bilayer. This is a combination, in which the gelatin acts as the active polymer that changes its volume (swells) upon immersion in warm water. The swelling is restricted by the paper layer and as a result, the bilayer does not uniformly expand, but folds.

Once in contact, the electrode is stable over a typical time scale of an operation and ensures applicability in electrical recording. The design allows an easy removal of the electrode after the surgery without any sharp cutting tool. Force tests performed indicated the efficacy of the removal. Ethylene oxide gas sterilization, the most widely used method for sterilizing medical equipment, did not impair the electrode properties and functionalities. In this regard, the proposed electrode offers a promising alternative for applications in neural electrical stimulation and recording.

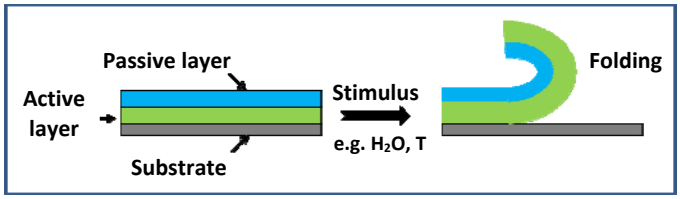


Fig. 1 Folding of a patterned polymer film caused by the shape change of the active polymer upon a certain stimulus.

This work was financially supported by AiF (KF2381604FRO)

II. EXPERIMENTAL SECTION

A. Fabrication of the electrodes

In order to fabricate the electrodes, at first the paper substrates (Pergamin type) were cut into pieces of 10x8 cm. On one side of the paper carbon paste (SD 2843 HAL, black, matt ordered from Lackwerke Peters GmbH + Co KG) as the conductive layer of the electrode structure was screen printed (MPM SPM screen printer, Precision Placement Machines Inc.) in the form of stripes having 50 mm in length and 2 mm in width with another 2 mm space between the stripes. After printing, the paste was dried at 150°C for 45 minutes.

On the other side of the paper, gelatin was screen printed. Gelatin with a relatively high Bloom index (meaning high renaturation level) was chosen for this study. Type B gelatin from bovine skin, bloom 225 (Sigma-Aldrich Ltd, G9382) was dispersed in distilled water and heated to 45°C under continuous steering at 500 rpm (IKA RCT Classic) for 30 minutes and the gelatin solution poured into 55 mm diameter polystirene Petri dishes (WVR). For stabilization of the gels, glutaraldehyde (Sigma-Aldrich, approved for biomedical purposes) dissolved in PBS pH=7.4 (Sigma Aldrich) was added and the solution quickly cooled down at 4 °C for 10 minutes allowing gelification. The gelatin layer was prepared by screen printing with a ceramic mask on a paper substrate. The optimal thickness of the gelatin layer was set for 1200 µm and attained in two subsequent screen printing steps (2x600 µm) with a drying step at 43°C in between. After screen printing the gelatin layer, the electrode structures were cut into single pieces of (8x50) mm^2 (containing two carbon stripes each) and characterized.

B. Characterization of the electrodes

For testing the folding ability in warm water, petri dishes were filled with distilled water and heated to 37°C. The electrodes were placed into the dishes and folding time recorded. Folding was also tested in cold distilled water and in artificial body fluid consisting of NaCl 124 mmol/L, KCl 3 mmol/L, Na_2HPO_4 1.24 mmol/L, $MgSO_4$ 1.3 mmol/L, $CaCl_2$ 2 mmol/L, $NaHCO_3$ 26 mmol/L and NaH_2PO_4 10 mmol/L.

The developed bipolar electrodes were subjected to mechanical i.e. force tests by introducing a hook of a force gauge (using DAGE BT 2400, DAGE Semiconductor GmbH) into the revelation of the electrode structure and applying a pulling force until complete unfolding i.e. straightening of the electrode.

The electrical resistance was measured with the four point method using the Grundig Electronics RLC 100 device.

The electrodes were packaged in gas permeable packaging and submitted to ethylene oxide gas sterilization at the University Hospital Carl Gustav Carus Dresden.

III. RESULTS AND DISCUSSION

Neural electrodes are designed to interface with the nervous system and provide control signals for neural prostheses. They interface with the central and peripheral nervous system by transducing ionic biological signals to electronic signals.

The primary requirements of stimulating electrodes include biocompatibility of the electrode material, good contact with the nerve, effective stimulation for a specific time period, no damage of the neural tissue, precise and accurate positioning on the nerve and possibly easy handling. Conventional metal e.g. platinum, gold and iridium electrodes used to date cause very frequently reactive tissue responses associated with a progressive increase in the impedance at the electrode/tissue interface [14, 15].

In this work, we developed a bipolar electrode based on new materials i.e. gelatin and paper. More importantly, these two materials enable the positioning of the electrode by a chemomechanical actuation i.e. self-folding over the nerve without the need for any wires or tethers.

Gelatin is a protein obtained by the partial hydrolysis of native collagen found primarily in connective tissue such as skin, bone and tendons. Collagen is comprised of linear, fiber-like structures called also fibrils that are crosslinked; treatment with dilute acid or alkali results in the partial cleavage of these crosslinks, giving rise to the “warm water soluble collagen” i.e. gelatin. Gelatin is commonly used as gelling agent in food, photography, pharmaceuticals and cosmetic manufacturing.

Gelatin, being a mixture of polymer chains of different length, forms colloidal solutions or sols with water. Upon cooling, these sols convert to gels and on warming (> 40°C) they revert to sols. This property of gelatin in combination with paper was used to fabricate the self-rolled electrode structure. The approach is based on thin bilayer configuration where gelatin functions as the active layer and changes its volume upon immersion in aqueous solution. The swelling is restricted by the inactive i.e. paper layer and as a result, the bilayer does not uniformly expand, but folds (Fig. 2). Rolling occurs due to the internal stress produced as a result of swelling. The stress in the film will be proportional to the swelling degree [16]. The higher the crosslinker amount, the less will swell. Therefore, we found that 1 ml 0.5% glutaraldehyde for a concentration of 30 mg/ml gelatin is optimal for folding of the electrode structure. Gelatin needs to be crosslinked in order to avoid loss of integrity upon longer exposure to water or body fluids. Although aldehydes are known to by cytotoxic, their use as crosslinking agents has been approved upon implantation evaluations [17].

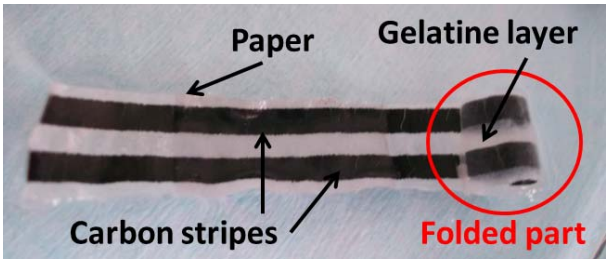


Fig. 2 The proposed bipolar electrode structure consisting of a paper and gelatin bilayer. The conductive track is made of carbon which was printed on top of the paper, while gelatin on the back side. Upon immersion in water, the electrode folds up with the carbon stripes pointing inwards.

Upon immersion in warm water (37°C), the electrode folds up within 1.5 minute making one single revelation with a diameter of 3 mm (Fig. 3) This hints to a strong and immediate swelling of the gelatin. Folding tests conducted in cold water showed faster rolling. Folding was even faster in the artificial body fluid, the electrode rolled up in just 1 minute. For assessment of nerve integrity during surgeries, a fast positioning by self-folding the electrode around the nerve i.e. within 1 minute will certainly help the surgeon in performing effectively and timely. Zakharchenko et al. fabricated self-rolled tubes based on polysuccinimide (PSI)/polycaprolactone (PCL) bilayers and reported a step-like character of the PSI film [18]. Rolling started only after 9 hours of incubation in buffer, phenomenon that in their case was ideal for encapsulating living yeast cells, since cells need a certain time for adjusting to the hosting conditions. However, in our case there is a need for rapid and efficient electrode placement during the medical procedure without any damage. Evidently, the rolling time has to fulfill the application requirements.

Rolling occurs on the short side and only in one direction with the carbon stripes pointing inwards. This is advantageous and desired since upon implementation, the electrode has to contact the nerve with the conductive track. The length of the conductor path was fixed to 20 mm and adjusted by a silicone layer, a polymer extensively used in applications that require high biocompatibility. On the remaining part outside the silicone layer, the electronic component generating the electronic pulses will be mounted, therefore it contains no gelatin. As a consequence, this part will remain unfolded to serve as a carrier.

The rolled part is quite stable, even after 16 hours, the electrode is still in the folded state and no delaminations of the layers (neither gelatin, nor carbon) have been observed.

The self-folding design presented here allows an easy removal of the electrode after surgery without any sharp cutting tool. It can be simply unrolled from the nerve by pulling gently from the unfolded side and since everything is in a humid atmosphere where the electrode due to the material combination gets softened, we suppose no damage of the nerve.

Fig. 3 The electrode in its folded state

In order to test the unfolding ability of the electrode and on the same time the efficacy of the removal, we performed mechanical pull tests. The force tests involved introduction of a hook of a force gauge into the electrode revelation and pulling until complete straightening. We measured a pulling force of 0.032 N, which is a relatively small force that does not imply a mechanical trauma to the neural tissue. For instance, in case of a nerve fiber of 2 mm in diameter and an electrode width of 8 mm, a pressure of 0.015 N/mm^2 should be applied to unfold it. Studies have shown that for rupturing the nerve, a pressure of 0.14 N/mm^2 is necessary which is much more than that necessary for unfolding our electrode [19].

For electrical characterization of the electrode, the electrical resistance was measured. As a result, we obtained 362 Ω, which is in accordance with our expectations.

Next, we subjected the electrode to steam sterilization. Afterwards the electrode was gradually cooled down to room temperature. No morphological changes have been detected on the electrode after the procedure only an increase of the resistance. The resistance increased from 362 Ω to 451 Ω. This value is still acceptable, however this sterilization procedure is known to damage electronics. Therefore we tested chemical sterilization with ethylene oxide gas. Ethylene oxide is widely used for sterilizing medical instruments because of its low temperature, low cost and effectiveness of neutralizing bacteria, spores and viruses. After sterilization, the electrodes appeared identical to unsterilized samples and showed no morphological changes only (as in the previous sterilization method) slight increase of the resistance, but this does not influence the functionalities of the electrode. In conclusion, EtO is an effective method for sterilizing the gelatin/paper/carbon electrode.

In comparison to the cuff design, the proposed self-folding electrode design assures a more secure mechanical enclosure and contacting of neurons. Not only the new design but also the material combination is innovative in this work. Carbon instead of metals is particularly interesting mainly because of its good conductivity and biocompatibility. Carbon in the form of graphene-based microelectrode has been used for recording action potential signals of lateral giant nerve fiber of the escape circuit of an American crayfish [20]. The study indicated a satisfactory biocompatibility and non-cytotoxicity. Consequently, we suppose our carbon paste behaving in a similar manner and plan toxicity tests according to DIN EU ISO 10993 regulations. Furthermore, currently there is a growing interest in using paper as platform for low-cost sensing devices for biochemical analysis, medical diagnostics and bio-detection [21-23]. Paper is manufactured from renewable resources on industrial scale, it is inexpensive, its surface can be easily manipulated trough coating, printing or impregnation, can be easily stored or disposed and is also suitable for biotechnological applications because cellulose is compatible with biological samples. To this end, using paper as electrode material for neural stimulation and recording might constitute an attractive alternative to the existing ones.
Regarding all these aspects, we believe that the proposed electrode represents a step toward development of more “bio-friendly”, minimally invasive, autonomous medical tools. Finally the principle used in the design of the electrode

highlight the benefit of thin film actuation and self-assembly in engineering medical systems and devices.

ACKNOWLEDGMENT

The authors thank Dr. Uwe Pliquett, Dr. Andreas Langer, Laura Diaz, Armin Taschwer and Daniel Rossbach for fruitful discussions regarding the electrode design.

M.V. thanks Frank Schumann for assistance with the force measurements and Sabine Gärn for sterilizing the electrodes.

REFERENCES

[1] Y. Hara, "Neurohabilitation with new functional electrical stimulation for hemiparetic upper extremity in stroke patients", J. Nippon Med. Sch., vol. 75, pp. 4-14, Februar 2008.

[2] M. R. Popovic, A. Curt, T. Keller, and V. Dietz, "Functional electrical stimulation for grasping and walking: indications and limitations", Spinal cord, vol. 39, pp. 403-412, August 2001.

[3] K. Vonck, V. de Herdt, M. Sprengers, and E. Ben-Menachem, Neurostimulation for epilepsy, in Handbook of Clinical Neurology, vol. 108 H. Stefan and W.H. Theodore, Eds. Elsevier B.V., 2012, pp. 955-970.

[4] C. A. Popeney, and K. M. Aló, "Peripheral Neurostimulation for the Treatment of Chronic, Disabling Transformed Migraine", Headache, vol. 43, pp. 369-375, April 2003.

[5] S.C. Della, A. Migliaccio, and A. Patel, "Electrical stimulation to restore vestibular function development of a 3-d vestibular prosthesis", Conference Proceedings IEEE Engineering in Medicine and Biology Society vol. 7, pp. 7380–7385, September 2005.

[6] R. L. Weiner, and K. L. reed, "Peripheral neurostimulation for control of intractable occipital neuralgia", Neuromodulation, vol 2. pp. 217-221, July 1999.

[7] S. Pierce, M. N. Orlin, R. T. Lauer, T. E. Johnston, B. T. Smith, and J. J. McCarthy, "Comparison of percutaneous and surface functional electrical stimulation during gait in a child with hemiplegic cerebral palsy", Am. J. Phys. Med. Rehabil., vol. 83, pp. 798-805, October 2004.

[8] W. Lamade, U. Meyding-Lamade, E. Hund, N. Senninger, and C. Herfarhrt, " Transtracheal monitoring of the reccurent laryngeal nerve. Prototype of a new endotracheal tube", Chirurg, vol. 68, pp. 193-195, February 1997.

[9] N. S. Korivi and P. K. Ajmera, "Self-closing cuff electrode for functional neural stimulation and recording", J. of Med. And Biol. Eng. vol. 31, pp. 353-357, August 2010.

[10] H. He, J. Guan, J. L. Lee, "An oral delivery device based on self-folding hydrogels", Journal of Controlled Release, vol. 110, pp. 339-346, January 2006.

[11] G. Stoychev, N. Puretskiy and L. Ionov, "Self-folding all-polymer thermoresponsive microcapsules", Soft Matter, vol.7, pp. 3277-3279, February 2011.

[12] K. E. Laflin, C. J. Morris, N. Bassik, M. Jamal, and D. H. Gracias, "Theteherless microgrippers with transponder tags", Journal of Microelectromechanical Systems, vol. 20, pp. 505-511, April 2011.

[13] T.G. Leong, A. M. Zarafshar, D. H. Gracias, "Three-dimensional fabrication at small size scales", Small, vol. 6, pp. 792-806, April 2010.

[14] R. B. Beard, B. N. Hung, and R. Schmukler, "Biocompatibility considerations at stimulating electrode interfaces", Ann. Biomed. Eng., vol. 20, pp. 395-410, 1992.

[15] V. S. Polikov, P. A. Tresco, and W. M. Reichert, "Response of brain tissue to cronically implanted neural electrodes", J. Neurosci. Methods, vol 148, pp. 1-18, October 2005.

[16] S. Timoschenko, "Analysis of bi-metal thermostats", J. Opt. Soc. Am., vol. 11, pp. 233-255, 1925.

[17] F.H. Lin, C.H. Yao, J.S. Sun, H.C. Liu, C.W.Huang, "Biological effects and cytotoxicity of the composite composed of tricalciumphosphate and glutaraldehyde cross-linked gelatin" Biomaterials, vol. 19, pp. 905-917, May 1998.

[18] S. Zakharchenko, E. Sperling, and L. Ionov, "Fully biodegradable self-rolled polymer tubes: a candidate for tissue engineering scaffolds", Biomacromolecules, vol. 12, pp. 2211-2215, April 2011.

[19] E. Marani, J. L. van Leeuwen, and C. W. Spoor, "The tensile testing machine applied in the study of human nerve rupture: a preliminary study", Clin Neurol Neurosurg. Vol. 95, pp. S33-35, 1993.

[20] C.H. Chen, C.T. Lin, J. J. Chen, W.L. Hsu, Y.C. Chang, S.R. Yeh, L. J. Li, and D.J. Yao, "A graphene-based microelectrode for recording neural signals", 16th International Solid-State Sensors, Actuators and Microsystems Conference, pp. 1883-1886, June 2011.

[21] A. W. Martinez, S. T. Philipps, M. J. Butte, and G. M. Whitesides, "Patterned paper as a platform for inexpensive, low-volume, portable bioassays", Angewandte Chemie-International Edition, vol. 46, pp. 1318-1320, February 2007.

[22] R.Pelton,"Bioactive paper provides a low-cost platform for diagnostics", Trac-trends in Analytical Chemistry, vol. 28, pp. 925-942, September 2009.

[23] A. W. Martinez, S. T. Philipps and G. M. Whitesides, "Diagnostics for the developing world: Microfluidic paper-based analytical devices", Analytical Chemistry, vol. 82, pp. 3-10, January 2010.

Magnetic Nanotherapeutics of Guerin Carcinoma

Orel V.E.[1,2], Nikolov N.A.[1,2], Rykhalskiy A.Y.[1,2], Dzyatkovskya N.N.[1], Romanov A.V.[1], Yaroshenko O.Yu.[2], Supruniuk D.A.[2], Orel V.B.[3], Khamrovskiy M. M.[4], Penievskyi M.I.[2], Radaguz A.V.[2]

[1]Medical Physics & Bioengineering Department, National Cancer Institute, MH Ukraine, Kyiv, Ukraine
[2]National Technical University of Ukraine 'Kyiv Polytechnic Institute', Kyiv, Ukraine
[3]Lyceum of Information Technologies N.79, Kyiv, Ukraine
[4]Specialized school № 304, Kyiv, Ukraine

***Abstract* — The research of animals with Guerin carcinoma has shown, that magneto-mechano-chemically synthesized magnetic nanocomplex on the basis of Fe_3O_4 nanoparticles with antitumor anthracycline row antibiotic doxorubicin and next local electromagnetic irradiation by spatially inhomogeneous constant magnetic and variable electromagnetic fields had a greater antitumor effect than conventional doxorubicin and magneto-mechano-chemically synthesized magnetic nanocomplex without electromagnetic irradiation.**

Keywords—electromagnetic field; magnetic-dipole applicator; magnetic nanocomplex; tumor

I. INTRODUCTION

Cancer is the second leading cause of death after heart diseases. Each year, on a worldwide basis, almost seven million people die of the consequences of cancer [1].

From molecular perspective and quantized level, cancer is not a specific disease. Cancer arises as a result of a succession of randomly occurring mutations. Tumors are inherently molecularly diverse. This complexity might give the wrong impression that cancer is not accessible to physics and engineering disciplines, which strives to describe nature by precise quantitative laws. Nevertheless, statistical physics has proven to be able to find the laws behind the stochastic process underlying thermodynamics, and nonlinear dynamics has even uncovered the principles that govern chaotic behavior in nature [2].

One of approaches for treatment of malignant tumor with the use of thermodynamics basic principles, and nonlinear dynamics is radio-frequency local hyperthermia with combination of magnetic nanotherapeutics. Magnetic nanotherapeutics is based on magnetic nanotechnology and utilized nanoparticles loaded with anticancer drugs and guided towards the tumor site to selectively destroy it in conjunction with electromagnetic excitation [3].

Experimental and clinical studies conducted in the last 10 years testify to the potential use of magneto-sensitive nanocomplex nanoparticles of iron oxide (magnetite) Fe_3O_4 and antitumor anthracycline row antibiotic doxorubicin (DOX) in quality antitumor drugs. The conversion of electromagnetic energy into heat by nanoparticles has the potential to be a powerful, non-invasive technique for biotechnology applications such as drug release, disease treatment and remote control of single cell functions. The attractiveness of this approach consists in enabling the controlled delivery of the magneto-sensitive complex by the use of permanent magnetic fields and subsequently strengthening its antitumor activity due to radiofrequency induced hyperthermia [4].

It's known as magnetic drug targeting, the method is based on the application of magnetic field to guide iron particles loaded with a therapeutic agent to a tumor and hold them there. In order to expose the particles to the highest possible traction force, magnetic drug targeting uses magnets with inhomogeneous fields. The goal of magnetic drug targeting is to concentrate the active ingredient specifically in the tumor region while at the same time minimizing the side effects of chemotherapy. In local hyperthermia, magnetic nanoparticles are heated by an external magnetic field, which weakens the tumor and enhances the cytotoxic effect of the chemotherapeutic drugs. In contrast to magnetic drug targeting, in which an inhomogeneous magnetic field is used, a homogeneous alternating field is needed for hyperthermia. This is the only way to induce motion in the nanoparticles containing iron, which ultimately generate heat [5].

Electromagnetic fields activate important biological processes in cells, including biosynthesis of stress proteins. Since the reactions occur at very low field strengths, the initiating forces are very weak. Electromagnetic fields interact with moving charges, and electrons may be especially influenced and cause changes in the biochemical reaction rate. The moving charge interaction model suggests that stress protein synthesis is initiated by interaction of electromagnetic fields with moving electrons in DNA. Electromagnetic fields accelerate electron transfer in cytochrome oxidase, and calculated speed of the moving charge in the Na, K-ATPase [6].

In nonuniform (asymmetric) electric field within tumor cells induces forces pushing all dipoles toward the furrow. The basis of these effects during cytokinesis was shown to be unidirectional forces induced by inhomogeneous fields at the bridge separating the daughter cells that interfere with spindle tubulin orientation and induce dielectrophoresis [7].

However there are problems during thermotherapy cancer patient using magnetic nanoparticles and electromagnetic field: difficult to homogenously distribute particles intratumorally by interstitial application; limited magnetic field strength applicable (< 4 kA/m, 100–300 kHz) for abdominal tumors;

high thermal gradients; direct temperature measurements are required initially; locoregional therapy, not suitable for disseminated tumors; target size limited to approximately 100 cm^3 [8].

To solve abovementioned problem for development molecular technology synthesis and tumor magnetic nanotherapeutics we used two well-known effects: disorder in the kinetics of free radical reactions are one of the potential causes cancer, that is, electronic problems at the molecular level for human [9]; magnetic properties of free radicals (the concentration of paramagnetic centers) influenced by mechanochemical effect and electromagnetic radiation in inhomogeneous field [10, 11]. To implement mentioned effects we produced magnetic-dipole applicator.

We believe that spatial orientation of magnetic-dipole applicator elements with constant magnetic field and alternating current flowing through them can be varied to produce asymmetric electromagnetic field patterns resulting in development of asymmetric eddy current patterns orthogonal to the magnetic lines of force. These asymmetric magnetic field patterns could be further focused by a tumor with nanocomplex of magnetic nanoparticles and antitumor drug. Field and eddy current patterns will subsequently correlate with three-dimensional heating patterns in heterogeneous patterns of malignant tumor. We propose that this approach to a problem of cancer treatment will increase antitumor effect for magnetic nanotherapeutics.

The aim of the paper is the research and bioengineering analysis of antitumor effects of magnetic nanotherapeutics initiated by nonuniform constant magnetic and electromagnetic fields on experimental tumor models of Guerin carcinoma.

II. Electromagnetic Irradiation by Magnetic-dipole Applicator

The present equipment relates to the general technical field of producing eddy currents excited by electromagnetic waves [12]. The eddy currents are obtained when the frequency of gyration of an electron in quasi-static magnetic field is equal to the frequency of the applied accelerating electromagnetic field. In Fig. 1 you can see the diagrammatic elevation view showing summation eddy currents production in tumor in accordance to the principle of magnetic-dipole applicator. This equipment allows implementing the method of amplified producing ring current in a tumor with magnetic nanocomplex.

The principal elements of magnetic-dipole applicator are shown on Fig. 2. To generate variable electromagnetic field producing eddy currents in tumor with magnetic nanocomplex (MNC) the first prototype of the device for medical treatment called “Magnetotherm” (Radmir, Ukraine) was used [13]. The frequency of electromagnetic irradiation (EI) was 70 MHz with an initial power of 75 W. Spatially inhomogeneous structure of electromagnetic field variables was printed in the paper [14]. For localization of magnetostatic field in the tumor we have used neodymium magnet (NdFeB). Intensity of the permanent magnetic field made on the end of magnetic flux dipoles was 30 mT.

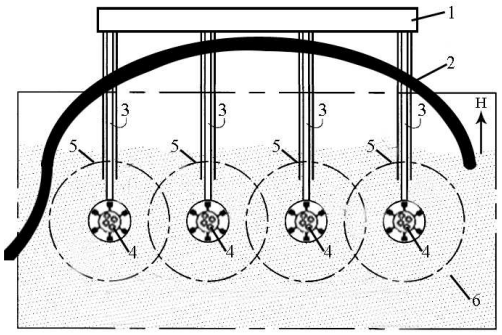


Figure 1. The diagrammatic elevation view showing eddy currents production in tumor with Fe_3O_4 nanoparticles in accordance to the principle of magnetic-dipole applicator: 1 – permanent magnet; 2 – curved loop; 3 – magnetic flux dipoles; 4 – magnetic nanocomplex; 5 – eddy currents; 6 – tumor.

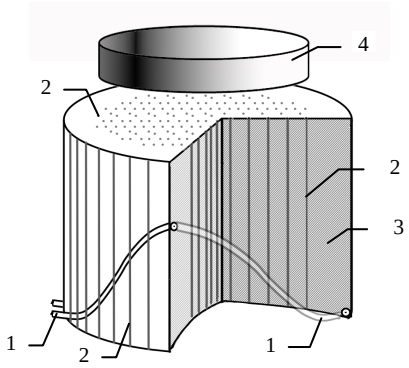


Figure 2. Magnetic-dipole applicator: 1 – curved loop; 2 – magnetic flux dipoles (diameter 0.5 mm, magnet pole separation 3 mm); 3 – dielectric; 4 – permanent magnet.

III. Magneto-mechano-chemical Synthesis of Magnetic Nanocomplex

Magneto-mechano-chemical technology is a new method for the synthesis of MNC; the method is based on the integration of two known methods: mechano-chemical synthesis [15] and synthesis of enzymes in a micro-reactor with magnetic elements [16]. The principle of the mechano-chemical synthesis is based on the ensemble fragmentation and formation of paramagnetic centers, such as free radicals in MNC structure. The increase in the concentration of free radicals of the MNC is suggestive for the potential increase of the drug antitumor effect under the influence of external radio frequency energy [17].

Magnetite Fe_3O_4 is a material with a high saturation magnetic moment m_s in a wide range of temperatures. When magnetite nanoparticles are used in the synthesis of MNC under the influence of external magnetic irradiation, strong local magnetic fields near the particle-components and other nanocomposites will appear. Effects such as, magnetic (or nuclear-spin) isotope effect, chemically induced magnetic polarization of electrons and nuclei, radiofrequency chemical maser, high-frequency magnetic-resonance modulation of the rates of physico-chemical processes involving paramagnetic

particles and magnetic effects in molecular physics, are all consequences of spin selection rules and spin evolution [18].

MNC consisted of the nanoparticles Fe_3O_4 (Sigma) with diameters < 50 nm, and DOX (Pfizer, Italy) were processed in high-precision magneto-mechano-reactor (NCI, Ukraine). Mechanical processing was performed by vertical vibrations of the chamber at a frequency of 36 Hz and an amplitude of 9 mm for 5 min using mechanical energy of 20 W/g. Simultaneously, electromagnetic irradiation at 42 MHz by an induction coil with 40W initial power and permanent magnets with magnetic field intensity 11 mT (in the centre) for 5 min, was applied. Mass concentration of DOX in MNC was 43%. MNC has saturation magnetic moment m_s= 30.3 emu/g and coercive force H_c= 45.7 Oe in an applied saturation magnetic field (H for m_s) H_{ms}=3000 Oe. The electron spin g-factor (dimensionless magnetic moment) =2.7 [19].

IV. Experimental Animals and Tumor Transplantation

In the study, 40 male rats weighing 100 ± 15g bred in the vivarium of National Cancer Institute were used. The transplantation of Guerin carcinoma was performed according to the established procedure. All animal procedures were carried out according to the rules of the regional ethic committee. Animals were housed in 7 groups: group 1 – control (no treatment); group 2 – treatment DOX; group 3 – treatment DOX + EI; group 4 – treatment MNC; group 5 – treatment MNC+ EI.

V. Treatment of Guerin Carcinoma

The animal tumors were irradiated locally (Fig. 3). Experimental animals were treated by DOX in the dose 1.5 mg/kg and MNC in the dose 3.5 mg/kg. The treatment was performed three times by DOX, EI from 9 to 13 days after tumor transplantation every 48 hours. Tumor size was measured by the caliper. Tumor volume before treatment was 0.82 ± 0.14 cm^3. Drugs were injected directly into the tumor. Preliminary research showed that 15 and 30 minutes of local EI on conventional Guerin carcinoma initiated practically identical strengthening of DOX antineoplastic activity. The temperature was reached up to 39.1°C after 15 min EI. Therefore, with aim of milder hyperthermic non-equilibrium effects at physiological temperatures the irradiation was being performed during 15 minutes at once after treatment by DOX or MNC.

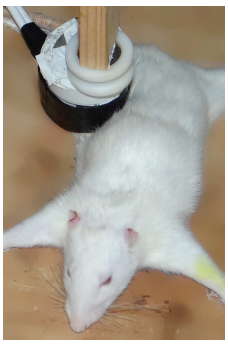


Figure 3. Electromagnetic irradiation of animal tumors.

VI. The Analysis of Nonlinear Kinetics of Tumor Volume

Nonlinear kinetics of tumor volume was evaluated by growth factor φ according to autocatalytic equation

$$\frac{dx}{dt} = \varphi(x + x_0)(1 - x), \qquad (1)$$

where $x = \frac{\Phi - \Phi_0}{\Phi_\infty - \Phi_0}$ is relative tumor growth by time t; $x_0 = \frac{\Phi_0}{\Phi_\infty - \Phi_0}$ is relative tumor volume at the moment of time $t = 0$; Φ_0 and Φ_∞ is initial and limiting tumor volume accordingly; Φ is tumor volume at the moment of time [20].

The solution of equation (1) is

$$\Phi = \Phi_0 + \Phi_0 \cdot \frac{e^{\varphi \frac{\Phi_\infty}{\Phi_\infty - \Phi_0} \cdot t} - 1}{1 + \frac{\Phi_0}{\Phi_\infty - \Phi_0} \cdot e^{\varphi \frac{\Phi_\infty}{\Phi_\infty - \Phi_0} \cdot t}}. \qquad (2)$$

The effect of EI and local inductive hyperthermia on nonlinear dynamics of the growth of animal tumors was evaluated with the braking ratio.

$$\kappa = \frac{\varphi_C}{\varphi_T}, \qquad (3)$$

where φ_C is growth factor for control group of animals, φ_T is growth factor for treated group.

VII. Statistical Analysis

Measured data were statistically processed with the Statistica 6.0 (StatSoft Inc.) software. For that purpose, parametric Student's t-test including preliminary verification of the hypothesis on the normal distribution law by Kolmogorov-Smirnov criterion was carried out [21].

VIII. Changes in Nonlinear Dynamic of the Growth for Guerin Carcinoma

The growth kinetics of animal tumors is shown in Table 1. The growth kinetics for the 2nd group had minimal response under the influence of DOX. At the same time a complete resorption was observed for the 5nd group (MNC + EI) which had maximal antitumor effect. The 3nd (DOX + EI) and 4nd (MNC) group of animals had an intermediate antitumor effect between the 2nd and the 5th group of animals.

Our research showed that antitumor effect of MNC was increased after treatment by MNC and EI. Antitumor effect of MNC didn't changed significantly after treatment by MNC without EI as compared to the effect of DOX.

TABLE I. THE GROWTH KINETICS OF GUERIN CARCINOMA

N	The Group of Animals	Growth Factor φ, Day^{-1}	Braking Ratio κ, Relative Units
1	Control (without treatment)	0.252 ± 0.019	1.00
2	Conventional DOX	0.201 ± 0.017	1.25
3	DOX + EI	0.183 ± 0.004*	1.38
4	MNC	0.185 ± 0.006*	1.36
5	MNC + EI	0.158 ± 0.007*+	1.60

*Statistically significant difference from control group, $p < 0.05$.

+Statistically significant difference from conventional DOX, $p < 0.05$.

IX. DISCUSSION

An external magnetic field can induce electron transitions in nanocomplexes composed of ferromagnetic magnetite and paramagnetic DOX. Since the reactivity of the magnetic particles depends on their spin state, it is possible to control the kinetics of free radical reactions by external magnetic field [22]. A possible mechanism can be described as follows. Under the influence of constant magnetic field the energy levels of iron atom, which is a part of magnetite, are splitting. An action of inhomogeneous alternating magnetic field which is generated by magnetic-dipole applicator initiates transition of iron ion in triplet state. Electron transfer at the heterojunction "iron oxide-DOX" from higher energy levels of the iron to aromatic rings of the DOX is possible by spintronic mechanism. Eddy currents originate in aromatic rings of DOX in external magnetic field (Fig. 4). The area of molecular currents in nanocomplex increases due to the effect of summation:

$$I = \oint \vec{M} \cdot \mathrm{d}\vec{l} = \int_S \vec{j} \cdot \vec{n}\, \mathrm{d}S\,, \qquad (4)$$

where I is the strength of molecular current in nanocomplex, $\vec{j}$ is bulk density of molecular current, l is a closed circuit, the surface S is spanned by this circuit, the direction of $\mathrm{d}\vec{l}$ and $\vec{n}$ vectors are coordinated, $\vec{M}$ is magnetization vector.

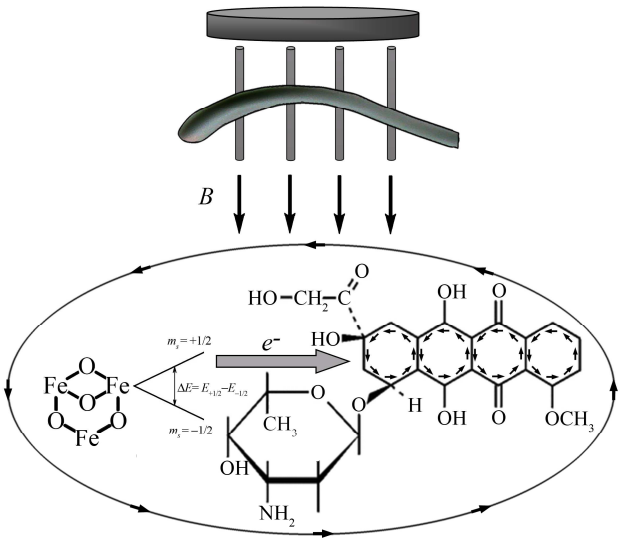


Figure 4. Electron transfer in nanocomplex from iron oxide to aromatic ring of doxorubicin.

It is possible to assume that antitumor effect of nanocomplex increased due to the phenomenon of spin-dependent electron transport between iron oxide and aromatic rings of DOX during electromagnetic irradiation. Varying strength of benzene ring current modulated the level of oxidative stress in malignant tumors. This increased the ability of anthracyclines in the presence of iron oxide to be metabolized by the cells with the formation of free radicals. Quinone part of DOX tetracycline ring transformed into semiquinone free radical involving mitochondrial, nuclear and microsomal NADP-redox enzyme cells. As a result the processes of apoptosis and necrosis are initiated in tumor cells [14].

X. CONCLUSION

In the research of animals with Guerin carcinoma was shown, that magneto-mechano–chemically synthesized magnetic nanocomplex on the basis of nanoparticles of Fe_3O_4 with antitumor anthracycline row antibiotic doxorubicin and next local electromagnetic irradiation by spatially inhomogeneous constant magnetic and variable electromagnetic fields had greater antitumor effect than conventional doxorubicin and magneto-mechano-chemically synthesized magnetic nanocomplex without electromagnetic irradiation. In the next steps of our research we are planning to extend pre-clinical researches on other models of tumor process for deeper investigations of oncological clinical potential of magnetic nanotherapeutics.

REFERENCES

[1] www-dep.iarc.fr

[2] WTEC panel report on assesment of physical sciences and engineering advances in life sciences and oncology (Aphelion) in Europe. 2012. http://wtec.org/aphelion

[3] http://www.cnt.uk.com

[4] M. Arruebo, R. Fernandez-Pacheco, and M. R. Ibarra, "Magnetic nanoparticles for drug delivery", J. Nanotoday, vol. 2, no. 3, pp. 22–32, 2007.

[5] M. M. Amiji, Nanotechnology for Cancer Therapy. Boca Raton: CRC Press, 2007.

[6] M. Blank and L. Soo, "Electromagnetic Acceleration of Electron Transfer Reactions", Journal of Cellular Biochemistry, vol. 81, pp. 278–283, 2001.

[7] E. D. Kirson, V. Dbaly, F. Tovarys, J. Vymazal, J. F. Soustiel, A. Itzhaki, et al., "Alternating electric fields arrest cell proliferation in animal tumor models and human brain tumors", PNAC, vol. 104, no. 24, pp. 10152–10157, 2007.

[8] A. Jordan, R. Scholz, P. Wust, H. Fahling, R. Felix, "Magnetic fluid hyperthermia: cancer treatment with AC magnetic field induced excitation of biocompatible superparamagnetic nanoparticles", J Mag. Mag. Mat., vol. 201, pp. 413–419, 1999.

[9] A. Szent-Gyorgyi, Electronic biology and cancer: a new theory of cancer. New York: Publisher, M. Dekker, 1976.

[10] E. Gutman, Mechanochemistry of materials. Cambridge: Int. Science Publishing, 1998.

[11] V. E. Orel and A. V. Romanov, "The Effect of Spatially Inhomogeneous Electromagnetic Field and Local Inductive Hyperthermia on Nonlinear Dynamics of the Growth for Transplanted Animal Tumors", in Nonlinear Dynamics, T. Evans, Ed. Chroatia: INTECH, 2010, pp. 285–308.

[12] T. Lagarde, J. Pelletier, and S. M. d'Heres, "Method of producing individual plasmas in order to greate a uniform plasma for a work

surface, and apparatus for producing such a plasma", Patent USA 6407359 B1, 2002.

[13] N. Nikolov, V. Orel; I. Smolanka, N. Dzyatkovskaya, A. Romanov, Y. Mel'nik et al., "Apparatus for Short-Wave Inductothermy "Magnetotherm", Proceedings of NBC 2008, A. Katushev, Yu. Dekhtyar, and J. Spigulis, Eds. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 294–298.

[14] V. E. Orel, I. B. Shchepotin, I. I. Smolanka et al., Radiofrequency hyperthermia of malignant neoplasm, nanotechnology and dynamic chaos. Ternopol: Ukrmedbook, 2012. (in Russian)

[15] E. M. Gutman, Mechanochemistry of materials. Cambridge: Int Science Publishing; 1998.

[16] H. Kuramitz, "Magnetic microbead-based electrochemical immunoassays", Anal. Bioanal. Chem., vol. 394, no. 1, pp. 61–69, 2009.

[17] V. Orel, A. Shevchenko, Y. Mel'nik et al., "Physicochemical characteristics of magnetosensitive nanocomplex obtained by mehanomagnetochemical technology of dry synthesis", Metallofiz. Noveishie Tekhnol., vol. 32, pp. 1157–1167, 2010.

[18] S. A. Pud, V. F. Kovalenko, and M.V. Petrichuk, "Enhancement of the magnetoresistive effect in nanocomposites based on manganite $La_{0.67}Sr_{0.33}MnO_3$ and magnetite Fe_3O_4", Technical Physics, vol. 55, no. 4; pp. 521–525, 2009.

[19] V. E. Orel, A. D. Shevchenko, I. I. Dzyatkovskaya, N. A. Nikolov, A. V. Romanov, A. Y. Rihalsky et al., "Comparative study of magnetic properties and antitumor effect of nanocomplexes of oxides iron nanoparticles and doxorubicin during radio-frequency hyperthermia of Walker 256 carcinosarcoma", Reports of the National Academy of Sciences of Ukraine, in press. (in Ukrainian)

[20] N. Emanuel, Kinetics of experimental tumor processes. Moscow: Nauka, 1977. (in Russian)

[21] R. J. Serfling, Approximation theorems of mathematical statistics. New York: John Wiley & Sons, 1980.

[22] S. Engstrem, "Magnetic field effects on free radical reactions in biology. bioengineering and biophysical aspects of electromagnetic fields", F. S. Barnes and B. Greenebaum, Eds. CRC Press, pp. 157–164, 2006.

Stochastic Properties of Nephroscintigraphic Images with ^{99m}Tc-DMSA

Nikolov N.A.
Medical Physics & Bioengineering Department
National Cancer Institute
Kyiv, Ukraine
nikolka_@ukr.net

Supruniuk D.A., Smetaniuk A.N.
Physical and Biomedical Electronic Department
National Technical University of Ukraine "KPI"
Kyiv, Ukraine

Abstract — **Physical phantom studies in nuclear medicine images have a number of shortcomings that makes it not possible to correctly estimate the parameters of computer image analysis, in particular renal ^{99m}Tc-DMSA. Therefore, developing a virtual phantom of scintigraphic kidney is important. This paper addresses the narrow task: the quantitative evaluation of the stochastic properties of scintigraphic images of kidney with ^{99m}Tc-DMSA and the evaluation of lumped and distributive parameters of noise. It is shown that the virtual image noising of renal ^{99m}Tc-DMSA should be used with Gauss white noise, and noise intensity increases proportionally with the increase of radioactivity in the kidney accounts.**

Keywords — image segmentation; scintigraphic phantom of kidney; stochastic properties of scintigraphic images; noise.

I. INTRODUCTION

One of the most informative and perspective methods of studies of functional state of kidney in children and adult is nephroscintigraphic with ^{99m}Tc-DMSA (dimercaptosuccinic acid) [1-3]. ^{99m}Tc-DMSA has a high level of cortical fixation in kidney because of active sulfhydryl ferments system of epithelial cells proximal canals and low excretion of radiopharmaceutical with urine. This feature of ^{99m}Tc-DMSA allows it to be long fixed in kidney [4, 6] which enables distinct visualization of cortical layer border of kidney. Such feature of fixation drug is defined using ^{99m}Tc-DMSA as one of the main radiopharmaceutical (RPh) for detecting of structure-functional changes of kidney [7-9]. Functional state of kidney parenchyma is estimated based on percent of including ^{99m}Tc-DMSA in kidney, whereas the structure changes are based on the analysis of a spatial distribution of RPh and availability area of hyperfixation RPh. Special meaning of scintigraphic with ^{99m}Tc-DMSA has in case of monitoring studies, estimate dynamics of disease kidney and the effectiveness of therapy. In the first instance in case of such studies an attention is paid on the appearance of new focuses of hyperfixation RPh, which is interpreted as sclerosis area of kidney tissues (except of cancer disorders). Under sclerosis changes understanding replacement/death of kidney functional tissues by connective tissue which is not functional related to appointment of kidney in organism. It is considered that such changes are irreversible. However, in practice change of spatial distribution of a preparation, change of an arrangement of focal changes is quite often noted. That is the quantity of the centers and their sizes can not only increase, but also decrease, some focus can disappear at all. Such changes should be interpreted not as sclerosis, but as local changing of functional state of parenchyma. For objective assessment of such process it is required to estimate specific percent of including RPh in pathologic area, to estimate form of focus and their quantity. However, software which is used by gamma-cameras allows only making of visual estimation of distribution of drug in kidney, that is performing quality estimate. The fundamental technical problem for solving these tasks is development of image segmentation algorithm. It is connected with that the scintigraphic images is stochastic and focus of sclerosis do not have clearly right form; they may be multiple and their size is lower than resolution of gamma-camera. Under stahostic image properties in this paper the noise properties should be understood. They are determined by a random radioactive decay, hardware noise, sparse matrix of the image (especially in gipofunctional state of renal parenchyma), and by nonlinear optical properties of gamma camera collimators. Noise properties of the images may also be dependent on renal function. The improvement quality of image filtering, segmentation and computer analysis of scintigraphic images devoted a lot of scientific works [10-13]. Algorithm of segmentation and analysis is developed, as usual, by using phantom images [14, 15, 16]. The main disadvantages of such phantoms are their extremely low prevalence and as a result high price. However, algorithms of image segmentation are multiparametric and these parameters strongly depend on individual features of gamma-camera. Another disadvantage is that they are not physiologic: phantoms, as usual, simulate large focus of changing with clear geometric forms or state which is close to normal. At the same time, general functional state of kidney can be completely different: hypo-, normal- and hyper-functional and focus can be minor focus with not irregular form. Third disadvantage to consider is that even on high contrast scintigraphic images it is difficult to identify profile of formation. There is a problem of an accurate assessment of correctness of allocation of interesting areas. At phantom physical studies current task is theoretically solved by strong axis bound of typical points/areas of phantom to coordinate system of gamma-camera detector. In contemporary gamma-camera it can be made only in manufacturing conditions. Disadvantages discussed above can be avoided by using virtual phantoms, generated by computer technique. In turn, computer model requires detailed analysis of stochastic properties of real scintigraphic images of kidney with different pathology and functional disorders. Our work group could not

to find detailed literature information about this theme, however some scientific works should be mentioned, for example [2, 17-19], which can be perspective in case of quantitative analysis of nephroscintigraphy with ^{99m}Tc-DMSA. That is why we think that current direction of research is perspective. The purpose of this work is researching stochastic properties of scintigraphic images of kidney with c ^{99m}Tc-DMSA, for further implementation of the virtual phantom adequate to renal ^{99m}Tc-DMSA

II. METHODS OF RESEARCH

The original data of nephroscintigraphic studies with ^{99m}Tc-DMSA of 30 patients were analyzed. The data base of Kyiv City Clinical Hospital №14 was used. Data for research was selected so that to most widely cover the different nosology disorders of kidney and different types of accumulation and distribution of drugs except anomalous growth of kidney and cancer diseases. Analyzed images were obtained for the patients in 10 to 30 years (21 ± 5 years) age group. The main clinical diagnoses for these patients were: pyelonephritis, chronic kidney failure, glomerulonephritis, vesico-ureteric reflux, dysmetabolic nephropathy. Studies were made by using gamma-camera SPECT-1 (Ukraine) in planar mode with resolution 512-512 (16 bit) and low-energy collimator general appointment:

- rear projection of kidney (1 frame for 1 min);
- rear projection of kidney (1 frame for 500 thousand impulse on detector);
- left skew projection of kidney (1 frame for 500 thousand impulse on detector);
- right skew projection of kidney (1 frame 500 thousand impulse on detector).

The obtained results were compared with equivalent images received by using Infinia Hawkeye™ gamma-camera produced by “General Electric” and located in Kyiv Heart Center.

Each frame was analyzed most part of whole scintigraphic image also left and right kidney was analyzed separately.

Computer analysis of images was made by using algorithm which was developed in Matlab 7.0. At first integral parameters of whole image were estimated:

- histogram of probability density of radioactive count,
- maximum and minimum value,
- sum of whole pixels,
- average value,
- dispersion (variance),
- entropy distribution.

On the second step, spatial distribution of RPh was estimated. Stochastic parameters were estimated in floating window with 5x5 pixels size in which were computed: average value of radioactive count, dispersion, asymmetry, excess, energy of distribution. It is known that point source of radioactivity has a normal (Gauss) distribution velocity of radioactive count. Connected with this were made testing can we consider that on some minimal square of scintigraphic image, for example, with size 5x5 pixels, the distribution tends to normal distribution? For checking this question the coefficient of correlation between experimental density of probability distribution of radioactive count and ideal normal distribution law with substituting average value and dispersion was computed. Another criterion of normal distribution was served by Kolmogorov-Smirnov test. As long as histogram of radioactive count is constructed based on results of analysis of matrix 5x5 pixels (in perfunctory experiments size varied to 50x50), distribution of probability density is always limited. At the same time, Gauss distribution law is not limited and argument varies from $-\infty$ to $+\infty$. That is the way for more accurate approximation of histogram of experimental data on the ending quantity of value idealized Gauss function multiplied by coefficient $1/dS$, where dS is area under the curve of ideal Gauss distribution which is out of argument of experimental distribution. Estimate of stochastic parameters in float window with minimal size of 5x5 pixels allows constructing parametric images of calculated values.

III. RESULTS AND DISCUSSION

Analysis of scintigraphic images obtained on SPECT-1 in whole scale of gamma-camera detector vision shows that the histogram of probability distribution of radioactive count tends to have the correlation coefficient of 0.97±0.03 with exponential appearance, as shown in “Fig. 1”. The side projections are characterized by a little smaller value of exponent constant. For rear projections, the exponent constant tends to -0.53±0.01 and side projections tends to -0.43±0.03. More detailed analysis of histogram shows that in interval of intensities 25÷60 % to maximal for them is typical quasi-uniform area, plateau. Typical histogram of probability distribution of radioactive count of image of kidney and their vicinity radius 1.5-2 cm is more clear view of plateau “Fig. 2”. In this situation quasi-constant level can be marked out in range from minimal radioactive count till level of 65 % from maximum count. In case of increasing contrast kidney/ background on area plateau appears characteristically maximum. At images from gamma-camera Infinia Hawkeye™ this phenomenon is observed more clearly. A characteristic feature of histogram of probability density of radioactive count of kidney image and their vicinity is sensitivity to exposition time or functional ability of kidney. In case of insufficient exposition time for clear visualization of kidney or in case of sudden reduction of parenchyma functional activity, approximation line of histogram tends to be the exponent. Three patients with "chronic renal failure" have a coefficient of correlation between trend of exponent line and experimental data equal to 0.95±0.03. Probability distribution of radioactive count directly of kidney without surrounding tissue tends to have high correlation coefficient (0.96±0.02) toward normal distribution, however, depending on availability of focus of sclerosis and their size, functional state of parenchyma and two-, three- modal distribution.

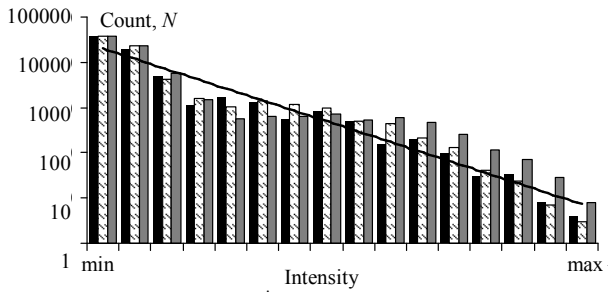


Figure 1. Typical histogram of radioactive count for whole image for: ■ – rear projection with exposure in 60 s, ▨ – rear projection with exposure in 500 thousand impulses on detector, ■ – skew side projection with exposure in 500 thousand impulses on detector.

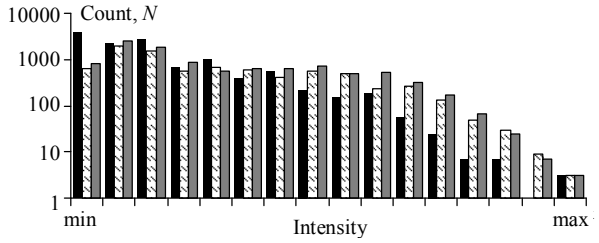


Figure 2. Typical histogram of radioactive count for image of kidney and surrounding tissue for: ■ – rear projection with exposure in 60 s, ▨ – rear projection with exposure in 500 thousand impulses on detector, ■ – skew side projection with exposure in 500 thousand impulses on detector.

In this case additional peaks (modes) have a little bit lower amplitude toward to primary mod, but they can locate as from left side and as from right side. If functional state of kidney becomes lower then dispersion of this distribution became higher. Also stochastic properties of the difference image was analyzed when from original image was subtracted relatively idealized (smoothed) image. Subtracted image is characterizes noise of system. In this case whatever on what gamma-camera was obtained images of probability distribution with validity (p>0.995) was described by Gauss function "Fig. 3".

Analysis of stochastic properties of specific distribution RPh on scintigraphic images of kidney shows that at scales lower then 0.2 см^2 function of radioactive distribution is described by normal (Gauss) law. This result is seen from parametric images which were plotted based on calculated Pirson coefficient of correlation between real distribution and approximation function, which has Gauss type, as well as Kolmogorov-Smirnov criteria. It is without doubt that existed areas where criteria gives low agreement between model and real distribution. However, it is impossible to ascertain visual correlation between pathology of kidney and current area. That is why such few areas of mismatch with normal distribution were interpreted as result of not enough statistics. Parametric images of specific distribution of RPh in kidney plotted based on estimation of asymmetry distribution and excess shows practically evenly noised distribution where kidney was difficult visualized. At the same time, largely part of pixels of parametric images of asymmetry distribution was >0 (average 0.5÷1.5), which is shown tendencies of dislocation of maximum distribution law to the left side in area of lowered values of radioactive count. Values parametric image of excess was trend to zero from the left side mainly (average -0.5). Parametric images of entropy and energy shows difference only between kidney tissue and background: average value of entropy at kidney parenchyma was exceeded background in 2-2.5 times, energy was lower in 5-7 times. These images do not characterized focuss changes in kidney. Row of regularities of RPh distribution at kidney characterized dispersion and relative standard deviation (σ/N_0) images. Current parametric images are quiet good mark out shape of kidney on the relation to surrounded tissue. Herewith the relative standard deviation was always lower than one as opposed to tissue background and average value was lying between 0.1 and 0.3. Value of dispersion distribution at kidney was higher than tissue background in several times and was depended on accumulation level of RPh. However it was mentioned that the significant difference between texture properties of parametric images of dispersion which was obtained at gamma-camera SPECT-1 and Infinia HawkeyeTM "Fig. 4".

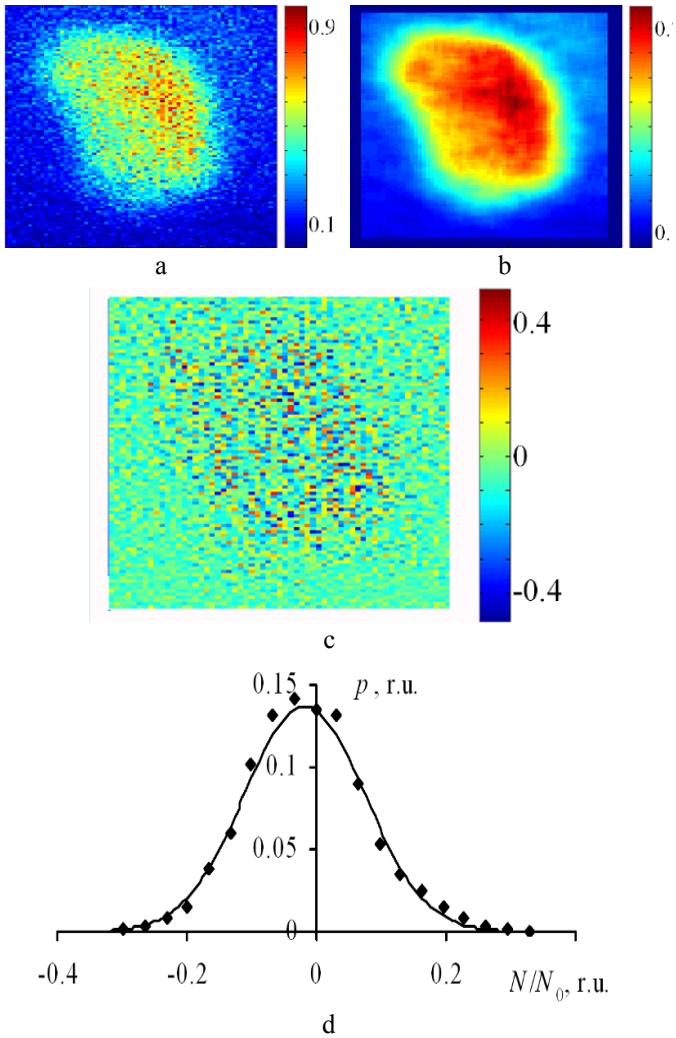


Figure 3. The distribution of noise in the image kidneys with ^{99m}Tc-DMSA: a – original image (A) kidney; b – smoothed original image (B); c – difference image (C=B-A); d – the probability distribution of the noise intensity in the image C (♦ – experimental data, solid line – approximation to a Gaussian function).

Thus dispersion distribution for images of kidney obtained by SPECT-1, characterized by even dispersion distribution at kidney parenchyma with decreasing absolute value at large areas with lower functional ability at kidney parenchyma, with increasing - with normal or hyper functional ability of kidney For images obtained by Infinia HawkeyeTM distribution of dispersion largely characterized contour of kidney and do not characterized sclerosis area with contrast "focus"/"normal tissue" less than 30%.

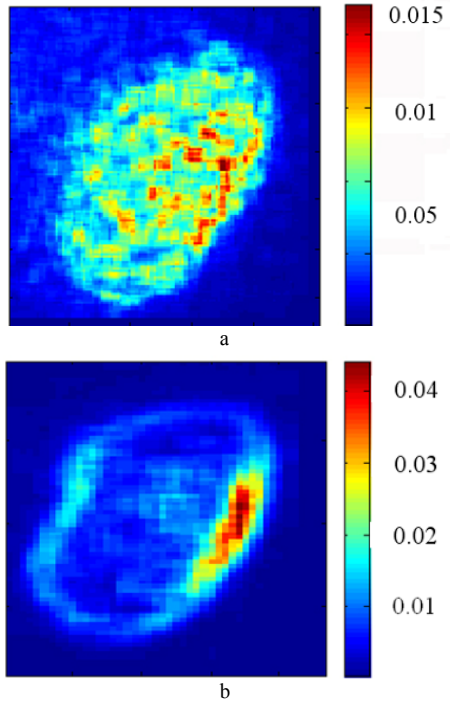


Figure 4. Parametric images of dispersion distribution at kidney of two different patient in case of large focuss changes in parenchyma with sufficient functional ability of kindey. Skew side projection, obtained on gamma-camera: a – SPECT-1, b – Infinia HawkeyeTM.

The differences in stochastic properties of images obtained at different gamma-camera also demonstrate dispersion depending on average value of radioactivity at analyzed window "Fig. 5".

At Fig. 5 it can be observed that dispersion as average is raising linearly with increasing level of accumulating drugs in parenchyma. Top border of points at Fig. 5 is defined predominantly by edge effects of image; more tight located points but is steel sparse characterized border of sharp transition of one structure to another; areas with tight packed points – accumulating of RPh in parenchyma, where accumulation of drug can be accepted as even. In this case for images obtained at Infinia Hawkeye^{TM} coefficient slope of liner approximation of dispersion from average level of radioactivity in kidney is little bit lower than for surrounded tissue. Geometric center mass of radioactive and dispersion distribution in normal trends to center of kidney with elliptic shape in case presence of focus changes and partly lowed functional ability of kidney – to external medial surface.

It should also be noted that in case of smooth of scintigraphic images of kidney with ^{99m}Tc-DMSA by medial filter, the maximal radioactive count is decreasing on 20-40 %. In case insufficient of frame exposition maximal count after smooth can be decreased more than in 2 times. Current feature can have a big role in computing modulation of kidney with sharply lowered functional activity of parenchyma.

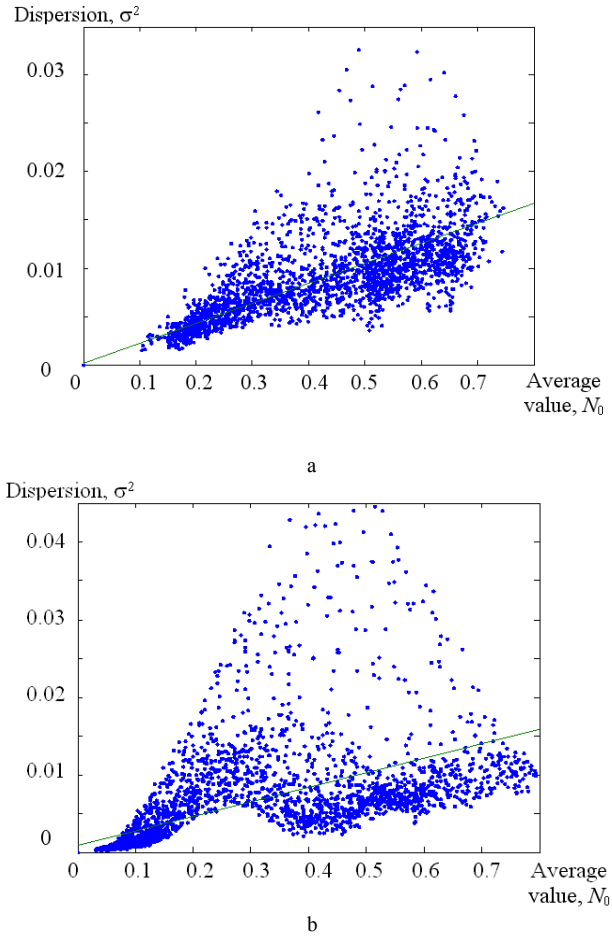


Figure 5. Typical depending of dispersion on average value of radioactive count at kidney for two different pattinet in case of large focuss changes in parenchyma with sufficient functional ability of kindey, obtained on gamma-camera: a – SPECT-1, b – Infinia HawkeyeTM.

CONCLUSION

1. Analysis of stochastic properties of scintigraphic images with ^{99m}Tc-DMSA shows that in case of computer modelling image with Gauss law of distribution can be used. In this case noise dispersion average quasi-linear depends on level of signal higher than 20-30 % from maximum value.
2. Among all analyzed parametric images of dispersion distribution and relative standard deviation is most perspective for scintigrams segmentation.
3. In case of development algorithm of image analysis, precise algorithm parameters should depend on type and model of particular gamma-camera.
4. With further conducting computer simulations of static nephroscintigraphic obtained in this study the adequacy of the virtual phantoms data to the stochastic properties of the image kidney with ^{99m}Tc-DMSA will be evaluated.

REFERENCES

[1] N. Girotto, A. Smokvina, S. Grbac Ivankovich, V. Licul, "Effects of background substraction on differential kidney function measured by static scintigraphy with DMSA and dynamic scintigraphy with MAG3", Nuklearmedizin, vol. 47, N. 1, pp. 43—47, 2008.

[2] A. Hitzel, A. Liard, J. N. Dacher, I. Gardin, et al., "Quantitative analysis of 99mTc–DMSA during acute pyelonephritis for prediction of long–term renal scarring", J. Nucl. Med., vol. 45, N. 2, pp. 285–289, 2001-2002.

[3] Society of Nuclear Medicine, "Procedure Guideline for Renal Cortical Scintigraphy in Children 2.0." Society of Nuclear medicine procedure guidelines manual, pp. 157–160, 2002.

[4] F. Heldrich, M. Barone, E. Spiegler "UTI : diagnosis and evaluation in symptomatic pediatric patients", Clin. Pediatr., vol. 39, N. 8, pp. 461–472, 2000.

[5] A. Hoberman, M. Charron , R. Hickey, et al., "Imaging studies after a first febrile urinary tract infection in young children", N. Engl. J. Med, vol. 16, pp. 195–202, 2003.

[6] L. De Guevara, P. Franken, C. De Sadeleer, H. Ham, A. Piepsz, "Interobserver Reproducibility in Reporting on 99mTc–DMSA Scintigraphy for Detection of Late Renal Sequelae", J. Nucl. Med., vol. 42, N. 4, pp. 564–566, 2001.

[7] A. Biggi, L. Dardanelli, G. Pomero, P. Cussino, C. Noello, O. Sernia et al., "Acute renal cortical scintigraphy in children with a first urinary tract infection," Pediatr. Nephrol., vol. 16, N. 9, pp. 733–738, 2001.

[8] V. Camacho, M. Estorch, G. Fraga et al., "DMSA study performed during febrile urinary tract infection : a predictor of patient outcome ?", Eur. J. Nucl. Med. Mol. Imaging., vol. 31, N. 6, pp. 862–866, 2004.

[9] A. Hitzel, A. Liard, P. Vera et al., "Color and power Doppler sonography versus DMSA scintigraphy in acute pyelonephritis and in prediction of renal scarring," J. Nucl. Med., Vol. 43, N 1, pp. 27–32, 2002.

[10] A. S. M. Sabbir Ahmed, M. Demir, L. Kabasakal, I. Uslu, "A dynamic renal phantom for nuclear medicine studies", Med. Phys., vol. 32, iss. 2, pp. 530–538, 2005.

[11] A. Kojima, Y. Ohyama, S. Tomiguchi at al., "Quantitative planar imaging method for measurement of renal activity by using a conjugate-emission image and transmission data", Med. Phys., vol. 27, iss. 3, pp. 608–615, 2000.

[12] A. Kojima, Y. Takaki, M. Matsumoto, et al., "A preliminary phantom study on a proposed model for quantification of renal planar scintigraphy", Med. Phys., vol. 20, iss. 1, pp. 33–37, 1993.

[13] C. Brambilla, G. Burkhardt, J. Borges, A. Marques da Silva, "Monte Carlo Simulations of GSF Family Voxel Phantoms for Quantification of Renal Planar Scintigraphy", Med. Phys., vol. 33, iss. 6, pp. 2015, 2006.

[14] G. Zubovsky, M. Devishev, E. Ivanov, O. Andreeva, A. Luchkov, "Radionuclide studies with dynamic kidney phantom," Medical radiology, vol. 28, N. 6, pp. 77–81, 1983.

[15] J. Heikkinen, "New automated physical phantom for renography", J Nucl Med., vol. 45, N. 7, pp. 495–499, 2004.

[16] Advertising Textiles International Trade Company PTW-Freiburg http://www.ptw.de/quality_control_nm.html.

[17] M. Goris, E. Gordon, D. Kim, "A stochastic interpretation of thallium myocardial perfusion scintigraphy", Invest Radiol., vol. 20, N. 3, pp. 253–259, 1985.

[18] K. Murase, Y. Yamazaki, T. Mochizuki, J. Ikezoe, "Renal uptake rate measurement of ^{99m}Tc-dimercaptosuccinic acid using spectral analysis", Nucl. Med. Commun., vol. 23, N. 3, pp. 265–273, 2002.

[19] M. Al-Rifaie, A. Aber, A. Oudah, "Utilising Stochastic Diffusion Search to Identify Metastasis in Bone Scans and Microcalcifications on Mammographs", Proceedings of 2012 IEEE International Conference on Bioinformatics & Biomedicine (BIBM 2012), Multiscale Biomedical Imaging Analysis (MBIA2012), Philadephia, United States, IEEE.

Saturation of Electroencephalogram Permutation Entropy for Large Time Lags

Anton Popov, Oleksii Avilov
Physical and Biomedical Electronics Department
National Technical University of Ukraine
"Kyiv Polytechnic Institute"
Kyiv, Ukraine
anton.popov@ieee.org

Oleksii Kanaykin
Department of Functional Diagnostics
Institute of Neurosurgery of Academy of Medical Sciences of Ukraine
Kyiv, Ukraine

***Abstract* — New approach to investigation of electroencephalogram (EEG) patterns over different time spans is proposed based on permutation entropy (PE) measure. PE for large time lags (1-100 samples) and orders from 2 to 8 was calculated for EEG signals of three types: EEG from healthy subject, EEG containing high-magnitude bursts and EEG with epileptiform complexes. PE value saturation phenomena is revealed for large time lags (approximately 10 or 30 samples for different EEG types) and for whole range of orders used in this study. This effect might be due to the inherent EEG property, namely that the samples in each EEG pattern are spaced too far from each other and they are like samples from some arbitrary stochastic process.**

Keywords—EEG; permutation entropy; epilepsy; EEG pattern analysis; permutation entropy saturation

I. INTRODUCTION

Over the past decades nonlinear biomedical signal analysis attracts much attention from researchers. The reasons for this could be, first, enhancing and growing knowledge about the nature of human body organization as an adaptive network of multiple nonlinear feedback subsystems aimed to sustain homeostasis and to interact with environment. These subsystems are connected by nonlinear pathways and thus analyzing them with conventional linear approach is not enough to reveal all characteristics. The second reason is the extreme development of sophisticated mathematical methods of uni-, bi-, and multivariate analysis which enabled deeper analysis of dynamic nonlinear systems and more precise imaging of underlying processes. All these are strengthened by utilization of powerful computer technology, in particular parallel and cloud computations for biomedical data analysis.

One of nonlinear dynamic system analysis techniques is entropy data analysis. Recent publications prove that entropy concept, from classical Shannon Entropy to entropy characteristics adapted to finite sampled signal description, adequately describes complex nonlinear and stochastic processes accompanying the activity of human body. Entropy analysis (Shannon entropy, spectral entropy, sample entropy etc.) is used for analysis of brain electrical activity, heart rate variability and other biomedical signals. Different types of entropy analysis reflect unpredictability, chaoticity, complexity and nonlinearity of brain activity, and thus are fruitfully employed in electroencephalogram (EEG) analysis. In our work the permutation entropy (PE) measure was used for investigating time characteristics of brain electrical activity reflected in electroencephalogram.

Many researches focused on EEG entropy analysis during anesthesia and for the classification of the depth of narcosis [1-3]. In [4] combined multiscale permutation entropy is proposed for estimating the anesthesia depth, and in the paper [5] this measure is used for automated sleep stage classification. PE is used in [6] for brain activity classification in brain-computer interface system for estimation of EEG complexity. In [7] PE is applied for discrimination between normal and epileptic EEG with 97% of true positives rate for denoised EEG. PE can also be applied to estimation of EEG complexity measure in ictal and interictal periods [8] or for dynamical characterization of EEG [9].

In this paper we addressed the question whether PE is capable to reflect similar behavior of EEG signal over time intervals of different duration. In particular, main purpose was to show PE employment for finding signal patterns of similar shapes and to obtain the PE dependence on the time lag and permutation order. As an example, PE of EEG signals of three classes: from healthy subject, EEG with epileptiform complexes, and EEGs with bursts of high-magnitude activity inherent to brain structures irritation was calculated.

II. METHOD OF PERMUTATION ENTROPY

Permutation entropy measure was introduced for the first time in the seminal work of Brandt and Pompe [10] and now is thoroughly mathematically developed, in particular for applications in EEG analysis [11-12].

The PE value can be calculated for sampled signals only. Let $x[n]$, $n = 0...N-1$ denotes the samples of one EEG channel in the time window of $N-1$ samples duration. To calculate PE, the order *m* (integer number, $m \geq 2$) and time lag l, $l \in N$, $l \geq 1$ should be chosen. Starting from the first sample in the time window of interest, *m* successive samples of $x[n]$ with time step *l* are selected according to:

$$v_l[i] = x[i], x[i+l], x[i+2l], x[i+3l], \ldots, x[i+(m-1)l],$$

where $i = 1 \ldots N - (m-1)l$,

to obtain $N - (m-1)l$ embedded signal patterns $v_l[i]$.

Following authors of [10], it is assumed that the quantization step of analog-to-digital signal conversion is small, and sampling rate is high enough, so the signal samples are not equal to each other; at least the absence of equal samples is required to hold true within each signal pattern $v_l[i]$ for which PE is calculated.

Let's assume that the values from embedded pattern $v_l[i]$ are taken and arranged in ascending order, making the sequence $\widetilde{v_l[i]}$ of m different numbers. It's known that for m different numbers there are $m!$ permutations, and they are variations of arrangements of m numbers without repetitions. In other words, for m different numbers there are only $m!$ possible ordered patterns or motifs. Thus each initial $v_l[i]$ can be considered as one of $m!$ permutation of $\widetilde{v_l[i]}$. This permutation pattern will be denoted π_j, $j = 1 \ldots m!$

When analyzing signal in the time window, among all patterns $v_l[i]$ some arrangements π_j of signal samples may occur more often than other ones. For signal $x[n]$ the relative frequency of permutation π_j is defined as:

$$p(\pi_j) = \frac{q(j)}{N - (m-1)l},$$

where $q(j)$ is the number of occurrence of permutation π_j.

Permutation entropy of the signal is given [10] by:

$$PermEn(m,l) = -\sum_{j=1}^{m!} p(\pi_j) \log p(\pi_j).$$

This value is the measure of the amount of information contained in comparing m consecutive samples of the signal over some time interval.

$PermEn(m,l)$ has values in the range $0 \le PermEn(m,l) \le \log m!$ $PermEn(m,l) = 0$ corresponds to the case when there is only one pattern in the signal (the signal values are totally predictable, they are only ascending or descending). $PermEn(m,l) = \log m!$ corresponds to the case when signal contains all possible patterns with equal probability of occurrence (the signal values are totally random).

To avoid the PE dependence on the order for the same signal, normalization procedure was introduced:

$$PermEn_{norm}(m,l) = \frac{-\sum_{j=1}^{m!} p(\pi_j) \log p(\pi_j)}{\log m!}.$$

In this case $PermEn_{norm}(m,l)$ value always lies between 0 and 1.

III. Considerations on PE parameters and lag-dependent EEG analysis

It can be seen from the previous equations that $PermEn(m,l)$ values depend on the order and lag. Order m equals to the number of samples to be taken from the signal to construct one pattern. It controls the number of permutations and in this way it influences the number of unique patterns which could be observable in the signal. If the order is not large, the variety of diverse arrangements of samples in the patterns is not large too. For example, for $m = 2$ there are only two patterns, for $m = 3$ - six patterns. But with order increase, the number of patterns also increases very fast. Selecting large order allows finding more patterns in the signal, i.e. more variations of interdependencies in the values of successive samples. But when the number of patterns is very large, this rich variability of patterns can lead to the situation when patterns will be different insomuch that they will be just the realizations of noise.

If order is small, there will be only a few patterns that could be detected, but if it is large, the number of possible permutations can be so large that using all of them could be rather meaningless. The order can be interpreted as the parameter of diverseness, so it is worth to investigate if PE of the same signal varies significantly with order change.

The lag l with which the signal samples are picked to form embedded pattern $v_l[i]$, corresponds to the duration of each pattern. When lag equals to one, duration will be defined by the PE order m only: the interval between samples will be equal to one sampling interval. When lag increases, interval between samples in $v_l[i]$ increases too. Thus duration of each pattern will also increase while its shape will remain unchanged.

The investigation of PE for the same order but for different lags, and in the same time for different orders but same lags can provide researcher with information about whether there are typical distributions of patterns over different time spans in the signal. This approach was used in this work for studying of the presence of similar patterns in the EEG signal over different time spans, and investigating the entropy characteristics of signal as the function of the time "depth" to which particular pattern $v_l[i]$ extends. Investigation of such EEG characteristics is the purpose of experiments presented further in this paper.

IV. Experimental results and discussion

During experiments normalized PE characteristics of EEG from three classes, reflecting the time properties of permutations of similar

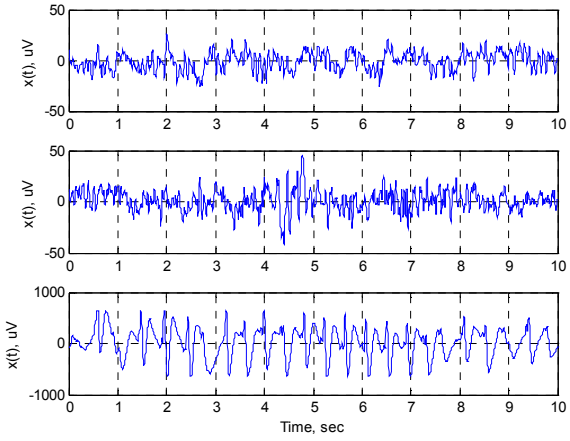


Fig. 1. Signals used in PE calculations, from top to bottom: from healthy subject, EEG containing high-magnitude bursts and EEG with epileptiform complexes.

order, but over different time intervals were obtained. EEG signals recorded in the Department of Functional Diagnostics of Institute of Neurosurgery of Ukraine were used in this study (256 Hz sampling rate, 10/20 derivation system with physical reference from two ears). All records were preliminary inspected by qualified physician and only artifact-free signal intervals were accepted for further analysis. Three types of EEG signals were used: from healthy subject, EEG containing high-magnitude bursts and EEG with epileptiform complexes. Examples of EEG signals under investigation are shown in Fig. 1. Four EEG signals of each type were used for PE calculations.

Although according to the recommendations given in [10] optimal PE orders are 3 to 5, for the sake of completeness and for obtaining general tendencies in PE changes, $PermEn_{norm}(m,l)$ values were calculated for the orders from 2 to 8. Figures 2-4 show the examples of PE calculations for lags from 1 to 100. Example of dependence of PE on lag for all types of signals is given in Fig. 5.

It can be seen from Fig. 2-4 that PE for particular order as a function of lag has very prominent property: large slope for small lags and then flat horizontal plateau for larger lags. This $PermEn(m,l)$ function shape is approximately the same for different orders. On Fig. 5 another representation of the plots from Fig. 2-4 is shown, and it can be seen that the change from inclined to horizontal shape of PE curve occurs at lags near 10 for EEG from healthy subject and EEG containing bursts. For EEG with epileptiform complexes PE growth slower with lag increase, and flat horizontal part starts at lags equal approximately 30. After some period of constant values (until lag 100), the PE value starts to decay very slowly and converge to one curve for all types of EEG signals. We think that this phenomenon can be interpreted as PE "saturation" for large time lags. To our knowledge, these types of PE saturation for large lags were rarely observed and analyzed by other researchers before for the case of EEGs of different types.

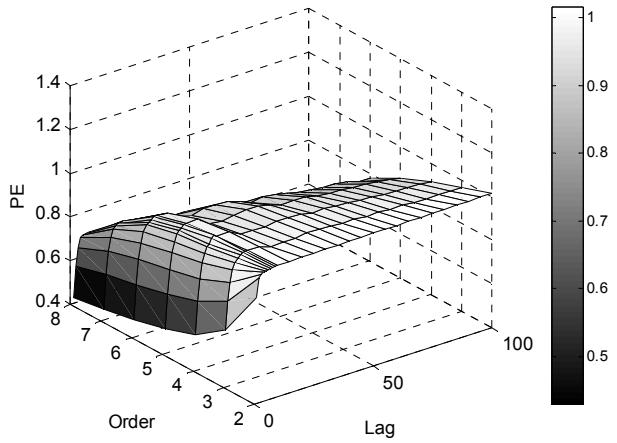


Fig. 2. Dependence of PE for orders $m = 2...8$ for different time lags, EEG from healthy subject.

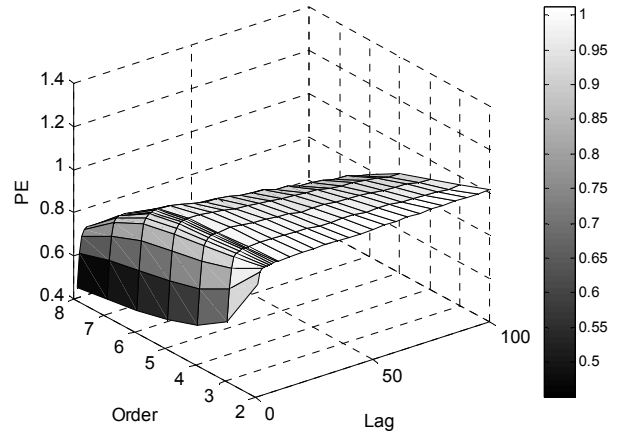


Fig. 3. Dependence of PE for orders $m = 2...8$ for different time lags, EEG containing high-magnitude bursts

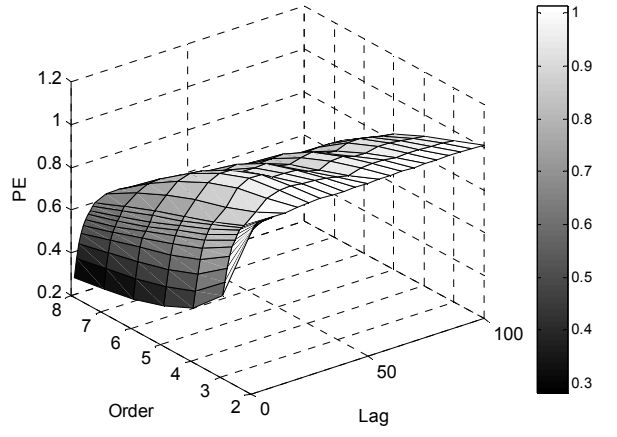


Fig. 4. Dependence of PE for orders $m = 2...8$ for different time lags, EEG containing epileptiform complexes.

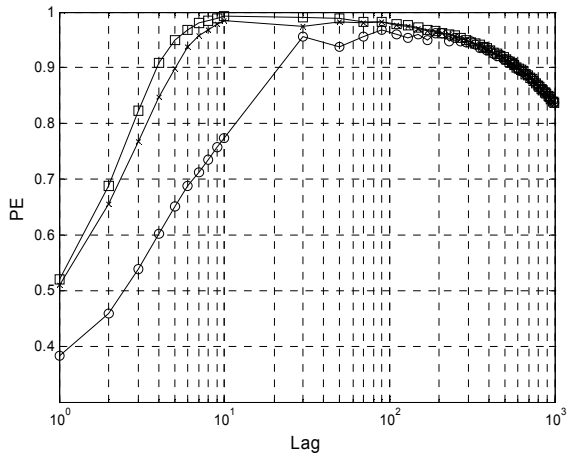


Fig. 5. PE as a function of time lag for order $m = 5$ for tree types of EEG signals: from healthy subject (×), containing high-magnitude bursts (□) and containing epileptiform complexes (○).

This phenomenon is to be explained, both from mathematical and physiological points of view. It's known that normalized PE close to 1 is inherent to completely random time series [10]. At the moment, we can speculate that PE saturation of EEG for large time lags might be because of the inherentproperties of brain electrical activity reflected in EEG signal. Lag corresponds to the time span of signal pattern, which lasts for $T = \frac{(m-1)l-1}{F_s}$ seconds. For order $m = 5$ saturation sets in for patterns of $T \simeq 0.15$ sec duration for first two classes and of $T \simeq 0.5$ sec for the epileptic EEG. This might be the time interval after which series of samples taken from EEG become close to random signal samples. It can be presumed that when lag reaches the bend value (in our case 10 or 30, depending on the signal types), the samples in each EEG pattern are spaced too far from each other and there are no meaningful connections between them. They doesn't reflect brain activity any more, and they are just like signal samples from some arbitrary stochastic process. Another evidence to the fact that pattern properties with lags larger than 100, do not represent EEG signal anymore, might be the fact that the decay of PE plot after this lag is the same for all types of EEG signals.

One of possible outcome of this phenomenon might be the recommendation not to use too large lags for PE calculation, because after some lag value the PE is almost the same and doesn't change with lag. We can define critical lag value, after which increasing of pattern length will result only in computational burden increase.

In the future studies the phenomenon of PE saturation can be considered from another point of view. EEG sampling rate may influence saturation. In the case considered here, signal sampling rate might be too high or too low to catch entropy properties of signals. The samples are too close or to far from each other for large lags which cause the saturation with following decay. High time-resolution EEG downsampling may be needed together with careful selection of time lags before calculation of PE. This issue is under investigation now. It is also of great interest to find out if there is similar behavior of other entropy measures employing the embedded vectors approach. If PE saturation reflects the essential property of brain activity or general property of biosignals, the deeper researches on this issue should be done. The theory behind saturation of PE for different lags and orders might be very intriguing.

V. CONCLUSION

The approach to analysis of permutation entropy of EEG signal over different time spans is proposed. Experimental studies on permutation entropy analysis of EEG from healthy subject, EEG containing high-magnitude bursts and EEG with epileptiform complexes have shown that the bending of PE-lag curve can be observed. The phenomenon of saturation of electroencephalogram permutation entropy is revealed and discussed.

REFERENCES

[1] A. Anier, T. Lipping, V. Jäntti, P. Puumala, and Ari-Matti Huotari, "Entropy of the EEG in transition to burst suppression in deep anesthesia: surrogate analysis," 32nd Annual International Conference of the IEEE EMBS, pp. 2790-2793, August 31 - September 4, 2010.

[2] N. Nicolaou, S. Houris, P. Alexandrou, and J. Georgiou, "Entropy measures for discrimination of 'awake' Vs 'anaesthetized' state in recovery from general anesthesia," Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, pp. 2598 – 2601, Aug. 30, 2011 – Sept. 3, 2011.

[3] Zhenhu Liang, Duan Li, and Xiaoli Li, "Comparison of permutation entropy index and Bispectral Index for monitoring effects of anesthetic drugs to brain activity," 2nd International Conference on Biomedical Engineering and Informatics BMEI'09, p. 5, 17-19 Oct. 2009.

[4] Duan Li1, Xiaoli Li, Zhenhu Liang, L.J. Voss, and J.W. Sleigh, "Multiscale permutation entropy analysis of EEG recordings during sevoflurane anesthesia," Journal of Neural Engineering, № 7, 6010, doi:10.1088/1741-2560/7/4/046010.

[5] Chih-En Kuo, Sheng-Fu Liang "Automatic stage scoring of single-channel sleep EEG based on multiscale permutation entropy," 2011 IEEE Biomedical Circuits and Systems Conference (BioCAS), pp. 448 – 451, 10-12 Nov. 2011.

[6] N. Nicolaou, J. Georgiou, "Permutation Entropy: A new feature for Brain-Computer Interfaces," 2010 IEEE Biomedical Circuits and Systems Conference (BioCAS), pp. 49 – 52.

[7] I. Veisi, N. Pariz, and A. Karimpour, "Fast and robust detection of epilepsy in noisy EEG signals using permutation entropy," 7th IEEE International Conference on Bioinformatics and Bioengineering BIBE 2007, pp. 200 – 203, 14-17 Oct. 2007.

[8] N. Mammone, F.C. Morabito, "Analysis of absence seizure EEG via permutation entropy spatio-temporal clustering," The 2011 International Joint Conference on Neural Networks (IJCNN), pp. 1417 – 1422, July 31, 2011-Aug. 5, 2011.

[9] Gaoxiang Ouyang, Chuangyin Dang, and Xiaoli Li, "Complexity analysis of EEG data with multiscale permutation entropy," Advances in Cognitive Neurodynamics (II), pp. 741-745, 2011.

[10] C. Bandt, B. Pompe, "Permutation Entropy: a natural complexity measure for time series," Physical Review Letters, vol. 88, no. 17, pp. 174102-1 - 174102-4, 2002.

[11] Yinhe Cao, Wen-wen Tung, and J.B. Gao "Detecting dynamical changes in time series using the permutation entropy," Physical Review Letters, vol. E 70, 046217, 2004.

[12] D. Labate, F. La Foresta, A. Bramanti, G. Morabito, and I. Palamara, "Multivariate multi-scale permutation entropy for complexity analysis of Alzheimer's disease EEG," Entropy, vol. 14, pp. 1186-1202; doi:10.3390/e14071186, 2012.

Noninvasive Method for Blood Glucose Measuring and Monitoring

Loshitskiy Pavel Pavlovich
Department of physical and biomedical Electronics
National Technical University of Ukraine "KPI"
Kiev, Ukraine
pepel@phbme.ntu-kpi.kiev.ua

Dmitriy Yurievich Mynziak
Department of physical and biomedical Electronics
National Technical University of Ukraine "KPI"
Kiev, Ukraine
minzyak@mail.ru

***Abstract* — This article is about new method of non-invasive measurement of glucose in human blood. The method is based on the properties of water to respond to subtle changes of the environment.**

Keywords — diabetes mellitus; concentration of solution; temperature fluctuations

I. INTRODUCTION

Diabetes prevalence and incidence has all features of the epidemic, covering most of economically developed countries. According to the WHO at present time world already has more than 347 million diabetic people and this number is growing steadily. WHO projects that diabetes deaths will increase by two thirds between 2008 and 2030 [1].

Diabetes mellitus – clinical syndrome of chronic hyperglycemia and glycosuria caused by absolute or relative insulin deficiency leading to metabolic disorders, affected vessels (various angiopathy), neuropathy and pathological changes in various organs and tissues [2].

For control and monitoring status of patient with diabetic disease is necessary to monitor the level of glucose in the blood. It allows to get the profile of glucose i.e. dependence between eating, exercising, other factors and level of blood glucose. At the present time, for determining of blood glucose the most widely used special test strips which are visually compared with the scale-standard or analyzed in portable devices – glucometers [3]. It's necessary to get blood sample in this cases. Received drop of blood is applied to the test strip and this raises a chemical reaction causing a change of its color. Degree of color changes depends on the level of blood glucose. Using the glucometer is more accurate and reproducible than visual assessment of the strips. Results of the required accuracy is guaranteed if you use the device following the manufacturer's instructions. This method is good enough for personal using. However diagnostic strips are quite expensive during the mass surveys. The main disadvantage of this method is the necessity of invasive impact on patient. There is a possibility of infection during blood sampling and after it considering of multiple blood sampling.

The purpose of this work is development of a new method of non-invasive measurement of glucose concentration in human blood, based on the properties of water to remember the slightest response of environmental changes.

II. THEORETICAL POSITIONS

Thermal motion of the atoms or molecules of the liquid far from the critical point is consists of irregular oscillations with an average frequency of $1/\tau_0$ so close to the frequencies of oscillations of atoms in crystalline solids and amplitude with size of the "free volume" given by its neighbors particles. Center of oscillations is determined by a field of neighboring particles and moves together with displacements of these particles. Therefore in contrast to crystal, the liquid has only temporary and unstable equilibrium. During the average time τ and $\tau >> \tau_0$, center of oscillation of each particle is displaced by the amount of the interatomic distances. This is, obviously, the time associated with the self-diffusion or mixing of particles of liquid. In liquids collective oscillations of whole system as a Debye waves describe only oscillating part of the thermal motion of particles but do not consider significant translational motion of the particles. Correspondingly spectrum of collective oscillations in the liquid breaks in longer wavelengths than in the crystals. The remaining part of the disordered motion of particles connected mainly with thermal translational motion and appears as a disorders of the order in the liquid. In the thermodynamic theory of light scattering this unshifted component is interpreted as the result of light scattering by the fluctuations of entropy [4]. The presence of such a motion in liquids allows to expect the presence of temperature fluctuations in any sufficiently small volume of liquid. Moreover these fluctuations depend on the extremely small external influences of high frequency.

N. Wiener discovered effect in oscillation spectrum of the water as well as in other non-linear substance: the line in the center of region of the spectrum is tapers and rises from dimple [5]. The presence of this effect is increases the accuracy of the relative frequencies of the lines. Therefore in this work we use UHF radiation, because reaction of solution to wideband noise of extremely high frequency of the low-intensity depends on the concentration.

III. MATERIALS AND METHODS

Distilled water with purity of 99.9% and reagents with chemical purity of 99% were used for investigations. Solutions of the desired concentration are prepared by dissolution of reagents in distilled water.

Wideband electromagnetic noise generator is used as a source of external influence that works in continuous mode in

range of 57 ... 68 GHz with a level of power spectral density 10^{-18} W/Hz. The noise signal is modulated in amplitude with random frequency in range of 6...18 Hz. Integral signal power is no more than 10^{-9} W/cm^2.

Criterion of changes in the state of the solution and influence of solute concentration is a fluctuation of temperature, as measured by a specially designed hardware, ensuring the accuracy of relative measurements 0.03 °C. Specifications and detailed measurement procedure of equipment complex are given in [6].

All of the following experiments are performed using next procedure: samples of water poured into two vessels with the same size – i.e. control and investigated samples. These vessels are placed into a rack of plastic foam to prevent convective exchange with the environment. Each test tube closed with a cork from foam in order to exclude the influence of the environment. Thermocouple leads threaded through the cork in such a way that depth of immersion in the solution is the same for both vessels. Positive thermocouple junction is in test tube with investigated water and negative – in the control. All rack of tubes placed in a galvanized metal box without a cap which is grounded separately from the computer for protection against electrical interference. Differential temperature sensor is designed as a differential thermocouple based on a copper-constantan. The measuring system provides input values of the differential temperature in the PC and criterion for the accuracy of the measuring path is selected mean-square error. Statistical processing of the measurement results allows to exclude the dispersion of random values associated with changes in the state of the measuring system. This is achieved by:

- selection of duration and limiting values of response time based on the specified accuracy depending on the level of external noise;

- definition of minimum standard deviation of the measured values of the signal in specified time interval;

- introduction of the calibration signal which allows to determine the actual value of transmission coefficient of the measuring path.

Software allows calculating dispersion and arithmetic mean of differential temperature noise on the selection record or whole fragment of the record.

Coefficients of the direct Fourier transform and values of the autocorrelation function of the differential temperature dispersion of the random process are determined by measuring of temperature fluctuations received for 30 minutes before, during and after action of external physical factors. The results are processed by standard methods in the application package of MatLab.

IV. Results and Discussion

Thermal motions of water solutions and accordingly noise temperature measured in a small volume have a character of a random process. During the measurement time number of dimensions are sufficient for operating on the average values because the measurement system calculates about 40 measurements per second i.e. the samples are representative.

The first phase of work is identifying characteristics dependence of random processes. This dependence describes fluctuations of the temperature for different concentrations of glucose received for 30 minutes before, during and after action of external physical factors. The measurements are performed directly in the glucose solution in this phase of work Fig. 1.

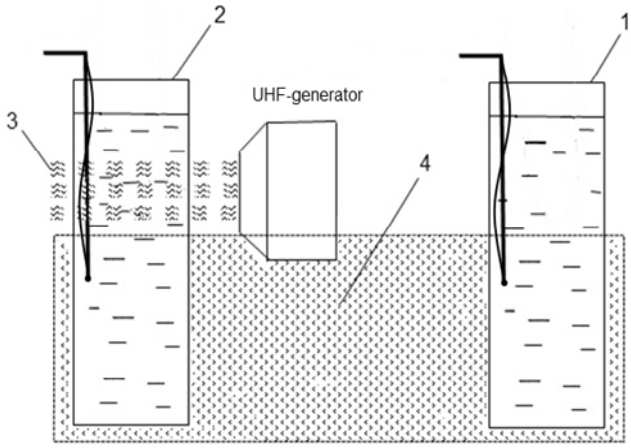


Fig. 1. The scheme of experiments which are conducted directly in the solution: 1 - investigated test tubes with glucose; 2 - control test tubes with glucose; 3- UHF – radiation; 4 - rack of plastic foam.

In the second phase of work investigations are conducted with solutions of glucose and alcohol placed in a closed vessel. In this way dependence on characteristics of random processes from the properties of solutions are determined. Scheme of remote measurement of concentrations on Fig.2.

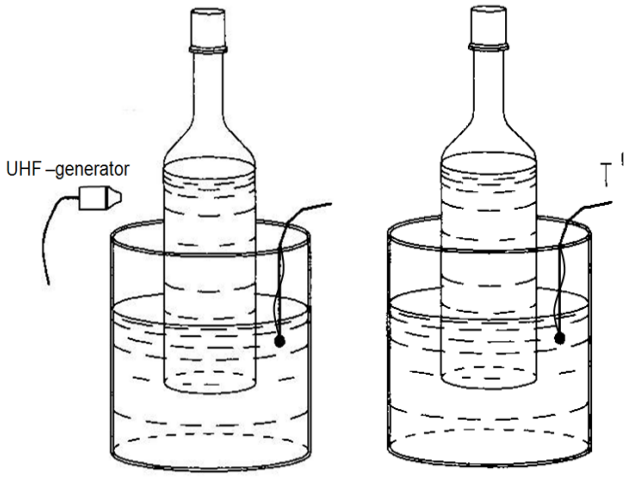


Fig. 2. Scheme of remote measurement of concentrations.

As model measurement objects used vessel of alcoholic drinks where content of ethyl alcohol and glucose is measured. These substances are selected for the reason of simplicity control of concentration of alcohol and sugar by standard techniques. The results obtained experimentally are compared with the results obtained by standard method. Part of experimental results are given at tab.1.

TABLE I. PART OF EXPERIMENTAL RESULTS FOR ALCOHOLIC DRINKS

Alcoholic Drinks	*Alcohol (measured by new method)*	*Sugar (measured by new method)*	*Alcohol (measured by standart method)*	*Sugar (measured by standart method)*
"Martini"1	13,8%	–	14,1%	–
"Martini"2	14,7%	–	14,9%	–
Wine "Kadarka"	11,1%	32 g/dm^3	11,5%	39 g/dm^3

The following table shows the dimensions of units that commonly used: in percent of alcohol and sugar in grams per cubic decimeter. The results of remote measurements are consistent with the direct measurement of solution parameters. The results obtained from the measurements are compared with a pre-calibration measurements carried out with solutions of known concentrations. Thus for realization remote method of measuring and monitoring concentration of solutions in closed vessels its necessary to impact on investigated solutions by external noise signal which does not change the temperature of the solution and measure fluctuations of the differential temperature of water that contains vessel with solution. Subsequently parameters of differential temperature fluctuations are calculated and compared with reference parameters [7].

After repeating these methods of measuring concentrations of solutions started research method noninvasive measuring glucose concentration in human blood by measuring noise differential temperature distilled water with dipped patient's finger. It allows to avoid necessity of blood sampling and consequently the possibility of infected patients. Noninvasive measurement scheme presented on Fig. 3.

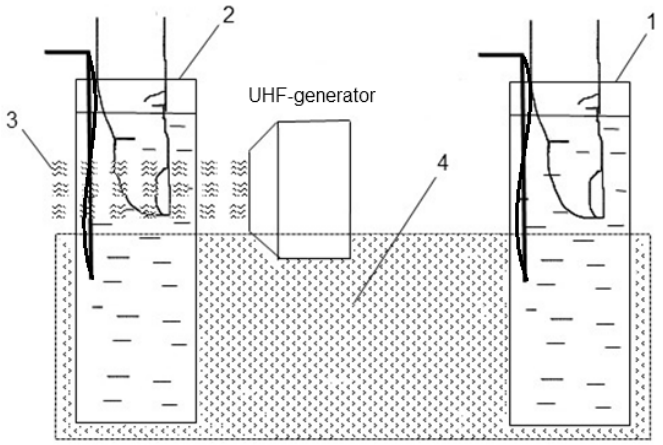


Fig. 3. Scheme of non-invasive measurement of glucose in human blood: 1- investigated test tubes with water; 2 - control test tubes with water; 3- UHF – radiation; 4 - rack of plastic foam.

It should be noted that differential temperature fluctuations of distilled water are measured when the finger of each hand sinks into a vessel with water under the action of external the physical factor of extremely low intensity. Parameters of a random process are calculated with values that exceed the set level value. The values obtained of the relative frequencies of Fourier transform coefficients or the relative periods of the autocorrelation functions are compared with previously obtained reference values for these quantities of glucose in the blood serum of respective concentrations. Concentration of glucose in the human blood is determined according to these results. The results obtained by the present method have difference about 0.3 mmol/l with the results obtained by electrochemical portable glucometer "OneTouch Select"[8] and are presented in tab. 2.

TABLE II. COMPARATIVE TABLE WITH RESULTS OF THE GLUCOSE CONCENTRATION IN HUMAN BLOOD

Patient Number	**Concentration of Glucose in the Human Blood**	
	Concentration of glucose in the human blood measured by glucometer "OneTouch Select", mmol/l	***Concentration of glucose in the human blood measured by new method, mmol/l***
№ 1	10.9	10.8
№ 2	7.7	7.5
№ 3	6.0	6.3
№ 4	5.5	5.5
№ 5	5.4	5.3

V. CONCLUSION

New non-invasive method for determining the concentration of glucose in the blood is developed.

Method is based on using properties of water and aqueous solutions change its spontaneous oscillations depending on the presence of impurities and their concentrations.

Accuracy is determined by calibration dependence obtained for the reference to appropriate solutions.

The reference dependence is based on the measured values of multiple measurements in order to reduce the overall error.

Relative error for the results presented in this paper does not exceed ±0.3 mmol/l.

REFERENCES

[1] WHO, Fact sheet N°312, September 2012 ftp http://www.who.int/mediacentre/factsheets/fs312/en/
[2] M. Balabolkin, Endocrinology, Moscow: "Universe Publishing", 1998.
[3] P. Watkins, "ABC of Diabetes," BMJ Books, Fifth edition, p. 108, December 15, 2002
[4] Z. Fisher, "Statistical theory of liquids", Moscow, publishing phys. & math. lit., p.280, 1961.
[5] N. Wiener "Nonlinear Problems in Random Theory", MIT Press, p. 138, August 1966.
[6] P. Loshitskiy, V. Mamaev "Investigation of changes in properties of water and aqueous solutions of sodium chloride under the action UHF - radiation of nonthermal intensity", Medical Informatics and Engineering, №1, 2008, pp. 53-60.
[7] P. Loshytski, D.Mynzyak "Investigation of Noninvasive Diagnostic Methods and Therapy", Medical Informatics and Engineering, №2, 2012, pp.56 – 59.
[8] O. Davies, L. Christopher, M. Alvarez-Icaza, Measurement of substances in liquids U.S. Patent 6,733,655, filed Mart 08. 2000 and issued May 11. 2004.

Influence of Non-Thermal Intensity EHF Radiation on Properties of Water and NaCl Aqueous Solutions

D.Y. Mynziak, P.P. Loshitskiy
Department of Physical and Biomedical Electronics
National Technical University of Ukraine "KPI"
Kiev, Ukraine
minzyak@mail.ru, pepel@phbme.ntu-kpi.kiev.ua

O. Korostynska, A. Mason and A. I. Al-Shamma'a
Built Environment and Sustainable Technologies (BEST)
Research Institute, School of the Built Environment,
Liverpool John Moores University
Byrom Street, Liverpool, L3 3AF, UK
o.korostynska@ljmu.ac.uk

***Abstract* — This paper reports on the investigation of concentration and time-dependent properties of water and NaCl solution under the effect of the electromagnetic fluctuations of extremely high frequencies (EHF). The maximum sensitivity of physiological solution to external physical effects is shown.**

Keywords—water; NaCl aqueous solution; EHF radiation; dielectric constant; biological systems

I. INTRODUCTION

Water and aqueous solutions of mineral salts, proteins and other organic compounds are the main building blocks of the biological objects. Their prevalence in nature and pivotal role in functioning of the living organisms make aqueous solutions the object of research in many scientific areas [1]. One of the important aspects is the investigation of the influence of the external physical phenomena on the properties of aqueous solutions.

NaCl solutions are one of the most studied systems from their thermodynamic and dielectric properties point of view [1-3]. Therefore, these solutions can be used as a model system for investigating influence of external physical phenomena, in particular, of the radiation with extremely high frequencies (EHF) of non-thermal intensity. Aqueous solutions of NaCl belong to the group of strong electrolytes are their behaviour is described by Debye–Hückel theory, that is based on the assumption of total dissociation of strong electrolytes into ions and accounts for overall interaction of each ion with all other ions sounding it [4, 5]. To account for the effect of the solvent, the theory contains the concept of macroscopic dielectric constant. Around an ion a cloud with the opposite charge is formed. Therefore each point of the solutions experiences total effect of all these ions. The radius of ionic cloud is a characteristic length, which determines some statistical sphere that surrounds central ion. The existence of ionic atmosphere (cloud) distinguished real solutions from the ideal ones. Characteristic length depends on dielectric constant, concentration and ion charge [6].

This theory is based on the assumption that ion has no dimensions and considers a solvent as non-structured media with average value of dielectric constant. It cannot describe the microstructure of electrolytes aqueous solutions as well as explain a number of properties of aqueous solutions, in particular abrupt decrease of NaCl diffusion coefficient in aqueous solutions of small concentration [7], which probably is due to structural changes in the solution.

This work aims to investigate the changes in properties of water and NaCl aqueous solutions in various concentrations under the influence of EHF radiation of non-thermal intensity by using a novel approach that allows determination of fluctuations of fluid differential temperature [8].

II. EXPERIMENTAL PROCEDURES AND METHODS

As a test media, reagents with at least 99% purity and deionised water with more than 99.9% chemical purity were used for the experimental studies. The solutions with required concentrations were prepared by dissolving certain amount of NaCl in a fixed volume of deionised water. Weight concentration (C) of NaCl in solution was calculated using (1):

$$C = {M_s}/{M_P} * 100\% \quad (1)$$

where $M_P = M_w + M_s$ is the mass of the solutions, with masses of water M_w and salt M_s measured by weighing on analytical scales with 0.001 g accuracy.

Experimental investigations on dependence of noise dispersion of differential temperature between the control and measured samples on NaCl aqueous solution concentration and assessment of influence of UHF radiation on these solutions were conducted using a specifically developed laboratory apparatus, which consists of a micro-thermocouple with digital processing of measured values. The apparatus has the absolute uncertainty of measurements is 0.03 °C and provides for recording data with 2-3 s period. Detailed description of measurements methods, metrological characteristic and calibration parameters of the apparatus are given in [9]. Measurements of aqueous solutions parameters were taken at a room temperature of 23 ±1 °C. Apparatus "Oratoria - IVM" was used as a source of UHF radiation and it provided generation of broadband noise in 53-68 GHz working range with noise intensity spectral density $S=10^{-19}$ W/Hz*cm^2 and integral power $P=10^{-9}$ W/cm^2, and has amplitude modulation with 6-18 Hz random frequency. The presence of random modulation allows reduction of irradiation time and, moreover, considerably increases repeatability and reproducibility of the results. The emitter (except for an antenna) and power supply and control unit were shielded and grounded.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

III. Results and Discussion

To determine which concentration of NaCl in aqueous solution is the most sensitive to the influence of UHF irradiation, the experiments were conducted for the solutions with the following concentrations: 0% (deionised water), 0.5%, 0.9% (physiological solutions), 1%, 2%, 3%, 4% and 5%. Each experiments in this series consists of three stages: (1) 10 min without irradiation (control sample), (2) 10 min under the UHF irradiation and (3) 10 min after the influence to assess the consequences of UHF irradiation influence. The results of recording of the differential temperature in a form of noise temperature dispersion are given in Fig. 1.

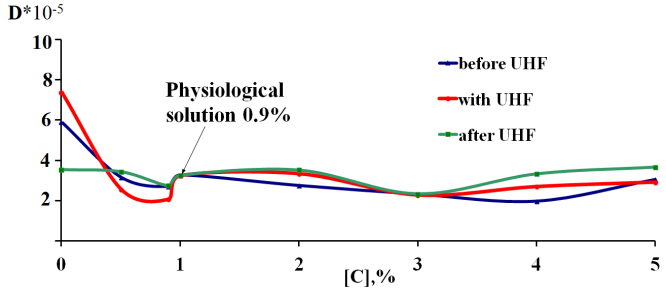


Fig. 1. Noise dispersion in NaCl solutions.

Analysis of the noise dispersion graphs showed that the biggest change in noise dispersion was demonstrated by 0.9 % NaCl (physiological solution).

To investigate the influence of UHF irradiation of various duration on 0.9% NaCl physiological solution, the measurements were taken as follows: 10 min before the irradiation, then for 2, 5, 10, 15, 20, 25 and 30 min UHF irradiation and finally after 10 min after the irradiation procedure. The average algebraic value of differential temperature dispersion in 5 min intervals was chosen as an assessment criterion of the UHF irradiation effect. Accounting for the fact that dispersion depends on the sampling range size, to exclude the results variations due to different size of sampling range, the following approach was used: each stage – before, during and after the irradiation – was separated into 5 min intervals in which the dispersion was calculated, and afterwards these values were processes as algebraic average. The results of noise dispersion processing are illustrated in Fig. 2 and Fig. 3.

As one can see from the above graphs, the noise dispersion curve reaches the maximum value at 15 min after beginning of irradiation (Fig. 2) and in another case at 20 min (Fig. 3).

To investigate the differential temperature noise dispersion under the influence of UHF irradiation on deionised water and 0.9% NaCl physiological solution, a number of experiments were conducted. The example of differential temperature recording in deionised water is given in Fig. 4, while Fig. 5 presents the graph of noise dispersion in NaCl solution.

From the analysis of average dispersion it can be concluded that the noise dispersion reaches its maximum value during UHF irradiation. To eliminate the influence of the external factors on the effect of noise dispersion, nine experiments were conducted were deionised water was effected with UHF. The recordings were made 40 min before UHF irradiation, 40 min during irradiation and 180 after the influence. The 40 min duration interval was chosen based on the previous experiments [8, 9] which revealed that the noise dispersion changed after 15 or 20 min from the start of UHF irradiation, while 180 min time period was chosen based on the assessment on behaviour of the water noise dispersion. All the results of temperature measurements were summed up and the average value of 9 recordings was plotted.

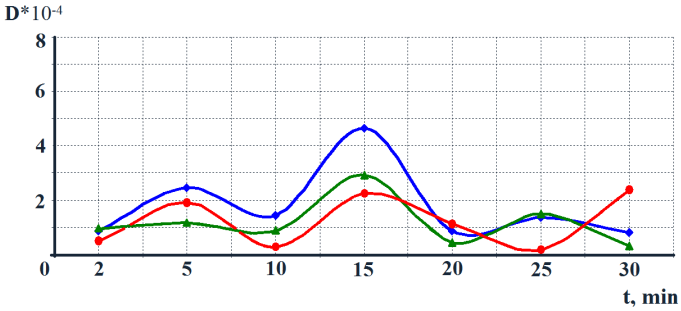


Fig. 2. Noise dispersion in NaCl solution with the maximum value of dispersion at 15 min of irradiation.

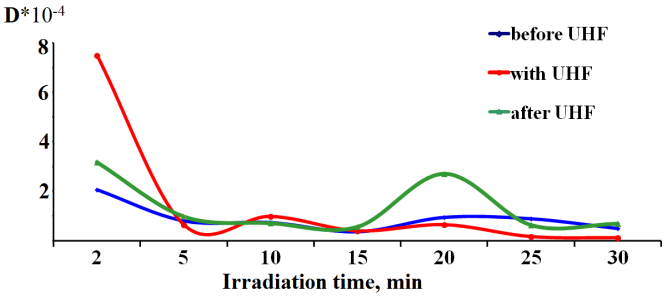


Fig. 3. Noise dispersion in NaCl solution with the maximum value of dispersion at 20 min of irradiation.

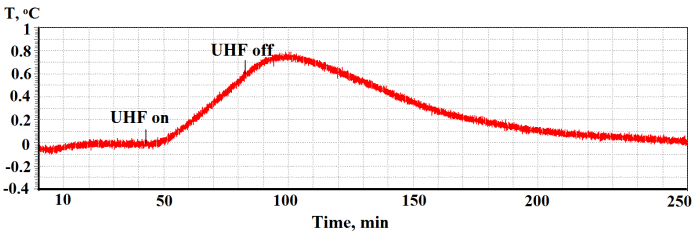


Fig. 4. Example of differential temperature recording in water for 260 minutes: for 40 min before irradiation, for 40 min during irradiation and for 180 min after the UHF was switched off.

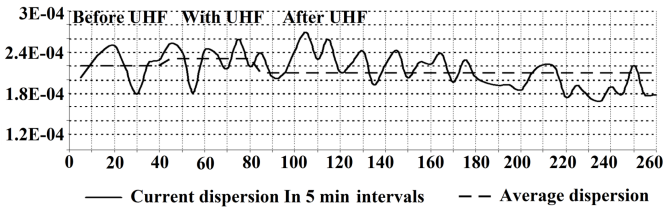


Fig. 5. Noise dispersion in 0.9% NaCl solution recorded for 260 minutes: for 40 min before irradiation, for 40 min during irradiation and for 180 min after the UHF was switched off.

Experiments with 0.9% NaCl solutions were conducted with various time intervals after the influence of UHF irradiation. In the first case, the recordings were made in the same manner as for the distilled water, namely 40-40-180 min. Fig. 6 presents the noise dispersion for 20 recordings. In the second case, for the assessment of noise dispersion dynamics, the time interval after UHF irradiation influence was set to 1000 min, in order to determine the time interval during which the information memory transferred to the solutions is preserved after the low intensity UHF irradiation. Fig. 7 depicts the curve of differential temperature for 1080 min duration, while Fig. 8 illustrates the differential temperature noise dispersion for the same time duration.

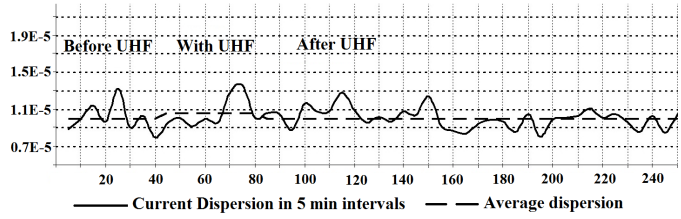


Fig. 6. Noise dispersion in 0.9% NaCl solution for 20 recordings.

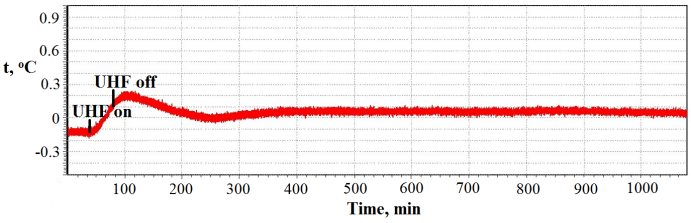


Fig. 7. Recording of differential temperature in 0.9% NaCl solution for 1080 min: for 40 min before irradiation, for 40 min during irradiation and for 1080 min after the UHF was switched off.

As one can see from the dynamics of noise dispersion, after over 13 hours after UHF irradiation the noise dispersion does not change. It can be assumed that "memory" of 0.9% NaCl solution, as well as of deionised water, exceeds this time period.

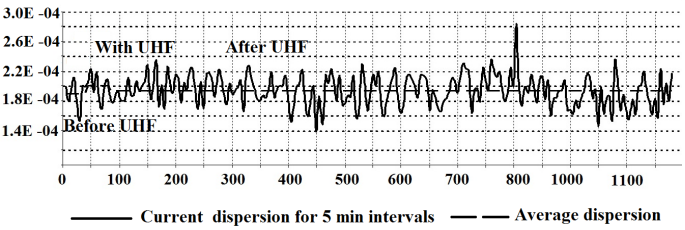


Fig. 8. Differential temperature dispersion in 0.9 % NaCl solution for 1080 min: for 40 min before irradiation, for 40 min during irradiation and for 1080 min after the UHF was switched off.

Therefore, long-term recordings are necessary to investigate the dynamics of water "memory", as supported by some authors in [10, 11], who suggest that water has a memory from few days to few months after the UHF irradiation. From the analysis of noise dispersion graphs it follows that noise dispersion during the influence of UHF irradiation reaches its maximum value, but after this influence the dispersion is less than before it.

The structure of electrolytes aqueous solutions is described by various models, the most realistic of which seems Franck-Evans's model, which accounts for existence of three concentric regions around the solution particles. In the first region, as a result of ion influence the water dipoles are oriented in such a manner that they form fair rigid sheath around an ion. External region is the one where the structure of standard water is completely preserved, and the influence of ion has the effect of dielectric polarisation. Between these regions intermediate region exists, where the structure of water is disturbed by a competitive tension of the two extreme regions. Their influence is mutually opposite. The ions, depending on their charge sign, try to reorient the water dipoles accordingly. The region of standard structure acts on the intermediate region and attempts to preserve the tetrahedral structure [12].

The structure of NaCl aqueous solutions was studied by investigating its dielectric characteristics and irradiation spectra in broad wavelength range from 10 μm to 10 cm [13-16]. These investigations didn't reveal any specific behaviour at small concentrations, including for physiological solutions, namely 0.9 % NaCl in deionised water, and is isotonic in relation to human cells. For the mass concentrations of 4.7 % and 21.7 %, specific behavioural features were observed, which were interpreted and accumulation of separate phases of water-salt solution [13].

Optimised model of hydro-separated ion pair of sodium chloride was given in [17], where the following bonds lengths have the values as given in Table 1.

Table 1. Bonds lengths values for ion pairs.

Ion pair	Bond length
Na^+ - O (H_2O)	0.211
Cl^- - H (H_2O)	0.222
O - H	0.099
O…. H	0.155
O(H_2O)…O(H_2O)	0.254
Na^+ - Cl^-	0.585

This means that in the nearest regions of ions there exists five water molecules, which are oriented according to the ion charge signs. Internal and intermediate sheath are within the radius of R = ½ [0.585 + 0.211 + 0.222] = 0.509 nm. Therefore, standard water surrounds hydro-separate ion pair of NaCl solution with radius R. Hexagonal plat mesh, which presents itself the structure of ice-type water frame, can be transformed into sphere surface or, more precisely, into "icosahedron" of any dimension, by substituting in it in a particular manner, the twelve hexagonal nods into the nods with five coordinations (pentamers) [18].

The surface of the system with six water molecules is $S_6 = A^2 \times 2.598$, while of the system with five molecules it equals $S_5 = A^2 \times 1.785$ (with A=0.254 nm), therefore, depending on the sphere radius R, it is easy to determine the number of molecules necessary to form such a sphere. The assessment of water molecules number, which are present in the second intermediate region, is made by accounting for a difference in volumes of hydro-separated ion pair and its surrounding

sphere. Using this method, each Na^+ - Cl^- ion pair needs 352-358 water molecules. The existence of the top and bottom values is due to the fact that correctness of the water molecules assembly and the presence of Bireme's defects were not accounted for. It was shown [18] that water molecules form fractal-cluster structure. Moreover, the order level of ideal diamond-type fractals is determined by pseudo-periods of triplet cores. The first-level cluster of the first order contains 356 water molecules. Therefore, for the "ideal" solution structure, each NaCl molecule (one ion pair) in solution needs 356 molecules or, in terms of the mass concentration, it corresponds to 0.9 % NaCl aqueous solutions, e.g. to physiological solution.

During the influence on physiological (ordered) solution with external physical factor, in particular with non-thermal intensity UHF irradiation, the processes of acceleration of the water dipoles combining into clusters – accumulation of the cluster energy – cluster breaking take place [19]. The ordered structure of the solutions is sensitive to disturbances, which considerably increase the fluctuations.

IV. Conclusions

Investigation of NaCl aqueous solutions in various concentrations has shown that the most sensitive to the external physical factors if the solutions with 0.9 % NaCl concentration, which corresponds to the physiological solution and concentration of NaCl in human blood. As the results of non-thermal intensity UHF irradiation influence on the water and physiological solution, the average values of differential temperature noise dispersion increases during the irradiation, and decreases after the irradiation, with the dispersion value being less as compared to non-irradiated solution. Water is very different from simple liquids formed by hydrate compounds of the sixth group of the periodic system, and most other liquids. One of the interesting properties of water is the "memory" effect, its means that water keeps properties for a long time after interaction with the external influences. Disordered motion of the liquid contributes to the temperature fluctuations, which depend on the impacts of high frequency. High frequency is able to increase the stability of the parameters of these oscillations.

Acknowledgment

This work is in part financially supported by the European Community's Seventh Framework Programme through the FP7-PEOPLE-2010-IEF Marie-Curie Action project 275201, Water-Spotcheck.

References

[1] K. M. Callahan, N. N. Casillas-Ituarte, M. Xu et al., "Effect of magnesium cation on the interfacial properties of aqueous salt solutions," Journal of Physical Chemistry A, vol. 114, no. 32, pp. 8359-8368, 2010.

[2] L. A. Ferreira, P. Parpot, J. A. Teixeira et al., "Effect of NaCl additive on properties of aqueous PEG-sodium sulfate two-phase system," Journal of Chromatography A, vol. 1220, pp. 14-20, 2012.

[3] E. Jager, A. Jager, T. Etrych et al., "Self-assembly of biodegradable copolyester and reactive HPMA-based polymers into nanoparticles as an alternative stealth drug delivery system," Soft Matter, vol. 8, no. 37, pp. 9563-9575, 2012.

[4] B. Maribo-Mogensen, G. M. Kontogeorgis, and K. Thomsen, "Comparison of the Debye-Huckel and the mean spherical approximation theories for electrolyte solutions," Industrial and Engineering Chemistry Research, vol. 51, no. 14, pp. 5353-5363, 2012.

[5] D. M. Zuckerman, M. E. Fisher, and S. Bekiranov, "Asymmetric primitive-model electrolytes: Debye-Huckel theory, critically, and energy bounds," Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, vol. 64, no. 1 I, pp. 011206/1-011206/13, 2001.

[6] A. I. Levin, Theoretical Fundamentals of Electrochemistry, Moskow: Mettalurgy, 1972.

[7] A. A. Furman, M. P. Beldu, and I. N. Sokolov, Salt, Moskow: Chemistry, 1989.

[8] P. Loshitsky, and V. Mamaev, "Investigation of changes properties of water and aqueous solutions of sodium chloride by the action EHF emission, nonthermal intensity," Medical Informatics and Engineering, vol. 1, pp. 53 - 60, 2008.

[9] P. P. Loshitsky, and V. N. Mamaev, "Investigation of low intensity thermal influence of stochastic electromagnetic fluctuations of extremely high frequencies on physiological solution," Electronics and Communications, vol. 2, no. Special Issue "Problems in electronics", pp. 46-50, 2007.

[10] V. D. Iskin, Y. V. Zavgorodnuy, N. M. Yatsenko et al., Biological effect of mirowaves, Moskow: Biophysics, 1987.

[11] L. B. Oriabinskaya, P. P. Loshitsky, V. C. Mosienko et al., "Structured water and its role in biological systems functioning," Electronics and Communications, vol. 1, no. 6, pp. 183-189, 1999.

[12] V. V. Silykov, Structure of single-atom fluids, water and aqueous solutions, Moskow: Science, 1976.

[13] A. N. Romanov, "Influence of mass mineral salts concentration on dielectric characteristics of their aqueous solutions in microwave region," Radiotechnology and Electronics, vol. 49, no. 10, pp. 1235 - 1242, 2004.

[14] A. Y. Zacetsky, A. C. Lileev, and A. K. Liaschenko, "Dielectric properties of NaCl aqueous solutions on EHF region," Journal of Inogranic Chemistry, vol. 39, no. 6, pp. 1035-1040, 1994.

[15] A. K. Liaschenko, T. A. Novskova, A. C. Lileev et al., "Rotational motion of water molecules in hydrant ions shealths and broadband dielectric spectra of the electrolytes solutions," Journal of Physical Chemistry, vol. 67, no. 8, pp. 1615 - 1622, 1993.

[16] B. M. Liberman, and V. I. Gayduk, "Calculations of the dielectric and illumination spectra of electrolytes aqueous solutions in broad wavelength range. Hybrid model.," Radiotechnology and Electronics, vol. 44, no. 1, pp. 97-103, 1999.

[17] V. A. Shaposhnik, "Kinetic theory of electrolytes aqueous solutions," Vestnik VGU: Chemistry, Biology, Farmocology, vol. 2, pp. 81 - 85, 2003.

[18] N. A. Bulenkov, "On the possible role of hydrotation as the leading integral factor in biosystem organisation at various hierarchy levels," Biophysics, vol. 36, no. 2, pp. 181 -243, 1991.

[19] P. P. Loshitsky, M. V. Kurik, and N. A. Nikolov, "Reaction of aqueous solutions to the influece of low-intensity physical factors," Electronics and Communications, vol. 16, pp. 80-84, 2002.

The Analysis of Technical Solutions for Medical Ozonators

Sokol E.I., Kipenskiy A.V., Kulichenko V.V., Tomashevskiy R.S., Barkhotkina T.M.
Dep. "Industrial and Biomedical Electronics"
The National Technical University "Kharkiv Polytechnic Institute"
Kharkiv, Ukraine
kavkpi@ukr.net

***Abstract* – The article deals with the basic principles of control systems construction in the medical ozone generators. The authors present and analyzes the specific diagram of the electrical and pneumatic part of ozonator. According to analysis results the authors give recommendations for the considered systems realization in the medical ozone generators.**

Keywords – ozonator, discharge chamber, control system, flow rate control, ozone concentration.

I. Introduction

Ozone (O_3) is an allotropic modification of oxygen with a pungent aroma, its molecule consists of three oxygen atoms and has an asymmetric structure of triangle. Ozone is highly soluble in water, it is a strong oxidant and quickly decomposes at ambient temperature. Impact of high concentrations of ozone is dangerous, it leads to the defeat of the respiratory tract, lungs and mucous membranes. Ozone is used in medical practice as an ozone-oxygen mixture (OOM) with different levels of ozone concentrations (OC) in it. The method of using OOM for medicinal purposes is called ozone therapy. Healthiness properties of ozone have been known at the beginning of the last century, but the development of the method has begun only 35-40 years ago. The OOM has a wide range of positive effects on humans; it has bactericidal, antiviral, antifungal, immunomodulatory, antihypoxic, cytostatic, disintoxicational properties. It is high tolerated and side effects are practically absent. Ozone therapy is used in surgery and therapy for eye diseases, urology, obstetrics and gynecology, dermatology, dentistry, and for sexually transmitted infections, diseases of the upper respiratory tracts and lungs, neurology, anesthesiology, pediatrics, reanimation, intensive care and cosmetics. Medical use of OOM shortens treatment time, reduces mortality and disability [1].

However, the use of OOM for medicinal purposes has a high profitability compared to drug treatments because consumable materials are oxygen and small electricity consumption. Methods of ozone therapy can be conventionally divided into three groups depending on the use of OOM: external impact, parenteral methods of exposure, enteral methods. [2].

Each of the methods is characterized not only by a certain sequence of actions, but also by the parameters of OOM, to which OC in OOM and flow of OOM are related [3]. For example, OOM with OC from 2.5 to 100 mg / l, oil – from 5 to 30 mg / l are used for bubbling the distillate and saline solutions. OOM with a concentration of 1 to 10 mg / l is used for flow gassing. OOM with a concentration of 10 to 40 mg / l is used in large and small autohemotherapy. OOM concentration of 5 to 25 mg / l, rectal insufflations from 10 to 60 mg / l is conducted with subcutaneous injections. The range flow control of OOM is from 0.1 to 2 l / min at preparing for the ozone therapy treatments. Therefore, OC in OOM should be regulated in the range of 0.1 to 100 mg / l, i.e. 1:1000, and the flow of OOM – from 0.1 to 2.0 l / min, i.e. 1:20 to assurance the implementation of all methods of ozone therapy.

The purpose of this paper is to analyze the various technical solutions of medical ozonators, that allow to regulate the output parameters of the ozone-oxygen mixture.

The principle of operation of the most medical ozonators is based on the receiving of ozone from medical oxygen with the help of electrophysical method that consists in creation of a volume barrier discharge in gas gap of the discharge chamber (DC), to which a high-voltage alternating current is attached. The sources of oxygen for ozonator are balloons with medical oxygen or oxygen network of medical institutions. Because at the output of ozonator there is a necessity to get OOM with two adjustable parameters, it is obvious that such regulation should be carried out in two control systems: electrical control system (ECS) to regulate OC in OOM and pneumatic control system (PCS) to control the flow of OOM. In this case, each system must contain informational part (IP) that collects information, its transformation, processing and presentation. In addition, each control system must be provided for the executive part (EP). In ECS EP provides to convert the parameters (voltage, current, frequency) of electrical energy from a power line, and its distribution. Thus, in the ECS EP manages energy flow. In PCS EP provides control parameters (pressure, flow) of gas coming from the cylinder or oxygen network, i.e. manages the real flow.

II. Electrical Control System

The analysis of the operation conditions of such control systems has shown that the stability of the OOM formation and its flow rate would affect a number of disturbing factors. With respect to the electrical control system, main factors are the deviation from the nominal voltage value u_{PS} of power supply, pressure and temperature of OOM in the discharge chamber. In relation to the pneumatic control system – the oxygen pressure

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

at the input of ozonator, resistance of its pneumatic channel and the external load.

In the process of regulation of OC in OOM it is necessary to modify the amplitude-frequency-time parameters of alternating voltage u_A at the output of energetic part of the electrical control system [4]. To stabilize OC in OOM with that having disturbing factors, the most simple solution, at first sight, is the solution to use a closed-loop automatic control with the organization of feedback exactly at OC in OOM (OC sensor (S_{OC}) in Fig. 1). Control signal u_C in this system is formed with the consideration of the setting signal u_S corresponding to OC, and feedback signal u_{FB} at the output from the S_{OC}.

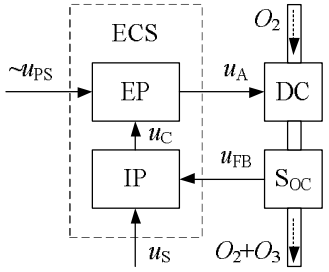


Fig. 1. Structure of ECS with ozone concentration sensor

The analysis of technical means according to the measurement of OC in OOM (optical gas analysis indicators "THE TSIKLON", CJSC "OPTEK", St. Petersburg; photometric meters of ozone concentration "MEDOZONE", "MEDOZONE LTD.", Moscow; optical analyzers of ozone concentration "ICO", JSC "LEPSE" Kirov, etc.), has shown that they mostly use photometric method for the ozone determination on resonance-absorption band at wavelength λ= 253.7 nm. This method is justified for measuring devices. However, it use in automatic control systems in real-time, it makes some difficulties. Firstly, it is inertia, and secondly – the need for periodic gas supply, which does not contain ozone in the optical cell to calibrate measuring means, and thirdly – the availability of sufficiently enough bulky resonant source of ultraviolet radiation with high temporal stability of power and wavelength λ = 253.7 nm.

Another, equally effective approach to the problem of stabilization of OC in OOM is the fact that the electrical system control provides compensation for disturbing factors impact. Of all the factors that have the greatest influence on the ozone formation, the most significant is the instability of voltage of power line [5].

In order to ensure the control alternating voltage on the discharge chamber in the energetic part of ECS, typically, the following types of electric energy conversion are used: the conversion of alternating voltage of power line to constant voltage, constant voltage regulation, and constant voltage conversion to alternating high voltage.

The conversion of the first type (from alternating voltage to constant voltage) is performed using uncontrolled rectifiers (R) or controlled rectifiers (CR). Capacitor filters (F) are commonly used in order to smooth the surge control of voltage capacitor. When using uncontrolled rectifier the regulation of rectified (constant) voltage is realized through pulse converter (PC) with the frequency or length modulation and LC-filters (F). The voltage inverters (VI) are applied to convert constant voltage to alternating voltage with frequency of 1-2 that is higher than frequency of power line voltage. The increase of the output voltage inverter to the desired level is achieved by step-up transformer (ST).

The easiest way to compensate the effect of changes in the alternating voltage of power line on average magnitude of voltage u_{PS} and at the discharge chamber consists of introducing a feedback from the output voltage of the inverter (output voltage sensor (S_{OV}) Fig. 2) in automatic control inverter. Whereas the signal u_C of control inverter will be formed in the informational part of ECS with the reference of the signal task u_S and feedback signal u_{FB} that come from the input voltage sensor (S_{IV}). This approach allows to solve the problem, but excludes the possibility of regulating the amplitude of the alternating voltage u_A, which does not allow to adjust OC in OOM at more than 1:50. It does not allow to implement the full range of methods of ozone therapy [2, 3].

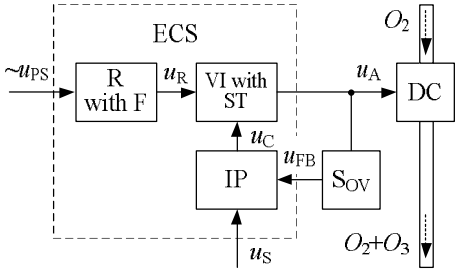


Fig. 2. Structure of ECS with output voltage feedback loop

The controlled rectifier with a filter (CR with F) is used to convert alternating voltage of power line to a constant voltage; the rectifier can have the regulation function of rectified voltage u_R (see Fig. 3). Here the control signal $u_{C1}$of rectifier is formed in respect to the signal u_{FB} coming from the input voltages sensor (S_{IV}) (for rectifier), that allows to compensate the effects of voltage changes u_{PS} at the input voltage inverter [6]. The control signal u_{C2} of voltage inverter with step-up transformer (VI with ST) is formed with the voltage task u_S that allows you to get at the output of the inverter u_A voltage with frequency-time parameters, providing the target values of OC in OOM.

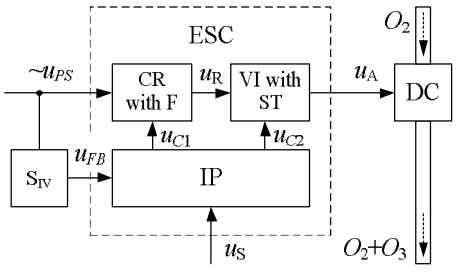


Fig. 3. Structure of ECS with input voltage sensor

However, with the need to regulate OC in OOM in the range 1:800, that is required for medical ozone generator, there is a need to control the rectifier not only to stabilized rectified voltage (i.e., compensate the impact of voltage instability of power line), but to change it in a certain range. This, in turn,

leads to an increase in ripple ratio voltage at the input of the inverter by 20-40% as compared to the ECS, which is shown in Fig. 2. The presence of pulsations complicating management processes, and consequently, reduces the effectiveness of ECS to achieve a given concentration of ozone in the OOM. Providing acceptable ripple ratio voltage ($r \le 10\ \%$), by reducing the range of the output voltage of the rectifier, control range of OC in OOM can increase only up to 1:100.

The control system with a pulse converter of constant voltage and filters circuit (PC with F) has larger possibilities to regulate and stabilize voltage u_{RS} on the discharge chamber, which is shown at Fig. 4. Here is information about the voltage changes of power line, and, consequently, the rectified voltage u_R, information enters the information part as the signal u_{FB} from the input voltage sensor (S_{IV}) (for pulse converter) [7]. The formation of the control signal u_{C1} with the reference to the setting signal u_S and the feedback signal u_{FB} at the input voltage allows you to regulate the output voltage pulse converter u_P and ensure that it is invariant to input voltage changes u_R. With ECS (Fig. 4) could be regulation of OC in OOM between 1:800 and more.

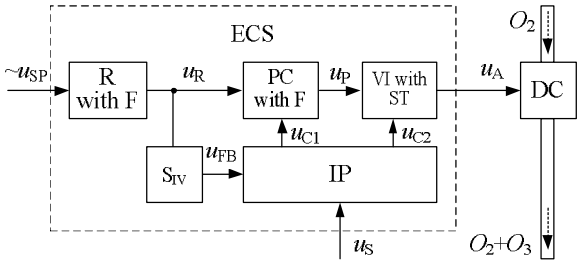


Fig. 4. Structure of ECS with pulse converter

The influence of such factors as OOM temperature can be significantly reduced, while ensuring the efficient cooling mode of the discharge chamber. OOM pressure changes are so insignificant that no essential changes occur in OC when those changes occur with the real values of the external load resistance of ozonator, especially with the use of degassers.

Thus, the main task of the electrical control system can be defined as the regulation and stabilization of amplitude-frequency-time parameters of the voltage on the discharge chamber, which can be achieved through the introduction of appropriate voltage feedback.

III. Pneumatic Control System

OOM flow through the ozonator in the general case can be defined by the expression

$$Q = k \cdot S^2 \Delta P / L \qquad (1)$$

where k – coefficient of proportionality;
S – cross-sectional area of pneumatic ozonator tract;
ΔP – pressure drop at the edges of pneumatic tract;
L – length of pneumatic tract.

From the above expression follows that to change the flow of OOM through pneumatic ozonator tract is possible with the help of two options. The first option is by changing the cross sectional area of pneumatic tract, the second option is due to changes in oxygen pressure at the input of DC. To solve such problems in pneumatic control systems pressure regulators are typically used to regulate and stabilize the gas pressure at its output, faucet, valves and chokes, which provide cross-sectional change in pneumatic tract.

The scheme of the pneumatic system that implements the first control version is shown in Fig. 5. Here oxygen from an oxygen cylinder (OC) via a gear (G) transferred to the input of the pneumatic control system (PCS). In this system executive part is in the form of the faucet (F), which allows you to manually regulate the cross sectional area of pneumatic tract of ozonator and informational part is in the form of a flow meter (FM). Provided that the oxygen pressure at the output gear will not change, the system will provide a given flow rate of OOM through the discharge chamber (DC) for the load with some constant pneumatic-resistance, which will determine the value of the pressure at the output of pneumatic tract of ozonator.

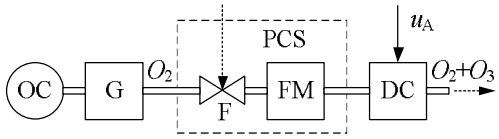


Fig. 5. Structure of PCS with manually regulation of the flow

However, changing the method of using ozone (eg, in the transition from ozonation of saline to fumigation of limb in a plastic bag) pneumatic resistance tract can vary by 4-5 times which increases the output pressure ozonator to 0.7-1 bar, therefore, to reduce the flow of OOM by 30-50%.

The solution of the regulation task and stabilization of the oxygen pressure at the input of the discharge chamber is possible through by performing in the form of IP an automatic pressure adjustment (APA) with a feedback (Fig. 6).

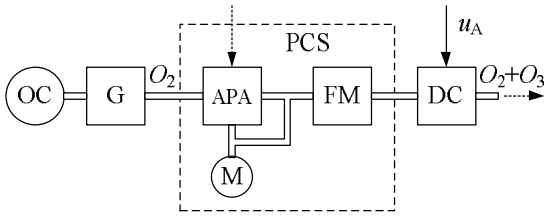


Fig. 6. Structure of PCS with automatic pressure adjustment

The velocity of OOM is set manually and monitored by FM. Manometer (M) is here an auxiliary tool for visual monitoring of oxygen pressure with its regulation in the manual mode.

The considered variants of PCS (Fig. 5-6) are quite viable. Some ozone generators (apparatus for the gas ozone therapy "KVAZAR", OC of the measuring devices of "KVAZAR", Nizhny Novgorod, synthesizer device of gas OOM a-s-GOKSF - 5-02-OZONE, OJSC "LEPSE" Kirov and others) are applied with such systems to this day. However, current trends of medical ozonizer-construction are aimed at the automation of all processes in the ozonator, including the control flow process of OOM.

The scheme of automatic flow control system is shown at Fig. 7 [8]. Here, the pressure regulator (PR) is switched on in consecutive ages with a pressure stabilizer (PS), which ensures the stability of the pressure drop at the edges of the pneumatic ozonator tract. Discrete regulation of the flow control of OOM according to binary principle is performed by throttle-valve block (TVB), which is controlled by the signal u_C from the informational part of OOM. Manual setting of regulator and pressure regulator, and also throttle-valve block is conducted at the manufacturer only during the system setup.

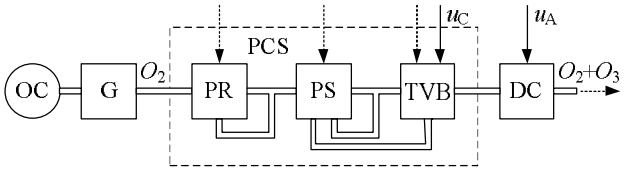


Fig. 7. Structure of PCS with throttle-valve block

In this case, by combining informational parts of pneumatic and electric systems, it is possible to set the flow rate of OOM and ensure its stabilization with an accuracy of 5%, as in deviation of oxygen pressure, as at a of oxygen pressure at the inlet to 50% and load resistance change in the whole real range [9]. However, the mass-dimensional parameters of throttle-valve block leave much to be desired. That is why such PCS should be used only in cases where, strict dosing of OOM flow is necessary to stabilize all significant disturbing factors.

It is possible to improve the mass-dimensional parameters of PCS by using methods of pulse modulation to control the electromagnetic valve (EMV), which is installed in the pneumatic ozonator tract (Fig. 8). The control signal u_C of EMV is formed in the informational part of the electrical system control with the reference of signal task of OOM flow coming from the flow sensor, which is used as a differential pressure sensor (S_{DP}). Receiver (Rec) serves here to smooth the pressure fluctuation at the output of EMV and thus the assurance of flow uniformity of OOM.

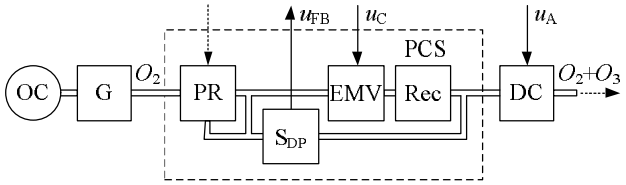


Fig. 8. Structure of PCS with pulse modulation

The tests of the system have shown that it is able to regulate and stabilize the flow of OOM with an accuracy of 10 % when the oxygen pressure at the inlet to the ozonizer to 50% and change the air resistance throughout the whole range.

In any variant of the realization of the pneumatic control system, it is appropriate to use emergency release valve at the input of ozonator to limit the pressure at the input of PCS and force reset valve for oxygen venting from the pneumatic tract before disconnecting the ozonator from the source of oxygen. In addition, it is useful to apply a filter to clean the oxygen of solids and humidity [10].

IV. Conclusions

In conclusion, it should be noted that in choosing of any construction variant of control systems by medical ozonator, is necessary to proceed from the conditions of relevancy and feasibility. During the universal ozonator development which is intended for research and development and new methods of ozone therapy, it should be provided a considerable range of opportunities for regulation and stabilization of OC in OOM and its flow rate. For this purpose the electric control system with a pulse transmitter is most suitable (see Fig. 4). Pneumatic system in this case is better to construct according to the third or fourth variants (Fig. 7 and 8, respectively). For more simple and repetitive manipulations in ozonators, it is better to use simpler technical solutions, such as the second or third variants of ECS (Figs. 2 and 3) and the second variant of PCS (Fig. 6). This approach will significantly reduce the cost of the ozonator. To regulate the OC in OOM it may be justified to use the first variant of ECS (Fig. 1), provided that the OC sensor will have good mass-dimensional parameters, adequate performance and reliability. It is also important, to our opinion, the production of specialized ozonators for physicians of various specialties, in order to avoid redundancy of technical solutions and thus, reduce the cost of products at high enough quality.

References

[1] “Nizhny Novgorod Medical Journal. Appendix: Ozone therapy. – N. Novgorod, 2003, - 344 p. (In Russian)

[2] O. Maslennikov, K. N. Kontorshchikova “Practical ozone therapy: The Handbook. - N. Novgorod: "Vector-TiS", 2003, - 52 p. (In Russian)

[3] O.V. Maslennikov, K.N. Kontorschikova. “Ozone therapy: The Internal Diseases,”N. Novgorod: “Vektor-TiS”, 2003, - 132 p (In Russian)

[4] E.I. Sokol, A.V. Kipenskiy, A.A. Lashin. “The Specifics of Controlling the Concentration of Ozone in Medical Ozone Generators” // Proceedings of the Conference "New technologies and preformed natural healing factors." - Kharkov: HMPA, 2002. - 229-231 p. (In Russian)

[5] E.I. Sokol, A.V. Kipenskiy, A.A. Lashin. “The compensation of the perturbations in the regulation of the parameters of the ozone-oxygen mixture in medical ozone generators // Nizhny Novgorod Medical Journal. - Nizhny Novgorod, 2003. - 313-315 p. (In Russian)

[6] A.V. Kipenskiy, E.I. Sokol, V.I. Ryabenkiy and others. Patent № 23060 UA, MKV N 02 M 7/12. “The method of the direct microprocessor control of the semiconductor power converter and device for its implementation” № 94043468; declared 11.05.94; published 30.06.98, bulletin № 3. (In Ukrainian)

[7] A.V. Kipenskiy, A.V. Kipenskaya. Patent № 1742964 RF, MKI 5 NO2 M 7/12. The method of controlling of the semiconductor converters DC voltage / № 4762580/07; declared 27.11.89; published 23.06.92, bulletin № 23. (In Russian)

[8] E.I. Sokol, A.V. Kipenskiy, A.A. Lashin and others. “The combined system of control the consumption of ozone-oxygen mixture in medical ozone generators” // International Medical Journal. Appendix: "Ozone therapy." - Kharkov, 2003. - 158-161p. (In Russian)

[9] A.V. Kipenskiy, E.I. Sokol, A.A. Lashin and others. “The characteristics and results of operation of the medical ozone OM 80/1” General Reanimatology. - M.: Research Institute of General Reanimatology, of the Russian Academy of Medical Sciences, 2006. - T. 2. - № 4/1. - 297-301p. (In Russian)

[10] A.V. Kipenskiy, E.I. Sokol, A.A. Lashin and others. “The development of medical ozone OM80/1 and the results of its tests” Herald of Physiotherapy and Health Resort. - Evpatoria: EIRITS, 2006. - № 5. - 76-78p. (In Russian)

The Impact of Magnetic Field on Blood Vessels

M.M. Baran, N.V. Kostevska, V.P. Virchenko, Y.S. Synekop
NTUU «KPI», FEL, Kyiv, Ukraine,
miwko@i.ua

A.I. Mukhomor
Clinical Hospital "Feofaniya", Kyiv, Ukraine

Abstract — **This article deals with the approach of modeling of magnetotherapy influence on biological objects particularly blood vessels. The block scheme and the advantages of trial device sample for magnetic therapy are described.**

Keywords—magnetotherapy; magnetostimulation; magnetic field

I. INTRODUCTION

The usage of magnetic field for different diseases' treatment has been practiced for hundred years, especially during last decades technical tools for magnetic therapy are being developed [1, 2, 3]. Researches have shown that the alternating magnetic fields up to 30 mT impact on vascular system and/or blood microcirculation. They are used successfully in clinical practice in treatment for surgical, gynecological and dental diseases. Magnetic fields (MF) have beneficial effect on muscular and vessel system, decrease the inflammatory processes, and improve blood circulation, facilitate the fixation of several drugs in particular place in organism; increase the effectiveness of such treatment methods as electrophoresis, phonophoresis and photophoresis. Especially they are effective in magnetic therapeutic particles' transportation inside human body, called drug address delivery. Such particles consist of capsule inside which is ferromagnetic or superparamagnetic material covered with medication. With the help of MF capsules can be delivered to the most remote cells overcoming different biochemical barriers, nanoparticles penetrate in depth of cartilage or bone. Therefore it's a great of importance to carry a set of medical and biological model analyze for blood circulation system characterization.

II. THE MAGNETIC FIELDS' IMPACT ON HUMAN BODY

MFs influence on human physical and mental functions impacting the structure of ions movement. In the case of being injured and damaging tissues positive ions go to the injured area and ionic composition ratio changes appears. It causes pain and swelling. For therapeutic effect on this area restoring ions natural negative electromagnetic charge is needed. Then it is important to get rid of pain and electrochemical matters relating to swelling and also to provide with oxygen and other vital substances supply of this area. Normal electromagnetic balance restores by the way of MF influence on injured area.

MF helps to loosen capillary cells and nearest muscles and binding tissues providing additional blood supply. Damaged area gets much oxygen and more nutrients, the painful effect disappears and all chemical substances are removed, formed as a result of swelling. So on this area normal electromagnetic balance is settled, pain relieves and swelling is removed, which in the final analysis forwards recovering process acceleration. Billions of positive ions of such mineral as potassium, natrium and magnesium solving in the hemolymph (blood plasma) go to human vessel system and take part in blood circulation in straight position. Placing the blood circulation system in MF we change the dynamic of ions' movement in blood, making them move in vessels in perpendicular way according to the direction of their movement. it causes "the pushing effect" of blood flow from left and right, what ,in turn, leads to dilatation of vessel cells and as a result to stimulation of blood circulation [4, 5].

III. THE ADVANTAGES AND BLOCK-SCHEME FOR MAGNETIC THERAPY

Formation of MF carries out with the help on trial sample of new magnetic stimulator (MS), designed by authors [6], the prototype of which is MS-92M. Main advantages of trial sample are following:

- modern elements' base;
- small dimensions, weight, low power consumption;
- signal linear distortions are considerably reduced on the device output;
- polar transistors switched on using half-bridge scheme with partial advancing pumping are used instead of bipolar transistors with emitter follower (MS - 92M);
- feedback is implied and it gives an opportunity to control constantly and adjust output signal according to "standard" i.e. recoded in microcontroller (MC) taking into account the enhancement factor;
- Microcontroller allows to:
  - connect MS and PC, and then control system;
  - give commands and specify their parameters;
  - control channels independently;
  - display data;
  - receive oscillations of different frequency, and forms (porosity);
  - send different signals to channels with the phase displacement (with different phase, frequency, and amplitude).

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

On Fig. 1 block-scheme of magnetic stimulator is represented. It consists of power supply unit (PSU) 1, stabilization unit (SU) 2, control unit (CU) 3, amplification and signal forming unit (AU and SF) 4, indication unit (IU) 5.

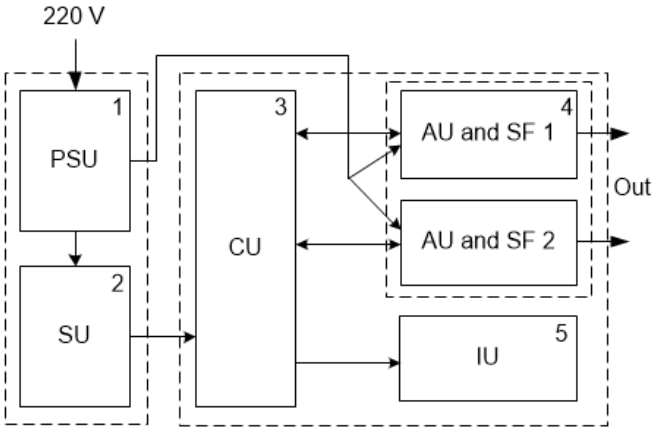


Fig. 1. Block-scheme of magnetic stimulator

IV. Relationship between magnetic field and changes in blood system

The influence of magnetic field on biological objects can be estimated implementing and using as a model basic structural elements organisms consist of, i.e. liquid and solution. Also the influence of liquid itself on biological tissues should be taken into account, for example magnesium sulphate is vasorelaxant, and as sodium chlorine is in blood and tissues' liquid with concentration of 0.5% serves for maintaining osmotic pressure, its solution can be used as model medium in carrying experiments on magnetic field influence on biological liquids [7].

For accurate estimation of magnetic field influence on biological liquids, influence of magnetic field on blood vessels, vascular wall indeed. Let us assume that this vessel is a certain conductor which is put in magnetic field [8]. The resultant force F which influences on boundary by current in magnetic field with magnetic induction B is calculated in this way

$$F = I\oint[dl, B], \qquad (1)$$

where integration is performed along a close path with current I.

Let us consider the influence of this force on human organism, in particular on blood vessels. To the vessel where blood circulates we can easily apply elastic vessel model. One of the main characteristics of vascular system of arterial route is elastic model. Now let find the relationship between elastic module E end volume elastic coefficient of vessel K. In this case we consider vascular wall as the example elastic solid body [9]. Elastic features of wall will be described by Hooke's law for elastic core:

$$(L - L_0) / L_0 = F / ES, \qquad (2)$$

where L_0 – starting length of elastic core, L – final length, S – square of cross-section, F – force which elongate core. If E is bigger than the relative length boost is less. When length of core increases, its thickness decreases ρ in concordance with dependence $\Delta\rho/\rho = \sigma\Delta L/L_0$, where σ - Poisson's ratio. If core volume doesn't change then $\sigma = 1/2$.

Let us consider cylindrical radius tube r_0 and length L (Fig. 2). The thickness of tube equals ρ_0 ($\rho_0 \ll r_0$) and material is characterized by elastic module E. When pressure is absent the tube volume is $V_0 = \pi r_0^2 L$, than $r_0 = \sqrt{V_0 / \pi L}$. If we apply to inner wall pressure P, than tube will widen; radius will increase to r, and thickness of the wall ρ will decrease. This case can be compared with the case of analysis of elastic plate deformation with its length $L_0 = 2\pi r_0$, thickness ρ and width L. In analysis Hooke's law is used.

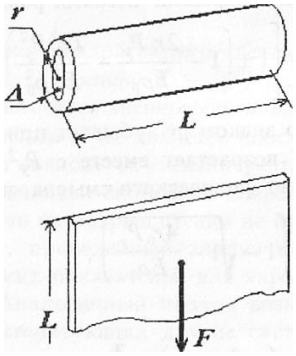


Fig. 2. Model of elastic vessel

Blood pressure on tube walls causes the appearance of tangential force as a result of elastic reaction of wall material. Everybody knows if tangential force f per length unit influences on cylindrical radius membrane r then pressure appears $P_y = f / r$. Pressure P_y applies on tube's inner wall that's why it will widen wall until elastic force is the same as force which influences it. Under pressure P_y length of rectangular plate will change: $L = 2\pi r$.

Relative change of length calculates in this way:

$$(L - L_0) / L_0 = (2\pi r - 2\pi r_0) / 2\pi r = (r - r_0) / r_0. \qquad (3)$$

Force F which influences on all length of plate L calculates as $F = Lf = P_y rL$. Square of plate cross-section also changes $S = \rho L$. Then in accordance with Hooke's law:

$$(r - r_0) / r_0 = P_y rL / EL\rho; r - r_0 = r_0 P_y r / E\rho. \qquad (4)$$

This formula is valid when $\Delta r \ll r_0$, and when r increases wall width ρ decreases.

To simplify the process the tube is fixed on its corners (for preventing changes of L) and $\sigma = 1/2$, we can consider that the whole tube wall volume is constant under the process of widening. Initial wall volume is $V_0 = 2\pi r_0 \rho_0 L$. For others - $r > r_0$ $V = 2\pi r L \rho$. If the volume hasn't been changed then $r_0\rho_0 = r\rho$ *and* $\rho = r_0\rho_0 / r$

We replace ρ on ρ_0 then $r - r_0 = P_y r^2 / E\rho_0$ and

$$r^2 - (E\rho_0 / P_y) r + E\rho_0 r_0 / P_y = 0 . \quad (5)$$

Equation has two roots:

$$r_{1,2} = \frac{E\rho_0}{2P_y} \pm \sqrt{\left(\frac{E\rho_0}{2P_y}\right)^2 - \frac{4E\rho_0 r}{P_y}} = \\ = \frac{E\rho_0}{2P_y} \pm \frac{E\rho_0}{2P_y}\sqrt{1 - \frac{4E\rho_0 r}{P_y}} \quad (6)$$

If P_y not enough then we can decompose root. We will limit just only with three first members of decomposing:

$$r_{1,2} = \frac{E\rho_0}{2P_y}\left[1 \pm \left(1 - \frac{2r_0 P_y}{E\rho_0} - \frac{2r_0^2 P_y^2}{E^2 \rho_0^2}\right)\right]. \quad (7)$$

First root with index "+" decreases when P_y^2 increases, second with index "-" increases together with P_y^2. So in this way first root doesn't have physical sense, that's why:

$$r = \frac{E\rho_0}{2P_y}\left(1 - \sqrt{1 - \frac{4r_0 P_y}{E\rho_0}}\right). \quad (8)$$

and when P_y is small:

$$r = r_0\left[1 + \left(r_0 / E\rho_0\right) P_y\right]. \quad (9)$$

Now let's define the plate volume after changing r:

$$V = \pi r^2 L = \pi r^2 L (1 + P_y r_0 / E\rho_0)^2 = V_0 (1 + P_y r_0 / E\rho_0)^2 . \quad (10)$$

If P_y isn't enough then quadratic term can be neglected, and then :

$$V \approx V_0 + \frac{2\pi r_0^3 L}{E\rho_0} P_y . \quad (11)$$

From carried analysis we understand that under adequate low pressures (and, consequently, low deformations) between P_y and V exists approximate linear dependence. comparing this formula with main formulas of elastic tank theory we can easy get for defining parameters of K such formula:

$$K = E\rho_0 / 2\pi r_0^3 L . \quad (12)$$

it shows that volume elasticity of K is proportional to Young's modulus E the material of blood vessel wall and it depends on initial width of ρ_0 and initial volume V_0. And if then $E = const$ but K isn't. If two blood vessels are analyzed with equal E but differ in ρ_0 and V_0 then their volume elasticities will be different. For example, two vessels with the same volume elasticity K but different sense of Young's modulus.

Let's admit the accounting of quadratic term and formulas for V lead to the opinion that curve $V = f(P)$ must have convexity. Therefore Young's modulus E isn't constant for wall vessel material and depends on the level of wall stretching. If E is bigger then the pressure in vessel is increasing with lower speed until tensile strength won't be exceeded.

CONCLUSION

Thereby the analysis of model allows us to carry a set of medical and biological rates for system characteristics of blood circulation system. Similar approach is possible for data entry which characterizes other organism's systems, respiration system. But according to this great difficulties can appear in choosing appropriate model of this or that functional system or organ. That's why on practice the indexes are entered without special background of theirs reasonability but their self-descriptiveness can be proved experimentally. Proposed by authors device allows us to conduct the research under considerable changes of output signal characteristics.

REFERENCES

[1] G.R. Solovjeva. Magnetotherapeutic equipment. - M: Medicine, - 1991, - 167p. (in Russian)

[2] V.I. Zhulev, O.V.Kirjakov. Management of magnetotherapeutic impact in terms of the activity of regulatory system. // Biomedical technologys and radioelectronics. –2001.–№10.–P.33–39 (in Russian)

[3] E.P. Lobkaeva. Theoretical justification of goal seeks of pulsed magnetic field for achievement stable therapeutic effect. // Biomedical technologys ana radioelectronics. – 2006.-№ 1-2. - P.12-20. (in Russian)

[4] System of complex magnetotherapy : [tutorial for high school] /Edited by A.M. Berkutov, V.I. Dguleva, G.A. Kurajeva, E.M. Proshina. — M. : Laboratory of basic knowledge, 2000. — 376 p. (in Russian)

[5] McKay JC, Prato FS, Thomas AW., "A literature review: the effects of magnetic field exposure on blood flow and blood vessels in the microvasculature", Bioelectromagnetics. 2007 Feb; 28(2):81-98.

[6] M.M. Baran. Development of apparatus for magnetic therapy / M.M. Baran, V.P. Virchenko, Y.S. Synekop, A.I. Mukhomor // Medical technology and engineering. – 2012. - № 2. P. 49 – 52. (in Ukrainian)

[7] O.S. Melnik., Y.I.Yakimenko. Biomedical matherials: Tutorial. – K.: NTUU "KPI", 2000. –228 p. (in Ukrainian)

[8] I.E. Irodov. Main laws of electromagnetics. - M: High school., - 1991, - 288 p. (in Russian)

[9] E.P. Popetchitelev. Methods of biomedical research. System aspects. Zhitomir. 1997. -186 p. (in Russian)

Barrier Properties of Polymer Encapsulation Materials for Implantable Microsystems

Sabine Kirsten, Jakob Wetterling, Jürgen Uhlemann, Klaus-Jürgen Wolter
Technische Universität Dresden, Faculty of Electrical and Computer Engineering, Electronics Packaging Laboratory
Dresden, Germany
kirsten@avt.et.tu-dresden.de

Sergej Zigler
Microelectronic Packaging Dresden GmbH
Dresden, Germany

***Abstract* — Biocompatible polymers utilized for flexible encapsulations of implantable microsystems provide many advantages compared to widely used rigid titanium or ceramic packages. However, polymers alter their properties due to interactions with their environment. As a result the protective function of these materials especially for long-term implants is not reliable. Therefore, we investigated barrier properties against water vapor of silicone and Parylene C membranes. And we combined these polymers to a multi-layer membrane to enhance the protective function of such an encapsulation system. Applying a bonding agent between two polymer layers increases the strength of the sample as well as the barrier properties significantly.**

Keywords—polymer; encapsulation; implant; silicone; parylene C; water vapor transmission rate

I. INTRODUCTION

Biocompatible polymers are increasingly used as encapsulation material for implantable electronic medical devices due to their advantages in weight, space, processing, various chemical and mechanical properties and electromagnetic permeability [1-3]. Especially for systems, which require a wireless communication between implant and control unit outside the human body or a mechanical adaptation of the implant interface to the surrounding tissue, polymer encapsulations are preferred [4-6]. Typical examples are retinal implants, neural prostheses, permanent monitoring systems for vital parameters and smart implants.

Widely used inflexible titanium or ceramic packages do not comply with the new emerging requirements. However, they provide a hermetic housing of the electronic device and therewith a reliable protective function. The protective function of an encapsulation implies on the one hand that the host tissue is not affected by the toxic material used for electronic devices and the encapsulation itself. On the other hand it means that the implanted electronic device is protected against the aggressive biological environment during the intended lifetime. Thus, harmful interactions between the tissue and the implant are suppressed. Metals and ceramics are nearly impermeable to fluids and gases as well as chemical resistant [7]. Their material properties are consistent during sterilization process, storage and implantation, which is another significant reason for using such inflexible encapsulations especially for long-term implants like cardiac pacemakers and cochlear implants. However, polymers interact with their environment altering their chemical, mechanical and electrical properties [8]. They are semipermeable for molecules and ions. Especially moisture uptake leads to swelling, chemical degradation and reduction of electrical insulation and decrease of mechanical as well as adhesive strength [9-10], which impairs the protective function. If the encapsulation material delaminates from the electronic components, water absorbed by the polymer matrix can condensate in this cavities at the interface. As a consequence, corrosion and leakage current could occur and cause failure of the electronic device. Furthermore, toxic degradation products may permeate through the polymer encapsulation affecting the surrounding tissue and activating an inflammation. Especially for long-term implants one should consider the mentioned failure mechanism.

According to this, one approach to enhance the protective function of encapsulation materials was to decrease the permeability of water molecules through the encapsulation. Therefore we investigated the water vapor transmission rate (WVTR) of silicone and parylene C mono-layers. Both materials, separated or combined, are currently used as flexible encapsulation materials for implantable devices [11]. Combining them to a flexible double or triple-layer system, plasma treatment of the interface between the polymer layers was utilized to improve bonding process. We also applied titanium dioxide as a bonding agent to increase the adhesion strength between the polymers in the multi-layer system.

II. MATERIAL AND METHODS

A. Production of polymer membranes

In order to characterize the permeability of water molecules through polymer layers we produced membranes by means of a sacrificial layer process (see Fig. 1). Afterwards, the samples were fixed in a frame of aluminium foil with a 40 mm by 50 mm opening.

For an 80 nm thick sacrificial layer, 80 µl of gelatin solution (RUF Lebensmittelwerk KG, Germany, gelatin concentration in purified water 18 g/l) were spin-coated with Spin-coater CPS 10 (SM Elektronik GmbH, Germany, 15 s at

This work was supported by Europäischer Sozialfond ESF and Microelectronic Packaging Dresden GmbH.

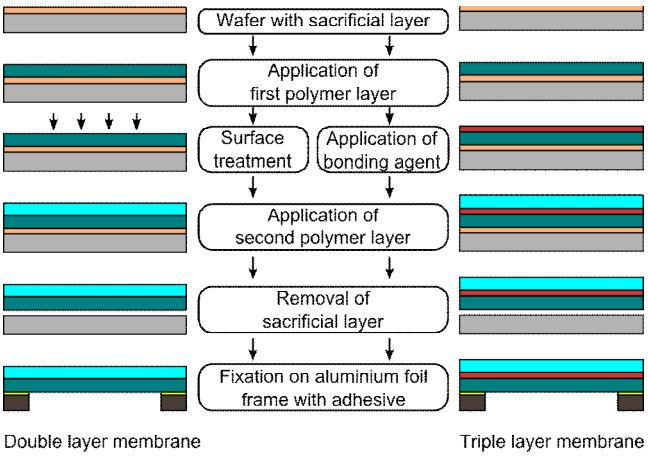


Fig. 1. Steps of multi-layer production of a douple layer (left side) and a triple layer membrane (right side) by means of sacrificial layer process

500 rpm and 30 s at 2000 rpm) on a glass or silicon wafer and dried. Afterwards the particular polymer was deposited on the gelatin covered wafer. First polymer was USP Class VI certified silicone (700 µl, MED6-6606, NuSil Technology, USA, 15 % polymer solution) which was spin-coated (Spin-coater CPS 10, SM Elektronik GmbH, Germany, 15 s at 500 rpm and 30 s at 3000 rpm) and dried for 72 h at ambient conditions. The thickness of this mono-layer was 3.8 µm. Parylene C as the second polymer (di-chloro-di-p-xylylene, Plasma Parylene Systems GmbH, Germany) was deposited in a physical vapor deposition (PVD) process with parylene coating system 2000-150 LV (Plasma Parylene Systems GmbH, Germany) utilizing the adhesion promoter silane A174 (Plasma Parylene Systems GmbH, Germany). Layer thicknesses reached from 5.7 µm to 6.1 µm. To obtain the polymer membrane, the gelatin layer was removed by placing the covered wafer in purified water at 50 °C. Thus, the polymer layer was separated from the wafer and afterwards fixed in an aluminium foil frame using MED1-4013 adhesive (NuSil Technology, USA).

The production of a multi-layer membrane also based on the described sacrificial layer process with the difference that a second polymer layer was applied on the first one. According to the low surface energy of polymers involving a worse bonding between these materials, a surface treatment of the first polymer layer is required. First oxygen plasma (Dreva Clean 450, Germany, P = 100 W, t = 60 s, p = 20 Pa) was used to enhance surface energy and to improve strength of a double layer membrane. A bonding agent applied between the polymer layers has the additional effect of a moisture barrier which we used alternatively to plasma treatment. Reactive pulse magnetron sputtering of titanium with LS730S Sputter System (von Ardenne Anlagentechnik GmbH, Germany, P = 1500 W, 75 sccm argon, 25 sccm oxygen) provides a 23 nm titanium dioxide layer whereon the second polymer was coated.

B. *Surface analysis*

Surface energy measurement was done to characterize the polymer surface of coated wafers before and after surface treatments. This was done before the sacrificial layer was removed to enable a better handling of the samples. Diiodomethane and purified water were used as test liquids for static sessile-drop technique. The contact angle was recorded with the OCA 15 (DataPhysics Instruments GmbH, Germany) 3 s after placing a droplet on the polymer sample surface.

C. *Water vapor transmission rate measurements*

WVTR of polymer membrane was measured by means of permeability measurement system HiBarSens® (Sempa Systems GmbH, Germany) which is based on diode laser spectroscopy (Fig. 2). The system consists of two chambers (pre-chamber with 90 % humidity and measurement chamber with a constant dry inert gas flow) separated by the membrane to be tested [12]. It operates at 38 °C. Increasing water vapor concentration c in the measurement chamber decreases the laser intensity $I_0(\lambda)$ according to Beer–Lambert law (ε_λ - absorption coefficient, d – path length):

$$I(\lambda) = I_0(\lambda) \cdot e^{-\varepsilon_\lambda \cdot c \cdot d} \quad (1)$$

WVTR is directly proportional to the detected water vapor concentration [12].

III. Results

A. *Surface analysis*

Knowing the measured contact angle θ and surface tension σ_L of the test liquids, the reversible adhesion energy W_a could be simply calculated using the *Young-Dupré* equation [13]:

$$W_a = \sigma_L \cdot (1 + \cos\theta) \quad (2)$$

The surface energy arises from different intermolecular forces which are divided into dispersive (van der Waals interactions) and polar ones (hydrogen bonds, forces between dipoles). Therewith surface energies of solid surfaces σ_S and liquids σ_L are described with a dispersive (σ_S^D, σ_L^D) and a polar energy component (σ_S^P, σ_L^P):

$$\sigma_S = \sigma_S^D + \sigma_S^P \quad (3)$$

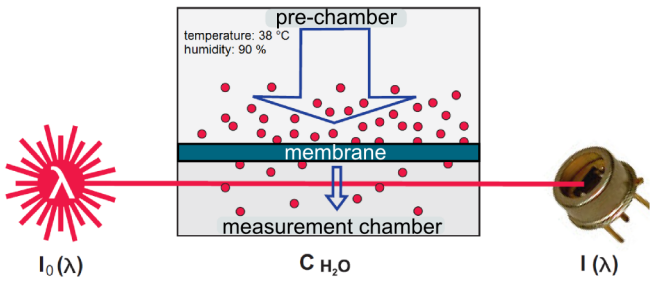


Fig. 2. Measurement system HiBarSens® based on diode laser spectroscopy for determining of water vapor transmission rate of membranes [12]

$$\sigma_L = \sigma_L^D + \sigma_L^P \quad (4)$$

Relating the *Young-Dupré* equation and the dispersive and polar energy components of surface energy *Owens and Wendt* derived the following formula which has the form of a linear equation $y = mx + n$ with $y = W_a/(2(\sigma_L^D)^{1/2})$, $m = (\sigma_S^P)^{1/2}$ and $n = \sigma_S^D)^{1/2}$.

$$\frac{W_a}{2\sqrt{\sigma_L^D}} = \sqrt{\sigma_S^D} + \sqrt{\sigma_S^P} \cdot \frac{\sqrt{\sigma_L^P}}{\sqrt{\sigma_L^D}} \quad (5)$$

TABLE I. SURFACE ENERGY WITH ITS DISPERSIVE AND POLAR ENERGY COMPONENT OF TEST LIQUIDS USED FOR SESSILE-DROP METHOD AND MEASURED SURFACES

Test liquids			
Liquid	***Surface tension*** $[\sigma_L] = mN/m$	***Dispersive energy component*** $[\sigma_L^D] = mN/m$	***Polar energy component*** $[\sigma_L^P] = mN/m$
CH_2I_2	50.8	46.6	4.2
H_2O	72.8	21.8	51.0
Measured surfaces			
Treat-ment	***Surface energy*** $[\sigma_S] = mJ/m^2$	***Dispersive energy component*** $[\sigma_S^D] = mJ/m^2$	***Polar energy component*** $[\sigma_S^P] = mJ/m^2$
Silicone MED 6-6606			
None	10.2	8.8	1,4
Plasma	37.5	20.3	17.2
Parylene C			
None	24.5	22.8	1.7
Plasma	70.0	25.2	44.8
Titanium dioxide (applied on MED 6-6606)			
None	41.6	27.7	13.9
Plasma	75.8	7.6	68.2

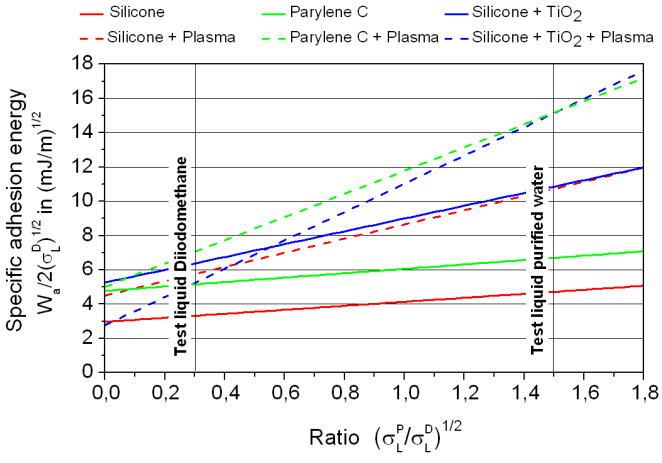


Fig. 3. Specific adhesion energy of silicone, parylene C and titanium dioxid surface (untreated and oxygen plasma treated) depending on ratio of polar to dispersive energy component of two test liquids, illustrates the dispersive polymer surface energy as y-intercept and the polar component as line slope

We chose Diiodomethane (CH_2I_2) as disperse and water (H_2O) as polar test liquid, to investigate the present intermolecular forces on the polymer surface (Tab. 1).

According to (5), polymer surface energy is plotted in Fig. 3 as a linear function illustrating the dispersive component as the y-intercept and the polar one as the slope of the line. We found that untreated silicone surfaces have the lowest surface energy compared to parylene C and titanium dioxide. Oxygen plasma treatment increases the surface energy for all investigated materials with the greatest effect on silicone which is plasma-treated 3.4 times higher. For parylene C we found a 2.8 times increased surface energy. Thus, oxygen plasma treatment of polymer surfaces is a good opportunity to increase their wettability enabling a reliable bonding of two polymers. Sputtering of titanium on silicone leads to a 4.1 times higher surface energy. This effect is improved by an additional plasma treatment.

B. Water vapor transmission rate measurements

Four membranes were produced and measured in the described way (see Tab. 2). WVTR was calculated of the silicone and parylene C double layer (sample 3) utilizing measured relevant monolayer parameters as followed:

$$WVTR_3 = \left(\frac{1}{WVTR_1} + \frac{1}{WVTR_2} \right). \quad (6)$$

TABLE II. MATERIALS AND PROCESSING TECHNOLOGY USED FOR PRAPARED SAMPLES FOR WVTR MEASUREMENTS

Material	**Layer thickness**	**Processing**	**Sample**			
			1	**2**	**3**	**4**
Silicone	3.8 µm	Spin-coating	✓		✓	✓
Titanium dioxide	0.023 µm	Sputtering				✓
Parylene C	5.9 µm	PVD		✓	✓	✓

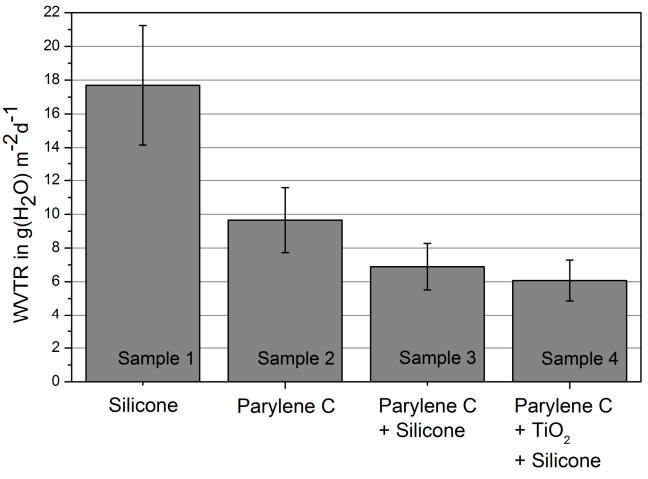


Fig. 4. Water vapor transmission rate of silicone (Sample 1) and parylene C (sample 2) monolayer and two multilayer membranes (sample 3: parylene C and silicone; sample 4: parylene C and titanium dioxide and silicone)

One obtains the WVTR of sample 3 with the value $WVTR_3 = 6.2\ g\ m^{-2}\ d^{-1}$ which is conform to the measured result. We found that the multilayer membrane composed of parylene C, titanium dioxide and silicone (sample 4) has threefold better barrier properties towards water molecules compared to the silicone monolayer (sample 1).

IV. Conclusion

Surface treatments like the plasma process or applying a bonding agent to the polymer interface facilitate combining polymers to a multilayer membrane. Best bonding properties were obtained using titanium oxide as a bonding agent followed by an additional plasma treatment. Barrier properties of flexible and biocompatible encapsulations can be increased by a multilayer system comparing polymers such as silicone and parylene C and an inorganic bonding agent like a titanium dioxide layer. Thus, protective function of polymer encapsulations can be improved significantly.

Acknowledgment

The authors would like to thank Harald Beese from Fraunhofer-Institut für Werkstoff- und Strahltechnik IWS Dresden for measuring water transmission rate of our membranes.

References

[1] S. Weber, T. C. Lueth, “Concept and realisation of an implantable EMG and motion measuring device”, Second International Conference on Pervasive Computing Technologies for Healthcare, 2008, pp. 165 – 168.

[2] N. Amanat, N. L. James, D. R. McKenzie, “Welding methods for joining thermoplastic polymers for the hermetic enclosure of medical devices”, Medical Engineering & Physics, Vol. 32, No. 7, 2010, pp. 690 – 699.

[3] J. Riistama, J. Väisänen, S. Heinisuo, H. Harjunpää, S. Arra, K. Kokko, M. Mäntylä, J. Kaihilahti, P. Heino, M. Kellomäki, O. Vainio, J. Vanhala, J. Lekkala, J. Hyttinen, “Wireless and inductively powered implant for measuring electrocardiogram”, Medical & Biological Engineering & Computing, Vol. 45, No. 12, 2007, pp. 1163 – 1174.

[4] U. Marschner, H. Grätz, H., B. Jettkant,D. Ruwisch, G. Woldt, W.-J. Fischer, B. Glasbrummel, “Integration of a wireless lock-in measurement of hip prosthesis vibrations for loosing detection”, Sensors and Actuators A: Physical, Vol. 156, No. 1, 2009, pp. 145 – 154.

[5] C. Hassler, R. P. von Metzen, P. Ruther, T. Stieglitz, “Characterization of parylene C as an encapsulation material for implanted neural prostheses”, Journal of Biomedical Materials Research Part B: Applied Biomaterials, Vol. 93B, 2010, pp. 266 – 274.

[6] F. Graichen, R. Arnold, A. Rohlmann, G. Bergmann, “ Implantable 9-Channel Telemetry System for In Vivo Load Measurements With Orthopedic Implants”, IEEE Transactions on Biomedical Engineering, Vol. 54, No. 2, 2007, pp. 253 – 261.

[7] R. Traeger, “Nonhermeticity of Polymeric Lid Sealants”, IEEE Transactions on Parts, Hybrids, and Packaging, Vol. 13, No. 2, 1977, pp. 147 – 152.

[8] Jui-Mei Hsu, L. Rieth, R. A. Normann, P. Tathireddy, F. Solzbacher, “Encapsulation of an Integrated Neural Interface Device with Parylene C“, Biomedical Engineering, IEEE Transactions on, Vol. 56, No. 1, 2009, pp. 23 – 29.

[9] T. Schanze, L. Hesse, C. Lau, N. Greve, W. Haberer, S. Kammer, S., T. Doerge, A. Rentzos, T. Stieglitz, “An optically powered Single-Channel Stimulation implant as Test system for chronic biocompatibility and biostability of miniaturized retinal vision prothese“,IEEE Transactions on Biomedical Engineering, Vol. 54, No. 6, 2007, pp. 983 – 992.

[10] C. Zhou, E. Greenbaum, “Implantable Neural Prostheses 2“, Springer, 2010.

[11] Stieglitz, T., “Biomedical microimplants for sensory and motor neuroprostheses”, IEEE International Symposium on Circuits and Systems, 2006, pp. 2189 - 2192

[12] SEMPA SYSTEMS GmbH, Fraunhofer Institut für Werkstoff- und Strahltechnik “HiBarSens: Ultra-high barrier properties are finally measurable”, Newsletter, 2012, URL: http://www.sempa.de/pdf/products/tech_HiBarSens.pdf

[13] A. W. Adamson, A.W., “Physical Chemistry of Surfaces“, Wiley-Interscience, ed. 4, New York, 1982, Chapter 10.

Experimental Investigation of Aqueous Solutions Superradiance

Nikolov N.A.[1,2]
[1]Medical Physics & Bioengineering Department
National Cancer Institute
Kyiv, Ukraine
nikolka_@ukr.net

Solyar A.G.[2], Yaroshenko O.Yu.[2]
[2]Physical and Biomedical Electronic Department
National Technical University of Ukraine "KPI"
Kyiv, Ukraine

***Abstract* — It is experimentally shown that the superradiance effect can be obtained using as a source for photosensitivity light flux from semiconductor light-emitting diode. It can be assumed that manifestation of the superradiance effect in water solution is by the mechanism closed to stochastic of synchronization cooperative properties of the system and spontaneous chemiluminescence.**

Keywords — superradiance; water solution; luminescence of solutions.

I. INTRODUCTION

The phenomenon of superradiance of water solutions was previously shown [1]. Our work group was interested in this phenomenon because it explored sensitivity of efficiency against external physical factors such as low-intensity electromagnetic field. There exist a considerable number of experimental and theoretical researches about the effect of low-intensity physical factors on aqueous and biological systems [2, 3]. However, the results of such experiments often have not been repeated in different laboratories and have specific features, not the least due to their nonlinearity and non-monotone response reaction, long delay of feedback and absence of dosimetric devices. Therefore, developing measuring systems and appropriate methods that would record short-time changes in object caused by the action of low-intensity radiation and would give an opportunity to control initial conditions and predict the results, is necessary to solve this problem. As an example of these measuring systems, the research reported in [4] should be mentioned, where about state of the aqueous solution is judged by the change in the variance differential temperature. Another group of authors trying to directly assessed on the basis scattering Mie theory of the formation water clusters on the scale of nanometers [5]. Our working group is looking for a solution to the above-described problem by analyzing the signals in spontaneous and photosensibilizated (PS) luminescence of solutions. So, in [1, 6] was studied the alteration of spontaneous and photosensitized luminescence of NaCl aqueous solutions and anticancer anthracycline antibiotic of Doxorubicinum after they were exposed to radiation with spatially heterogeneous electromagnetic field. Large amount of experimental and theoretical basis for the explanation of the observed effects in the monograph A.I. Zhuravlev [7], as well as in [8]. Exploring the quantum effects of biological solutions try to use the evidence base controlled biochemical reactions. This allows correct and pragmatic talk about the optical activity of biopolymers and electron transport [9]. However, the role of the solvent, water, in these effects, its spaces structural-dynamic properties of the relatively little known. Although it is water as condensed matter largely determines the physical and chemical properties of biological systems [10].

The outline of the experiments reported in [1] was as follows. The low-intensity optical noise radiation was sent through the measuring cell. The light intensity from the measuring cell was measured by a photomultiplier in chemiluminometer. As a source of noise backlight a radiocarbon calibration element with luminescent crystals was used. It was expected that the light intensity would decrease due to the light dissipation and absorption with water, but photomultiplier had been recording an increase in the intensity of light of 2.5 times. At the same time, depending on the preprocessing of an aqueous solution, low intensity electromagnetic field amplification factor of light varies by 10-30%. With increasing volume of liquid in the measuring cell (height of liquid column) the gain of the light increases with the quadratic dependence. That is the reason why we have assumed that the mechanism of the observed light is Dicke's superradiance [11]. While discussing this work, an argument was raised that the phenomenon of superradiance is linked with low-intensity electromagnetic field. It should also be mentioned that luminescent crystals of radiocarbon source accumulate optical energy of environment and parameters of radiation are changed when working at experimental stand. This could be the cause that affects the results. In addition, the practical use of light sources containing radionuclides is often limited by legal regulations. That is why the experimental setup was modified, where the light source was a semiconductor light-emitting diode (LED). The aim of this work is experimental confirmation of the results and assessment of the interaction of external low-intensity electromagnetic fields with aqueous solutions.

II. MATERIALS AND METHODS

Measurements were performed at the stand with medical chemiluminometer ХЛИМЦ-01 (Ukraine). Operating voltage of the photoelectronic multiplier was 1600 V. Effectiveness factor of superweak luminescence was at least 0.005 imp/quantum. The physical dimensions of a quartz cylindrical measuring cell were as follows: diameter – 25 mm, height – 30 mm.

PS was done on a gleam. LEDs were used as the light source. LEDs emit very intense light for photomultiplier during work. The effect of the superradiance was decreased when the intensity of the external light was brighter. Also, the effect of the superradiance was not seen in the directed focused radiation, such as the laser radiation. That is why the light from optical fiber was directed through a dispersing glass. Aqueous solutions were distilled water, 0.9% NaCl in aqueous solution and aqueous solution of doxorubicin.

The interaction of external low-intensity electromagnetic fields with aqueous solutions was tested using a special noise generator of electromagnetic waves, which emits waves at frequencies from 78

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

to 100 GHz and has an integral output power of 50 mW. It was developed by P.P. Loshitsky and A Pavlyuchenko [12].

III. Results and Discussion

Superradiance was observed in all tested aqueous solution when using LED as a PS-source of light. In this case spectral range was not so important. We used LEDs with green ranges because these were the most preferable for generating a signal from the photomultiplier. The most interesting spectral dependence by PS has been achieved with a white diode. However, understanding the change in the spectral dependences is ambiguous because of difficult nonlinear spectrum and nonlinearity amplification of the photomultiplier at different frequency ranges. Usual spectral dependence of the original spectrum and the output spectrum of the emission of the green LED are shown in the Fig. 1 below, where I_0 is the intensity of light passing through the empty measuring cell.

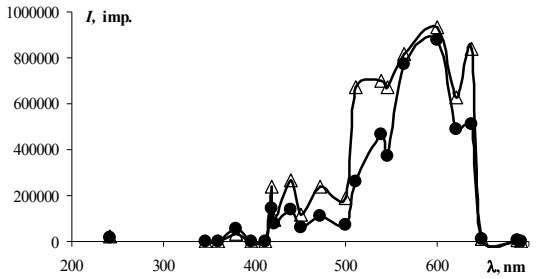
Of course, each LED has its own form of spectral dependences (more / less wideband and different irregularity). But general trend is the following. Signal is amplified as short waves and long waves. Maximum of the signal is shown in the short-wave part of the spectrum. At the point of maximum the original spectrum corresponds to the minimum point of the gain.

Usual dependences of the light gain on the intensity of photosensitivity are shown in the Fig. 2.

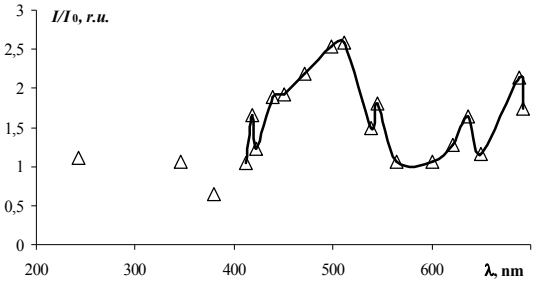
Analysis of dependences shows that for investigated aqueous solutions the gain of the light (superradiance) does not change until some threshold at 8000 pulses per 7 second (counting rate = 11 000 imp/s); when the counting rate equals about 7000 imp/s, almost invisible maximum is observed; at a counting rate of more than 14 000 imp/s – a significant decrease in the gain is recorded. In the case of 0.9% NaCl after UHF-irradiated superradiance effect was increased by 5-10%, but it was decreasing for distilled water. Keeping NaCl solution or distilled water in the measuring cell for more than two days leads to 20-30% reduction in the effectiveness of SR compared with freshly prepared samples. The superradiance effect was also decreased for the doxorubicin solution after UHF- irradiation. If doxorubicin was dissolved in distilled water, the effect decreased by 5-10%, whereas if doxorubicin was dissolved in 0.9% NaCl, the effect is decreased by 3-5%. However, it should be noted that these regularities are valid for doxorubicin within long photosensitivity of it. Doxorubicin increases counting rate of up to two times within a 7-30 seconds after the source PS was switched on without pre-processing UHF field and it increases counting rate of up to 2.5-3 times after it has been previously processed with UHF field.

Comparing these data with previous research [1], it should be noted that the superradiance effect can be achieved when the light outputs from the LEDs are used as a source for PS. However, this system would be less sensitive to aqueous solutions which had been preprocessed with low-intensity external factors. It may depend on the fact that in radioluminescent sources of PS the signal can be considered as noisy. Anyway, the use of LED is not considered as random despite the fact that it has more sophisticated emission spectrum. It is well known that "exciton density (dipoles) may be made by methods of injection or optical pumping in heterostructures with quantum well. In a situation where the distance between the dipoles is less or equal the wavelength of the radiation, but is large enough that it would neglect the dipole-dipole interaction in the field of an electromagnetic waves spontaneous phasing may take place which leads to the Dicke's superradiance" [13]. Liquid water solutions may be considered as clusters dynamical systems [6, 14] with a relatively stable hydration of ions and dissolved impurities in the Stern's layer [15] and with high instability of the Gouy diffusion layer. The foregoing statement and theory could explaing these experimentally received results. In other words, an analogy can be drawn between the heterostructures in solid and dynamic cluster formations in the water. Thus, if externally low-intensity electromagnetic field can change the dynamic spatial structure of the water system, then in the case of lower molecular weight ionic impurities distant order will be increased in the water. In other words, it is ion which stabilizes temperature fluctuations of dipole water molecule with increase in thickness Gouy's layer (hydrate fur) due to:

- hydration shells interaction with each other,
- self-diffusion of water molecules,
- there is always a non-equilibrium and distributed in space temperature,
- formation of gas-phase oxygen and hydrogen.



a



b

Fig. 1. Typical spectral dependences of emission of aqueous solutions during PhS: a – measured spectra; b – amplification factor; • – initial spectrum of green LED; Δ – output spectrum luminescence of 7 ml of 0.9% NaCl solution

Therefore, increasing the cluster formation of an aqueous solution must necessarily create increasing tensions in mechano-chemical bonds. Conditions should be created for the formation of excited molecules. In the experimental data preprocessing of 0.9% NaCl by a broadband noise with low intensity UHF field has led to increased superradiance effect; whereas a distilled water that has little amount of stabilizing its clusters and it has led to decreasing effect.

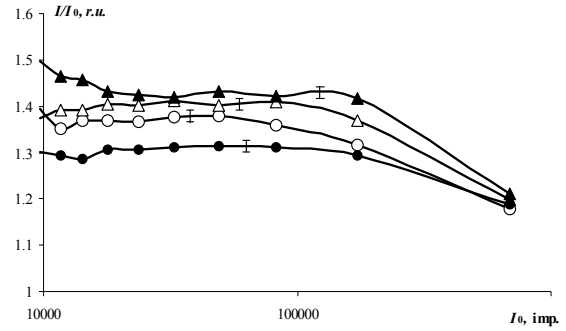


Fig. 2. Typical dependence of the amplification factor of light depending on the intensity of photosensitivity: Δ – 0.9 % of NaCl solution; ▲ – 0.9 % of NaCl solution after 15 min UHF-irradiated; • – distilled water; o – distilled water after 15 min UHF-irradiated

However, according to the Dicke's general theory [16, 17], the cooperative phenomena may occur where the response intensity of the light may exceed the exciting pulse and spontaneous emission by a few orders of magnitude. At the same time Dicke's superradiance pulse duration is several orders of magnitude smaller than the spontaneous radiation after PS (ms - s). The following should be noted:

- the experimental spectra observed Stokes and anti wider range expansion of the range (in red and ultraviolet scattering);
- much better results of the superradiance effect when noise source radioluminescent PS is used;
- complex dynamic fractal structure of the water system;
- SR reduction with increasing intensity of the PS;
- SR reduction if the solution is in equilibrium (solution was kept for a long time in the measuring cell).

Considering the above, it can be suggested that as opposed to classical theory of Dicke where the timing of the quantum states occurred by the front of exciting pulse of light, in this case, synchronization of system occurs due to effects of which are close stochastic synchronization mechanism [18-20]. This happens despite spontaneous luminescence making quite a large contribution to the enhancement of light.

CONCLUSION

1. Superradiance effect can be obtained using a flux from semiconductor diodes as a light source for photosensibility.

2. After the aqueous solution has been irradiated with the noise of broadband high-frequency electromagnetic field the following were observed: trend of increasing superradiance effect in 0.9% NaCl solution; decreasing the effect in distilled water; in water solution of doxorubicin – short-term increasing of radiation with the subsequent decreasing of superradiance as compared with control samples.

3. The sensitivity of the system is higher when random noise signals rather than LEDs are used as a source of photosensitivity.

4. It can be assumed that in the experiment described synchronization system and, as a consequence, the manifestation of the superradiance effect is by the mechanism closed to stochastic synchronization.

REFERENCES

[1] N.A. Nikolov, A.G. Solayr, "Change in aqueous solutions' superradiance as a result of their interaction with spatially-heterogeneous electromagnetic field", Electronics and nanotechnology. Proceedings of the XXXII international scientific Conference ELNANO 2012, Kiev, Ukraine, pp. 144-145, April 10-12, 2012.

[2] V.N. Binhi, A.V. Savin, "The effects of weak magnetic fields on biological systems: physical aspects", Physics-Uspekhi (Advances in Physical Sciences), vol. 173, № 3, pp. 256-300, 2003.

[3] Yu. Trushkov, L. Makarova, A. Kamenskih, A. Trushkov, "To the question of structural modifications of the water under the influence of external and internal environment", Vestnik State National Research Polytechnical University of Perm, vol. 12, № 4, pp. 146-159, 2010.

[4] Loshitskiy P.P., Minziak D.K. "Remote determination of parameters of water solution", Electronics and nanotechnology. Proceedings of the XXXII international scientific Conference ELNANO 2012, Kiev, Ukraine, pp. 194, April 10-12, 2012.

[5] V.V. Goncharuk, A.E. Bulanova, E.A. Orehova, V.V. Malyarenko, "The Water as a Heterogeneous Cluster Disperse System", Polish J. of Chemistry, vol. 82, N. 1-2, pp. 403-409, 2008.

[6] N.A. Nikolov, "Cheng structure-dynamics properties of water solution under interaction with spatially non-uniform electromagnetic fields", Electronics and Communications, vol. 62, N. 3, pp. 131-139, 2011. (Rus).

[7] A.I. Zhuravlev, "Quantum biophysics animals and humans", Moskow: «БИНОМ», 398 p., ISBN 978-5-9963-0448-6. (Rus)

[8] I.L. Fabelinskij, "Seventy years of combinational scattering", Physics-Uspekhi, vol. 168, iss. 12, pp. 1341–1360, 1989. (Rus)

[9] S.R. Hameroff, J.A. Tuszynski, "Quantum states in proteins and protein assemblies: The essence of life?", Proc. SPIE 5467, Fluctuations and Noise in Biological, Biophysical, and Biomedical Systems II, 27 (May 25, 2004), doi:10.1117/12.548520.

[10] L. N. Gall , N. R. Gall, "The mechanism of intermolecular energy transmission and perception of ultraweak actions by chemical and biological systems", Biophysics, vol. 54, iss. 3, pp. 563 – 574, 2009.

[11] R.H. Dicke, "Coherence in spontaneous radiation processes", Physical Review, vol. 93, № 1, pp. 99-110, 1954.

[12] Loshitski P.P., Pavlyuchenko A.V. "Research of influence of supply mode of double-drift avalanche diodes on Q-factor of 3-mm band noise oscillator ", 21st Int. Crimean conference "Microwave & Telecommunication Technology", CriMiCo'2011, 12 sept., Sevastopol, p.183-184, 2011.

[13] A.I. Klimovskaya, E.G. Gule, Yu.A. Dryha, "Dicke superradiation in quantum heterostructures at optical pump", Semiconductors, vol. 36, № 2, pp. 232-233, 2002.

[14] P. Loshickiy, M. Kurik, N. Nikolov, "The reaction of water systems to the action of low-intensity physical factors", Electronics and Communications, №16, pp. 80-84, 2002. (Rus)

[15] Adamson A. W. "Physical Chemistry of Surfaces". Third Edition . A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York/London/Sydney/Toronto, 1979.

[16] A.V. Andreev, V.I. Emel'yanov, Yu.A. Il'inskii, "Collective spontaneous emission (Dicke superradiance)", Sov. Phys. Usp., V. 23, pp. 493–514, 1980.

[17] A.V.Andreev, V.I.Emelyanov, Yu.A.Iliinskii, "Cooperative Effects in Optics", Bristol and Philadelphia, IOP Publ, p. 467, 1993.

[18] E. Simonotto, M. Riani, Ch. Seife, M. Roberts, J. Twitty, F. Moss, "Visual Perception of Stochastic Resonance", Physical review letters, vol. 78, № 6, pp.1186-1189, 1997.

[19] P. James, H.J. Clemens, "Carmichael Stochastic initiation of superradiance in a cavity: An approximation scheme within quantum trajectory theory" Physical review A, vol. 65, p. 023815-1, 2002.

[20] D. Witthaut, "Stochastic resonance driven by quantum shot noise in superradiant Raman scattering", J. Phys. B: At. Mol. Opt. Phys., vol. 45, № 22, p. 225501, 2012.

Computer-Aided Software for Early Diagnosis of Eerythemato-squamous Diseases

Bekir Karlık
Department of Computer Engineering
Selçuk University Konya, Turkey
bkarlik@selcuk.edu.tr

Güneş Harman
Department of Computer Engineering Yalova University
Yalova, Turkey
gunes.guclu@gmail.com

***Abstract*— Early diagnosis and appropriate treatment remain a necessary challenge. Dermatologic emergencies have insufficient attention by the general population and by physicians from other specialties. The differential diagnosis of erythemato- squamous diseases is a real problem in dermatology. They all share the clinical features of erythema and scaling with very little differences. These diseases are psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis, and pityriasis rubra pilaris. Usually a biopsy is necessary for the diagnosis but unfortunately these diseases share many histopathological features as well. In this study, computer-aided software was developed to diagnosis dermatological diseases by using artificial neural networks. The supervised back-propagation algorithm is used to train the networks. Classification of the average value of sensitivity (or recognition percentage) was found as 98% for six erythemato-squamous diseases.**

Keywords—Early diagnosis; Dermatology diseases; Software; Artificial neural networks

I. INTRODUCTION

The software development exploits the potential of human intelligence such as reasoning, making decision, learning and many others. Many intelligent systems have been developed for the purpose of enhancing health-care and provide better health care facilities, reduce cost and etc. As express by many studies, intelligent system was developed to assist users (particularly doctors and patients) and provide early diagnosis and prediction to prevent serious illness such as breast cancer [1], leukemia cancer [2], prostate cancer [3], diabetes [4], hepatitis [5], epilepsy [6], retinopathy [7], and some heart failures [8,9].

Although ANN have been developed in the last few decades for many different applications in medical science, their use in the field of dermatology remains relatively limited. ANN has the potential to be applied in various fields of dermatology, there is a need for further appropriately designed clinical trials before these emergent techniques can be widely used in the clinical practice to help to medical doctors. So far, the main use of ANN in dermatology is the dermatology problem related to the well-known erythemato-squamous diseases is a difficult problem in medicine, since all these diseases share the clinical features of erythema and scaling, with very little differences [10]. The diseases regard psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis, and pityriasis rubra pilaris. Usually a biopsy is necessary for the diagnosis but unfortunately these diseases share many histopathological features as well. Another difficulty for the differential diagnosis is that a disease may show the features of another disease at the beginning stage and may have the characteristic features.

The aim of the dermatology classification problem considered in this work is to determine the type of eryhemato-squamous diseases when 34 clinical and histopatological features are defined as inputs. We did not use the age feature which represents the age of the patient. The literature examining the feature space of erythemato-squamous diseases is relatively recent-most studies have deployed the well-known UCI dermatology dataset which was employed in this work. To solve this problem different studies have been proposed in the literature in order to define classifier systems such as voting feature intervals [11], adaptive neuro-fuzzy inference (ANFIS) [12], k-Means clustering[13], a novel neuro-genetic approach[14], and a novel decision tree classifier[16]. Several authors have used a support vector machine (SVM) and LSVM approach to classifying this dataset [17-19]

In this work, we present a supervised Back-Propagation training algorithm to optimize a particular kind of neural networks that would work as multi-classifier systems in this dermatological application. Used ANN architecture consists of 3 layers as 33 nodes of input layer, 33 nodes of hidden layer, and 6 nodes of output layer which is class distribution of dermatological diseases which are psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis, pityriasis rubra pilaris.

II. MATERIALS AND METHODS

Artificial neural networks (ANN) are not a new concept, yet it has been accepted as a new technology in computer science. ANN has been proved extremely useful in pattern recognition problems. ANN is a network of many very simple processors named nodes (individual neurons), each possibly having a local memory. The nodes are connected by unidirectional connections named weights which carry numerical data. The nodes operate only on their local data and on the inputs they receive for the connections. The weights associated with the interconnections between nodes, and the activation function of each node. The weights of ANN are initialized to random values. We used the back-propagation training algorithm which has supervised generalized learning

rule. It is an iterative gradient algorithm designed to minimize the root mean square error (MSE) between the actual output of a multilayered feed-forward neural networks and a target output. Each layer is fully connected to the previous layer, and has no other connection. The most common choice of activation functions for multi-layered perceptron (MLP) is used as transfer functions in research and engineering. Among the reasons for this popularity are its boundedness in the unit interval, the function's and its derivative's fast computability, and a number of amenable mathematical properties in the realm of approximation theory. However, considering the huge variety of problem domains MLP is applied in, it is intriguing to suspect that specific problems call for single or a set of specific activation functions. According to Karlik and Olgac experimental study, we can say that the Tanh and Sigmoid activation functions can be used in the vast majority of MLP applications as a good choice to obtain high accuracy [20].

In this study, computer-aided software was developed to diagnosis dermatological diseases. This method has the characteristic features at the following stages. Patients were first evaluated clinically with 11 features. Afterwards, skin samples were taken for the evaluation of 22 histopathological features. The values of the histopathological features are determined by an analysis of the samples under a microscope the dataset constructed for this domain, the family history feature has the value 1 if any of these diseases has been observed in the family, and 0 otherwise. Data consists of 355 examples with 33 features of each example and 6 classes in total. One will notice that we deleted 11 samples which have missing values from the original datasets of erythemato-squamous diseases. Because there is no age attributes in these data. In this study, 150 of them were for training, the others were used for testing. Attributes information for used data has been divided into Clinical attributes and Histopathological atributes;

Clinical Attributes: (take values 0, 1, 2, 3, unless otherwise indicated)

1: erythema
2: scaling
3: definite borders
4: itching
5: koebner phenomenon
6: polygonal papules
7: follicular papules
8: oral mucosal involvement
9: knee and elbow involvement
10: scalp involvement
11: family history, (0 or 1)

Histopathological Attributes: (take values 0, 1, 2, 3)

12: melanin incontinence
13: eosinophils in the infiltrate
14: PNL infiltrate
15: fibrosis of the papillary dermis
16: exocytosis
17: acanthosis
18: hyperkeratosis
19: parakeratosis
20: clubbing of the rete ridges
21: elongation of the rete ridges
22: thinning of the suprapapillary epidermis
23: spongiform pustule
24: munro microabcess
25: focal hypergranulosis
26: disappearance of the granular layer
27: vacuolisation and damage of basal layer
28: spongiosis
29: saw-tooth appearance of retes3
30: follicular horn plug
31: perifollicular parakeratosis
32: inflammatory monoluclear inflitrate
33: band-like infiltrate [11].

Every other feature (clinical and histopathological) was given a degree in the range of 0 to 3. Here, 0 indicates that the feature was not present, 3 indicate the largest amount possible, and 1, 2 indicate the relative intermediate values. All data was normalized between 0,2 to 0,8 (0 as 0,2; 1 as 0,4; 2 as 0,6 and 3 as 0,8). These values were applied as inputs of ANN. Used Multi-Layered Perceptron (MLP) architecture consists of 3 layers as 33 nodes of input layer, 33 nodes of hidden layer, and 6 nodes of output layer. As shown in Figure.1, the output of ANN is class distribution of dermatological diseases such as psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis, pityriasis rubra pilaris.

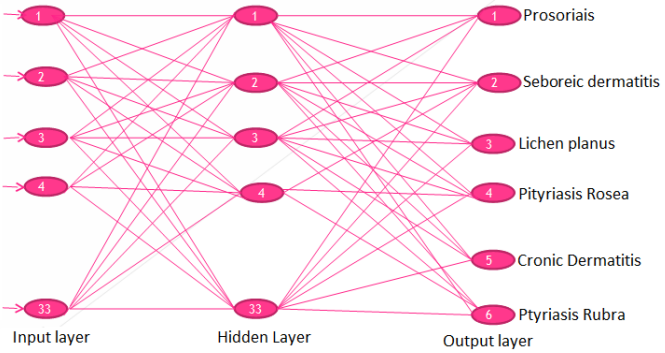


Fig. 1. Used MLP architecture

III. RESULTS AND CONCLUSION

This study has demonstrated computer-aided software for early diagnosis of dermatological diseases. Proposed software user interface is also able to generalize with very low MSE (0.0205) as seen in Fig. 2. Table I shows that the labels for the six decision classes and the number of exemplars for each for testing. For testing 205 data was used.

TABLE I. CLASS LABELS AND NUMBER OF EXEMPLARS

Class Label	Number of instances
1 psoriasis	60
2 seboreic dermatitis	36
3 lichen planus	46
4 pityriasis rosea	24
5 cronic dermatitis	27
6 pityriasis rubra pilaris	12

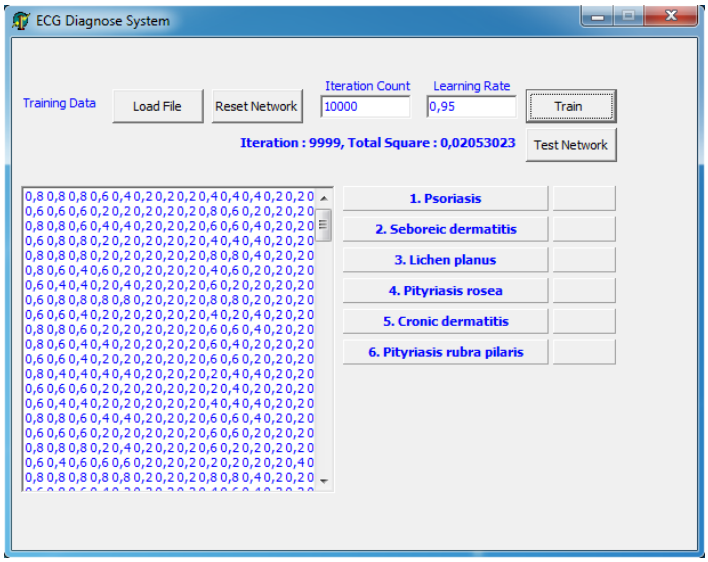


Fig. 2. USER INTERFACE OF THE SOFTWARE

We observed very high recognition rate from the other literature studies. Table 1 shows that the number of false recognition for each erythemato- squamous diseases. As it can be seen in Table II, there is no any false recognition for Lichen Planus, Cronic Dermatitis, and Pityriasis Rubra Pilaris.

TABLE II. TESTING RESULTS OF SIX ERYTHEMATO- SQUAMOUS DISEASES

Class	Number of True Recognition	Number of False Recognition
P	1SD, 2PR	1SD, 2PR
SD	5PR	5PR
LP	0	0
PR	3SD	3SD
CD	0	0
PRP	0	0

Where P: Psoriasis, SD: Seboreic Dermatitis, LP: Lichen Planus, PR: Pityriasis Rosea, CD: Cronic Dermatitis, PRP: Pityriasis Rubra Pilaris

We summarized comparison between the optimal performance of this study and the other studies from literature in Table III. The best performance is that the testing accuracy can achieve 100% with features selected from data. After 10000 iterations, the average training accuracy was found approximately as 98% which is almost perfect classification.

In this study, we gave the best testing accuracy and the average one of this run. However, there are many aspects in this work to be improved. The accuracy was computed by the ratio of correctly classified exemplars to the exemplars under considering, which did not consider the classification accuracy of each class.

In the future, we will try to use the other neural networks algorithms such as SVM, RBF, LVQ, and the other hybrid algorithms using our normalization method.

TABLE III. COMPARISON OF CLASSIFICATION ACCURACIES

Authors	*Method*	*Classification accuracy (%)*
Guvenir, Demiroz, Ilter	voting feature intervals	96.20
Ubeyli	ANFIS	95.50
Nanni	LSVM	97.22
Azzini and Marrara	neuro-genetic approach	94.62
Xie1, Xie1, Wang, Gao	SVM	97.58
Polat and Güneş	decision tree classifier	96.71
This study	Back-Propagation	97.95

REFERENCES

[1] B. Karlık and S. Ünlü, "Practical Artificial Neural Networks Software for Breast Cancer Diagnosis", Ukrainian Journal of Telemedicine and Medical Telematics, v. 6, 2, pp. 63-68, 2008.

[2] B. Karlık and E. Öztoprak, "Personalized Cancer Treatment by Using Naïve Bayes Classifier", Inter. Journal of Machine Learning and Computing, vol.2(3), pp.339-344, 2012.

[3] B. Karlık and E. Öztoprak, "Web-Based Tele-Medical Consultation and Diagnosis Model by Multiple Artificial Neural Networks," Ukrainian J. of Telemedicine and Medical Telematics, vol. 5(2), pp. 156-160, 2007.

[4] O. Karan, C. Bayraktar, H. Gümüşkaya, and B. Karlık, "Diagnosing Diabetes Using Neural Networks on Small Mobile Devices", Expert Systems with Applications, vol. 39, pp. 54-60, 2012.

[5] B. Karlık, "Hepatitis Disease Diagnosis Using Backpropagation and the Naive Bayes Classifiers", BURCH Journal of Science and Technology, vol. 1(1), pp. 49-62, 2011.

[6] B. Karlık and T.G. Aparı, "Development of Effective Telemedicine Software for Epileptic Seizure Recognition", Inter. Journal of Computing & Information Technology, vol. 1(2), pp. 91-99, 2009.

[7] A. Okatan, B. Karlık and F. Demirezen, "Detection of Retinopathy Diseases Using Artificial Neural Network Based on Discrete Cosine Transform", Neural Network World, 19(2): pp. 215-221, 2009.

[8] R. Ceylan, Y. Özbay and B. Karlık, "Telecardiology and Teletreatment System Design for Heart Failures Using Type-2 Fuzzy Clustering Neural Networks", International Journal of Artificial Intelligence and Expert Systems vol. 1(4), pp: 100-110, 2011.

[9] Y. Özbay, R. Ceylan, and B. Karlık, "Integration of Type-2 Fuzzy Clustering and Wavelet Transform in a Neural Network Based ECG Classifier", Expert Systems with Applications, vol. 38, pp. 1004-1010,2011.

[10] J.M. Renders and T. Simonart, "Role of Artificial Neural Networks inDermatology", Dermatology, vol. 219, pp. 102–104, 2009.

[11] H. A. Guvenir, G. Demiroz and N. Ilter, "Learning differential diagnosis of eryhemato-squamous diseases using voting feature intervals", Aritificial Intelligence in Medicine, vol. 13, pp. 147–165, 1998.

[12] E. Ubeyli and I. Guler. "Automatic detection of erythematosquamous diseases using adaptive neuro-fuzzy inference systems", Computers in Biology and Medicine, vol. 35, pp. 421–433, 2005.

[13] Ubeyli, E.D. & Dogdu, E., Automatic Detection of Erythemato-Squamous Disease using k-Means Clustering, J Med Syst, vol. 34(2), pp.179-184, 2010

[14] K. Polat and S. Güneş, "A novel hybrid intelligent method based on c4.5 decision tree classifer and one-against-all approach for multi-classifi cation problems", Expert Systems with Applications, vol. 36, pp. 1587-1592, 2009.

[15] A. Azzini and S. Marrara, "Dermatology Disease Classification via Novel Evolutionary Artificial Neural Network", 18th International Workshop on Database and Expert Systems Applications, pp. 148-152, Regensburg, Germany, 3-7 Sept. 2007.

[16] J. Xie1, W. Xie1, C. Wang, and X. Gao, "A Novel Hybrid Feature Selection Method Based on IFSFFS and SVM for the Diagnosis of Erythemato-Squamous Diseases", Journal of Machine Learning Research - Proceedings Track 01/2010; vol. 11, pp. 142-151, 2010.

[17] L. Nanni, "An ensemble of classifiers for the diagnosis of erythemato-squamous diseases. Neurocomputing", vol. 69, pp. 842-845, 2006.

[18] E.D. Ubeyli, "Multiclass support vector machines for diagnosis of erythemato-squamous diseases", Expert Systems with Application, vol. 35, pp.1733-1740, 2008

[19] E.D. Ubeyli, "Combined neural networks for diagnosis of erythemato-squamous diseases", Expert Systems with Applications, 36:5107{5112, 2009.

[20] B. Karlık and A.V. Olgaç, "Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks, Inter. Journal of Artificial Intelligence and Expert Systems, vol. 1(4), pp. 111-122, 2011.

Nanoparticles of Ferromagnetic Materials and Possibilities of their Application in the Hyperthermia of Malignant Tumors

A. Belous, S. Solopan, A. Yelenich,
V.I. Vernadskii Institute of General and Inorganic Chemistry, NAS of Ukraine, Palladina avenue 32/34, 03680 Kyiv, Ukraine; solopan@ukr.net

L. Bubnovskaya, S. Osinsky
R. E. Kavetsky Institute of Experimental Pathology, Oncology and Radiobiology, NAS of Ukraine, vul. Vasylkovska 45, 03022, Kyiv, Ukraine

***Abstract* — Slightly agglomerated, crystalline nanoparticles of ferromagnetic compounds with spinel and perovskite $La_{1-x}Sr_xMnO_3$ structure have been synthesized by the sol-gel method, precipitation from glycol solutions and using microemulsion technique. "Core/shell" $La_{1-x}Sr_xMnO_3/SiO_2$ stucture have been developed. Magnetic properties of the synthesized nanoparticles have been investigated. It was shown that the obtained nanoparticles exhibit superparamagnetic properties. It has been found that the synthesized nanoparticles have promise in hyperthermia of cancer cells.**

Keywords—spinel, perovskite, nanoparticle, precipitation, hyperthermia.

I. INTRODUCTION

It is known that during the transition to nanoscale, most physical and chemical properties of materials change. Nanoparticles of ferromagnetic materials, which can be used to create a new generation of permanent magnets [1], storage systems [2], are the object of considerable interest. Of particular interest are ferromagnetic nanoparticles in medicine, e.g. for nanoparticle-based diagnosis [3] and hyperthermia in the treatment of cancer diseases [4].

One of the promising aplications of nanoparticles for combined treatment of oncological patients is the method of local hyperthermia (HT). The use of nanosized magnetic particles, which are able to generate heat by the action of variable HF and MW magnetic field, offers much ampler scope for the use of hyperthermia in the therapy of malignant neoplasms. The nanoparticles of ferromagnetic materials have to meet various requirements. In particular, they must be slightly agglomerated, have small size, be single-domain ones, easily excretable from the body, release a significant amount of heat under the influence of electromagnetic fields and not interact with each other after removing the magnetic field. To increase the stability of properties of nanoparticles, as well as to better excrete ferromagnetic nanoparticles from the body, it is desirable to cover them with a dielectric phase, e.g. SiO_2.

A large number of methods for the synthesis of ferrite nanoparticles is based on the precipitation of slightly soluble compounds from aqueous solutions [5]. The complex and uncontrollable mechanism of such reactions involves crystal nucleation and growth, which occur simultaneously. This often results in the agglomeration of nanoparticles and low controllability of parameters of obtained products.

The objects of investigation were compounds AFe_2O_4 (A = Mn, Fe, Co, Ni, Zn) and $La_{1-x}Sr_xMnO_3$ crystallizing in the spinel and perovskite structures respectively, which in the nanosize form may exhibit superparamagnetic properties [6], and which are characterized by the controllable phase transition temperature [7].

Therefore, the aim of these studies was synthesis of slightly agglomerated crystalline nanoparticles of ferromagnetic compounds with spinel AFe_2O_4 (A = Mn, Fe, Co, Ni, Zn) and perovskite $La_{1-x}Sr_xMnO_3$ structure by the sol-gel method, precipitation from glycol solutions and microemulsions and study of magnetic properties and elucidation of their possible use in hyperthermia of cancer cells.

II. EXPERIMENTAL DETAILS

A. Synthesis of $La_{1-x}Sr_xMnO_3$ nanoparticles by sol-gel method

The synthesis of ferromagnetic nanoparticles of substituted lanthanum manganites $La_{1-x}Sr_xMnO_3$ was carried out at different pH values of gel according to the scheme presented in Fig. 1. The heat treatment of obtained precipitates ("precursors") was carried out in the temperature range 400-1100°C (in air) during 2-4 hours by the method described in [8].

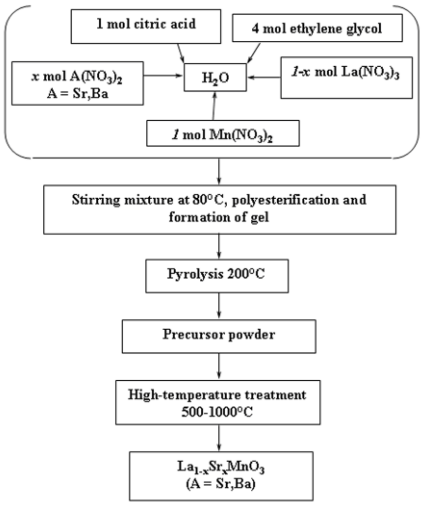


Fig. 1. Scheme of sol-gel synthesis of $La_{1-x}Sr_xMnO_3$ nanoparticles.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

B. Synthesis of $La_{1-x}Sr_xMnO_3$ and AFe_2O_4 (A = Mn, Fe, Co, Ni, Zn) nanoparticles by precipitation from nonaqueous solutions

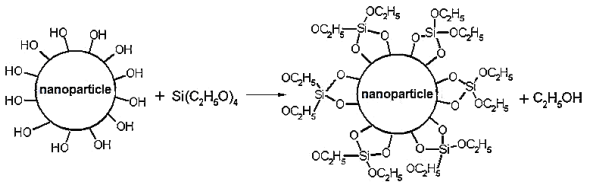
As starting reagents in the synthesis of $La_{1-x}Sr_xMnO_3$ and AFe_2O_4 nanoparticles (A = Mn, Fe, Co, Ni, Zn) metal nitrates and diethylene glycol as a nonaqueous medium were used. According to the method described in [9], synthesis was performed in an argon atmosphere.

C. Synthesis of $La_{1-x}Sr_xMnO_3$ and Fe_3O_4 nanoparticles from microemulsion solutions

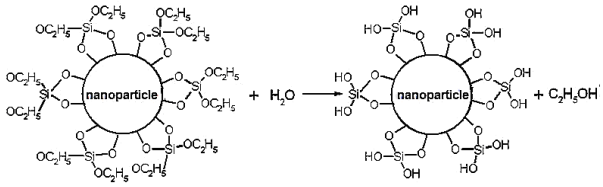
The synthesis was carried out using a procedure developed by the authors of [10-11]. For synthesis two types of microemulsions based on CTAB and Triton X-100 surfactants were selected. The synthesis of nanoparticles of $La_{1-x}Sr_xMnO_3$ and Fe_3O_4 compounds was carried out in an agron atmospere. In order to obtain crystalline product of $La_{1-x}Sr_xMnO_3$ nanoparticles, additional heat treatment was carried out in the temperature range 200-800°C.

D. Synthesis of $La_{1-x}Sr_xMnO_3/SiO_2$ nanocomposites of "core/shell" type.

The synthesis of $La_{1-x}Sr_xMnO_3/SiO_2$ nanocomposites of "core/shell" type was carried out by layer-by-layer deposition. The essence of this method is the controllable deposition onto the surface of particles of layers of a different material through the creation of certain functional groups on the surface. An advantage of this method is the control of coating thickness. The functional hydroxide groups were created on the surface of nanoparticles using a procedure developed by the authors of [12].The SiO_2 layer was deposited in accordance with reaction (1). To deposit several SiO_2 layers, the surface of the preceding layer was treated in accordance with reaction (2), and repeated procedure of SiO_2 deposition was carried out.



(1)



(2)

The particles obtained after synthesis were dried in air at 150°C.

E. Preparation of magnetic fluids

Fluids based on synthesized nanoparticles of compounds AFe_2O_4 (A = Mn, Fe, Co, Ni, Zn) and $La_{1-x}Sr_xMnO_3$ with spinel and perovskite structures respectively were prepared using 0.1 % aqueous agarose solutions. Magnetic fluids were prepared by the method described previously [13].

F. Methods of analysis

Synthesized nanoparticles of compounds with spinel and perovskite structure were analyzed by differential thermal analysis (DTA), differential scanning calorimetry (DSC) and IR spectroscopy. Prepared nanopowders were investigated by X-ray phase analysis (XRD) and full-profile X-ray phase analysis on a DRON-4 diffractometer (CuKα radiation). The size and morphology of powder particles have been determined by means of a JEM-1230 scanning electron microscope. Magnetic properties of nanopowders were measured in the magnetic field of 20 and 50 Oe with the Quantum Design (QD) Magnetic Properties Measurement System MPMS.

III. Results and Discussion

Sol-gel synthesis of $La_{1-x}Sr_xMnO_3$ ferromagnetic nanoparticles has been carried out, which showed that by changing gel pH during synthesis, one can affect the crystallinity, size of particles and their aggregation (Fig 2). It has been found that the use of the proposed sol-gel procedure makes it possible to reduce the temperature of preparation of single-phase product. In this case, crystalline nanoparticles of 40-50 nm size can be obtained.

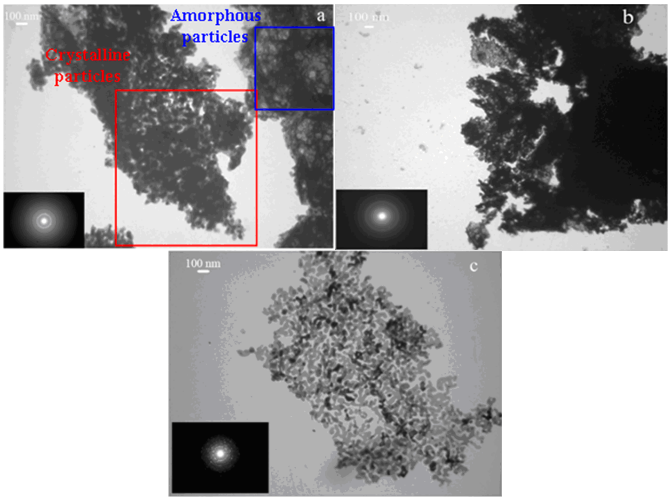


Fig. 2. TEM images of $La_{1-x}Sr_xMnO_3$ nanoparticles synthesized by sol-gel method at pH ~ 1.5 (a), pH ~ 6 (b), pH ~ 9(c) after heat treatment at 800°C

When fabricating $La_{1-x}Sr_xMnO_3$ and AFe_2O_4 (A = Mn, Fe, Co, Ni, Zn) nanoparticles ***from nonaqueous solution,*** the chemical reactions that occur during the synthesis were investigated. It has been shown that slightly agglomerated crystalline nanoparticles are formed during the synthesis from nonaqueos solution (diethylene glycol) already at 200°C. In this case, for example, the size of $ZnFe_2O_4$ nanoparticles is less than10 nm (Fig 3a). Studies have shown that heat treatment of particles of compounds with spinel structure leads to an increase in particle size to ~ 20-25 nm (Fig. 3b).

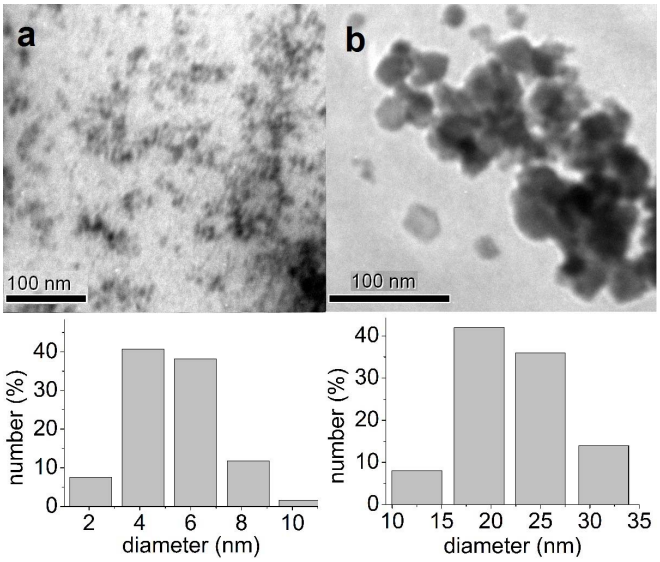


Fig. 3. TEM images of $ZnFe_2O_4$ nanoparticles synthesized from nonaqueous solutions at 200°C(a) and calcined at 500°C (b)

At the same time, it was shown that in the case of $La_{1-x}Sr_xMnO_3$ solid solutions at synthesis temperatures of 200-220°C, only amorphous precipitates are formed. The heat treatment in a wide temperature range has shown that the formation of the crystalline structure of solid solutions $La_{1-x}Sr_xMnO_3$ is observed at 600°C. In this case nanoparticles are crystalline, nanosized (≈ 25 nm) and slightly agglomerated (Fig 4).

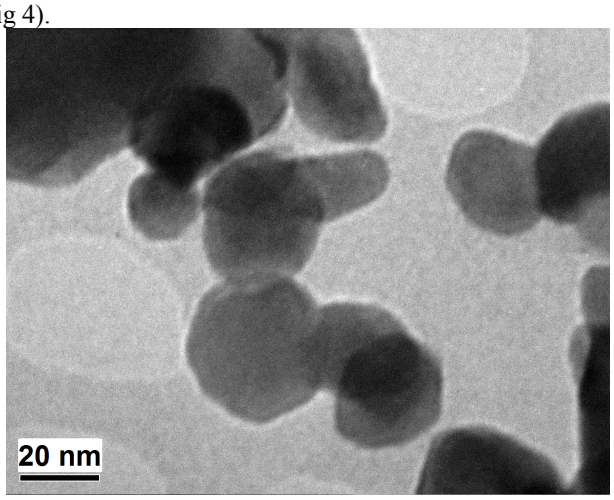


Fig. 4. TEM images of $La_{1-x}Sr_xMnO_3$ nanoparticles synthesized by precipitation from nonaqueous solutions after heat treatment at 600°C.

Of particular interest in fabricating slightly agglomerated nanoparticles is ***synthesis in microemulsions*** [11], which provides chemical interactions in nanoreactors, which are isolated in the organic matrix. The use of this method for the synthesis of nanoparticles of ferromagnetic materials based on $La_{1-x}Sr_xMnO_3$ and AFe_2O_4 (A = Mn, Fe, Co, Ni, Zn) allowed us to prepare crystalline particles in the size range 20-40 nm (for $La_{1-x}Sr_xMnO_3$) and 2-15 nm (for AFe_2O_4) depending on the choice of surfactant (Fig. 5).

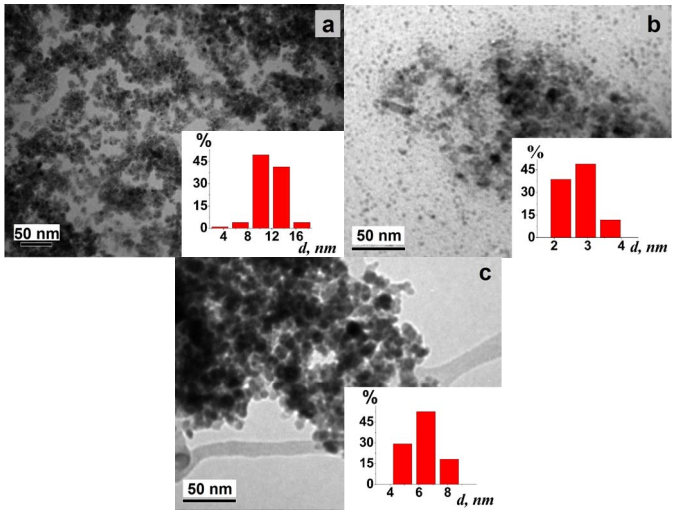


Fig. 5. TEM images and size distribution diagrams of Fe_3O_4 nanoparticles of compound synthesized from microemulsion solution based on cetyltrimethylammonium bromide (a), Brij-35 (b), and Triton X-100 (c)

It is known [14] that to ensure the stability of nanoparticles in time and to improve their excretion from the organism, it is desirable to cover synthesized nanoparticles with some materials such as a layer of silicon oxide (SiO_2). For this purpose, the method of layer-by-layer deposition was used [12,15]. As a result, ***core/shell structures of $La_{1-x}Sr_xMnO_3/SiO_2$*** have been obtained, in which the thickness of the SiO_2 layer can be changed between 5 and 10 nm (Fig. 6).

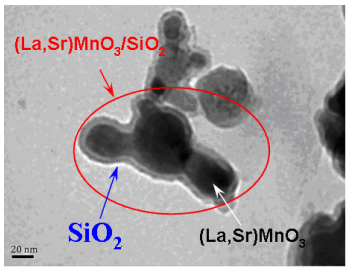


Fig. 6. TEM image of $La_{1-x}Sr_xMnO_3/SiO_2$ core/shell structure

Magnetic ***investigations*** were carried out in a wide temperature range (Fig. 7).

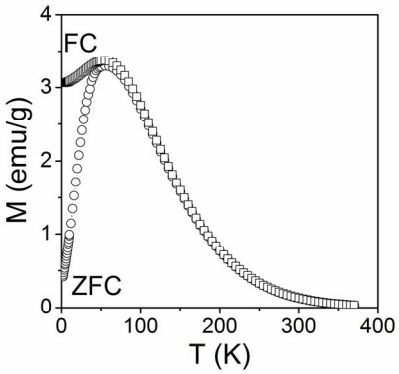


Fig. 7. Plots of magnetization against temperature for $ZnFe_2O_4$ nanoparticles obtained in 20 Oe probing field after cooling in zero magnetic field (ZFC) and in magnetic field $H = H_{measur}$ (FC).

It was shown that nanoparticles are characterized by low coercivity with low saturation fields as indicated by hysteresis loops. In addition, it was also shown that the obtained curves are described by the Langevin formula, which indicates a superparamagnetic behavior of the nanopowders (Fig. 8).

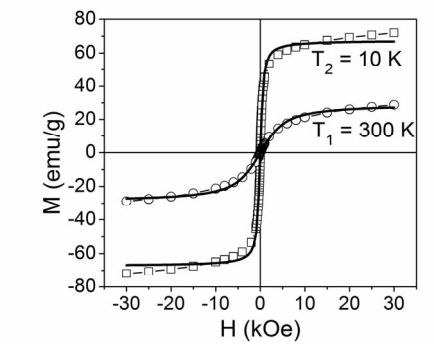


Fig. 8. Hysteresis loops measured at 300 K and 10 K for $ZnFe_2O_4$ nanopowder (circles and squares, respectively) and also fitted hysteresis loops obtained by approximation of experimental data by Langevin function (solid lines).

The investigation of magnetic properties of slightly agglomerated nanoparticles of $La_{1-x}Sr_xMnO_3$ and core/shell nanostructures ($La_{1-x}Sr_xMnO_3/SiO_2$) showed that they are also characterized by low coercive force and exhibit properties close to superparamagnetic ones. ***The properties of ferrofluids*** developed on the basis of synthesized slightly agglomerated nanoparticles of $La_{1-x}Sr_xMnO_3$, AFe_2O_4 (A = Mn, Fe, Co, Ni, Zn), $La_{1-x}Sr_xMnO_3/SiO_2$ core/shell structures and aqueous agarose solution have been studied. As a result of the investigations carried out, the optimal concentrations of nanoparticles in organic-matrix solutions at which the fluid remains stable and can not separate into layers have been determined.

For ferrofluids based on nanoparticles of $La_{1-x}Sr_xMnO_3$ it has been shown that heating of magnetic fluids under the influence of an external electromagnetic field in the frequency range 100-400 kHz occurs up to the required temperature (≈ 45°C) for 5-20 min. Also it was found that the temperature of the synthesis of magnetic particles affects the temperature of the heating of magnetic fluids. This regularity may be accounted for by a change in the Mn^{3+}/Mn^{4+} ratio, which leads to a change in Curie temperature and conductivity, and by a change in particle size.

Fig. 9 shows the results of heating temperature measurement of ferrofluids created on the basis of $La_{1-x}Sr_xMnO_3$ nanoparticles synthesized by sol-gel method at a heat treatment temperature of 800°C and coated with three SiO_2 layers. The properties of magnetic fluids based on particles without SiO_2 coating are presented for comparison. As is evident from the results obtained, the coating of nanoparticles does not affect the heating temperature of magnetic fluid in comparison with the as-prepared sample and affects only the time during which the magnetic fluid is heated to a constant temperature.

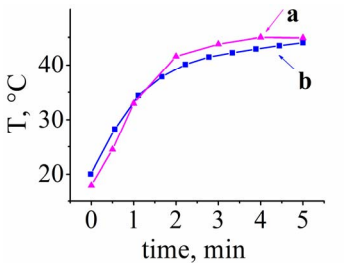


Fig. 9. Plots of heating temperature against time for magnetic fluids based on $La_{1-x}Sr_xMnO_3$ lanthanum-strontium manganite (a) and core/shell $La_{1-x}Sr_xMnO_3/SiO_2$ nanostructures (b).

It has been shown that heating of magnetic fluid occurs under the influence magnetic field to a certain temperature, the value of which can be controlled by changing the chemical composition and properties of ferromagnetic nanoparticles. Using conditions in vivo, the toxicity of magnetic fluids based on $La_{1-x}Sr_xMnO_3$ nanoparticles was studied. As a result of the investigations carried out, it has been found that maximum tolerable dose (recalculated as ferromagnetic nanoparticles) is 300 mg/kg of body weight [13].

Using the developed magnetic fluid, studies of the effect of nanohyperthermia on the Guerin carcinoma was performed. Guerin carcinoma was transplanted into the femur muscle. When tumors have reached the volume of 1.0-1.5 cm^3 rats were unrolled into the experiment. Magnetic fluid was given into the tumor at a dose of 0.6 ml followed by alternating magnetic field (300 kHz, 7.7 kA/m) that was affected during 20-30 min (the treatment was repeated 2-4-times with 1-3 days break). Such treatment has resulted in the delay of tumor growth – tumor volume on the 13th day after the first treatment was: in treated rats – 2.7 cm^3, in the untreated rats - 47 cm^3. Temperature in the tumor that was injected by magnetic fluid and treated by alternating magnetic field was in average 44-45°C at the 30th min of the influence. 24 rats were used in this experiment. For the more clear comparison the tumor growth after such treatment with untreated one the Guerin carcinoma was transplanted both in the left and right legs of the same rat in the separate experiments (total number of rats – 8). The results of one of such experiments was presented in Fig.10. The control groups were used: 1) untreated tumor-bearing rats; 2) rats with tumor treated with alternating magnetic field alone; 3) rats with tumor treated with magnetic field alone. The results were presented in Table. It has to be noted that tumor growth was not affected in all control groups.

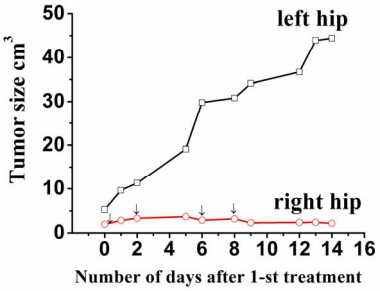


Fig. 10 Growth of Guerin carcinoma after action by magnetic fluid combined with alternating magnetic field (treatment at the 0, 2nd, 6th, 8th days) and growth of intact tumor in the same rat (tumor was transplanted into the both femurs). Red line – growth of treated tumor, black line – growth of intact tumor. Treatment were indicated by arrows.

TABLE Results of nanohyperthermia influence on Guerin carcinoma

Tumor	***Index of tumor growth inhibition (by n-fold)***	***Tumor growth delay (day)***	***Number of rats with complete tumor regression (%)***	***Number of cured rats (%)****
Untreated tumor	0	0	0	0
Nanohyperhermia	3,2	7,5	25*	25*

* - rats with complete tumor regression and without tumor recurrence within 120 days after complete regression of the tumor.

IV CONCLUSION

Thus, ferromagnetic nanoparticles synthesized by different methods (sol-gel, precipitation from nonaqueous solutions and microemulsion solutions) and ferromagnetic fluids based on them exhibit properties that allow one to consider them as potentially promising inducers of hyperthermia of deep-seated tumors and require further research.

REFERENCES

[1] S.P. Gubin, Y.A. Koksharov, G.B. Khomutov, G.Y. Yurkov, "Magnetic nanoparticles: preparation methods, structure and properties," Uspekhi Khimii, vol. 74, pp. 539-574, 2005.

[2] T. S. Cho, S. J. Doh, and J. H. Je, and D.Y. Noh, "Thickness dependence of the crystallization of Ba-ferrite films," J. Appl. Phys. vol. 86, pp. 1958-1964, 1999.

[3] J.S. Weinstein, C.G. Varallyay, E. Dosa, S. Gahramanov, B. Hamilton, W.D. Rooney, L.L. Muldoon, E.A. Neuwelt, "Superparamagnetic iron oxide nanoparticles: diagnostic magnetic resonance imaging and potential therapeutic applications in neurooncology and central nervous system inflammatory pathologies, are view," J. Cereb. Blood Flow Metab., vol. 30, pp. 15-35, 2010.

[4] Ibrahim Sharifi, H. Shokrollahi, S. Amiri, "Ferrite-based magnetic nanofluids used in hyperthermia applications," Journal of Magnetism and Magnetic Materials, vol. 324, pp. 903-915, March 2012.

[5] Jian Li, Dalin Dai, Baogang Zhao, Yueqiang Lin and Cenye Liu, "Properties of Ferrofluid Nanoparticles Prepared by Coprecipitation and Acid Treatment," Journal of Nanoparticle Research, vol. 4, pp. 261-264, 2002.

[6] S. Laurent, S. Dutz, Urs O. Häfeli, M. Mahmoudi, "Magnetic fluid hyperthermia: Focus on superparamagnetic iron oxide nanoparticles," Advances in Colloid and Interface Science, vol. 166, pp. 8-23, 2011.

[7] S.A. Solopan, O.I. V'yunov, A.G. Belous, A.I. Tovstolytkin, L.L. Kovalenko "Magnetoelectric effect in composite structures based on ferroelectric–ferromagnetic perovskites," Journal of the European Ceramic Society, vol. 30, pp. 259-263, 2010.

[8] A.G. Belous, O.I. V'yunov, E.V. Pashkova, O.Z. Yanchevskii, A.I. Tovstolytkin, A.M. Pogorelyi, "Effects of Chemical Composition and Sintering Temperature on the Structure of $La_{1-x}Sr_xMnO_{3\pm\gamma}$ Solid Solutions," Inorganic Materials, vol. 39, pp. 161-170, 2003.

[9] D. Caruntu, Y. Remond, Nam Hawn Chou, Moo-Jin Jun, G. Caruntu, Jibao He, G. Goloverda, Charles O'Connor, and V. Kolesnichenko, "Reactivity of 3d transition metal cations in diethylene glycol solutions. Synthesis of transition metal ferrites with the structure of discrete nanoparticles complexed with long-chain carboxylate anions," Inorganic Chemistry, vol. 41, pp. 6137-6146, 2002.

[10] S.A. Solopan, Ye.D. Fateev, A.G. Belous, "Peculiarities of the synthesis of low-agglomerated nanoparticles Fe_3O_4 from microemulsions," Ukrainian Chemistry Journal, vol. 78, pp. 3-8, 2012.

[11] Brian L. Cushing, Vladimir L. Kolesnichenko, and Charles J. O'Connor, "Recent Advances in the Liquid-Phase Syntheses of Inorganic Nanoparticles," Chem. Rev., vol. 104, pp. 3893-3946, 2004.

[12] S. Mornet, C. Elissalde , V. Hornebecq, O. Bidault, E. Duguet, A. Brisson, and M. Maglione, "Controlled Growth of Silica Shell on $Ba_{0.6}Sr_{0.4}TiO_3$ Nanoparticles Used As Precursors of Ferroelectric Composites," Chem. Mater., vol. 17, pp. 4530-4536, 2005.

[13] S. Solopan, A. Belous, A. Yelenich, L. Bubnovskaya, A. Kovelskaya, A. Podoltsev, I. Kondratenko, S. Osinsky, "Nanohyperthermia of malignant tumors. I. lanthanum-strontium manganite magnetic fluid as potential inducer of tumor hyperthermia," Exp. Oncol., vol. 33, pp. 130-135, 2011.

[14] L. Harivardhan Reddy, Jose L. Arias, Julien Nicolas, and Patrick Couvreur, "Magnetic nanoparticles: design and characterization, toxicity andbiocompatibility, pharmaceutical and biomedical applications," Chem. Rev. vol. 112, pp. 5818-5878, 2012.

[15] W. Stober, A. Fink, E. Bohn, "Controlled Growth of Monodisperse Silica Spheres in the Micron Size Range," J. Colloid Interface Sci., vol. 26, pp. 62-69, 1968.

Combined Method for Detection of Atrial Late Potentials

Matveyeva N. A., Ivanushkina N. G., Ivanko K. O.
National Technical University of Ukraine "Kyiv Polytechnic Institute"
Physical and Biomedical Electronics
Kyiv, Ukraine
matveyeva.nataliya@gmail.com

***Abstract* — The work is devoted to improvement of methods for noninvasive identification of low-amplitude components of electrocardiogram (ECG) – atrial late potentials (ALP) which are markers of potentially dangerous heart rhythm disorders.**

A combined method for ALP detection based on wavelet analysis, decomposition in the basis of eigenvectors and classification by neural network is proposed. As the result of the numerical experiment the dimension of ALP feature vector was minimized that made it possible to distinguish between 2 classes "norm − no ALP" and "pathology − ALP are present" with minimum error classification.

Keywords* — *atrial late potentials; eigenvectors basis; neural network; high-resolution electrocardiography; markers of atrial tachyarrhythmia; wavelet analysis.

I. Introduction

Development of means for digital processing of biomedical signals enables to improve the diagnostic methods in cardiology. Special attention should be paid to non-invasive methods of myocardial electrical instability detection in order to further cardiac arrhythmias prediction, since arrhythmia is the most common disease of the cardiovascular system.

It is believed that the first signs of electrical instability of the heart reflect the depletion of regulatory systems at the level of individual cells of the myocardium. These primary changes cannot be detected by traditional diagnostic methods in clinical cardiology.

High-resolution electrocardiography (HR ECG), which is based on special signal averaging techniques and digital processing tools, is able to provide information about the primary electrical changes in the myocardium [1,2]. With the advent of the HR ECG detection of low-level signals on the order of 1-10 μV, so-called "late potentials", attracted widespread clinical attention. Atrial late potentials (ALP) are known as markers of cardiac tachyarrhythmia. In surface recordings ALP appears as low-amplitude (about several μV) high-frequency potentials which are localized in the terminal part of ECG *P* wave. These signals contain frequency components in band 40–250 Hz. The emergence of ALP in ECG signal reflects the slow and fragmented atrial depolarization associated with the "re-entry" circles and development of atrial tachyarrhythmia.

To reveal ALP, considerable attention should be paid to appropriate signal processing. ALP detection could be based on applying of a broad class of algorithms: from the traditional analysis in both time and frequency domains up to the complex pattern recognition procedures [3, 4].

The most common methods in electrocardiology are time-domain analysis with measurement of amplitudes and durations of ECG components, spectral analysis with ECG frequency content estimation and wavelet analysis which allows data in both domains (time and frequency) to be analyzed at the same time. The diagnostic value of these methods is low. More accurate results are obtained by a combination of different methods such as wavelet analysis and decomposition in the basis of the eigenvectors [2-3, 5, 6].

II. Description of the Proposed Combined Method

The work aims to improve information and algorithmic support for noninvasive ALP detection in HR ECG systems using an integrated method: a combination of the methods of electrocardiosignals analysis [2-5, 7-9] and neural networks [10]. General scheme of the proposed combined method for ALP detection is shown in Figure 1.

The proposed method includes the following stages:

1. ECG preprocessing (removing baseline wander and powerline interference, accumulation of cardiocycles for ECG signal averaging);
2. Wavelet decomposition up to the 5^{th} levels and selection of the first level detail coefficients cD_1 (Fig. 2);
3. Decomposition in the basis of eigenvectors and the main vectors selection;
4. ALP features extraction and training set formation;
5. Neural network training and ECG classification using neural network algorithm.

According to the method of HR ECG signal preprocessing includes registration and averaging from 100 to 400 ECG complexes [1, 2]. Assuming the independence of noise

samplings in N averaged cardiocycles, the noise variance in the signal-averaged ECG decreases in N times, i.e. signal to noise ratio is improved in $\sqrt{N}$ times.

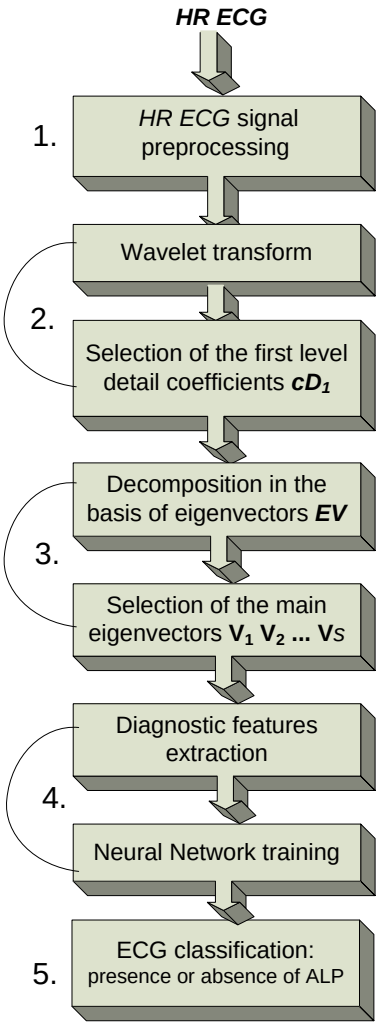


Fig. 1. General scheme of complex method of ECG analysis to identify ALP as the combination of the wavelet transform, decomposition in the coordinate basis of eigenvectors and classification by neural network

The proposed combined method takes into account that the multiresolution wavelet analysis is effective for the study of ECG signals containing extended high amplitude low-frequency components as the main waves and high frequency components of short duration as ALP. It was revealed that detail coefficients cD_1 of the first level discrete wavelet decomposition reflect the pattern of high-frequency low-amplitude potentials of *P*-wave and thus contain the signs of ALP appropriate for their recognition [4, 7, 9].

In the next step of the complex method of ECG analysis the detail coefficients cD_1 were decomposed in the coordinate basis of eigenvectors [3, 8]. It was done with the purpose to separate diagnostically important signal and noise components which appear at the same level of wavelet decomposition of *P*-wave. In eigenvectors basis the subspace of signal contains the information about ALP and the noise subspace corresponds to noise components.

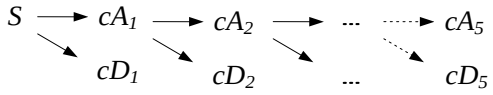


Fig. 2. Wavelet decomposition tree up to the 5th levels

The ensembles of analyzed *P*-waves were formed in the way that each *i*-th *P*-wave was presented by the vector of detail wavelet coefficients cD_1. Such the set of detail wavelet coefficients cD_1 is a matrix:

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1N} \\ w_{21} & w_{22} & \dots & w_{2N} \\ \vdots & \vdots & & \vdots \\ w_{M1} & w_{M2} & \dots & w_{MN} \end{bmatrix}, \qquad (1)$$

where w_{ij} – the value of the *j*-th wavelet coefficient of the first level of detail for the *i*-th *P*-wave, N – the number of elements in the vector of detail coefficients corresponding to the first level of wavelet decomposition.

Covariance matrix of wavelet coefficients corresponding to the ensemble of *P*-waves is:

$$\mathbf{C}_w = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1N} \\ C_{21} & C_{22} & \dots & C_{2N} \\ \vdots & \vdots & & \vdots \\ C_{N1} & C_{N2} & \dots & C_{NN} \end{bmatrix}, \qquad (2)$$

where the elements C_{ij} of the matrix $\mathbf{C}_w$ can be defined as

$$C_{ij} = \sum_{k=1}^{N} (w_{ki} - m_k)(w_{kj} - m_k), \qquad (3)$$

where m_k – the average value of wavelet coefficients cD_1 calculated for the *k*-th *P*-wave.

The sets of eigenvalues $\lambda_1, \lambda_2, ..., \lambda_N$ and eigenvectors $\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_N]$ were obtained for each covariance matrix in order to define features that enable to discriminate between ECG signals with normal *P*-waves and *P*-waves including ALP.

The coefficients of decomposition in the coordinate basis of eigenvectors (projections on the corresponding coordinate axis) are defined as

$$b_{ij} = \left\langle \mathbf{W}_i \cdot \mathbf{V}_j \right\rangle = \sum_{k=1}^{N} W_{ik} V_{jk}. \qquad (4)$$

The reconstructed set of wavelet coefficients for *P*-waves can be defined as the sum of the components which are determined by the eigenvectors and the coefficients of the decomposition:

$$\mathbf{W}_i^* = \sum_{j=1}^{S} b_{ij}\mathbf{V}_j, \qquad (5)$$

where $\mathbf{W}_i^*$ – recovered set of detail wavelet coefficients for the i-th P-wave; S – selected number of main eigenvectors $\mathbf{V}$.

To separate diagnostically important signal with the signs of ALP and noise component it is proposed to use eigenvectors and eigenvalues of decomposition in the coordinate basis of eigenvectors. To estimate the ECG signal with the desired information about ALP and remove noise components S main eigenvectors $\mathbf{V}_1$, $\mathbf{V}_2$…, $\mathbf{V}_S$ which correspond to the largest eigenvalues λ_1, λ_2 ..., λ_S should be used. Other vectors of the eigenvector basis characterize the noise subspace which is a mixture of the less significant ECG signal components with the physiological disturbances and uncorrelated measurement noise. This approach allows combining the effective noise suppression and ALP extraction by the division of high-frequency ECG components into the diagnostically important signal and noise.

Hence, ALP feature selection procedure consists of evaluation of the eigenvalues and eigenvectors of the covariance matrix found for the set of wavelet coefficients cD_1 which are defined for each P-wave in the ensemble of ECG realizations. To minimize the dimension of ALP feature vector the largest eigenvalues λ_1, λ_2, …, λ_S ($\lambda_1>\lambda_2>\ldots>\lambda_S$) are supposed to be used for ECG classification.

Considering the complexity of ALP detection it is suggested to construct a classifier for HR ECG system with neural networks using (NN) [10] which based on flexible algorithms for solving complex and nonlinear tasks.

Design of NN implies the processes of a choice of NN architecture and algorithm for NN training, the input training set formation and the creation of the objective function. For ECG classification (Fig. 3) it is proposed to use NN architecture which has the input, hidden and output layers of neurons with weights Q_{ij} that determine the value of i-th synaptic connection of j-th neuron.

The learning process includes the adjustment of Q_{ij} weights. The vectors from the training set are fed into the neural network in order to approximate the network outputs to their target values defining the different classes of diseases in cardiology:

$$\mathbf{R} = f\left(\mathbf{IQ}+\mathbf{D}\right) = f\left(\mathbf{T}\right), \qquad (6)$$

where $\mathbf{R}$ – vector of output values of NN; $\mathbf{I}$ – training set; $\mathbf{Q}$ – weighting matrix; $\mathbf{D}$ – displacement vector; f – sigmoid activation function applied to the components of the vector $\mathbf{T}$.

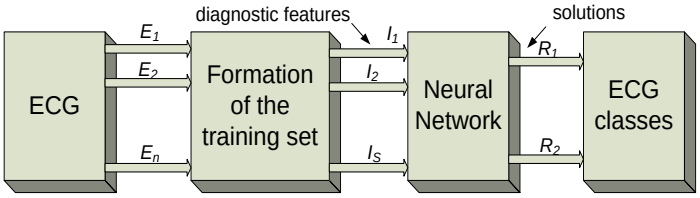


Fig. 3. ECG classification by neural network

The number of test patterns in the training set can be defined as:

$$2\left(n_{inp}+n_{hid}+n_{out}\right) \le N \le 10\left(n_{inp}+n_{hid}+n_{out}\right), \qquad (7)$$

$$n_{hid} = \sqrt{n_{inp}\cdot n_{out}} \qquad (8)$$

where n_{inp}, n_{hid}, n_{out} – the number of neurons in the input, hidden and output layers.

In the neural network for ECG classification the number of input neurons n_{inp} corresponds to the dimension of feature vector and the number of output neurons n_{out} corresponds to the number of classes.

III. Numerical Experiment for Detection of Atrial Late Potentials

In order to detect the signs of slow atrial depolarization in HR ECG signal with noise components the experiment was carried out in MATLAB system (Fig. 4).

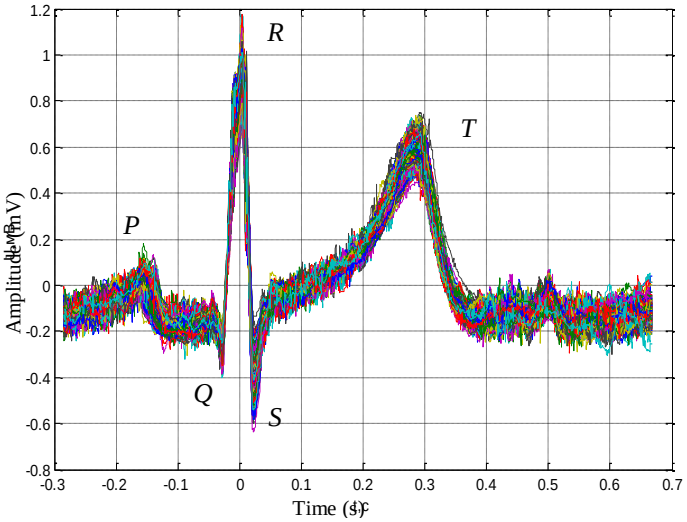


Fig. 4. Accumulation of cardiocycles synchronized on R-wave

340 P-waves extracted from ECG signals of several patients were used as data set for classification. 170 P-waves represented the norm (Fig. 5, a). 170 other P-waves represented the pathology and contained simulated ALP which were different in their amplitude, shape and duration (Fig. 5, b). To simulate ALP in some cases the Hodgkin-Huxley equations were solved and the action potentials with low amplitude were added to the terminal part (30 ms) of P-waves (Fig. 6, a). In another cases ALP activity was simulated as the sum of sinusoids with frequency range from 40 to 250 Hz and amplitudes 1-10 μV (Fig. 6, b):

$$x(t) = A_1 Sin(2\pi f_1 t + \varphi_1) + A_2 Sin(2\pi f_2 t + \varphi_2) + \ldots + A_n Sin(2\pi f_n t + \varphi_n), \qquad (9)$$

where A_1 - A_n – amplitudes of the harmonic components of ALP; f_1 - f_n and φ_1 - φ_n – their frequencies and phases.

The noise in HR ECG recordings was simulated by superimposing *P*-waves and a Gaussian additive noise signal with signal to noise ratio of 65 dB.

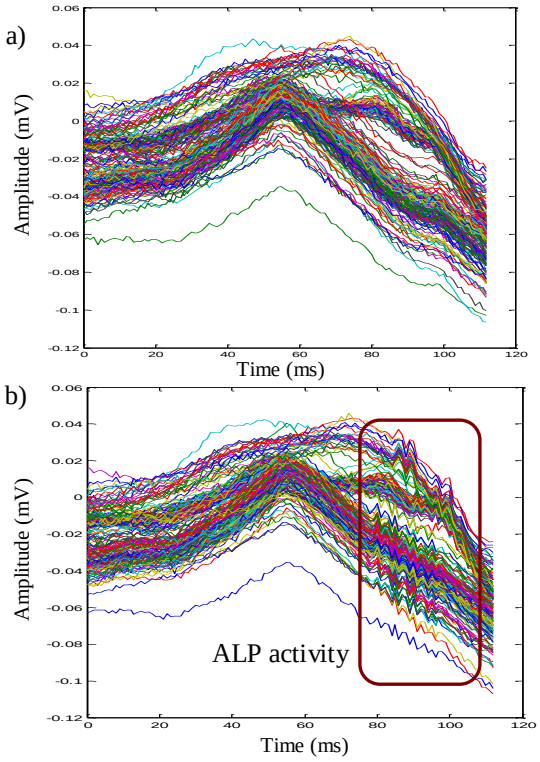


Fig. 5. *P*-waves extracted from real ECG (signal to noise ratio is 65 dB): a) without ALP, b) with simulated ALP

As a result of discrete wavelet decomposition of *P*-waves up to the 5th levels the approximation coefficients that represent the smoothed ECG signals and detail coefficients that describe high frequency oscillations were obtained (Fig. 7). Data dimension can be reduced by removing large-scale wavelet-coefficients which responsible for the form of high-amplitude (in comparison with ALP) *P*-wave. Model experiments showed that the signs of ALP can be found at d_1 level of detail representing the high-frequency low-amplitude bursts. The other levels of discrete wavelet decomposition reflect the features of lower frequency components. For the purpose of further compression of information and separating ALP from noisy background the method of eigenvectors was used.

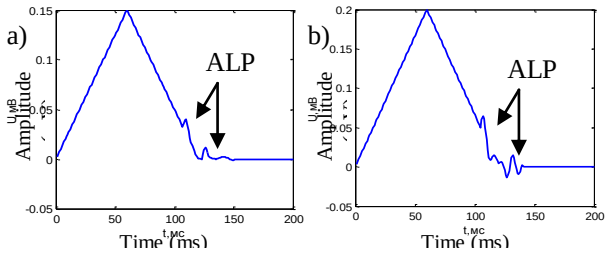


Fig. 6. ALP simulation: a) low amplitude action potentials in the terminal part of *P*-wave; b) sum of sinusoids in the terminal part of *P*-wave

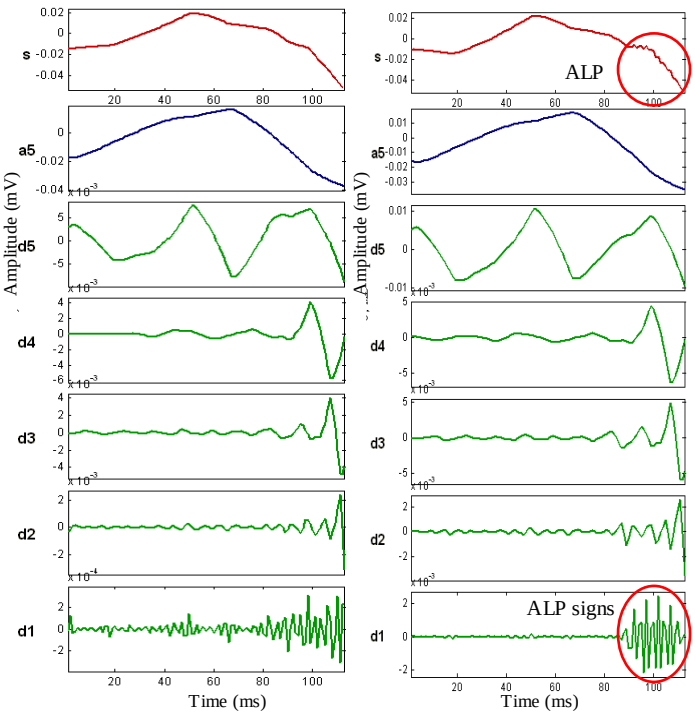


Fig. 7. Wavelet-analysis of *P*-waves: a) *P*-wave without ALP; b) *P*-wave with simulated ALP

Eigenvalues of covariance matrix (2) were obtained in result of wavelet detail coefficients cD_1 decomposition in eigenvectors basis. Using these eigenvalues as feature vector for ALP pattern recognition the training samples for neural network were formed.

Differences between the eigenvalues for *P*-waves with ALP and without late potentials are shown in Fig.8.

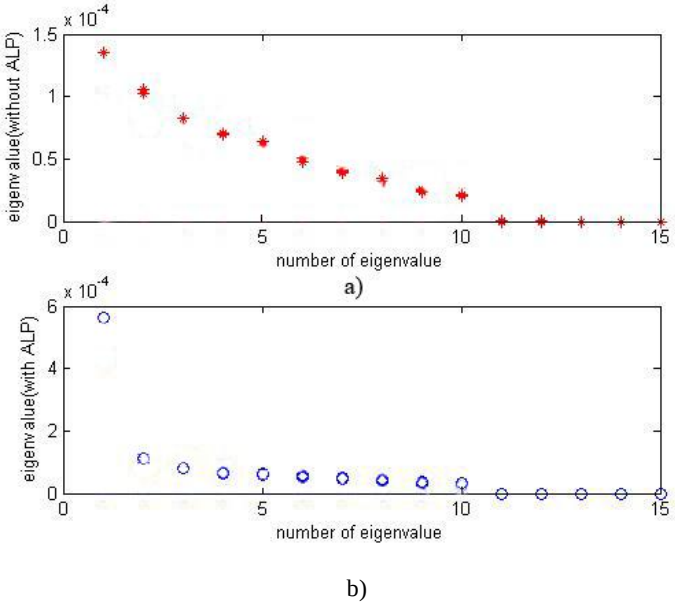


Fig. 8. Eigenvalues for ensembles of analyzed *P*-waves: a) without ALP; b) with ALP

Diagrams show the eigenvalues for ensembles of *P*-waves without ALP (Figure 8 (a)) and with ALP (Fig. 8 (b)). The values of the main eigenvalues (first 5-10) for the case with the presence of ALP is approximately 4 times greater than in the case without ALP.

At the stage of ECG classification based on ALP detection a multilayer perceptron can be used. Multilayer network consists of three layers (Fig. 9): input, hidden and output. The output layer has two neurons (2 outputs). If the active level (1) is detected on the first output this means that ALP are present in ECG, if the active level appears on the second output then ECG signal doesn't contain ALP in *P*-waves.

The input layer has *S* neurons what corresponding to the number of the elements which has ALP feature vector in the training set. Hidden layer (middle layer) has a variable number of neurons depending on the number of inputs. In our experiment for detection of atrial late potentials *S* was determined as 5.

To train the neural network 140 *P*-waves with ALP and 140 *P*-waves without ALP were taken. To check the results of classification 30 *P*-waves with atrial late potentials and 30 *P*-waves without ALP, which were not included in the training set, were used.

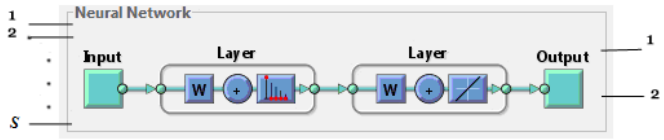


Fig. 9. Architecture of the neural network for HR ECG classification into 2 classes: "pathology − ALP are present" and "norm − no ALP"

The results of neural network training with input vector formed from eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_5$ are presented in the Figure 9.

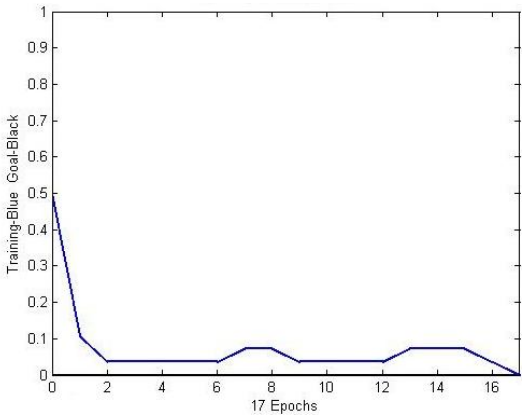


Fig. 9. Neural network training

The result of neural network training in case of using $\lambda_1, \lambda_2, \ldots, \lambda_5$ as inputs was stable for the classification of real ECG signals into 2 classes: "pathology − ALP are present" and "norm − no ALP".

The method was tested in numerical experiments and showed the correct classification results. For the experiments, we used the results of clinical trials (real HR ECG recordings with a sampling frequency of 2 kHz).

To check the validity of the classification results the method based on the discriminant functions was used as a reference [3].

Conclusion

In this study the combined method for ECG analysis and its further classification into 2 classes "pathology − ALP are present" and "norm − no ALP" was proposed. The method is based on the wavelet transformation, decomposition in eigenvectors basis and neural networks application. The main points of the work are: revealing of ALP signs on the detail level of wavelet decomposition, ECG separation into diagnostically important signal and noise components by decomposition in eigenvectors basis and finding the features of ALP suitable for neural network training.

The results obtained in this work are encouraging and make it possible to improve the ability to classify ECG signals despite the low amplitude of atrial late potentials. The further research will include analysis of the proposed method application in large- scale clinical trials for early noninvasive diagnostics of cardiac arrhythmias.

References

[1] L. A. Bokeria, E.Z. Golukhova, A.V. Ivanitsky , "Functional diagnostics in cardiology" , in 2 parts , Publishing SCCS Bakulev RAMS, 2002, p. 427. (in Rus)

[2] N.G. Ivanushkina, V.A. Fesechko, "Technologies in high resolution electrocardiography", a tutorial, NTUU "KPI", 2007, p. 116. (in Ukr)

[3] K. O. Ivanko, N.G. Ivanushkina , "The Detection Method for Atrial Late Potentials Based on Eigenspaces Analysis of ECG Wavelet Pattern ", Scientific news, 2010, № 6, pp. 11-18. (in Ukr)

[4] K. Ivanko, N. Ivanushkina, N. Matveeva, S. Sporevoi, "Atrial late potentials pattern recognition: approaches to classification", Proceedings of the International Conference "Electronics and Nanotechnology", Kyiv, 2011. p. 194.

[5] Tou J. T., R. C. Gonzalez , "Pattern recognition Principles", Addison-Wesley Publishing Company, p. 411.

[6] Wiley Encyclopedia of Biomedical Engineering [ed. M. Akay], 2006, p. 4152.

[7] C. Saritha, V. Sukanya, Y. N. Murthy, " ECG Signal Analysis Using Wavelet Transforms", Bulg. J. Phys., № 35, 2008, pp. 68- 77.

[8] F. Castells, J. Lorenz, A. M. Climent , "Morphological Descriptors Based on Eigen Value Decomposition for P-Wave Analysis" [et al.], Computers in Cardiology, 2008, № 35, pp. 245-248.

[9] Time-Frequency and Wavelets in Biomedical Signal Processing [ed. Metin Akay], IEEE Press, 1997,p. 652.

[10] A. Mousa, A. Yilmaz , "Neural Network Detection of Ventricular Late Potentials in ECG Signals Using Wavelet Transform Extracted Parameters", IEEE Proceedings of the 23rd Annual EMBS International Conference, 2001, pp. 1668-1671.

Heat Production, Heat Transfer and Heat Exchange in Man During Water Immersion: Mathematical Modeling

IrenaYermakova, Yuliia Solopchuk
Department of Information Technologies of Physiology
International Scientific - Training Centre for Information Technologies and Systems NAS of Ukraine
Kyiv, Ukraine
irena.yermakova@gmail.com, julia.solopchuk@gmail.com

Lyudmila Khudyakova
Physical and Biomedical Electronics Department
National Technical University of Ukraine "KPI"
Kyiv, Ukraine

Abstract — **The multicompartmental mathematical model of heat production, heat transfer and heat exchange in human during water immersion and air environment has been developed for prediction of dynamic changes of human temperature and other physiological parameters.**

Keywords — heat production; cooling water; model.

I. INTRODUCTION

In case of a human getting into water, survival time is limited due to very quick hypothermia. It is explained by thermal conductivity of water that is twenty four times as many as air. That is why evaluation and prediction of human thermal state is of high importance.

The first mathematical model describing the physiological thermal state of a human in water was developed by Montgomery in 1974 [1]. From that time this model is improved and adapted to specific tasks of authors. Tikuisis and Keefe [2] developed mathematical model for evaluation of thermal physiological state of soldiers in water for the USA military forces. Xu et al. [3, 4] made thermoregulatory model for evaluation of human thermal state in water and developed software based on this model for prediction of survival time of victims for rescue-and-recovery service. Although there are some studies deal with water protection clothing [5, 6]. In order to estimate the insulating properties of clothing and help in clothing designing mathematical thermoregulation model needs to be used.

II. AIM AND METHODS

The aim of this study is the development of computer model for prediction of dynamic changes of human temperature and other physiological parameters in water immersion and air environment.

A. Limitations and assumptions

Current mathematical model is based on the class of multicompartmental dynamic models of thermoregulation developed by I. Yermakova [7].

In order to develop mathematical model limitations and assumptions should be stated. These limitations and assumptions are below.

The anthropometrical shape of the human body is approximated as a set of cylindrical compartments. Head is approximated by sphere. A cylinder is divided into a few layers corresponding to organs and tissues of the human body. Two indices *ij* are used, where *i* corresponds to the cylinder number, and *j* corresponds to the cylinder layer.

Conduction heat transfer between the end faces of cylinders is not taken into account as the end face of elementary cylindrical volume is incommensurably smaller than the lateral surface. So heat flow from the end faces in the heat balance equation can be neglected. And this won't introduce significant inaccuracy.

Blood flows in each compartment are described as a system of radiated concurrent branches. Heat flows transferring by blood mix in the blood pool. This blood pool includes great veins of circulation heart. The inflow blood temperature in all compartments is equal to temperature of mixed venous blood. Arterial and venous blood flows are the same.

Free and forced water convection is taken into consideration during immersion of human.

B. Heat production.

Metabolic heat production for *ij* – compartment is determined by oxygen consumption and caloric equivalent of substrate. The formula is:

$$M_{ij} = k_{ij}^{eq} qO_2 \qquad (1)$$

where qO_2 is the oxygen consumption; k^{eq}_{ij} is the caloric equivalent of oxygen of the *ij* - compartment.

C. Heat conduction.

Conductive flux through the lateral surface of a cylinder is described as following:

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

$$Q_{ij}^{K} = -2\pi r L \lambda \left. \frac{dT_{ij}}{dt} \right|_{r} \quad (2)$$

where λ is the heat conductivity coefficient, L is the length of a cylinder, r is the direction of conduction, T_{ij} is the compartment temperature.

Equation (2) can be written as finite-difference equation. If R_{ij+1} is the mean radius of $j+1$-layer, and R_{ij} is the mean radius of j -layer, then the heat conductive flux from ij-compartment to compartment of $ij+1$ is:

$$Q_{ij}^{K} = \lambda \alpha (T_{ij+1} - T_{ij}) \quad (3)$$

where

$$\alpha = \frac{2\pi R_{ij} L_{ij}}{R_{ij+1} - R_{ij}} \quad (4)$$

D. Convective heat transfer by blood.

Heat transfer in the human body takes place mainly via forced convection by blood flow heat and conductive properties of the tissues and the organs.

Heat transfer by blood is defined as following:

$$Q_{ij}^{B} = c_b \rho_b w_{ij} (T_b - T_{ij}) \quad (5)$$

where c_b is the blood specific heat; ρ_b is the blood density; w_{ij} is the volume blood flow of ij-compartment, T_b is the blood temperature.

E. Heat exchange of human body in water

The main way of heat exchange of human body in water is free and forced convection.

Heat transfer by free convection takes place if the movement of fluid particles under the temperature difference evokes the difference of densities. Heat transfer by forced convection occurs under the condition of pressure difference.

Convection heat transfer of ij-compartment is defined by:

$$Q_{ij}^{C} = h_{ij}^{C} A_{ij} (T_{ij} - T_{ie}) \quad (6)$$

where h_{ij}^{C} is the convective heat transfer coefficient; A_{ij} is the skin surface; T_{ie} is the environment temperature.

Computation of convective heat transfer coefficient is essential problem for water.

Convection heat transfer coefficient for water environment is calculated as [8]:

$$h^{CW} = \frac{K_w Nu}{d} \quad (7)$$

where Nu is the Nusselt number, K_w is the thermal conductivity of water, d is the diameter.

Free and forced convection takes place depending on water conditions and human behavior. The Nusselt number has different meanings for free and forced convection.

F. Heat exchange in air

1) Convection heat transfer. Convection heat transfer coefficient for air environment is calculated as:

$$h^{CA} = h^{CA*} \left(\frac{v_a}{0.1} \right)^{0.5} \quad (8)$$

where v_a is the air velocity, h^{CA*} is the heat transfer coefficient measured for naked human at air velocity $v_a=0.1$ *m/s* and thermoneutral ambient temperature:

$$h^{CA*} = 0.53 \left(\frac{\rho^2 \beta g c}{\mu d} \right)^{0.25} \lambda^{0.75} \quad (9)$$

where ρ is the air density, g is acceleration of gravity, β is the thermal expansion coefficient of the air; c is the specific heat of air; μ is the kinematic viscosity of air; d is the diameter of cylinder.

2) Heat radiation. Radiation heat exchange between human body and environment is calculated by the Stefan-Boltzmann law.

If $(T_{ij} - T_{iA}) < 20^{\circ}C$ heat radiation flow is calculated by a linear dependence. Therefore, heat radiation is calculated by empirical formula:

$$Q_{ij}^{R} = h_{ij}^{R} A_{ij} (T_{ij} - T_{ie}) \quad (10)$$

where h_{ij}^{R} is the radiation heat exchange coefficient determined by the Stefan-Boltzmann constant, emissive ability of skin and temperature spreading factor.

3) Evaporation from skin. Evaporation heat loss depends on the heat transfer coefficient, area of wetted surface, vapor pressure saturated at skin and environmental temperatures:

$$Q_{ij}^{E} = h_{ij}^{E} A_{ij} (P_{ij} - P_{ie}) \quad (11)$$

where h_{ij}^{E} is the evaporation heat transfer coefficient for human body surface; P_{ie} is the vapor pressure saturated at environment temperature.

4) Evaporation heat loss from upper respiratory ways.

Evaporation heat loss from upper respiratory ways contains inspired and expired air and evaporation heat loss. Respiratory heat loss is 13% approximately of total heat loss of human:

$$Q_{ij}^{RS} = \dot{V} \cdot \nu \cdot (y_e - y_b) + \dot{V} \cdot c_e \cdot (T_e - T_b) \quad (12)$$

where y_e, y_b are the moisture content of inspired and expired air, respectively; $\dot{V}$ is the pulmonary ventilation, ν is the blood viscosity.

Dynamics of heat production, heat transfer and heat exchange during water immersion for *ij*-compartment is following:

$$\begin{aligned} & c_{ij} m_{ij} \frac{dT_{ij}}{dt} = k_{ij}^{eq} q O_2 + \alpha_{ij-1}\lambda_{ij-1}(T_{ij-1} - T_{ij}) - \\ & -\alpha_{ij}\lambda_{ij}(T_{ij} - T_{ij+1}) + c_b \rho_b w_{ij}(T_b - T_{ij}) - \\ & -h_{ij}^{CW} A_{ij}(T_{ij} - T_{iW}) - h_{ij}^{CA} A_{ij}(T_{ij} - T_{ie}) - \\ & -h_{ij}^{R} A_{ij}(T_{ij} - T_{ie}) - h_{ij}^{E} A_{ij}(P_{ij} - P_{ie}), \quad i = \overline{1,N}, j = \overline{1,K} \end{aligned} \quad (13)$$

where N – the number of cylinders; K - the number of layers.

Number of equations (13) is the same as number of compartments.

G. Blood temperature

Heat transfer by blood in human body has principal effect on human temperature and heat balance. Heat exchange in large veins is determined by heat flows transported by blood from all *ij*-compartments. Blood temperature equation is:

$$\begin{aligned} & V_b \rho_b c_b \frac{dT_b}{dt} = \sum_{i=1}^{N}\sum_{j=1}^{K} w_{ij}\rho_b c_b T_{ij} - W\rho_b c_b T_b - \\ & -\dot{V}\rho_e r(\rho_{ex} - \rho_{in}) - \dot{V}\rho_e c_e (T_b - T_e) \end{aligned} \quad (14)$$

where V_b is blood volume in large veins; W is cardiac output, w_{ij} is compartment blood flow.

The mathematical model of heat production, heat transfer and heat exchange of human in water is the set of ordinary differential equations including equations of *ij*-compartment (13) and blood temperature (14). The order of equation system is determined by quantity of compartments. The number of cylinders and layers are determined by the aim of study. Human body is approximated by 38 compartments in this model.

III. Conclusions

The model has been implemented in MATLAB R2011b. There were performed model experiments for prediction of thermal changes in human body in water and air environment.

Modeling experiments were performed for two different combinations of air and water temperatures.

1) Thermoneutral environment.

It means that metabolic heat production in human body is equal to heat loss to environment. For nude man air temperature is 30 ˚C, water temperature is 34 ˚C

Modeling results showed that practically there is no difference in human temperatures for both cases. According to experiments blood temperature is 36.36 ˚C in water, 36.35 ˚C in air, brain temperature is 36.32 ˚C, 36.41 ˚C respectively, thigh muscle temperature is 35.70 ˚C, 35.82 ˚C respectively, mean skin temperature is 35.40 ˚C, 35.16 ˚C respectively. The same tendency is for all body temperatures.

2) Water temperature is equal to air temperature. Water temperature is 30 ˚C.

Modeling results showed fundamental differences among human temperatures during stay in water and air environment.

During water immersion blood temperature decreased to 32.59 ˚C (36.36 ˚C in air) less by 3.77 ˚C. Brain temperature decreased to 32.61 ˚C (36.41 ˚C in air), less by 3.8 ˚C. Thigh muscle temperature decreased to 31.82 ˚C (35.82 ˚C in air), less by 4.0 ˚C. Mean skin temperature decreased to 30.6 ˚C (35.16 ˚C in air) less by 4.56 ˚C.

It was shown the essential cooling effect of water environment on human body in comparison to air environment.

Analysis of modeling results allowed to make conclusions that model of heat production, heat transfer and heat exchange in human is valid for water and air environment.

References

[1] L. D. Montgomery, “A model of heat transfer in immersed man,” Annals of Biomedical Engineering, vol. 2, pp. 19-46, 1974.

[2] P.Tikuisis and A. A. Keefe, “Prediction of sea survival time,” Defence and Civil Insitute of Environmental Medicine DCIEM No.96-R-12, 1996.

[3] X. Xu, J. W. Castellani, W. R. Santee, and M. Kolka, “Thermal responses for men with different fat compositions during immersion in cold water at two depths: prediction versus observation.,” European Journal of Applied Physiology, vol. 100, pp. 79–88, May 2007.

[4] X. Xu, C. a. Turner, and W. R. Santee, “Survival time prediction in marine environments,” Journal of Thermal Biology, vol. 36, pp. 340–345, August. 2011.

[5] P. Tikuisis, “Prediction of the thermoregulatory response for clothed immersion in cold water,” European Journal of Applied Physiology and Occupational Physiology, vol. 59, pp. 334–341, 1989.

[6] B. S. Shender and W. Todd, “Predicted water immersion survival times for anti- exposure ensembles,” in Proceedings of the Forty Third Annual SAFE Association Symposium, 2005, pp. 176–184

[7] I. Yermakova, “Mathematical modeling of thermal processes in man for development of protective clothing,” Journal of. Korean Society of Living Environmental system, vol. 8, pp.127-134, 2001.

[8] P. Tikuisis, R. R. Gonzalez and K. B. Pandolf, “Thermoregulatory model for immersion in cold water,” Journal of Applied Physiology, vol. 64, pp. 719–727, 1988.

Automatic Pattern Recognition of Epileptiform Discharges Using Morphological Descriptors and Linear Discriminant Analysis

Christine Fredel Boos, Fernando Mendes de Azevedo
Biomedical Engineering Institute
Electrical Engineering Department, Federal University of Santa Catarina
Florianópolis, Brazil
cfboos@gmail.com; azevedo@ieb.ufsc.br

Abstract— **This paper presents the performance analysis of a methodology for automated recognition of epileptiform patterns using morphological descriptors and Linear Discriminant Analysis. Morphological descriptors, in this paper, are parameters related to the morphology of the signal's waveform and Linear Discriminant Analysis (DA) is a method of multivariate statistical analysis commonly used for classification, size reduction and/or feature extraction. Thus, the main purpose of this paper is to analyze the classification performance of the discriminant functions and examine the applicability of Discriminant Analysis in reducing the number of independent variables (in this case morphological descriptors) necessary to obtain a discriminant function with acceptable classification performance. Simulations showed that the best functions exhibited efficiency greater than or equal to 85%, sensitivity of 85-90% and specificity between 80 and 84%.**

Keywords— EEG signal; epileptiform patterns; morphological descriptors; linear discriminant analysis.

I. INTRODUCTION

Epilepsy is a chronic condition or a group of diseases of high prevalence, poorly explained by science and whose diagnosis is predominantly clinical, mainly based on the medical history of the subject. However, the analysis of electroencephalogram (EEG) records can be an important tool for confirmation of clinical diagnosis, support in defining the type of syndrome, planning the drug therapy and deciding on surgical feasibility [1].

The analysis of EEG records is made by a thorough visual analysis performed by a specialist in the area with the objective of locating a very specific type of electrographic activity, called epileptiform patterns or discharges [2]. In this study, the term epileptiform pattern will be used to represent sharp-and-slow-wave and spike-and-slow-wave complexes. However, since the records for this analysis are acquired from long term monitoring of the subject this evaluation task can be a very tiresome. Over the years, several studies have proposed methodologies and/or algorithms for automating this type of analysis, thus facilitating the work of the specialists and reducing the time required to perform the revision of the EEG records. However, there is still no methodology and/or algorithm that is widely employed and has a performance in accordance with the needs of specialists in the field [3]. Nevertheless, so far one of the most used and successful systems is Gotman's [4], which is based on the morphological analysis of the signals.

The application of morphological analysis of the signals can allow the extraction of parameters related to morphology features of the EEG waveform. After the extraction of these parameters – which in this study are called morphological descriptors – a statistical analysis technique or an Artificial Intelligence tool can be used for the pattern recognition portion of a methodology for automating the process of EEG record review.

Therefore, the purpose of this paper is to analyze the performance of a methodology for automated recognition of epileptiform patterns using morphological descriptors and Linear Discriminant Analysis. Furthermore, it is also examined the applicability of Discriminant Analysis in reducing the number of independent variables (in this case morphological descriptors) necessary to obtain a discriminant function with acceptable classification performance.

II. MATERIALS AND METHODS

A. EEG signal database

The EEG signals used in this work are a part of a database[1] containing nine records acquired from seven adult subjects with proven epilepsy diagnosis. The signals have a sampling frequency of 100Hz and were acquired through 24 and 32 channels using a bipolar electrode assembly of the type zygomatic-temporal.

In order to perform the necessary simulations we selected two sets of 300 signal segments: one containing epileptiform patterns (EEP) and the other containing non-epileptiform patterns (NEP). According to literature, epileptiform patterns are electrographic signals with typical amplitude around 200μV

This study was co-sponsored by the National Council for Scientific and Technological Development (CNPq).

[1] The study's use of the EEG records was approved by the University's Ethics Committee (protocol number 144.469).

and duration between 20 and 200ms. They normally are morphologically characterized by a peak (spike or sharp wave) that can sometimes be immediately followed by low frequency activity (slow wave) [3]. In turn, the non-epileptiform patterns are all the other types of activities found in EEG signals. In this study, the selection of these patterns was restricted to those that are generally considered as responsible for the appearance of false positives in the automatic review process and are illustrated in Fig. 1:

- eye blinks;
- normal background activity;
- alpha wave activity;
- artifacts generated by various sources of noise.

B. Morphological descriptors

The set of descriptors used in this study was originally thought of by Pereira [5] in her doctoral thesis and later refined in some studies of our research group [6][7]. The development of the set of descriptors was based on extensive review of literature on feature extraction and detailed analysis of the morphological characteristics of epileptiform discharges.

From the literature on automatic detection of epileptiform patterns it was selected, the following descriptors: duration of the discharge, the maximum amplitude (Amax), the patterns' variation coefficient, the entropy and crest factor. The other descriptors were created in an effort to further highlight the epileptiform discharges from other patterns and thus increasing the power of discrimination of the set.

The developed descriptors can be separated in four categories with relation to be basic nature of the characteristic they represent:

- amplitudes and duration;
- vertex angle of the peak;
- initial and final regions;
- statistical indices.

The basic descriptors related to the amplitudes and duration are, as illustrated in Fig. 2, the minimum (Bmin) amplitude, the amplitude difference between Amax and Bmin (DifAB) and the difference between their points of occurrence (Tdif).

Considering that in our analysis the EEG signals are divided into adjacent epochs of one second and that more than one pattern (epileptiform or not) can appear in this time frame, a second subset of descriptors had to be created. These descriptors are the same as mentioned above but they are calculated within a segment of 300ms around the principal amplitude contained in the epoch as can be observed in Fig. 3.

From the visual observation that epileptiform discharges generally have more acute peaks when compared to the obtuse peaks of others patterns such as eye blinks and alpha waves, the second category of descriptors was developed considering the analysis of the vertex angle of the signal's peaks (as can be seen in Fig. 4). Based on the vertex angles (θp and θn) we derived the calculation of the tangent values (tgp and tgn), thae value of the base of the peaks (dpos and dneg) the ratio between the base of a pattern and its positive or negative maximum amplitude (trp and trn), and the value of the base of the principal peak of the waveform (dbase).

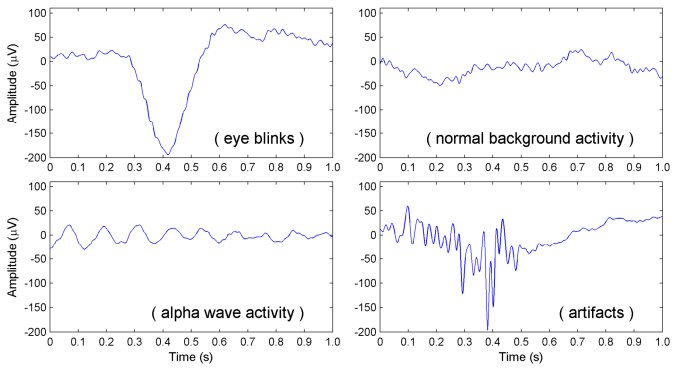


Fig. 1. Illustration of the four main types of non-epileptiform patters used in the simulations.

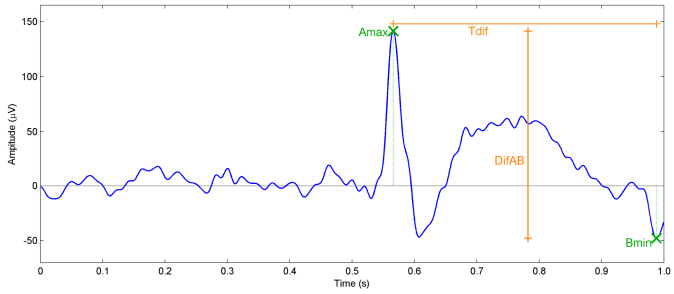


Fig. 2. Illustration of the basic descriptors based on amplitudes and duration of the signals waveform.

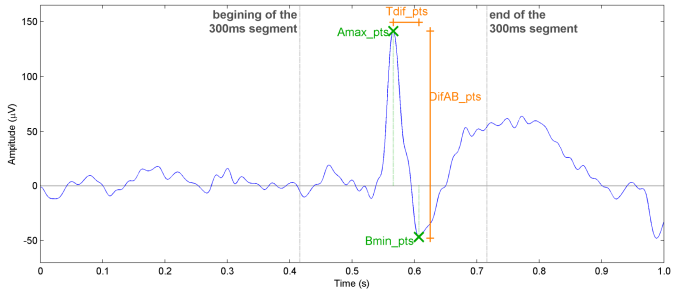


Fig. 3. Illustration of the descriptors calculated from the 300ms segment within the one second epoch in which the descriptor analysis is based.

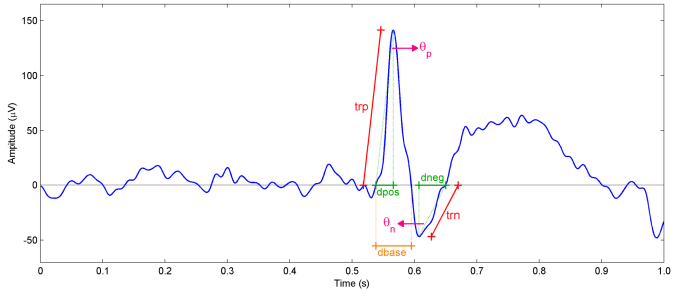


Fig. 4. Illustration of the descriptors based on the analysis of the vertex angle of the signal's peaks.

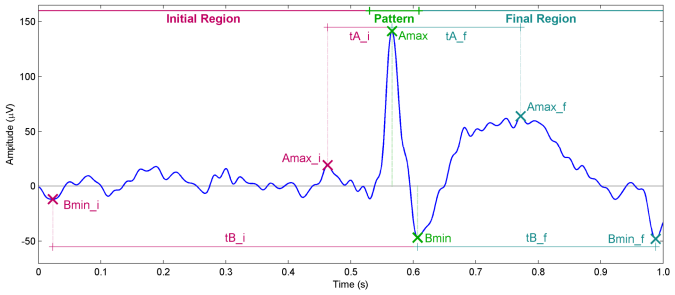


Fig. 5. Illustration of the descriptors developed to aid the discrimination between epileptiform discharges and EEG signal artifacts.

The third category of descriptors was designed to aid the differentiation between artifacts and epileptiform discharges. The artifacts have typical waveforms that consist of sequential high amplitude points separated by short periods of time whereas the discharges usually are more spaced because one of their primary characteristic is that they are distinguishable and prominent from the background activity [3]. The descriptors proposed to make this distinction were based on dividing the epoch in two regions (initial and final) and calculating time and amplitude differences between the amplitude peaks within the epoch.

Initially the determination of the two regions was made considering the principal amplitude as the common point for the segments but this was later adjusted so that the division is made by the duration of the pattern. As can be seen in Fig. 5, the initial and final regions correspond directly to the segments before and after the duration the pattern present in the epoch under analysis and the time and amplitude differences are all in relation to Amax and Bmin.

Finally, the last category of descriptors was composed by classical statistical parameters such as average, standard deviation and variance. These parameters were all calculated for the epoch under analysis and the previously described 300ms segment.

The complete set of descriptors has a total of 45 numerical independent variables that are calculated automatically by an algorithm developed specially for this task. This algorithm begins by interpolating the signal to 1 kHz (so that the sampling frequency is increased thus improving the precision of the subsequent calculations). After this the algorithm performs digital signal filtering to remove baseline deviation (high-pass Butterworth filter with cutoff frequency of 0.5Hz) and power line interference (notch filter with 60Hz cut-off frequency). Thereafter, the 45 descriptors are calculated for all 600 signal segments and stored for subsequent analysis.

C. Discriminant Analysis

Linear Discriminant Analysis (DA) is a method of multivariate statistical analysis commonly used for classification [8], size reduction and/or feature extraction [9]. DA uses a set of independent variables to create a function that can classify the observations of a given problem between two or more classes determined a priori [10].

As can be observed in (1), the discriminant function (DF) is a linear weighted (w_1, w_2, …, w_n) combination of n independent variables (v_1, v_2, …, v_n) that can be used to predict the classification of new observations of the problem under analysis.

$$DF = w_1 \cdot v_1 + w_2 \cdot v_2 + \ldots + w_n \cdot v_n \quad (1)$$

In our study, we used Discriminant Analysis to create a DF derived from the morphological descriptors to differentiate between epileptiform and non-epileptiform patterns. For our DA simulations we applied logarithmic normalization on the calculated descriptors and we used the IBM® SPSS Statistics software.

In order to test discriminant functions with different sizes we opted for the stepwise method of analysis and used Mahalanobis distance. The stepwise approach consists in adding and/or removing the independent variables from the DF, one at a time, according to significance levels (f) specified in the beginning of the analysis. These f levels are two default values that work as a reference point for adding and removing an independent variable. During the DA simulations we changed the f values with the intention of obtaining functions with different lengths and constituted by different groups of descriptors.

III. Results

The 600 segments of EEG signals extracted from the database and used for the simulations were divided in two groups: model and test. The “model group” was composed by 400 patterns (equally divided between EEP and NEP) and was used to create the discriminant functions. The “test group” was composed be 200 patterns (also equally divided between EEP and NEP) and was used to test and analyze the DF created.

Several f values were used and after all the simulations we selected eight functions that achieved the best results, which can be observed in following tables:

- TABLE I. – presents the number of variables in each DF (Size), the value of the centroid of the two classes under analysis (EEP and NEP), the amount of variable reduction achieved, the significance of the DF (Sig.) and the variability explained by the DF;
- TABLE II. – presents the indices used for the analysis of the classification performance.

TABLE I. Results of the best discriminant functions obtained after simulations with 45 morphological descriptors

DF	Size	Centroids		Variable reduction	Explained variability	Sig.
		EEP	*NEP*			
1	3	1.0899	-1.0899	93.33%	65.26%	1.3×10^{-21}
2	7	1.2963	-1.2963	84.44%	63.08%	1.3×10^{-21}
3	9	1.3591	-1.3591	80.00%	65.26%	1.3×10^{-21}
4	12	1.4087	-1.4087	73.33%	66.87%	1.1×10^{-20}
5	22	1.5572	-1.5572	51.11%	54.71%	7.1×10^{-20}
6	24	1.5707	-1.5707	46.67%	71.50%	4.3×10^{-17}
7	31	1.6076	-1.6076	31.11%	72.44%	1.9×10^{-14}
8	36	1.6105	-1.6105	20.00%	54.71%	7.1×10^{-20}

TABLE II. RESULTS OF THE CLASSIFICATION PERFORMANCE OF THE BEST DISCRIMINANT FUCNTIONS

DF	Size	TP	TN	FP	FN	SE	SP	EF
1	3	83	81	19	17	0.83	0.81	0.82
2	7	85	84	16	15	0.85	0.84	0.85
3	9	90	80	20	10	0.90	0.80	0.85
4	12	90	81	19	10	0.90	0.81	0.86
5	22	89	77	23	11	0.89	0.77	0.83
6	24	90	74	26	10	0.90	0.74	0.82
7	31	85	75	25	15	0.85	0.75	0.80
8	36	86	77	23	14	0.86	0.77	0.82

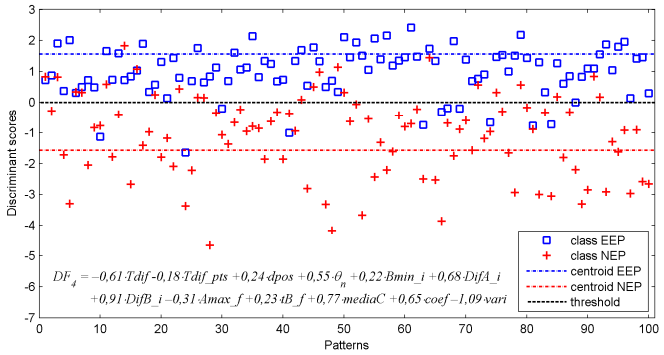


Fig. 6. Illustration of the classification of the patterns made by the fourth discriminant function on Table I and Table II.

The performance analysis was based on the results obtained from patterns of the "test group" and was made using indices easily found in literature [11] such as the number of True Positives (TP) and Negatives (TN), the number of False Positives (FP) and Negatives (FN), value of Sensibility (SE), Specificity (SP) and Efficiency (EF).

The information about the centroids of the two classes was used to determine a threshold (Fig. 6) for the classification based on the discriminant scores (i.e. the output of the discriminant function).

IV. CONCLUSIONS AND DISCUSSIONS

The results obtained with the Linear Discriminant Analysis can be considered relatively good. The best functions exhibited efficiency greater than or equal to 85%, sensitivity of 85-90% and specificity between 80 and 84%. The use of the Discriminant Analysis of allowed a good reduction of the number of variables necessary to perform the classification between epileptiform patterns and non-epileptiform without any significant loss of performance when compared to similar previous studies that used other classification and reduction methodologies.

From the eight functions selected only three achieved less than 50% variable reduction (which in this study is the minimum desired value). From the point of view of variable reduction the best function results where within the range of 80-93%. From a statistical point of view all of the functions can be considered highly significant ($p<<0.05$) and the average variability explained by the functions is over 64% with be best DF explaining approximately 72.5% of the group variation.

The performance analysis showed that the average indices for all functions were relatively good and when considering only the five functions with good variable reduction (size ≤ 22) the average for all performance indices was above 81%. The best classification performance achieved a 90% sensibility, 81% specificity and 86% overall efficiency. The number of 19 false positives (i.e., 10%) is a bit high and can still be improved. However, when compared to other similar studies the performance results can be considered compatible.

Thus, although some aspects of this study can be discussed in much greater depth, it is observed that the proposed objectives were achieved. Promising results have been achieved in the use of DA as a pattern classifier using morphological descriptors and as a tool to reduce the amount of descriptors necessary to perform the automated pattern recognition of epileptiform discharges in EEG signals.

REFERENCES

[1] S. Sanei and J. A. Chambers, EEG Signal Processing, 1st ed. West Sussex: John Wiley & Sons Ltd, 2007, p. 312.

[2] J. Pillai and M. R. Sperling, "Interictal EEG and the diagnosis of epilepsy.," Epilepsia, vol. 47 Suppl 1, pp. 14–22, January 2006.

[3] D. L. Schomer and F. L. da Silva, Eds., Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, 6th ed. Philadelphia: Lippincott Williams & Wilkins, 2011, p. 1296.

[4] L. Ver Hoef, R. Elgavish, and R. C. Knowlton, "Effect of detection parameters on automated electroencephalography spike detection sensitivity and false-positive rate.," Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society, vol. 27, no. 1, pp. 12–6, February 2010.

[5] M. do C. V. Pereira, "Avaliação de Técnicas de Pré-Processamento de Sinais de EEG para Detecção de Eventos Epileptogênicos Utilizando Redes Neurais Artificiais," Thesis (Phd), Federal University of Santa Catarina, 2003.

[6] C. F. Boos, M. do Carmo Vitarelli Pereira, F. I. Marques Argoud, and F. Mendes de Azevedo, "Morphological descriptors for automatic detection of epileptiform events.," in Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2010, vol. 2010, pp. 2435–8.

[7] C. F. Boos, F. M. de Azevedo, G. R. Scolaro, and M. do C. V. Pereira, "Automatic Detection of Paroxysms in EEG Signals Using Morphological Descriptors and Artificial Neural Networks," in Biomedical Engineering, Trends in Electronics, Communications and Software, A. Laskovski, Ed. InTech, 2011, pp. 387–402.

[8] J. Brankack, V. I. Kukushka, A. L. Vyssotski, and A. Draguhn, "EEG gamma frequency and sleep-wake scoring in mice: comparing two types of supervised classifiers.," Brain research, vol. 1322, pp. 59–71, March 2010.

[9] M. M. López, J. Ramírez, J. M. Górriz, I. Alvarez, D. Salas-Gonzalez, F. Segovia, and R. Chaves, "SVM-based CAD system for early detection of the Alzheimer's disease using kernel PCA and LDA.," Neuroscience letters, vol. 464, no. 3, pp. 233–8, October 2009.

[10] A. Subasi and M. Ismail Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," Expert Systems with Applications, vol. 37, no. 12, pp. 8659–8666, December 2010.

[11] A. J. Casson, E. Luna, and E. Rodriguez-Villegas, "Performance metrics for the accurate characterisation of interictal spike detection algorithms.," Journal of neuroscience methods, vol. 177, no. 2, pp. 479–87, March 2009.

Adaptive Algorithm for Continuous Monitoring of Blood Pressure Using a Pulse Transit Time

Ali Jadooei, Zaderykhin O.
Department of Aircraft Radio-Electronic Systems Design
National Aerospace University
Kharkov, Ukraine
e-mail : xai.ali.501@gmail.com

Shulgin V.I.
Scientific and Technical Center of Radio-Electronic Medical Equipment and Technologies XAI-MEDICA
National Aerospace University
Kharkov, Ukraine
e-mail : vyacheslav.shulgin@gmail.com

***Abstract* — this paper attempts to present the problem of continuous Blood Pressure (BP) estimation using the Pulse Transit Time (PTT) cuff-less method. PTT is measured by synchronous electrocardiogram and photoplethysmograph (PPG) registration and the difference time is the basis parameter for blood pressure estimation. One of the problems to use PTT method is the necessity to calibrate with the traditional methods of Blood Pressure (TMBP) measurements. Adaptive Kalman filter (AKF) algorithm and maximum-likelihood estimation (MLE) methods were used in this paper to reduce the number of calibration measurements. The results clearly showed that the AKF used to estimate continuous BP, reduces the frequent use of the TMBP apparatus for calibration of BP measurement with PTT.**

Keywords— pulse transit time; blood pressure; maximum-likelihood estimation; adaptive kalman filter algorithm

I. INTRODUCTION

Blood pressure (BP) is an important physiological indicator, determining both the patient's cardiovascular system (CVS) and the whole-body state. The existing methods of BP measurement are divided into invasive and non-invasive.

The invasive method of BP measuring is carried out by inserting a catheter into the blood vessel to measure the blood pressures accurately has still the disadvantages of having pain and contamination, and the side effects arising from inserting the catheter-tip into a patient's blood vessel [1].

The most commonly used non-invasive methods of BP measurement are the oscillometric and auscultatory methods. However, they have disadvantages such as lower accuracy and stability of measurements on the basis of occlusion of blood vessels with the aid of the cuff. So they cause significant inconvenience for patient during its pumping. In addition, they don't allow to perform continuous measurements [1].

As a result, the researchers, aimed at developing of new BP assessment methods, based on use of connection of BP with other measurable physiological signals and their parameters have considerably intensified [2,3,4].

One of the new methods for BP measuring is using pulse transit time (PTT). PTT is the time taken for the arterial pulse pressure wave to travel from the aortic valve to a peripheral site.

II. BACKGROUND

Blood pressure is just one of the many indicators for the patient's CVS performance, which represents a closed-loop system with a multiple-loop feedback. Therefore, there is a direct connection between all its parameters. Using this connection allows an indirect assessment of BP with the same accuracy as in TMBP device [1,2,3].

Functional dependency between BP and pulse wave velocity were established long ago, and were based on the well-known relation obtained by A. I. Moens and D. J. Korteweg [2,3,4,5]. According to it, the rate at which the BP pulse wave distributed through the arteries was defined by the following relation:

$$\mathrm{PWV} = \frac{(\Delta \mathrm{d})}{\mathrm{PTT}} = \sqrt{\frac{\mathrm{g}}{\rho} \cdot \frac{Ea}{d}} \qquad (1)$$

where PWV – pulse wave velocity, g – length of an artery section, E – elasticity of the vascular walls, a – thickness of vascular walls, d - diameter of a lumen, ρ - density of blood.

Later, this expression was complemented by L. A. Geddes [2,3,4,5], who showed that the elasticity of the vascular wall is not constant and depends on the current BP:

$$E = E_0 . \mathrm{e}^{\alpha \mathrm{P}} \qquad (2)$$

where E_0 – E value at zero pressure, α – some constant that depends on the state of vessels, P - value of BP:

Assuming that the pulse wave velocity in the artery section in length of L is constant, and substituting an expression (2) into (1), one can obtain the relation of PTT to the BP value:

$$(PTT)^2 = \frac{\mathrm{L}^2 . \rho \mathrm{d}}{g . E_0 . a} \times (\mathrm{e}^{\alpha \mathrm{P}})^{-1} \qquad (3)$$

And its inverse relation, connecting BP with PTT:

$$P = -\frac{2}{\alpha} \ln(PTT) + \frac{1}{\alpha} \ln \frac{\mathrm{L}^2 . \rho \mathrm{d}}{g . E_0 . a} \qquad (4)$$

Finally, considering the relatively small changes of PTT in the range of BP changes, the expression (4) can be expanded in a Taylor series and restrict oneself to the linear approximation. As a result, we shall obtain the relation:

$$P = -\frac{2}{\alpha}\left(\Delta PTT\right) + \frac{1}{\alpha}\ln\frac{L^2.\rho d}{g.E_0.a} = \\ = -C_1\left(\Delta PTT\right) + C_0 \qquad (5)$$

This indicates a linear connection between changes in BP and PTT. In this expression, C_1 and C_0 - are some constants ratios for a certain patient, determined by the properties and the state of patient's vessels. As it is has been seen from (5), PTT decreases while BP increases [6].

III. MEASUREMENTS

In the present study, 10 people were involved for experimental test. All participants were young, aged 20 to 33, and were in healthy state. No subject has been diagnosed with cardiovascular diseases. ECG and PPG were registered by the portable digital electrocardiograph (ECG) and photoplethysmograph (PPG) with Bluetooth communication to personal computer. Each signal was sampled at 500 Hz by an AD converter. SBP and DBP were measured by TMBP device (Microlife BP A100) from the left arm of the subject. All the measurements of BP were made periodically, at intervals of one minute along with continuous recording of ECG and PPG. Recorded data processing algorithm is shown in Fig. 1.

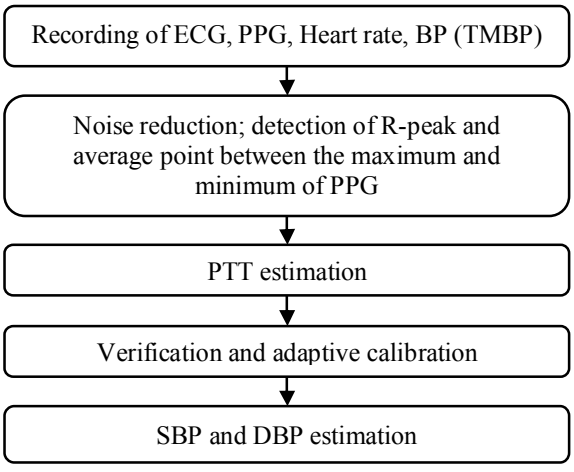


Fig. 1. Flowchart for estimation of the systolic and diastolic blood pressure.

In this study, PTT was measured on the basis of simultaneously recorded ECG and PPG signals. Time interval from maximum the R-peak of ECG and the average point between maximum and minimum of PPG signal has taken as PTT [2,3,4]. The principle of PTT measurement could be represented as shown in Fig 2.

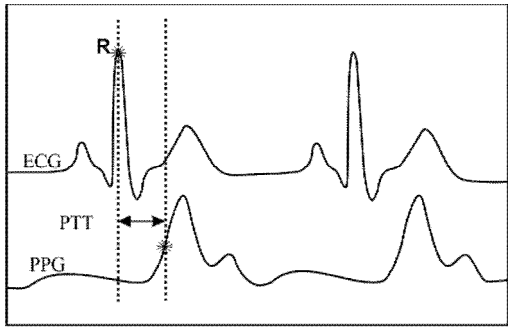


Fig. 2. PTT measured between R peak of ECG and the average point between the maximum and the minimum of PPG.

IV. METHODOLOGY

According to conducted researches [2,6] heart rate (HR) is one of the parameters which is correlated directly with BP. With respect to this, in this research, the equation (5) was used to estimate systolic blood pressure (SBP) and diastolic blood pressure (DBP).

$$SBP = C_{01} + PTT \cdot C_{11} + HR \cdot C_{21} \qquad (6)$$

$$DBP = C_{02} + PTT \cdot C_{12} + HR \cdot C_{22} \qquad (7)$$

where $C_{01} \cdots C_{22}$ are some coefficients for per patient.

One of the problems to estimate SBP and DBP is that the estimated $C_{01} \cdots C_{22}$ parameters must be re-calibrated after some time. Calibration measurements of SBP and DBP were carried out by using a TMBP device.

In order to estimate the coefficients $C_{01} \cdots C_{22}$ two algorithms were considered.

The first approach provides solution for the problem, using the method of maximum likelihood (MLE) [7], and the second approach provides solution for the problem using the adaptive Kalman filter (AKF) algorithm. In both cases linear models of initial (PTT, HR) and evaluated (SBP and DBP) parameters' connection were used.

A. Problem solving using the maximum likelihood estimation method.

We will consider the signals to be discrete functions of counting sample number $k = 1, 2, 3, \ldots$. Then the initial measured parameters of PTT, HR and etc. can be put down in a vector configuration:

$$\vec{M}(t_k) = \vec{M}_k = \begin{bmatrix} PTT_k \\ HR_k \\ \vdots \end{bmatrix}$$

t_k – these are the moments of discrete time, when initial measures are carried out.

To simplify the problem solving, we will consider estimation of one of SBP or DBP parameters. We will mark the evaluated parameter as Y_k. Then replacing Y_k with SBP or DBP, we will get the algorithm for estimation of respectively SBP or DBP. Estimation of the required parameter will be marked with $\tilde{Y}_k$, and calibrating measurement of the required parameter – Y_r. As calibration measurements are carried out more rarely, we will use another time index for them – *r,* using the linear connection model of the measured and initial parameters, the observation equation will be the following:

$$Y_k = c_0 + c_1 \cdot PTT_k + c_2 \cdot HR_k + \cdots + n_k = \\ = \vec{C}^T \cdot \vec{X}_k + n_k \quad (8)$$

where $\vec{X}_K = \begin{bmatrix} 1 \\ PTT_k \\ HR_k \end{bmatrix}$, $\vec{C} = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix}$, n_k – noise, normally distributed with zero mean and dispersion of σ_n^2.

The required estimation of the parameter can be calculated in the following way:

$$\tilde{Y}_k = \vec{\tilde{C}}^T \cdot \vec{X}_k \quad (9)$$

where $\vec{\tilde{C}}$ is estimation of $\vec{C}$ coefficients, which appears in the result of determination of likelihood function extremum:

$$p\left(Y_r / \vec{C}\right) = \alpha \cdot \left(\sigma_n^2\right)^{-\frac{R+1}{2}} \cdot \exp.... \\ ....\left\{ -\frac{1}{2\sigma_n^2} \cdot \sum_{r=0}^{R} \left(Y_r - \vec{C}^T \cdot \vec{X}_r\right)^2 \right\} \quad (10)$$

where R is the number of calibration measurements of parameter Y, α is some constant . After finding the logarithm, differentiation at $\vec{C}$ and zero equalization of (10) we will get the following optimal in terms of MLE algorithm determination of coefficients' $\vec{C}$ evaluations:

$$\vec{\tilde{C}} = A^{-1} \cdot \vec{a};\ A = \sum_{r=0}^{R} \vec{X}_r \cdot \vec{X}_r^{\,T};\ \vec{a} = \sum_{r=0}^{R} \vec{X}_r \cdot Y_r \quad (11)$$

B. Problem solving using the adaptive Kalman filter algorithm.

Using the same assumptions which were made for solving the problem, using MLE method, we will get an algorithm of coefficients estimation using AKF algorithm:

$$\vec{C}^*_r = \vec{C}^*_{r-1} + \frac{\sum_r . \vec{X}_r}{\sigma_n^2} \left[Y_r - \vec{C}^*_{r-1} . \vec{X}_к\right] \quad (12)$$

where $\vec{C}^*_r$ is Kalman estimation of $\vec{C}$ coefficients made in the moment of time t_r (these are the moments of discrete time, in which calibration measurements of BP are carried out); $\sum_r$ – evaluation precision (estimation error variance):

$$\sum_r = \left(\left(\sum_{r-1} + \sum_c\right)^{-1} + \frac{\vec{X}_r . \vec{X}_r^T}{\sigma_n^2} \right)^{-1} \quad (13)$$

$$\sum_{r=0} = \frac{\vec{X}_0 . \vec{X}_0^T}{\sigma_n^2} \quad (14)$$

where $\sum_c$ – covariance matrix of coefficients perturbation $\vec{C}_K$, modeled as Markov process; $\vec{C}^*_0$ is initial approximation, it is evaluated by using the MLE.

Like in the case when we use the MLE, we will get the required continuous estimation of the parameter Y_k in the following way:

$$\tilde{Y}_k = \vec{C}^{*\,T}_r \cdot \vec{X}_k \quad (15)$$

where $\vec{C}^*_r$ is coefficients, calculated at the moment of time t_r, that corresponds to the time of the last calibration measurement of the parameter Y_k.

Testing the method performed on real data. At the beginning of each test, coefficients were estimated using first 4 measurements. The MLE algorithm was used to estimate the initial coefficients. These initial coefficients characterize the state of the patient at the beginning of the test. Over time, the patient's condition change. This leads to the fact that the estimated coefficients over time become not relevant. Using non relevant coefficients lead to errors in the estimation of BP. Therefore it is necessary to adjust periodically coefficients. Adjustment procedure is called calibration. To perform the calibration of estimated BP, it is necessary to measure with TMBP device.

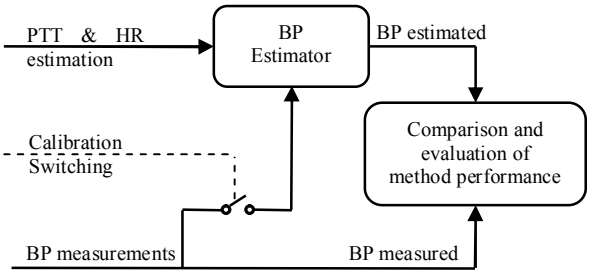


Fig. 3. Flowchart of experimental methodology.

In addition, the measured BP is needed to investigate the quality the two used methods. Therefore simultaneous measurement of BP and PTT were carried out, but to calibrate BP only some measurements of BP were used.

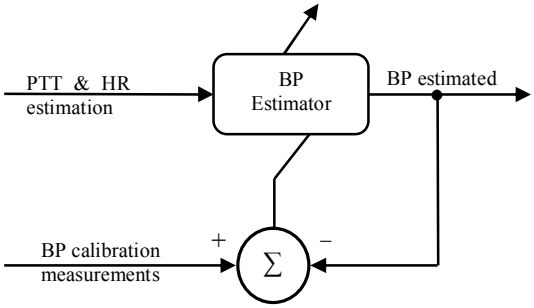


Fig. 4. Flowchart of adaptive Kalman estimator.

To determine the quality of the method, comparison of the measured and estimated BP parameters were used. A quantitative comparison was performed using RMS and correlation coefficients.

V. RESULTS

The results of investigation of two BP measurement algorithms on the basis of MLE and AKF are discussed below. The measurements were performed in two periods, first time it was between 10:00 to 14:00 o'clock (1st to 29th points of BP measurement). And the second period was between 16:00 to 22:00 o'clock (from 30th to 75th points of BP measurement). Each point of estimated BP was obtained from the mean of 350 samples of PTT and HR measurements per 5 minutes and the mean of 5 samples of actual BP measured by the TMBP device every minute.

The comparison between MLE-based estimated SBP, DBP and actual BP data is represented in Fig. 5.

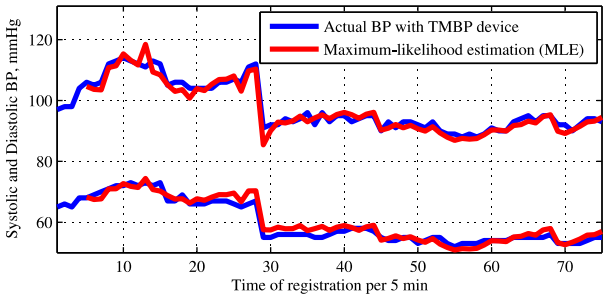


Fig. 5. Maximum likelihood estimation method for estimation SBP and DBP using PTT.

The comparison between estimated SBP and DBP by AKF algorithm and actual BP data were shown in Fig. 6. It can be see that the actual BP measured by TMBP device and forecast BP by both algorithms has negligible differences, and the estimated waveform follows closely the trends of the measured BP.

We emphasize here, that using MLE method the coefficients has calibrated on every point of measurement. Using AKF the coefficients has calibrated only 3 times for the entire measurement period. The time of calibrations has shown by stars 1 to 3.

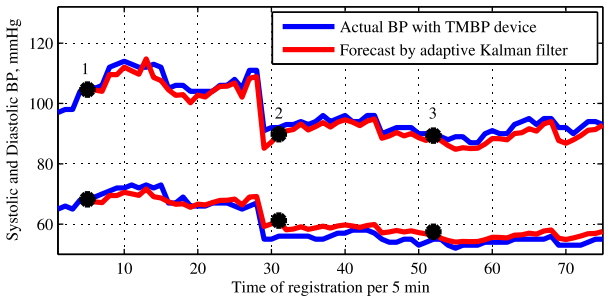


Fig. 6. Adaptive kalman filter algorithm for estimation SBP and DBP using PTT.

The root-mean-square error (RMSE) and correlation coefficients between the actual BP measured by the TMBP device and the BP estimated with MLE and AKF algorithms are summarized in TABLE 1.

According to Table I, correlation coefficient between the actual measurements and their estimated values are very high since the statistical signification is lower than 0.05 for SBP and DBP. Also p-value for both cases is < 0.0001.

TABLE 1.

Subject	***MLE***		***Kalman adaptive filter algorithm***	
	SBP	***DBP***	***SBP***	***DBP***
RMSE	1,80	1,57	1,69	1,04
correlation coefficient	0,978	0,974	0,976	0,989

The results of SBP and DBP estimation by AKF algorithm and LME method in this work were satisfactory in accordance with the regulation of American National Standards of the Association for the Advancement of Medical Instrumentation (ANSI/AAMI) that is the test device must not differ from the mercury standard by a mean difference more than ±5 mmHg or a standard deviation more than 8 mm Hg [3]. Standard deviation of the SBP and DBP estimation errors, averaged over all 75 records, for 10 hour, by LME method and AKF algorithms which has shown in fig.7and fig.8.

The RMSE, when MLE method was used, for SBP is about 1.80 mmHg and for DBP is about 1.57 mmHg.

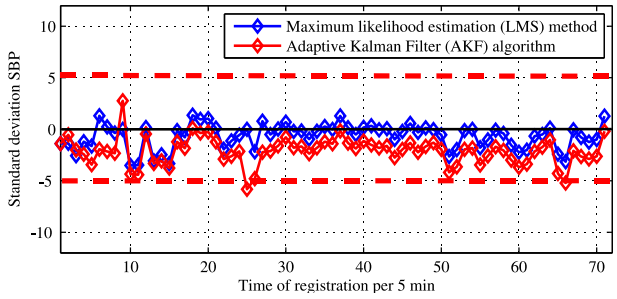


Fig. 7. Standard devition for MLE method and AKF algorithm of SBP estimation using PTT.

When AKF algorithms is used, RMSE for SBP is about 1.69 mmHg and DBP is about 1.04 mmHg as shown in Fig. 8.

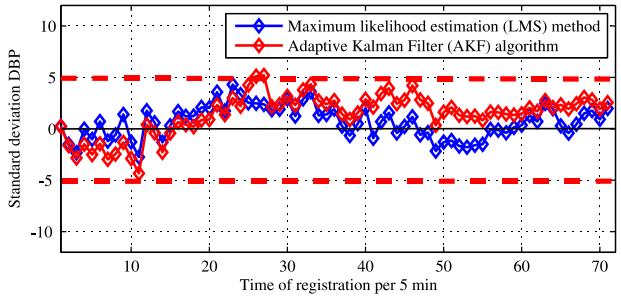


Fig. 8. Standard devition for MLS method and AKF algorithm of DBP estimation using PTT.

VI. CONCLUSIONS

This paper proposes an algorithm to estimate SBP and DBP using PTT and HR, that allows reduce continuous use of TMBP device for calibration. Both algorithms showed approximately the same accuracy of SBP and DBP measurement, but using the AKF, we can significantly reduce the number of calibration measurements.

REFERENCES

[1] Hannu Sorvoja, "Noninvasive blood pressure pulse detection and blood pressure determination" Kirsti Nurkkala, Ed. Oulu, Finland: Oulu University, 2006, pp. 17-33.

[2] Federico S., Harinath Garudadri, "Noninvasive Cuffless Estimation of Blood Pressure from Pulse Arrival Time and Heart Rate with Adaptive Calibration" in Proc. IEEE BSN Conf., Cambridge, 2009, pp.114-115.

[3] Soo-young Ye, Gi-Ryon Kim, Dong-Keun Jung, Seong-wan Baik, and Gye-rok Jeon. "Estimation of Systolic and Diastolic Pressure using the Pulse Transit Time". J. World Academy of Science, Engineering and Technology, vol. 43, 2010, pp.726-731.

[4] R. A. Payne, C. N. Symeonides, D. J. Webb, and S. R. J. Maxwell, "Pulse transit time measured from the ecg: an unreliable marker of beat-to-beat blood pressure," J. Appl. Physiol., vol. 100, pp. 163–141.

[5] Heiko Gesche, Detlef Grosskurth, Gert Kuchler, "Continuous blood pressure measurement by using the pulse transit time: comparison to a cuff-based method," J. J. Appl. Physiol., 2011, pp.726-731.

[6] V. I. Shulgin, Ali Jadooei, D.I. Shulga, K.V. Nasedkin, V.V. Fedotenk "Measurement of arterial blood pressure, on the basis of co-processing of a set of physiological parameters, ". J. Clinical Informatics and Telemedicine, vol. 9, 2012, pp.45-50.

[7] Zaderykhin O., "Nonlinear electrocardiographic leads system transformations" in Proc. Modern Problems of Radio Engineering, Telecommunications, and Computer Science Conf., Lviv-Slavske, Ukraine, February 2010, pp.136.

Stimulating Light System to Create a Controlled Effect Relaxation

Barylo G.I., Gotra Z.Yu., Zazulyak A.M., Kozhuhar O.T., Kus N.I.
Lviv Polytechnic National University,
UKRAINE, Lviv, S. Bandery street 12, 79013,
akozhuha@rambler.ru

***Abstract* - A fully automated OES of light-stimulation radiation was developed with the view of creating a controlled relaxation effect with the possibility of its operational control in accordance with the space-time characteristics of the irradiated object.**

Keywords – phototechnologies, optoelectronic systems, biomedical objects, microcontroller PSoC, interactive mode, relaxation effect.

I. INTRODUCTION

On the basis of scientific and practical of researches is established, that rhythmic light stimulation (light stimuli) can influence brain activity and dependence of activity of the cerebral cortex on the stimulation frequency is associated with frequencies of biorhythms [1-3]. Thus, β-waves - over 13 Hz - support and activate the wakeful state whereas α-waves - 8 ... 13 Hz debilitate the wakeful state and τ and δ-waves - less than 8 Hz cause slight or profound trance, or relaxation. Light-stimulation modes including those with light pulse repetition frequencies in accordance with frequencies of space-time processes in biomedical objects (BO) can cause the bio-resonance effect in these objects. One can reach the effect of profound relaxation using optical programmable triggering of brain waves in the direction of low-frequency τ and δ-waves.

Obtained reactions can be explained by changes in activity of the somatosensory areas of the cerebral hemispheres' cortex under the influence of the photo-stimulation program. The moment of stabilization of these changes indicates that the cerebral cortex centers responsible for relaxation are stimulated sufficiently. These changes are reflected by influencing stem vegetative and somatic nerve centers and reducing the tone of skeletal muscles and peripheral vessels. The result is the registered optical and thermal effects and this fact agrees with physiology of processes in a human body during stimulation of photoreceptors. It's proposed to use appropriate contactless sensors for getting information about achievement of the required state of the patient's organism.

II. MAIN PART

The main problem of wide using light-stimulation and development of controlled relaxation effect based on light-stimulation consists in the absence of principles of developing light-stimulation tools with systems of controlling light-stimulation processes and providing their interactive conducting.

One promising solution to this problem is to formulate the basic medical and technical requirements and principles of development and application in medical practice photos medicines, the basis of which would be assigned new advances in research in optoelectronics and the design of specialized OES for specific FT. Such OES based on thermal, plasma-discharge and semiconductor sources of directional incoherent radiation can create on BO stimulation exposure with programmably controlled dynamics (according to the space-frequency characteristics of the irradiated BO) and continuous non-destructive control of a session for the analysis of change from his optical parameters BO.

Especial attention is to be paid to light stimulation through visual analyzers. They perceive a given light stimulation program communication channels excited centers of the cerebral cortex, which is introduced and because the state relaxation. Particularly, providing guidelines of photo-pulse to the very retina causes induced electrical potentials. It is clear from these reactions that retinal neurons convert photo-stimuli benched by force to electrical signals benched by amplitude. This is a relatively low-frequency modulation of the resting potential leading to relaxation[4].

This effect is accompanied by increased thermogenesis caused by the increase of the blood flow in small vessels and capillaries. As a result, of this optical properties of tissues, namely the coefficients of transmission and reflection of the light beam in the interaction with certain connective tissues change, that confirmed the results of medical research conducted (Fig.1,2)[1].

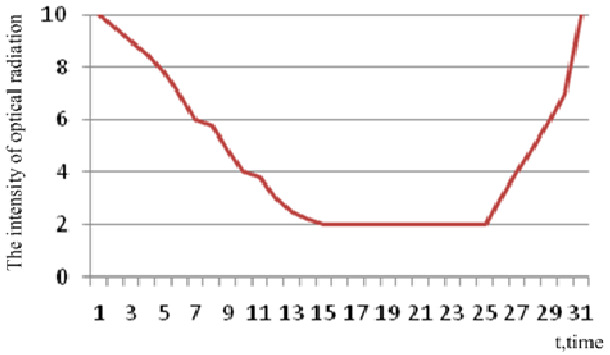


Fig. 1. Changing the intensity of optical radiation device during the time of the treatment session.

Taking into account studies of the impact of stimulating light on the human body is proposed opto-electronic system (IPS) based on incoherent optical emitters, which holds stimulating light through visual analyzers. The system consists of a generator, a generator's frequency control unit, a power supply, an output stage amplification, a LED matrix, a LED matrix control unit, a light guide. A change of modes is provided by the generator, and controlling the frequency and output power that can be set to choose medical program exposure modes (Fig.3).

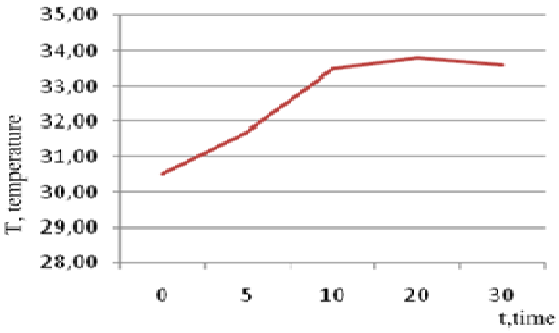


Fig. 2. The temperature variation of tissue (skin) of the patient during the treatment session.

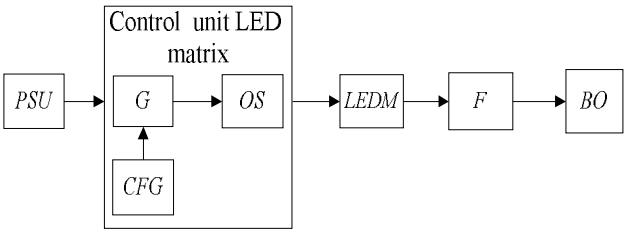


Fig. 3. Block diagram of the opto-electronic system for radiation.

In order to provide information density and interactive conduction of photo-stimulation we propose OES for noninvasive and harmless for BO recording changes in the psychophysical state based on three converters using the method of simultaneous comparison of changes in optical properties and heat flows caused by changes in peripheral BO's temperature (Fig. 4).

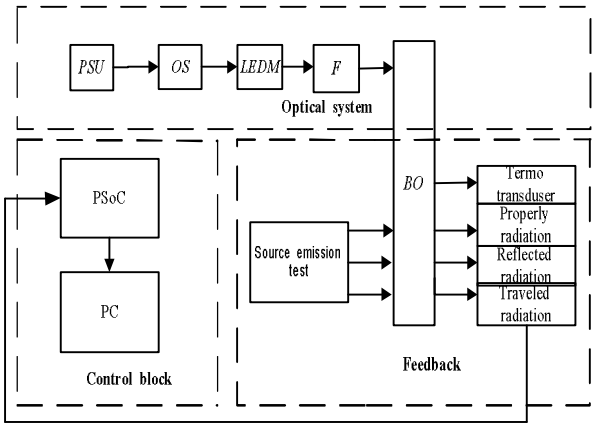


Fig. 4. Block diagram of stimulating light.

The system consists of a block of optical radiation, the control block and control unit. Block optical radiation consists of the following components: a power supply (PSU), the output stage (OS), the radiation source (LEDM) with light guide (F) and creates required for irradiation luminous flux BO in the process of treatment session. Under the influence of this radiation changes the optical and thermal parameters of BO, which is fixed by the circle feedback control unit implemented. Controlled optical signals about the state of BO registered respective transducers and processed microcontroller PSoC. [5,6] Based on the results of the analysis carried control output stage radiation source. Additionally, obtained results are transmitted via a serial interface to a PC for software processing and displaying on screen in interactive mode work.

Upon the experimental investigations it was detected that the BO's measured signals change during 30–40 min of a treatment session. The obtained dependencies underlie the algorithm of the BO's state monitoring system's functioning. The limits of controlled intensity parameters of BO's own radiation and radiation reflected and passed through a BO were determined. On the basis of these three parameters one can evaluate changes in the total organism's physiological state.

In the process of monitoring session selected dominant parameter of the most expressed changes of dynamics. The remaining parameters are used as additional. Select one of the three options occurs software and by means determining the value of maximum change (Fig.5).

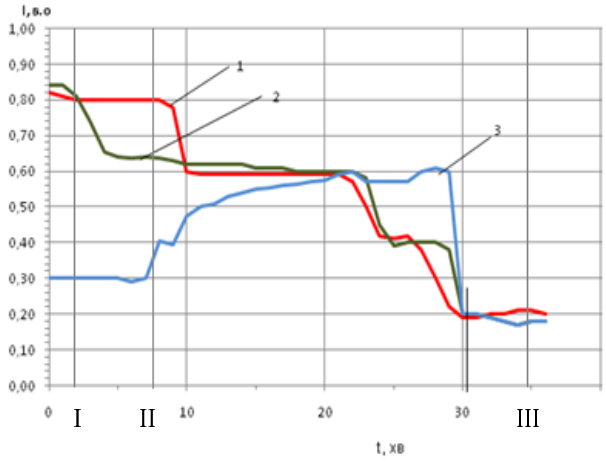


Fig. 5. Experimental results dependences of the intensity controlled light streams: 1 - passed through BO luminous flux test 2 - test the reflected light beam, 3 - own stream of optical radiation.

As seen from the experimental results at the I phase is the beginning of treatment session and the dynamics of change between absorbed and reflected radiation tissue does not. After stimulating light for 5-7 min. (phase II) change the dynamics maximum values and $\Delta I = 0.15$ r.u., therefore, the patient is in a state of relaxation. Next 15 minutes. occurs time stabilization change, signaling to achieve the desired excitement centers of the brain, that there is an effect relaxation. Last 2 minutes treatment session dynamics observed $\Delta I = 0.1$ r.u. This shows

the state of vigil centers of the brain and the patient is in a state of relaxation. As seen from the graph the intensity of absorbed and reflected radiation at the third stage, after 30 min session almost identical and very low and ≈ 0,2 r.u. and is a condition for termination of treatment session.

In order to implement interactive operation mode the following algorithm for analysis of the parameters of controlled optical was proposed (Fig. 6).

Process management is carried out using a PC through an informative user interface. Before starting work in automatic or manual mode one has to set the parameters of the LED matrix: duration of light signals, amplitude, and limits of the spectrum of test signals. After starting the process each of the input parameters I1 - I4 is controlled continuously. At the same time each parameter is checked for compliance with the values ranges defined in a database. In the case when the controlled parameter In exceeds the allowable limits a message about the relaxation effect's achievement is formed and light stimulation session is automatically terminated. Checking the values of all parameters In is continuous throughout the session. After the session on the basis of the results a diagnostic message is formed according to a combination of optical parameters In set in the system's database system.

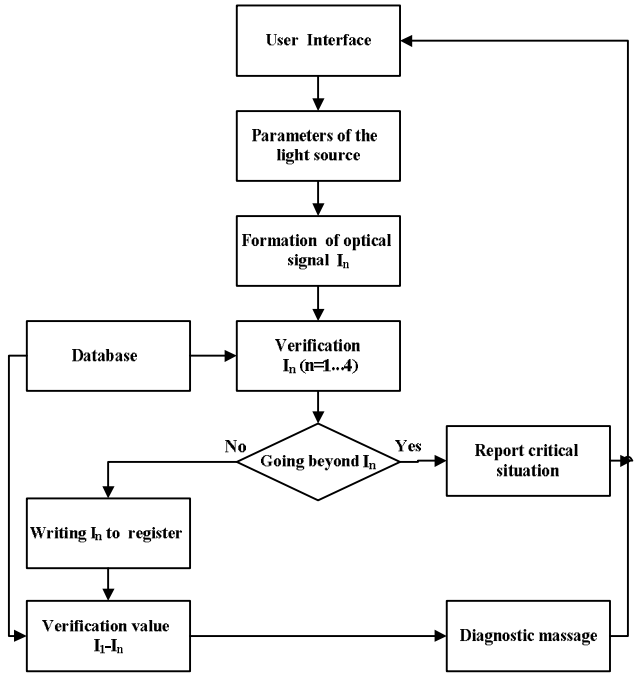


Fig. 6. The algorithm interactive mode of a light stimulated session.

The view of the test sample of the controlled light stimulation device is shown in Fig. 7. Channel 1 provides programmable light pulses to visual analyzer via a light guide's cable to dimming glasses. Channel 2 provides programmable photo-stimulation through other receptor areas.

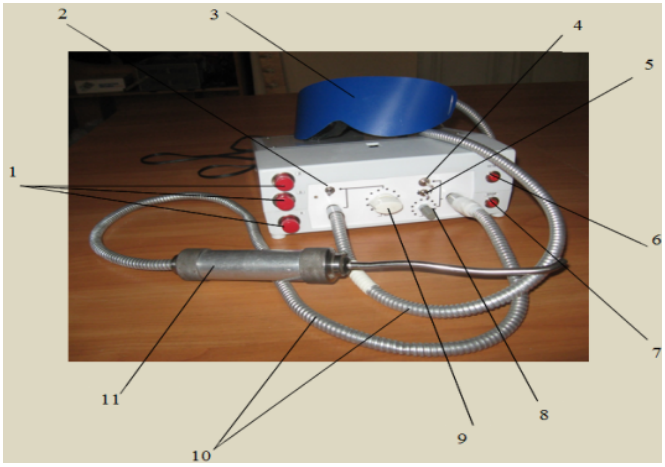


Fig. 7. 1– on / off RGB 1st – channel, 2 – mode switching manual / automatic 1-st channel, 3 – glasses for a patient, 4 – mode switching manual / automatic 2-channel, 5 – choice programs in incorporating LEDs 2nd channel, 6 –RESET reset switch program of the 2nd channel 7 – STOP program stop switch of the 2nd channel, 8 – switching frequency control of the 2nd channel in manual mode, 9 – frequency control of the 1st channel, 10 – fiber optics cables, 11 fiber-tip – instrument.

III. CONCLUSION

In the developed OES implemented a hardware and software system that allows contactless obtaining effective results of the current changes in the organism's the psychophysical state during the session of the light- stimulation radiation program.

Software and hardware of the developed OES with the current values of change in indicators reflect the results of analysis and a notification of preschedule termination of the current session or unreasonableness of further sessions.

REFERENCES

[1] S.Tonnies, Entspannung fur Tinnitusbetrofenne durch Photostimylation // Springer Medizin Verlag 2006,№ 54, c.481-486

[2] O. Kozhuhar The programmable system is for photomedic technologies/ O. Kozhuhar, Didych, A. Zazulyak// Materialy konferencyjne XIII Miedzynarod. Szkoly Komputerowego wspomaganja projektovanja. Wytwarzania I eksploatacji. Jurata. 11-15 05.2009. , Warszawa, Poland.-2009.ISBN 978-83-61486-13-8. p. 101-106.

[3] Kozhukhar, O.Tkachenko Therapeutic and diagnostic devices based on encephalographic frequencies programmed low intensity photostimulation on visual Receptors/ Zazuljak A, Kozhukhar O., Tkachenko O.// Proceedings of the 2 nd Forum Science&Technology Days Poland - East. Forum Catalogue. Bialystok. IICoE. Poland.-2009. p. 42.

[4] Kozhukhar O.T. The influence fotostymuliv with frequencies on human biorhythms body/ Kozhukhar O.T, Zazuliak A.M., Kuchak E.V.// "Electronics" Bulletin ou "LP", № 680, publ ou "LP", Lviv, 2010. - S. 225-230.

[5] Robert Ashby. Designer's Guide to the Cypress PSoC (Embedded Technology).

[6] D. Kilochek Design on programmable systems-on-chip PSoC Cypress / Kilochek D. // Components and technology. 2006. № 4.

Differentiation of Moist Fine Rales and Crackles by Polyspectral Analysis

Anna Poreva, Vladimir Fesechko
Physical and Biomedical Electronics Department
National Technical University of Ukraine "Kyiv Polytechnic Institute"
Kyiv, Ukraine
porevanna@gmail.com

***Abstract*—The problem of moist fine rales (MFR) and crackles differentiation is considered. New method MFR and crackles separation, based on polyspectral analysis, is proposed. In this method bispectrum is calculated at first and bifrequency corresponding to the maximal spectral amplitude is detected. The specific frequency bands for MFR and crackles are identified. If the frequency is in the uncertainty zone, then parametric bispectrum is calculated. This method allows us to separate MFR from crackles for more accurate diagnosis.**

Keywords—moist fine rales; crackles; bispectrum; parametric bispectrum

I. INTRODUCTION

At present a rapid growth of respiratory system diseases is observed in the world. This makes the relevance of timely diagnosis of respiratory disease. Auscultation of breath noise (sounds) during respiratory patient examination remains the main non-invasive method of respiratory disease diagnosis. Advances in signal processing allow comprehensive analysis of respiratory noise. Objectification of auscultation features is improved due to quantitative evaluation of the registered sounds with the functional state of body systems. Identification of such correspondence is a major task to be solved by physician [1].

At present number of approaches to the analysis of respiratory sounds are used, such as frequency analysis [2, 3], the time-frequency [4, 5], wavelet analysis [6], etc. Each of them has its advantages, but no one possesses absolute universality, and therefore research in this area is continuing.

Possibility of higher-order cumulant analysis to identify sounds crepitations (called crackles) and the effectiveness of this approach were considered in earlier works [7, 8].

Lung-sound nomenclature has long suffered from imprecision and ambiguity. Until the last few decades, the names of lung sounds were derived from the originals given by Laennec and translated into English by Forbes. These names carried the implication of the pathologic mechanism of their production, e.g., humid or dry rales, or the character of the sound, e.g., hissing rale. The need for a more objective naming system has long been recognized. In 1985, at the 10th meeting of the International Lung Sounds Association, an ad hoc committee agreed on a schema that included fine and coarse crackles, wheezes, and rhonchi. Each of these terms can be described acoustically and does not assume a generating mechanism or location. These terms are now widely accepted, although the term "rale," generally meaning "crackle," is still frequently used.

Further classification of lung sounds is still vague. No single characteristic distinguishes perfectly between the fine and coarse crackles although combinations of features provide adequate discrimination. The fact that fine and coarse crackles tend to appear at different times within the inspiratory cycle assists in their differentiation [9]. Another, more extensive classification of lung sounds adopted in the countries of CIS [10]. Presented study is based on this terminology. According to this classification rales and crackles are two different sounds and they have different acoustic and physiological properties. Although the terms crackles, wheezes and rhonchi entered into wide practice to refer to the three main types of additional respiratory noise, the obsolete terms rales (crackles) or crepitus remain in use as shown by surveys of physicians and pulmonologists and analysis of clinical descriptions of a number of medical journals [11]. To avoid ambiguity, we will use the terms *fine rales*, adding the word *moist* to clarify the origin of the sound and *crackles*. Recall that terms fine rales and crackles are rather synonyms in the classification adopted by International Lung Sounds Association, but in fact these two sounds are different in their properties. To further understand what it is about, we introduce the definition of these abnormal breath sounds, which can indicate pathology.

Fine rales are additional sounds occurring in the respiratory tract as a result of movement in them, in addition to air, liquid or viscous contents. Crackles are abnormal breath sounds which are formed by alveoli walls when they are more humid than usual and lost elasticity [12, 13].

Sometimes crackles according to their acoustic properties are similar to moist fine rales (MFR) that are formed when a small bronchi or bronchioles accumulated liquid secret. For this reason it is important to distinguish the crackles from MFR. Resistant crepitation indicates the presence of pneumonia, and MFR can indicate inflammation of the bronchi (bronchitis) Physicians` observations indicate, that some moist rales are easy to confuse with crackles during auscultation. This leads to misdiagnosis and therefore the differentiation of these sounds is important [14].

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

The aim of the study was to find a method to separate MFR from crackles.

II. POLYSPECTRAL ANALYSIS

There are three motivations behind the use of high order spectrum (HOS) in signal processing: 1) to suppress Gaussian noise of unknown mean and variance; 2) to reconstruct the phase as well as the magnitude response of signals or systems; and 3) to detect and characterize nonlinearities in the data [15]. The solution of the above problems in relation to the sound of lung is an important task because the real sound lung signals are influenced by various noise, and the respiratory system is very heterogeneous structure.

Some crackles and MFR were analyzed by a higher-order cumulant and spectral analysis.

This paper proposes an iterative approach to the sounds separation. In the first stage non-parametric estimation is used to obtain bispectrum of the signal. The bispectrum is analyzed to decide, whether the sound belongs to the class of crackles or MFR. If such decision can`t be taken at this stage than move to the second stage of analysis. At this stage the non-parametric bispectrum is calculated, and final decision on the differentiation of the studied signals is performed.

A. Non-parametric estimation

At first the third-order spectrum (bispectrum) was found for all investigated sounds using the direct method (non-parametric estimation). This method consists of the following steps:

1) Segment the data into L records or M samples each. Let $X^i(k)$, $k = 1,\ldots,M$, represent the ith record.

2) Subtract the mean of each record.

3) Compute the Discrete Fourier Transform $F_x^i(k)$ of each segment, based on M points, i.e.,

$$F_x^i(k) = \sum_{n=0}^{M-1} X^i(n) e^{-j\frac{2\pi}{M}nk}, k = 0,1,...,M-1, i = 1,2,...,L \quad (1)$$

4) The third-order spectrum of each segment is obtained as

$$C_3^{x_i}(k_1,k_2) = \frac{1}{M} F_x^i(k_1) F_x^i(k_2) F_x^{i*}(k_1+k_2), i = 1,...,L \quad (2)$$

Due to the bispectrum symmetry properties, $C_3^{Xi}(k_1,k_2)$ need to be computed only in the triangular region $0 \le k_2 \le k_1$, $k_1+k_2 < M/2$.

5) In order to reduce the variance of the estimate additional smoothing over a rectangular window of size ($M_3 \times M_3$) can be performed around each frequency, assuming that the third-order spectrum is smooth enough, i.e.

$$\tilde{C}_3^{x_i}(k_1,k_2) = \frac{1}{M_3^2} \sum_{n_1=-M_3/2}^{M_3/2-1} \sum_{n_2=-M_3/2}^{M_3/2-1} C_3^{x_i}(k_1+n_1, k_2+n_2) \quad (3)$$

6) Finally, the third-order spectrum is given as the average over all third order spectra, i.e.

$$\hat{C}_3^{x_i}(\omega_1,\omega_2) = \frac{1}{L} \sum_{i=1}^{L} \bar{C}_3^{x_i}(\omega_1,\omega_2), \omega_i = \frac{2\pi}{M} k_i, i = 1,2 \quad (4)$$

The final bandwidth of this bispectrum estimate is $\Delta = M_3/M$, which is the spacing between frequency samples in the bispectrum domain [16,17].

B. Parametric estimation

Phase relationships of signals are lost in the power spectrum and in the autocorrelation function. Detection and quantitative description of nonlinearities in signals (not only stochastic) is possible with higher order spectra. Such signals arise, when they are passed through the systems with nonlinear characteristic. The human body due to its inhomogeneity can be an example of such nonlinear system.

Assuming that a real-valued lung sounds sequence $\{x(n)\}$ with zero mean is expressed by a Pth-order Autoregressive (AR) model:

$$x(n) = \sum_{k=1}^{P} a_k x(n-k) + \nu(n) \quad (5)$$

where $\nu(n)$ is third-order stationary non-Gaussian white noise with zero mean. It satisfies $E[\nu(n), \nu(n+m)] = Q\delta(m)$, and $E[\nu(n)\nu(n+k)\nu(n+m)] = \beta\delta(k,m) \neq 0$, where both Q and β are constant.

Owing to higher-order cumulants being insensitive to Gaussian process, we assume third-order moment of $\nu(n)$ is not zero that is $\beta = E[\nu^3(n)] \neq 0$, then we can get

$$C_{3,x}(-l,-m) + \sum_{k=1}^{p} a_k C_{3,x}(k-l,k-m) = \beta\delta(l,m), l,m \ge 0 \quad (6)$$

where l and m are arbitrary time delays, $\delta(l,m)$ is the 2-D unit impulse function. Choose $(p+1)$ slices of third-order cumulant $C_{3,x}(-l,-m)$ at lags $l = m = 0,1,2,...,p$ and obtain the matrix equation:

$$\begin{bmatrix} C_{3x}(0,0) & C_{3x}(1,1) & \cdots & C_{3x}(P,P) \\ C_{3x}(-1,-1) & C_{3x}(0,0) & \cdots & C_{3x}(P-1,P-1) \\ \vdots & \vdots & \ddots & \vdots \\ C_{3x}(-P,-P) & C_{3x}(-P+1,-P+1) & \cdots & C_{3x}(0,0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \beta \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (7)$$

Solve it to get the p parameters $(a_1,a_2,...,a_p)$ of AR model and then obtain the parametric bispectrum estimation:

$$S_{3x}(\omega_1,\omega_2) = \beta H(\omega_1) H(\omega_2) H^*(\omega_1+\omega_2) \quad (8)$$

where $H(\omega) = 1 \Big/ (1 + \sum_{k=1}^{p} a_k \exp(-j\omega k)), |\omega| \le \pi$, and ω_i are normalized frequencies (bifrequencies) [16].

If we consider the AR-model for three harmonics with frequencies f_k, k=1,2,3, that harmonics are quadratically phase coupled if $f_3 = f_1 + f_2$. In this case ideally $H(\omega)$ contains impulses at $\pm f_k$, and $S_{3x}(\omega_1, \omega_2)$ is given by

$$S_{3x}(\omega_1,\omega_2) = \left(\sum_{k=1}^{3}\delta(\omega_1 \pm f_k)\right)\left(\sum_{k=1}^{3}\delta(\omega_2 \pm f_k)\right)\left(\sum_{m=1}^{3}\delta(\omega_1+\omega_2 \pm f_m)\right) \quad (9)$$

As discussed earlier due to properties of bispectrum it is enough to consider the bispectrum within $0 \le \omega_2 \le \omega_1 \le 1/2$. If $f_2 \neq 2f_1$, the parametric bispectrum will show a single impulse at (f_1, f_2), corresponding to $k = 1$, $l = 2$, $m = 3$ in (9), indicating the existence of the quadratically phase-coupled triplet (f_1, f_2, f_1+f_2). If $f_2 = 2f_1$, the bispectrum will have an additional peak at $(\omega_1, \omega_2) = (f_1, f_2)$, corresponding to $k = 1$, $l = 1$, $m = 2$ in (9), which can also result from the phase-coupled triplet (f_1, f_2, $2f_1$). [17].

III. Experimental results

The 23 crackle sounds and 30 MFRs were analyzed in this paper. At first bispectrum contour representation were obtained for test sounds of crackles and MFR (Fig.1 and Fig.2). According to bispectrum properties it is fully defined in the whole plane (ω_1, ω_2), if there are given values in the range $0 \le \omega_2 \le \omega_1$ (shaded area of I in Fig.1,a). [15,18]. (ω_1, ω_2 - bifrequencies normalized by sampling frequency 8000 Hz). All MFR have bispectrum peaks near the point of intersection of frequencies $\omega_1 = \omega_2 = 0{,}02$. (The peaks appear in the topogram as central contour line to each segment). Some examples of bispectral analysis of MFR are shown in Fig.1 a)-d). Crackles have other polyspectral characteristics and don't have distinct numerical characteristics to assign them to a particular class of sounds. (Fig. 2 a)-d)). As shown in [7, 8] the third-order cumulant analysis is better to identify of crackles.

Then the bispectrum slice along the main diagonal was analyzed. Figure 3 shows a plot of the diagonal slice of the MFR and Figure 4 shows the diagonal slice of the crackles

The relation of the bispectrum amplitude from bifrequency is symmetric around zero, so below we will consider only the positive frequency. We analyzed the bifrequencies at which the diagonal slice peak has maximum amplitude Am. It was determined that frequency distribution with the maximum bispectrum amplitude for crackles is in the range of f_{aver_cr} = (0,02225 ± 0,00505), and for MFR f_{aver_MFR} = (0,03381 ± 0,0104). The probability distribution of frequencies are shown in Figure 5.

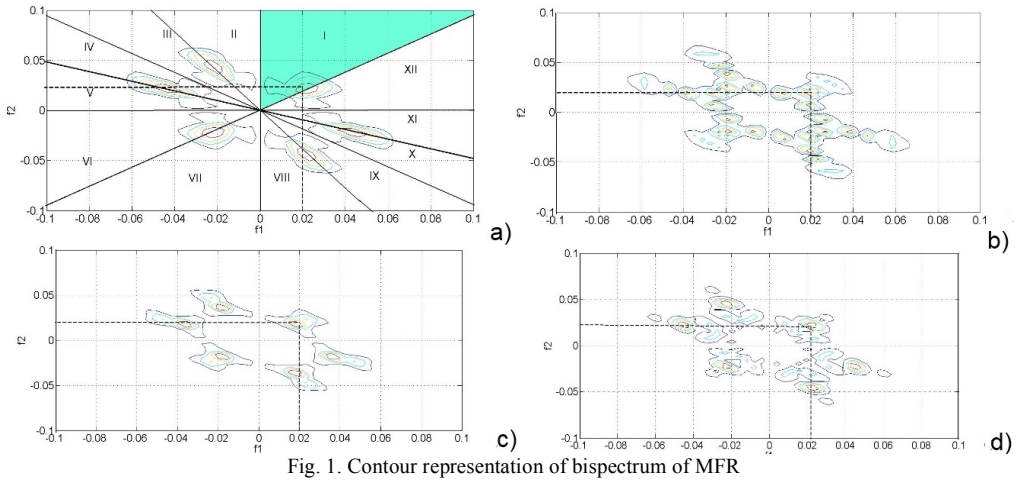


Fig. 1. Contour representation of bispectrum of MFR

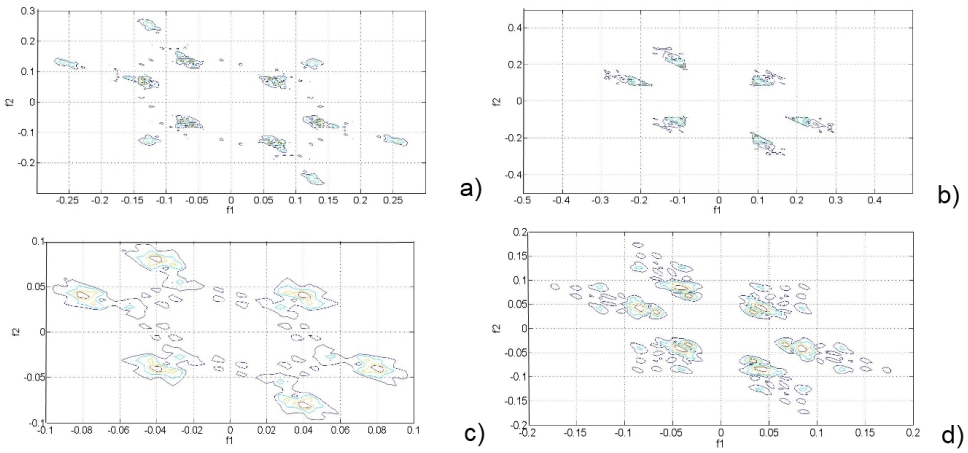


Fig. 2. Contour representation of bispectrum of crackles

If the found frequency falls in the indeterminate range, marked by shading in Figure 5, then parametric bispectrum is calculated according to (8).

Analyzing parametric bispectrum, we can conclude that the crackles' maximum of bispectrum is at the intersection of frequencies f1=f2 and for MFR maximum value is determined by f2 = 0 (Fig.6, Fig.7).

Thus, method for differentiation crackles and MFR is proposed. At first bispectrum calculated by direct method estimation. Diagonal slice is analyzed in frequency domain, which means the sound may be classified under a particular class. If it is impossible to make a clear conclusion at this stage of the analysis that parametric bispectrum is calculated to clarify. As a result a decision about classification of sounds to class of crackles or MFR is accepted.

IV. CONCLUSIONS

We investigated the results of lung sounds bispectral analysis. Use of bispectral analysis in the early stages patient examination provides reliable and simple mechanism to differentiate crackles and MFR sounds. Their identification is important for correct diagnosis and treatment. For more detailed information on the properties and parameters of the study of sounds and for the next stage of processing one can use other types of signal analysis: cumulant, spectral or spectral-time. The proposed method allows physician to make an accurate diagnosis and therefore assign appropriate treatment.

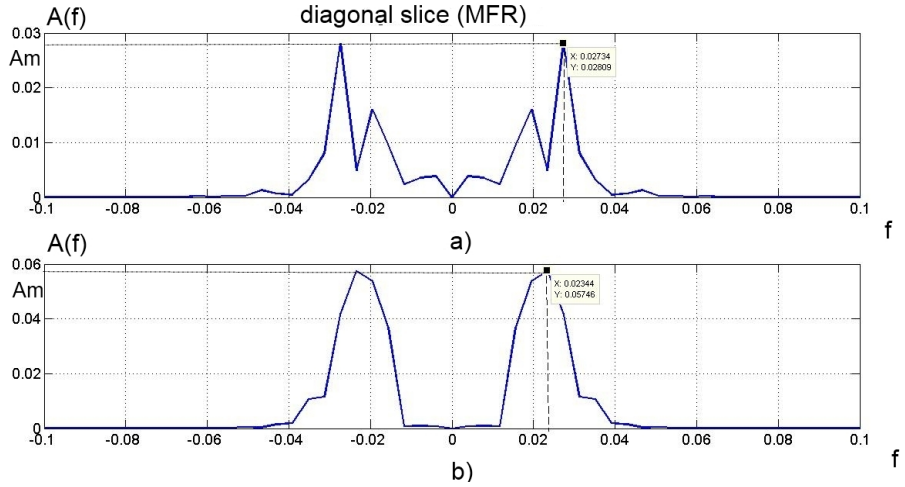


Fig. 3. The diagonal slice of the MFR.

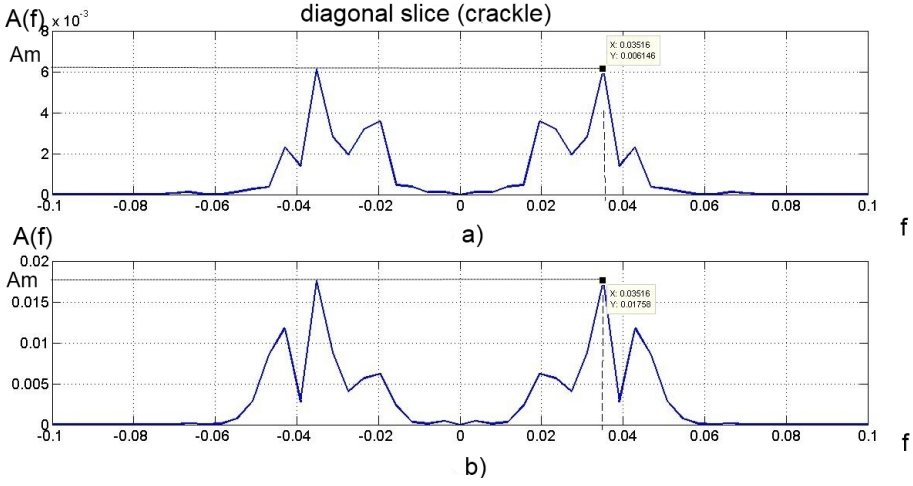


Fig. 4. The diagonal slice of the crackles.

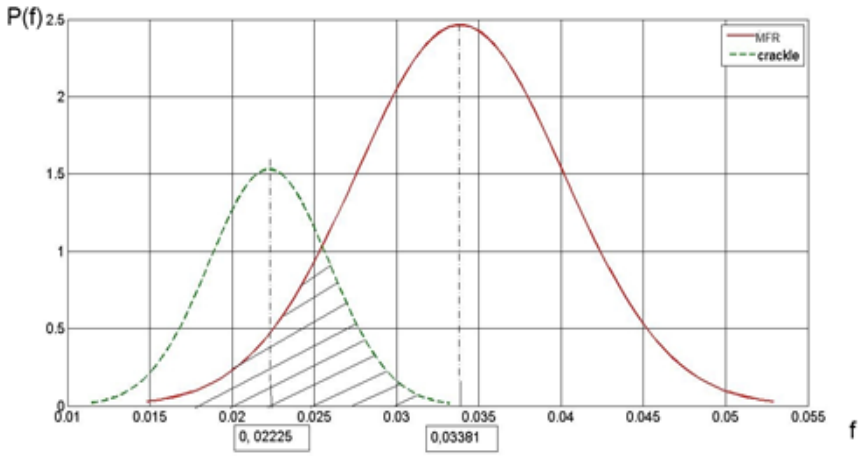


Fig. 5. The probability distribution of frequencies of the MFR and crackles

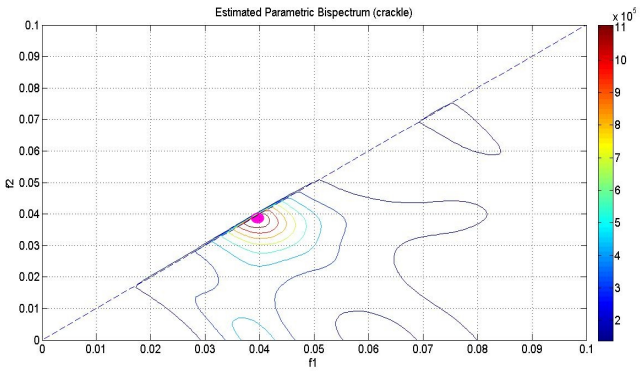


Fig. 6. Parametric bispectrum of crackles.

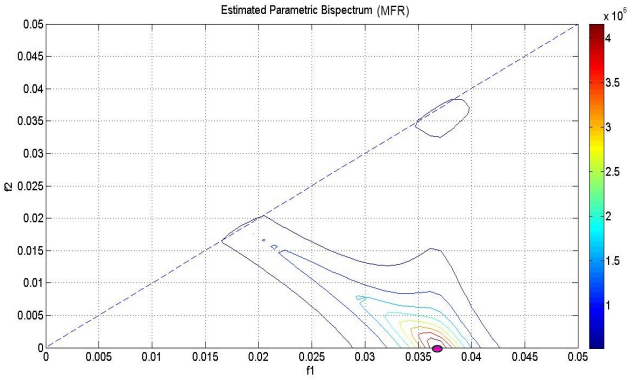


Fig. 7. Parametric bispectrum of MFR

REFERENCES

[1] Basanets A., Yermakova O., Makarenkov A., Makarenkova A. “Objective assesment of breath sounds in patients with COPD” Ukrainian journal on the problems of occupational medicine. - 2010. №3(23). - P.47-55. (In Russian)

[2] Murphy RL, Vyshedskiy A, Power-Charnitsky VA, Bana DS, Marinelli PM, Wong-Tse A, Paciej R. “Automated lung sound analysis in patients with pneumonia” Respir. Care.– 2004.– 49 (12).– P. 1488–1489.

[3] Zislin DM, Rosenblatt VV, Likhachev EI “Objective study of respiratory sounds with a frequency analysis” Therapist. archive. - 1969. - 41, № 11. - P. 108-112. (In Russian)

[4] Makarenkova A. “Acoustic characteristics of breath sounds in patients ill with pneumoconiosis” ISSN 1028 -7507 Acoustic bulletin. - 2008. V.11, №1, P.51-59. (In Russian)

[5] Alice Jones, Douglas Jones, Kevin Kwong,Yvonne Burns. “Effect of Positioning on Recorded Lung Sound Intensities in Subjects Without Pulmonary Dysfunction” in Physical Therapy. Volume 79. Number 7. July 1999

[6] J. Knocikova, J. Korpas, M. Vrabec, M. Javorca. “Wavelet analysis of Voluntary Cough Sound in Patients with Respiratory Diseases” in Journal of Phisiology and Pharmacology. – 2008 – 59 (6). – P. 331-340.

[7] Fesechko V.A., Poreva A.S. “Analysis of respiratory sounds by higher-order spectra” Electronics and Nanotechnology 2011. №2(61). C. 119-125 (In Russian)

[8] A. Poreva, P. Danilov, V. Fesechko. “Cumulant Analysis for the Detection of Crackles in Lung Sounds” Proceedings of ELNANO' 2012. P.188-189

[9] Hahs Paeterkamp, Steve S. Kraman, and George R. Wodicka “Respiratory Sounds. Advances Beyond the Stethoscope” American Journal of Respiratory and Critical Care Medicine. – 1997. vol.156. P.974-987

[10] The Great Soviet Encyclopedia. - M. Sov. Encycl., 1975. - V.7, P. 532-534 (In Russian)

[11] S. Mandzhoni. “Secrets of clinical diagnosis” M.: "BINOM", 2004. - 608 p., Ill. pp. 381-383 (In Russian)

[12] Fundamentals of Pulmonology. Guide for Physicians / ed. A Cocosov - M.; "Medicine", 1976. (In Russian)

[13] Gavriely N. “Breath sounds methodology” / N. Gavriely, D. Cugell. – Boca Raton, Fla : CRC Press,1995.– 481 p. – [ISBN]

[14] S. Mandzhoni. “Secrets of clinical diagnosis” M.: "BINOM", 2004. - 608 p., Ill. pp. 386-387 (In Russian)

[15] Petropulu, A. P. “Higher-Order Spectral Analysis. The Biomedical Engineering Handbook”: Second Edition. Boca Raton: CRC Press LLC, 2000

[16] LI Sheng-jun, LIU Yi “Feature Extraction of Lung Sounds Based on Bispectrum Analysis” Transactions of the IRE Professional Group on Vol. 30, No. 6 Third International Symposium on Information Processing, P. 393-397

[17] Ananthram Swami, Jerry M. Mendel, Chrysostomos L. (Max) Nikias “Higher-Order Spectral Analysis Toolbox. User’s Guide Versio”

[18] Bochkov G.N., Gorohov G.V. “Polyspectral analysis and synthesis of signals”: Educat.-methodol. Nizhny Novgorod State University. Nizhny Novgorod. 2007

Automated Parallel Detection Based Analyzer for Integrated Bioassays in Digital Microfluidic Biochip

Pranab Roy, Mahua Raha Patra, Hafizur Rahaman
School of VLSI Technology
Bengal Engineering and Science University,Shibpur
Howrah,West Bengal,India
ronmarine14@yahoo.co.in, mahua75@yahoo.co.in,
rahaman_h@yahoo.co.in

Parthasarathi Dasgupta
Indian Institute of Management,Calcutta
Kolkata,West Bengal,India
partha@iimcal.ac.in

Abstract — **In this paper we have proposed a design of an intelligent detection and analysis system that has been interfaced with a digital microfluidic biochip. The system is capable of performing automated detection and analysis of multiple bioassay protocols executed simultaneously within the same microfluidic device. The proposed design has been tested for both heterogeneous as well as homogeneous detection simultaneously within a scalable and integrated DMFB system for human physiological samples.**

Keywords — Digital Microfluidic Biochips, Automated detection and analysis, droplet manipulation, Intelligent system, Embedded design, digital analyzer, Photodiodes and optical detection.

I. INTRODUCTION

The recent advancement in microfabrication and microfluidic technology has made it possible to shrink an entire laboratory based protocol into a single lab-on-chip system. The development of such devices has made great impact in the domains of biotechnology, pharmacology, medical diagnostics, forensics, environmental monitoring and other basic research. A popular class of commercially available microfluiidic device termed as Biochip is based on continuous fluid flow being controlled by micropumps, microvalves, electro kinetics, or electroosmosis. In late 1990s an alternative class of lab-on-chip system termed as Digital microfluidic biochip (DMFB) has been emerged. In DMFBs the micro-drops are moved and treated individually as digital entities on a planar surface allowing for the scope of developing a highly reconfigurable system. The process of manipulation of nanolitre samples are carried out through different mechanisms, viz. electrowetting [1-2], dielectrophoresis [3], thermo capillary transport [4], and surface acoustic wave transport [5].

Electro Wetting on Dielectric (EWOD) is currently used as one of the best actuation methods for droplet manipulation in digital microfluidic systems [7]. Generally, the mechanism of electro wetting phenomenon is based on control of surface tension of a liquid-solid interface by applying electrical potential at the interface. Whenever a droplet is kept in contact with a solid electrode, a wetting force can arise upon application of an electric field through these electrodes [9]. Hence by application of a programmed sequence of electric potential at consecutive electrodes a strength disparity of this wetting force is created .This results in movement of droplets from one electrode contact to an adjacent one.

The microfluidic array contains a set of basic cells that is made up of two parallel glass plates (Figure 1).The bottom plate contains a patterned array of individually controllable electrodes, and the top plate is coated with a continuous ground electrode. The filler medium, such as silicone oil, along with the sample droplet is sandwiched between the two plates. By assigning time-varying voltage values to turn on/off the electrodes on the digital microfluidic biochip, the interfacial tension of the droplets are modulated- resulting in their transportation around the entire 2-D array and execution of fundamental microfluidic operations for different bioassays. The operations performed by actuating control voltages at the electrodes are also called reconfigurable operations because of their flexibility in location and in execution time. There is another variation termed as coplanar designs where both the buried activation electrode and the exposed electrodes that ground the droplet are located on the bottom surface (Figure 2).

On completion of the entire protocol for a specific sample the resultant droplet is transported to an optical detection site comprised of a cell dedicated for optical sensing using the integrated optical detection mechanism. Detectors such as photodiodes and LEDs are either integrated or aligned in conjunction (for earlier designs) with the biochips to perform the task of optical detection [10]. For colorimetric or chemo luminescence based tests a separate detailed analysis has to be carried out based on the results obtained from optical detection.

With the present development of microfabrication techniques, it was possible to accomplish the intimate integration of optical components with microfluidics technology for the digital microfluidic platform. This chip scales integration of digital microfluidics with active compound semiconductor devices has the potential to develop next generation portable Lab-on-Chip systems that may completely contain optical generation, processing, and detection [17]. In DMFBs, the optical detection is performed on the droplet under test through a photodiode integrated in the upper glass plate along a plane perpendicular to the 2D grid. Based on the intensity of light based on droplet luminescence incident on the photodiode an output current is measured. The absorbance of incident light by the test droplet is directly related to the concentration and presence of other parameters under consideration within the sample. The measured value of the detected current provides major suggestive and confirmatory evidence towards conclusive diagnosis involving clinical and biomedical applications.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

However the analysis based on the optical detection is highly dependent on human intervention where accuracy and precision may become a major factor for clinical diagnosis. In this work, we have proposed a new design of a photodetection based digital analyzer circuit that operates both for heterogeneous as well as homogeneous sample based operations. The purpose of this design is to perform an intelligent detection and analysis of the results obtained from DMFBs through comparison with predefined standard values stored in an integrated memory. In this design we initially partitioned a biochip into separate regions for execution of separate sets of bioassay protocols and integrated a group of photodetection analyzer for concurrent detection and automated analysis within a single DMFB device. The overall combination of results obtained from multiple detection sites can enhance the process of conclusive detection related to exquisite medical diagnostic applications and treatment as well as formation of disease profiles.

The organization of the paper is as follows: Section II detailed the prior works regarding optical detection in DMFBs. Section III states the data acquisition and significance of the diagnostics performed in DMFB respectively. Section IV briefs about the two modes of detection analyzer circuits to be integrated with the Biochip in order to form the integrated analyzer system. Section V describes the final integrated system design together with its operation in parallel mode. Section VI displays the experimental results and concluding remarks being stated in section VII.

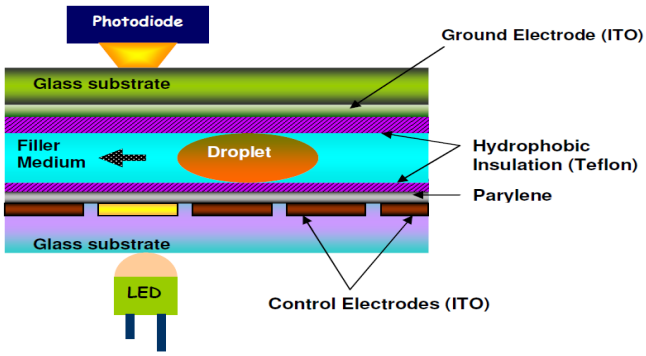


Fig. 1. Schematic diagram for a digital microfluidic biochip

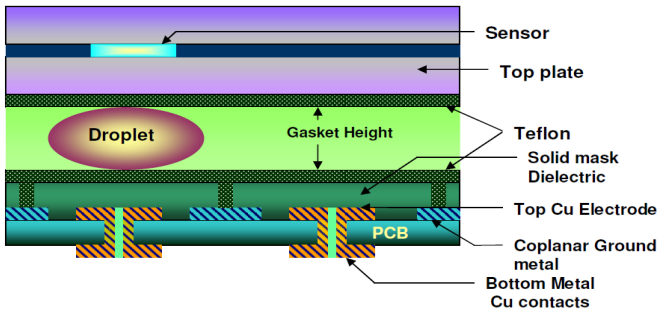


Fig. 2. Schematic diagram for a Coplanar DMFB

II. Prior Works

So far very few research works have been reported in the areas of optical detection and integration of optical detectors in digital microfluidic biochips. A colorimetric enzyme-kinetic method based on the Trinder's SZ reaction used for the determination of glucose concentration is demonstrated in [10]. The on-chip glucose assay is performed in three steps—dispensing, mixing and detection. At the end of the mixing phase, the absorbance is measured for at least 30 s, using an LED-photodiode setup comprised of a green light emitting diode, a photodiode (TSL 257) which is combined with a trans impedance amplifier integrated on the same monolithic device. Major drawback of the design lies in the fact that optical absorbance detection scales poorly with miniaturization, since Beer's law incorporates path length dependence [11]. Regarding the detectors reported in [10] the optical path length typically was 100-300µm which is 30 to 100 times smaller than conventional systems (10mm). Such small path length poses major sensitivity problems, and limits the use of absorbance to assays with very high analyte concentrations [12].In [13] a point of care testing (POCT) device has been proposed for PCR manipulation that contains a miniature fluorimeter, comprising of an LED and a photodiode which was aligned to specific areas on the chip to enable real-time detection for the PCR reaction. The excitation wavelength was 475 nm and the emission filter's wavelength was centered on 525 nm. It was used with a long-pass dichroic mirror. Chemiluminescence's based detection has been found to be highly compatible with digital microfluidic platform [14-15] together with diagnostic applications as well as DNA sequencing by synthesis [16].A major contribution has been reported in [14] where a compound semiconductor thin film metal-semiconductor-metal photodetector (MSM photodetector) was integrated in to an electrowetting microfluidics system so that an optical detection function could be integrated with a fluidic sample processing system. The InGaAs-based MSMs has been used for this work. The thin film photodetectors were heterogeneously integrated onto the glass substrate by transferring the devices from the diaphragm to the metal pad on the glass substrate. The optical detection function of the integrated photonic and microfluidic system was tested by mixing two droplets that participate in a chemiluminescent reaction. The reaction was reported as the oxidation of pyrogallol (1,2,3-trihyroxibenzne) in the presence of formaldehyde, in an alkaline solution, was chosen for its luminescent intensity and its peak wavelength.

A new cyber-physical DMFB design based on the optical detector integration and the corresponding sensor data has been proposed in [18].This method develops a "physical-aware" system reconfiguration technique that uses sensor data at checkpoints to dynamically reconfigure the biochip. In a more recent work [19] a digital microfluidic architecture is introduced for micron-scale localized fluid actuation and in in-situ optical sensing. The on-chip sensing technique proposed in the design incorporates a micro lens in a folded-cavity arrangement . In [20] another major work has been reported for automated detection based analysis and multilevel Biochip integration (based on a set of test sequence) for homogeneous samples. The circuit uses an ADC as well as a photodiode integrated with TI amplifier for optical detection.

III. Data Acquisition

A. *Absorption spectra for blood*

Knowledge about the micro-optical parameters of human blood i.e. absorption coefficient µa, scattering coefficient µs and anisotropy factor g, and their dependence on physiological parameters such as haematocrit, oxygen saturation, osmolarity and haemolysis plays an important role in many diagnostic and therapeutic applications in laser medicine and medical diagnostics.[21] The optical parameters were

determined in the wavelength region 400-1000 nm for circulating blood for different concentrations of red blood cells (0,2-5,0 mill/µl), various degrees of oxygen saturation (60-100%), osmolarity (225-400 mosmol) and haemolysis (0-100%).

Colorimetric analysis used for determination of blood parameters in DMFBs is a method of determining the concentration of a chemical element or chemical compound in a solution with the aid of a color reagent. It is applicable to both organic compounds and inorganic compounds and may be used with or without an enzymatic stage. The in-vitro measurement of glucose and other metabolites, such as lactate, urea, glutamate, pyruvate and others in human physiological fluids (that involves enzymatic analysis) is of great importance in clinical diagnosis of metabolic disorders. For instance, the change of regular metabolic parameters in the patient's blood can signal organ damage or dysfunction prior to observable microscopic cellular damages or other symptoms.[22-23].

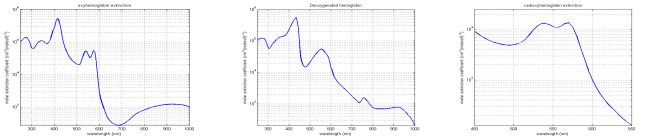
B. Optical detection and data acquisition process

We have used the same techniques for data acquisition as mentioned in [10],[23],[24]. the setup consists of a light emitting diode (LED) and a photodiode((TSL257). The Photodiode provides a light to voltage conversion which is obtained through a transimpedence amplifier on the same monolithic device. A 12-bit analog data acquisition board (PCI-DAS08) has been used for ADC conversion. The analog data acquisition board has a measurable range of −5 to 5V and an analog-to-digital (A/D) resolution of 2.44mV. The voltage V(t) is directly proportional to the light intensity incident on the photodiode and is related to the absorbance by the following equation:

$$A(t) = \ln\left(\frac{V_0 - V_{dark}}{v(t) - V_{dark}}\right) \quad \ldots (1) \quad [24]$$

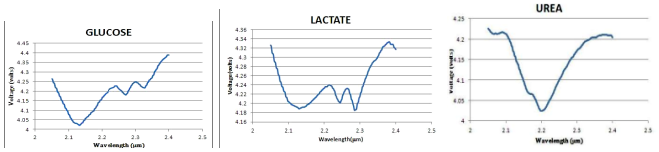
$$A = \frac{\varepsilon\left(\frac{1/cm}{moles/litre}\right).c\left(\frac{gm}{litre}\right).l(cm)}{gm\text{-}molecular\text{-}weight\left(\frac{gm}{mole}\right)} \quad \ldots . (2) \quad [24]$$

Where A = molar extinction co-efficient and C = concentration of substance .V_0 corresponds to zero absorbance (or 100% transmittance) and V_{dark} corresponds to the voltage output of the photodiode under dark conditions. The absorbance is measured for 45 s, using the LED-photodiode setup. The output is connected to the proposed analyzer for detection analysis. The test cases for Hemoglobin, Blood Glucose, Urea, Lactate and Pyruvate have been used in our proposed design for automated detection and analysis for medical diagnosis applications. The molar extinction coefficient based spectra for Hb and the Voltage vs. wavelength curves for Glucose, Lactate and Urea have been shown in figure 4.



a) Oxyhemoglobin b) Deoxyhemoglobin c) Carboxyhemoglobin

Fig. 3. Molar-ext-coefficient vs. wavelength curve for Hemoglobin [25]



a) Glucose b) Lactate c) Urea

Fig. 4. Voltage vs. wavelength curve for Glucose, Lactate and Urea [25]

IV. PROPSED DESIGN OF DETECTION ANALYZER

A. Design of Homogeneous detection analyzer

A specific protocol for same blood samples of a given type has been carried out repeatedly – for determination of the result on the basis of the majority occurrence of the specific parameter. So whenever the number of sample droplets detected to be positive are higher than remaining others -the final detection result is analyzed as positive and can be transferred as digital signal to another controller circuit (if any) or directly to the display unit. The procedure for analysis is as follows:

- Multiple droplets (in our case - 4) of a given sample are placed on the bioassay as per presynthesized schedule for a specified protocol.
- The test sequence in the microcontroller memory to execute the protocol on the first sample is carried out and the digital output from the photo detector is stored in a register stack.
- The individual results are compared with a reference result pre characterized from standard samples earlier stored in an LUT.
- The result from the comparator is added to an accumulator using a 2 bit full adder – being enabled on completion of the comparator function.
- A two bit counter increments the execution sequence number and actuates the test sequence for the next sample.
- The process continues until the protocol is completed on all four samples. An end of computation signal is generated that disables the operation of the adder.
- The final result from the accumulator is compared with a predefined threshold value K using a second comparator.
- If comparator output is found to be less than K – result is analyzed as negative ,otherwise as positive and is conveyed to the display unit.

The circuit diagram for the proposed analyzer is shown in figure 5.

B. Design of Heterogeneous detection analyzer

The design for heterogeneous detection analyzer is proposed for simultaneous execution of multiple bioassay protocol for different droplet samples taken one for each type. The results obtained for each type simultaneously at different detection sites are evaluated by referencing them with standard values for each type stored in a Look Ahead Table within the device memory and a digital output for each detection is combined together to formulate a binary string. Based on the binary pattern of this string- intelligent detection for multiple droplet samples (heterogeneous type) is achieved. The procedure for analysis is as follows:

• Droplets for multiple samples (in our case - 3) scheduled to be mixed with respective reagents for execution of different pre-sequenced Bioassay protocols are placed on the bioassay as per pre-synthesized schedule.

• The detection result obtained from Photo detectors for each droplet from the respective detection site is stored in a separate register stack termed as result storage for detection.

• The individual result for first detector is compared with the corresponding low reference result (pre-characterized from standard samples earlier) being stored in an Low reference LUT.

•If the test result is found to be less in comparison to the reference result, the test is considered to be negative and transferred to the display. Otherwise a second comparator is enabled for higher precision based analysis.

•On enabling second comparator(if necessary at all), the test results are compared with high reference value stored in High reference LUT and the final result being detected as high or moderate is transferred to the display unit.

On completion of analysis for first detection site – address for the next detector result storage is selected through a 2 bit counter .The same process is continued unless all storage locations are addressed. An End of Computation signal is generated that disables the shifting operation of the display register as well as the first comparator.

• Thereby the heterogeneous analyzer computes results for each individual droplet scheduled for different bioassay protocol and combines them together to form a binary pattern that generates the conclusive detection with intelligent integration of multiple Bioassays within a single biochip.

The circuit diagram for the proposed analyzer is shown in figure 6.

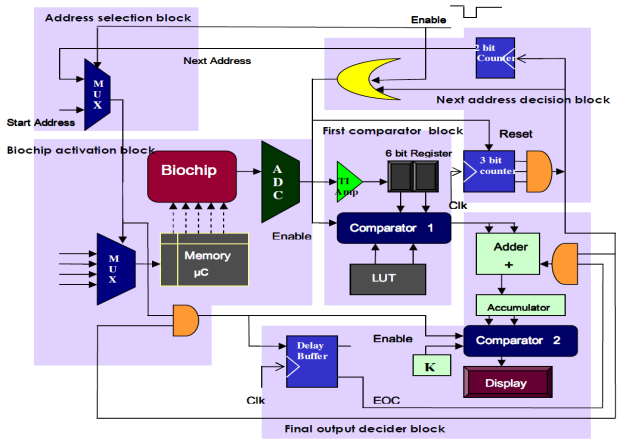


Fig. 5. Homogeneous sample based droplet detection analyzer circuit

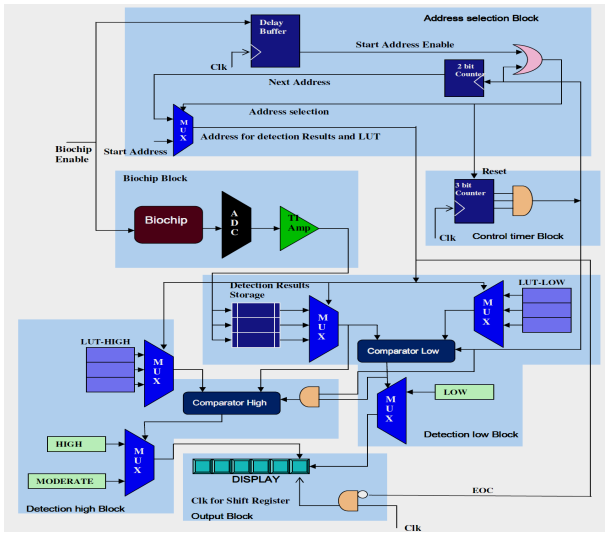


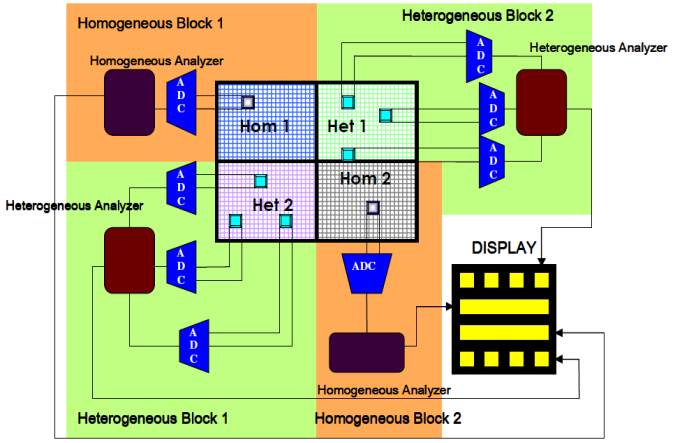
Fig. 6. Heterogeneous sample based droplet detection analyzer circuit

V. DESIGN OF INTEGRATED DETECTION SYSTEM

The two modes of analyzer as described in the previous section can be used separately for executing bioassay protocols in individual modes exclusively. In this design we formed an integrated system for automated detection and analysis within a scalable design of a biochip – capable of executing multiple bioassay simultaneously in different modes. The integrated system design is described as follows:

- Partition the biochip into a number of separate regions – each region being operated in a single mode (concurrently).
- For each homogeneous region a single photodetector is sufficient for sequential execution of repeated protocols. Whereas for heterogeneous regions a set of protocols are to be executed concurrently – and it becomes necessary to employ multiple photodiodes for these regions.
- Each region is connected through the corresponding photodiodes to the respective types of analyzers through the corresponding ADCs.
- On application of enables for each regions simultaneously the execution of bioassays for each region is started simultaneously.
- Homogeneous regions requires larger times for sequential operation of each protocol repeatedly – to obtain the final analysis results.
- However for heterogeneous regions – although the protocols are executed simultaneously and the results stored in the respective detection storage – the analysis for each storage location takes place sequentially and may also require certain amount of time. Although these timing requirements for heterogeneous regions are found to be less as compared with homogeneous regions. (As the execution time within DMFB is found to be considerably larger in comparison to the gate level applications.)

The block level schematic diagram for the integrated system is shown in figure 7.



Hom1 and Hom 2 - Homogeneous Regions
Het 1 and Het 2 - Heterogeneous Regions

Fig. 7. Schematic for the integrated detection and analyzer system

VI. EXPERIMENTAL RESULTS

For the homogeneous sample based region namely Hom 1 the measurement of the oxygen saturation in Hemoglobin using absorbance for Oxyhemoglogin and Deoxyhemoglobin for Four droplets of similar blood samples has been taken. While for the heterogeneous sample based region Het 1 four droplets:- two for determination of metabolite concentration of Glucose and the other two for Lactate have been taken. For the homogeneous region Hom 2, the measurement of blood glucose for four droplets of similar blood samples have been taken. Finally for heterogeneous region Het 2, four droplets with two for determination of Urea and the other two for Pyruvate concentration has been used. A separate simulation for homogeneous analyzer is implemented using

FPGA Spartan 3E architecture (Target device - xc3s500e-4fg320).The corresponding simulation (using ,Xilinx 9.1 ISim) results has been displayed in figure 8. Similary the heterogeneous analyzer is implemented using FPGA Virtex 4 architecture (target device – xc3s500e-4fg320). The corresponding simulation (using Xilinx 9.1 ISim) results has been displayed in figure 9.

In figure 10 the simulation result(using Xilinx 9.1 ISim) for the automated integrated system has been displayed. The design is implemented using FPGA Virtex 4 architecture (target device – xc3s500e-4fg320).The device utilization summary and the corresponding timing details is given in table 1 and 2.Table 3 gives the sample concentration and corresponding digital threshold values for glucose,Lactate,Urea and Pyruvate and the detection results are displayed in table 4.

Signal *lutout[6:0],*Signal *lutout1[6:0],* Signal *lutout2[6:0],* Signal *lutout3[6:0]* represent Biochip output of heterogeneous Block 1 and 2 and homogeneous block 1 and 2 respectively. lut_bio_out for heterogeneous module1 and heterogeneous module2, homogeneous module1 and homogeneous module2 are transferred concurrently to result storage stack . The address selection output for each module triggers the reference values of all four modules at the same time. *Signal result [1:0]* and *Signal result2 [1:0]* represents final output for first and second heterogeneous block respectively. *Signal result3 and Signal result4 represents final* output for first and second homogeneous block respectively.

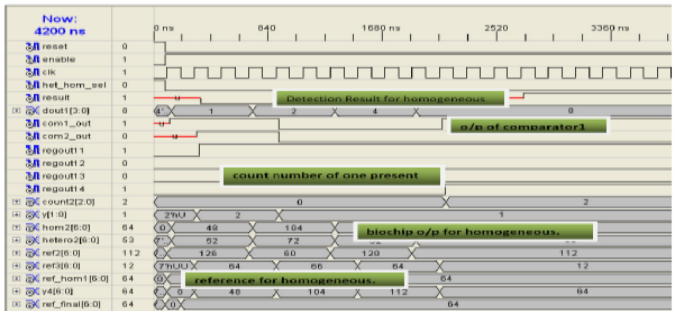


Fig. 8. Simulation result for homogeneous sample based analyzer

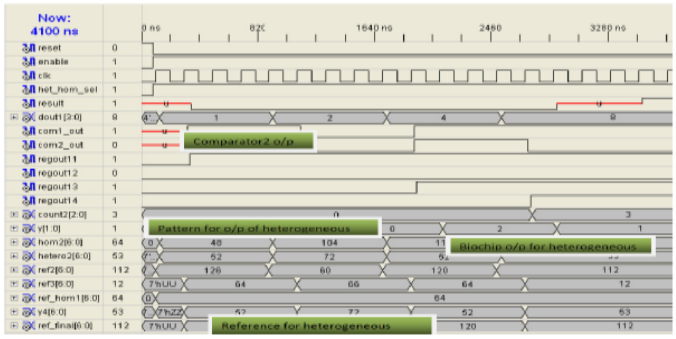


Fig. 9. Simulation result for heterogeneous sample based analyzer

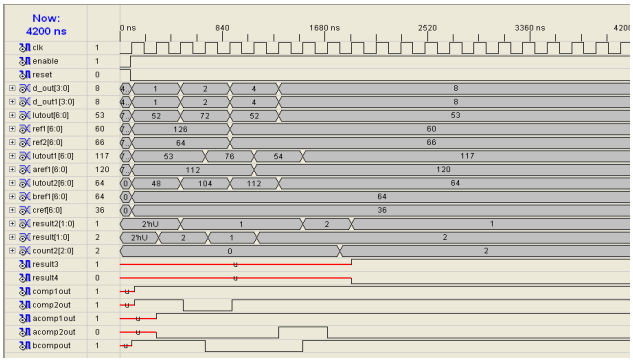


Fig. 10. Simulation result for integrated analyzer system

TABLE I. DEVICE UTILIZATION FOR INTEGRATED DETECTION SYSTEM

Device Utilization Summary (estimated values)			
Logic Utilization	Used	Available	Utilization
Number of Slices	59	960	6%
Number of Slice Flip Flops	90	1920	4%
Number of 4 input LUTs	102	1920	5%
Number of Bonded IOBs	58	66	87%
Number of GCLKs	2	24	8%

TABLE II. TIMING DETAILS FOR INTEGRATED DETECTION SYSTEM

Timing parameters	Values
Minimum period	10.024ns (Maximum Frequency: 99.761MHz)
Minimum input arrival time before clock	5.355ns
Minimum output required time after clock	6.763ns
Maximum combinational path delay	5.67 ns

TABLE III. SAMPLE CONCENTRATION DETAILS AND DIGITAL THRESHOLD VALUES

Metabolites	Normal concentration in human blood [mg/dl]	Gram molecular weight [gm/mole]	Digital value(normal)
Glucose (fasting)	< 100	180.15588	1000000
Urea	7-21	60.05526	1000000
Lactate	4.5 – 14.4	120.6	1000010
Pyruvate	300 – 900 (µg/dl)	NA	0001100

TABLE IV. FINAL DETECTION RESULTS OBTAINED - REGION WISE

Detection	Glucose	Lactate	Pyruvate	Urea	Hemoglobin
Reference High	1111110	0111100	1110000	1111000	
Reference Normal	1000000	1000010	0001100	1000000	0100100
Heterogeneous Outputs	Het 1	Het 1	Het 2	Het2	
	1111110, 1001000	0110100, 0110101	0110101, 1001100	0110110, 1110101	
	HIGH MODERATE	MODERATE MODERATE	MODERATE MODERATE	LOW MODERATE	
Homogeneous Outputs	Hom 1				Hom 2
	0110000, 1101000, 1110000, 1000000				0110000, 1101000, 1110000, 1000000
	HIGH				HIGH

VII. CONCLUSION

In this paper we have proposed an integrated analyzer system using simultaneous detection and analysis in two modes within a single scalable DMFB array. This analyzer system is highly applicable for on line clinical diagnosis and other detection based biomedical applications which provides a much higher precision and accuracy in comparison with human intervention based methods. The homogeneous analysis is used for implementation of a single Bioassay on multiple samples of similar type. Whereas the heterogeneous analysis is employed for intelligent detection using multiple bioassay protocols integrated in a single zone. There remains ample scope for further enhancement of the design by optimizing the hardware for sequential operation of the regions using the same analyzer as well as the storage memory.

REFERENCES

[1] M.G. Pollack, R.B. Fair, and A.D. Shenderov, "Electrowetting-based actuation of liquid droplets for microfluidic applications," Appl. Phys. Letters, 77 pp. 1725- 1727 (2000).

[2] J. Lee, H. Moon J. Fowler, C.-J Kim and T.Schoellhammer, "Addressable micro liquid handling by electric control of surface tension," Proc. of 2001 IEEE 14th International Conference on MEMS, Interlaken,Switzerland 499-502 (2001).

[3] R.C.Gascoyne, and J.V Vykoukal,"Dielectrophoresis based sample handling in general-purpose programmable diagnostic Instruments," Proc. IEEE 92 .22-42 (2004).

[4] D.A. Anton, J.P. Valentino, S.M. Trojan, and S Wagner,"Thermocapillary actuation of droplets on chemically patterned surfaces by programmable microheater arrays," J Microelectromechanical Sys. 12 .873-879 (2003)

[5] A. Renaudin, P. Tabourier , V. Zhang, C. Druhon and J.C. Camart "Plateforme SAW dédiée à la microfluidique discrète pour applications biologiques", 2ème Congrès Français de Microfluidique,Société Hydrotechnique de France, Toulouse, France, 14-16 (2004).

[6] S.K. Cho, H. Moon, and C.J. Kim, "Creating, transporting, cutting, and merging liquid droplets by electrowettingbased actuation for digital micro fluidic circuits," Journal of Microelectromechanical Systems, vol. 12, 2003.

[7] Alborz Arzpeyma, Shan Bhaseen, Ali Dolatabadi, Paula Wood-Adams,"A coupled electro-hydrodynamic numerical modeling of droplet actuation by electrowetting," Colloids and Surfaces A: Physicochem. Eng. Aspects, vol. 323, pp. 28-35,2008

[8] F Su, Jun Zeng," Computer aided design and test for digital microfluidics",IEEE Design and test for computers, pp. 60-70. Feb, 2007.

[9] J. Berthier, Microdropes and Digital Microfluidics, William Andrew Inc,Norwich, NY, 2008.

[10] V. Srinivasan et al., "An integrated digital microfluidic lab-on-a-chip for clinical diagnostics on human physiological fluids", Lab on a Chip, vol. 4, pp. 310-315, 2004.

[11] M.J. Madou and R. Cubicciotti, "Scaling issues in chemical and biological sensors." Proc. Of IEEE 91 830-838 (2003).

[12] R.B.fair," Digital Microfluidic Chips for Chemical and Biological Applications" 31st Annual International Conference of the IEEE EMBS Minneapolis, Minnesota, USA, September 2-6, 2009.

[13] Ramakrishna Sista , Zhishan Hua, Prasanna Thwar Arjun Sudarsan, Vijay Srinivasan, Allen Eckhardt ,Michael Pollack and Vamsee Pamula,"Developm ent of a digital microfluidic platform for point of care testing"Lab on Chip Journals 2008, 8, pg. 2091–2104.

[14] Randall Evans, Lin Luan, Nan Marie Jokerst, Richard B. Fair,"Optical Detection Heterogeneously Integrated With a Coplanar Digital Microfluidic Lab-on-a-Chip Platform" proceedings of IEEE SENSORS 2007 Conference,pp423-426.

[15] Lin Luan, Randall D. Evans, Nan M. Jokerst, Richard B. Fair, "Integrated Optical Sensor in a Digital Microfluidic Platform" IEEE Sensors Journal,vol.8 no. 5, pp. 628-635, 2008.

[16] P. Thwar, J.L. Rouse, A.E. Eckhardt, P. Griffin, M.G. Pollack ,R.B.Fair , "Digital Microfluidic DNA Sequencing," Advances in Genome Biology and Technology (AGBT) Meeting, Marco Island,Florida, Feb., 2009.

[17] L. Luan, R.D. Evans, D. Schwinn, R.B. Fair, and N.M. Jokerst, "Chip Scale Integration of Optical Microresonator Sensors with Digital Microfluidics Systems," LEOS-2008, Newport Beach, CA, Nov. 9-13, 2008.

[18] Yan Luo , Chakrabarty, K. ; Tsung-Yi Ho ," A cyberphysical Synthesis approach for error recovery in digital microfluidic biochips "- Proc. of Design, Automation and Test in Europe conference & exhibition(DATE).

[19] Jacqueline Nichols, Emily L. Landry, Brandon Born, Michael Wiltshire, Christopher M. Collier, & Jonathan F. Holzman ," Optical sensing for on-chip digital microfluidics " Microfluidics,BioMEMS, and Medical MicrosystemsX, Proceedings of the SPIE, Volume 8251, pp. 82510L-82510L-10 (2012).

[20] P.Roy,M.Shohid,S.Chakrabarty,H.Rahaman and P.S.Dasgupta," System on Biochips:a new design for integration of multiple DMFBs".Proc. of IEEE International symposium on Electronic System Design,2012,Kolkata,India.

[21] M. Meinke, M. Friebel, A. Lutter, I. Gersonde, A. Roggan ,"Determination of optical properties of human blood to develop new devices for blood monitoring"Journal of Biomedical Optics ,May/June 2006.

[22] V. Srinivasan, V. K. Pamula, M. G. Pollack and R. B. Fair, "A digital microfluidic biosensor for multianalyte detection'', Proc. IEEE Conf MEMS, pp. 327-330, 2003,

[23] V. Srinivasan, V. K. Pamula, M. G. Pollack and R. B. Fair," Clinical diagnostics on human whole blood, plasma, serum, urine, saliva,sweat, and team on a digital microfluidic platform", Proc. μTAS, pp.1287-1290,2003.

[24] Vijay Srinivasan, Vamsee K. Pamula, Richard B. Fair," Droplet-based microfluidic lab-on-a-chip for glucose detection",Transactions of Analytica Chimica Acta ,507- pp 145–150,2004.

[25] Scott Prahl " Optical absorption of Haemoglobin"- Oregon Medical Laser center – SAP 4th march,1998

On the Spirometry and Spirometers Standardization

Lopata V., Sinekop Yu., ElShabbah M.
Electronics Faculty,
National Technical University of Ukraine
"Kyiv Polytechnic Institute",
Kiev, Ukraine
v.lopata@meta.ua

Myasnyi I.
Center of pulmonology, allergology
and clinical immunology,
Clinical Hospital "Feofaniya",
Kiev, Ukraine
IvanLanzerx@gmail.com

***Abstract*— The problem of spirometry standardization is considered. We analyzed the standard of ATS/ERS (2005) and proposed the prospective requirements for spirometers of three levels. The possibility of flow sensors to meet these requirements is considered.**

***Keywords* – spirometry, spirometer, standardization.**

I. INTRODUCTION

The solving of spirometry problems require that spirometers meet a lot of conditions for adequately perform their functions. These conditions have been developed over many years and defined now as medical / technical requirements (MTR) for main features of devices:

1. Resistance to breathing, created by spirometer
2. Volume of airflow sensor "dead space"
3. Response threshold of on airflow rate
4. Range of airflow rate measuring
5. Limits of measurements error
6. Non-linearity of the frequency response in the range of relevant frequencies.

II. AIM AND METHODS

The aim of this study is a formulation of MTR for spirometers. As is known, from 2005 these MTR has taken the form of mutual standard (Table 1) of American Thoracic Society (ATS) and the European Respiratory Society (ERS) [1].

TABLE I. Parameters of MTR for spirometers, standardized in 2005

Standardized parameter	Values
Resistance to breathing, Pa·s/l	≤ 150
Threshold of response on airflow rate, l/s	≤ 0,050
Upper limit of volume measurements, l	8,0
Upper limit of airflow rate measurements, l/s	14,0
Limits of volume measurement errors, %	± 3
Limits of airflow rate measurement errors, %	± 3,5
Nonlinearity of dynamic characteristics in frequencies range till 15 Hz , %	± 5

As can be seen from this standard, spirometers are not differentiated by levels of the hierarchy for spirometric testing , which can be identified at least three [2]:

1. In Family medicine
2. In Hospitals
3. In Diagnostic Centers or Research Institutes.

Analysis of the ATS/ERS standard identifies the trends in the formulation of future tasks for spirometers design and normalization of their characteristics, differentiation of requirements for them according to the devices hierarchy levels.

1. Resistance to breathing is a main component of spirometry methodical errors. Therefore normalizing of resistance in limits of 150 Pa·s/l don't corresponded to up-to-date requirements. We consider the necessary to standardize it in limits of 20 Pa·s/l for children till 12 years aged and 40 Pa·s/l for adults.

2. "Dead space" volume of sensor must be ≤ 30-50% of the tidal volume [3].

3. Standardized threshold of response on airflow rate is sufficient, but should be differentiated according to the age of patients - the ratio of these values for children and adults should be at least 1:2.

4. Standardized upper limit of airflow measurements at 14 l/s is a physiological limit for the normal human respiration. For special spirometers modifications (surveys of trained athletes and breathing by special gas mixtures) this limit should be increased to 20 l/s [3]. The upper limit of airflow measurements for children till 12 years aged should be limited by the value of 8 l/s, corresponding to their anatomical and physiological data [4].

5. In discussed standard normalized error of ± 3,5% applies to any spirometers, but taking into account the purpose and levels of the devices hierarchy, this value should be much greater range. The nature of diagnostic problems at different levels, the reproducibility and repeatability of the breathing parameters [5-7], achievable accuracy of calibration equipment requires to assign a range of airflow rate measurements errors within ± (1,5-7,5)%.

6. Considerable divergence of opinions on the frequency range, which has normalized dynamic characteristics of spirometers (from 4 Hz to 30 Hz), requires a more precise definition of this parameter. Study of the forced expiratory frequency spectra [8,9] determined the maximum frequency of airflow rates significant harmonic content by the values

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

of 4,0-6,5 Hz. Based on these results and taking into account a limits of measurements errors in forced expiratory process within ± (1,5-7,5)%, the non-linearity of spirometer dynamic characteristics should be normalized in limits ± (1-5)% for frequency range till 20 Hz.

Based on the analyzed requirements for spirometers and considering differentiation of devices in purpose and hierarchy levels, we can offer prospective standards of MTR for all levels of spirometers, having in mind both achieved technological level and the potential for improvement of equipment (Table 2).

TABLE II. Prospective MTR for spirometers

Parameter	Application of spirometer					
	For adults			For children till 12 years		
	Levels of hierarchy					
	1	2	3	1	2	3
Resistance to breathing, Pa·s/l	40,0			20,0		
"Dead space" volume, ml	100,0			50,0		
Response threshold on volume, l	0,05			0,02		
Response threshold on airflow rate, l/s	0,1	0,07	0,05	0,05	0,03	0,02
Upper limit of volume measurements, l	8,0		10,0	5,0		
Upper limit of airflow rate measurements, l/s	12,0	15,0	20,0	8,0		
Limits of volume measurements error, %	± 7,5	± 3,0	± 1,5	± 7,5	± 3,0	± 1,5
Limits of airflow rate measurements error, %	± 7,5	± 3,0	± 1,5	± 7,5	± 3,0	± 1,5
Nonlinearity of dynamic characteristics in frequencies range till 20 Hz , %	± 5,0	± 2,0	± 1,0	± 5,0	± 2,0	± 1,0

III. DISCUSSION

We must clearly understand the technology of these standards implementation in spirometry, firstly - the capabilities of airflow sensors, used in spirometry.

For a primary testing of ventilatory lung function (the level 1) offers a simple device for measuring 1-3 of forced expiratory parameters (peak expiratory flow, forced vital capacity, forced expired volume in first second). Spirometer for level 1 are quite specific instrument, which should be very simple in their design and operation, with minimum weight and sizes, taking into account their use as well as "home medicine" device [10]. The need for such devices is justified by highly informative measured parameters in the preliminary assessment of the function.

For the diagnosis of bronchial obstruction (the level 2) need automated device that measure all standard parameters of forced expiration. Such device should:

- automatically evaluate the correctness of maneuvers and choose the best of them for further processing;
- display the real-time "flow-volume" loop of the forced expiration and calculate its parameters;
- provide printing of the measurements results and calculations.

Spirometers, correspond to hierarchical level 2, contains programmable microprocessor controller with the function of long-term memory, information output to the LCD at least of 128 x 64 pixels format and availability of a wireless communication interface [1,11]. Described configuration can be a base for the design of the whole spirometry equipment system [12], including the devices of level 3.

Obviously, for devices of level 1 the most appropriate is use of turbine sensor, almost perfectly coincide with MTR to spirometers of level 1 for both adults and children. These sensor are a resistance to expired airflow in the range of 25 - 50 Pa·s/l, the "dead space" volume under 50 ml and response threshold on airflow rate about 0,03 l/s, the upper limit of the measuring range up to 12 l/s. An important argument for this choice should be assumed minimal dimensions and weight of the turbine sensor, containing the inverter of the blades turn frequency - a optical pair "photodiode-LED". Also, spirometer represented as turbine sensor equipped with an electronic pulse counter as a computing device and LCD as an imaging unit.

We also see good prospects for adaptation to the MTR for spirometers jet and vortex flow sensors with high dynamical characteristics and the absence of mobile elements [7].

Level 2 of hierarchy focuses mainly on clinical testing and screening of large populations, as well as for preventive surveys of respiratory function, including in the field conditions. In solving these problems by measuring vital capacity and forced expiratory parameters essential factors of medical process are:

- the time spent on examination of the patient;
- portability and autonomy of used spirometers;
- the ability of receiving the diagnostic conclusions for tests and its automated archiving.

In view of these factors for the level 2 can be recommended, except the turbine, sensors with aperture or Pitot tubes. Inapplicable here are sensors with hydraulic resistance, because they are sensitive to clogging by products of breathing and require significant effort and time for cleaning.

Sensors with aperture provide possibility to study the processes of respiration with static characteristics stability for both inspiration and expiration. The relationship between aperture diameter and pressure drop on it makes possible to

get a output signal of high level (and hence high sensitivity), the minimum response threshold within the resistance to breathing no more than 50 Pa·s/l and the "dead space" volume within 24 ml [13]. Elements of sensor with the aperture having simple forms, are not affected by the products of breathing contamination and do not require a corps heating. Full access to the internal surfaces and cavities of sensor determine the usability of their cleaning.

Sensor with Pitot tubes also provide to obtain an enough output signal of pressure when the resistance to breathing not more than 50 Pa·s/l.

On level 3 the area of solved methodological and metrological tasks expanded significantly by study with maximum achievable accuracy and sensitivity of spontaneous breathing pattern, vital capacity structure and maximal voluntary ventilation parameters. For these purposes, the most rational is use of sensors with aperture, Pitot tubes, hot-wire sensors. Since the time required to conduct researches in the laboratory are not critical, the known disadvantages of sensors with hydraulic resistance are minimized, while their individualized linear static characteristics is crucial factor of adequate implementation in completed clinical and scientific researches of respiratory function.

Hot-wire sensor have a resistance to breathing ≤ 40 Pa·s/l, the response threshold within 0,05 l/s [3] and the opportunity to reduce these values at 20 Pa·s/l and at 0,025 l/s [14]. Among the advantages of hot-wire sensors allocated a small "dead space" volume, excellent dynamic characteristics, indifference to vibration and orientation in space and unsurpassed sensitivity [14].

Modern ultrasound sensors [14] have characteristics of resistance to breathing 30 Pa·s/l, the response threshold 0,005 l/s, limit of airflow rate measurements error ± 2% and are also suitable for use in spirometry on level 3.

I. CONCLUSION

1. The main characteristics of modern flow sensors meet the requirements of the actual standards and have certain reserves to meet more stringent future standards.

2. The improvement of spirometers should be directed at decrease the resistance of the breathing and increase the accuracy of measurements, so good opportunity for using in spirometry have hot-wire and ultrasonic flow sensors.

3. For improve the spirometers standards, they must be structured for other purposes (for adults and children), levels of the hierarchy, contain requirements for the "dead space" volume and put more stringent requirements on the devices dynamic characteristics.

4. It should be expected in near future using for spirometry jet and vortex flow sensors.

REFERENCES

[1] M. Miller, J. Hankinson, V. Brusasco, F. Burgos, R. Casaburi, A. Coates, R. Crapo, P. Enright, C. van der Grinten, P. Gustafsson, R. Jensen, D. Johnson, N. MacIntyre, R. McKay, D. Navajas, O. Pedersen, R. Pellegrino, G. Viegi, J. Wanger, "Standardization of spirometry", Eur. Resp. J., vol. 26, pp. 319-338, August 2005.

[2] V. Lopata, "Four flow-spirometer generations in respiratory diagnostics", J. of Physiol. & Pharm., Suppl. 2 "Eight Polish-German Conference on Advances in Experimental and Clinical Pneumology, 9-11 May 2002, Warsaw, Poland", 2002, p. 39.

[3] V. Lopata, M. Khvorov, O. Grudin, G. Frolov, L. Pisarenko, Yu. Sinekop, I. Katsan, B. Lupina, "Choice of measuring channels designs for spirometric telemedical systems", in Tagungsband 12.Workshop "Mikrotechniken und Mikrosensoren fur die Medizin, Biologie und Umwelt", 6-7 Oktober 2000, Iena, p. 27.

[4] H. Eigen, "Pulmonary function testing: a practical guide to its use in pediatric practice", Pediatrics in review , vol. 7, pp. 235-245, February 1986.

[5] B. Cooper, "Spirometry standards and FEV1/FVC repeatability", Prim. Care Resp. J., vol.19, pp. 292-294, September 2010.

[6] J. Hankinson, J. Stocks, R. Peslin, "Reproducibility of lung volume measurements", Eur. Resp. J., vol. 11, pp. 787-790, October 1998.

[7] R. Jensen, J. Teeter, R. England, H. White, E. Pickering, R. Crapo, "Instrument accuracy and reproducibility in measurements of pulmonary function", Chest, vol. 132, pp. 388-395, July 2007.

[8] R. Lemen, C. Gerdess, M. Wegmann, K. Perrin, "Frequency spectra of flow and volume events of forced vital capacity", J. Appl. Physiol., vol. 53, pp. 977-984, October 1982.

[9] R. Peslin, P. Jardin, A. Bohadana, B. Hannhart, "Contenu harmonique de signal de debit pendant l'expiration forcee chez l'homme normal", Bull. Eur. Physiopath. Resp., vol. 18, pp. 491-500, September1982.

[10] M. Miller, "Peak expiratory flow meters", ERS Buyer's uide, vol. 3, pp. 12-14, August 2000.

[11] A. ivachev, Systematic pproach o he omenclature of equipment for functional pulmonary diagnosis", Med. Tekh., pp. 13-14, September-October 994.

[12] V. Lopata, V. Tkachenko, "State of art and development prospects of diagnostic technologies in medicine", in National Technical University of Ukraine "Kyiv Polytechnic Institute", Electronics and nanotechnology. Proceedings of the XXXI International Scientific Conference ELNANO 2011, April 12-14, 2011, Kyiv, Ukraine, p. 7.

[13] V. Lopata, "Choice of efficient dimensions of constricting devices for tachospirometric transducers", Med. Tekh., pp. 14-16, January-February 1991.

[14] V. Lopata, "Optimum choice of flow-rate sensors for spirometry", in IX Polish-German Conference on Advances in Experimental and Clinical Pneumology, Zakopane, 8-10 May 2003, Programme and Abstracts, 2003, p. 29.

[15] A. Torres, F. Burgos, "Infection control and spirometry", ERS Buyer's Guide, vol. 3, p. 39, August 2000.

The Enhancement of the Linear X-Ray Tomography with Digital Tomosynthesis Algorithms

Sergiy Senchurov
Physics Faculty
Taras Shevchenko National University of Kyiv
64/13, Volodymyrska Street, City of Kyiv, Ukraine, 01601
Email: sergsenc@mail.univ.kiev.ua

Olexander Motolyga
Physics Faculty
Taras Shevchenko National University of Kyiv
64/13, Volodymyrska Street, City of Kyiv, Ukraine, 01601
Email: motor.sasha@gmail.com

***Abstract*—The experimental comparison of the performance of the medical X-Ray laminography and digital tomosynthesis algorithms was carried out. The results of comparison show the promising future of the integration of the digital tomosyntesis algorithms into the currently produced medical X-Ray laminography complexes**

***Keywords*—Computed tomography, X-ray applications, Reconstruction algorithms, Biomedical imaging**

I. Introduction

Before the invention of computed tomography (*CT*), the linear tomography (or *laminography*) [1] was the only way of the visualization of a single cross-section. The invention of CT posed a task of comparison of the performance of CT and laminography. A number of researchers [2], [3], [4] reported the higher clical efficiency of CT compared to laminography due to the three-dimensional nature of CT. These advantages of CT caused the common transition from laminography to CT in medical practice [5].

The availability of high-resolution digital X-Ray receivers provided their usage in modern clinical laminography equipment.

On the other hand today there is a number of X-Ray techniques, called *digital tomosynthesis*, mechanical realization of which is similar to the laminography with digital X-Ray receivers. Tomosynthesis shows a better performance than CT and common X-Ray imaging in some applications [6], [7], [8], [9].

Therefore there is a good reason to introduce a digital tomosynthesis algorithms into the already manufactured laminography equipment with digital X-Ray receivers. The estimation of resolution which can be achieved should be done for each model of laminography equipment.

The scheme of X-Ray laminography is shown in Figure 1. The X-Ray tube and X-Ray film or X-Ray receiver move in the opposite directions around the common rotation axis. The cross-section of the patient by the plane containing this axis remains unblurred, while the other layers, lying above and below the axis are motion blurred. So we get a sharp image of the relatively thin layer parallel to the detector plane and containing the rotation axis.

The implementation of tomosynthesis algorithms allows to obtain the tomographic sections in X-Rays rooms from one hand and to achieve manifold patient dose reduction compared to common – axial – computed tomography from the other hand [10].

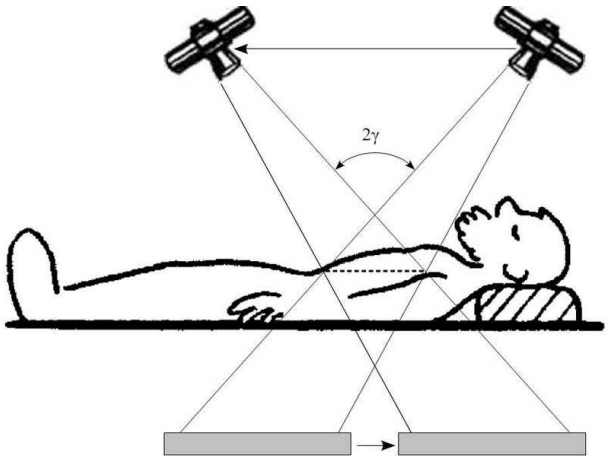


Fig. 1. Schematic representation of the X-Ray laminography

The purpose of our study was to implement a tomosynthesis algorithm based on simultaneous algebraic reconstruction technique (SART) [11], suitable for the implementation in the industrially produced medical X-Ray lamingoraphy equipment and to compare the performance of the X-Ray laminography with the possibilities provided by the digital tomosynthesis. The SART algorithms are among the most efficient in the digital tomosynthesis [12].

II. Experimental Technique

To analyze the performance of X-Ray laminography and digital tomosynthesis the custom phantom shown in Figure 2 was created. This phantom was a number of plexiglass plates with the X-Rays test patterns attached to them.

The plates was separated by plexiglass washers. Some of plates contained a small lead pellets. The full construction was fixed by screws.

The X-Ray laminography equipment produced by KVOMA, Kyiv, with the digital X-Ray receiver produced by Teleoptic-PRA was used for the analysis of the performance of the common X-Ray laminography. The minimal distance between the X-Ray tube and the receiver is 1,00 m. A number

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

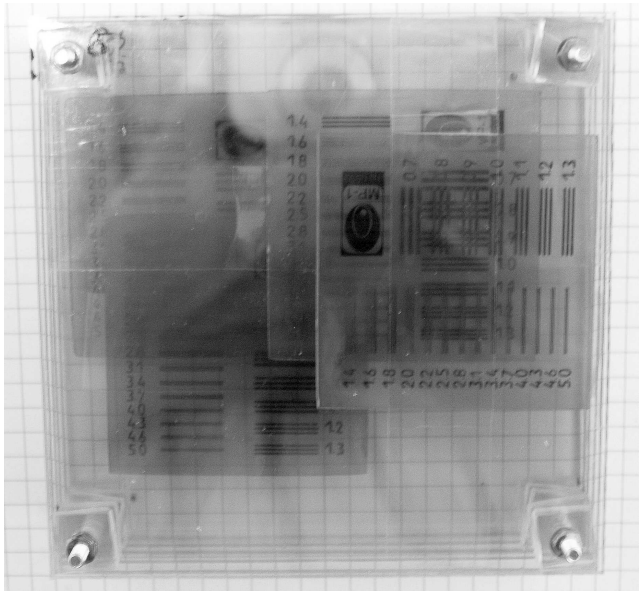


Fig. 2. Phantom used to measure the characteristics of the X-Ray laminography and digital tomosynthesis algorithms

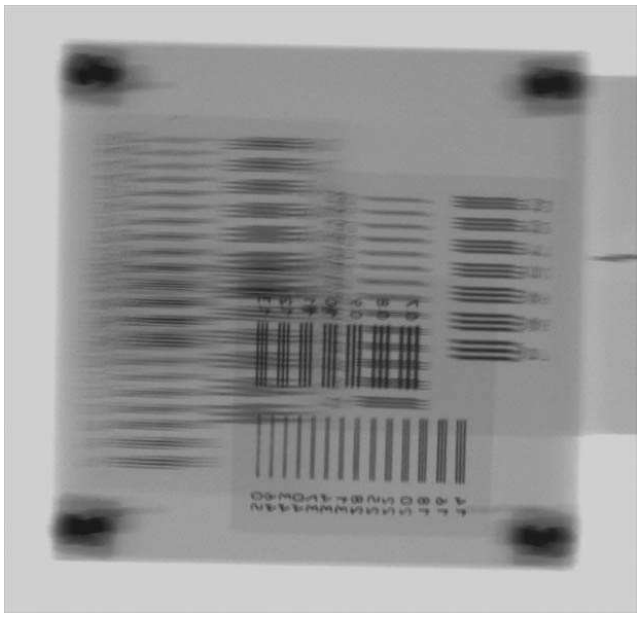


Fig. 3. A part of the X-Ray laminography image taken at the γ value 40° and the minimal speed of the X-Ray tube and receiver

of images was taken every 3 mm with various speeds of movement, kV's, mAs's and 2γ values. The values of the rotation half-angle γ can be set to 8°, 20°, 30° and 40°. The speed of the tube and receiver movement has three fixed values, allowing to make one image in $\approx$ 1,3; 1,8 and 2,3 s.

To analyze the performance of the X-Ray digital tomosyntesis the same experimental arrangement and phantom were used, but the number of still images has been taken at a number of angles covering the 2γ range from -48° to +48° with the intervals 2°, 3° and 5°.

The tomosynthesis algorithm was based on the ideas from [11] and utilized the SART. It was implemented in C++ using Qt, Vtk and openMP open-source libraries. The software was compiled to be run on Microsoft Windows x64 and utilizes the possibility of multiprocessor parallel calculation and usage of large memory arrays up to 16 Gb. This allows to reconstruct a volume containing up to 2500x2500x600 voxels with a potential resolution up to 5 lp/mm in the plane parallel to X-Ray receiver. Number of iterations varied from 3 to 500.

III. Experimental Results

The part of the X-Ray laminography image is shown in Figure 3.

The example of X-Ray digital tomosinthesis is shown in Figure 4. The reconstruction was based on 13 images taken each 5° in the 2γ range from -30° to +30°.

The resolution in the plane of the image can be seen directly from the images. To measure the resolution in the direction perpendicular to the plane of the receiver, we calculated how the contrast of the image reduces with the distance between the axis of rotation and the test objects. The contrast K dependence on the distance d was calculated according to [13] as

$$K(d) = \frac{I_{max} - I_{min}}{I_{backgr}}, \qquad (1)$$

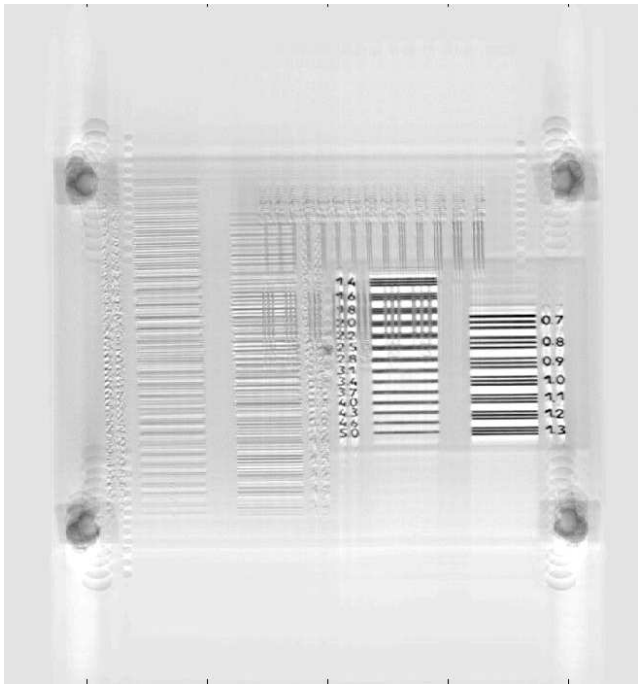


Fig. 4. A tomosynthesis cross-section calculated from 13 images taken each 5°

where I_{max} and I_{min} are the maximum and minimum intensities of the test object image and I_{backgr} is the background intensity. The contrast values were adjusted to $K(0) = 1$. The contrast-distance dependencies are shown in Figure 5 for X-Ray laminography and in Figure 6 for digital tomosynthesis. The dependencies were fitted by Lorentz peaks according to

$$y(d) = y_0 + \frac{2A}{\pi}\frac{\omega}{4d^2 + \omega^2}, \qquad (2)$$

where y_0, A and ω are fitting parameters. Fitting results are

shown with lines.

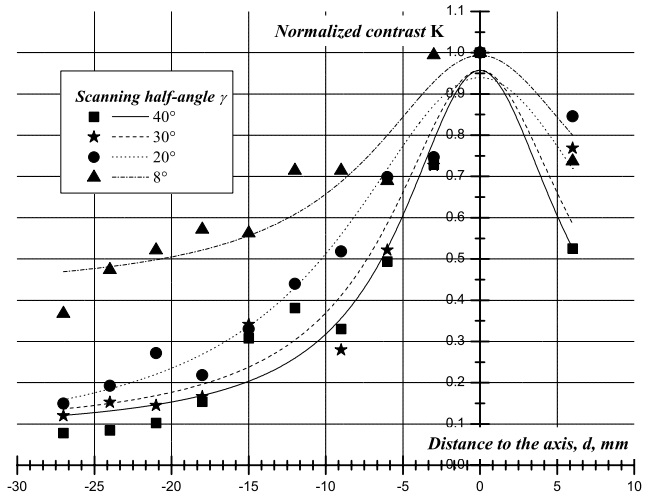


Fig. 5. The result of fitting of the X-Ray laminography relative contrast decrease with the distance from the scanning plane

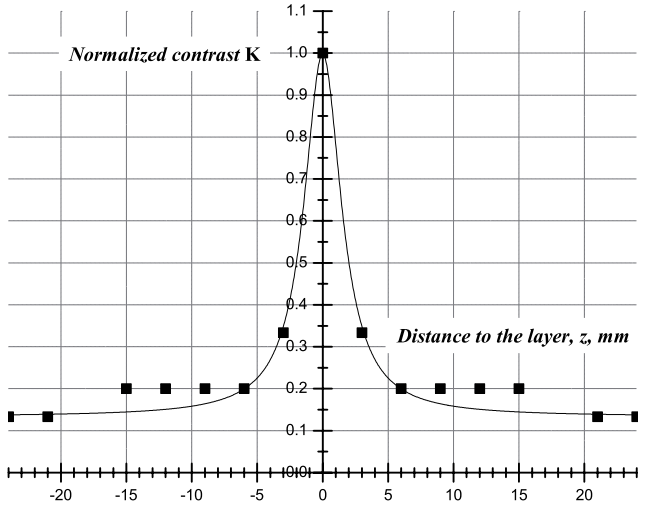


Fig. 6. The result of fitting of the digital tomosynthesis relative contrast decrease with the distance from the cross-section plane

IV. Discussion

As one can see from Figure 5, the contrast decreases more rapidly for wider 2γ angle values and is nearly the same for $\gamma = 30°$ and for $\gamma = 40°$. We can estimate the thickness of the layer by the estimation of the distance where the contrast reduces to the half of the initial value.

One can conclude from Figures 5 and 6 that the thickness of the cross-section for X-Ray laminography is about 6 mm while the thickness of X-Ray digital thomosynthesis cross-section is about 3 mm. We can see the more rapid decrease of the contrast in the digital thomosynthesis compared to laminography. On the another hand we have to improve the implemented algorithm for better artifacts reduction.

The spatial resolution in the plane parallel to X-Ray receiver for laminography and digital tomosynthesis can potentially achieve up to 5 lp/mm which is far beyond the resolution of existing X-Ray computed tomography equipment. The resolution in transverse direction perpendicular to the plane of the X-Ray tube and receiver movement, is much lower that the resolution of the X-Ray computer tomography. The strong advantage of the X-Ray laminogrphy is low patient dose, which is equivalent to the dose of a single X-Ray still image.

V. Conclusions

The utilization of digital tomosynthesis algorithms in the X-Ray laminography equipment is very promising both from the point of view of simplicity and and the availability of hardware implementation and from the poin of view of the possibility of patient's X-Ray dose reduction.

References

[1] T. M. Buzug, *Computed Tomography*. Berlin Heidelberg: Springer-Verlag, 2008.

[2] E. Schaner, A. Chang, J. Doppman, D. Conkle, M. Flye, and S. Rosenberg, "Comparison of computed and conventional whole lung tomography in detecting pulmonary nodules: a prospective radiologic-pathologic study," *American Journal of Roentgenology*, vol. 131, no. 1, pp. 51–54, 1978.

[3] A. E. Chang, E. G. Schaner, D. M. Conkle, M. W. Flye, J. L. Doppman, and S. A. Rosenberg, "Evaluation of computed tomography in the detection of pulmonary metastases. a prospective study," *Cancer*, vol. 43, no. 3, pp. 913–916, 1979.

[4] S. Siegelman, E. Zerhouni, F. Leo, N. Khouri, and F. Stitik, "Ct of the solitary pulmonary nodule," *American Journal of Roentgenology*, vol. 135, no. 1, pp. 1–13, 1980.

[5] A. M. Parsons, F. C. Detterbeck, and L. A. Parker, "Accuracy of helical ct in the detection of pulmonary metastases: Is intraoperative palpation still necessary?" *Ann Thorac Surg*, vol. 78, no. 6, pp. 1910–1918, 2004.

[6] A. Smith, "Full-field breast tomosynthesis," *Radiology Management*, vol. 27, pp. 25–31, 2005.

[7] S. P. Poplack, T. D. Tosteson, C. A. Kogel, and H. M. Nagy, "Digital breast tomosynthesis: Initial experience in 98 women with abnormal digital screening mammography," *American Journal of Roentgenology*, vol. 189, no. 3, pp. 616–623, 2007.

[8] D. Gur, G. S. Abrams, D. M. Chough, M. A. Ganott, C. M. Hakim, R. L. Perrin, G. Y. Rathfon, J. H. Sumkin, M. L. Zuley, and A. I. Bandos, "Digital breast tomosynthesis: Observer performance study," *American Journal of Roentgenology*, vol. 193, no. 2, pp. 586–591, 2009.

[9] I. T. P. Wells, V. M. Raju, B. K. Rowberry, S. Johns, S. J. Freeman, and I. P. Wells, "Digital tomosynthesisa new lease of life for the intravenous urogram?" *British Journal of Radiology*, vol. 84, no. 1001, pp. 464–468, 2011.

[10] European Commission Directorate-General for the Environment. (2000) Referral guidelines for imaging.

[11] K. Mueller, "Fast and accurate three-dimensional reconstruction from cone-beam projection data using algebraic metods," Ph.D. dissertation, The Ohio State University, 1998.

[12] Y. Zhang, H.-P. Chan, B. Sahiner, J. Wei, M. Goodsitt, L. Hadjiiski, J. Ge, and C. Zhou, "A comparative study of limited-angle cone-beam reconstruction methods for breast tomosynthesis," *Medical Physics*, vol. 33(10), pp. 3781–3795, Oct. 2006.

[13] E. Krestel, *Imaging Systems for Medical Diagnostics*. Berlin: Siemens, 1990.

Geometrical measuring at 3D stereoscopic radiography

Miroshnychenko S.I., Nevhasymyy A.A.
"Teleoptic" PRA" Ltd
Kiev, Ukraine

Volkov E.V.
National Aviation University
Kiev, Ukraine
Eugene.Volkov@teleoptic-pra.com.ua

***Abstract*—Stereoscopy is one of the ways for creating 3D images in radiography. Such technique can be realized on widespread X-ray equipment. Geometrical object measurement on stereo images is realized with the help of synthesized stereo cursor, which is displayed superimposed with X-ray stereo image.**

Keywords—3D, X-ray, stereo effect, digital receptor, anaglyph, geometrical measuring

I. INTRODUCTION

Radiography is one of the most significant diagnostic methods in medical practice. With the help of new technologies it continues to develop. Nowadays the most important directions are dose reduction and 3D technologies.

There are few ways to provide 3D images in radiography. One of them is well-known computed tomography. For 3D object reconstruction a large series of cross-section images is used. They are taken during rotation of X-ray tube and detectors around a patient.

Another way to form 3D X-ray images is to use stereoscopic effect. The important advantage of stereoscopic method that for providing 3D X-ray image it needs fewer doses for patient than during computed tomography. Because only two images of a patient are acquired at least and for every of such initial images only 1/200 of CT radiation dose is used [1]. Another preference of stereoscopic method that it's creates the illusion of depth. But for medical purposes not only perception of 3D depth is useful, geometrical measuring is also needed. In the article the method for measuring geometrical parameters of objects in stereo images is offered.

II. PROVIDING OF STEREOSCOPIC EFFECT

For stereoscopic effect two images of an object are taking with some parallax and every of them are shown separately to the proper eye. So they combined in the brain to give the perception of 3D depth [2]. The most common methods of separate showing are special eyeglasses (Fig. 1).

One of them used light polarization effect. In such system two opposite polarizing filters in eyeglasses are used. For personal use polarized filters are mounted in screen [3]. So as each filter of the eyeglasses only passes light which is similarly polarized and stops the opposite polarized light, for each eye the proper image is shown.

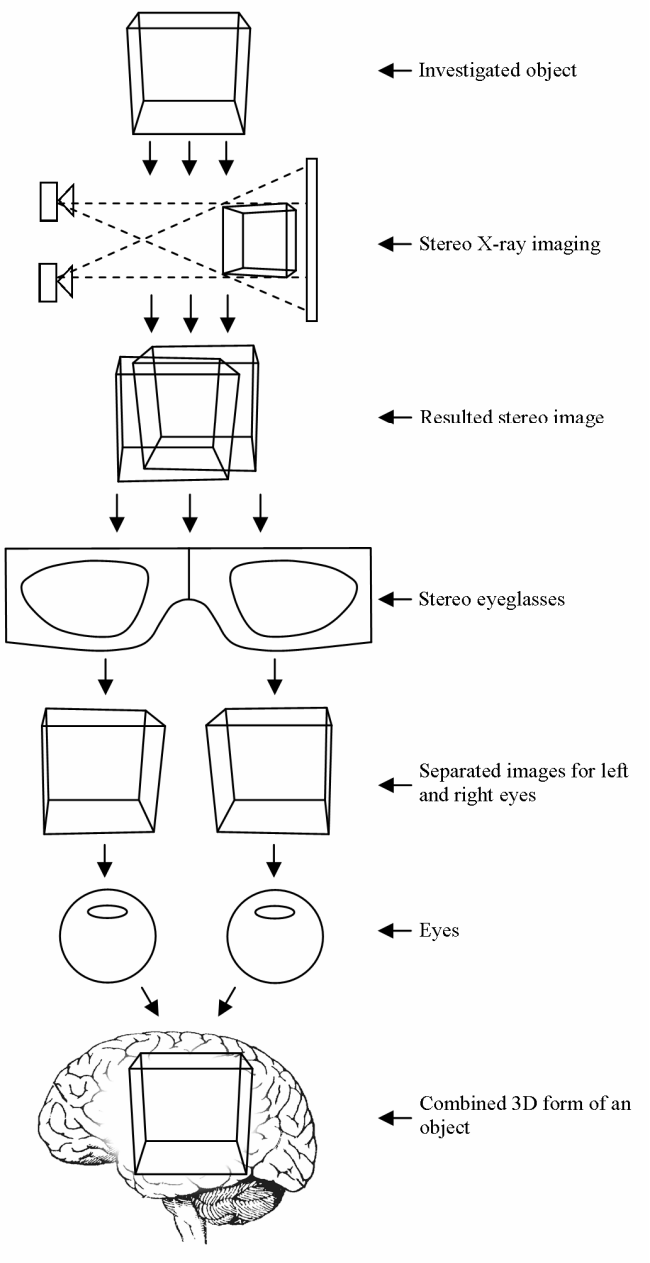


Fig. 1. Process of stereo X-ray image reception

Another system using color coding and doesn't need special screens. This system is known as anaglyph [4]. Two offset images are shown together at the screen. One of them is shown in red channel of RGB color system. Another in green and blue channels also named cyan channel. According to this scheme one of the filters of eyeglasses is red and another is cyan. In such way every from two images is shown to the proper eye, and the 3D effect is achieved.

The basic lack of anaglyph method is incomplete color rendering. But there is no such problem for grayscale images, which are roentgenograms. So it's the easiest way to provide 3D image in radiography taking into account that anaglyph eyeglasses are low-cost and any computer or TV display is suitable.

Another method is shutting system where synchronization between screen and eyeglasses is used. In such eyeglasses liquid crystal layer are mounted [5]. This layer is opaque when voltage is applied and otherwise transparent. At the same moment of time one of the eye's glasses is open and another is shut.

At the same time 3D X-ray imaging can be static (stereo radiography) or dynamic (stereo radioscopy). Only the difference in the number of acquired images. For stereo radioscopy more then two offset images are needed. The dynamic X-ray film is formed by combining nearby images into separated frame of such film.

III. STEREO IMAGING GEOMETRY

That means that object measuring can be provided both on static or dynamic 3D X-ray. Also geometrical measuring of the object during X-ray stereo imaging can be provided for all stereo systems. Because the all have the same principle – two offset images are shown separately to the left and right eye of the viewer. For stereo imaging, how it was mentioned above, minimum two images of an object is needed. For getting such offset X-ray images different positions or movement of X-ray unit and detector can be used.

One of the schemes is rotating system. It looks like con beam CT with flat panel detector [6]. But for stereo 3D more wide spread and simple digital X-ray equipment can be used. For example such experimental assembly on the basic X-ray unit and digital X-ray receptor produced by "Teleoptic" Pra" Ltd was set (Fig. 2). In such system X-ray tube and detector can rotate around a patient and needed number of images for static or dynamic X-ray can be collected [7]. Rotating scheme can be also realized with the help of C-arm X-ray system that is widespread and provide great flexibility in the examination process.

Particular case can be a system where digital receptor fastened stationary and X-ray source moves on an arc trajectory in limited angle. Similar assemblies are used in tomosynthesis. Another way is to use linear tomography that can be also upgraded with 3D X-ray imaging. Disadvantage of rotating scheme or linear tomography system is inaccuracy of geometrical measurements that are caused by backlash errors of rotating joints.

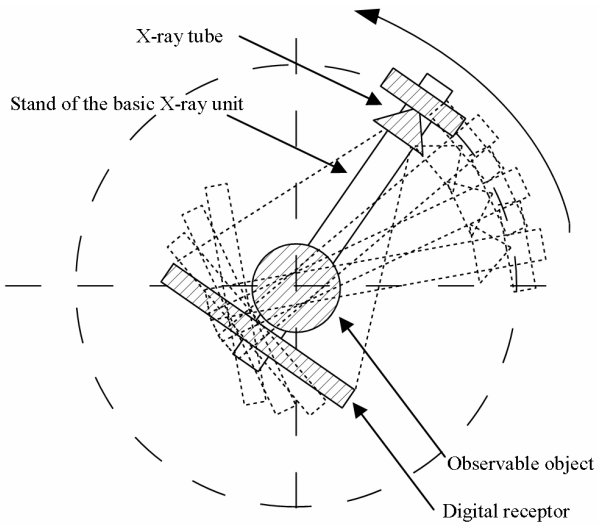


Fig. 2. Expiremental assembly with basic X-ray unit

The most simplest scheme for stereo X-ray imaging with no rotation joints and that can provide short time of patient's checkup consist of one digital receptor and two X-ray sources. Both sources and receptor are mounted stationary. Patient or object is placed to the digital receptor. Every offset image is taking with one of the X-ray sources.

The distance between two X-ray sources is named base line of stereo image (Fig. 3). Also distances AA' and BB' are known, because it's the distance between X-ray tube and digital receptor. Line m lies in plane of intensifier screen of digital receptors. Let's take some point C of observable object. Geometrical measuring of an object is provided if coordinates of such few points can be found. Point A" and B" are projections of point C at digital receptor from X-ray sources, that are placed in points A and B. That means that distances A'A" and B'B" are known. Because coordinates of points A" and B" can be found from received offset images. Position of point C can be found. Will build line A"'B. It goes through point B and parallel to the line AA". Two equal triangles ΔAA'A" and ΔBB'A"' are obtained. Fro this two equal triangles

$$A'A'' = B'A''' \tag{1}$$

We will construct line n through point C, which will be parallel to the line m. That gave two similar triangles ΔBB"A"' and ΔBCC"'. From this two similar triangles

$$BB' / BC'' = A'''B' / CC''' \tag{2}$$

$$BC'' = BB' - C''B' \tag{3}$$

Because ABC"'C is a parallelogram it will be correct statement

$$CC''' = AB \tag{4}$$

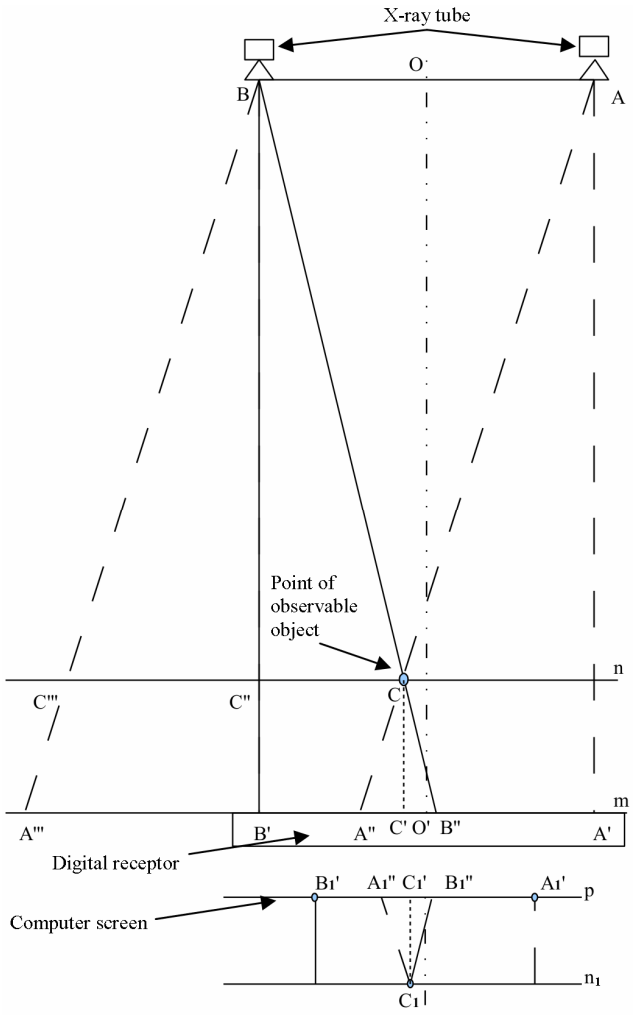


Fig. 3. Geometrical measuring based on X-ray stereo imaging

Being based on (1) it will be correct

$$A'''B'' = A'A'' + B'B'' \quad (5)$$

Paste (3), (4) and (5) in (2) will receive

$$BB' / (BB' - C''B') = (A'A'' + B'B'') / AB \quad (6)$$

Because of

$$CC' = BC'' \quad (7)$$

from (6) is

$$CC' = (\mathrm{AB} \cdot BB' / (A'A'' + B'B'')) - B'C \quad (8)$$

We will consider forming of point C_1 of obtained stereo image between viewer and computer screen that correspond to the point C. A certain scale between image on computer screen and obtained image on intensifier screen of digital receptor exist. Will name this scale – k. Then

$$A'B' = k \cdot A_1'B_1' \quad (9)$$

$$A'A'' = k \cdot A_1' A_1'' \quad (10)$$

$$B'B'' = k \cdot B_1' B_1'' \quad (11)$$

So if over displayed X-ray stereo image the image of stereo cursor will be formed, coordinates of such cursor will correspond to similar point of object and placed between digital receiver and X-ray sources. As a result distance between looked point and digital receptor is calculated with

$$CC' = BB' - (\mathrm{AB} \cdot BB' / k \cdot (A_1' A_1'' + B_1' B_1'')) \quad (12)$$

Synthesized image of cursor creates with the same technology, which used for X-ray stereo image. In such system with fixed digital receptor and X-ray sources inaccuracy of geometrical object measurements will defined by spatial resolution of digital receiver and inaccuracy of distance measurement between X-ray sources and receiver.

IV. CONCLUSION

Stereo X-ray imaging can be realized with the help of widespread equipment, such as basic X-ray unit, C-arm systems, tomosynthesis units or linear tomography.

Geometrical measuring of objects on stereo X-ray images is realized by synthesized image of cursor. Position of this 3D cursor will correspond to point that lay in volume in front of digital receptor, where patient is based. Such cursor must be displayed simultaneously with X-ray stereo image. The next stages of the research must be experiments to determine the accuracy of measurement and then clinical tests, which will allow estimating the diagnostic value of such object measuring at finding out pathologies during radiography.

REFERENCES

[1] C.I. Lee, J.G. Elmore, "Radiation-related risks of image studies" http://uptodate.com/contents/radiation-related-risks-of-imaging-studies

[2] Valus N.A., Stereoscopy. – M.: AS USSR, 1962.

[3] Usencov D.U, "On the wave of polarization" // 3D world. Number 33 2012. p. 9-15

[4] Usencov D.U, "Anaglyphic technology" // 3D world. Number 2, 2012. p. 4-22

[5] Usencov D.U, "Shutter technique of displaying stereoscopic images" // 3D world. Number 4, 2012. p. 4-12

[6] Baba R., Ueba K., Using a flat-panel detector in high resolution cone beam CT // Dentomaxillofacial radiology. Number 33, 2004. p. 285-290.

[7] Miroshnychenko S.I., Nevhasymyy A.A., Volkov E.V. 3D image generation in digital X-ray detectors. // I International Workshop "Medical physics – the current status, problems, the way of development. Innovation technologies": workshop proceedings. – K. : Modern printed technologies "Bavok", 2011. – p. 132-134.

The Possibilities of the Contrast Improvement in Fluorography Image with Digital Processing

Aslamova L., Melenevska N., Miroshnichenko N.,Grabovska E.
Training-research center of radiation safety
Taras Shevchenko National University of Kyiv
Kyiv, Ukraine
radbez@gmail.com

Miroshnichenko S., Radko D.A.
Production-Research Association "Teleoptic"
Kyiv, Ukraine

***Abstract*—The annual fluorography chest screening profoundly contributes to the total radiation dose delivered to the patient. The application of additional image contrast improving methods (high tube potential technique, anti-scatter grid method) increases the radiation dose to several times. X-ray image digitalprocessing with special softwareis proposed as alternative method for visual contrast improvement to minimize the patient dose.**

Keywords—reducing radiation dose, chest screening, chest examination, digital image processing, Bucky factor, X-ray image, image quality, entrance surface dose, water chest phantom, contrast, anti-scatter grid, software image processing.

I. INTRODUCTION

In Ukraine the most common type of patient X-ray examination is a fluorography chest screening. The calculated amount of the tuberculosis and lungs cancer diagnostics procedures in Ukrainian clinical facilities is 20 million per year [1]. It significantly contributes to the total dose annually deliveredto the patient.

The usual set of X-ray image contrast improvement methods during the pathology examination in routine clinic practice increases the patient dose up to several times. Among them the high tube potentialtechnique based on setting of high tube potential and low tube current values and the low tube potential technique based on vice versa principle [2]. The high tube potential leads to generation of high energy photons. This result in the increase of radiographic image noise caused by scatteredphotons. Usually in clinic practice the application of anti-scatter grid allows to avoid such adverse effects. At the same time the grid use requires the increase of photons quantity via tube current changes, leading to dose rise. The higher grid ratio is used the higher mAs value should be set. As it was measured with thermoluminescent dosimeter by Fung [3], it leads to X-ray exposure of not only examined area, but also adjacent organs and tissues.

The low tube potential technique allows receivingof the images with good radiographic contrast, but the disadvantage of this method is the risk of high dose delivered to the patient due to weak penetration ability of the low energy photons as a resultdeposited in soft tissues.

Another way to reduce the patient exposure dose during the organs examination simulated on phantoms and computed with Monte-Carlo method proposed Vassileva [4,5], to find optimal technical parameters due to application of high potential technique, 1mm Al and 0,1 mm Cu filters and 25 cm air gap with a proper focus-to-film distance. Theentrance surface dose achieved due to these parameters setting is four times less compared to clinical routine methods with anti-scatter grid. Also Mizuta [6] searched for optimal selection standard for anti-scatter grids application due to the proper X-ray unit technical parameters, demonstrating the high potential technique performed using grids with different ratio.

In the present experiments we studied the possibilities of digital X-ray image processing improving thevisual contrast compared to the effects of anti-scatter grid application to find the optimal way to reduce thepatient exposure dose, usingthe water phantom, simulating the chest organs fluorography examinationwith digital X-ray unit.

II. MATERIALS AND METHODS

The study was performed with self-protected digital X-ray unit at the base of LTD "Teleoptic PRA" laboratory, Kyiv, Ukraine. The aim of the study was the examination of the possibility to reduce the X-ray exposure dose with a special software image processing (ContextVision CVIE – Teleoptic – XR2 - ADI). The software ContextVision CVIE based on the principle of non-linear signal filtering is certified in Europe and widely used in clinical facilities of the EU and Ukraine [7].

In the experiments the X-ray screening system was simulated. It was equipped with X-ray dosimeter (behind the anti-scatter grid, on the beam entrance side of the digital receiver assembly). The mode of tube operation was regulated due to the water phantom presence and thickness in a way to keep the constant brightness of X-ray image on the beam entrance side of the digital receiver (Fig 1.) The X-ray generator by Listem Corporation was used. The photon beam was generated with X-ray tube by Toshiba. The measuring was performed at constant tube potential 81 kV which is common for fluorography chest screening in clinic practice. The focus-to-tube distance was 125 cm. The anti-scatter grid SOYEE 11104 with 10:1 ratio, 51 lines/cm frequency and 120 cm focus distance was used during the simulation of clinic routine

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

examination. All images were obtained with digital receiver (Iona – R4000) produced by Ltd "Teleoptic PRA".

The patient chest was simulated with water phantom of 3-12 cm water thickness. Each set of measurement was performed with phantom thickness by 3 cm thicker than previous. According to recommendations of regulatory agencies, the chest anthropomorphic phantoms are mainly used in clinical practice to adequately assess the radiation dose during chest fluorography examination. The phantom structure is similar to the phantom developed by the Center for Devices and Radiological Health (CDRH phantoms) [8].These phantoms are boxes up to 25.4 cm thick in total, including 7.6 cm thick layer of polymethyl methacrylate (PMMA), 4.6 mm thick aluminum plate (on the X-ray beam input side), and 7.7 cm air gap between the phantom output side and the image detector input. The chest phantom (with air gap) corresponds to 10.5 cm thick layer of water or 10 cm acrylic layer, and it is considered an equivalent of the average size patient by the X-ray density.

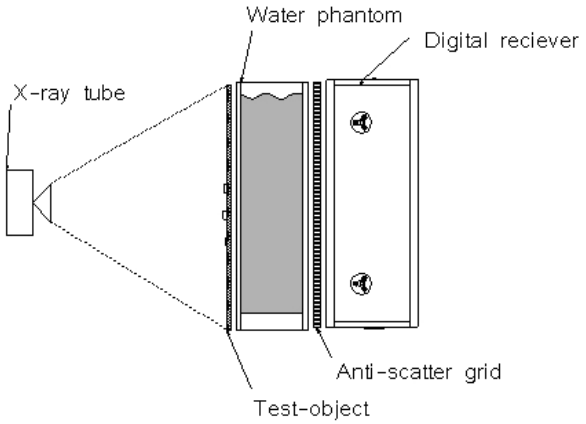


Fig. 1. Experimental installation scheme.

The phantom overlapped the entrance plane of X-ray digital image receptor assembly (Fig.1). Aluminum discs were installed in front of the phantom (from the X-ray tube side). The surface entrance dose was measured with dosimeter Radcal Corporation, 2026 Radiation Monitor, ionizing chamber (20X-60) positioned in front of the phantom. The exposure-meter operation in diagnostics X-ray unit was simulated with brightness stabilization of the received digital images via tube potential and time exposure selection. In the experiments the image brightness was stabilized at approximately 2000 units of a gray scale.

The contrast of the test-object – aluminum disk with 0.5 mm thickness and 15 mm diameter – was set as the effectiveness criterion of the image conversion.

The contrast, K, calculation:

$$K=(B_o - B_b)/(B_b - B_a) \quad (1)$$

B_o– an averaged brightness of the object in image, B_b–an averaged background brightness, B_a– an averaged background brightness value on image received without operating X-ray tube.

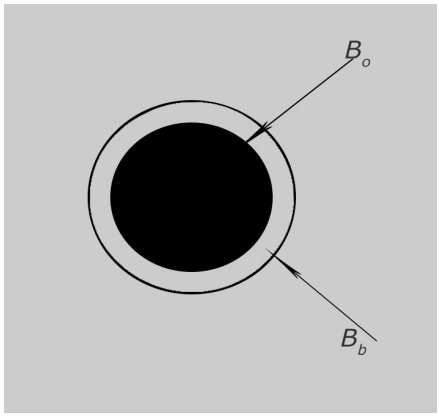


Fig. 2. The brightness measurment: the light ring – the background brightness measurment, the dark circle – the test-object area brightness.

In the experiments $B_a$was equal to 1000 units. The background brightness on image was detected via background brightness value over all pixels within the ring around the 0,5 mm test-object image with the next averaging of the received data (Fig.2). The test-object brightness was detected via brightness value over all pixels in the object field with the next averaging of the data.

III. Results and Discussion

All experiments were performed at constant tube potential 81 kV (Table I). The first measurement performed with empty phantom with 20 mA during 50 ms exposure time detected 5.5% test-object contrast value. After the digital processing of the received radiographic image with software the contrast value increased to 13.9%, which is 2.5 times higher compared to original image.

As a next step, the X-ray image contrast of the test-object on phantom with 3 cm water layer thickness was examined under 63 mA with exposure time 50 ms. The image contrast was 2.9%, which is less compared to 5% visual contrast threshold. The digital processing increased the contrast value to 8.8%. The application of anti-scatter device increased the entrance surface dose in Bucky factor to 36.0 mR and improved the contrast value of the test-object image to 4.2%. After digital processing of this image the contrast increased up to 10.5%.

The contrast value for test-object image received using 6 cm thickness water phantom was 3.2% (less than 5% threshold). The image digital processing improved this value up to 8.9%. The anti-scatter grid application improved the contrast value up to 4.8%, but it also raised the exposure dose

$$\alpha + \beta = \chi. \quad (1) \quad (1)$$

TABLE I. TEST-OBJECT IMAGE CONTRAST MEASUREMENT (CONSTANT TUBE POTENTIAL 81 KV)

No	*I*, mA	*t*, ms	*d*, cm	*ESD*, mR	*R*, p.l./mm	*K*, %	K_{dp}, %	Conditions	*Bucky factor*
1.	20	50	0	9.17	4.0	5.5	13.9	Water	
2.	20	50	0	22.93	4.0	5.9	20.5	Water and grid	2.5
3.	63	50	3	14.19	3.7	2.9	8.8	Water	
4.	160	50	3	36.0	3.7	4.2	10.5	Water and grid	2.5
5.	40	50	6	18.3	3.4	3.2	8.9	Water	
6.	125	50	6	57.2	3.4	4.8	10.5	Water and grid	3.1
7.	320	80	9	139	3.1	2.2	5.0	Water	
8.	320	250	9	432	3.1	3.6	8.6	Water and grid	3.1
9.	160	100	12	84.3	3.1	1.5	4.1	Water	
10.	320	160	12	270	3.1	2.6	7.5	Water and grid	3.2

[a] *I- tube current, t – exposure time, d- phantom thickness , ESD – entrance surface dose, R – resolution , K - test-object image contrast value, K_{dp} –test-object image contrast value after digital processing*

18.3mR in Bucky factor 3.1 up to 57.2 mR. The additional software processing increased the visual contrast of the test-object image to 10.5%.

The phantom with 9 cm water thickness allowed receiving the test-object X-ray image contrast 2.2%. The anti-scatter grid increased the contrast value up to 3.6%, but it was still less than comparedto visual contrast threshold. It also raised the exposure dose in Bucky factor 3.1 up to 432 mR. The digital processing of the original test-object image improved the contrast value up to 5.0%, which was more effectively than anti-scatter device. The additional processing of the test-object image obtained with a grid increased the contrast value up to 8.6%.

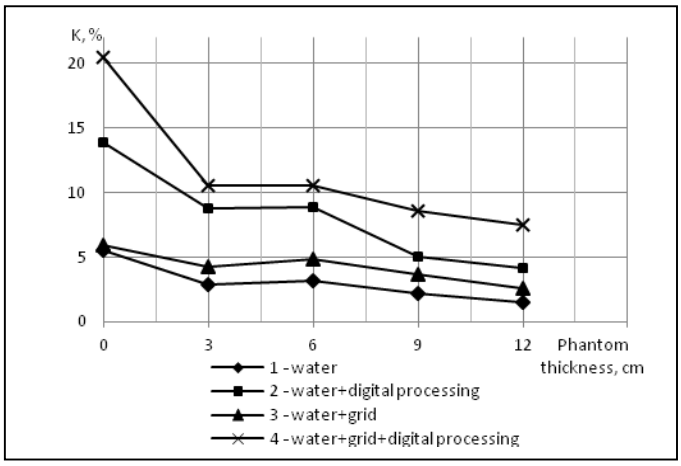


Fig. 3. Graphic representation of experimental study results, K – the test-object image contrast.

The contrast value gradually decreased along with phantom thickness up to 12 cm (Fig.3). The original test-object X-ray image contrast value was estimated 1.5%. The digital processing increased this value up to 4.1%. Compared to processed image, the X-ray test-object image contrast value obtained with anti-scatter grid application was only 2.6%. The additional image processing with software improved this value up to 7.5%.The most actual problem in clinical radiology is achieving of the reasonable balance between sufficient imagecontrast, brightness, accuracy to detect the pathology and maximal dose reduction during the examination procedure. For the patients with high body mass index the problem ofhigh X-ray dose delivering is especially urgent. The experimental results demonstrate the visual contrast reduce due to the rise of water phantom thickness. The anti-scatter device improved the contrast value, but increased significantly the exposure dose in Bucky factor (several times).

IV. CONCLUSION

The experimental data demonstrate the test-object X-ray image contrast value reduced from 5.5% to 1.5% due to the water phantom thickness rise from 3 cm to 12 cm. The application of grid removing scatter improved the contrast value, but the threshold of visible image contrast 5% was not achieved. Also it led to increase of exposure dose.

The image digital processing improves the radiographic image contrast value up to 2-3 times. This allows to obtain the better results compared to application of anti-scatter grid without additional increase of entrance surface dose.

REFERENCES

[1] The indicators of radiology service function in Ukraine in 2008-2009 (reference book), 2010, 80 p.(in russian)

[2] J.T.Bushberg, J.A. Siebert, A.M.J.R. Leidholdt, J.M.Boone, The Essential Physics of Medical Imaging. 2nd ed. Philadelphia: Lippincott Williams& Wilkins, 2002.

[3] K.K.L. Fung, W.B. Gilboy, The effect of beam tube potential variation on gonad dose to patients during chest investigated using high sensitivity LiF:Mg,Cu,P thermoluminescent dosimeters, The British Journal of Radiology, 2001, No. 74, pp.358-367.

[4] J. Vassileva, A phantom approach to find the optimal technical parameters for plain chest radiography, The British Journal of Radiology, Vol.77, 2004, P.648–653.

[5] J. Vassileva, A phantom for dose-image quality optimization in chest radiography,The British Journal of Radiology,Vol.75, 2002, pp.837–42.

[6] M. Mizuta, S. Sanada., H. Akazawa, Comparison of grids for digital imaging with use of a direct-conversion flat-panel detector, Radiology Physics Technology, No.5, 2012, pp.46-52.

[7] www.contextvision.com

[8] AAMP Report No.31, Standardized methods for measuring diagnostics X-ray exposures, Report of task group 8 diagnostic X-ray imaging committee, 1990.

Efficiency and Pattern Noise Supression in Sensor Array X-Ray Digital Receptors

Miroshnychenko S.I., Nevgasimy A.A
PRA "Teleoptic"
Kyiv, Ukraine

***Abstract* - The use of many partial picture forming sensors (sensors array - SA), gives the optimal parameters of X-ray radiography, mammography and fluoroscopic receptors. At the same time, the pattern noise of multi-channel interference occurs. The effective noise suppression is possible not always. Analysis efficiency of sensor arrays, justification options for their implementation, and noise reduction is the subject of this article.**

Keywords - digital X-ray detectors, CCD and CMOS receptors, flat panel detectors, sensor array receptors

I. INTRODUCTION

By the end of the XX century three lines of design of digital X-ray image detectors were formed. Chronologically, the first were the receptors on large CCD. That was digital system for general radiography company Swissray (Switzerland) [1] and system for the chest examination firm Imix (Finland). The receptors of this type have a large size, weight, and low efficiency of optical coupling (Table 1, road 1). For the first generation of CCD receptors resolution was not high - 2.5 ... 3.0 line pairs per millimeter (lp/mm).

TABLE I. TYPE OF X-RAY DETECTORS

Receptor / company	Registra-tion method	Field of view, mm x mm	Resolution lp/mm	DQE (0)	Proc. time
Add-On Bucky (ñSwissrayò)	CsI + 4 x CCD sensors	350 x 430	3,0	0,3*	10
Pixium 4600 (ñTrixellò)	CsI + a-Si readout	426 x 432	3,5	0,7	4
DirectRay (ñDirect Radiologyò)	aSe	350 x 430	3,6	0,35*	12
Xplorer 1700 (ñIDCò)	CsI + 1 x CCD sensor	430 x 430	4,6	0,45	15
Iona-R-4343 (ñTeleoptic)	CsI + 48 x HAD sensors	430 430	4,6	0,4	10
Rad Eye (ñRadiconò)	CsI + 8 tiled CMOS sensors	200 x 300	5,0	0,45*	1

* - calculated using formula (3).

Compare limitations of the first generation of receptors on the CCD, the design of detectors in the form of flat panels (Flat Panel Detectors - FPD), based on amorphous silicon and CsI or amorphous selenium screens – seems as an ideal solutions. This was a second direction in building digital X-ray detectors.

Development of flat panels (Table 1, road 2 and 3) required investment of $ 100 million for each projects. The flat panel resolution increases up to - 3.5 ... 3.6 lp/mm and detective quantum efficiency was high - DQE = 0,4 ... 0,7 at low frequencies. Unfortunately, both projects had serious internal contradiction - inconsistency features of integrated technology (making hundreds of thousands units per year) and a relatively small volume of orders. This has led to the high cost of the panels. In addition, significant internal noise, the presence of defective pixels and degradability of amorphous materials, made difficulties in panels operation. Despite these limitations, digital receptors in a form of flat panels, starting from 2005, are the leaders in the world market.

In 1997, at the exhibition RSNA (Chicago, USA) at the booth of the company Caresbuilt (USA) was demonstrated X-ray receptor prototype, having field of view 43x43 cm and twice compared to flat panel better resolution - 7lp/mm. The total digital matrix of the receptor was 6Kx6K pixels and total area of the crystal silicon was 60x55 mm. Record receptor settings was due to the 99-matrix SA design, developed in the company Teleoptic (Ukraine). The receptor used auto calibration for sensors field of view overlapping areas.

In this technology the assembled image has all the pixels of the original X-ray image, and the boundaries of the partial images were not visually detectable. In February 2002, FDA has authorized the use of multi sensor detectors in the U.S. market. [1] This meant an official recognition of the third line in design of digital X-ray detectors - the use of SA structures with autocalibration.

Example of the SA design of mammography receptor without sensors field of view overlapping is known [2]. The receptor consisted of a tightly fitted mechanically fiber optical tapers (FOT). The big end of FOT was connected to the intensifying screen, while for the smaller end was glued CCD. The main drawbacks of this design were losses of the original X-ray image on the borders of the FOTs and the lack auto calibration. As a result of sensors temperature difference, transmission rates changed. At the picture appears a "chessboard" noise. The failure of such SA design has created false opinion of the sewing algorithms inefficiency in multi sensor image structures.

From 1998 to 2004, multi sensor technology of digital X-ray detectors with auto calibration was developed by company Teleoptic. In 2004, sales of SA receptors of this company amounted to about 5% of the world production of all types of digital receptors.

In 2000...2008 the development of large CCD receptors continued. Canadian firm IDC Ltd., on the base of large Kodak CCD, began production the receptor Xplorer 1700. Its resolution was 4.6 lp/mm (Table 1, road 4), optical path length of about 1 m and weight - about 75 kg. In 2007, help to the good image quality, the receptor had the highest rates of growth in sales. Similar receptors are currently produces by South Korean and Chinese companies.

Starting in 2005, there were new developments of SA radiographic receptors with auto calibration: the 192 sensors panel Naomi, developed by company RF System lab. (Japan)[2], and Korean - 9 sensors receptor Tradix-100p firm Star V-Ray co., Ltd. In 2010 company Teleoptic started selling SA receptor Iona-P-4343 (Table 1, road 5) with the same resolution - 4.6 lp/mm as receptor Xplorer 1700, but thinner - 22 cm and less in weight - 23 kg.

In 2005 - 2008 a principally new generation of SA detectors, based on large area three-side-tileable CMOS sensors and fiber optics where developed [3]. These design has the best resolution 5,0 lp/mm, high DQE, but no sensors field of view overlapping. On the borders of chips 1 or 2 roads or columns of image pixels are loosed and no auto calibration can be used.

Comparison SA receptors characteristics and other lines of X-ray receptors design for the efficiency of electro-optical path, detective quantum efficiency (DQE) and "chessboard" noise suppression is the aim of this article.

II. Materials and Methods

Different types of digital X-ray detectors can be compared by calculating the efficiency of electro-optical path, which together with the intensifying screen form the basis of main types of the detectors.

To assess the efficiency of optical path, can be used formulas [4] for efficiency q_{2L} of lens coupling

$$q_{2L} = p_G/4F^2(m+1)^2$$

and the efficiency q_{2f} of fiber optics coupling

$$q_{2f} = p_G NA^2/m^2,$$

where

p_G – transparency of the glass;

F – optical power of the lens;

NA - fiber numerical aperture;

m – demagnification.

Detective quantum efficiency of the detector [5]:

$$DQE = p/(1+1/q_1\, q_2+1/q_1q_2q_3+Na^2/(q_1q_2q_3)^2Ni^2), \quad (1)$$

where

q_1 - conversion factor of X-ray photons into visible light photons in the intensifying screen;

q_2 - optical path efficiency;

q_3 - quantum efficiency of the photo sensor;

p - quantum efficiency of the intensifying screen;

Na - the number of pixel reading noise electrons;

Ni - the number of the interacted photons per screen pixel.

In the case of ideal electro-optical receptors path will be $q_2 = 1$, $q_3 = 1$ and Na = 0. Formula (1) is transformed

$$DQEid = p/(1 +2/q_1). \quad (2)$$

The relative detective quantum efficiency of the real optical-electronic channel can be written as

$$Qu = DQE / DQEid =$$
$$=1/(1+1/q_1\, q_2+1/q_1q_2q_3+Na^2/(q_1q_2q_3)^2Ni^2), \quad (3)$$

On the basis of formula (3) may by calculated the efficiency of electro-optical path of digital X-ray detectors for various lines of design.

One of the central questions of the design of SA-structures is to ensure the suppression of structural noise. In 1996 was proposed [6] the optical path as a combination of intensifying screen - an array of optical channels on lenses and sensors with overlapping fields of view. The distinction of the proposed approach is the procedure of image reconstruction by combining the partial image of optical channels without loss of information on the boundaries of the visual fields of individual sensors and eliminating geometric image distortion and brightness in each channel.

In terms of the correct setting of the full restoration of the image this approach certainly compares favorably with designs that provide the absence of overlapping zones, as it allows recovering the correlations at the boundaries of the visual fields of adjacent sensors.

Model image distortions in SA systems can be represented by two successive sections describing the spatial distortion of the partial images Gs {x, y, k} and effects associated with changes in the intensity of the signal brightness G_B{B,t,k; B(x,y,k)}:

$$G(x,y,t,k) = G_S\{x,y,k, G_B [B,t,k, B(x,y,t,k)]\} \quad (4)$$

where

x, y - the spatial coordinates;

t - time;

k - number of optical channel;

B - the intensity of the signal brightness.

The reconstruction problem is to find the inverse transform and eliminate the distortions introduced in the image during its formation

$$B'(x,y,t) = G^{-1}[x,y,t,k, B(x,y,t,k)], \quad (5)$$

where B'(x, y, t) - the reconstructed image.

In this case, the aim is to obtain images suitable for analysis by the operator (radiologist).

The analysis reveals the following types of distortions: determined in time and depending of time. The first type includes geometric and brightness distortions, depending from performance of the optical channels.

Spatial (geometric) distortion caused by the following factors: the geometric distortion of the lens, not perpendicularity to the optical axis of the lens and the plane of the matrix, a geometric heterogeneity of the screen (light source). To restore the form of the operator Gs{x,y,k} the method of a priori model based on the analysis of distortion of test objects. As test objects can be used linear grid (Fig. 1), where each line segment in the test is observed in at least two adjacent sensors.

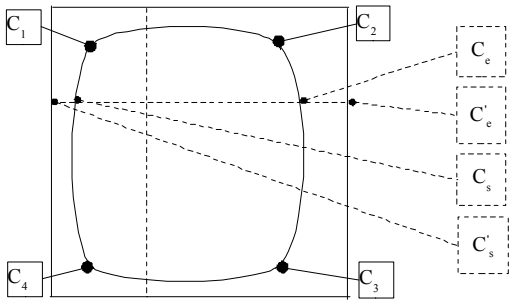


Fig. 1 Artifacts of grid cell

Correction procedure is to restore the original rectangular grid, and is described by (6).

Conversion of image coordinates:

$$C = C'/(C'e - C's)*(Ce - Cs)+Cs, \quad (6)$$

where

C - pixels coordinates in the input image;

C'- the output image pixel coordinates

Transformation in a rectangle is made by using bilinear interpolation or higher order polynomial. Quality of recovery depends on the accuracy of geometric distortion positioning of the test object. To improve the accuracy of reconstruction the form of the operator uses geometric distortion polynomial approximation coordinates of the centers of the grid lines, calculated from the center of mass of the search algorithm.

The second type of distortion is determined by the different image brightness in SA. It depends of deterministic characteristics of the sensor, amplifier output sensor signal dependence, of the transmittance of the lens in the field of view and the non-homogeneity of the X-ray flux. This feature is good enough (accuracy of 1 grayscale in 12-bit representation of the signal) described by a polynomial order of 4.

The task of restoring the form of the operator G_B may be solved by the analysis of series of images of test objects. Test objects are the images obtained for different doses, in the absence of objects within the field of view. At each point in the image brightness value sets the site of the approximation.

Distortion correction is performed in accordance with the expression

$$Bout(x,y)=(Bin(x,y)-CF_1(x,y,n))CF_2(x,y,n)+Bd(n), \quad (7)$$

where

CF_1-indent;

CF_2 - slope of the line;

n - number of the test;

Bd(n) - the value of the reference brightness correction for the n-th test;

Bin - the brightness of the input image;

Bout - the brightness of the output image.

The procedure is implemented as follows: the maximum brightness for images of test objects chosen as reference point. The task of the conversion of brightness is "alignment" test images to a reference point on the brightness of each test. Arbitrary image brightness is recalculated according to the formula (7).

The temporal variability of the signal parameters of optical channels can be restored on the basis of a posteriori model in the analysis of image parameters neighboring optical channels in the overlapping areas. Distortion correction parameters are calculated by solving a system of linear equations, which include estimates of the statistical characteristics of the image in areas of overlap.

The presence of zones of overlapping allows us to solve the problem of image reconstructing by the images of the partial output optical channels and closing gaps function at the points of joining partial images. For these purposes, we use a local spatial filtering within the overlapping areas. This takes into account features of the visual system of the operator performing the subsequent image analysis

III. Results and Discussion

Table 2 shows the calculations for the expression (3) parameters for the different design of X-ray digital receptors, where the size of the pixel array $N_X * N_Y$, the size of the pixel Lpix, q_2 - optical path efficiency, as well as parameters q_3 and Na. In the calculations the parameters were reduced to the size of the field of receptor 381 * 381mm (15 "* 15") and the dose at the receptor entrance D = 1 mR. Quality of the electro-optical detector path shown in Table 2 by resolution R and the relative detective quantum efficiency Qu.

For comparison, in the last roads of the table 2 is shown the parameters of FPD and CMOS detectors.

The calculation results indicate that the relative quantum efficiency of detection Qu, electro-optical detector path of FPD is higher than constructions with lenses. This fact, in the late 90's caused a big optimism on the prospects of receptors on flat panels. Unfortunately, the high cost of flat panels, sensitivity to external electromagnetic fields, the relatively low resolution, bad pixels in the image, and the absence of maintainability significantly reduce the initial optimism.

In contrast to the FPD receptors, the use of larger CCD matrices leads to structures with low relative detective quantum

TABLE II. PARAMETERS OF X-RAY DIGITAL RECEPTORS

Detector type	Nx* Ny	Lpix , um	q_2	q_3	Na	Resolution lp/mm	Qu
1 CCD	2048 * 2048	14	8*10-4	0.3	25	2,7	0,25
1 CCD	3000 * 3000	10	9*10-4	0.3	25	3,9	0,26
1 CCD	4000 * 4000	10	2*10-3	0.35	25	4,6	0,45
4 CCD	2200 * 2200	25	3*10-3	0.3	25	3,0	0,54
6*8 SA (1/3ò)	2900 * 2900	10	2*10-3	0.65	25	3,8	0,54
6*8 SA (1/2ò)	4500 * 4600	9	3*10-3	0.65	25	4,6	0,48
7*10 SA (1/3ò)	3300 * 3300	10	2*10-3	0.65	25	4,3	0,60
7*10 SA (1/2ò)	3100 * 3100	17	5*10-3	0.65	25	4,0	0,82
FPD	2700 * 2700	140	0,5	0.1	1000	3,6	0,95
CMOS (200 x 300 mm)	2000 * 3000	96	0,7	0,45	250	5,0	0,98

efficiency - Qu = 0,25 ... 0,26 (see roads 1 and 2, Table 2), i.e. to a four times radiation dose excess to the patient relative to the ideal receptor. Significantly higher relative quantum efficiency of detection - Qu = 0,54 corresponds to 4-matrix structure of the receptor on the CCD with large pixels (road 4 Table 2). The disadvantages of such receptor is low resolution and high cost.

A natural way to increase the resolution of SA technology is to increase the number of sensors and their area. For the field 381x381 mm array of 48 standard 1/3 " sensors (road 5 Table 2) with a pixel size of 10x10 um authorizes 3.8 lp/mm at Qu = 0,54. This model has more than 1200 installations. A further increase in resolution to 4.3 lp/mm in the standard 1/3 " 70 sensors – SA model (about 150 installations) at the same time accompanied by the growth of the relative quantum efficiency of detection Qu = 0,6 (road 7 Table 2). The 70 - SA model has superior the resolution of the flat panels, slightly losing to them in the detective quantum efficiency.

The maximum resolution - 4.6 lp/mm for X-ray detectors in market has 48 sensors array (road 6 Table 2) help the transition to double the area by 1/2" sensors (about 100 installations). The aggregate 4500 * 4600 digital matrix is implemented with a detective quantum efficiency Qu = 0,48, which is higher than a similar resolution one CCD receptor (road 3 of Table 2).

A further increase in the density of 1/2" 70-sensors array model (road 8 Table 2) allows, on the one hand, to realize mammography receptors resolution R = 7,3 lp./mm and high quantum efficiency of detection (20 installations). On the other hand, it is possible to create very sensitive receptors. By using the 2x2 binning in matrices forming virtual pixel of size 17 microns was obtained the receptor with a resolution of 4.0 lp/mm and a relative detective quantum efficiency of 0.82. Aligning a receptor with a micro focus X-ray generator implemented a set for neonatology. This unit confirmed experimentally the possibility of observing the phantom with a 1% contrast at a dose of 0.2 mR. Last is significantly higher than the sensitivity of X-ray film.

Using a 4x4 binning is obtained sensitivity, which required for fluoroscopy - 2% contrast at dose rate 0.03 mR/s and the resolution 2.0 lp/mm for the receptor field of view 205x215 mm (25 installations). This dynamic receptor replaces the X-ray image intensifier.

Image processing to suppress the noise in SA was carried out in accordance with formulas (4) - (6). The error in the reconstruction of geometric distortion is typically less than 1/4 pixel, and the error in the reconstruction of the brightness does not exceed 0.1-0.5%. Processing is carried out at a rate of five million pixels per sec. Run-time correction for a single image in a standard PC is 2-3 seconds. Residual structural noise in images doctors radiologists visually do not observed.

The best characteristics have SA detectors (line 10 Table 2) based on large area side-tileable CMOS sensors and fiber optics. Principal disadvantages of such design are border pixel losses and brightness auto calibration absents.

IV. CONCLUSION

Lens coupled SA receptors for main characteristics surpass the characteristics of large CCD receptors. They are close to FPD for detective quantum efficiency, but have better resolution. Effective "chessboard" noise suppression may be dun by sensors field of view overlapping.

For the detectors, based on large area side-tileable CMOS sensors and fiber optics, highly reasonable to make a design with sensors field of view overlapping.

REFERENCES

[1] M. Palasio, Shart Smart Digital Radiology.//Advance for Imaging and Oncology. July 2002. p. 80-81.

[2] The world's first "Multi-CCD Detector Method" DR System. RF SYSTEM lab. Press Relise. 31.03.2009.

[3] M. Farrier, T.G. Achterkirchen, G.P. Weckler, A. Mrozack Very large area CMOS active-pixel sensor for digital radiography. IEEE Trans. Electron Devices, vol. ED-56, no. 11, pp. 2623-2630, Nov. 2009.

[4] A. Maidment, M. Yaffe Analysis of the spatial-frequency-dependent DQE of optically coupled digital mammography detectors, Medical Physics, 21 (6), June 1994, p.721-729.

[5] H. Liu, A.Karellas, L. Harris, Metods to calculate lens efficiency in optically coupled CCD x-ray imaging systems, Medical Physics, 21 (7), July 1994, p.1193-1195.

[6] S. Miroshnychenko, A. Nevgasymyi, X-ray converter, Patent No.: US 7,496,177 B2, Feb. 24, 2009.

SECTION III

POWER ELECTRONICS AND SMART GRID

Electronic System With Series Organization For Connection the Distributed Generators of Renewable Energy to Low Voltage Industrial Network

Sokol E.I., Goncharov Y.P., Eresko A.V., Zamaruiev V.V., Krivosheev S.Y., Ivakhno V.V., Ilina O.V., Malyarenko E.A., Styslo B.A., Upyrenko K.G.
Dep. "Industrial and Biomedical Electronics"
National Technical University "KhPI"
Kharkov, Ukraine
Sokol@kpi.kharkov.ua

Abstract – **The article deals with the structure of the converters which connects the one-dimensional (linear) distributed energy sources that are installed along the streets and highways, to low voltage distribution network without the use of individual transformer isolations**

Keywords – photovoltaic system, wind powered generator; series transformerless connection with the industrial network

I. INTRODUCTION

Nowadays, the concept of building the photovoltaic and wind generators of electric energy in respect to large power stations with subsequent connection to the united power grid through matching converters is the most developed [1-5]. This approach is consistent with the idea of centralized electricity production, which underlies modern electrical system and therefore has a number of obvious advantages. As objects of distributed power generation such power stations may be called two-dimensional (planar), because they consist of a large number of generating cells, which are distributed in the plane. However, a lot of space is occupied by such objects for purely energetic use and therefore for the countries and regions with a high population density the use of existing regional facilities for the electric energy generation (streets, highways connecting settlements, railways etc.) is advanced. These objects occupy much of the area and, at the same time, can be used to generate electric energy, provided that connection with low-voltage distribution systems as the main receivers of electric energy is organized. A positive feature of the streets and highways as the generating facilities is that while installing generators on electric transmission line (TL) poles, the use of galvanic isolation for generators and TL under the terms of electrical safety are not required. As distributed generators such energetic objects can be called one-dimensional (linear), because they consist of generating cells which are distributed on the not necessary direct line.

These facts may be taken into account while developing the concept of "Power-generated highway" or "Power-generated street", which uses semiconductor converters as binding elements of the system. This report attempts to develop possible variants of this concept realization.

The objects of the system are installed by the distributed principle along the streets and highways on lighting poles through DC transmission line, which can be air or cable. The integration of distributed generating objects with the system of street and highway lighting at night with the help of energy-conserving semiconductor lighting is already being announced as the lighting system without the additional cost of electricity, for which it must have batteries [7]. However, when effective communications with low-voltage industrial network are organized, it is possible to get along without expensive electrochemical batteries by using means of organizing the balance between consumption and generation of energy available in the network. The advantage of these objects with network communication is that solar energy is received in the afternoon, when the network load is at maximum, and the return of energy for lighting occurs at night, when the other load in the network is reduced. The fact that galvanic isolation is not required under the terms of electrical safety is used for series organization of electrical system, according to which all of its objects are connected in series with each other and through a general rather extended DC transmission line and through a regenerative converter (RC) are added to the existing alternating current system. An object of the system is generally understood as a set of photogenerators, light-emitting diode (LED), and wind powered generators, which through matching converters, and in some cases without them, are connected in series. In steady mode the regenerative converter maintains the current at TL at a desired reference value, and operates mainly in the generator mode as the current source inverter. If the power of distributed generators is not enough to supply the lighting load, then RC switches into the rectifier mode, taking the missing power from the power line.

The polarity of voltage at the input of TL changes on reversed polarity and the current keeps the same direction. Fig. 1 illustrates the overall structure of the system.

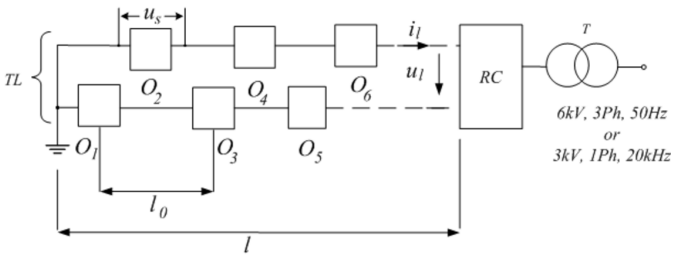


Fig. 1. The overall structure of the system

The objects must be installed on the top of the lighting poles (on the base stands), and be isolated toward the ground. In case of two-wire performance of TL with grounded endpoint the base stand potentials increase as one moves from the endpoint, and at the reaches the value $u_l/2$ at the input of TL. This two-wire organization allows twice reducing the maximum potential of the base.

As the voltage u_l is approximately equally divided between objects, the system objects are low-voltage, which can be used to simplify and reduce the cost of their design. But the main advantage of the structure is that galvanic isolations between objects and TL are not required, which would require a low-power isolated transformers. Individual matching converters are performed as pulse width converter (PWC) and transformer isolation is not required, which would be inevitable with traditional parallel methods of connecting to the TL. Total isolation transformer (T) is therefore sufficiently powerful that allows achieving acceptable characteristics even at 50 Hz. But the greatest effect can be obtained by applying the single-phase alternating current in low voltage distribution networks with a frequency of about 20 kHz instead of the traditional three-phase low frequency of 50 Hz [6].

In the following sections of the paper the main properties of the structure Fig. 1 are analyzed and suitable schematics solutions of converters are considered.

II. Determination of Transmission Line Parameters

Let's assume that the total number of objects in TL is equal to n, and the greatest voltage on each of them is equal to v_s. Then we obtain the following basic relations:

$$u_l = n \cdot u_s, \tag{1}$$

$$(1-\eta) \cdot u_l = 2 \cdot \rho \cdot j \cdot l, \tag{2}$$

$$l = n \cdot l_0, \tag{3}$$

where η is the estimated efficiency of TL at maximum load, j and ρ — current density in the TL conductor and specific resistance of the conductor, l — distance between the grounded endpoint and the input of TL, l_0 — distance between two objects (see Fig. 1).

Equation (2) gives the total voltage drop in TL that can be associated with the estimated efficiency and full voltage u_l at the input of line.

For example, take the permitted voltage at the input of TL $u_l = 6$ kV, which corresponds to half the voltage of ±3 kV on the isolated base relative to ground. We also assume $l_0 = 50$ m, $\eta = 0.9$, $\rho_{(Al)} = 2.5 \cdot 10^{-8}$ Ω·m, $j = 2 \cdot 10^6$ A/m^2. Then we find that $l = 6$ km, $n = 120$, $u_l = 50$ V. The distance $2l = 12$ km is comparable to the average distance between the settlements, where it is appropriate to locate the converter of connection with electric system. This distance can be considered as acceptable for highways outside of settlements. With smaller distance between settlements it is possible to use shorter distance of TL and respectively, with less voltage u_l on it. While changing the interval l_0 only the number of poles and working voltage u_s at the object is changed.

The control strategy of the proposed system, the overall structure of which is shown on Fig. 1, should be organized in such a way that the voltage at the input of TL is maintained close to the maximum, regardless of the transmitted power. Then, if the transmitted power declines, TL current instantly proportionally reduces. Assuming that the average power is equal to half of the maximum power and the fact that power losses at TL are proportional to the square of current, we find that the average power losses are equal to $1/3 \cdot (1 - \eta) = 3.3\%$ of the maximum power, which is acceptable.

III. Individual Matching Converters

Fig. 2 shows a possible structure of the object of Fig. 1 system, that has the most extensive control possibilities.

The use of back-boost converter as a matching converter for photo and light modules allows to regulate the voltage both downwards and upwards from the supply voltage. For the photovoltaic module supply voltage is the output voltage of the photovoltaic element and for the light module it is the voltage on the capacitor C_c.

Converters allow to maximize the power utilization of photovoltaic modules and to increase the efficiency of TL.

It is expedient to operate the matching converters at frequency $f = 1/T$ of 50 kHz or higher and in discontinuous current mode of filter inductors L (Fig. 2b). That allows obtaining the following advantages: small inductor size, the absence of shoot-through currents during switching interval, the omission of the photogenerator's current sensor in the part of control structure that implements the maximum power point tracking (MPPT) algorithm.

The last is implementable while the amount of energy transferred to the load during the PWM cycle T is known:

$$A = L \cdot I_m^2/2. \tag{4}$$

The current amplitude I_m is measured by a simple DC current transformer, which is installed in the control switch circuit.

With a decrease of solar radiation intensity, the converter reduces the voltage on filter capacitor C_f. The duration of the current pulse in the diode VD1 increases, as shown at Fig. 2b by hatching, and in the limit, the capacitor C_f is completely shunt by the diode through the inductor wiring. The desired current in the light-emitting diode is set similarly. The measurement of LED's current is also not required.

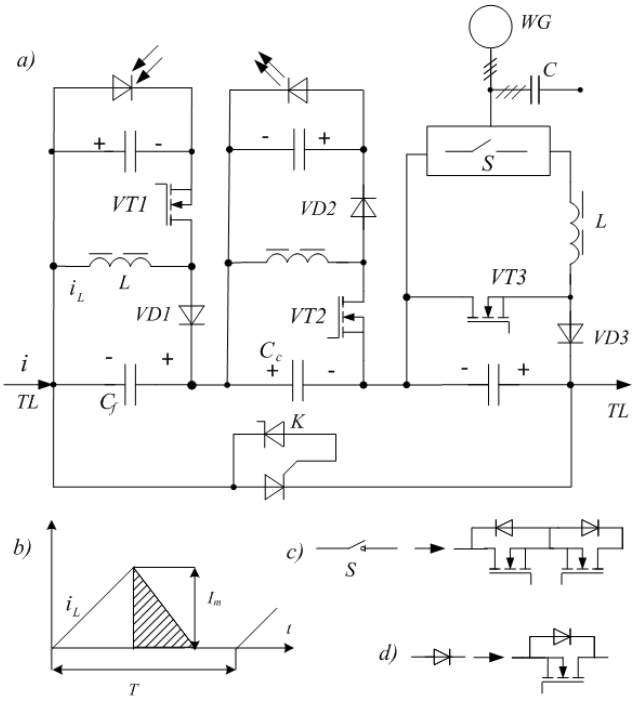


Fig. 2. The structure of the system′s object with elements installed on the same base stand

The main converter of wind generator, which is performed as asynchronous or permanent magnet synchronous [1], is in general regenerative. The converter acts as a current source inverter with PWM in starter mode of wind generator, which is suitable for rapid start and breaking of the wind turbine, and as a rectifier with PWM in operating mode.

To implement these algorithms at low input voltage that can range from zero to 50V the basic switch S is advisable to perform as a series connection of two low-voltages MOSFETs (Fig. 2c). Forward voltage is perceived by the left transistor and the back voltage – by the right diode. Right transistor shunts the diode in its conducting state in order to reduce the direct voltage drop and the left diode which is naturally presented in the structure of the MOSFET, in this case is not used. In the low-voltage MOSFET it is possible to get the direct voltage drop of order (0.1 ÷ 0.2) V, which provides high efficiency at low input voltage.

When using PWM in the current source inverters, the input filter is composed by self-inductance of the generator and the parallel capacitor C. Application of MOSFETs makes it possible to implement PWM frequency in the range of 20 - 50 kHz. Therefore, the required capacity is low and the resonance frequency of the input filter is certainly beyond the bounds of generator's operating frequency range. The presence of input capacitor combined with low voltage practically eliminates the problem of high-frequency isolation that exists in high-voltage machines, which are controlled by voltage source inverters.

When using a permanent magnet synchronous generator the main three-phase converter can be simplified by replacing the MOSFETs on the Schottky diodes. However, the application of controlled converter with rather complex switches implementation according to Fig. 2c is paid off by the number of advantages: the expensive permanent magnet synchronous generator is replaced by the cheaper asynchronous generator, it is possible to form sinusoidal currents in the windings and to use starter mode through power grid, the expensive mechanically operated controller of turbine blades location may be replaced by a simple wind vane with restrictions on the turning angle. The latter is associated with the possibility to control the magnetic flux of machine at abnormally high wind speeds and greater stability of the small wind turbine against a possible mechanical failure. Small wind turbine, being a high-speed turbine, provides a given torque on the generator shaft without the use of permanent magnets. This requires a shift to wind turbine construction with axial magnetic flux [1] or with radial magnetic flux, but with the use of disk shaped construction. This also creates the possibility to increase the actual frequency and as a result to reduce the generator size by increasing the number of pole pairs.

The feature of the current source inverter is that it operates on the TL side as a step-down converter. At the same time, the maximum efficiency of TL is achieved at the highest voltage v_l on it that corresponds to the minimum current and minimum energy loss in the line. According to this factor, it is desirable to supplement the wind turbine module by step-up converter, as it is shown in Fig. 2.

Another feature of the series structure is that the open-circuit mode in any of the modules is an emergency for it, as it disables the entire system. In conventional parallel structure the short-circuit mode in any of the modules has similar property. For protection against open-circuit fault it is necessary to shunt the damaged object by switch K, for example, by thyristor, which switches on at voltage increase to a certain point, which is specified by Zener diode in a gate circuit. This protection at maximal voltage is dual to protection of maximal current in case of parallel principle of system organization. To reduce the voltage drop in the on-state, diodes VD1 – VD3 can be replaced by MOSFETs according to Fig. 2d.

IV. Regenerative Converter for Grid-connection

At long and high-voltage TL it is appropriate to perform RC as several series-connected sections, which are attached to separate secondary windings of the power transformer T (Fig. 3 a). Separate secondary windings of transformer enable sections voltage balancing. Within the settlement, it is possible to get along without galvanic isolation from the network by using short TL (Fig. 3 b).

When connecting to the distribution network 380 V, 50 Hz the maximum possible voltage u_l at the input of TL is 500 V. Under the same initial data, as in the example from the second part of this paper, we obtain that l=0.5 km so that the distance $2l$ between adjacent points of connection to the three-phase network will be 1 km. At each point of connection to the network there are generally two separate sections of converter and their TLs go down along the street in opposite directions. The feature of this variant is that the opposite end of TL can not be grounded, as is the case at Fig. 3a, because of the presence of a grounded neutral wire in the network.

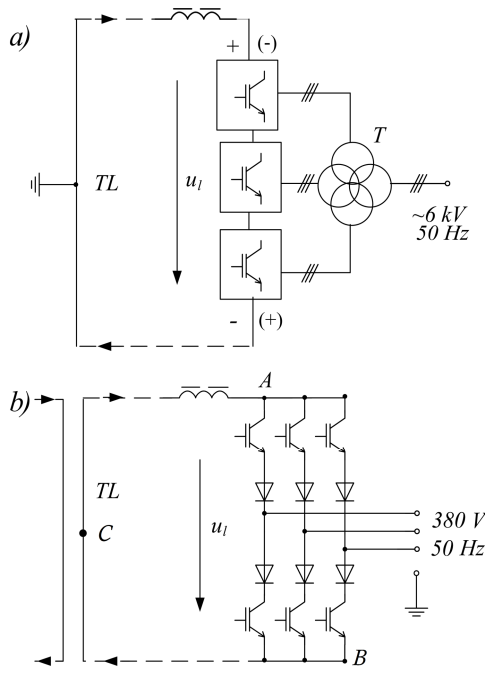


Fig. 3. Connection options of RC: a - with galvanic isolation from the power line, b - without galvanic isolation. Separate secondary windings provide voltage dividing at sections.

Real control algorithms of three-phase current source inverter do not provide symmetry of potential in points A, B relative to the grounded point and therefore the potential of point B at the opposite end of TL is not equal to zero, even with a symmetric connection of filter inductor in the forward and return conductors. The proposed regenerative converter topology with the control algorithm given in [8] has been simulated in Matlab / Simulink. As simulation results the potentials of the points A, B are shown on Fig. 4.

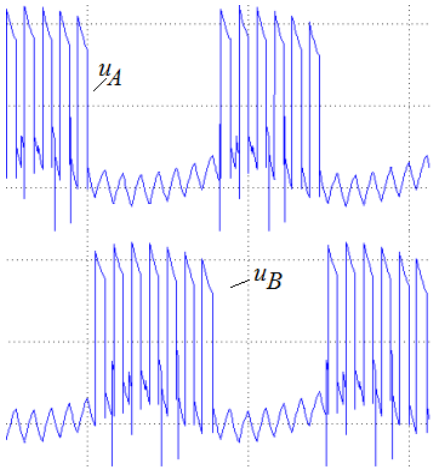


Fig. 4. Potentials of output points at scheme Fig 3b, measured on its computer model

Given the intermittent voltage change, some influence will cause the wire capacity of TL relative to the ground. However, the evaluation shows that at short TL and low voltage in it this effect is negligible.

Since in the current source inverter two-quadrant switches are required, whether the IGBTs should have the ability to block reverse voltage or additional free-wheeling diodes are needed. MOSFETs can also be used, but devices with low operating voltage require more sections.

CONCLUSIONS

This paper has presented the series organization of distributed power generation system, which is integrated with the lightning load through DC transmission line and connected to the grid by means of regenerative converter based on voltage source inverter or current source inverter topology. When applied to one-dimensional (linear) objects such as streets and highways, railways, etc. the proposed system presents several advantages in comparison with conventional (parallel) system of distributed generation, such as the elimination of the need for individual isolation transformers and simplification of matching converters schemes. Reliability of the proposed series distributed power generation system is assured by electronic switch, which shunts damaged object in the case of open-circuit fault and by the fact that short-circuit mode of object operation does not affect on system operability as a whole.

REFERENCES

[1] Mukand R. Patel Wind and Solar Power Systems. Design, Analysis, and Operation Second Edition. Taylor & Francis. 2006, 433p.

[2] Ahmed F. Zobaa, Ramesh C. Bansal. Handbook of renewable energy technology. Copyright © 2011 by World Scientific Publishing Co. Pte. Ltd., 847 p.

[3] S.B. Kjacr, J.K. Pedersen, F. Blaabjerg. A Review of Single–Phase Grid–Connected Inverters for Photovoltaic Modules // IEEE Trans, on Industry Applications, v.41, 2005, N5, pp. 1292–1306

[4] W. Kramer, S. Chakraborty, B. Kroposki, H. Thomas. Advancer Power Electronic Interfaces for Distributed Energy Systems. Technical Report NREL/TP–581–42672, March 2008

[5] Rashid M.H. Power Electronics Handbook. – San Diego. Academic Press, 2007, 1153 p.

[6] Sokol E.I., Goncharov Y.P., “Ispol'zovanie silovoj elektroniki v nizkovol'tnyh raspredelitel'nyh elektrosetjah kommunal'nyh ob’ektov” (“Applying of power electronics in low voltage common power line”). Sbornik nauchnyh trudov instituta jelektrodinamiki NAN Ukrainy. Kiev, 2011, str. 101-111 /In Russian/

[7] Solar Wind Hybrid Street Lights http://www.sunecochina.com/, http://www.rentechno.com.ua/index.php

[8] Peretvoruvalna tehnika (“Power conversion equipment”) Goncharov Y.P., and others. – Kharkiv: Folio, 2000. – 360 p.

Method for Determining Energy Losses in Switched-Mode Converter Power Switches Depending on Storage Inductor Current Modes

Oleksandr M. Afanasyev
Donbas State Technical University
Lenin Ave., 16, Alchevs'k, UA-94204, Luhans'k reg., Ukraine
alexandr.afanasyev@gmail.com

Anatoliy A. Shcherba
Institute of electrodynamics of NAS of Ukraine,
Peremogy Ave. 56, UA-03680 Kyiv, Ukraine

Abstract – **The analytical method for evaluation of switched-mode converters energy efficiency depending on storage inductor current modes using experimental values of IGBT-modules parameters from datasheets is proposed.**

Key words – switched-mode converter; discontinuous and continuous current; IGBT-module; energy efficiency; analytical method.

I. INTRODUCTION

Switched-mode converters are widely developed nowadays. Increasing their energy efficiency is up-to-date trend. Energy saving effect can be achieved by increasing conversion frequency, by using fast power switches and correct choice of storage inductor current mode. Depending on the current mode of storage inductor of switched-mode convertor the continuous current mode (CCM), the discontinuous current mode (DCM) and the boundary current mode (BCM) are distinguished.

Fig. 1 illustrates the basic circuit diagrams of the switched-mode converters.

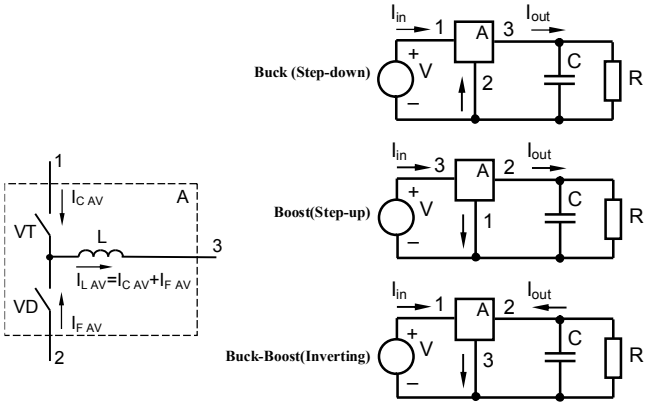


Fig. 1. The basic circuit diagrams of the switched-mode converters.

Each circuit includes the following elements: L – storage inductor, VT – transistor power switch, VD – diode. They are connected by the "Y" configuration in unit A. Different variants of connection of the unit A with DC voltage source V, the load resistor R and the filtering capacitor C allow to implement three types of switched-mode converters: Buck, Boost and Buck-Boost. Input and output currents of those circuits are shown in TABLE I.

TABLE I. INPUT AND OUTPUT CURRENTS OF SWITCHED-MODE CONVERTERS

Converter	Buck	Boost	Buck-Boost
I_{in}	$I_{C\,AV}$	$I_{L\,AV}$	$I_{C\,AV}$
I_{out}	$I_{L\,AV}$	$I_{F\,AV}$	$I_{F\,AV}$

Fig.2 illustrates typical current diagrams of unit A for CCM and DCM of switched-mode converters.

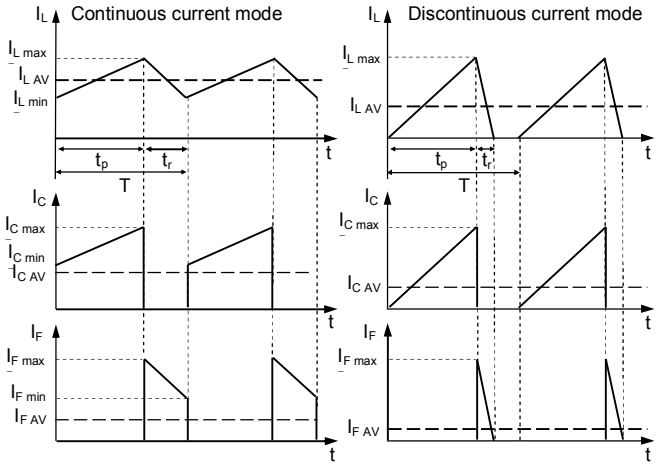


Fig. 2. Typical current diagrams for CCM and DCM of switched-mode converters.

The average value of storage inductor current $I_{L\,AV}$ is

$$I_{L\,AV} = I_{CAV} + I_{FAV}. \quad (1)$$

The average and RMS values of converter currents (fig. 1) determine the conductivity energy losses in power switches. The instant values of current, corresponding to turn-on and turn-off moments, determine the switching losses.

For CCM the transistors hard switching (HS) and great energy losses E_{on}, E_{off} are typical. In BCM and DCM the turn-on energy losses (E_{on}) are lower due to zero current switching (ZCS) transistors, but the turn-off energy losses (E_{off}) are greater than in CCM.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

The analysis of influence of storage inductor current ripple on switched-mode convertor energy parameters may be performed using the normalized time constant $T_L = f \cdot L/R$ (f is frequency of commutation, L is inductance of storage inductor, R is resistance of load) or using load coefficient $\beta = 2 \cdot f \cdot L/R$. To determine the current modes it is necessary to introduce either relative normalized time constant T_L^* (the ratio of T_L to the time constant for BCM T_{LB}) or relative load coefficient β^* (the ratio of β to the load coefficient for BCM β_B). The relation between the inductor peak-to-peak ripple current and those coefficients has nonlinear character and is described by two different functions depending on the current mode that makes the analysis more difficult [1, 2]:

- for CCM and BCM

$$\Delta I_L = 2 \cdot I_{L\,AV} / \beta^*, \quad (2)$$

- for DCM

$$\Delta I_L = 2 \cdot I_{L\,AV} \cdot \sqrt{\beta^*}, \quad (3)$$

where $\Delta I_L = (I_{Lmax} - I_{Lmin})$ is the peak-to-peak inductor ripple current, D is transistor duty cycle, $I_{L\,AV}$ is the average value of storage inductor current.

To simplify analysis of switched-mode converter and define the current mode of storage inductor without regarding load resistance the ripple factor of storage inductor current K_{RL} is proposed for use. The relation between the peak-to-peak inductor ripple current and K_{RL} has linear character and is described in the same way for all current modes (CCM, DCM, BCM) [3, 4]:

$$\Delta I_L = 2 \cdot K_{RL} \cdot I_{L\,AV}. \quad (4)$$

The values $0 < K_{RL} < 1$ correspond to CCM, the value $K_{RL} = 1$ corresponds to BCM, the values $K_{RL} > 1$ correspond to DCM.

The aim is to develop the method of switched-mode convertor energy efficiency evaluation depending on ripple factor K_{RL}.

II. The Energy Efficiency Evaluation Method

The turn-on transistor current is equal to minimal inductor current $I_{Con} = I_{Lmin}$, the turn-off transistor current is equal to maximal inductor current $I_{Coff} = I_{Lmax}$, that are calculated considering (4) as follows:

- for CCM and BCM ($0 < K_{RL} \leq 1$)

$$I_{L\,min} = I_{L\,AV} \cdot (1 - K_{RL}); \quad (5)$$

$$I_{L\,max} = I_{L\,AV} \cdot (1 + K_{RL}); \quad (6)$$

- for DCM ($K_{RL} > 1$):

$$I_{L\,min} = 0; \quad (7)$$

$$I_{L\,max} = 2 \cdot I_{L\,AV} \cdot K_{RL}. \quad (8)$$

The transistor's average current value $I_{C\,AV}$ is determined as:

$$I_{C\,AV} = \frac{1}{T} \int_0^{t_p} i_L(t)\,dt = \frac{1}{T} \int_0^{t_p} I_{Lmin} + (I_{Lmax} - I_{Lmin}) \frac{t}{t_p}\,dt. \quad (9)$$

The solution of (9) is

$$I_{C\,AV} = [I_{Lmin} + 0.5 \cdot (I_{Lmax} - I_{Lmin})] \cdot D. \quad (10)$$

Considering (4) the transistor's average current value $I_{C\,AV}$ for $K_{RL} = [0..1]$ is calculated as

$$I_{C\,AV} = I_{Lmax} \cdot D \cdot (1 + K_{RL})^{-1} = I_{LAV} \cdot D, \quad (11)$$

where $D = t_p \cdot T^{-1}$ is transistor duty cycle, T is cycle time, t_p is the transistor's conductivity time.

For $K_{RL} = (1..\infty)$, when $I_{Lmin} = 0$, considering (8) the transistor's average current value $I_{C\,AV}$ is

$$I_{C\,AV} = 0.5 \cdot I_{L\,max} \cdot D = I_{LAV} \cdot D \cdot K_{RL}. \quad (12)$$

The average current value of diode I_{FAV} is determined similarly, considering that the duty cycle of diode is $D' = t_r \cdot T^{-1}$.

For CCM and BCM $D' = 1 - D$, but for DCM the duty cycle of diode D' is calculated using volt-second integral of inductor. Formula (13) is correct for Buck converters, formula (14) is correct for Boost converters, formula (15) is correct for Buck-Boost converters.

$$t_p \cdot (V_{in} - V_{out}) = t_r \cdot V_{out} \Rightarrow D' = D \cdot (1 - M) \cdot M^{-1}; \quad (13)$$

$$t_p \cdot V_{in} = t_r \cdot (V_{out} - V_{in}) \Rightarrow D' = D \cdot (M - 1)^{-1}; \quad (14)$$

$$t_p \cdot V_{in} = t_r \cdot V_{out} \Rightarrow D' = D \cdot M^{-1}. \quad (15)$$

The transistor's RMS current value $I_{C\,RMS}$ is determined as:

$$I_{CRMS} = \sqrt{\frac{1}{T} \int_0^{t_p} [i_L(t)]^2\,dt} = \sqrt{\frac{D}{3}(I_{Lmax}^2 + I_{Lmax} \cdot I_{Lmin} + I_{Lmin}^2)}. \quad (16)$$

Considering (5) and (6) the transistor's RMS current value $I_{C\,RMS}$ for $K_{RL} = [0..1]$ is

$$I_{C\,RMS} = I_{LAV} \sqrt{D \cdot (K_{RL}^2 + 3)/3}. \quad (17)$$

The diodes' RMS current value I_{FRMS} is determined similarly:

$$I_{F\,RMS} = I_{LAV} \sqrt{D' \cdot (K_{RL}^2 + 3)/3}. \quad (18)$$

The obtained relationships for calculating transistor currents defining energy losses are assembled in TABLE II.

TABLE II. Transistor Currents

$K_{RL} = [0..1]$	$K_{RL} = (1..\infty)$
$I_{Con} = I_{L\,AV} \cdot (1 - K_{RL})$	$I_{Con} = 0$
$I_{Coff} = I_{L\,AV} \cdot (1 + K_{RL})$	$I_{Coff} = 2 \cdot I_{LAV} \cdot K_{RL}$
$I_{C\,AV} = I_{L\,AV} \cdot D$	$I_{C\,AV} = I_{LAV} \cdot D \cdot K_{RL}$
$I_{C\,RMS} = I_{LAV} \sqrt{D \cdot (K_{RL}^2 + 3)/3}$	$I_{C\,RMS} = I_{LAV} \cdot K_{RL} \sqrt{4 \cdot D/3}$

The conversion ratio M is determined as:

$$M = I_{in} \cdot I_{out}^{-1}. \quad (19)$$

Considering (11)-(15) and (19) for the basic converter circuits (Fig. 1) relationships for transistor duty cycle D and diode duty cycle D' are obtained. The relationships are represented in TABLE III.

TABLE III. TRANSISTOR AND DIODE DUTY CYCLES

Converter	K_{RL}=[0..1]	K_{RL}=(1..∞)
Buck	$D = M$ $D' = 1 - M$	$D = M \cdot K_{RL}^{-1}$ $D' = (1 - M) \cdot K_{RL}^{-1}$
Boost	$D = (M - 1) \cdot M^{-1}$ $D' = M^{-1}$	$D = (M - 1) \cdot (M \cdot K_{RL})^{-1}$ $D' = (M \cdot K_{RL})^{-1}$
Buck-Boost	$D = M \cdot (M + 1)^{-1}$ $D' = (M + 1)^{-1}$	$D = M \cdot ((M + 1) \cdot K_{RL})^{-1}$ $D' = ((M + 1) \cdot K_{RL})^{-1}$

The obtained analytical relationships may be used for energy efficiency analysis of switched-mode converters and for determining optimum current mode and optimum commutation frequency.

The method of piecewise linear approximation is proposed to obtain the analytical relationships describing current-voltage (I-V) characteristics.

In general, the energy of conductivity losses in transistor can be presented as

$$E_{VTcond} = \int_0^{t_p} v_{CE}(t) \cdot i_C(t) dt. \quad (20)$$

According to I-V characteristic approximation results the relationship for collector-emitter saturation voltage V_{CE} depending on I_C is obtained:

$$V_{CE} = V_{CE0} + I_C \cdot r_{CE\,dif}, \quad (21)$$

$$r_{CE\,dif} = \left(V_{CE(2)} - V_{CE(1)}\right) \cdot \left(I_{C(2)} - I_{C(1)}\right)^{-1}, \quad (22)$$

where V_{CE0} is initial collector-emitter voltage with I_C=0; $V_{CE(1)}$, $V_{CE(2)}$, $I_{C(1)}$, $I_{C(2)}$ are voltage and current values taken from the datasheet graphs; r_{CEdif} is transistor's differential resistance.

The solution of (20) considering (21) is

$$E_{VTcond} = f^{-1} \cdot [V_{CE0} \cdot I_{C\,AV} + r_{CEdif} \cdot I_{C\,RMS}^2]. \quad (23)$$

The relationship for determining the conductivity energy losses in inverse diode is similarly obtained as

$$E_{VDcond} = f^{-1} \cdot [V_{F0} \cdot I_{F\,AV} + r_{Fdif} \cdot I_{F\,RMS}^2], \quad (24)$$

where $r_{F\,dif}=\Delta V_F/\Delta I_C$ is diode's differential resistance.

The switching losses E_{sw} of IGBT module are defined as the sum of turn-on losses E_{on} and turn-off losses E_{off} in transistor and also turn-off losses in inverse diode E_{rr}.

The regression analysis is proposed for approximation of datasheet graphs of energy for considering their nonlinearity. The data averaging is performed by least squares method using MathCAD.

For the most correct approximation of datasheet graph of E_{on} depending on I_C the following relationship is proposed:

$$E_{on} = b \cdot I_{Con}^n + c \cdot I_{Con}, \quad (25)$$

where b, c, n are approximation indexes.

The relationships $E_{off}=F(I_{Coff})$, $E_{rr}=F(I_{Con})$ for approximation of datasheet graph of E_{off} and E_{rr} depending on I_C are obtained similarly to (25).

The similar methods using approximation of datasheet characteristics are considered in [5 - 7].

The total power losses in IGBT module is defined as follows:

$$P_{tot} = \left(E_{VTcond} + E_{VDcond} + E_{VTsw} + E_{VDsw}\right) \cdot f, \quad (26)$$

where E_{VTsw} – switching energy losses in transistor;

E_{VDsw} – switching energy losses in diode.

For convenience the concept of relative power losses in IGBT module is introduced. They are determined by formula

$$P_{tot}^* = P_{tot} / P_{out}, \quad (27)$$

where P_{out} is power in the load.

The following assumptions are adopted: the energy losses in all circuits' elements equal zero except for in transistors and diodes, storage inductor is the linear element, voltage ripple on filtering capacitor are negligibly small.

III. THE ANALYSIS OF ENERGY EFFICIENCY OF BUCK CONVERTER WITH DIFFERENT TYPES OF IGBT MODULES

The parameters of different types Semikron IGBT modules obtained as the result of datasheet graphs approximation are represented in TABLE IV.

TABLE IV. THE RESULT OF APPROXIMATION OF REFERENCE DEPENDENCE OF SWITCH POWER LOSSES ON COLLECTOR CURRENT AND APPROXIMATION OF STATIC VOLT-AMPERE CHARACTERISTICS

Parameter		SKM200GB12V	SKM200GB12E4	SKM200GB125D	SKM200GB12T4
$E_{on}(I_{Con})$	b	0.13	$1.018 \cdot 10^{-9}$	$1.136 \cdot 10^{-11}$	$2.254 \cdot 10^{-12}$
	c	0	0.103	0.166	0.11
	n	0.75	4	5	5
$E_{off}(I_{Coff})$	b	0.237	0.304	0.187	0.23
	c	0	0	0	0
	n	0.85	0.85	0.8	0.85
$E_{rr}(I_{Con})$	b	0.658	0.667	0.269	1.429
	c	0	-0.074	0	0
	n	0.6	0.7	0.6	0.4
$V_{CE}(I_C)$	V_{CE0}	0.797	0.705	1.406	0.789
	r_{CEdif}	$7.094 \cdot 10^{-3}$	$7.579 \cdot 10^{-3}$	0.017	$6.836 \cdot 10^{-3}$
$V_F(I_F)$	V_{F0}	0.712	0.736	0.995	0.771
	r_{Fdif}	$8.314 \cdot 10^{-3}$	$7.257 \cdot 10^{-3}$	0.771	$7.143 \cdot 10^{-3}$

The Buck Converter (Fig. 1) was simulated and energy efficiency characteristics were obtained using the proposed method. The following parameters of converter were accepted: load resistance R=6 Ohm, input DC voltage V=600V and conversion ratio M=0.75.

Fig.3 illustrates the buck converter simulation results. The currents on the graphs are relative values obtained through division by maximal load current.

Fig. 4 illustrates the relations between the energy losses E_{on}, E_{off}, E_{rr}, E_{VTcond}, E_{VDcond}, E_{tot} and the ripple factor K_{RL}. As we see the BCM is the most effective mode because total energy losses E_{tot} are minimal at K_{RL}=1.

Fig. 5 shows the relative total power losses in IGBT module change depending on the ripple factor K_{RL} at commutation frequencies 1 kHz and 100 kHz. High energy efficiency at low frequency is achieved in CCM and BCM (K_{RL}≤1). At high frequency the maximal energy efficiency is achieved in BCM (K_{RL}=1).

Fig. 6 illustrates the relative total power losses in IGBT module change depending on the commutation frequency at K_{RL}=0.1, K_{RL}=1 and K_{RL}=2. The maximal energy efficiency is achieved at frequencies more than 20kHz for IGBT module SKM200GB125D which has the lowest turn-off energy E_{off} compared to other modules.

At low frequencies (up to 10 kHz) the converter energy efficiency mostly depends on conductivity energy losses E_{VTcond}, E_{VDcond} in IGBT module. Therefore, high efficiency is achieved using semiconductor modules with low conductivity losses (SKM200GB12V,SKM200GB12T4, SKM200GB12E4) for all values of K_{RL}.

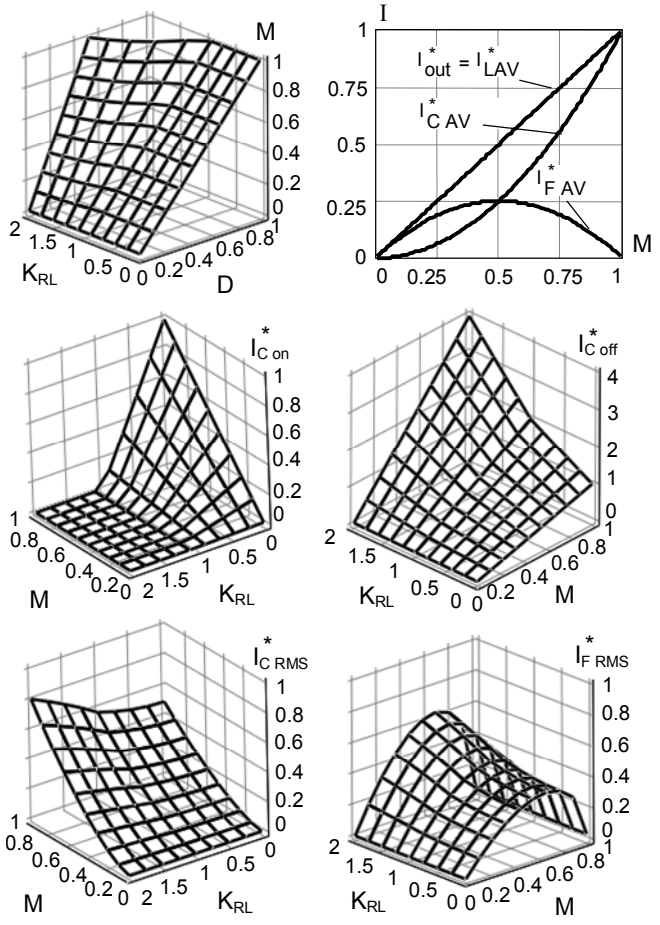


Fig. 3. Simulation results of the buck converter.

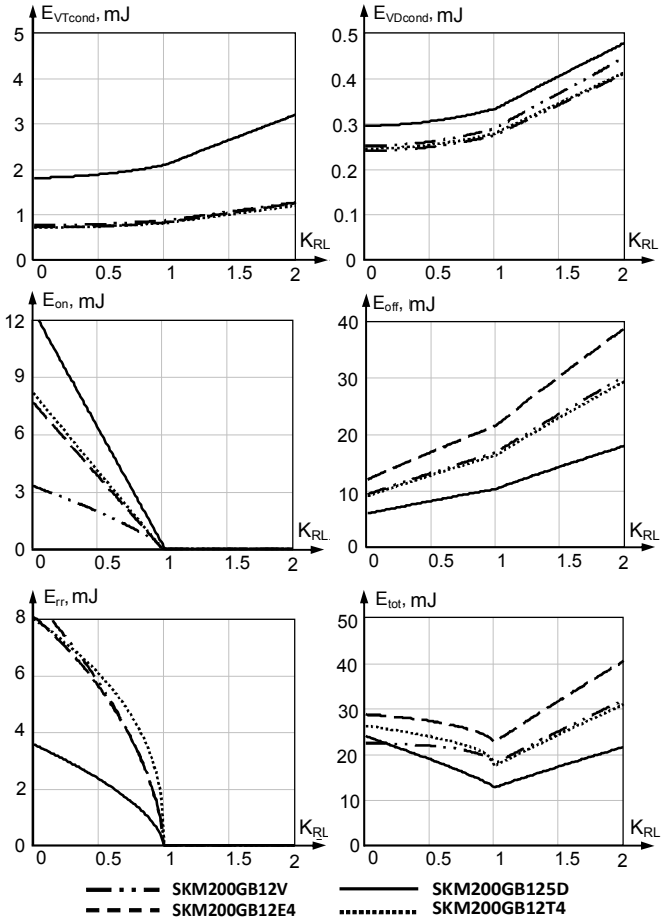


Fig. 4. Simulation results of the buck converter at f = 100 kHz.

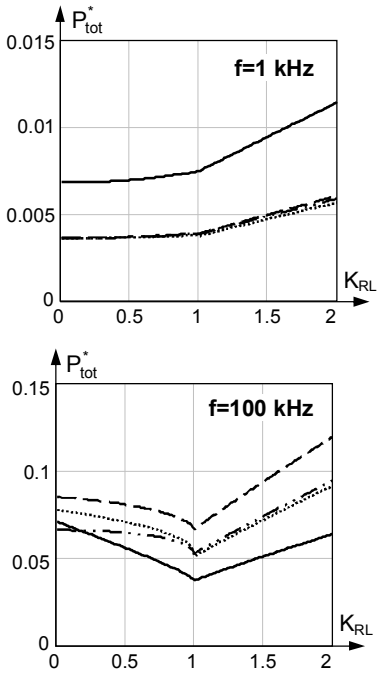


Fig. 5. The relative total power losses change depending on the ripple factor K_{RL}

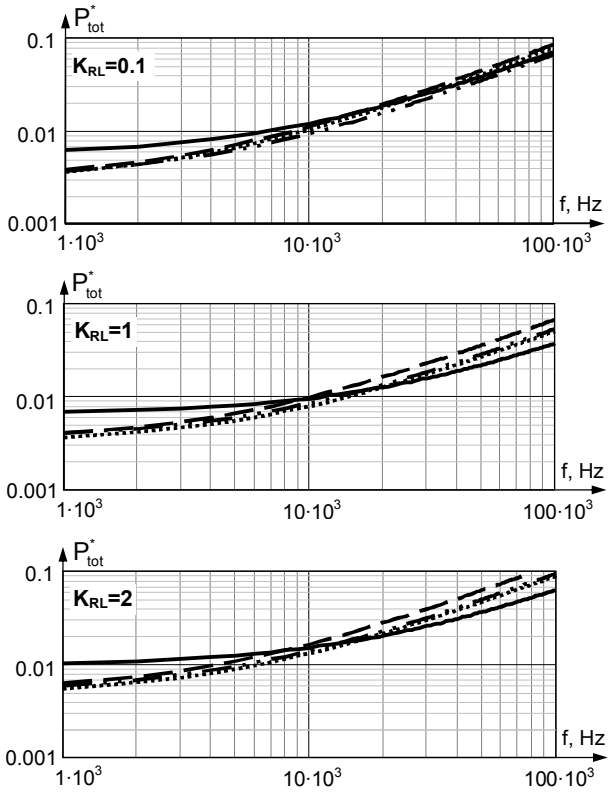


Fig. 6. The relative total power losses change depending on the commutation frequency.

CONCLUSION

The proposed analytical method of evaluation of switched-mode converters energy efficiency using experimental values taken from datasheets allows to choice optimal current mode and frequency range for achieving minimal energy losses in semiconductor modules.

Using the ripple factor K_{RL} allows easy determining current modes (CCM, BCM or DCM) and contributes to simplifying energy characteristics analysis of switched-mode converters.

REFERENCES

[1] R. Severns, G. Blum, Impulsnyye preobrazovateli postoyannogo napryazheniya dlya sistem vtorichnogo elektropitaniya, M.: Energoatomizdat, 1988.

[2] V. S. Moin, Stabilizirovannyye tranzistornyye preobrazovateli, M.: Energoatomizdat, 1986.

[3] A. M. Afanasyev, A.A. Shcherba, A.N. Baranov, "Metod analiza rezhimov impulsnykh preobrazovateley napryazheniya s primeneniyem koeffitsiyenta pulsatsiy toka v nakopitelnoy induktivnosti", Tekhnichna yelektrodinamika. Temat. vip. Silova yelektronika ta yenergoyefektivnist, 2012, № 4, pp. 62-67.

[4] A. M. Afanasyev, A.V. Yeryomina, "Zavisimost energeticheskikh pokazateley silovykh klyuchey impulsnykh preobrazovateley ot koeffitsiyenta pulsatsii", Sbornik nauchnykh trudov DonGTU, 2012, №38, pp. 284-289.

[5] Infineon Technologies, Application Note ANIP9931E—Calculation of Major IGBT Operating Parameters. Germany: Infineon Technologies, 1999.

[6] A. Blinov, D. Vinnikov, T. Jalakas. "Loss Calculation Methods of Half-Bridge Square-Wave Inverters", Electronics and Electrical Engineering. – Kaunas: Technologija, 2011, No. 7(113), pp. 9–14.

[7] R. Schnell, U. Schlapbach, "Realistic benchmarking of IGBT-modules with the help of a fast and easy to use simulation-tool", PCIM proceedings, 2004.

Conversion of Shaded-Pole Induction Motor to Switched Reluctance Motor and Effects of Pole Shoe and Notch on SRM Noise

A. Dadpour, K.Ansari,
Ferdowsi University of Mashhad,
Mashhad, Iran
a.dadpour@gu.ac.ir, k.ansari@um.ac.ir

M. Divandari, M. R. Hosseinkhah
Department of Electrical Engineering,
Aliabad Katoul branch, Islamic Azad University,
Aliabad Katoul, Iran
mohammad.divandari@gmail.com,
m.r.hosseinkhah@gmail.com

***Abstract*—The switched reluctance motor (SRM) is well known for its higher acoustic noise, caused by stator vibrations. This paper describes a method for noise reduction in a new stator pole face SRM having pole shoe and notch. This motor has been made by conversion shaded-pole induction motor to SRM. Static characteristic of the motor is analyzed and calculated by using the method of the finite element (FE) analysis and compared with SRMs having conventional shape but without pole shoe and notch. The acoustic noise intensity and torque ripple for the base motor were calculated by taking the static profile of radial force and torque. The results showed that the amplitude of radial force and torque ripple in converted SRM were decreased in comparison with the base motor but there was no change in torque.**

Keywords — Switched reluctance motor, radial force, torque ripple, acoustic noise

I. INTRODUCTION

Switched reluctance motors (SRM) have a wide range of applications in industries, mainly due to the special properties of this motor. High reliability, high torque to power ratio, robust contracture, low cost, high efficiency, high torque in low speed are only a few advantages in SRM, because of these SRM has undergo rapid development in hybrid vehicles, aircrafts, starter/generator systems, washing machines and other automotive applications [1].

We know SRM torque ripple and radial force are the basic reasons of acoustic noise generation. Noise sources in electric machines are broadly categorized as magnetic, mechanical, electronic and aerodynamic as far as the production of direct airborne noise is concerned, the acoustic noise measured around an SRM is found to be directly correlated to the radial vibration of the stator yoke due to the radial magnetic force. Radial force produced by the magnetic flux crossing the air gap in an approximate radial direction excites various mode shapes [2].

Exciting the stator winding in the SRM with a constant voltage, will make radial force, lateral force, tangential force and torque. The current in the stator windings of the SRM could interact with the local magnetic field to produce a force on the windings. This force could excite winding vibrations that would emit acoustic noise. The winding vibrations could also excite stator vibrations that would emit acoustic noise [3].

However, if the SRM is already built, the acoustic noise reduction can be done by the optimization of drive parameters. Also, the radial force is the function of the phase current and the rotor position. Obtaining non-linear relations of the radial force with the phase current and for vibration and acoustic noise study [4].

Radial force is dependent on length of air gap, rotor position, rotor length and radius of rotor, square of phase current and number of turns in one phase of the machine. Also, sound power radiated by a SRM is the function of radial force square [5].

Review in the researches of a recent decade about of acoustic noise reduction SRM drive shows most of researchers worked on the torque ripple reduction while motor design parameters and radial magnetic force are effective at acoustic noise and SRM vibrations. In this paper, with keep of maximum torque, torque ripple and maximum radial force has been reduced.

II. CALCULATION OF SRM NOISE AND TORQUE

Radial force is dependent on the many parameters in the SRM and its drive. In the SRM, magnetic radial force is based on the assumption of independent pole pairs. Magnetic non-linearity, machine geometry and material properties are all taken into account while calculating the radial force. For more study on the radial force and acoustic noise both in the SRM design and drive parameters we need the fundamental equations and dimensions of SRM.

The Fig. 1 illustrates the various dimensions involved in the derivation of radial, tangential, and lateral forces.

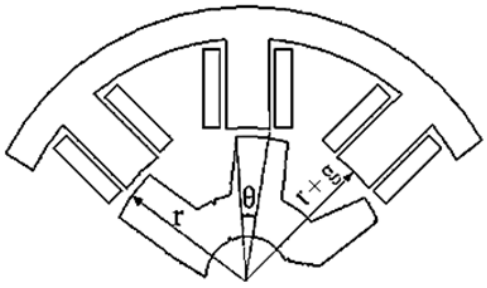


Fig. 1. Relevant dimensions of the SRM.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

The air gap flux density at a given stator and rotor pole overlap angle overlap angle θ , air gap l_g , and current i is given, is given as $B_g(\theta,i,l_g,L)$ [5].

Let *r* be the outer radius of the rotor, *L* be the stack length or iron length in the *z* direction, T_{ph} be the number of turns in one phase of the machine, *i* be the current in the winding, H_g be the magnetic field strength, φ be the flux, and μ_0 be the permeability of air. Then the following variables are derived as:

$$B_g(\theta,i,l_g,L)=\frac{\varphi}{Lr\theta}=\mu_0 H_g=\mu_0\frac{T_{ph}i}{l_g} \tag{1}$$

$$T_{ph}=\frac{l_g}{\mu_0 Lri}\frac{\varphi}{\theta} \tag{2}$$

The incremental electrical input energy (dW_e) is given by:

$$dW_e=id\Psi=id(T_{ph}.\varphi)=T_{ph}id\varphi=\frac{l_g}{\mu_0 Lr}\frac{\varphi}{\theta}d\varphi \tag{3}$$

The stored energy in the magnetic field (W_S) is:

$$W_S=\frac{l_g Lr\theta}{2\mu_0}B_g{}^2(\theta,i,l_g,L)=\frac{l_g}{2\mu_0 Lr}\frac{\varphi^2}{\theta} \tag{4}$$

where the term $l_g Lr\theta$ gives the volume of the air gap contained by the stator and rotor pole overlap. Note that $r\theta$ gives the arc of the overlapping region between the stator and rotor poles. The energy balance equation neglecting losses is:

$$dW_e=dW_S+dW_m \tag{5}$$

where dW_m is the incremental mechanical energy and dW_S is the incremental field energy. To calculate the tangential force, the incremental field energy is obtained from Eq. (4) as:

$$dW_S=-\frac{l_g}{2\mu_0 Lr}\frac{\varphi^2}{\theta^2}d\theta+\frac{l_g}{\mu_0 Lr}\frac{\varphi}{\theta}d\varphi \tag{6}$$

By substituting Eqs. (3) and (6) in Eq. (5), the incremental mechanical energy is obtained as:

$$dW_m=\frac{l_g}{2\mu_0 Lr}\frac{\varphi^2}{\theta^2}d\theta \tag{7}$$

Consequently, electromagnetic torque is obtained as:

$$T=\frac{dW_m}{d\theta}=\frac{l_g}{2\mu_0 Lr}\frac{\varphi^2}{\theta^2}=\frac{l_g Lr}{2\mu_0}B_g{}^2(\theta,i,l_g,L) \tag{8}$$

Also, the tangential force is obtained by dividing tangential torque by the radius of the rotor pole, yielding:

$$F_t=\frac{T}{r}=\frac{l_g L}{2\mu_0}B_g{}^2(\theta,i,l_g,L) \tag{9}$$

Similarly, the normal force, which is the radial force in the direction of the air gap, is obtained from considering the incremental field energy with respect to the air gap and the incremental mechanical energy as:

$$F_r=\frac{dW_m}{dl_g}=\frac{1}{2\mu_0 Lr}\frac{\varphi^2}{\theta}=\frac{Lr\theta}{2\mu_0}B_g{}^2(\theta,i,l_g,L) \tag{10}$$

And the lateral force can be derived as:

$$F_y=\frac{dW_m}{dL}=\frac{l_g r\theta}{2\mu_0}B_g{}^2(\theta,i,l_g,L) \tag{11}$$

The ratio between the radial and tangential force from Eqs. (9) and (10) is:

$$\frac{F_r}{F_t}=\frac{r\theta}{l_g} \tag{12}$$

Therefore, the radial force is always multiple times that of the tangential force in the SRM with substituting Eq. (1) in Eqs. (8) and (10) the torque and the radial force is obtained as:

$$T=\frac{\mu_0 rLT_{ph}{}^2}{2l_g}i^2 \tag{13}$$

$$F_r=\frac{\mu_0 Lr\theta T_{ph}{}^2}{2l_g{}^2}i^2 \tag{14}$$

It can be seen from Eq. (13) the torque proportional to square of current and inverse of air gap width. From Eq. (14) we find that the radial force proportional to square of current and square inverse of air gap width.

Sound power radiated by an electric machine can be expressed as [6]:

$$P=k\,F_{r\,per}{}^2 \tag{15}$$

Where *k* depend on the traveling speed of sound in the medium, the density of air, dimensions of the stator, the excitation frequencies and the equivalent stiffness and mass of stator, and $F_{r_{per}}$ is the Amplitude of radial force.

The equation (15) shows that sound power radiated is proportional to the square amplitude of radial force. Therefore, to decrease the sound power radiated we must find a way to decrease the radial force.

III. TORQUE RIPPLE

When motor is running and torque is not constant, we have torque ripple and noise. The expression for the torque ripple is [6]:

$$Torque\ \ Ripple=\frac{T_{max}-T_{min}}{T_{av}} \tag{16}$$

To noise reduction, we must decrease the torque ripple. The average torque is found out over that portion of the torque profile, which is actually utilized in the motor. In a 6/4 SRM, one phase will be excited for 30 degree or may be a little bit more. The authors would like to select that 30 degree where the torque is nearly constant. This choice will ensure maximum average torque with minimum ripple.

IV. EXPERIMENTAL SYSTEM

The experimental model of shaded pole induction motor was l-A, 220-V, 1/3-hp, 50-Hz, six-pole. For conversion of this motor to SRM, the following stages were done, respectively:

1- Bring out the squirrel-cage rotor and the shading coils.

2- Open the connections between coils.

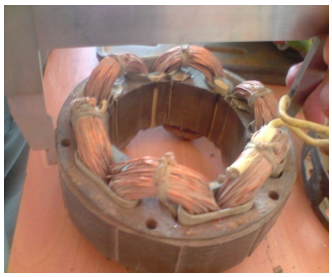
3- Manufacture the rotor by steel layer so that rotor outer diameter was equal to squirrel-cage rotor. This rotor had four poles, and its pole arcs were 2 degree greater than stator pole arcs.

4-Assemble rotor in the stator on the front and back plates.

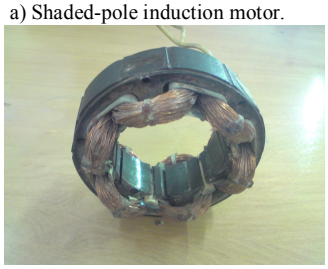
5-The coils of opposite poles were connected in series and then the coils interconnected to form a Y connection.

The stages of conversion shaded pole induction motor to SRM were shown in figure 2.

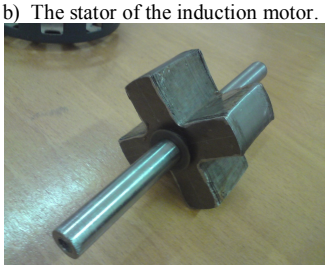
a) Shaded-pole induction motor.



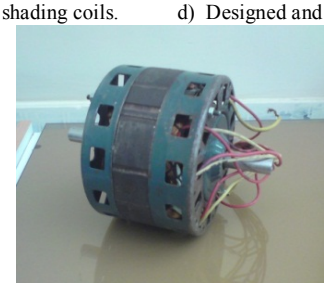
b) The stator of the induction motor.



c) Stator without shading coils.



d) Designed and manufactured rotor



e) Designed and manufactured SRM.

Fig. 2. Stages of conversion shaded pole induction motor to SRM

The SR motor is excited by a switching inverter under microprocessor control. Rotor position and phase currents are fed back to the microprocessor for control purposes. A personal computer provides a development environment for microprocessor programming and serves as the host for the microprocessor by downloading control programs and commanding their operation. Figure 3 shows the circuit of SRM drive.

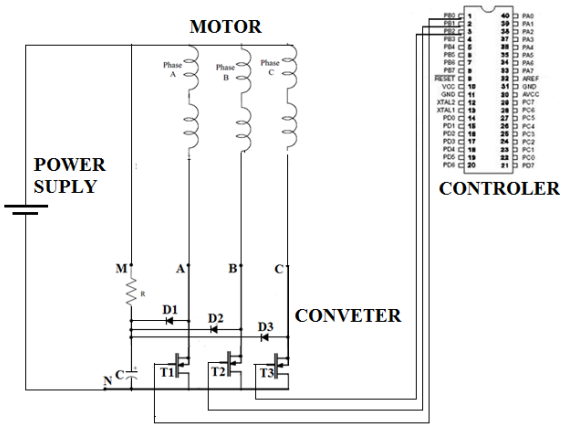


Fig. 3. Circuit of SRM drive

The current is measured in phase A during experiment using a current probe. Acoustic noise is measured in an anechoic chamber using sound level meter that is normally located to the side of the SRM. At different times, a microphone for detecting of the sound level is located elsewhere as described below; however, its distance from the motor center always remains at 20 cm. The results on a practical SRM confirmed some of the predictions with reasonable accuracy.

V. SRM SIMULATION BY USING FE METHOD

Table I shows dimensions of a three phase 6/4 SRM that has been used. The materials properties defined by using BH magnetizing curve have been showed in Fig. 4.

TABLE I
DIMENSIONS AND MATERIAL DATA OF THE ORIGINAL MOTOR

Number of stator poles	6
Number of rotor poles	4
Stator outer diameter	12.4 cm
Stator inner diameter	7.16 cm
Air gap length	0.04 cm
Number of turns/phase	180
Stator pole height	1.62 cm
Stator back iron thickness	1 cm
Stator pole arc	40 deg
Rotor pole arc	42 deg
Rotor outer diameter	7.12 cm
Rotor pole height	2.1 cm
Stator and rotor core material	DBII steel
Stack depth	3.74 cm

A Finite Element Analysis Software, ANSYS, has been used to build and solve nonlinear magnetic 2-D model of this SRM. A non-linear radial force and torque simulation model of the SRM is constructed based on the finite element model. Magnetic parameters of the SRM are calculated at different positions at different phase currents.

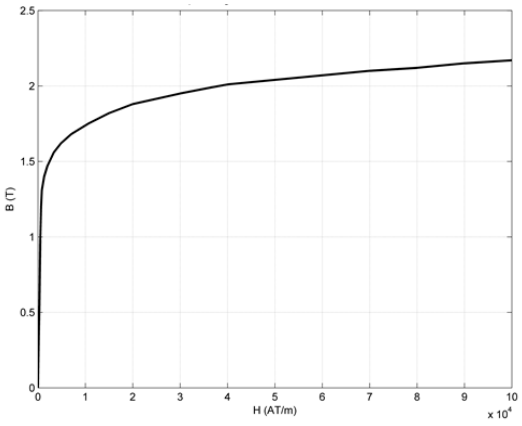


Fig.4. BH characteristic of DBII steel.

The effect of pole shoe and notch

Figures 6, 7, and 8 show the section of SR motor without pole shoe and notch, with pole shoe and without notch, and, with pole shoe and notch, respectively.

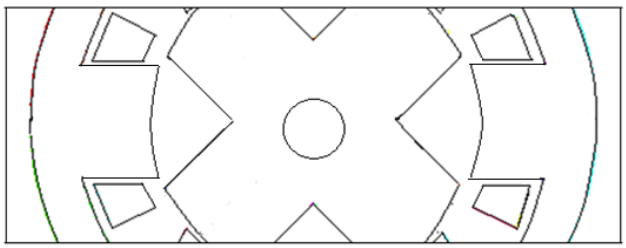


Fig. 6. Section of motor without pole shoe and notch

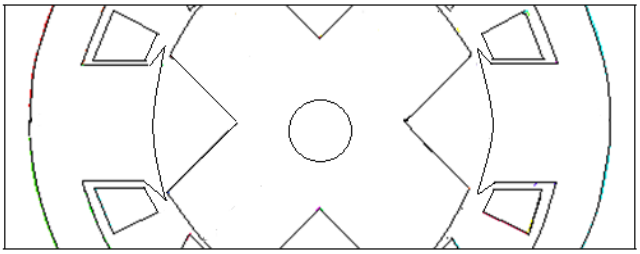


Fig. 7. Section of motor with pole shoe and without notch

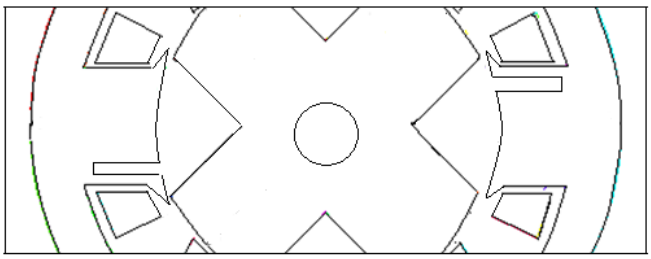


Fig. 8. Section of motor with pole shoe and notch

To figure out the effect of pole shoe and notch on the stator pole, the mentioned three models of SRM were simulated in ANSYS to obtain torque and radial force by keeping constant the phase current to the rated current. Figure 9, 10, and 11 show the flux density, torque, and radial force of SRM for various stator poles, respectively when phase current is equal to 2A.

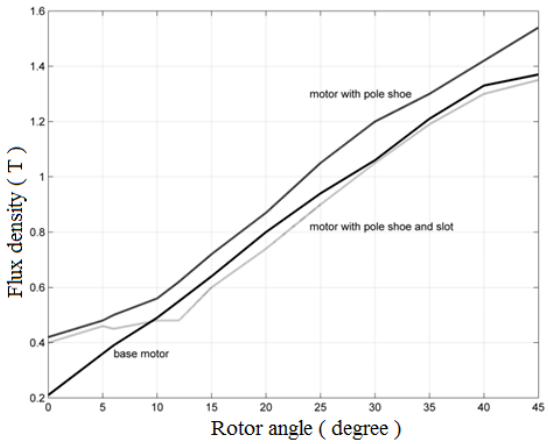


Fig. 9. Computed flux density profiles for various stator poles

These figures draw for half cycle, but the average torque and maximum radial force is found out over that portion of the torque profile which is actually utilized in the motor. One phase is excited for 30 degree in this SRM. The average torque and maximum radial force were found by changing the rotor angle which current in phase winding flow for 30 degree motor rotating. Table II shows the average torque, torque ripple, and maximum radial force for various rotor angles which current is in phase winding flow.

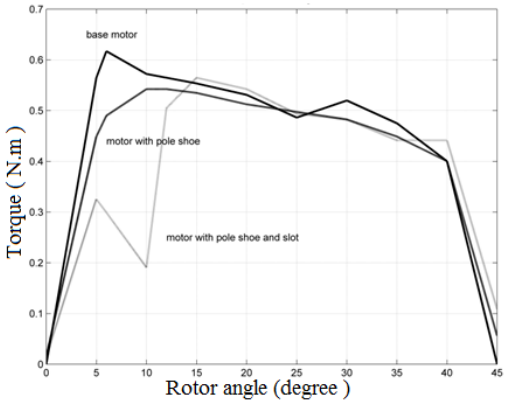


Fig. 10. Computed static torque profiles for various stator poles

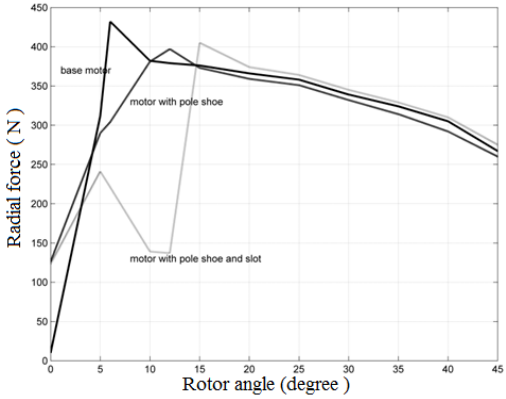


Fig. 11. Computed radial force profiles for various stator poles

TABLE II.
Maximum radial force, average torque and torque ripple for various rotor angles which current in phase winding flow

Θ (°)	T_{av} (N.m)			Torque Ripple			$F_{r\,max}$ (N)		
	*	**	***	*	**	***	*	**	***
0	0.499	0.466	0.408	1.24	1.13	1.35	432	397	405
5	0.535	0.505	0.457	0.27	0.18	0.82	432	397	405
6	0.531	0.504	0.461	0.30	0.20	0.81	432	397	405
10	0.509	0.491	0.487	0.34	0.29	0.77	382	397	405
12	0.492	0.477	0.489	0.66	0.59	0.52	379	397	405
15	0.448	0.439	0.456	1.23	1.08	1.00	376	373	405

* SRM without pole shoe and notch (base motor)
** SRM with pole shoe and without notch
*** SRM with pole shoe and notch

At this moment, we assume that the motor without pole shoe and notch is the base motor and T_{base} and Fr_{base} are torque and radial force of the base motor. Then, the percent of torque error, torque ripple error and radial force error are equal to:

$$\%\Delta T = \frac{T - T_{base}}{T_{base}} \times 100 \tag{17}$$

$$\%\Delta TR = \frac{TR - TR_{base}}{TR_{base}} \times 100 \tag{18}$$

$$\%\Delta F_r = \frac{F_r - F_{r\,base}}{F_{r\,base}} \times 100 \tag{19}$$

TABLE III.
Error of static torque, torque ripple, and radial force for various stator poles

Poles	Θ (°)	T (N.m)	ΔT (%)	TR	ΔTR (%)	F_r (N)	ΔF_r (%)
*	5	0.535	0	0.27	0	432	0
**	5	0.505	-5.6	0.18	-33	397	-8.10
***	12	0.489	-8.6	0.52	92	405	-6.25

Table III shows these errors for each motor on the rotor angle which average torque is maximum.

Table III also shows that in SRM with pole shoe and notch (***), torque decreased %8.6, torque ripple increased %92, but radial force decreased %6.25 compared with the base motor. In SRM with pole shoe and without notch (**), torque, torque ripple, and radial force decreased %5.6, %33, and %8.10, respectively compared with the base motor. Hence, SRM with pole shoe and without notch (**) will be a good motor to noise reduction. Improvement of SRM with pole shoe and notch will be accessible if the notch of stator is filled by iron powder core.

VI. CONCLUSION

In this paper, the effects of pole shoe and notch on the noise emission of SRM have been studied. The mentioned design was simulated with FEM software. In the case of the stator with pole shoe and notch, torque decreased %8.6 and torque ripple increased %92. However, in SRM with pole shoe and without notch, torque, torque ripple, and radial force decreased -%5.6, %33, and %8.10, respectively compared with the base motor.

REFERENCES

[1] Y. CUI, X. YU, H. FAN, J. FAN "The Simulation Study for Switched Reluctance Motor Drives Based on MATLAB6.5" , IEEE Proceeding of Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, PP. 1076-1081, 18 − 21th August 2005.

[2] M. N. Anwar, I. Husain "Radial force Calculation and Acoustic Noise Prediction in Switched Reluctance Machines," IEEE Transactions on Industry Applications, Vol. 36, No.6, PP. 1589-1597, November/December 2000.

[3] Derrick E. Cameron, Jeffrey H. Lang , and Stephen D. Umans ," The origin and reduction of acoustic noise in doubly salient variable-reluctance motors," IEEE Trans. on Industry Applications, Vol. 28, No. 6, pp.1250 – 1255, 1992

[4] J. Zhang, L. Wang, H. Zhang, R. Gao, " Non-linear Radial Force Simulation of Switched Reluctance Motors Based on Finite Element Model" IEEE International conference, Vol. 1, PP. 1678-1682, February 2009

[5] R. Krishnan, Switched Reluctance Motors Drives, (Modeling, Simulation, Analysis, Design, and Application), U.S.A., Florida, CRC Press LLC, 2001.

[6] M. Divandari, A. Dadpour, "Radial Force and Torque Ripple Optimization for Acoustic Noise Reduction of SRM Drives via Fuzzy Logic Control", IEEE International conference, PP. 1-6, Nov. 2010.

A Friendly Approach to Increasing the Frequency Response of Piezoelectric Generators

Sam Ben-Yaakov, Gil Hadar, Amit Shainkopf, Natan Krihely
Power Electronics Laboratory,
Department of Electrical and Computer Engineering
Ben-Gurion University of the Negev,
P.0.Box 653, Beer-Sheva 84105, ISRAEL

***Abstract* — A wide bandwidth piezoelectric generator (PZG) was constructed and tested experimentally. The PZG was characterized using chirp and wideband random excitations. The experimental results showed that by proper shaping of an attached cantilever beam it is possible to increase the number of vibration modes of the PZG and, hence, to improve the effectiveness of the energy harvesting as compared with a conventional cantilever PZG configuration. The proposed structure was designed to produce 3 vibration modes, which extend the operation around the frequencies 35Hz, 57Hz and 76Hz. As a result, the bandwidth was widened by a factor of 2.81 as compared with a conventional harvester.**

Keywords—Harvesting, piezoelectric devices, resonant power conversion.

I. INTRODUCTION

Ambient mechanical vibration energy is characterized by different frequency spectra. A piezoelectric harvester is a high Q resonant device that can effectively pick up ambient vibration energy only at or near a particular resonant frequency. This is also evident from the expression of the output average power Pavg as function of vibration frequency [1],

$$P_{avg} = \frac{m\zeta\delta^2(f/f_{res})^3}{[1-(f/f_{res})^2]^2+(2\zeta f/f_{res})^2} \quad (1)$$

where f_{res} is the resonant frequency and m, ζ and δ are parameters of the PZG. In applications where the frequency of ambient vibration varies periodically, such as in vehicles and with human motion, the PZG might not always be tuned to the resonance condition. As a result, the efficiency of a PZG with one fixed resonant mode at frequency f_{res} drops significantly (1).

Increasingly, efforts are being made to develop broadband energy harvesters that can harvest energy over a large frequency interval. Typically, this is achieved by electrically and/or mechanically connecting energy harvesters whose operating frequencies are slightly different from, but very close to each other [2]. However, these assembled generators must be carefully designed so that each individual generator does not effect the others. This makes such a configuration more complex to design and fabricate and the final volume of the device depends on the number of PZGs utilized. It has also been shown that bandwidth widening can be achieved using mechanical stoppers, nonlinear springs or bi-stable structures [3]-[5]. These strategies can be broadly classified as: 1) bandwidth widening along with increased output power, 2) bandwidth widening without power enhancement.

A conventional PZG harvester usually consists of piezoelectric cantilever beam (Fig. 1a) with or without a proof mass attached to the free end. The other end of the harvester is bonded to a vibrating base. All commercial PZGs exhibit this elementary configuration [6]. However, a limitation of this approach lies in the fact that the generator is, by definition, designed to work at a single frequency (1).

This study demonstrates a simple and friendly solution to extend the frequency response of the generator, based on the work presented in [7]. The behavior of energy harvesting systems subject to a random broadband excitation is less obvious than the classic sinusoidal case [8]. Therefore, the efficacy of the proposed configuration has been demonstrated under the effect of abnormal vibrations.

II. PROPOSED STRUCTURE

In many piezoelectric harvesting applications, a conventional PZG includes one or more piezoelectric transducer (PZT) patches that are attached to a flexible beam [9]. Such a concept underpins the conventional cantilever-based generator shown in Fig. 1a. As can be seen in the figure, an additional beam has been clamped to the PZG using a clamping support and screws. In our work, this rectangular beam-based harvester serves as a reference structure, denoted henceforth as “conventional PZG”. A commercial piezoelectric bimorph V20W (Mide Technology) [6] has been selected as the basic platform to be improved.

The proposed PZG depicted in Fig. 1b is targeted to the following key objectives [7]:

1) simplified configuration construction approach,
2) cost- and volume- effectiveness,
3) matching practical ambient vibration frequencies around 30Hz-200Hz [1].

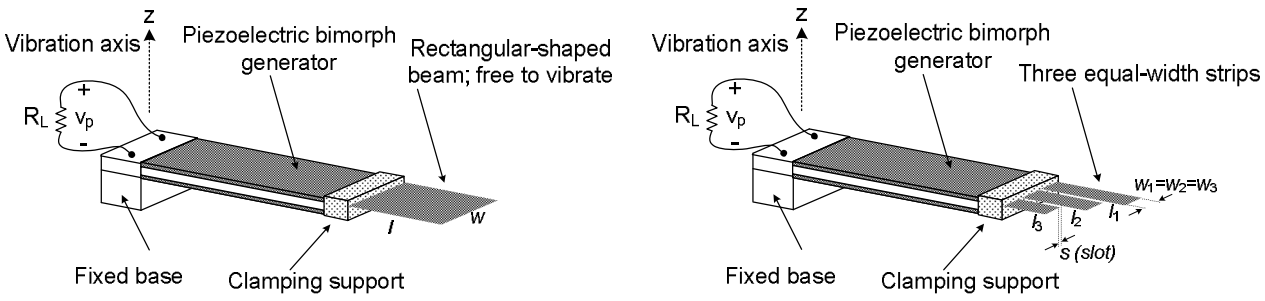


Fig. 1. Schematic of the piezoelectric generator with a cantilever beam: (a) rectangular beam, (b) proposed beam with three equal-width fingers.

TABLE I
STRUCTURAL PARAMETERS OF BOTH STRUCTERS

Conventional PZG		Proposed PZG	
Parameter	Value	Parameter	Value
Beam width	w=43mm	Finger width	w_1=w_2=w_3=13.5mm
Beam length	l=30mm	1th finger length	l_1=44mm
-		2nd finger length	l_2=30mm
-		3rd finger length	l_3=21mm
-		Slot	s=0.8mm
Beam area	$l \cdot w$=1290mm^2	Beam area	$l_1 \cdot w_1+l_2 \cdot w_2+l_3 \cdot w_3$=1282mm^2
Material	Brass	Material	Brass
Thickness	0.1mm	Thickness	0.1mm

It should be noted that the surface of the extra beam that extends beyond the bonded area of either structure was not covered by any PZT layers or patches. It is known that adding additional mass or, alternatively, extending the supporting beam, alters the frequency response of the system [10]. This causes a decrease in the working frequency, which is usually preferred for many practical applications with low-frequency ambient vibrations [1].

Subject to the volume criteria (objective 2, above), the dimensions of the attached beam were constrained to the dimensions of the hosting platform. The attached beam of the proposed configuration (Fig. 1b) has been shaped to include an array composed of three fingers of different lengths, l_1, l_2, l_3, but with the same width, w_1=w_2=w_3. The parameters l_1, l_2, l_3 describe the lengths extending out beyond the clamping support. A residual section of the beam penetrates between the clamping support and the given platform. Thereby, the three parts overlap for 10mm along the beam. A very small gap, or slot, s, has been introduced between the fingers to allow free movement. Apart from shifting down the fundamental resonant frequency of the original structure, the aim of this concept is to create new vibration modes due to the individual mechanical responses of each finger strip. The geometrical properties of the proposed configuration are summarized in Table I. In both structures, the two attached beams had approximately the same total surface area of 1290mm^2 and the same material thickness.

III. PERFORMANCE EVALUATION

In this section we investigate the performance of the proposed PZG. In the experimental setup, the energy harvesting systems were excited vertically by a LING laboratory shaker with a bandwidth of 9KHz (model V406). The input base acceleration level was monitored using an Analog Devices ADXL327 accelerometer with sensitivity and bandwidth of 462mV/g and 1.6KHz, respectively.

In the following study we conducted a comparison between the two considered configurations (Fig. 1) in terms of the frequency response and the average output power under chirp and random excitations.

A. Chirp excitation

The dominant flexural modes of each prototype were first determined by applying a chirp excitation signal over the frequency range 1-200Hz with a period of 10sec and by measuring the open circuit voltage output with respect to the 1MΩ input resistance of the oscilloscope. The amplitude of the acceleration was maintained at a fixed level of 0.45g. The time domain output voltage, v_p, of both structures was recorded (Fig. 2). It is evident from Fig. 2 that when the input frequency approaches a certain vibration mode (resonant frequency), the voltage increases significantly and then drops as the input frequency changes. Fig. 2a reveals that the prototype with a rectangular beam yields one dominant frequency at 58Hz. On the other hand, the configuration with the three finger beams produced locally maximal open circuit voltages at 35Hz, 57Hz and 76Hz (Fig. 2b). Hence, two additional dominant vibration modes were created as a result of splitting the beam into 3 different length fingers. This comprises the widening effect of the frequency response, which can be further controlled by adjusting the properties of the beam (length, shape, proportion etc.).

Let us define a local bandwidth as the frequency range from the first -3dBV occurrence up to the second one for each frequency mode. The bandwidth of the rectangular beam cantilever is found to be Δf_{res}=4.4Hz. Fig. 2b of the proposed structure reveals bandwidths of Δf_{r1}=3.8Hz, Δf_{r2}=4.1Hz, Δf_{r3}=4.5Hz.

$$\frac{\Delta f_{r1} + \Delta f_{r2} + \Delta f_{r3}}{\Delta f_{res}} \approx 2.81 \qquad (2)$$

Hence, (2) implies a bandwidth widening factor of 2.81 compared to the conventional case.

In addition, the input capacitance, C_p, of the PZG platform was measured to be 34.5nF by an LCR meter working at an operating frequency of 1KHz, (well above the dominant modes of the structures). It should be noted that the internal capacitance, C_p, is common to both structures because the attached beams are free of any PZT material.

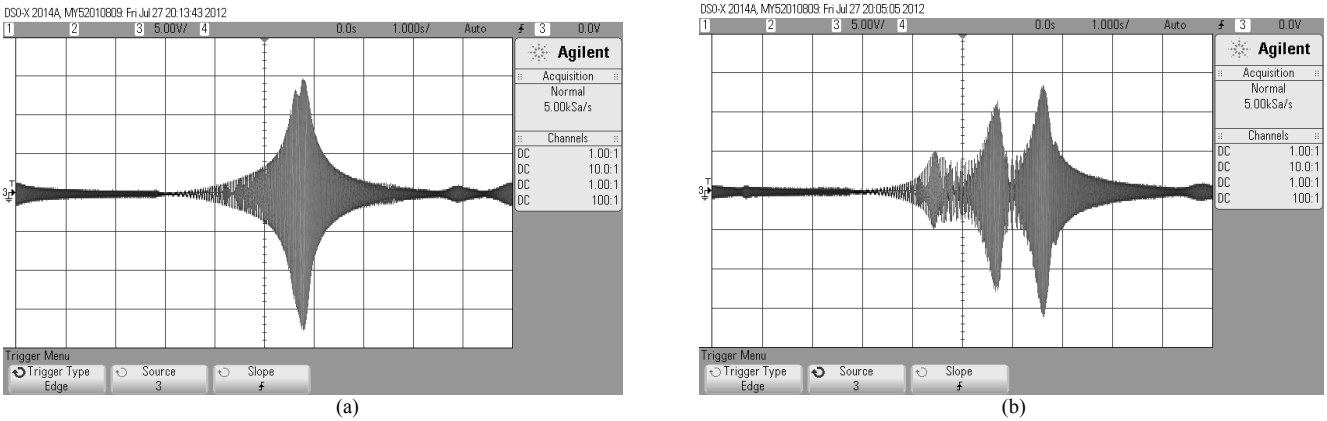
B. Random excitation

Most previously reported research has focused on an idealized single sinusoidal input, which is valuable for obtaining insight into the behavior of the system but is not sufficient for understanding how a real vibration energy harvesting device would behave when operated under broadband ambient vibration excitation. To this end, a filtered PSD (power spectral density) of a Gaussian white noise excitation signal with a 300Hz bandwidth was constructed in order to cover most of the sensitivity range of the generators. According to the experiment described in the previous sub-section, the useful frequency range of the two configurations was identified to exist around the 100Hz frequency band. The random driving force was generated by an Agilent 33220A arbitrary waveform generator. The generated signal was constructed with a time period of 10sec and a sampling frequency of 6.4KHz.

The average powers are derived from the steady-state time response in a 50sec interval (Fig. 3),

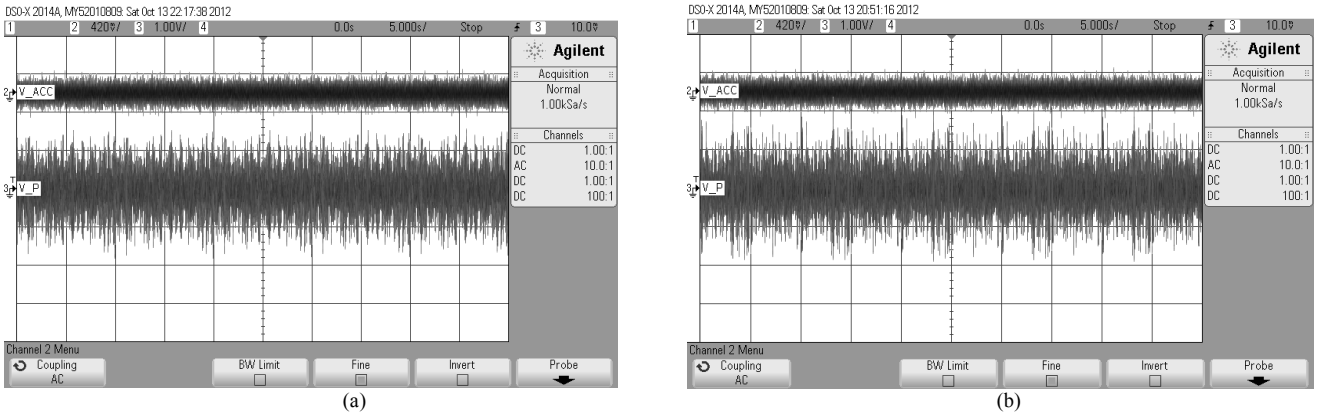
$$P_L = \frac{V_{p(rms)}^2}{R_L} = \frac{1}{R_L 50}\int_0^{50} v_p^2 dt \qquad (3)$$

where P_L, v_p, R_L and $V_{p(rms)}$ are, respectively, the average output power, the instantaneous output voltage of the generator, the output load and the rms value of v_p.



(a) (b)

Fig. 2. Experimentally obtained waveforms of the open circuit output voltages under chirp excitation between 1-200Hz with a fixed amplitude input acceleration of 0.45g; input resistance of the scope -1MΩ: (a) rectangular beam (b) proposed beam.



(a) (b)

Fig. 3. Time domain experimentally obtained waveforms of the output voltage, v_p, (lower trace) under random excitation at input acceleration of 0.145g (rms) and 74KΩ output load: (a) rectangular beam (b) proposed beam. The upper trace is the ADXL327 accelerometer output.

The rms value of the input base acceleration amplitude was also extracted for the same time interval as:

$$a_{acc(rms)} = \sqrt{\frac{1}{50}\int_0^{50} a_{acc}^2 dt} \quad (4)$$

where a_{acc} and $a_{acc(rms)}$ are the instantaneous base acceleration (Fig. 3 upper traces) and the rms value of a_{acc}, respectively. Fig. 3 shows the random fluctuations of a_{acc} and v_p versus time.

The measurement was performed at a fixed base acceleration of 0.145g (rms) for each configuration and repeated for different resistive loads, R_L. For each configuration, three successive measurements were made to verify consistency. In each measurement the supporting beam was released and clamped again. Fig. 4 depicts the average power of the two configurations against load resistances. As can be seen, the trends of the results are indeed similar. In particular, the resultant maximum output power of the conventional rectangular beam (P_{conv1}, P_{conv2}, P_{conv3}) was found to be ~4.6μW at an optimal load of R_L=74KΩ. While the proposed system tends towards the same peak average power of ~4.6μW, it is achieved at a slightly lower optimal load (P_{cut1}, P_{cut2}, P_{cut3}). This can be explained by the fact that the dominant power exists between the 57Hz and the 76Hz frequencies (Fig. 2b), which is higher than the resonant frequency of the conventional structure. The systems were also tested under a higher acceleration of 0.24g (rms). In this case, the behavior of the systems was consistent with the previous setup and yielded the same optimal load, but with a peak output power of ~12μW.

Fig. 5 depicts the normalized FFT spectrum of the random output voltage, v_p, time responses. It is evident that Fig. 5b shows three dominant voltage peaks, corresponding to the respective resonant modes arising from the three mechanical fingers.

Fig. 6 plots the statistical distribution of the proposed PZG output voltage, v_p, at an optimal load of $R_{L(opt)}$= 74KΩ with applied accelerations of 0.145g (rms) and 0.24g (rms). It is evident that the distribution exhibits normal behavior, characterized by a standard deviation of σ=0.5784Vrms and a zero mean, μ=0, for the 0.145g (rms) case, while for an excitation of 0.24g (rms) we get σ=0.9378Vrms and μ=0. Hence, we confirm that the system responds linearly to a Gaussian random excitation input. This is probably due to the fact that the excitation level was set low enough to induce a small signal linear equilibrium condition in the system. Using the statistical

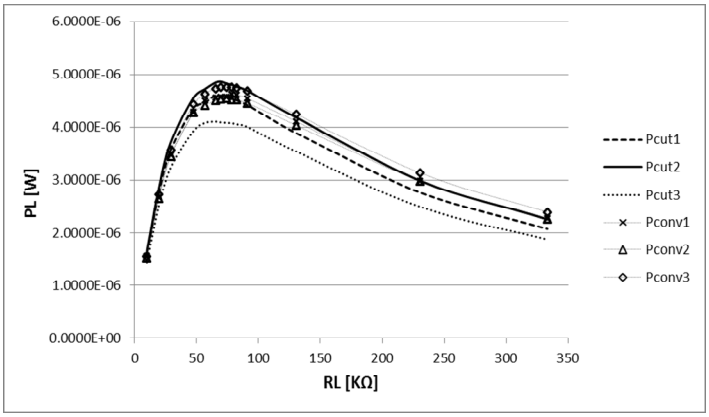
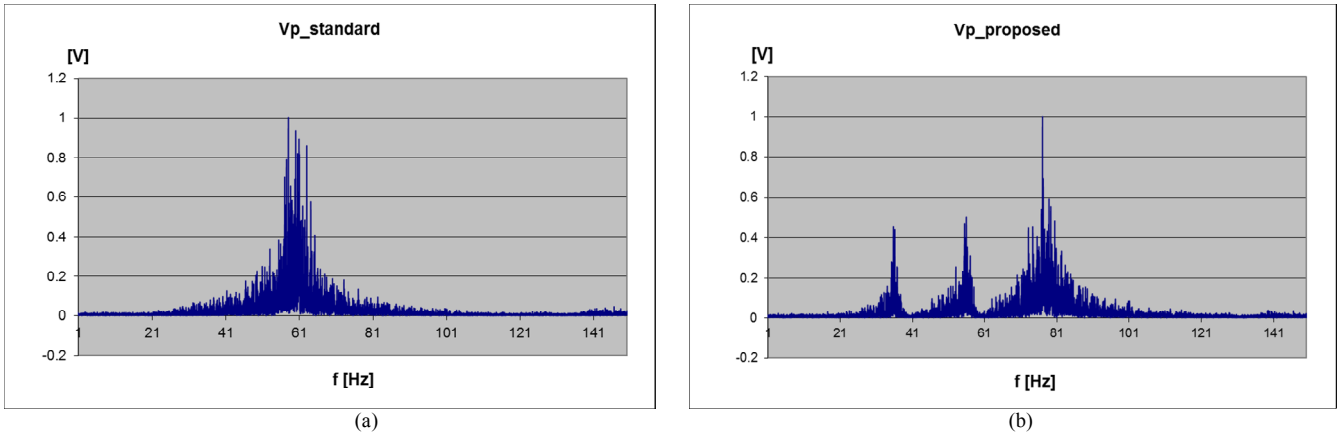


Fig. 4. Average output power vs. resistive load R_L for random excitation at a fixed input base acceleration of 0.145g (rms). P_{conv1}, P_{conv2} and P_{conv3} are the results of a three-measurement set carried out with the conventional configuration. P_{cut1}, P_{cut2} and P_{cut3} describe the three-measurement set results for the proposed generator. The experimental curves approach a maximum power point around R_L=74KΩ.



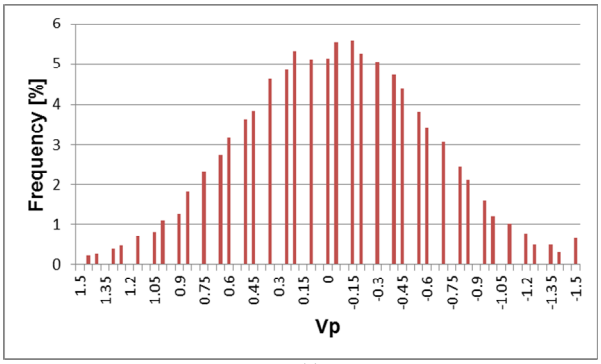
(a) (b)

Fig. 5. Normalized FFT spectrum of the output voltage, v_p, under random excitation and with the same parameters as used in Fig. 3: (a) rectangular beam (b) proposed beam. Vertical units are volts [V] and horizontal units are Hertz [Hz].

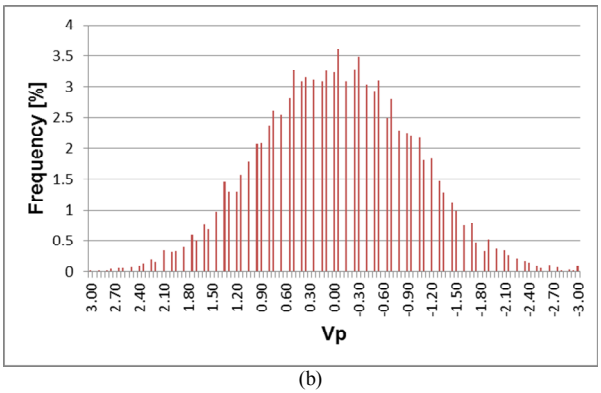
parameter σ we may calculate the average output powers through,

$$P_L = \frac{\sigma^2}{R_{L(opt)}} \quad (5)$$

and we obtain 4.52μW and 11.88μW, respectively. This, of course, matches the results obtained using the time domain averaged power over a 50sec interval as calculated by (3).



(a)



(b)

Fig. 6. Gaussian distribution of 50,000 time domain samples of the proposed generator output voltage, v_p, at an optimal load of 74KΩ. The standard deviation was found to be: (a) σ=0.5784Vrms @ $a_{acc(rms)}$=0.145g rms, (b) σ=0.9378Vrms @ $a_{acc(rms)}$=0.24g rms.

IV. Discussion and Conclusions

In this paper, the bandwidth and the output power behavior of two PZG structures are assessed experimentally under weak wideband excitation. In the proposed method, the bandwidth is widened by attaching a shaped beam composed of three strips of increasing length. Each strip (finger) was chosen with a different length and, consequently, several frequency modes were obtained. The proposed structure does, indeed, widen the frequency band as seen in Fig. 2b. However, the magnitude of each mode is decreased compared with the results from the structure with a rectangular beam (Fig. 2a). In addition, it was also found that for the same beam area both systems exhibit similar peak and average output power under the same applied rms acceleration level (Fig. 4).

There are more ways to realize additional resonant modes [3]. This method, however, utilizes a given PZG platform and only the parameters of the attached beam are changed. It has been shown that the resultant frequency modes are discrete, in accordance with the number of employed strips. In order to achieve a higher number of overlapping frequency modes one should carefully divide the beam into a higher number of narrow, equally spaced strips with appropriate proportions.

The harvesting concept is attractive because it is simple and cheap to construct and volume-effective. Specifically, the approach is valuable for shaping the frequency response of a commercial PZG before installing it into a time-varying environment. As previously mentioned, the average output power remained approximately the same as for the conventional PZG. Nevertheless, because in a realistic working environment the material properties of the PZG and the driving source tend to drift as a result of environmental changes such as temperature, the proposed solution can be utilized to broaden the bandwidth of the initial system by addressing the expected variations while continuing to provide useful output power.

It is challenging to process the low output levels associated with wideband vibrations using previously reported power converters [1], [2], [5], [8], [11], [12]. Future work will consider designing low-loss power conditioning circuits suitable for broadband weak excitations.

Acknowledgment

This research was supported by THE ISRAEL SCIENCE FOUNDATION (Grant No. 517/11).

References

[1] S. Roundy, P. K. Wright, and J. Rabaey, "A study of low level vibrations as a power source for wireless sensor nodes," *Comput. Commun*., vol. 26, no. 11, pp. 1131–1144, July 2003.

[2] I. C. Lien and Y. C. Shu, "Array of piezoelectric energy harvesting by the equivalent impedance approach," *Smart Materials and Structures*, vol. 21, no. 8, in press.

[3] D. Zhu, M. J. Tudor and S. P. Beeby, "Strategies for increasing the operating frequency range of vibration energy harvesters: a review," *Meas. Sci. Technol.*, vol. 21, (2010) pp. 1-29.

[4] H. Liu, et al., "Piezoelectric MEMS energy harvester for low-frequency vibrations with wideband operation range and steadily increased output power," *Journal of micromechanical systems*, vol. 20, no. 5, pp. 1131–1142, 2010.

[5] C. Luo and H. F. Hofmann, "Wideband energy harvesting for resonant piezoelectric devices," in *Proc. IEEE Energy conversion congress and exposition (ECCE)*, USA, 2010, pp. 4171–4178.

[6] Mide Technology Corporation, V20W piezoelectric vibration energy harvester, New-Volture data sheet.

[7] A. Shainkopf , G. Hadar and S. Ben-Yaakov, "Harvesting system for a piezoelectric generator", final project report in Hebrew 2011.

[8] L. J. Blystad, E. Halvorsen and S. Husa, "Piezoelectric MEMS energy harvesting systems driven by harmonic and random vibrations," *IEEE Trans. Ultrason. Ferroelect., Freq. Control.*, vol. 57, no. 4, pp. 908–919, April 2011.

[9] S. Mehraeen, S. Jagannathan and K. A. Corzine, "Energy harvesting from vibration with alternate scavenging circuitry and tapered cantilever beam," *IEEE Trans. Ind. Electron.*, vol. 57, no. 3, pp. 820-830, March 2010.

[10] W. G. Li, S. He and S. Yu, "Improving power density of a cantilever piezoelectric power harvester through a curved L-shaped proof mass," *IEEE Trans. Ind. Electron.*, vol. 57, no. 3, pp. 868-876, March 2010.

[11] N. Krihely and S. Ben-Yaakov, "Piezoelectric harvesting circuit with extended input voltage range," in *Proc. IEEE 26 Convention of Electrical and Electronics Engineers in Israel (IEEEI)*, Eilat, Israel, 2010, pp. 670–674.

[12] A. Tabesh and L. G. Frechette, "A Low-Power Stand-Alone Adaptive Circuit for Harvesting Energy From a Piezoelectric Micropower Generator," *IEEE Trans. Ind. Electron.*, vol. 57, no. 3, pp. 840-849, March 2010.

A Novel Fuzzy Dynamic Observer for High Speed BLDC Motor

M. Divandari, F. Shahsavari, M. Rezaeimoghadam
Department of Electrical Engineering,
Aliabad Katoul branch,Islamic Azad University,
Aliabad Katoul, Iran
mohammad.divandari@gmail.com,
fashahsavari@aol.com, m.rezaie.moghaddam@gmail.com

A.Dadpour
Golestan University,
Gorgan, Iran
ahmad.dadpour@yahoo.com,

***Abstract*—In this paper, a high performance brushless DC (BLDC) motor drive based on a fuzzy dynamic observer (FDO) is investigated. The FDO acts on the motor current and its gains are corrected by estimating current, rotor position and speed by fuzzy logic control (FLC). FLC is correcting gain's FDO via real time. A PI speed control was chosen due to its low processing time and fast control. In order to reduce the model complexity, the back-EMF is assumed as being trapezoidal in a simplified machine model. The presented drive has been simulated by the MATLAB/SIMULINK software on the high speed BLDC motor model. Simulation results show that the proposed drive is able to estimate the rotor position and speed with high precision when high speeds are considered. Simulation results also show the reliability, fast computation and excellent dynamic performance with using fuzzy logic for high speed BLDC motor.**

Keywords — BLDC motor, fuzzy dynamic observer, fuzzy logic.

I. INTRODUCTION

Permanent-magnet (PM) motor drives have been widely employed in industrial applications in recent years. PM motor with trapezoidal back-EMF shape for instance the brushless direct current motor has several benefits such as: high power density, efficiency, power factor, silent operation, compact form, reliability, low maintenance, and better controllability. These characteristics justify the BLDC rapid development in many industrial applications and instrumentation fields [1,2].

In order to commutate the phase currents in the correct order the BLDC motor needs a rotor position detector. A magnetic position sensor is widely used to detect the flux position leading to a reliability decrement and a limitation on the motor coupling with load [3,4]. A sensor-less drive decreases costs and increases reliability by extracting the shaft position information indirectly from electric signals of motor terminals.

As the motor dynamics depends on the current control, B. Lee et al. [7] proposed a direct current controlled PWM scheme to produce the desired dynamic and static speed-torque characteristics in a low cost four-switch BLDC motor drive. A control strategy based on a single current sensor is proposed by C. Xia et al. [8] for a four-switch three-phase BLDC motor system to lower costs and improve performance. The proposed strategy showed satisfactory self-adapted track ability with low current ripple and strong robustness to the given speed reference model.

Also, a new starting method without position sensor by Wook-Jin Lee and et al. [9] presented. It was based on the back electromotive force could work reasonably. The principle of the estimation was described on the variation of the current response caused by the magnetic saturation of the stator core of the BLDC motor when the current flows along the magnetic axis.

The sliding mode observer method has inherent robustness to face parameter uncertainties. Its computational simplicity and high stability provide a powerful approach to implement sensor-less schemes [5]. As observers are model-based, variation on model parameters will increase estimation errors. An approach using online dynamic gain corrections can increase the observer speed and reduce errors.

ost observers for sensor-less SRM drive which have so far been expressed were static, meaning once their gains have been regulated. Since observers are based model and model parameters variable, will increase estimation errors. Dynamic gains are able to conceal their difficulties. That is to say, online gain correction, is one significant progress has been made in based model observer. This new approach will increase observer speed and precision notify and it can make a revolution for all observers having regulation gains. This new procedure will recommend a new generation of observer named fuzzy dynamic observers (FDO). FDOs can be used in tracking and drive systems.

FDO by fuzzy logic control is proposed for a nonlinear model of BLDC motor. It has been developed to estimate the rotor position at high speeds. The FDO estimates the rotor position and speed using the combination of three-phase current sliding surfaces.

The organization of this paper is as follows. First, mathematical non-linear model of BLDC motor and switching sequence are presented in Section II. Next, the proposed sensor-less control drive based on FDO by fuzzy logic control is described in detail in Section III. The simulation results with speed control by FDO are given to verify the effectiveness of the proposed sensor-less control method in Section IV. Finally, some concluding remarks are given in Section V.

II. MATHEMATICAL NONLINEAR MODEL OF BLDC MOTOR

Modeling of a three-phase BLDC motor can be developed like in what in the three-phase synchronous machine is operated, by employing a constant excitation field [6]. The three-phase BLDC motor is worked in a two-phase-on style. By using two phases which to be energized depends on the shaft position. As it can be seen in Fig.1 the signal from position sensors produce a change in every 60º (H1, H2, and

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

H3). Also, the ideal current and PM back-EMF waveforms is shown in this Fig. 1.

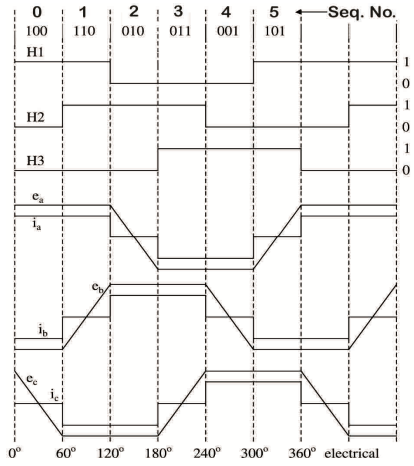


Fig. 1. Ideal back-EMF, phase currents, and position sensor signals.

In this study, the BLDC motor model was based on a star connection reducing the electronic costs is investigated.

For maximizing the torque, the current needs to be introduced in the stator phase winding during the interval with maximum PM back-FEM. In a three phase Y-connected winding, this interval lasts 120º starting at the moment 60º, before the rotor and the stator field lines are perpendicular. As it can be seen in table I we summarize the controller logic. Fig. 1 and 2 that shows the currents in each interval represented in and a cross section of the three phase star-connected motor respectively.

TABLE I

SWITCHING SEQUENCE	Seq.	Pos. sensor			Phase current		
Switching interval	No.	H1	H2	H3	A	B	C
0º - 60º	0	1	0	0	+	-	OFF
60º - 120º	1	1	1	0	+	OFF	-
120º - 180º	2	0	1	0	OFF	+	-
180º - 240º	3	0	1	1	-	+	OFF
240º - 300º	4	0	0	1	-	OFF	+
300º - 360º	5	1	0	1	OFF	-	+

The BLDC motor can be expressed by the following four equations:

$$v_{ab} = R(i_a - i_b) + L\frac{d}{dt}(i_a - i_b) + e_a - e_b \quad (1)$$

$$v_{bc} = R(i_b - i_c) + L\frac{d}{dt}(i_b - i_c) + e_b - e_c \quad (2)$$

$$v_{ca} = R(i_c - i_a) + L\frac{d}{dt}(i_c - i_a) + e_c - e_a \quad (3)$$

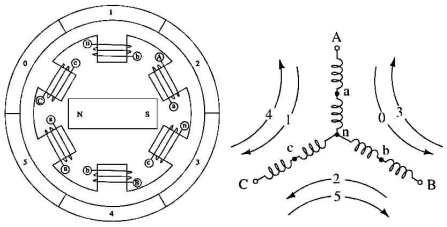


Fig. 2. BLDC motor cross section and phase energizing sequence.

where the symbols *v, i* and *e* are the phase to phase voltages, phase currents and phase back-EMF in the three phase *a, b* and *c,* respectively. The resistance *R* and the inductance *L* are per phase values. The back-EMF and electrical torqueses are:

$$e_a = \frac{k_e}{2}\omega_m F(\theta_e) \quad (4)$$

$$e_b = \frac{k_e}{2}\omega_m F(\theta_e - \frac{2\pi}{3}) \quad (5)$$

$$e_c = \frac{k_e}{2}\omega_m F(\theta_e - \frac{4\pi}{3}) \quad (6)$$

The function F(.) which gives the trapezoidal waveform of the back-EMF can be represented as:

$$F(\theta_e) = \begin{cases} 1, & 0 \le \theta_e < \frac{2\pi}{3} \\ 1 - \frac{6}{\pi}(\theta_e - \frac{2\pi}{3}) & \frac{2\pi}{3} \le \theta_e < \pi \\ -1 & \pi \le \theta_e < \frac{5\pi}{3} \\ 1 - \frac{6}{\pi}(\theta_e - \frac{5\pi}{3}) & \frac{5\pi}{3} \le \theta_e < 2\pi \end{cases} \quad (7)$$

Torque equations are:

$$T_e = \frac{k_t}{2}[F(\theta_e)i_a + F(\theta_e - \frac{2\pi}{3})i_b + F(\theta_e - \frac{4\pi}{3})i_c \quad (8)$$

$$T_e - T_L = k_f\omega_m + J\frac{d\omega_m}{dt} \quad (9)$$

where T_e and T_L are the electromagnetic torque and the load torque respectively. J is the rotor inertia, k_f is a motor friction constant and ω_m is the rotor speed . And k_e and k_t are the back-EMF and the torque constants. The electrical angle, θ_e, is equal to the rotor angle times the number of pole pairs ($\theta_e = \theta_m . p/2$) for a better implementation in MATLAB/SIMULINK, equations (1, 2, 3 and 9) must be written in state space form.

These equations can be modified to allow a simpler state space representation. Only two equations are needed since each voltage equation is a linear combination of the other two voltages. The same applies to current since

$$i_a + i_b + i_c = 0 \quad (10)$$

Thus, the voltage equation becomes

$$v_{ab} = R(i_a - i_b) + L\frac{d}{dt}(i_a - i_b) + e_a - e_b \quad (11)$$

$$v_{bc} = R(i_a + 2i_b) + L\frac{d}{dt}(i_a + 2i_b) + e_b - e_c \quad (12)$$

and finally, the complete model via state space is obtained as

$$\begin{pmatrix} \dot{i}a \\ \dot{i}b \\ \dot{\omega}m \\ \dot{\theta}_m \end{pmatrix} = \begin{pmatrix} \frac{-R}{L} & 0 & 0 & 0 \\ 0 & \frac{-R}{L} & 0 & 0 \\ 0 & 0 & \frac{-k_f}{J} & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} ia \\ i_b \\ \omega_m \\ \theta_m \end{pmatrix} + \begin{pmatrix} \frac{2}{3L} & \frac{1}{3L} & 0 \\ \frac{-1}{3L} & \frac{1}{3L} & 0 \\ 0 & 0 & \frac{1}{J} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} v_{ab} - e_{ab} \\ v_{bc} - e_{bc} \\ T_e - T_L \end{pmatrix} \tag{13}$$

$$\begin{pmatrix} i_a \\ i_b \\ i_c \\ \omega_m \\ \theta_m \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} i_a \\ i_b \\ \omega_m \\ \theta_m \end{pmatrix} \tag{14}$$

Figures 3 show block diagram of the simulated BLDC speed control.

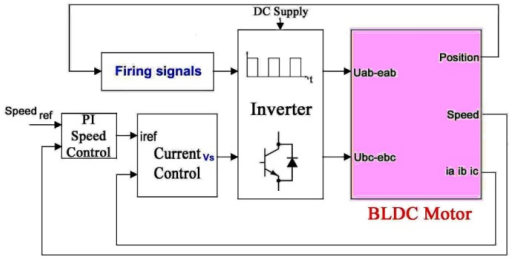


Fig. 3. Block diagram of speed control of a three-phase BLDC motor.

III. Fuzzy Dynamic Observer (FDO)

A speed and shaft position fuzzy dynamic observer (FDO) for BLDC motor was constructed based on the measurements of phase voltage and current. Fig. 4 shows block diagrams of simulated FDO. The difference between the actual phase currents and the estimated ones are used to calculate shaft speed and position. According to the system, differential equations were derived from the nonlinear model of the BLDC.

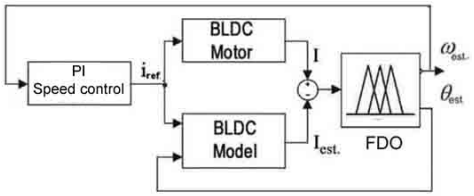


Fig. 4. Block diagram of FDO.

Sliding surface is represented as:

$$S(t) = i_j - i_{j\,est.} \tag{15}$$

$$S_{cont.} = \text{sgn}. \sum_{j=1}^{n} S(t) \tag{16}$$

Differential equations of FDO become:

$$\dot{\theta}_{est.} = \omega_{est.} + \alpha \ .S_{cont.} \tag{17}$$

$$\dot{\omega}_{est.} = T_{est.} + \beta . S_{cont.} \tag{18}$$

The errors is defined as

$$e_\theta = \theta_{est.} - \theta, e_\omega = \omega_{est.} - \omega \tag{19}$$

$$\dot{\omega} = \dot{\theta} \tag{20}$$

Consequently, by subtracting Eq. (17) from (20) and Eq. (18) from (9), the error dynamics were obtained

$$\dot{e}_\theta = e_\omega + \alpha . S_{cont.} \tag{21}$$

$$\dot{e}_\omega = -\dot{\omega} + \dot{T}_{est.} + \beta . S_{cont.} \tag{22}$$

$$\alpha = \alpha_1 + \alpha_{e\theta} \tag{23}$$

$$\beta = \beta_1 + \beta_{e\omega} \tag{24}$$

By choosing the α_1, β_1, $\alpha_{e\theta}$, and $\beta_{e\omega}$ can make,

$$\dot{e}_\theta e\theta \langle 0 \Rightarrow (e_\theta \Rightarrow 0) \Rightarrow (\theta_{est.} \Rightarrow \theta) \tag{25}$$

$$\dot{e}_\omega e\omega \langle 0 \Rightarrow (e_\omega \Rightarrow 0) \Rightarrow (\omega_{est.} \Rightarrow \omega) \tag{26}$$

α_1, β_1 and $\alpha_{e\theta}$, $\beta_{e\omega}$ are static and dynamic gains, respectively. Fig. 5 shows the block diagram of FDO. First we set α_1 and β_1, but $\alpha_{e\theta}$, $\beta_{e\omega}$ will be correct due to the position and speed error value by fuzzy logic control. Fundamental of gain correction shows in Fig. 6. Fuzzy logic gain correction acts on the $\alpha_{e\theta}$ and $\beta_{e\omega}$ with position error and also speed error. Table II shows the fuzzy logic rules. As it can be seen in the Fig. 7, Fuzzy logic gain correction has two inputs and two outputs. Inputs are position error and speed error, and outputs are micro gains $\alpha_{e\theta}$ and $\beta_{e\omega}$.Due to the fast computation we used of three level PS, Z, NS (positive small, zero, negative small) for all of inputs and outputs.

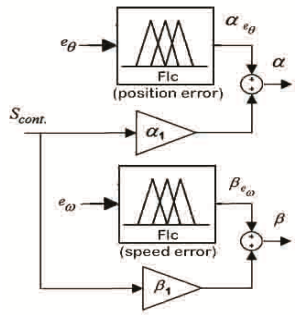


Fig. 5. Block diagram of dynamic gains using fuzzy logic.

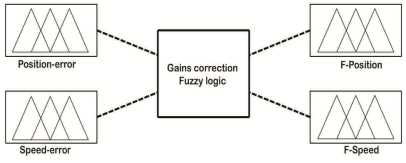


Fig. 6. Fundamental of gains correction.

TABLE II
FUZZY LOGIC RULES

$\frac{F-position}{F-speed}$	Speed (Error)			
		NS	Z	PS
Position (Error)	NS	$\frac{NS}{NS}$	$\frac{NS}{Z}$	$\frac{NS}{PS}$
	Z	$\frac{Z}{NS}$	$\frac{Z}{Z}$	$\frac{Z}{PS}$
	PS	$\frac{PS}{NS}$	$\frac{PS}{Z}$	$\frac{PS}{PS}$

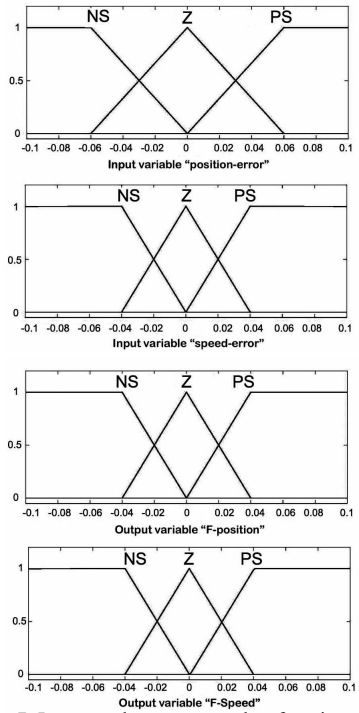


Fig. 7. Inputs and outputs member functions.

IV. Simulation Results

The FDO has been run by MATLAB/ SIMULINK, where the parameters of the high speed BLDC motor have been listed in table III.

TABLE III
BLDC MOTOR PARAMETERS

P=2	Number of poles
W=1.2	Nominal Power(W)
Vs=6	Nominal voltage(V)
N=40000	No load speed(RPM)
R=12.50	Terminal resistance phase to phase(Ohm)
L=0.091e-3	Terminal inductance phase to phase(H)
In=.5	No load current(A)
ke=1.05e-3	Back-EMF constant (V/(rad/s))
kt=1.05e-3	Torque constant (N.m/A)
J=0.005e-7	Rotor inertia (kg.m^2)
Kf=1.38e-8	Friction constant N/(rad/s)

A. Speed control BLDC motor

Fig.8 demonstrates the starting process of the BLDC motor. Also Fig. 8 shows that the setting time of speed is 0.01 seconds. Consequently after the motor's start up, the speed is kept constant at N=30,000 rpm.

Figures 9 and 10 show back-EMFs three-phases and total torque in 0.1 second.

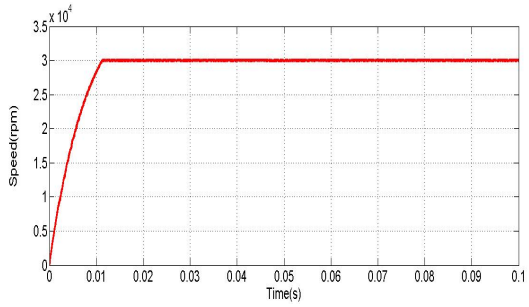


Fig. 8. Speed of rotor (N=30,000 rpm).

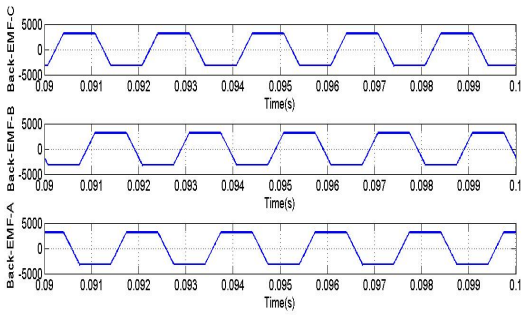


Fig. 9. Back EMF three-phases (N=30,000 rpm).

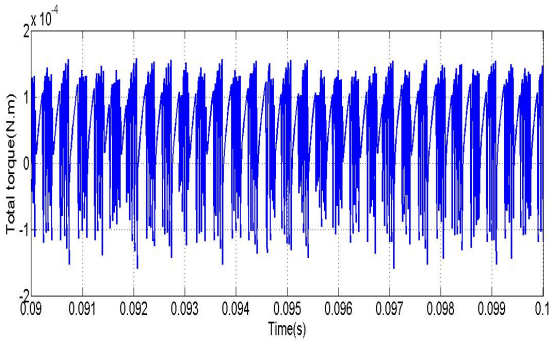


Fig. 10. Total torque (N=30,000 rpm).

B. Speed control BLDC motor with FDO

By using FDO the speed control BLDC motors at speed of 30,000 rpm has been simulated. Figs.11, 12 and 13 demonstrate the speed, estimated speed, rotor position, estimated position, and speed estimated error, respectively. The results shows that the FDO has a better predict in speed and position with high precision for high speeds.

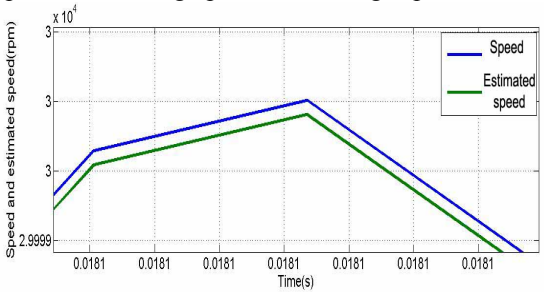


Fig. 11. Speed and estimated speed with FDO (N=30,000 rpm).

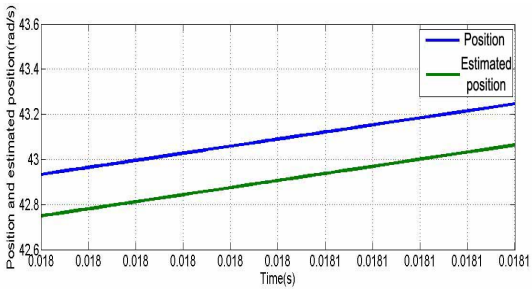


Fig. 12. Rotor position and estimated position with FDO (N=30,000 rpm).

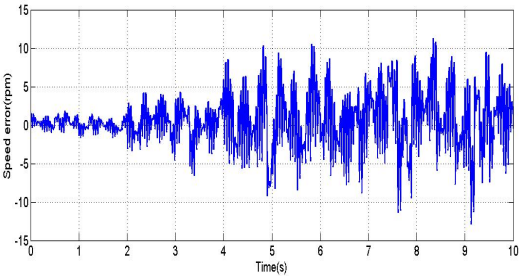


Fig. 13. Speed estimated error with FDO (N=5,000 rpm, N=10,000 rpm and N=30,000 rpm).

V. CONCLUSION

In this study, a fuzzy dynamic observer (FDO) scheme for a high speed BLDC motor drive is investigated. FDO micro gains are regulation by using fuzzy logic control (FLC) via real time. This approach has been simulated on a high speed BLDC motor nonlinear model. The FDO acts on the phase currents and also the micro gains will be quickly regulated real time according to error values by FLC. Also, the use of PI speed control accelerates the calculations of the rotor position estimation and speed. In this study, simulation results show that FDO are suitable for high speed BLDC motors and torque ripple is one of the indirect factors affecting the errors estimation. Nevertheless, these results show that the FDO is suitable for high speeds. In addition, as torque ripple is one of the main estimation error this parameter can be decrease by the torque ripple optimization.

REFERENCES

[1] Lei Hao, H. A. Toliat, "BLDC Motor Full-Speed Operating Using hybrid Sliding Mode Observer "Applied Power Electronics Conference and Exposition, APEC '03. Eighteenth Annual IEEE, vol. 1, pp. 286 - 293, February 2003.

[2] S. M. M. Mirtalaei, J. S. Moghani, K. Malekian, B. Abdi, "A Novel Sensorless Control Strategy for BLDC Motor Drives Using a Fuzzy Logic-based Neural Network Observer "International Symposium on Power Electronics, Electrical Drives, Automation and Motion, SPEEDAM 2008. IEEE, vol. 1, pp. 1491 - 1496, July 2008.

[3] Li Qiang, W. Ruixia, "Study on Rotor Position Detection Error in Sensorless BLDC Motor Drives "5th International Conference Power Electronics and Motion Control, IPEMC 2006. IEEE, vol. 3, pp. 1-5, August 2006.

[4] J. Lee, S. Sathiakumar, Y. Shrivastava, "A novel speed and position estimation of the brushless DC motor at low speed "Power Engineering Conference, AUPEC 2008. IEEE, vol. 3, pp. 1-6, December 2008.

[5] M. Divandari. R. Brazamini, A. Dadpour, M. Jazaeri, "A Novel Dynamic Observer and Torque Ripple Minimization via Fuzzy Logic for SRM Drives "International Symposium on Industrial Electronics, ISIE 2009. IEEE, vol. 1, pp. 847 - 852, July 2009.

[6] S. Poonsawat and T. Kulworawanichpong, "Speed Regulation of a Small BLDC Motor Using Genetic-Based Proportional Control "WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY, vol. 30, pp. 849-854, July 2008.

[7] Byoung-Kuk Lee Tae-Hyung Kim , M. Ehsani, "On the feasibility of four-switch three-phase BLDC motor drives for low cost commercial applications: topology and control " IEEE Transactions on Power Electronics, vol. 18, pp. 164, Jan. 2003.

[8] Changliang Xia Zhiqiang Li Tingna Shi, "A Control Strategy for Four-Switch Three-Phase Brushless DC Motor Using Single Current Sensor " IEEE Transactions on Industrial Electronics, vol. 56, pp. 2058, Jun. 2009.

[9] Wook-Jin Lee, and Seung-Ki Sul, "A New Starting Method of BLDC MotorsWithout Position Sensor", IEEE Transaction on Industry Applications, vol. 42, no. 6, Dec. 2006.

Minimization of Cable Losses in Three-Phase Four-Wire Systems by means of Instantaneous Compensation with Shunt Active Filters

M.Yu. Artemenko, L.M. Batrak
Department of Industrial Electronics, National Technical University of Ukraine "Kyiv Polytechnic Institute"
Kyiv, Ukraine
artemenko_m_ju@ukr.net

S.Y. Polishchuk, V.M. Mykhalskyi, I.A. Shapoval
Institute of Electrodynamics of the National Academy of Sciences of Ukraine
Kyiv, Ukraine
mikhalsky@ied.org.ua

***Abstract*—The three-phase four-wire network with the presence of a zero-sequence component in the reference voltage vector is considered. The feature of considered network is the significant difference between resistivity of the neutral wire and resistivity of phase wires. The control strategy of shunt compensators based on forming the network current with partial attenuation of zero-sequence component is proposed to obtain the minimum of power losses in the four-wire cables.**

Keywords—control strategy; instantaneous compensation; zero-sequence component; cable active power losses

I. INTRODUCTION

The growth in number of nonlinear consumers leads to the increasing of harmonic distortions of currents and voltages in three-phase electrical power networks caused by the reactive currents and the active power losses in four-wire networks. To reduce the power losses and to improve the quality of currents the special means are needed.

Instantaneous power theory developed in [1]–[5] allows to devise efficient algorithms for instantaneous reactive power compensation and related reactive components of currents in three-phase four-wire networks using active filtering. However, the unbalanced phase voltages lead to the appearance of the zero-sequence component increasing the active power losses. To minimize of power losses two main control strategies [6, 7] of shunt compensators are using at present time.

According to the first strategy the instantaneous reactive current is maximally compensated in each phase wires. However it does not exclude the current flow in the neutral wire. This control strategy is based on the forming of instantaneous vector current that is proportional to instantaneous phase voltage vector [4]. The second control strategy [6] uses the vector of instantaneous phase voltages without zero-sequence component to form the network currents. It allows the removing of current completely from the neutral wire. As it is shown in [7] the total losses of instantaneous active power in four-wire system are smaller in this case than for the first control strategy. The similar results are obtained by using of the average value of active power losses on the period interval [7].

The current vector norm of the consumed current is used as a criterion of power efficiency for the parallel active filtering in three phase network. This criterion is usually applied to power supply with three-wire network to provide the minimization of power losses. The neutral wire makes a significant difference from the three-wire network due to the presence of a zero-sequence current in the reference voltage vector. Therefore, minimization of energy losses in the four-wire cable can be the energy efficiency criterion of parallel active filtering in the case of three-phase four-wire network.

The control strategy based on forming the network current proportional to the instantaneous phase voltage vector with partial attenuation of zero-sequence component during the forming of network current is proposed in this paper. In comparison with the above mentioned strategies it gives the advantage in power losses for the four-wire system.

II. THEORETICAL BACKGROUND

The analysis of active power losses in the three-phase four-wire network, load and reactive power compensator (Fig. 1) is performed based on the next assuming: all the instantaneous values of current and voltage are periodic with period T of the network voltage; the voltage drop across all network wires is significantly less than the network voltage; the resistances of phase cable conductors are identical and equal to r, resistance of the neutral wire is equal to r_n, other resistances and losses of active power will be neglected.

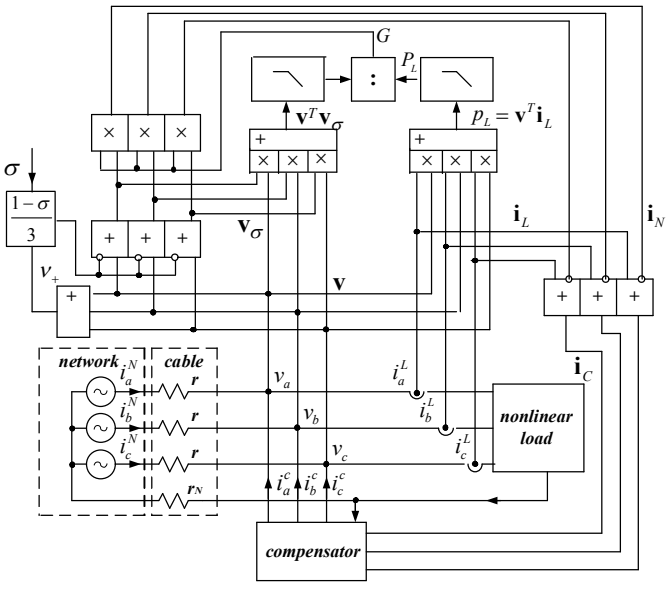


Fig 1. Block diagram of the three-phase system with control unit.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

Let's present the vector of instantaneous values of phase voltages $\mathbf{v} = \|v_a \;\; v_b \;\; v_c\|^T$ as the sum of two vector orthogonal components

$$\mathbf{v} = \mathbf{v}_0 + \mathbf{v}_+, \tag{1}$$

where $\mathbf{v}_+ = (v_+ / 3)\mathbf{j}$ is zero-sequence component directed along the unit vector $\vec{o}$ of $\alpha\beta o$ −coordinate system [5]; $v_+ = \mathbf{j}^T\mathbf{v} = v_a + v_b + v_c$; $\mathbf{j} = \|1 \; 1 \; 1\|^T$; $\mathbf{v}_0 = \mathbf{v} - \mathbf{v}_+ = = \|v_a - v_+ / 3 \;\; v_b - v_+ / 3 \;\; v_c - v_+ / 3\|^T$ is the component of vector of instantaneous values of phase voltages, localized in the $\alpha\beta$ −plane, because the sum of its coordinates is zero. The scalar product of these components is zero $(\mathbf{v}_0{}^T\mathbf{v}_+ = \mathbf{v}_+{}^T\mathbf{v}_0 = 0)$.

Then the voltage vector with the partial attenuation of zero-sequence component is specified by

$$\mathbf{v}_\sigma = \mathbf{v}_0 + \sigma\mathbf{v}_+ = \mathbf{v} - (1-\sigma)\mathbf{v}_+ = \mathbf{v}_0 + \sigma\frac{v_+}{3}\mathbf{j}, \tag{2}$$

where σ is scalar attenuation coefficient of zero-sequence component of the instantaneous phase voltage vector that lies within $0 \le \sigma \le 1$.

Let's assume that the network current vector is proportional to the vector $\mathbf{v}_\sigma$:

$$\mathbf{i}_N = \|i_a^N \;\; i_b^N \;\; i_c^N\| = G\mathbf{v}_\sigma, \tag{3}$$

where G the coefficient of proportionality that could be obtained from the condition of absence of average active power consumption by compensator

$$\frac{1}{T}\int_0^T \mathbf{v}^T\mathbf{i}_C dt = 0 \tag{4}$$

in the case of control by the average value of power consumption or from condition that the compensator instantaneous active power is zero in the case of control by instantaneous value of power consumption. Let consider the first version of control strategy. The obtained results will be easily extended to the second version of control by instantaneous value.

For the first control strategy $\sigma = 1$ and $\mathbf{v}_\sigma = \mathbf{v}$ [4], and it is obtained under condition that network current vector is proportional to the vector of instantaneous values of phase voltages. If $\sigma = 0$ and $\mathbf{v}_\sigma = \mathbf{v}_0$ then second control strategy [6] is applied, and the vector of instantaneous values of phase voltages is used to devoid of zero-sequence component. Next, it will be found an intermediate value of σ that provides minimum of cable active power losses.

Let's represent the compensator current vector as the difference of instantaneous values of load current vector and network current vector respectively

$$\mathbf{i}_C = \mathbf{i}_L - \mathbf{i}_N. \tag{5}$$

Taking into account (2) and (3) the condition (4) will become as follow

$$\frac{1}{T}\int_0^T \mathbf{v}^T\mathbf{i}_L dt = \frac{1}{T}\int_0^T \mathbf{v}^T G\mathbf{v}_\sigma dt. \tag{6}$$

Based on (6) considering orthogonal components $\mathbf{v}_0$, $\mathbf{v}_+ = (v_+ / 3)\mathbf{j}$ the considering coefficient G may be obtain

$$G = \frac{\frac{1}{T}\int_0^T \mathbf{v}^T\mathbf{i}_L dt}{\frac{1}{T}\int_0^T \mathbf{v}^T\mathbf{v}_\sigma dt} = \frac{P_L}{\frac{1}{T}\int_0^T (\mathbf{v}_0 + \frac{v_+}{3}\mathbf{j})^T(\mathbf{v}_0 + \sigma\frac{v_+}{3}\mathbf{j})dt} = = \frac{P_L}{\frac{1}{T}\int_0^T \mathbf{v}_0{}^T\mathbf{v}_0 dt + \frac{\sigma}{9}\mathbf{j}^T\mathbf{j}\frac{1}{T}\int_0^T v_+{}^2 dt} = \frac{P_L}{V_0{}^2 + \sigma V_+{}^2 / 3}, \tag{7}$$

where $P_L = \frac{1}{T}\int_0^T \mathbf{v}^T\mathbf{i}_L dt$ is an average load power; $V_0{}^2 = \frac{1}{T}\int_0^T \mathbf{v}_0{}^T\mathbf{v}_0 dt$; $V_+{}^2 = \frac{1}{T}\int_0^T v_+{}^2 dt$.

Instantaneous cable power losses are

$$\begin{aligned} p_r &= [(i_a^N)^2 + (i_b^N)^2 + (i_c^N)^2]r + (i_a^N + i_b^N + i_c^N)^2 r_n = \\ &= \mathbf{i}_N{}^T\mathbf{i}_N r + (\mathbf{i}_N{}^T\mathbf{j})^2 r_n = \\ &= G^2\mathbf{v}_\sigma{}^T\mathbf{v}_\sigma r + (G\mathbf{v}_\sigma{}^T\mathbf{j})^2 r_n = G^2 r(\mathbf{v}_0{}^T\mathbf{v}_0 + \sigma^2 v_+{}^2 / 3\sigma_r), \end{aligned} \tag{8}$$

where $\sigma_r = r / (3r_n + r)$ is the parameter that characterizes the ratio between resistance of the cable's phase and neutral conductors.

The average power losses will be as follow

$$\begin{aligned} P_r &= \frac{1}{T}\int_0^T p_r dt = \frac{G^2 r}{T}\int_0^T (\mathbf{v}_0{}^T\mathbf{v}_0 + \sigma^2 v_+{}^2 / 3\sigma_r)dt = \\ &= \frac{P_L{}^2 r(V_0{}^2 + \sigma^2 V_+{}^2 / 3\sigma_r)}{(V_0{}^2 + \sigma V_+{}^2 / 3)^2}. \end{aligned} \tag{9}$$

The extremum analysis of (9) gives the optimal attenuation coefficient of zero-sequence component:

$$\sigma_0 = \sigma_r = r / (3r_n + r). \tag{10}$$

With this coefficient value the average of power loss becomes to minimal

$$P_{rMIN} = \frac{P_L{}^2 r(V_0{}^2 + \sigma_0{}^2 V_+{}^2 / 3\sigma_r)}{(V_0{}^2 + \sigma_0 V_+{}^2 / 3)^2} = \frac{P_L{}^2 r}{V_0{}^2 + \sigma_0 V_+{}^2 / 3}. \tag{11}$$

The analysis of (10) shows that the optimal coefficient of σ_0 is only determined by the ratio of specific resistance values of the phase and neutral conductors, i.e. there is the optimal value $\sigma_0 = \sigma_r$, that provides a minimum of active power losses for each brand of four-wire power cable of three-phase network. Thus, in comparison with the control strategies both with complete preservation of zero-sequence component $(\sigma = 1)$ and under full removing $(\sigma = 0)$ the proposed control strategy provides an advantage in power losses with any proportion between r and r_n corresponding to value $0 < \sigma_0 < 1$.

III. Power Loss Comparison

Let's estimate the quantitative gain in power losses compared with the first strategy. If $\sigma = 1$ then (9) becomes as follow

$$P_{r(\sigma=1)} = \frac{P_L^2 r(V_0^2 + V_+^2/3\sigma_r)}{(V_0^2 + V_+^2/3)^2}.$$

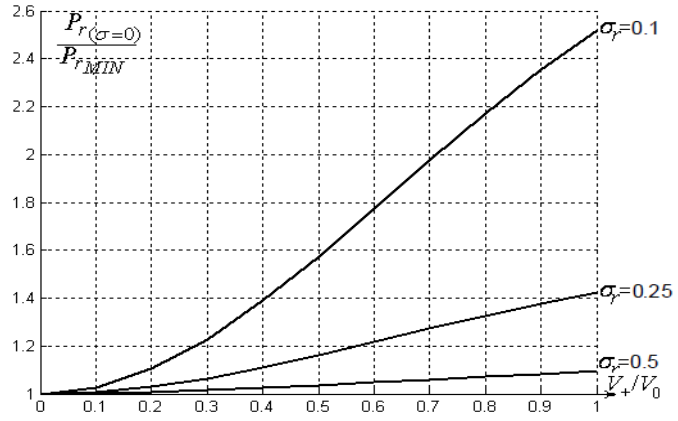
Then quantitative benefit of the proposed control strategy $(\sigma = \sigma_0 = \sigma_r)$ may be compared with the first strategy

$$\frac{P_{r(\sigma=1)}}{P_{rMIN}} = \frac{(V_0^2 + \sigma_0 V_+^2/3)(V_0^2 + V_+^2/3\sigma_r)}{(V_0^2 + V_+^2/3)^2} = \\ = \frac{[1+\sigma_r(V_+/V_0)^2/3][1+\sigma_r^{-1}(V_+/V_0)^2/3]}{[1+(V_+/V_0)^2/3]^2}. \quad (12)$$

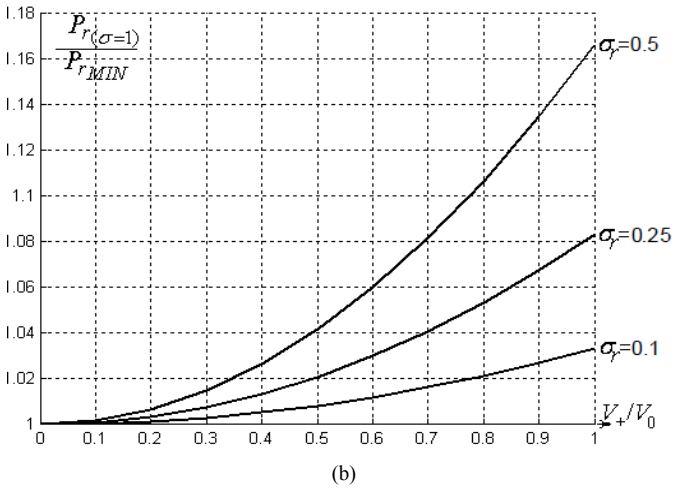
If $\sigma = 0$ then (9) becomes to $P_{r(\sigma=0)} = P_L^2 r / V_0^2$. The quantitative gain compared with the second strategy is

$$\frac{P_{r(\sigma=0)}}{P_{rMIN}} = \frac{V_0^2 + \sigma_0 V_+^2/3}{V_0^2} = 1 + \sigma_r (V_+/V_0)^2/3. \quad (13)$$

Fig. 2 shows the graphs obtained by (12) and (13).



(a)



(b)

Fig 2. The diagram of power losses ratio for the known and proposed control strategies.

The advantages of control strategy with partial attenuation of zero-sequence component will rise together with increasing of the phase voltages unbalance (the growth of V_+ / V_0). At the same time active power losses can be considerably reduced in comparison with both the first strategy (under decreasing of σ_r) and the second strategy (under increasing of σ_r).

IV. Simulation Results

The block diagram of the three-phase system with a control unit that provides a proposed strategy with partial attenuation of zero-sequence component is shown in Fig. 1. The simulation was performed in Matlab based on the structure model of Fig.1. The half-wave rectifier with resistive load was used as a three-phase nonlinear load. The goal of simulation was to validate the active power losses ratio in the power cable of three-phase four-wire network with the coefficient σ obtained based on (9). Let's represent this ratio in the normalized form

$$\frac{P_r V_0^2}{P_L^2 r} = \frac{1 + \sigma^2 \sigma_r^{-1} V_+^2/3V_0^2}{(1+\sigma V_+^2/3V_0^2)^2}. \quad (14)$$

The integrated three-phase network asymmetry is completely determined by the parameter $\Delta = V_+^2/3V_0^2$. Lets' express it through the rated amplitude of voltage V_m of the three phase network and its instability U. For sinusoidal voltages and given instability the maximum value of zero-sequence component will be observed at $\mathbf{v} = \begin{Vmatrix} (V_m+U)\sin\omega t \\ (V_m-U)\sin(\omega t - 2\pi/3) \\ (V_m-U)\sin(\omega t + 2\pi/3) \end{Vmatrix}$. In this case the average value of the square modulus of this vector per period is

$$V^2 = \frac{1}{T}\int_0^T \mathbf{v}^T \mathbf{v} dt = 0.5 \times (3V_m^2 - 2V_m U + 3U^2). \quad (15)$$

The total amount of instantaneous values of phase voltage vector is $v_+ = 2U\sin\omega t$. The another values may be found as

$$V_+^2 = 2U^2;$$
$$V_0^2 = V^2 - V_+^2/3 = 0.5 \times (3V_m^2 - 2V_m U + 5U^2/3).$$

The required parameter is

$$\Delta = \frac{2U^2}{3 \times 0.5 \times (3V_m^2 - 2V_m U + 5U^2/3)} = \frac{4\delta^2}{(9 - 6\delta + 5\delta^2)}, \quad (16)$$

where $\delta = U/V_m$ is relative instability of the amplitude.

The diagram of theoretical ratio (14) and result of the computer simulation with different ratios between resistances of phase and neutral conductor σ_r = 0.1 ($3r$=r_n), 0.25 (r=r_n), 0.5 (r=$3r_n$) and relative instability $\delta = 0.2$ ($\Delta = 0.02$) are shown in Fig. 3 (a) and relative instability $\delta = 0.5$ ($\Delta = 4/29$) are shown in Fig. 3 (b).

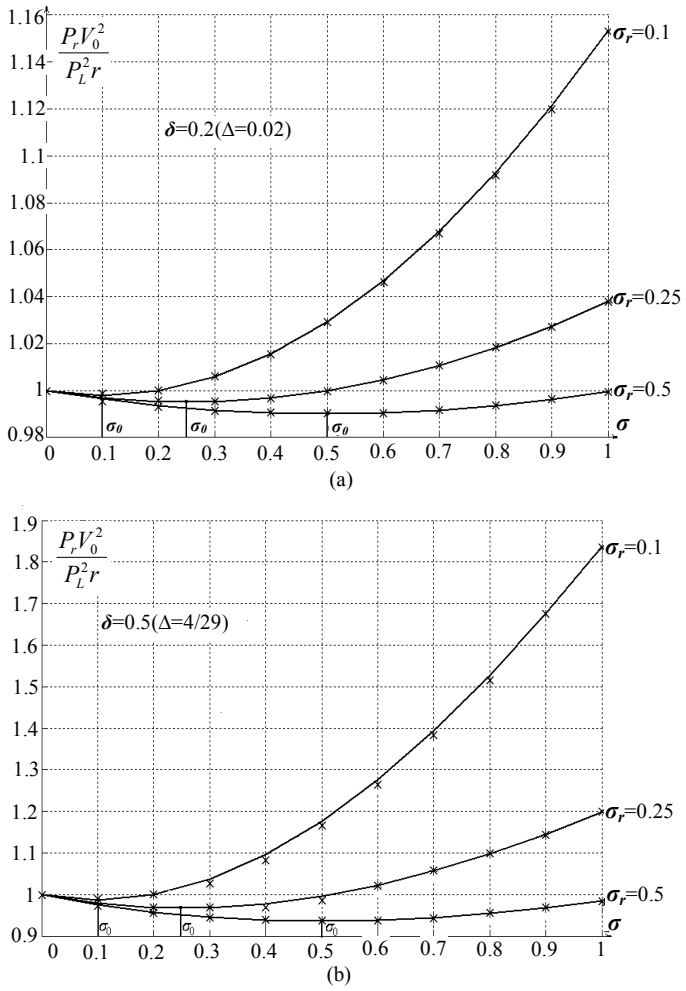


Fig 3. The diagram of theoretical ratios of relative active power losses with the attenuation coefficient of zero-sequence component (solid lines) and computer simulation results (x marked lines).

The simulation confirmed the presence of minimums in all curves whose coordinates correspond to (10). Also, it shows that minimums of values are less than unity, which confirms the decreasing of cable active power losses with the proposed control strategy. Since the values of these minimums differ insignificantly, then for the cable with an unknown brand the value of σ_r, $\sigma_0^* = 0.25$ should be accepted as for the most widespread case $r = r_n$. Then, the σ parameter estimation (Fig. 1) on the control system block diagram is $\frac{1-\sigma_0^*}{3} = \frac{1}{4}$ i.e. the value of each coordinate of partially decreased zero-sequence component is $v_+ / 4$.

V. CONCLUSIONS

The control strategy of shunt active filter in three-phase four-wire network at which instantaneous currents vector of three-phase network is proportional to the instantaneous phase voltages vector with partial attenuation of zero-sequence component has been proposed. At the optimal value of the attenuation coefficient of zero-sequence component and a large of phase voltages unbalance this strategy provides certain advantages of active power losses in power cable in comparison with two other control strategies: with saving the zero-sequence current and with removing of zero-sequence current.

The optimal attenuation coefficient is only determined by the ratio of specific phase resistance values and neutral conductors, i.e. for each brand of four-wire power cable of three-phase network the optimal value is $\sigma_0 = \sigma_r = r/(3r_n + r)$, that provides a minimum of active power losses. If the cable brand is an unknown and it is impossible to establish the exact value of σ_r, $\sigma_0^* = 0.25$, then ratio $r = r_n$. should be applied as for most common case. The value of each coordinates of partially compensated zero-sequence current is equal to the sum of instantaneous values of the phase voltage vector divided by 4.

The simulation confirmed the presence of minimums in all diagrams of relative active power losses in the power cables It is shown, that minimum value depends on the compensation coefficient of zero-sequence current whose coordinates correspond to (10). The values minimums are less than unity, which confirms the decreasing of active power losses in the cable due to the proposed control strategy.

REFERENCES

[1] H. Akagi, Y. Kanazawa, and A. Nabae, "Generalized theory of the instantaneous reactive power in three phase circuits," Proceedings IPEC83 Conference, Tokyo (J8), pp. 1375-1386, Sept. 1983.

[2] H. S. Kim, and H. Akagi, "The instantaneous power theory based on mapping matrices in three-phase four-wire systems," In/ Proc. PCC"97 Conf., Nagaoka, Japan, vol.1, pp. 361-366, Aug., 1997.

[3] H. S. Kim, S. Ogasawara, and H. Akagi, "The theory of instantaneous power in three-phase four-wire systems," In/ Proc. IEEC/IAS'99, Annu. Metting, pp.431-439, Oct.,1999.

[4] F. Z. Peng, and J. S. Lai, "Generalized instantaneous reactive power theory of three-phase power systems," IEEE Trans. Instrum. Meas., vol. 45, no 1, pp.293-297, Feb. 1996.

[5] H. Akagi, E.H. Watanable, and M. Aredes, Instantaneous power theory and applications to power conditioning. Piscataway, NJ: IEEE Press / Wiley Interscience, 2007.

[6] J. C. Montano Asquerino, and P. Salmeron Revuelta, "Compensation in Nonsinusoidal, unbalanced three-phase four-wire systems with active power-line conditioner," IEEE Trans. Power Delivery, vol. 17, no. 4, pp. 1079–1084, Oct. 2002.

[7] J. C. Montano, P. Salmeron, and J. P. Thomas, "Analysis of power losses for instantaneous compensation of three-phase four-wire systems," IEEE Trans. Power Electron., vol. 20, no. 4, pp. 901–907, July 2005.

The use of the Dirichlet Kernel in the Control Systems of Active Filters for Industrial Power Line

V.V. Zamaruiev
The Department of Industrial and Biomedical Electronics
National Technical University "Kharkiv Polytechnic Institute"
Kharkiv, Ukraine

Abstract — **Optimization of power converter topologies for power factor converters, active filters, active rectifiers etc. requires dividing converter's output signal spectrum between several power modules. These include the need to harmonize the frequency and phase responses and to synchronize input and output signals of several filters that process the same signal on different frequency intervals, on the boundary of their bandwidth in real time. The transfer of well known solutions for the digital filter synthesis from the field of radio, audio and other signal processing to the control systems of power converters faces several challenges. The article discusses the construction of a digital filter using the Dirichlet kernel that eliminates these problems.**

Keywords — power active filter; band-pass filter; control system; Dirichlet kernel; digital filter; distributed active filter; power quality.

I. INTRODUCTION

With the development of technology of power semiconductor devices, power converters have been increasingly used in industrial applications, such as electric drive, power supply of technological equipment and uninterruptible power supply. In most cases, diode or thyristor rectifiers are used for the conversion from AC to DC. Using such converters leads to a significant reduction of power factor. To improve dissipation coefficient the Static Var Compensators (SVC) are widely used. Recently power active filters (PAF) are used for the suppression of non-active power components [1]. Typical industrial power system is shown in Fig. 1. Application of SVCs and PFs cases serious problems in utility power distribution systems caused by harmonic resonance between line inductances and shunt capacitors for power factor correction, which contributes to a significant amplification of voltage harmonics in a distribution line [2].

The feature of the active filters work in power systems is the need to suppress the voltages (currents) with predetermined frequencies - harmonics of the fundamental frequency of power supply system voltage. Wide range of suppressed harmonics makes it reasonable to use a distributed active filter [3] consisting of several PAF modules [4], each of which suppresses harmonics of their own frequency range (Fig. 2). The joint work of several PAF allows to optimize the operating

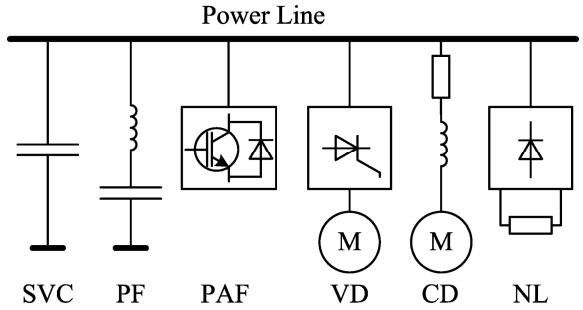


Fig. 1. Typical industrial power system. SVC – static VAR compensator; PF – passive filter; PAF – power active filter; VD – variable speed (controlled) drives; CD – conventional drives; NL – nonlinear load.

frequency range, installed capacity, schematics solutions and applied element base for each module on criteria of minimum losses, costs, etc. [5].

The use of a system with distributed PAF is useful when strict requirements to electric power quality in electric power distribution networks [6] and power networks for responsible consumers are made. For example, in the railway power system the demands to limit the voltage ripple cased by harmonic distortion to ensure electromagnetic compatibility with automation and telecommunications systems are made [7]. PAFs have to operate in a wide frequency range to mitigate the content of harmonics from 1-st to 111-th in power supply voltage according to standards.

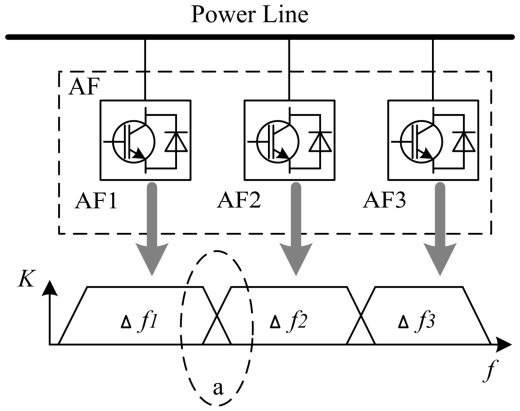


Fig. 2. Distributed active filter and its frequency response.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

With interoperating of several PAF the coordination of their frequency and phase responses in the overlapping areas of their bandwidths (region "a" in Fig. 2) is an essential requirement. The frequency response control for PAF modules is implemented by filters in microprocessor control system. The use of digital filters requires an estimation of phase shift of the control system output sygnals with respect to the measured voltage (current), that distinguishes it, for example, from the sound recording applications. It is necessary to use mathematical methods that reduce this shift to zero. The use of the Dirichlet kernel in the joint control system for several active filters in a distributed system of active filtering is presented and discussed.

II. FILTERS IN THE CONTROL SYSTEM

A. Requirements to Digital Filter

The requirements to the frequency responses of filters, which are part of a distributed PAF system, are determined by their joint work [8]. Since the disturbing signals have determined frequencies, further filters with specified gains and phase responses at discrete points will be considered. These filters can be called "filters of discrete frequencies".

Let us consider the overlapping bandwidths of adjacent filters (region "a" in Fig. 2). Standard methods of filter synthesis are generating frequency response similar to the form shown in Fig. 3. Interference of filters leads to occurrence of harmonics with frequencies f_i in the frequency range f_n - f_m (cut-off frequencies of both filters). Assume the equal gains of both filters, the total gain of two filters on the i-th harmonic K_i must be equal to the gain K_1 of single filter in the passband, with the identical phase shift of the harmonics of the same numbers. In the case of different phase shifts it is necessary to use the rules of vector addition.

To simplify the coordination of PAF modules control systems it is desirable to use a filter with close to ideal frequance and phase responces (Fig. 4). The increase of conventional filter frequance responce squareness leads to its order raise, the phase shift increasing of the harmonics in bandwidth, poor performance in the stopband. However, it is possible to get the desired frequency response of "filter of discrete frequencies" while using the Fourier transform, which operates with the harmonics of the fundamental. A partial sum

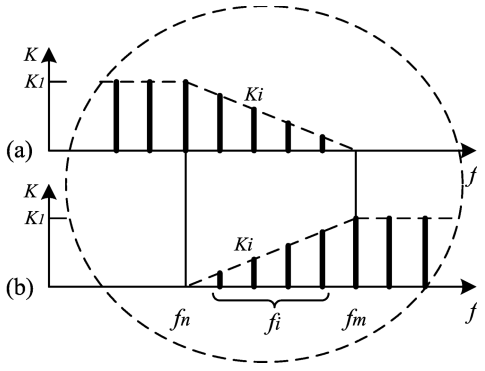


Fig. 3. Frequency response of PAF modules: (a) – PAF1, (b) – PAF2.

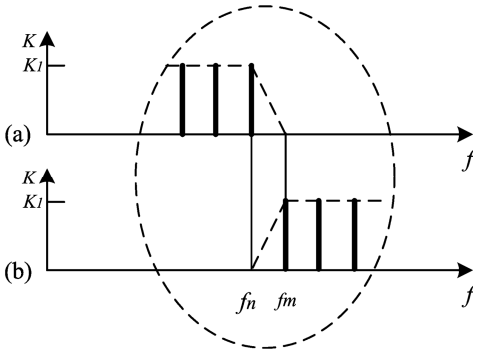


Fig. 4. Frequency response of PAF modules: (a) – PAF1, (b) – PAF2.

of Fourier series corresponds to the use of a low pass filter. On the other hand, the implementation of a filter with a finite impulse response (FIR filter) corresponds to the integral transformation with kernel which is defined by impulse response of filter.

B. Dirichlet kernel

Kernel of integral Fourier transform is a function $D_n(z)$. The convolution of D_n with any function q of period 2π is the n th-degree partial sum of the trigonometric Fourier series $S_n(x)$ (1). This function is called the Dirichlet kernel of order n [9].

$$S_n(x) = \int_0^{2\pi} q(t) D_n(x-t)dt \qquad (1)$$

Dirichlet kernel is 2π periodic function defined by the formula:

$$D_n(z) = \frac{1}{2\pi}\sum_{k=-n}^{n} e^{ikz} = \frac{1}{2\pi}(1 + 2\sum_{k=1}^{n}\cos(kz)), \qquad (2)$$

or, in short form,

$$D_n(z) = \frac{1}{2\pi}\frac{\sin((n+1/2)z)}{\sin(z/2)}. \qquad (3)$$

In the form of (3) the function must be defined in the points z=0, $D_n(0)=(2n+1)/(2\pi)$. The constant coefficient $1/(2\pi)$, which is present in (2) and (3), is reduced to 1/2 or 1 in a number of sources [10], respectively, the integral in (1) is multiplied by $1/\pi$ or $1/(2\pi)$. The value of the Dirichlet kernel of the fifth order on it's period is shown in Fig. 5.

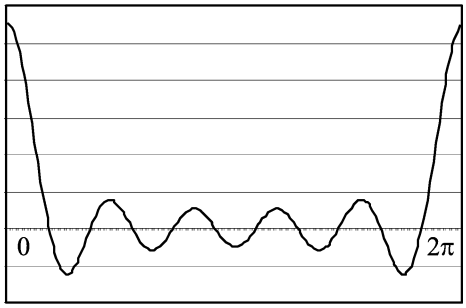


Fig. 5. Graph of Dirichlet kernel $D_5(z)$ *(impulse response)*.

The use of the Dirichlet kernel allows to get an FIR low pass filter with a unit value of gain at frequencies that are multiple of the first harmonic. There is zero phase shifts between input and output signals of the filter at these frequencies. Frequency and phase response of the synthesized filter is shown in Fig. 6. The frequency of the signal is given in the value relative to the first harmonic of the kernel $f^*=f/f_1$. Characteristics correspond to an ideal low pass filter for discrete frequencies. While this method of filter synthesis limits the gain and phase shift only for the frequencies that are multiple of fundamental harmonic, the gain of the filter for fractional frequencies are not normalized.

The digital filter allows using the filter kernel with a phase shift (Fig. 7). It is expedient to estimate filter properties in this mode. Substituting the variable x for $x + \Delta$ in (1) we will get

$$S_n(x+\Delta) = \int_0^{2\pi} f(t) D_n(x-(t-\Delta))dt \,. \qquad (4)$$

As follows from (4), the time shift of the kernel of the integral transformation leads to a corresponding time shift of image relative to the original function without changing its value.

The frequency and phase responses of the FIR filter based on the Dirichlet kernel shifted by half a period are shown in Fig. 8. The gain's absolute value at frequencies multiple of fundamental harmonic has not changed. Phase shift of the odd harmonics became equal to π, which corresponds to (4). Generally, in AC systems there are no odd harmonics in current and voltage waweforms. In such systems, the transition to this version of the filter requires changing the sign of the feedback. A slight change in the gain of the filter in the neighborhood points of the integer value of the relative frequency f^* eliminates fundamental voltage (current) frequency tracking in the control system of PAF.

One of the main characteristics of FIR filter is its step response. Fig. 9 shows the step response of the FIR filter of 150-th order with a different shift of the kernel depending on the number of calculation step. As follows from Fig. 9 step response of filter with shift of kernel on π corresponds to the response of typical FIR filter.

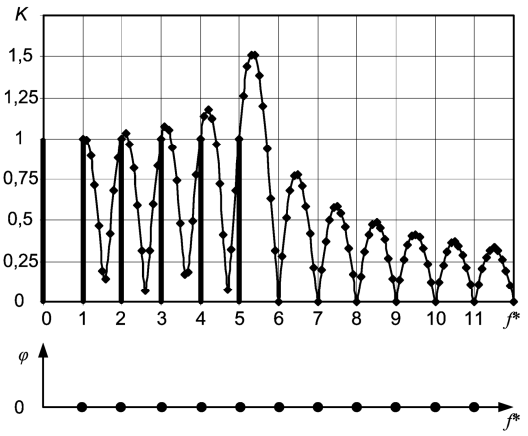


Fig. 6. Frequency response and phase response.

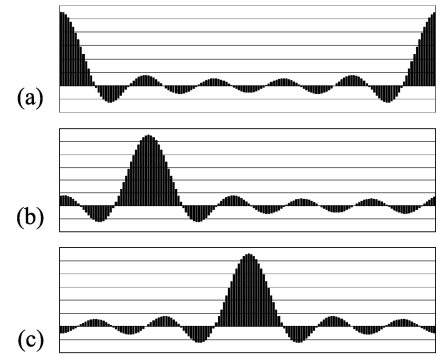


Fig. 7. Dirichlet kernel: (a) without phase shift, (b) phase shift π/2, (c) phase shift π.

With smaller shift of kernel the reaction of filter accelerates and jump occurs in the time, on which the kernel has been shifted, and that improves the operating speed of the filter. At zero shift of the kernel step response is atypical and is divided into two steps each of half-amplitude. Such behavior of the filter in the control circuit of PAF allows eliminating 50% of the disturbance immediately after its occurrence and completely suppressing the disturbance in the time corresponding to the filter's impulse response length.

C. *Band-pass Filter*

When using the Dirichlet kernel to receive a partial sum of the Fourier series (2), in which only a few harmonics are taken into consideration, we will get a modified kernel of the filter. For example, the kernel of the bandpass filter that transmits 6th, 7th and 8th harmonics is specified by expression:

$$D_{6,7,8}(z) = \frac{1}{2\pi}(2\sum_{k=6,7,8}\cos(kz)) \,. \qquad (5)$$

The frequency and phase responses of the received filter are shown in Fig. 10. The received characteristics correspond to discrete frequency bandpass filter, since the gain at specified harmonics is equal to unity, and phase shift is equal to zero.

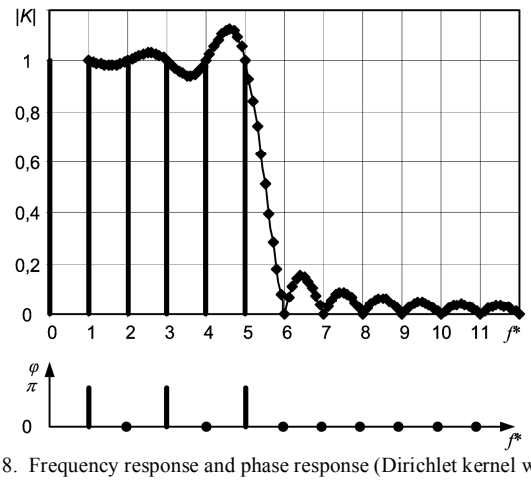


Fig. 8. Frequency response and phase response (Dirichlet kernel with π phase shift).

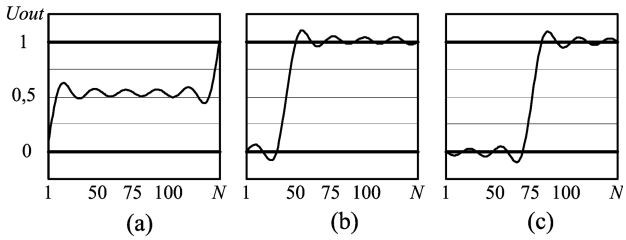


Fig. 9. Step response of filter based on Dirichlet kernel: (a) without phase shift, (b) phase shift π/2, (c) phase shift π.

The appliance of filters and kernels (2) for $1 \leqslant n \leqslant 5$, and (5) for n = 6, 7, 8 allows to obtain the active filters with the form of frequency response as shown on Fig. 4, which excludes their mutual influence.

III. Experimental Results

The computational experiment was carried out on a model of the active filter shown in Fig. 11. The structure of the active filter consists of two discrete frequency bandpass filters with kernels (5) for n=3, 4, 5 and n=6, 7, 8 respectively. The input signal S_{in} of the active filter contains only first harmonic and has unity amplitude (Fig. 12 a). In the instant time N_1 the disturbance was added to original signal as the sum of the 5th and 6th harmonics with unity amplitude.

Output signal S_{out} of active filter is shown on Fig. 12 b. It can be seen from the analysis of the figure, that the disturbance on the output of the system is less than half of its own value over time period N_1-N_2, which is equal to the filter impulse response length. After this period the disturbance is fully compensated. The results of numerical experiments confirmed the theoretical calculations and the lack of mutual influence of filters on adjacent frequencies.

IV. Conclusion

The use of the Dirichlet kernel as the kernel for the FIR filter allows synthesizing low pass filter with a unity gain and zero phase shift for discrete frequencies.

The time shift of kernel of the integral transformation leads to a corresponding time shift of function of image which is relative to the original function without changing its value.

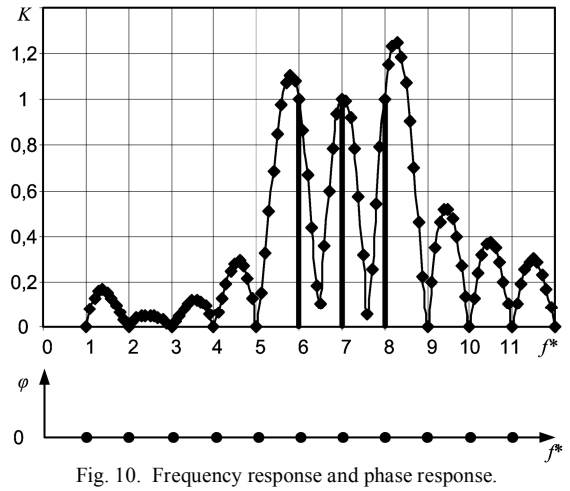


Fig. 10. Frequency response and phase response.

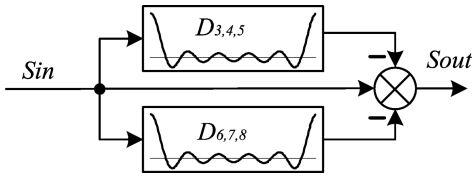


Fig. 11. Simple active filter.

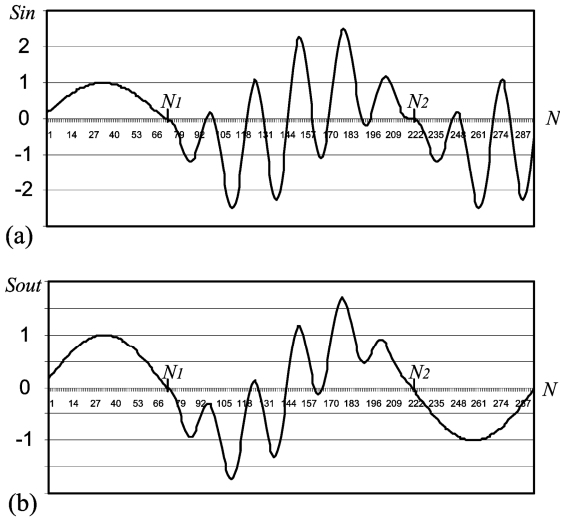


Fig. 12. Input (a) and output (b) signal of simple active filter.

The use of the modified filter kernel provides synthesis of a bandpass filter of discrete frequencies.

When FIR filter based on the Dirichlet kernel is used in active filter, the disturbance effects at the output of the system is less than half of its value during the time period that is equal to the filter impulse response length.

References

[1] V.K. Sood, "HVDC and FACTS Controllers: Applications of Static Converters in Power Systems," Springer, 2004, 296 p.

[2] H. Akagi, "Instantaneous Power Theory and Applications to Power Conditioning," IEEE Press, 2007, 379 p.

[3] Po-Tai Cheng, Tzung-Lin Lee, "Distributed Active Filter Systems (DAFSs): A New Approach to Power System Harmonics" *IEEE Trans. Ind. Appl.*, vol. 42, no. 5, pp. 1301-1309, Sept./Oct. 2006.

[4] J. Tlusty, V. Valouch, "Harmonic Voltage Mitigation in Industrial Power Systems by Using Multiple Active Power Filter Configuration," *EPE-PEMC 2006*, Portoroz, Slovenia, 2006, pp. 580 – 584

[5] V.V. Zamaruiev, "Raspredelennaja sistema aktivnoj filtracii (Distributed system of active filtering)," Tehnichna electrodynamika. Sylova electronika ta energoefektyvnist', vol. 3, Kyiv, 2012, pp. 18-21. /in Russian.

[6] "Harmonic Limits" IEEE Std 519-1992.

[7] J. Claxton, "Signalling and telecomms compatibility issues," IET Professional Development course on "Electric Traction Systems", 2008, pp. 323–329.

[8] P. Jintakosonwit, H. Akagi, H. Fujita, S. Ogasawara, "Implementation and performance of automatic gain adjustment in a shunt-active filter for harmonic damping throughout a power distribution system," *IEEE Trans. Power Electron.*, vol. 17, pp. 438–447, Jan. 2002.

[9] A.M. Bruckner, J.B. Bruckner, B.S. Thomson, "Real Analysis," (2nd Edition), ClassicalRealAnalysis.com, 2008. Available: http://classicalrealanalysis.info (URL)

[10] R. J. Beerends "Fourier and Laplace Transforms" Cambridge University Press, 2003, 447 p.

The Method of Determining the Duration of Transients in Semiconductor Devices for Current-Voltage Characteristics Measurement

Ievgen O. Iermolenko, Oleksandr F. Bondarenko, Kristina I. Iermolenko
Electronic Systems Department
Donbas State Technical University
Alchevs'k, Ukraine
ermolenkoea@gmail.com, bondarenkoaf@gmail.com, chi.ermolenko@gmail.com

***Abstract* — The improved method of automatic determining the duration of oscillating transients in semiconductor devices for correct current-voltage characteristics measurement is proposed. This method allows increasing the accuracy of current-voltage characteristics measurement. It is applied in measuring device for current-voltage characteristics analysis.**

Keywords—semiconductor device; transient; current-voltage characteristic; measurement

I. INTRODUCTION

Technology development demands improvement of semiconductor converters characteristics, which directly depend on correct choice of electronic components. As a rule, selection of semiconductor devices means determining their electrical parameters. The most complete information about electrical parameters of semiconductor device may be obtained by its current-voltage (IV) characteristics measuring. Precise IV-characteristic measuring is very important for obtaining true information about parameters of semiconductor device under test (DUT).

It is known that the accuracy of current-voltage characteristic measurement significantly depends on parameters of electrical signals used for measurement. In particular, the prolonged influence of measuring signals on semiconductor devices can lead to heat accumulation in its structure because of self-heating effects. The pulse method of measurement of current-voltage characteristics is more effective for reducing the self-heating. The highest accuracy of measurement can be achieved with the right choice of parameters of measuring pulse sequence.

II. ADAPTIVE METHOD OF MEASURING THE IV-CHARACTERISTICS

The automatic method of measuring IV-characteristics is proposed in [1]. This method allows forming the parameters of measuring pulses adaptively, considering the tested semiconductor device reaction time and thermal properties. According to the proposed method (Fig. 1) the duration t_{pulse} of each pulse P_i should be minimal to avoid the excessive heating

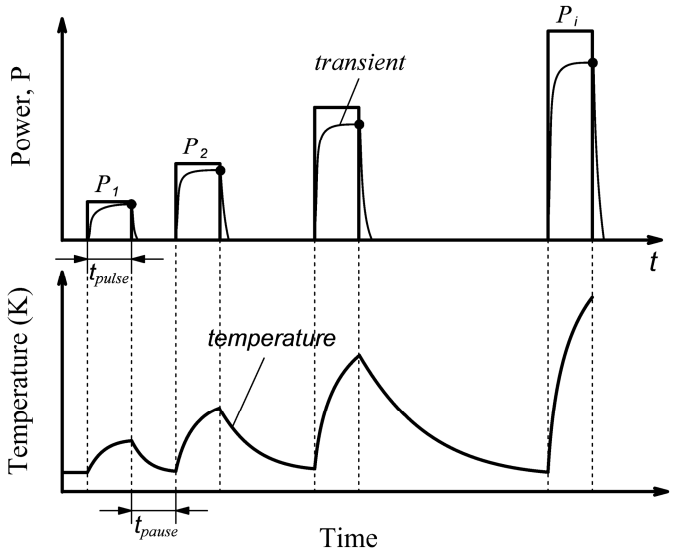


Fig. 1. Adaptive method of measuring the IV-characteristics

of DUT, but sufficient to steady the transient response of the DUT. Too short duration of the pulse can lead to wrong registration of the DUT-response due to incomplete transient.

The pause duration t_{pause} following the measuring pulse should be variable while measuring the IV-characteristic: greater power P_i and longer pulse time t_{pulse} cause greater heating the DUT and require its longer cooling off.

Therefore, the aim is to develop the method of accurate determining the DUT transient time for adaptive measuring the IV-characteristics implementation.

III. THE METHODS OF DETERMINING THE TRANSIENT TIME

A method of determining the parameters of the transient is described in [2]. According to the method, four instant values of DUT response after regular time intervals are measured and then transient parameters are calculated. The disadvantage of this method is its applicability only for aperiodical transients. As a rule, the construction of a DUT contains parasitic inductance (e.g. external pins). Together with internal capacity

of DUT the inductance forms a resonant circuit. That is why the transients in most semiconductor devices have oscillatory form.

Another algorithm of determining the duration of transient is described in [3]. According to this algorithm, the response U_i of the DUT is measured periodically during the action of a rectangular pulse. The time interval between samples is defined as minimal as possible. Simultaneously, the difference ΔU_i between the just measured value U_i and the value measured in the previous step U_{i-1} is calculated. This difference characterizes the dynamics of DUT response at any time. If this difference is small enough, the transient can be considered as steady. Thus, the termination of the DUT transient is the moment when the difference ΔU_i reaches a certain limit.

To verify the efficiency of the algorithm [3] for the oscillatory form of the transient the simulation and calculations were carried out. The results are shown in Fig. 2.

Fig. 2 (a) illustrates the form of an acting pulse and the form of the transient of a DUT response. Fig. 2 (b) shows the difference ΔU_i calculated at each time point. Fig 2 (c) illustrates an output rectangular signal Q, determining the duration of the transient in accordance with the method [3].

The results of simulation show that the duration of transient with oscillatory form is determined incorrectly because of reaching the value ΔU to zero in time moments of transient maximums. So, it is necessary to monitor the change of ΔU of the DUT response for a certain time to develop the universal algorithm, which allows determining the duration of the transient of any form.

There is no information in [3] about the limit of ΔU_i enough to consider the transient completed. Besides, the comparison of ΔU_i with some absolute value can lead to wrong determining the transient termination. For example, the value of ΔU_i can be smaller than the limit if the fast-acting meter, which allows to sample the responses of U_i with the small discreteness, is used. Therefore, in order to determine the transient duration correctly it is necessary to analyze a relative value of ΔU_i.

In order to exclude the described disadvantages, the improved method of determining the duration of the DUT transient is proposed. The method is represented in Fig. 3 as the algorithm.

According to the method, test pulse with longer duration than time of transient applies the DUT. Analysis of the characteristics of modern semiconductor devices from Fairchild Semiconductor, ON Semiconductor, Texas Instruments, NXP, showed that the maximum duration of the transient of these devices does not exceed 10 μs. That is why the test pulse duration is proposed to set equal to 10 μs.

Similar to the initial algorithm the values of DUT response are cyclically measured with a period Δt during the test pulse. The difference ΔU_i between the just measured value U_i and the value measured at the previous step U_{i-1} is calculated and saved to the RAM at each iteration of the cycle.

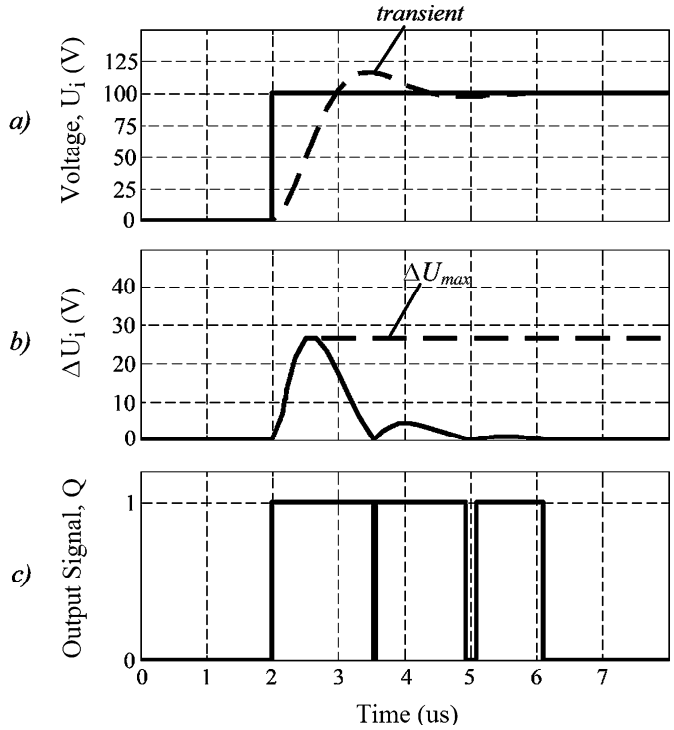


Fig. 2. Simulation of the algorithm [3] for the oscillatory form of the transient

Simultaneously, the maximum deviation ΔU_{max} is calculated. These operations are repeated until the test pulse stops.

Thus, when the test pulse stops the i number of ΔU_i values will be stored in the RAM. These values characterize the change of DUT response at each iteration of the cycle. Then the memory is scanned in reverse order to find the iteration number k corresponding the value ΔU_k that exceeds the preset limit – 1% of ΔU_{max}.

Since the duration of a single iteration is Δt, the total time of the transient t_{tr} is calculated by the formula:

$$t_{tr} = k \cdot \Delta t .$$

To implement the adaptive method [1] the device for measuring the current-voltage characteristics (Fig. 4) and software (Fig. 5) for this device were developed. The device interacts with the PC via a data acquisition board PCI-1711 from Advantech. The device measuring limits are ±10 V and ±1 A. The software allows to control the device and to observe the measurement results on a computer screen.

The proposed algorithm of determining the transient duration is used in this measurement system. This algorithm allows to set the duration of measuring pulses automatically on the basis of the dynamic properties of DUT and to increase the accuracy of the IV-characteristic.

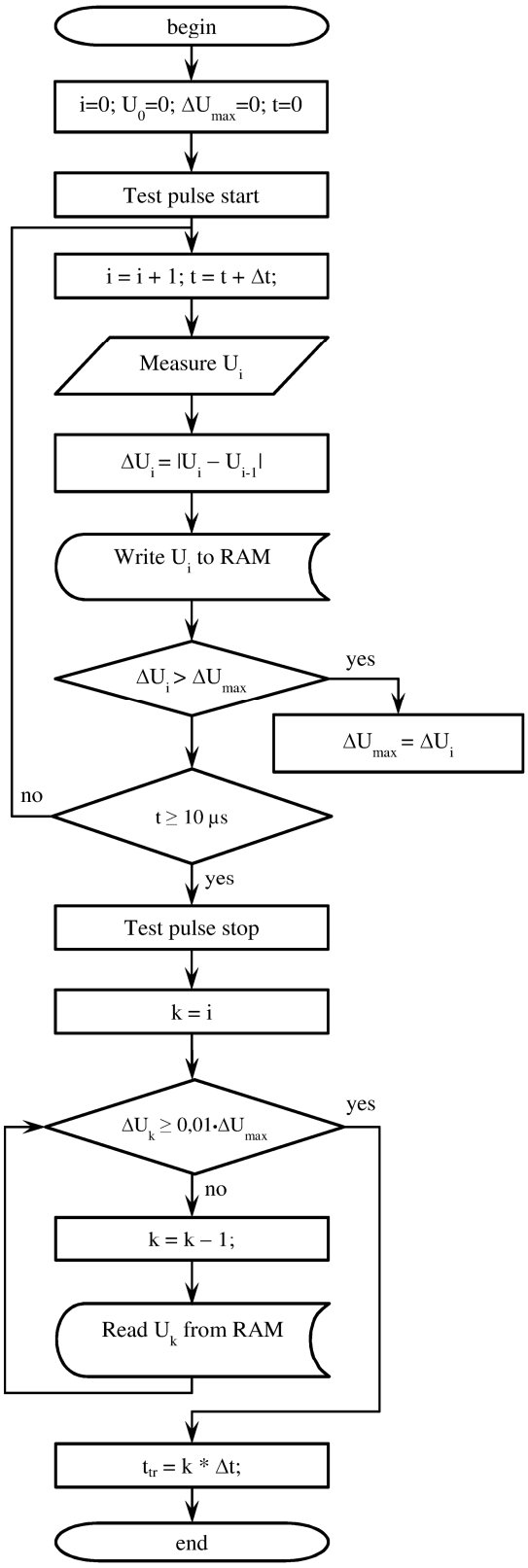


Fig. 3. Algorithm of determining the transient duration

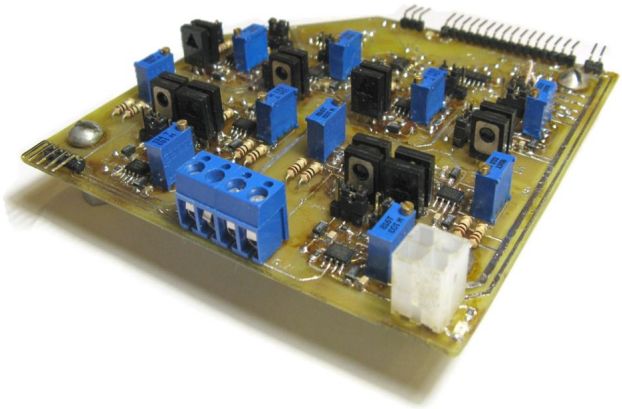


Fig. 4. Device for measuring the IV-characteristics of semiconductor devices

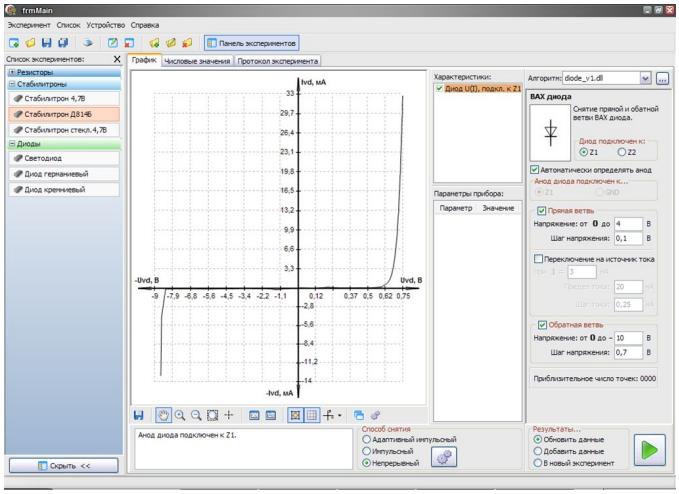


Fig. 5. Screenshot of software for measuring the IV-characteristics

IV. Conclusion

The analysis of methods for determining the parameters of DUT transients showed that the known methods allow correct determining the duration of aperiodical transients only.

The improved method for determining the duration of oscillating transients allows forming measuring pulse sequence in accordance with adaptive method for measuring IV-characteristics. Use of proposed method can increase the accuracy of IV-characteristic measurement.

References

[1] Patent № UA96998, G01R 23/16, 19/32. Method for automated measurement of volt-ampere characteristics of semiconductor devices / Bondarenko O.F., Yermolenko Ye.O. (UA). Publication date: Dec. 26, 2011.

[2] Patent № RU2187822, G01R 23/16. Way determining parameters of transient process / Safarov M.R., Sarvarov L.V. (RU). Publication date: Aug. 20, 2002.

[3] Patent № US5389990, G01R 31/28. Method for measuring DC current/voltage characteristic of semiconductor device / Osamu Nakamur, Kabushiki Kaisha Toshiba (JP). Publication date: Feb. 14, 1995.

Multilevel Inverter as VAR-compensator

Vladimir Bezhenar, Dmitry Mykolaets, Vitaly Mykytyuk, Tatiana Tereshchenko
Faculty of Electronics
National Technical University of Ukraine "Kyiv Polytechnic Institute"
Kyiv, Ukraine
vl.bezhenar@gmail.com

***Abstract*—Multilevel inverter consists of base cells that have the same structure as active power line conditioner. This allows to use the inverter cells as the VAR-compensator in uninterruptible power supply systems. The advantage of this system is increasing the allowable current range that can be correctly compensated. All the necessary changes for such mode provides control system algorithm only.**

Keywords—uninterruptible power supply; multilevel inverter; compensation; active power line conditioner.

I. INTRODUCTION

Active implementation of the energy saving technologies imposes high requirements for power supply systems, such as efficient use of the renewable energy sources, ensuring the smooth operation of loads with specified parameters.

To provide uninterruptible power supply for responsible consumers [2], the systems with redundant power sources such as photocells are used. Thus the important characteristics are working time without network voltage, a form of output voltage (current), the cost of additional equipment that provides these parameters improving. In this paper the economic solution of the problem by combining the functions of multilevel inverter and VAR-compensator is proposed.

II. UNINTERRUPTIBLE POWER SUPPLY SYSTEM

As an example, an uninterruptible power supply system for responsible equipment is considered. Existing systems provide trouble-free operation time of such equipment without the supply voltage [1, 2]. But at the same time they don't match power supply higher quality. To ensure this in power system (Fig. 1) additional measures are used.

Using of standalone multilevel inverter (SAMI) provides a required THD level while photocell operation. While it operates with AC line, the SAMI will be used as the VAR-compensator. It provides high quality operation for different loads with significant voltage parameters deviations.

The base principle of the device (Fig.1) is similar to the "line-interactive" type [2]. Switches S1-S5 provide the required scheme configuration, depending on the availability of AC line. Charging unit (CU) charges the accumulator battery (AB). Photocell (PC) ensures system's standalone operational mode and allows maintaining AB charging regardless of mains availability. Stabilizer (ST) is used to improve the output voltage (current) parameters and connected directly to the load (L).

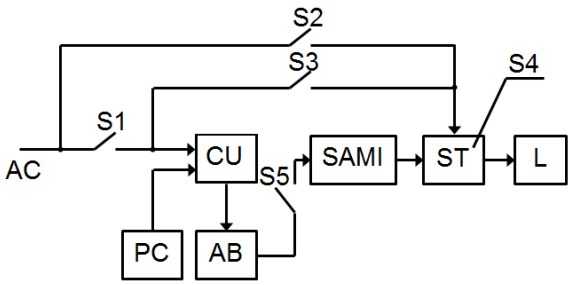


Fig. 1. Block scheme of responsible equipment' power supply

The AB is charged from the PG and the AC line if it is available. In this case, S2 and S5 are in open state. In case of AC line voltage failure or excessive lowering switches S1, S3 are turned off, and S5 is opened. This ensures supply of L with energy stored in AB. Nevertheless the charging process continues from PG that prolongs battery life.

With the AC line voltage failure SAMI operates as multilevel inverter or VAR-compensator otherwise. This solution provides power supply higher quality with capacity rate at least 0.99 and THD in inverter mode is significantly lower than 1% [3].

III. ACTIVE POWER LINE CONDITIONER

Nowadays there is a great variety of methods and schemes for reactive power compensation, which can be divided into two main groups: passive compensation devices (PCV) and active compensation devices (ACV). Passive reactive power compensation is implemented by using a set of capacitors [4, 5].

The system's advantages are the simpler implementation, comparing to active devices, and small specific losses 0,25-0,35 % [4]. On the other hand the disadvantages are large dimensions, the ability to compensate only the fundamental frequency and significant voltage and current surges during transistors' switching, which leads to system unstable operation. Active compensation of reactive power has significant advantages compared to passive, such as the stability of the system. One of the most widely used devices for reactive power compensation is active filter [6, 7], that operates at high frequency and can compensate the reactive power not only on the fundamental but also on higher frequencies.

Recently the active power line conditioners (APLC) showed on Fig. 2 were proposed [8]. The advantages of this

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

device the reactive power compensation ability with load character changing, and also that compensator is designed only for a part of the devices total power. The devices power section is similar to bridge inverter and due to control system's algorithms it operates as a reactive power compensator [8]. Obviously, such system can operates as: bridge inverter, APLC or as one of the multilevel inverter cells.

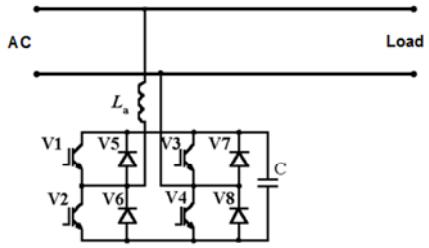


Fig. 2. Filtering VAR-compensator schematic

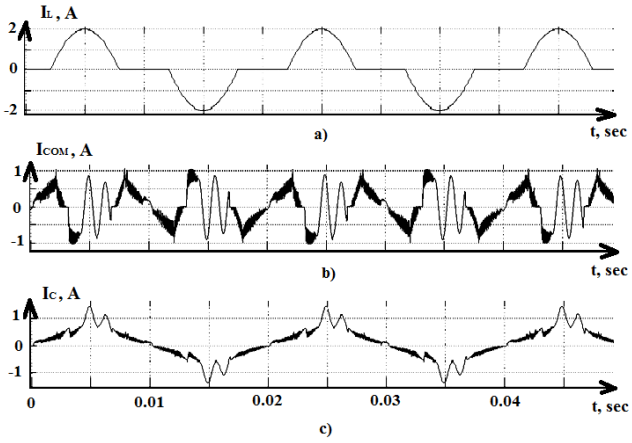


Fig. 3. APLC overpowered diagrams

While the load changes in the wide range, there are cases when the APLC is unable to compensate all the assigned power (Fig. 3a). This leads to compensation (Fig. 3b), and consumption (Fig. 3c) current distortions and can cause device malfunction while prolonged running in this mode. Therefore such use of the APLC leads to additional AC line voltage distortions and should be eliminated by more powerful or parallel APLC as suggested.

The circuit with two parallel connected APLCs is simulated in Simulink software. The test model is showed on Fig.4. It corresponds to power supply system with typical rectifier and high power RC-load.

Two multilevel inverter cells that operates as APLC are represented as Universal Bridge1 and Universal Bridge2. These elements are powered by capacitors C1 and C2 and parallel connected to the AC line and load. On Fig.5a load current, Fig.5b total compensation current and Fig.5c the source resulting current.

When APLC's rated power is less than load power it leads to source current distortion. The parallel connection of several APLCs provides no distortion reactive power compensation.

APLC is operating performing power balance:

$$W_2 = W_1 + W_3, \quad (1)$$

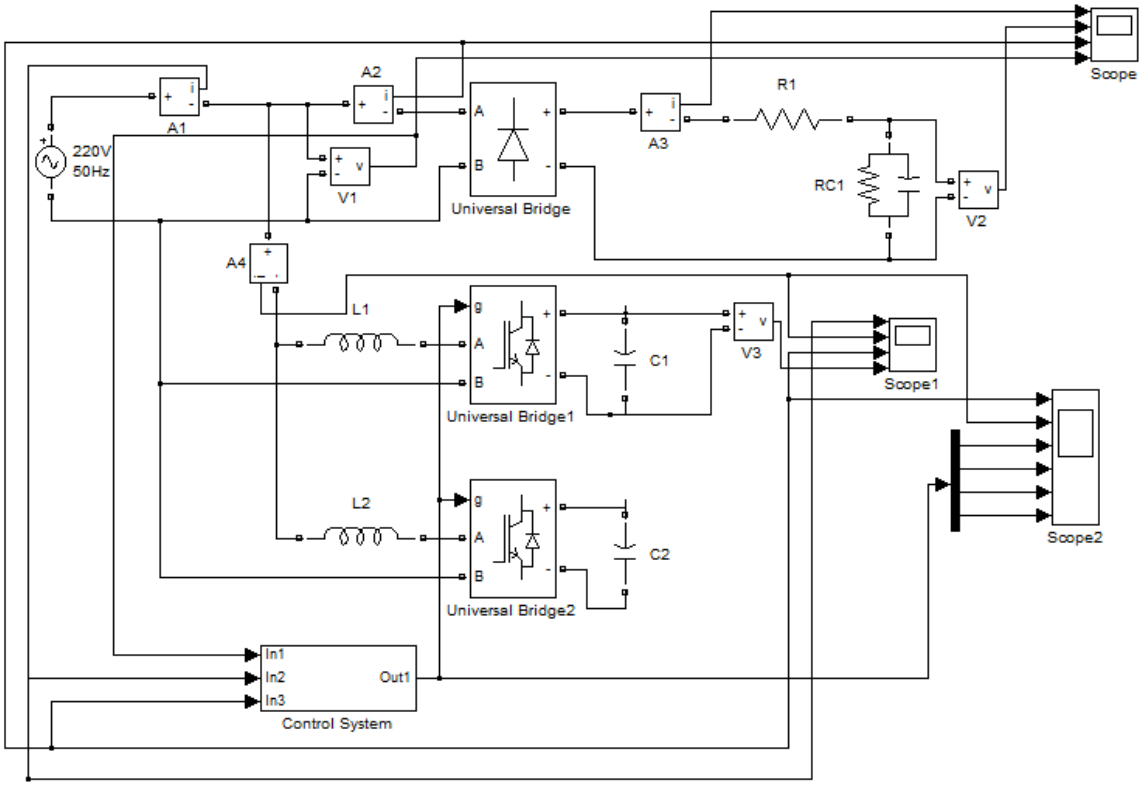


Fig. 4. Simulink modelling schematic

where W_1 and W_3 – the energy accumulated in APLC, W_2 – the energy that APLC returns to the network [8]. On this basis the control system connects the necessary number of cells, following the law:

$$n = [(W_1 + W_3) / W_2], \quad (2)$$

where n is the number of cells. Thus, the allowable range of the load current is increased.

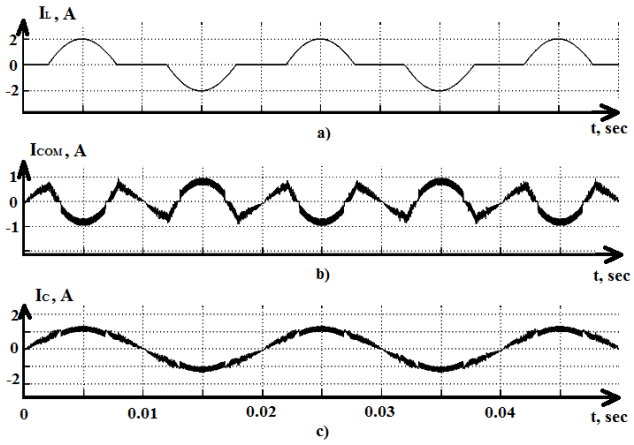


Fig. 5. APLC normal operating diagrams

The two parallel APLC using reduces the input current THD from 0.48 (Fig. 5a) to 0.11 (Fig. 5c), which corresponds to the standard's maximum allowable values [9]. Compared to the current form shown on Fig. 3c, THD decreases by more than 3.5 times – from 0.4-0.6 to 0.11.

IV. Multilevel Inverter Operating Modes

Fig.6 represents current-source multilevel inverter as a reactive power compensator. This solution allows not only to increase the compensator rated power, but also provide it operation with no output current distortion by connecting an inverter cells combination of appropriate power.

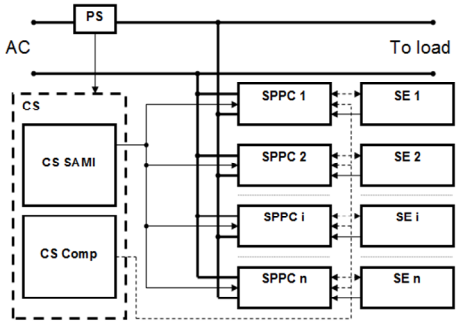


Fig. 6. Multilever current-source inverter block scheme and connection

Multilevel inverter consists of n single-phase power converters (SPPC), which are connected to power source elements (SE). Accumulators, alternative sources or capacitors could operate as SE. In operation mode of each inverter cell with the corresponding output filter the standard sinusoidal PWM-technique ensuring THD about 3-5% is used. This rate can be reduced to <0.1-1% with the same smoothing filter as in the case of single-level inverter by using at least three-level inverter [3].

The circuit operates in two modes depending on the availability of AC line voltage: as VAR-compensator when stable AC line is available and as standalone multilevel inverter otherwise. Control system (CS) realizes two operation modes for the same power scheme. The CS SAMI control system part realizes multilevel inverter control algorithms, while the VAR-compensation is made through CS Comp.

In the inverter mode CS generates control pulses (continuous arrows on Fig.6) for switching elements of power converters SPPC, which provides the sinusoidal current.

In the VAR-compensation mode SPPC operates as of buck or boost PWM-converter. The signal from power sensor (PS) impacts the CS, which define the number of parallel inverter cells. It provides no distortion operation as showed on Fig. 5.

In order to change the system's mode it is necessary to change the CS algorithm only.

V. Conclusions

The multilevel inverter's cells usage depends on the AC line voltage availability. While AC line voltage is available the multilevel inverter's cells are used as the APLC and eliminate the distortion by inverter's cells required number connecting. Thus no distortion load current allowable range increasing is achieved.

In the other mode (no AC line voltage) all the multilevel inverter's cells form output sinusoidal current. THD of current <1% in comparison to 3-5% of typical inverters. This technology reduces the dimensions and the cost of the uninterruptible power supply system by reducing the inverter's output filter rated power and using less powerful switching elements. The same control system provides two different operation modes by using appropriate control algorithms.

References

[1] M. Lesovaya, E. Pichkalev, T. Tereshchenko Y. Hohlov, "Power system with alternative sources modeling", Elektronika i svyaz, 2010, vol.4, pp. 91-95.

[2] A. Garganeev, "Emergency power supply for AC responsible consumers", Tomsk, Uchebnoe posobie: Tomskiy politehnicheskiy universitet, 2010, p.190.

[3] V.Bezhenar, T.Tereshcenko, J.Yamnenko, "Multilevel inverters based on m-ary system", Kyiv, Electronics and nanotechnology, 2012, pp.23-24.

[4] V. Glushkov, V Gribin, "Power plants VAR-compensation", Moskva: Energiya, 1975.

[5] B. Konstantinov, G. Zaycev, "VAR-compensation", Moskva: Energiya, 1976, pp.56-59.

[6] A. Volkov, V. Metelsky, V. Volkov, "Research of the AVC for four wire three-phase AC mains", Tehnicheskaya elektrodimamica, 2010, vol.6, pp.61-68.

[7] F. Govorov, V. Govorov, A. Ganus, O. Tereshyn, "VAR-compensation in illuminating power linesconsidering nonlinear parameters", Silovaya elektronika i energoefektivnost, 2010, vol.2, pp. 179.-184.

[8] V. Zhuikov, Nguyen Hai Tung, "Paralel type high frequency converter VAR-compensation", Elektronika i svyaz, 1998, vol.5, pp. 82-85.

[9] GOST 13109-97

Modeling of Electrical and Vibration Signals of Transformers with Different Magnetic Properties

Valeriy Zhuikov, Nikolay Kuznietsov
Department of Industrial electronics, Faculty of Electronics
National Technical University of Ukraine "KPI"
Kyiv, Ukraine
nikolay_kuznyetsov@i.ua

***Abstract*—The modeling of currents and vibration signals of two-winding transformers which have different magnetic characteristics are considered in the article. The impact of different types of magnetization curves on the electrical and vibration transformer signals is shown. The possible applications for the transformers, which cores are characterized by different magnetization curves, on the basis of ensuring the least vibration are noted.**

Keywords—transformer model; magnetization curve; vibration signals

I. INTRODUCTION

While calculating the vibroacoustic diagnostic and control systems of electric power semiconductor converters it is necessary to model the electrical and vibration signals of the converter's power transformer [1, 2, 3]. Transformer signals depend on the magnetic characteristics of its core [4], which allows evaluating the converter operation mode by using vibration signals. In modern engineering more attention is given to transformers vibrations of medium and high power, designed mainly for low frequency sinusoidal currents, and not designed for operation in the saturation mode. The vibration signals of such transformers are well described [4]. The model of low power transformer designed for pulsed high-frequency currents is presented in the article. The model simulates the transformers operation in the power semiconductor converter and allows estimating the waveform and amplitude of the transformers electrical and vibration signals.

II. MATHEMATICAL RELATIONS

For electrical signals modeling the circuit T-model of two-winding transformer [5] showed on Fig. 1 was used.

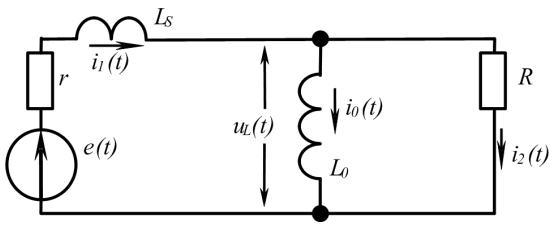


Fig. 1. Transformer's T-model.

On the transformer's circuit model showed on Fig. 1 L_S is the leakage inductance, r – the primary winding resistance, L_0 – the magnetizing inductance, R – the load resistance, $e(t)$ – the voltage source model, $i_0(t)$, $i_1(t)$, $i_2(t)$ – the magnetizing current, currents of primary and secondary windings, $u_L(t)$ – the inductance L_0 voltage.

The general form of the equation for voltages and currents while modeling the voltage source $e(t)$ with linear functions is as following [3]:

$$\begin{aligned}
&e(t)=E/t_f\cdot t\Big|_{t=0\ldots t_f};\\
&e(t)=E\Big|_{t=t_f\ldots t_p-t_f;t=T/2+t_p\ldots T/2+t_p+t_z};\\
&e(t)=E-E/t_f\cdot\left(t-t_p+t_f\right)\Big|_{t=t_p-t_f\ldots t_p};\\
&e(t)=-E\Big|_{t=t_p\ldots t_p+t_z;t=T/2+t_f\ldots T/2+t_p-t_f};\\
&e(t)=0\Big|_{t=t_p+t_z\ldots T/2;t=T/2+t_p+t_z\ldots T};\\
&e(t)=-E/t_f\cdot\left(t-T/2\right)\Big|_{t=T/2\ldots T/2+t_f};\\
&e(t)=-E+E/t_f\cdot\left(t-T/2-t_p+t_f\right)\Big|_{t=T/2+t_p-t_f\ldots T/2+t_p};
\end{aligned}\tag{1}$$

where E is the voltage amplitude, T is the signal period, t_f is front duration, t_p is the voltage pulse duration, t_z is the reverse voltage pulse duration, simulating the flow of current in the reverse diodes of source $e(t)$ prototype – half-bridge inverter – under the influence of the transformer inductance;

$$\begin{aligned}
&i_0(t)=A+B\cdot(t-t_1)^2+C\cdot(t-t_1)+\\
&+D\cdot e^{-\frac{R\cdot(Ls+L_0)+L_0\cdot r}{LsL_0}(t-t_1)}\Big|_{t=t_1\ldots t_2};\\
&i_1(t)=(A+O)+B\cdot(t-t_1)^2+(C+P)\cdot(t-t_1)+\\
&+(D+Q)\cdot e^{-\frac{R\cdot(Ls+L_0)+L_0\cdot r}{LsL_0}(t-t_1)}\Big|_{t=t_1\ldots t_2};\\
&i_2(t)=O+P\cdot(t-t_1)+Q\cdot e^{-\frac{R\cdot(Ls+L_0)+L_0\cdot r}{LsL_0}(t-t_1)}\Big|_{t=t_1\ldots t_2};
\end{aligned}\tag{2}$$

where A, B, C, D, O, P, Q – the numerical coefficients, individual for each time interval $t_1 \ldots t_2$. Currents in the circuit model showed on Fig. 1 can be easily found with a fitting method [6] by using (1) and (2), as the source $e(t)$ voltage and the inductance L_0 values are described by linear functions on a finite time intervals.

The transformer core's material magnetic characteristics are defined by magnetization curves [7] and accounted for the nonlinear inductance L_0 as showed on Fig. 1. Value L_0 is defined by the magnetization curves of Fig. 2 with the current values of the magnetizing current $i_0(t)$ and/or the integral of the inductance L_0 voltage $u_L = i_2(t) \cdot R$. $L_0 = \mu \cdot \mu_0 \cdot N^2 \cdot S / l_{av}$ [5], where μ is the core material magnetic permeability; μ_0 is the magnetic constant; N is the number of turns in the winding forming the core magnetic flux; S is the core cross-section, l_{av} is the effective magnetic path length.

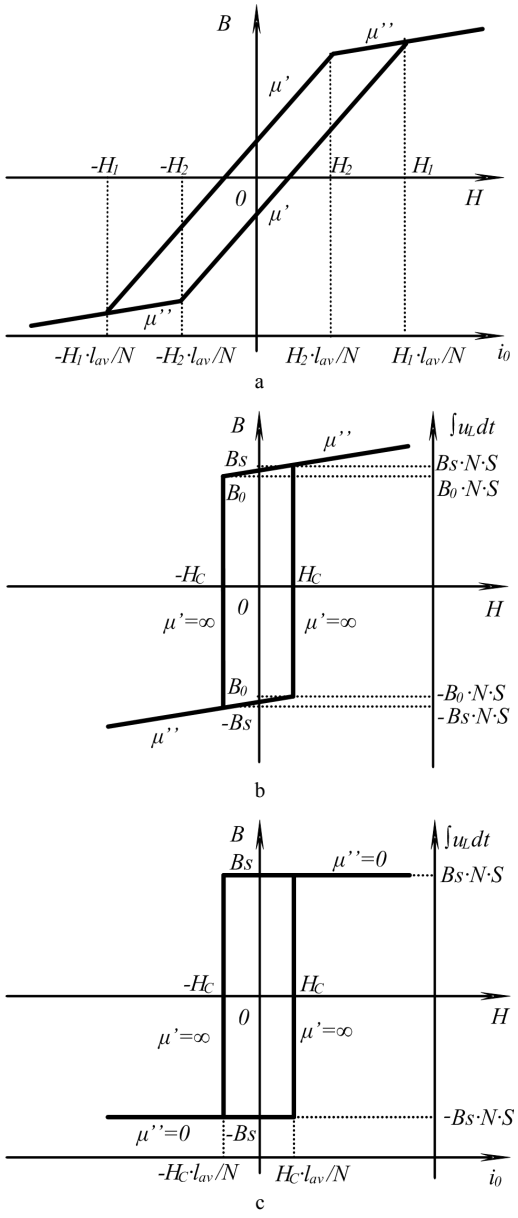


Fig. 2. Core magnetization curves. **(a)** – curve "1", **(b)** – curve "2", **(c)** – curve "3".

Magnetization curve "1" on Fig. 2 **(a)** is characterized by finite non-zero values of the core material magnetic permeability μ and, consequently, of the inductance L_0. In this case the condition $\mu' >> \mu''$ is satisfied. Determining the magnetic permeability current value in the circuit model on Fig. 1 for the curve "1" is convenient as for magnetizing current $i_0(t)$ magnitude on Fig. 2 **(a)**, so for the integral $\int u_L(t)dt$ magnitude. Such magnetization curve is usual for the materials used in power transformers of power semiconductor converters [5, 8]. Because of the finite nonzero values of the magnetic permeability μ and inductance L_0 during the whole signals period, the currents calculation in the circuit model on Fig. 1 with using (2) does not allow simplifications on individual intervals of the considering process.

Magnetization curve "2" on Fig. 2 **(b)** is characterized by finite non-zero values of the core material magnetic permeability $\mu = \mu''$ with $|\int u_L(t)dt| > N \cdot S \cdot |B_0|$ and infinitely large magnetic permeability $\mu = \mu' = \infty$ for other values of the integral $\int u_L(t)dt$. Determining the inductance L_0 current value in the circuit model Fig. 1 for this curve is carried out by the integral of the voltage $u_L(t)$ value on Fig. 2 **(b)**. For the curve areas with the finite magnetic permeability ($\mu = \mu''$) it is also convenient to use the magnetizing current $i_0(t)$ magnitude. Magnetization curve "2" on Fig. 2 **(b)** is typical for the materials used in cores of saturable transformers and throttles [9]. In cases when it is possible to neglect the waveform of the magnetizing current $i_0(t)$, the curve "2" is used as a simplification of the curve "1". With this the calculation of L_0 and transformer currents from (2) for intervals $\mu = \mu'' = \infty$ is simplified.

Magnetization curve "3" on Fig. 2 **(c)** is different from the curve "2" on Fig. 2 **(b)** with material magnetic permeability zero values $\mu = \mu'' = 0$ with $|\int u_L(t)dt| > N \cdot S \cdot |Bs|$ and infinitely large magnetic permeability $\mu = \mu' = \infty$ for other values of the integral $\int u_L(t)dt$. Determining the magnetic permeability current value and corresponding inductance L_0 in the circuit model on Fig. 1 for the curve "3" is carried out by both the voltage $u_L(t)$ integral and magnetizing current $i_0(t)$ values. This magnetization curve is used as a further simplification of the curve "2", and also when the value of the magnetic permeability μ with $i_0(t) > H_C \cdot l_{av}/N$ is negligible. Using the properties of curves "2" and "3" $\mu = 0$ and $\mu = \infty$, simplifies equation (2) by turning to zero some coefficients.

With the known magnetic and structural parameters of the transformer, as well as time currents changes the core surface movement function $x(t)$ is described as following [3,10]:

$$x(t) = h \cdot \frac{3 \cdot \lambda_S}{2 \cdot \left(B_S / \mu_0 - H_S \right)^2} \cdot \left(i_0(t) \cdot N / l_{av} \right)^2 \cdot (\mu(t) - 1)^2, \quad (3)$$

where λ_S is the elongation in saturation, B_S and H_S are the induction and the magnetic field in saturation, $\mu(t)$ is the function of changes of core material magnetic permeability μ in time, h is the length of the core in a perpendicular direction to the moving surface.

The function of coil surface movement $y(t)$ is described by equation [3]:

$$y(t) = \mu_l \cdot \mu_0 \sum_{j=1}^{2} \frac{D_j / K_1 + d_j / K_2}{(D_j + d_j) D_j} \cdot (i_j(t))^2 \cdot (n_j - 1), \quad (4)$$

where μ_l is the magnetic permeability of the lacquer turns insulation, n_j is the number of turns layers in the j winding, D_j is the diameter of the j winding wire, d_j is the thickness of insulation between the wires of j windings, K_1 is the elastic modulus of the turns material, K_2 is the elastic modulus of insulation material between the windings. Note that equations (3) and (4) do not take into account the inertia of the core and windings.

The core vibration is caused by magnetostriction [4,10], and depends on the magnetizing current, geometric and magnetic core parameters, the number of winding turns. The coil movement is caused by the interaction of windings turns under Lorentz force [4] and depends on the number of windings, number of turns in them, diameters and wire material, insulation, flowing currents.

As follows from (3) and (4) the core and windings movements are determined by the squares of the flowing currents, and the vibration period is two times less than the currents change period.

III. Modeling of Electric and Vibration Signals

For currents $i_0(t)$, $i_1(t)$, $i_2(t)$, core $x(t)$ and coil $y(t)$ surface movement functions modeling a transformer with the following parameters was used: **1)** the core standard size is EI33; **2)** parameters of the magnetization curve "1" (Fig. 2 **(a)**) Bs = 0.53 T, B_0 = 0.5282 T, H_1 = 51.2778 A/m, H_2 = 25.1480 A/m; **3)** parameters of the magnetization curve "2" (Fig. 2 **(b)**) Bs = 0.5274 T, B_0 = 0.5256 T, H_C = 13 A/m; **4)** the parameters of the magnetization curve "3" (Fig. 2 **(c)**) Bs = 0.5274 T, H_C = 13 A/m; **5)** two-winding transformer, wound with lacquer insulation copper wires tightly pressed together; **6)** the seal between the windings is inelastic and diamagnetic; **7)** the primary winding has 18 turns, wrapped in 2 layers, the diameter of wire without insulation is 1.12 mm, the outer wire diameter is 1.247 mm; **8)** the secondary winding has 7 turns, wrapped in 2 layers, the diameter of wire without insulation is 2 mm, outer wire diameter is 2.172 mm.

Transformer's input is connected to the voltage source $e(t)$ with the following parameters: **1)** the voltage is bipolar and symmetric; **2)** the pulse duration $t_p = 0.8 \cdot T/2$; **3)** the fronts duration $t_f = 0.2 \cdot t_p$; **4)** the amplitude E = 97 V; **5)** the period of voltage in the normal operation mode is T = 0.05 ms, in the saturation operation mode T = 0.1429 ms, **6)** the limitation level for primary current is 22 A. The transformer's output is connected to the resistive load R with resistance of 2.88 Ohm. The transition between normal and saturated transformer operation modes is achieved by changing the voltage $e(t)$ (1) period T.

A. Transformer Currents in Normal Operation Mode

Currents of primary $i_1(t)$ and secondary $i_2(t)$ windings, magnetizing current $i_0(t)$ for magnetization curves on Fig. 2 are showed on Fig. 3, where $i_{01}(t)$, $i_{02}(t)$, $i_{03}(t)$ are magnetizing currents for transformer models with magnetization curves "1", "2", "3" Fig. 2 **(a)**, **(b)**, **(c)** respectively.

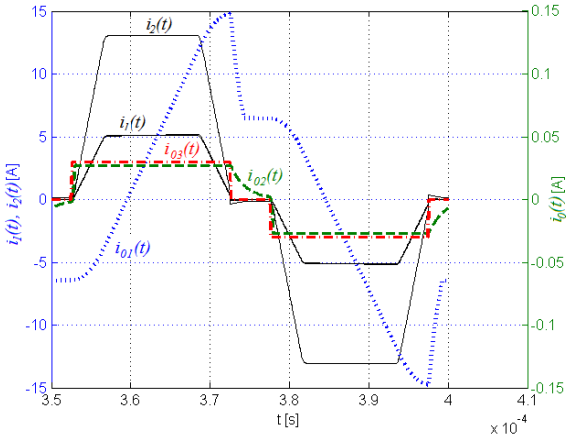


Fig. 3. Currents of the transformer in normal operation mode.

The current $i_0(t)$ amplitude on Fig. 3 is determined by the magnetic field variation range in the core, which for the magnetization curve "1" is wider than for curves "2" and "3" Fig. 2. Constant current $i_0(t)$ values on Fig. 3 for transformer models with curves "2" and "3" on the intervals 3.52...3.72 and 3.78...3.98 x10^{-4} s are caused by circuit model's (Fig 1) infinitely large magnetic permeability μ and inductance L_0. Differences in the magnetization currents waveforms for models with magnetization curves "2" and "3" for the range 3.72...3.78 x10^{-4} s are caused by different values of magnetic permeability μ and corresponding inductance L_0. Since the magnetization currents amplitudes are significantly less than windings currents ones, then the last are of little dependence on the core magnetization curve and their waveform repeats the voltage source $e(t)$ waveform.

B. Transformer Vibration in Normal Operation Mode

The core $x(t)$ and coil $y(t)$ surface movement functions for different magnetization curves are showed on Fig. 4, where $x_1(t)$, $x_2(t)$, $x_3(t)$ – the core surface movement functions for transformer models with magnetization curves ""1", "2", "3" Fig. 2 **(a)**, **(b)**, **(c)** respectively.

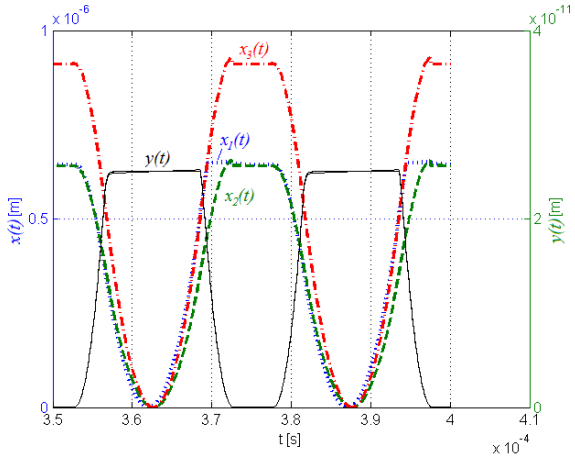


Fig. 4. Vibration signals of the transformer in normal operation mode.

Differences in the core and coil surface movement functions' amplitudes and waveforms on Fig. 4 for models with different magnetization curves are caused by differences in the corresponding magnetization currents Fig. 3. Since the core movement function depends not only on the magnetizing current, but also on the magnetization curve parameters (3), the differences in the core vibration amplitudes for the models with different magnetization curves are not proportional to the corresponding differences in the magnetizing currents amplitudes. Coil movement functions $y(t)$ Fig. 4 for models with different magnetization curves are different insignificantly.

C. Transformer Currents in Saturated Operation Mode

Primary winding current $i_1(t)$ and transformer magnetizing current $i_0(t)$ for different magnetization curves are showed on Fig. 5 **(a)**, secondary winding current $i_2(t)$ – on Fig. 5 **(b)**. $i_{01}(t)$, $i_{02}(t)$, $i_{03}(t)$ are magnetizing, $i_{11}(t)$, $i_{12}(t)$, $i_{13}(t)$ are primary winding and $i_{21}(t)$, $i_{22}(t)$, $i_{22}(t)$ are secondary winding currents for the transformer models with magnetization curves "1", "2" and "3" Fig. 2 **(a)**, **(b)**, **(c)** respectively.

Windings and magnetization currents on the intervals 1...1.038 and 1.072...1.108 ms on Fig. 5 **(a)** and **(b)** correspond to the transformer currents in the normal operation mode. On intervals 1.038…1.072 and 1.108…1.142 ms the difference between currents for models with magnetization curves "1" and "2" is inessential since the values of magnetic permeability $\mu = \mu''$ and inductance L_0 on these intervals are equal. For the model with magnetization curve "3" on the specified intervals currents have steep fronts, close currents $i_1(t)$ and $i_0(t)$ amplitudes Fig. 5 **(a)**, as well as the current $i_2(t)$ zero value Fig. 5 **(b)**, due to the influence of inductance L_0, which has zero value. Current $i_2(t)$ pulses on the interval 1.065…1.08 ms appears due to the influence of inductances Ls and L_0 Fig. 1 on the voltage source $e(t)$ (1).

D. Transformer Vibration in Saturated Operation Mode

Core $x(t)$ and coil $y(t)$ surface movement functions for magnetization curves Fig. 2 are showed on Fig. 6 **(a)** and **(b)** respectively. $x_1(t)$, $x_2(t)$, and $x_3(t)$, $y_1(t)$, $y_2(t)$, and $y_3(t)$ are core and coil surface movement functions for transformer models with magnetization curves "1", "2" and "3" Fig. 2 **(a)**, **(b)**, **(c)**.

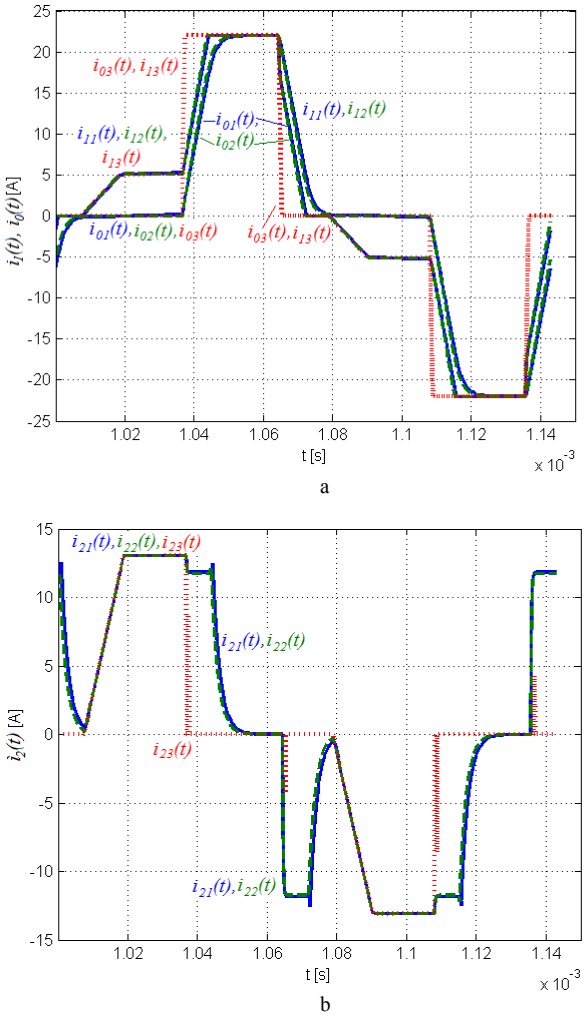


Fig. 5. Currents of the transformer in saturated operation mode.

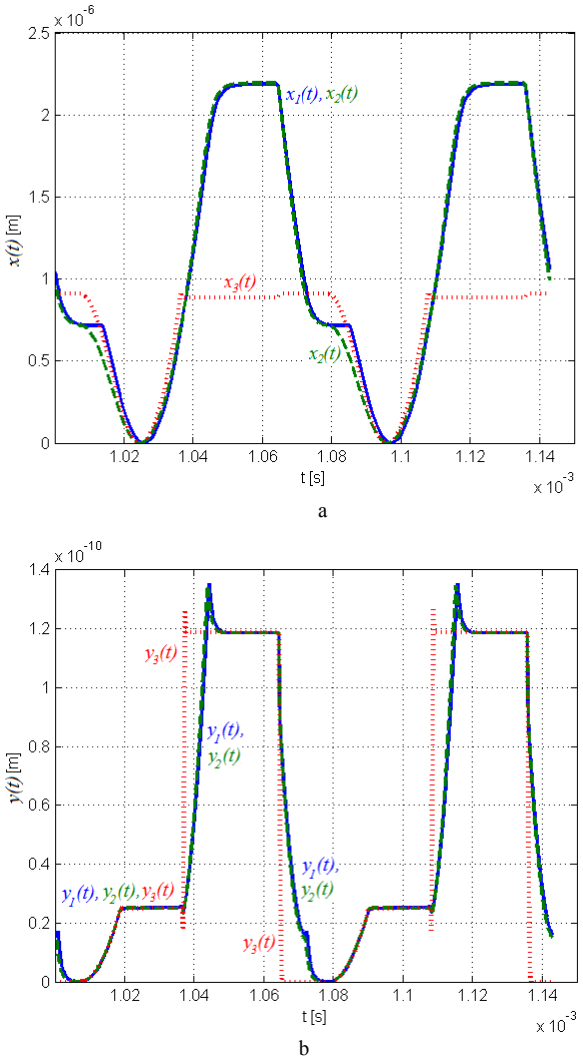


Fig. 6. Transformer vibration signals in saturated operation mode.

The core and windings surface vibration functions for models with magnetization curves "1" and "2" have slight differences showed on Fig. 6, which are caused by the close transformer current values on Fig. 5. The shape of core surface vibration functions for model with curve "3" is different from the shapes of core surface vibration functions for models with other curves because of the influence of the curve on Fig. 2 **(c)** parameters on equation (3). The shape of winding surface vibration functions for this model on the intervals 1.038…1.072 ms and 1.108…1.142 ms is characterized by steep fronts because of inductance L_0 zero value.

The part of the magnetization curve with high magnetic permeability values significantly affects the transformer signals in normal operation mode, with low magnetic permeability values – in saturation mode. This statement remains true for vibration signals as well as for electrical ones.

Consider the transformers model vibration signals. The vibration signals amplitudes of windings $y(t)$, $y_1(t)$, $y_2(t)$, $y_3(t)$ on Fig. 4, Fig. 6 **(b)** for all of the magnetization curves on Fig. 2 **(a)**, **(b)**, **(c)** increase significantly during the transformers transition from the normal to saturated operation mode. The signals amplitudes of windings $y(t)$, $y_1(t)$, $y_2(t)$, $y_3(t)$ are much lower than the amplitudes of the core vibration signals $x(t)$, $x_1(t)$, $x_2(t)$, $x_3(t)$ on Fig. 4, Fig. 6 **(a)**.

In transformers normal operation mode the highest vibration signal amplitude $x_3(t)$ on Fig. 4 has the core from the material with a magnetization curve "3" on Fig. 2 **(c)**. The cores from materials with curves "1" on Fig. 2 **(a)** and "2" on Fig. 2 **(b)** generate signals with almost the same amplitudes. The area under the vibration signal $x_3(t)$ on Fig. 4 of the core from the material with magnetization curve "3" on Fig. 2 **(c)**, is also bigger than for other reviewed materials. This indirectly indicates the greater core losses and then generally the greater transformers noise. For transformers that operate mainly in normal mode, it is rational to use cores from materials with magnetization curve "2" on Fig. 2 **(b)** as the cores correspond to a smallest generated signal $x_2(t)$ bounded area on Fig. 4. Such cores magnetization curves are characterized by the highest values of the magnetic permeability μ' and μ''. Transformers with such cores are less noisy and more energy efficient.

In the transformers saturated operation mode the amplitude of the vibration signal $x_3(t)$ in Fig. 6 **(a)** of the core from material with a magnetization curve "3" on Fig. 2 **(c)**, is significantly less than the vibration amplitude of the cores from other materials. The area under the signal $x_3(t)$ on Fig. 6 **(a)** is also smaller that indicates smaller cores losses. It is advisable to carry out the transformers cores, designed to operate in saturated mode, from materials with magnetization curve "3" on Fig. 2 **(c)**. Due to a steeper signals front $y_3(t)$ shown on Fig. 6 **(b)**, the losses in the transformers windings with core from material with magnetization curve "3" on Fig. 2 **(c)**, may be greater than in the transformers windings with cores from materials with other magnetization curves. The analysis of the working point position on the transformers cores magnetization curves that operate in normal mode, shows that the greatest influence on vibration signal is made by the magnetic curves area with magnetic permeability μ''. The smaller is the value of μ'', the lower is the amplitude of the cores vibration signal.

Comparing the transformers vibration signals waveform in the normal and saturated operation modes Fig. 4 and Fig. 6 **(a)**, the amplitude of the cores signal $x_3(t)$, which corresponds to the magnetization curve "3" on Fig. 2 **(c)**, is stored while the transformers transition from the normal to saturated operation mode. For the cores from materials with other magnetization curves in such a transition the amplitude of vibration signal $x_1(t)$, $x_2(t)$ is substantially. The transformers with cores from materials with magnetization curve "3" on Fig. 2 **(c)**, are justified to apply in the case of transformers transition from the normal to the saturated operation mode and conversely, ensuring the low vibration.

IV. Conclusion

The areas of core magnetization curves with high magnetic permeability ($\mu = \mu'$) are crucial for transformers signals in normal operation mode. Areas with low values ($\mu = \mu''$) – in a saturated mode. This allows considering the magnetic permeability only for certain areas of the magnetization curves in the normal and saturated modes. The analysis of the transformer vibrations shows that with a small number of tightly pressed to each other turns the core is the main vibration source of a low-power high-frequency transformer. In normal operation mode, it is appropriate to use cores from materials with high magnetic permeability μ' and μ'' (curve "1" and "2" on Fig. 2 **(a)**, **(b)**), and in a saturated operation mode, or with the transition from normal mode to saturated, – cores from materials with low values of permeability μ'' (curve "3" on Fig. 2 **(c)**).

References

[1] A.A. Bagaiev, "Vibroacoustic diagnostics as monitoring subsystem of power equipment", Sbornik nauchnykh trudov. Sieviero-Kavkazkoie otdielieniie Akadiemii tiekhnologichieskikh nauk Rossiiskoi Fiedieratsyy, Vol. 3. Yestiesvtonauch., 2000, pp.71-74.(Rus)

[2] P. Micheau, E. Leboucher, A. Berry, "Implementation of decentralized active control of power transformer noise", Université de Sherbrooke, Canada.-2004

[3] N. N. Kuznietsov, "Vibroacoustic signals modeling of semiconductor converter's power transformer", Tekhnichna elektrodynamika, tematychnyi vypusk Sylova elektronika ta enerhoefektyvnist, chastyna 3, Kharkiv: NTU «HPI», 2012, pp. 172–177.(Rus)

[4] M. van der Veen, "Measuring Acoustic Noise Emitted by Power Transformers", 109th convention, Los Angeles, 2000

[5] B. Yu. Siemonov, "Power electronics for hobbyists and professionals", Moskva: Solon-R, 2001, 327 p.(Rus)

[6] Ye. P. Popov, "The theory of nonlinear automatic regulation and control systems", Moskva: Nauka, 1988, 256 p.(Rus)

[7] K.P. Bielov, "Magnetostrictive phenomena and their technical applications", Moskva: Nauka, 1987, 160p.(Rus)

[8] R. Mack, "Switching power supplies. The theoretical basis of the design and guidance on the practical applications", Moskva: Dodeka -XXI, 2008, 272p.(Rus)

[9] I.M. Chizhenko, V.S. Rudenko, V.I. Sienko, "Fundamentals of power conversion technique ", Moskva: Vysshaia shkola, 1974, 430p. (Rus)

[10] R. Szewczyk, "Magnetic and magnetostrictive properties modeling of high permeability Mn-Zn ferrites", Pramana, Vol. 67, No. 6, December 2006, pp. 1165-117.

Optimal Control of Uninterruptible Power System with Diesel Generator

Ievgen Pichkalov, Anna Kyselova
Department of Industrial Electronics, College of Electronics
National Technical University of Ukraine
"Kyiv Polytechnic Institute"
Kyiv, Ukraine
pichkalyov@ieee.org

Olga Kyselova[1], Gennadiy Kyselov[2]
[1] Intercollegiate College of Medical Engineering
[2] ESC "Institute for Applied System Analysis"
National Technical University of Ukraine
"Kyiv Polytechnic Institute"
Kyiv, Ukraine
olga.mmif@gmail.com

***Abstract* — Conditions to minimize fuel consumption in uninterruptible power system with diesel generator at variable load are obtained. Estimation of fuel consumption at acceleration and braking stages of diesel generator was executed taking into account the cost of battery charge. Minimization of fuel consumption required for the transition process is achieved by adjusting the position of the fuel valves. Minimization of fuel consumption is shown for diesel generator HATZ 1D90-Supra.**

Keywords— storage battery; uninterruptible power system; fuel consumption; diesel generator; profit

I. INTRODUCTION

Uninterruptible power systems (UPS) based on using multiple uninterruptible power supply additionally to power network in common scheme allow to provide reliable power supply for the loads [1]. The most common in UPS is combination of diesel generator (DG) and storage battery (SB) [2]. In such case realization of control allowing the most efficient use of DG resources, i.e. minimization of fuel consumption, is an important task.

II. OPERATION OF POWER SYSTEM WITH DIESEL GENERATOR

Typical structure of UPS with DG shown in Fig. 1.

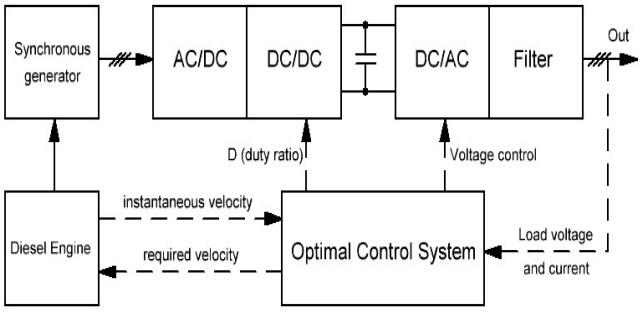


Fig.1. Structure of uninterruptible power systems with diesel generators

Let's consider operation of uninterruptible power supply system with DG and SB (Fig.2).

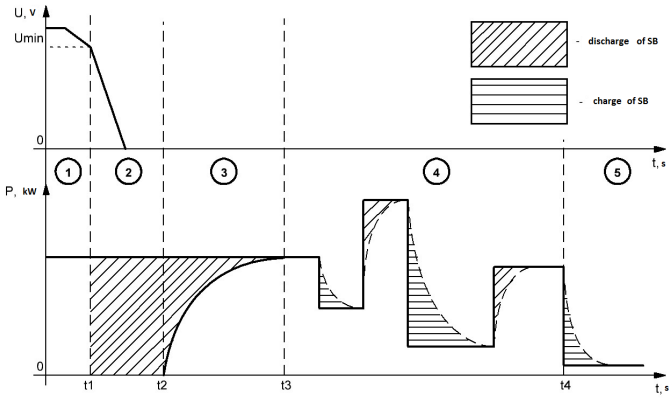


Fig.2. Operation steps of the DG and SB after disconnecting of power network (horizontal hatching marks SB charge, angle hatching marks SB discharge intervals)

In the moment t_1 power network voltage is reduced below acceptable level. Then at the interval $[t_1;t_2]$ the load is supplied by SB. At the same time DG performs all necessary actions for it's starting. If power network voltage is not restored to the required level, then in the moment t_2 DG starts working and achieves necessary power P_o in the moment t_3. At the same time SB discharges to the load (Fig.1). At next interval $[t_3;t_4]$ DG accelerates or decelerates depending on load requirements, and SB discharges or charges accordingly. At the last interval $[t_4;\infty]$ diesel generator decelerates to some minimum value of power in order to maintain the charge level of SB, or turns off if power network voltage appears.

At each of the considered time intervals it is important to maintain optimal work of diesel generator with minimization of fuel consumption.

III. Optimal Control For Minimization Of Fuel Consumption

To solve the problem of fuel consumption minimizing let's consider interval $[t_3; t_4]$ where diesel generator accelerates and decelerates (Fig.1). DG is described by the equation of first order with mechanical constant τ of the system "DG-synchronous generator", P_m is the maximum power that can be obtained at assigned value of relative fuel consumption $B'(P_m)$ [2, 3]. Relative fuel consumption is described by the equation:

$$B'(P_m) = a \cdot P_m^2 + b \cdot P_m + c \quad [l/(kW \cdot h)]$$

where the coefficients a, b, c are determined by the type of diesel engine. Total fuel consumption for DG acceleration taking into account SB charge is defined as:

$$B_{roz} = B'(P_m) \cdot P_m \cdot T_0 - \frac{1}{\eta} \cdot B'(P_m) \cdot (P_m \cdot \ln\left(\frac{P_m - P_0}{P_m}\right) - P_0) \cdot \tau,$$

where T_0 – time of transient process, η – efficiency of charge controller for SB.

Fuel consumption at DG deceleration interval is defined similarly. Dependences of fuel consumption for DG HATZ 1D90-Supra from acceleration power (curve 1) and deceleration power (curve 2) are presented in Fig. 2. As it's shown on Fig. 2, curves have minimums. It allows to determine the condition of fuel consumption minimizing. To achieve minimum fuel consumption for HATZ 1D90-Supra acceleration power should be equal $P_m = 7,91 kW$, and deceleration power - $P_m = 8,05 kW$.

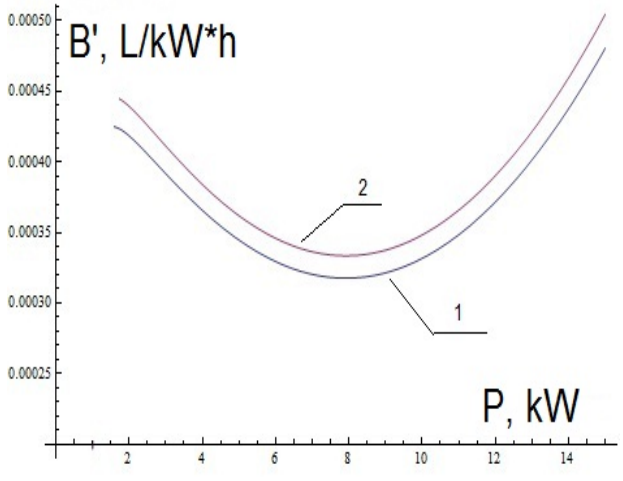


Fig. 3. Fuel consumption according to the acceleration (deceleration) power

IV. Optimal Control For Maximization Of Profit

An important task is to obtain maximum economic benefits for each interval (Fig.4) of the system [4]. The economic benefit is described as the profits.

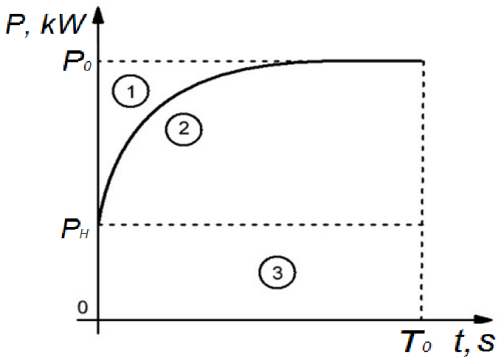


Fig. 4. Power acceleration of diesel generator

Profit for the acceleration interval described by the equation:

$$\Pi = \Pi_{SB} + \Pi_{DG} = \int_0^{T_0} P_0 dt \cdot Q_{SB} +$$

$$+ \int_0^{T_0} P_2(t) dt \cdot (C_2 - Q_{SB}) + T_0 P_3(t) - Q_{DG}$$

where Π – profit; Π_{SB} – profit of SB; Π_{DG} – profit of DG; T_0 – time of transient process; Q_{SB}, Q_{DG} – cost factors.

Task of control system is maximizing profits by changing the value of the fuel consumed by diesel generator.

Profit for the deceleration interval described by the equation (Fig.5):

$$\Pi = \int_0^{T_0} P_0 dt \cdot Q_{SB} + T_0 P_H C_2(t) - T_0 P_2(t) \cdot Q_{DG}$$

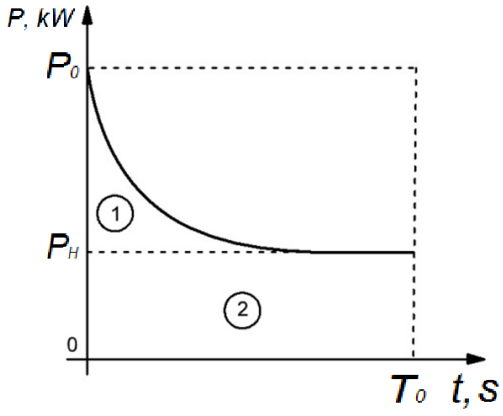


Fig. 5. Power deceleration of diesel generator

To calculate the criteria necessary to repeatedly calculate the integral of power, that results to time delays and loss of stability. Order to eliminate these drawbacks in proposed digital control system to make an analog part, which would have accelerated the computation of the integral of power.

V. Analogue-Digital Control System

Analog-digital control system consists of a digital (microcontroller) and analog system. The analog part consists of two integrators and a comparator. With a digital system by the DAC in the analog system entered the current values of power, acceleration and power output to be achieved.

Analog system calculates the time value of transients and the integral of power. Through the ADC these values are entered in the microcontroller (Fig.6).

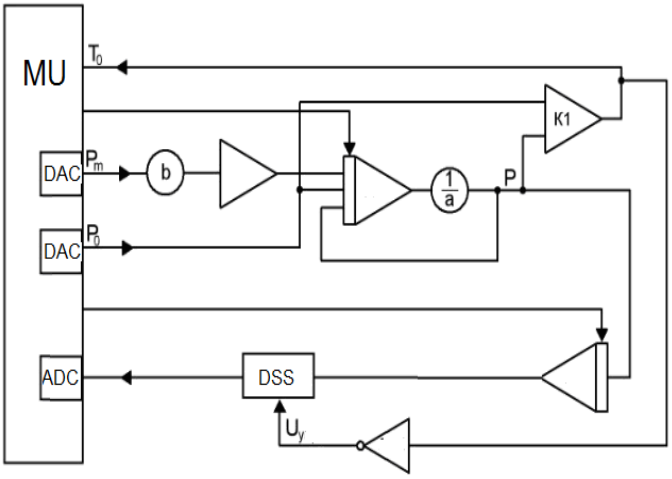


Fig. 6. Analog-digital control system

Microcontroller iterates the values of acceleration and deceleration power for as long as profits on current iteration does not exceed the value of profits in the previous iteration (Fig.7). The value of time transients obtained by comparing the current power and power to be achieved. UPS control system defines position of fuel valve providing maximization of profit at each interval of the transient process.

VI. Conclusion

Conditions of fuel consumption minimization for diesel generator HATZ 1D90-Supra are obtained. Realization of these conditions in control system allows to reduce fuel consumption by 17% at acceleration and deceleration intervals. Control system provides minimization at all intervals of increasing or decreasing of the load power.

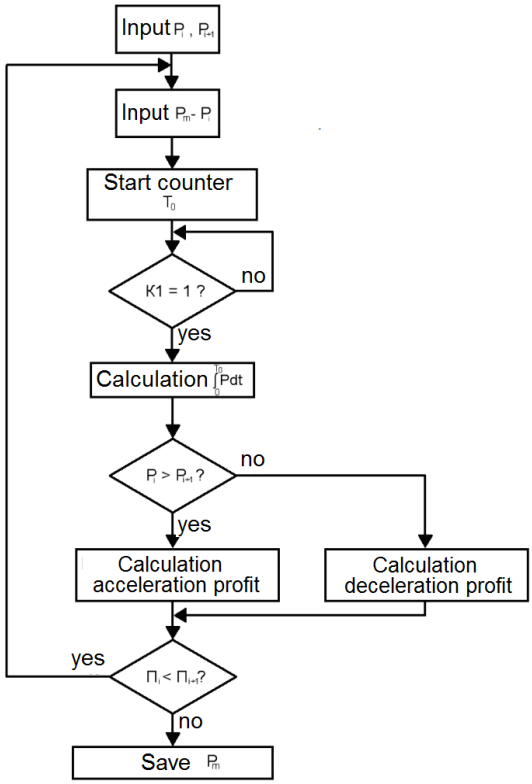


Fig. 7. Working algorithm of analog-digital control system

UPS control system defines position of fuel valve providing minimization of fuel consumption at each interval of the transient process.

Conditions of profit maximization for diesel generator and storage battery are obtained. Control system provides maximization profit at all intervals of increasing or decreasing of the load power. Proposed control systems based on digital and analog parts. The combination of analog and digital systems increases performance computing systems.

References

[1] Kamaev V.S., Petergerya Y.S. Optimization of operation of backup power system with diesel generator // Pratsi Instytutu elektrodynamiky Natsionalnoi Akademii Nauk Ukrainy. – 2006. - №1(13). – p. 38-41. (Rus.)

[2] Khyzhnyak T.A., Pichkalyov E.S., Nevmerghiskiy O.V. Power management of a diesel generator with criterion of cost minimizing // "Tekhnichna elektrodynamika", Special Issue of "Power electronics and energy efficiency". – 2010. – Ch.2. – p. 122–127. (Rus.)

[3] Pichkalyov E.S. Power management of the local object taking into account of cost factors // "Tekhnichna elektrodynamika", Special Issue of "Power electronics and energy efficiency". – 2010. – Ch.1. – p. 126–130. (Rus.)

[4] Petergerya Y.S., Zhuikov V.Y., Tereschenko T.O. Intellectual systems of energy saving homes. Textbook. – K.: Media-PRESS, 2008. – p. 256 (Rus.)

Design an Algorithm to Syntheses DC-DC Converters

Murad Ahmed Ali Taher
Department of Computer Engineering
Hodeidah University, Yemen
dr_muradabsi@yahoo.com

Abstract — **This paper proposes an algorithmic method for generating DC–DC converters with galvanic isolation. This method uses the Main Topological Matrix MTM. The definition and attributes of the MTM are stated. The common properties for single-switch DC-DC converters are described. The proposed method has unique advantages over other methods. This method is fully algorithmic. In addition this method is based on the design by parts that facilitates the design of power electronic circuits. And by some examples will be shown how this method is used for generating all sub-class (sub-family) of DC-DC converters with predetermined characteristics.**

Keywords — DC-DC converter, galvanic isolation, MTM, algorithmic method.

I. INTRODUCTION

The main part of any electronic device and/or electric vehicle is its power supply. In recent year DC-DC Converters are the power parts in the major of many electronic or electric devices (Computers, Embedded Systems, etc). But the development and minimization of these converters do not happen at the same speed as digital parts of these devices. This is due to the fact that the design (generation and development) of the DC-DC converter is based only on the skills and experience of the designer [1] [2], unlike in digital parts that have several mathematical and algorithmic methods for their implementations and simplifications [3]. There are new designed DC-DC converters which their implementation and topology are presented in [4][5][6][7][8] and [9]. The different loads require different features and characteristics of DC-DC converters. The common trends for all DC-DC converters are: high efficiency, increasing the switching frequencies for reducing the total size and weight, improved power density and dynamic performance, etc.

There are different series of designed DC-DC converters, but their design based only on the skills of the designers. There are no fully algorithmic methods that can be translated into computer program for designing and generating these power electronic converter circuits.

This work proposes algorithmic method that can be translated into computer program for generating and implementing sub-families of DC-DC converters with predetermined features.

II. MAIN TOPOLOGICAL MATRIX "MTM": DEFINITION & FEATURES

Any electrical circuit with a number of b elements can be represented by graph [10] [11][12][13], containing b- branches and n- nodes. By the graph theory branches are divided into two topological groups: b_T- tree branches and b_L-links. A tree of the graph is a subset of the branches such that all graph nodes are connected by branches but without forming a closed path (b_T= n-*1*). Then these branches are the tree branches. The remaining branches (collectively called a co-tree b_L) are the links (b_L= b-b_T=b-n+*1*). Given a network graph with b branches and n nodes select a tree, we can obtain the Matrix of Independent Meshes (MIM)- KVL equations, and Matrix of Independent Nodes (MIN)- KCL equations. There are b–(n-1) independent KVL equations and n-1 independent KCL equations.

A. Definition of MTM

For the topological relations between MIM and MIN, we will consider basic step-up/ step- down converter shown in Fig. 1. In Fig. 2 the graph for this circuit is shown, in which the bold lines show tree branches b_T and dashed lines show the links b_L. Matrix of Independent Meshes and Matrix Independent Nodes MIM and MIN are shown in Fig. 3 and Fig. 4 respectively. For the undirected graph, MIM and MIN are matching (Matrix transportations) and can be combined as shown in Fig. 5 (a).

The Fig. 5 (a) shows the Matrix of Independent Meshes (KVL equations) for the horizontal view (rows) and Matrix of Independent Nodes (KCL equations) for the vertical view (columns). Unity sub- matrixes in the left side and in the down of Fig. 5 (a) can be removed and we will obtain the Matrix that we named the Main Topological Matrix (MTM) as shown in Fig. 5 (b). This MTM will be the tool for generating converter circuits; MTM will be used for generating converter circuits (backward using MTM).

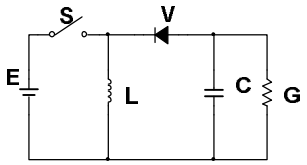


Fig. 1. The basic step-up/ step-down converter

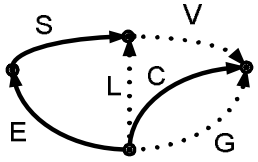


Fig. 2. The graph of the converter

L	V	G	S	C	E
1			1		1
	1		1	-1	1
		1		-1	

Fig. 3. The MIM matrix of the converter.

L	V	G	S	C	E
1	-1		1		
	1	1		1	
1	-1				1

Fig. 4. The MIN matrix of the converter.

L	V	G	S	C	E	
1			1		1	L
	1		1	1	1	V
		1		1		G
			1			S
				1		C
					1	E

a

	S	C	E	
L	1		1	1
V	1	1	1	2
G		1		3
	1	2	3	

b

Fig. 5. a - The MIM and MIN for step-up/ step- down converter
b - The structure of MTM for basic step-up/ step-down converter.

B. Features and descriptions of MTM

Now, we will formulate the general features and characteristics of MTM matrix that will be used for formulation of the logical equations for the syntheses of our converters:

- Equality (identity) of MTM 's columns satisfies the corresponding elements in the series (*E* in series with *S*- see Fig. 5);
- Equality (identity) of MTM's row satisfies the corresponding elements in parallel. Also the existence unity one in the row of MTM satisfies the corresponding elements of branch and link are linked in parallel (*G* in parallel with *C* - see Fig. 5);
- Any other meshes (dependent mesh) can be obtained from logical addition (logical operation - XOR) two or more MTM's rows.
- Any other nodes (dependent node) can be obtained from logical addition (logical operation XOR) two or more MTM's columns.
- Isomorphic circuits have identical MTM. This means we have also a tool for the identification of isomorphism (this problem is called: subgraph isomorphism problem - isomorphic circuits have identical MTM)
- The MTM Matrix will be designated by *m*, and its cells by *m[i,j]* or *mi,j,* and *m[i,j]*. These elements ∈ {0,1}: 1- element is present (exists) in the corresponding mesh or node, 0 - means the corresponding element is absent from the corresponding mesh or node (also we can designate it as $m_{i,j}$) .
- There are Column and Row matrixes for the MTM. For example in Fig. 5, column *C*=[0 1 1] means *m[1,2]=0, m[2,2]=1 and m[3,2]=1*. Column matrix *E= [1 1 0]* means *m[1,3]=m[2,3]=1* and *m[3,3]=0*. Row Matrix *L=[1 0 1]* it means: *m[1,1]=m[1,3]=1* and *m[1,2]=0*. Row Matrix *G=[0 1 0]* means *m[3,1]=m[3,3]=0* and *m[3,2]=1*. The sign ^ will designate the logical XOR operation.

III. FORMULATING THE COMMON PRINCIPLES OF DC-DC CONVERTERS: OPERATION AND THEIR DESIGN APPROACH BY USING MTM

The common operation principles for all DC-DC Converters (Single-Switch Switched Mode Power Supply SMPS) are the two timing intervals: some switches are "ON" in the first timing interval and others are "OFF". In the second timing interval will be vice versa. This means the switches that were "ON" in first interval will be "OFF" in the second timing interval and the switches that were "OFF" in first interval will be "ON" in the second timing interval (we can distinguish two equivalent circuits). *S* - designates the switch that will be "ON" in the first timing interval (*S*-type switch). The number of *S*-type switches is designated by *s*. *V* - will designate the switch that will be "ON" in the second timing interval (*V*-type switch) and their number will be designated by *v*.

For Single-Switch DC-DC Converters the number of Capacitors (*c*) should equal the number of inductive elements (*l*): *c*=l and this parameter will be designated by *q*.

For Single-Switch DC-DC Converter's graph, the number of the independent meshes [9] is $k= l+ v+ 1 –$ (*1* - corresponding the load resistance) in the first timing interval of operation, and the number of independent meshes is $k= l+ s+ 1$ in the second timing interval of operation. Since the graph contains a fixed number of independent meshes, so *v=s* (number of S-type switches should be equal to the number of *V*- type switches). And this parameter will be designated by *p*.

From the above the number of the Independent Meshes KVL equation for the graph circuits for Single-Switch DC-DC Converters: $k= q+ p+1$, and the number of the independent nodes (KCL equations) $u= b- k$, where *b* - is the number of circuit's elements.

We will follow the following approaches for all DC-DC Converters:

- Neglecting the losses in all the components of DC-DC converter - all switches, coils, transformers (multi-winding coils) are assumed ideal elements.
- *S* –type switch - the switches that transfer energy from DC source (switches that absorb energy from the energy source , these switches are "ON" in the first interval of the operation),
- *V* -type switch - the switches that delivered the energy which stored in magnetic elements to the load.
- *E* – Input Voltage Source.
- *L1, L2, Ll* –Inductive elements (transformers, single or multi-winding coils)
- *C1, C2, Cc* – Capacitors.
- The load resistance *R =1/G* - assumed in parallel with output capacitor(*s*) won't be shown in MTM.

- Element MTM *m[i.j]* of our converter will satisfy*: i≤ k, and j≤u.*

IV. GENERATING SUBCLASSES OF SINGLE-SWITCH DC-DC CONVERTERS BY USING MTM

A. Subclass DC-DC Converters with two switches

Generating subclass DC-DC converters with two switches *(s=v=1)* and two reactive elements (one Inductive element L and one output Capacitor *C: c=l=1*) will be as the following.

The structure of MTM for these sub-class converters are shown in Fig. 6, the inductive element *L* - assumes a single-winding coil. From the above discussion the number of the independent Meshes is *k= p+q+1=3.* As the links of the graph we will choose the elements *L, V,* and *G.* The remaining elements *S, C* and *E* will be the tree branches. Since the load *G* conductance is always in parallel with the output capacitor *C*, it can be unconsidered (removed) in the MTM in this phase (generating phase procedure).

	S	C	E	
L	x	x	x	1
V	x	x	x	2

Fig. 6. Structure of MTM for basic DC-DC Converters.

	S	C	E	
L	1	x	1	1
V	1	x	x	2
	1	2	3	

Fig. 7. Possible MTM's for basic converters (see text).

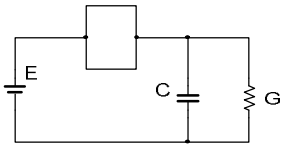


Fig. 8. This connection should be eliminated.

Now we will write the logical equations that determine the completion of MTM for generating these sub-class converters:

- *m [1, 1] =m [1, 3] = 1* - first timing interval (Energy is absorbed from E i.e. forming the loop *ELS..*, .., this equation can be rewritten: *m11* AND *m12 =1*).
- *m[1,1] ^ m[2,1]=0* - Second timing interval (forming the loop *LV*...,that not contents the switch *S* which will be "off" in this timing interval).
- *m[2,1]=1* - from the above Equation.
- *m12 ^ m22=1* - Condition to be the loop *LV*... contains the capacitor C (forming the loop *LVC*...). The sign ^ is the XOR logical operation.
- *C ^ E≠ 0* - Column Matrix *C=[m[1,2] m [2,2]]* and Column Matrix *E=[[1,3] [2,3]]* should be different if not so then *C* and *E* will be in series and the equivalent circuit will be as shown in Fig. 8. In this case the output voltage will be equal to the input, or the output voltage will be zero. This connection should be eliminated.

Taking into account the logical equations above and filling the MTMs we will obtain the three alternatives of MTMs corresponding to the three alternative circuits familiar as the basic DC-DC converters. These converters are shown in Fig. 9 (a), (b), (c) with their MTMs "The load *G*- is assumed in parallel with the output capacitor *C*"

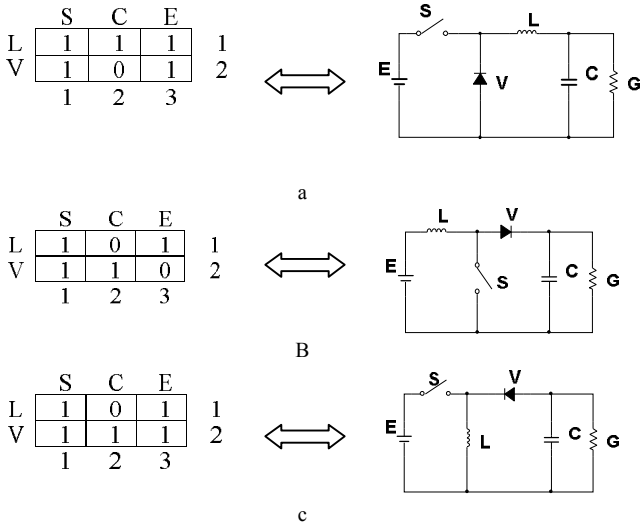


Fig. 9. The basic DC-DC converters with their MTM.
a - Step-down Converyer. b - Step-up Converyer.
c - Step-up/ Step-down Converyer (Flyback Convterter)

.

V. USING MTM FOR GENERATING DC-DC CONVERTERS WITH GALVANIC ISOLATION

In the majority of applications, it is required to incorporate a transformer (multi-winding inductor) into the switching converter, to obtain dc isolation between the converter input and output. However, since transformer (multi-winding inductor) size and weight vary inversely with frequency. This transformer operates at the converter switching frequency of tens or hundreds of kilohertz. These high frequencies lead to dramatic reductions in transformer size and weight (modern ferrite power transformers is much minimized). In addition some applications are required the converters with high output voltage.

A. Generating Subclass Converter with Four Switches and Four Reactive Elements

The MTM structure of these Subclass converters with high output voltage and Galvanic Isolation: converters with 4 Switches (*q*=2) and four reactive elements (*p*=2) is shown in Fig. 10. In Fig. 10 (a) is shown the MTM structure of the converters in which galvanic isolation (GI) is accomplished by one multi-winding inductor - *L2*. And the Fig. 10 (b) shows the MTM structure of the converters that galvanic isolation is accomplished by two multi-winding inductors - *L1* and *L2*. This GI (Galvanic Isolation) attribute is reflected (encoded) by zero Sub-matrixes in our MTM as shown in the Fig. 13.

For accomplishing the high output voltage the input current could be continuously, so the section (super node) of the input E with one or two inductors should be exist in this converters: this feature will be encoded in the logical equations for MTM. In addition the tow capacitors will be in the output part of converters to be accomplished voltage multiplication in the output: also this

feature will be encoded in the following logical equations. As in the previous examples the electrical Load not shown in this phase: in the design phase. And the load will be in parallel with output capacitor/s.

		In		GI		Out	
		E	L1	L2	C1	C2
In	S1	1	1	0	0	0
	V1	1	1	1	0	0
Out	S2	0	0	1	1	0
	V2	0	0	1	1/0	1

a

		In		GI		Out	
		E	L1	L2	C1	C2
In	S1	1	1	0	0	0
	V1	1	0	1	0	0
Out	S2	0	1	1	1	0
	V2	0	1	1	1/0	1

b

Fig. 10. a - Structure of MTM for Converters with 4 switches and 4 reactive element- GI is accomplished by one L
b - Structure of MTM for Converters with 4 switches and 4 reactive element- GI is accomplished by two L

B. *Sub-class converters that galvanic isolation GI is accomplished by one multi-winding inductor:*

The logical equations for the MTM in Fig. 10 (a) - converters in which the galvanic isolation GI is accomplished by one multi-winding inductor are:

- *m11=m12=1* - Condition for energy absorbing from E (forming the loop ESL.., this equation can be rewritten: m11 AND m12 =1).
- *m43=m45=1* - Condition for energy delivered to the output (forming the loop *V2L2C2...*).
- *m31=m41=m32=m14=m24=m15=m25=0* – the condition for GI.
- *m21 = m22=1* – To be accomplished that *E* and *L1* will be in series (condition for continuous input current and high voltage at the output).
- *m33=1* – To remove the series connection *C1* and *C2*.
- *m13 ^ m23=1* – To be accomplished different values for m13 and m23: to remove the parallel connection S1 and V1.

Taking into account the logical equations above the alternatives combinations of MTMs can be obtained as shown in Fig. 10 (a). And the sub-class of this DC-DC converter circuits is shown in the Fig.11 (a), (b) (two different converters a and b).

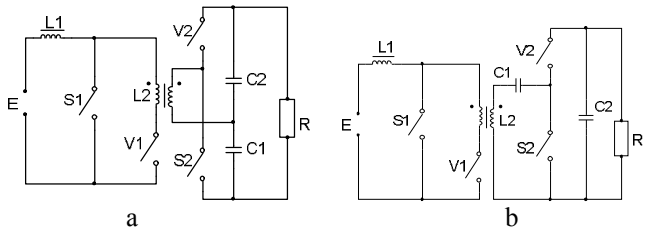


Fig. 11. Converters with one L for GI

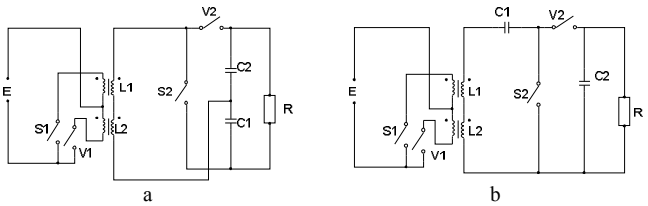


Fig. 12. Converters with two L for GI

C. *Sub-class converters that galvanic isolation GI is accomplished by two multi-winding inductor:*

The logical equation for the MTM in Fig. 10 (b): converters in which the galvanic isolation GI is accomplished by two multi-winding inductor *L1* and *L2*:

- *m11=m12=1* - Condition for energy absorbing from *E* (forming the loop ESL.., this equation can be rewritten: *m11* AND m12 =1).
- *m43=m45=1* - Condition for energy delivered to the output (forming the loop *V2L2C2...*) as in the above.
- *m31=m41=m32=m14=m24=m15=m25=0* –The condition for GI.
- *E^L1^L2= 0* - for forming the node with *E,L1*, and *L2* (condition for continuous input current and high voltage at the output).
- *m32=m33=m42=1* and *m13=0* – from the above equation.
- *m22 ^ m23 = 1*: To remove the parallel connection *S1* and *V1*.

With respect to the logical equations above the alternatives combinations of MTMs can be obtained as shown in Fig. 10 (b). And the sub-class of this DC-DC converter circuits is shown in the Fig. 12 (a) , (b) (two converter circuits a and b).

VI. Algorithm and Software Implementation

The main tasks of the software were to:

1. Receive the parameters *p* and *q* to determinate the dimension of MTM.
2. Generate all possible combinations of MTM's matrixes.
3. Print all possible topological MTM's combination matrixes that satisfy the logical equations for the determined sub-class of the circuits.

The Flowchart Algorithms for this task is shown in Fig. 13. It is beginning to receive the parameters *p* and *q,* generate all possible combinations of binary numbers to MTM (Count_v - count variable incremented from …0000 to ….1111). The mask variable (mask_v) each time is shifted left so we can complete each cells in the MTM by 0 or 1. Next each completed MTM has to be checked for satisfying the logical equations, if so the software will print the MTM (the topological matrix of the circuit). These procedures are repeated until all combinations will be finished.

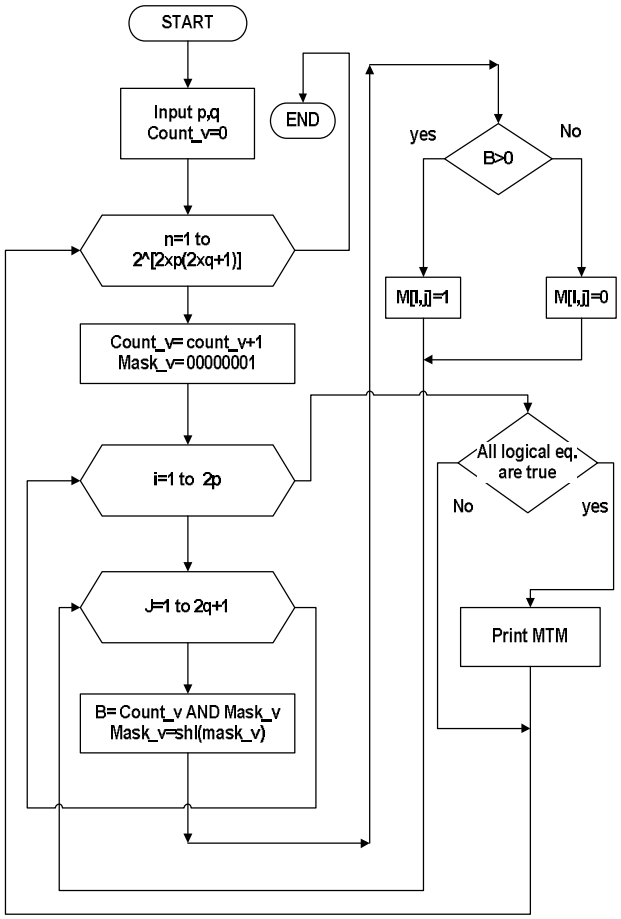


Fig. 13. Flow chart of the proposed algorithm for generating all members of sub-class circuits of DC-DC converters.

VII. CONCLUSION

This paper develops and proposes a fully algorithmic method for synthesizing and generating DC–DC converters by using MTM. By using the topological matrix MTM we generate all sub-class circuits of the DC-DC converters with predetermined characteristics. These predetermined characteristics for converters are encoded (implied) in the logical equations of the designed circuits. So we can call this method – the design by parts for synthesizing and generating converter circuits with predetermined characteristics.

Isomorphic circuits have identical MTM, so we have also a tool for identification of isomorphism of the circuits.

This method can be developed for synthesizing and designing not only DC-DC converters, but also another circuits like filters circuits... etc. And the proposed algorithm can be translated to computer program for synthesis different classes of power electronic circuits.

This method gives ability to cooperative design and the ability to build library of logical equations and/or classes for different power electronic circuits.

ACKNOWLEDGMENTS

Author wish to thank all his teachers and colleagues in NTUU "KPI".

REFERENCES

[1] Daniel M. Mitchell: DC-DC Switching Regulator Analysis. McGraw-Hill Book Company, USA (1988)

[2] . R. W. Erickson: Fundamentals of Power Electronics. Cluwer Academic Publishers (2004)

[3] David Calcutta, Fred Cowan, Hassan Parchizadeh: 8051 Microcontrollers - An Applications Based System Introduction. Newnes, Linacre House, Jordan Hill, Oxford OX2 8DP (2004)

[4] G. J. Su and F. Z. Peng: A Low Cost, Triple-Voltage Bus DC-DC Converter for Automotive Applications. The IEEE Applied Power Electronics Conference and Exposition (APEC), vol. 2, pp. 1015--1021, March 6–10, Austin, Texas (2005)

[5] Jaroslav Dudrik, Juraj Oetter : High-Frequency Soft-Switching DC-DC Converters for Voltage and Current DC Power Sources. Acta Polytechnic Hungarian, Vol. 4, No. 2.pp.29--46. Hungary (2007)

[6] N. Rajarajeswari, K. Thanushkodi: Design of an Intelligent Bi-Directional DC-DC Converter with Half Bridge Topology. Euro Journals Publishing, ISSN 1450-216X Vol.22 No.1 (2008), pp.90-97(2008)

[7] R. Kelley, M. Mazola, W. Draper and J. Cassidy: Inherently Safe DC-DC Converter Using a Normally-On SiC JFET. In Proc. of IEEE APEC,pp.1561--1565. (2005)

[8] P. James, A. Forsyth, G. Calderon-Lopez, V. Pickert: DC-DC converter for hybrid and all electric vehicles. EVS24 University of Manchester, UK, May (2009)

[9] Jong-pil lee, Byung-duk min, Tae-jin kim, Dong-wook yoo and Ji-yoon yoo: Input-Series-Output-Parallel Connected DC/DC Converter for a Photovoltaic PCS with High Efficiency under a Wide Load Range. Journal of Power Electronics, Vol. 10, No. 1, pp. 9-13,January (2010)

[10] Artemenko, M.E., Taher, M.A.: Synthesis transistor's Converters with determined characteristics. Technical Electrodynamics, Kiev, N4. pp. 43--47 (1994).

[11] Narsingh Deo: Graph Theory with Applications to Engineering and Computer Science. PHL learning Prt, ltd (2004)

[12] Reinhard Diestel: Graph theory, Third Edition. Springer –Berlin (2005)

[13] Murad Ahmed Ali Taher: "Algorithmic Method for Generating DC-DC Converter Circuits by Using Topological Matrix" , International conference DEIS 2011 London, UK, Communications in Computer and Information Science (CCIS), Vol. 194, pp. 714- 723.

Integration of Context-Aware Control System in Microgrid

Valerii Zhuikov
National Technical University of Ukraine
'Kyiv Polytechnic Institute', Kyiv, Ukraine

Anna Kyselova
National Technical University of Ukraine 'Kyiv Polytechnic Institute', Kyiv, Ukraine
a.g.kyselova@gmail.com

***Abstract* — In the last decades the energy demand is continuously increasing (both in industrialized and in emerging countries) and control system for renewable source and electrical loads are becoming more and more sophisticated. The last years we have seen an increase interest in the area of control system for microgrid. The paper focuses on the development context-aware system of microgrid convertors for control with minimization of user's costs and ensuring the most comfortable conditions in microgrid. It is carried out using new method which consists of reasoning module and context data processing module.**

Keywords— context, microgrid, control system, ontology.

I. INTRODUCTION

A microgrid comprises medium- and/or low-voltage distribution systems with distributed energy sources, storage devices and controllable loads. They can operate either if connected to the main power network or if isolated (islanded) in a controlled and coordinated way. Frequently we refer to a self-sufficient interconnection of distributed generation, residential and industrial load in a low-voltage network without a persistent connection to a larger grid. We develop the control system for microgrid is a complex information processing system with integrate level organization such as: renewable sours, electrical devices with embedded digital sensors, only sensors, users and the whole system. So, today, we know a lot about control system of power switches and microprocessors control system of power converters of individual levels but much less about complex relationships across levels of investigation [1-4]. Understanding function and building mathematical and logical models is a major challenge with implications for prevention a problem situation, for example, a critical power consumer.

II. CONTEXT-AWARE CONTROL SYSTEM

A. The architecture of microgrid

In order to fulfill the above listed requirements, a suitable automation system is needed. It should be intelligent enough to correctly take into account generation profiles that may change with the weather and the time (like wind or photovoltaic generation).

The result is a continuously changing distribution of power flow and direction, instead of the relatively stable, unidirectional power flow of a today distribution network. The develop context-aware control system of microgrid such as theoretical basis for energy-effective control of power supply system at microgrid on the principles decentralized strategies with hierarchic distribution of electrical devices and minimization of user's costs and ensuring the most comfortable user's conditions in micrigrid.

The comfort is a subjective measure experienced by the inhabitants, and hence quite difficult to derive analytically. In microgrid climate, specifically temperature, plays an important role in defining comfort. Moreover, the number of manual operations and the time spent by the users in performing daily activities also contribute to the inhabitants' comfort. The defined terms comfort as a joint function of temperature deviation, $\Delta(\theta)$, number (M) of manual device operations, number (K) of manual renewable energy source operation, and time (τ) spent for those activities by the users. Thus,

$$Comfort = f\left(\frac{1}{\Delta(\theta)}, \frac{1}{M}, \frac{1}{K}, \frac{1}{\tau}\right)$$

Proposed context models of generation, consumption, common balance and welfare economics take into account both electrical characteristics of devices at microgrid, and cost factors. It creates the basis for control with minimization of user's costs. The architecture of microgrid includes a loads, sensors, renewable and alternative energy sours (AB – accumulating battery, WG - wing generator, DG - diesel generator) - as it is presented in Fig. 1 two levels – physical and logical levels [5].

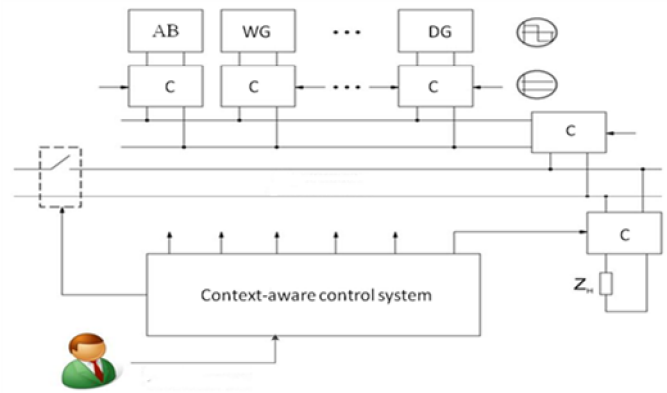
A context-aware control system of microgrid consists of a raw data from variously sensors (low level), Data processing module, ontology module, context processing module.

Summaries about the role of the constituents as follows:
Low level: It obtains context information from various physical sensor which provides a context-aware energy conversation system of microgrid with the information about physical environment.

The Raw data-processing module to evaluate the reliability of context information and predicted the context information that Context Processing Packages needs. That it consist of algorithms that smoothed data because the context information influenced by environment status may be imperfect. If a context-aware application responds to the user's requirements

using the incorrect context information, this will lead to undesirable result.

Physical level



Logical level

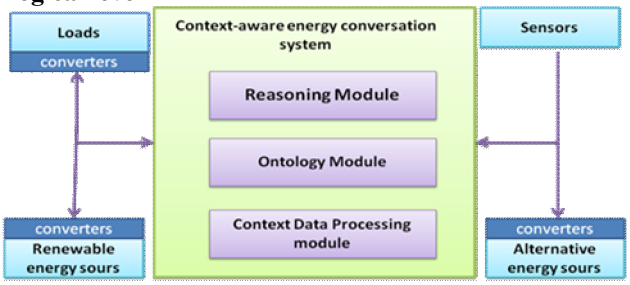


Fig. 1. The architecture of microgrid

III. CONTEXT DATA PROCESSING MODULE

In literature the term context-aware appeared in Schilit and Theimer (1994) the first time [6,7,8].

Context is defined in our context-aware control system as information used for identifying the status of an entity [9].

Attributes of context:

- Physical (Lighting, compass, temperature, acceleration, file ID);
- Computing entity (connectivity, network bandwidth, network latency);
- Equipment (loads, energy sours, sensors etc.);
- Time of synchronization (time of day, weekday, month, season of year);
- Action at previous action time (get, set, subscribe, unsubscribe).
- Architecture (zone or location)

Context prediction requires the consideration of the time dimension. A set of observations $Tt_1...Tt_n$ with Tt_l being recorded at a specific time interval t_i is called a time series [10].

A discrete time series in one in which the observatations Tt_l are taken at discrete interval time.

There are three distinct groups of smoothing methods - as it is presented in Fig. 3.

- Averaging Methods
- Exponential Smoothing Methods
- Kalman filter

The purpose of the context processing data module is to identify algorithms that are best suited to compromise to all the requirements in a microgrid environment that constitutes a context-aware control system community.

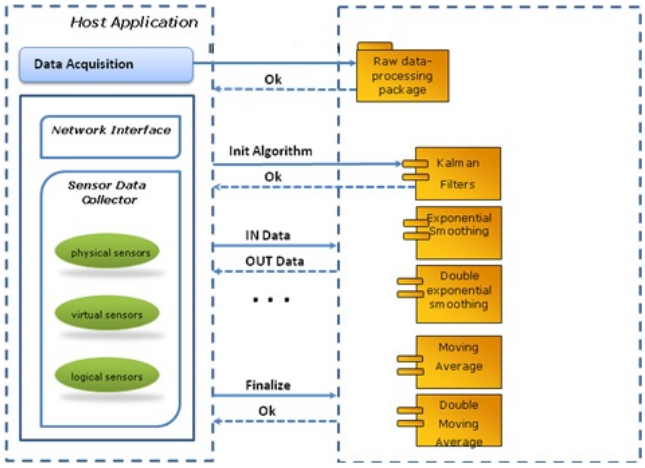


Fig. 2. The struck of raw data-processing module

The goal of testing is to estimate algorithm's quality, trace functionality issue, and estimate resources consumption and estimate of difference between the data after the algorithm with original data.

Correctness is the minimum requirement of software.
Test objective – ensure proper target of test functionality.
Technique – execute each use case flow to verify the following:

- Expected result occurs when valid flow is used
- Exception and warning message are displayed when invalid data is used .

To test the software filtering and smoothing time series was selected noise generator based on the source of random numbers with a normal (Gause) distribution with the specified mean and variance of the noise components. The unit of comparison test environment compared smoothed noisy time series with the original test time series. Characterize the quality of the results comparing smoothing algorithms.
The results of testing programs (Fig.4.) for all the methods of filtering and smoothing time series are shown in Table 2 for examples of tests 1-6. The program was tested on a computer with a processor with a clock speed of 2.26GHz.

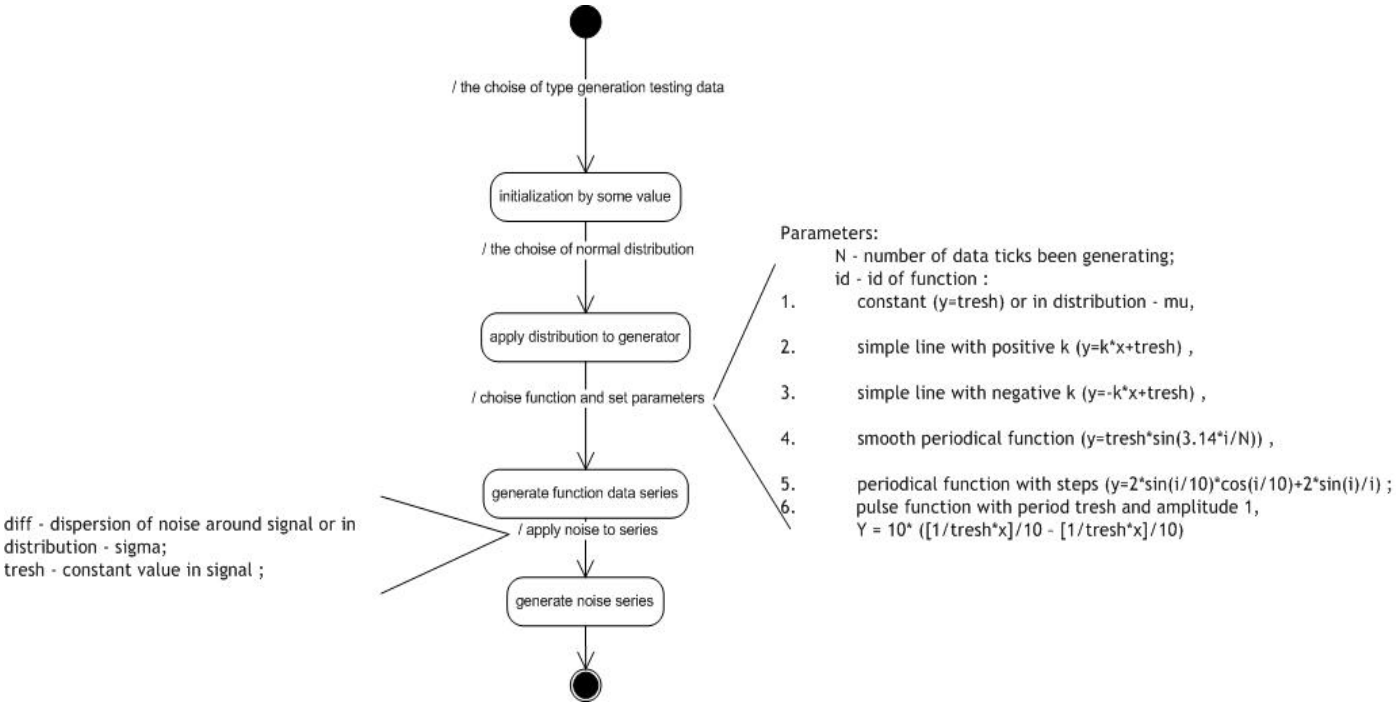


Fig. 3. Test approaches

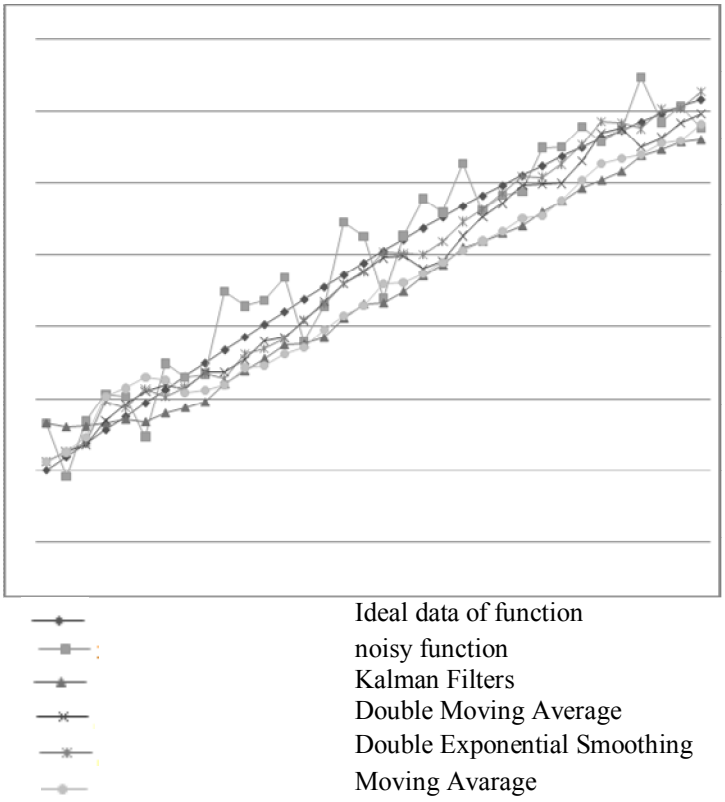


Fig. 4. Smoothing the initial time series with application of the KF, DMW, DES, MA.

TABLE I. TEST RESULT OF CONTEXT DATA PROCESSING MODULE

Test	Flow of the Test Case	Test Item	Result
Double Moving Average Model (WMA)	a) generate noised signal (1000 ticks) b) process 1000 ticks (smooth) c) compare with original data	Difference range between smoothed and original data Difference range between smoothed and original data Difference range between smoothed and original data	value entered in 4 %range
Moving Average	a) generate noised signal (1000 ticks) b) process 1000 ticks (smooth) c) compare with original data		value entered in 3% range
Kalman Filter	a) generate noised signal (1000 ticks) b) filter each next value based on previous		value entered in 3% range
Double Exponential Smoothing	a) generate noised signal (300 ticks) b) process 12 noised ticks c) smooth 12 - noised tick		value entered in 2-3% range

IV. ONTOLOGY

An ontology module: It preserves and shares context knowledge on behalf of the energy consumption, restricted by resources, and a computing entity. We propose a context ontology model for efficient knowledge sharing, knowledge reuse, context querying and reasoning in microgrid.

Microgrid ontology (describing the information space microgrid) includes a description of the following classes: Activity, the user, the environment, the room (zone), network components (transformers, load, alternative sources), a service platform.These are based around the most important aspects in context information.

- User: The user plays a main role in microgrid environments.
- Environment: The environment of microgrid in which the user interacts is an important aspect of the context specification.
- 3) Service: Services provide specific functionality to the user.
- 4) Location: This concept provides the location of the microgrid.
- 5) network components: CompEntity stands for computational entities. It includes all type of loads, renewable sours, all tupe of sensors etc.
- 6) Activity: A environment may have different kinds of activities like deduced activities [11, 12].

In our prototype implementation, we try to define rules for the microgrid space. We define rules using SWRL (Semantic Web Rule Language). All rules are expressed in terms of OWL concepts (classes, properties, individuals, literals etc.).

$$Rule\ A_n\text{: } C_1(?x) \wedge C_2(?y) \wedge P_1(?x, ?y) \wedge C_3(?x, ?z) \rightarrow C_2(?z, ?y)$$

where C –classes of ontology, P, x, y – instances or variablese, z – variables or values.

V. REASONING MODULE

It reasons out the context, interpreting context information. It coordinates and provides directions based on context information through Context Data Processing Module.

Existing algorithms for the process of ontology-based reasoning does not allow for analysis and comparison of the properties of instances of types. Developed a method for finding and implementing its algorithm will take into account both the standard procedures for the withdrawal of the ontology and the search for analogies, which allows you to compare your current Problem Situation [13] with precedents in the ontology precedents Problem Situation and choose the best of them.

When initializing the search procedure is queried q, representing a set of instances or classes corresponding to the given parameters (ai). First is the detection of the Problem Situation using the developed classification rules based on the standard procedure of referring the request to one of the predefined classes. Next, select all items (cq I) already defined the Problem Situation. Then - a comparison with the precedent request instances that are stored in the database with the relations defined in the ontology.

To compare properties in accordance with an algorithm to determine the measure of similarity for these commonly used as the means used comparison (character by character string comparisons, comparing numbers).

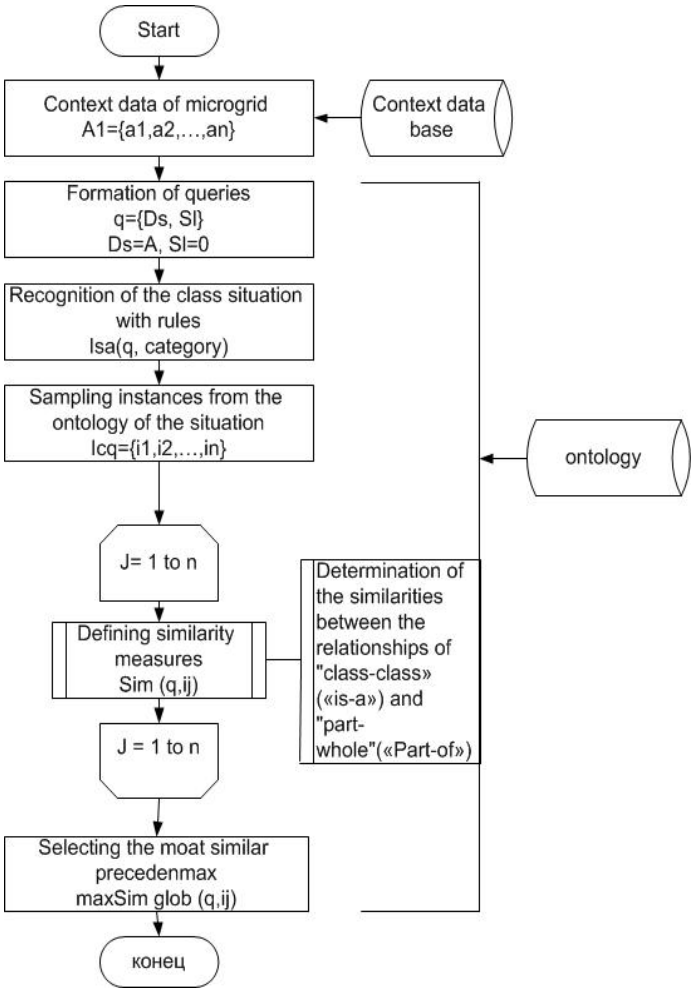


Fig. 5. Algorithm for desiciom making im microgrid

TABLE II. CALCULATING THE SIMILARITY MEASURES

Components of similarity measures	The formulas for calculating the similarity measures

Semantic similarity measure of attribute values Problem Situation and precedents	$sim_{is-a}(c_i, c_j) = \frac{\left\|c_s(c_i) \cap c_s(c_j)\right\|}{\left\|c_s(c_i) \cup c_s(c_j)\right\|}$
Attribute similarity measure attribute values Problem Situation and precedents	$sim_{po}(q, i_j) = \frac{\sum_{j=1}^{l}\sum_{k=1}^{m} fsim_t(a_j^q, a_k^{i_j}) * w_j}{l+m}, t \in T, a \in A$

The main differences between context-aware system in microgrid and existing control systems:

1) Saturation of the network (as in its nodes and geographically distributed) active elements, could be change the topological parameters.
2) Quite a number of sensors that measure the current value of the current state of the environment assessment microgrid in normal and problem situations.
3) The collection of contextual data, as well as the transmission and processing of information (mathematical, information and software for context-dependent control of power semiconductor converters local object based on object-cognitive analysis).
4) The ability to automatically (programmatically) assess current and predict the future situation in the energy system and its individual parts.
5) The high performance of the control system (when choosing a method of processing the context information is taken into account while processing the time series (ms) is the used memory (B)), the organization of the cyclic control system.

CONCLUTION

The necessity of ontology for the establishment of context-aware control systems is broadly acknowledged. In this paper, we have presented a context ontology model for microgrid. The major concepts can be reused for other microgrid too.

In this paper we have presented a general framework for calculating similarity among ontology. In order to define a general framework, we have based our work on an abstract ontology model that allowed adhering to various existing and evolving ontology standards.

The main characteristic of the framework is its layered structure: We have defined three levels on which the similarity between two entities could be measured: data layer, ontology layer, and context layer, which cope with the data representation, ontology meaning, and the usage of these entities, respectively.

REFERENCES

[1] J. Petergerya, A.Kyselova "Constructing models of local objects distributing computing networks", "Electrinics and Communications", №1-2, 2008, pp. 213

[2] J. Petergerya, A.Kyselova. "object-oriented modeling for system to control of electric energy of local object", "Electrinics and Communications", №4-5, 2009, pp. 269

[3] A.Kyselova "Local object loadings management based on the predicting triangle", "Electrinics and Communications", №1-2, 2010, pp. 83.

[4] Kyselova A., Shalaginov Intelligent model for Dealing with Contexts Application // CAD in Machinery Design. Implementation and Educational Problems : 18-th Ukranian–Polish Conference «CADMD'2010», October 14-16, 2010, Lviv, Ukraine: Proc. – Lviv: Vezha& Co , 2010. – P. 20–22.

[5] Arbanowski, S., Van der Meer,S., Steglich S. and Popescu-Zeletin R. The Human Communication Space: Towards Icentric Communications (2001)

[6] Abowd, G. D., Atkenson, C. G., Hong, J., Long, S., Kooper,R. and Pinkerton, M. Cyberguide: A mobile context-aware tour guide (1997)

[7] Bardram, J. E. and Bærbak, H. Christensen. Supporting Pervasive Collaboration in Healthcare – An Activity-Driven Computing Infrastructure (2004)

[8] Beale, R., Lonsdale, P. Mobile Context-aware Systems: The Intelligence to Support Tasks and Effectively Utillise Resources. MobileHCI (2004)

[9] Bornträger, C., Cheverst, K., Davies, N., Dix, A., Friday, A. and Seitz, J. Experiments with Multi-modal Interfaces in a Context-Aware City Guide (2003)

[10] Firat A. Information Integration Using Contextual Knowledge and Ontology Merging [Electronic resource]: PhD Thesis: 27.08.03: defended Sep. 2003. MIT, 2003. 151 p. URL: http://www.mit.edu/~bgrosof/paps/phd-thesis-aykut-firat.pdf (access date: 29.05.2009).

[11] Gruber, T.R.: Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In Guarino, N., Poli, R., eds.: Formal Ontology in Conceptual Analysis and Knowledge Representation, Deventer, The Netherlands, Kluwer Academic Publishers (1993)

[12] Stumme, G., Ehrig, M., Handschuh, S., Hotho, A., Maedche, A., Motik, B., Oberle, D., Schmitz, C., Staab, S., Stojanovic, L., Stojanovic, N., Studer, R., Sure, Y., Volz, R., Zacharias, V.: The Karlsruhe view on ontologies. Technical report, University of Karlsruhe, Institute AIFB (2003)

[13] Handschuh, S., Staab, S., Maedche, A.: CREAM - creating relational metadata with a componentbased. In: Proceedings of the First International Conference on Knowledge Capture K-CAP 2001. (2001)

[14] Haase, P., Siebes, R., Harmelen, F.: Peer selection in peer-to-peer networks with semantic topologies. In: International Conference on Semantics of a Networked World: Semantics for Grid Databases, 2004, Paris. (2004)

[15] Kyselova A., Shalaginov A. Hidden Markov Model for Dealing with Contexts Application// CAD in Machinery Design. Implementation and Educational Problems: 18-th Ukranian–Polish Conference «CADMD'2010», October 14-16, 2010, Lviv, Ukraine: Proc. – Lviv: Vezha& Co , 2010. – P. 20–22.

[16] Kyselova A.G., Kyselov G.D., Sergeev A.A., Shaloginov "Data processing for multimodal application" "Electrinics and Communications", №2, 2011, pp. 86-92.

Modeling of Power-supply Subsystems of Microgrid Using Petri Nets

Tetiana Khyzhniak, Viktor Kolesnyk
Faculty of Electronics
National Technical University of Ukraine "KPI"
Kyiv, Ukraine
tatjana.khizhnjak@gmail.com

***Abstract* — The modeling of power-supply systems such as photovoltaic system and wind-power system by using Petri nets was considered. The models were created for different inputs conditions such as states of photovoltaic elements and of accumulator battery, wind power, power of loads and for different combinations of these conditions. These models are the base for creating of flexible algorithms for control of electric energy flows in microgrid subject to some optimization criterions.**

Keywords—power system, modeling, Petri nets; photovoltaic system; wind-power system, model, microgrid

I. INTRODUCTION

In recent years theoretical and practical aspects of using different types of energy sources together in microgrid have increasingly developed. It causes more attention to requirements for efficient conversion and consumption of electric energy in microgrid. These requirements can be provided and considered with diagnosis, forecasting of states and optimisation of microgrid control using enough simply and visual modeling approaches. The modeling using Petri nets is an appropriate approach [1] for the fulfillment of these.

II. MICROGRID AND ITS CONTROL SYSTEM

A. Main structural blocks and its characteristics

Microgrid is a fully integrated, self-regulatory and self-repairing electricity system that has the network topology and includes all generating sources, main and distribution network, and all types of electricity consumers, managed a single network of information and control devices and systems in real time. So, the main parts of this system are: array of generators, energy conversion systems, which consist of different semiconductors converters, storage battery for accumulation of energy, load, which can have different nature, and control system (Fig.1). Since, different types of renewable energy used in Microgrid together with common electrical network (external network), so local grid is formed to supply the load from renewable sources.

The structure of energy conversion system depends on type of energy source, of load, which connecting with this source by using semiconductor converters. The modeling of power sources together with semiconductor converters and load is the step for building of optimal structure of conversion system.

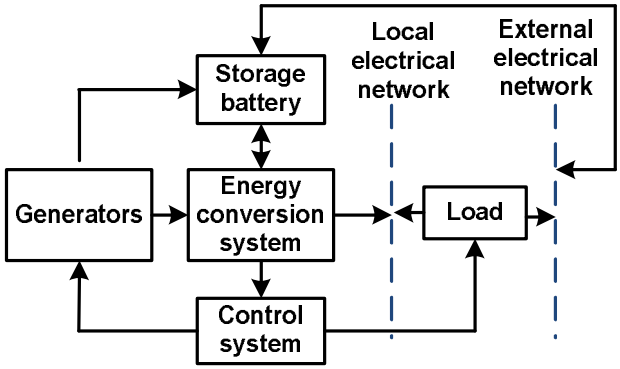


Fig.1. Generalized structure of microgrid.

The load can be divided into three types: uninterruptible load, interruptible load and load with back-up generators. So, type of load must be taken into account when modeling the energy conversion system and the control system.

The control system provides analysis and monitoring of all components and generates control signals according to given criteria [2, 3]. The structure of microgrid control system divided into three levels (Fig.2): the first level represented by the central control unit, the second level is the interaction between subsystems of microgrid, the third level is the set of processes within subsystems.

The central control unit generates control signals in a accordance with a special algorithm and information signals from the third and second levels.

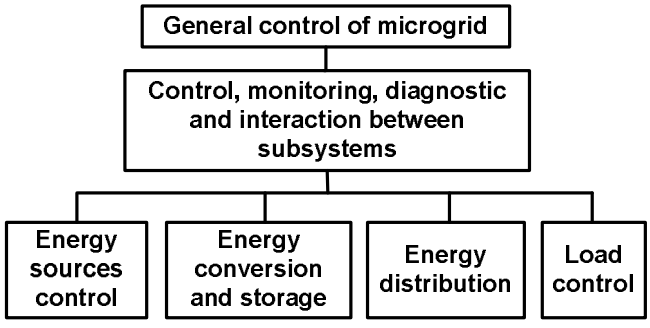


Fig.2. Structure of microgrid control system.

The second level is used for accumulation, conversion and analysis of date from subsystems. Results of analysis are used for diagnostic and monitoring of state. The important step of these procedures is modeling of sources, convertors and loads. Models are used for prediction of state of different subsystems and forming appropriate control algorithms. Information about current states and about main parameters which used in models is formed on third level.

The deterministic local tasks solves at the third control level such as the direct regulation of energy generation and conversion according to control signals from first and second level, management and distribution of energy flow, the protection of electrical equipment against fault operating modes and the process of exchange information with a higher control level. All sensors, executive devices and blocks for previous conversion of date are the elements of this part of control system.

B. Researching tasks

Microgrid is self-regulating system. So, coordination between necessities of load and capabilities of sources is the main task. Because, different energy sources are presented must be defined the set of operations for realisation of a microgrid energy management, i.e. must be studied a sequence of events and processes in subsystems. This task can be provided using modeling.

The model of power supply subsystems allows to define the set of discrete states for different input conditions. Combination of all models in one general model helps to implement an optimal energy distribution from the power-supply subsystems with different types of energy sources. Petri nets is the modeling approach for solving previous task that allows to determine cause-effect relations in working modes of elements and system in whole.

All possible operating modes of the power-supply subsystems, the behavior and structure of load, conditions of using an external electrical network and directions of energy distribution from renewable sources must be studied in process of modeling.

A synthesis of subsystem models should take into account initial conditions (initial marking) for each model, possibility of parallel processes and combining model elements that duplicate events in a separated operation.

III. Modeling of Power Supply Subsystems by Petri Nets

The set of energy sources in microgrid represented with renewable energy sources (connection with load through the local electrical network) and external electrical network. This paper introduces the Petri nets models of power supply subsystems that are based on photovoltaic and wind-power systems [4, 5].

A. Petri nets model of photovoltaic system

The structure of the photovoltaic system, see Fig.3, consists of photovoltaic panels; DC-DC converter, which implements the maximum power point tracking (MPPT); storage battery (SB) for reserving redundant power from photovoltaic panels; bidirectional DC-DC converter (BDC), which used for charging and discharging SB; inverter and load.

The initial conditions in the model of photovoltaic system are voltage of the photovoltaic panels (V_{pv}), voltage of SB (V_{bat}) and charging or discharging current of SB (I_{bat}). An operating mode of the photovoltaic system depends on a combination of these parameters. So, there are four operating modes for introduced on Fig.3 power-supply subsystem.

The next parameters used in description of operating modes and models of photovoltaic system:

a) voltage V_{bat_min} is the voltage of full discharge of SB;

b) voltage V_{bat_max} is maximum allowed voltage of SB;

c) current I_{bat_max} is the muximum allowed charge/discharge current of SB;

d) voltage V_{pv_min} correspond to the least power of photovoltaic panels.

The *first operating mode* described by next parameters: battery voltage have middle value $V_{bat_min}<V_{bat}< V_{bat_max}$, so energy of it can be used for supply of loads with current $I_{bat}<I_{bat_max}$ together with photovoltaic panels, which energy which is more than some minimal level, so it voltage is $V_{pv}>V_{pv_min}$. The DC-DC converter works in MPPT mode and BDC in boost mode for these conditions.

The *second operation mode* starts when V_{bat} reaches over-charged point voltage or I_{bat} reaches the maximum charge current, so $V_{bat} \geq V_{bat_max}$ and $I_{bat} \geq I_{bat_max}$. When in this case $V_{pv}>V_{pv_min}$, so BDC must change from boost to buck mode, which control V_{bat}. Meanwhile the operation mode of DC-DC converter must also change from MPPT to constant-voltage mode.

There are adverse weather conditions and in this case voltage $V_{pv} \leq V_{pv_min}$. The output power of photovoltaic panels is almost zero. The DC-DC converter stops working and only BDC works in boost mode, when $V_{bat_min}< V_{bat} < V_{bat_max}$ or $V_{bat} \geq V_{bat_max}$ with $I_{bat}<0$ (SB is discharged in load). There is the *third operation mode*.

During third mode the SB may come into over-discharge state for continuously supplying power to load. There is the emergency state. This state should not be allowed because it leads to a rapid deterioration of battery characteristics. So, control system must track it and *fourth operating mode* is considered in model with next parameter:

$$V_{bat} \leq V_{bat_min},\ V_{pv} \leq V_{pv_min},\ I_{bat} \geq I_{bat_max}.$$

BDC immediately stop working in forth mode to protect the SB and the whole system stops.

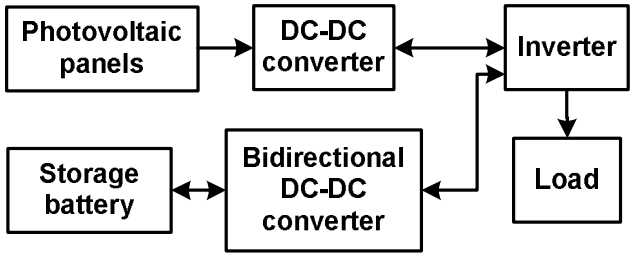


Fig.3. Structure of photovoltaic system.

All considered conditions and operating mode of photovoltaic system are implements in Petri net model (Fig.4). The description of places and transitions of model are shown in TABLE I and TABLE II respectively.

The places P_j in model describes input conditions of power sources (P_1-P_8), modes of converters and battery (P_9-P_{13}, P_{21}) and condition for reach of MPPT mode (P_{14}-P_{20}).

The increment or decrement of power ΔP and voltage ΔV_{pv} of photovoltaic panels is selected according to algorithm of the maximum power point tracking, which can be implements in inverter.

Input dates for real power supply subsystem are received from sensors, which measure voltage and current of photovoltaic panels, storage battery, converters and loads, which connection with it. So, we consider the third level of control system (see Fig.2) in this case.

Simulation results are used in the second level of control system as a basis for the development of algorithms for monitoring, diagnosis and management of all subsystems and loads.

B. Petri nets model of wind-power system

The generalized structure of wind-power system with external electrical network is shown on Fig.5.

Initial conditions for modeling are power from wind generator (P_w), SB state and power which demands for load (P_L). Five operation modes form by a combination of these initial conditions.

The *first operating mode* is characterized of next parameters: $P_w \geq P_L$ and SB is discharged. So, the wind generator supplies load and the redundant power from the wind generator transfers to SB.

There is the *second operating mode* if $P_w<P_L$ in case when SB is discharged. The power from the wind generator is insufficient to supply the load, and because SB is discharged, then the external electrical network compensates a lack of power.

The *third operating mode* occurs when $P_w<P_L$, but SB is charged. SB compensates the lack of power to the load and system work without external electrical network.

TABLE I. THE PLACES IN MODEL OF PHOTOVOLTAIC SYSTEM

Index	Place Pi
1	$V_{pv} > V_{pv_min}$
2	$V_{pv} \leq V_{pv_min}$
3	$V_{bat_min} < V_{bat} < V_{bat_max}$
4	$V_{bat} \geq V_{bat_max}$
5	$I_{bat} < I_{bat_max}$
6	$I_{bat} \geq I_{bat_max}$
7	$I_{bat} < 0$
8	$V_{bat} \leq V_{bat_max}$
9	Condition for MPPT mode
10	Condition for DC-DC converter constant-voltage mode
11	Condition for BDC boost mode
12	Condition for BDC buck mode
13	Condition for SB over-discharge state
14	$\Delta P>0$
15	$\Delta P=0$
16	$\Delta P<0$
17	$\Delta V>0$, when $\Delta P>0$
18	$\Delta V\leq 0$, when $\Delta P>0$
19	$\Delta V>0$, when $\Delta P<0$
20	$\Delta V\leq 0$, when $\Delta P<0$
21	Condition for next step of MPPT

TABLE II. THE TRANSITIONS IN MODEL OF PHOTOVOLTAIC SYSTEM

Index	Transition Tj
1	First mode
2	Second mode
3	Third mode
4	Fourth mode
5	Mode of MPPT
6	DC-DC converter constant-voltage mode
7	BDC boost mode
8	BDC buck mode
9	SB over-discharge state
10	Measure V_{pv}, when $\Delta P>0$
11	Obtaining maximum power point
12	Measure V_{pv}, when $\Delta P<0$
13	Voltage reference increment
14	Voltage reference decrement

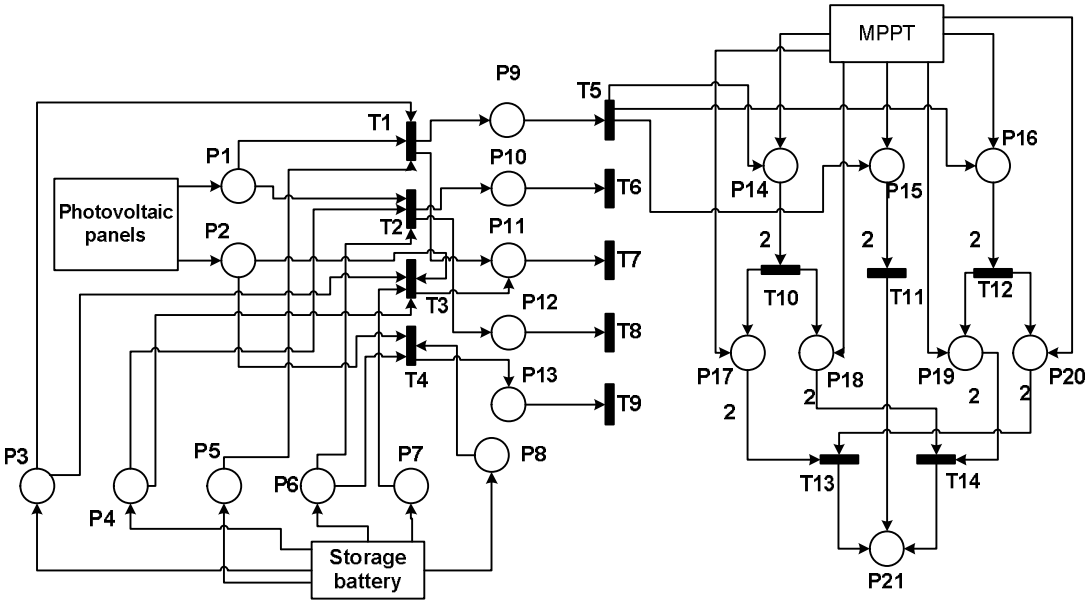


Fig.4. Petri nets model of photovoltaic system.

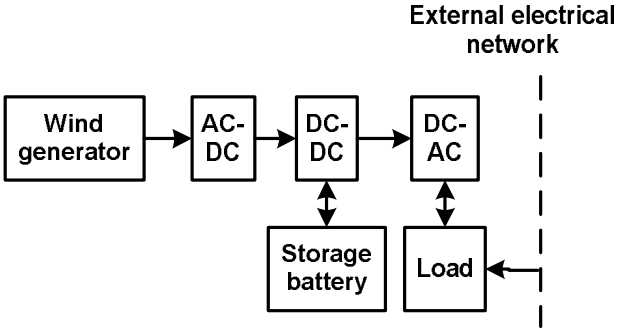


Fig.5. Structure of wind-power system.

The *fourth operating mode* is similar to third mode, but $P_w = 0$ and SB is charged. In the absence of wind-generated power, SB directly supplies the load.

External electrical network is the once source for the *fifth operating mode*, because $P_w = 0$ and SB is discharged in this case. The external electric grid supplies the load and used for charging SB if necessary.

The fourth and fifth modes changed each other in the context of the charge/discharge process of SB.

The description of places and transitions of Petri nets model (Fig.6) are shown in TABLE III and TABLE IV respectively.

The power-supply subsystems work based on concepts such as the maximum energy extraction and ensuring continuous load power supply. The sequence of actions in power-supply subsystems is formed in accordance with the external conditions, including weather conditions for renewable energy sources.

Petri nets models allow take into account necessary input conditions and operating modes of power converters for operation algorithms of power-supply subsystems.

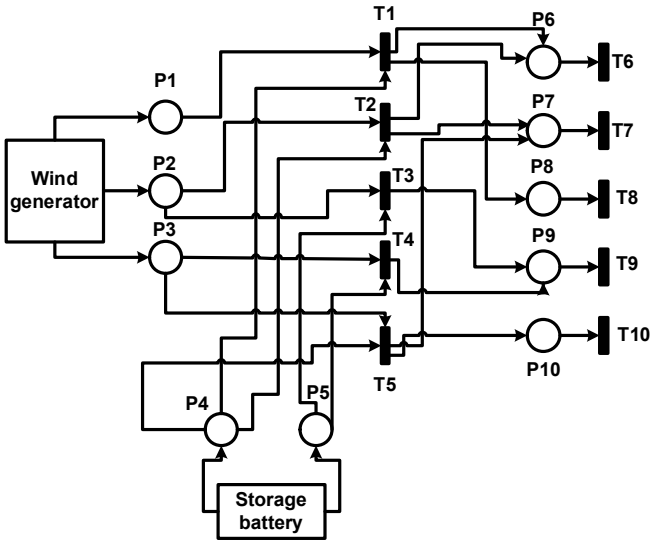


Fig.6. Petri nets model of wind-power system

CONCLUSIONS

Petri nets models of power supply subsystems allow form the set of different operating modes of converters, show the procedure of maximum power point tracking for photovoltaic system and condition of work the wind-power system with the external electrical network. Such modeling is the necessary stage for future study and optimization of microgrid. Developed models will be used to build the generalized model of microgrid. It was allow to investigate compatible operations of all power-supply subsystems and optimizing an energy flow distribution.

TABLE III. THE PLACES IN MODEL OF WIND-POWER SYSTEM

Index	Place Pj
1	$P_w \geq P_L$
2	$P_w < P_L$
3	$P_w = 0$
4	SB is discharged
5	SB is charged
6	Condition for power transfer from wind generator to load
7	Condition for power transfer from external electric grid to load
8	Condition for SB charging from wind generator
9	Condition for power transfer from SB to load
10	Condition for SB charging from external electric grid

TABLE IV. THE TRANSITIONS IN MODEL OF WIND-POWER SYSTEM

Index	Transition Tj
1	First mode
2	Second mode
3	Third mode
4	Fourth mode
5	Fifth mode
6	Transmission power to load from wind generator
7	Transmission power to load from external electric grid
8	SB charging from wind generator
9	Transmission power to load from SB
10	SB charging from external electric grid

REFERENCES

[1] Pyterson J. Teoryya setey Petry y modelyrovanye system. M.: 1984 - 264.s.

[2] Kyrylenko O.V., Peterherya Yu.S., Tereshchenko T.O., Zhuykov V.Ya. Intelektual'ni systemy keruvannya potokamy elektroenerhiyi u lokal'nykh ob"yektakh. K.:2005 – 212s.

[3] Yakymenko Yu.I., Sokol Ye.I., Zhuykov V.Ya., Peterherya Yu.S. Vidnovlyuval'ni dzherela enerhiyi u lokal'nykh obyektakh. K.:2001 – 114s.

[4] D. Lu, H. Fakham, T. Zhou, B. Francois. Application of Petri nets for the energy management of a photovoltaic based power station including storage units // Renewable energy. – 2010. – №35. – p. 1117-1124.

[5] Muhammad Shahid Khan. Supervisory hybrid control of a wind energy conversion and battery storage system: A dissertation submitted in conformity with the requriments for the degree of Doctor of Philosophy Graduate Department of Electrical and Computer Engineering University of Toronto. – Toronto, 2008. – 230.

Simulation of Electromagnetic Environment of Class D Amplifier

Y. Onikienko, V. Pilinsky, D. Probity, V. Shvaychenko
Audio Engineering and Registration of Information Department, Faculty of Electronics
NTUU "KPI"
Kyiv, Ukraine
vbs2011@ukr.net

***Abstract* — In this article the equivalent circuit of class D amplifier with elements that create and distribute conducted electromagnetic interferences is suggested. SPICE simulation of class D amplifier inclusive parasitic parameters with connection to LISN have been realized. The results of computer simulation with experimental data of the interferences measuring are compared. The theoretical and experimental analysis of the influence of input signal amplitude on the noise level of class D amplifiers are confirmed. Simulation results are generally coincide with the measurement ones.**

Keywords — class D amplifiers, electromagnetic compatibility, EMI modeling, radiofrequency interference, simulation

I. INTRODUCTION

Modern class D power amplifiers have a high efficiency and audio performance and are highly competitive with a class A, B, AB linear amplifiers [1]. Switch mode operation of class D amplifiers supplies highest efficiency (up to 100 percent theoretically), but has an essential disadvantage namely considerable level of the electromagnetic interferences (EMI). EMI are spreading in the conducting lines as electromagnetic emission.

Therefore it is necessary to reach at the design stage an electromagnetic compatibility (EMC) of class D amplifiers with other parts of audio equipment, mains and various devices. Assurance of the compliance of the class D amplifiers to the requirements of EMC standards [2,3] can be attained by complex application of class D amplifier special methods of the electromagnetic noises suppression [4,5,6] along with the well-known methods of EMI reducing [7], applicable for all switch mode devices. The review of sources of EMC and methods of their suppression is given in [8]. An origin and a distribution of EMC in the modern power supplies and class D amplifiers have a similar nature due to the switch mode of operation, although they have the additional factors related to modulation of the audio signal influence on the EMI level of class D amplifiers. Such factors can be an amplitude of audio signal and a total harmonic distortion (THD).

This paper aims to consider the equivalent circuit of the class D amplifier with the parasitic parameters of assembly, wires, capacitors, transistors etc. In this paper the simulation circuit of the class D amplifier with parasitic parameters influence on the EMI level is suggested. The results of simulation of the class D amplifier connected to Line Impedance Stabilization Network (LISN) have been analysed and compared with the experimental data of EMI measurement, obtained previously.

II. ESTIMATION OF EMI LEVEL DEPENDING ON THE INPUT SIGNAL LEVEL

It is known that spectrum of the sequence of rectangular pulses contains harmonics that are multiples of the fundamental frequency. The spectrum of pulse-width modulated signal contains harmonic with frequency of useful signal, harmonics of switching frequency and multiple to it, and also a combination of harmonics mentioned above. Amplitudes of harmonics since the second has a large difference between switching and signal frequencies and considerably less than amplitude of the first harmonic. One can assume that EMI level of the class D amplifier with an input signal will be less, than without it, because of more "spread" spectrum of electromagnetic noises.

The estimation of the input signal level, THD and other factors influencing on the EMI level of the class D amplifier was performed during experimental researches. The results of researches are presented in [9]. Class D automotive power amplifier based on TPA3100D2 chip from Texas Instruments was tested in accordance to the standard methodology of conducted EMI measuring in the frequency range from 0,15 MHz to 30MHz without additional noise suppression in two versions (with an input signal and without it). Test sample was powered from unipolar nonstabilized linear power supply. The curves of EMI levels and permissible limits are taken from [9] and presented in Fig. 1.

It is evident from Fig. 1 that almost in the whole range of measuring frequencies the level of conducted noises without input signal is higher than in case of output power 10 W. At frequencies of 600 kHz, 6 MHz and 30 MHz noises level exceeds permissible limits. In a band of frequencies higher than 10 MHz a noise level also increases at 10-15 dB because of influence of the PCB mounting parasitic parameters. But on frequency of about 20 MHz a peak noise level is higher with input signal than without it.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

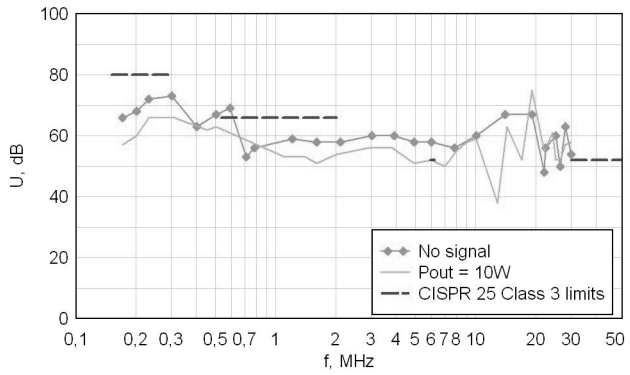


Fig. 1. EMI levels of TPA3100D2 and CISPR 25 limits.

It should be noted that a standard regulates measuring of EMI levels of professional audio apparatus during the noise-emission tests at power equal to 1/8 from nominal power [2]. Therefore it is necessary to take into account that during the professional class D amplifiers testing, that the generated noise level without input signal on some frequency bands can exceed a noises measured with standard methodology with input signal.

The computer model of the class D amplifier connected to LISN was developed in order to compare the obtained experimental results to the simulation data.

III. Modeling of Class D Amplifier Considering Parasitic Parameters

The computer simulation of class D amplifier with parasitic parameters can be used for predicting of the generated EMI levels at the development stage. EMI simulation of class D amplifiers appears a useful instrument at the development stage and will considerably simplify and reduce in the cost a process of EMI providing at a stage of production and finishing.

In papers [10, 11] the methods of simulation of conducted EMI in class D amplifiers are considered. The research of integrated amplifier in TQFP case were performed in the paper [10]. The EMI of two amplifiers was compared: one amplifier with a carrier-based PWM, other with a self-oscillating modulation based on sliding-mode control. The results of simulation are close to the measured, but the simulation of EMI levels was performed only for the amplifier output signal. In the article [11] the method of passive components presentation (output filters and loudspeaker) is used as impedance matrix with their subsequent association in single one and simulation of output currents of class D amplifier. This method has high accuracy too, but requires measuring of passive elements impedance by means of high-priced equipment. As in paper [10] the simulation of EMI levels was performed only for the amplifier output signal.

In presented paper the procedure of simulation in the Multisim environment of EMI levels measured by a selective microvoltmeter on the output of LISN is undertaken. LISN is plugged in the power supply circuit of class D amplifier in accordance with the standard (CISPR 25 for automotive electronics). Predicted level of conducted EMI is depending on supply voltage, output power, parasitic parameters values etc.

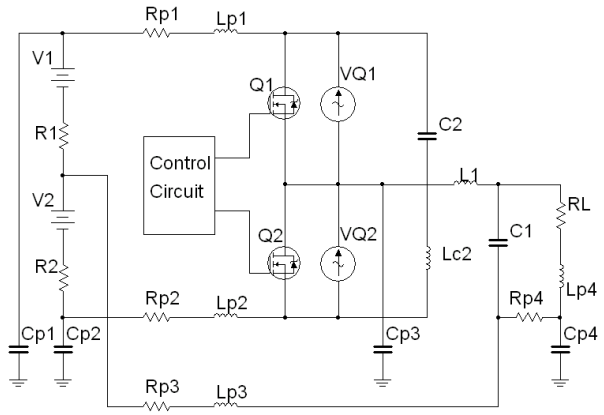


Fig. 2. Equivalent circuit of class D amplifier containing EMI elements.

Equivalent circuit of class D amplifier for the EMI analysis is presented in fig. 2. There are components of signal amplification and EMI creating and distributing elements (components and parasitic parameters) on the circuit. Real components are transistors Q1 and Q2, decoupling capacitor C2, filter elements L1, C1 and load resistor RL. The EMI receptor is internal resistor R1 of power source V1. The basic sources of electromagnetic interference are transistors Q1 and Q2 which are replaced by voltage pulse generators VQ1 and VQ2. These sources are modeling pulse voltages on transistors outputs. EMI propagate in all wires, thus resistances and inductances of power lines Rp1, Lp1 and Rp2, Lp2 are taken into account. Resistance Rp3 and inductance Lp3 of wires to the loudspeaker, are also taken into account. Parasitic inductance Lc2 is connected in series with the decoupling capacitor C2, which in case of long pins of C2 or long PCB board connectors and at the wrong place of setting can increase a generated noise level. Capacitors Cp1 - Cp4 are parasitic components of wires. Cp1 and Cp2 is a capacity between each of power wires and "ground" or car chassis.

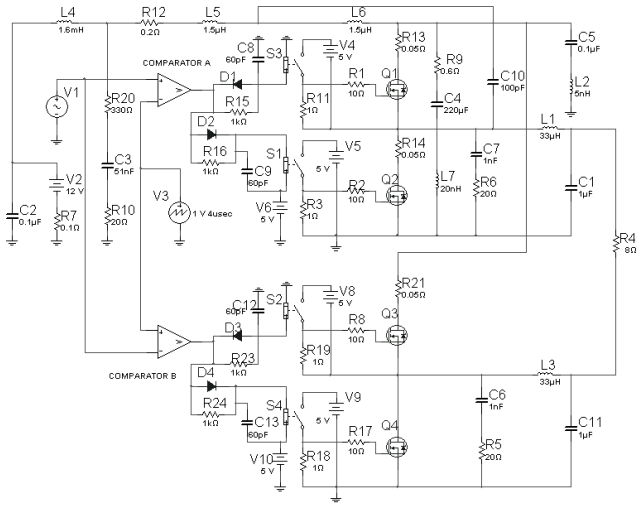


Fig. 3. Class D amplifier simulating circuit containing parasitic elements.

Cp3 is a capacitance between the output of amplifiers and "ground", and Cp1 is a capacitance between the wires of loudspeaker and "ground". The interturn capacity of filter inductance coil is not considered in this equivalent circuit because the number of turns is too small. This equivalent circuit is taken as basis of the class D amplifiers model containing EMI components.

Simplified simulation circuit of class D amplifier connected to LISN is shown on Fig.3. In this circuit there are elements of output transistors "dead time" adjustment. Pin inductance of capacitors, inductance and resistance of wires are taken into account. Parasitic inductances and capacities have the distributed nature, though considering, that analysed band is up to 30 MHz, they can be treated as the undistributed elements L and C [12].

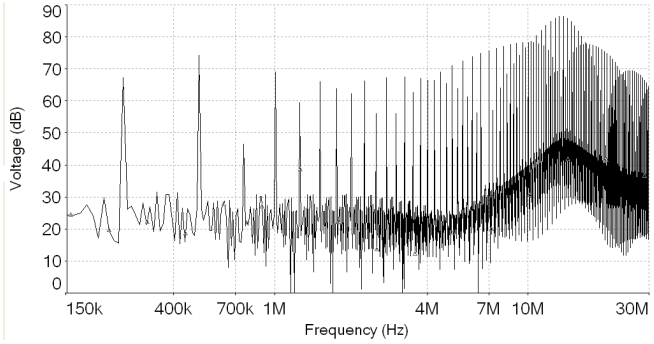


Fig. 4. EMI level of class D amplifier without input signal.

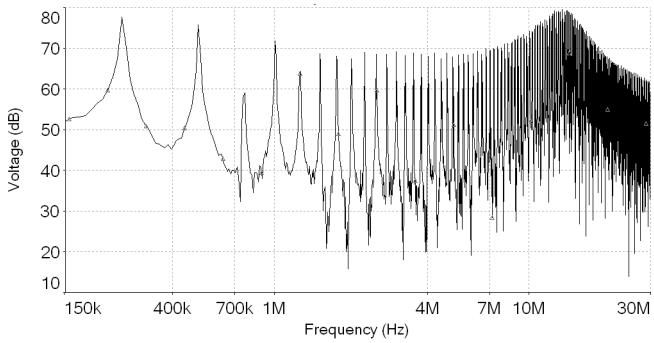


Fig. 5. EMI level of class D amplifier with output power 10W.

The parasitic capacitances between the power supply lines and high current components of amplifier are presented as a capacitor between the output of amplifier and one of the power lines. The parasitic parameters of output transistors (drain-to-source resistance, output capacity of transistors, parameters of output diode etc.) are set by a SPICE model.

In Fig.4 and Fig.5 the simulation curves of EMI levels on the exit of LISN are presented. In Fig. 4 curve corresponds to the case when there is no signal on the amplifier's input. Fig. 5 shows the curve that corresponds to the case when output power on the amplifier is equal to 10W. It is easy to see that at frequencies above 4MHz noise level for some frequencies is smaller up to 6-8 dB in the case when amplifier has input signal. The increase of EMI levels at frequencies above 10 MHz is caused by the influence of the parasitic parameters shown in the model of amplifier for the simulation of the real surges of noise voltages on this frequency band. Simulation curves have rises at frequencies below 1MHz and at frequencies of 15-20MHz. Experimental curves in Fig.1 have the similar behavior. Simulation curves differ from experimental ones no more than 8-10 dB. It confirms the validity of the parasitic parameters choice for this model.

IV. Conclusion

The equivalent circuit of class D amplifier contains components of generation and distribution of conducted EMI is offered. Simulations of class D amplifier considering parasitic parameters are realized. Frequency dependences of simulated and experimental curves generally are the same. The computer simulation and experimental data showed that the EMI level with input signal is less than without a signal. Though the noise level increases at separate frequencies due to the influence of the parasitic parameters. In further research the increased number of parameters influencing on the EMI level will be taken into account for prediction of electromagnetic environment of different types of class D amplifier.

References

[1] Karsten Nielsen "A novel Audio Power Amplifier Topology with High Efficiency and State-of-the-art performance". 109th *AES* Convention, Los Angeles, USA, - 22 - 25 September 2000.

[2] EN 55103-1: 1996 Electromagnetic compatibility - Product family standard for audio, video, audio-visual and entertainment lighting control apparatus for professional use - Part 1: Emission.

[3] CISPR 25:2008. Radio disturbance characteristics for the protection of receivers used on board vehicles, boats, and on devices – Limits and methods of measurement.

[4] http://pdfserv.maxim-ic.com/en/ an/AN3973.pdf Maxim's Active-Emissions-Limiting Circuitry Demystified.

[5] http://pdfserv.maxim-ic.com/en/ an/AN3977.pdf Class D Amplifiers: Fundamentals of Operation and Recent Developments.

[6] http://pdfserv.maxim-ic.com/en/ an/AN3977.pdf and /AN3973. pdf

[7] H.W. Ott, Electromagnetic Compatibility Engineering, New York: Wiley, 2009

[8] Onikienko Y., Pilinsky V., Shvaychenko V. "Features of Electromagnetic Compatibility of Class D Power Amplifiers"/ Technical electrodynamics. Special Issue. Power Electronics and Energy Efficiency. 2011, Pt.1 – pp.39-42. (In Russian)

[9] Onikienko Y., Pilinsky V., Shvaychenko V. "Analysis of the Influence of the Power Supply and Input Signal Parameters on the Level of the Conducted Emissions of the Class D Amplifier" / Technical electrodynamics. Special Issue. Power Electronics and Energy Efficiency. 2012, Pt.1 – pp.70-74. (In Russian)

[10] S.E. Adami, R. Mrad, F. Morel, C. Vollaire, G. Pillonnet, R. Cellier, "Conducted EMI of Integrated Switching Audio Amplifier for Mobile Phone Applications", EMC Compo, nov. 2011

[11] R. Mrad, F. Morel, G. Pillonnet, C. Vollaire, A. Nagari, "Conducted EMI Prediction for Integrated Class D Audio Amplifier", ICECS, dec. 2011

[12] The Suppression of Electromagnetic Interference in the Power Supply Circuits / G.S.Veksler, V.S.Nedochotov, V.V. Pilinsky etc. – Kyiv: Tekhnika, 1990. – 167pp. (In Russian)

FPGA Based Capacitive Touch Pad/Interface

Radovan Stojanović, Nedjeljko Lekić, Zoran Mijanović and Jovan Kovačević
Faculty of Electrical Engineering
University of Montenegro
Podgorica, Montenegro
stox@ac.me

***Abstract* — This paper reports a sensing principle and FPGA design of a capacitive touch pad/interface where the sensing pad is connected to the I/O pin via an external resistor. The circuit transforms the change in pad capacitance into voltage amplitude during charging, discharging and sharing phases. By using multiple pins and resistors, a multipad system is achieved. The sensing algorithm is implemented in VHDL code. The read-out cycle is parallel and short, what results in a high noise immunity in low frequency range. The silicon/hardware requirements are minimal. The interface can be easily embedded into a system-on-chip and used for human-machine, bio-chemical and mechanical sensor interfacing.**

Keywords— FPGA; capacitive pad; sensor; touch interface

I. INTRODUCTION

The capacitive sensors have long history of use in industrial measurements, but only during the last two decades they started to make inroads into human-computer interaction. The expansion of multifunctional and multimedia digital devices and gadgets, such as smart phones, iPads and different mobile and desktop instruments have opened huge opportunities for commercialization and further development of these types of interaction.

In the open literature and available to us applications different capacitive sensing techniques can be identified: oscillator-based, resistor-based, approaches based on switched capacitors and programmable current sources etc. [1, 2]. The read-out circuits are usually implemented in microcontroller [3, 4] or ASSP (Application Specific Standard Products) chips [5].

Essentially, none of the elaborated approaches and systems can fully and satisfactorily answer main requirements of today's and future capacitive sensing interfaces: i) serving almost infinite number of sensor pads; ii) reading pads quickly; iii) achieving high noise immunity; iv) embedding interface in same chip with the rest of digital system and v) reaching a competitive cost.

FPGA (Field Programmable Gate Arrays) technology with full parallelism, huge capacity of digital circuits, high operating speed, on-chip signal processing possibilities and affordable price, shows a potential to solve some of the above issues.

This paper presents an FPGA based methodology for capacitive sensing which includes minimal number of external components and appears to be suitable for human-machine, bio-chemical and mechanical multi sensor interface.

II. SENSING PRINCIPLE AND SYSTEM DESIGN

A. *Sensing Principle*

In single-touch configuration the sensing pad is connected to the bidirectional FPGA pin Pn via resistance R, Fig. 1. The direction of pin, input or outpud, is determined by DIR register, while the input state is read by IN register. C_p is a pin capacitance and for FPGAs has value of about 4pF. C_b is body-ground capacitance, typically in the range 100pF-150 pF, but can be as high as 400pF. C_{x0} and C_{x1} are the capacitances of untouched and touched pad, respectively. Usually, $C_b>>C_{x1}$, expressing the equivalent capacitance pad-finger-ground system as $C_x \approx C_{x0}+C_{x1}$. C_{xo} has the value of 6.5pF for D=10m, while C_{x1} decreases from 7pF-1.5pF for $0.1\text{mm}<d_{x1}<1\text{mm}$.

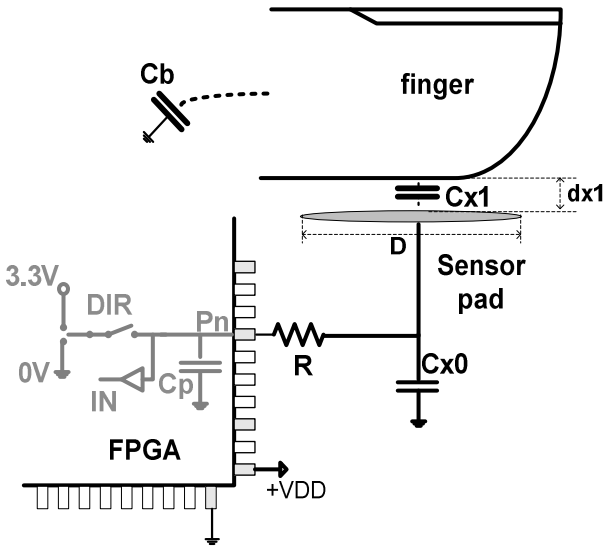


Fig. 1. FPGA based capacitive pad/interface

The sensing process starts with "CHARGE" phase, Fig. 2, by setting logical "1" (3.3V) to the pin P_n for time t_c=0.75us-2us, long enough to charge C_p and C_x to the values $V_1=V_2$=3.3V. Dashed line indicates situation for increased value of Cx, pad touched. During the "DISCHARGE" phase, C_p discharges totally, because of V_1=0V, while V_2 declines by time constant RC_x reaching the end value $V_2(t_d)$ after the discharging time t_d. t_d may be chosen between 200ns-750ns.

The remaining charge on C_x, after time td, is:

$$Q_{C_x}(t_d) = V_2(t_d)C_x = C_x V_{DD} e^{\frac{-t_d}{RC_x}} \quad (1)$$

During the "SHARE" phase, with duration of t_s, pin P_n is switched to the high impedance mode (Z mode) and $Q_{Cx}(t_d)$ shares proportionally to C_p and C_x. At the end of the sharing $V_1 \approx V_2$, $i \approx 0$ and voltage onto pin Pn has analytical value:

$$V_1 = (V_{DD} e^{\frac{-t_d}{RC_x}})/(1 + \frac{C_p}{C_x}) \quad (2)$$

"MEASURE" phase begins with delay t_m, when IN register is read. As seen, for known C_p, t_d and R, the V_1 is a function of C_x and further of dx_1 – which is a finger distance. V_1 can reach the level bellow or above the pin's threshold V_{TR} depending on whether the pad is touched or not. Translated into digital domain it means logical "1" or "0" of the corresponding bit within the register IN.

Finding an optimal value for R is an additional issue. Fig 3 shows graphical illustration of (2) for t_d=200us, V_{DD}=3.3V, C_p=4pF and R as parameter. As seen, for R=68K, V_{TR} corresponds to $V_{TR} \approx (V_1(\text{touched}) - V_1(\text{untouched}))/2$, and that is an optimal case. Additional purpose of R is to protect the FPGA pin from the electrostatic discharges and with 68K it can go up to1000V.

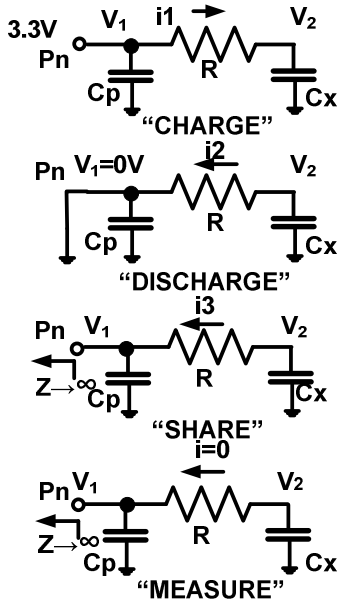


Fig. 2. The sensing algorithm

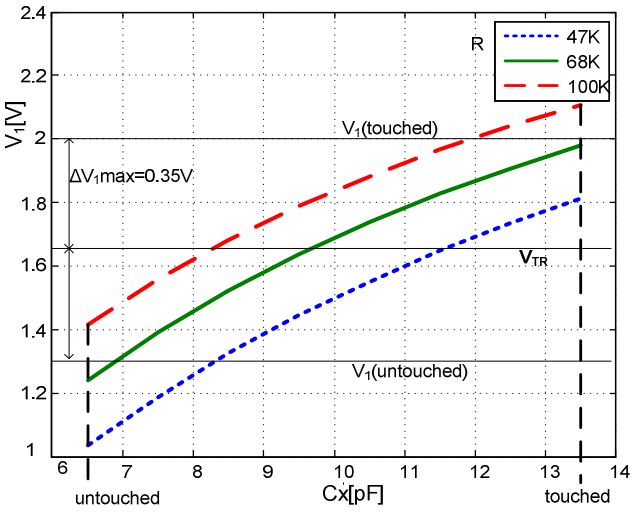


Fig. 3. Selection of external resistor R

B. FPGA Design Methodology

The Single-Pad Capacitive Interface (SPCI) presents a basic design core of the sensing interface, Fig. 4. It is a three terminal port with clock (CLK), sensor input (S) and sensor status (SS) signals. In essence, it is a counter with comparator. The generic constants CHARGE_TIME_NS, DISCHARGE_TIME_NS, START_READ_NS, correspond to the t_c, t_d and t_s+t_m respectively, while PERIOD_NS marks a reading cycle. These constants are parameterized, but can alternatively be stored in configuration registers. The VHDL code of SPCI unit, Listing 1, presents the software implementation of sensing protocol.

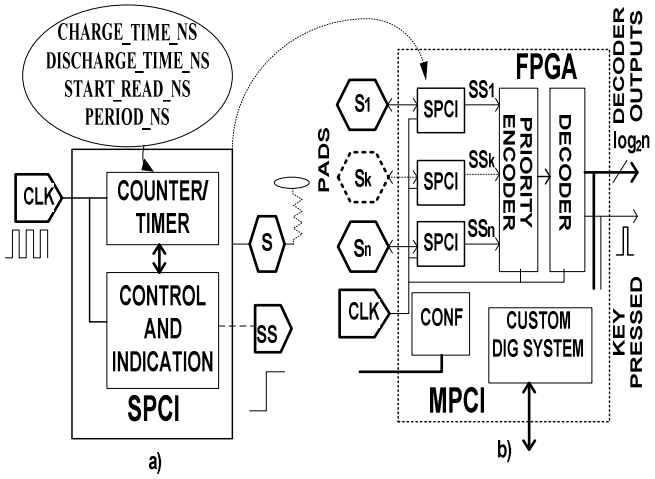


Fig. 4. SPCI) and MPCI FPGA architectures

In the case of Multi-Pad Capacitive Interface MPCI, SPCIs are "copied" *n* times in parallel. The Si (i=1..n) signals are sensor inputs. SS_i (i=1..n) present internal outputs that are fed to the PRIORITY ENCODER and DECODER producing KEY CODE and KEY PRESSED outputs. KEY CODE is the length of $\log_2 n$, while KEY PRESSED is a binary signal which is active when one of pads is pressed and can be used for producing interrupt. Both are carried out or forwarded to

the CUSTOM DIGITAL SYSTEM unit. In our case CLK has frequency of 48 MHz.

```
-- SPCI CODE
LIBRARY ieee;
USE ieee.std_logic_1164.all;
USE ieee.numeric_std.all;

ENTITY FPGA_tri_state IS
generic (CHARGE_TIME_NS       : integer := 144405;
         DISCHARGE_TIME_NS    : integer := 4545;
         START_READ_NS        : integer := 10000;
         PERIOD_NS            : integer := 400000
         );
    PORT(
            CLK    : in std_logic;
            S  : inout std_logic;
            SS : out std_logic
        );
END FPGA_tri_state;

ARCHITECTURE arh OF FPGA_tri_state IS

constant clock_cycle  : integer := 21;
constant one_time: integer := CHARGE_TIME_NS / clock_cycle;
constant zero_time    : integer := (CHARGE_TIME_NS + DISCHARGE_TIME_NS) /
                                    clock_cycle;
constant read_time    : integer := (START_READ_NS + CHARGE_TIME_NS +
                                    DISCHARGE_TIME_NS)/ clock_cycle;
constant max_time     : integer := PERIOD_NS / clock_cycle;

SIGNAL  b  : STD_LOGIC; -- DFF that stores feedback value.
shared variable counter : integer range 0 to PERIOD_NS/clock_cycle :=0;
signal dir      : std_logic;

BEGIN
    PROCESS(CLK)
    BEGIN
        IF rising_edge(CLK) THEN  -- Creates the flipflops
            if counter < max_time then
                counter := counter + 1;
                if counter < zero_time then
                    dir <= '1';
                else
                    dir <= '0';
                end if;
            else
                dir <= '0';
                counter := 0;
            end if;
        END IF;

        SS <= b;

        IF( dir = '0') THEN
            if counter = zero_time then
                S <= 'Z';
            else
                if counter = read_time then
                    b <= S;
                end if;
            end if;
        ELSE
            if counter < one_time then
                S <= '1';
            elsif counter >= one_time and counter < zero_time then
                S <= '0';
            end if;
        END IF;
    END PROCESS;
END arh;
```

Listing 1. VHDL implementation of sensing algorithm

III. EXPERIMENTAL RESULTS

On-chip verification of the proposed approach was performed on EP1C6Q240C8 FPGA chip from Altera Cyclone Series [6]. After the code design, compilation and debugging, the chip is configured in SPCI configuration. The pad of 10mm diameter is isolated with 0.1mm paper layer. In real measurements, Fig. 5, the oscilloscope probes were connected to the ends of resistor R, introducing the additional capacitances C_{op} of 15pF in parallel to C_x and C_p, Agilent Tech. DSO3102A and Agilent Tech. probes N2862A. It significantly reduces the sensing effects having in mind the capacitance C_{x1}, about 7pF, see Fig.1. Thus, in order to emphasize visual effects, the charging, discharanging and sharing times are increased to t_c=24us, t_d=2.5us and t_s+t_m=15us, respectively, Fig. 6. As can be seen, the traces match very well the theoretical prediction, Fig. 2, and the presence of finger is detected. In real situation the distinction effect is even more pronounced.

The second criterion was checking of occupancy of silicon resources, Table I. SPCI occupies only 60 LEs, (Logical Elements) which is mere 1% of total number of available LEs for chip EP1C6Q240C8, which has 5980 LEs in total. MPCI in 2, 4, 8, 16, 32 and 64 input configurations occupy 140, 151, 168, 213, 297 and 462 LEs. It can be noted that the design is very silicon-effective and, as an example, 64 pads configuration occupies only 7.7 % of the EP1C6Q240C8 resources and 72 pins out of 240. Practically, N inputs MPCI may even be embedded in a low-capacity FPGA. The only limitation is a number of available FPGA pins, usually several hundreds.

Fig. 5. Oscilloscope probes introduce the additional capacitances C_{op}.

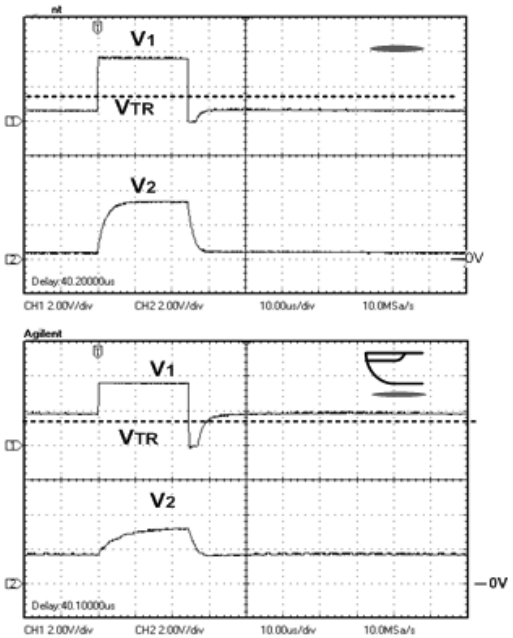


Fig. 6. Real waveforms for untouched (up) and touched (down) pad, D=10mm and d_{x1}=0.1mm, paper insulator.

The third test criterion was noise immunity. Because of the parallelism, the sampling period does not increase with the number of pads, and can be adjusted from 5us to 20us in a real situation. Additional benefit in terms of noise is that the input

impedance of pin Pn has a low value during phases "CHARGE" and "DISCHARGE" and high value (Z), which is noise sensitive, only during "SHARE" and "MEASURE". This makes the sensor being very robust regarding to the low-frequency noise such is AC noise of 50Hz or 60Hz, because it can affect very shortly, only during "SHARE" and "MEASURE" phases, whose total duration is about 1µs. For example, AC noise with induced amplitude of 100V has an increment of ΔV_{ACmax}= 22mV, during 1µs on steepest interval. It is much less than the margin $|V_1(\text{touched})-V_{TR}| \approx |V_1(\text{untouched})-V_{TR}|$, see Fig. 3, for R=68K the margin is above 0.35V. Thus, mains voltage can be considered as DC signal which does not influence the sensing mechanism. The influence of radiated noise (high frequency) can be reduced by implementing software filters in VHDL, while environmental noise has a slow increment over time.

TABLE I. HARDWARE OCCUPATION IN LES

# of pads	# of LEs	% of available LEs	# of occupied pins
1	60	1	5
2	140	2.4	5
4	151	2.5	8
8	168	2.8	13
16	213	3.5	22
32	297	5.0	39
64	462	7.7	72

IV. CONCLUSIONS

A capacitive sensing technique is presented. It is fully digital and as such very suitable for implementation in system-on-chip like FPGA. The sensing principle and design architectures are presented together with the core code. The testing results against set criteria are elaborated. The system is very versatile and can be used in low cost embedded human-machine, bio-chemical and mechanical sensor interfacing.

ACKNOWLEDGMENT

This paper presents a part of the research performed in the projects: "Development and implementation of embedded systems for medical applications", MESI. The authors are grateful for their support.

REFERENCES

[1] L. K. Baxter, Capacitive sensors: Design and Applications, IEEE Press, 1996.

[2] X. Li and G. C. M. Meijer, "An Accurate Interface for Capacitive Sensors", IEEE Trans. Instrum. Meas., vol. 51, No. 5, pp. 935–939, October 2002.

[3] Z. Albus, "PCB-Based Capacitive Touch Sensing With MSP430", Texas Instruments, Dallas, Application Report SLAA363A, 2007.

[4] N. Lekic, R. Stojanovic, Z. Mijanovic , "An approach of using microcontroller port as capacitive keypad", *Proceedings of the IEEE 20th International Electrotechnical and Computer Science Conference, ERK-2011,* September 2011, Portoroz, Slovenija, pp. 43–46.

[5] Fujitsu Microelectronics America, "Interfacing the FMA1127 Touch Sensor Controller with Fujitsu Microcontrollers", Application Note, 2009.

[6] *Altera Corporation, Cyclone Device Handbook, Volume 1, 2008,* http://www.altera.com/literature/hb/cyc/cyc_c5v1.pdf .

SECTION IV

ELECTRONIC SYSTEMS

A 1.2V Wideband CMOS Mixer with High Conversion Gain and Low Flicker Noise

Chao Chen
ASIC research center
Southeast University
Nanjing, China
chen_seu@yahoo.cn

Jianhui Wu
ASIC research center
Southeast University
Nanjing, China
wjh@seu.edu.cn

***Abstract*—This paper presents the design of a wideband CMOS mixer (frequency convertor) for direct-conversion receivers. The mixer consists of a transconductance stage, a passive switching stage and a current amplifier. The current amplifier with gm-boosted structure realizes a low input resistance. It absorbs the down-converted IF current at the output node of the switching pair and amplifies it through a current mirror. The low noise characteristic of the current amplifier helps to improve the noise performance of the proposed mixer. In order to improve the power efficiency, current-reusing technique is adopted in the transconductance stage. The proposed mixer is designed in 0.13µm RFCMOS process. Simulation results show a 20dB conversion gain, a 10dB SSB noise figure and a 4.5 dBm IIP3 over the wideband operation frequency band from 1GHz to 3GHz while dissipating 2.5mA under a 1.2V supply voltage.**

Keywords—CMOS, frequency convertor, mixer, wideband, RF, direct-conversion receiver, gm-boosted, current amplifier.

I. INTRODUCTION

The direct conversion architecture, which down-converts the RF signal directly to baseband signals, have many benefits such as high integration level, simplicity in structure and low power and low cost [1], [2]. For those direct conversion receivers such as handset DTV receivers, achieving low flicker noise and low power dissipation while providing high performance is the major goal to design the down-converted mixer in the receiving chain. Active mixers such as the traditional Gilbert mixers, often suffers from high flicker noise which restricts its usage in direct-conversion receivers. Furthermore, due to the limited input conductance seen into the switching pair in a Gilbert mixer, the parasitic capacitances at the source of switching pair bypass a portion of RF current generated by gm stage, causing notable degradation of conversion gain and linearity performance at high frequency band [3], [4].

To meet the low flicker noise requirement of the mixer in direct-conversion receivers and to achieve flat conversion gain in wide frequency band, the passive mixer is often chosen by some previous works [5], [6], [7]. A passive mixer employs a double balanced switching pair working in the triode region and cut-off region to commutate the polarity of RF currents. In order to provide enough conversion gain and sufficient isolation between IF and RF ports, a RF transconductance stage and IF transimpedance stage are used in cooperation with the passive switching pair [6], [7]. The block diagram of a traditional passive mixer is shown in Fig. 1 (a). The transconductance (gm) stage converts the RF input voltage into RF current. The switching pair conducts and commutates the RF current into the transimpedance amplifier (TIA). The down-converted IF current flows into the input of the TIA and is converted into output IF voltage by the TIA. TIAs in most of traditional passive mixers are built by an op-amp with feedback resistors. However, there are two issues concerning the usage of the TIA. The first problem is the noise issue. A two stage op-amp is generally noisy. Thus large dc bias current in the gm stage is needed to suppress the noise induced by the op-amp. The power efficiency is therefore affected. The other issue is the RF feed through at the IF output node. Because of the relatively high output impedance of the op-amp and the limited loop gain at high frequency, the RF current at the input node of op-amp can feed through the capacitor CL, flowing through the output impedance of the op-amp. Thus a RF voltage can be seen at the output node.

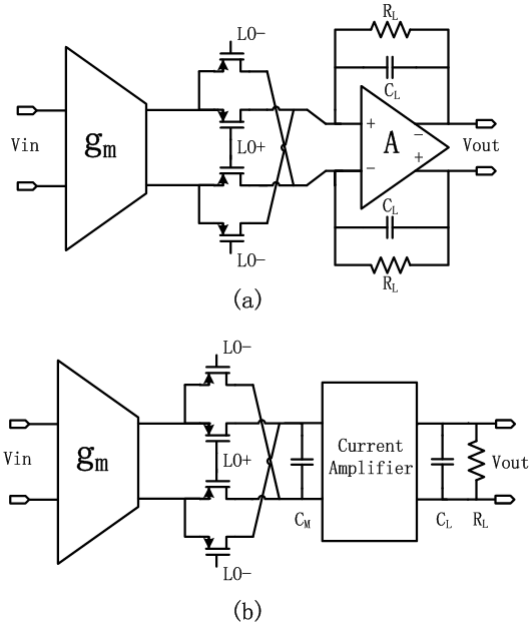


Fig. 1. (a) Block diagram of a traditional passive mixer. (b) Block diagram of the proposed passive mixer

For the purpose of improving the noise performance and RF feed through effect caused by TIA in the traditional passive mixer, a novel passive mixer structure shown in Fig. 1 (b) is proposed in

this paper. The proposed mixer replaces the TIA with a current amplifier with load resistors. The current amplifier employs gm-boost technique in the input stage, realizing low input impedance at the output node of the switching pair. It absorbs and amplifies the IF current, turning it into output voltage at the load resistors. The limited bandwidth of the current amplifier and the existence of C_L helps to diminish the RF current at the output node of the mixer. The current amplifier also has better noise performance than op-amp based TIA. So the noise figure of the proposed mixer can be lower than traditional passive mixer under the same dc bias current.

This paper is arranged as follows. Section II describes the structure and working principles of the proposed mixer. In Section III, the simulation results of the proposed mixer are presented and discussed. Finally, conclusions are stated in Section IV.

II. CIRCUIT DESIGN

The block diagram of the proposed mixer is shown in Fig. 1 (b). The transconductance stage converts the RF input voltage into RF current. The RF current's polarity is switched by the double balanced switching pair under the LO signal. The working status of the switching transistors change from cutoff region to triode region according to the LO signal. The switching transistor acts like a small resistor when the switch is on and acts like an open circuit when the switch is off. At the output node of the switching pair, the small input resistance of the current amplifier provides low impedance at low frequency band. The parallel connected capacitance CM provides low impedance at high frequency. Therefore, despite of the frequency conversion effect of the switching pair, low input impedance is obtained at the output node of the gm stage at both RF and IF frequency. This can prevent the RF current from bypassing through the parasitic capacitors and the limited output resistance of the gm transistors. The conversion gain and linearity performance is thus improved.

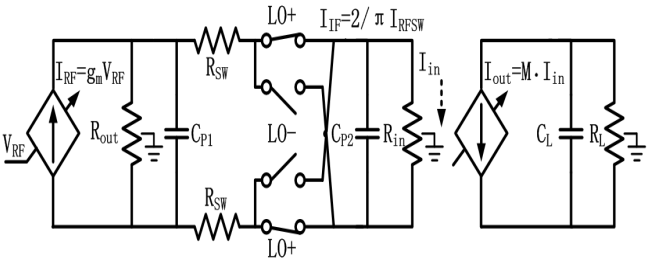


Fig. 2. Simplified equivalent circuit of the proposed mixer

The simplified equivalent circuit of the proposed mixer is indicated in Fig. 2. The transconductance stage is modeled as a voltage-controlled current source which has an output resistance of R_{out}. The switching pair can be represented as four ideal switches in series with a R_{SW} which denotes the on-resistance of the switching transistors. C_{P1} and C_{P2} are the capacitance at the input and the output nodes of the switching pair. The capacitor C_{P2} at the output node of the switching pair eliminates all of the mixing outcomes at RF frequencies while reserving the IF signals. The current amplifier is modeled as a current-controlled current source with an input resistance of R_{in}. The input current is amplified by M through a current mirror. R_L and C_L are the load resistor and capacitor which convert the current into output voltage and also provide low-pass filtering for the down converted signal.

When the switch is on, the resistance between the output and input node is small. Down-conversion and up-conversion are performed on both sides of the switching pair. The down-converted IF signal with the frequency of $\omega_{IF} = \omega_{RF} - \omega_{LO}$ is up-converted to ω_{RF} and counteract with the input RF signal. Meanwhile, an RF signal with the image frequency of $\omega_{RFim} = \omega_{LO} - \omega_{IF}$ appears on the input side of the switching pair. The image signal is also down-converted to ω_{IF} with opposite polarity following $\omega_{IF} = \omega_{LO} - \omega_{RFim}$. At the input node of the ideal switching pair shown in Fig. 2, the existence of the RF voltages at both ω_{RF} and ω_{RFim} diminish the IF current flowing into the current amplifier. In order to suppress this counteraction effect, R_{IN} should be kept sufficiently low to limit the amplitude of the IF voltage. R_{OUT} should be high to limit the RF current at the image frequency. In addition, a low R_{SW} is also needed to ensure low impedance at the output node of the gm stage, absorbing most of the RF current into the switching pair. Another benefit of low R_{SW} is that, the variation of R_{SW} does not change the current flowing through the switching pair. The linearity performance is thus improved.

The overall voltage conversion gain of the proposed mixer can be approximately presented as (1).

$$\frac{V_{out}(\omega_{out})}{V_{in}(\omega_{IN})} = \frac{2}{\pi} g_m \left(\frac{M}{1+j\omega_{IF}C_{P2}R_{in}(\omega_{out})}\right)\left(\frac{R_L}{1+j\omega_{out}C_L R_L}\right) \quad (1)$$

At the output IF frequency, the input resistance $R_{in}(\omega_{IF})$ of the current amplifier is as low as several ohms. And the 3-dB bandwidth of the output RC network is beyond the output IF bandwidth. The conversion gain within the entire output bandwidth can be written as (2).

$$\frac{V_{out}(\omega_{IF})}{V_{in}(\omega_{RF})} = \frac{2}{\pi} g_m \cdot M \cdot R_L \quad (2)$$

For the unwanted high frequency signals, $R_{in}(\omega_{out})$ increases because of the limited GBW of the feedback loop gain of the gm-boosted input stage of the current amplifier. As indicated by (1), the proposed mixer thus shows second-order low-pass filtering effect to provide sufficient attenuation.

Fig. 3 shows the simplified circuit schematic of the transconductance stage and switching stage of the proposed mixer. The NMOS transistors M1, 2 and PMOS transistors M3, 4 constitute the complementary structure in which both NMOS transistors and PMOS transistors provide transconductance. M1 and M2 are biased through current mirror while M3,4 is self-biased through a large resistor RB3,4 which is connected between its gate and drain. Capacitors C0/1 connects the gate of M1/2 to the gate of M3/4. In order to suppress noise from the current amplifier, large gm of the transconductance stage is required. Compromising between NF and power dissipation, a moderate bias current of 1.6mA is chosen. Capacitors C2 and C3 prevent the DC bias points of gm stage and the input node of the current amplifier from intervening with each other. The source and drain bias voltage of

the switching pair is determined by the DC voltage at the node IF±. This voltage is one V_{dsat} lower from the supply voltage, giving adequate voltage headroom for high amplitude of the LO signal.

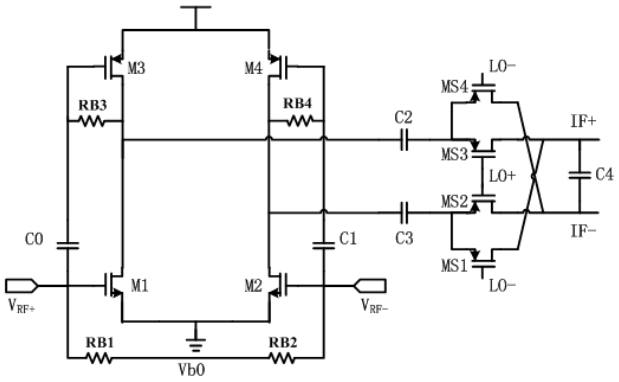


Fig. 3. Simplified circuit schematic of the transconductance stage and switching stage of the proposed passive mixer.

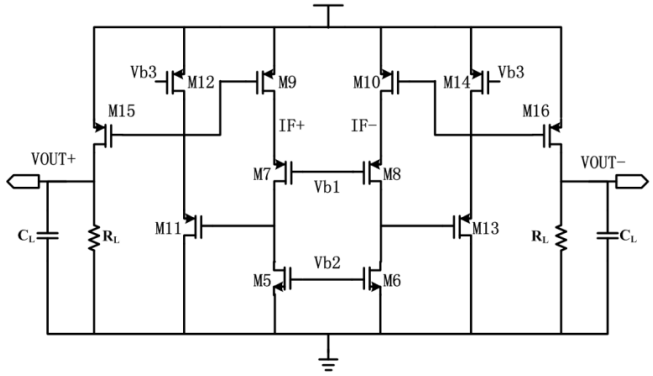


Fig. 4. Fig. 4. Simplified circuit schematic of the current amplifier

Fig. 4 indicates the circuit schematic of the current amplifier. The fully differential current amplifier consist a gm-boosted input stage and current mirror output stage. Taking one half of the differential circuit for analysis, the gm-boosted input stage consists of NMOS transistor M6, PMOS transistors M8, M10, M13 and M14. M6 is used as a current source which set the current of M8 fixed. The drain of M8 is connected to the voltage follower consists of M13 and M14. The output of the voltage follower is connected to the gate of M10 to form a negative feedback loop. The input current is injected at the source of M8 and is absorbed by M10. The negative feedback boosts the effective gm of M8, as indicated in (3).

$$g_{meff}(\omega) = \left(1 + \frac{g_{m10} r_{o6}}{1 + j\omega r_{o6} C_{P13}}\right) \cdot g_{m8} \quad (3)$$

Within the bandwidth of the feedback loop, the effective gm of M8 is boosted by times of its original value. A low input resistance at IF frequency is achieved. M16 copies the current of M10 and magnify it by 2. Thus the down-converted IF current is copied and amplified in M16. The current flows through the output resistor R_L, being converted into output voltage. The noise performance of the current amplifier is much better than an op-amp based trans-impedance amplifier. If the same NF with the op-amp based passive mixer is to be achieved, the gm requirement of the transconductance stage can be lessened. Thus the power consumption of the proposed mixer is reduced.

The component parameters of the proposed mixer are illustrated in Table I.

TABLE I
THE COMPONENT PARAMETERS OF THE PROPOSED MIXER

Device	Parameter	Device	Parameter
C0,1	2pF	C4	2pF
C2,3	1.5pF	CL	2pF
RB1-4	20KΩ	M9,10	40μm/1μm
M1,2	22.4μm/0.13μm	M11,13	160μm/2μm
M3,4	104μm/0.13μm	M12,14	32μm/1μm
MS1-4	16μm/0.13μm	M15,16	80μm/1μm
M5,6	80μm/4μm	RL	500Ω
M7,8	192μm/0.3μm		

III. SIMULATION RESULTS

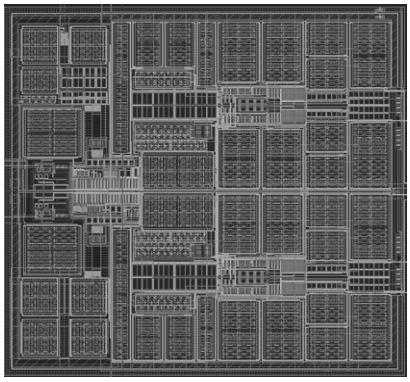


Fig. 5. Layout of the proposed passive mixer

The proposed mixer is designed in TSMC 0.13μm RFCMOS process with 1.2V supply voltage. The overall DC bias Current is 2.5mA excluding the current dissipated by LO buffer circuit. The layout of the proposed passive mixer is shown in Fig. 5 which occupies a compact active area of 352μm*348μm.

Fig. 6 shows the simulation results of conversion gain and NF versus input frequency from 1GHz to 3GHz at a fixed output IF frequency of 2MHz. The conversion gain at 1GHz is 20.5dB; this value drops by only 1.3dB to around 19.2dB at 3GHz. The NF is simulated at the output IF frequency of 2MHz. The NF value under 1GHz input frequency is 10dB. When the input frequency increases, the NF increases by 1.2dB to 11.2dB under the input frequency of 3GHz. The degradation of NF performance attributes to the RF current leakage under high input frequency. As Fig. 6 indicates, the RF current leakage due to parasitic capacitance also leads to the conversion gain decrease.

In Fig. 7, simulated NF versus output frequency under a fixed 1GHz LO frequency is shown. Because of the low noise characteristic of the current amplifier, low noise figure is achieved in the output frequency band from 10 KHz to 10 MHz. The NF is 11.5dB and 10dB at the output frequency of 10 KHz and 1 MHz respectively. Flicker noise dominates in the low frequency band below 10 KHz. Since there is no dc current in the

switching pair, the proposed passive mixer has a low flicker noise corner frequency of 200 KHz.

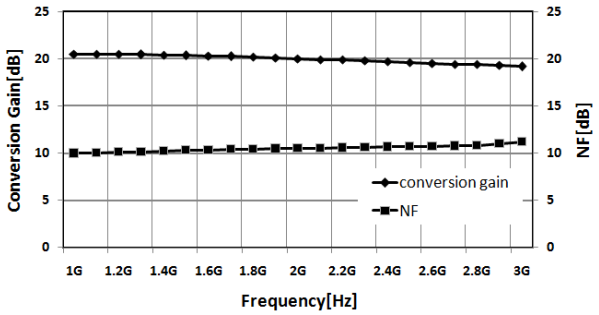


Fig. 6. Simulated conversion gain and NF of the proposed mixer

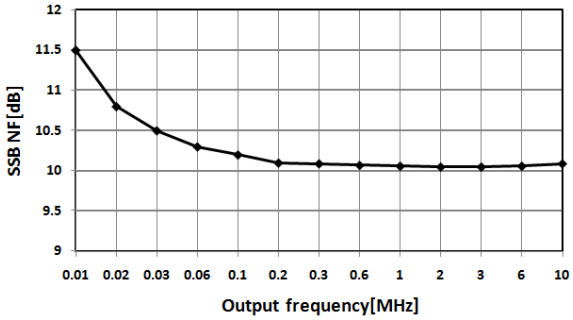


Fig. 7. Simulated NF versus output frequency under 1GHz LO frequency

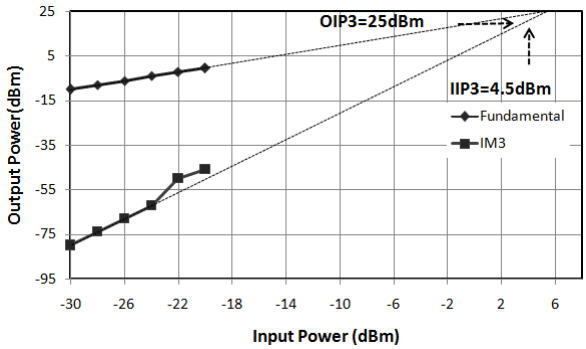


Fig. 8. Simulated IIP3 and OIP3 of the proposed mixer at the LO frequency of 1GHz.

The linearity of the proposed mixer is determined by the transconductance stage. The switching pair only commutates and conducts the RF current from the gm stage. Thus the fluctuation of the small on-resistance of switches has little impact to the overall linearity. The current amplifier working in low frequency band is based on negative feedback and has a good linearity performance. To optimize the linearity of the gm stage, the NMOS and PMOS transistors are biased at the V_{GS} value where the second derivation coefficients of gm are zero.

Fig. 8 shows the simulation results of IIP3 and OIP3 in two tone test procedures under a LO frequency of 1GHz. The RF input sine-signals locate at frequencies of 1001MHz and 1002MHz. Output signal and IM3 locates at 2MHz and 3MHz respectively. Fig. 8 indicates that, a 4.5dBm IIP3 and a 25dBm OIP3 can be observed.

Table II compares the performance of the proposed mixer with other state-of-the-art mixers. With the help of the current reusing in the gm stage and the low noise characteristic of the current amplifier, the proposed passive mixer shows better performances in power efficiency and noise performance compared with the other mixers. Meanwhile, a high conversion gain is also achieved while keeping a moderate IIP3 value.

TABLE II.
PERFORMANCE COMPARED WITH OTHER STATE-OF-THE-ART MIXERS

	Freq [GHz]	VDD [V]	Gain [dB]	IIP3 [dBm]	NF [dB]	Current [mA]
This work	**1-3**	**1.2**	**20**	**4.5**	**10**	**2.5**
[3]	**2.4**	**1.8**	**15.7**	**1**	**12.9**	**4.5**
[4]	**5.2**	**1**	**4.5**	**-6**	**16**	**3.8**
[6]	**1.5-2.3**	**2**	**22**	**7**	**12.5**	**5**
[7]	**0.05-0.86**	**1.2**	**20**	**8.5**	**12**	**4.5**

IV. CONCLUSION

A wideband CMOS passive mixer with high conversion gain and low flicker noise is proposed in this paper. To improve the power efficiency, current-reuse technique is adopted in the transconductance stage. In the proposed mixer, the op-amp based trans-impedance amplifier used in the traditional passive mixer is replaced with a current amplifier, which has a better noise performance. The current amplifier with gm-boosted structure realizes a low input resistance at the output node of the switching pair. The down-converted IF current from the mixing core is absorbed and amplified. The proposed mixer is designed in 0.13μm TSMC RFCMOS process. Simulation results show a 20dB conversion gain, a 10dB SSB noise figure and a 4.5 dBm IIP3 over the wideband operation frequency band from 1GHz to 3GHz are achieved. The proposed mixer dissipates 2.5mA current under a 1.2V supply voltage.

ACKNOWLEDGEMENT

This work was supported by The Natural Science Foundation of Jiangsu under the project number BK2011334 and supported by the Scientific Research Foundation of Graduate School of Southeast University under project number 196058395.

REFERENCES

[1] P. Antoine, P. Bauser, H. Beaulaton, "A direct-conversion receiver for DVB-H," IEEE J. Solid-State Circuits, vol. 40, no. 12, pp. 2536–2546, Dec. 2005.

[2] M. Valla, G. Montagna, "A 72-mW CMOS 802.11a direct conversion front-end with 3.5-dB NF and 200-kHz 1/f noise corner," IEEE J. Solid-State Circuits, vol. 40, no. 4, pp. 970–977, Apr. 2005.

[3] V.Vidojkovic, "A Low-Voltage Folded-Switching Mixer in 0.18μm CMOS", IEEE J. Solid-State Circuits, vol.40, pp. 1259-1264, June, 2005.

[4] Chih-Hao Chen, Pei-Yuan Chiang, "A Low Voltage Mixer with Improved Noise Figure", IEEE Microwave and Wireless Components Letters, vol. 19, pp.92-95, Feb., 2009.

[5] Sining Zhou, and Mau-Chung Frank Chang, "A CMOS Passive Mixer With Low Flicker Noise for Low-Power Direct-Conversion Receiver", IEEE J. Solid-State Circuits, vol. 40, no. 5, pp. 1084–1094, May. 2005.

[6] Namsoo Kim, Vladimir Aparin, and Lawrence E. Larson, "A Resistively Degenerated Wideband Passive Mixer with Low Noise Figure and High IIP2", IEEE Transactions on Microwave Theory and Techniques, vol.58, no.4, PP.820-830, April 2010.

[7] Viet-Hoang Le, Hoai-Nam Nguyen, "A Passive Mixer for a Wideband TV Tuner", IEEE Transactions on Circuits and Systems—II: Express Briefs, vol. 58, no. 7, pp.398-401, July 2011.

A 1.2V Current Mirror Double Balanced Mixer with Current Reusing Technique

Xiao Shi, Jianhui Wu, Zhilin Liu
National ASIC research center
Southeast University
Nanjing, China
xiaoshiseu@yahoo.cn, wjh@seu.edu.cn, sdlzl887@163.com

Abstract — **This paper presents the design of a double balanced current-mirror mixer with current reusing technique. The proposed mixer is based on high linearity current mirror structure which guarantees a high IIP3. To improve the power efficiency of the current mirror structure, a current reusing technique is used to share the bias current of the transconductance stage and the output stage. Compared with the traditional current mirror mixer, a power decrease of 36% can be achieved. The mixer is designed in 0.13-µm 1P8M RFCMOS process; operating at the frequency of 1GHz. Simulation results indicate a peak conversion gain of 12.6dB, a relatively high IIP3 of 4.0dBm and a moderate NF of 16dB. The whole mixer has a power consumption of 3.84mW under 1.2V supply voltage.**

Keywords- CMOS, current mirror, current reusing, folded structure, high linearity, mixer.

I. INTRODUCTION

High linearity is required in many modern wireless communication systems such as CDMA and OFDM receivers. According to the linearity cascading theory, the linearity of back-end module dominates the linearity of the whole receiver. For RF circuits such as LNA and mixer, design tricks to improve linearity is limited because of the wide bandwidth of the circuit. However, as the last module of the RF front-end, the mixer plays a key role to achieve high linearity.

There are several approaches to improve the linearity performance of mixers such as the multi-gate technique [1], [2], the pre-distortion technique [3] and the current mirror mixer [4]. Multi-gate technique involves using two transconductance transistors which are biased in different V_{GS} with opposite polarity of g_m''. The third order component in output current of the gm stage can be counteracted. Because of the narrow flatten area of the counteracted g_m'', the improvement in linearity can only be observed when the input power is low. The pre-distortion technique linearizes the down-converted IF signal with the pre-distorted RF signals. The IMD generated from the mixer in IF band was canceled by the controlled RF error signal, which is generated by the pre-distorter. The magnitude and phase of the RF error signal were properly adjusted through a vector modulator. The pre-distortion technique can improve the IIP3 by 3.5dB whereas at the cost of great increment in circuit complexity.

The current mirror mixer is another choice to realize a mixer which can achieve high linearity and good adaption to low supply voltage. The current mirror mixer is based on a pair of differential current mirrors in which a double balanced switching pair is embedded to duplicate and down- convert the RF current from gm stage into IF current in the output stage. This structure avoids using of stacked structure that is unsuited to low supply voltage, and also avoids the linearity deterioration due to the cascading structure. The high linearity characteristic of current mirror prevents linearity deterioration when the current from gm stage is copied to the IF stage. The current mirror structure set the bias current of the output stage to be equal with the current of the gm stage which is usually above 1mA. The power efficiency is the main shortcoming of that structure.

In this paper, a current-reusing technique is applied in the current mirror mixer to improve the power efficiency. This approach shares the bias current of the gm stage and the output stage. Compared with the traditional current mirror mixer, a power decrement of 36% can be achieved without degrading the IIP3 notably.

This paper is arranged as follows. Section II presents the principle and circuit structure of the proposed current reusing current mirror mixer. In Section III, simulation results are presented and discussed. Finally, conclusions are stated in Section IV.

II. CIRCUIT DESIGN

In this Section, two mixers are introduced. The first one, which proposed previously, is utilized to interpret how to realize the high linearity by current mirror. Then the second one is a newly proposed mixer which uses the current reusing technique to reduce the power dissipation. The basic structure of the mixers is built on two pairs of current mirrors, while the two sides of the current mirrors, the gm stage and the load stage, are connected by a double balanced passive switching pair. The current mirrors duplicate the differential gate voltage from the input stage to the output stage, so the RF current change could be fully copied to the IF output stage. The switching pair will just switch over the polarity of the differential gate voltage of the current mirror, where a time-vary system is formed. Thus at the load of the IF output stage, the down-converted current is proportional to the RF current. With the current flowing through the load resistors, the ultimate output voltage is obtained. In this paper, as shown below, both the previously and newly proposed mixers use the folded structure, which is suited for the low supply voltage.

Fig.1 shows the principal structure of the basic current mirror mixer. The common source differential pair M1, 2 with degraded

This work was supported by the Natural Science Foundation of Jiangsu Province, China under Grant BK2011334.

resistors R1 consist the gm stage which converts the input RF voltage into current. The current is then transformed into voltage again on the diode-connected M3, M4. When the gate voltage of M12, 13 is low, the switch transistors M12, 13 are on, so the voltage signals from the M3 and M4 at the gm stage will be duplicated to M6 and M5 separately. When the gate voltages' polarity reserved, the result will also be inverted. While the double balanced switching pair switching on or switching off, the gate voltages of M5, 6 will change with the alternation of M3, 4's gate voltages. In order to copy the current precisely, the channel length and width of M5, 6 are set to the same as M3, 4 to form a current mirror. Because of the relatively high gate impedance of M5, 6 compared with the on-resistance of the switch transistors, gate voltages of the diode-connected M3, 4 passing through the mixer core are free of attenuation.

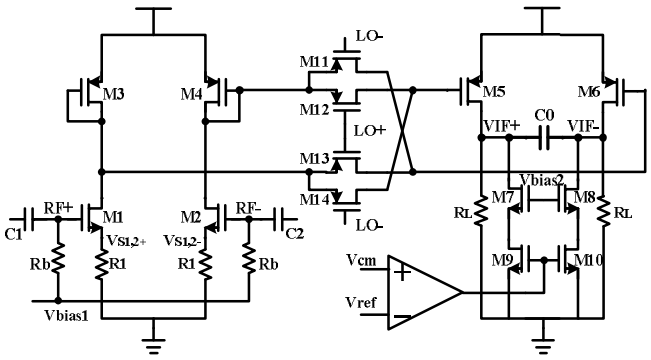


Fig. 1. Circuit diagram of the previously proposed mixer based on current mirrors.

The linearity performance of the current mirror mixer will be analyzed as follows.

For most conventional mixers, the linearity is mainly decided by the gm stage. Considering this, source degraded resistors R1 are added to improve the linearity of gm. However, the conventional mixers' linearity will degrade more or less when the RF current flowing through the switching stage or load stage.

In the current mirror mixer, the linearity is almost completely decided by the gm stage. The power spectrum at the source of M1, 2 and the power spectrum at the output node under two tone test condition are illustrated in Fig.2. The IIP3 of the common drain amplifier constructed by M1, 2 and R1 can be calculated by using the voltage gain A_1, the fundamental frequency element P1 and IM3 element. The $IIP3_{S1,2}$ and A_1 are presented in (1) (2). All the parameters mentioned above are in terms of dB.

$$IIP3_{s1,2} = \frac{3P1 - IM3}{2} - A_1 \quad (1)$$

$$A_1 = \frac{V_{s1,2}}{V_{RF}} \quad (2)$$

As shown above, the current values of M5, 6 are mirrored from the gm stage after being modulated by a sgn function. The signal spectrum of the fundamental elements and IM3 elements in the gm stage is down-converted to the IF band. For the current mirror structure, the currents flowing through M5, 6 keep the same current with that of M3, 4 and M1, 2. The voltage gain A2 from $V_{S1,2}$ to V_{IF} is thus the ratio of the load resistor R_L to the source degenerated resistor R1 (4). As the current mirror can copy the signal with little decrease, each spectrum element of the output power spectrum at the drains of M5, 6 is increased by the same amount of A_2 consistently. The final expression of the overall IIP3 is presented in (3). Clearly, it reveals that the IIP3 of the entire mixer and the IIP3 at the node of $V_{S1,2}$ are equal.

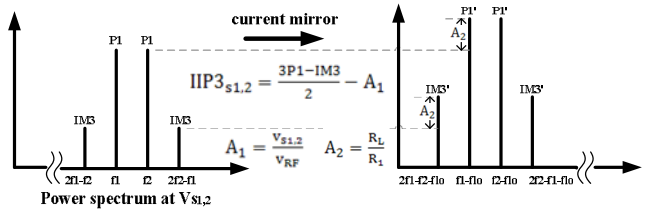


Fig. 2. The current mirror mixer's non-linearity characteristic under two-tone test.

$$IIP3_{all} = \frac{3(P1 + A2) - (IM3 + A_2)}{2} - A_2 - A_1 = IIP3_{s1,2} \quad (3)$$

$$A_2 = \frac{R_L}{R_1} \quad (4)$$

With the help of the source degenerated resistor R1, IIP3 on the gm stage can easily reach a relatively high value, which ensures the entire mixer can obtain a considerable IIP3.

Fig.3 illustrates a newly proposed current reusing mixer based on the circuit of current mirror mixer. The gm stage uses the cascode structure to convert the input RF voltage into current. The current will be multiplied by $1/g_{m3,4}$, which is the equivalent impedance of the diode-connected M3, M4. The cascode structure can shield the input devices from voltage variations at the output, so in this circuit it means the source voltages of M9, 10 will not be affected by the drain voltages of M3, 4. This can be helpful to use the static current reusing technique analyzed later. The double balanced switching pair conducts and commutates M3, 4's gate voltages to M5, 6 according to LO frequency. In the circuit, many factors especially the size of the transistors should be carefully designed. The channel length and width of M5, 6 are set to the same as M3, 4 for the consideration of current mirror. To obtain a moderate voltage gain at the gate of the current mirror to suppress the noise, the small size of M3-6 is helpful. However, considered the V_{GS} under low supply voltage and the output flicker noise from M5, 6, the size of M3-6 cannot be designed too small. In this design, a moderate size is chosen to guarantee a 5-6 dB gain at the gate of current mirror. Meanwhile, the sizes of the switch transistors M11-14 are set to relatively small values to keep the parasitic capacitors small. Hence the current mirror can duplicate the differential gate voltages from the input stage to the output stage more precisely by small parasitic capacitance. Through polarity commutating of the mirrored currents according to LO signal, the RF currents at the gm stage are modulated by a $\mathrm{sgn}[\cos(\omega_{LO})]$ function which is produced by the switch pairs, thus the RF currents can be converted from RF band to IF band. With a pair of current source connected in parallel, two 900Ω load resistors are applied to enhance the conversion gain. At the output stage, a common mode feedback circuit is adopted to stabilize low frequency CM voltages. Using this structure can realize the high linearity.

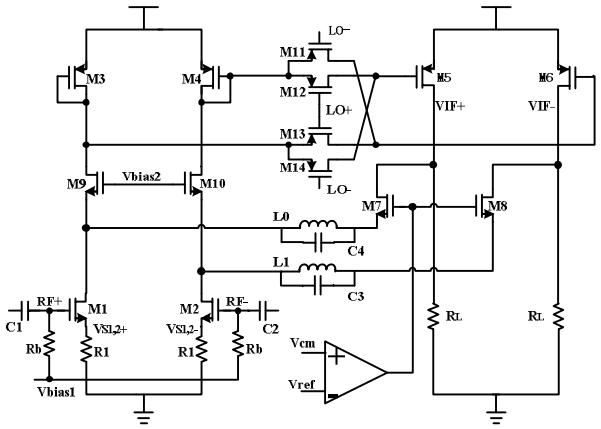


Fig. 3. Circuit diagram of the newly proposed mixer with current reusing technique.

The common mode feedback circuit is much different from the one described in Fig.1. In Fig.3, the current of the common feedback circuit reflowed to the gm stage through M7, 8 and two LC networks. The values of the capacitance and the inductance are specially selected to assure that both the L0, C4 network and L1, C3 network resonate in the value of RF high frequency, so the high frequency from the gm stage cannot pass through to the load stage. The LC networks eliminate the affect from input to output. Meanwhile, the IF frequency at the output terminal can also be resisted to occur in the gm stage, because the impedance seen at the drain of M7 or M8 is very high. Since the analysis above, only DC current can pass through M7, 8 and the LC networks. Here the current reusing technique is used to share the bias current of the gm stage and the output stage. However, the deterioration of noise performance using this current reusing technique cannot be neglected. Considering the low frequency noise of the gm stage, especially at the source of M9, 10, it has two paths of propagation, transistors M9, 10 or LC networks. While if the transconductance of M9, 10 is much larger than that of M7, 8, the noise will almost completely be conducted to M9, 10, which will improve the noise performance notably. The reason is that the low frequency noise current flowing through M9, 10 will be modulated to high frequency by the switching core and then has little impact to the load stage, while the noise current flowing through M7, 8 will add to the output noise directly. So in this circuit, the size of M7, 8, 9, 10 should be designed appropriately to make sure that $g_{m9,10}$ is much larger than $g_{m7,8}$.

This new structure with current reusing technique has the advantage of circuit in Fig.1, the high linearity, moderate conversion gain and noise figure. Besides, the outstanding quality of this circuit is that it has very low power consumption because of the feedback paths formed by LC networks and M7, 8. The whole mixer has a power dissipation of 3.84mW under 1.2V supply voltage.

III. SIMULATION RESULTS

The proposed mixer is designed in 0.13-μm 1P8M RFCMOS process with 1.2V supply voltage. During the simulation, balun is used to convert the single-ended input RF and LO signal to differential signal. A 50-Ω resistor is connected in parallel with the input terminal of RF and LO port to provide input matching. The DC bias Current is 3.2mA excluding the current dissipated by LO buffer circuit.

Fig.4 indicates the simulation result of the conversion gain versus LO power. From the figure, when the LO signal's power ranges from 2dBm to 8dBm, the simulated CG at 1.002GHz maintains at the value 12.6dB. Then CG versus the input LO power is shown.

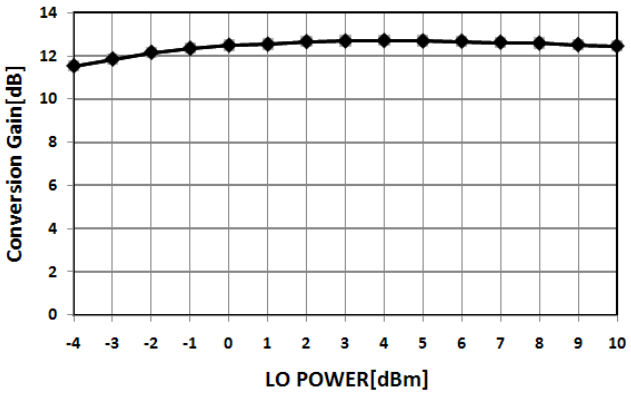


Fig. 4. CG versus LO power at the LO frequency of 1 GHz.

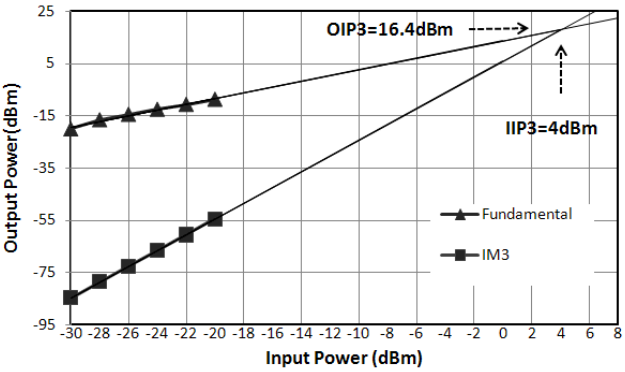


Fig. 5. Simulated IIP3 and OIP3 of the proposed mixer at the LO frequency of 1 GHz.

In Fig.6, the simulated double-sideband (DSB) noise figure of the current-mirror mixer is shown. The DSB NF is simulated at the output frequency from 1MHz to 10MHz. At the IF frequency of 2 MHz, the DSB NF value is 16dB. Thus after optimizing the size of the transistors, the DSB NF can achieve a moderate value. Its impact on the receiver's sensitivity could be mitigated by a preceding low noise amplifier with 20- to 25-dB gain.

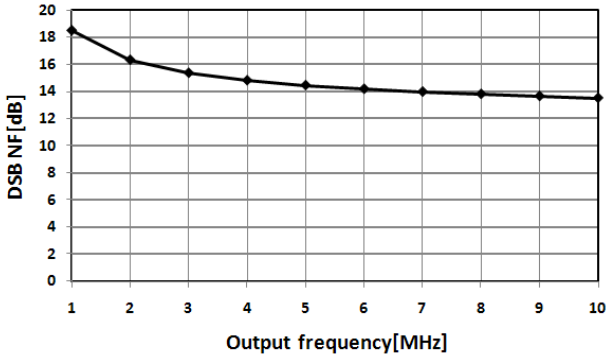


Fig. 6. Simulated NF of the proposed mixer.

IV. CONCLUSION

A new double balanced mixer based on current mirror structure is proposed in this paper. For the linear duplication characteristic of a current mirror, high linearity can be easily guaranteed. A moderate conversion gain is also achieved by using suitable load resistors with current reusing branch in the IF stage. The folded structure makes it suitable for low supply voltage. Furthermore, due to the current reusing technique, the circuit power consumption is very low. This mixer is designed by using the 0.13-µm 1P8M RFCMOS process, operating in the LO frequency of 1GHz. The DSB NF of the proposed mixer can be a moderate value of 16 dB. A conversion gain of 12.6dB and an IIP3 of 4dBm suggest the mixer's potential applications in linearity stringent communication systems such as OFDM and CDMA receivers.

REFERENCES

[1] Tae Wook Kim, Bonkee Kim, Kwyro Lee, "Highly linear receiver front-end adopting MOSFET transconductance linearization by multiple gated transistors," Solid-State Circuits, IEEE Journal of , vol. 39, no. 1, pp. 223- 229, Jan 2004.

[2] Kung-Hao Liang, Chi-Hsein Lin, Hong-Yeh Chang, Yi-Jen Chan, "A New Linearization Technique for CMOS RF Mixer Using Third-Order Transconductance Cancellation," Microwave and Wireless Components Letters, IEEE , vol. 18, no. 5, pp. 350-352, May 2008.

[3] Youngwook Kim, Youngsik Kim, Soonghak Lee, "Linearized mixer using predistortion technique," Microwave and Wireless Components Letters, IEEE,vol. 12,no. 6, pp. 204-205,June 2002 doi: 10.1109 / LMWC. 2002. 1009995.

[4] Long Xing Shi, Chao Chen, Jian Hui Wu, Meng Zhang, "A 1.5-V Current Mirror Double-Balanced Mixer With 10-dBm IIP3 and 9.5-dB Conversion Gain," Circuits and Systems II: Express Briefs, IEEE Transactions on , vol. 59, no. 4, pp. 204-208, April 2012.

[5] V. Vidojkovic, J. van der Tang, A. Leeuwenburgh, A. H. M. van Roermund, "A low-voltage folded-switching mixer in 0.18-µm CMOS," Solid-State Circuits, IEEE Journal of , vol. 40, no. 6, pp. 1259- 1264, June 2005.

[6] S.K.Hampel, O. Schmitz,M. Tiebout, I. Rolfes, "Inductorless low-voltage and low-power wideband mixer for multistandard receivers," Microwave Theory and Techniques, IEEE Transactions on , vol. 58, no. 5, pp. 1384-1390, May 2010.

[7] G. Theodoratos, Y. Papananos, G. Vitzilaios, "A low-voltage 5-GHz downconversion mixer employing a second harmonic injection linearization technique," Circuits and Systems II: Express Briefs, IEEE Transactions on , vol. 54, no. 11, pp. 964-968, Nov 2007.

[8] E. Bruun, "Analytical expressions for harmonic distortion at low frequenciesdue to device mismatch in CMOS current mirrors," IEEE Trans.Circuits Syst. II, Analog Digit. Signal Process., vol. 46, no. 7, pp. 937–941,Jul 1999.

[9] S. Zhou, "A CMOS passive mixer with low flicker noise for low-powerdirect-conversion receiver," *IEEE J. Solid-State Circuits*, vol. 40, no. 5, pp. 1084–1093, May 2005.

[10] DaeHoon Na, Tae Wook Kim, "A 1.2 V, 0.87–3.7 GHz Wideband Low-Noise Mixer Using a Current Mirror for Multiband Application," Microwave and Wireless Components Letters, IEEE , vol. 22, no. 2, pp. 91-93, Feb 2012

Sustaining the Lower Bound of QoS in MIMO Ad-Hoc Network with Concurrent Virtual Streams

Viktor V. Zaharov
Electrical and Computer Engineering Dept.
Polytechnic University of Puerto Rico
San Juan, PR 00918, USA
Email: vzaharov@pupr.edu

Olexander B. Kokhanov
Electrical Engineering Dept.
Odessa Polytechnic State University
Odessa, Ukraine
Email: okokhanov@gmail.com

Angel Lambertt Lobaina
Electrical Engineering Dept.
Universidad Anahuac Norte
Huixquilucan, Edo.de Mexico
Email: alambertt@anahuac.mx

***Abstract*—The ad hoc wireless network that operates in the rich multipath propagation environment is considered. It is supposed that the node to node communication links are performed with multi-element antennae and MIMO spatial multiplexing transmitting-receiving strategy in use. The signal to interference plus noise ratio (SINR) in the presence of interpath interference is analyzed. The cooperation among transmitted nodes is used to avoid the stream collision through the concurrent paths when the performance criteria is the quality of service (QoS) of peer to peer links. The revised water pouring algorithm (RWPA) that helps to redistribute the power among the transmitted antennae according to QoS performance criteria is discussed. The simulation part analyzes scenario with two pairs of node to node communication links.**

***Keywords* — Ad hoc networks, array signal processing, channel state information, MIMO, multiuser communications, spatial multiplexing, water pouring algorithm.**

I. Introduction

Wireless ad hoc network is basically peer-to-peer network of hosts that have not fixed communication infrastructure [1]. Extending MIMO concept to ad hoc network applications by using the medium access protocols such as IEEE 802.11n helps to achieve much higher spectral efficiencies than that of the traditional single-input-single-output (SISO) omnidirectional antennas nodes networks providing higher data throughput and improving system performance [2], [3].

A serious challenge is how to efficiently allocate the transmission powers among multiple antenna elements for each peer to peer link maximizing network throughput in the presence of mutual interfering paths (interpath interference). A lot of efforts in the literature have been undertaken to solve this problem. For example, multiuser water-pouring algorithm (WPA) is described in [4], where the interference path at the receiver can be suppressed with "whitening of the channel matrix" by matrix inversion operation. The noncooperative maximization of mutual information in the vector Gaussian interference channel via game theory is considered in [5]. Among other, some suitable solutions have been also discussed in [1], [2], [6]. The rigorous solution, as a rule, accompanies the high-priced computational procedure. However, sometimes less tense criteria than overall throughput are more essential for practical applications, for example, when the QoS of peer to peer links plays the dominant role.

In this paper we consider the performance of node to node communication link in the presence of interpath interference for spatial multiplexing MIMO physical layer when multiple independent data streams are transmitted using channel state information on the transmitter part (CSIT). We propose and analyze inexpensive approach that helps to avoid the stream collision through the concurrent paths and in the same time to maintain the acceptable QoS for each peer to peer communication pare. The revised water pouring algorithm (RWPA) that helps to succeed our criteria is considered.

II. SINR in the presence of interpath interference

A typical scenario in MIMO ad hoc network is where the numerous nodes communicate simultaneously by using the same spectrum band and exploiting the propagation channel diversity. Let us consider an ad hoc network with simultaneously communicating transmitter-receiver node pairs, as Fig. 1 shows. Each node is equipped with M_t transmit antennas and M_r receive antennas. The transmitted signal $\mathbf{s} \in \mathbb{C}^{M_t \times 1}$ is column vector with i.i.d. standard Gaussian entries $s_i,\ i = 1, 2, \ldots, M_t$, and its covariance matrix is $E\{\mathbf{s}\mathbf{s}^{\dagger}\} = (P/M_t)\boldsymbol{\Upsilon}$, where $\boldsymbol{\Upsilon} = diag\{\gamma_1, \gamma_2, \ldots, \gamma_{M_t}\}$, is a power allocation matrix, P is a total power allocated to a transmitted node, and $\sum_{k=1}^{K} \gamma_i = M_t$.

All nodes communicate in the Rayleigh fading propagation channel with a rich scattering environment, and each transceiver-receiver pair attempts to avoid the interference through the use of multiple receiver antennas.

We introduce the channel matrix for the arbitrary transmitter-receiver pair as

$$\mathbf{H} = \mathbf{R_r}^{1/2}\mathbf{H}_{\mathcal{N}}\mathbf{R}_t^{1/2}, \tag{1}$$

where $\mathbf{H}_{\mathcal{N}} \in \mathbb{C}^{M_r \times M_t}$ is the i.i.d. complex values $\mathcal{CN}(0, 1)$, which are the collection of all channel propagation coefficients, $\mathbf{R}_r$ and $\mathbf{R}_t$ are receiver and transmitter correlation matrices, respectively [7]. Assuming that $\mathbf{R}_r$ is the identity matrix, we consider the case with correlation at the transmitter side only. For the uniform linear array, the correlation coefficient r_{ij} between ith and jth transmitting antennas is [3]

$$r_{ij} = J_0\left[2\pi(i-j)d/\lambda\right], \tag{2}$$

where $J_0(x)$ is the zero order Bessel function of the first kind, and d/λ is an inter-element distance to the carrier wavelength ratio.

Furthermore, we ignore the large scale propagation attenuation of the received signal, assuming that $\sum_{j=1}^{M_t} E\{|h_{ij}|^2\} = M_t, i = 1, 2, \ldots, M_r$, where $E\{\cdot\}$ is the expectation operator, and h_{ij} are the elements of the matrix $\mathbf{H}$. This implies that each of the receiver antenna receives a power, which is equal to the total transmitted power P.

Let us consider the communication scenario with desired link, where one of the node transmits the data (desired transmitted node) and some other node receives the data (desired receiving node), and *K* interferer links around. The desired receiving node receives the signal of interest $\mathbf{s}^{(d)}$ as well as K interferer signals $\mathbf{s}^{(k)},\ k = 1, 2, \ldots, K$.

The set of the channel matrices related to the desired receiving node is $\{\mathbf{H}^{(d)}, \mathbf{H}^{(1)}, \mathbf{H}^{(2)}, \ldots, \mathbf{H}^{(k)}, \ldots, \mathbf{H}^{(K)}\}$. Let the singular value decompositions (SVD) of the matrices $\mathbf{H}^{(d)}$ and $\mathbf{H}^{(k)}$ are

$$\mathbf{H}^{(d)} = \mathbf{U}^{(d)}\mathbf{\Sigma}^{(d)}\mathbf{V}^{(d)\dagger};\ \ \mathbf{H}^{(k)} = \mathbf{U}^{(k)}\mathbf{\Sigma}^{(k)}\mathbf{V}^{(k)\dagger}, \qquad (3)$$

where $\mathbf{U}^{(d)}, \mathbf{U}^{(k)} \in \mathbb{C}^{M_r \times M_r}$, $\mathbf{V}^{(d)}, \mathbf{V}^{(k)} \in \mathbb{C}^{M_t \times M_t}$, and $\mathbf{\Sigma}^{(d)}$, $\mathbf{\Sigma}^{(k)}$ are corresponding receiver and transmitter precoding matrices with the orthonormal properties, and the singular values matrices.

The desired receiving node output vector in the presence of K interferer signals is

$$\begin{aligned}\mathbf{y} =& \sqrt{P^{(d)}}\mathbf{U}^{(d)\dagger}\mathbf{H}^{(d)}\mathbf{V}^{(d)}\mathbf{s}^{(d)} \\ &+ \sum_{k=1}^{K}\sqrt{P^{(k)}}\mathbf{U}^{(d)\dagger}\mathbf{H}^{(k)}\mathbf{V}^{(k)}\mathbf{s}^{(k)\dagger} + \mathbf{U}^{(d)\dagger}\mathbf{n}.\end{aligned} \qquad (4)$$

Substituting (3) into (4) and afterwards using the orthonormal properties of precoding matrices we rewrite (4) as

$$\begin{aligned}\mathbf{y} =& \sqrt{P^{(d)}}\mathbf{\Sigma}^{(d)}\mathbf{s}^{(d)} \\ &+ \sum_{k=1}^{K}\sqrt{P^{(k)}}\mathbf{U}^{(d)\dagger}\mathbf{U}^{(k)}\mathbf{\Sigma}^{(k)}\mathbf{s}^{(k)} + \mathbf{U}^{(d)\dagger}\mathbf{n},\end{aligned} \qquad (5)$$

where the first term is the desired signal and the second term is an interfering signal in the output, and $\mathbf{n}$ the AWGN vector with the variance σ_n^2. Taking into account that $E\left\{\mathbf{s}^{(d)}\mathbf{s}^{(k)\dagger}\right\} = E\left\{\mathbf{s}^{(m)}\mathbf{s}^{(k)\dagger}\right\}_{m\neq k} = 0$, the overall average SINR in the output of the MIMO receiver in the spatial multiplexing mode regarding to each partial virtual subchannel yields

$$SINR_i = \frac{P^{(d)}\gamma_i^{(d)}\lambda_i^{(d)}}{\sum_{k=1}^{K} P^{(k)}\mathbf{u}_i^{(d)\dagger}\mathbf{U}^{(k)}\mathbf{\Lambda}^{(k)}\mathbf{\Upsilon}^{(k)}\mathbf{U}^{(k)\dagger}\mathbf{u}_i^{(d)} + \sigma_n^2}. \qquad (6)$$

where $\mathbf{u}_i^{(d)}$ and $\lambda_i^{(d)}$ are *i*th eigenvector and *i*th eigenvalue of the matrix $\mathbf{H}^{(d)}$, respectively.

As (6) shows, the level of interfering signal in the output of the *i*th virtual subchannel is completely determined by the product $\mathbf{u}_i^{(d)\dagger}\mathbf{U}^{(k)}$. When $\mathbf{u}_i^{(d)\dagger}\mathbf{U}^{(k)} = \mathbf{0}$, the vector $\mathbf{u}_i^{(d)}$ is orthonormal to all columns of the matrix $\mathbf{U}^{(k)}$, meaning no stream collision, i.e., the non-interfering scenario is running.

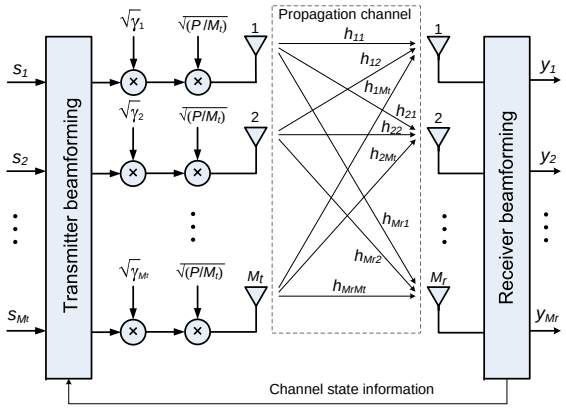


Fig. 1. Transmitter-receiver node pair in MIMO ad hoc network

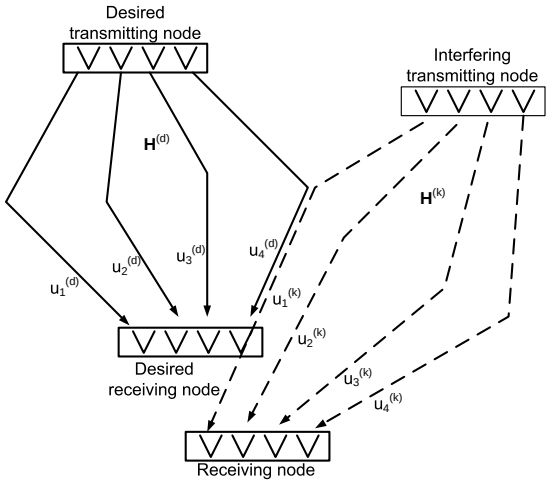


Fig. 2. MIMO communication scenario with collision of two streams

However, $\mathbf{u}_i^{(d)\dagger}\mathbf{U}^{(k)} \neq \mathbf{0}$ indicates that some paths of desired link concur with some paths of interferer link. For example consider scenario shown in the Fig. 2, where $\mathbf{u}_i^{(d)\dagger}\mathbf{U}^{(k)} = \mathbf{0}$, $i = 1, 2, 3$, and $\mathbf{u}_4^{(d)\dagger}\mathbf{U}^{(k)} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$, meaning that fourth path of desired link is concurs with the first path of interferer link, and corresponding streams collision occurs.

Hence, when the number of collision is L, the expression of average SINR for *i*th virtual channel (6) can be rewritten as

$$SINR_i = \frac{P^{(d)}\gamma_i^{(d)}\lambda_i^{(d)}}{\sum_{j=1}^{L} P^{(k)}\gamma_j^{(k)}\lambda_j^{(k)} + \sigma_n^2}, \qquad (7)$$

where *L* is a number of nonorthogonal paths between $\mathbf{u}_i^{(d)}$ and $\mathbf{U}^{(k)}$, or number of stream collisions.

As follows from (6) and (7), the more jointly nonorthogonal virtual channels the higher the level of interference and as a consequence, the lower ratio of SINR in the output of *i*th virtual channel. The total impact of *k*th interfering link into overall average SNIR of the desired link can be estimated by forming the product $\mathbf{U}^{(d)\dagger}\mathbf{U}^{(k)}$. For example, for the scenario shown in Fig. 2 the result is 4×4 matrix with all zero elements

except one with the unit value in the crossing of first column and fourth row.

III. Stream allocation Strategies in the presence of interpath interference

Obviously that the non-interferer scenario for the desired link is when the matrix $\mathbf{U}^{(d)\dagger}\mathbf{U}^{(k)}$ has all zero elements. It could be achieved by suppression of all interferer paths on the desired receiver side using spatial-temporal processing [8], [9]. However, it is almost useless when interfering and desired streams are not spatially resolved; but even if they are spatially resolved the spatial-temporal processing requires enormous computational work diminishing the power efficiency of affected node.

We propose other solution, where the nonzero elements of the matrix $\mathbf{U}^{(d)\dagger}\mathbf{U}^{(k)}$ are forced to zero by discarding those virtual channels which causes the stream collision. This approach does not require big computational efforts (that could be rigorously found in the full version of paper), however, it requires the transmitted power redistribution in the transmitter side of the link. We discuss the virtual path discarding strategy based on the game-theoretic framework [6], where the payoff function of the game is the certain level of QoS of each player, i.e. $C_j \geq C_{min}$, where C_{min} is a link capacity that still able to maintain a certain level of QoS. The players are the peer to peer links with corresponding channel eigenvalues entries, and available transmitted power (further without loose of generality we set $P^{(d)} = P^{(k)} = 1$). Furthermore, some sort of cooperations among transmitted nodes are implicated.

Let us consider the scenario where desired transmitting and receiving nodes try to establish the communication link in the presence of already operating K interfering links. To avoid the stream collisions the following operations should be done on the desired transmitter side to satisfy the payoff function:

1. Send the pilot signal and receive back the matrices $\mathbf{U}^{(k)}$ through the reverse channels, then compute the product $\mathbf{U}^{(d)\dagger}\mathbf{U}^{(k)}$, $k = 1, 2, ..K$, and estimate which virtual paths already in use.

2. Using WPA compute the power allocation and channel capacity $C^{(d)}$ if only "clear" virtual paths will be involved to transmit the data.

3. If $C^{(d)} \geq C_{min}$, then discard the eigenvalues that cause the stream collisions and start to transmit the data.

4. If $C^{(d)} < C_{min}$, estimate which virtual sub-channel can be borrowed from some interferer links to achieve C_{min}. For this, send demand to interferer link through concurrent paths, inquiring about possibility to borrow the concurred path.

5. Interferer node receives the demand and recompute its capacity excluding the requested path. If $C^{(k)} \geq C_{min}$ the kth node send the positive confirmation through the reciprocal concurrent path to the desired node, then discard this path from the transmission, and recalculate new power allocation coefficients. If $C^{(d)} \leq C_{min}$ the request is denied.

6. If desired node obtains the required path, it starts to transmit data, if request is denied, it tries to send the request to other interferer links if someone. If all concurrent paths have denied the request the desired node is getting idle, till propagation condition would be more favorable.

When some virtual path is discarded, the power distribution strategy should be revised. It is well know that for the MIMO non-interfering scenarios to distribute the transmitted power optimally between transmitted antennae the WPA provides the optimum solution [10]. In the presence of interpath interference, and criteria QoS in use the conventional WPA should be slightly revised including the following improvements:

1. The eigenvalues of interferer virtual paths should be determined and discarded.

2. The matrix of eigenvalues $\mathbf{\Lambda}$ should be rearranged, and new rank, $r = r - L$ is redefined, where L is a number of discarded eigenvalues.

Further, it is running similarly to WPA. Revised version of WPA we call the revised WPA or RWPA. The computational work to adjust the power allocation factors using RWPA is just about $O(M_t L)$. The power redistribution among transmitted antennae either desired (d) or (k)th interferer node yields to the following link capacity

$$C = \max_{\sum_{j=1}^{r-L}\gamma_j = M_t} \sum_{j=1}^{r-L} \log\left(1 + \frac{P}{M_T \sigma_n^2}\gamma_j \lambda_j\right). \qquad (8)$$

The proposed approach is a simple way to avoid the interpath interference. Despite the total network throughput could be far from the optimum (because the virtual sub-channels with the good propagation condition could be discarded), it helps to maintain the QoS of each link above the lower limit, C_{min}.

IV. Simulation Result

We consider scenario with two transceiver pairs of nodes, all equipped with uniform linear array (ULA) antennae, $M_t = M_r = 4$, $\lambda/2$ space element distance, and the signal to noise ratio in the receiver side is 20 dB. The desired link operates in the presence of kth interfering node, as Fig. 2 shows. The channel matrices $\mathbf{H}^{(d)}$ and $\mathbf{H}^{(k)}$ have been generated and their eigenvalues matrices are computed. The result is $\mathbf{\Lambda}^{(d)} = diag\{35.5181 \;\; 3.4111 \;\; 1.0148 \;\; 0.1743\}$ and $\mathbf{\Lambda}^{(k)} = diag\{32.1907 \;\; 2.1954 \;\; 0.6920 \;\; 0.1597\}$ that indicate that both matrices have a full rank. The potential capacities of desired and interfering links in the non-interferer scenario according to (8) are $15.4303\ b/s/Hz$ and $14.2346\ b/s/Hz$, respectively.

We discuss scenario when interferer link already in use, but desired link still idle. The power allocation coefficients for kth interfering node are computed using the conventional WPA in the non-interferer scenario. WPA converged after the first iteration, and the power allocation diagram depicted in Fig. 3, where $n = 1, 2, 3, 4$ is a virtual channels, and the dark areas related to the positive values of γ_n.

In the first interfering scenario we suppose that desired nodes pair try to establish the link. Firstly, the transmitter transmit the pilot signal, and the stream collision occurs between fourth desired and first interferer paths as Fig. 2

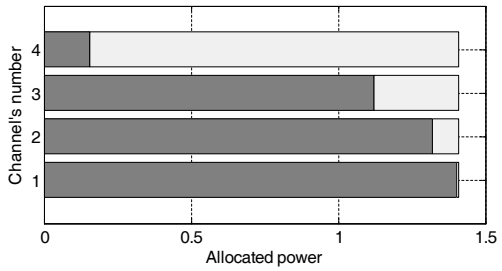


Fig. 3. MIMO link power allocation for non-interfering scenario.

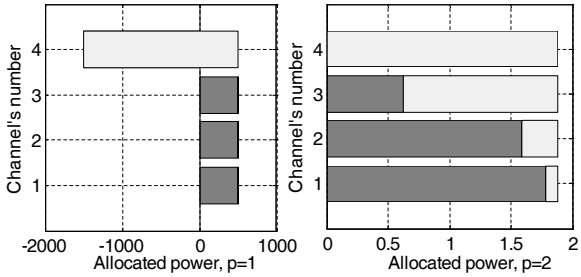


Fig. 4. MIMO interferer link power allocation with RWPA.

shows. The interferer receiving node detect the collision, and, because of the reciprocity of the channel, transmit the matrix $\mathbf{U}^{(k)}$ to the desired link. The desired transmitter find the product $\mathbf{U}^{(d)\dagger}\mathbf{U}^{(k)}$ and determines that fourth path could cause the stream collision with the first interferer path, and estimate $C^{(d)}$ assuming that fourth path is excluded from its own transmission, and the result is $C^{(d)} = 15,4267\ b/s/Hz$. If $C^{(d)} \geq C_{min}$ then the fourth path is discarded, the transmission power is rearranged among antennae with SWPA, and desired link start to transmit by using three "clear" paths.

However, if $C^{(d)} < C_{min}$, then the desired transmission node try to borrow the first path from the interferer node sending the request by using its fourth path. The interferer node receives the request and recalculate the capacity assuming that the first path would be discarded, and the result is $C^{(k)} = 7.6490\ b/s/Hz$. If request is succeeded, i.e., $C^{(k)} \geq C_{min}$ in the absence of the first path, then desired node start to transmit the data. The kth interfering transmitted node discards the first path of its link in favor of fourth path of desired link and then optimally (in terms of channel throughput) redistributes the power among the transmitted antenna elements with SWPA. It turns to the modified eigenvalue matrix $\mathbf{\Lambda}^{(k)} = diag\{2.1954\ \ 0.6920\ \ 0.1597\ \ 0.0\}$, i.e, the first eigenvalue is discarded. SWPA converged after the second iteration, as Fig. 4 shows. However, if request is denied ($C^{(k)} < C_{min}$), then desired link can not be established due to unable to satisfy the payoff criteria.

In the second interfering scenario we consider the case when the interferer node borrows the fourth path from the desired link. It can be happened when interfering link try to transmit the data when desired link already in use. Firstly, the interferer node compute the potential capacity without first interferer path, and it is $C^{(k)} = 7.6490\ b/s/Hz$, and if $C^{(k)} < C_{min}$ then it try to borrow the fourth path from the desired link. If discarding of the fourth path results $C^{(d)} \geq C_{min}$, then the request succeed. RWPA recalculate the resulting eigenvalues matrix of the desired link as $\mathbf{\Lambda}^{(d)} = diag\{35.5181\ \ 3.4111\ \ 1.0148\ \ 0.0\}$ and desired and interferer link capacities become $C^{(d)} = 15,4267\ b/s/Hz$ and $14,2346\ b/s/Hz$, respectively.

Because C_{min} is a value that very dependable on applications we just estimate the partial and joint capacities for three presented scenarios. Using (8) for both desired and interfering link the total capacity of non-interferer scenario is $C^d + C^k = 15.4303 + 14.2346 = 29.6648\ b/s/Hz$, the first scenario yields $C^d + C^k = 15.4303 + 7.6490 = 23.0792\ b/s/Hz$, and the second one give us $C^d + C^k = 15.4267 + 14.2346 = 29.6613\ b/s/Hz$. The worst case is when the dominant eigenvalue is discarded. However, the QoS of both links could be still acceptable.

V. Conclusions

We consider the stream management strategy for MIMO ad hoc network links in the presence of interpath interference where cooperative transmission strategy among nodes is used and the performance criteria is QoS of peer to peer links. We analyzed the SINR for the each virtual channel of the receiving node. It leads to the receiver precoding matrices $\mathbf{U}^{(d)\dagger}\mathbf{U}^{(k)}$ product that provides information about the concurrent virtual transmission data streams. The inexpensive way to avoid the interpath interference when QoS criteria in use is proposed. The node transmitting power is adjusted by revised water pouring algorithm (RWPA). Proposed approach could simplify significantly the ad-hoc nodes complexity comparing with the case when the whole network capacity should be optimized, in the same time maintaining the accepted QoS of each peer to peer link.

References

[1] B. Tavli and W. Heinzelman, *Mobile Ad Hoc Networks Energy-Efficient Real-Time Data Communications*. Springer, 2006.

[2] B. Chen and M. J. Gans, "MIMO communications in ad hoc networks," *IEEE Trans. Signal Proc.*, vol. 54, no. 7, pp. 2773–2783, July 2007.

[3] G. Tsoulos, Ed., *MIMO System Technology for Wireless Communications*. Taylor & Francis Group, 2006.

[4] W. Yu, W. Rhee, S. Boyd, and J. Cioffi, "Iterative water-filling for gaussian vector multiple access channels," *IEEE Trans. on Information Theory*, vol. 50, no. 1, pp. 145–151, Jan 2004.

[5] G. Scutari, D. P. Palomar, and S. Barbarossa, "The MIMO iterative waterfilling algorithm." *IEEE Transactions on Signal Processing*, pp. 1917–1935, 2009.

[6] Z. Ho, M. Kaynia, and D. Gesbert, "Distributed power control and beamforming on MIMO interference channels," in *EW2010,16th European Wireless Conference, April 12-15, 2010, Lucca, Italy*.

[7] M. Jankiraman, *Space-Time Codes and MIMO Systems*. Artech House, 2004.

[8] J. C. Mundarath, P. Ramanathan, and B. D. V. Veen, "A cross layer scheme for adaptive antenna array based wireless ad hoc networks in multipath environments," *Wireless News*, vol. 13, pp. 597–615, 2007.

[9] V. Zaharov and H. Kettani, "MIMO ad hoc network performance in the presence of co-channel interference," in *the 2nd Intern. Conf. on Computer Research and Development (ICCRD'10)*, May 2010.

[10] T. Cover and J. Thomas, *Elements of Information Theory*. New York: John Wiley and Sons, 1992.

Peculiarities of Multichip Micro Module Frameless Design with Ball Contacts on the Flexible Board

D.V. Vertyanov, K.S. Tikhonov, S.P. Timoshenkov, V.S. Petrov
Department of Microelectronics
National Research University of Electronic Technology
Moscow, Zelenograd, Russia,
evolutionden@mail.ru, tihonov-ks@mail.ru, spt@miee.ru, petrovvasiliy.s@yandex.ru

G.A. Blinov
Scientific-research Institute Microdevices&Technology
Moscow, Zelenograd, Russia

***Abstract* — New design and technological solution of three-dimensional multi-chip modules has been considered. Installation of bare chips on a flexible board has been done by printed circuit board assembly and external module contacts have performed in the matrix ball pins form. Test program was conducted in order to quantify the limits of the process. Application of flip-chip process to a wide range of customers particularly for 3D packaging has been discussed.**

Keywords— multichip module, flexible board, flip-chip, golden ball pins; three-dimensional assembly.

INTRODUCTION

Creating advanced spacecraft for Earth satellite has been required solving a number of structural and technological aspects that can enable dramatically reduce the module size, weight and power and also on-board equipment, as well as to increase its reliability and active lifetime.

ASIC SRAM (static random access memory) large memory crystals, being established even by surface mounting in standard ceramic-metal or lighter plastic cases have limited the prospects of application because of the sharp constructive volume increasing and parasitic couplings too. High-capacity storage systems (over 2 or 3 GB) built on the hull of VLSI RAM and installed on a standard printed circuit board have been inferior in mass and size and characteristics even the hard disk drive, especially in the speed of information exchange.

Thus, becoming urgent problem is to develop a promising compact design and manufacturing technology of SRAM base module with improved memory management and high-speed radio links for modern spacecraft with the approach speed of the module to the signal speed limit in the crystal.

EXPERIMENTAL PART

Unlike traditional design embodiment of peripheral outputs location [1], when the crystal was mounted on a board with a flexible polyimide carrier, the authors proposed a new design and technology option. It consists of that after the formation of the 100 μm gold balls diameter on the findings it is necessary to mount crystals directly to the board, which increases the speed of the module.

Instead of outputs which were located on the periphery of the module, the authors proposed the matrix pins location. Such design has several advantages and the main ones are: (1) the ability to accommodate more pins in a limited area; (2) the best conditions to ensure the functional performance of the electronic modules with less parasitic effects on high-speed operations; (3) simplification of surface mount technology for the use of group processes reflow [2].

A characteristic feature of the crystal ball pins is that all pins were formed within the crystal on its front side, and installation has been performed face down by the PCBA or «flip-chip» installation. The main mounting disadvantages of the Printed Circuit Board Assembly (PCBA) is the time consuming quality control elements of pin connections with pads board and also rather rigid crystal connection to the board. Under certain conditions these can influence the reliability of the connection due to differences in thermal linear expansion factor of the crystal and the board. The use of a flexible polyimide board and fill with epoxy space between the chip and the board allows you to align crystal and substrate deformation, and thus significantly improve assembly reliability [3].

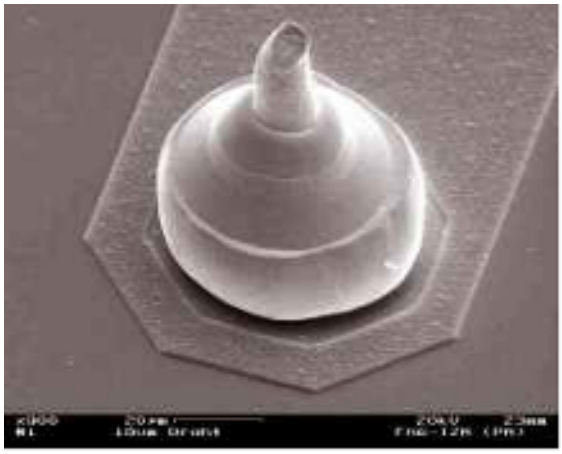
Golden ball pins were formed by gas flame fusion of the burner top gold wire. Molten gold sphere was created be the surface tension. Once the ball was attached to the pad, the wire came off at the base of the ball, leaving unsuitable surface for mounting (Fig. 1 a). The process, known as «coining» (or chasing) should be used to create a flat surface and alignment pins for height, as shown in Fig. 1(b) [1], [4].

As we see, base was a double sided flexible board in a forthcoming version of design and technology of three-dimensional multi-crystalline memory micro-module. Polyimide Kapton was selected as a board material. It has good electrical, physical and mechanical properties compared with other materials and more suitable for high-speed three-dimensional assemblies [3].

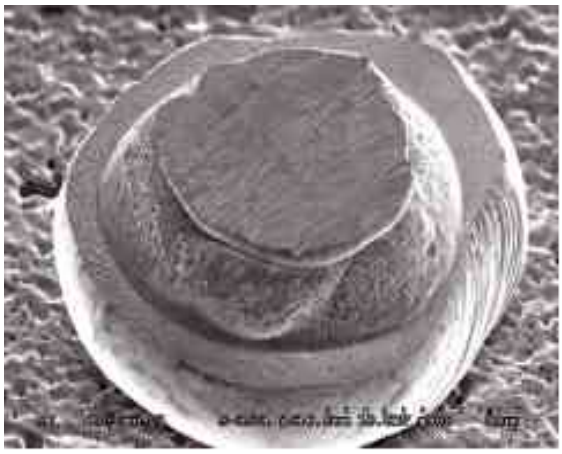
"Saddle" pads for mounting the crystal ball pins (four seats) were formed on the top side of the board and on the lower board side outer ball joints were performed in the full matrix form. The use of "saddle" pads allow for more reliable contact connection.

In our investigation the SRAM crystals of "Milandr" 1645RU4U (16 MB) recommended in the equipment of special purpose have been used. Golden Ball pins diameter 100 μm, located at the crystal periphery in the steps of 250 to 750 μm (Fig. 2) were used on the chip.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE



a



b

Fig. 1. Ball output on crystal: (a) after gold wire avulsion; (b) after «coining» process.

External module contacts were arranged in a matrix with a 1.27 mm pitch. Ball pins from eutectic alloy have been formed on them. It enables soldering at 160-170 °C with a deposition and subsequent melting Cr-Cu-SnBi layers (Fig. 3).

Simulation and a comparative analysis of two structural and technological MCM options have been performed (TABLE I.) to confirm the prospects of using FCOF-technology to create three-dimensional multichip micromodules.

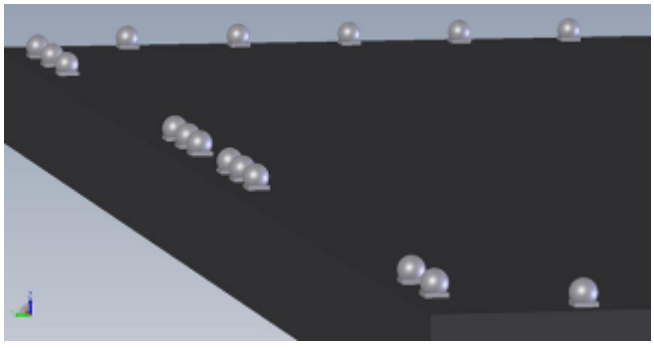


Fig. 2. Three-dimensional crystal models with pin balls.

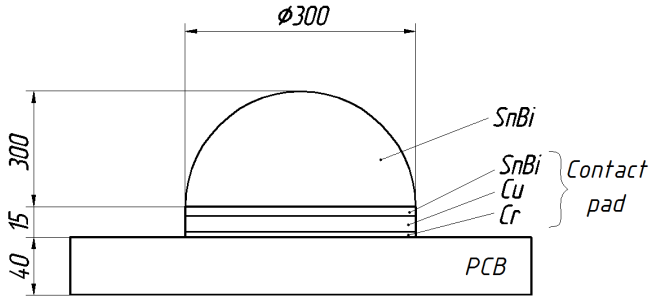
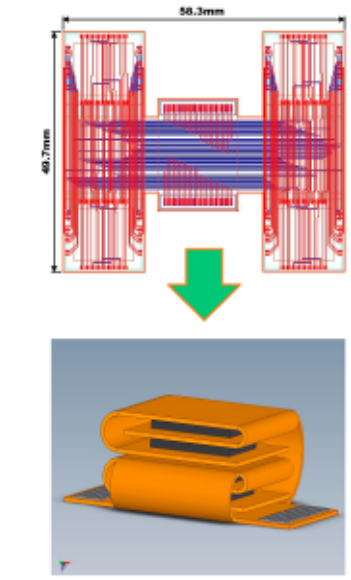


Fig. 3. Scheme of ball pins to the flexible board.

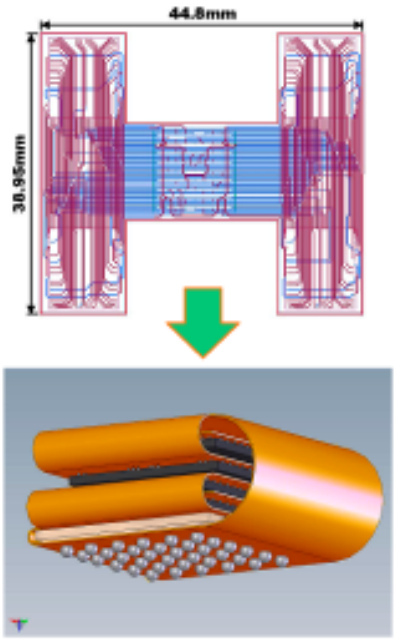
TABLE I. COMPARATIVE ANALYSIS DATA OF TWO STRUCTURAL AND TECHNOLOGICAL OPTIONS MULTICHIP MICROMODULES

Structural and technological options	**MCM-P64M-PV**	**MCM-P64M-MB**
Number of micro-transitions	192	221
Board dimensions in 2D, mm	58,3 x 49,7	44,8 x 39,1
Dimensions of mounting area, mm	23,2 x 16,5	11,5 x 11,5
Placement method of external pins	peripheral	matrix
Minimum conductors width, µm	100	100
Minimum diameter of transferring hole, µm	70	70
Relative signal quality	1	3
High-speed performance for high-frequency crystal	1	3
High-speed performance for crystals of low frequencies	1	1,1
The maximum pins number	46	81
Relative mass	1.2	1

In this table MCM-P64M-PV is the multi-chip memory module of 64 MB on a flexible carrier body with tape pins and peripheral external pins distribution (Fig. 4a); and MCM-P64M-MB is the multichip memory module of 64 MB based on crystals with ball pins and matrix distribution of the external pins (Fig. 4b).



a



b

Fig. 4. Flexible Circuits for crystals mounting in the MCM: a) MCM-P64M-PV, b) MCM-P64M-MB.

Usage of integrated development in software company Mentor Graphics (Route Expedition Enterprise) and three-dimensional CAD software SolidWorks allowed us to realize three-dimensional model of highly integrated multi-chip micromodules Designed flexible board with two-dimensional version of the constructive data board such as dimensions, bend radius has been assembled in a three-dimensional structure in SolidWorks.

The proposed design and technological solutions will reduce the volume requirements, improve the quality of the electrical signals (the integrity of signals, noise immunity, reduced parasitic effects), and high-speed modules performance. Also, the use of FSOF technology makes it easier to automate installation of crystals on the board and reduce the cost of the modules. But it is worth noting that the effectiveness of the design and technological solutions of high-speed performance will be only if the crystals will be used with high performance frequency.

In order to improve the reliability of the assembly space between the crystal and the polyimide compound with organosilicon SIEL has been filled. We think it will align the crystal deformation and flexible substrates.

Using design and research results sample prototype of MCM-P64M-MB (Fig. 5) has been performed. Its characteristics within the error satisfy similar modules parameters.

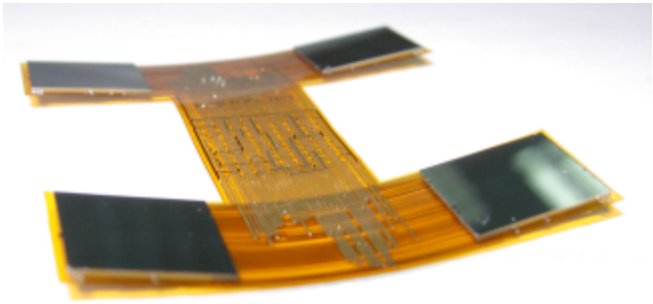


Fig. 5. Sample prototype of MCM-P64M-MB.

Conclusion

The complex development of design and technology for creation three-dimensional multichip micromodule memory for advanced space vehicles board equipment intended for the information collection, storage and distribution have been performed. Construction of three-dimensional multichip memory module has been developed. Bend radius circuit boards were designed. For scheme design and flexible circuit board topology MCM-64 M-MB device modern CAD Company Mentor Graphics was used. Profile model topcoat pad flexible circuit board for durable micro-site, and stress-strain state of micro contact nodes solder balls with the pad were created The algorithm of the assembly process and installation of three-dimensional multichip memory module based on crystals with ball pins has been developed. Flip chip mounting on flex is a logical way to provide best-in-class product density at minimum cost, at maximum production rates.

References

[1] Joseph Fjelstad. Flexible Circuit Technology. Fourth Edition. Publishing Inc. USA. 2011. – pages 613

[2] George Riley, "Doctor FlipChip" http://www.drflipchip.com/

[3] Grushevskiy A.M. Assembly and installation of multi-chip modules: Ouch. allowance. Ed. L.A. Koledova - MIEE, 2003. - pages 196.

[4] Medvedev A., Mylov G. The technology in the electronics industry. Part 3. Flexible printed circuit boards. - M: "IDT Group", 2008 . – 488 p.

Investigation of Electron-Ion Optics of Pulse Technological Glow Discharge Electron Guns

S.V. Denbnovetskiy, V.G. Melnyk, I.V. Melnyk, B.A. Tugay, S.B. Tugay
National Technical University of Ukraine "KPI", Electronic Faculty, Electron Devices Department,
Kyiv, Ukraine
e-mail: imelnik@edd.ntu-kpi.kiev.ua

Abstract— **This paper is devoted to theoretical and experimental analyze of self-maintained electron-ion optics of pulse high voltage glow discharge electron guns, which are an advanced instruments to using in modern production. Theoretical analyze is based on solving of equation of particles balance in anode plasma, as well as on defining of plasma boundary form and position by computer analyze of discharge photographs with using image recognizing technique. Experimental investigations have been provided for designed pilot model of electron gun.**

Keywords—high voltage glow discharge, triode electron gun, pulse regime, anode plasma, electron-ion optic

I. INTRODUCTION

High voltage glow discharge electron guns age widely used in modern electron technologies for welding and annealing of items parts, deposition of thing films, and for vacuum refining of refractory materials [1]. Main branches of application of this perspective guns are: microelectronics and nanoelectronics, production of precision devices, mechanical engineering and metallurgy.

The advantages of high voltage glow discharge electron guns are following [1].

1. Possibility of operation in the soft vacuum, range of 1 – 10 Pa, in the medium of different technological gases, including noble and active (hydrogen, oxygen, helium, argon etc.).

2. Using of simple evacuation equipment for maintaining glow discharge electron guns' operation.

3. Simplicity of control of electron beam power by changing operation pressure in the gun's volume with applying electronic control system. For this purpose specially elaborated electromagnetic valves have been used [2].

4. Relative simplicity of guns constructions and its maintenance in technological equipment.

5. Relative simplicity and low price of electron beam industrial equipment, based on glow discharge electron guns.

However, minimal time regulation constant for changing pressure in gun volume, which is usually lying in the range of few hundredths millisecond, is not satisfied to modern electron-beam technologies [3]. For obtaining high production rate and best quality of produced items time regulation constant for beam power must be significantly smaller, in the range of tenths or hundredths microsecond. Only in this case the cost, spent for expensive electron beam equipment, is compensated. Low value of time regulation constant for control of electron beam power can be obtained in the triode high voltage glow discharge electron guns. The distinguishing feature of this type of guns is lighting of the additional, low voltage nonself-sustained discharge. Namely this additional discharge is used for regulation the current of main, high voltage glow discharge [4]. Using of electric control of beam power is lead to future possibility for elaboration new generation of high voltage glow discharge electron guns, which operated in pulse regime [5]. This regime is very perspective for modern electron beam technologies, particularly for deposition of polymetals films and for refining of alloys in vacuum. This advanced technologies opening the new possibilities for development such branches of production, as nanoelectronic and nanometallurgy.

Therefore, future investigation of pulse operation regime for high voltage glow discharge is actual scientific and engineering problem. This article is devoted to investigation of complex electron-ion optics of pulse high voltage glow discharge electron guns.

II. MATHEMATIC SIMULATION OF ELECTORN-ION OPTIC PROPERTIES OF PULSE HIGH VOLTAGE GLOW DISCHARGE ELECTRON GUNS

A. Calculation of Plasma Boundary Position For One-Dimentional Discharge Model

Defining of plasma boundary form and position is the first very important problem for analyzing complex electron-ion optics of high voltage glow discharge. It's connected with physical properties of anode plasma in such type of discharges, where plasma boundary is usually considered as the surface with low potential and as ions source [6]. For diode high voltage glow discharge one-dimensional gap all formulas to defining the position of plasma boundary relatively to the cathode surface have been obtained in work [7]. This mathematical model is based on analyzing the charges particles balance in anode plasma with taking into account the equation of self-consistent for high-voltage discharge. The same model can be created for triode high-voltage glow discharge systems, but influence of additional low-voltage discharge to gas ionization must be taking into account. Therefore, equation for particle balance in anode plasma generally written in such form:

$$z_f + z_s + z_{dis} = z_{dif}\,, \qquad (1)$$

with z_f – efficiency of gas ionization by fast electrons of beam, z_s – efficiency of gas ionization by slow electrons, emitted from anode surface after its bombarding by fast electrons, z_{dis} – efficiency of gas ionization in low-voltage discharge, z_{dif} – decreasing of ions concentration in anode plasma as a result of its leaving by diffusion process. The equation, described the self-consistent conditions of high voltage glow discharge lighting can be written in such form [7]:

$$n_i = \frac{j_{ec}\sqrt{\frac{kT_e}{2m_e}}}{e\left(A_i U_c^{-a_i} + 1\right)}, \quad (2)$$

with n_i – ions concentration in anode plasma, j_{ec} – density of electron current at the cathode surface, T_e – temperature of slow electrons in anode plasma, U_c – acceleration voltage of high voltage glow discharge, k – Boltsman constant, e and m_e – electron's charge and mass correspondently, A_i and a_i – empirical constants. In equation (1) components for efficiency of gas ionization can be defined by following four relations [7, 8].

$$z_f = \frac{j_{ec} d_p p_{a0} A_i U_c^{-a_i}}{e} \times \times \left(1 + \eta_i\left(1 - f\left(1 - d_p p_{a0} Q_{ep0}\right)\right)\eta_U^{-a_i} k_e\right), \quad (3)$$

with d_p – anode plasma longitudinal dimension, p_{a0} – reduced pressure in the discharge region, η_i and η_U – coefficients of electrons reflection from the anode by the current and by the voltage correspondently, f – transparent coefficient for anode plasma, Q_{ep0} – cross-section of ions scattering at the residual gas molecules, k_e – coefficient of electrons' trajectories longitude.

$$z_s = \frac{d_p p_{a0} n_e}{4}\sqrt{\frac{8kT_e}{\pi m_e}}\overline{Q}_{ep0};\ \overline{Q}_{ep0} = N_0\alpha_i\left(U_i + \frac{2kT_e}{e}\right)\exp\left(-\frac{eU_i}{kT_e}\right), \quad (4)$$

with n_e – concentration of free electrons in anode plasma, U_i – potential of gas ionization, N_0 – Loschmidt constant, α_i – empirical constant for given operation gas.

$$z_{dis} = \frac{4\pi^3 e^4 n_i \mu_{i0} \gamma_a d_l U_d \ln\left(1{,}5 - \sqrt{\frac{kT_e\left(eU_d + kT_e\right)^3}{\pi p_{a0}}}\right)}{9kT_e d_p\left(eU_d + kT_e\right)}, \quad (5)$$

with U_d – voltage of additional discharge, d_l – longitudinal dimension of discharge gap, μ_{i0} – mobility of ions' in anode plasma, γ_a – secondary ion-electron emission coefficient from anode surface [8].

$$z_{dif} = \frac{\pi^2 n_i \mu_{i0} kT_e}{d_p p_{a0} e}. \quad (6)$$

Equation system (1 – 6) can be transformed into cubic equation for parameter d_p and solved analytically, corresponded analytical transforming are given in work [8]. Therefore, the distance between anode plasma boundary and cathode surface for one-dimensional discharge model simply calculated as:

$$d_{cp} = d_l - d_p. \quad (7)$$

Obtained dependences of cathode-plasma distance, calculated by equation (1 – 7), from the reduced pressure and the voltage of additional discharge, are presented in Fig. 1. Nitrogen is considered as operation gas, cathode material – aluminum, anode material – cooper. Other simulation parameters are depended from many factors and generally given in work [8].

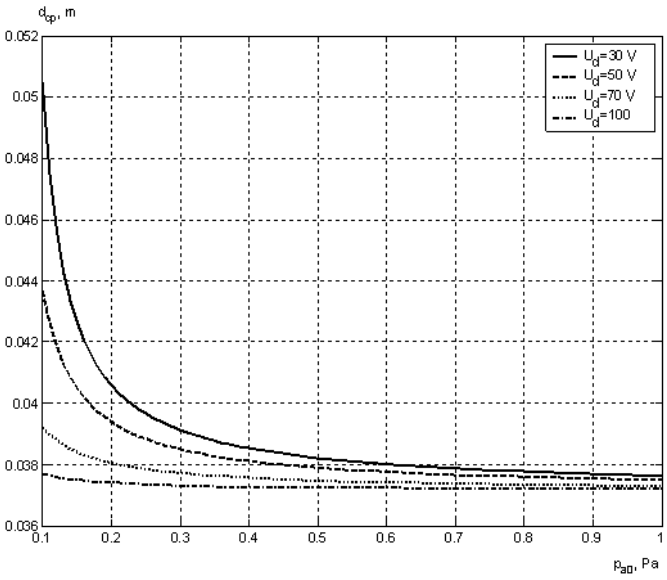


Fig. 1. Dependences of plasma boundary position from reduced pressure in discharge gap for different voltage of additional discharge. Acceleration voltage – 10 kV, d_l = 0.035 m

Obtained simulation results, which characterized dependence of the plasma boundary position from parameters of high voltage and additional discharges, is very important for future analyze of self-maintained optics of pulse glow discharge electron guns. Theoretical and experimental basis of this analyze will be considered in the next subsections of this article.

B. Approximation of Plasma Boundary Geometry

As have been pointed out in pervious subsection, for analyzing self-maintained optics of triode glow discharge electron guns defining plasma boundary form and position is necessary. In the first step of simulation equations (1 – 7) are used. But in real electrodes systems with the spherical cathode and conic anode [1], defining of complex plasma boundary geometry, which influences to electric field distribution and charged particles trajectories, is also necessary. It is clear, that for pulse guns plasma boundary position is changed for maximal and minimal values of beam current, therefore future analyze of discharge optics properties provided only for maximal value

of current. The technique of defining plasma boundary geometry is based on analyzing the brightness of high voltage discharge photographs and using functional approximation. For approximation of plasma boundary empirical function

$$z(r)=\left(Ar^2+B\right)\exp\left(-\frac{r^2}{C}\right)+D \qquad (8)$$

have been used. Here empirical coefficients A, B, C and D are defined by approximation technique with using least-squares method [9]. Plasma boundary points are defined in the region, where brightness of discharge lightings is increased. Main distinguishing feature of photograph analyze is necessity of wires shadows deleting. These wares are used in additional electrodes for limitation of plasma volume and for visualization [9]. For lighting of additional discharge cylindrical and conic electrodes with negative potential have been used [4]. Schemes of corresponded electrodes systems are presented at Fig. 2. The details of provided experiments are described in the section III.

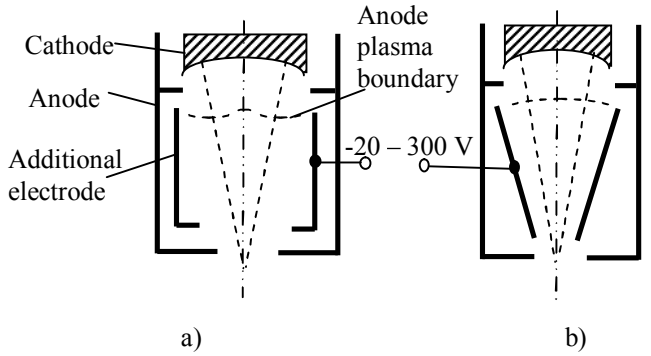


Fig. 2. Schemes of electrodes systems of pulse glow discharge electron guns with cylindrical (a) and conic (b) additional electrodes

Obtained approximation results for conic electrode are presented in Fig. 3. These simulation results were obtained for main discharge current 100 mA and accelerating voltage 10 kV, voltage of additional discharge is 100 V.

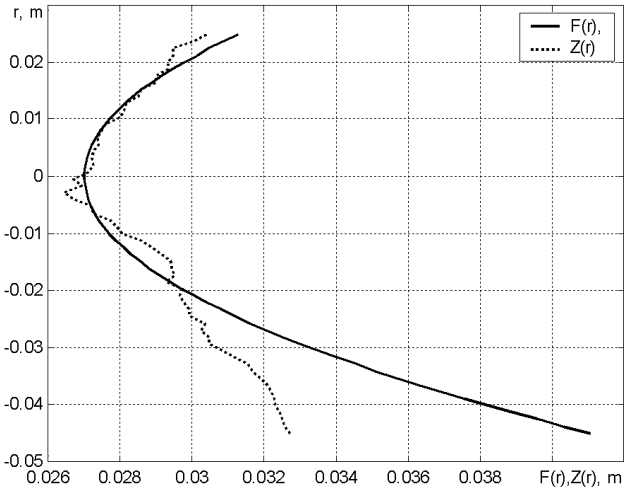


Fig. 3. Approximation of plasma boundary geometry. Z(r) – points from photograph, F(r) – approximation curve

Obtained simulation results for plasma boundary form and position is used for analyzing the electric fields distribution in discharge gap and the trajectories of charged particles.

C. *Numerical Simulation of Self-Maintained Optics of Pulse Glow Discharge Electron Guns*

Distribution of electric field in discharge gap was obtained by numerical solving of such integral equation:

$$\varphi(r,z)=\frac{\sum_{l=1}^{L}\frac{\sigma_l K(t)}{\sqrt{(r+r_l)^2+(z-z_l)^2}}}{2\pi\varepsilon_0},\quad K(t)=\int_0^{\pi/2}\frac{d\beta}{\sqrt{1-\sin^2(\beta)\cdot t^2}},$$

$$\left\|\varphi_{mi}\right\|\cdot\left\|q_i+q_m\right\|=\left\|U_m\right\|, \qquad (9)$$

with σ – density of surface charge in electrodes, r_l, z_l – radial and longitudinal distances from observing point to the point, where the potential $\varphi(r,z)$ is calculated, $K(t)$ – full elliptic integral, $q_i=h_i^2\cdot\sigma_i$ – the surface charge distribution, $q_m=U_m\cdot\rho_m$ – the volume charge distribution, V_m – volume of elementary space cell, h_i – steps for radial and longitudinal coordinates [10]. Obtained distribution of electric field in the discharge gap is presented at Fig. 4.

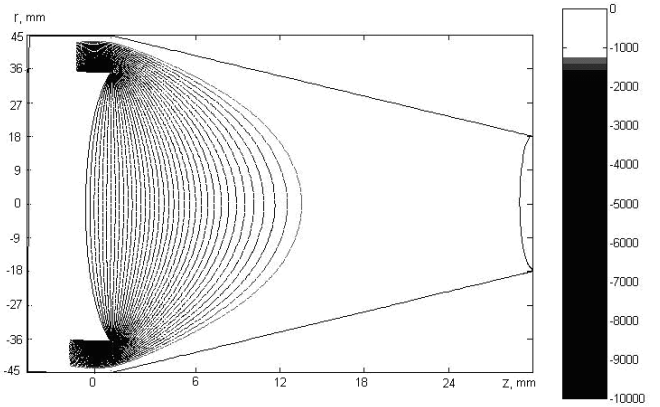


Fig. 4. Simulation of distribution of electric field in electrode system of pulse glow discharge electron gun with conic additional electrode

Ions current density from plasma boundary calculated by using Bohm equation [1, 6, 10]:

$$j_i=0{,}004en_i\sqrt{\frac{2kT_e}{m_i}}. \qquad (10)$$

With known current density, Gumel method is used for calculation electrons' and ions' trajectories in electric field with taking into account the space charge. Trajectories of charges particles with mass m_s and charge q_s calculated by known form electron optics equation [10]:

$$\frac{d^2r}{dz^2}=\frac{\left(\frac{\partial\varphi(r,z)}{\partial r}-\frac{\partial\varphi(r,z)}{\partial z}\cdot\frac{dr}{dz}\right)\left(1+\left(\frac{dr}{dz}\right)^2\right)}{\frac{m_s v_s}{2}+2\left(\varphi(r,z)-U_c\right)}, \qquad (11)$$

with $\varphi(r,z)$ – potential in the considered point. Emission of electrons from the cathode surface under ions' bombarding calculated by such formulas [6, 10]:

$$\frac{j_{ec}}{j_{ic}} = \gamma_i \left(1 + \frac{\exp \int\limits_L Q_{in}(l) \cdot dl}{\exp \int\limits_L Q_{ii}(l) \cdot dl} \right); \quad \gamma(\varphi) = C(\varphi) \cdot \varphi^{\alpha(\varphi)},$$

$$Q_{ii} = A_i \varphi(r,z)^{-a_i}; \qquad Q_{in} = A_{in} \left(\frac{E(r,z)}{E_0} \right), \tag{12}$$

with j_{ic} – ions' current density on the cathode surface, j_{ec} – electrons' current density here, γ – coefficient of electron-ion emission from the cathode, Q_{ii} and Q_{in} – cross-sections of gas ionization and recharging of ions in gas atoms respectively, $E(r,z)$ – strength of electric field, C, α, E_0, A_{in} – empirical coefficients for specified operation gas. With applying Gumel method for equation system (9 – 12) space charge in discharge gap is calculated by using equations [6, 10]:

$$\Delta N_i = 2\pi N_a (1-\xi) U_m Q_{in} \frac{\Delta r l_t h_z p_0}{\Delta z p_a r_t^2},$$

$$U_m = \left(U_{k,l} + U_{k-1,l} + U_{k,l-1} + U_{k-1,l-1} \right) / 4,$$

$$\rho_c^{i+1} = \frac{\rho_c^{i\,2}}{\rho_c^i - 2\pi \Delta n_i r_t^2 \frac{\Delta r l h_z}{\Delta z}}, \quad I_t^{i+1} = I_t^i - \frac{\rho_c^i}{\rho_c^i - 2\pi \Delta n_i r_t^2 \frac{\Delta r l h_z}{\Delta z}},$$

$$\xi = \frac{a(U_c - U_i) \exp(-b(U_c - U_i))}{p_0 \sqrt{\Delta r^2 + \Delta z^2}}, \tag{13}$$

with i – iteration number, I_t – current of considered tube, ρ_c – calculated space charge density in the considered cell, $U_{k,l}$ – potentials in the nodes of considered cell, U_m – medium potential in considered cell, N_a – Avagadro constant, ξ – specific ionization coefficient correspondent to Morgulius model, Δr and Δz – changing of radial and longitudinal coordinate for current tube in considered cell, l_t – length of current tube in considered cell, r_t – thickness of current tube in considered cell, p_a – atmospheric pressure, U_i – potential of gas ionization, a and b – empirical constant for specified operation gas.

The electron beam trajectories in anode plasma is calculated by its' extension inside the plasma volume and taking into account Rutherford scattering [6, 10]. Then destination of focus of electron beam is calculated numerically by finding minimal thickness of formed electrons' flow [6, 10]. Distribution of beam current density in the focal plane is presented at Fig. 5.

From presented in Fig. 5 dependences, which have been obtained as simulation results, it is clear, that in pulse glow discharge electron guns focal beam diameter is usually smaller, then 2 mm. Corresponded parameter to guns', operated in uninterrupted regime, is usually grater then 3 mm [6, 10]. Therefore, taking into account, that electric control of beam power provided its fast regulation [4], using of pulse glow discharge electron guns in the modern electron beam technologies is very perspective.

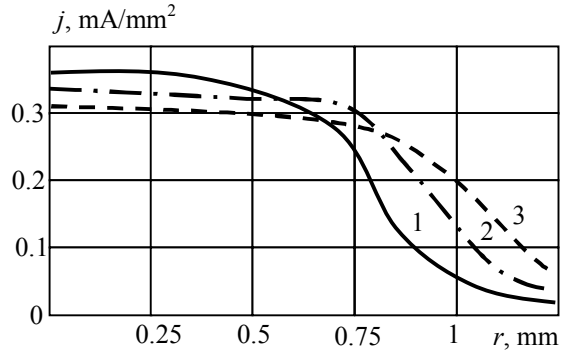


Fig. 5. Simulation results for current density distribution in the beam focus in pulse glow discharge electron guns. 1 – U_d = 50 V; 2 – U_d = 70 V; 3 – U_d = 100 V. U_c = 10 kV, p_0 = 0.5 Pa

Particularities of experimental investigations of elaborated pilot construction of the pulse high voltage glow discharge electron gun and its' electron optic parameters are described in the next section of this article.

III. Experimental Investigations of Pulse Gun and Its Parameters

Pilot pulse high voltage glow discharge electron gun has been elaborated by the authors in the Laboratory of Electron Beam Technological Equipment of National Technical University of Ukraine "KPI". Investigations of physical conditions of electron beam forming and focusing in electrodes systems with cylindrical and conic additional electrodes, presented in Fig. 2, were realized on specialized experimental electron-beam vacuum installation, in which pilot model of pulse gun, after assembling, have been installed. Experimental installation included following blocks.

1. Evacuated vacuum camber, in which elaborated gun have been installed.

2. High voltage power supply with regulated output voltage range of 10 – 25 kV.

3. Generator of impulses with maximal voltage up to 300 V and frequency up to 500 Hz.

4. Necessary control and measurement devices.

For visual investigation of discharge parameters anode and additional electrode was made transparent, for example, as composition of wares or as grid. Also for this purpose evacuated chamber have large illuminators. Such experimental equipment allows make the photographs of anode plasma, located in gun volume, and providing computer analyze of this photographs with using method of approximation, described in section I.B. Focal diameter of obtained electron beam have been measurement with using well-known calibrated diaphragms method and rotated probe method [3]. Focal beam distance is measurement visually with using the ruler, located in chamber.

Obtained dependences of focal beam distance f from electrodes geometry are presented in Fig. 6, a, R_c is the radius of cathode sphere here. Dependences of focal beam diameter d_b from the beam current in impulse I_{eb} for different geometry of additional electrode are presented in Fig. 6, b.

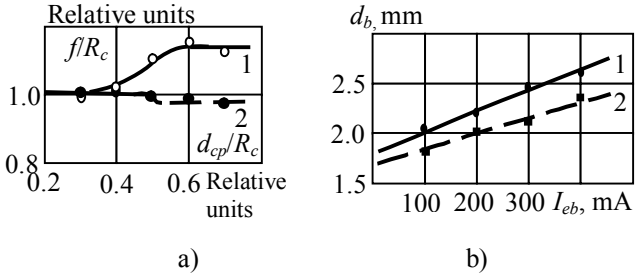


Fig. 6. Experimental dependences of beam focus position (a) and focal beam diameter (b) from beam current and electrodes geometry, 1 – for cylindrical additional electrode, 2 – for conic additional electrode. U_c = 20 kV, p_0 = 0.5 Pa, U_d = 100 V.

Change of focal distance and focal beam diameter with increasing of beam current caused by changing the geometry of anode plasma boundary and by its focusing properties. It is clear from Fig. 2, that with cylindrical additional electrode plasma boundary is convex and its influence to electrons' flux is defocusing. In contrary, for conic electrode plasma boundary is concave and it acted to electrons' flux as focusing electrode. Therefore, analysing of complex optic of pulse high voltage glow discharge electron guns is connected with consecutive using of numerical simulation and analysing of experimental results. Making of discharge photograph and defining plasma boundary form and position is necessary for complex investigation of electron-ion optics of pulse high voltage glow discharge electron guns. In proposed method position of beam focus also defined from discharge photographs and compared with theoretical results. Divergence between experimental and theoretical data was nearly 30%, and its cause by complex physical processes, usually taking place in high voltage glow discharge [6, 7, 10]. Some of these processes can not be described mathematically.

IV. Conclusion

Provided theoretical and experimental investigations shown, that pulse high voltage glow discharge electron guns can be used for forming electron beams with current density in impulse range of 0,3 A/mm^2 and greater with impulse frequency up to 500 Hz and greater. Optimization of electrodes geometry allow obtaining the beams with focal power density range of 10^5 – 10^6 W/cm^2 and total power up to tens kW. Durability of impulses can be changed from few second to tens microsecond.

Provided investigation allows optimize the electrodes geometry and regimes of operation of pulse guns for obtaining the best focal beam parameters. First estimation of plasma boundary position is obtained from simple one-dimensional model. Then, for obtaining the higher accuracy of simulation, modern computer image recognizing technique is used for analyzing the brightness of experimental discharge photographs. With known plasma boundary position Gumel method is used for numerical calculation of charges particles trajectories with taking into account its' space charge, and at the same time potential distribution in the discharge gap calculated by numerical solving of integral equation. Divergence between theoretical and experimental results was nearly 30%.

As shown provided theoretical and experimental investigations, elaboration of industrial samples of pulse high voltage glow discharge electron guns with required parameters is really possible. Such type of guns is characterized by high power density of formed electron beam and by simplicity of fast electric control of beam power. These guns can find application in industry and given new possibilities for improving the quality of produced materials and items. Especially perspective is pulse guns using in such advanced branches, as nanoelectronic and nanometalurgy. Its also can be successfully applied in production of precision devices and in mechanical engineering.

References

[1] Denbnovetskiy S.V., Melnyk V.G., Melnyk I.V. "High voltage glow discharge electron sources and possibilities of its application in industry for realising of different technological operations". – IEEE Transactions on plasma science. – Vol. 31, #5, October, 2003. – P. 987 – 993.

[2] Denbnovetskiy S.V., Melnyk V.I., Melnyk I.V., Tugay B.A. "Model of control of glow discharge electron gun current for microelectronics production applications". – Proceedings of SPIE. Sixth International Conference on "Material Science and Material Properties for Infrared Optoelectronics". – Vol. 5065 (2003). – P. 64 – 76.

[3] Shiller S., Heisig Ul., Panzer S. Elektronenstrahltechnologie. – VEB Verlang Technik, Dresden, 1976. (Publisched in German language).

[4] Denbnovetskiy S.V, Melnik V.I., Melnik I.V., Tugay B.A. "Investigation of forming of electron beam in glow discharge electron guns with additional electrode". – XVIII-th International Symposium on Discharges and Electrical Insulation in Vacuum, August 17-21, 1998, Eindhoven, The Netherlands, Proceedings, vol.2. - P. 637 – 640.

[5] Denbnovetskiy S.V., Melnyk V.G., Melnyk I.V., Tugay B.A., Tugay S.B. "Generation of electron beams in high voltage glow discharge in pulse regime". // Electronics and nanotechnology. Proceedings of the XXXII International Scientific Conference ELNANO 2012. April, 10-12, 2012. Kiev. Ukraine. – P. 40 – 41.

[6] Denbnovetskiy S.V., Felba J., Melnik V.I., Melnik I.V. "Model of Beam Formation in a Glow Discharge Electron Gun With a Cold Cathode". – Applied Surface Science, 111 (1997). - P. 288 – 294.

[7] Novikov A.A. "The high voltage glow discharge electron sources with anode plasma". – Moscow, Energoatomizdat, 1983. – 96 p. (Publisched in Russian language).

[8] Melnik I.V., Tugay S.B. "Simulation of current-voltage characteristics of nonself-sustained additional discharge in triode gas discharge electron guns". – Electronic Modelling. – Vol. 34, #5, 2012. – P. 103 – 114. (Publisched in Russian language).

[9] Melnik I.V., Tugay S.B. "Simulation of anode plasma boundary geometry in high voltage triode glow discharge electrodes systems". Electronic Modelling. – Vol. 34, #1, 2012. – P. 15 – 28. (Publisched in Russian language).

[10] Melnik I.V., Tugay S.B. "Methodic of simulation of technological high voltage glow discharge electron sources". Electronic Modelling. – Vol. 32, #6, 2010. – P. 31 – 43. (Publisched in Russian language).

Waveguide Variable Attenuator Suitable for Electromechanical Control

Victor Kazmirenko, Irina Golubeva, Yuriy Prokopenko
Physical & Biomedical Electronics Department
National Technical University of Ukraine "Kyiv Polytechnic Institute"
Kiev, Ukraine
prok@phbme.kpi.ua

Abstract — **The paper continues research of microwave devices consisting of rectangular waveguide section with dielectric slab in it. The air clearance exists between dielectric slab and wide wall of the waveguide. Variation of the clearance by means of external actuator influences on propagation constant of the wave travelling through the device. This way the device can produce variable phase shift. Under certain conditions waveguide section goes beyond cutoff, and variable clearance produces controllable attenuation. Combined with high speed actuator device can operate as fast tunable attenuator or switch. The paper presents electromagnetic field problem and simulation results.**

Keywords—attenuator; phase shifter; waveguide; dielectric; piezoelectric actuator

I. INTRODUCTION

Mechanical reconfiguration is one of the basic tuning techniques in microwave systems. It could be implemented in rather static way, when adjustments are made time to time using tuning screw or similar part. Systems with high sensitivity to mutual positioning of their parts may have automated control with high speed mover. For example, split dielectric resonators (SDRs) placed in rectangular waveguide operate as tunable filters [1]. They exhibit resonant frequency change over 20% with displacement in range up to 100 μm. Such small displacement can be achieved using compact piezoelectric actuator. Piezoelectric actuators of "Cymbal" [2] and "Moonie" [3] type can produce such displacement at operating frequencies up to 100 kHz.

Similar concept was further developed in electromechanically controlled waveguide phase shifter [4], Fig. 1. Dielectric slab placed inside rectangular waveguide with variable air clearance is used to control propagation constant in control region and thus results in variable phase delay of the travelling wave.

Common feature that made both devices feasible is extremely high sensitivity to control factor, namely air clearance variation. To achieve it air gap is arranged perpendicularly to electric field lines, providing strong perturbation of electromagnetic field distribution.

Thorough study of waveguide phase shifter also demonstrates that waveguide height and consequently

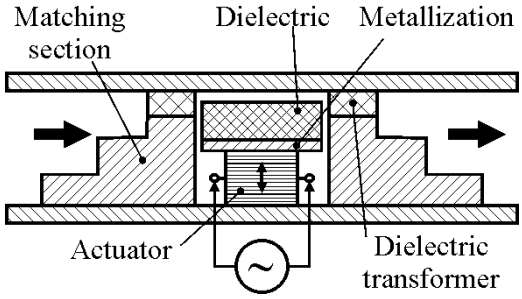


Fig. 1. Waveguide phase shifter with piezoelectric control

dielectric slab thickness must be reduced comparing to height of air filled section of the waveguide [5]. This requirement is dictated mainly by two factors. First, dielectric inset reduces cutoff frequencies of higher order modes

$$f_c = \frac{c}{2\pi\sqrt{\varepsilon}}\sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2},$$

where a, b are waveguide width and height respectively, ε is the dielectric inset permittivity, m, n are wave mode indices. Excitation of higher order modes becomes possible and that process introduces undesirable loss.

Secondly, further studies demonstrate, that specific phase shift increases as waveguide height reduces near dielectric inset [6]. For phase shifter operation it is important to remain above cutoff. This paper presents analysis of waveguide phase shifter in beyond cutoff conditions.

II. ELECTROMAGNETIC FIELD PROBLEM

Consider rectangular waveguide with dielectric inset in it, Fig. 2. When dielectric material completely fills waveguide's cross section, its dispersion equation is:

$$\varepsilon k^2 = \beta_x^2 + \beta_y^2 + \beta_z^2, \quad (1)$$

where k is the wavenumber of free space, β are wavenumbers along coordinate axes, ε is the permittivity of dielectric inset.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

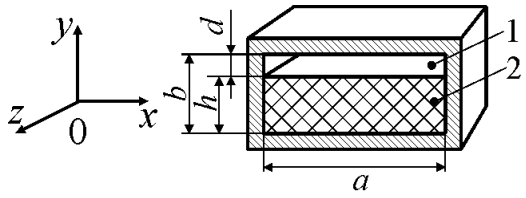


Fig. 2. Dielectric inset in the rectangular waveguide. 1 – air, 2 – dielectric

For the basic mode wavenumbers $\beta_x = \pi/a, \beta_y = 0$, and cutoff frequency is found from (1) as minimum frequency at which wavenumber β_z is still real, i.e.

$$\varepsilon k^2 \geq \left(\frac{\pi}{a}\right)^2. \qquad (2)$$

In fixed device requirement (2) is decided at design time. However, shiftable dielectric inset with variable air clearance may deliberately validate or invalidate condition (2) during device operation. If (2) is not satisfied, then β_z becomes imaginary, thus waveguide section does not allow traveling waves and attenuate incident power.

Electromagnetic field problem can be solved using longitudinal waves approach. When air clearance appears between dielectric and wide wall of the waveguide, basic TE_{10} mode reconfigures to longitudinal LM wave with β_y being nonzero. Applying boundary conditions on the waveguide walls and requiring continuity of tangential field components, electromagnetic problem can be reduced to complex nonlinear equation on transverse wavenumbers:

$$\beta_{ya} \tan\left(\beta_{ya} d\right) + \frac{\beta_{y\varepsilon}}{\varepsilon} \tan\left(\beta_{y\varepsilon} h\right) = 0,$$

where β_{ya}, $\beta_{y\varepsilon}$ are wavenumbers of 0y direction in air and dielectric, d is the air gap between dielectric and waveguide wall, h is dielectric height. This equation is solved numerically and propagation constant is found from (1). This approach was proven in experiments with waveguide phase shifter [5].

III. Simulation Results

Research of presented idea was performed by simulation. Calculations were performed at the frequency of 10 GHz in reference to X-band waveguide of 10 by 23 mm dimensions. Parameters of high quality microwave ceramics used for dielectric inset are listed in Table I.

TABLE I. Parameters of materials used in simulations

Material	ε	tanδ
Al_2O_3	11.6	$0.7 \cdot 10^{-4}$
$(Mg,Ca)TiO_3$	21	$2 \cdot 10^{-4}$
$BaTi_4O_9$	37	$3 \cdot 10^{-4}$
BLT	85	$2 \cdot 10^{-3}$

TABLE II. Reduced waveguide dimensions

Material	*a*, mm	*b*, mm
Al_2O_3	4.84	2.06
$(Mg,Ca)TiO_3$	3.60	1.53
$BaTi_4O_9$	2.71	1.15
BLT	1.79	0.76

As it is shown in [6], maximum sensitivity is observed, when propagation conditions are close to cutoff. Critical width of the waveguide a can be found from (2) for given material permittivity. To allow device operation slightly above cutoff waveguide width was assigned as 110% of critical width.

In [6] it is proven, that maximum tunability is observed, when device height is reduced to satisfy

$$h \leq \frac{\pi}{2k\sqrt{\varepsilon - 1}}.$$

Accounting these considerations, set of reduced dimensions was defined, as listed in Table II.

Fig. 3 presents simulation of transverse wavenumbers β_y in air gap and dielectric slab respectively. At zero gap basic TE_{10} mode exists and $\beta_{y\varepsilon} = 0$. When air clearance is created the $\beta_{y\varepsilon}$ becomes nonzero and grows as air gap increases. The wavenumber β_y of the air does not exist when there is no air gap. When air gap introduced, β_{ya} exhibits dependency inverse to that of $\beta_{y\varepsilon}$. At large values of the air clearance $\beta_{y\varepsilon}$ dominate, that means electromagnetic energy concentration in the dielectric slab, also known as dielectric effect.

Fig. 4 presents simulation of effective propagation constant β_z. At zero clearance and while waveguide operates above the

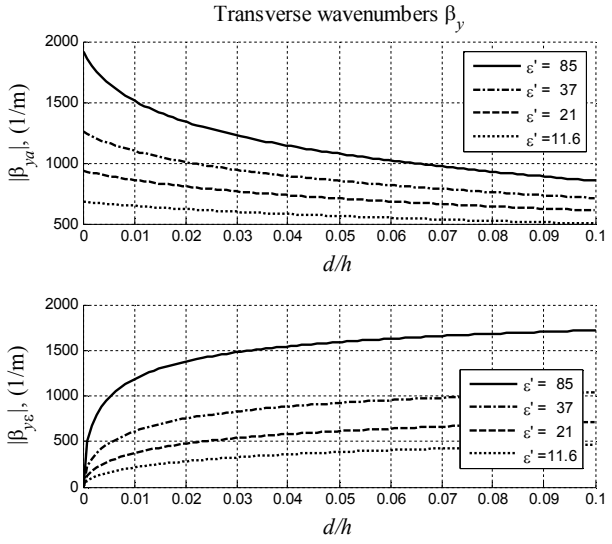


Fig. 3. Transverse wavenumbers β_y. Note that no β_{ya} exists when $d/h = 0$, i.e. when there is no air gap

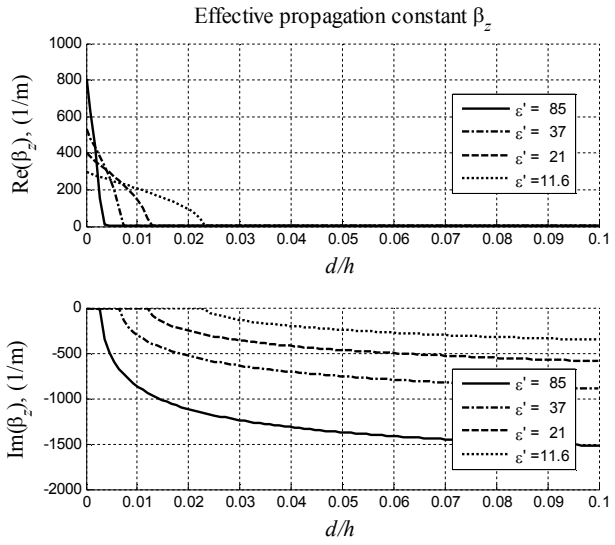


Fig. 4. Effective propagation constant

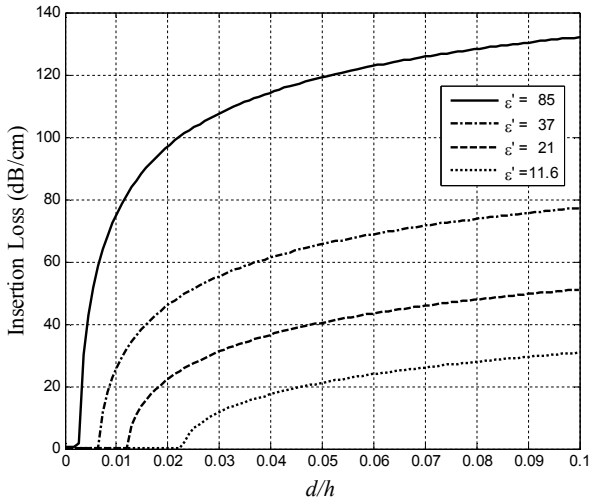


Fig. 5. Specific insertion loss of tunable attenuator

cutoff, imaginary part of the propagation constant remains small despite of dielectric loss. As air gap increases, real part of the propagation constant reduces and becomes zero at cutoff point according to (1). Then imaginary part of propagation constant grows as air gap increases. Large imaginary propagation constant is observed as strong attenuation of the incident electromagnetic energy.

Fig. 5 presents simulation of specific insertion loss. As long as waveguide operates above cutoff, insertion loss is determined mainly by energy dissipation in dielectric slab, so it remains relatively small due to low loss of materials used. However, beyond cutoff insertion loss increases a lot and may reach well above 100 dB/cm, so device can operate not only as attenuator, but practically as switch as well.

IV. Discussion

Sensitivity suitable for piezoactuator employment is achieved when waveguide dimensions are reduced considerably. Also higher sensitivity and stronger attenuation are achieved using materials with higher permittivity.

Dielectric inset and reduced waveguide dimensions make characteristic impedance of control area of the device quite different from the rest air filled part of the waveguide. Reflection loss can contribute to reduction of transmission coefficient, which is primary function of attenuator. However, return loss may create adverse condition for microwave source and load, degrading their performance or tuning. At any fixed frequency these complications can be avoided using quarter wavelength matching section. That obviously does not work if operation frequency changes. In the same time, variable air clearance between dielectric slab and waveguide wall varies impedance of controlled section as well. That breaks single frequency matching, so some kind of rather broadband impedance transformers are required to match attenuator to microwave source and load. Because waveguide width is reduced, some kind of two dimensional stepped structure is anticipated. The larger is device's impedance difference against regular waveguide one, the longer matching section might be required. So certain trade-off between sensitivity and mismatch may give minimum device dimensions. This problem yet to be researched.

V. Conclusion

Presented analysis demonstrates, that well studied waveguide phase shifter can have more applications as tunable attenuator or switch. At small values of the air gap insertion loss may be kept within fraction of decibel per centimetre thanks to high quality dielectric materials used. Large attenuation introduced with rather small displacement of dielectric slab in the waveguide, so fast piezoactuator can be used. Due to permittivity and dimensional discontinuity device requires matching to the regular part of the waveguide.

References

[1] Y. Poplavko, Y. Prokopenko, V. Molchanov, A. Dogan, "Frequency-tunable microwave dielectric resonator," IEEE Trans. on Microwave Theory and Techniques.– 2001. Vol. 48, Issue 6.– P. 1020–1027.

[2] A. Dogan, K. Uchino, and R. E. Newnham, "Composite piezoelectric transducer with truncated conical endcaps, cymbal," IEEE Trans. Ultrason., Ferroelect., Freq. Contr., vol. 44, pp. 597–605, May 1997.

[3] K. Onitsuka, A. Dogan, J. F. Tressler, Q. Xu, S. Yoshikawa, and R. E. Newnham, "Metal–ceramic composite transducer, the Moonie,"J. Intell. Mater. Syst. Structures, vol. 6, pp. 447–455, July 1995.

[4] M. Jeong, V. Kazmirenko, Y. Poplavko, B. Kim, S. Baik, "Electrically Tunable Phase Shifters With Air-Dielectric Sandwich Structure," Material Research Society Symposium Proceedings. Vol. 720.–2002.– P. H3.12.1–H3.12.6.

[5] V. Kazmirenko, Y. Poplavko, M. Jeong, S. Baik, "Waveguide dielectric phase shifter with fast piezoelectric control," IEEE MTT-S International Microwave Symposium Digest.– 2005.– P. 1059–1062.

[6] Y. Prokopenko, "Tunability limits of dielectric discontinuity between metallic planes," Technology and designing in electronic apparatus.– 2012, Vol.6.– P. 16–20. [in Russian].

Simulation Parameters of the Microwave Synthesizer for Radar System

Gramarchuk Y., Kobak N., Petrenko V., Slesarenko S.
Faculty of Electronics
National Technical University of Ukraine «KPI»
Kyiv, Ukraine
it.kotefi@gmail.com, slavapp@ukr.net, s.slesarenko@gmail.com

***Abstract* — Now radiolocation is one of the fastest growing industries in electronics. One of the characteristics of radiolocation (radar) technology is its rapid continuous improvement, due to the continuous development of new elements, theory and practice of radar. Modern radar uses the latest achievements of information theory and high-frequency electronics. The main object of our work is the assistance in researches of synthesizer of frequency in the block of input generator, which must give frequency 8.5 Ghz to provide the next block of the device the frequency 34 Ghz with the help of multiplier. Important part of our work is the simulation for synthesizer of frequency using software package "ADIsimPLL".**

Keywords— radiolocation; radar, simulation for synthesizer, microwave synthesizer.

I. RADAR

Important role in development of the radar system is its purpose. Radar is in most cases is the installation of special purpose. Its design is optimized for a particular purpose. The radar system millimeter waves for observation of objects and their motion in a certain area is a fixed ground installations. It is used in systems moving target, working in conditions of intensive clutter from the earth's surface.

II. PARAMETERS

The use of millimeter waves is fundamental for the achievement of high precisionpositioning and high-resolution radiological munication systems, the range of not exceeding 5-10 kilometers. One of the important tasks in building radar is to create a receiver-transmitter unit (RTU), the parameters which satisfy the electric-cerned and dimensions and mass characteristics of modern radar.

III. RECEIVE – TRANSMITTER UNIT

The main part of the radar is a receiver – transmitter unit (RTU), which functionally and structurally finished product.

More recently it became possible to create a fully solid-state radar in L-and S-wave bands with mirror antennas and active phased array antennas. In modern radar is used up to thousands of units, and in each of the transmitting unit up to 10 high-power microwave transistors. The advantage of solid-state radar is inextricably linked with the transition from concentrated to distributed sources of microwave power in the form of modules. The use of a large number of microwave modules increases the survivability of the station as a whole, since the refusal of even 10-20% of the total number of modules does not deprive the radar performance. Thus it becomes possible through the use of spare (redundant) modules and repair of failed radars provide continuous availability throughout the service life (10-25 years)[1].

In the RTU implemented modular design, where the RF - blocks are located in isolatedcompartments. All the schematics in this paper are implemented as a hybrid-integratedassemblies, placed in a sealed housing. As an active and passive components used surface mounting components

Transceiver unit (RTU) millimeter-wave band is designed to work as part of a coherent pulsed radar. During the operation of RTU performs the following operations:

- *in receive mode:*
  - Gain antenna received signal at microwave frequencies for a given level of intrinsic noise given to the input flange of the RTU;
  - Conversion of the total bandwidth in the low-frequency region of the spectrum for later the amplification and processing of high-frequency signals;
  - The formation of coherent signals synchronized to the internal subsystems RTU;
  - Simultaneous detection of the quadratic;
  - Signal generation operational control, reflecting the quality of functioning of the basic elements and subsystems RTU;
  - Development of secondary power supply voltages.
- b) transfer mode:
  - Gain and modulation of the output microwave signal;
  - The formation of highly stable in amplitude and phase of the pulse signal output power of 20 watts.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

Unit RTU Millimeter Wave structurally consists of two parts: a transceiver module (RTM) and the power amplifier unit (PAU) united by a microwave input tract (the circulator).

For optimization and synthesis of microwave monolithic integrated device needed fast enough and while the exact model elements. Mentioned qualities have a model based on equivalent electrical circuits. They can have acceptable accuracy and require large computational cost. The equivalent electrical circuit is easily integrated into existing design. However, to maintain the accuracy should be performed verification of these models in a given frequency range and geometry elements.

A common approach to the design of microwave monolithic integrated device is as follows. First, select the original or synthesized circuit on ideal elements. Followed by electrical parameters of ideal elements defined design parameters (geometrical dimensions) of the corresponding monolithic integrated elements (for example in the form of equivalent circuits and electromagnetic models). After that, in order to ensure the optimization of the desired characteristics of the resulting circuit in the space of design parameters of the monolithic integrated elements.

One of the drawbacks of this approach is to a great complexity. In addition, the use of accurate models of the monolithic integrated elements of optimization of the monolithic integrated circuits requires a time-consuming and may be ineffective or impossible. Not the fact that originally chosen circuit structure by replacing the real ideal elements will satisfy the requirements. Effective methods of designing the microwave monolithic integrated circuit solely responsible for the parasitic parameters of the elements taken into account at the stage of selection (fusion) chain structure. [4]

Based on the study of circuit design and manufacturing technology of solid-state microwave devices for domestic and foreign publications, working with RF equipment from the USA and the leading Russian companies, as well as from the experience of the design and operation with radio in the UHF and EHF bands proposed classification process and methods production units depending on the frequency range of their work. The dependence of the technologies used and the choice of hardware components from the operating frequency in the table.

TABLE I. DIFFERENCES BETWEEN DEVICES, WAVELENGTH RANGE AND KIND OF MINIATURIZATION

	Wavelength range			
	VHF and UHF	Centimetric	MM	SubMM
Transmission line	Coaxial wire	Coplanar line	Lines with a suspended substrate	Coplanar, dielectric, semi-open
Passive circuits	Scheme on lumped elements	Semi-lumped and distributed circuit	The lines on the material with a small ε	Monolithic integrated device volume assembly of structural elements
Active devices	Suspended elements, large-scale integrated circuits	Suspended base elements	Suspended single crystals	Monolithic integrated device volume assembly of crystals
Finished device	Printed circuit Boards	Hybrid integrated circuits	Hybrid integrated circuits	Monolithic integrated device
	Constructive miniaturization	Scheme - topological miniaturization		structural miniaturize -tion
	Kind of miniaturization			

IV. SIMULATION

In this part we want to show the results of our simulations and explain destination of our researches. First of all, modeling of processes, which take place in synthesizer of frequencies of millimeter range, is a big part of projecting and has a big importance for better way to get the right components. In this work computer modeling conducted to confirm or refute the appropriateness of the assumptions of a particular structure. For simulating we used a program "ADIsimPLL". This program give us a opportunity to view, which parameters and components we need to make a synthesizer. Because our frequency of synthesizer is 8,5 GHz, we took a microchip, which can work on this frequency ADF 41020 . Now it's only one microchip, which could work on this frequency. Examining different models of filters, we stopped at the operation amplifier with feedback, because we get the lowest level of phase noises of filter (-103.8 dBC/Hz) and the same phase noises (-103.8 dBC/Hz) (from the table 1. on the frequency 100 kHz).

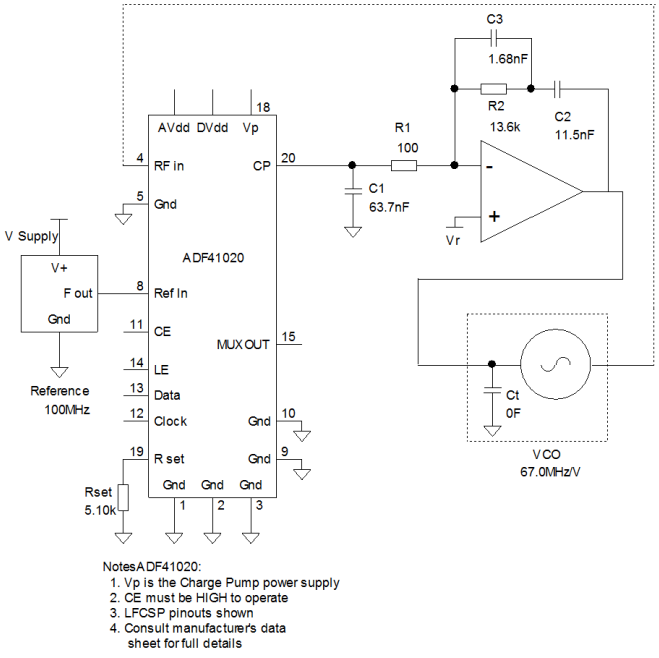


Fig. 1. Frequency synthesizer

In the process of projecting we received a master oscillator-based voltage-controlled oscillator [2], and covered system phase-locked loop frequency control (PLL) to the reference generator.

The main elements of the PLL is a phase detector, on one of the inputs connected control signal, and a controlled generator, whose input is connected to the output of the phase

detector and the output connected to another input of the phase detector. An important feature of the PLL, which distinguishes it from most other systems of automatic control, is that the output variable of the controlled oscillator

is frequency and input variable - difference between the phase of the control signal and the feedback signal coming from the output of the controlled generator. [3]

TABLE II. PHASE NOISE TABLE

Frequency	Total	Chip ADF41020	Filter
100	-67.32	-67.35	-89.11
1k	-64.09	-65.52	-69.62
10k	-72.15	-83.85	-72.45
100k	-103.8	-135.1	-103.8
1M	-143.5	-194.8	-143.5

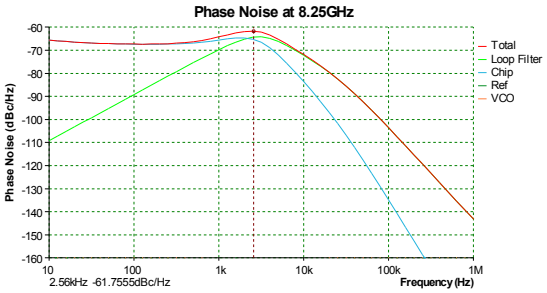


Fig. 2. Phase Noise at 8.25 GHz

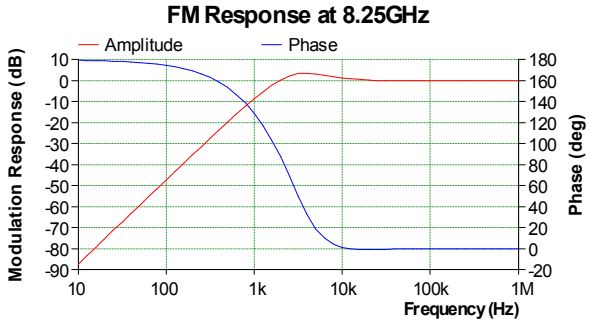


Fig. 3. Modulation and phase response

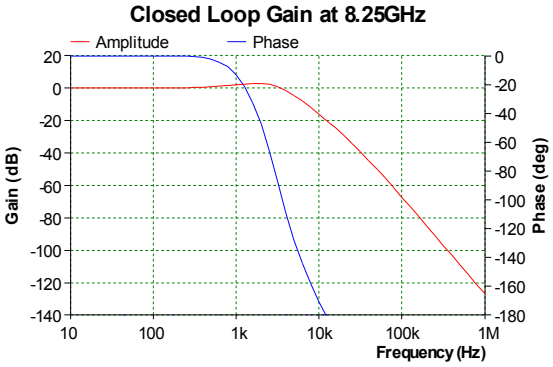


Fig. 4. Closed Loop Gain at 8,25 GHz

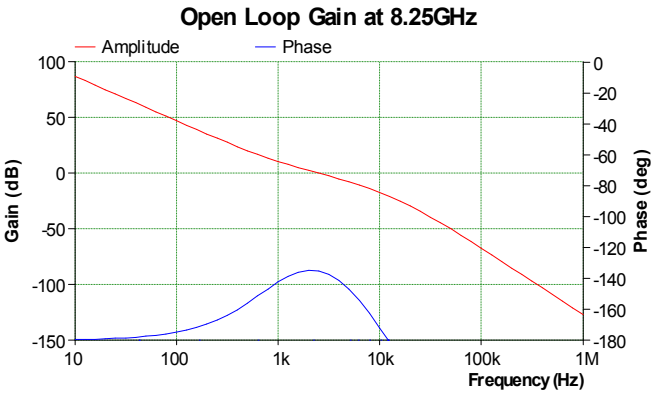


Fig. 5. Open Loop Gain at 8,25 GHz

V. CONCLUSIONS

With the help of simulating of the our synthesizer with software package “ADIsimPLL” with parametres, which contain our task, we took optimal chip ADF 41020, we get chracteristics of the noise and response.

Це get the optimal operation amplifier with feedback with minimal (for our task) level of the phase noises (-103.8 dbc/Hz).

This simulating shows, that we can include this model to the real device in the millimeter-wave range synthesizer.

REFERENCES

[1] Grishakov M. N., “Failure analysis and development of technical measures to improve the reliability of solid-state microwave modules for radar”, synopsis, M., 2010, p. 3;

[2] Kobak N.N., Petrenko V.P., Slesarenko S.S. Noise Immunity in Data Channel of Multifunction Information Systems // Proc. XXXII International Conference Electronics and Nanotechnology "ELNANO-2012", Kyiv, April 10-12, 2012, PP.98-99.

[3] Magazine “Components and technologies”, #8, 2003, “A few words about system PLL: phase-locked loop frequency control”, Golub V.

[4] Projecting microwave monolithic integrated systems based on model transformation elements, dissertation, Sheerman F. I, Tomsk, 2007;

[5] Complex miniaturization of components and devices transmitter - receiver modules microwave, dissertation, Babunko S. A., N. Novgorod, 2006;

Distributed Feedback Analysis

Oleg Vityaz
Physical and Biomedical Electronics Lab.
National Technical University of Ukraine "Kyiv Polytechnic Institute"
Kyiv, Ukraine
ovityaz@hotmail.com

***Abstract* — A new technique for circuit feedback analysis is presented. A formula for the return ratio is derived using a strict definition of feedback. The technique may be applied in both laboratory-based and computer-aided analysis of feedback in electronic circuits and other dynamic systems. Presented simulation results demonstrate the effectiveness of the developed technique.**

Keywords—distributed feedback; feedback relation; excitation signal; impedance; one-port system; return ratio.

I. Introduction

Stability is a key property of dynamic systems. Stability means steady state which may be a rest, or a continuous motion. Analysis of stability should be a part of the system design process. It should be a design-oriented analysis [1].

Stability of a circuit may be examined using the feedback theory where a small-signal AC model is analyzed at a DC operating point. According to Bode, a transfer function K of a system with feedback is related to a transfer function K_0 of the system without feedback as

$$K = \frac{K_0}{F} \tag{1}$$

where F is a feedback factor [2]. The locations of the feedback factor zeroes on a complex plane determine stability properties of a circuit but provide no information about stability margins.

Computation of a return ratio is used as an alternative approach to stability analysis. Over the past decades, several methods that use return ratio in stability analysis were developed. They may be grouped in two sets: techniques based on injecting of test signals in a circuit [1], [3] and techniques based on using two-port parameters [4]-[6]. A special place is taken by the technique based on a signal-flow graph [7]. Sometimes these techniques lead to different results, and the reasons are discussed in [8], [9]. To avoid inconsistencies, our return ratio calculation technique is based on a strict definition of feedback.

II. Feedback definition

There is a broad class of entities that may be called systems. Any physical system occupies a limited space and interacts with its environment via ports. There are systems with one or several ports.

Feedback in a system is a continuous dynamic process caused by external excitations and internal interactions between the parts constituting the system. The excitation signal X_s from the environment is applied to the input port. The feedback signal X_f coming back from the system also contributes to the input. The resulting signal is X_i as illustrated graphically in Fig. 1 and defined mathematically in the expression that follows:

$$X_i = X_s + X_f . \tag{2}$$

Signals X_s, X_f, and X_i may represent various physical domains. In electronics they may be currents or voltages in time or frequency domain. In the further discussion, signals are denoted as X, V, and I, component impedances as Z, and admittances as Y unless specified otherwise.

Assuming that excitation signal X_s is independent and known, the feedback analysis requires identification of any of the two remaining signals X_f and X_i. Relationship between the feedback signal X_f and the resulting signal X_i depends on properties of the system and its environment. The return ratio is defined as a ratio of feedback and resulting signals [2]:

$$T = \frac{X_f}{X_i} . \tag{3}$$

A stability margin is defined as a distance between the return ratio T and unity [10].

III. Feedback in a one-port system

The simplest feedback system is a one-port system (two-terminal network) presented by its impedance Z_i or admittance Y_i. The environment is modeled by a non-ideal signal source with impedance Z_s or admittance Y_s. Fig. 2 (a) shows a one-port system and its environment in terms of their impedances, and Fig. 2 (b) shows the same system and its environment in terms of their admittances. Voltage across a branch and current through the branch are marked by curved and dark arrows, respectively. The voltage and current arrows are defined to point in the opposite directions.

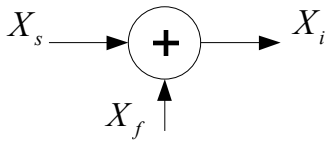


Fig. 1. Graphical illustration of the feedback relation.

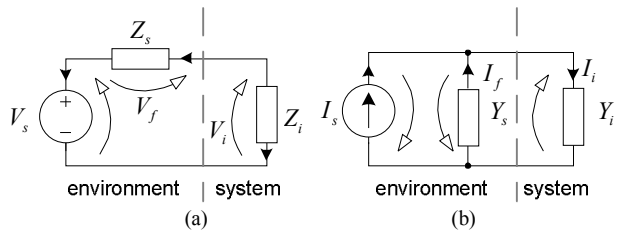


Fig. 2. The simplest feedback systems: (a) voltage divider; (b) current divider.

The feedback relation (2) may be interpreted either as Kirchhoff's voltage law for the loop in Fig. 2 (a)

$$V_i = V_s + V_f \tag{4}$$

or as Kirchhoff's current law for one of the nodes in Fig. 2 (b)

$$I_i = I_s + I_f . \tag{5}$$

The feedback in a voltage divider will be called a voltage feedback, and its return ratio T_V may be obtained according to (3) and (4):

$$T_V = \frac{V_f}{V_i} = -\frac{Z_s}{Z_i} . \tag{6}$$

Similarly, the feedback in a current divider will be called a current feedback, and its return ratio T_I may be obtained with the use of (3) and (5):

$$T_I = \frac{I_f}{I_i} = -\frac{Y_s}{Y_i} . \tag{7}$$

If the source is ideal, i.e., $Z_s = 0$ or $Y_s = 0$, it follows from (6) and (7) that the return ratio becomes equal zero, hence, the feedback vanishes. All the energy produced by the source is consumed by the system.

On the other hand, a trivial one-port system with $Z_i = 0$ or $Y_i = 0$ shown in Fig. 4 has infinite return ratio. In this case all the energy produced by the source is consumed by the environment.

If the voltage and current source impedances are the same, $Z_s = 1/Y_s$ and considering the system with a finite impedance $Z_i = 1/Y_i$, it follows from (6) and (7) that

$$T_V \cdot T_I = 1 . \tag{8}$$

Under the same conditions, the following relation may be derived from (2), (6) and (7)

$$\frac{T_V - 1}{T_I - 1} = \frac{Z_s}{Z_i} . \tag{9}$$

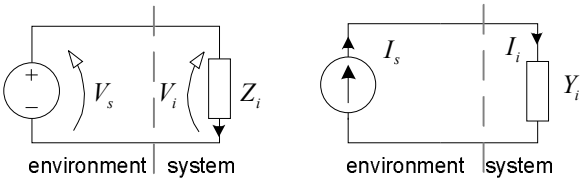


Fig. 3. The systems without feedback.

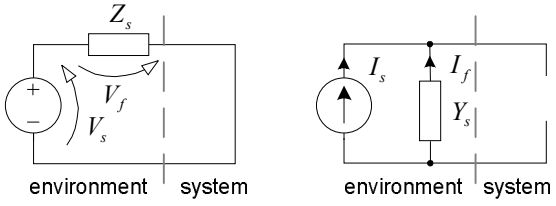


Fig. 4. The systems with infinite feedback.

IV. Feedback in a Two-Port System

Consider a linear or linearized two-port system with a non-ideal input signal source connected to the first port with the impedance Z_i or admittance Y_i. The port and the source constitute one of the two networks shown in Fig. 2. An output response signal X_o, observed at the second port and shown in Fig. 5 (a), is proportional to the excitation signal X_s. Mathematically this may be written as

$$X_o = K \times X_s \tag{10}$$

where X_o can either be an output voltage V_o or an output current I_o, and K is a transfer function of the system. It is important to note that the second port of the system may be chosen arbitrarily.

All other signals are also proportional to the excitation signal X_s and, hence, may be expressed similarly:

$$\begin{aligned} X_f &= h_f \times X_s \\ X_i &= h_k \times X_s . \end{aligned} \tag{11}$$

From (10) and (11) the following relationships may be derived:

$$\begin{aligned} \frac{X_o}{X_i} &= \frac{K}{h_k} = K_0 \\ \frac{X_f}{X_o} &= \frac{h_f}{K} = \beta \end{aligned} \tag{12}$$

where K_0 is a transfer function of a two-port, β is a feedback transfer function.

A signal-flow graph of a two-port system shown in Fig. 6 (a) is obtained with the use of the cause-and-effect relations (10) and (12) [7]. A loop in the graph may be mistakenly interpreted as a unilateral signal flow in a circuit. However, it only represents a relationship between signals in a cause-and-effect notation. The presence of the unilateral closed-loop feedback in an actual circuit is rather an exception than the

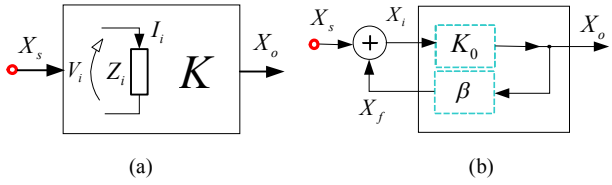


Fig. 5. A block-diagram representation of a two-port system: (a) feedback signal is inside; (b) feedback signal is outside.

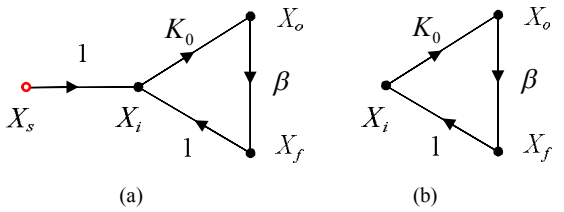


Fig. 6. A signal-flow graph: (a) of a two-port system and a source; (b) of an oscillator.

rule. Nevertheless, it has become common to use the block-diagram shown in Fig. 5 (b) for graphical representation of feedback in a two-port system.

Transfer function K may be derived from the signal-flow graph using Mason's rule [7]:

$$K = \frac{X_o}{X_s} = \frac{K_0}{1 - \beta K_0} = \frac{K_0}{1 - T}. \quad (13)$$

The expression (13) shows that when $T = 1$ the transfer function K is infinite. The output response may be finite only when $X_s = 0$. This means that oscillation occurs in the system. A signal-flow graph of an oscillator is shown in Fig. 6 (b). The oscillation frequency ω may be obtained by solving the following equation:

$$T(j\omega) = 1. \quad (14)$$

Thus, the return ratios T_V and T_I may be used for stability analysis and for stability margin calculation of a feedback system. When system is stable, the stability margin may be obtained from frequency characteristics of its return ratio.

The oscillation condition may be expressed in terms of the source and system impedances with the use of (6), (7) and (14)

$$Z_i = -Z_s. \quad (15)$$

The return ratio may be written using (2) and (3) in the following way:

$$T = \frac{X_f}{X_i} = \frac{X_i - X_s}{X_i} = 1 - \frac{X_s}{X_i} \quad (16)$$

which simplifies to

$$T = 1 - \frac{1}{X_i} \quad (17)$$

when $X_s = 1$. Thus, the return ratio of a system may be obtained by applying a test input signal $X_s = 1$ and measuring or calculating the system's input signal X_i.

V. Numerical Results

A simple LC-network is depicted in Fig. 7 (a). Its voltage return ratio may be derived by connecting an ideal voltage source in series as depicted in Fig. 7 (b). Then according to (6) the return ratio is

$$T_V(j\omega) = -\frac{Z_c}{Z_L} = \frac{1}{\omega^2 LC}. \quad (18)$$

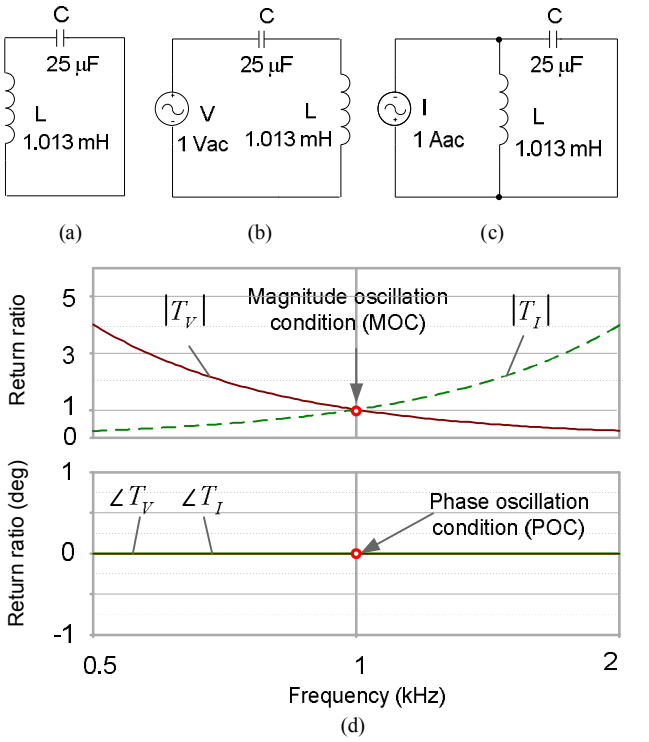


Fig. 7. A simple LC-network : (a) original network; (b) feedback loop; (c) feedback node; d) return ratio magnitude and phase.

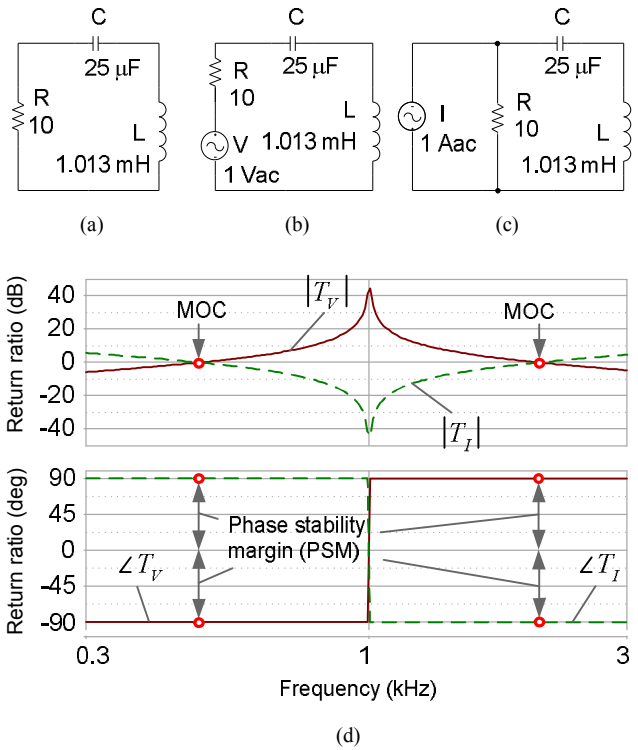


Fig. 8. A simple RLC-network : (a) original network; (b) feedback loop; (c) feedback node; (d) return ratio magnitude and phase.

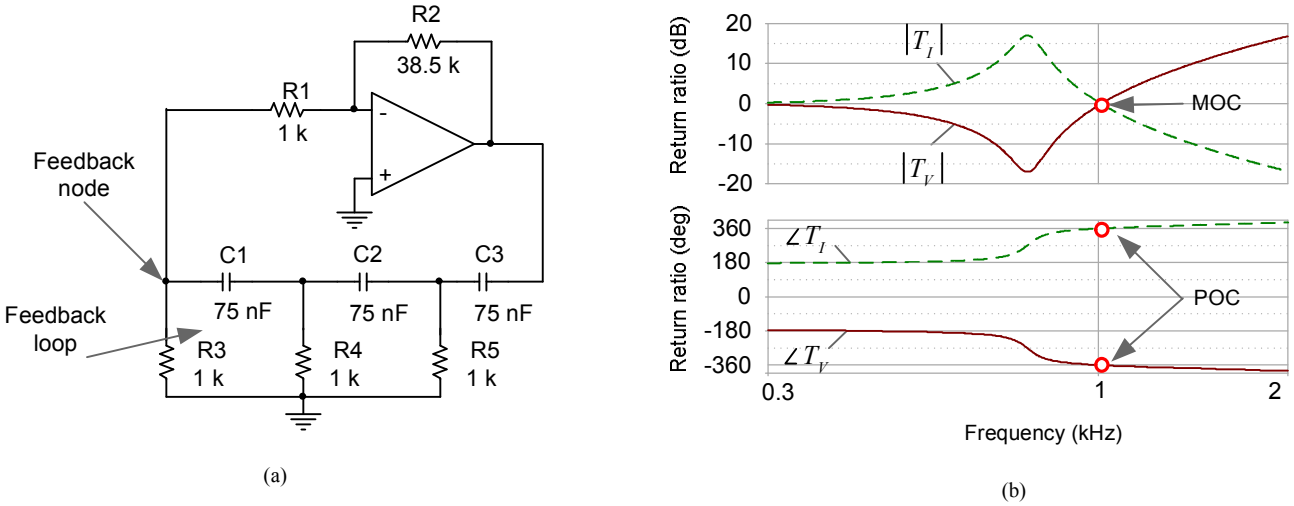


Fig. 9. RC phase-shift oscillator: (a) oscillator schematic; (b) return ratio magnitude and phase.

From the equation $T_V(j\omega)=1$ the oscillation frequency f_o may be derived:

$$f_o=\frac{\omega_o}{2\pi}=\frac{1}{2\pi\sqrt{LC}}=1\ kHz\,. \tag{19}$$

The AC current source connected in shunt, as shown in Fig. 7 (c), allows us to compute the current return ratio with the use of (7):

$$T_I(j\omega)=-\frac{Y_L}{Y_C}=\omega^2 LC. \tag{20}$$

Similarly, by equating (20) to unity the oscillation frequency (19) may be derived.

A simple RLC-circuit is depicted in Fig. 8 (a). Its voltage and current return ratio plots show that there are two frequencies at which the magnitude condition for oscillation $|T(j\omega)|=1$ is satisfied. As there is a 90 degrees phase stability margin at these frequencies, oscillation doesn't occur.

The same technique may be applied to define the oscillation frequency of RC-oscillator depicted in Fig. 9. An ideal AC voltage source and an ideal AC current source were connected, respectively, in series and in shunt to resistance $R3$ to obtain the voltage and current return ratios. At frequency around $1\ kHz$ the oscillation conditions both for magnitude and phase are satisfied and, hence, oscillation occurs.

A Wien bridge oscillator is shown in Fig. 10. The voltage

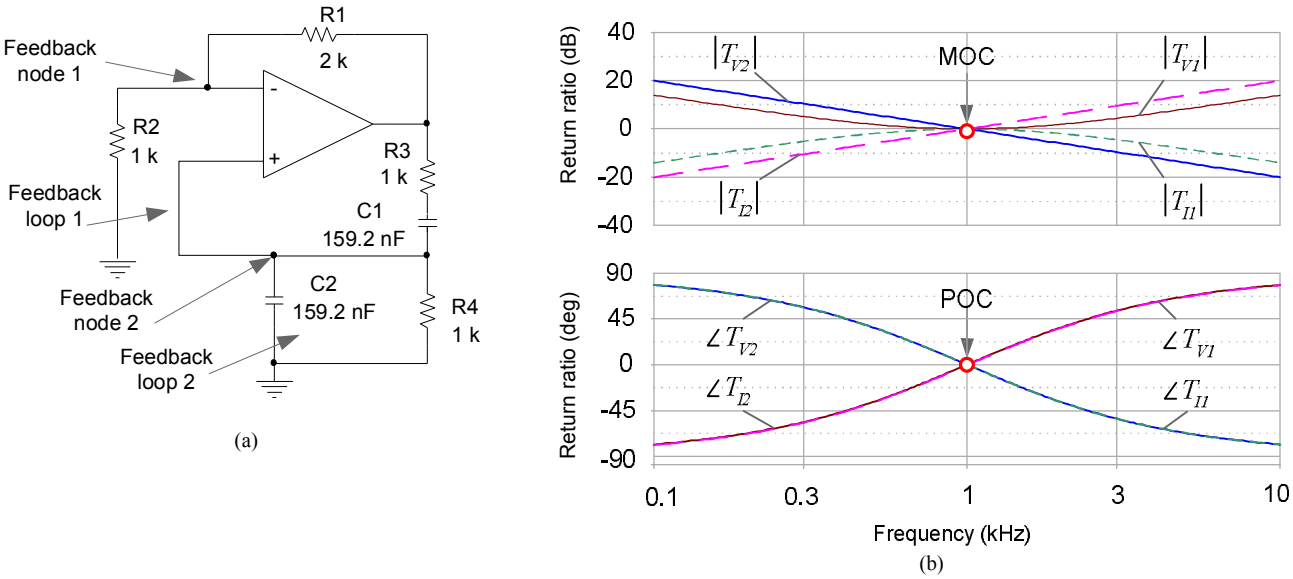


Fig. 10. Wien-bridge oscillator: (a) oscillator schematic; (b) return ratio magnitude and phase.

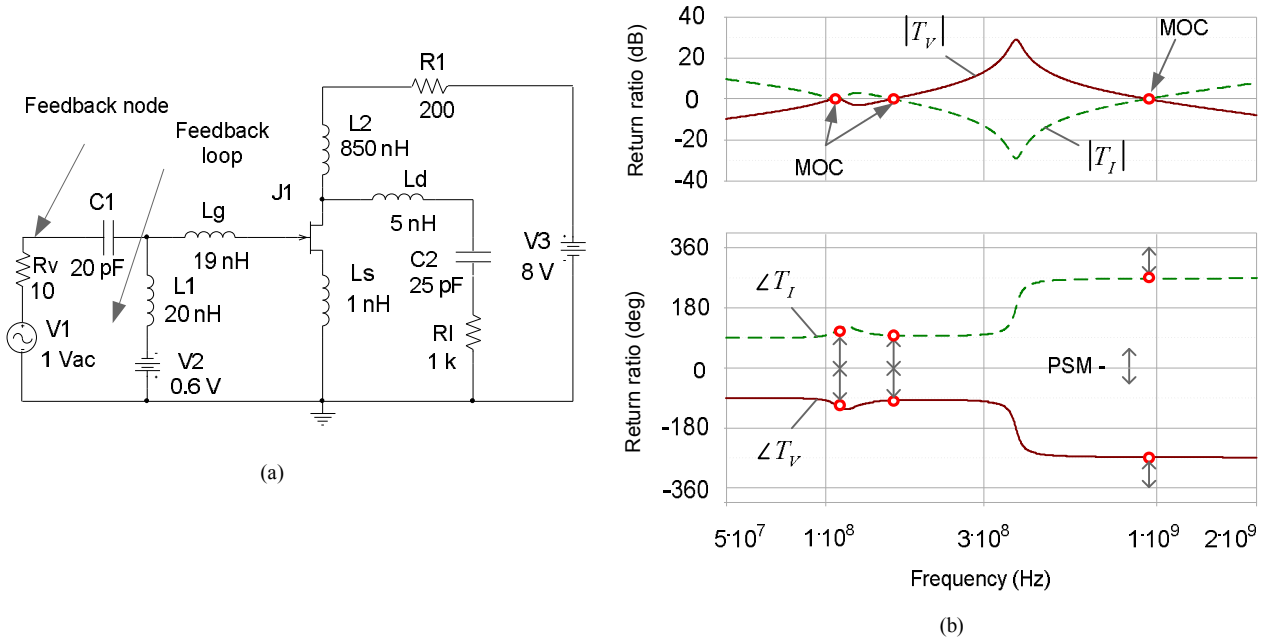


Fig. 11. An amplifier: (a) amplifier schematic; (b) return ratio magnitude and phase.

return ratios T_{V1} and T_{V2} were obtained by connecting in turn an ideal AC voltage source in the loops containing, respectively, resistance $R2$ and capacitance $C2$. Similarly, the current return ratios T_{I1} and T_{I2} were obtained by connecting in turn an ideal AC current source in shunt, respectively, to resistance $R2$ and capacitance $C2$. All the ratios indicate the same oscillation frequency.

An amplifier network is depicted in Fig. 11. The return ratio plots show that the amplifier is not oscillating in spite of the magnitude oscillation condition is satisfied at several frequencies. This is due to the phase stability margin at these frequencies.

Oscillation conditions obtained from presented return ratio frequency-domain characteristics are in agreement with time-domain analysis.

Conclusions

Feedback as a unilateral signal flow from the system's input via the feedback loop back to the system's input is rather a theoretical assumption than a physical reality. The feedback signal flow is distributed over all possible paths in the circuit. In the proposed technique there is no need to identify feedback loops and points for test signal injection. The return ratio may be determined by connecting either an AC voltage source to a system's port and measuring the input voltage or by connecting of an AC current source and measuring the input current. These measurements provide sufficient data for detection of stability margin. The technique may be applied in both laboratory-based and computer-aided analysis of feedback in electronic circuits and other dynamic systems.

References

[1] R.D. Middlebrook, "The general feedback theorem: A final solution for feedback systems", *Microwave magazine, IEEE,* vol. 7, no. 2, pp. 50-63, Apr. 2006.

[2] H.W. Bode, *Network Analysis and Feedback Amplifier Design*. New York: D. Van Nostrand, Inc., 1945, pp. 31-102.

[3] P.J. Hurst and S.H. Lewis, "Determination of stability using return ratios in balanced fully differential feedback circuits", *IEEE Trans. Circ. Syst. II*, vol. 42, no. 12, pp. 805 - 817, Dec. 1995.

[4] M. Tian, V. Visvanathan, J. Hantgan, K. Kundert, "Striving for smal-signal stability", *IEEE Circuits and Devices Mag,* vol. 17, no. 1, pp. 31-41, Jan. 2001.

[5] Bellone, S., Licciardo, G.D., "Extension of the TPA method for the exact analysis of feedback circuits in terms of the return ratio", Proc. of *Environment and Electrical Engineering (EEEIC), 10th International Conference,* 2011, pp. 346-349.

[6] B. Nikolic and S. Marjanovic, "A general method of feedback amplifier analysis", *Proc. of ISCAS,* Monterey, CA, 1998, pp. 415-418.

[7] J. Mason, "Feedback theory: Some properties of signal-flow graphs", *Proc. IRE,* vol. 41, no. 9, pp.1144-1156, Sep.1953.

[8] P.J. Hurst, "A comparison of two approaches to feedback circuit analysis", *IEEE Trans. Educ.,* vol. 35, no. 3, pp. 253 - 261, Aug. 1992.

[9] O.A. Vityaz, "Parametrical Identification of the System Regularity", *Electronics and communication,* (rus.), vol. 5, pp. 55-59, Apr. 2010.

[10] W.K. Chen, *Active Network and Feedback Amplifier Theory*. New York: McGraw-Hill, and Washington, DC: Hemisphere, 1980, pp. 185-246.

Optimal Synthesis of Electronic System for Gyroscopic Nautical Compass Stabilisation

O.A. Sushchenko
National Aviation University
Kiev, Ukraine
sushoa@ukr.net

***Abstract* — The problem of design of optimal electronic system for stabilisation of the gyroscopic nautical compass taking into consideration uncertainties of the mathematical description and external disturbances is considered. The basic modes of operation of such system are listed. The modern methods of the parametrical and structural robust optimization suitable for both modernization and creation of the new systems are represented. The simulation results are given.**

Keywords—Electronic system, H_2 / H_∞ ***-optimization,*** H_∞ ***-synthesis.***

I. INTRODUCTION

Now processes attending exploitation of the nautical vehicles are sufficiently complicated [1]. To provide the precision stabilisation of the gyroscopic nautical compass it is necessary to use the electronic systems. Synthesis of these systems is implemented in conditions of uncertainties caused by both inaccuracies of the mathematical description of the real system and influence of the internal and external disturbances. Mainly, the systems designed for exploitation at the nautical vehicles are subjected to influence of the disturbances caused by the sea irregular waves. The modern approach to these systems synthesis is creation of the robust systems able to operate in conditions of both the parametrical structured and the external coordinate disturbances.

II. OPTIMAL SYNTHESIS FEATURES

The studied system allows to solve the following tasks of the gyroscopic nautical compass stabilisation:

1) the coarse stabilisation of sensors and electronic devices in the mode of previous leveling;

2) the precision stabilisation of sensors and electronic devices in the mode of precision leveling;

3) the initial alignment;

4) the simultaneous precision stabilization and heading determination;

5) the calibration of sensors during operation and berthing.

The structural scheme of the electronic system for the gyroscopic nautical compass stabilisation is represented in fig. 1.

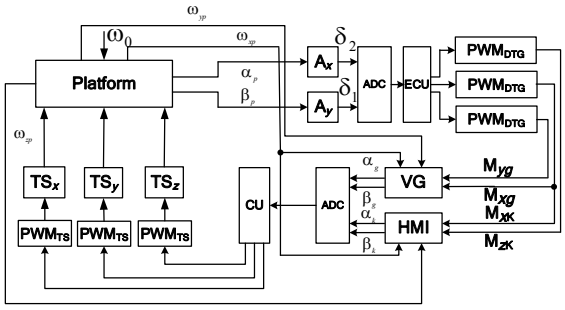


Fig. 1. The structural scheme of the electronic system: A is the accelerometer; ECU is the electronic-calculating unit; PWM is the pulse width modulator; ADC is the analog-to-digital converter; DTG_g is the vertical gyro; DTG_h is the directional gyro; TS is the torque sensor

Improvement of the studied electronic systems may be implemented in two directions including operational system development and creation of the new systems. The operational system development in conditions of uncertainties may be realized by means of the robust parametrical optimization. The modern approach to design of the new systems foresees use of the H_∞-synthesis or the robust structural optimization.

One of the modern approaches to definition of the robust optimization quality criteria is use of the H_∞-norm of the system complementary sensitivity function [2, 3]. In this case the quality criterion and the formalized problem statement may be represented in the following form

$$J_{H_{\infty p}}(K) = \| \Phi(K) \|_\infty , \quad (1)$$

$$K^* = \arg \inf_{K \in D} J_{H_\infty}(K) , \quad (2)$$

if the condition

$$K(s) \in D ,\ D : \mathrm{Re} \,|\, \mathrm{eig}(I + L(s)) | < 0 \quad (3)$$

is satisfied. In the general case the optimization criterion $J_{H_\infty p}(K)$ represents the H_∞-norm of the system complementary sensitivity function $\Phi(K,s)$; $K(s)$ is the controller matrix transfer function; $\| \cdot \|_\infty$ is the notation of the

H_∞-norm; $W(s)$ is the stabilisation object transfer function; $L(s) = K(s)W(s)$.

The mathematical description of the studied system represents the set of the models in the different operation modes. For example, in the mode of previous levelling the system model may be described by the set of the differential equations

$$\dot{\omega}_{xp} = [-(J_z - J_y)\omega_{yp}\omega_{zp} - (f_x + k_7)\omega_{xp} - k_1\delta_1 - \\ - k_5(-\delta_1 + k_3\beta)/T + H\omega_{0x}]/J_x;$$

$$\dot{\omega}_{yp} = [-(J_x - J_z)\omega_{xp}\omega_{zp} - (f_y + k_8)\omega_{yp} - k_2\delta_1 - \\ - k_6(-\delta_2 + k_4\alpha)/T + H\omega_{0y}]/J_y;$$

$$\dot{\omega}_{zp} = [-(J_y - J_x)\omega_{xp}\omega_{yp} - (f_z + k_9)\omega_{zp} + H\omega_{0z}]/J_z;$$

$$\dot{\alpha} = (\omega_{xp}\sin\gamma + \omega_{yp}\cos\gamma)/\cos\beta;$$

$$\dot{\beta} = \omega_{xp}\cos\gamma - \omega_{yp}\sin\gamma;$$

$$\dot{\gamma} = \omega_{zp} + \mathrm{tg}\beta(\omega_{xp}\sin\gamma + \omega_{yp}\cos\gamma);$$

$$\dot{\delta}_1 = (-\delta_1 + k_3\beta)/T;$$

$$\dot{\delta}_2 = (-\delta_2 + k_4\alpha)/T,$$

where J_x, J_y, J_z are the platform inertia moments; f_x, f_y, f_z are the viscous friction torques; $\grave{I}_0$ is the dry friction torque; $\omega_{x0}, \omega_{y0}, \omega_{z0}$ are the external angular rates; $\delta_{x\ddot{i}}, \delta_{y\ddot{i}}$ are the accelerometers signals, T_A is the time constant; k_A is the accelerometer transfer constant; $k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8, k_9$ are the stabilisation laws coefficients; $\omega_{xp}, \omega_{yp}, \omega_{zp}$ are the platform angular rates.

The represented model allows to carry out simulation of the designed systems. But the modern calculating procedures of the parametrical and structural optimization require use of the state space models. The above stated model after linearization may be represented in the space of states in the following way

$$\dot{x} = Ax + Bu;$$
$$y = Cx + Du,$$

where x, u, y are the state, control, observation vectors, A, B, C, D are the state space matrices.

$$A = \begin{bmatrix} -\frac{f+k_7}{J_x} & 0 & 0 & -\frac{k_1 - k_5/T}{J_x} & 0 & -\frac{k_3 k_5/T}{J_x} & 0 & 0 \\ 0 & -\frac{f+k_8}{J_y} & 0 & 0 & -\frac{k_2 - k_6/T}{J_y} & 0 & -\frac{k_4 k_6/T}{J_x} & 0 \\ 0 & 0 & -\frac{f+k_9}{J_z} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{T} & 0 & \frac{k_3}{T} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{T} & 0 & \frac{k_4}{T} & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{H}{J_x} & 0 & 0 \\ 0 & \frac{H}{J_y} & 0 \\ 0 & 0 & \frac{H}{J_z} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}; C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix};$$

$$D = 0.$$

III. PARAMETRICAL OPTIMIZATION

To increase efficiency of the robust parametric optimization it is necessary to use the mixed H_2/H_∞ approach. In this case, the complex quality criterion may be described by the following expression

$$J_{H_2/H_\infty} = \lambda_2^{\mathrm{nom\,d}} \| \Phi(K) \|_2^{\mathrm{nom\,d}} + \lambda_2^{\mathrm{nom\,s}} \| \Phi(K) \|_2^{\mathrm{nom\,s}} + \\ + \sum_{i=1}^{n} \lambda_{2_i}^{\mathrm{dis\,s}} \| \Phi(K) \|_{2_i}^{\mathrm{dis\,s}} + \sum_{i=1}^{n} \lambda_{2_i}^{\mathrm{dis\,d}} \| \Phi(K) \|_{2_i}^{\mathrm{dis\,d}} + \\ + \lambda_\infty^{\mathrm{nom}} \| \Phi(K) \|_\infty^{\mathrm{nom}} + \sum_{i=1}^{n} \lambda_\infty^{\mathrm{dis\,i}} \| \Phi(K) \|_\infty^{\mathrm{dis\,i}} + PF,$$

where $\|\cdot\|_2^{\mathrm{nom\,d}}$, $\|\cdot\|_2^{\mathrm{nom\,s}}$, $\|\cdot\|_{2_i}^{\mathrm{dis\,s}}$, $\|\cdot\|_{2_i}^{\mathrm{dis\,d}}$ are H_2-norms of the transfer functions of the nominal system and the system disturbed by the parametrical structured disturbances for the deterministic and stochastic cases; $\|\cdot\|_\infty^{\mathrm{nom}}$, $\|\cdot\|_\infty^{\mathrm{dis\,i}}$ are the H_∞-norms of the transfer functions of the nominal system and the system disturbed by the parametrical structured disturbances; $\lambda_2^{\mathrm{nom\,d}}, \lambda_2^{\mathrm{nom\,s}}, \lambda_\infty^{\mathrm{nom}}, \lambda_{2_i}^{\mathrm{dis\,d}}, \lambda_{2_i}^{\mathrm{dis\,s}}, \lambda_\infty^{\mathrm{dis\,i}}$ are the weighting coefficients of the appropriate norms; n is a number of the models disturbed by the parametrical structured disturbances; PF is the penalty function satisfying the system stability conditions during the optimization execution.

The statement of the H_2/H_∞-optimization problem (1), (2) under condition (3) becomes

$$K^* = \arg \inf_{K \in D} J_{H_2/H_\infty}(K).$$

Such approach to the optimization is called the multi-purpose as it allows to find a compromise between the system accuracy and robustness [2].

The parametrical H_2/H_∞-optimization includes creation of the mathematical descriptions of the studied electronic system both in the space of states and taking into consideration

non-linearities peculiar to the real system, determination of the minimal and balanced realizations of the linearized model and use of the optimization genetic algorithm. Checking of the synthesis results is implemented by means of the non-linear model as much as possible close to the real system. In the case of the unacceptable results the parametrical optimization must be repeated after change of the initial conditions or the weighting coefficients.

The results of the synthesized system simulation are represented in fig. 2, 3. These representations allow to carry out the comparative analysis of the gyroscopic nautical compass stabilization by means of the non-optimized and robust designed systems under disturbance actions.

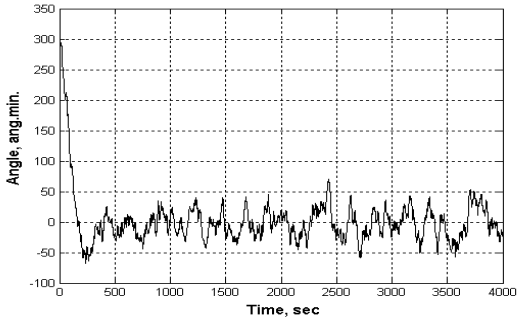


Fig. 2. The angle of deviation from the plane of horizon in the mode of precise leveling for the non-optimized system

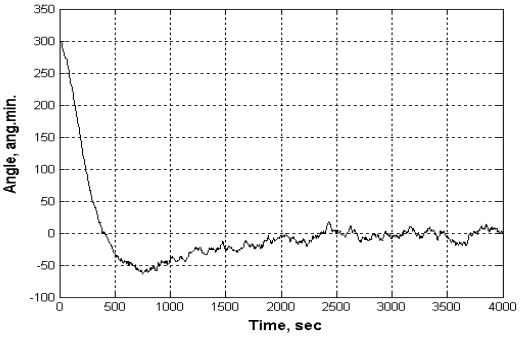


Fig. 3. The angle of deviation from the plane of horizon in the mode of precise leveling: for the optimized system

The results of the parametrical optimization in the mode of the initial alignment are represented in fig. 4, 5.

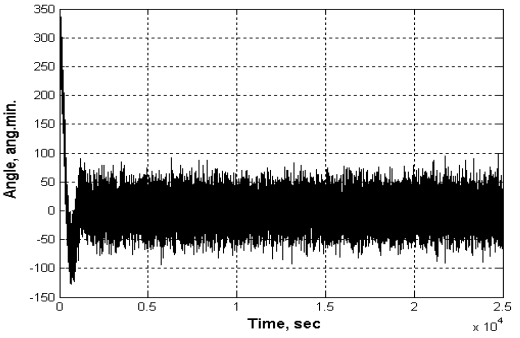


Fig. 4. An angle of the azimuth in the mode of the initial alignment for the non-optimized system

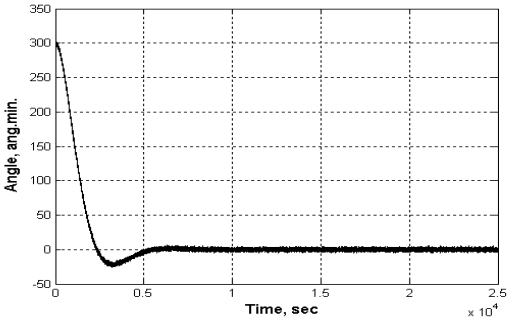


Fig. 5. An angle of the azimuth in the mode of the initial alignment for the optimized system

The robust parametrical H_2 / H_∞-optimization results represent the vectors of the stabilisation laws coefficients. For the mode of previous levelling the vector of the optimal stabilization law coeffients is equal to the following values

$$k_1 = k_2 = 5,\ k_3 = k_4 = 4,\ k_5 = k_6 = 0{,}2,$$

$$k_7 = k_8 = k_9 = 1{,}5\,.$$

IV. STRUCTURAL OPTIMIZATION

The robust structural optimization of the electronic system assigned for stabilisation of the gyroscopic nautical compass is based on solution of two Riccati equations, checking of some conditions [3] and minimization of the H_∞-norm of the mixed sensitivity function of the system which includes the plant G and the controller K. This system is characterized by the vector of outputs z, which defines the system quality, by the vectors of inputs r, controls u and observations y [3].

The modern approach to solution of the robust structural H_∞-optimization is based on shaping of the desirable frequency characteristics of the designed system.

Such approach is implemented by means of forming of the augmented plant due to connection of the weighting transfer functions.

The H_∞-norm of the mixed sensitivity function of the augmented system is used as the optimization criterion

$$J_{H_\infty S} = \left\| \begin{bmatrix} W_1 S \\ W_2 R \\ W_3 T \end{bmatrix} \right\|_\infty ,$$

where W_1, W_2, W_3 are the weighting transfer functions, S, R, T are the sensitivity function, the control sensitivity function and the complementary sensitivity function.

The statement of the structural H_∞-optimization by the method of the mixed sensitivity may be explained by the structural scheme which is represented in fig. 6.

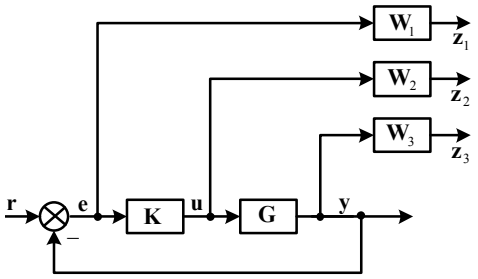


Fig. 6. H_∞-optimization by the method of the mixed sensitivity

Such instrument of the computer-aided design as the Robust Control Toolbox includes functions for the structural synthesis by the method of the mixed sensitivity. These functions (*augtf*, *hinfopt*) provide creation of the augmented model and execution of the H_∞-synthesis calculating procedure. The results of the structural optimization of the electronic system for the gyroscopic nautical compass stabilisation by the method of the mixed sensitivity are represented in fig. 7, 8.

The success of the H_∞-synthesis by the method of the mixed sensitivity essentially depends on the choice of the weighting transfer functions. In many cases the choice of these functions is carried out by the empiric methods taking unto account experience of the similar systems development.

As a rule, the controller synthesized by the method of the mixed sensitivity represents a system of the high order. There are some approaches [4] to reduction of the synthesized controllers such as reduction of the system model before the H_∞-synthesis procedure, reduction of the system model after the H_∞-synthesis procedure and use of the special methods which provide creation of the system with the reduced order.

The comparative analysis had been shown that the reduction of the synthesized controller is convenient to carry out after the H_∞-synthesis procedure termination. The reduction of the controller of the studied system may be carried out by means of the function *balmr*.

The structure of the synthesized controller reduced from 10 to 7 order looks like

$$A=\begin{bmatrix} -1{,}36 & 54{,}17 & 1{,}11 & -98{,}9 & 10{,}75 & -0{,}07 & 3{,}55 \\ 78{,}7 & -221{,}9 & 12{,}42 & 512{,}3 & -145{,}5 & 96{,}18 & -0{,}76 \\ 11{,}62 & 28{,}72 & 1{,}58 & -50{,}17 & -0{,}54 & 7{,}06 & 3{,}29 \\ -198{,}4 & 162{,}4 & -68{,}61 & -551{,}1 & 140{,}1 & -194{,}8 & -4{,}01 \\ 84{,}63 & 192{,}1 & 59{,}36 & -293{,}5 & 59{,}6 & 17{,}05 & 20{,}66 \\ -4{,}2 & -125{,}3 & -10{,}33 & 224{,}1 & -52{,}15 & -36{,}82 & -5{,}11 \\ 55{,}2 & 226{,}3 & 39{,}1 & -172{,}9 & 116{,}2 & 157{,}9 & -105{,}8 \end{bmatrix}$$

$$B^T=\begin{bmatrix} 0{,}23 & 0{,}77 & -0{,}32 & -1{,}04 & 0{,}3 & -0{,}07 & 0{,}57 \\ -7{,}33 & -104{,}4 & -13{,}36 & 208{,}8 & -55{,}92 & 14{,}23 & -87{,}19 \end{bmatrix}$$

$$C=\begin{bmatrix} 0{,}88 & 4{,}48 & 0{,}84 & -8{,}31 & 2{,}68 & -0{,}53 & 0{,}22 \end{bmatrix};$$

$$D=[0 \quad 0].$$

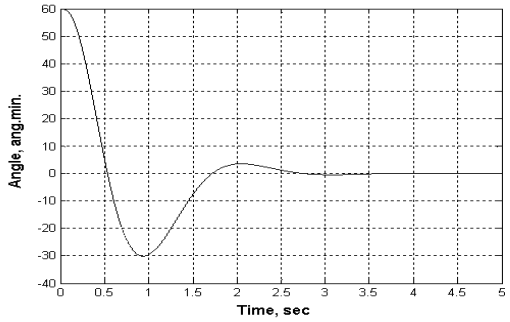


Fig. 7. The platform angle stabilisation process

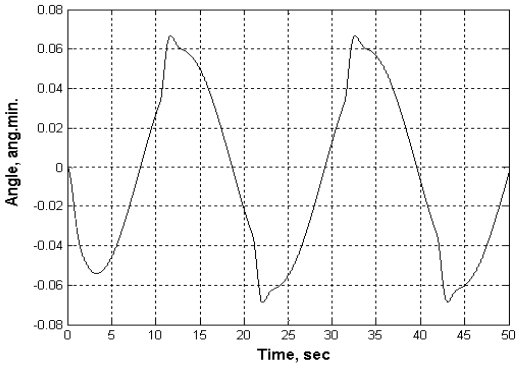


Fig. 8. The stabilisation dynamical error

CONCLUSIONS

The basic scientific results lie in consideration of the robust parametrical and structural optimization of the electronic system for the gyroscopic nautical compass stabilization including creation of the mathematical description, optimality criterion choice and stabilisation laws and controller synthesis.

Analysis of the obtained results shows that the synthesized system is resistant to action of the external disturbances in comparison with the non-optimized system. It follows from the graphs represented in fig. 2, 3 and 4, 5 correspondingly, where simulation of the stabilisation processes taking in consideration influence of the sea irregular waves is represented. The action of the irregular sea was simulated by means of the forming filters.

The obtained results (fig. 7, 8) show that the synthesized system provides high speed and accuracy of the stabilization. The small dynamic error (fig. 8) proves the high dynamic characteristics of the designed system.

REFERENCES

[1] J.M. Hilkert, "Inertially stabilized platform technology", IEEE Control Systems Magazine, no. 1, pp. 47-64, 2008.

[2] A.A. Tunik, H. Rye, H.C. Lee, "Parametric optimization procedure for robust flight control system design", KSAS International Journal, vol. 2, no 2, pp. 95 – 107, 2001.

[3] S. Skogestad, I. Postlethwaite, Multivariable Feedback Control, New York: Jonh Wiley, 1997.

[4] D. Gu, P. Petkov, M. Konstantinov, Robust Control Design with MATLAB, London: Springer-Verlag, 2005.

Real-time Implementation of the Optimal Two-Dimensional Filter for Highly Nonstationary Frequency-Modulated Signal Estimation

Veselin N. Ivanović, Nevena Radović, Zdravko Uskoković
Department of Electrical Engineering, University of Montenegro, 81000 Podgorica, MONTENEGRO,
very@ac.me, nevenar@ac.me, zdravko@ac.me

Srdjan Jovanovski
Faculty of Information Technology, Mediterranean University,81000 Podgorica, MONTENEGRO
srdjaj@t-com.me

***Abstract*— Real-time implementation of the optimal filter for highly nonstationary two-dimensional (2D) signal estimation is designed. It is based on the real-time results of 2D nonstationary signal analysis, on the correspondence of the filter's region of support (FRS) to the signal's local frequency (LF) and on the real-time LF estimation, implemented by the sliding matrix operation. Multiple FRS detection in the observed 2D signal point is provided by the proposed system, enabling very efficient real-time estimation of nonstationary monocomponent and multicomponent frequency-modulated (FM) noisy signals. The design is proven through the verification of the sliding matrix operation, that represents central part of the designed filter. As well, it is tested by estimation of the nonstationary monocomponent and multicomponent FM signals exposed to the high white noise.**

Keywords—Space/spatial-frequency filter, Filter's region of support, Local frequency, Estimation, Real-time implementation.

I. Introduction and Theory Background

Optimal nonstationary two-dimensional (2D) estimating definition, obtained by expanding the corresponding 1D definition, [1-5], related to the 2D Wigner distribution (2D WD), [1-5], and used to overcome distortion of the filtered 2D frequency-modulated (FM) signal, [5], is:

$$(Hx)(\vec{n}) = \sum_{\vec{k}=-N/2+1}^{N/2} L_H(\vec{n},\vec{k}) STFT_x(\vec{n},\vec{k}) \qquad (1)$$

where $L_H(\vec{n},\vec{k})$ is the filter's region of support (FRS, Weyl symbol of the filter's impulse response, [1-5]), $\vec{n}=(n_1,n_2)$,

$$STFT_x(\vec{n},\vec{k}) = DFT_{\vec{m}}[w(\vec{m})x(\vec{n}+\vec{m})]$$

is the 2D short-time Fourier transformation (2D STFT) of the noisy q-component signal $x(\vec{n}) = \text{sum}_{i=1,..,q}(f_i(\vec{n})) + \varepsilon(\vec{n})$, $DFT_{\vec{m}}[]$ is the operator of the discrete Fourier transformation in $\vec{m}$, $w(\vec{m})$ is real-valued 2D lag window, and $N\times N$ is the 2D signal duration. Following the procedure for the stationary Wiener filter design, [6], the optimal nonstationary 2D filter, in the case of signal not correlated with the additive noise $\varepsilon(\vec{n})$, is defined by:

$$L_H(\vec{n},\vec{k}) = 1 - \frac{E\{WD_\varepsilon(\vec{n},\vec{k})\}}{E\{WD_x(\vec{n},\vec{k})\}}. \qquad (2)$$

$E\{WD_\varepsilon(\vec{n},\vec{k})\}, E\{WD_x(\vec{n},\vec{k})\} = E\{WD_f(\vec{n},\vec{k})\} + E\{WD_\varepsilon(\vec{n},\vec{k})\}$ are cross-terms-free expected values of 2D WDs of noise $\varepsilon(\vec{n})$ and noisy signal $x(\vec{n})$, respectively. In the case of wide class of nonstationary FM signals $f_i(\vec{n}), i=1,..,q$, highly concentrated in the space/spatial-frequency (S/SF) area around their local frequencies (LFs), and of a widely spread white noise, the FRS (2) corresponds to the combination of LFs of signals $f_i(\vec{n})$, [5]. Then, the optimal filtering of these signals can be reduced to the LF estimation in a noisy environment. This approach can be applied to the highly concentrated 2D FM signals, such as pattern images (i.e., interferograms in optics). Therefore, here we will test our design on this kind of 2D signals.

In practice, the LF estimation has to be performed based on a single realization of noisy signals. In the S/SF analysis framework, this is done by determining frequency-frequency (FF) points $\vec{k}_i, i=1,...,q$ where S/SF distribution (S/SFD) of the noisy 2D signal has local maximum, [5],

$$LF_i(\vec{n}) = \arg[\max_{\vec{k}\in Q_{\vec{k}_i}} S/SFD_x(\vec{n},\vec{k})]. \qquad (3)$$

$Q_{\vec{k}_i}$ is the basic region in FF plane around $f_i(\vec{n})$, the LF of which is $LF_i(\vec{n})$. Recently defined 2D cross-terms-free WD (2D CTFWD), [7], preserves optimal (2D WD) auto-terms representation, but without cross-terms presence. Additionally, in comparison to the commonly used S/SFDs, it minimizes calculation complexity and simultaneously optimizes LF estimation characteristics, [8]. Besides, 2D CTFWD is defined based on the same 2D STFT elements used in (1) and has been already implemented in real-time, [8].

Real-time design of an optimal nonstationary 2D filter (1) will be developed here to estimate nonstationary 2D FM signal exposed to the high white noise. It will be based on the available 2D CTFWD real-time design, developed in [7, 8] to provide the improved S/SF representation of input 2D noisy signal, and on

the real-time algorithm for the 2D CTFWD-based LF estimation, implemented through the sliding matrix operation. The implemented algorithm includes, as a special case, the corresponding algorithm used in the 1D signal case, [9].

II. IMPLEMENTATION

The 2D CTFWD-based Wiener filter real-time design, principally following (1), is presented in Fig.1. Input 2D STFT elements of the design (STFT_IN elements in Fig.1), are produced outside, in well-known 2D STFT, [5, 10], or 2D FFT, [6], modules. They are imported to the convolution window operation file by each STFTLoad/CTFWDStore cycle. The 2D STFT-based 2D CTFWD elements are produced in the already available STFT-to-CTFWD gateway, [8]. By each STFTLoad/CTFWDStore cycle, they are moved into the FRS detection module, implemented to generate $L_H(\vec{n},\vec{k})$ by following the algorithm, graphically shown in Fig.2. The algorithm tests an LF existence in the FF point (k_1,k_2), k_1,k_2= $-N/2+1,\ldots,N/2$, for the observed signal point (n_1,n_2), by grouping frequency-only-dependent 2D CTFWDs, symmetrically distributed around (k_1,k_2), $CTFWD_x(\vec{n},k_1+L,k_2+L)$, …, $CTFWD_x(\vec{n},\vec{k})$,…, $CTFWD_x(\vec{n},k_1-L,k_2-L)$, into matrix Q, implemented through the sliding matrix register block, sized $(2L+1)\times(2L+1)$ locations, Fig.1. In this way, the basic region $Q_{\vec{k}}$ (from (3)) is created, enabling LF estimation. The LF is detected in the FF point (k_1,k_2) corresponding to the maximum matrix element, but only if the maximum matrix element is:

(i) the central matrix element, $CTFWD_x(\vec{n},\vec{k})$,

(ii) greater than the introduced spectral floor R, and

(iii) if the matrix size satisfies:

$$2\cdot\max_{1\le i\le q}\{A_i\}\le 2L+1<\sqrt{2}\cdot\min_{1\le i,j\le q,i\ne j}|LF_i(\vec{n})-LF_j(\vec{n})| \quad (4)$$

where A_i, $i=1,\ldots,q$ are different widths of the non-overlapping 2D CTFWD auto-terms.

The condition (i) combined with the first inequality from (4) has to be met to ensure that:

(a) All points (k_1,k_2) from the observed auto-term, including the true LF, have the corresponding 2D CTFWDs inside the matrix Q when the LF existence in each of these points is investigated, making the minimal (noise-only-dependent) estimation error inside the 2D CTFWD auto-terms' domains;

(b) For each auto-term and each signal point (n_1,n_2), only one point (k_1,k_2) can be detected as an LF, reducing the influence of the frequency discretization on the estimation quality, [5].

The condition (i) combined with the second inequality from (4) ensures detection of an LF of each signal component separately in the observed signal point (n_1,n_2), enabling the FRS estimation in the multicomponent signal case.

The condition (ii) has to be met to suppress the noise influence outside the 2D CTFWD auto-terms' domains. Spectral floor, a few percent of the maximum 2D CTFWD value, is selected based on a priory knowledge about the digital signal's range, determined by the optimal usage of the available input A/D converter and memory locations, [8].

After the execution in the point (k_1,k_2), this procedure is repeated for the next FF point (k_1,k_2+1) from the same signal point (n_1,n_2). The $CTFWD_x(\vec{n},k_1+L,k_2+L+1)$ element (CTFWD signal in Fig.1) is imported to the matrix Q area, shifting the existing matrix elements for one position right. In this way, matrix Q slides for one position right over frequency-only-dependent 2D CTFWDs, creating the next basic region $Q_{(k_1,k_2+1)}$, as principally illustrated in Fig.2 and, for a simple real situation, presented in Fig.3. $2L$ FIFO delays from the sliding matrix operation file (Fig.1) provide real-time sliding of the matrix Q and, therefore, the real-time FRS detection.

In general, for each signal point (n_1,n_2), the real-time FRS detection is performed by sliding matrix Q over all frequency-only-dependent 2D CTFWD input elements and by FRS computing according to the input elements and the described algorithm. The COMP block tests conditions (i)–(ii) of the algorithm and, if they are satisfied, detects an FRS in $(\vec{n},\vec{k})$, determined by $L_H(\vec{n},\vec{k})=1$. $L_H(\vec{n},\vec{k})=1$ enables inclusion of the $STFT(\vec{n},\vec{k})$ element in the output signal generation through the summation into output cumulative adder. By the next STFTLoad/CTFWDStore cycle (after the $L_H(\vec{n},\vec{k})$ value calculation), the $STFT(\vec{n},k_1+L,k_2+L+1)$ element is imported into the convolution window operation file. Therefore, the FIFO delay block from Fig.1, sized NL_m+L_m+1 locations, is used to hold the imported 2D STFT data, such that it's output corresponds, in frequencies, to the central element of the sliding matrix register block.

Wide frequency range (4), obtained in the case of highly concentrated, non-overlapping 2D FM signals, implies the robustness of the FRS estimation with respect to the L value, that is proven by the simulation, Table 1. Then, application of the matrix Q of several locations is usually sufficient.

Execution time of the proposed design is determined by the cycle of the STFTLoad/CTFWDStore signal, produced in the STFT-to-CTFWD gateway, [8]. In this way, the proposed design retains calculation speed of the STFT-to-CTFWD gateway without introducing additional time delays.

III. VERIFICATION OF THE SLIDING MATRIX FUNCTION

The basic part of the architecture shown in Fig. 1 is the sliding matrix area implementation. It is used twice. The first time it is named convolution window operation file and is used to provide movement through the STFT samples and to create address order of the STFT-to-CTFWD gateway's inputs. The second time it is named sliding matrix operation file and is used to provide movement through the CTFWD samples, to create address order of the COMP block's inputs, and to enable the FRS detection (the signal FRS_k generation) for the observed

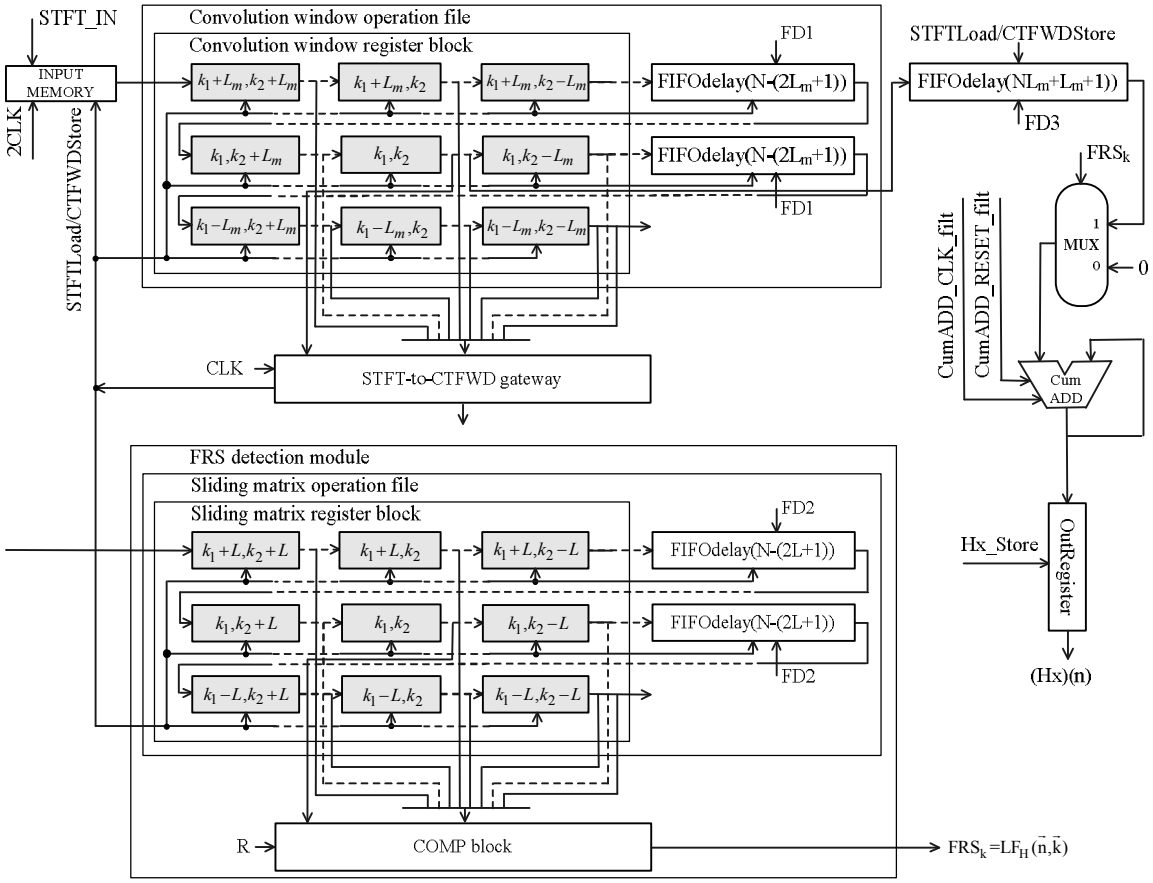


Fig. 1. Optimal nonstationary 2D filter real-time design.

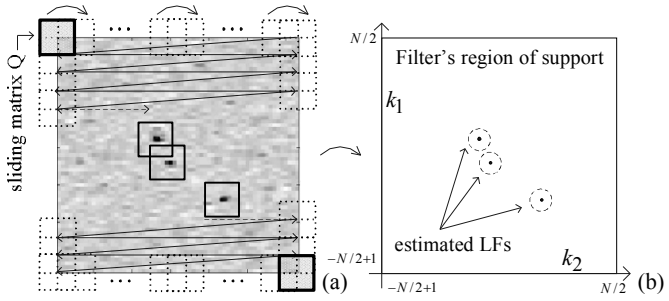


Fig. 2. Procedure of the matrix Q sliding (a) and of the FRS producing (b).

1.0	-3.05	6.2	1.1	7.8	15.0	-3.2
4.4	-8.0	16.0	-13.2	3.0	19.2	8.9
-4.0	7.3	6.22	8.18	-2.09	5.12	-3.5
8.2	19.0	-7.2	16.8	1.0	-8.2	9.2
3.0	5.2	21.9	6.0	14.2	7.1	-3.0
-2.04	3.2	9.5	-5.1	8.2	-5.0	-9.0
-3.19	-6.5	-3.5	12.9	7.4	7.3	3.1

Fig. 3. Test CTFWD matrix. Sliding of the 3×3 matrix is illustrated.

time and frequency instant. Other parts of the architecture from Fig.1 are already implemented (STFT-to-CTFWD gateway in [11, 12]), or are well known from the literature (such as STFT module, comparators, cumulative adders etc.). This is the main reason why we will explain the sliding matrix function in details and present the results of its simulation via ModelSim Altera 6.3g_p1 software.

Sliding matrix operation file consists out of sliding matrix register block and 2*L* FIFO delays. Registers from the sliding matrix register block contains the 2D CTFWDs, created in the STFT-to-CTFWD gateway. As seen in Fig.1, CTFWD signal is loaded in the first element (register) of the sliding matrix register block. At the same time, each element of the sliding matrix register block row k_1+L (as well as rows k_1+L-1, $k_1+L-2,\ldots,k_1-L$) will be shifted by FIFO registers, to generate the 2D CTFWD elements in time index ($k_2+L-1,k_2+L-2,\ldots$, k_2-L). The 2*L* FIFO delays are used to generate the 2D CTFWDs of the sliding matrix register block column k_2+L in time index (k_1+L-1,k_1+L-2, …, and k_1-L). It means that, for each described step, a new 2D CTFWD element will be shifted into the sliding matrix register block area, that imply sliding of the actual window from Fig.2 for one position right.

The proposed design is verified by the real-time simulation of the 2D CTFWD test matrix given in Fig.3. Fig.4(a),(b) demonstrates the FRS detection for the observed test matrix. To enable representation and understanding of the simulation results, the small matrix dimensions of $N\times N=7\times 7$ and the small sliding matrix areas of $(2L+1)\times(2L+1)=3\times 3$ (Fig.4(a)) and of 5×5 (Fig.4(b)) are observed. Note that the derived conclusions remain valid for arbitrary *N*, *L* values.

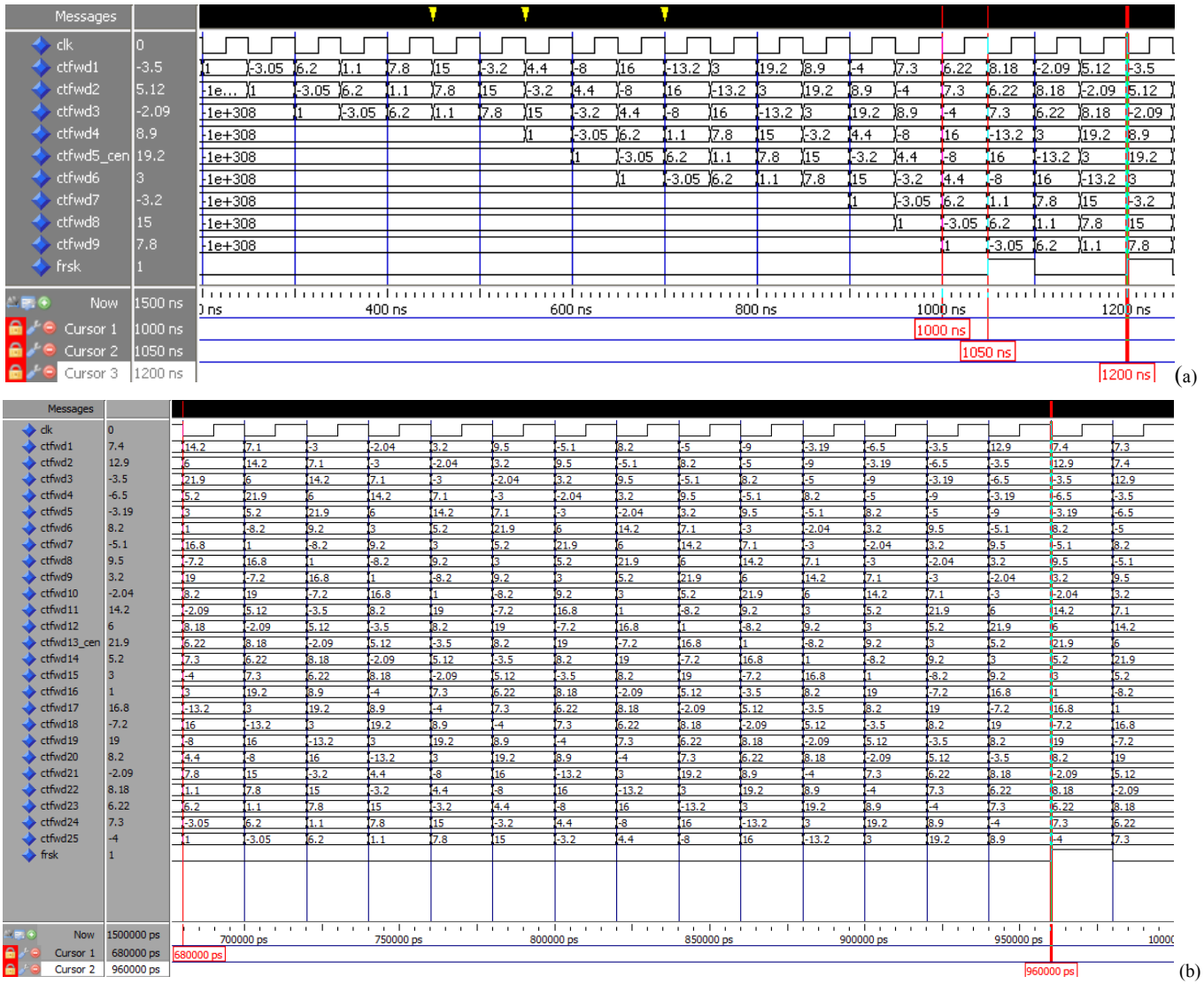


Fig. 4. Results of the FRS detection for the test CTFWD matrix, implemented in real-time and presented for (a) $2L+1=3$ and (b) $2L+1=5$. Cursor 1 denotes the time instant when the first data window is read in. Cursors 2 and 3 respectively denote the time instants when the first FRS and the second FRS are detected.

IV. Numerical Example

The verification of the proposed design is performed on the real life nonstationary 2D signals. In that order, monocomponent and 3-component 2D test signals:

$$f(n_1T,n_1T)=1.4e^{j(96\pi((n_1T)^2+(n_2T)^2))}e^{-1.5((n_1T)^2+(n_2T)^2)}, \quad (5)$$

$$\begin{aligned} f(n_1T,n_1T)=&\cos(-100\cos(\pi x/2)\\ &+100\cos(\pi y/2))\\ &+1.2\cos((30\pi(x+2)^2\\ &+30\pi(y+2)^2))+1.5\cos(96\pi x^2), \end{aligned} \quad (6)$$

have been considered within the time intervals $0.1\le n_1T,n_2T\le 0.6$ and $0.4\le n_1T,n_2T\le 0.9$, respectively. They are masked by high white noises such that the corresponding input signal-to-noisy ratios (SNRs) are $SNR_{in}=10\log(P_f/P_\varepsilon)=$–1.42[dB] and $SNR_{in}=10\log(P_f/P_\varepsilon)=$–0.46[dB], respectively. A sampling interval of $T=0.5/128$, the Hanning lag window, the signals' duration of $N\times N=64\times 64$, the siding matrix Q area of 9×9, the spectral floor of $R=0.1\cdot\max_{\vec{n},\vec{k}}\{CTFWD_x(\vec{n},\vec{k})\}$, and the reference level used in the 2D CTFWD calculation, [7], of $0.05\cdot\max_{\vec{n},\vec{k}}\{|STFT_x(\vec{n},\vec{k})|^2\}$ are applied.

The S/SF filtering is performed by using the proposed real-time design. Results are presented in Figs.5, 6. Output SNRs of $SNR_{out}=36.56$[dB] and of $SNR_{out}=29.29$[dB], have been obtained in the observed cases, respectively. Corresponding SNR improvements, achieved in these cases, are SNR_{out}–$SNR_{in}=35.14$[dB] and SNR_{out}–$SNR_{in}=29.75$[dB], respectively.

The achieved improvements are very high, since maximum SNR improvements of $10\log(N\times N/1)=36.12$[dB] and of $10\log(N\times N/3)=31.35$[dB] can be (but, only theoretically) achieved in the monocomponent and 3-component signal cases, respectively.

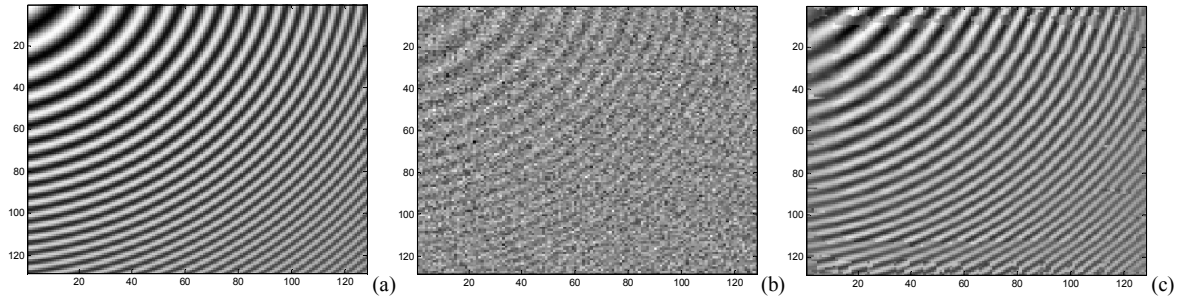


Fig. 5. (a) Non-noisy monocomponent test signal (5), (b) Noisy signal (5), (c) Output signal of the proposed S/SF filtering design.

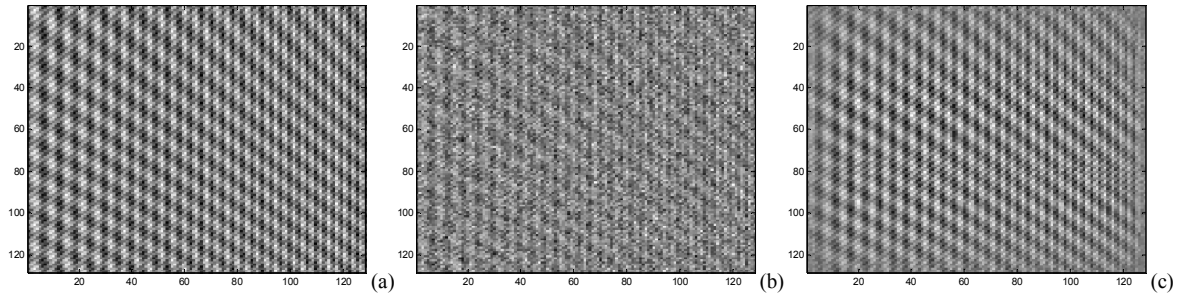


Fig. 6. (a) Non-noisy 3-component test signal (6), (b) Noisy signal (6), (c) Output signal of the proposed S/SF filtering design.

TABLE I. OUTPUT SNRS AND CORRESPONDING IMPROVEMENTS (IMPS) FOR DIFFERENT INPUT SNRS AND DIFFERENT VALUES OF SLIDING MATRIX Q SIZE FOR 3-COMPONENT TEST SIGNAL (6) FILTERING

SNR_{in} [dB]	$2L+1=5$		$2L+1=7$		$2L+1=9$	
	SNR_{out} [dB]	IMP [dB]	SNR_{out} [dB]	IMP [dB]	SNR_{out} [dB]	IMP [dB]
1.00	31.15	30.15	31.34	30.34	31.77	30.77
0.00	28.28	28.28	29.03	29.03	30.11	30.11
-1.00	24.67	25.67	25.38	26.38	25.91	26.91

V. COMPARATIVE ANALYSIS AND CONCLUSION

The proposed design increases complexity of the 2D CTFWD real-time design by additional inclusion of only $(2L+1)^2+2L^2+2L+1$ memory locations, a cumulative adder, and a set of comparators. In this way, the design retains characteristics of the 2D CTFWD real-time design, [8], regarding several critical design performances (calculation and implementation complexity, as well as the execution time).

Our design enables real-time S/SF filtering. In comparison to this, the available S/SF filtering approach [5], based on a statistical method that compares biases and variances of S/SFDs with extremely different numbers of filtered signal elements, requires quite computationally complex and highly time consuming FRS detection and, therefore, cannot be applied in real-time. Further, the proposed design provides high quality filtering, based on the highest quality signal adaptive 2D CTFWD-related LF estimation. The designs based on non-adaptive algorithms, as well as the stationary 2D filters, cannot produce so high quality results, [1-5].

REFERENCES

[1] G. Matz and F. Hlawatsch, "Linear time-frequency filters: Online algorithms and applications", in Applications in Time-Frequency Signal Process. (A. Papandreou-Suppappola, Ed.), CRC Press, 2002, pp.205-271.

[2] F. Hlawatsch, G. Matz, H. Kirchauer, and W. Kozek, "Time-frequency formulation, design and implementation of time-varying optimal filters for signal estimation", IEEE Trans. Signal Process., Vol. 48, No. 5, pp. 1417-1432, May 2000.

[3] R.G. Shenoy and T.W. Parks, "The Weyl correspodence and time-frequency analysis", IEEE Trans. Signal Process., Vol. 42, No. 2, pp. 318-331, Feb. 1994.

[4] G. Matz and F. Hlawatsch, "Linear time-frequency filters", in Time-Frequency Signal Analysis and Processing, (B. Boashash, Ed.), Pretice Hall, 2002.

[5] LJ. Stanković, "Space/spatial-frequency analysis based filtering", IEEE Trans. Signal Process., Vol. 48, No. 8, pp. 2343-2352, Aug. 2000.

[6] A. Papoulis, Signal Analysis, McGraw-Hill, New York, USA, 1997.

[7] V. N. Ivanović and S. Jovanovski, "Signal adaptive method for improved space/spatial-frequency representation", Electronics Letters, Vol. 45, No. 19, pp. 1003-1005, Sept. 2009.

[8] V. N. Ivanović and S. Jovanovski, "Signal adaptive system for space/spatial-frequency analysis", EURASIP J. Advances in Signal Process., vol. 2009, pp. 1-15, 2009.

[9] S. Jovanovski and V. N. Ivanović, "Signal adaptive pipelined hardware design of time-varying optimal filter for highly nonstationary FM signal estimation", J. Signal Process. Syst., Vol. 62, No. 3, pp.287-300, 2011.

[10] K.J.R. Liu, "Novel parallel architectures for short-time Fourier transform", IEEE Trans. CAS II, Vol.40, No.12, pp.786-790, Dec.1993.

[11] V. N. Ivanović, R. Stojanović and LJ. Stanković, "Multiple clock cycle architecture for the VLSI design of a system for time-frequency analysis", EURASIP J. Appl. Signal Process., Special Issue Design Methods for DSP Syst., pp. 1-18, 2006.

[12] V.N. Ivanović and S. Jovanovski, "A signal adaptive system for time-frequency analysis", Electronics Letters, Vol. 44, No. 21, pp. 1279-1280, Oct. 2008.

Electronic Hardware Design of a Low Cost Tactile Sensor Device for Physical Human-Robot Interactions

Ganna Pugach, Viacheslav Khomenko, Artem Melnyk
Electrical Engineering Department,
National Technical University of Donetsk,
Donetsk, Ukraine
anna.pougatch@gmail.com
slava.khomenko@gmail.com
artemmelnyk@gmail.com

Alexandre Pitti, Patrick Henaff, Philippe Gaussier
ETIS, UMR 8051, CNRS-UCP-ENSEA,
University of Cergy-Pontoise,
Cergy-Pontoise, France
alexandre.pitti@ensea.fr
patrick.henaff@u-cergy.fr
gaussier@ensea.fr

***Abstract*—This paper proposes a low-cost system, based on the method of Electrical Impedance Tomography (EIT), for data acquisition from soft conductive fabric, for the purposes of designing of robots artificial skin. A simple multiplexer/ demultiplexer circuit is used for retrieving the resistance field from the pair-wised electrodes which inject the electrical current and the electrodes which measure the output voltage from the conductive fabric. A microcontroller governs the injection of current, voltage output patterns and the analog-digital conversion, from the tactile material. After explanation of the EIT method, the electronics corresponding to the data acquisition is presented and the material characteristics are analyzed. The results show that the spatial patterns of the tactile contact can be acquired and localized in real time.**

Keywords—tactile sensors, artificial skin, Electrical Impedance Tomography, EIDORS.

I. INTRODUCTION

With respect to vision and audition modalities, tactile modality is now considered to be an important feature to achieve in robots in order to enable them with intelligent and adaptive behaviors like grasping an object or touching people [1]. This aim in mind, nowadays hardware technology permits the development of new artificial skins devices [2–4]. Combined with the visual and auditory senses and proprioception, artificial skins would provide additional sensory perception to robots and demultiply their capabilities.

Several techniques have been proposed for which most of them are using isolated unitary devices that cover the specified surface and for which each is reacting to a small receptive field [5–8]. Another promising method is the use of large conductive fabric sheets that reacts to electrical variations on its surface. Using inversion methods like the well-known Electrical Impedance Tomography (EIT), it is possible to reconstruct the spatial location on its surface [9]. Compared to more classical methods, this technique is (1) inexpensive in the design, (2) allows to cover large surface portions, (3) allows to be nearest the design of the human-skin. Despite these advantages, fast acquisition and cheap reconstruction algorithms are still a drawback [10, 11].

The paper is organized as follows. The second section describes the current state of art in artificial skins based on imaging reconstruction methods and presents the Electric Impedance Tomography (EIT) method. The third section explains the design and the test results from our robotic tactile feedback device. Finally, in the fourth section, draws a conclusion and our future works.

This work is supported by the French/Ukrainian Dnipro Project 24316WA, the French embassy in Ukraine (program "master in partnership"), the INTERACT French project ANR-09-CORD-014.

II. STATE OF ART IN ARTIFICIAL SKINS BASED ON IMAGING RECONSTRUCTION METHODS

Kato et al. [12], Nagakubo et al. [13] realized a "skin"–like device for robots. They applied the Electrical Impedance Tomography (EIT) method and inverse problem analysis (IPA) to sensitive skin: by injecting currents and measuring voltages from connected electrodes on the borders of a rubberized conductive fabric, they reconstructed the local resistivity changes response from any applied pressure on the material.

The IPA enables to have a non-invasive estimation of the internal state of an object by performing measurements only on the object boundaries. These inverse techniques are implemented in different technologies: Computed tomography (CT) [14] and CT Scan (CTS) for X-rays analysis, Magnetic Resonance Imaging (MRI) [15], Positron Emission Tomography (PET) [16] to detect pairs of gamma rays emitted as a result of the collision between positron-emitting radio-isotope, Electrical Capacitance Tomography (ECT) [17] and Electrical Impedance Tomography (EIT) [18] that estimate the capacitance and impedance distribution inside a conductive material by injecting currents from different locations on the boundary and measuring the resulting electric potential distribution on it.

A. Electrical Impedance Tomography

The EIT aims at reconstructing a 2D or 3D cross-sectional image of the internal spatial distribution of conductivity from electrical measurements by injecting small currents (DC or AC) via electrodes array placed on the surface of the investigated object. The advantages of EIT compared to conventional imaging techniques such as CT or MRI is that (1) it is considered safer by merit of the small alternating currents required; (2) it can be used as a long-term, continuous imaging method; and (3) the system can be constructed at low-cost and in a very portable size [19].

The EIT method has been applied to many domains. In medicine, it has been used for gastric emptying and pH measurements [20], cancer detection [21] and tissue characterization [22]. In industry, it has been used for the non-destructive testing of polycrystalline diamond [23], 3D malformations detection in industrial pressure filtration systems [24]. In geophysics, it has been used for measuring the polarization properties of soils and sediments [25] and the analysis of gravel-air-seawater mixture [26].

Different EIT reconstruction methods was proposed. The neighboring method [27], applies a current to neighboring electrodes and measures the voltage from the other electrode pairs. Fig. 1 depicts the idea of this method for a cylindrical volume conductor with 16 symmetrically spaced electrodes. The current is applied to electrodes pair 1-2, for which the current density is now highest in the material between them. The current density is decreasing rapidly as a function of distance.

The voltage difference is measured between the other pairs of electrodes, which corresponds to 13 measurements, for reconstructing the impedance between the equipotential lines that intersect the measurement electrodes. At the next round, we inject the current to the electrodes pairs 2-3, and measure the successive set of 13 potential differences, till the end of the sequence.

Another method, named cross method [28] does not have sensitivity in the periphery as good as the neighboring method, but it has better sensitivity over the entire region. The opposite method for the impedance measurement [28] gives more uniform current distribution and has a good sensitivity.

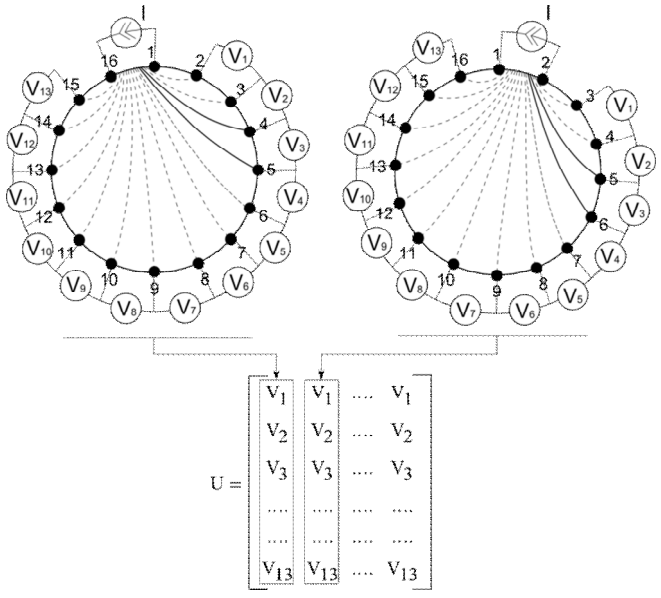


Fig. 1. Neighboring method of data collection show for a cylindrical volume conductor and 16 symmetrically space electrodes.

In order to reconstructed sensorial image from EIT, several methods were proposed such as the back-projection method, the layer-stripping method and various iterative methods [29]. The most popular method for image reconstruction is the Newton's iterative method [30].

B. *Image reconstruction with Newton's iterative method*

The goal of the Newton's method is to minimize the error between the voltages V_i obtained by forward modeling and those V_m measured from the conductive sensor. The forward model is obtained from the Maxwell's equation and from a Finite Element Method (FEM) model of the sensor's body (see [30] for more details):

$$V_i = F(\gamma_i, c) \quad (1)$$

γ_i is the resistance distribution and c the injected current.

A Jacobian Matrix represents the effect on the boundary voltages for a slight change in the resistivity of the FEM elements. A resistance distribution allows to calculate the Jacobian Matrix and to use it to estimate accurately the resistance distribution. An iterative process, starting from an initial resistance distribution, converges towards the true resistance distribution. Equation of the i^{th} step of the method are as following:

$$\min\|V_m - V_i\|^2 + h^2\|R\gamma\|^2 \quad (2)$$

$$\text{where } \gamma_{i+1} = \gamma_i + \delta\gamma_i \quad (3)$$

$$\text{and } \delta_{\gamma i} = \left(J_i^T J_i + h^2 R^T R\right)^{-1} J_i^T (V_m - V_i) \quad (4)$$

$$J_i = \frac{\partial F}{\partial \gamma_i}$$

is the gradient of the forward model with respect to the resistance distribution, and $h^2\|R\gamma\|^2$ is a regularization term introduced to stabilize the response. h defines the smoothness of the estimated resistance distribution. R is the Laplace image prior or the unity matrix.

III. PROPOSED TACTILE SENSOR BASED ON VELOSTAT™

Conductive rubbers are conceived with carbon as conductive particles. The rubber material used in this paper is the 3M™ Velostat™ [31]. The initial resistance of the material is approximately 500 Ohms/m and it is sensitive to both pressure and stretch. Nonetheless, since the material is not extremely deformable, the sensibility towards pressure is quite feeble. Those materials have been applied in tactual sensing during a period of 10-20 years. This is the most common used material when examining EIT-based tactile distribution sensors.

C. *Analysis of rubber sensitivity*

In the case of most conductive rubber, at the time of load removal, the conductivity of the region overshoots to a peak and then restores to the original value. Especially, fast load removal causes larger hysteresis while a slower load removal causes smaller hysteresis. The dynamic characteristics of conductive rubber with respects to fast or slow pressure removal are less well investigated.

The response of an 80 mm x 25 mm Velostat™ conductive sheet connected to 4 electrodes is investigated, with one pair for current injection and the other for voltage measurement (Fig. 2).

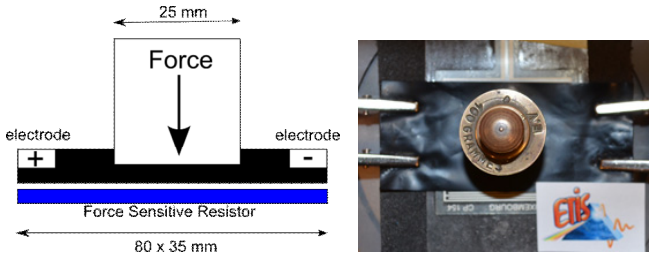


Fig. 2. Experimental device to measure tissue properties in physical contact with the object.

The force at the center of the material was applied, and it's voltage response was measured. The experiments were performed with conductive weights (Fig. 3) and with non-conductive weights (Fig. 4).

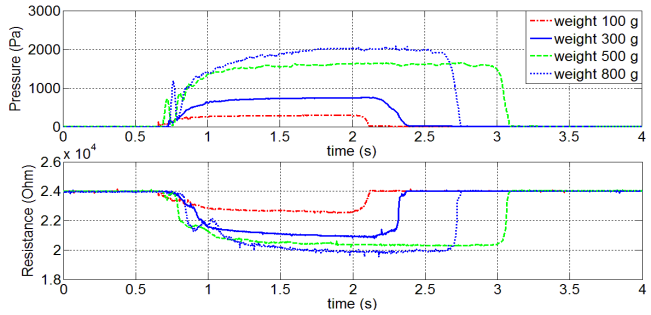


Fig. 3. Properties of the conductive rubber material to various conductive weights. Top: Pressure in [Pa]; Bottom: Resistance in [Ohm].

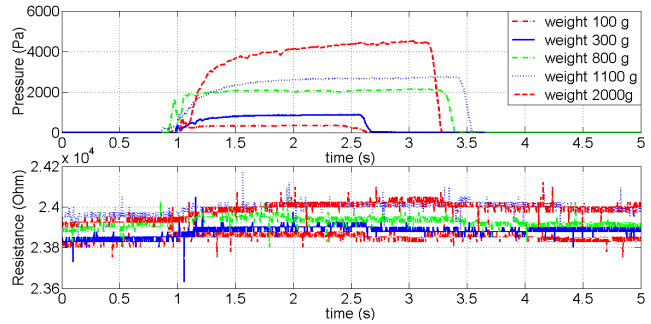


Fig. 4. Properties of the conductive rubber material to various non-conductive weights. Top: Pressure in [Pa]; Bottom: Resistance in [Ohm].

Various conductive objects of different weights were placed on the conductive material and its resistivity was measured (Fig.3). It was observed that the conductive material responds almost proportionally to the weights (Fig. 5), which permits to discriminate distinctly various resistance profiles.

In comparison to its previous behavior, the conductive rubber weakly for the case of non-conductive objects responds very poorly. Here, its resistance varies only slightly to different pressure values and no clear discrimination is possible (Fig.4). Therefore, it can be concluded that the conductive rubber works in a quasi-linear regime with the possibility to discriminate the pressure applied by conductive objects. For the case of non-conductive objects, conductive rubbers possess a high hysteresis which does not make them the most suitable material because its resistance does not return back steadily to its initial value.

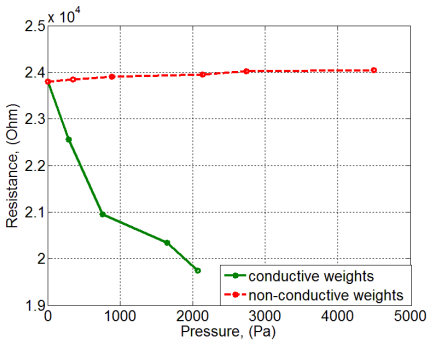


Fig. 5. Dependence of the resistance on the pressure.

Finally, one conductive layer (for example conductive aluminum laminated fabric) was added between a non-conductive object and rubber, and the same type of the characteristics (Fig.5.) as in the interaction with a conductive object was obtained.

D. Data acquisition principle

Fig. 6 illustrates the components of a basic EIT system. The system typically has 16 electrodes placed on the boundary of the conductive material. The device is used to inject current and measure the voltages. It then sends a data frame to the PC where the resistance distribution of the material is estimated.

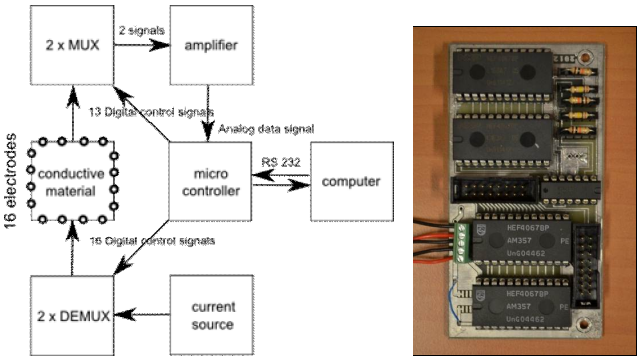


Fig. 6. Architectural overview of the proposed wireless impedance analyzer supporting 16 data acquisition channel.

The system developed by our team comprises a base-station and a scan-head. The scan-head has a double demultiplexers circuit used for applying the excitation current between any 2 electrodes and the double multiplexers circuit used for switching between other 13 electrodes of tissue surface for connect to the base-station. If the injected current is a DC current, then the electric potentials are measured on all of the boundary electrode pairs. If the injected current is an AC current, then the voltage amplitude and lag (relative to the AC current) are measured. Also the scan-head has an instrumental amplifier for the current source.

The base-station has a 8-bit micro-controller for data sampling using a 10-bit analog-to-digital converter (ADC). It has also a versatile multiplexor interface with a wireless impedance analyzer to offer 16 independently addressable channels to which the current can be injected and to which the sensing of the boundary potential can be simultaneously measured.

Although many methods can be used among the different measurement strategies [27, 28], the neighboring method was chosen for its robustness. Our system was originally intended for 35 Hz sampling. Moreover, preliminary studies on the material have shown the feasibility to work with currents from 50 to 200 µA.

Despite its benefits, one of the disadvantage of the neighboring method is that it is impossible to provide a very high sensitivity since the potential difference between the electrodes near the current source is very high (up to 0.5V for a given maximum current of 200 µA), while the potential difference between electrodes which are far from the current source is rather small (20 mV). This drawback limits the maximum rate of the differential amplifier about ten times its sensitivity.

E. Experimental results

The spatial location of the conductive weights and the resistance density distribution was reconstructed with the MATLAB™ toolbox EIDORS (Electrical Impedance and Diffused Optical Reconstruction Software) [32]. This software is composed of three software for mesh generation, forward problem and the inverse problem [33].

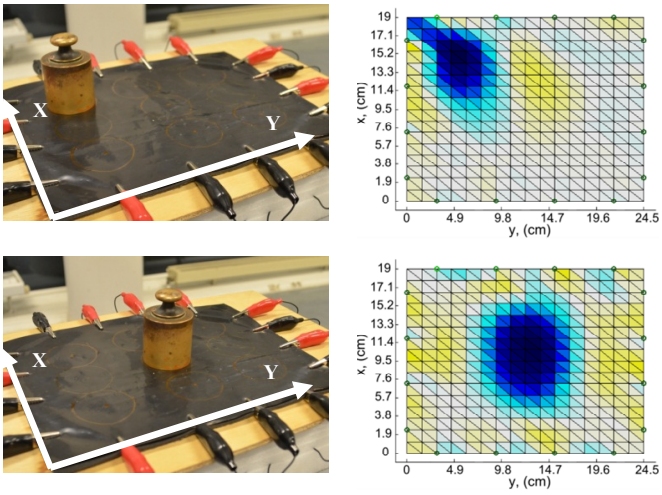


Fig. 7. Image reconstruction with use EIDORS.

The experimental setup for two different locations in the rubber surface in coordinate (X,Y) and the image reconstruction for the top-left corner and the center position are depicted in Fig. 7. This figure shows respectively the boundary effects on the surface and the localization error on the center.

In order to measure the accuracy of the EIT image reconstruction, the relative localization error in centimeters (relative to the X and Y axis) is plotted on Fig. 8. The {green/blue/red} colors indicate the relative errors depending on the weight location and the relative distance to the X or Y axis; resp. 5.5cm/12.5cm/19cm. As one can observe, the maximum localization error on the fabric is less than 3.5% at the center position, which means that the software EIDORS can reconstruct with high accuracy, with approximately a distance error of 7mm. On the borders, the error diminishes to less than 1mm of error.

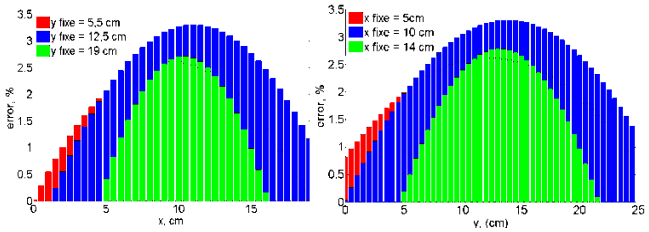


Fig. 8. Relative error of the determination of the position of the object on the conductive fabric.

CONCLUSIONS AND PERSPECTIVES

In this paper, a low-cost method of EIT data acquisition from soft conductive fabric for the design of a robotic artificial skin was presented. The electronic designed data allows realizing the neighboring method of impedance tomography. The advantages are following (1) excellent noise characteristics due to the high proportion of the load impedance (10 ohms) to the internal resistance of the multiplexers (not more than 10 Ohm), (2) high performance data, (3) low power consumption, which is a priority for self-contained use of the data, (4) low cost, high reliability and integration.

The classical reconstruction approaches for estimating the resistivity distribution on the tactile sheet were used. Experiments showed some good performances for accurately determining the contact surface with an object. The research done on the conductive rubber Velostat™ characteristics showed that its resistance varies only slightly with a non-conductive object, which requires increasing the material's sensitivity.

In future applications, increased spatial resolution with machine learning methods will be investigated for more spatial accuracy to the material properties. In EIT method, for instance, the reconstruction is analytical, which requires (1) to know exactly the tissue dimension and (2) a compute time costly matrix inversion. Learning algorithms, such as neural networks, will be considered to provide more robustness and more adaptivity to spatial localization (independently on its dimension) and to provide faster response for estimating the resistance distribution on the tissue's surface.

Furthermore, the bandwidth resolution of the acquisition device is limited by the high voltage range retrieved from the electrodes, which is in the interval [10^{-3} mV, 10^{0} mV]. New electronic devices should investigate how to reduce this range for higher precision.

ACKNOWLEDGMENT

The authors thank to Pr. V. Borissenko, prof. Kovaliev A., National Technical University of Donetsk, Ukraine.

REFERENCES

[1] R. Alami, A. Albu-Schaeffer, A. Bicchi, R. Bischoff, R. Chatila, A. De Luca, A. De Santis, G. Giralt, J. Guiochet, G. Hirzinger, F. Ingrand, V. Lippiello, R. Mattone, D. Powell, S. Sen, B. Siciliano, G. Tonietti and L. Villani, "Safe and Dependable Physical Human-Robot Interaction in Anthropic Domains: State of the Art and Challenges", Proc. IROS'06 Workshop on pHRI - Physical Human-Robot Interaction in Anthropic Domains, 2006.

[2] H. Alirezaei, A. Nagakubo, and Y. Kuniyoshi, "A tactile distribution sensor which enables stable measurement under high and dynamic stretch," IEEE Symp. 3D User Interfaces, pp. 87–93, 2009.

[3] M.Inaba, Y.Hoshino, K.Nagasaka, T.Ninomiya, S.Kagami, and H.Inoue, "A full-body tactile sensor suit using electrically conductive fabric and strings," In Proc. of the RSJ/IEEE Int. Conf. on Intelligent Robots and Systems vol.2, pp. 450- 457, November 1996.

[4] Y.Ohmura, Y.Kuniyoshi and A.Nagakubo, "Conformable and Scalable Tactile Sensor Skin for Curved Surfaces," in Proc. of Int. Conf. on Robotics and Automation,vol. 6, pp.1348-1353,May 2006.

[5] H.-K. Lee, S.-Il Chang, and E. Yoon, "A flexible polymer tactile sensor: fabrication and modular expandability for large area deployment," Jour. Microelectromec. Sys., vol. 15, pp. 1681-1686, Decembre 2006.

[6] E. Meister, I. Zilberman, and P. Levi, "Fuzzy Logic Based Sensor Skin for Robotic Applications," In Proc. of Intl. Conf. on Artificial Intelligence, Las Vegas, USA, July 2012.

[7] X. Lamy, F. Colledani, F. Geffard, Y. Measson and G. Morel, "Robotic skin structure and performances for industrial robot comanipulation," IEEE/ASME Intern. Conf. Advanc. Intellig. Mechatron., vols.1-3, pp.427-432, July 2009.

[8] G. Cannata, M. Maggiali, G. Metta and G. Sandini, "An embedded artificial skin for humanoid robots," Proc. of the IEEE Intl. Conf. on Multisensor Fusion and Integration for Intelligent Systems, pp. 434-438, August 2008.

[9] D. S. Tawil, D. Rye and M. Velonaki, "Touch Modality Interpretation for an EIT-Based Sensitive Skin,"in Proc. of the IEEE Intl. Conf. on Robot. and Autom., pp. 3370-3376, August 2008.

[10] T. Miyashita,, T. Tajika,H. Ishiguro,K. Kogure and N. Hagita, "Haptic Communication between Humans and Robots ," Spring. Tracts in Advanc. Robotics, vol. 28, pp. 525-536, 2007.

[11] P. Maiolino, A. Ascia, L. Natale, G. Cannata, G. Metta, "Large scale capacitive skin for robots," Smart actu. sens. sys. - Recent advances and future challenges. InTech, Ed. Berselli, G. R Vertechy, G. Vassura, ISBN:979-953-307-990-4, (pp.16), 2012.

[12] Y. Kato, T. Mukai, T. Hayakawa, and T. Shibata, "Tactile Sensor without Wire and Sensing Element in the Tactile Region Based on EIT Method," IEEE Sensors, pp. 792–795, October 2007.

[13] A. Nagakubo, H. Alirezaei, and Y. Kuniyoshi, "A deformable and deformation sensitive tactile distribution sensor," in Proc. IEEE Int. Conf. Robotics and Biomimetics, 2007, pp. 1301–1308.

[14] R.M. Lewitt and S. Matej, "Overview of methods for image reconstruction from projections in emission computed tomography," Proc. of the IEEE, vol. 93, pp.1588-1611, October 2003.

[15] Bohare M. D, Cheeran A. N and Sarode V, "Analysis of Breast MRI images using Wavelets for Detection of Cancer" ICEICE 2011, Proceedings published by International Journal of Computer Applications (IJCA).

[16] J. M. Ollinger, and J.A. Fessler, "Positron-emission tomography," IEEE Signal Processing Magazine, vol. 14, pp. 43-55, January 1997.

[17] W. Q. Yang. "Hardware design of electrical capacitance tomography systems", Measurement Science Technology, vol. 7, 225, 1996

[18] K. Okazaki, A. Tangoku, T. Morimoto, K. Hattori, R. Kotani, E. Yasuno, M. Akutagawa, Y. Kinouchi, "Basic study of a new diagnostic modality by non-invasive measurement of the electrical impedance tomography (EIT) on localized tissues," ICBPE, pp. 1-5, December 2009.

[19] Webster J. G, "Medical Instrumentation Application and Design", Third edition, Wiley India, 2007.

[20] N. Vaisman, N. Weintrob, A. Blumental, Z. Yosefsberg, and P. Vardi, "Gastric Emptying in Patients with Type I Diabetes Mellitus," Annals of the New York Academy of Sciences, vol. 873, no. 1, pp. 506–511, 1999.

[21] Y. Ultchin, U. Nachaliel, and A. Ori, "Indirect calculation of breast tissue impedance values," Physiolog. Meas., vol. 23, pp. 177-182, January 2002.

[22] J. Conway, M. Hawley, Y. Mangnall, H. Amasha, and G. C. van Rhoon, "Experimental assessment of electrical impedance imaging for hyperthermia monitoring," Clin. Phys. and Physiolog. Meas., vol. A13, pp. 185-189, April 2007.

[23] G. Bogdanov, R. Ludwig, J. Wiggins, and K. Bertagnolli, "Non destructive testing of polycrystalline diamond cutters using dc electrical conductivity imaging," Meas. Sci. Technol., vol. 20 095707 (10pp) July 2007.h

[24] L. M. Heikkinen, J. Kourunen, T. Savolainen, P. J. Vauhkonen, J. P. Kaipio, and M. Vauhkonen, "Real time three-dimensional electrical impedance tomography applied in multiphase flow imaging," Meas. Sci. and Technol., vol. 17, pp. 2083-2087, July 2006.

[25] E. Zimmermann, A. Kemna, J. Berwix, W. Glaas, and H. Vereecken, "EIT measurement system with high phase accuracy for the imaging of spectral induced polarization properties of soils and sediments," Meas. Sci. Technol., vol. 19, 094010 (9pp.), July 2008.

[26] G. Teague, J. Tapson, and Q. Smit, "Neural network reconstruction for tomography of a gravel-air-seawater mixture," Meas. Sci. Technol., vol. 12, pp. 1102-1108, July 2001.

[27] B.H. Brown and A.D. Seagar, "The Sheffield data collection system," Clin. Phys. Physiol. Meas., vol. 8A, pp. 91-97, 1987.

[28] P. Hua, "Effect of the measurement method on noise handling and image quality of eit imaging," Proc. Annu. Int. Conf. Engng Med. and Biol. Soc., vol. 9, pp. 1429–1430, 1987.

[29] W. Lionheart, N. Polydorides, and A. Borsic, "The reconstruction problem," in Electrical Impedance Tomography: Methods, History and Applications, D. S. Holder, Ed. Bristol and Philadelphia: Institute of Physics Publishing, 2005, pp. 3–63.

[30] H. Alirezaei, A. Nagakubo, and Y. Kuniyoshi, "A highly stretchable tactile distribution sensor for smooth surfaced humanoids," in Proc. IEEE-RAS International Conference on Humanoid Robots (ICHR '07), Pittsburgh, Pennsylvania, November 2007, pp. 167–173.

[31] http://www.plugandwear.com/default.asp?mod=product&product_ID=136&cahttp://www3.3m.com/catalog/uk/en002/electronics_mfg/static_control_packaging/node_JV20CN7XFBbe/root_K3BHNB8005gv/vroot_5KR9K2C2KZge/gvel_JQGPTT9723gl/theme_uk_staticcontrolpackaging_3_0/command_AbcPageHandler/output_html

[32] A. Adler and W. R. B. Lionheart, "Uses and abuses of EIDORS: an extensible software base for EIT," Physiological Measurement, vol. 27, no. 5, pp. 25–42, 2006.

[33] Vauhkonen M., Lionheart W.R.B, Heikkinen L.M.,. Vauhkonen P. J and Kaipio J. P,"A MATLAB Package for the EIDORS project to reconstruct two-dimensional EIT images", ISSN 0788-4672, September 2000.

Wavelet Transform at Oriented Basis for Network Traffic Forecasting

Tereshchenko T.,Yamnenko Y., Veretiuk A., Veretiuk S.

Faculty of Electronics

National Technical University of Ukraine "Kyiv Polytechnic Institute"
Kyiv, Ukraine

Abstract – **In this article the network traffic forecasting preprocessing based on wavelet transforms is considered. The analysis of network traffic as time series was made. The data preprocessing using new wavelet transform was proposed. Application of the two types of wavelet transforms for data preprocessing (Oriented Basis and Haar) was analyzed. Results regarding the feasibility of wavelet transform type depending on application are presented.**

Keywords – wavelet; Haar; Hurst index; preprocessing; forecasting; oriented basis.

I. INTRODUCTION

Nowadays one of the main issues in telecommunication engineering is to ensure sustainable management and information exchange between the elements of complex telecommunication systems.

Modern multiservice networks serve diverse and previously undefined traffic type. This is caused by the fact that depending on the services provided by the network and the information flows' type the traffic time characteristics will vary considerably. In order to use the network elements (like switches, routers, etc.) effectively it is necessary to have information about the future traffic behavior, i.e., to predict it. Sometimes the situations may arise when the load to some elements increases while others are rarely used. This leads to the appearance of congestion in some parts of the network, and therefore to the integrity violation, information losses, service denial and slowdown of the network operation as a whole.

Development of mechanisms for traffic forecasting is relevant and important task of network management. Forecasting enables evenly load distribution, prevention the problems' emergence in the network, and takes measures to avoid them. This ensures the safe network operation and reliability.

For today not all known methods can be used for the traffic forecasting. Extrapolation forecasting methods [1] have high result correction complexity when receiving the current values of the numerical series and calculation for determining the analytical expressions' parameters. Forecasting model has low adaptive capacity to the changes in the time series. Statistical forecasting methods [2] use a relatively large data amount to adjust the model, have low model adaptive capacity to the changes in time series. While using these methods it is impossible to predict values changes not included in the data, which was conducted while model configuration.

Complicated non-stationary nature of time dependent traffic, availability of telemetry errors, terms of changes in the telecommunications market make forecasting task uncertain and require the application of new techniques to solve it. One of such technique is the use of artificial neural networks because of their ability to learn and adapt. In order to improve the neural network preprocessing of input data is applied. Traffic is represented as a time series. The preprocessing is based on wavelet transform that enables to eliminate uninformative fluctuations of the time series and squeeze it in several times. Therefore, number of input neurons can be reduced, total neural network structure can be simplified, and the predicted value delivery speed can be increased.

To be the subject of wavelet transform the initial function must meet two following conditions:

1) the level of similarity to the original function, with saving its characteristic features should be sufficient;

2) the number of digits should be small, since larger input vector dimension leads to more network training duration and database performance processing time increasing, and as a consequence – to predicting possibility reducing and computation complexity increasing.

Since these conditions are in contradiction it is necessary to look for a compromise between forecasting accuracy and computation costs [3].

II. TRAFFIC TIME SERIES ANALYSIS

Network traffic describes the dependence of telecommunication system load on the time. In general, the time series that characterizes the traffic can be represented by the sum:

$$X(t) = T(t) + P(t) + M(t) + R(t),$$

where $T(t)$ – trend that shows the change of subscriber base, increasing of traffic, network applications, etc., $P(t)$ – periodic component, which shows the daily and monthly traffic changes, $M(t)$ – traffic fluctuations caused by the network resources' distribution, routers' buffer overflow,

temporary overloads, network elements failure, etc., $R(t)$ – stochastic component [4].

As an example real traffic through a central switch of one of the Ukraine Internet providers is shown for different periods (Fig.1).

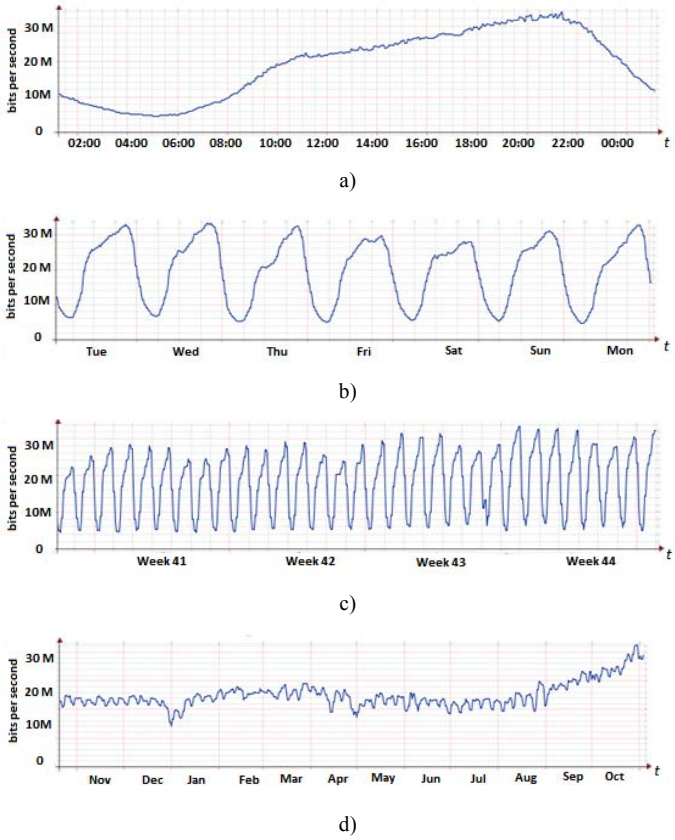


Fig. 1. Switch load graphics: a) day; b) weekly; c) monthly; d) annual

The monthly and annual graphics (see Fig.1,c,d) show gradually increasing of traffic which can be explained by increasing the number of network users, the launch of new network applications, etc. These processes are described by the component $T(t)$. The periodic component $P(t)$ can be traced on three graphics (Fig.1,a,b,c). It is noticeable that in daytime the switch load gradually increases, and in night time – decreases. On Fig.1,b,c,d some fluctuations are present, which may be caused either by network resources operation (incorrect operation, overload), that is component $M(t)$, or the users operation result (random large number of simultaneous network connections) – $R(t)$ component.

III. WAVELET TRANSFORM

Wavelets are promising mathematical apparatus for many applications. They are used in cases when the result of the signal analysis does not only contain simple characteristic frequencies (scales), but also identifies local coordinates' data on which these frequencies are revealed.

Let's investigate two types of discrete wavelet transforms – Haar and Oriented Basis (OB) in order to detect their advantages and disadvantages for preprocessing of traffic time series. The objective is to find the optimal data preprocessing method to accelerate network traffic forecasting and improve the computer networks safety and efficiency.

Haar wavelet is described quite fully in literature sources [6]. Then let's consider wavelet construction on the base of discrete spectral transform at oriented basis [5].

A. OB-transform

Discrete transform at oriented basis (hereafter OB-transform) is the base for extremely effective and fast wavelet transform that is used for preliminary processing. This transform operates with the discrete functions at finite interval $N=m^n$ where m and n are integer positive numbers, and besides m is simple.

OB-transform was initially constructed by the combination of Hartley and Wilenkin-Krestenson spectral transforms with further orientation of transform axis to the one of basic function vectors at complex coordinate plane.

The goal of OB-transform construction was minimization of different values of basic functions. In partial case when m is chosen equal to 3 and interval $N=3^n$ all multiplications are eliminated and basic functions are integer. It's important while realization of DSP algorithms at integral processors and microcontrollers.

Direct transform at oriented basis is follows:

$$Y(\nu) = \sum_{x=0}^{N-1} y(x) \cdot \varphi_d(\nu, x),$$

where x, $\nu = 0, 1, \ldots, N-1$;

$$\varphi_d(\nu, x) = \cos\left[\frac{2\pi}{m}\sum_{s=1}^{n} \nu^{(s)} x^{(s)}\right] + A\sin\left[\frac{2\pi}{m}\sum_{s=1}^{n} \nu^{(s)} x^{(s)}\right]$$

- basic functions of direct transform; $y(x)$ – discrete original function assigned at the interval $N = m^n$; $A = tg\,\alpha$; $\alpha = \frac{2\pi i}{m}, i = \overline{1, m-1}$ – orientation angle of transform axis; $x^{(s)}, \nu^{(s)}$ – digit components in m-ary representation of the numbers x and ν.

Reverse transform is chosen from the condition of orthogonality of direct and reverse basic function matrices:

$$y(x) = \frac{1}{N}\sum_{\nu=0}^{N-1} Y(\nu) \cdot \varphi_r(\nu, x),$$

where

$$\varphi_r(v, x) = \cos\left[\frac{2\pi}{m}\sum_{s=1}^{n} v^{(s)} x^{(s)}\right] + \frac{1}{A}\sin\left[\frac{2\pi}{m}\sum_{s=1}^{n} v^{(s)} x^{(s)}\right]$$

- basic functions of reverse transform.

OB-transform can be adapted to the concrete application. Such adaptation can be achieved by changing of orientation angle. Taking into account that changing $i = \overline{1, m-1}$ leads to m-1 values of the angle α, and direct transform formula can

be replaced by reverse one and vice versa, the amount of possible transforms at the same interval is 2(*m*-1). Depending of processed signal type, waveform or other factors the most appropriate modification can be chosen in order to provide most flexibility and effectiveness for control algorithms.

B. Wavelet-OB transform

Wavelet transform at oriented basis is also non-symmetrical, i.e. has different equations for the direct and reverse operation. Construction of direct wavelet-OB transform is follows. As scaling function $\varphi_d(0,x)$ is chosen, and as parent wavelet - $\phi_d(\nu,x), \quad \nu \neq 0$.

Basic functions of reverse transform $\phi_r(0,x)$, $\phi_r(\nu,x), \quad \nu \neq 0$ are constructed analogically.

Let's consider the construction of wavelet-OB transform for *m*=3. Then scaling function and two parent wavelets will be as follows:

$$\phi_d(x)=\varphi_d(0,x)=\{1;1;1\} \quad \psi_d(x)=\varphi_d(1,x)=\{1;-1;0\},$$
$$\gamma_d(x)=\varphi_d(2,x)=\{1;0;-1\}$$

These functions are shown at Fig.2.

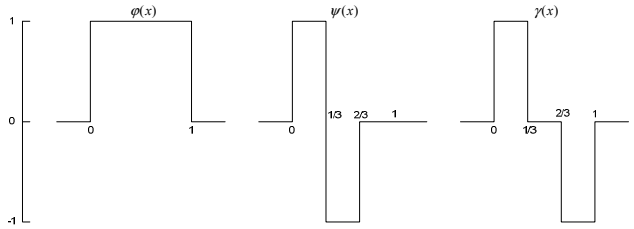


Fig.2. Basic function of direct wavelet-OB transform for *m*=3

Scaling function and parent wavelets for the reverse transform are defined by equations:

$$\phi_r(x)=\varphi_r(0,x)=\{1;1;1\}, \ \psi_r(x)=\varphi_r(1,x)=\{1;-2;1\},$$
$$\gamma_r(x)=\varphi_r(2,x)=\{1;1;-2\}$$

The basis wavelet function is produced from scaled functions of direct and reverse OB-transform:

$$\phi_{j,k}=3^{j/2}\phi\left(3^j x-k\right)$$
$$\psi_{j,k}=3^{j/2}\psi\left(3^j x-k\right)$$
$$\gamma_{j,k}=3^{j/2}\gamma\left(3^j x-k\right)$$

where ϕ is replaced by ϕ_d or ϕ_r, ψ is replaced by ψ_d or ψ_r and γ is replaced by γ_d or γ_r depending on the transform type (direct or reverse).

Let original function *f(x)* be defined by amount of digits 3^{j_m}:

$$f(x)=\sum_{k=0}^{3^{j_m}} s_{j_m,k}\, j_{j_m,k}(x)$$

where $s_{j_m,k}=\dfrac{f(k/3^{j_m})}{3^{\frac{j_m}{2}}}$

$\varphi_{j_m,k}$ is defined as the step function with unit norm and width $\dfrac{1}{3^N}$. It's non-zero only at *k*-th segment.

Decomposition of original function will be fulfilled using three types of coefficients that are calculated by formulas:

$$s_{j-1,k}=\frac{1}{\sqrt{3}}\left[s_{j,3k}+s_{j,3k+1}+s_{j,3k+2}\right],$$
$$d_{j-1,k}=\frac{1}{\sqrt{3}}\left[s_{j,3k}-s_{j,3k+1}\right],$$
$$l_{j-1,k}=\frac{1}{\sqrt{3}}\left[s_{j,3k}-s_{j,3k+2}\right]$$

Such calculation is iterative procedure when function analysis is executed at different decomposition levels *j* with gradual enlargement of considered interval from minimal that includes three digits, up to maximal equal to the interval of function definition.

While calculating at all levels, decomposition coefficients will be obtained.

The reconstruction function is based on the coefficients obtained at various decomposition levels *j* and scaled reverse transform basis functions.

$$f(x)=s_{0,0}\phi_{0,0}(x)+d_{0,0}\psi_{0,0}(x)+l_{0,0}\gamma_{0,0}(x)+\sum_{k=0}^{2}d_{1,k}\psi_{1,k}(x)+$$
$$+\sum_{k=0}^{2}l_{1,k}\gamma_{1,k}(x)+\sum_{k=0}^{8}d_{2,k}\psi_{2,k}(x)+\sum_{k=0}^{8}l_{2,k}\gamma_{2,k}(x)$$

Thus, signal can be characterized by its average (for some sub-interval) value – trend and its oscillations around the trend. These oscillations around the average values are called fluctuations independently on the reasons of their appearance. Usually during signal processing value of such fluctuations at different scales define its origin.

For practical application of wavelet transform it is necessary to solve two problems – to choose the samples' required number and to find the decomposition sufficient depth.

C. The samples' required number choice

The wavelet transform application for data preprocessing requires determining the wavelet decomposition maximum level j_m. This parameter determines the time series' number of elements to be predicted and, correspondingly, the neural network's number of elements.

The preprocessing first step is to determine the time series' predicting possibility. For this operation the Hurst index (*H*) is used [7]. Hurst index has been widely used due to its stability. It contains minimal system assumptions and

can classify time series. Hurst index characterizes the trend ratio (deterministic factor) to the noise (random factor) [8].

While calculating the Hurst index for different number of samples (N), it is possible to make a conclusion about the time series' behavior and the possibility to use it for forecasting [9], and then, based on the found value of N, set the wavelet transformation decomposition maximum level.

For the Hurst index calculation the following algorithm is used:

• the numeric series average value is calculated $X_j, (j = 1,.., N)$:

$$M = \frac{1}{N}\sum_{j=1}^{N} X_j ,$$

• the sample variance is calculated:

$$S_N = \sqrt{\sum_{j=1}^{N}\left(X_j - M\right)^2 / N} ,$$

• the integral deviation is calculated:

$$D_j = \sum_{k=1}^{j} X_k - jM$$

• the deviation scale is calculated:

$$R_N = \max_{1<j<N} D_j - \min_{1<j<N} D_j$$

• the Hurst index is defined as:

$$H = \log(\frac{R}{S}) / \log(\frac{N}{2}).$$

The Hurst index can be interpreted as follows:

• when $H = 0.5$ the obtained data is white noise, i.e. there is some independent random process;

• when $0 < H < 0.5$ the obtained data have the past and future values negative correlation, this series is called nonpersistent and its prediction is impossible;

• when $0.5 < H \le 1$ the obtained data have the past and future values positive correlation, this series is called persistent and its prediction is possible.

Let's calculate the Hurst index for two daily switch load graphic (Fig.3). The load values are measured after 5 minutes, that is the maximum samples' number N=(48·60)/5=576.

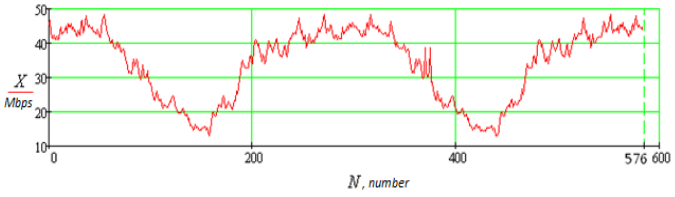


Fig. 3. Traffic through the switch within 48 hours

On Fig. 4 the dependence of Hurst index, calculated by the given algorithm, for different N is shown.

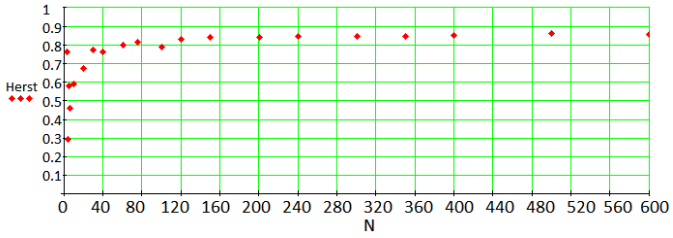


Fig. 4. The dependence of Hurst index on the samples' number

As shown on Fig.4 the Hurst index value becomes sustainable for the number of samples $N > 160$. Since wavelet-OB transform operates with $N = 3^{j_m}$ samples where j_m decomposition level, it is advisable to choose N=3^5=243 samples with $j_m = 5$. For the Haar wavelet transform these values are respectively equal to N=2^8=256 and $j_m = 8$. Then the Hurst index takes values within 0.8 – 0.85, indicating the pronounced series' persistent and the expediency selected values as maximum decomposition levels.

D. Depth of decomposition choice

One of the main ideas of wavelet transform is the signal's decomposition into two components – approximated and detailing – with the possibility of further fragmentation to change the signal's decomposition level. The wavelet transforms comparison should be performed with two parameters: the number of decomposition levels required to achieve the desired time series' filtration degree and the time needed for the decomposition to one level, i.e. the operations' number.

For the depth filtration numerical determination the approximated time series to the original function correlation coefficient R is calculated, which shows the statistical dependence degree between two variables [10], according to the formula:

$$R = \sum_{i=1}^{n}\left(y_i - \overline{y}\right)\left(y_{Ai} - \overline{y_A}\right) / \sqrt{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 \sum_{i=1}^{n}\left(y_{Ai} - \overline{y_A}\right)^2}$$

where y_i is the time series' original value; y_{Ai} – the approximated time series' value; n – time series' counting number; $\overline{y} = \frac{1}{n}\sum_{1}^{n} y_i$; $\overline{y}_A = \frac{1}{n}\sum_{1}^{n} y_{Ai}$.

The correlation coefficient's calculation results for the wavelet-OB transform (N=243) and Haar (N=256) are shown on Fig.5. The correlation coefficient's limits $R1$=0.98 and $R2$=0.965 (there is a high coupling degree between approximated and original function) were specified. The correlation coefficient's boundary level depends on the required accuracy, according to the problem statement. Various transforms allow achieving this value using different decomposition levels' number. In order to achieve a given similarity coefficient for OB-transform decomposition levels

1 and 2 are required, and for Haar – levels 2 and 3. The correlation coefficient value R=1 corresponds to the original function, i.e. zero decomposition level.

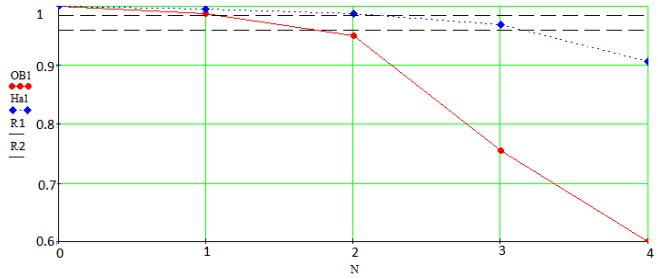


Fig. 5. The dependence of correlation coefficient on decomposition depth

With $R1$=0.98 the number of digits in Haar wavelet transform is 20% less than in OB, while with $R2$=0.965 another situation is observed – the number of counts in OB is 15% less than in Haar, which in turn causes the input neurons number decreasing. Note that both coefficients $R1$ and $R2$ show a high original and approximated function dependency.

In order to determine the wavelet transform's optimal method the performance of each are compared. For example for $R2$=0.965 the number of operations for 2-level OB and 3-level Haar wavelet transforms is calculated. The required number of operations at all decomposition levels is equal:

1) for wavelet-OB transform:

$$Q_A = \sum_{i=0}^{v} Q_{v_A} = \sum_{i=0}^{v} \left(\frac{1}{3^i} N \cdot (1+k) \right) = 702 \quad operations$$

2) for Haar wavelet transform:

$$Q_A = \sum_{i=0}^{v} Q_{v_A} = \sum_{i=0}^{v} \left(\frac{1}{2^i} N \cdot (1+k) \right) = 960 \quad operations$$

Where Q_{v_A} - the number of operations for the time series approximation on decomposition level v; k - the number of operations required for time series counts normalization; N - the number of samples in the original time series. Coefficient k, which depends on the microprocessor's type, determines the addition and multiplication cycles ratio, microprocessors' for modern types accepted value $k = 1$ [11].

IV. Conclusions

The expediency of the time series preprocessing that describes the communication traffic using wavelet transforms is shown. Preprocessing allows removing uninformative fluctuations from the time series, squeezing it in several times, that will lead to a reduction of the input neurons number, leading to a neural network simpler scheme and improving the predicted value issuance performance.

The advantage of wavelet-OB transform is reducing of the approximated series samples number that is notable for the correlation coefficient boundary values R<0.97. With $R1$=0.98 the number of digits in Haar wavelet transform is 20% less than in OB, while with $R2$=0.965 another situation is observed – the number of counts in OB is 15% less than in Haar, which in turn causes the input neurons number decreasing. Note that both coefficients $R1$ and $R2$ show a high original and approximated function dependency. In addition, wavelet-OB transform has advantages in performance: for considered example with interval values N=243 and correlation coefficient R=0.965, number of operations decreased by 25% compared with Haar wavelet transform.

These advantages allow receiving the necessary information about the traffic future behavior sooner, i.e. to accelerate the forecast, which in turn will improve the network equipment control quality.

References

[1] Kagramanzade A.G. Prediction and design of telecommunications networks. - Baku, 1998. – 242 p. (In Russian)

[2] Lukashin Y.P. Adaptive methods of short-term time series prediction. - Moscow: Finance and Statistics, 2003. - 415 p.(In Russian)

[3] Petergerya Y.S., Zhuikov V.Y., Tereschenko T.A. Intelligent energy saving houses. Textbook. - K.: Media-Press, 2008. - 256 p. (In Ukrainian)

[4] Bezrukavnyi Dmitry Sergeevich. Analysis and management of data in the form of time series: dis. phd: 05.13.01 Moscow, 2007 - 114 p. (In Russian)

[5] Zhuikov V.Y. Discrete spectral conversion functions on finite intervals: studies manual / V.Y. Zhuikov, T.A. Tereschenko Y.S. Petergerya. - K.: NTU "KPI", 2010.-244 p. (In Russian)

[6] Smolentsev N.K. Fundamentals of the theory of wavelets. Wavelets in MATLAB. - Moscow: DMK Press, 2005. – 304 p. (In Russian)

[7] Kasimov R.R. . Improved algorithms for QoS routing in networks with technology IP / MPLS-based traffic forecast: dis. phd: 05.12.02 / Rustam R. Qasim. - K., 2011. - 179 p.(In Ukrainian)

[8] Petrov V.V. Teletraffic structure and algorithm to ensure quality of service at the impact of self-similarity effect: dis. phd: 05.12.13/Petrov Vitaly Valerevich.-M., 2004.-199 p. (In Russian)

[9] D.V. Belkov, E.N. Edemskaya, L.V. Nezamova. Statistical analysis of network traffic // Proceedings of Donetsk National Technical University, series "Informatics, Cybernetics and Computer Science", vol. 13 (185), Donetsk, Donetsk National Technical University, 2011. - pp. 66-75. (In Russian)

[10] Eliseeva I.I, Yuzbashev M.M. General Theory of Statistics: Textbook / ed. I.I. Eliseev. - The 4th edition, revised and enlarged. - Moscow: Finance and Statistics, 2002. – 480 p. (In Russian)

[11] Kolotov M.V. Manage transducer orientation system electronic platforms: dis. phd: 05.09.12 / Kolotov Mukola Viktorovich. - K., 2010. - 194 p. (In Ukrainian)

Slot Allocation Model and Data Burst Scheduling in Downlink WiMAX Technology

Lemeshko Olexandr
department of Telecommunications Systems
Kharkiv National University of Radioelectroniks
Kharkiv, Ukraine
avlem@mail.ru

Garkusha Sergiy
department of Information Activities in Economic Systems
Poltava University of Economics and Trade
Poltava, Ukraine
sv.garkusha@mail.ru

Abstract – **Current approaches to slot allocation among data bursts in downlink of WiMAX technology were analyzed. As a result of the completed analysis drawbacks of known decisions were distinguished and basic requirements to the slot allocation problem among data bursts were formed. A mathematical model of slots allocation and data burst scheduling in the downlink of IEEE 802.16 standard is proposed. The proposed model is focused on the formation of one data packet in downlink for each subscriber station that minimizes the number of overhead information sent over the used frequency communication channel. The dependence of the amount of slots, forming one data packet from required transmission rate of a subscriber station and type of modulation and coding scheme were analyzed.**

Keywords – slot allocation; data burst scheduling; mathematical model; modulation and coding system.

I. INTRODUCTION

In WiMAX technology an effective way to increase productivity and to improve the basic quality of service (QoS), is to use structural and functional principles of self-organization. A high level of self-organization can be achieved through improved network of protocols and mechanisms responsible for the allocation of the available network resources. This kind of resources includes first and foremost the network traffic (information resource), the bandwidth of communication channels (channel resource), the queue (buffer resources), as well as the frequency subcarriers (frequency resource), that is especially important for wireless networks [1, 2]. IEEE 802.16 standard defines mechanisms for planning and allocation of network resources for uplink, leaving the right to choose for telecom operators and equipment manufacturers in the downlink [3].

Within WiMAX, built on OFDMA technology, it is possible to control the frequency and time resources. The subcarriers and subchannels are used as the frequency resources. Slots are used as a temporary resource of OFDMA technology which always occupies one logical subchannel in the frequency domain and from one to three OFDM-symbols – in the time resource. Slots allocated for the transmission of information to a single user and united into one area, form a burst data [4].

Thus, the problem of allocation of the frequency and time resource of WiMAX should be formulated as the problem of allocation of slots between the subscribe station (SS) of the wireless network and their relevant allocation into the bursts data based on the stated rate and QoS parameters.

II. THE ANALYSIS OF THE KNOWN SOLUTIONS

Analysis of the known solutions [5-21] showed that the increase of WiMAX productivity and ensuring of QoS can be achieved by both separate and coherent problem solutions as to allocation of frequency and time resources. Thus, modification of the allocation problem of frequency resources is presented in [5-15]. The approach proposed in [16], refers the problem of temporary resource allocation. Furthermore, studies [17-21] proposed the approaches to the solution of a joint distribution of frequency and time resource allocation problem formulated as slots allocation and data burst scheduling in the downlink. However, the approaches proposed in [17-21] are heuristic by nature.

Regarding the shortcomings of the known solutions [5-21], the following requirements to the promising solutions of slot allocation and data burst scheduling in downlink of WiMAX technology have been formulated: focus on the efficient use of the frequency and time resources; control for the SS rate and quality of service; minimization of the number of service data transmitted over the communication channel; focus primarily on the dynamic nature of the problem of slots allocation and data burst scheduling; focus on maximizing network performance in the whole and improving the quality of other services; consideration of the technological features of the network (operating mode, bandwidth, number of subchannels, frame duration); consideration of geographically remote stations (it determines the choice of modulation and coding system for signal transmission to SS signal).

Considering the analysis, as well as the requirements, a unified model, which describes the joint distribution procedure both for frequency and time resources, is proposed. This model is presented in the form of the solution of the slot allocation and their scheduling into the data bursts in the downlink of WiMAX technology based on indicators of QoS (SS required rate) and geographically remote of SS network (the modulation and coding system for signal transmission to SS).

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

III. Initial Data for the Problem of the Slot Allocation and Data Burst Scheduling in the Downlink

In the proposed model, the following inputs are assumed to be known:

1) N – the number of user stations in a wireless network;

2) K – the number of subchannels in the frequency channel specified by the submodes and making 32 subchannels for DL FUSC, and for DL PUSC - 60 subchannels;

3) L – the number of symbols in the frame. It should be noted that under WiMAX frame duration may vary and take values equal to 2, 2.5, 4, 5, 8, 10, 12.5, 20 ms. Regarding the fact that the useful part of the symbol has a fixed duration T_b=89.6 ms, the number of symbols in the frame will take the value of 19, 24, 39, 49, 79, 99, 124, 198 according, to specified duration of the frame. However, there is a guard interval between the symbols, which can take four values concerning the length of useful part of the symbol: $T_g = T_b / 4 = 22.4$ ms; $T_g = T_b / 8 = 11.2$ ms; $T_g = T_b / 16 = 5.6$ ms; $T_g = T_b / 32 = 2.8$ ms. As a result, the exact duration of the frame can be calculated as $(T_b + T_g)L$ [4, 22, 23];

4) R_{req}^n – the required data rate for serving the n-th SS (Mbps);

5) S – the number of symbols that form a slot. The size of the slot depends on the submode [4] and can be set for DL FUSC - one OFDMA-symbol in one subchannel; DL PUSC - two OFDMA-symbols in one subchannel; UL PUSC - three OFDMA-symbols on one subchannel, AMC - 6, 3 or 2 OFDMA-symbols in one subchannel;

6) the types for the modulation and coding system (MCS), depending on geographically remote SS;

7) the number of slots in one subchannel of the downlink to transmit useful information:

$$M = \left\lfloor \frac{UL - L_{pr}}{S} \right\rfloor, \quad (1)$$

where $\lfloor\ \rfloor$ – operation of rounding to the nearest integer in the lower side, U is a ratio of the number of symbols in the downlink to the total number of symbols in the frame (the ratio of the number of symbols in the downlink to the number of symbols in the uplink varies from 3:1 to 1:1 to support different traffic profiles [4]), L_{pr} – the number of symbols to transmit preamble (equal to one).

The mathematical model records the useful information transmitted in a single frame, and consists of a fixed and a variable part. The fixed part consists of a preamble and Frame Control Header (FCH). The variable part includes DL-MAP and UL-MAP, which contain downlink MAP information element (DL-MAP_IE) and uplink MAP information element (UL-MAP_IE) respectively. One DL-MAP_IE corresponds to each data burst in the downlink, and one UL-MAP_IE - to each burst in the uplink. Fixed and variable parts have the following dimensions [4, 22, 23]:

- DL-MAP – 104 bits;
- DL-MAP_IE – 60 bits;
- UL-MAP – 64 bits;
- UL-MAP_IE – 32 bits.

In order to obtain better recovering of the overhead data at the receiver's side in WiMAX technology the number of repetitions of fixed and variable parts can have the following values: w=2, w=4 and w=6. As a result, the number of slots needed used for transmission of the overhead information will be calculated as follows:

$$Q_{DL-MAP} = \left\lceil \frac{104 + 60N}{SR_c^{\mathrm{MAP}} k_b^{\mathrm{MAP}} K_s} \right\rceil w; \quad (2)$$

$$Q_{UL-MAP} = \left\lceil \frac{64 + 32N}{SR_c^{\mathrm{MAP}} k_b^{\mathrm{MAP}} K_s} \right\rceil w, \quad (3)$$

where in the numerator there is the total number of bits of the overhead information, in the denominator there is the number bits of the overhead information transmitted in a single slot, $\lceil\ \rceil$ – operation of rounding to the nearest larger integer; R_c^{MAP} – coding rate used for encoding the signal of overhead information; k_b^{MAP} – bit symbol load in an OFDM symbol used for encoding a signal of overhead information; K_s – the number of subcarriers for data transmission in a single subchannel (DL PUSC submode is used for sending the overhead information, resulting in K_s=24).

The total number of slots for the overhead information transmission can be calculated from the expression

$$Q = Q_{DL-MAP} + Q_{UL-MAP} + Q_{FCH}, \quad (4)$$

where Q_{FCH}=4 – the number of slots to transmit a frame control header.

IV. The Mathematical Model of Slots Allocation for the Data Bursts in the Downlink

During the solving of the slot allocation problem and data bursts scheduling for transmission of useful information in the downlink by each SS of the network, it is necessary to determine a boolean control variable ($x_{k,m}^n$) in the framework of the proposed model, ensuring the allocation of slots:

$$x_{k,m}^n = \begin{cases} 1, \text{ if the } m\text{-th slot is used in the } k\text{-th} \\ \quad \text{subchannel allocated by the } n\text{-th SS;} \\ 0, \text{ otherwise.} \end{cases} \quad (5)$$

The total number of control variables depends on the number of SS in the network, the number of used subchannels and the number of slots in one downlink subchannel and will be determined by the expression $N \times K \times M$. The calculation of the variables (5) results in binding of SS subchannels and allocation of slots which will be used to transfer the data in the downlink. In addition, the calculation of the variables $x_{k,m}^n$ must be performed considereing a number of important conditions/constrains:

1) Fixing the k-th subchannel for the m-th slot transmission for no more than one SS

$$\sum_{n=1}^{N} x_{k,m}^n \le 1 \ (k = \overline{1,K}\,; m = \overline{1,M}\,); \quad (6)$$

2) Fixing the number of slots for the n-th SS providing the necessary transmission rate when using MCS

$$R_S^n \sum_{m=1}^{M} \sum_{k=1}^{K} x_{k,m}^n \ge R_{req}^n \ \ (n = \overline{1,N}\,), \quad (7)$$

where $R_S^n = \dfrac{SR_c^n k_b^n K_s}{(T_b + T_g)L + T_{RTG} + T_{TTG}}$ is the slot bandwidth assigned to the n-th SS, that depends on the MCS determining the number of bits transmitted in a time equal to the length of the slot; R_c^n – code rate used when encoding the signal of the n-th SS (for example, for 16-QAM 1/2 modulation the parameter is R_c=4); k_b^n is a bit symbol load in an OFDM symbol of the n-th SS (for example, for 16-QAM 1/2 modulation the parameter is k_b^n=1/2); K_s – the number of subcarriers for data transmission in a single subchannel (for DL FUSC submodes K_s=48, and for DL PUSC K_s=24); $\Delta f \approx 11{,}16$ KHz – frequency separation between subcarriers; $T_{TTG} = 105{,}7$ μs – transmit/receive transition gap (TTG) interval time; $T_{RTG} = 60$ μs – receive/transmit transition gap (RTG) interval time [4, 22, 23].

3) Scheduling conditions of a single burst for the n-th SS are:

$$x_{k,i}^n x_{k,z}^n (i - z + 1) - \sum_{u=z}^{i} x_{k,u}^n \le 0, \quad (8)$$

at ($z = \overline{1,M-1}\,;\ i = \overline{2,M}\,;\ n = \overline{1,N}\,;\ k = \overline{1,K}\,;\ i > z$);

$$x_{j,m}^n x_{r,m}^n (j - r + 1) - \sum_{s=r}^{j} x_{s,m}^n \le 0, \quad (9)$$

at ($r = \overline{1,K-1}\,;\ j = \overline{2,K}\,;\ n = \overline{1,N}\,;\ m = \overline{1,M}\,;\ j > r$).

4) Scheduling condition for the "rectangular" bursts are:

$$x_{k,m}^n \sum_{d=1}^{M} x_{k,d}^n \sum_{b=1}^{K} x_{b,m}^n = x_{k,m}^n \sum_{g=1}^{K} \sum_{h=1}^{M} x_{g,h}^n, \quad (10)$$

at ($n = \overline{1,N}\,;\ k = \overline{1,K}\,;\ m = \overline{1,M}$).

5) Conditions for the required number of slots reservation for the overhead transmission

$$\sum_{k=1}^{K} \sum_{n=1}^{N} x_{k,m}^n = 0, \ (m = \overline{1, m_{oh} - 1}\,, \lceil Q/K \rceil \ge 1); \quad (11)$$

$$\sum_{n=1}^{N} x_{k,m_{oh}}^n = 0, \ (k = \overline{1, k_{oh}}\,, \lceil Q/K \rceil \ge 1); \quad (12)$$

$$\sum_{n=1}^{N} x_{k,1}^n = 0, \ (k = \overline{1, k_{oh}}\,, \lceil Q/K \rceil < 1), \quad (13)$$

where $m_{oh} = \lceil Q/K \rceil$ – the number of slots allocated for the overhead information transmission, occupying the entire width of the frequency channel (available after preamble);
$k_{oh} = Q - K(m_{oh} - 1)$ – the number of slots allocated for the overhead information transmission, occupying only a part of the width of the frequency channel.

Conditions (11) and (12) are used, if the number of slots needed to transmit overhead data is greater than or equal to the number of subchannels in the frequency channel $\lceil Q/K \rceil \ge 1$. Besides, condition (11) allows to make a reservation of all subchannels for transmission of the number of slots equal to m_{oh}. Conditions (12) and (13) reserve only a part of subchannels (k_{oh}) for the duration of one slot.

Calculation of the variables (5), in accordance with the conditions/constraints (6)-(13) is better to carry out in the course of solving the optimization problem by selecting minimum or maximum quality criteria to solve the problem of slots allocation and bursts scheduling for data transmission by the SS wireless network of IEEE 802.16 standard.

The problem of SS slots allocation and data bursts scheduling can be solved by using the optimality criterion, aimed at saving the frequency and time resources, and reducing the time spent by SS in the active state, that will reduce power for SS. Thus, the optimality criterion becomes:

$$\min \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{m=1}^{M} x_{k,m}^n, \quad (14)$$

taking into account the conditions/constraints (6)-(13).

The problem formulated from the mathematical point of view is the problem of Mixed Integer Non-Linear Programming (MINLP). In the model the variables $x_{k,m}^{n}$ (5) are the booling ones. Variable used in the optimality criterion (14) is integral, limits on the unknown variables (6), (7), (11)-(13) are linear, and the constraints (8)-(10) are non-linear.

V. Simulation and Results

To analyze the obtained solutions for the slot allocation problem among data bursts in downlink within the proposed model (5)-(13) let us consider an example of the problem. In this example MatLab R2011a system in which the minlpAssign program is involved, being a part of optimization pack TOMLAB was applied to the solutions formulated in an optimization problem.

In the solving of the slot allocation problem among data bursts in downlink we used the following initial values: the number of user stations – N=9; used submode OFDMA – DL PUSC; the number of subchannels in the frequency channel – $K=60$; types MCS SS 1 – QPSK 3/4, SS 2 – 16 QAM 3/4, SS 3 – QPSK 1/2, SS 4 – QPSK 1/2, SS 5 – 16 QAM 1/2, SS 6 – 64 QAM 3/4 , SS 7 – QPSK 3/4, SS 8 – 16 QAM 3/4, SS 9 – QPSK 3/4; the required data rate for serving each SS (Mbps) – R_{req}^{1}=1.5; R_{req}^{2}=1.25; R_{req}^{3}=2.15; R_{req}^{4}=1.75; R_{req}^{5}=0.2; R_{req}^{6}=0.25; R_{req}^{7}=1.5; R_{req}^{8}=0.5; R_{req}^{9}=1; frame duration – 5 ms; guard interval duration – $T_g = T_b / 8 = 11.2$ ms; the number of symbols in the frame – L=49, resulting in precise frame duration is $(T_b + T_g)L$=4.9392 ms; the number of symbols that form a slot – S=2 (determined submode OFDMA); the ratio of the number of symbols in the downlink to the number of symbols in the uplink – 3:1 ($U = 3/4$); the number of slots in one subchannel of the downlink to transmit useful information – $M = 18$.

As we can see from the obtained solution (Fig. 1) nine data bursts (DL-packets) are formed in downlink. The amount of slots in each data burst satisfies the transmission rate ofa subscriber station.

In addition, the analysis showed that the use of (8) and (9), ensuring the scheduling of a single data bursts for each substation, enables to minimize the amount of overhead information transmitted in the downlink, since the scheduling of each additional DL-bursts leads to the DL-MAP_IE formation.

Besides the analysis showed that the problem is solved more effective in case of using the joint allocation of frequency and time resource of WiMAX technologies downlink. Fig. 2 shows the number of times SS slots can be reduced while solving the problem of joint frequency and time resources allocation, compared to the problems solved separately for frequency resource (B_{freq}) and time resource (B_{time}) allocation, depending on the desired SS rate transmission.

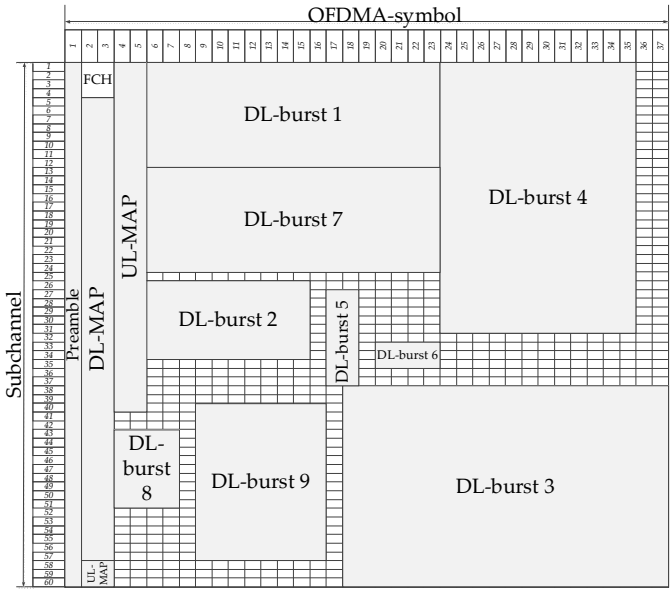


Fig. 1. The example of the slot allocation problem among data bursts in downlink

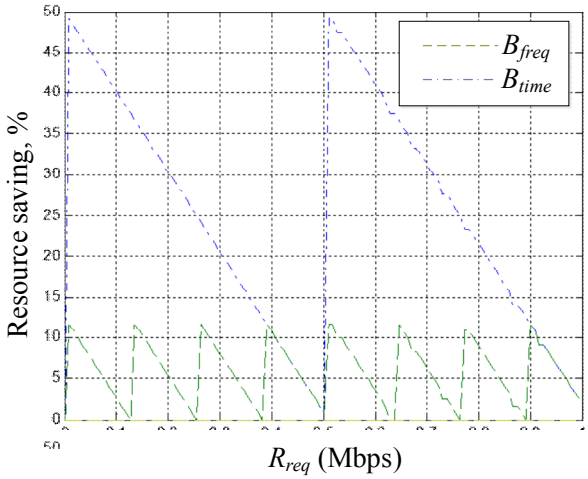


Fig. 2.The advantage of the joint frequency and time resource allocation problem under effective use of downlink bandwidth in WiMAX technology

The curves shown in Fig. 2 are characterized by sharp fluctuation. The presence of fluctuations is explained by the need of the next subchannel allocation (in case of solving the problem of frequency resource allocation) or a sequence of symbols, divisible to the number of symbols in the slot (in case of solving the problem of time resources allocation).

VI. Conclusion

It was stated that one of the main challenges in the WiMAX wireless networks is the problem of ensuring the required quality of service in terms of includes the required transmission rate in the downlink for the SS network. Ensuring the necessary rate of WiMAX can be achieved by solving the problem of slot allocation among the data bursts of

one frame in the downlink. In this regard, existing approaches for slot allocation between the data bursts in the downlink of wireless network in WiMAX technology were analyzed. The analysis identified the drawbacks of the known solutions and formulated the requirements for determining the way of solving slots allocations and data bursts scheduling problem.

On the basis of formed requirements to slot allocation problem and their scheduling into data busts the mathematical model, presented by the linear and nonlinear conditions/constraints was proposed. The novelty of the model is the formulation of the problem concerning slots allocation and data bursts scheduling as a problem of balancing the available downlink bandwidth of WiMAX technology for the transmission of useful information by SS, taking into account their geographical distance (type of modulation and coding system). The increase of network productivity was caused by the decrease in overhead information transmitted in the downlink, that was achieved by allocating one data burst for each SS.

The formulated problem of slot allocation and data bursts scheduling in the downlink from the point of view of physics processes in the wireless network refers to the class of balancing the time-frequency resources – the number of slots in the data burst, and from the mathematical point of view it is a problem of Mixed Integer Non-Linear Programming (MINLP). As an example we used the solution formulated in the optimization problem, where we applied MatLab R2011a system, within which the minlpAssign program was involved as a part of optimization pack TOMLAB.

REFERENCES

[1] Gogolieva M., Garkusha S., Ahmed H. Abed A Mathematical Model of Channel Distribution in Multichannel Mesh Networks 802.11 // Proc. of 11-th International Conference the Experience of Designing and Application of CAD Systems in Microelectronics. – Polyana-Svalyava-(Zakarpattya), UKRAINE, 23-25 February 2011: Publishing House of Lviv Polytechnic, 2011. – P. 71-73.

[2] Lemeshko A., Garkusha S., Abed A.H. Two-index Mathematical Model of Channels Distribution in Multichannel Mesh Networks 802.11 // Modern Problems of Radio Engineering, Telecommunications and Computer Science. Proceedings of the international Conference TCSET'2012. – Lviv-Slavske, Ukraine, 21-24 February 2012: Publishing House of Lviv Polytechnic, 2012. – P. 279-280.

[3] Shabani A.H., Beg M.T., Khade A.A.-H. Survey of Down Link Data Allocation Algorithms in IEEE 802.16 WiMAX // International Journal of Distributed and Parallel Systems. – 2012. – Vol.55, No.6. – P. 197-207.

[4] Andrews J.G., Ghosh A., Muhamed R. Fundamentals of WiMAX Understanding Broadband Wireless Networking – Prentice Hall Communications Engineering and Emerging Technologies Series, 2007 – 449 p.

[5] Garkusha S.V. Design and analysis of model sub-channel allocation in the network standard IEEE 802.16 // Bulletin of the National University "Lviv Polytechnic": Radioelectronics and Telecommunications. – 2012. – № 738. – C. 177-185.

[6] Garkusha S.V. Design and analysis of scalable model sub-channel allocation in the network standard IEEE 802.16 // Proceedings of Kharkov University of Air Force. – 2012. – Vol. 4(33). – C. 68-74.

[7] Niyato D., Hossain E. Adaptive fair subcarrier/rate allocation in multirate OFDMA networks: Radio link level queuing performance analysis // IEEE Transactions on Vehicular Technology. – 2006. – Vol. 55, № 6. – P. 1897-1907.

[8] Letaief K., Zhang Y. Dynamic multiuser resource allocation and adaptation for wireless systems // IEEE Wireless Communications Magazine. – 2006. – Vol. 13, № 4. – P. 38–47.

[9] Ergen M., Coleri S., Varaiya P. QoS aware adaptive resource allocation techniques for fair scheduling in OFDMA based broadband wireless access systems // IEEE Transactions on Broadcasting. – 2003. – Vol. 49, № 4. – P. 362–370.

[10] Jang J., Lee K. Transmit power adaptation for multiuser OFDM systems // IEEE Journal on Selected Areas in Communications. – 2003. – Vol. 21, № 2. – P. 171–178.

[11] Rhee W., Cioffi J. Increase in capacity of multiuser OFDM system using dynamic subchannel allocation // Proceeding of IEEE Vehicular Technology Conference (VTC 2000). – 2000. – Vol. 2. – P. 1085-1089.

[12] Cai J., Shen X., Mark J. Downlink resource management for packet transmission in OFDM wireless communication systems // IEEE Transactions on Wireless Communications. – 2005. – Vol. 4, № 4. – P. 1688-1703.

[13] Seong K., Yu D., Kim Y., Cioffi J. Optimal resource allocation via geometric programming for OFDM broadcast and multiple access channels // Proceeding of IEEE Global Telecommunications Conference (GLOBECOM '06). – 2006. – P. 1-5.

[14] Mehrjoo M., Awad M.K., Shen X.S. Resource Allocation in OFDM-Based WiMAX // CRC Press, Wireless Networks and Mobile Communications: WiMAX network planning and optimization. – 2009. – P. 113-131.

[15] Zhang Y., Letaief K. Multiuser adaptive subcarrier-and-bit allocation with adaptive cell selection for OFDM systems // IEEE Transaction on Wireless Communications. – 2004. – Vol. 3, № 5. – P. 1566-1575.

[16] Ohseki T., Morita M., Inoue T. Burst construction and packet mapping scheme for OFDMA downlinks in IEEE 802.16 systems // IEEE Global Telecommunications Conference (GLOBECOM). – 2007. – P. 4307-4311.

[17] Lodi A., Martello S., Monaci M. Two-dimensional packing problems: A survey // European Journal Operational Research. – 2002. – Vol.141. – P. 242-252.

[18] Ben-Shimol Y., Kitroser I., Dinitz Y. Two-dimensional mapping for wireless OFDMA systems // IEEE Transactions on Broadcasting. – 2006. – Vol. 52, No. 3. – P. 388-396.

[19] Jin X., Zhou J., Hu J., Shi J., Sun Y., Dutkiewicz E. An Efficient Downlink Data Mapping Algorithm for IEEE802.16e OFDMA Systems // IEEE Global Telecommunications Conference (GLOBECOM). – 2008. –P. 5233-5237.

[20] Sheu S.T., Tsai M.H., Tsai T.-Y., Tsai Y.-H. Condensed Downlink MAP Structures for IEEE 802.16e Wireless Metropolitan Area Networks (MANs) // Vehicular Technology Conference. – 2010. – P. 237-244.

[21] Kim J.Y., Cho D.-H. Piggybacking Scheme of MAP IE for Minimizing MAC Overhead in the IEEE 802.16e OFDMA Systems // Vehicular Technology Conference. – 2007. – P. 284-288.

[22] IEEE Standard for Local and Metropolitan Area Networks: Part 16: Air Interface for Broadband Wireless Access Systems, IEEE Std 802.16-2009. –2009. – 2094 p.

[23] Fundamentals of WiMAX: A Technology Primer. – Telesystem Innovations Inc, 2010. – 35 p.

Topological Analysis of Active Networks Containing Pathological Mirror Elements

Vladimir Filaretov
Department of Electrical Engineering
Ul'yanovsk State Technical University
Ul'yanovsk, Russia
vvfil@mail.ru

Konstantin Gorshkov
Department of Electrical Engineering
Ul'yanovsk State Technical University
Ul'yanovsk, Russia
K.Gorshkov@ulstu.ru

***Abstract* — The new topological rules providing symbolic analysis of the networks with pathological mirror elements have been proposed. The approach is based on generalized parameter extraction method and it doesn't require presentation of the circuit in matrix or oriented graph form. The main advantage of this method is that there are no reductions of the generated terms of symbolic expressions. The method has been implemented in the computer program SCADS. A discussion of theoretical basics as well as results of the calculation is included.**

Keywords — symbolic analysis; nullor; voltage mirror; current mirror; pathological elements, network function.

I. INTRODUCTION

The pathological elements nullor [1-3] and voltage mirror-current mirror (VM-CM) pair [4] are quite useful in analog behavioral modelling and circuit synthesis [3, 5-7]. The nullor is an ideal operational amplifier which is composed of a nullator connected in the input-port and a norator connected in the output port. The grounded mirror elements VM and the CM make a pair which is also an ideal element and is composed of a VM at the input port and a CM at the output port.

The nullor-equivalents for the VM and CM are presented in Fig. 1 (a) and Fig. 1 (b) respectively [8]. If any terminal of the VM or CM is connected to the ground, it is equivalent to a nullator or norator element, respectively.

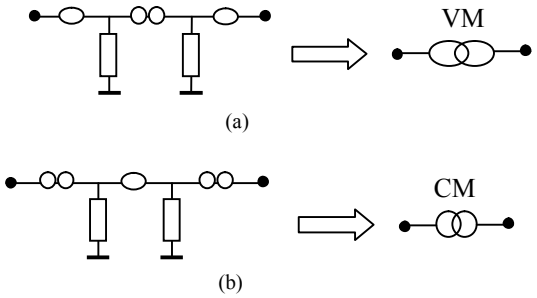


Fig. 1. The nullor-equivalents for the VM (a) and CM (b)

This models provide the implementation of the standard nodal analysis methods for calculation of the symbolic circuit functions of the active networks [8]. Also the construction of the nodal analysis matrix direct from networks containing the pathological mirror elements is possible as shown in [9]. Unfortunately, there is no report about usage of topological approach advantages of VM-CM based network analysis.

In this paper the special rules and the algorithm for symbolic analysis of active networks with pathological mirror elements are proposed. The approach is based on the generalized parameter extraction method, an effective tool for network symbolic analysis [10-12] and synthesis [13-15].

II. THE TOPOLOGICAL APPROACH TO SYMBOLIC ANALYSIS OF NETWORKS WITH PATHOLOGICAL ELEMENTS

A. Orientation of Pathological Elements

In contrast to symbolic analysis of nullor-based equivalent circuits by means of nodal analysis methods the implementation of topological approach requires the oriented nullor model presented in Fig. 2 [16]. Note, that the orientation of the nullator and norator corresponds to arrows shown as triangles instead of standard ovals [10].

Fig. 2. Oriented norator (a) and nullator (b)

The orientation of the nullors provide the direct formulation of circuit determinant by means of parameter extraction formulae without special presentation of the network as a number of algebraic sets or topological graphs [10]. If we use the oriented nullors for pathological mirror modelling we can obtain the oriented VM and CM as it shown in Fig. 3.

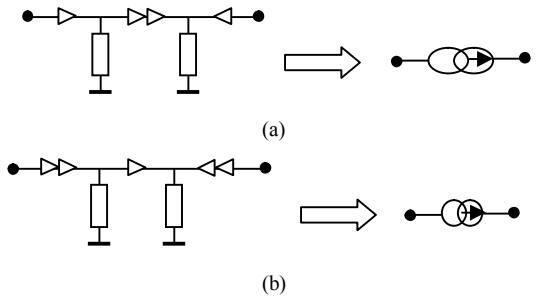


Fig. 3. Oriented VM (a) and CM (b)

Let us suppose that the orientation of the VM (CM) corresponds to the middle norator (nullator) in the nullor equivalent network. The nullor elements from the right and left sides of the equivalent circuits in Fig. 3 are connected in opposite orientation.

B. Equivalent Transformations of Pathological Elements-Based Circuits

The orientation of mirror elements provides the generalization of well-known topological rules of equivalent transformations of nullors-based circuits (Fig. 4) [10, 16, 17] for pathological mirror-based circuits. As shown in Fig. 5 (a) in case of series connection the VM-CM pair will be deleted from the circuit, because the nullor marked by index "1" corresponds to an open loop.

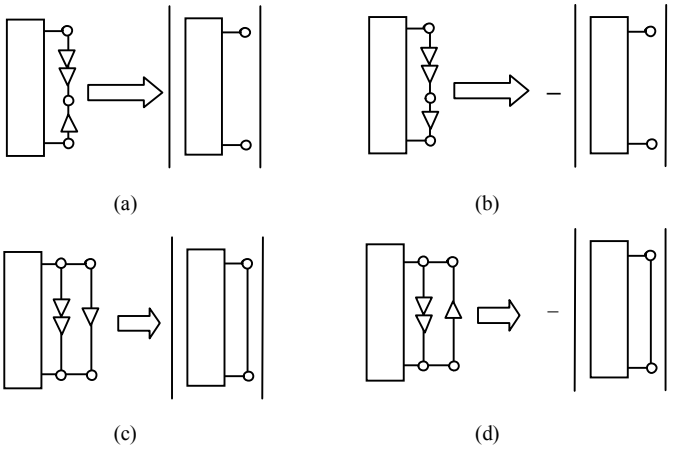


Fig. 4. Equivalent transformations of nullors

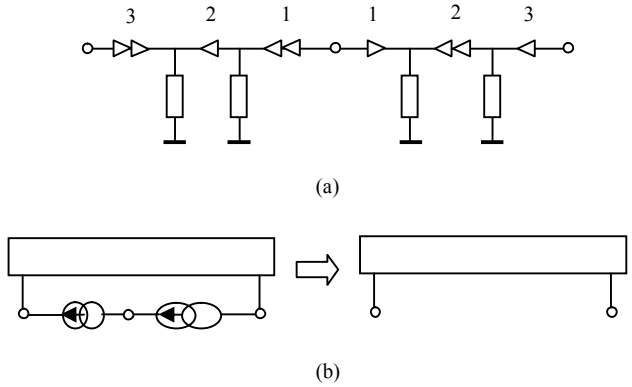


Fig. 5. Transformation of the oriented VM-CM pair connected in series

It is obvious that the same thing will happen if we have a similar connection of the VM and norator or CM and nullator in series. Note, that the opposite connection of VM-CM, VM-nullator or CM-norator pair leads to the change of a determinant expression sign.

The transformation of mirrors in parallel connection will make sense only if one of their common nodes will be grounded (Fig. 6 (a)) or isolated (Fig. 6 (b)). In that case all of three nullors are deleted from the circuit and the mirrors loop is shorted. Note, that in case of a parallel connection the sign of the circuit determinant is still the same if the pathological elements have the different orientation and change if the orientation is similar.

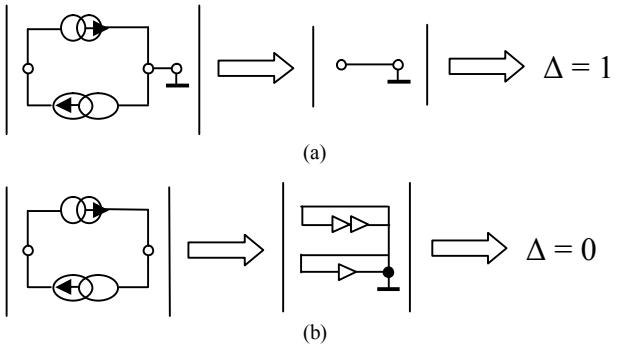


Fig. 6. Transformation of the oriented VM-CM pair connected in parallel

C. Special Cases of Circuit Elements Connection

It is very important for a circuit analysis by means of the generalized parameter extraction method that there are some special cases of elements connection providing sufficient simplicity of a network function calculation process. In addition to equivalent transformations of pathological elements considered above there are many topological rules that can be helpful for a symbolic analysis [11, 14, 19, 20]. They are presented in Table I.

TABLE I. THE CIRCUIT ELEMENTS IN SHORT-CIRCUIT AND IN OPEN LOOP

Element	**Special Case**	
	Short-circuit	***Open Loop***
Impedance	Parameter extracted – element deleted	Element shorted
Admittance	Element deleted	Parameter extracted – element shorted
Controlled voltage-source branch	$\Delta = 0$	Element shorted
Current-sensor branch	$\Delta = 0$	Element shorted
Controlled current-source branch	Element deleted	$\Delta = 0$
Voltage-sensor branch	Element deleted	$\Delta = 0$
Nullator	$\Delta = 0$	$\Delta = 0$
Norator	$\Delta = 0$	$\Delta = 0$
VM	$\Delta = 0$	$\Delta = 0$
CM	$\Delta = 0$	$\Delta = 0$

These special cases are the consequences of the implementation of the parameter extraction formula which generalizes the Feussner's equations [21, 22]:

$$\Delta = \chi\, \Delta(\chi \to \infty) + \Delta(\chi = 0), \qquad (1)$$

where χ is a parameter of arbitrary circuit element, $\Delta(\chi \to \infty)$ and $\Delta(\chi = 0)$ correspond to the determinants of the circuit matrix in which the parameter of extracted elements $\chi \to \infty$ or $\chi = 0$ respectively.

In case when $\chi \to \infty$ the selected element must be omitted from the circuit if χ is an impedance, and replaced by a short or by a nullor if χ is an admittance or a controlled source

respectively. In case when $\chi = 0$ the selected element must be replaced by a short if χ is an impedance, and deleted if χ is an admittance. If the selected element is a controlled source (CS) then in case when $\chi = 0$ a voltage source and a controlling current must be replaced by a short, but a current source and a controlling voltage must be deleted.

In special cases presented in Table I one of the obtained determinants $\Delta(\chi\to\infty)$ or $\Delta(\chi = 0)$ will be equal to zero in accordance with the topological conditions for solvability of linear networks proposed in [11, 19, 20]. If the first summand of (1) is omitted from the expression the extracted parameter χ will be omitted too. If the second summand is omitted from the expression the extracted parameter χ will become a multiplier of a corresponding circuit determinant.

D. Pathological elements extraction

The procedure of nullor extraction can be formalized by the following steps [10]:

1. The choice of a supporting node. As a supporting node is the only node connected to singular elements can be chosen. The supporting node corresponds to the common node of a nullator and a norator of the extracted nullor.

2. The supporting node is splitted into two nodes for the first of them to be connected only with nullators while the second node must be connected only with norators.

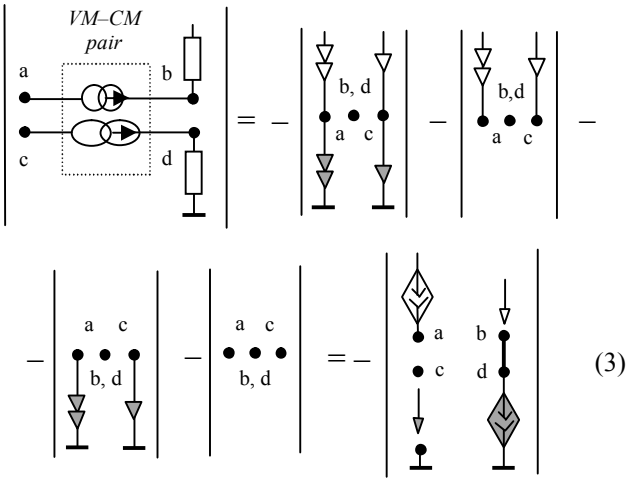
3. The branches corresponding to a nullator and norator of the extracted nullor are short-circuited.

The extraction rules of the nullor number n can be expressed by the following formula:

$$\Delta = \pm \Delta_n, \qquad (2)$$

where Δ_n is the determinant of the circuit after the procedure of the nullor number n extraction has been executed. Generally speaking, this equation means that the nullor extraction will change the sign of the initial determinant. The choice of the sign depends on the orientation of the nullor elements. If the norator and the nullator have the same orientation with respect to the supporting node the sign will be positive. In the opposite case the sign will be negative.

The generalization of formula (2) for the extraction of the pathological mirrors is illustrated below:



(3)

The meaning of the presented circuit transformation becomes clear if we compare it with the summation of two rows and two columns of a circuit matrix. It is obvious that these two procedures are isomorphic.

III. The Algorithm for a Symbolic Analysis of the Pathological Element-Based Circuits

In accordance with [16-18] the network function of a linear electronic circuit can be expressed as $\Delta N / \Delta D$, where ΔN is the determinant of the circuit, in which the independent source and an arbitrary response are replaced by a nullor, and ΔD is the determinant of the circuit, in which the input excitation and the arbitrary response are zero.

The procedure of the determinant calculation of the circuit with pathological elements is based on the following recursive algorithm:

1) equivalent transformation of two-ports elements (impedances in series and admittances in parallel);

2) search for special cases of elements connection (Table I);

3) equivalent transformation of the pathological elements in accordance with Fig. 4, Fig. 5 and Fig. 6;

4) extraction of nullors by means of formula (2);

5) extraction of pathological mirrors by means of formula (3);

6) parameter extraction by means of formula (1).

Note, that active networks must be checked by the rules presented in Table I after every change of circuit topology to avoid the violation of network solvability conditions.

The equivalent transformations of impedances in series and admittances in parallel must be done if possible to simplify the analysis process. The search for special cases of circuit elements connections (Table I) can simplify the calculation of circuit determinant even more. If circuit contain many pathological elements formulas (2) and (3) will be handy. Parameter extraction formula (1) must be used only when further topology simplicity is not possible. Eventually all parameters will be extracted from the initial network until we obtain one of the residual circuits presented in Fig. 7 [17].

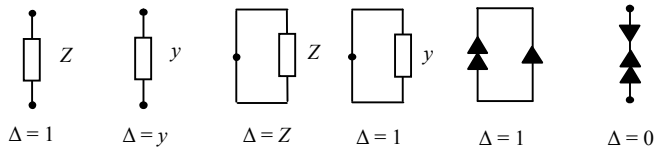


Fig. 7. The residual circuits and their determinants

Since the capacitors and the inductances can be presented as operational admittances and impedances respectively, so by means of the proposed method we can obtain the compact expression of the network function with a rational polynomial numerator and a denominator.

The approach discussed has been implemented in the program CIRSYM by V. Filaretov as a part of the software tool for the computer aided circuit design system SCADS. The demo-version of the program can be downloaded from the internet-site www.intersyn.narod.ru.

IV. Analysis Examples

To illustrate the proposed method, let us consider the symbolic analysis of the ICCII+-based inverting low-pass filter shown in Fig. 8 (a). The modified circuits for calculation of determinants of numerator ΔN and denominator ΔD of the network function $V_3/V_{in} = \Delta N / \Delta D$ are presented in Fig. 8 (b) and (c), respectively.

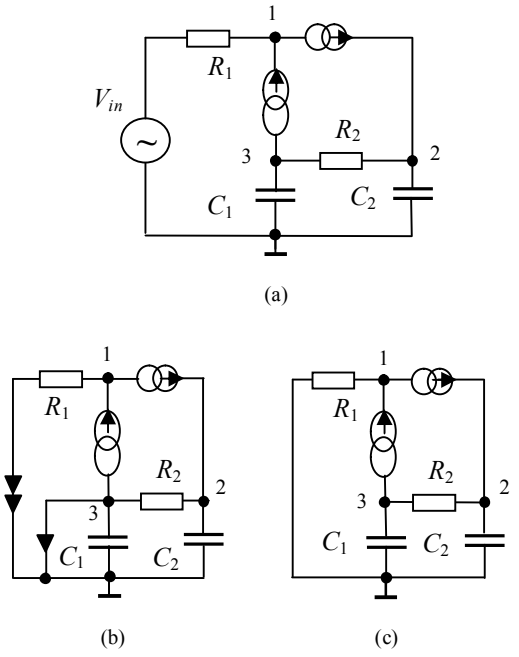


Fig. 8. Pathological element-based model of an inverting low-pass filter

The modified circuit for numerator3 calculation in Fig. 8 (b) will simplify because the topology including the resistor R_1 is connected in series with a norator and capacitor C_1 is connected in parallel with a nullator. In accordance with (1) and Table I a resistor must be shorted and capacitor must be deleted. Now we have to use the formula (1) for a parameter extraction of elements R_2 or C_2. In any case, both of the parameters will be omitted from determinant expression. Then we obtain the circuit consisting of only pathological elements. The determinant of this circuit can be easily calculated by means of a nullor extraction (3). The discussed procedures are illustrated by Fig. 9.

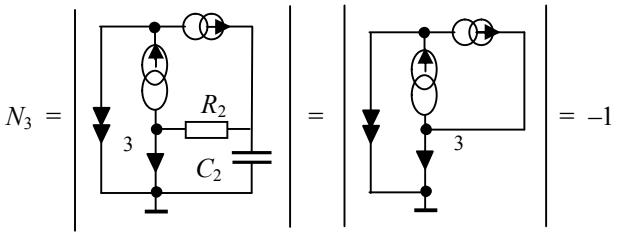
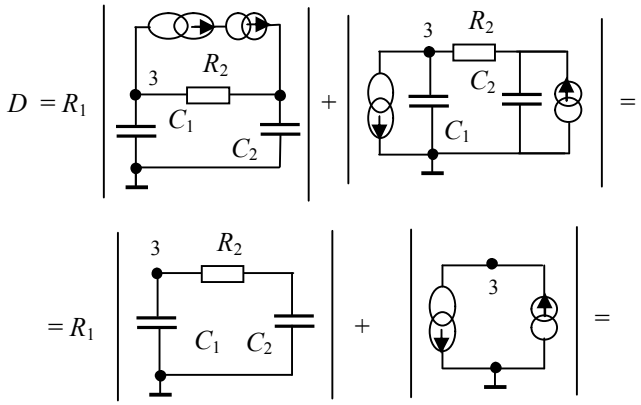


Fig. 9. The calculation of a numerator N_3

The extraction of R_1 from the modified circuit in Fig. 8 (c) leads to two subcircuits that satisfy the special cases of pathological elements connections. In the first one $R_1 \to \infty$ and the circuit can be simplified by deleting of VM-CM pair connected in series. The second circuit will include the operational admittance connected in parallel with pathological mirrors. After deleting of them VM and CM will be connected in series with R_2 and the circuit becomes a VM-CM pair grounded loop. So, the whole process of the denominator determinant calculation is illustrated by Fig. 10.



$$= R_1(R_2(p^2C_1C_2+p(C_1+C_2)))+1$$

Fig. 10. The calculation of a denominator D

The obtained expression of network function corresponding to the symbolic nodal analysis solution of the ICCII+-based inverting low-pass filter are presented in [8, 9]:

$$V_3/V_{in} = (-1)/(R_1(R_2(p^2C_1C_2+p(C_1+C_2)))+1). \quad (4)$$

The algorithm effectiveness has been tested by the analysis of the circuit presented in Fig. 11. The circuit consists of eight nodes and three VM–CM pairs.

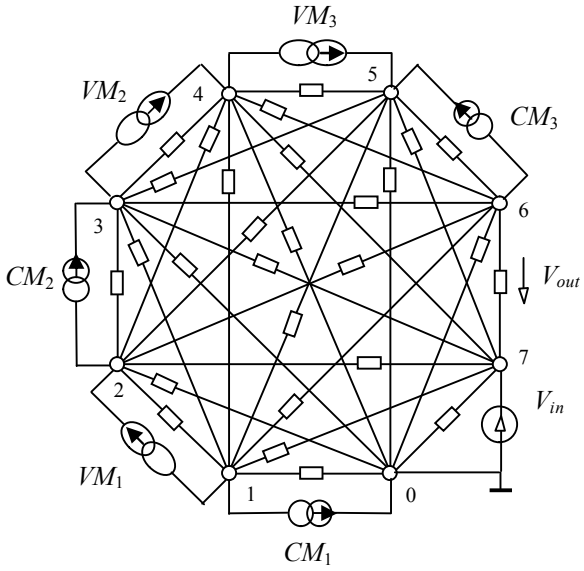


Fig. 11. The circuit for the analysis algorithm test

The circuit net-list in Spice-like format is presented in Table II. Note, that VM-CM pairs are marked by "M".

TABLE II. CIRCUIT NET-LIST

```
Vin   7   1
M1    1   0   1   2
M2    2   3   3   4
M3    6   5   4   5
g1    0   1   1
g2    0   2   2
….
 g28  6   7   28
.end
```

The numeric results of the circuit analysis will be as follows:

$$V_{out}=N/D = (-3521142) / 3144968= -1.1196113919124137.$$

The calculation of V_{out} of circuit presented in Fig. 11 by means of a matrix approach done by doctor G. V. Mayko employing mathematic software Matlab leads to the same numeric results. The sufficient indexes of network function symbolic expressions obtained by different circuit analysis methods are presented in Table III.

TABLE III. EXPERIMENTAL RESULTS OF THE SCADS NETWORK FUNCTION GENERATION, COMPARED TO THE MATLAB

Indexes	Symbolic expressions calculated by	
	SCADS	Matlab
*	472	2383
+	951	700
–	182	189
(	726	0
)	726	0
number of terms	7794	11479

The proposed algorithm has been used for a symbolic analysis of many other pathological mirror-based networks. These analysis solutions can be downloaded from the internet-site www.intersyn.narod.ru.

V. CONCLUSIONS

The new topological method of network function calculation of networks with voltage mirror-current mirror has been proposed. The new equivalent transformations of pathological element-based network providing the simplicity of a symbolic circuit analysis are presented. The main advantage of the method is that it is cancellation free. There are no limitations related to the element types of linear networks. The proposed method has been implemented in the computer aided circuit design system SCADS.

REFERENCES

[1] B.D.H. Tellegen, "La recherche pour una sêrie complête d'êlêments de circuit ideaux non–linêaires," Rendiconti del seminario matematico e fisico di Milano: Sotto gli auspice dell'universitâ e del politecnico, vol. 25, Milano (1953–1954), 1955, pp. 134–144.

[2] H.J. Carlin, "Singular network elements," In IEEE Trans. Circuit Theory, vol. CT–11, no. 3, 1964, pp. 67–72.

[3] P. Kumar and R. Senani, "Bibliography on nullors and their applications in circuit analysis, synthesis and design," in Analog integrated circuits and signal processing, vol. 33, no. 1, Oct. 2002, pp. 65–76.

[4] A.M. Soliman and R.A. Saad, "On the voltage mirrors and the current mirrors," in Analog Integr. Circuits signal process, vol. 32, no. 1, Jul. 2002, pp. 79–81.

[5] A.M. Soliman and R.A. Saad, "Use of mirror elements in the active device synthesis by admittance matrix expansion," in IEEE Trans. on circuits and systems–I: regular papers, vol.55, no.9, 2008, pp. 2726–2735.

[6] A.M. Soliman and R.A. Saad, "The voltage mirror-current mirror pair as a universal element," in Int. J. circuit theory Appl., vol. 38, no. 8, Oct. 2010, pp. 787–795.

[7] A.M. Soliman, "Synthesis of oscillators using limit variables and NAM expansion," In Hindawi Publishing Corporation: Active and passive electronic components, vol. 2011, Article ID 131546, 13 p.

[8] E. Tlelo-Cuautle, C. Sánchez-López, E. Martínez-Romero, and S.X.-D. Tan, "Symbolic analysis of analog circuits containing voltage mirrors and current mirrors," Analog Integr. Circuits Signal Process, vol. 65, no. 1, Oct. 2010, pp. 89–95.

[9] C. Sanchez-Lopez, F.V. Fernandez, E. Tlelo-Cuautle, S.X.-D. Tan, "Pathological element-based active device models and their application to symbolic analysis," In IEEE Trans. on circuits and systems–I: regular papers, vol. 58, no. 6, 2011, pp. 1382–1395.

[10] V.V. Filaretov and A.S.Korotkov, "Generalized parameter extraction method in symbolic network analysis," in Proc. ECCTD, Kraków, Poland, vol. 2, pp. 406–409, Sept. 2003.

[11] V.V. Filaretov and A.S. Korotkov, "Generalized parameter extraction method in case of multiple excitation," in Proc. 8th Int. Workshop on Symbolic methods and applications to circuit design (SMACD`04), Wroclaw, Poland, pp.8–11, Sept. 2004.

[12] V.V. Filaretov and K.S. Gorshkov, "The Generalization of the Extra Element Theorem for Symbolic Circuit Tolerance Analysis," Journal of Electrical and Computer Engineering, vol. 2011, Article ID 652706, 5 p, April 2011.

[13] V.V. Filaretov and K.S. Gorshkov, "Transconductance Realization of Block-diagrams of Electronic Networks," in Proc. of International Conference on Signals and Electronic Systems (ICSES`08), Krakow, Poland, pp. 261–264, Sept. 2008.

[14] V.V. Filaretov and K.S. Gorshkov. Network synthesis by means of circuit approach. LAP LAMBERT Academic Publishing GmbH & Co. KG, 2011.

[15] V.V. Filaretov and K.S. Gorshkov, "A circuit synthesis technique based on network determinant expansion," in Proc. of International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD`12), Seville, Spain, pp. P. 293–296, Sept. 2012.

[16] J.Braun, "Topological analysis of networks containing nullators and norators," Electronics letters, vol.2, no. 11, pp. 427–428, Nov. 1966.

[17] M.E.Parten and R.H.Seacat, "Topological analysis of networks containing nullators and norators using residual networks," in Proc. 23th annual Southwestern IEEE conference and exhibition, Houston, Texas, USA, pp. 39–42, April 1971.

[18] R.Hashemian, "Symbolic representation of network transfer functions using norator-nullator pairs," Electronic circuits and systems, vol.1, no.6, pp.193–197, Nov. 1977.

[19] M.M. Milic, "General passive networks – solvability, degeneracies, and order of complexity," In IEEE Transactions on circuits and systems, vol. CAS–21, no. 2, March 1974, pp. 177–183.

[20] T. Ozawa, "Topological conditions for the solvability of linear active networks," In Circuit Theory and Applications, vol. 4, 1976, pp. 125–136.

[21] W. Feussner, "Ueber Stromverzweigung in netzformigen Leitern," Annalen der Physik, vol. 9, no. 13, pp. 1304–1329, 1902.

[22] W. Feussner, "Zur Berechnung der Stromstarke in netzformigen Leitern," Annalen der Physik, vol. 15, no. 12, pp. 385–394, 1904.

Non-Destructive Volume and Thickness Measurements with Planar Microwave Sensors

M. Al Jader, O. Korostynska, A. Mason and A. I. Al-Shamma'a
Built Environment and Sustainable Technologies (BEST) Research Institute
School of the Built Environment, Liverpool John Moores University
Byrom Street, Liverpool, L3 3AF, UK
o.korostynska@ljmu.ac.uk

***Abstract*—This paper reports on the development of novel microwave sensors for non-destructive volume and thickness measurements in real-time. The sensors are in the form of metal patterns printed on planar substrates, and the type of both the metal and the substrate used alters the sensitivity of the sensor. In particular, changes in the response of these sensors in the microwave region of the electromagnetic spectrum when in contact with deionised water of 10-350 µl volumes and polyvinyl cellulose in 0.15-0.60 mm thicknesses are used as an indicator of volume/thickness respectively. Resonant peaks change their frequency and amplitude depending on the type and amount of material brought in contact with the sensor. This method offers increased sensitivity, real-time response and non-destructive evaluation of material type and profile, which can be used for structural health monitoring, for concealed target detection and security, and in a range of biomedical applications.**

Keywords—microwave sensor; real-time measurements; silver and gold printed patterns; resonant frequency shift; signature spectra.

I. INTRODUCTION

Traditional methods of measuring ultra-small volumes and thickness of materials with high accuracy for quality control are lab-based and include standard UV-Vis measurements, mass spectrometry, amperometric sensors, fibre-optic sensors and MEMs [1]. However, when it comes to an in-situ real-time method capable of non-destructive assessment of material thickness profile, there is none readily available.

Microwave sensing is a novel but rapidly developing technology which has been successfully used as a sensing method for various industrial applications including solution concentrations [2], fluid level measurements [3], material moisture content [4, 5], for continuous process monitoring of biogas plants [6], for the determination of moisture content in soil [7] and in the healthcare industry, for example for real-time monitoring of glucose in diabetic patients [8, 9] and for non-invasive monitoring of bodily fluids [10].

Microwaves are largely used for material characterisation since they easily propagate through low-loss dielectrics and the amplitude of the electromagnetic wave reflected by or transmitted through a material obstacle strongly depends on the dielectric properties of the material itself.

A microwave system is able to adequately trace multiple phases in a complex fluid flow. For example, an electromagnetic waves cavity resonator was used to monitor the percentage volumes of each of the two phases (oil and gas) in the pipeline using the resonant frequencies shifts that occur within the resonator [11].

However, the microwave planar printed patterns for various sensing applications are increasingly used due to their versatility, flat profile and low weight. Their design can be tailored to suit particular applications, coupled with reliability and cost-efficiency. They are easily manufactured using common methods for printed circuit board production, and their impedance can be matched to the input line by altering the micro-strip line feed configuration. The patch antenna represents the frequency-selective element of a phase shift transistor oscillator. The active integrated antenna frequency of operation is determined by both the patch geometry and the electrical loads connected at the two microstrip ends [12].

For example, a coaxial-fed patch antenna suitable for non-destructive porosity measurements in low-loss dielectric materials has been reported [13]. The variation of the patch resonant frequency when it is put on the surface of the material under test was used to estimate the dielectric permittivity at 2.4 GHz ISM frequency bands. The estimated porosity was in good agreement with that obtained by the conventional mechanical measurements, and the mean percent error was less than 13.5%.

To clarify the principle behind electromagnetic wave sensing suggested in this work for volume/thickness measurements, it is worth mentioning that microwave sensors in the form of planar printed patterns operate based upon the fact that an object under test, e.g. a water sample or piece of plastic, when placed into the vicinity or in direct contact with a microwave sensor, interacts with the electromagnetic waves in a unique manner, which can be specifically correlated with the properties of this material. In particular, the sensing is based on interaction of propagating or resonating modes within the material under test. By considering how reflected (S_{11}) microwave signals vary at

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

discrete frequency intervals, the change in the signal can be linked to the amount and type of material under test. The response of the sensor manifests itself as a resonant frequency change, attenuation of the signal or as a phase shift.

II. EXPERIMENTAL PROCEDURE

A. *Microwave Sensors Structure*

The design of any microwave sensor is the key parameter that regulates its performance. Sensors with structure shown in Fig. 1 (a) operating at microwave frequencies were chosen for their versatile design that combines ease of manufacturing with the desired functionality. In this work, Au and Ag patterns in the form of interdigitated electrodes were printed on FR4 substrate with 1.6 mm thickness and on DuPont™ Pyralux® AP polyimide flexible laminate, with 2 mil thickness, substrates respectively to make the microwave sensors. The thickness of both Ag and Au layers in all the sensors was 35 µm, and the bottom layers acted as ground plane. Notably, Ag patterns are suitable for measuring chemically inert substances, but Au may be preferable when measuring solutions with biologically active components, as it is the intention to develop a sensing system for biomedical and diagnostic applications. The sensing area was 5×8 mm^2.

Fig. 1 (b) and (c) illustrate optical images of the manufactured prototype microwave sensor on polyimide laminate substrate and on a standard FR4 substrate respectively. The barrier/well was also constructed from FR4 material to provide a housing for water samples, with a maximum capacity of 350 µl.

A distinct feature of the microwave planar sensors is their superior sensitivity to change close to the sensor surface, with this sensitivity decaying rapidly with distance away from the surface, as shown in Fig. 2, which illustrates the distribution of electromagnetic field emitted by the sensor as simulated using HFSS software. This is advantageous as it reduces significantly the chance of undesirable external factors influencing the sensor response.

B. *Measurement Setup*

The microwave sensors were attached to a Rohde and Schwarz ZVA24 vector network analyser (VNA) via a coaxial cable. Molex edge type SMA connectors [14] where used to connect the sensors via coaxial cable to the VNA. The sensor and associated equipment were all specified for 50Ω impedance.

The VNA used for the purposes of data acquisition from the sensors, with this unit being appropriately calibrated according to manufacturer specifications, is shown in Fig. 3. The data (60,000 points for each measurement) was captured in the frequency range of 1-15 GHz for the reflected (S_{11}) signals. All the measurements were performed at a constant temperature of 18 °C. Each sample was measured at least 5 times and the results were repeatable with less than 5% deviation and thus were deemed reproducible. Notably, average sensor responses are depicted in the graphs shown in following section.

The response of each sensor was measured for air and then when in contact with varying volumes of deionised water (DIW) and polyvinyl cellulose (PVC) plastic material to evaluate if the sensors respond differently. The spectra from each sensor are plotted on common graphs to illustrate that each sample has a unique response to the microwave signal resulting in resonant peaks occurring at different frequencies and this particular feature makes the developed sensors an attractive option for real-time monitoring of water volume and material thickness brought in contact with it.

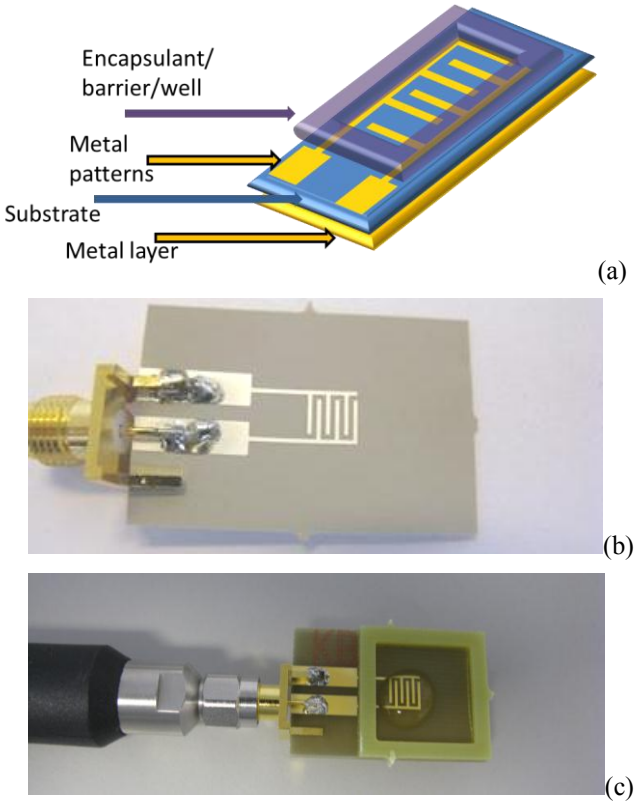


Fig. 1. Microwave sensor with metal pattern: (a) schematic structure of the sensor; (b) Ag pattern on polyimide flexible laminate substrate; (c) optical image of Au pattern printed on FR4 substrate with SMA connector.

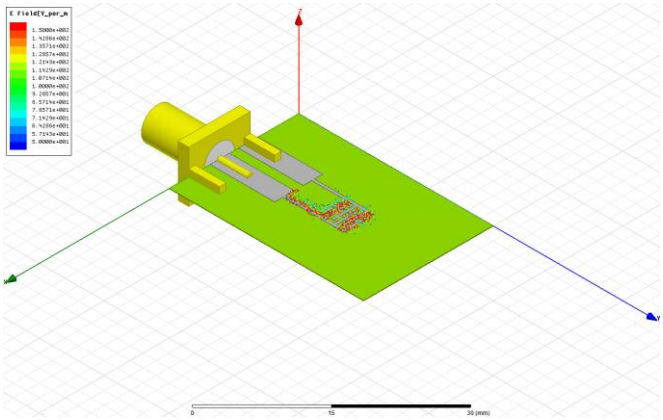


Fig. 2. HFSS model of microwave sensor field.

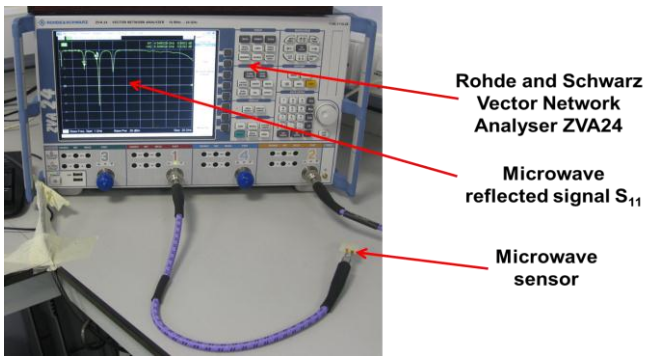


Fig. 3. Measurement setup showing VNA and a microwave sensor connected via coaxial cable.

III. Results and Discussion

Experimental trials of the planar type sensor printed on DuPont™ Pyralux® AP Polyimide Flexible Laminate substrate verified its capability to accurately determine the amount of fluid, in this case deionised water, placed on top of the sensor. Fig. 4 presents the S_{11} response of this sensor to increasing volume of water in 10-100 µl volume, focusing on 1.5-2.5 GHz range, as it shows the change in the resonant peak frequency and amplitude. As one may see, there is a distinct monotonic decrease in both the resonant peak amplitude and frequency shift recorded in the 2-3 GHz frequency range. Water volume was controlled by two approaches: (1) by gradually adding 10 µl volume till 100 µl was reached and measuring the microwave spectra each time, and alternatively (2) by placing the exact amount of fluid, having previously removed any residue water. Each time the response was identical. Notably, the sensor is extremely sensitive to microliter volumes of fluid, but the response becomes saturated at around 50 µl.

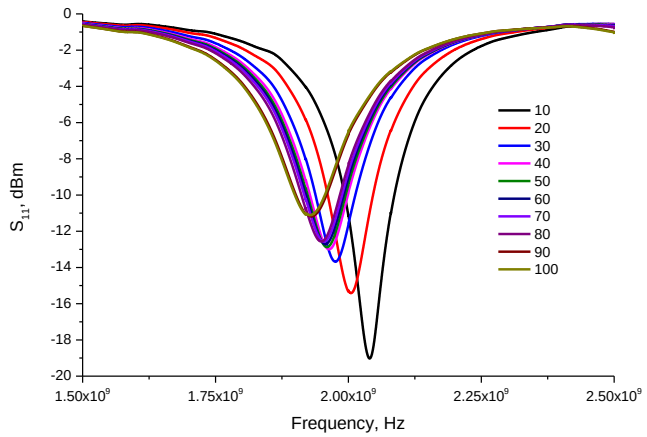


Fig. 4. Dependence of S_{11} signal on the volume of deionised water, given in µl, brought it contact with Ag pattern printed on flexible microwave sensor, 1.5-2.5 GHz range.

Microwave sensors manufactured on FR4 substrate were also tested for their ability to measure the amount of solution in the vicinity of the electromagnetic field emitted by Au patterns. The change of substrate material allowed increasing the amount of liquid which can be determined using this method. In particular, the volume of deionised water was gradually increased from 50 µl to 350 µl. A well around the Au pattern was used to keep the fluid over the sensor and 350 µl was the maximum amount that it would hold. The microwave spectra were recorded in 1-15 GHz frequency range, but for measuring purposed it is easier to trace the change in the resonant peaks behaviour, as for example the change in resonant peak frequency or its amplitude can be used as a calibration curve for computerised real-time measurement, providing superior precision. Fig. 5 illustrates the dependence of S_{11} signal on the volume of deionised water, given in µl, brought it contact with Au pattern on FR4 substrate, focusing on 2.0-3.0 GHz range. As in the case with the sensor on flexible substrate, gradual monotonic decrease was recorded for the amplitude of the resonant peak with increasing amount of water, accompanied by a distinct shift in resonant frequency towards lower values. Similarly, at high volumes of fluid the sensor response was saturated and became less pronounced.

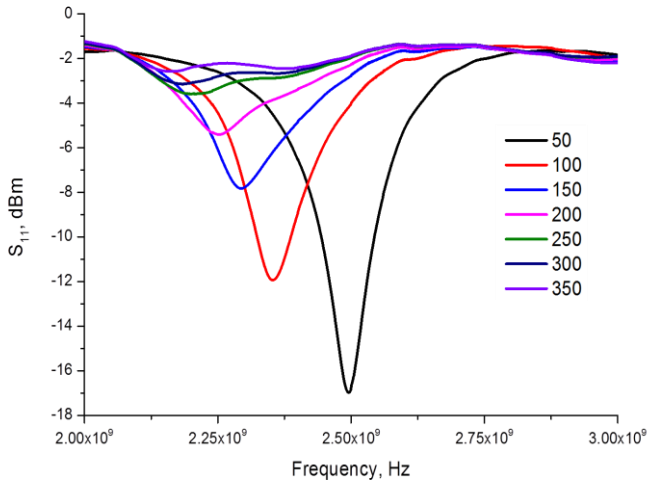


Fig. 5. Dependence of S_{11} signal on the volume of deionised water, given in µl, brought it contact with Au pattern on FR4 substrate, 2.0-3.0 GHz range.

The ability of the developed microwave sensors to non-destructively measure the thickness of the material, having the access from one side only, was verified by placing pieces of PVC on top of the sensors, with their thickness ranging from 0.15 mm to 0.60 mm. Having examined 1-15 GHz range, two distinct areas were focused on, since they gave the most pronounced change in the response, namely the focus was on the two resonant peaks in 1.05-1.15 GHz and 5.0-8.0 GHz frequency ranges, as shown in Fig. 6 and Fig. 7 respectively. Substantial changes in the microwave spectra caused by the varying PVC thickness were recorded and these can be used as an accurate measure of unknown material profile. Such sensor would be invaluable in accessing the integrity of geometrically complex structures, as the method is non-destructive, it provides real-time responses and the microwave power used is in the order of 1 mW, which is non-ionising radiation, less than commonly used even in mobile phones, and thus has negligible deleterious effect on the material being measured. There is simply no alternative to the proposed approach that would compete in terms of accuracy, efficiency and reliability of measurements.

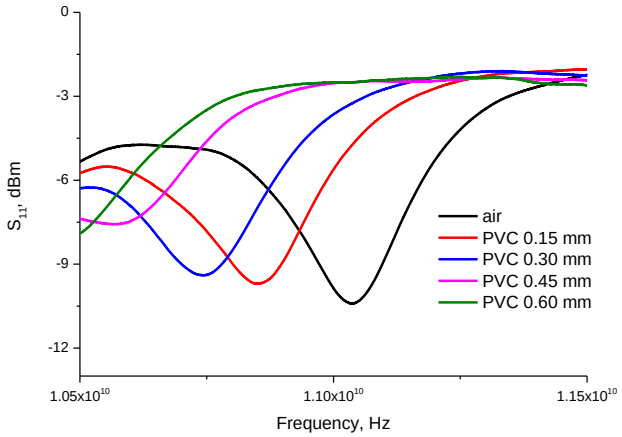


Fig. 6. Dependence of S_{11} signal on the thickness of PVC, placed on top of the sensor with Au pattern on FR4 substrate, 1.05-1.15 GHz range.

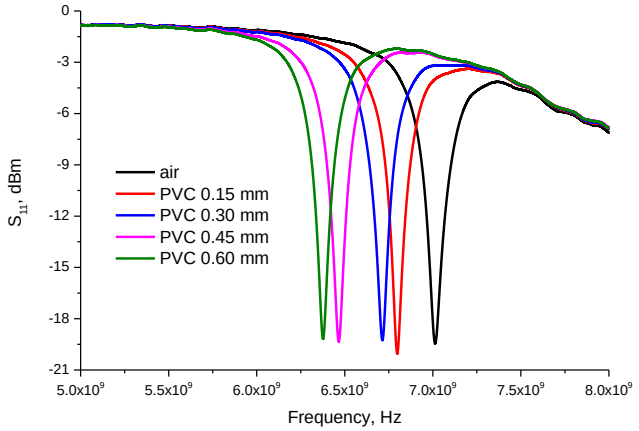


Fig. 7. Dependence of S_{11} signal on the thickness of PVC, placed on top of the sensor with Au pattern on FR4 substrate, 5.0-8.0 GHz range.

Notably, the sensors' responses returned to their original positions, namely air spectra, after each water volume / PVC thickness measurement, confirming that the developed microwave sensors are reliable, re-usable and thus a sustainable solution for precise in-situ real-time method for monitoring structural integrity of material, thickness profile, or volume of fluids, essential for a broad range of applications.

As was mentioned earlier, the design of the sensor, including its dimensions and substrate composition, plays a pivotal role in determining its sensitivity. Thus, if higher volumes of fluid are required to be measured, a microwave sensor with different dimensions and pattern layout can be constructed to suit the needs of a particular application. For the thickness measurements, the response of the sensors to other materials is being explored to build a database which can later be used not only to examine a profile of complex structures, but to reveal the material type itself, as each material produces a unique signature pattern in microwave spectrum.

IV. Conclusion

This research was driven by the industrial need for a novel real-time non-destructive method of measuring extra small fluid volumes and material thickness profile. Microwave sensors were developed for this purpose and successfully tested in the 1-15 GHz frequency range. It was clearly shown that the resonant peaks shift and their amplitude decreased once the deionised water and PVC of increasing volumes / thickness were placed in contact with the sensor pattern. Notably, these sensors are reliable and re-usable since their response, represented by reflected signals S_{11} parameters was identical.

Acknowledgment

This work is financially supported by the European Community's Seventh Framework Programme through the FP7-PEOPLE-2010-IEF Marie-Curie Action project 275201, Water-Spotcheck. The authors would like to express their gratitude to Mike Garner, Garner Osborne Circuits Limited, UK and to Mabruk Farrah, Mechan Controls Plc, for manufacturing sensor prototypes.

References

[1] O. Korostynska, A. Mason, and A. I. Al-Shamma'a, "Monitoring of Nitrates and Phosphates in Wastewater: Current Technologies and Further Challenges," *International Journal on Smart Sensing and Intelligent Systems,* vol. 5, no. 1, pp. 149-176, March, 2012.

[2] B. Kapilevich, and B. Litvak, "Microwave sensor for accurate measurements of water solution concentrations." pp. 1-4.

[3] J. D. Boon, and J. M. Brubaker, "Acoustic-microwave water level sensor comparisons in an estuarine environment." pp. 1-5.

[4] B. Jackson, and T. Jayanthy, "A novel method for water impurity concentration using microstrip resonator sensor." pp. 376-379.

[5] C. Bernou, D. Rebière, and J. Pistré, "Microwave sensors: a new sensing principle. Application to humidity detection," *Sensors and Actuators B: Chemical,* vol. 68, no. 1–3, pp. 88-93, 2000.

[6] T. Nacke, A. Barthel *et al.*, "Continuous process monitoring for biogas plants using microwave sensors." pp. 239-242.

[7] K. Y. You, J. Salleh, Z. Abbas *et al.*, "A rectangular patch antenna technique for the determination of moisture content in soil," *PIERS 2010 Cambridge - Progress in Electromagnetics Research Symposium, Proceedings.* pp. 850-854.

[8] O. Korostynska, A. Arshak, P. Creedon *et al.*, "Glucose monitoring using electromagnetic waves and microsensor with interdigitated electrodes." pp. 34-37.

[9] A. Mason, S. Wylie, A. Thomas *et al.*, "HEPA Filter Material Load Detection Using a Microwave Cavity Sensor," *International Journal on Smart Sensing and Intelligent Systems,* vol. 3, no. 3, pp. 16, Sep, 2010.

[10] A. Al-Shamma'a, A. Mason, and A. Shaw, *Patent: Non-Invasive Monitoring Device*, US2012150000 (A1), WO2010131029 (A1), EP2429397 (A1), 2012.

[11] M. A. Al-Kizwini, S. R. Wylie, D. A. Al-Khafaji *et al.*, "The monitoring of the two phase flow-annular flow type regime using microwave sensor technique," *Measurement: J. Int. Measurement Confederation,* vol. 46, no. 1, pp. 45-51, 2013.

[12] F. Völgyi, "Microstrip Transmission- and Reflection-Type Sensors Used in Microwave Aquametry," *Electromagnetic Aquametry*, K. Kupfer, ed., pp. 243-256: Springer Berlin Heidelberg, 2005.

[13] A. Zucchelli, M. Chimenti, E. Bozzi *et al.*, "Application of a Coaxial-fed Patch to Microwave Non-Destructive Porosity Measurements in Low-Loss Dielectrics," *Progress In Electromagnetics Research M,* vol. 5, pp. 1-14, 2008.

[14] D. Guha, and Y. M. M. Antar, *Microstrip and Printed Antennas: New Trends, Techniques and Applications*: Wiley, 2010.

Fractal Spline Signals: Synthesis and Estimation

Ihor Shelevytsky
Informatics and applied mathematics department
Kryvyi Rih National University
Kryvyi Rih, Ukraine, 50086
sheleviv@gmail.com

Olga Novikova
Electronics department
National Aviation University
Kiev, Ukraine, 03058
greeneyed.kr@gmail.com

Abstract — **In this work the fractal signal constructed on the base of linear and cubic Hermite splines is considered. Spline shape at a large scale and shape of its fragments are the same. Small number of parameters allows to obtain a complex waveform. The methods of constructing of such signals and their evaluation at the presence of noise are considered.**

Keywords— DSP; fractal; spline; signal

I. INTRODUCTION

Improving of radio systems creates new requirements of the signals. Compound wideband signals are used for sensing, communication, protection and masking. Wideband fractal signals are such kind of signals. A.A.Potapov [5] has considered the features and advantages of fractal signals in RF technology in details. P. Abry and others [1] have represented the relationship between fractals and wavelets. M.A.Navascues [2] has investigated the mathematical aspects of fractal splines interpolation. Also advantages of fractal signals application in the condition of intense noise were investigated [6].

There are a lot of kinds of fractal signals. Such as signals modulated by fractal sequences (A.A.Potapov [5]), fractal wavelets (V.M.Bolotov, Y.V.Tkach, Y.Y.Tkach [7]), fractal ultra wideband RF pulses (O.V.Lazorenko, L.F.Chernohor [4]) and others. Fractal splines developed by different authors can be distinguished by the way of construction, types of splines or wavelets, different areas of application. Wavelets and fractals are similar, because they both maintain shape similarity for different scales as a result of scaling. Spline wavelets are the kind of wavelets. However, wavelet decomposition and synthesis are based on the interpolation, making them difficult to use for small signal to noise ratio and hard obtain statistical estimates.

II. PROBLEM STATEMENT

Fractal spline is a function that consists of spline functions of various scales that maintain the property of self-similarity. As a regular spline, spline fractal is characterized by power, number of nodes, boundary conditions. As a fractal, spline fractal has such characteristics as the number of scales and fractal dimension.

Let's choose Hermite algebraic splines of first and third order to construct a fractal spline. Such splines are simple in construction and their coefficients coincide with the values at the nodes. Linear spline is a polygonal curve and provides a continuity of values. Cubic Hermite spline provides continuity of values and first derivative.

Let's consider the number of enclosed levels of self-similar splines as a scale. In order to become a fractal it is necessary for each of the R fragments to be also splines that are similar in shape to the original. Then the spline of the k-th scale consists of R^{k+1} fragments. Division of each spline fragment maintains proportion of fragments of zero scale. For simplicity let's assume that fragments of splines have identical width. To ensure continuity of fractal spline boundary conditions are periodic.

III. SYNTHESIS OF FRACTAL HERMITE SPLINE

Suppose that Hermite spline at the interval $x \in [a;b]$ is given by the polynomial [8]:

$$S(x) = \sum_{j=0}^{R} a_j H_j(x), \quad (1)$$

with a_j as the numerical coefficients that are equal to the values of the spline nodes, $H_j(x)$ as the basis Hermite splines, R as the number of spline fragments.

${}^0S(x)$ is a spline of zero scale calculated by (1). Then the spline of k-th scale on an arbitrary fragment is:

$${}^kS(x) = {}^0S(m_k x), \quad (2)$$

with m_k as scale factor.

Fractal spline on the interval $[a,b]$ of k-th scale is a sequence R^k of splines (2). Fractal spline as a whole is the sum of splines of different scales, but a fragment of such spline has the shape of any other fragment except the zero scale. Any fragment of such spline is similar to the spline of zero scale.

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

So:

$$ {}^{k}S(x) = \begin{cases} {}^{k}S_1(x), x \in [a, h_k), \\ {}^{k}S_2(x), x \in [h_k, 2h_k), \\ \dots \\ {}^{k}S_R(x), x \in [(R-1)h_k, b), \\ 0, x \notin [a, b] \end{cases} . \quad (3) $$

Then the fractal spline of the k-th scale is:

$$ {}^{k}Sf(x) = \sum_{j=0}^{k} {}^{j}S(x) . \quad (4) $$

As at every scale and at every fragment the numerical coefficients a_j of (1) remain the same, expression (4) can be transformed as:

$$ {}^{k}Sf(x) = \sum_{j=0}^{R} a_j \, {}^{k}Hf(x) , \quad (5) $$

with ${}^{k}Hf(x)$ as fractal Hermite basis spline with k-th scale depth.

Let's consider the solution of the problem with a discrete set of points: $x_i = i \cdot Ts$, with Ts as sampling period. This situation occurs at the processing of digital signals and corresponds to the interpolation of node values by the fractal spline. At the same the grid of interpolated values is agreed with grid of spline nodes of the smallest scale. That is, at any spline scale and at any fragment integer number of sampling periods is located. Then N points of fractal spline are defined by R parameters which are the node values of the spline at zero scale:

$$ N = (b-a)/Ts \quad (6) $$

In matrix form a column vector of values of fractal spline is:

$$ \mathbf{S} = {}^{0}\mathbf{S} + \dots + {}^{k}\mathbf{S} = ({}^{0}\mathbf{P} + \dots + {}^{k}\mathbf{P}) * \mathbf{A} = \mathbf{P} * \mathbf{A} , \quad (7) $$

with $\mathbf{P}$ as a planning matrix of $(N \times R)$ dimension, A as a matrix of coefficients of $(1 \times R)$ dimension.

Values of columns of the matrix $\mathbf{P}$ are the values of fractal basis spline ${}^{j}Hf(x)$. Because of the local properties of Hermite basis spline matrices ${}^{0}\mathbf{P}, \dots, {}^{k}\mathbf{P}$ are block-diagonal. Obviously, these matrices are also self-similar (fig.1).

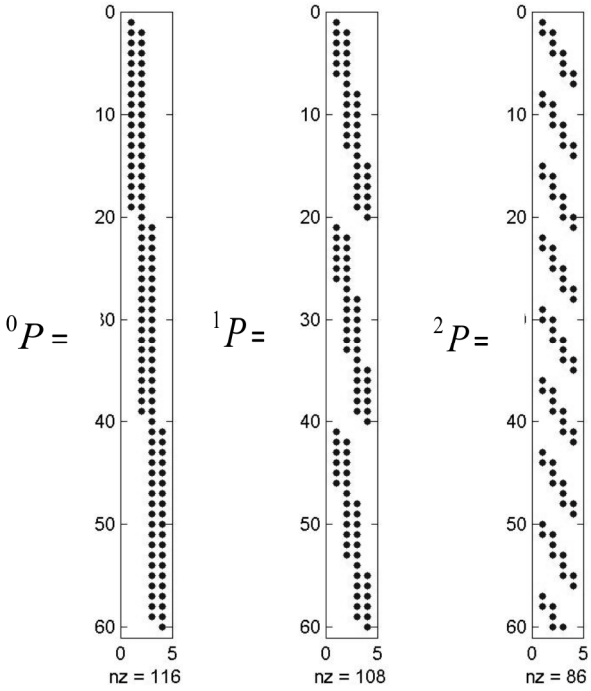


Fig. 1. Structure of planning matrices (N=60, R=5): a – zero scale, b – first scale, c – second scale (nz – amount of non-zero elements).

General view of fractal basis is shown on figure 2.

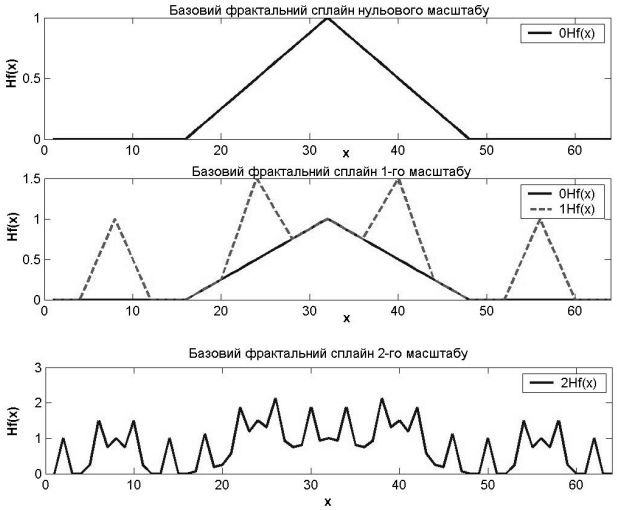


Fig. 2. Fractal basis splines: a – zero scale, b – first scale, c – second scale.

Let's consider as an example the construction of fractal splines for spline of zero scale with four fragments R = 4 on the interval [0; 64] with node points (0; 0), (16; 1), (32; -1), (48; 1), (64; 0). Constructed fractal spline has a scale depth scale equal to two. Applying the above method for linear spline fractal spline shown on figure 3 was constructed. Figure also illustrates the process of shape complication of fractal spline during scale increasing.

Fractal dimension for the linear spline ${}^2S(x)$ is calculated by the method described in [3, p.418] and it is equal to 1.5284.

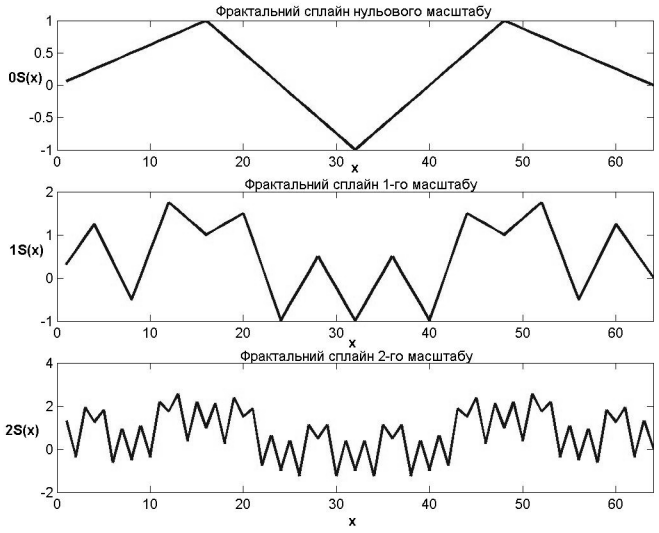


Fig. 3. Linear fractal spline with scale depth k = 2.

If to take as not linear but cubic spline as a basis, then the result is shown on figure 4. Fractal dimension for cubic fractal spline ${}^2S(x)$ is equal to 1.3635.

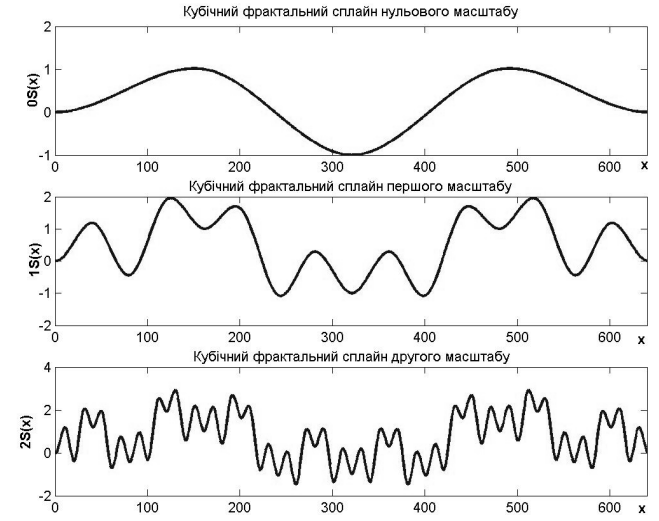


Fig. 4. Cubic fractal spline with scale depth k = 2.

Process of scale increasing for fractal can be continued and practically it is limited only by the ratio between the length of the spline of zero scale (signal duration) and by sampling period of discrete data at the conditions (6).

IV. Estimation of Fractal Hermite Splines

Estimation of parameters of fractal spline can be considered as a problem inverse to the interpolation. Knowing values at the points of interpolation (perhaps with an error) values in nodes that are parameters should be found. Taking into account practical existence of additive noise that with sufficient accuracy can be considered as white noise the method of least squares can be used.

Assume that the empirical data is the sum of the values of a fractal θ and uncorrelated random component ε:

$$Y = \theta + \varepsilon . \tag{7}$$

Type of fractal functions (system of basis functions) is known or with reasonable accuracy is approximated by selected fractal spline:

$$\| \theta - \mathrm{PA} \| \le \delta . \tag{8}$$

Solution of "inverse problem" with the help of least-squares method is based on well-known method of estimates of parameter vector of spline finding:

$$\hat{\mathbf{A}} = (\mathbf{P}^{\mathrm{T}}\mathbf{P})^{-1}\mathbf{P}^{\mathrm{T}}\mathbf{Y} . \tag{9}$$

Fundamentally evaluation by LSM for fractal spline doesn't differ from the general case. Feature is a specific form of the matrix $\mathbf{P}$. Sparsity of the matrix and its specific shape potentially can organize calculations effectively.

Specificity of fractal spline signal takes into account not only the approximation of the data in general at the maximum scale, but equally takes into account each individual scale.

Let's replace the requirement of minimization:

$$(\mathbf{Y} - \mathbf{PA})^T (\mathbf{Y} - \mathbf{PA}) \to \min , \tag{10}$$

by:

$$\left(\begin{bmatrix}\mathbf{Y}\\ \mathbf{E}\end{bmatrix} - \begin{bmatrix}{}^0\mathbf{P}\\ {}^1\mathbf{P}\end{bmatrix}\mathbf{A}\right)^T \left(\begin{bmatrix}\mathbf{Y}\\ \mathbf{E}\end{bmatrix} - \begin{bmatrix}{}^0\mathbf{P}\\ {}^1\mathbf{P}\end{bmatrix}\mathbf{A}\right) \to \min , \tag{11}$$

with

$$\mathbf{E} = \mathbf{Y} - {}^0\mathbf{PA} . \tag{12}$$

Then

$$\begin{bmatrix}\mathbf{Y}\\ \mathbf{E}\end{bmatrix} = \begin{bmatrix}\mathbf{Y} - \mathbf{ZA}\\ \mathbf{Y} - {}^0\mathbf{PA}\end{bmatrix} = \begin{bmatrix}\mathbf{Y}\\ \mathbf{Y}\end{bmatrix} - \begin{bmatrix}\mathbf{Z}\\ {}^0\mathbf{P}\end{bmatrix}\mathbf{A} = \tilde{\mathbf{Y}} - \mathbf{XA} , \tag{13}$$

with $\mathbf{Z}$ as non-zero matrix of proper dimensions.

Then:

$$\left(\begin{bmatrix}\mathbf{Y}\\ \mathbf{E}\end{bmatrix} - \begin{bmatrix}{}^0\mathbf{P}\\ {}^1\mathbf{P}\end{bmatrix}\mathbf{A}\right) = \left(\begin{bmatrix}\mathbf{Y}\\ \mathbf{Y}\end{bmatrix} - \begin{bmatrix}\mathbf{Z}\\ {}^0\mathbf{P}\end{bmatrix}\mathbf{A} - \begin{bmatrix}{}^0\mathbf{P}\\ {}^1\mathbf{P}\end{bmatrix}\mathbf{A}\right),$$

$$\left(\begin{bmatrix}\mathbf{Y}\\ \mathbf{E}\end{bmatrix} - \begin{bmatrix}{}^0\mathbf{P}\\ {}^1\mathbf{P}\end{bmatrix}\mathbf{A}\right) = \begin{bmatrix}\mathbf{Y}\\ \mathbf{Y}\end{bmatrix} - \begin{bmatrix}\mathbf{Z} + {}^0\mathbf{P}\\ {}^0\mathbf{P} + {}^1\mathbf{P}\end{bmatrix}\mathbf{A} . \tag{14}$$

Then expression (10) can be transformed as:

$$(\tilde{\mathbf{Y}} - \tilde{\mathbf{P}}\mathbf{A})^T (\tilde{\mathbf{Y}} - \tilde{\mathbf{P}}\mathbf{A}) \to \min , \tag{15}$$

with

$$\tilde{\mathbf{P}} = \begin{bmatrix} {}^0\mathbf{P} \\ {}^0\mathbf{P} + {}^1\mathbf{P} \end{bmatrix}.$$

This criterion minimizes the mean square deviation of the spline of zero scale and deviation of the spline of first scale from the remains of the approximation. Process described by formulas (11-15) can be extended to k scales. Taking into account the next scale the following matrix can be obtained:

$$\tilde{\mathbf{P}} = \begin{bmatrix} {}^0\mathbf{P} \\ {}^0\mathbf{P} + {}^1\mathbf{P} \\ {}^0\mathbf{P} + {}^1\mathbf{P} + {}^2\mathbf{P} \end{bmatrix}. \quad (16)$$

The proposed criterion can be modified, taking into account the importance of certain scales. For example, a specific scale can be important according to the level of noise in its frequency.

For example, let's estimate fractal cubic Hermite spline with depth scale of three by LSM. Fractal spline of zero scale has nodes {0,16,32,48,64} and values at the nodes {1.10, -10,10,1}. The obtained values were summed with uncorrelated random numbers normally distributed with unit variance. The results are represented on figure 5.

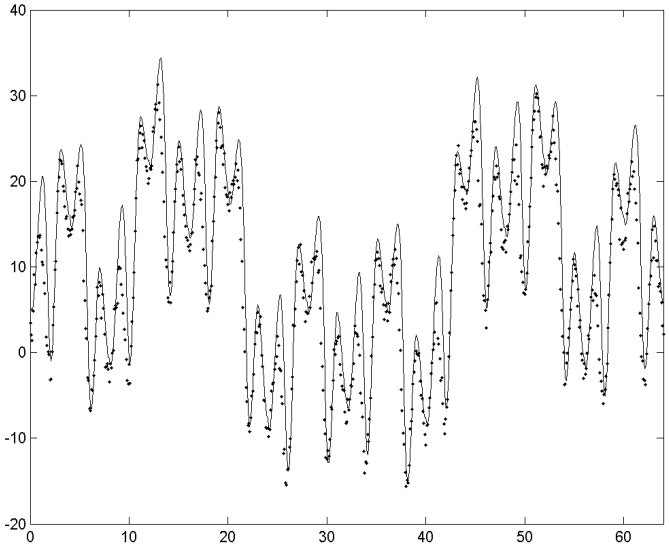


Fig. 5. Evaluation of model fractal spline with normal additive white noise. The points denote the initial data, solid line - resulting fractal spline (estimate).

RMS error of estimation evaluation priori defined parameters was about 0.3. Numerical experiments show that the accuracy of estimation of fractal signal is several times higher than estimates obtained for the same signal - spline.

V. Conclusion

The proposed model of the fractal spline signal is easy to synthesize and allows to specify the signals of complex shape with a small number of parameters. Linear parameters pendence of spline fractal allows to use LSM for parameter estimation of fractal signals with noise. Simulated experiments show better robustness of fractal spline signals.

The proposed fractal splines are simpler for synthesis and evaluation than fractal signal based on the Weierstrass function and similar functions.

Numerical experiments were conducted for ideal conditions. Possible mismatch between the location of nodes of model signal and nodes evaluated by fractal spline was not considered. Mismatch of nodes may be the result of signals delay. There also may be differences in the amplitudes of the splines on different scales at their distribution in different conditions for different frequencies. This requires further research.

It should be noted that only the simplest version of fractal spline with identical parameters on all scales and fragments of equal width was considered. Construction of fractal splines with arbitrary width of fragments with different types of bases at different scales, with different parameters on different scales is possible. All this allows to adapt the proposed model to a wide range of possible practical applications.

References

[1] Abry P., Goncalves P., Vehel J.-L. Scaling Fractals and Wavelets. – Wiley, 2009.

[2] Navascues M.A., Sebastian M.V. Fractal Splines. Monografias del Seminario Matematico Garcia de Galdeano 33 (2006), p. 161-168.

[3] Owen D.J., Yuan Shen. Estimating the Hurst index of a self-similar process via the crossing tree. Signal Processing Letters, 11, 2004, pp. 416-419.

[4] Chernogor L. F., Lazorenko O. V. and Potapov A. A. Wavelet Analysis of Model Fractal Ultra-Wideband Signals // Proc. 6 th Int. Conf. "Ultrawideband and Ultrashort Impulse Signals" (17 – 21 September, 2012, Ukraine, Sevastopol).- Kharkiv: Kharkiv IEEE Joint Chapter end KNU, 2012. P. 291 – 293.

[5] Potapov A.A. Fractals in RF physics and radars.- M.: Logos, 2002.- 664 p.

[6] Potapov A.A. Bolotov V.N., Tkach Y.V. The selection of fractal signals in a complex electromagnetic environment. // Electromagnetic Phenomena, Part 3. - №2, 2003. – p. 211-227.

[7] Bolotov V.N., Tkach Yu.V.,Tkach Ya.Yu. Synthesis and generation of fractal information signals. 12th Int. Conference "Microwave & Telecommunication Technology" (CriMiCo'2002). 9"13 September, Sevastopol, Ukraine. 2002. P. 247-248.

[8] I.V. Shelevytsky "Methods and means of spline-technology of digital signal processing", Kryvyi Rih European University, 2002 (In Ukrainian)

Short-Range C-Band Noise Radar for Meteorological Application

Lukasz Maslikowski and Krzysztof Kulpa
Warsaw University of Technology
Institute of Electronic Systems
Nowowiejska, 15/19, 00-665 Warsaw, POLAND
Email: l.maslikowski@stud.elka.pw.edu.pl

Dmitriy Glushko and Felix J. Yanovsky
National Aviation University
Electronics Department
Prospekt Komarova 1, 03-680 Kiev, UKRAINE
Email: yanovsky@nau.edu.ua

***Abstract*—This paper concerns a ground-based continuous-wave noise radar system that was built from Commercial-Off-The-Shelf measurement equipment (Vector Signal Analyser, Arbitrary Waveform Generator, amplifiers, antennas) and operated at 5.6 GHz in the C-band. This electronic system was tested as a meteorological radar in the experiment devoted to atmospheric precipitation sensing. The result of the processing is a set of Range-Doppler plots containing visible echo of rain, based on which its intensity can be calculated. There are the system setup, signal processing and results described in the paper.**

Keywords*—noise radar, software radar, solid state radar, signal processing, meteorological radar.

I. Introduction

During many years meteorological objects and phenomena were studied using conventional pulse radars and Doppler radars (pulse [1] or CW FM [2]). Recently polarimetric Doppler radar [3] and Doppler-polarimetric radar is considered as a powerful instrument for remote sensing of the atmosphere [4]. The simplest but perhaps one of the most important characteristics of precipitation is the intensity of rain or snow. It is well-known that rain intensity can be estimated from measured Radar Reflectivity Factor (RRF) Z (measured at single polarization, for example at horizontal polarization Z_h) which is directly proportional to the average power of receiving signal. One of disadvantages related with this way of rain rate estimation is the necessity to perform absolute measurements of the reflected power that requires complicated calibration and radar parameters stabilization. Otherwise the results become inaccurate.

Dual-polarization radar, which is naturally a multi-parameter system, offers some new parameters that are related with rain rate. First of all such parameters are differential reflectivity Z_{DR} and specific differential phase K_{DP}. Differential reflectivity can be defined as $Z_{DR} = 10\log(Z_h/Z_v)$, where Z_h and Z_v are RRF of the same resolution volume measured at two orthogonal polarizations, say, horizontal (h) and vertical (v) correspondingly. Specific differential phase is defined as $K_{DP} = [\phi_{DP}(r_2) - \phi_{DP}(r_1)]/2(r_2 - r_1)$. Here $\phi_{DP} = \phi_h - \phi_v$ is a differential phase, and ϕ_h is the phase of the horizontally-polarized pulse at a given point in the propagation path, ϕ_v is the phase of the vertically-polarized pulse at the same point in the propagation path, r_1 and r_2 refer to measurements at range 1 and range 2 from the radar, $r_2 > r_1$, $\phi_h \geq 0$, $\phi_v \geq 0$. This definition shows that differential phase ϕ_{DP} is simply the difference in phase between the horizontally- and vertically-polarized pulses at a given range along the propagation path. Naturally, ϕ_{DP} will increase with range from the radar, so we can take the range derivative over $\Delta r = r_2 - r_1$ to determine where along the propagation path phase changes are occurring. This derivative is called the specific differential phase, or K_{DP}. In the denominator a multiplier 2 appears at $\Delta r = r_2 - r_1$ because there is a phase shift on both, the outbound trip and the return trip.

Detailed analysis and physical explanation of the relationships between above-mentioned measurable variables and rain rate R can be found, for example, in [3]. Physically based radar rain rate estimation algorithms can be broadly classified into four categories, namely, $R(Z)$, $R(Z, Z_{DR})$, $R(K_{DP})$ and $R(K_{DP}, Z_{DR})$. Each of these algorithms has advantages and disadvantages. Important to mention that polarimetric parameters Z_{DR} and K_{DP} are relative parameters, so they do not suffer from problems with calibration that are characteristic for absolute value measurements. Of course, suitable engineering implementation of rain rate algorithm can be done using statistical approach or on the basis of fuzzy logics or/and neural network. However this is beyond the scope of this paper.

Thus, narrowband Doppler radars both pulse and CW FM, including radars with polarization diversity, became operational tools for remote sensing of meteorological objects. Nevertheless, they have some disadvantages. The range cell size depends on the length of the pulse or bandwidth, so it is difficult to achieve very good resolution in range with adequate power and velocity resolution. Another problem is related to ambiguous returns that cause a severe limitation when studying widely distributed targets as weather objects. One of the first mentions of noise radar for meteorological application was made in [5]. Meteorological application of UWB including noise waveform was discussed in [6]. In [5] it was shown that the frequency band of transmitting signal and the filter band in meteorological noise radar can be set and controlled independently; thus, there is possible to control simultaneously both resolution on distance and velocity. Moreover, the transmitting signal is not a periodical one, so there are no ambiguities in distance measuring, that is, it is possible to set parameters to provide unambiguous velocity range which is interesting for given application of the noise radar. That means that noise radar can measure very wide Doppler spectra. The last advantage is especially important for

978-1-4673-4672-6/13/$31.00 ©2013 IEEE

the case of Doppler-polarimentric measurements because they need measurable information on the spectrum itself but not only the spectrum width [7].

Nowadays, a noise radar can be implemented as a software-defined system. Due to growing computational powers, even radar with large bandwith and resulting abundant stream of digital data can be implemented to work in real-time. The restrictions on the frequency band or the central frequency of the transmitter depend mainly on the analogue front-end and the ADC capabilities. Experimental implementation of noise sounding waveform for sounding the atmosphere was checked using acoustic radar (sodar) as was described in [8]. The noise radar seems to be an interesting option to implement a weather radar. Taking into account the experience on noise radar development accumulated in institutions like Warsaw University of Technology [9] or Institute for Radiophysics and Electronics, NAS of Ukraine [10] we can say that the technology is mature enough to be used in practice. In addition, an important advantage of the noise radar is that it is highly immune to noise and unintentional jamming which can be a significant problem as the occupation of radio frequencies increasingly grows. The purpose of this paper is to describe an experiment on atmospheric precipitation sensing with a C-band ground-based CW noise radar.

II. System setup

The system constructed in order to carry out an experiment was a software defined noise radar. It was built from COTS (Commercial-Off-The-Shelf) elements. It used an Arbitrary Waveform Generator (AWG - NI-PXIe 5673) as the source of the sounding waveform and a an Vector Signal Analyser (2 channel VSA - NI-PXIe 5663) as the receiver. The transmitted waveform was a realization of pseudo-noise white, complex Gaussian noise generated in Matlab environment and uploaded to the memory of the AWG. It was synthesized in the baseband at a sampling clock frequency of 10 MHz and modulated to the desired carrier of 5.6 GHz. The waveform contained 10 MS repeated in a continuous loop. Due to the fact that the time of a single integration was set to 1 second and covered 10 MS, there were no ambiguities in the resultant crossambiguity functions. In a ready-made noise radar the waveform should not rather be periodic, however for this kind of experiment it did not bring any inconveniences but simplified the measurement itself. The generator offered maximum output power of 5 dBm so an additional power amplifier on the transmitting site was needed. As a result the power supplied to a commercially available 5.6 GHz grid antenna was about 20 dBm. A noise radar needs a reference signal for correlational range compression to be performed. In the given case such signal was captured by the second, time-synchronized and phase-coherent channel of the VSA. This provided also the synchronization between transmission and reception that needs particular hardware support in conventional pulse radars. To improve the SNR, a bandpass filter and low-noise amplifiers were put after the receiving antenna. In order to evaluate the power balance in the system, a measurement in calibration connection with attenuator instead of antennas was also carried out. The diagram of the system setup is presented in Fig. 1.

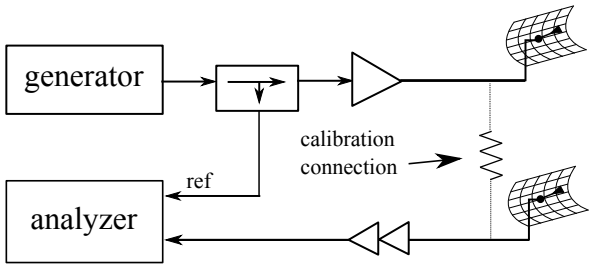


Fig. 1. A diagram of the system setup.

III. Signal processing

All the signal processing of the received signals was done offline in the MATLAB environment. It could be divided in two main steps. The first was cancellation of the direct-path interference resulting from CW operation mode. In case of the described experiment, the echo signal was at least 70 dB weaker than the direct one. For the bandwidth-time product of 10^7 samples, the echoes of such power are covered under residual fluctuations resulting from the strongest signal. In order to suppress the direct signal together with the disturbances it introduces to the crossambiguty function, a block lattice filter was used [11]. It reduced the residual fluctuations level by approximately 30 dB, which revealed the useful echoes. Then range-Doppler correlation between the reference and surveillance signal was done to obtain the crossambiguity function. The surveillance signal was correlated with Doppler-shifted copies of the reference signal with removed interferences. The results were then normalized to the system reference power level. It was obtained from range-Doppler plot (with no lattice filtering) calculated for data collected in calibration setup (with a known attenuator connecting Rx and Tx antenna cables).

With the knowledge of antenna gain and other radar specifications the calculation of precipitation intensity can be done. Different algorithms can be used for this purpose. As an example, in this paper we use the simplest algorithm which was mentioned in the introductory part of the paper. Radar equation in case of meteorological target observation can be written as [12]:

$$P_r = Z\frac{\pi^4 P_t G_t G_r c|K|^2 k_{fill}}{128\lambda^2 B L_0 r^2}, \qquad (1)$$

where P_r and P_t are received and transmitted powers, G_r and G_t are gains of receiving and transmitting antennas, λ is the wavelenght of the carrier, B is the spectrum width of the sounding noise waveform, c is the lightspped, L_0 is a coefficient of losses, $|K|^2$ is a coefficient depending on the material of scatterers (e.g. for water $|K|^2 = 0.93$, for ice $|K|^2 = 0.19$), k_{fill} is a fill factor that shows which part of the resolution volume is filled with scatterers ($0 \leq k_{fill} \leq 1$), r is a distance between the radar and the resolution volume under observation (radar range), Z is radar reflectivity factor which shows the reflection power of the object and depends on the microstructure of the resolution volume under observation.

Let us assume $k_{fill} \approx 1$ (this is practically always true in case of high radar resolution, that is small resolution volume), combine all radar specifications, except the transmitting power,

Fig. 2. Antennas during the measurement.

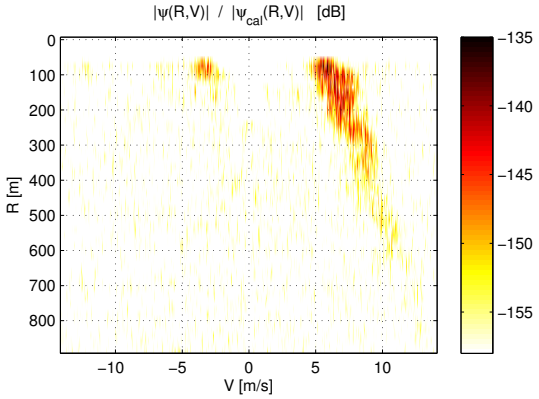


Fig. 3. Echo of the rain visible at positive Doppler velocities.

as a dimensional factor Ξ, and solve equation (1) regarding Z. Then we receive:

$$Z = \frac{\Xi r^2}{|K|^2} \cdot \frac{P_r}{P_t}, \tag{2}$$

where $\Xi = (128\lambda^2 B L_0)/(\pi^4 G_t G_r c)$.

Thus, radar reflectivity factor Z can be estimated from measured P_r/P_t ratio with assumption of $|K|^2$ value (say water, ice, snow) and known parameters of the radar (Ξ).

In turn, radar reflectivity factor is related with different important meteorological parameters (rain rate, water content, etc). In case of rain, the intensity of rainfall (or rain rate) R can be estimated from:

$$R(Z) \approx \alpha Z^{\beta} \ [\mathrm{mm} \cdot \mathrm{h}^{-1}], \tag{3}$$

where Z is in units of $[\mathrm{mm}^6 \cdot \mathrm{m}^{-3}]$.

Parameters α and β in the empirical expression (3) depend on different factors. The range of these parameters reported in the literature is very large. In moderate latitudes one can use $\alpha = 0.041$ and $\beta = 0.625$ (taken from $Z \approx 200R^{1.6}$ [13]). For US weather radar WSR-88D according to [3] $R_{\mathrm{WSR}}(Z) \approx 0.017Z^{0.714}$.

Of course, application of polarimetric measurements of Z_{DR} and K_{DP} can make results more accurate, however the purpose of this work is just to show possibility of meteorological measurements with a CW noise radar.

IV. Measurement campaign and its results

The measurement campaign was conducted in October 2012 at the balcony of the DSP Laboratory of Warsaw University of Technology at the 5^{th} floor over the ground. During rain observation two grid antennas with directivity of 25 dBi were mounted on a rotator. The rain intensity was maximum 25 dBz (according to weather radar data available online). Several measurements including azimuth and elevation scans were done. Fig. 2 shows a photo of the measurement setup while in the Fig. 3 an exemplary result for rain observation is presented. During this measurement antennas were directed approximately at 45° in elevation. It can be seen that at further distances and thereby bigger heights, the raidrops moved faster in the radial direction.

V. Conclusions

The paper has shown the potential of use of a low power, COTS-based noise radar demonstrator as a source of weather data. In the future the authors plan to perform multichannel, polarimetric trials with a noise radar to obtain data that can make full use form accurate methods of recalculation of collected results into meaningful meteorological quantities.

References

[1] R. Doviak and D. Zrnic, *Doppler Radar and Weather Observations*. San Diego: Academic Press, 1993.

[2] L. Ligthart and L. Nieuwkerk, "An X-band solid-state FM-CW weather radar," *Radar and Signal Processing, IEE Proceedings F*, vol. 137, no. 6, p. 41, dec 1990.

[3] V. N. Bringi and V. Chandrasekar, *Polarimetric Doppler Weather Radar: Principles and Applications*. Cambridge, UK: Cambridge University Press, 2001.

[4] F. Yanovsky, H. Russchenberg, and C. Unal, "Retrieval of information about turbulence in rain by using doppler-polarimetric radar," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 53, no. 2, pp. 444–450, Feb. 2005.

[5] F. J. Yanovsky, "Potential of a noise radar for meteorological applications," in *First International Workshop on the Noise Radar Technology (NRTW 2002)*, Sep. 2002, pp. 195–199.

[6] F. Yanovsky, V. Kharchenko, and L. Ligthart, "Wideband radar remote sensing of the atmosphere: a data interpretation problem," in *Ultra-wideband and Ultrashort Impulse Signals, 2004 Second International Workshop*, Sept. 2004, pp. 42 – 47.

[7] F. Yanovsky, "Phenomenological models of doppler-polarimetric microwave remote sensing of clouds and precipitation," in *Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International*, vol. 3, june 2002.

[8] R. Sinitsyn and F. Yanovsky, "Signal processing algorithms for ultra-wideband noise radar," in *Proc. International Radar Symposium IRS 2007*, Sep. 2007, pp. 83–88.

[9] L. Maslikowski, M. Malanowski, K. Kulpa, and C. Contartese, "Preliminary results of ground-based noise SAR experiments," in *Proc. EUSAR 2010*, June 2010, pp. 596–599.

[10] K. Lukin, P. Vyplavin, V. Palamarchuk, O. Zemlyaniy, V. Kudriashov, and S. Lukin, "Capabilities of noise radar in remote sensing applications," in *Advances in Radar and Remote Sensing (TyWRRS), 2012 Tyrrhenian Workshop on*, Sept. 2012, pp. 10–17.

[11] K. Kulpa and Z. Czekala, "Ground clutter suppression in noise radar," in *Int. Conference RADAR 2004*, Oct. 2004, pp. 236–240.

[12] F. Yanovsky, *Meteorological and Navigation Radar Systems of Aircraft (in Ukrainian)*. Kiev, Ukraine: NAU Press, 2003.

[13] V. Stepanenko, *Radar in Meteorology (Radiometeorology) (in Russian)*. Leningrad, form. USSR: Gidrometeoizda, 1966.