Elena Velichko · Viktoria Kapralova ·
Platon Karaseov · Sergey Zavjalov ·
Pablo Angueira · Sergey Andreev *Editors*

International Youth Conference on Electronics, Telecommunications and Information Technologies

Proceedings of the YETI 2021, St. Petersburg, Russia

Springer

Springer Proceedings in Physics

Volume 268

Indexed by Scopus

The series Springer Proceedings in Physics, founded in 1984, is devoted to timely reports of state-of-the-art developments in physics and related sciences. Typically based on material presented at conferences, workshops and similar scientific meetings, volumes published in this series will constitute a comprehensive up-to-date source of reference on a field or subfield of relevance in contemporary physics. Proposals must include the following:

– name, place and date of the scientific meeting
– a link to the committees (local organization, international advisors etc.)
– scientific description of the meeting
– list of invited/plenary speakers
– an estimate of the planned proceedings book parameters (number of pages/articles, requested number of bulk copies, submission deadline).

Please contact:

For Americas and Europe: Dr. Zachary Evenson; zachary.evenson@springer.com
For Asia, Australia and New Zealand: Dr. Loyola DSilva; loyola.dsilva@springer. com

More information about this series at https://link.springer.com/bookseries/361

Elena Velichko · Viktoria Kapralova ·
Platon Karaseov · Sergey Zavjalov ·
Pablo Angueira · Sergey Andreev
Editors

International Youth Conference on Electronics, Telecommunications and Information Technologies

Proceedings of the YETI 2021, St. Petersburg, Russia

Springer

Editors
Elena Velichko
Peter the Great St. Petersburg Polytechnic
University
St. Petersburg, Russia

Viktoria Kapralova
Peter the Great St. Petersburg Polytechnic
University
St. Petersburg, Russia

Platon Karaseov
Peter the Great St. Petersburg Polytechnic
University
St. Petersburg, Russia

Sergey Zavjalov
Peter the Great St. Petersburg Polytechnic
University
St. Petersburg, Russia

Pablo Angueira
Department of Communication Engineering
Univ Basque Country UPV EHU
Bilbao, Spain

Sergey Andreev
Unit of Electrical Engineering
Tampere University
Tampere, Finland

Organization

Program Committee

Sergey Makarov, Peter the Great St. Petersburg Polytechnic University
Luis Conde, Polytechnic University of Madrid
Dong Ge, Tsinghua University
Viktoria Kapralova, Peter the Great St. Petersburg Polytechnic University
Yevgeni Koucheryavy, Tampere University
Sergey Andreev, Tampere University
Hayk Sarkisyan, Russian-Armenian University
Pablo Angueira, University of the Basque Country
Wei Xue, Harbin Engineering University
Elena Velichko, Peter the Great St. Petersburg Polytechnic University
Platon Karaseov, Peter the Great St. Petersburg Polytechnic University
Sergey Zavjalov, Peter the Great St. Petersburg Polytechnic University
Maxim Vinnichenko, Peter the Great St. Petersburg Polytechnic University
Elina Nepomnyashchaya, Peter the Great St. Petersburg Polytechnic University

Preface

The International Youth Conference on Electronics, Telecommunications and Information Technologies (YETI 2021) was organized by Peter the Great St. Petersburg Polytechnic University (Saint Petersburg, Russia) and took place during April 22–23, 2021. The YETI 2021 conference provided an opportunity to young researchers and early career scientists to share recent achievements, current trends and knowledge as well as future prospects in the fields of development and design of nanoelectronic and telecommunication devices, signal processing, material science and nanotechnology, photonics, optical and information technologies. This conference attracted 82 papers from eight countries, which generated lively discussions and fruitful debates. As usual at Polytechnic University conferences, a contest for the best paper was held to encourage the commitment of young scientists in conference activities.

The official conference Website is http://yeti.spbstu.ru/.

St. Petersburg, Russia	Elena Velichko
St. Petersburg, Russia	Viktoria Kapralova
St. Petersburg, Russia	Platon Karaseov
St. Petersburg, Russia	Sergey Zavjalov
Bilbao, Spain	Pablo Angueira
Tampere, Finland	Sergey Andreev

Contents

Contents   xiii

Contributors

Elizaveta Agapova Institute of Experimental Medicine, St. Petersburg, Russia

Arseniy Alekseev Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Sergey Alexandrov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Dmitriy P. Andreev Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

Elena I. Andreeva The Bonch-Bruevich Saint-Petersburg State University of Telecommunications, St. Petersburg, Russian Federation

Alexey V. Andropov Bonch-Bruevich Saint-Petersburg State University of Telecommunications, St. Petersburg, Russia

Aleksei Anisimov Institute of Experimental Medicine, St. Petersburg, Russia; Saint Petersburg Electrotechnical University, St. Petersburg, Russia

Kseniya Arinushkina Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Ara Abdulsatar Assim Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia;
Salahaddin University, Erbil, Iraq

Hamlet Ayvazyan Engineering and Physics Institute, Russian-Armenian University, Yerevan, Armenia

Alexander Azarov Centre for Materials Science and Nanotechnology, University of Oslo, Oslo, Norway

Evgenii Balashov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Evgenii V. Balashov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Maksim Baranov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Andrey Belmesov Institute of Problems of Chemical Physics (IPCP RAS), Chernogolovka, Russia

Polina Bespalova Alferov St. Petersburg National Research Academic University, St. Petersburg, Russia;
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Yuri Y. Bleyan Russian–Armenian University, Yerevan, Armenia

Yulia O. Bobrova Saint Petersburg Electrotechnical University "LETI", St. Petersburg, Russia

Daniil Borisov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Alexey Borodkin ITMO University, St. Petersburg, Russia

Linda Boudjemila Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia;
Department of Mechanical Engineering, University of Science and Technology, Algiers, Algeria

Anatolii A. Briushinin Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Dmitry O. Budanov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Wei Cao Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Aleksandra Chekireva Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Luis Conde Department of Applied Physics, ETS Ingeniería Aeronáutica Y del Espacio, Universidad Politécnica de Madrid, Madrid, Spain

Maxim N. Davydov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

Vadim V. Davydov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia;
All-Russian Research Institute of Phytopathology, Moscow Region, Russia

Mark Denisenko Southern Federal University, Taganrog, Rostov-on-Don Region, Russia

Dmitry Dicky Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Ge Dong School of Aerospace Engineering, Tsinghua University, Beijing, China

Dmitry Dyubo Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Egor V. Egorov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Yakov Enns Alferov St. Petersburg National Research Academic University, St. Petersburg, Russia;
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Artur A. Ermolaev The Bonch-Bruevich St. Petersburg State University of Telecommunications, St. Petersburg, Russia

Olga Ezhova Southern Federal University, Taganrog, Rostov-on-Don Region, Russia

Olga A. Ezhova LLC «MEMS Design», Taganrog, Russia

Kseniya V. Filipenko Saint Petersburg Electrotechnical University "LETI", St. Petersburg, Russia

Stanislav A. Fyodorov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Marina Galeeva Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Dong Ge Tsinghua University, Haidian District, Beijing, China

Birgit Glasmacher Institute for Multiphase Processes, Leibniz University Hannover, Garbsen, Germany

Alexander Glukhov Institute of Problems of Chemical Physics (IPCP RAS), Chernogolovka, Russia

Daniela S. Gogova Central Lab of Solar Energy at the Bulgarian Academy of Sciences, Sofia, Bulgaria

Jorge Gonzalez DIFFER-Dutch Institute for Fundamental Energy Research, Eindhoven, The Netherlands

Igor N. Gorbunov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Anna Grevsteva Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Irena M. Gureeva Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Gesine Hentschel Department for Biomedical Engineering, Linköping University, Linköping, Sweden;
Institute for Multiphase Processes, Leibniz University Hannover, Garbsen, Germany

Babken Hovhannisyan Radiophysics and Electronics, Yerevan State University, Yerevan, Armenia

Alexander Y. Ivanov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Sergei I. Ivanov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Johannes Johansson Department for Biomedical Engineering, Linköping University, Linköping, Sweden

Olga N. Kapranova Saint Petersburg Electrotechnical University "LETI", St. Petersburg, Russia

Konstantin V. Karabeshkin Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia;
CJSC "ELAR", St.-Petersburg, Russia

Platon Karaseov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia;
Alferov University, St. Petersburg, Russia

Platon A. Karaseov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Khachik S. Khachatryan Russian–Armenian University, Yerevan, Armenia

Mahdi Khadem Institute of Microelectronic Technology and High Purity Materials, Chernogolovka, Russia

Igor Khodos Institute of Microelectronic Technology and High Purity Materials, Chernogolovka, Russia

Nikolai F. Kidyaev Southern Federal University, Taganrog, Russia

Dae-Eun Kim Department of Mechanical Engineering, Yonsei University, Seoul, South Korea

Nikita Kirpanev Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Lyubov Kiseleva Russian Research Center for Radiology and Surgical Technologies, St. Petersburg, Russia

Roman Kleimanov Alferov University, St. Petersburg, Russia

Evgeni Kolodeznyi ITMO University, St. Petersburg, Russia

Ilya Kolokolnikov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Ivan Komarevtsev Alferov St. Petersburg National Research Academic University, St. Petersburg, Russia;
Ioffe Physical-Technical Institute of the Russian Academy of Sciences, St. Petersburg, Russia

Dmitrii V. Kondakov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Anastasia Kondrateva Alferov St. Petersburg National Research Academic University, St. Petersburg, Russia;
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Irina A. Kondratyeva Saint Petersburg Electrotechnical University 'LETI', Saint Petersburg, Russian Federation

Boris G. Konoplev Southern Federal University, Taganrog, Russia

Konstantin O. Korovin Bonch-Bruevich Saint-Petersburg State University of Telecommunications, St. Petersburg, Russia

Andrey Korshunov Alferov University, St. Petersburg, Russia

Lida Kouhalvandi Electronics and Communication Engineering Department, Istanbul Technical University, Istanbul, Turkey;
Department of Electronics and Telecommunications, Politecnico di Torino, Torino, Italy

Yakov Kovach ITMO University, St. Petersburg, Russia

Andrey V. Kovalev Southern Federal University, Taganrog, Russia

Kirill Krainov PAO Severstal, Cherepovets, Vologda Region, Russia

Alexander S. Krasichkov Saint Petersburg Electrotechnical University 'LETI', Saint Petersburg, Russian Federation

Yury E. Krivenko The Bonch-Bruevich Saint-Petersburg State University of Telecommunications, St. Petersburg, Russian Federation

Vladislav A. Kruglov Institute of Analytical Instrumentation of the Russian Academy of Sciences, St. Petersburg, Russia

Aleksei Krylov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Daria Kulik Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Maria Kuropatenko Institute of Experimental Medicine, St. Petersburg, Russia

Sergey V. Kuzmin Bonch-Bruevich Saint-Petersburg State University of Telecommunications, St. Petersburg, Russia

Maxim Kuznetsov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Ilya Lavrenyuk Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Alexander P. Lavrov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Leonid Liokumovich Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Semen E. Logunov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

Maksim A. Lopatin Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Roman Lozov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Nikita Lukashev Bonch-Bruevich St. Petersburg State University of Telecommunications, St. Petersburg, Russia

Igor Lysenko Southern Federal University, Taganrog, Rostov-on-Don Region, Russia

Igor E. Lysenko Southern Federal University, Taganrog, Russia

Nikita Makarov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Sergey Makarov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Sergey B. Makarov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Armen Makaryan Radiophysics and Electronics, Yerevan State University, Yerevan, Armenia

Victor M. Malyshev Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Vladislav G. Malyshkin Ioffe Institute, St. Petersburg, Russia

Aleksandr Markvart Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Ekaterina Mashkova Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Ladislau Matekovits Department of Electronics and Telecommunications, Politecnico di Torino, Torino, Italy;
Department of Measurements and Optical Electronics, Politehnica University Timisoara, Timisoara, Romania;
Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni, National Research Council, Turin, Italy

Alexey Mayzel Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Andrei Medvedev Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Eugene Mirkin Computer Information Systems, International University of Kyrgyzstan, Bishkek, Kyrgyz Republic

Maxim Mishin Alferov St. Petersburg National Research Academic University, St. Petersburg, Russia

Mher A. Mkrtchyan Russian–Armenian University, Yerevan, Armenia

Azam Moradi Federal State Autonomous Educational Institution of Higher Education, National Research University ITMO, Saint-Petersburg, Russia

Mojtaba Mousavi Khatam Al-Nabieen University, Kabul, Afghanistan

Mojtaba Najafi Oshnari Saint Petersburg State University, St. Petersburg, Russia

Samaneh Navvabi Department of Mathematics and Mechanics, Saint-Petersburg State University, Saint-Petersburg, Russia

Peter A. Nekrasov Smorodintsev Research Institute of Influenza, Ministry of Healthcare of the Russian Federation, St. Petersburg, Russia

Elina Nepomnyashchaya Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Canh Minh Nguyen University of Transport and Communications, Ha Noi, Viet Nam

Dac Cu Nguyen Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Tan Ngoc Nguyen Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Anastasiya Nitkina Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Boris Novikov Saint Petersburg State University, St. Petersburg, Russia

Tatiana Novikova Institute of Experimental Medicine, St. Petersburg, Russia

Sergey P. Nuzhny Special Technological Center Ltd, Saint-Petersburg, Russia

Mahdi Orooji Department of Electrical Engineering, Tarbiat Modares university, Tehran, Iran;
Department of Electrical Engineering, University of California, Davis, USA

Mojtaba Najafi Oshnari Department of Mathematics and Mechanics, Saint-Petersburg State University, Saint-Petersburg, Russia

Valeriia Osochenko Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Victor Osovskikh Russian Research Center for Radiology and Surgical Technologies, St. Petersburg, Russia

Anna Ovsyannikova Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Anna S. Ovsyannikova Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Liudmila Pankova Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Vitalii A. Pavlov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Valerii Perhun Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Artem Petrenko ITMO University, St. Petersburg, Russia

Alexander Petrov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Ivan Pleshakov Ioffe Institute, St. Petersburg, Russia

Anastasiya A. Polyakova Saint Petersburg Pediatric Medical University, Saint Petersburg, Russian Federation

Vladimir Pukha Institute of Problems of Chemical Physics (IPCP RAS), Chernogolovka, Russia

Andrey Rashich Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Vladislav S. Reznik Institute of Analytical Instrumentation of the Russian Academy of Sciences, St. Petersburg, Russia

Stanislav Rochas ITMO University, St. Petersburg, Russia

Sergey Rozov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Vasiliy Yu. Rud' All-Russian Research Institute of Phytopathology, Moscow Region, Russia;
Ioffe Institute, St. Petersburg, Russia

Tigran Sargsian Russian-Armenian University, Yerevan, Armenia

Ekaterina A. Savchenko Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Elena Savchenko Computer Information Systems, International University of Kyrgyzstan, Bishkek, Kyrgyz Republic

Danila D. Savin Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Anastasiia I. Semenova Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Timofey Sergeev Institute of Experimental Medicine, St. Petersburg, Russia;
Saint Petersburg Electrotechnical University, St. Petersburg, Russia

Dmitry Y. Sevostyanov Southern Federal University, Taganrog, Russia

Alexander Shakhmin Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Faridoddin Shariaty Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia;
Department of Electrical Engineering, University of Qom, Qom, Iran

Maria A. Shevchenko The Bonch-Bruevich St. Petersburg State University of Telecommunications, St. Petersburg, Russia

Daniil Shiryaev ITMO University, St. Petersburg, Russia

Irina I. Shpakovskaya Saint Petersburg Electrotechnical University 'LETI', Saint Petersburg, Russian Federation

Van Sinian Xi'an Shiyou University, Xi'an, China

Alexander S. Sinyukin Southern Federal University, Taganrog, Russia

Eduard Sivolenko Radiophysics and Electronics, Yerevan State University, Yerevan, Armenia

Alexey N. Skvortsov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Gleb I. Stesev Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Andrei I. Struchkov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Nikolay Suvorov Institute of Experimental Medicine, St. Petersburg, Russia

Konstantin S. Sychev Institute of Analytical Instrumentation of the Russian Academy of Sciences, St. Petersburg, Russia

Nguyen Ngoc Tan Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Valentina Temkina Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Andrei I. Titov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Olga I. Tkach Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Alexey Tkachenko Southern Federal University, Taganrog, Rostov-on-Don Region, Russia

Alena Tomchuk Saint Petersburg Electrotechnical University, St. Petersburg, Russia

Dina V. Trotsyuk Federal State Autonomous Educational Institution of Higher Education, Belgorod State University, Belgorod, Russia

Kseniya V. Trusova JSC RDC St. Petersburg Branch, St. Petersburg, Russia

Oleg Tsybin Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Kirill Tyshuk Saint Petersburg State University, St. Petersburg, Russia

Nikolai Ushakov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Anton Valov Bonch-Bruevich St. Petersburg State University of Telecommunications, St. Petersburg, Russia

Elena Velichko Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Elena N. Velichko Peter the Great St Petersburg Polytechnic University, Saint Petersburg, Russia

Sergey Volvenko Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Alexandr Vorobyev Alferov St. Petersburg National Research Academic University, St. Petersburg, Russia

Kira Vyatkina Saint Petersburg State University, St. Petersburg, Russia;
Alferov University, St. Petersburg, Russia;
Saint Petersburg Electrotechnical University "LETI", St. Petersburg, Russia

Christina Winkler Institute for Multiphase Processes, Leibniz University Hannover, Garbsen, Germany

Karin Wårdell Department for Biomedical Engineering, Linköping University, Linköping, Sweden

Wei Xue College of Information and Communication Engineering, Harbin Engineering University, Harbin, China

Alexander Yafarov Institute of Experimental Medicine, St. Petersburg, Russia

Zulfia A. Zaripova The Pavlov First Saint Petersburg State Medical University, Saint Petersburg, Russian Federation

Sergey Zavjalov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Sergey V. Zavjalov Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Aleksandr V. Zhila Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Dmitry Zubko Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Part I
Information Technologies and Signal Processing

Chapter 1
Synthesis and Analysis of Avionics Functions Digital Twins Using Machine Learning Classification Algorithms

Kseniya V. Trusova

Abstract Different industries have started to apply digital twins as digital copies of physical objects or processes to provide the best solutions. The Aerospace industry also has many best practices in this field especially for aircraft systems, Integrated Vehicle Health Management processes, and others. The intelligent application of digital twins is able to optimize different processes of the industry including failure safety and make them more evident and qualitative. This paper describes a data-driven approach to digital twins synthesis of avionics functions and their analysis by machine learning classification algorithms for such problems as function loss, behavioral indication, and detection of function loss causes at the design and development stages of an avionics lifecycle. "Avionics function" term is considered in this work as the main part of the man–machine interface between flight crew and aircraft. Avionics functions definition is the basis of the design stage of the avionics lifecycle. The results of a decision tree, random forest, support vector machine, kernel support vector machine, naive Bayes, and logistic regression classification are considered. The classification results show which of the algorithms and dataset types match the objective. The presented approach will be able to improve the understanding of avionics design and development complex processes and become the basis for further streamlining of data-driven digital twins design approach for avionics.

Keywords Avionics · Functions of avionics · Digital twin · Data-driven · Machine learning classification · Function loss

K. V. Trusova (✉)
JSC RDC St. Petersburg Branch, St. Petersburg 195009, Russia

1.1 Introduction

1.1.1 General Approach

The concept of digital twins is widely applied in different spheres as a part of Industry 4.0 and a digitalization process. With the growth of design, development, and manufacturing opportunities in view of advanced technologies, the digital twins concept is able to optimize processes of a product lifecycle including a qualification assessment process.

The aerospace industry has also started using digital twins as digital copies of physical products at different stages of a lifecycle including the design stage [1] as well as a part of Integrated Vehicle Health Management (IVHM) to enable Condition-Based Maintenance (CBM) [2].

Data-driven, operation principle and combined approach are possible methods for the creation of a digital twin. A data-driven digital twin can be developed by machine learning including consideration of such problems as health monitoring and fault detection problems [3].

Machine learning finds wide application in the aerospace industry as well as digital twins [4].

This paper focuses on digital twins synthesis of avionics functions based on the data-driven approach at the design and development stages of an avionics lifecycle. Machine learning classification algorithms were considered as a data analysis approach.

Digital twins help to research the different variants of a functions architecture at the design and development stages of a life cycle before the creation of a mock-up and physical asset saving the resources. The results can be used for hardware and software requirements specification definition and correction. This flexible approach to the development process allows designers and developers to arrange the more integrated picture of possible object complex behavior and get more data.

Another highly important side of this approach is failure safety. The process of failure safety follows through every stage of a life cycle. Failure safety requirements influence the complexity of design and development processes. Development Assurance Level (DAL) specifies the design and development program. Different forms of failure safety assessments are implemented at the stages of a life cycle such as Functional Hazard Assessment (FHA), System Safety Assessment (SSA), Fault Tree Analysis (FTA), Failure Mode and Effects Analysis (FMEA), Common Cause Analysis (CCA), etc. [5]. Clear functions architecture and sufficient data of functions behavior allow analytics to consider more possible outcomes and get more qualitative assessments.

1.1.2 The Difference in Operation Between Flight Crew and Service Technicians

The operation of avionics systems is divided into operation by flight crew and operation by service technicians.

The service technicians handle avionics equipment as a Line Replaceable Unit (LRU). The service technicians use the failure and faults information of the LRU to implement onboard maintenance (troubleshooting). If it is not enough to recover the LRU, the service technicians remove LRU for deeper diagnostics and repair.

The flight crew handles aircraft functions such as a barometric altitude indication, radio navigation aids adjustment, the input of landing heading, air speed indication, radio communication adjustment, etc. An avionics system is a set of functions combined by similar features. Functions are implemented by distributed hardware and software. Therefore, several LRU connected by a network can implement one function. It should be pointed out that the health of the entire system has little significance for the flight crew. The analysis of every function is more relevant in terms of operation because it gives information about which particular function the flight crew can or cannot use on every flight stage.

The features of internal hardware and software health were considered in the problem of LRU health detection [6–8]. The external network connections state was not taken into account.

The early detection of a condition, when a loss of a function happens, is highly significant for the flight crew in view of considering functions. Features related to hardware and software health, network connections state, processed and transmitted information state, which are used in the function implementation, are considered to be indicators of function loss.

It should be noticed that not every hardware and software component health of LRU testifies that the function is performed correctly. Some parts of hardware and software components of LRU can be unused for some functions execution. It describes a difference between function loss detection and LRU health detection apart from network connections and information state.

This paper describes an approach to the synthesis of digital twins for avionics functions and analysis of its capabilities by machine learning classification algorithms. The objective of the research has been defined as to learn how to build the digital twin of an avionics function at the design and development stages of the avionics lifecycle, and to explore its relevance for solving such problems as function loss, behavioral indication, and detection of function loss causes.

The paper is organized as follows. Section 1.2 describes the barometric correction enter function architecture, its digital twin, and problems, which were researched by machine learning classification algorithms. Section 1.3 presents the results of problems research for the barometric correction enter and "pop-up" functions, and a discussion of the results.

1.2 Models and Methods

1.2.1 The Description of the Barometric Correction Enter Function

The function of barometric correction enter is performed by three connected LRUs and includes several operations (Fig. 1.1).

The flight crew enters the value of barometric correction by baro correction control unit (BCU) from Instrument and Control Panels of Indicating/Recording Systems. The BCU processes input value using the results of its own built-in-test procedure (BCU health), and defines data validity by the results. On the next step, the BCU transmits the input barometric correction value, BCU health, and data validity through ARINC 429 aviation interface to the air data system (ADS).

The ADS from Flight Environment Data of Navigation verifies the BCU input connection state and processes the received data. The ADS also uses the results of its own built-in-test procedure (ADS health), and defines data validity by the results. It is a standard procedure for avionics devices. If the received value is valid, the ADS uses it for computation of its own parameters and transmits the received value, the BCU health and data validity, the ADS health and data validity through ARINC 429 to the multifunctional display (MFD).

The MFD from Central Display Systems of Indicating/Recording Systems verifies the ADS input connection state and processes the received data. This block is highlighted in red in Fig. 1.1 because it selects the final type of indication depending on the BCU and ADS health data (LRU health, data validity, and transmitter input connection state), and MFD health. The flight crew sees this final indication as the result of data processing performed by connected LRUs.
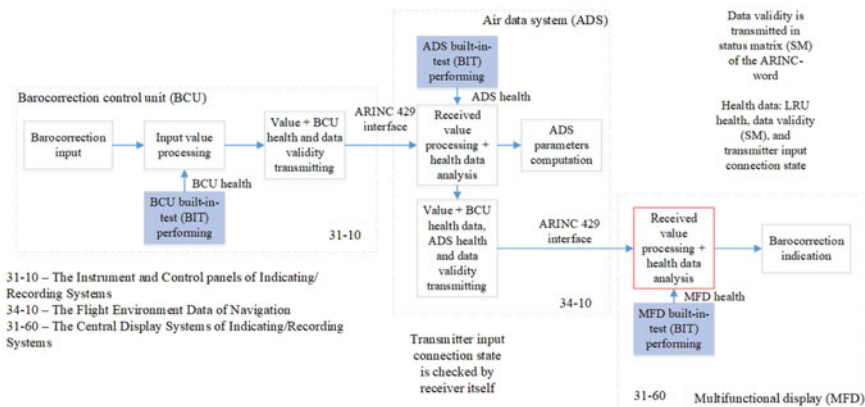
Fig. 1.1 The architecture of the barometric correction enter function

The BIT blocks of the LRUs in Fig. 1.1 were the subject of the works [7, 8], and were used as an accomplished part of the research.

The described architecture was considered as a framework for the digital twin of the barometric correction enter function. The function has a relatively simple structure that makes digital twin synthesis and analysis more obvious for the research.

A framework based on The Open Standard Architecture—Condition Based Maintenance (OSA-CBM) was considered as a generalised methodology for developing diagnostics for different aircraft systems [9]. The diagnostics for aircraft systems followed the data-driven approach. This work researched digital twins of aircraft systems as a part of OSA-CBM and used the profiles of continuous features for solving classification problems.

Data-driven approach finds applications for problems solving in different technical areas including the classification of objects using channel state information [10].

1.2.2 The Digital Twin of the Barometric Correction Enter Function

Data-driven digital twins can be built by different methods. Practical experience shows that two methods are usually used at the development and testing stages of an avionics system life cycle. The operating system software version of LRU (usually Windows version) or operating system version combined with LRU emulator are able to verify only some operations.

The above mentioned methods consider only the series of input value conversions excluding health information. For some simple functions, it can be enough and correct when extra code is not added. Complex functions require more detail consideration. For example, the more computation of correlated parameters received from different transmitters they perform, the more variants of final states they have. The neglecting of health information leads to errors at the next stages of a life cycle and increases the final cost of the product. Therefore, the digital twin of a complex function should include health data.

Figure 1.2 demonstrates flow diagrams for either variations of the digital twins.

The health data on the flow diagram is highlighted in green including blocks. This flow diagram shows how health data is involved in the decision-making process.

The "Data processing, state detection, indication type selection" block is highlighted in red and has the same meaning as "Received value processing+health data analysis" block shown in Fig. 1.1. Possible final health states are specified under the LRUs.

The digital twin including health data seems significantly more complicated than the digital twin, which does not include this data. It can create an impression that this complexity requires much more resources. In this case, resource expenditure should be compared with the possible benefit of a digital twin implementation at the considered stage of a life cycle and failure safety requirements.
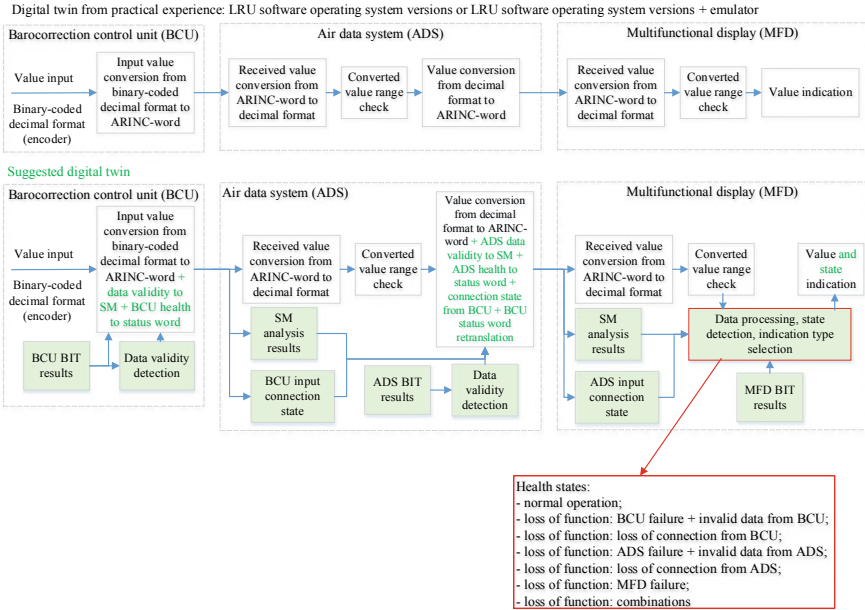
Fig. 1.2 The digital twins flow diagram of the barometric correction enter function excluding and including health data

The considered digital twins flow diagram describes the function explicitly enough to research such relevant problems of avionics development as function loss, behavioral indication, and detection of function loss causes.

1.2.3 Function Loss, Behavioral Indication, and Detection of Function Loss Causes

The function loss. The function loss can be recognized by features extracted from the digital twins flow diagram of the function including health data. The combinations of discrete features describe the classes of the function state (Table 1.1).

The state of the function includes three classes. The features take discrete values "0" or "1". "Partial function loss" is detected by features with values "1", which are not used for "Total function loss detection". "Normal operation" is detected by all features with values "0".

The research of this problem helps to analyze the influence of every feature and emphasize the more significant features. The results also are useful for statistical analysis of particular functions loss in view of failure safety processes.

The behavioral indication. The behavioral indication means that the final indication for the flight crew depends on the detected function state (the block "Health states"

Table 1.1 The features of the barometric correction enter function loss

Features	The barometric correction enter function state
BCU uC[a] reset, BCU uC checksum, BCU secondary power supply, MFD FPGA[b] checksum, MFD secondary power supply, MFD CPU reset, MFD CPU[c] checksum	Partial function loss
BCU encoder fault: key stuck, BCU encoder fault: signal error, BCU primary power supply, MFD LCD[d] power supply, MFD LED[e] break, MFD LED short, MFD heat break, MFD memory: data read/write, No data from BCU to ADS, No data from BCU to MFD, Invalid data from BCU to ADS, Invalid data from BCU to MFD, ADS failure	Total function loss
All abovementioned features	Normal operation

[a]Micro Controller Unit
[b]Field-Programmable Gate Array
[c]Central Processing Unit
[d]Liquid Crystal Display
[e]Light-Emitting Diode

in Fig. 1.2). The problem uses the same features as function loss and different classes of indication (Table 1.2).

The flight crew sees this final indication at the MFD and initiates corresponding actions. The results of this problem investigation have a wide range of applications at the design and development stages of a life cycle and for failure safety analysis. For instance, this information is used for the process of man–machine interface design.

The detection of function loss causes. The detection of function loss causes is utilized for statistical analysis in failure safety processes. Table 1.3 presents features grouped by types of operations on information.

Table 1.2 The features of the barometric correction enter function for behavioral indication

Features	The type of indication
BCU encoder fault: key stuck, BCU encoder fault: signal error, BCU primary power supply	BCU P0 failure
MFD LCD power supply, MFD LED break, MFD LED short, MFD heat break, MFD memory: data read/write	MFD failure
No data from BCU to ADS, Invalid data from BCU to ADS, ADS failure	No P0[a] to ADS, standard P0 value is indicated with yellow colour
No data from BCU to MFD, Invalid data from BCU to MFD	No P0 to MFD, P0 value is not indicated ("--")

[a]The designation of barometric correction

Table 1.3 The features of barometric correction enter function grouped by function loss causes

Features	The cause of function loss
MFD LCD power supply, MFD LED break, MFD LED short, MFD heat break	Indication issue
MFD FPGA checksum, MFD memory: data read/write, No data from BCU to ADS, No data from BCU to MFD	Information exchange issue
BCU uC reset, BCU uC checksum, BCU secondary power supply, BCU primary power supply, MFD secondary power supply, MFD CPU reset, MFD CPU checksum, Invalid data from BCU to ADS, ADS failure, Invalid data from BCU to MFD	Input value processing issue
BCU encoder fault: key stuck, BCU encoder fault: signal error	Control or input value processing issue

The function loss, behavioral indication, and detection of function loss causes operate in the same categories as the problems considered in works [6–8]. They consist of sets aggregated by features, which are equivalent to classes. Therefore, an approach based on aggregated models [6] can be employed for function loss, behavioral indication, and detection of function loss causes.

1.2.4 Aggregated and Classification Models

The considered problems can be easily converted to aggregated models applied for avionics objects description. Taking into account that the problems description was considered in detail in 2.4, and its structure is closer to aggregated models structure, the aggregated models themselves are not represented in this paper.

The function loss detection and behavioral indication are implemented by conditional logic in physical assets. Conditional logic is easy to realize if a function is uncomplicated. Avionics functions have many connections between each other and they are not independent. Machine learning algorithms are going to become a good alternative at least for researching such problems at the different stages of a life cycle within the framework of digital twins.

Aggregated models are adapted to classification models for machine learning algorithms implementation by sets transformation [7].

1.2.5 The Researched Algorithms of Machine Learning

The research of function loss, behavioral indication, and detection of function loss causes problems by machine learning algorithms was implemented by resources of

Python programming language and Google Colaboratory notebook using built-in libraries.

The following machine learning classification algorithms were researched: decision tree, random forest, support vector machine (SVM), kernel SVM, naive Bayes, logistic regression. The mathematical background of the mentioned methods was briefly described in the work [7] and is not considered in this paper.

1.3   Research Results

1.3.1   Research Explanation

One of the research objectives consisted in the analysis of machine learning classification algorithms implementation for the function loss, behavioral indication, and detection of function loss causes problems.

The aggregated and classification models of the barometric correction enter function were considered. The models of a "pop-up" function were considered additionally. The architecture of this function and its digital twin flow diagram were not described in this paper. They are based on the same approach as the architecture and digital twin of the barometric correction enter function.

Several types of datasets were implemented for the classification algorithms research. Separated dataset with a strictly determined training part was created to investigate how determined combinations influence classification results. A determined combination is a combination where only one feature with the value "1" corresponds to a particular class. Therefore, the determined part completely consisted of combinations, which had one feature with value "1" and other features with values "0". Semi-determined data means that not every feature corresponding to a particular class was included in the dataset.

This work also attends the training part using the paradigm of "interval teacher" in the classical scheme of training neural networks [11].

The behavioral indication deals with detailed LRU health features, which are not combined by LRU into one integral feature, as well as integral features themselves. Therefore, both cases were considered in the research.

The test data in the separate dataset mostly consisted of combinations, which were absent in training data. The combined dataset also included these unknown combinations.

The accuracy of the classification and the number of incorrect recognitions (false negatives and false positives) were used as the metrics of the classification.

Table 1.4 The dataset description of the function loss problem

Number of classes	Number of features	Type of dataset	Combinations specifics	Sample size
Barometric correction enter function				
3	20	Separated training and test data	Strictly determined training data	13 (test)
		Combined	-	44 (25%)
"Pop-up" function				
2	27	Separated training and test data	Strictly determined training data	28 (test)
		Separated training and test data	Semi-determined training data	30 (test)
		Combined	Strictly determined part is included	64 (25%)
		Combined	Semi-determined part is included	67 (25%)

1.3.2   The Function Loss Research Results

Table 1.4 describes the dataset of the barometric correction enter and "pop-up" functions for the function loss problem.

Figure 1.3 shows the functions states detection results of the barometric correction enter and "pop-up" functions.

For the barometric correction enter function, all algorithms showed better accuracy in the determined training data case than in the combined dataset. The naive Bayes algorithm gave better accuracy in the combined dataset case.

For the "pop-up" function, the decision tree and random forest showed good results only in the separated dataset case. Logistic regression, SVM, and kernel SVM did not detect "Normal operation" function states in every case. These algorithms work better on the combined dataset than on the separated dataset. Naive Bayes was robust to any dataset changes.

1.3.3   The Behavioral Indication Research Results

The behavioral indication problem was researched only for the barometric correction enter function loss in view of data insufficiency for the "pop-up" function. Types of indication were presented in Table 1.3. The combinations of these types were also considered in the research. Table 1.5 describes the dataset of the barometric correction enter function for the behavioral indication problem.

Figure 1.4 shows the indication types detection results of the barometric correction enter function.
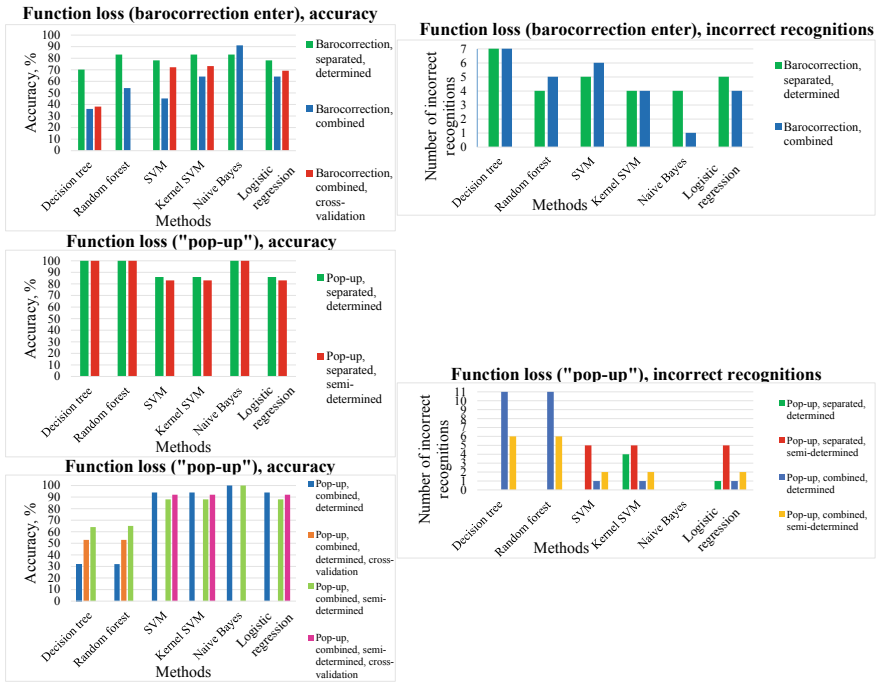
Fig. 1.3 The functions states detection results of the barometric correction enter and "pop-up" functions

Table 1.5 The dataset description of the behavioral indication problem

Number of classes	Number of features	Type of dataset	Combinations specifics	Sample size
Detailed features				
9	20	Separated training and test data	Strictly determined training data	10 (test)
		Combined	–	42 (25%)
Integral features				
11	7	Separated training and test data	Strictly determined training data	10 (test)
		Combined	–	24 (25%)

In the separated dataset case, test data included two new classes, which were absent in training data. The algorithms did not recognize the new classes. However, naive Bayes, SVM, decision tree, and random forest were closer to detecting correct classes, which were defined by the similarity of the classes. This observation of algorithms learnability is useful for further research. The methods showed relatively low
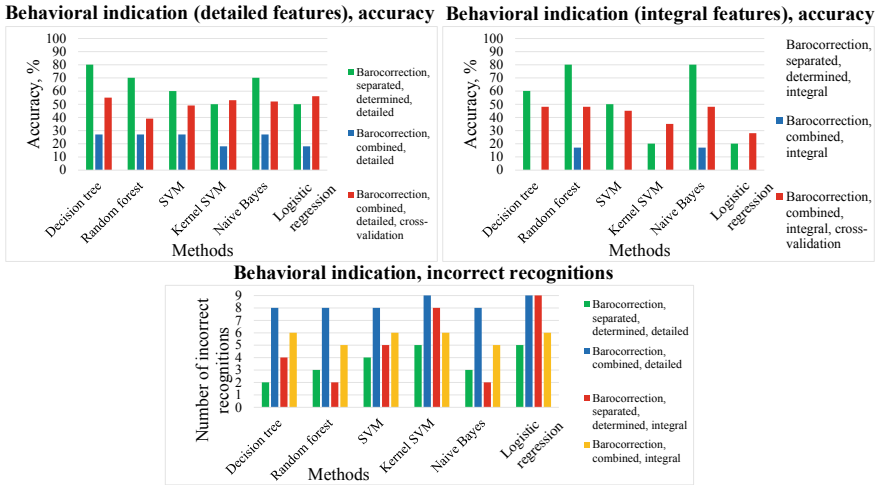
Fig. 1.4 The indication types detection results of the barometric correction enter function

accuracies and a large number of incorrect recognitions for the behavioral indication problem.

1.3.4 The Results of Function Loss Causes Detection

Table 1.6 describes the dataset of the barometric correction enter and "pop-up" functions for the detection of function loss causes.

Figure 1.5 shows the results of function loss causes detection of the barometric correction enter and "pop-up" functions.

For the barometric correction enter function, all algorithms showed better accuracy in the determined training data case than in the combined dataset as well as for

Table 1.6 The dataset description for the detection of function loss causes

Number of classes	Number of features	Type of dataset	Combinations specifics	Sample size
Barometric correction enter function				
4	20	Separated training and test data	Strictly determined training data	23 (test)
		Combined	–	43 (25%)
"Pop-up" function				
7	27	Separated training and test data	Strictly determined training data	17 (test)
		Combined	–	53 (25%)
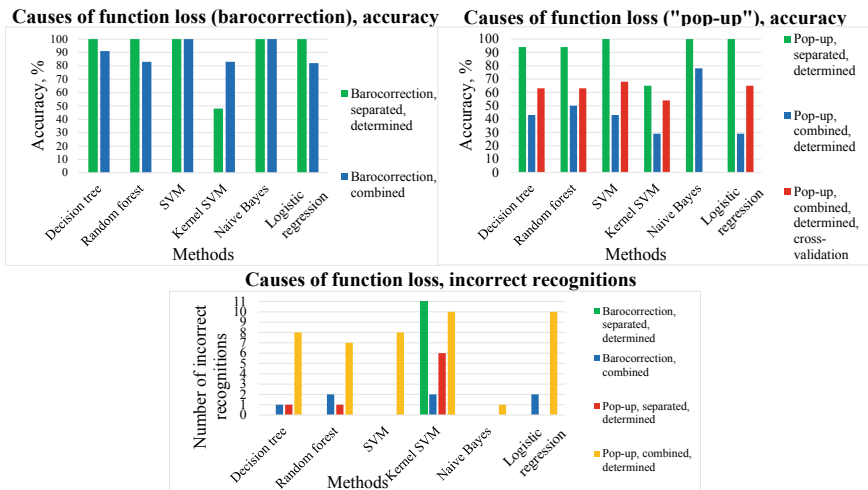
Fig. 1.5 The results of function loss causes detection of the barometric correction enter and "pop-up" functions

the "pop-up" function. SVM, naive Bayes, and logistic regression showed the best accuracy in the determined training data case for either function. SVM and naive Bayes also gave 100% accuracy in the combined dataset case. For the "pop-up" function, the algorithms showed low values of accuracy in the combined dataset case.

1.3.5 Discussion

The research results showed that the algorithms handled separated and combined datasets in different ways. The determinateness of feature combinations also influenced the results. Table 1.7 contains a summary of recommendations that provide information about which type of dataset and algorithm are suitable for the particular problem.

Kernel SVM showed weak results that can be explained by the possible incorrect choice of coefficients. The algorithm showed good results in works [7, 8], and should be investigated more accurately in the terms of considered problems.

Some observations were pointed out during the research but they were not analyzed in detail. The sample of the dataset also influenced the accuracy and the number of incorrect recognition as well as the number of classes and corresponding features. These unconfirmed conclusions require additional research.

Neural networks find a successful application not only in aerospace [12, 13] but also in different spheres [14–17] including fault localization problem [18]. These

Table 1.7 Summary recommendations

Problem	Type of dataset	Recommended algorithms
Function loss	Combined	Naive Bayes
Behavioral indication	Separated, determined training data	Decision tree, naive Bayes, logistic regression as an assumption. Further investigation is required
Detection of function loss causes	Separated, determined training data	Decision tree, random forest, logistic regression, SVM, naive Bayes
	Combined	SVM, naive Bayes. Further investigation is required

algorithms were not considered in this paper. It should be pointed out that the application of adaptive software as neural networks is of limited use to date in the aviation domain. However, Federal Aviation Administration (FAA) has accomplished a research study of software assurance requirements for adaptive systems [19].

Therefore, neural networks are going to become one of the points of further research.

1.4 Conclusion

The paper describes the digital twin synthesis of the barometric correction enter function of avionics and machine learning classification results analysis of function loss, behavioral indication, and detection of function loss causes problems of the digital twin.

The obtained classification results clarified what particular algorithm and the type of dataset are appropriate to work on mentioned problems. The significance of the problems also was emphasized for the design and development stages of an avionics lifecycle.

As was mentioned before, conditional logic has a simple implementation for the function loss and behavioral indication problems in physical assets if a function is uncomplicated. In other circumstances, the implementation of conditional logic leads to errors. Avionics equipment is going to become more complicated, have significantly more features, and its behavior is going to become more unpredictable. As a result, the conditional logic is likely to malfunction because of implementation complexity. Machine learning algorithms application expects additional research for such challenging problems according to FAA requirements.

The digital twin of the barometric correction enter function is a relatively simple example. Avionics and aircraft have many functions that are more complicated

because they have large computation parts and a lot of crossfeeds between their operations and aggregated components. Maneuvering computation and hazards computation are the direct examples of such functions. The research of more complicated avionics functions is a possible subject for further work.

Applying the formulated approach for consideration of every avionics or aircraft function, the common parts between created functions digital twins can be defined. These common parts will be able to become a basis for integrated digital twins, optimizing created models and design, development, and failure safety processes.

Another advantage of digital twins application is increasing the remaining life of LRU at the debugging and testing stages of a lifecycle. Therefore, a digital twin can be used most of the debugging and testing time instead of a physical LRU.

The presented approach can become a flexible platform for design, development, failure safety, and quality assessment processes in view of the further increasing complexity of avionics.

It should be noticed, that the process of digital twin design and development can expend a large number of resources, require more data, and increase the product development cost. The required DAL, fidelity level of a digital twin, complexity of the development, and its estimated cost should be analyzed in the decision-making process.

References

1. J. Ríos, F. Mas, M. Oliva, J.C. Hernandez-Matias, Framework to support the aircraft digital counterpart concept with an industrial design view. Int. J. Agile Syst. Manage. **9**(3), 212–231 (2016)
2. C.M. Ezhilarasu, Z. Skaf, I. Jennions, Understanding the role of a digital twin in the field of integrated vehicle health management (IVHM), in *IEEE International Conference on Systems, Man, and Cybernetics*, pp 1500–1507. (Bari, Italy 2019)
3. University of Minnesota Homepage. https://www.ima.umn.edu/2017-2018/SW3.6-8.18/26842. Last accessed 10 Mar 2021
4. A. Maheshwari, N. Davendralingam, A.D. DeLaurentis, A Comparative STUDY OF MACHINE LEARNING TECHNIQUES FOR AVIATION APPLICAtions, in *Aviation Technology, Integration, and Operations Conference.* Atlanta, GA, USA (2018). https://doi.org/10.2514/6.2018-3980.
5. SAE ARP4754, Guidelines for Development of Civil Aircraft and Systems, Rev.A. Revised 2010–12
6. K. Trusova, Health analysis and diagnostic program synthesis for avionics systems, in *IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*, pp 103–107. (IEEE, Saint Petersburg, Russia, 2019)
7. K.V. Trusova, Machine learning methods application for the avionics systems health analysis and faults localization challenges, in *International Youth Conference on Electronics, Telecommunications and Information Technologies (YETI-2020)* ed. by E. Velichko, M. Vinnichenko, V.

Kapralova, Y. Koucheryavy. Springer Proceedings in Physics, vol 255 (Springer, St. Petersburg, Russia, 2020), pp. 383–397

8. K. Trusova, Comparison of two classification machine learning models of avionics systems for health analysis, in *IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*. IEEE, Saint Petersburg, Russia, 2020, pp. 172–175

9. C.M. Ezhilarasu, Z. Skaf, I. Jennions, A Generalised methodology for the diagnosis of aircraft systems. IEEE Access, IEEE Reliab. Soc. Section 11437–11454, (2021). https://doi.org/10.1109/ACCESS.2021.3050877

10. M.A. Lopatin, S.A. Fyodorov, S.V. Zavjalov, D. Ge, Object classification based on channel state information using machine learning, in *International Youth Conference on Electronics, Telecommunications and Information Technologies (YETI-2020)* ed. by E. Velichko, M. Vinnichenko, V. Kapralova, Y. Koucheryavy, Springer Proceedings in Physics, vol 255 (Springer, St. Petersburg, Russia, 2020), pp. 367–374

11. E. Mirkin, E. Savchenko, Configuring the interval target in a multilayer feed forward neural network on the example of the problem of medical diagnostics, in *International Youth Conference on Electronics, Telecommunications and Information Technologies (YETI-2020)* ed. by E. Velichko, M. Vinnichenko, V. Kapralova, Y. Koucheryavy. Springer Proceedings in Physics, vol 255 (Springer, St. Petersburg, Russia, 2020), pp. 463–475

12. D. Rengasamy, H.P. Morvan, G.P. Figueredo, Deep learning approaches to aircraft maintenance, repair and overhaul: a review, in *21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, HI, USA (2018), pp 150–156

13. J. Ríos, Unsupervised anomaly detection in flight data using convolutional variational auto-encoder. Aerospace **7**(8), 115 (2020). https://doi.org/10.3390/aerospace7080115

14. A. Grevtseva, K. Dominges, M. Dominges, Neural network for processing ultrasonic signals in flaw detection control systems, in *IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)* (IEEE, Saint Petersburg, Russia, 2020), pp. 180–183

15. I.R. Aliev, V.A. Pavlov, S.V. Zavjalov, Y. Sadovaya, Application of a convolutional neural network for detection of ignition sources and smoke, in *International Youth Conference on Electronics, Telecommunications and Information Technologies (YETI-2020)* ed. by E. Velichko, M. Vinnichenko, V. Kapralova, Y. Koucheryavy. Springer Proceedings in Physics, vol. 255 (Springer, St. Petersburg, Russia, 2020), pp. 421–427

16. S.B. Makarov, V.A. Pavlov, A.K. Bezborodov, A.I. Bobrovskiy, D. Ge, Multiple object tracking using convolutional neural network on aerial imagery sequences, in *International Youth Conference on Electronics, Telecommunications and Information Technologies (YETI-2020)* ed. by E. Velichko, M. Vinnichenko, V. Kapralova, Y. Koucheryavy. Springer Proceedings in Physics, vol 255 (Springer, St. Petersburg, Russia, 2020), pp. 413–420

17. E. Mirkin, E. Savchenko, E. Savchenko, Use of the method of configuring the interval target in the problem of synthesis of the neural network classifier for diagnosing cardiovascular diseases, in *IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*, (IEEE, Saint Petersburg, Russia, 2020), pp. 140–144

18. W. Zheng, D. Hu, J. Wang, Fault localization analysis based on deep neural network. Math. Prob. Eng. 2016, Article ID 1820454, 11 pages (2016). https://doi.org/10.1155/2016/1820454.

19. Federal Aviation Administration Homepage. https://www.faa.gov/aircraft/air_cert/design_approvals/air_software/media/TC-16-4.pdf. Last accessed 12 Mar 2021.

Chapter 2
On the Application of a Fully Connected Neural Network for Multifrequency Signals Demodulation

Anastasiia I. Semenova and **Sergey V. Zavjalov**

Abstract The paper presents a research of the effectiveness of using neural networks to demodulate SEFDM signals when passing through a channel with AWGN. The questions of choosing the architecture, the number of layers and the number of neurons in the layers of a neural network for processing SEFDM signals with different bandwidth compression factor between subcarriers are considered. Simulation modeling and comparison with "classical" signal demodulation algorithms are carried out.

Keywords SEFDM · Neural networks · Demodulation

2.1 Introduction

5-6G standards are focused on reducing the consumed and radiated power and increasing the spectral efficiency, which is calculated as the ratio of the value of the information rate to the occupied frequency band [1–4]. A decrease in the value of the peak-to-average power ratio is associated with an increase in efficiency [5–7], an increase in spectral efficiency can be achieved by using the SEFDM (Spectrally Efficient Frequency Division Multiplexing) technology of signals [8–11]. The main idea of this technology is a closer arrangement of the subcarrier frequencies, which leads to an increase in spectral efficiency. However, in this case, bit error rate (BER) performance degrades [12–16]. Multi-frequency spectral-effective signals are described by the following formula [10–16]:

$$x(t) = \frac{1}{\sqrt{T}} \sum_{i=-\infty}^{\infty} \sum_{n=0}^{N-1} s_{i,n} e^{\left(\frac{j2\pi n\alpha(t-iT)}{T}\right)}$$

A. I. Semenova (✉) · S. V. Zavjalov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: semenova.aigor@gmail.com

where N is the number of subcarrier, $\alpha = \Delta f T$ is the bandwidth compression factor, Δf is interval between subcarriers, T is the period of one SEFDM symbol, $s_{i,n}$ is symbol modulated on the n-th sub-carrier.

To solve the problem of degradation of BER performance of SEFDM signals, it is necessary to use more complex algorithms, in the limit-full search algorithms [12–17]. However, the computational complexity of these algorithms is high, which requires large computational resources.

The application of neural networks for demodulating SEFDM signals can allow us to obtain BER performance better than the algorithm of element-by-element processing [17, 18]. However, as opposed to the full search algorithms, the computational complexity of neural network approaches is lower.

As a result of the analysis of the literature [19–22], the prospects for using full connected neural networks were shown. This architecture has the property of taking into account interference between subcarriers, since all neurons of this architecture are connected to each other [23]. Thus, it is possible to take into account interference in the frequency domain to improve the BER performance.

The aim of the work is to evaluate the features of the use of neural networks for demodulating SEFDM signals.

2.2 The Structure of Neural Network

Two structures were used to study the demodulation of multi-frequency signals using a full connected neural network. The first network structure for demodulating multi-frequency signals without intersymbol interference (ISI) is shown in Fig. 2.1. This structure has one hidden layer with one neuron. The input of the network consistently receives samples of the signals from the subcarrier frequencies.

Figure 2.2 shows the structure of a fully connected neural network to combat intersymbol interference. The peculiarity of this architecture is the number of hidden layers and neurons: three hidden layers of 5 neurons in each layer.
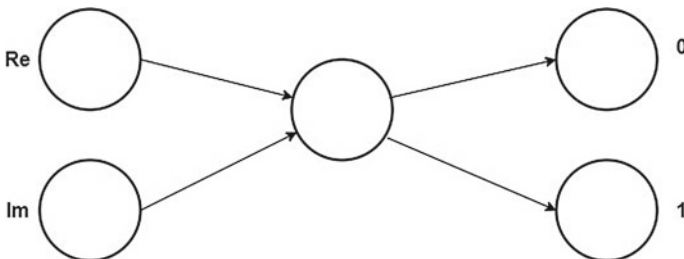
Fig. 2.1 Network structure with one hidden layer and one neuron
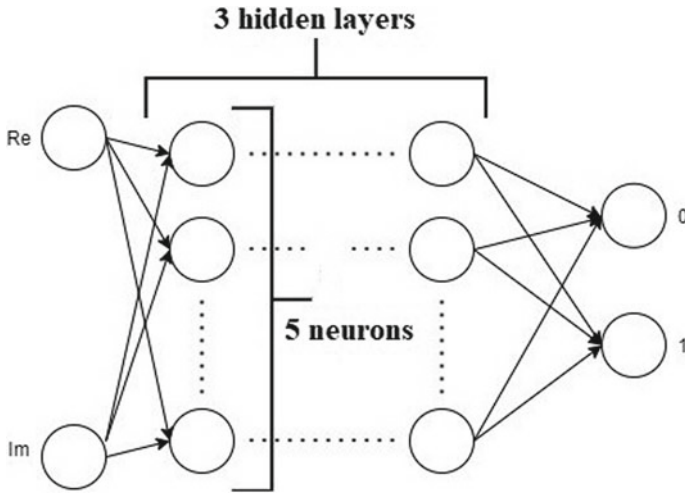
3 hidden layers

Fig. 2.2 Network structure with three hidden layers and five neurons

Figure 2.3 shows a modified structure of a fully connected neural network that takes into account the influence of a neighboring subcarrier. In addition to the demodulated signal, the signal samples from the neighbor subcarrier frequency are fed to the network input. First, a direct neural network pass is implemented. The input is supplied with signal samples divided into real and imaginary components. Next, the weighting and summation are performed.
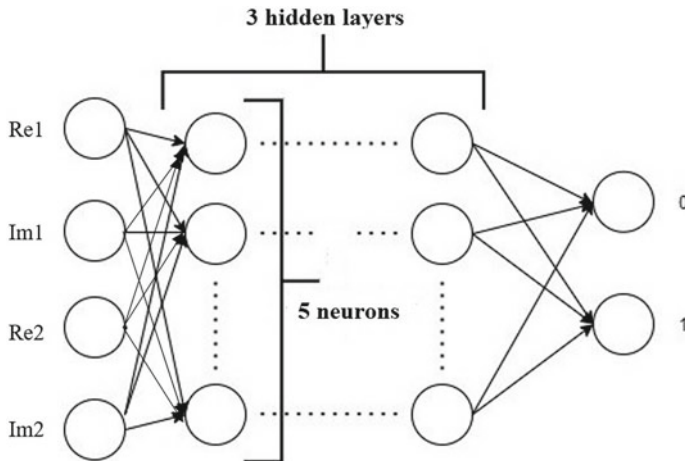
3 hidden layers

Fig. 2.3 The structure of the network with three hidden layers and five neurons, taking into account the samples of the signals of the neighboring subcarrier
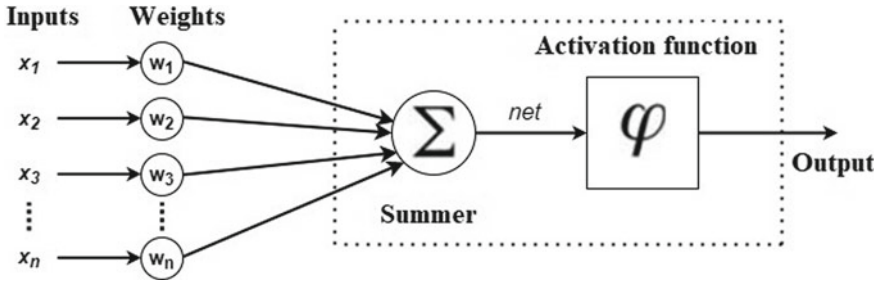
Fig. 2.4 A model of an artificial neuron. Direct passage
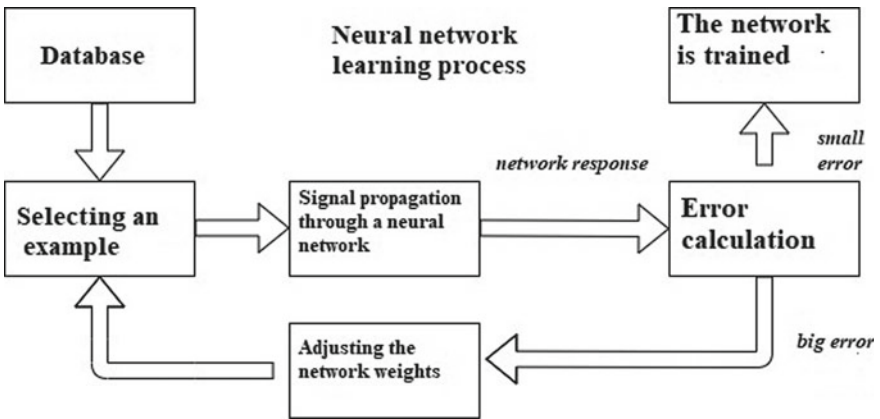
Fig. 2.5 The process of training a neural network. The example in the figure indicates a complex count

The activation function calculates the output value and feeds it to the network output in this example (Fig. 2.4). Note that an example is given without hidden layers, for the convenience of understanding how a direct pass occurs. In this paper, the hyperbolic tangent activation function was selected. The selection was made taking into account the range of input data values in the range from -1 to $+1$.

To train the neural network, the built-in toolbox Neural Network MATLAB library was used. The training process is performed until the minimum error is reached (Fig. 2.5). The minimum error criterion is set before the start of training and is the value of the error function (mse) $\leq 10^{-5}$.

2.3   Simulation Model

Firstly, procedure of training of the neural network (NN) is performed. The network input is fed samples of the signal from the subcarrier frequencies sequentially after
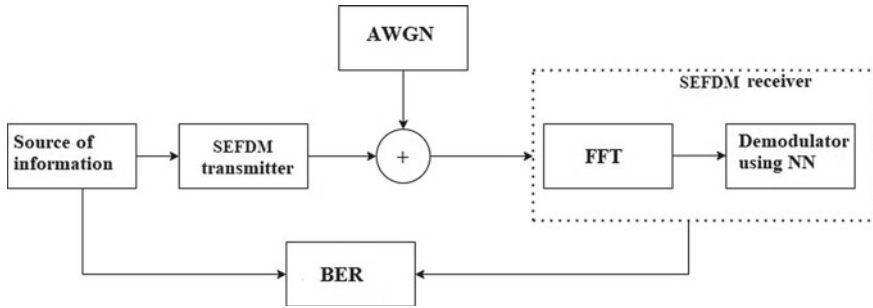
Fig. 2.6   Block diagram of the simulation model

the fast Fourier transform (FFT). Before entering the NN, the samples are divided into real and imaginary components (Fig. 2.6). After the direct passage of the neural network, the learning process takes place, which was carried out using the method of back propagation of the error. During this process, the weight coefficients are corrected. Training using the back propagation method of error consists in correcting the weight coefficients so that the error at the output of the neural network decreases. A network is considered trained if the minimum error at the network output is the same over several iterations. The weight coefficients are adjusted in the opposite direction, i.e., the weight coefficients are adjusted from the output layer to the input layer. The weight coefficients can be any real numbers. The procedure for teaching this architecture is described in detail in the following literature [23].

When analyzing the BER performance, the SEFDM signal is formed at the first stage, which is superimposed with additive white Gaussian noise (AWGN) in the range from 0 to 20 dB. At the input of the receiver, a direct Fourier transform is performed, after which the samples are fed to the input of the neural network for demodulation and error calculation. At least 10^6 information bits were transmitted to each point according to the error probability.

2.4   Simulation Results

Figure 2.7 shows the BER performance curves of the SEFDM signal for different values of the bandwidth compression factor α for a single-beam channel with AWGN with the first network structure (consisting of one hidden layer and one neuron). Due to a decrease in the compaction coefficient, the BER performance deteriorates, which is caused by the influence of neighboring subcarriers on each other and the network structure with one neuron does not cope with the ISI.

As can be seen in Fig. 2.7, in the case of a compaction coefficient of 0.9, the energy loss is 5 dB for error probability equal to 10^{-3}. Thus, it is necessary to complicate the network structure, that is, to use the second presented type of structure, consisting of three hidden layers, each with 5 neurons.
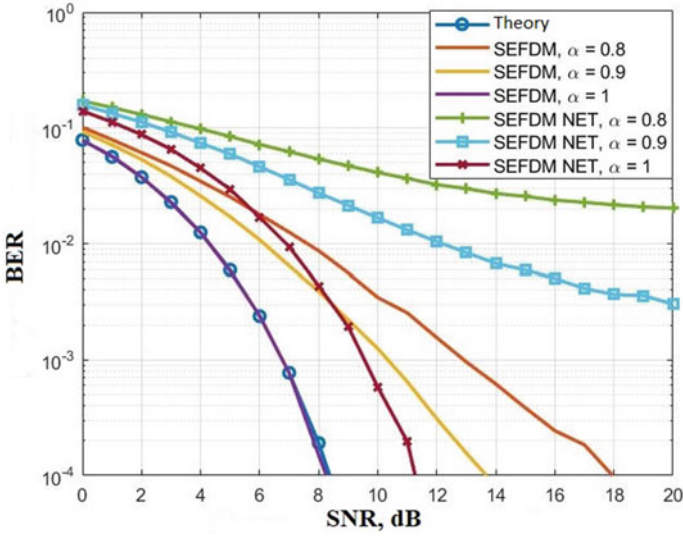
Fig. 2.7 The BER performance of SEFDM signals in a channel with AWGN (one hidden layer and one neuron)

Figure 2.8 shows the BER performance curves of the SEFDM signal for different values of the bandwidth compression factor α for a single-beam channel with AWGN
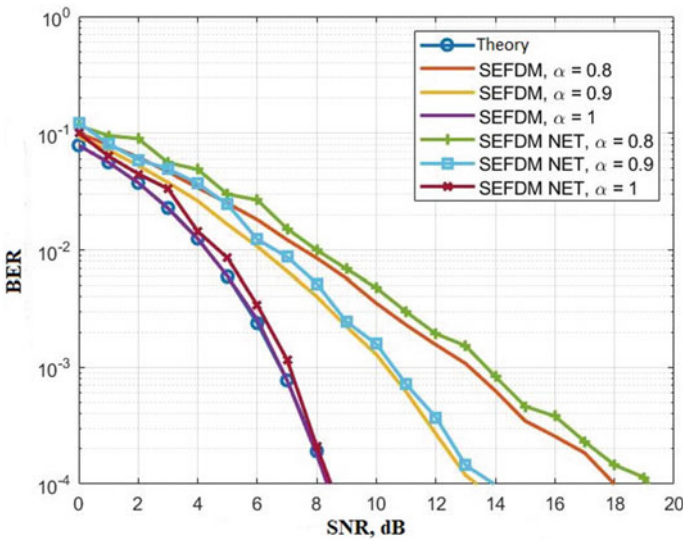
Fig. 2.8 The BER performance of SEFDM signals in channel with AWGN (three hidden layers and five neurons)
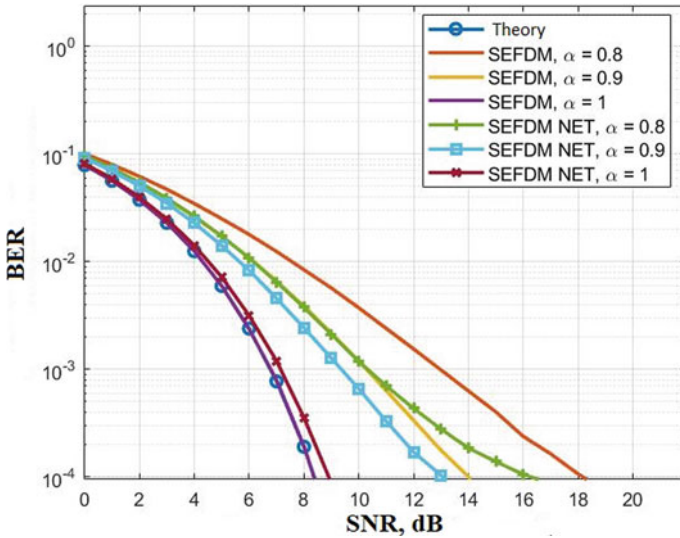
Fig. 2.9 The BER performance SEFDM signals in channel with AWGN taking into account the samples of the signals of the neighboring subcarrier (three hidden layers and five neurons)

with a second network structure (consisting of three hidden layers and five neurons in each layer). It can be seen that the BER performance in the case of using NN has improved, the energy loss for the case $\alpha = 1$ is practically absent.

Figure 2.9 shows the curves of the BER performance of the SEFDM signal for different values of compression ratio for single-beam channel AWGN given counts of neighbor subcarrier signals. The network structure of such reception is shown in Fig. 2.3. Based on the result in Fig. 2.9, we note that when taking into account the samples of the neighboring subcarrier, this version of the demodulation algorithm is more stable to the ISI compared to the algorithm of the element-by-element reception. The energy gain at alpha 0.8 was 3 dB in terms of the probability of error 10^{-3}.

Figure 2.10 shows the dependence of the error probability on the number of subcarriers (1 demodulated and 1/2/3/4/5/6 neighboring) at a fixed SNR = 15 dB. If 5 and 6 neighbor subcarriers are considered, for the case $\alpha = 0.9$, the error probability is equal to 10^{-5}. For the case $\alpha = 0.8$, when using 5 neighbor subcarriers, the error probability tends to 10^{-5}.

Figure 2.11 shows the dependence of the error probability on the number of subcarriers (1 demodulated and 1/2/3/4/5/6 neighboring) at a fixed SNR = 20 dB. When 5 neighbor subcarriers are considered, for the case $\alpha = 0.9$, the error probability tends to 10^{-7}. For the case $\alpha = 0.8$, when using 5 neighbor subcarriers, the error probability tends to 10^{-6}.
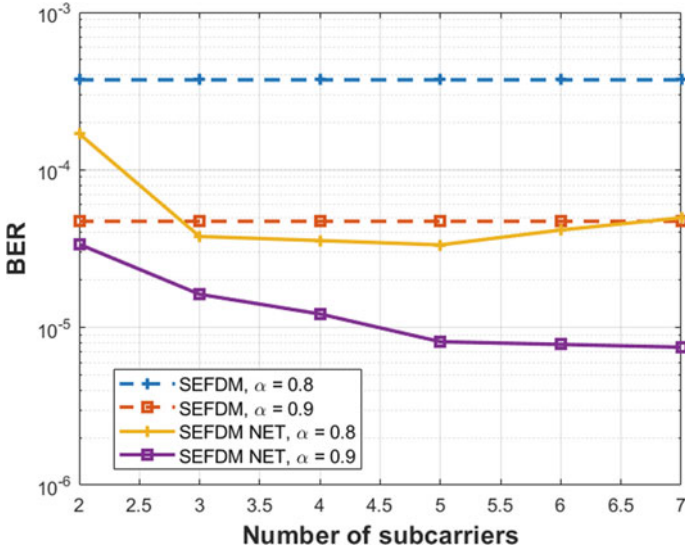
Fig. 2.10 The dependence of the error probability on the number of subcarriers at a fixed SNR = 15 dB
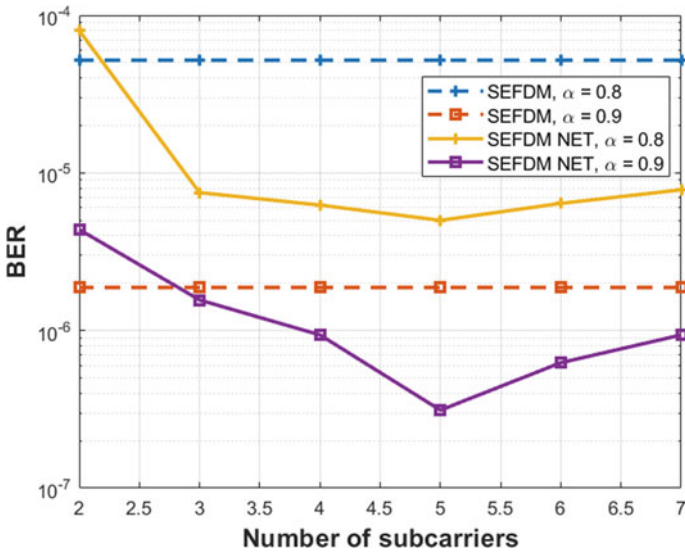
Fig. 2.11 The dependence of the error probability on the number of subcarriers at a fixed SNR = 20 dB

2.5 Conclusions

In this paper, we have studied options for demodulating a multi-frequency signal using a neural network, and compared the results with two types of network structures and with demodulated signals without considering the neural network. A structure consisting of 3 hidden layers of 5 neurons each is more resistant to ISI than a network structure consisting of 1 hidden layer and 1 neuron.

The BER performance at the element-by-element reception is less resistant to the ISI than the result considering the samples of the neighboring subcarrier. As the network structure becomes more complex, the energy loss for the compaction coefficient $\alpha = 1$ decreases.

The most promising NN for combating ISI is the modified NN structure, which considers the influence of signals from neighboring subcarrier frequencies (Fig. 2.3). In the future, it is planned to implement the demodulator operation using neural networks in a multipath channel.

References

1. S.B. Makarov, M. Liu, A.S. Ovsyannikova, S.V. Zavjalov, I.I. Lavrenyuk, W. Xue, J. Qi, Optimizing the shape of faster-than-nyquist (FTN) signals with the constraint on energy concentration in the occupied frequency bandwidth. IEEE Access **8**, 130082–130093 (2020)
2. S.B. Makarov, A.M. Markov, Incoherent reception "in a whole" of spectrally effective signals formed on the basis of phase pulses of the shape $\sin^p x$. Radioengineering. **84** 12(24), 43–51 (2020)
3. M. Irfan, S. Aïssa, Generalization of index-modulation: breaking the conventional limits on spectral and energy efficiencies, in *IEEE Transactions on Wireless Communications* (2021), pp.1–1
4. Y. Sadovaya, A. Gelgor, Synthesis of signals with a low-level of out-of-band emission and peak-to-average power ratio, in *2018 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*, St. Petersburg (2018), pp. 103–106
5. V. Sorotsky, R. Zudov, Application of neural networks to reduce distortion of RF Signals in switch mode power amplifiers, in *2020 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*, St. Petersburg (2020), pp. 76–79
6. A. Pergushev, V. Sorotsky, A. Ulanov, Criteria for selection envelope tracking power supply parameters for high peak-to-average power ratio applications, in *2019 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*, St. Petersburg (2019), pp. 13–16
7. A. Pergushev, V. Sorotsky, Signal distortion decreasing in envelope tracking power amplifiers, in *2018 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*, St. Petersburg (2018), pp. 44–47

8. V.I. Slyusar, K.A. Vasil'ev, Potential limits of frequency division multiplexing of N-OFDM signals based on Hartley's basis functions. Radioelectron. Commun. Syst. **51**(3), 129–133 (2008)

9. V.I. Slyusar, V.G. Smolyar, Communication channels frequency multiplexing on the basis of superrayleigh signals resolution. Izvestiya Vysshikh Uchebnykh Zavedenij. Radioelektronika **46**(7), 30–39 (2003)

10. I. Darwazeh, H. Ghannam, T. Xu, The first 15 years of SEFDM: a brief survey, in *2018 11th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP)*, Budapest (2018), pp. 1–7

11. H. Ghannam, I. Darwazeh, SEFDM: spectral efficiency upper bound and interference distribution, in *2018 11th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP)*, Budapest, Hungary (2018), pp. 1–6

12. I. Kanaras, A. Chorti, M. Rodrigues, I. Darwazeh, Analysis of sub-optimum detection techniques for bandwidth efficient multi-carrier communication system, in *Cranfield Multi-Strand Conference: Creating Wealth Through Research and Innovation* (2008), pp. 505–510

13. A. Kislitsyn, A. Krylov, A. Rashich, Experimental evaluation of SEFDM trellis based demodulator, in *2020 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*, St. Petersburg (2020), pp. 110–113

14. I. Kanaras, A. Chorti, M.R.D. Rodrigues, I. Darwazeh, Spectrally efficient FDM signals: bandwidth gain at the expense of receiver complexity, in *2009 IEEE International Conference on Communications*, Dresden (2009), pp. 1–6

15. A. Rashich, S. Gorbunov, ZF equalizer and trellis demodulator receiver for SEFDM in fading channels, in *2019 26th International Conference on Telecommunications (ICT)*, Hanoi, (2019), pp. 300–303

16. A. Gelgor, A. Gorlov, V.P. Nguyen, Performance analysis of SEFDM with optimal subcarriers spectrum shapes, in *2017 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, Istanbul, (2017), pp. 1–5

17. T. Xu, I. Darwazeh, Design and prototyping of neural network compression for non-orthogonal IoT signals, in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, Marrakesh (2019), pp. 1–6

18. T. Xu, T. Xu, I. Darwazeh, Deep learning for interference cancellation in non-orthogonal signal based optical communication systems, in *2018 Progress in Electromagnetics Research Symposium (PIERS-Toyama)*, Toyama (2018), pp. 241–248

19. A. Liu, Y. Yang, Q. Sun, Q. Xu, Deep fully convolution neural network for semantic segmentation based on adaptive feature fusion, in *2018 5th International Conference on Information Science and Control Engineering (ICISCE)*, Zhengzhou (2018), pp. 16–20

20. C. Catanese, R. Ayassi, E. Pincemin, Y. Jaouen, A fully connected neural network approach to mitigate fiber nonlinear effects in 200G DP-16-QAM transmission system, in *22nd International Conference on Transparent Optical Networks (ICTON)* (2020), pp 1–4

21. M. Mazumdar, V. Sarasvathi, A. Kumar, Object recognition in videos by sequential frame extraction using convolutional neural networks and fully connected neural networks, in *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, Chennai (2017) pp. 1485–1488

22. X. Fu, H. Qu, Research on semantic segmentation of high-resolution remote sensing image based on full convolutional neural network, in *2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE)*, Hangzhou (2018), pp. 1–4

23. S. Haykin, in *Neural Networks and Learning Machines*. 3rd edn, (Pearson, 2008), p. 936

Chapter 3
Automated Power Amplifier Design Through Multiobjective Bottom-Up and Particle Swarm Optimizations Using Neural Network

Lida Kouhalvandi and Ladislau Matekovits

Abstract This study presents an automated power amplifier (PA) design process by optimizing the topology and values of design parameters, sequentially. The automated optimization environment is created with the combination of an electronic design automation tool and a numerical analyzer. As a first step, the configuration of the PA is generated using the bottom-up optimization (BUO) method, then the values of the components are optimized using the particle swarm optimization (PSO) algorithm that is employed with a shallow neural network. The PSO method is applied for optimizing PA in terms of output power, power gain, and efficiency leading to obtain optimal design parameters. The proposed optimization process is automatic and compact leading to reduce interruptions of designers during the process. In order to verify the effectiveness of the presented method, one lumped element PA including GaN HEMT transistor is designed and optimized. The optimized PA reveals higher than 45% power added efficiency with the linear gain performance between $10 \div 14.6$ dB in the frequency band of $1 \div 2.3$ GHz.

Keywords Automated · Bottom-up optimization (BUO) · Multiobjective · Particle swarm optimization (PSO) · Power amplifier · Shallow neural network (SNN)

L. Kouhalvandi
Electronics and Communication Engineering Department, Istanbul Technical University, 34467 Istanbul, Turkey

L. Kouhalvandi (✉) · L. Matekovits
Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Torino, Italy
e-mail: lida.kouhalvandi@ieee.org

L. Matekovits
e-mail: ladislau.matekovits@polito.it

L. Matekovits
Department of Measurements and Optical Electronics, Politehnica University Timisoara, 300223 Timisoara, Romania

Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni, National Research Council, 10129 Turin, Italy

3.1 Introduction

Future mobile networks supporting the sixth generation (6G) technology require high performance systems that can manage high amount of data. For this case, power amplifiers (PAs) play an important role in the wireless communication systems [1, 2] and they must be of high performance in terms of output power (P_{out}), power gain (G_p), and power added efficiency (PAE) [3]. Concurrently achieving these three significant specifications is not straightforward and requires multiobjective optimization-based methods.

Optimization-based approaches, opposite to the Knowledge-based approaches, are divided into three techniques, namely: equation-based, simulation-based, and learning-based approaches [4]. In the field of radio frequency (RF), there is a large amount of data to be dealt with due to the non-linearity of active devices and harmonic effects of passive and active components. Hence, learning-based method gets the attention of researchers as this method can pave the ways of RF designs by modeling the circuits. This type of learning provides the relationships between the input and the output data that are design parameters and design specifications, respectively [5, 6] by properly weighting the available design parameters. This method has been recently applied for sizing the included design parameters (input) [7] and is employed to optimize various design specifications (output) of PAs. From another point of view, this type of learning provides an automated environment and reduces the dependency to the designer's experiences.

This work presents a two-phase optimization process that is performed automatically and sequentially aiming to both configuring the structure and also predicting the design variables. The first phase optimization is applied using the bottom-up optimization (BUO) algorithm where the lumped element input and output matching networks (MNs) are constructed. Then the multiobjective particle swarm optimization (PSO) algorithm is employed for optimizing the design parameters of the constructed PA in the first phase. This algorithm is employed using the shallow neural network (SNN) includes one hidden layer and optimizes three important specifications of PA that are output power, gain, and efficiency. As Fig. 3.1 represents, the proposed method paves the way of RF designers to reduce the human interruptions during the optimization process where by selecting the transistor model, both the structure and values of design parameters are optimized automatically.

This paper is organized as follows. Section II provides in detail descriptions about the optimization process. The validation of the method is provided in Sec. III by designing and optimizing a wideband lumped element PA. Finally, Sec. IV concludes this work.
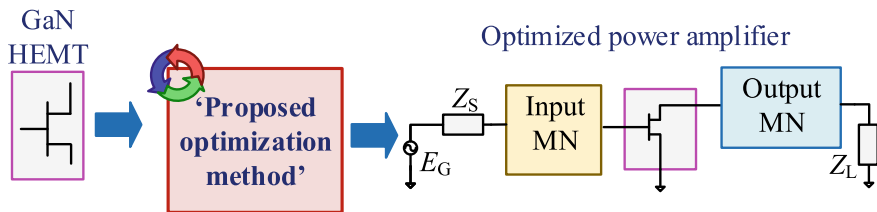
Fig. 3.1   General flowchart of proposed optimization method

3.2   Proposed Automated Optimization Process

Our presented method is devoted to concurrently optimize i) the general structure of the PA that includes reactive components (i.e., capacitor (C) and inductor (L)), and ii) the optimal values of involved components. The first phase is performed by applying the BUO algorithm and the second phase is executed using the multiobjective PSO algorithm. The PSO method is employed using a shallow neural network (SNN). More importantly, the entire optimization process is performed automatically in the created environment that is the combination of electronic design automation (EDA) tool, as ADS, and numerical analyzer, as MATLAB. Interested reader in further details can refer to [8] for getting the view point of constructed optimization platform. In this section, the detailed description of an automated optimization-oriented process are given that leads to optimize the lumped element PA in terms of P_{out}, G_p, and PAE.

3.2.1   Bottom-Up Optimization (BUO) Algorithm

Providing that the first structure of any PA includes input and output MNs, is an important step in RF designs. In order to obtain the optimal structure and configuration of the PA, the BUO method is employed in this paper. This method starts with one unit cell (that includes one capacitor and one inductor) in both input and output sides of MNs; sequentially increases in the number of unit cells as Fig. 3.2 shows is then performed. Regarding the depicted Smith chart in [9], inductor-capacitor networks that are normalized to 50 Ω are used as the unit blocks of MNs.

As mentioned above, the BUO starts with one unit cell in the input and output MNs and sequentially increases their number on both sides up to achieving the determined design goals. The initial values of each C and L are determined by employing the *random optimization (RO)* method presented in [8]. This optimization process leads to generate the optimal configuration of the PA that consists of passive L and C components and also provides an effective way of synthesizing the MN topologies.
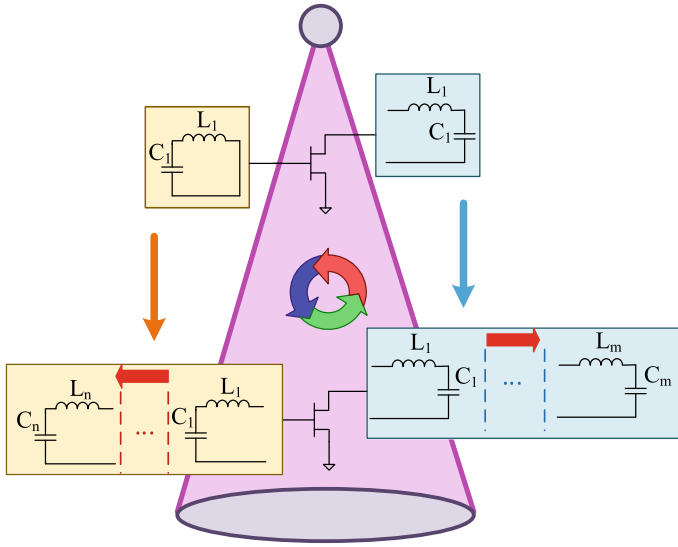
Fig. 3.2 BUO algorithm for configuring the structure of the PA design

3.2.2 Particle Swarm Optimization (PSO) Algorithm

After constructing the PA's structure using the BUO algorithm, the optimal values of design parameters (i.e., Cs and Ls) must be determined. For this case, the multi-objective PSO algorithm [10] is employed for achieving the optimal variables. This algorithm is based on the Pareto Optimal Front (POF) and aims to find the optimal values achieved by trading-off between two functions namely f_1 and f_2. For our problem, these two functions are represented as $f_1 = G_p(P_{out})$ and $f_2 = PAE(P_{out})$ where both G_p and PAE are functions of output power.

For employing this algorithm, we get benefit of learning-based approach that is using the SNN. Figure 3.3 shows the implementation of multiobjective PSO algorithm using the SNN. This network consists of one hidden layer with several neurons. For obtaining the optimal number of neurons, the *rule of thumb* approach is employed [11].

3.3 Designed and Optimized Lumped Element PA

For validating the proposed optimization process, one lumped element 10 W PA in the operational frequency band from 1 GHz to 2.3 GHz is designed and optimized. The used transistor model is Wolfspeed CGH40010 Gallium Nitride (GaN) high-electron mobility transistor (HEMT) and the designed PA is implemented on Rogers RO4350B substrate with relative dielectric constant $\varepsilon_r = 3.66$ and a thickness of

Fig. 3.3 Employment of the multiobjective PSO algorithm through shallow neural network for predicting the optimal design parameters; $f_1 = G_p(P_{out})$ and $f_2 = PAE(P_{out})$
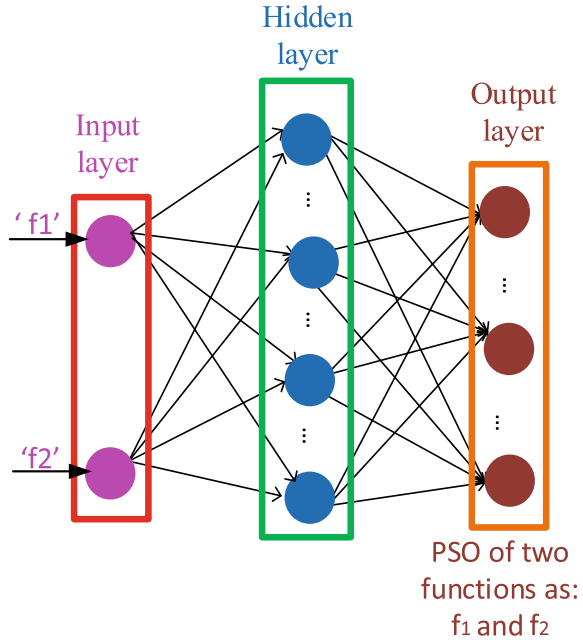
0.508 mm. The optimization process is performed on CPU execution environment that has specifications as: Intel Core i7-4790 CPU @ 3.60 GHz with 8.0 GB RAM.

Figure 3.4 shows the optimized PA by using the BUO and PSO algorithms leading to generate optimal structure with sizes of included components. As can be seen, the BUO algorithm provides four unit cells in the input MN side and three unit cells in the output MN sides. The initial reactive component values are achieved using the RO method where the values are randomly increased or decreased.

After constructing the PA, it is time to achieve the optimal component values. For this case, suitable amount of data must be created for training SNN. Randomly the component values are changed and the corresponding PA responses (i.e., P_{out}, G_p, and PAE) over the interested frequency band are achieved. In total 1000 training
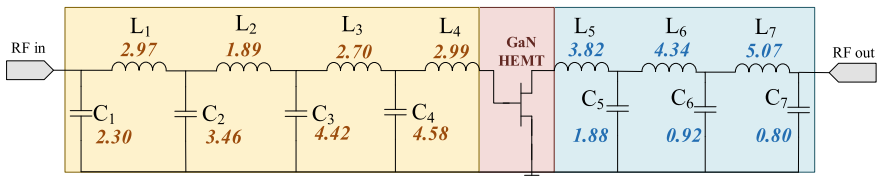
Fig. 3.4 Optimized lumped element PA using the BUO for constructing the PA structure and also the PSO method for optimizing the design parameters sequentially and automatically; The units for each inductor and capacitor are nH and pF, respectively

Fig. 3.5 S-parameter simulation results of optimized PA in Fig. 3.4
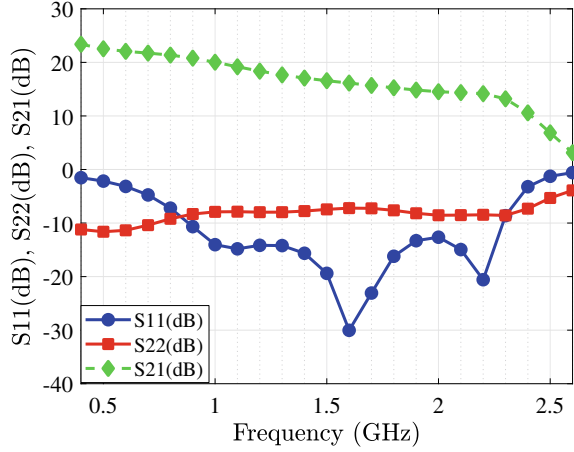
data and 500 testing data are obtained and the rule of thumb approach is applied for predicting the optimal neuron numbers. The number of neurons are achieved by using the training/testing data and also applying the rule of thumb approach where the simulation results reveal that when there are 100 neurons in the hidden layer, the training and testing accuracy are 95.54% and 93.02%, respectively.

The optimized values of passive components are shown in Fig. 3.4 and the related simulation results are presented in Figs. 3.5, 3.6, 3.7. Figure 3.5 presents the S-parameter simulation results in terms of S_{11}, S_{22}, and S_{21}. For the optimized PA, the S_{11} is lower than -30 dB in the whole operation frequency band. The PA achieves a simulated S_{21} of 20 dB at 1 GHz and it remains flat and higher than 14 dB in the operation frequency band. The optimized three specifications (i.e., P_{out}, G_p, PAE) are shown in Fig. 3.6 that are extracted at 3-dB gain compression. The automatically optimized PA reveals around 40 dBm output power in the operation frequency band. The minimum and maximum PAE values appears at 1.9 GHz and 1 GHz that are 45.28% and 61.54%, respectively. Additionally, the stability of the optimize PA is reported in Fig. 3.7 that illustrates its well performance in the large signal operation frequency band.

3.4 Conclusion

This work presents an automated optimization process for designing a lumped element PA using the BUO and PSO algorithms sequentially. Firstly, the BUO method is employed for configuring the structure of the PA. Then, the multiobjective PSO algorithm is employed with a SNN for achieving the optimal values of reactive components. The presented method is used for optimizing concurrently the output power, gain, and efficiency specifications of the PA. This learning-based optimization

Fig. 3.6 P_{out}, G_p, and PAE results of optimized PA @ 3-dB gain compression
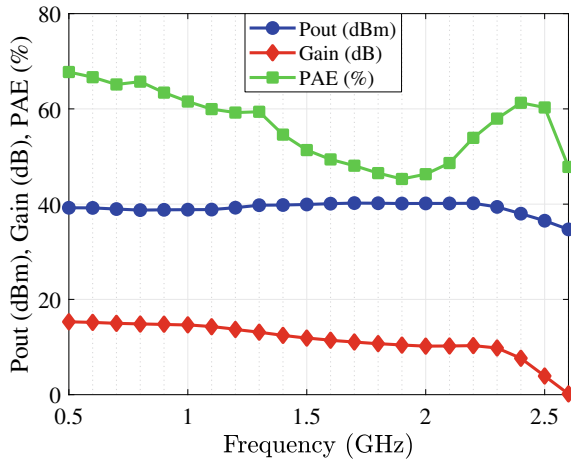
Fig. 3.7 Stability factor of optimized PA
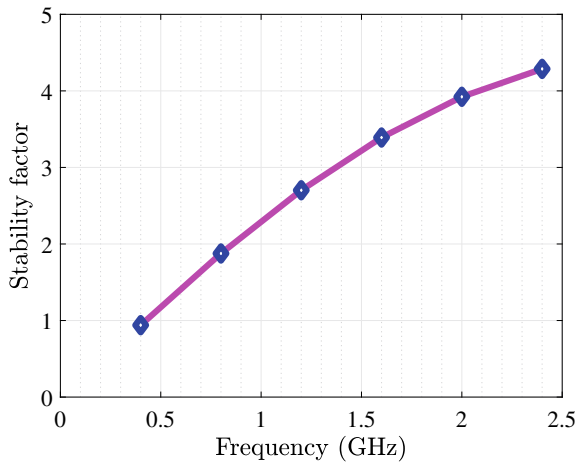
method is performed with the combination of ADS and MATLAB which provides a fully automated platform. For validating the proposed method, one PA in the frequency band $1 \div 2.3$ GHz is designed and optimized. The simulation results illustrate that in the operation frequency band, the achieved power gain is flat and higher than 10 dB and also the efficiency can reach maximum 61.5%.

References

1. N. Poluri, M.M. DeSouza, Designing a broadband amplifier without loadpull. IEEE Microw. Compon. Lett. pp. 1–1 (2021)
2. N.L.K. Nguyen, B.T. Nguyen, T. Omori, D.P. Nguyen, R. Moroney, S. D'Agostino, W. Kennan, A.V. Pham, A wideband sige power amplifier using modified triple stacked-hbt cell. IEEE Microw. Wireless Compon. Lett. **31**(1), 52–55 (2021)
3. H. Wang, P.M. Asbeck, C. Fager, Millimeter-wave power amplifier integrated circuits for high dynamic range signals. IEEE J. Microw. 299–316 (2021)
4. K. Settaluri, A. Haj-Ali, Q. Huang, K. Hakhamaneshi, B. Nikolic, *Autockt: Deep reinforcement learning of analog circuit designs, Design, Automation Test in Europe Conference Exhibition (DATE)* (Grenoble, France, IEEE, 2020), pp. 490–495
5. J. Cai, C. Yu, L. Sun, S. Chen, J.B. King, Dynamic behavioral modeling of RF power amplifier based on time-delay support vector regression. IEEE Transact. Microw. Theo. Tech. **67**(2), 533–543 (2019)
6. X. Yu, X. Hu, Z. Liu, C. Wang, W. Wang, F.M. Ghannouchi, A method to select optimal deep neural network model for power amplifiers. IEEE Microw. Wireless Compon. Lett. **31**(2), 145–148 (2021)
7. L. Kouhalvandi, O. Ceylan, S. Ozoguz, Automated RF power amplifier optimization and design: from lumped elements to distributed elements, In *27th Telecommunications Forum (TELFOR)* (IEEE, Belgrade, Serbia, 2019), pp. 1–4
8. L. Kouhalvandi, O. Ceylan, H.B. Yagci, Power amplifier design optimization with simultaneous cooperation of EDA tool and numeric analyzer, in *18th Mediterranean Microwave Symposium (MMS)* (IEEE, Istanbul, Turkey, 2018), pp. 202–205
9. L. Kouhalvandi, O. Ceylan, S. Ozoguz, Automated matching network modeling and optimization for power amplifier designs, in *11th International Conference on Electrical and Electronics Engineering (ELECO)*, (IEEE, Bursa, Turkey, 2019), pp. 510–513
10. Multi-objective particle swarm optimization (MOPSO). http://www.mathworks.com. Last accessed 09 Dec 2020
11. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (MIT Press, 2016)

Chapter 4
Investigation of Super-Resolution Methods for Processing a Sequence of Aerial Photographs

Alexander Y. Ivanov ⓘ **, Vitalii A. Pavlov** ⓘ **, and Sergey P. Nuzhny** ⓘ

Abstract Today, one of the main problems in the field of small object detection is the low resolution of the image because in most cases an object can only occupy a small area of the frame (areas of 15×15 pixels or 30×30). Until recently, the main way to improve the quality of images was to upgrade the technical characteristics of the equipment installed on board the aircraft. However, the limitations created by the element base increase the cost of such equipment. With a dramatic leap in the performance of computing resources, it has become possible to apply digital image processing algorithms. The challenges posed by the constraints imposed on the installed hardware on board the aircraft are being overcome by machine learning techniques. With the advent of the possibility of training neural networks, variational methods and classical interpolation methods have faded into the background and are practically not used due to their shortcomings. Algorithms that are used to improve the visual quality of an image are called super-resolution methods. The task of the methods is to search in a space of higher dimension for such an instance, which, upon de-citation, is transformed into an original image of low quality. This article compares some super-resolution methods in terms of quality / speed. Examples of the best algorithms among the considered ones are given. Based on the analysis of the results obtained, a conclusion is made about the advisability of using the investigated methods aboard. Conclusions are given on the advisability of using neural network approaches, at best based on compensation between personnel movements. The main advantages and disadvantages of existing approaches are noted.

Keywords Super-resolution · Variational method · Neural network methods

A. Y. Ivanov (✉) · V. A. Pavlov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: ivanov.ayu@edu.spbstu.ru

V. A. Pavlov
e-mail: pavlov_va@spbstu.ru

S. P. Nuzhny
Special Technological Center Ltd, Saint-Petersburg, Russia
e-mail: snuzhny@stc-spb.ru

4.1 Introduction

Super-resolution is a relatively new direction that began to actively develop with the advent of convolutional neural networks. The use of super-resolution technology makes it possible to analyze and process [1–8] images with improved quality in computer vision systems. Analysis of the literature [9–12] showed that researchers are mainly guided by the qualitative indicators of the developed method, in comparison with other approaches that have proven themselves well on the same data set. Whereas the assessment is performed on a small number of datasets. In addition, since in most cases the dataset contains high quality images, it is difficult to say how well a particular algorithm will perform on a sequence of aerial photographs. To date, there are no studies aimed at a comprehensive study of the behavior of various super-resolution methods in relation to video sequences received from an aircraft. The purpose of this work is to study super-resolution methods for processing a sequence of aerial photographs.

4.2 Existing Approaches

4.2.1 Classical Methods

Classical methods [13, 14] (interpolation) super-resolution are simple to implement and do not require long processing [15]. However, processing time and simplicity are the only advantages of this approach. The main problem of these methods is to introduce negative effects into the image, in addition to those available after the formation process:

(1) Distortions associated with image sampling are called aliasing or the effect of superimposing high frequencies on low frequencies. If the sampling is incorrect, the high-resolution image will have noticeable artifacts - the aliasing effect [16]. This effect is especially characteristic of images obtained by the nearest neighbor method.

(2) The second negative artifact is the Gibbs effect, which consists in the appearance of false lines near the object, repeating its contour, arises near contrasting differences in the image. This effect manifests itself with bicubic interpolation and when using Lanczos kernels [17].

(3) The third negative effect that interpolation brings (typical for interpolation with a Gaussian kernel) is blurring. However, in addition to the negative effect, Gaussian interpolation can reduce the noise level [18]. The variance parameter sets the amount of blur.

4.2.2   The Variational Approach

The variational super-resolution method consists in finding a solution to an incorrectly posed problem by introducing a regularizing stabilizer Bilateral Total Variation [19] to minimize the L_1 or L_2 norm between the original and super-resolved images. This approach is iterative, and after each pass of the cycle, the output image is refined by calculating a certain stabilizing factor [20]. One of the fundamental tasks in this method is the calculation of the operator for the movement of points, since the accuracy of the super-resolved image directly depends on the deformation operation.

4.2.3   Neural Networks Approaches

Neural network methods [21] are by far the most popular ways to improve resolution. If in classical approaches the result of super-resolution was only forecasting, then in the case of deep learning methods the result is a complex mapping. The greater the depth of the neural network, the more complex the mapping, and the more likely it is to improve the accuracy of reconstruction [22]. The methods can be divided into two groups according to the way the problem is solved:

(1) The first group - methods [23] based on models of convolutional neural networks, the purpose of which is to find a function of mapping one space to another of a higher order. The main problem of this approach is the smoothness of the results obtained, since the loss function is often initialized with a per-pixel function. Minimizing the error allows you to find the average probability solution for each of the pixels. However, the results of this approach often have a poor quality of perception [24].

(2) The second group - methods based on generative adversarial networks. The idea of this approach is to introduce adversarial learning and obtain images close to a real set (photorealistic frames) [25]. The network consists of two competing subnets - a generator and a discriminator. The generator's task is to apply such a transformation to the input image so that the discriminator cannot distinguish it from a real high-quality image. The discriminator, in turn, learns to distinguish forgery.

The following approaches to spatial information recovery are possible: single image super-resolution (SISR); Ultra-high-resolution imagery (MISR). The idea of the SISR method [23, 25, 39, 40, 42] is to use information inherent in a frame independently of other frames in the video sequence. However, in this case, the neural network learns to predict the missing information that the image contains. Since the information that was removed during compression disappears without a trace. The second approach is MISR [19, 26, 41], the idea of which is to search for an optical stream (however, there are other approaches to implementing the search for interframe displacement [27–29]) between a pair of frames (or more) of a video sequence.

Thus, the redundant information contained in the sequence of low-resolution frames is introduced by sub-pixel shifts between them. This approach is distinguished by the plausibility of the restoration results, since the algorithm has the ability to find information in a sample, and not in a single image. However, the disadvantage is the significant processing time of the sequence, especially when the optical flow is calculated with deep dependencies.

4.3 Research

4.3.1 Research Setup

To carry out the research, a universal software was implemented for the experimental evaluation of various super-resolution methods. The software is implemented in C++ with support for the OpenCV [30] and TensorFlow [31] libraries. The program downsamples the original dataset. In the future, the task is reduced to super-resolution, comparing the original dataset with the super-resolved one and estimating the execution time. For comparison, the following similarity measurement methods were selected: MSE, PSNR, SSIM [32], GMSD [33], BRISQUE [34].

The dataset [35], on which the testing took place, consists of 13 video sequences (each video contains from 1000 frames to 7000), received from the aircraft cameras, the duration of the video is on average one and a half minutes. Features of the dataset: compression artifacts, blur due to camera shake, resizing of objects. Mostly the dataset [35] consists of image sequences with a resolution of 640×480 pixels. Also, in the set there are many scenes with the presence of small objects. Calculations were performed on a platform with the following characteristics: Core i7, 3.7 GHz, RAM 32 Gb, GPU RTX 2080 TI.

4.3.2 Results

Table 4.1 shows the results of processing video sequences, averaged over the entire dataset.

The analysis of the results obtained allowed us to conclude that neural network approaches are suitable for solving the problem. However, the metrics MSE, PSNR have a weak correlation with human perception. The resulting super-resolution images are blurry because the dataset was initially affected by negative effects. It is advisable to analyze the results obtained, focusing on the results of the GMSD and BRISQUE metrics. Low-weight networks SRCNN (9–1-5), FSRCNN, ESPCN, LapSRN have shown themselves well, if you look for a compromise between quality, speed and computing resources, this will be the best choice. SRCNN (9–5-5), SRCNN (9–3-5) and EDSR also performed well, but the problem is the low processing speed.

Table 4.1 The results of processing video sequences, averaged over the entire dataset

Method	MSE	PSNR, dB	SSIM	GMSD	BRISQUE	Time, sec
AREA [14]	13.18	37.55	0.92	0.0256	36.94	0.0003
CUBIC [14]	8.50	39.54	0.94	0.0249	50.36	0.0005
LANCZOS4 [14]	8.38	39.59	0.94	0.0249	49.26	0.0019
LINEAR [14]	9.18	39.15	0.94	0.0264	54.91	0.0003
NEAREST [14]	13.18	37.55	0.92	0.0256	36.94	0.0002
PYRLK [36]	1268.13	18.35	0.43	0.1822	23.09	0.6202
BROX [37]	1231.54	18.54	0.46	0.1839	85.43	0.6537
FARNEBACK	1195.06	18.67	0.44	0.1815	40.72	0.5474
TVL1 [38]	1234.69	18.53	0.45	0.1821	53.54	1.839
LapSRN [39]	8.32	39.57	0.94	0.0246	49.01	0.4789
FSRCNN [40]	8.36	39.54	0.94	0.0244	44.80	0.1001
ESPCN [41]	8.52	39.42	0.94	0.0244	48.96	0.0397
EDSR [42]	7.96	39.76	0.94	0.0243	49.59	24.712
SRCNN (9–1-5)	8.19	39.64	0.94	0.0241	49.29	0.2669
SRCNN (9–3-5)	8.22	39.62	0.94	0.0241	49.63	0.8847
SRCNN (9–5-5)	8.23	39.62	0.94	0.0241	49.57	1.3203

The image obtained by the ESPCN method (Fig. 4.1) has become less contrasting; its shape can distinguish the vehicle. However, there are difficulties with the definition of the class of the object. The image obtained by the FSRCNN method (Fig. 4.2) is also less contrasting, as in the previous case, because of this, the shape of the car's surface becomes poorly distinguishable, the silhouettes of small-sized parts (headlights) disappeared.

The EDSR method (Fig. 4.3) retained all contrasting drops and contours, however, the image turned out to be less saturated, visually the surface texture became blurred.

The image reconstructed by SRCNN (9-5-5) (Fig. 4.4.) has a strong blurring of the contour of cars, the red car becomes poorly distinguishable. The result of the SRCNN method (9-1-5) (Fig. 4.5.) is similar to the original, but the shadows and brightness of the image are lost, the contours become less clear.

4.4 Conclusion

For the problem of super-resolution and specifically for the restoration of aerial photographs, the classical methods show low efficiency, since there is no information content of the introduced data. Also, the appearance of negative effects on the image, as a result of processing, such as: aliasing, blurring, false contouring.

Stabilizer-based method Bilateral full parameterization provides high performance with optimal parameter selection. The advantage is the fact that the algorithm

(a)   (b)

(c)   (d)

Fig. 4.1 On the left are original images (**a**, **c**), on the right are images (**b**, **d**) recovered using the ESPCN method
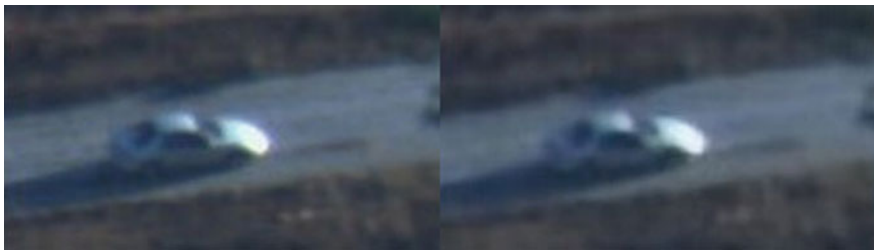
Fig. 4.2 On the left is the original image, on the right is the image reconstructed by the FSRCNN method

works with a selection from a video sequence, and not specifically with an image, as a result of the re-construction is more plausible. Also, the disadvantages are: the use of interpolation at the heart of the algorithm, as a result of which negative effects are introduced (blur, false contouring); the accuracy of the reconstruction depends on the accuracy of the calculation of the optical flow.

Neural network methods show the best results in terms of quality assessment metrics, however, the more complex the network, the longer processing time is required. Based on the results of the experiments, we can conclude that the best
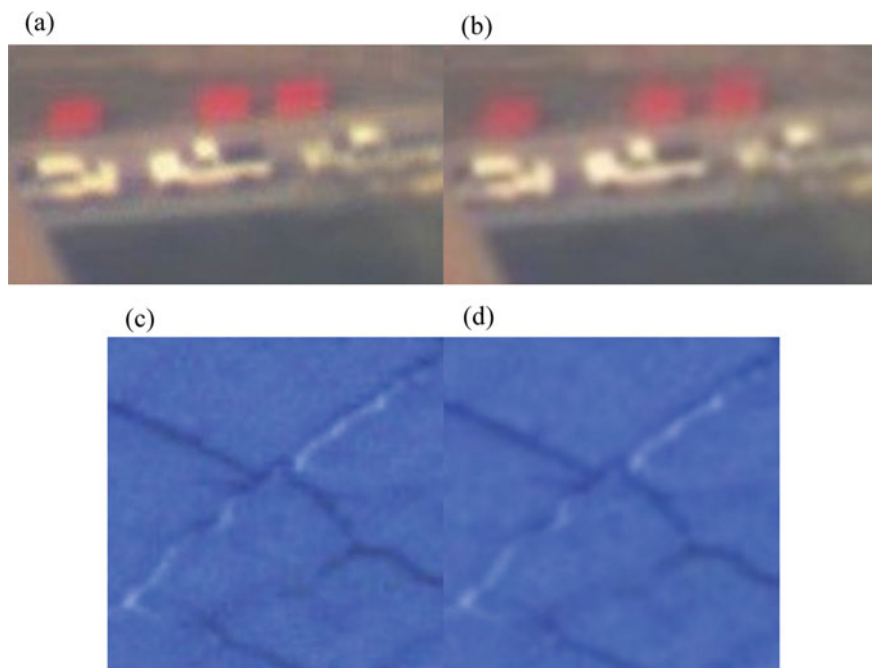
Fig. 4.3 On the left are original images (**a**, **c**), on the right are images (**b**, **d**) reconstructed by the EDSR method
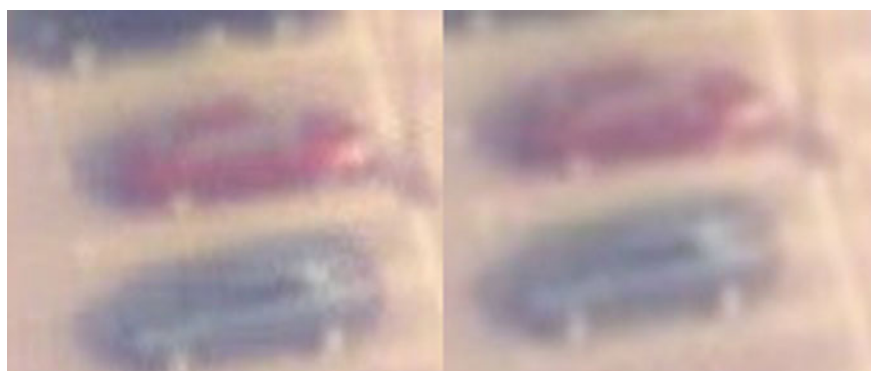
Fig. 4.4 On the left is the original image, on the right is the image reconstructed by the SRCNN method (9-5-5)

options (in terms of quality/speed) are: SRCNN (9-1-5), FRSCNN, ESPCN. If we compare the visual quality of restoration, then preference can be given to SRCNN (9-5-5) and EDSR. These approaches do a good job of enhancing contrast and detail.

Fig. 4.5 On the left is the original image, on the right is the image reconstructed by the SRCNN method (9-1-5)

When applying the filter to the image, the influence of objects located outside the field of view of the lens is not taken into account, the brightness at the edges of the distorted image may be unstable. This circumstance leads to the appearance of edge effects (false halos of high brightness) on the reconstructed image, which is typical for all super-resolution methods and is a disadvantage.

References

1. V.A. Pavlov, M.A. Galeeva, Detection and recognition of objects on aerial photographs using convolutional neural networks. J. Phys: Conf. Ser. **1326**, 1–6 (2019)
2. A.I. Bobrovsky, M.A. Galeeva, A.V. Morozov, V.A. Pavlov, A.K. Tsytsulin, Automatic detection of objects on star sky images by using the convolutional neural network. J. Phys. Conf. Ser. **1236**, 1–7 (2019)
3. M. Kots, V. Chukanov, U-Net adaptation for multiple instance learning. J. Phys. Conf. Ser. **1236**, 1–6 (2019)
4. V.A. Pavlov, A.A. Belov, A.A. Tuzova, Implementation of synthetic aperture radar processing algorithms on the Jetson TX1 platform. In: Proceedings of the 2019 IEEE International Conference on Electrical Engineering and Photonics, EExPolytech (2019), pp. 90–93. Peter the Great St. Peterburg Polytechnic University (SPbPU), Peterburg, Russia (2019)
5. N. Bakir, V. Pavlov, S. Zavjalov, S. Volvenko, A. Gumenyuk, M. Rethmeier, Development of a novel optical measurement technique to investigate the hot cracking suscep-tibility during laser beam welding. Welding in the World **63**, 435–441 (2019)
6. A.K. Tsytsulin, A.I. Bobrovsky, A.V. Morozov, V.A. Pavlov, M.A. Galeeva, Using convolutional neural networks to automatically select small artificial space objects on optical images of a starry sky. J. Opt. Technol. **86**(10), 627–633 (2019)
7. M.A. Baranov, E.N. Velichko, A.A. Andryakov, Image Processing for Analysis of Bio-Liquid Films. Optical Memory and Neural Networks **29**(1), 1–6 (2016)
8. O.B. Kuznetsova, E.A. Savchenko, A.A. Andryakov, E.Y. Savchenko, Z.A. Musakulova, Image processing in total internal reflection fluorescence microscopy. J. Phys: Conf. Ser. **1236**(1), 1–6 (2019)

9. A. Chadha, J. Britto, M. Roja, ISeeBetter: Spatio-temporal video super-resolution using recurrent generative back-projection networks. Springer Journal of Computational Visual Media **6**(3), 1–11 (2020)
10. Z. Wang, J. Chen, S.C. Hoi, Deep learning for image super-resolution: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1–24 (2020)
11. M.S. Sajjadi, R. Vemulapalli, M. Brown, Frame-recurrent video super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (IEEE, Salt Lake City, UT, USA, 2018), pp. 6626–6634
12. S. López-Tapia, A. Lucas, R. Molina, A.K. Katsaggelos, A Single Video Super-Resolution GAN for Multiple Downsampling Operators based on Pseudo-Inverse Image Formation Models. Digital Signal Processing **104**, 1–12 (2020)
13. S.C. Park, M.K. Park, M.G. Kang, Super-resolution image reconstruction: a technical overview. IEEE Signal Process. Mag. **20**, 21–36 (2003)
14. W. Burger, M. Burge, *Principles of Digital Image Processing* (Springer, New York, 2009)
15. A.S. Glassner, Graphics Gems I (Academic Press, Massachusetts, USA 1990)
16. R. Gonzalez, R. Woods, *Digital Image Processing* (Pearson, New York, 2018)
17. C. Duchon, Lanczos Filtering in One and Two Dimension. J. Appl. Meteorol. **18**(8), 1016–1022 (1979)
18. A.A. Andryakov, Image filtering for the nanosatellite vision system. J. Phys: Conf. Ser. **1326**(1), 1–7 (2019)
19. S. Farsiu, D. Robinson, M. Elad, P. Milanfar, Fast and Robust Multi-Frame SuperResolution. IEEE Trans. Image Process. **13**(10), 1327–1344 (2004)
20. A.G. Korobeynikov, A.Y. Grishentsev, E.N. Velichko, C.C. Korikov, S.A. Aleksanin, M.E. Fedosovskii, I.B. Bondarenko, Calculation of regularization parameter in the problem of blur removal in digital image. Optical Memory and Neural Networks.**25**(3), 184–191 (2016)
21. Y. Jianchao, H. Thomas, *Image Super-Resolution: Historical Overview and Future Challenges* (CRC Press, Super-Resolution Imaging, 2017), pp. 1–33
22. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (The MIT Press, Massachusetts, USA, 2016)
23. C. Dong, C.C. Loy, K. He, X. Tang, Learning a deep convolutional network for image superresolution. In: Fleet D., Pajdla, T., Schiele B., Tuytelaars T. (eds.) Computer Vision - ECCV 2014. Lecture Notes in Computer Science, LNCS, vol. 8692 (Springer, Cham, 2014), pp. 184–199
24. J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution. In: Leibe B., Matas J., Sebe N., Welling M. (eds.) Computer Vision - ECCV 2016. Lecture Notes in Computer Science, LNCS, vol. 9906 (Springer, Cham, 2016), pp.694–711
25. C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition 2017, (IEEE, Honolulu, Hawaii, USA, 2017), pp. 4681–4690
26. Multi-frame methods Homepage, https://paperswithcode.com/task/multi-frame-super-resolution, last accessed 2021/02/10
27. H. Wang, D. Su, C. Liu, L. Jin, X. Sun, X. Peng, Deformable Non-local Network for Video Super-Resolution. IEEE Access **7**, 177734–177744 (2019)
28. T. Yapeng, Z. Yulun, F. Yun, X. Chenliang: Tdan, Temporally deformable alignment network for video super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), (IEEE, Seattle, WA, USA, 2020), pp. 3360–3369
29. J. Caballero, C. Ledig, A. Aitken, A. Acosta, J. Totz, Z. Wang, W. Shi, Real-time video super-resolution with spatio-temporal networks and motion compensation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (IEEE, Honolulu, HI, USA, 2017), pp. 4778–4787
30. OpenCV Homepage, https://opencv.org last accessed 2021/02/10
31. Tensorflow Homepage, https://www.tensorflow.org last accessed 2021/02/10
32. Z. Wang, E.P. Simoncelli, A.C. Bovik, Multiscale structural similarity for image quality assessment. In: 2003 The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, (IEEE, Pacific Grove, CA, USA, 2003), pp. 1398–1402

33. W. Xue, L. Zhang, X. Mou, A.C. Bovik, Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index. IEEE Trans. Image Process. **23**(14), 684–695 (2014)
34. A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. **21**(12), 4695–4708 (2012)
35. Vivid Dapra Homepage, https://vision.cse.psu.edu/data/vividEval/datasets/datasets.html, last accessed 2021/02/10
36. J.-Y. Bouguet, Pyramidal implementation of the Lucas Kanade feature tracker description of the algorithm. Intel Corporation, Microprocessor Research Labs, OpenCVDocuments, 1–10 (1999)
37. T. Brox, A. Bruhn, N. Papenberg, J. Weickert: High accuracy optical flow estimation based on a theory for warping. In: T. Pajdla and J. Matas (eds.) 8th European Conference on Computer Vision, LNCS, vol. 4 (Springer, Prague, Czech Republic, 2004), pp. 25–36
38. C. Zach, T. Pock, H. Bischof, A duality based approach for realtime TV-L1 optical flow. In: Hamprecht F.A., Schnorr C., Jahne B. (eds.) Pattern Recognition. DAGM 2007. Lecture Notes in Computer Science, LNCS, vol. 4713 (Springer, Berlin, Heidelberg, 2007), pp. 214–223
39. L. Wei-Sheng, H. Jia-Bin, A. Narendra, Y. Ming-Hsuan, Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (IEEE Honolulu, HI, USA, 2017), pp. 624–632
40. C. Dong, C. Change Loy, X. Tang, Accelerating the super-resolution convolutional neural network. In: Leibe B., Matas J., Sebe N., Welling M. (eds.) Computer Vision - ECCV 2016. Lecture Notes in Computer Science, LNCS, vol. 9906 (Springer, Cham 2016), pp. 391–407
41. W. Shi, J. Caballero, F. Huszar, J. Totz, A.P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (IEEE, Las Vegas, NV, USA, 2016), pp. 1874–1883
42. L. Bee, S. Sanghyun, K. Heewon, N. Seungjun, M.L. Kyoung, Enhanced deep residual networks for single image super-resolution. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), (IEEE, Honolulu, HI, USA, 2017), pp. 1132–1140

Chapter 5
On the Classification of OFDM Signals Modulation Types Using Neural Networks

Igor N. Gorbunov , **Vitalii A. Pavlov** , **Sergey V. Zavjalov** , **and Dong Ge**

Abstract Automatic classification of modulation types is an important step in the development of a cognitive transceiver. In this paper, an algorithm was proposed for the automatic classification of six types of OFDM signal modulations: BPSK, QPSK, 16-, 64-, 256-, 1024-QAM. The method is based on the extraction of features, for which the statistics of the deviation of the instantaneous value of the amplitude, cumulants and adaptive wavelet entropy were used. In order to identify the algorithm that gives the optimal result in terms of the classification quality and the work speed, various combinations of used features were investigated. For each combination of features, simulation modeling was carried out, and characteristic curves of the classification accuracy were obtained.

Keywords Neural networks · OFDM · Classification

5.1 Introduction

The active development of «Beyond 5G» and «Pre-6G» systems leads to the research of various signal processing technologies. Variants of such development are the ability to automatically classify the applied modulation schemes [1] and machine learning methods. This is used in the construction of intelligent communication systems, such as cognitive and adaptive radio receivers, as well as radio frequency monitoring systems. Methods for automatic classification of modulation types can be divided into two groups.

The first group includes methods based on calculating the likelihood function (likelihood based). To make a decision on the type of modulation, a likelihood ratio is calculated and compared with a threshold value. These methods make it possible

I. N. Gorbunov · V. A. Pavlov (✉) · S. V. Zavjalov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: pavlov_va@spbstu.ru

D. Ge
Tsinghua University, Haidian District, Beijing, China

to obtain an optimal solution [2] about the type of modulation, but they have high computational complexity [3, 4]. This fact makes it difficult to implement, especially on the SDR platform. An additional disadvantage of this method is the requirement for a priori information about the signal parameters, which is undesirable for an adaptive receiver. The second group includes methods based on the extraction of features (feature based). Such algorithms, as a rule, consist of two stages: extraction of characteristic features from the received signal and their classification. This approach shows good productivity with low computational complexity [5–11]. An additional advantage is that there is no requirement for a priori information about signal parameters and channel statistics.

Analysis of the literature has shown that in recent decades, various algorithms for classifying modulations have been developed for single-frequency information transmission systems [2, 5–7]. However, there is not much research on the development of classification algorithms for multifrequency signals [12]. Multifrequency signals, in particular OFDM signals, are widely used in broadband and local communication systems Wi-Fi, WiMax, LTE, in digital television [13–19].

In this paper, we will focus on the methods for the classification of multifrequency signals based on feature extraction.

A method for classifying QAM modulations based on statistical features was proposed in [8]. The authors used such OFDM signal features as mean value, variance, skewness and kurtosis. This algorithm was tested on a multipath channel model, but it was not robust to time and frequency synchronization errors. In [9], for the classification of QAM modulations, a method was proposed that uses the amplitude moments and correlation between the subcarriers of the OFDM signal. The authors have shown that their method remains reliable in the presence of a frequency shift and its efficiency improves with an increase in the order of the amplitude moments. An example of using a normalized fourth-order cumulant for the classification of BPSK, QPSK, OQPSK, MSK and 16-QAM modulations is considered in [10, 11]. The method proposed by the authors is resistant to shifts in time, frequency and phase and does not require a priori information about the signal parameters and channel statistics.

The sets of basic types of modulations considered in these works are limited and do not exceed six varieties. Also, the use of wavelet entropy as a feature for OFDM systems remains unexplored. Meanwhile, this feature has shown good results in single-frequency systems [20]. The purpose of this work is to develop a method for classifying the types of modulations BPSK, QPSK, 16-, 64-, 256-, 1024-QAM in OFDM systems.

5.2   Description of Features for the Classification of Modulation Types

A review of the literature has shown the promise of using features based on adaptive wavelet entropy [20], statistics of the deviation of instantaneous amplitude values [1] and cumulants of various orders [1]. Let's take a closer look at these features.

5.2.1   Adaptive Wavelet Entropy

The features are based on adaptive wavelet entropy is charged in accordance with the formula [20]:

$$E(s) = \frac{1}{N} \sum_{i=1}^{N} |s_i|^p,$$

(5.1)

where p is the power that is selected in the range $1 \le p < 2$, s_i are the approximation or detail coefficients of discrete wavelet transform of OFDM symbol, N is the size of OFDM symbol.

5.2.2   Statistical Features

In the following statistical features were used:

(1)   The maximum spectral power density of the normalized and centered value of the instantaneous amplitude:

$$\gamma_{\max} = \frac{\max|DFT(A_{cn})|^2}{N}; \quad A_{cn} = \frac{A}{\mu_A} - 1,$$

(5.2)

where A_{cn} is the normalized and centered value of the instantaneous amplitude of OFDM symbol, μ_A is the expected value of the instantaneous amplitude of the OFDM symbol, N is the size of OFDM symbol.

(2)   The logarithm of the variance of the absolute value of the normalized and centered instantaneous signal amplitude:

$$L\Delta_{aa}^2 = \lg\left(\frac{1}{N} \sum_{n=1}^{N} A_{cn}^2[n] - \left(\frac{1}{N} \sum_{n=1}^{N} |A_{cn}[n]| \right)^2 \right).$$

(5.3)

(3)   The standard deviation and the absolute value of the normalized instantaneous amplitude of the signal centered:

$$\sigma_{aa} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} A_{cn}^2[n] - \left(\frac{1}{N}\sum_{n=1}^{N}\left|A_{cn}[n]\right|\right)^2}. \tag{5.4}$$

5.2.3 Cumulants

In [21] fourth-order cumulants of a complex signal were proposed as features for the classification of PAM, PSK, and QAM modulations:

$$\hat{C}_{20} = \frac{1}{N}\sum_{n=1}^{N} r^2[n], \ \ \hat{C}_{21} = \frac{1}{N}\sum_{n=1}^{N}|r[n]|^2,$$

$$\hat{C}_{40} = \frac{1}{N}\sum_{n=1}^{N} r^4[n] - 3\hat{C}_{20}; \quad \hat{C}_{41} = \frac{1}{N}\sum_{n=1}^{N} r^3[n]*r[n] - 3\hat{C}_{20}\hat{C}_{21};$$

$$\hat{C}_{42} = \frac{1}{N}\sum_{n=1}^{N}|r[n]|^4 - \left|\hat{C}_{20}\right|^2 - 2\hat{C}_{21}^2. \tag{5.5}$$

5.3 Description of the Simulation Model

Consider the principle of the simulation model. On the receiving side, after the fast Fourier transform, features are extracted, which are then transmitted to the input of the classifier. As a classifier were tested varieties fully connected neural networks of direct distribution, consisting of the three hidden layers. The number of neurons in the input layer was chosen based on the number of input features used for training. The training was carried out according to the Levenberg–Marquardt method [22]. This method has a high convergence rate and allows you to achieve the smallest classification error. The hyperbolic tangent was used as the activation function for the hidden layers of the neural network. The output layer has a linear activation function applied. The root-mean-square error (MSE) method was used as the error measurement function. The sizes of the training sample ranged from 3200 to 32,000 bits, depending on the type of modulation.

The training and testing of the network was carried out in accordance with the scheme shown in Fig. 5.1. Testing of the trained neural network was carried out on an information sample of at least 5×10^6 bits. Each OFDM symbol was fed to the input of the classifier.
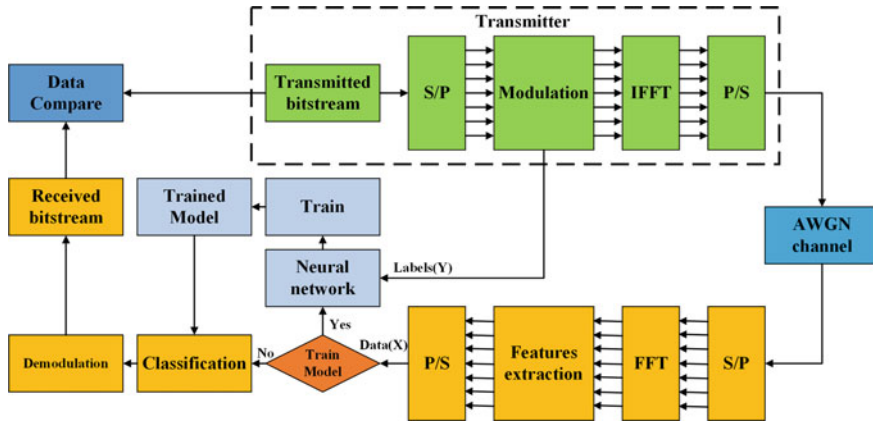
Fig. 5.1 Block diagram of the simulation model

5.4   Simulation Results

In the work, a series of experiments was carried out for each of the five neural networks trained on different sets of features extracted from the OFDM signal: (1) 14 dwt, (2) 14dwt + 3stat, (3) 14dwt + 3stat + 5cum, (4) 7dwt(A) + 5cum и (5) 7dwt(D) + 5cum, where.

- 14 dwt is 7 coefficients of approximation and 7 coefficients of detail of the discrete Wavelet transform. 7dwt (A) and 7dwt (D)—coefficients of approximation and detail separately;
- 3stat is 3 statistics of deviation of instantaneous amplitude values;
- 5cum is 5 cumulants of different orders.

At $E_b/N_0 = 5$ dB (E_b is the transmitted bit energy, N_0 is the spectral density of the average noise power) three of the five networks under study show a classification accuracy of more than 80% for all types of modulations (see Table 5.1).

With a value of $E_b/N_0 = 15$ dB (see Table 5.2), only classifier 1 shows in some cases an accuracy of less than 90%—for modulation types: QPSK, 16-QAM and

Table 5.1 Classification accuracy at 5 dB

	BPSK	QPSK	16-QAM	64-QAM	256-QAM	1024-QAM
14 dwt	0.6388	0.4947	0.9347	0.8801	0.9784	0.9855
14dwt, 3stat	0.7489	0.8966	0.9705	0.9802	0.9720	0.9755
14dwt, 3stat, 5cum	0.9951	0.9967	0.9670	0.9728	0.9792	0.9824
7dwt(A), 5cum	0.9995	0.9970	0.9914	0.9910	0.9949	0.9869
7dwt(D), 5cum	0.9992	0.9970	0.9908	0.9925	0.9891	0.9885

Table 5.2 Classification accuracy at 15 dB

	BPSK	QPSK	16-QAM	64-QAM	256-QAM	1024-QAM
14 dwt	0.9975	0.8100	0.8990	0.8616	0.9690	0.9820
14dwt, 3stat	0.9994	0.9597	0.9864	0.9800	0.9699	0.9766
14dwt, 3stat, 5cum	1	0.9993	0.9906	0.9716	0.9797	0.9879
7dwt(A), 5cum	1	0.9987	0.9836	0.9849	0.9961	0.9901
7dwt(D), 5cum	1	0.9988	0.9816	0.9923	0.9888	0.9927

64-QAM. In other cases, all networks demonstrate classification accuracy above 95%.

Analysis of the results shows that the use of cumulants gives a good increase in the classification accuracy, but an increase in the number of extracted features increases the complexity of the algorithm, and it is not advisable to use the full set of features. A classifier using cumulants with a truncated set of discrete wavelet transform coefficients shows the result not worse (the difference is less than 3%) compared to a classifier trained on a full set of features.

Comparing classifiers 4 and 5 (variants with separated coefficients of approximation and detail), we can conclude that the results of their work are almost the same (in some cases, the accuracy values differ by less than 1%). Starting from $E_b/N_0 = 5$ dB, the probability of making the right decision regarding the type of modulation is ~99%. The result of the work of the best of the studied classifiers for the range from −5 to 20 dB is shown in Fig. 5.2.
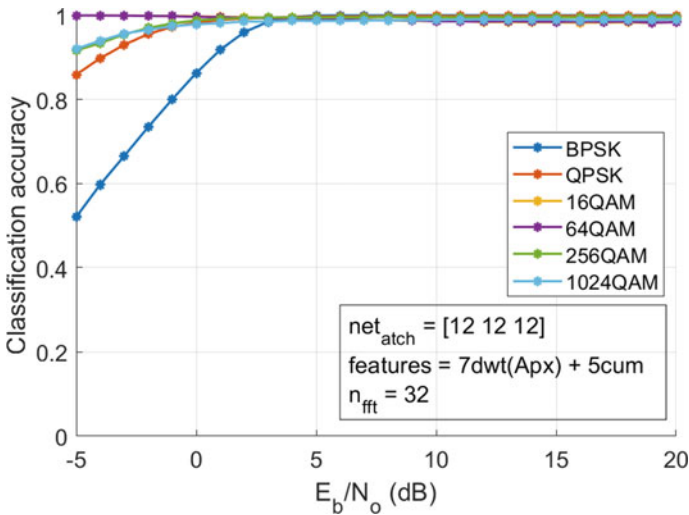
Fig. 5.2 Dependence of classification accuracy on SNR for the 4th set of features

Table 5.3 Number of different operations needed for each classifier (for one OFDM symbol)

	$+/-$	$\times$	$\div/\sqrt{\ }$	$\log()$
14 dwt	$35\,N$ - 14	$35\,N$	$14\,N+14$	0
14dwt, 3stat	$2N\log_2 N + 42\,N$ - 14	$N\log_2 N + 40\,N$	$19\,N+17$	1
14dwt, 3stat, 5cum	$2N\log_2 N + 49\,N$ -13	$N\log_2 N + 52\,N + 10$	$21\,N+23$	1
7dwt(A), 5cum	$24.5\,N$ - 6	$25.5\,N + 10$	$9\,N+13$	0
7dwt(D), 5cum	$24.5\,N$ - 6	$25.5\,N + 10$	$9\,N+13$	0

The probability of a false alarm when using this classifier, already starting from 5 dB, turned out to be less than 0.3% for each type of modulation from the studied set of signals.

Data on the computational complexity of the algorithms are given in the Table 5.3:

Here N is the size of the OFDM symbol.

5.5 Conclusions

In this work, a method was proposed for the automatic classification of OFDM signal modulations, based on the extraction of the following features: adaptive wavelet entropy (coefficients of approximation and DWT detail), statistics of the deviation of instantaneous amplitude values and cumulants of the second and fourth orders.

Analysis of the results of simulation modeling showed that the worst-case classification algorithm is performed by an algorithm trained on only one type of features. At the same time, the use of cumulants makes it possible to exclude the deviations of the instantaneous amplitude values and half of the used DWT coefficients from the set of statistics. The following combination of features turned out to be the best for the classification of BPSK, QPSK, 16-, 64-, 256-, 1024-QAM modulations: 7 approximation coefficients of the discrete wavelet transform, 5 cumulants of 2 and 4 orders. This method demonstrates, on average, an accuracy of 99% at a value of $E_b/N_0 = 15$ dB for the studied types of modulations. For comparison, the method proposed in [10] shows an average accuracy of 95% at $E_b/N_0 = 15$ dB.

References

1. Z. Zhu, A.K. Nandi, Automatic modulation classification: principles, algorithms and applications. Wiley, 2015, 184 p
2. W. Wei, J.M. Mendel, Maximum-likelihood classification for digital amplitude-phase modulations. IEEE Trans. Commun. **48**(2), 189–193 (2000)
3. L. Häring, Y. Chen, A. Czylwik, Automatic modulation classification methods for wireless OFDM systems in TDD mode. IEEE Trans. Commun. **58**(9), 2480–2485 (2010)
4. J. Zheng, Y. Lv, Likelihood-based automatic modulation classification in OFDM with index modulation. IEEE Trans. Veh. Technol. **67**(9), 8192–8204 (2018)
5. M. Oner, O.A. Dobre, On the second-order cyclic statistics of signals in the presence of receiver impairments. IEEE Trans. Commun. **59**(12), 3278–3284 (2011)
6. S. Majhi, R. Gupta, W. Xiang, Novel blind modulation classification of circular and linearly modulated signals using cyclic cumulants, in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC).* (IEEE, 2017), pp. 1–5
7. P. Daponte, G. Mercurio, S. Rapuano, A wavelet networks-based method for the digital telecommunication system monitoring. IEEE Trans. Instrum. Meas. **50**(6), 1773–1780 (2001)
8. A.D. Pambudi, S. Tjondronegoro, H. Wijanto, Statistical properties proposed for blind classification OFDM modulation scheme, in *IEEE International Conference on Aerospace Electronics and Remote Sensing Technology* (IEEE, 2014), pp. 89–93
9. D. Shimbo, I. Oka, A modulation classification using amplitude moments in OFDM systems, in *International Symposium On Information Theory & Its Applications,* (IEEE, 2010), pp. 288–293
10. R. Gupta, S. Kumar, S. Majhi, Blind modulation classification for OFDM in the presence of timing, frequency, and phase offsets, in *IEEE 90th Vehicular Technology Conference (VTC2019-Fall)* (IEEE, 2019), pp. 1–5
11. R. Gupta, S. Kumar, S. Majhi, Blind modulation classification for asynchronous OFDM systems over unknown signal parameters and channel statistics. IEEE Trans. Veh. Technol. **69**(5), 5281–5292 (2020)
12. J.G. Proakis, et al., *Communication Systems Engineering,* vol. 2 (Prentice Hall, New Jersey, 1994), 816p
13. A. Rashich, A. Kislitsyn, S. Gorbunov, Trellis demodulator for pulse shaped OFDM, in *IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)* (IEEE, 2018), pp. 1–5
14. V.O. Rashich, A.V. Rashich, Asymptotically optimal algorithm for OFDM-signal reception under AWGN and OFDM interference shifted in frequency, in *IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom).* (IEEE, 2015),
15. E.N. Smirnova, A.S. Ovsyannikova, S.V. Zavjalov, G. Dong, On features of implementation of SEFDM-transmitter with optimal shape of envelope. J. Phys. Conf. Ser. **1236**(1), 1–6 (2019), art. № 012067. https://doi.org/10.1088/1742-6596/1236/1/012067
16. E.N. Smirnova, Experimental research of receiver based on SEFDM-signals with optimal envelopes. J. Phys. Conf. Ser. **1326**(1), 1–5 (2019), art. № 012031. https://doi.org/10.1088/1742-6596/1326/1/012031
17. A. Kislitsyn, A. Krylov, A. Rashich, Experimental evaluation of SEFDM trellis based demodulator, in *IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*, St. Petersburg, Russia (2020), pp. 110–113. https://doi.org/10.1109/EExPolytech50912.2020.9243960
18. E. Vylegzhanin, A. Chilingarov, B. Khuc, D. Puzko, Y. Batov, A. Gelgor, Comparison of PAPR reduction techniques for OFDM transmission over underwater acoustic channel. Springer Proc. Phys. **255**, 665–674 (2021). https://doi.org/10.1007/978-3-030-58868-7_73
19. A. Gelgor, V.P. Nguyen, Outperforming conventional OFDM and SEFDM signals by means of using optimal spectral pulses and the M-BCJR algorithm, in *2019 26th International Conference on Telecommunications (ICT).* (IEEE, 2019), pp. 130–134

20. D. Avci, An intelligent system using adaptive wavelet entropy for automatic analog modulation identification. Digital Signal Process. **20**(4), 1196–1206 (2010)
21. A. Swami, B.M. Sadler, Hierarchical digital modulation classification using cumulants. IEEE Trans. Commun. **48**(3), 416–429 (2000)
22. M.T. Hagan, M.B. Menhaj, Training feedforward networks with the Marquardt algorithm. IEEE Trans. Neural Netw. **5**(6), 989–993 (1994)

Chapter 6
Use of the Method of Setting the Interval Target in the Problem of Synthesis of a Neural Network Classifier for Diagnosing Chronic Kidney Disease in Patients

Check for updates

Eugene Mirkin and Elena Savchenko

Abstract The paradigm of teaching ANN with an "interval teacher" is considered to solve the problems of medical diagnosis of chronic kidney disease. In this concept, the ANN implemented a mechanism for choosing a specific "teacher" from the deterministic interval set. A comparison of the learning results of the classical scheme and the proposed paradigm is given. The effectiveness of the proposed learning concept is shown. The accuracy of the classifier with the "interval teacher" is 84.6%.

Keywords Interval training set · Medical classifier · Training set tuning mechanism · Feedforward neural network · Interval teacher

6.1 Introduction

Creating machines or mechanisms endowed with intelligence is the idea of researchers from various fields of science and technology. Although the first artificial neural networks (ANN) were published more than 50 years ago, this item began to be deeply studied in the early 1990s and still possesses a huge research potential. Applications with the participation of systems considered intellectual cover various areas of knowledge, which is confirmed by the publications of researchers of numerous scientific journals. One of the directions where artificial neural networks are actively used—this is a medical diagnosis [1–6], as they have a powerful tool for analyzing, modeling and processing complex clinical data. At the same time, when developing intelligent medical diagnostic systems, researchers face typical problems of selecting the neural network topology, setting the coefficients of synaptic network connections, search for new methods and learning algorithms, leading to acceptable results of the operation of the ANN.

E. Mirkin (✉) · E. Savchenko
Computer Information Systems, International University of Kyrgyzstan, Bishkek, Kyrgyz Republic

The paradigm of teaching ANN with an "interval teacher" [7] was used to create a model for diagnosing chronic kidney disease in patients. Chronic kidney disease (CKD) is damage to the kidneys, or a decrease in their function for 3 months or more. The disease is classified into 5 stages, which differ in the tactics of patient management and the risk of developing end-stage renal failure and cardiovascular complications [5, 6, 8–10].

Chronic kidney disease is a slow and progressive loss of kidney function over several years, often resulting in irreversible renal failure.

Recognition of chronic kidney disease in patients will be the task of the neural network: determining whether a patient belongs to one of the classes 0- "no CKD disease", 1- "CKD present". To train the classifier, we used data from a retrospective sample of diagnostics of chronic kidney disease in Indians. This dataset will be used to predict chronic kidney disease and was obtained over a two-month period from Apollo Reach Hospital Karaikudi, India [9]. The structure of the database of 282 patients is presented in Table 6.1.

Here is a description of the fields characterizing the patient's condition at the time of diagnosis:

1. Age.
2. Blood pressure, diastolic pressure is a numerical indicator of the work of the heart, it is the pressure that the blood flowing in the blood vessels exerts on the wall of the blood vessel. Blood pressure readings are influenced by the strength of the heart muscle and the tone of the blood vessels [1].
3. Specific gravity of urine is a urinalysis parameter commonly used to assess kidney function and can aid in the diagnosis of various kidney diseases.
4. Albumin is a protein found in the blood. A healthy kidney does not allow albumin to pass from the blood to the urine. The damaged kidney passes some albumin in the urine. The less albumin in the urine, the better.
5. Sugar (glucose) indicates glucosuria, a condition when glucose presents in the urine due to high blood sugar or kidney damage. Glucosuria is a symptom of diabetes mellitus.
6. Accumulations of purulent cells or pyuria can be a symptom of a huge amount of diseases and commonly caused by an infection in any area of the urinary system, including the kidneys, ureters, urethra, or bladder.
7. Bacteria: detection of bacteria in the urine indicates a possible infection of the urinary system (cystitis, urethritis, pyelonephritis, etc.).
8. The volume of packed cells is a value equal to the height of the column of erythrocytes, directly measured by centrifuging blood in a test tube. It is the fastest and most affordable way to measure red blood cells in your blood.
9. Hypertension, arterial hypertension or elevated blood pressure: a serious condition, significantly improving the risk of developing diseases of the cardiovascular system, brain, kidney and other diseases.
10. Diabetes mellitus is a chronic disease, developing in cases where the pancreas does not produce enough insulin or when the body cannot effectively use produced insulin.

Table 6.1 Patient database structure

No.	Parameter	Unit of measurement	Database record 1	Database record 2	Database record 282
1	Age	age (number)	48	62	…	42
2	Blood presure	bp (number)	80	80		70
3	Urine specific gravity	sg (number) -(1.005,1.010,1.015,1.020,1.025)	1.020	1.010		1.020
4	Albumin	al (number) - (0,1,2,3,4,5)	1	2		0
5	Sugar (glucose)	su (number) - (0,1,2,3,4,5)	0	3		0
6	Accumulations of purulent cells	pcc notpresent = 0 present = 1	0	0		0
7	Bacteria	ba notpresent = 0 present = 1	0	0		0
8	The volume of packed cells	pcv (number)	44	31		43
9	Hypertension	htn yes = 1, no = 0	1	0		0
10	Diabetes mellitus	dm yes = 1, no = 0	1	1		0
11	Ischemic heart disease	cad yes = 1, no = 0	0	0		0
12	Appetite	appet good = 1, poor = 0	1	0		1
13	Edema	pe yes = 1, no = 0	0	0		0
14	Anemia	ane yes = 1, no = 0	0	1		0
15	Presence of disease	class ckd = 1 nock d = 0	1	1		0

11. Ischemic heart disease (IHD) or coronary disease: the heart pathology, which is based on the damage of myocardium due to its insufficient blood supply due to atherosclerosis and usually arising in its background with thrombosis or spasm of coronary heart arteries.
    It is one of the most common diseases and one of the main causes of mortality, as well as the temporary loss of working capacity of the population in developed countries of the world [10].
12. Appetite is a feeling associated with the need for food, as well as a physiological mechanism that regulates the intake of nutrients into the body.
13. Swelling of the foot and ankle is one of the main signs of congestive heart failure (HF) but can also result from other systemic or local conditions, including

chronic kidney disease, liver disease, thyroid disease, venous insufficiency, and venous thrombosis.
14. Anemia is a condition in which you do not have enough healthy red blood cells to deliver enough oxygen to your body tissues.
15. Presence of disease: 0 means "no CKD disease", 1 means "CKD presents"

6.2 Creating a Neural Network Model

To create a medical classifier for the diagnosis of chronic kidney disease in patients, we will use the topology of the ANN, with the "interval teacher" training paradigm, according to the method described in [7]:

1. Define the target (training) set in the form of a blurred set with clearly designated interval boundaries, defined by the final training result. Table 6.2 shows the training set $(x_1, x_2 \ldots x_{14})$, as well as classic $(\hat{y})$ and interval (y_{min}, y_{max}) target sets for the synthesis of a medical classifier for the diagnosis of chronic kidney disease in patients. As can be seen from the Table 6.2, the training selection consists of 282 entries.
2. We will produce determining the rules for changing the boundaries of the target (training) set, depending on the parameters of the learning process and the final learning outcome, according to the algorithm:

$$\hat{y} = \begin{cases} y_{min}, & if \quad y \leq y_{min} \\ y_{max}, & if \quad y \geq y_{max} \\ y, & if \ y_{min} < y < y_{max} \end{cases}, \tag{6.1}$$

where y_{max}, y_{min} diapason of "interval teacher"; y—neural network training result; $\hat{y}$—teacher for the next training epoch.

To confirm the effectiveness of the method proposed in [7], we will conduct a comparative analysis of the two models of the neural network medical classifier based on:

- a classic neural network that implements the paradigm of the training of an ANN classical (traditional) "teacher."
- neural network that implements the paradigm of ANN training with the interval teacher.

For two models a two-layer feed forward neural network was created with two neurons in the first layer and the activation function of a sigmoidal, and one output neuron with a linear activation function (Fig. 6.2). For both models Levenberg–Marquardt algorithm was chosen (Fig. 6.1).

The initial initialization of the weight and threshold coefficients of the neural network with the following random values, identical for the two models, was performed (Table 6.3):

Table 6.2 Data for the synthesis of a medical classifier to recognize chronic kidney disease in patients

Training set

№	Age	Blood pressure	Urine specific gravity	Albumin	Sugar (glucose)	Accumulations of purulent cells	Bacteria	The volume of packed cells	Hypertension	Diabetes mellitus	Ischemic heart disease	Appetite	Edema
1	48	80	1.020	1	0	0	0	44	1	1	0	1	0
2	7	50	1.020	4	0	0	0	38	0	0	0	1	0
3	62	80	1.010	2	3	0	0	31	0	1	0	0	0
280	42	80	1.020	0	0	1	0	44	0	0	0	1	0
281	35	80	1.020	0	0	0	0	45	0	0	0	1	0
282	30	80	1.020	0	0	0	0	45	0	0	0	1	0

Target set

Anemia	Classical teacher	Interval teacher $y_{min} \in [y_{min}\ y_{max}]$	Interval teacher $y_{max} \in [y_{min}\ y_{max}]$
0	1	0.51	1.49
0	1	0.51	1.49
1	1	0.51	1.49
0	0	−0.49	0.49
0	0	−0.49	0.49
0	0	−0.49	0.49

Fig. 6.1 The architecture of the created neural network
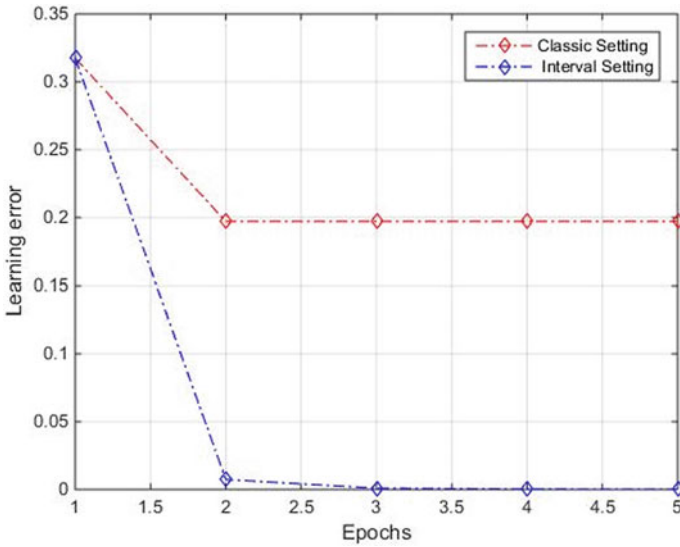
Fig. 6.2 The dynamics of changes in the error of a classical neural network and a neural network with setting interval targets

The initial data were divided as follows: the training sample was 80%, validation was 10% and test was 10%. The initial data were normalized in the traditional way in the range from zero to one:

$$x_i = \frac{x_i - x_{min}}{x_{max} - x_{min}},$$ (6.2)

where x_i is the current value of the variable, x_{min} and x_{max} are the minimum and maximum values of the variable respectively. The results of setting the models are shown in Figs. 6.3 and 6.4, respectively. As can be seen from the figures, the dynamics of error change and linear regression of goals with respect to the results of the proposed model of a neural network medical classifier with an ANN training paradigm with an "interval teacher" showed a high learning rate with a guaranteed

Table 6.3 The weight and threshold coefficients of the neural network

weight coefficients w_1 matrix (2*14)

First layer	−0.00958	−0.77168	10.72008	2,39,931	0.363531	−0.45932	2.269027
	0.304216	0.397394	6.122102	2,09,133	−1.22196	7.935835	7.060963
	−0.01585	5.84248	−1.22346	−11,593	3.061286	1.18897	8.21363
	−0.07522	3.821831	−699,956	6,12,473	−6.72447	3.346557	4.36549

weight coefficients w_2 matrix (1 × 2)

Second layer	0.62204 0.72268

Threshold values of neurons of the first layer

First layer	2.89931 6.398203

Threshold values of neurons of the output layer

Second layer	0.763798
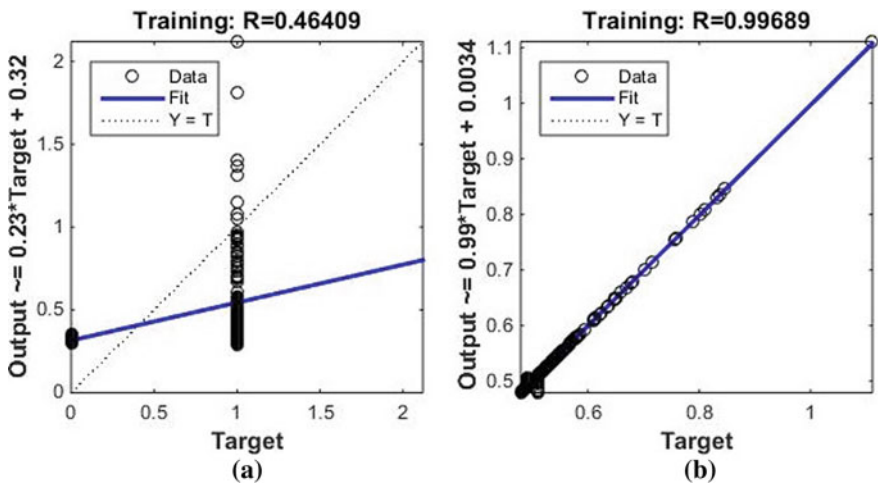
Fig. 6.3 Linear regression of targets for results: **a** classical neural network (coefficient of determination R = 0.46), **b** neural network with setting interval targets (coefficient of determination R = 0.99)

result in comparison with the traditional method of training a neural network, both in terms of the convergence rate of the process training, and the quality of recognition.

Let us show the efficiency of the operation of two models of a neural network medical classifier on a test set of 26 examples, on which the classifier was not trained. The classifier error is presented in the form of a Confusion Matrix (Fig. 6.4).

Fig. 6.4 Data classification accuracy for a classic teacher (**a**) and an interval teacher (**b**)
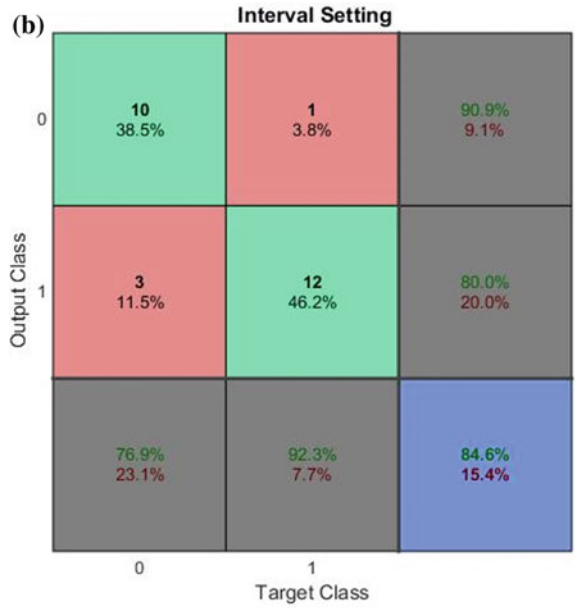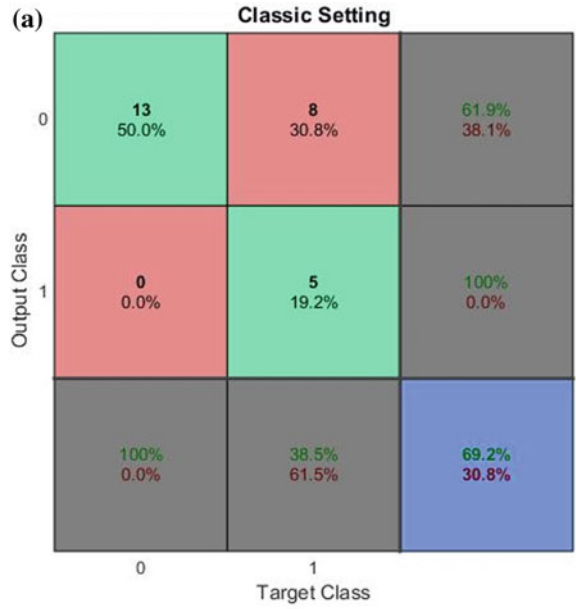
From Confusion Matrix (Fig. 6.4) we see that the accuracy of the classifier of the neural network with an interval teacher shows the best result. The classification accuracy was 69.2% for a classical teacher and 84.6% for an interval teacher.

6.3 Conclusion

The method of ANN synthesis proposed in [7] is implemented here in the task of creating a medical classifier for the diagnosis of chronic kidney disease in patients. A comparative analysis of two models of a neural network medical classifier based on a classical neural network, which implements the paradigm of learning an ANN of a classical (traditional) "teacher" and a paradigm of learning an ANN with an "interval teacher" showed the effectiveness of the latter, both in terms of the convergence rate of the learning process and the quality of recognition. The accuracy of the classifier with the "interval teacher" is 84.6%.

References

1. https://www.regionaalhaigla.ee/sites/default/files/documents/Vererohk_ja_korgvererohu tobi_ehk_hupertoonia_-_vene_keeles.pdf. Last accessed 30 Mar 2021
2. A.A. Aidaraliev et al., Neural network prediction of difficult tracheal intubation risk by using the patient's face image. HERALD of North-Western State Med. Univ. named after II Mechnikov **11**(3), 23–32 (2019)
3. E.K. AbuSharekh et al., Diagnosis of hepatitis virus using artificial neural network. (2018)
4. M.U. Emon, A.M. Imran, R. Islam, M.S. Keya, R. Zannat, Ohidujjaman, Performance analysis of chronic kidney disease through machine learning approaches, in *2021 6th International Conference on Inventive Computation Technologies (ICICT),* Coimbatore, India (2021), pp. 713–719
5. P. Chittora et al., prediction of chronic kidney disease—a machine learning perspective. IEEE Access **9**, 17312–17334 (2021)
6. N. Bhaskar, M. Suchetha, N.Y. Philip, Time series classification-based correlational neural network with bidirectional LSTM for automated detection of Kidney Disease. IEEE Sens. J. **21**(4), 4811–4818 (2021)
7. E. Mirkin, E. Savchenko, Configuring the Interval target in a multilayer feedforward neural network on the example of the problem of medical diagnostics, in *International Youth Conference on Electronics, Telecommunications and Information Technologies*. (Springer, Cham, 2021), pp. 463–475
8. E. Mirkin, E. Savchenko, E. Savchenko, Use of the method of configuring the interval target in the problem of synthesis of the neural network classifier for diagnosing cardiovascular diseases, in *2020 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*, (IEEE, 2020), pp. 140–144
9. http://archive.ics.uci.edu/ml//datasets/Chronic_Kidney_Disease. Last accessed 30 Mar 2021
10. Z. Musakulova, E. Mirkin, E. Savchenko, Synthesis of the backpropagation error algorithm for a multilayer neural network with nonlinear synaptic inputs, in *2018 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*. (IEEE, 2018), pp. 140–144

11. I. Kolokolnikov, I. Lavrenyuk, E. Savchenko, M.A. Baranov, E. Savchenko, Photometry setup for dynamic dye concentration measurement, in *International Youth Conference on Electronics, Telecommunications and Information Technologies.* (Springer, Cham, 2021), pp. 277–282
12. T. Nguyen et al., Medical data classification using interval type-2 fuzzy logic system and wavelets. Appl. Soft Comput. **30**, 812–822 (2015)
13. V. L. Wiggins, S.K. Engquist, L.T. Looper, Neural network applications: a literature review (1992)
14. O.B. Kuznetsova, E.A. Savchenko, A.A. Andryakov, E.Y. Savchenko, Z.A. Musakulova Image processing in total internal reflection fluorescence microscopy. J. Phys. Conf. Ser. IOP Publ. **1236**(1), 012039 (2019)
15. A. Prieto et al., Neural networks: An overview of early research, current frameworks and new challenges. Neurocomputing **214**, 242–268 (2016)

Chapter 7
Semi-automatic Segmentation of COVID-19 Infection in Lung CT Scans

Faridoddin Shariaty ⓘ**, Mojtaba Mousavi, Azam Moradi, Mojtaba Najafi Oshnari, Samaneh Navvabi, Mahdi Orooji, and Boris Novikov**

Abstract As the prevalence of COVID-19, concerns about the treatment of the disease and its impact on communities' future have increased sharply. The best way to prevent the spread of COVID-19 disease is to quickly diagnose patients and prevent them from coming into contact with healthy people. Computer methods are very effective in finding patients with COVID-19 and speed up the diagnosis. These methods are also widely used to assess a patient's condition, for example, to assess the disease's progression over time and to measure the rate of spread of the virus in the lungs. In this article, a segmentation method is introduced to segment the infected parts of the lung in CT scans. This method is based on Lazy-Snipping and Super-pixel algorithms. As a result of segmentation, the performance of algorithm is presented and compared with other methods using Dice score which was 80%.

Keywords COVID-19 · CT scan · Computer Aided Detection (CAD) · Super-pixel algorithm · Lazy-snipping algorithm

F. Shariaty (✉)
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: Shariati.f@edu.spbstu.ru

M. Mousavi
Khatam Al-Nabieen University, Kabul, Afghanistan
e-mail: s.mojtaba.mousavi@knu.edu.af

A. Moradi
Federal State Autonomous Educational Institution of Higher Education, National Research University ITMO, Saint-Petersburg, Russia

M. N. Oshnari · S. Navvabi
Department of Mathematics and Mechanics, Saint-Petersburg State University, Saint-Petersburg, Russia

M. Orooji
Department of Electrical Engineering, Tarbiat Modares university, Tehran, Iran

Department of Electrical Engineering, University of California, Davis, USA

B. Novikov
Saint Petersburg HSE University, Saint-Petersburg, Russia

7.1 Introduction

Assessing the disease's progression is very important in medicine, which helps to accurately analyze the type of disease. For example, in COVID-19 disease, one of the critical components for each patient is to check the level of progression of the virus and infection in the lungs, according to which the patient's treatment should be determined. Obtaining the level of disease progression by physicians is a difficult task and requires a lot of time, while the diagnosis and treatment of COVID-19 patients require high-speed methods. This is due to the rapid spread of the disease, and in some cases, the lack of empty beds in hospitals, which complicates treatment and requires more speed [1].

The development of computer systems in the field of medicine [2] helps a lot in the initial diagnosis of the disease quickly as well as following the progress of the disease. In recent years, these systems have found many applications in medicine, for example, in the field of automatic detection of lung cancer [3, 4] or segmentation of nodules in the lungs from surrounding areas. Segmentation of infectious lung tissues from the surrounding areas is one of the most important steps in the processing of medical data [5, 6]. This is also very important for patients with COVID-19, as the segmentation of infected tissues can determine the percentage of virus spread in the lungs. On the other hand, it is very difficult to separate the infected tissues of the lungs because the infected parts of the lungs have a high intensity and are very similar to the environment around the lungs and also very similar to the lungs. Therefore, computer systems usually do not recognize these parts like lungs, and the separation is incomplete [7].

Segmentation methods are divided into three categories: image-based, model-based, and hybrid methods. Model-based methods such as Active Appearance Model and Active Shape Model [8]. These methods use the feature information [9] in the image to segment it. Image-based methods use only the information in the image appearance to segment it. Among these methods, we can mention morphological operation, thresholding [10], watershed, level set, active contour and region growing [11].

This article, reviewed standard methods of separating different lung tissues in CT scan images in a categorized manner. Also, a new method of isolating infections caused by COVID-19 in the lung has been proposed using CT scan images that can segment the infected tissue with high accuracy. This method is based on Lazy-Snipping and Super-pixel algorithms, which have many applications for the segmentation of lung nodules.

7.2 Review of Literature

As the lungs are complex organs, the separation of lung components is one of the essential steps in computer-aided diagnosis systems. Segmentation methods, as men-

tioned, are divided into three categories: model-based methods, image-based methods, and hybrid methods. In following presented an overview of the literature that has used each of these methods.

CT images provide high-quality 3D images for COVID-19 detection, and deep learning techniques are widely used to segment ROI in these images [12–14]. Frequently used deep networks for patients with COVID-19 include classic U-Net [15–17], UNet++ [18], VB-Net [19, 20] (Table 7.1). Compared to CT, X-ray imaging is easily accessible worldwide. However, splitting X-ray images is even more challenging because of the 2D projections projected onto soft textures and thus confuse the image contrast [21].

7.3 Materials and Methods

7.3.1 Patients

The procedure for selecting patients with COVID-19 is such that to ensure that the patient has COVID-19 and not another similar disease (such as edema), a PCR test is taken from the patient twice. If the result is positive, a CT scan of the lung is taken six days after the disease's onset.

7.3.2 Data Acquisition

To achieve the desired results, 85 cases with COVID-19 disease have been used. These CT scan images were provided from Shariati Hospital and Taleghani Hospital in Tehran, Iran. For each patient, the ground-truth annotation by physicians and CT imaging data were provided. The number of CT scan image layers in these images is from 126 to 135, and their thickness is from 1 to 6 mm.

7.3.3 Methods

The segmentation method used in this article is based on Lazy-Snipping and Superpixel algorithms for segmentation of infected areas in patients with COVID-19 in CT scan image. These two algorithms are described separately below [34, 35].

Table 7.1 The review of literature about lung segmentation for COVID-19 patients

Literature	Modality	Method	Target ROI	Application
Zheng et al. [22]	CT	U-Net	Lung	Diagnosis
Cao et al. [23]	CT	U-Net	Lung	Quantification
			Lesion	
Huang et al. [24]	CT	U-Net	Lung	Quantification
			Lesion	
			Lung lobes	
Qi et al. [25]	CT	U-Net	Lung lobes	Quantification
			Lesion	
Gozes et al. [26]	CT	U-Net	Lung	Diagnosis
			Lesion	
Chen et al. [27]	CT	U-Net++	Lesion	Diagnosis
Jin et al. [28]	CT	U-Net++	Lung	Diagnosis
			Lesion	
Shan et al. [20]	CT	VB-Net	Lung	Quantification
			Lesion	
			Lung lobes	
			Lung segment	
Tang et al. [29]	CT	Commercial Software	Lung	Quantification
			Lesion	
			Trachea	
			Bronchus	
Shen et al. [30]	CT	Threshold-based region growing	Lesion	Quantification
Li et al. [31]	CT	U-Net	Lesion	Diagnosis
Wang et al. [32]	CT	Noise-robust deep learning	Lesion	Quantification
Zhou et al. [13]	CT	3D U-net	Lesion	Quantification
Amyar et al. [33]	CT	Multi-task deep learning	Lesion	Quantification

7.3.3.1 Super-Pixel Method

To make segmentation easier and more efficient, as well as to create more inclination between the infected area as the foreground and the rest of the components as the background, the image of the lung has been turned into several sub-regions. To convert the lung image into several subregions, the simple linear iterative clustering (SLIC) Super-pixel method has been used [35], in which the pixels are classified into a group based on one pixel, which makes it easy to calculate the image properties and the next steps of image processing efficiently. The advantages of the Super-

pixel algorithm include the following: This algorithm connects the borders of the infected area precisely, and when the pre-processing is done, the processing speed and system efficiency increase. In addition, Super-pixel increases the speed and quality of segmentation results.

To use Super-pixels, the k-mean method was used. In this work, we selected the number of Super-pixels arbitrarily and experimentally, then the number of real Super-pixels is selected by the SLIC method. After that, the image is clustered and segmented for the next step.

7.3.3.2   Lazy-Snipping Method

The Lazy-Snipping method is a controllable and flexible method that segments the objects inside the image well, even if they are small or large and have blurred or low contrast borders. The Lazy-Snipping method has two steps. In the first step, the background and foreground are specified by the user. Then in the second step, the system boundaries are separated, which again allows the user to manually correct the borders that are not separated properly. This method is region-based and boundary-based method. Segmentation is such that the image is considered as a graph $G =< v, \xi >$ where v is a set of nodes considered as pixels and ξ as arcs with four or eight neighboring pixels. Then each x_i pixel corresponding to node $i\,(i \in v)$ is made easy as a label for labeling $X = x_i$ by minimizing Gibbs energy $E(X)$.

$$E(X) = \sum_{(i \in v)} E_1(x_i) + \lambda \sum_{((i,j) \in \xi)} E_2(x_i, x_j) \tag{7.1}$$

That $E_1(x_i)$ is likelihood and when the label i is x_i, the valuation is done and $E_2(x_i, x_j)$ is the previous energy and determines that the values of nodes i, j are related to x_i, x_j. The energy terms E_1 and E_2 are specified by the user input, which in Eq. (7.1) E1 encodes the same term and specifies whether it is a foreground or a background. E_2 is used as an indicator of energy along boundaries.

In this step, the image that has been converted to sub-region by the SLIC method is divided by the Lazy-Snipping method so that the image is divided into two parts, foreground, and background. In this method, the background and foreground are specified manually by the user and then using the image obtained in the previous step, which is the number of sub-regions, the image is segmented and the infected regions are recognized.

7.3.4   Statistical Analysis

In order to compare the performance of CNN with the radiologist, the five performance indicators were calculated as follows:

$$Dice = \frac{2 \times N_{TP}}{2 \times N_{TP} + N_{FN} + N_{FP}} \qquad (7.2)$$

$$Sensitivity = \frac{N_{TP}}{N_{FN} + N_{TP}} \qquad (7.3)$$

$$Positive\,Predictive\,Value(PPV) = \frac{N_{TP}}{N_{TP} + N_{FP}} \qquad (7.4)$$

In this study, positive and negative were assigned to the infected and non-infected areas, respectively. Thus, N_{TP} and N_{TN} show the correctly segmented infections and non-infection areas. N_{FP} and N_{FN} show the number of non-correct segmented infections and non-infection areas, respectively.

7.4   Results and Discussion

Figure 7.1 shows the visual performance of the algorithm presented in the paper. In this figure, the points are specified by the user, and the selection of two or three points from the background and one point in each area suspected of COVID-19 is enough to run the algorithm and separate the infected areas. Patients with COVID-19 can be acute or non-acute, as can be seen from Figs. 7.1 and 7.2, where Fig. 7.2 shows that the infection fills most of the lung volume. If the infection occupied a large area of the lung, more than one seed point should be selected for more accurate separation.

To check the accuracy of the proposed algorithm, a radiologist segmented the infected areas due to COVID-19, and finally, the results of the algorithm were compared with the results of manual segmentation. As can be seen in Table 7.2, the performance of the algorithm is very good, and the Dice criterion is 80%.

In this paper, in addition to the Dice criterion, the criteria of sensitivity and Positive Predictive Value (PPV) are also calculated. According to the TP and FN, the sensitivity of the proposed algorithm is 0.77, which shows the high power of the algorithm for segmentation of the areas containing the disease in patients with Covid-19. Also, the Positive Predictive Value (PPV) Criterion is equal to 0.84, which indicates that the algorithm has segmented most of the virus-contaminated areas in the lung on average.
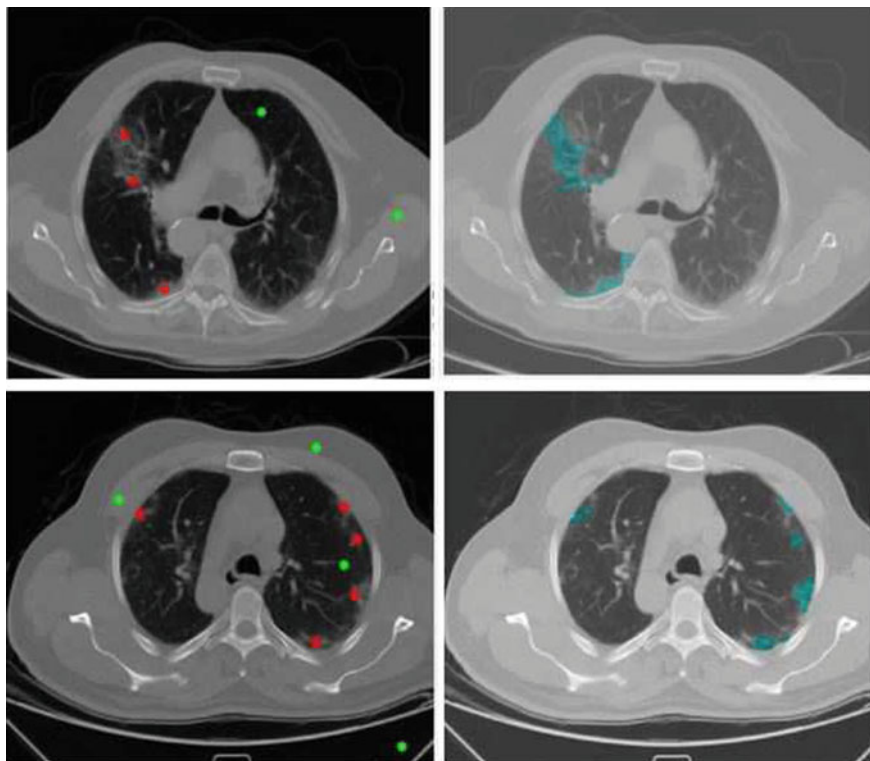
Fig. 7.1 Choose seed points
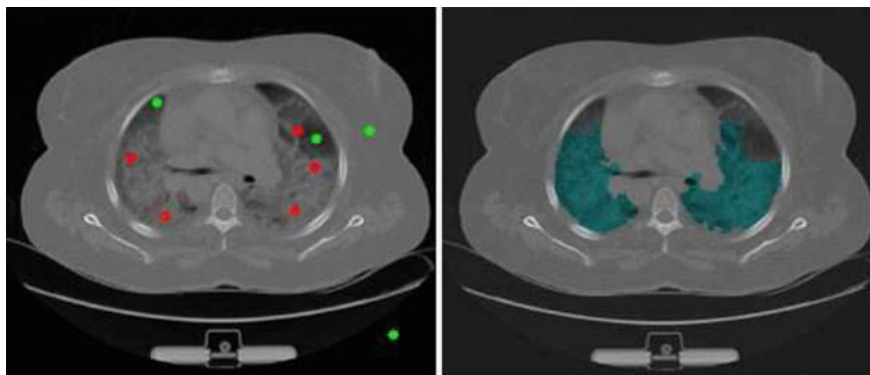
Fig. 7.2 Choose seed points

Table 7.2 A quantitative comparison between our proposed model and state of the art for the segmentation task

	Method	Dice
Proposed Method	Semi-Automatic	0.80
Zhou et al. [17]	U-NET	0.78
Zhou et al. [17]	U-Net + DL	0.61
Zhou et al. [17]	U-Net + FTL	0.66
Zhou et al. [17]	AU-Net + DL	0.68
Zhou et al. [17]	AU-Net + FTL	0.69
Fan et al. [5]	Backbone + PPD + RA + EA	0.73
Wu et al. [36]	JCS	0.77
Wu et al. [36]	JCS	0.78

7.5 Conclusion

In this paper, we present a semi-automatic method for the isolation of COVID-19-induced infections in CT images. Our method can improve the results of segmentation even if we do not have much ground truth of segmentation. Our method shows very promising results and can be of great help to physicians in laboratories.

References

1. E.E.D. Hemdan, M.A. Shouman, M.E. Karar, Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. arXiv preprint arXiv:2003.11055 (2020)
2. M. Ebrahimi, B. Novikov, E. Ebrahimie, A. Spilman, R. Ahsan, M.R. Tahsili, M. Najafi, S. Navvabi, F. Shariaty, The first report of the most important sequential differences between covid-19 and mers viruses by attribute weighting models, the importance of nucleocapsid (n) protein (2020)
3. F. Shariaty, M. Mousavi, Application of cad systems for the automatic detection of lung nodules. Inf. Med. Unlocked **15**, (2019)
4. F. Shariaty, V. Davydov, V. Yushkova, A. Glinushkin, V.Y. Rud, Automated pulmonary nodule detection system in computed tomography images based on active-contour and svm classification algorithm. J. Phys. Conf. Ser. **1410**, 012075 (2019). (IOP Publishing)
5. D.P. Fan, T. Zhou, G.P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, L. Shao, Inf-net: Automatic covid-19 lung infection segmentation from ct images. IEEE Trans. Med. Imaging **39**(8), 2626–2637 (2020)
6. S. Hassantabar, M. Ahmadi, A. Sharifi, Diagnosis and detection of infected tissue of covid-19 patients based on lung x-ray image using convolutional neural network approaches. Chaos, Solitons Fractals **140**, (2020)
7. M. Baranov, E. Velichko, A. Andryakov, Image processing for analysis of bio-liquid films. Optical Memory and Neural Networks **29**, 1–6 (2020)

8. F. Shariaty, M. Orooji, M. Mousavi, M. Baranov, E. Velichko, Automatic lung segmentation in computed tomography images using active shape model, in *2020 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*. (IEEE, 2020), pp. 156–159

9. F. Shariaty, M. Baranov, E. Velichko, M. Galeeva, V. Pavlov, Radiomics: extracting more features using endoscopic imaging, in *2019 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*. (IEEE, 2019), pp. 181–194

10. F. Shariaty, S. Hosseinlou, V.Y. Rud, Automatic lung segmentation method in computed tomography scans. J. Phys. Conf. Ser. **1236**, 012028 (2019). (IOP Publishing)

11. M. Mousavi, F. Shariaty, M. Orooji, E. Velichko, The performance of active-contour and region growing methods against noises in the segmentation of computed-tomography scans, in *International Youth Conference on Electronics, Telecommunications and Information Technologies*. (Springer, 2021), pp. 573–582

12. X. Chen, L. Yao, Y. Zhang, Residual attention u-net for automated multi-class segmentation of covid-19 chest ct images. arXiv preprint arXiv:2004.05645 (2020)

13. L. Zhou, Z. Li, J. Zhou, H. Li, Y. Chen, Y. Huang, D. Xie, L. Zhao, M. Fan, S. Hashmi et al., A rapid, accurate and machine-agnostic segmentation and quantification method for ct-based covid-19 diagnosis. IEEE Trans. Med. Imaging **39**(8), 2638–2652 (2020)

14. F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, D. Shen, *Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19* (IEEE Rev. Biomed, Eng, 2020)

15. N. Saeedizadeh, S. Minaee, R. Kafieh, S. Yazdani, M. Sonka, Covid tv-unet: Segmenting covid-19 chest ct images using connectivity imposed u-net. arXiv preprint arXiv:2007.12303 (2020)

16. D. Müller, I.S. Rey, F. Kramer, Automated chest ct image segmentation of covid-19 lung infection based on 3d u-net. arXiv preprint arXiv:2007.04774 (2020)

17. T. Zhou, S. Canu, S. Ruan, An automatic covid-19 ct segmentation based on u-net with attention mechanism. arXiv preprint arXiv:2004.06673 (2020)

18. Y. Lu, X. Qin, H. Fan, T. Lai, Z. Li, Wbc-net: a white blood cell segmentation network based on unet++ and resnet. Appl. Soft Comput. **101**, (2021)

19. Y. Wang, Y. Chen, Y. Wei, M. Li, Y. Zhang, N. Zhang, S. Zhao, H. Zeng, W. Deng, Z. Huang et al., Quantitative analysis of chest ct imaging findings with the risk of ards in covid-19 patients: a preliminary study. Annal. Trans. Med. **8**(9), (2020)

20. F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, D. Shen, Y. Shi, Lung infection quantification of covid-19 in ct images with deep learning. arXiv preprint arXiv:2003.04655 (2020)

21. E. Velichko, E. Nepomnyashchaya, M. Baranov, M.A. Galeeva, V.A. Pavlov, S.V. Zavjalov, E. Savchenko, T.M. Pervunina, I. Govorov, E. Komlichenko, A concept of smart medical autonomous distributed system for diagnostics based on machine learning technology, in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*. (Springer, 2019), pp. 515–524

22. C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, X. Wang, Deep learning-based detection for covid-19 from chest ct using weak label. MedRxiv (2020)

23. Y. Cao, Z. Xu, J. Feng, C. Jin, X. Han, H. Wu, H. Shi, Longitudinal assessment of covid-19 using a deep learning-based quantitative ct pipeline: illustration of two cases. Radiol. Cardiothoracic Imaging **2**(2), (2020)

24. L. Huang, R. Han, T. Ai, P. Yu, H. Kang, Q. Tao, L. Xia, Serial quantitative chest ct assessment of covid-19: a deep learning approach. Radiol. Cardiothoracic Imaging **2**(2), (2020)

25. X. Qi, Z. Jiang, Q. Yu, C. Shao, H. Zhang, H. Yue, B. Ma, Y. Wang, C. Liu, X. Meng, et al., Machine learning-based ct radiomics model for predicting hospital stay in patients with pneumonia associated with sars-cov-2 infection: a multicenter study. Medrxiv (2020)

26. O. Gozes, M. Frid-Adar, H. Greenspan, P.D. Browning, H. Zhang, W. Ji, A. Bernheim, E. Siegel, Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection and patient monitoring using deep learning ct image analysis. arXiv preprint arXiv:2003.05037 (2020)

27. J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, Q. Chen, S. Huang, M. Yang, X. Yang et al., Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography. Sci. Rep. **10**(1), 1–11 (2020)
28. S. Jin, B. Wang, H. Xu, C. Luo, L. Wei, W. Zhao, X. Hou, W. Ma, Z. Xu, Z. Zheng et al., Ai-assisted ct imaging analysis for covid-19 screening: building and deploying a medical ai system in four weeks. MedRxiv (2020)
29. L. Tang, X. Zhang, Y. Wang, X. Zeng, Severe covid-19 pneumonia: assessing inflammation burden with volume-rendered chest ct. Radiol. Cardiothoracic Imaging **2**(2), (2020)
30. C. Shen, N. Yu, S. Cai, J. Zhou, J. Sheng, K. Liu, H. Zhou, Y. Guo, G. Niu, Quantitative computed tomography analysis for stratifying the severity of coronavirus disease 2019. J. Pharm. Anal. **10**(2), 123–129 (2020)
31. L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song et al., Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. Radiology (2020)
32. G. Wang, X. Liu, C. Li, Z. Xu, J. Ruan, H. Zhu, T. Meng, K. Li, N. Huang, S. Zhang, A noise-robust framework for automatic segmentation of covid-19 pneumonia lesions from ct images. IEEE Trans. Med. Imaging **39**(8), 2653–2663 (2020)
33. A. Amyar, R. Modzelewski, H. Li, S. Ruan, Multi-task deep learning based ct imaging analysis for covid-19 pneumonia: Classification and segmentation. Comput. Biol. Med. **126**, (2020)
34. Y. Li, J. Sun, C.K. Tang, H.Y. Shum, Lazy snapping. ACM Trans. Graph. (ToG) **23**(3), 303–308 (2004)
35. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2274–2282 (2012)
36. Y.H. Wu, S.H. Gao, J. Mei, J. Xu, D.P. Fan, R.G. Zhang, M.M. Cheng, *Jcs: an explainable covid-19 diagnosis system by joint classification and segmentation* (IEEE Trans, Image Process, 2021)

Chapter 8
On the Experience of Using NI ELVIS III in Remote Laboratory Practice During Pandemic Lockdown

Alexey Mayzel ⓘ**, Andrei Medvedev** ⓘ**, Valentina Temkina** ⓘ**,
Liudmila Pankova** ⓘ**, and Eduard Sivolenko** ⓘ

Abstract By the beginning of COVID pandemic, remote education technology have was being implemented at some level by almost every educational institution, but the following lockdown forced the world to accelerate this process. In these circumstances, a lack of laboratory practice in online engineering education was revealed. The existing lab equipment, even quite modern, was focused on offline laboratory experience, and required enough efforts to be reachable from the outside. The International Scientific and Educational Center "National Instruments – Polytechnic" possessed a number of virtual instrumentation lab stations called NI ELVIS II, which had been being used in education for more than 10 years and allowed the replacement outdated bulky instruments with computer-based all-in-one virtual measurement solutions. But even these devices required additional development to allow students to work remotely, did not allow them to work in groups and manage access control. The new product applied by our professors called NI ELVIS III appeared to be a perfect solution to establish online lab practice. We managed to run several labs in electronics course with almost zero configuration, group collaboration and possible real-time teacher supervision. The paper describes the experience of applying NI ELVIS III in remote electronics lab, the benefits and caveats of using this platform.

Keywords Remote educational laboratory · NI ELVIS III

A. Mayzel · A. Medvedev · V. Temkina (✉) · L. Pankova
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: valenssiya@bk.ru

A. Medvedev
e-mail: medvedev@rphf.spbstu.ru

L. Pankova
e-mail: pankova_lv@spbstu.ru

E. Sivolenko
Radiophysics and Electronics, Yerevan State University, Yerevan, Armenia

8.1 COVID Lockdown Remote Education Issues

8.1.1 Prior Circumstances

As for the beginning of 2021 one can take an easy outlook at the evolution of COVID pandemic lockdown issues the world faced with in 2020. It was a major challenge for almost all the economy industries, especially for uninterruptible production processes, but finally for everyone.

Before the beginning of the pandemic most of the academic institutions used the remote educational technologies as a helpful tool supporting the offline course program [1–6]. Even if the institution did not adopt any centralized platform like Moodle e.g., the professors and students used cloud storage, messengers, and video meetings to communicate, interchange the materials and works and reduce the need of meeting in person. For some educational programs, like philosophy, law, history, etc. remote education platforms allowed the totally remote higher education without even visiting the University campus. But for natural sciences and engineering it was never an option to substitute the offline lab practice with remote education technology. But lockdown has shifted priorities, demanding a quick solution to the absence of an offline laboratory practice [7–18].

8.1.2 Using Virtual Instruments

Institute of Physics, Nanotechnology and Telecommunications of Peter the Great St. Petersburg Polytechnic University applied the virtual instrumentation teaching labs more than a decade before. It was relying on the NI ELVIS II [19] base stations with interchangeable lab circuit boards and engineering software like NI LabVIEW, Multisim etc. (Fig. 8.1) [20, 21]. It was a big leap from analogue bulky instruments with fixed functionality to a modern digital programmable flexible solution, allowing the educational process to expand from just observing physical phenomena to advanced experimental technology, measurement automation and process control, as in large physics facilities such as the LHC, ITER and others.

But those virtual instruments were not created for connected world and are not 100% ready for remote access. However, being a part of computer-based system, ELVIS II can be used as remote lab base with some commercial PC remote access solution, like AnyDesk or TeamViewer. This would require certain additional efforts to manage the access control, information security, and, of course, the third-party software licenses.

Fig. 8.1 Computer-based virtual instrumentation workstation using NI ELVIS II platform

8.2 NI ELVIS III for Remote Education Tryout

8.2.1 The Paradigm

In the end of 2019 the International Scientific and Educational Center "National Instruments – Polytechnic" at the Institute of Physics, Nanotechnology and Telecommunications of Peter the Great St.Petersburg Polytechnic University purchased a unit of new NI ELVIS III [22] platform (Fig. 8..2). This platform includes both the evolution of NI ELVIS II and totally different hardware and software architecture, that makes it ideal base for remote online education.

While NI ELVIS II platform relied a on data acquisition (DAQ) board, which required a PC with the ELVISmx driver installed (3.45 GB distributive size as for 2019 version), NI ELVIS III contains an NI RIO standalone controller inside with a webserver running on it. It results in browser-based instruments interfaces with no software installation requirements and no platform limitations.
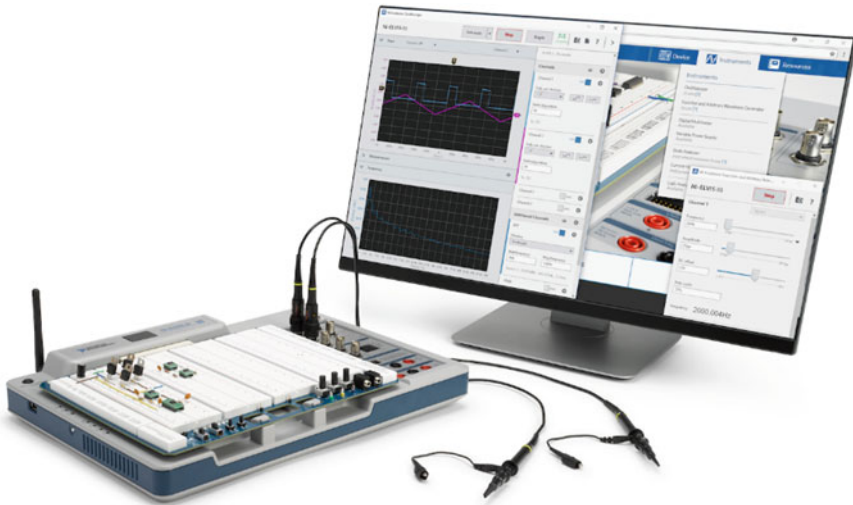
Fig. 8.2 NI ELVIS III with online web-based interface

8.2.2 Hardware

Compared to NI ELVIS II as an instrument, the new ELVIS III brings several improvements like more measurement channels, wider range, and new measurement types. The comprehensive comparison is shown in Table 8.1.

Additionally, the onboard real-time controller and FPGA allow the implementation embedded systems, such as IoT prototypes, fast control systems, long-time process registrations and many more [23]. This requires additional competencies in FPGA and RT programming, which National Instruments—Polytechnic International Science and Education Centre teaches students and engineers, but these capabilities are beyond the scope of this paper.

8.2.3 Basic Remote Access

The main advantages for fast rollout of the online educational laboratory are web-based interfaces, rich connectivity, standalone functioning, and multiple user simultaneous access to a single device. So, in the default approach the NI ELVIS III requires no additional configuration, no software installation on teacher's or students' workstations. While being connected to the internet, NI ELVIS III establishes a secure connection to https://measurementslive.ni.com/. The tiny display on the top of the device shows a temporary code, which a professor sends to students no matter where they are, and they can connect to the instruments of the station in groups, controlling or observing the measurements.

Table 8.1 The comparison of two generations of the NI Education platform

	Legacy NI ELVIS II/II+	NI ELVIS III
Oscilloscope	2 ch, 1.25 MS/s, 10 bits (100 MS/s for II+)	4 ch, 400 MS/s, 14 bits
Function generator	1 ch, 5 MHz, 10 bits	2 ch, 100 MS/s, 15 MHz, 14 bits
Logic analyzer/pattern generator	–	16 ch, 100 MS/s
IV Analyzer	–	$\pm$ 10 V, $\pm$ 30 mA, 15 MHz
Digital multimeter	5½ digits	4½ digits
Variable power supply	$\pm$ 12 V, 500 mA	$\pm$ 15 V, 500 mA
Processor FPGA	–	Xilinx Zynq-7020
AI/AO	16 ch, 16 bits/2 ch, 16 bits	16 ch, 16 bits/4 ch, 16 bits
DIO	24 DIO, 15 PFI	40 ch
OS Support	Windows	Windows, Mac, Web
Connectivity	USB 2.0	USB Type-C, Ethernet, Wi-Fi
Programming Language support	LabVIEW	LabVIEW (Including Real-Time and FPGA), Python, C, Simulink

8.2.4   Advanced Remote Access Management and the Experience

If this zero-config access does not satisfy the needs of the laboratory in terms of access control, there is a possibility to get more flexible solution. It requires to install and configure a tiny Linux-based daemon called Signaling Server and edit configuration files on the ELVIS III stations. The Signaling Server gives a remote control of the whole lab of ELVIS III to a professor, setting the time of access, resetting the temporary access codes for students etc.

We applied this approach at the International Scientific and Educational Center "National Instruments – Polytechnic" and ran a series of lab practices in electronics course in 2020. That experience verified the described approach consistency. The work in student teams was easy and efficient. No software was installed on the student workstation, and no software platform dependency was noticed. The students even could connect and participate in lab practice using their smartphones, which was not expected to happen.

8.3   Conclusion

The pandemic lockdown has shown that laboratory practice in science and engineering education is not quite ready for online learning. The existing lab instruments, even the digital ones, are not perfectly suitable for remote access. Even with certain additional configuration, the access in groups remains challenging.

The tryout of NI ELVIS III lab station has shown that it is a perfect way to transform many labs into online. It is a standalone, basically zero-config, platform-independent solution, that we managed to implement easily in lockdown circumstances. Our experience should be expanded to other labs of our Institute and even to other Universities, because the remote education is supposed to take bigger part in upcoming years.

References

1. M.A. Babaeva, Online course (MOOC) "concepts of Modern Natural Science" on the national platform of open education: experience in teaching students. J. Phys. Conf. Ser. **1348** (2019).
2. M. Avdeeva, N. Leonova, F. Gomazov, E. Strelcova, The automated system of models of management of information resources of higher education institution, in *IOP Conference Series: Materials Science and Engineering*, vol 666 (2019)
3. Y.V. Bozhko, A.I. Maksimkin, G.K. Baryshev, A.I. Voronin, A.S. Kondratyeva, Digital transformation as the key to synthesis of educational and innovation process in the research university. Commun. Comput. Inf. Sci. **674**, 386–391 (2016)
4. L. He, N. Yang, L. Xu, F. Ping, W. Li, Q. Sun, Y. Li, H. Zhu, H. Zhang, Synchronous distance education vs traditional education for health science students: a systematic review and meta-analysis. Med. Educ. **55**(3), 293–308 (2021)
5. K.M. Moser, T. Wei, D. Brenner, Remote teaching during COVID-19: Implications from a national survey of language educators. System **97**, art. no. 102431 (2021)
6. N. Almazova, A. Rubtsova, E. Krylova, D. Barinova, Y. Eremin, N. Smolskaia, Blended learning model in the innovative electronic basis of technical engineers training, in *Proceedings of the 30th DAAAM International Symposium.* (DAAAM International, Vienna 2019), pp. 0814–0825
7. V.J. Bhute, P. Inguva, U. Shah, C. Brechtelsbauer, Transforming traditional teaching laboratories for effective remote delivery—a review. Educ. Chem. Eng. **35**, 96–104 (2021)
8. O.G. McGrath, Learning on and at the edge: Enabling remote instructional activities with micro controller and microprocessor devices, in *Proceedings ACM SIGUCCS User Services Conference, 48th ACM SIGUCCS Annual Conference, SIGUCCS 2021.* (Association for Computing Machinery, 2021), pp. 16–22
9. J. Kobylarczyk, D. Kuśnierz-Krupa, Student assessment of remote learning as an alternative to on-campus learning at technical universities during a pandemic. World Trans. Eng. Technol. Educ. **19**(1), 48–51 (2021)
10. H.S. Jo, R.S. Jo, Design and development of remote laboratory system to facilitate online learning in hardware programming subjects, in *2020 13th International UNIMAS Engineering Conference, EnCon* 2020, art. no. 9299326. IEEE (2020), pp. 1–5
11. F. Ouatik, M. Raoufi, F. Ouatik, M. Skouri, Online instrument systems in reality for remote wiring and measurement of electronic in e-learning from labVIEW+NI ELVIS II vs embedded system+web services. Int. J. Electr. Comput. Eng. **11**(2), 1178–1185 (2021)
12. I. Etier, A. Charles, G. Suresh Kumaar, N. Kannan, NI test stand based automated testing of robotic arm placed standard electrical elements. J. Green Eng. **11**(1), 327–335 (2021)

13. I.A. Rumyancev, M.A. Ivanov, M.O. Aivazova, Multi-user system for manual and automated online measurements, in *2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)*. (IEEE, 2021), pp. 1051–1054

14. M.V. Lagunova, L.A. Ivanova, N.V. Ezhova, Organization of remote education for higher mathematics. challenges and solutions, in *Advances in Mechanical Engineering, MMESE 2020* ed. by A.N. Evgrafov, Lecture Notes in Mechanical Engineering. (Springer, Cham, 2021).

15. N. Almazova, E. Krylova, A. Rubtsova, M. Odinokaya, Challenges and opportunities for russian higher education amid COVID-19: teachers' perspective. Educ. Sci. **10**, 368 (2020)

16. L.N. Borisoglebskaya, A.A. Fedotov, O.V. Pilipenko, S.M. Sergeev, O.J. Kravets, Multi-disciplinary educational information environment. J. Phys. Conf. Ser. **1691**, 012065 (2020)

17. A.I. Dashkina, L.P. Khalyapina, A.M. Kobicheva, M.A. Odinokaya, D.A. Tarkhov, Developing a model of increasing the learners' bilingual professional capacity in the virtual laboratory environment. Appl. Sci. **10**, 7022 (2020)

18. Z. Dvorakova, A. Kulachinskaya, How the COVID-19 made universities switch to distance education: the Russian and Czech cases, in *Proceedings of the 2nd International Scientific Conference on Innovations in Digital Economy: SPBPU IDE-2020 (SPBPU IDE '20)*, Article 50. Association for Computing Machinery, New York (2020), pp. 1–7

19. NI Educational Laboratory Virtual Instrumentation Suite II Series (NI ELVISTM II Series) User Manual. https://www.ni.com/pdf/manuals/374629c.pdf. Last accessed 10 Mar 2021

20. S.A. Molodjakov,, S.I. Ivanov, A.P. Lavrov, Optoelectronic pulsars processor and its real-time software, in *2017 IEEE II International Conference on Control in Technical Systems (CTS),* CTS 2017 (2017), pp. 59–62

21. V. Temkina, A. Medvedev, A. Mayzel, Computer modeling of fiber optic current sensor. Springer Proc. Phys. **255**, 263–276 (2021)

22. NI ELVIS III 1.0 MANUAL. Hardware Specifications. https://www.ni.com/documentation/en/ni-elvis-iii/1.0/appendix/specs/. Last accessed 10 Mar 2021

23. A. Mayzel, A. Medvedev, V. Temkina, CPU-Based FPGA algorithm model of fiber optic current sensor demodulator. Springer Proc. Phys. **255**, 455–461 (2021)

Chapter 9
Application of Deep Learning Techniques for Detection of COVID-19 Using Lung CT Scans: Model Development and Validation

Vitalii A. Pavlov, Faridoddin Shariaty, Mahdi Orooji, and Elena N. Velichko

Abstract Due to the worldwide spread of the COVID-19, efforts to combat the disease have intensified. Among these efforts, the only effective way to prevent further spread to communities and disease progression is to control the spread of the disease, which is done using public vaccination as well as repeated and rapid testing to diagnose and isolate sick people. In this regard, computer systems with the help of medical science can speed up the diagnosis of COVID-19 disease. This paper, proposed a review of the methods used in rapid and automatic detection of COVID-19 using CT scan images. Finally, by presenting a new method based on deep learning, the obtained results compared with the results of widely used algorithms such as VGG-16 and MobileNet.

9.1 Introduction

The outbreak of the COVID-19 began in 2019 in Wuhan, China, and has had a major impact on the global economy and health system. Following the outbreak, the World Health Organization identified the disease as a global public threat. By March 2021, the disease had killed about 2 million people and infected 90 million, with a relatively

V. A. Pavlov (✉) · F. Shariaty · E. N. Velichko
Peter the Great St Petersburg Polytechnic University, Saint Petersburg, Russia
e-mail: pavlov_va@spbstu.ru

F. Shariaty
e-mail: shariati.f@edu.spbstu.ru

F. Shariaty
Department of Electrical Engineering, University of Qom, Qom, Iran

M. Orooji
Department of Electrical Engineering, Tarbiat Modares University, Tehran, Iran

Department of Electrical Engineering, University of California, Davis, USA

high mortality rate among various diseases. Given medical advances and available statistics, currently, ways to deal with COVID-19 include preventing the spread of the virus with a rapid diagnosis of the disease and general vaccination. To prevent the spread of the virus with a rapid diagnosis of patients, an automatic COVID-19 diagnosis system is needed to detect the disease in its early stages.

Early detection of COVID-19 disease is very important and is directly related to the spread of the disease and even mortality. If a person is diagnosed with the disease in the early stages, it will reduce their contact with ordinary people and prevent the spread of the disease. In addition, if the disease is diagnosed in the early stages, the treatment of the disease becomes much easier and the probability of recovery is greatly increased. If COVID-19 is diagnosed late, the infection will spread to the lungs and will have irreversible consequences such as shortness of breath and severe coughing.

Due to the development of computer systems, these systems can be used in medicine and the diagnosis of disease [1]. In recent years, many advances have been made in computer systems and have been widely used to diagnose lung cancer [2]. Recently, due to the prevalence of COVID-19 disease [3], these systems can be used for automatic and early diagnosis of this disease. Many systems have recently been designed for this purpose, most of which use machine learning.

In this paper, a review of the literature has been conducted to introduce the methods used based on machine learning for the early identification of COVID-19. Next, a machine learning-based system has been designed and simulated that can automatically detect COVID-19 disease in its early stages using CT scan images. The system, which uses a custom structure, consists of eight deep layers, two connected layers, and an output layer. The results of the presented method are compared with conventional methods based on deep learning.

9.2   Review of Deep Learning Approaches for COVID-19 Determination

Using X-ray and Computed Tomography (CT) images are the main diagnostic methods, which play an important role in the global dealing with the COVID-19. In the last years, using Artificial Intelligence (AI) technologies further enhance the capabilities of visualization tools and helps professionals process data [4, 5]. AI image processing can greatly help automate the diagnostic process by accurately separating infection on X-ray and CT images, which facilitates subsequent quantification and diagnosis. Thus, computer platforms help healthcare professionals make clinical decisions: diagnose, track, and predict diseases.

Given the high rate of false-positive RT-PCR test results, many patients must be tested several times at several days intervals before a definitive diagnosis is made. Thus, imaging results play a critical role in containing the spread of the virus, as well

as in dealing with COVID-19. The imaging-based diagnostic procedure for COVID-19, as exemplified by a CT scan of the chest, includes three main steps [6, 7]

1. preliminary preparation for scanning,
2. obtaining images
3. diagnosis of diseases.

Artificial intelligence (AI), new technology in medical imaging, is being actively used to combat COVID-19 [8]. Compared to traditional imaging processes that are heavily dependent on humans, AI provides safer, more accurate, and efficient imaging solutions. AI applications for COVID-19 over the past two years have consisted primarily of a dedicated imaging platform, segmentation of lungs and areas of infection, clinical evaluation, and diagnosis. X-rays and CT scans of the lungs are often used to diagnose lung diseases, including pneumonia and COVID-19 [9–14].

9.2.1 COVID-19 Detection Using AI Methods in CT Images

Dynamic lung CT of COVID-19 described and summarized in 4 steps [12]. In summary, the first four days after the occurrence of early symptoms, the early stage is considered and GGO can be seen subpleural in the lower lobes unilaterally or bilaterally. Progressing stage is 5–8 days when it is possible to find diffuse GGO spreading over bilateral polylobed. At the peak stage (9–13 days), dense consolidation becomes more common. When the infection is controlled, absorption occurs (usually after 14 d) and consolidation is gradually absorbed and only GGO remains. These x-ray patterns are important evidence for CT-based classification and COVID-19 severity assessment.

There are several studies aimed at classifying patients with COVID-19 from patients without COVID-19 (which includes patients with advanced pneumonia and subjects without pneumonia). Chen et al. [15] used lung CT scans of 51 COVID-19 patients and 55 patients with other diseases for training a UNet++-based segmentation model which as a result segments lesions associated with COVID-19. The results of this method were: accuracy = 95.2%, sensitivity = 100%, and specificity = 93.6%. In another dataset of 16 patients with viral pneumonia and 11 patients without pneumonia, this model identified all patients with viral pneumonia and 9 patients without pneumonia. Time of reading for radiologists reduced by 65% using AI results.

A model based on U-Net + 3D CNN (DeCoVNet) proposed by Zheng et al. [14]. In this method, lungs are segmented using U-Net, and the result of segmentation is used as 3D CNN input to predict the likelihood of COVID-19. 540 lung CT scans (i.e. 313 COVID-19 patients and 229 healthy) were used as data for deep learning. This model achieved a sensitivity of 90.7%, a specificity of 91.1%, and an AUC of 0.959.

Gene et al. [16] used chest computed tomography of 496 cases of COVID-19 and 1385 cases without COVID-19. To segment the lungs and then to cut out positive cases of COVID-19, a two-dimensional CNN-based model is proposed. According to the results, this model achieves a specificity of 95.5%, a sensitivity of 94.1%, and an AUC of 0.979.

9.2.2 Classification of COVID-19 from Other Pneumonias

Due to the radiological similarity of COVID-19 to common pneumonia and viral pneumonia, differentiation in facilitating the screening process would be more useful in clinical practice. Thus, Wang et al. [17] proposed a CNN model for the classification of these two diseases using 99 lung CT scans in which exist 44 COVID-19 patients and 55 typical viral pneumonia). The test dataset shows an overall accuracy, specificity, and sensitivity of 73.1, 67.0, and 74.0%.

Ying et al. [18] proposed deep learning computed tomography system (called Deep Pneumonia using ResNet50) for identifying COVID-19 patients in patients with bacterial pneumonia and healthy people. Chest CT data from 88 COVID-19 patients, 101 bacterial pneumonia patients, and 86 healthy people are used as data for learning the network. Pieces of full lungs from CT images are obtained of the chest as input data of deep learning. The model achieved good results with 86.0% accuracy for classifying between patients with COVID-19 or patients with bacterial pneumonia and 94.0% accuracy for diagnosing pneumonia (COVID-19 or healthy).

Xu et al. [19] use lung CT scans of 219 patients with COVID-19, 224, and 175 influenzas A and healthy cases. A V-Net-based deep learning model at first used for the segmentation of infected areas. Infected area patches were then sent to the Resnet-18 network along with indications of the relative infection distance from the edge, and as output, they had 3 groups. The overall accuracy of the model was 86.7%.

Shi et al. [20] used chest CT scan of 2685 patients, consist of 1658 patients with COVID-19 and 1027 patients with generalized pneumonia. At the preprocessing step, VB-Net [21] used for segmentation of images into different parts like right and left lung. Various handcrafted elements were designed using a trained random forest model. Based on the results of experiments the sensitivity, specificity, and accuracy are 90.7, 83.3, and 87.9%. In addition, test results are grouped by the size of the infection, indicating a low sensitivity in patients with minor infections.

9.3 Materials and Methods

9.3.1 Patients

The procedure for selecting patients with COVID-19 is such that to ensure that the patient has COVID-19 and not another similar disease (such as edema), a PCR test is taken from the patient twice. If the result is positive, a CT scan of the lung is taken 6 d after the onset of the disease.

9.3.2 Data Acquisition

To achieve the desired results, 85 cases with COVID-19 disease and 100 cases without the disease have been used. These CT scan images were provided from Shariati Hospital and Taleghani Hospital in Tehran, Iran. For each patient, the confirmed result of being sick or healthy is known and examined by a doctor. The number of CT scan image layers in these images is from 126 to 135 and their thickness is from 1 mm to 6 mm.

9.3.3 Study of Deep Learning

In this article, a custom deep learning network is designed that, while simple, has a good efficiency compared to other widely used networks. Figure 9.3 shows the architecture of this network which has 5 convolutional layers (input size of 224 × 224 pixels, 32 filters with filter size of 3×3, activation function ReLU, batch normalization), 5 max-pooling, and 2 fully connected (containing 128 neurons, activation function ReLU, and Dropout(0.5)) and Soft-max layer. Among the advantages of this network can distinguish simple structure and hence low computational complexity (Fig. 9.1). The Custom network was trained using the following parameters: Adam optimizer, Categorical cross-entropy Loss Function, 16 batches, and 30 epochs.

VGG-16 consists of five convolution blocks (13 convolution layers) and three FC (fully connected) layers. This deep network is trained on one million images from 1000 classes. VGG-19 consists of 19 layers including five convolution blocks (16 convolution layers) and three FC layers. VGG-19 is CNN's deeper architecture with more layers compared to VGG-16 [22].

The MobileNet model is separable depth convolutions based. Deep convolution in MobileNets applies to each input channel a single filter. Then a 1 × 1 convolution applies for a combination of outputs with the depth convolution. This layer is split into two layers (using a separable convolution splitter): a separate layer for filtering and a separate layer for blending, which leads to reduce the calculation and the size

Fig. 9.1 The architecture of proposed custom network
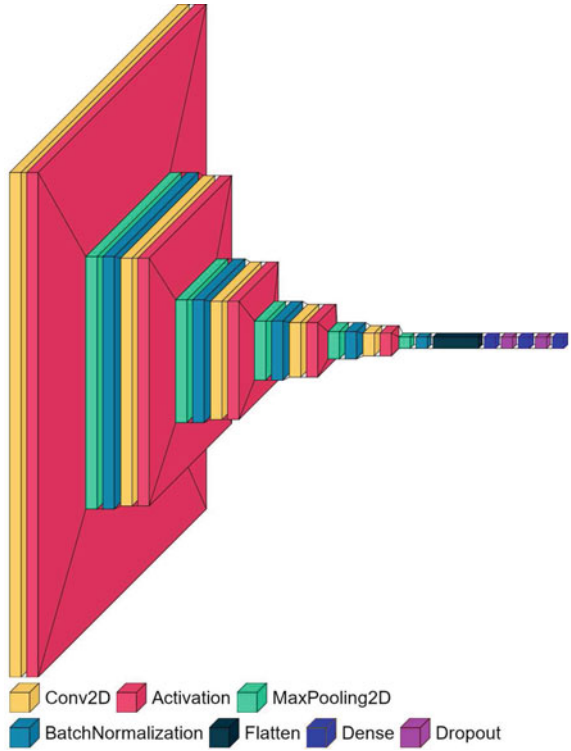
of the model. Figure 9.2 shows how a standard convolution 2 (a) factors into a depth convolution 2 (b) and a 1 × 1 point convolution 2 (c) [23].

In this work, a neural network was used in the VGG-16 and MobileNet networks as a classifier, consisting of two fully connected layers (512 neurons, activation function ReLU, Dropout (0.5)) and a SoftMax classification layer.

9.3.4 Statistical Analysis

In order to compare the performance of CNN with the radiologist, the five performance indicators were calculated as follows:

$$Accuracy = \frac{N_{TN} + N_{TP}}{N_{TN} + N_{FN} + N_{TP} + N_{FP}} \tag{9.1}$$

$$Sensitivity = \frac{N_{TP}}{N_{FN} + N_{TP}} \tag{9.2}$$
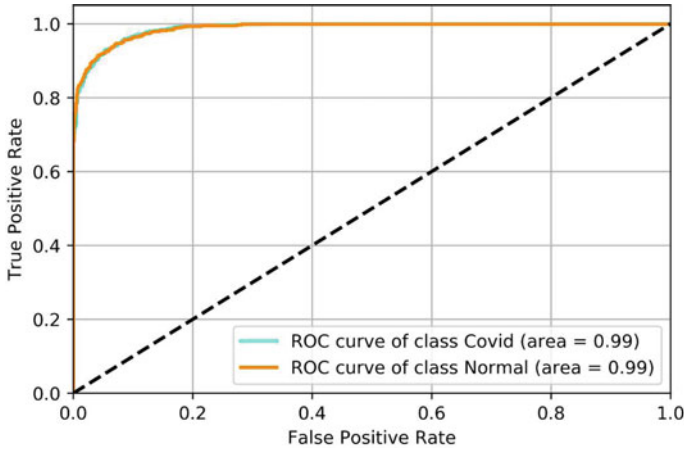
Fig. 9.2 The ROC plot of custom network

$$Specificity = \frac{N_{TN}}{N_{TN} + N_{FP}} \tag{9.3}$$

$$Positive\ Predictive\ Value\ (PPV) = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{9.4}$$

$$Negative\ Predictive\ Value\ (NPV) = \frac{N_{TN}}{N_{TN} + N_{FP}} \tag{9.5}$$

In this study, positive cases were assigned to COVID-19 and negative cases were assigned to normal cases, respectively. Thus, N_{TP} and N_{TN} show the number of correctly diagnosed COVID-19 infections and normal cases, respectively. N_{FP} and N_{FN} show the number of misdiagnosed COVID-19 infections and normal cases, respectively. To calculate these parameters, the number of slices identified by the algorithm as COVID-19 and normal in the results section was used, but the tables in the results are mentioned according to the number of patients because this way the results are better understood.

9.4 Results

For the implementation of the algorithm, the distribution of train and test and validation data is shown in Table 9.1. As indicated, for the evaluation of the proposed model, confusion matrix, the area under the curve (AUC) of the receiver operating characteristics (ROC) are calculated and plotted (Fig. 9.2).

Table 9.1 The distribution of dataset

	COVID-19	Normal
Train set	153	84
Validation set	51	28
Test set	51	28

Table 9.2 Confusion table using custom network for prediction of COVID-19

		Predicted classes	
		COVID-19	Normal
Actual classes	COVID-19	TP = 49	FN = 2
	Normal	FP = 2	TN = 26

The confusion matrix consists of the main parameter: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) (Table 9.2). In Table 9.2 the rows and attributes are the 'Actual class values' and 'Predicted/detected class values'. In order to evaluate the models' performance, this has been used. The table of confusion of test data for the proposed custom network has represented a sensitivity and specificity of 95.8 and 92.8%, in which sensitivity refers to the true positives (TP) number or in this case, is COVID-19 positive patients. Hence, the custom network correctly predicted 49 patients with COVID-19 out of 51 cases which giving a probability of 96%. This shows that a custom network can predict COVID-19 among patients with only 4% error.

Specificity mentions the measurement of true negative (TN) patients which in our work is normal cases. It means, using a custom network, 26 cases were correctly detected as normal (without COVID-19) out of 28 cases which presents a probability of 92.8%. This shows custom network can predict COVID-19 in the CT scan of infected patients with a 7.3% error rate. By calculating the proposed model precision using the confusion table, the custom network shows overall precision of 94% given the limited amount of used training data.

9.5 Discussion

As mentioned, the amount of data used is the same for all methods, and the distribution of training, testing, and validation data is the same. In this work a reduction algorithm based on deep learning is presented, the advantage of which is the simplicity of the algorithm. In this section, the results of the presented algorithm are compared with the results of the widely used algorithms of VGG-16 and MobileNet. For the implementation of these algorithms, the distribution of train and test and validation data is the same as a custom network (Tables 9.1, 9.3 and 9.4).

Table 9.3 Confusion table using VGG-16 for prediction of COVID-19

		Predicted classes	
		COVID-19	Normal
Actual classes	COVID-19	TP = 49	FN = 2
	Normal	FP = 2	TN = 26

Table 9.4 Confusion table using MobileNet for prediction of COVID-19

		Predicted classes	
		COVID-19	Normal
Actual classes	COVID-19	TP = 50	FN = 1
	Normal	FP = 1	TN = 27

Table 9.5 Comparison of the results of used methods

		Precision	Recall	F-score
Custom	Covid-19	0.95	0.95	0.95
	Normal	0.92	0.91	0.92
VGG-16	Covid-19	0.97	0.96	0.96
	Normal	0.93	0.95	0.94
MobileNet	Covid-19	0.98	0.96	0.97
	Normal	0.94	0.96	0.95

Tables 9.5 and 9.6 show the results of these three algorithms that the results of these three algorithms are approximately equal. However, according to the points stated, the presented algorithm in the article has significant advantages. These benefits can be simple structure and hence low computational complexity. The ROC of VGG-16 and MobileNet also illustrated in Fig. 9.3.

From the confusion matrix (Tables 9.3 and 9.4) of VGG-16 and MobileNet in the test data set, a sensitivity of 97% and specificity of 93% for VGG-16 and sensitivity of 98% and specificity of 94% for MobileNet are achieved. Sensitivity is related to the measurement of true positives (TP) or in this work, COVID-19 positive patients. Therefore, VGG-16 correctly detected COVID-19 in 49 cases out of 51 in test data which means the probability is 97% (considering the number of slices) and MobileNet correctly detected COVID-19 in 50 cases out of 51 in test data which means the probability is 98%. It means, VGG-16 can predict COVID-19 in cases with infection with 3% error and 2% error in MobileNet.

As stated specificity mentions the measurement of true negative (TN) patients which in our work is normal cases. It means, using VGG-16, 26 cases correctly detected as normal (without COVID-19) out of 28 cases which presents a probability of 92%, and using MobileNet, 27 cases correctly detected as normal (without COVID-19) out of 28 cases which present the probability of 94%. This shows, VGG-

Table 9.6 Comparison of the results of used methods

	Accuracy	Sensitivity	Specificity	PPV	NPV
Custom	0.94	0.95	0.92	0.96	0.92
VGG-16	0.95	0.97	0.93	0.96	0.93
MobileNet	0.96	0.98	0.94	0.96	0.93

Fig. 9.3 The ROC plot of MobileNet and VGG-16 networks
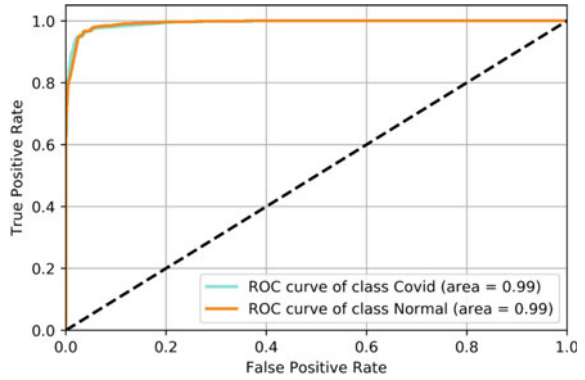
(a) MobileNet
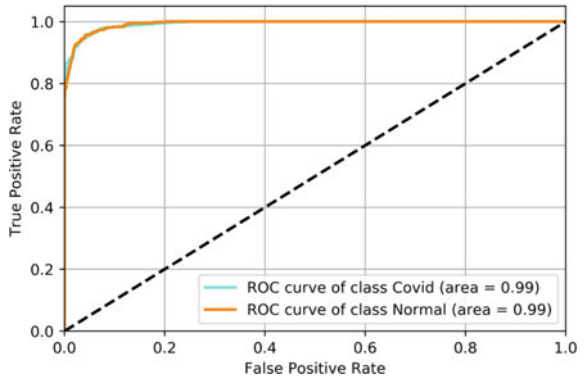
(b) VGG-16

16 can predict COVID-19 in the CT scan of infected patients with an 8% error rate, and MobileNet can predict COVID-19 in the CT scan of infected patients with a 6% error rate. By calculating the VGG-16 precision using the confusion table, VGG-16 shows overall precision of 95%, and MobileNet shows overall precision of 96% given the limited amount of used training data.

9.6 Conclusion

Prompt and timely diagnosis of COVID-19 patients is essential to prevent and control the spread of the disease and is somehow the only way to combat the disease. This research has been done to identify COVID-19 patients from CT lung images in a simple and inexpensive way. In this work, a deep learning model is designed and assembled. The proposed model has reached 94% classification accuracy. More importantly, it is 95% sensitive, meaning that out of 214 COVID-19 patients, 203 can be correctly diagnosed by proposed model. The results obtained from the proposed algorithm are compared with other widely used algorithms such as MobileNet and VGG-16. The results show that while the proposed model is simple, the performance of this model is almost equal to the famous models. It is believed that this research work, together with the graphical user interface, will help physicians to diagnose patients with the help of computer analysis, in a matter of seconds. We believe that this has significantly added value in the field of medicine.

References

1. F. Shariaty, S. Hosseinlou, V.Y. Rud, Automatic lung segmentation method in computed tomography scans. J. Phys. Conf. Ser. **1236**,012028 (2019). (IOP Publishing)
2. M. Mousavi, F. Shariaty, M. Orooji, E. Velichko, The performance of active-contour and region growing methods against noises in the segmentation of computed-tomography scans, in *International Youth Conference on Electronics, Telecommunications and Information Technologies* (Springer, 2021), pp. 573–582
3. M. Ebrahimi, B. Novikov, E. Ebrahimie, A. Spilman, R. Ahsan, M.R. Tahsili, M. Najafi, S. Navvabi, F. Shariaty, The first report of the most important sequential differences between covid-19 and mers viruses by attribute weighting models, the importance of nucleocapsid (n) protein (2020)
4. A.S. Ovsyannikova, S.V. Zavjalov, S.B. Makarov, S.V. Volvenko, Choosing parameters of optimal signals with restriction on correlation coefficient, in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems* (Springer, 2017), pp. 619–628
5. M.A. Lopatin, S.A. Fyodorov, S.V. Zavjalov, D. Ge, Object classification based on channel state information using machine learning, in *International Youth Conference on Electronics, Telecommunications and Information Technologies* (Springer, 2021), pp. 367–374
6. A. Ur, K. Verma, Pulmonary edema in COVID-19-a neural hypothesis. ACS Chem. Neurosc. **11**(14), 2048–2050 (2020)
7. F. Shariaty, M. Mousavi, A. Moradi, M. Oshnari, S. Navvabi, M. Orooji, B. Novikov,: Semi-automatic segmentation of covid-19 infection in lung ct scans
8. H. Yue, Q. Yu, C. Liu, Y. Huang, Z. Jiang, C. Shao, H. Zhang, B. Ma, Y. Wang, G. Xie et al., Machine learning-based ct radiomics method for predicting hospital stay in patients with pneumonia associated with sars-cov-2 infection: a multicenter study. Ann. Transl. Med. **8**(14) (2020)
9. F. Shariaty, M. Mousavi, Application of cad systems for the automatic detection of lung nodules. Inf. Med. Unlocked **15** (2019)

10. F. Shariaty, V. Davydov, V. Yushkova, A. Glinushkin, V.Y. Rud, Automated pulmonary nodule detection system in computed tomography images based on active-contour and svm classification algorithm. J. Phys.: Conf. Ser. **1410**, 012075 (2019) (IOP Publishing)

11. F. Shariaty, M. Orooji, M. Mousavi, M. Baranov, E. Velichko, Automatic lung segmentation in computed tomography images using active shape model, in *2020 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech* (IEEE , 2020), pp. 156–159

12. F. Pan, T. Ye, P. Sun, S. Gui, B. Liang, L. Li, D. Zheng, J. Wang, R.L. Hesketh, L. Yang et al., Time course of lung changes on chest ct during recovery from 2019 novel coronavirus (Covid-19) pneumonia. Radiology (2020)

13. A. Bernheim, X. Mei, M. Huang, Y. Yang, Z.A. Fayad, N. Zhang, K. Diao, B. Lin, X. Zhu, K. Li et al., Chest ct findings in coronavirus disease-19 (COVID-19): relationship to duration of infection. Radiology **200463** (2020)

14. C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, X. Wang, Deep learning-based detection for COVID-19 from chest ct using weak label. MedRxiv (2020)

15. J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, Q. Chen, S. Huang, M. Yang, X. Yang et al., Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography. Sci. Rep. **10**(1), 1–11 (2020)

16. C. Jin, W. Chen, Y. Cao, Z. Xu, X.Zhang, L. Deng, C. Zheng, J. Zhou, H. Shi, J. Feng, Development and evaluation of an ai system for COVID-19 diagnosis. medrxiv 2020. preprint https://doi.org/10.1101/2020.03.20.20039834

17. S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng et al., A deep learning algorithm using ct images to screen for corona virus disease (COVID-19). MedRxiv (2020)

18. Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, H. Zhao, Y. Jie, R. Wang et al., Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with ct images. MedRxiv (2020)

19. X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Q. Ni, Y. Chen, J. Su et al., A deep learning system to screen novel coronavirus disease 2019 pneumonia. Engineering **6**(10), 1122–1129 (2020)

20. F. Shi, L. Xia, F. Shan, D. Wu, Y. Wei, H. Yuan, H. Jiang, Y. Gao, H. Sui, D. Shen, Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification. arXiv preprint arXiv:2003.09860 (2020)

21. F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, Y. Shi, Lung infection quantification of COVID-19 in ct images with deep learning. arXiv preprint arXiv:2003.04655 (2020)

22. Q. Guan, Y. Wang, B. Ping, D. Li, J. Du, Y. Qin, H. Lu, X. Wan, J. Xiang, Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study. J. Cancer **10**(20), 4876 (2019)

23. J. Sivasamy, T. Subashini, Classification and predictions of lung diseases from chest x-rays using mobilenet. Int. J. Analyt. Exper. Modal Anal. **12**(3), 665–672 (2020)

Chapter 10
Algorithm for the Joint Analysis of Beat-To-Beat Arterial Pressure and Stroke Volume for Studying Systemic Vasoconstrictor and Vasodilator Responses

Elizaveta Agapova ⓘ**, Aleksei Anisimov** ⓘ**, Maria Kuropatenko** ⓘ**,
Tatiana Novikova** ⓘ**, Nikolay Suvorov, Timofey Sergeev** ⓘ**,
and Alexander Yafarov**

Abstract Prolonged hypokinesia causes a decrease in functional reserves of the organism. The study of physiological mechanisms and the degree of their decrease and, in particular, the role of vasoconstrictor and vasodilator reactions is an urgent task. This paper describes an algorithm for studying these reactions on the basis of the beat-to-beat pressure and stroke volume signals. The electrodynamic model of the cardiovascular system (Windkessel model) was the theoretical basis for the algorithm development. The signals obtained during functional tests based on postural oscillatory loads were used as test signals. As a result, we obtained an algorithm for determining the total peripheral resistance (by linear regression) and the limits of vasoconstrictor and vasodilatation reactions. This, in turn, makes it possible to judge about the functional reserves of the whole organism.

Keywords Functional reserves · Algorithm · Electrodynamic model · Vasoconstriction · Vasodilation

10.1 Introduction

Reduction of motor activity leads to a significant decrease in functional reserves of the human organism, which is clearly manifested during tests with postural (ortho- and anti-orthostatic) loads. Such postural loads cause a complex of conjugate reactions from the cardiovascular system of human organism. Severity of these reactions

---

E. Agapova · A. Anisimov (✉) · M. Kuropatenko · T. Novikova · N. Suvorov · T. Sergeev · A. Yafarov
Institute of Experimental Medicine, Acad. Pavlov St. 12, 197376 St. Petersburg, Russia
e-mail: aaanisimov@etu.ru

A. Anisimov · T. Sergeev
Saint Petersburg Electrotechnical University, Popova St. 5, 197376 St. Petersburg, Russia

depends on the intensity of the effects (angles of elevation/inclination, speed of movement, sequence and duration of loads) and the initial state of organism [1, 2].

The hemodynamic response to the change of passive ortho- and anti-orthostatic positions is based on the alternating deposition of blood in vessels of the lower and upper body, accompanied by vasoconstrictor and vasodilator vascular reactions. At the same time, changes are occurred in the inotropic response of heart, cardiac output, arterial and venous pressures, heart rate, changes in the circulating blood volume [3, 4, 7, 8].

These changes are associated with the action of the cardiovascular system regulatory mechanisms, including the arterial baroreceptor reflex (ABR), acting according to the principle of negative feedback. For example, the result of decreasing arterial blood pressure (BP), and, consequently, the frequency of impulses from ABR coming to the vasomotor center, is a reflex reaction aimed at increasing ABP, i.e., decreasing the activity of parasympathetic nerve endings in heart and activating vasomotor sympathetic nerves. As a result, the heart rate and its contractility increase simultaneously with the narrowing of arterioles and veins in most body organs, except for the brain and heart [3, 4].

In conclusion, we can say that, for research and diagnostic purposes, it is rather important to consider degree of different components of ABR involvement in cardiovascular system regulation and, in particular, the role of vasoconstrictor and vasodilator reactions. It can be studied by the means of synchronous analysis of beat-to-beat values of ABP and stroke volume (SV) dynamics. Such analysis provides information about value of total peripheral vascular resistance (TPR) and its changes. At the same time, this paper does not consider the effect of ABR itself.

Thus, the aim of this study was to develop an algorithm for the joint analysis of beat-to-beat arterial blood pressure and stroke volume signals for studying systemic vasoconstrictor and vasodilator responses under various impacts, including postural loads. The development of such algorithm will make it possible to solve a number of tasks together with other techniques:

1. To determine the functional reserves of the whole organism, whose value is connected with the ability to regulate the TPR and hence the total volume of circulating blood;
2. To study the adaptive abilities of organism related to the rate of reactions to various influences;
3. To differentiate the diagnosis in pathologies of the cardiovascular system [5–7].

10.2   Materials and Methods

Proposed algorithm is based on the electrodynamic model of cardiovascular system, known as three-element Frank Windkessel (WK) model. Figure 10.1 shows schemes implementing this model: in hemodynamic (Fig. 10.1, a) and electrodynamic (Fig. 10.1, b) representations (according to [10]). There are also known variants of this simple model consisting of two, four and more elements. The history of this
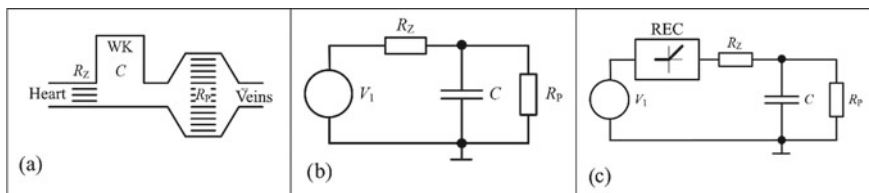
Figure. 10.1. Three-element WK model: a) in hemodynamic, b) and c) in electrodynamic representation, REC is single half-period rectifier

question is considered in details by N.Westerhof et al. [10]. In this scheme R_P corresponds to the total peripheral resistance, C to the total arterial compliance, R_Z to the characteristic aortic resistance.

As it was shown in earlier researches, the results of three-element WK model analysis have a significant contradiction with cardiovascular system functioning [11]. Current passing through R_Z resistance (similar to the flow entering the artery from heart) can have two directions: forward during the pulse simulating systole, and reverse during the pause simulating diastole (Fig. 10.1, b). This result contradicts cardiac function – there should be no return flow to the left ventricle.

This contradiction is eliminated by using an additional nonlinear element (Fig. 10.1, c), simulating the work of the heart left ventricle and corresponding to a single half-period rectifier (REC). At the same time, responses from the original circuit (Fig. 10.1, b) and the circuit with REC (Fig. 10.1, c), with equal values of R_Z, R_P and C in both circuits differ significantly. The main difference is expressed in the rise time and pressure drop. REC scheme provides a relatively constant P level and unidirectional flow. Estimated values of model parameters R_Z, R_P, C, and signal from V_1, obtained by using the scheme with half-period rectifier are more adequate [11], therefore used in further simulations.

As an example, let us consider the signals obtained from model on the 10 s simulation section (Fig. 10.2): ABP is voltage at the R_P resistor (red, green and blue
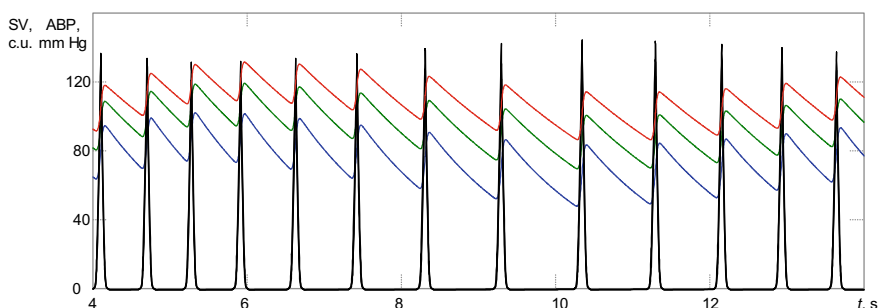

Figure. 10.2. Plot of signals: ABP is voltage across resistor R_P: $R_P = R_{P0}$ (green curve), $R_{P1} = 0.75\ R_{P0}$ (blue curve), $R_{P2} = 1.25\ R_{P0}$ (red curve); SV is current through resistor R_Z (three gray curves, made in conventional units and given for qualitative comparison with ABP curves)

curves) and stroke volume (SV) is current through resistor R_Z (three gray curves, they are visible only in the area of peaks, the scale of SV is chosen in conventional units). All these signals were obtained at different values of TPR –resistance of R_P resistor. If we take R_{P0} (green curve) as the reference value of R_P resistance, then $R_{P1} = 0.75\ R_{P0}$ (blue curve), $R_{P2} = 1.25\ R_{P0}$ (red curve). Values of arterial pressure, as well as systolic (SBP) and diastolic (DBP) blood pressures depend on the values of TPR, and the absolute values of SV also change.

10.3   Results

The first stage of proposed algorithm is the extraction of maximal beat-to-beat values from all the signals. For blood pressure, this is SBP, for stroke volume – peak values in the ejection phase. The next step is to compare these maximal values and plot them on the phase plane. At this stage, the assumption is made that these maximal values are synchronous. Figure 10.3, a shows an example of such construction for three different R_P (TPR) values: $R_P = R_{P0}$ (green markers), $R_{P1} = 0.75\ R_{P0}$ (blue markers), $R_{P2} = 1.25\ R_{P0}$ (red markers).

The third step is devoted to determining the limits of vasoconstriction and vasodilation (Fig. 10.3, b). For each of the R_P values, not a specific value but some set of values is obtained. These sets intersect at the selected change of R_P. The boundaries of the sets are found as straight lines passing through the extreme points of these sets and the origin of the coordinates. Accordingly, the limits of vasoconstriction and vasodilation are the extreme upper and extreme lower limits. The average value of TPR over the measurement period is determined by the standard formula.
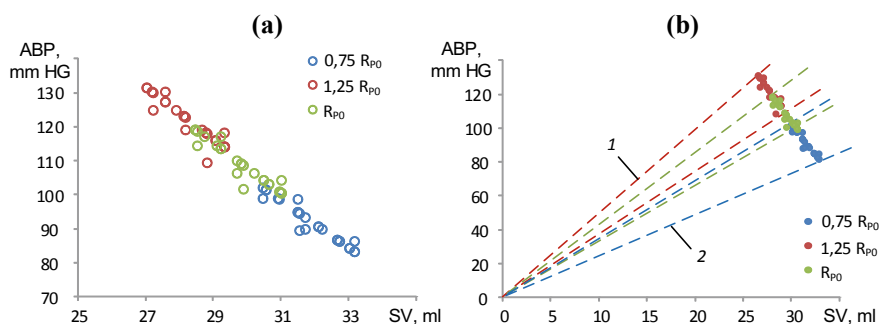

Figure. 10.3.   Phase plane of SBP, SV: (a) in enlarged and (b) normal scales (1 for vasoconstriction boundary, 2 for vasodilation boundary)

10.4 Discussion

It should be noted that the obtained scattergrams for three R_P values are not concentrated at certain points, but have a significant scatter. Moreover, as can be seen from Fig. 10.3, a, the same pairs of SBP-SV at different R_P values are possible (see distribution boundaries). This means that in the case of definition TPR values as the ratio of SBP to TRP, we would obtain different values despite the constancy of R_P (Fig. 10.4). This effect is related to the presence of C (total arterial elasticity) and R_Z (aortic characteristic resistance). Consequently, it is reasonable to determine non-specific (absolute) value of total peripheral resistance, but some set of values when assessing TPR. For this purpose the duration of ABP and SV records at calm position of test subject should exceed 2 min as it is related to durations of regulatory processes [12–14].

In order to induce meaningful changes in TPR, sufficient physiological loads are required. As it was mentioned in the introduction section, the most adequate and universal for this purpose are postural loads performed on a special tilt-table.

In addition, during calculation of TPR we must take into account values of C and R_Z. The development of appropriate methods for this challenge are the main directions for future work.
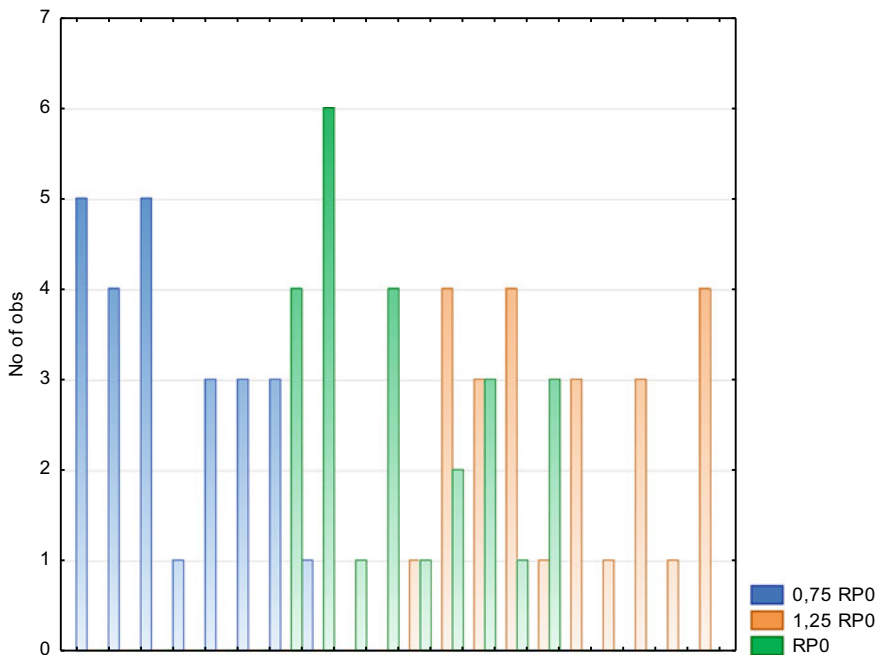

Figure. 10.4. The qualitative distribution of TPR values, calculated as the ratio of ABP to SV values

Thus, the use of the developed algorithm makes it possible to study the limits of systemic vasoconstrictor and vasodilator responses under various loads. When the algorithm described above and the one presented earlier (devoted to evaluation of the arterial baroreceptor reflex effectiveness [16]) are used together, it makes it possible to determine the types of response as well as the variants of functional reserves reduction, i.e., the ways for their recovery. The developed algorithm is intended for use in research and diagnostic systems.

References

1. G.A. Sofronov, N.B. Suvorov, P.I. Tolkachev, T.V. Sergeev, Influence of postural correction of hemodynamics on heart rhythm parameters. Medical Academic Journal **3**, 38–51 (2014)
2. E.M. Lesova, Individual differences of hemodynamic parameters during combination of hypoxic and orthostatic loads. Bulletin of Russian Military Medical Academy **1**(49), 157–163 (2015)
3. D. Morman, L. Heller, Physiology of the cardiovascular system. Piter, St. Petersburg, Russia (2000)
4. L.I. Osadchii, Postural systemic vascular reactions. Phys. J. Named After I.M. Sechenov **3**(79), 66–72 (1993)
5. K. Berganzo, B. Tijero, J.J. Zarranz, J.C. Gómez-Esteban, Postural orthostatic tachycardia syndrome, neurally mediated syncope, and joint hypermobility: a case report. Neurologia **29**, 447–449 (2014)
6. G. Cybulski, *Ambulatory Impedance Cardiography* (Verlag Berlin Heidelberg, The Systems and Their Applications Springer, 2011)
7. J.M. Stewart, Common Syndromes of Orthostatic Intolerance. Pediatrics **131**(5), 968–980 (2013)
8. D.D. Sheriff, I.-H. Nadland, K. Toska, Hemodynamic consequences of rapid changes in posture in humans. J. Appl Physiol **103**, 452–458 (2007)
9. K. Toska, L. Walloe, Dynamic time course of hemodynamic responses after passive head-up tilt and tilt back to supine position. J. Appl. Physiol. **92**, 1671–1676 (2002)
10. N. Westerhof, J.-W. Lankhaar, B.E. Westerhof, The arterial Windkessel. Med Biol Eng Comput **47**, 131–141 (2009)
11. A. Anisimov, T. Sergeev, Postural stress tests for evaluation of pulse wave propagation time based method for indirect blood pressure measurement, in *Young Researchers in Electrical and Electronic Engineering* (IEEE, 2018), pp. 1195–1198
12. C. Julien, The enigma of Mayer waves: Facts and models. Cardiovasc. Res. **70**, 12–21 (2006)
13. A.M. Whittam, R.H. Clayton, S.W. Lord et al., Heart rate and blood pressure variability in normal subjects compared with data from beat-to-beat models developed from de Boer's model of the cardiovascular system. Physiol Meas. **2**(21), 305–318 (2000)
14. R.M. Baevsky, Analysis of heart rate variability: history and philosophy, theory and practice. Clinical Informatics and Telemedicine **1**, 54–63 (2004)
15. A.V. Belov, A.A. Anisimov, T.V. Sergeev, N.B. Suvorov, Tunable band-pass filter for continuous spectral analysis of cardiointervalogram, in *Journal of Physics: Conference Series, International Scientific Conference "Metrological Support of Innovative Technologies - ICMSIT-2020"*, vol. 1515 (IOP Publishing Ltd, 2020)
16. A.A. Anisimov, N.B. Suvorov, N.L. Frolova, A.V. Belov, E.A. Agapova, T.V. Sergeev, Assessment of Baroreflex Mechanism by Joint Analysis of Arterial Blood Pressure and Heart Rate, in *International youth conference on electronics, telecommunications and information technologies*, ed. by E. Velichko, M. Vinnichenko, V. Kapralova, Y. Koucheryavy. Springer Proceedings in Physics, vol. 255 (Springer, Cham, 2020), pp. 565–571

Chapter 11
Development of Simple Model of the Arterial Baroreflex

Aleksei Anisimov , **Alena Tomchuk** , **and Timofey Sergeev**

Abstract It is extremely difficult to develop a full-fledged model of the dynamic operation of the cardiovascular system. This is primarily due to the fact that not only individual organs, but also the activities of the whole body contribute to this complex process. And the temporal scales of cardiovascular system oscillations include fractions of seconds (for pacing, baro- and chemoreceptors, respiratory activity), minutes (for renin-angiotensin system), hours (for renal regulation, ambient temperature influence, blood viscosity) and even a day. Rhythms in each of the above frequency ranges may demonstrate very complex behavior. At the same time, direct measurement of values of parameters characterizing rhythm regulation (which are described in more detail below) may be difficult and sometimes impossible without direct intervention into the cardiovascular system, which is practically unrealizable in standard investigations. That is why mathematical modeling of dynamic behavior of cardiovascular system is sometimes the only possibility to study the processes on qualitative level. Therefore, in this study we attempted to develop a simple model of baroreceptor control based on electrodynamic analogies, which still can provide valuable information for future experiments.

Keywords Model · Arterial baroreflex · Feedback · Arterial blood pressure

11.1 Introduction

Blood pressure (BP) is one of the main indicators of vital functions of the human body, reflecting the work of the entire cardiovascular system in an integral form [1]. Therefore, the construction of models that allow us to trace the interaction of various

A. Anisimov (✉) · A. Tomchuk · T. Sergeev
Saint Petersburg Electrotechnical University, Popova st. 5, 197376 St. Petersburg, Russia
e-mail: aaanisimov@etu.ru

A. Anisimov · T. Sergeev
Institute of Experimental Medicine, Acad. Pavlov Street 12, 197376 St. Petersburg, Russia

elements involved in the regulation of BP is a very urgent and complex task, which has not been fully solved even with the increased computing power [2, 3].

Currently, a very wide set of methods for processing cardiorhythmograms is used to assess heart rate (HR) fluctuations, which are described in many works. This diversity is largely due to the wide spread and introduction into clinical practice of Holter monitors, which allow recording electrocardiogram (ECG) signals for several days [4–7].

Much less attention is paid to the analysis of BP fluctuations and their relationship with the HR and respiratory rate. This is due to the insufficient number of appropriate tools (the existing systems of long-term BP registration work on the basis of cuff measurement methods which excludes the registration of a continuous curve BP), and considerable difficulties encountered in the attempts of theoretical model of the interaction of all the elements involved in the regulation and maintenance of optimal BP.

In addition to the average (static), the dynamic properties of blood pressure also have a significant impact on the state of the cardiovascular system. For example, inhibition of the baroreceptor control mechanism increases the daily level of BP and increases the intensity of its short-term oscillations. Thus, the need to build adequate models of the cardiovascular system that allow us to trace the dynamic behavior and regulation of BP through a complex chain of interacting organs and systems comes to the fore. The construction of mathematical models based on the numerical values of specific experimentally obtained parameters makes it possible to correctly assess the changes occurring in the body for those cases when direct measurements are not possible or involve a risk for the subject [1].

The development of a full-fledged model of the dynamic work of the cardio-vascular system is extremely difficult. This is primarily due to the fact that not only individual organs contribute to this complex process, but also the activity of the entire body as a whole. At the same time, the time scales of the oscillations of the cardio-vascular system include fractions of a second or seconds (for pacemakers, baro- and chemoreceptors, respiratory activity), minutes (for the renin-angiotensin system), hours (for renal regulation, the influence of ambient temperature, blood viscosity), and even days and months (due to the season activity).

Rhythms in each of these frequency ranges can exhibit very complex behavior. At the same time, direct measurement of the values of parameters that characterize the regulation of rhythms (which are described in more detail below) is difficult, and sometimes impossible without direct intervention in the cardiovascular system, which is practically unrealizable in standard studies. That is why mathematical modeling of the dynamic behavior of the cardiovascular system is sometimes the only way to study the processes at a qualitative level.

11.2 Materials and Methods

Studies of complex multicomponent systems are accompanied by attempts to build models of these objects. As new knowledge about such objects is gained, their models develop, improve and become more complex, starting with representations in the form of high-quality structural diagrams. Modeling is of particular importance in physiology and medicine, since the study of the functioning and interaction of the systems of regulation of the human body, in particular, the circuits of the nervous regulation of the work of the cardiovascular system, is of fundamental and applied importance [8].

Figure 11.1 shows a conceptual physiological model of the behavior of the baroreceptor regulation circuits for increasing (solid lines) and decreasing (dotted lines) blood pressure (BP). Let's consider in more detail the process of rising BP. When the pressure in the artery is significantly reduced, the intensity of baroreceptor stimulation also decreases. This leads to a decrease in afferent impulses in the solitary nucleus (SN) of the medulla oblongata. In this case, the following processes occur.

Increased tone of sympathetic innervation from the vasomotor center (VC). Stimulation of VC leads, firstly, to an increase in the stimulation of vasoconstrictors, which leads to vasoconstriction—narrowing of veins and arterioles and, consequently, to an increase in total peripheral vascular resistance (TPVR), decrease in volumetric blood flow (VBF) [9], as well as the movement of a large volume of blood from peripheral blood vessels to the heart cavities, what causes them to stretch. As a result, heart rate (HR) and cardiac output (CO) increase, allowing the heart to pump more blood;
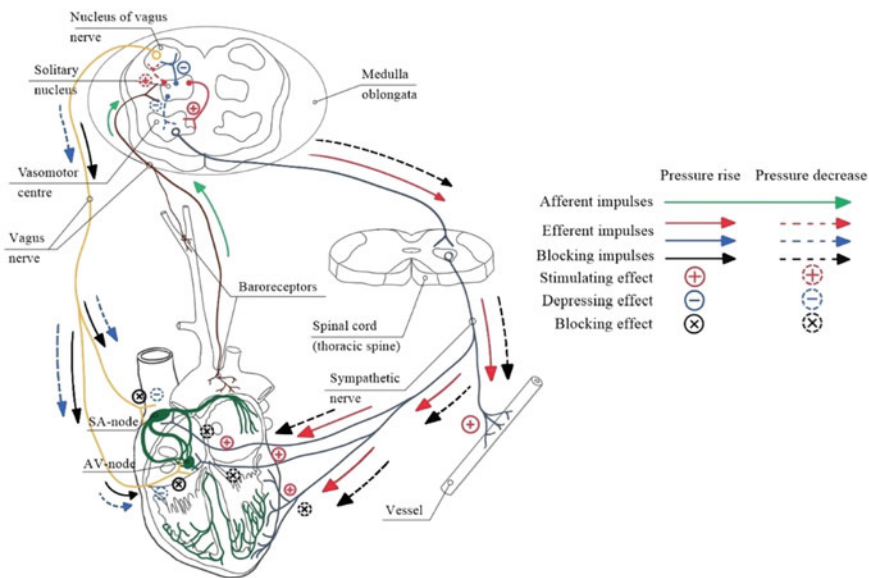

Fig. 11.1 Conceptual physiological model of baroreceptor regulation circuits

secondly, an increase in the stimulation of nerve nodes and fibers that innervate the heart, which leads to an increase in the HR [10].

Decreased tone of parasympathetic innervation from the nucleus of the vagus nerve (NVN). The depressing signal from the SN to the NVN helps to reduce the activity of the parasympathetic nervous system (PNS), which sends inhibitory signals to the heart, which allows the sympathetic nervous system (SNS) to increase the HR without hindrance. Thus, until BP rises to a normal value, control over BP regulation is completely transferred to the SNS. Thus, with increasing pressure, the following parameters change:

- Heart rate increases;
- Cardiac output increases;
- Total peripheral vascular resistance increases;
- Volumetric blood flow decreases;
- Blood pressure increases.

When BP rises strongly, the intensity of baroreceptor stimulation also increases. This leads to an increase in afferent impulses in the SN of the medulla oblongata. In this case, the opposite processes occur.

To build a model describing the relationship between arterial pressure and pulse wave velocity, a transition was made from Frank's Windkessel hydrodynamic model [11, 12] to Frank's electrical model. The simplest two-element model consists of resistance R, which corresponds to the total peripheral resistance (TPR) of the arteries, and capacity C, which acts as an elastic reservoir. Models based on electrical analogies are an effective way to analyze hemodynamic processes in the human body. In this case, the known equations of the electrical circuit theory can be applied, effective for calculations using modern circuit modeling programs, which allow to visualize the solutions of differential equations [13–15]. The result of solving problems related to the analysis of circulatory processes by means of equivalent electrical circuits, with the proper degree of adequacy of the model used and the choice of boundary and initial conditions known from experimental data, sufficiently correspond to the real hemodynamic processes occurring in the human body. The program of circuit modeling MicroCap 12, intended for analysis and modeling of processes in electric circuits, has been used for the analysis of processes of the developed electrical model.

Figure 11.2 shows diagrams, explaining the model, known as Frank's Windkessel (WK): in hemodynamic (Fig. 11.2a) and electrodynamic (Fig. 11.2b) representations. In these schemes, resistance R corresponds to TPR, which is taken as the sum of all individual resistances of different microcirculation sections, which gives the resistance of the entire vascular bed, with the main resistance to flow in the arterial system set by the smallest arteries and arterioles. Capacity C is the total arterial elasticity (or arterial compliance), which is determined mainly by the elasticity of large (conducting) arteries—the smaller the artery, the more insignificant its elasticity and greater resistance. R_Z is the characteristic impedance of the aorta, which allows relating concentrated two-element model and pulsatile pressure wave propagation through the arterial system.
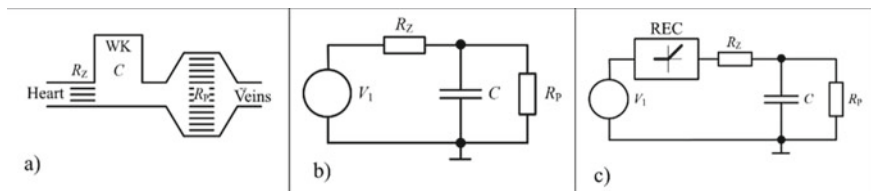
Fig. 11.2 Explanation of the WK model: in hemodynamic **a**, standard **b**, and modified electrody-namic **c** representation

The three-element model, given the statistics accumulated in recent years, is much more accurate in reflecting the relationship between arterial blood pressure and blood flow, especially in the lower frequencies.

Windkessel model assumes that during diastole, when the arterial valve is closed, pressure will decrease exponentially with a time constant $\tau = RC$. Diastolic aortic pressure P_{DIA} (when the valve is closed) can be described by an exponential relationship:

$$P_{DIA}(t) = P_{SIS} \cdot e^{-\frac{t}{RC}}, \tag{11.1}$$

where P_{SIS} is the maximum systolic pressure.

The problem of return flow to the left ventricle (LV) was solved using an additional nonlinear link (Fig. 11.2c), simulating LV valve operation of the heart and corre-sponding to a single half-period rectifier (REC) with the function of the following form:

$$\begin{cases} (P - P_V) \leq 0, \, K_{REC} = 0, \\ (P - P_V) \geq 0, \, K_{REC} = 1. \end{cases} \tag{11.2}$$

where PV is the value of pressure at the input of the rectifier, P is the value at its output, K_{REC} is its transfer coefficient.

11.3   Results

The developed model allows us to study the effect of heart rate variability on blood pressure values without any feedbacks (arterial baroreflex, regulation of total periph-eral resistance, etc.). This is a significant simplification of the conceptual model given earlier, but it still allows a qualitative assessment of the processes taking place. Figure 11.3a shows the section of a 10s cardiorhythmogram (CRG) used in the simulation.

Heart rate here varies between 67 and 116 beats per minute, with a total of 11 complete cardiac cycles. This CRG corresponds to a series of consecutive heartbeat
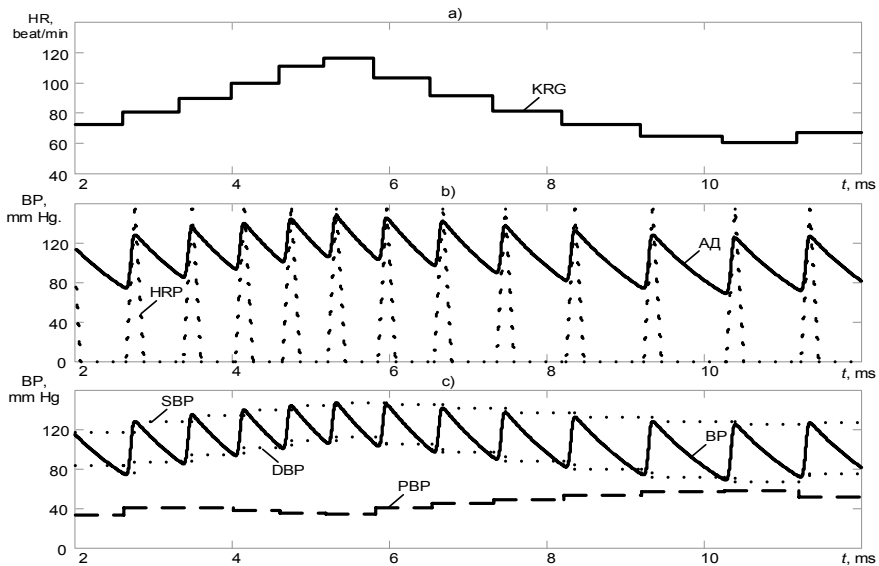
Fig. 11.3 Simulation results: **a** section of the cardiorhythmogram (CRG); **b** heart rate pulse (discontinuous curve) and BP signal (solid curve); **c** BP signal (solid curve), envelope of local maxima—graph of BP beat-to-beat values (discontinuous curve), envelope of local minima—graph of BP beat-to-beat values (discontinuous curve), graph of PBP values (dashed curve)

pulses (HBP) (Fig. 11.3b, discontinuous curve). For clarity, they are presented with normalized values, in the model their source is a voltage generator. As a result of the HBP action on the RC circuit with a nonlinear element the ABP signal is formed, presented in Fig. 11.3b (solid curve). Local maxima of this curve correspond to systolic blood pressure (SBP) beat-to-beat values, minima—to diastolic blood pressure (DBP). The plots corresponding to these beat-to-beat SBP and DBP values are shown in Fig. 11.3c (discontinuous curves). Beat-to-beat values of SBP and DBP, like beat-to-beat values of heart rate, have variable character and vary in the following ranges: from 117 to 147 mm Hg for SBP, from 67 to 112 mm Hg for DBP. In the same figure there is a graph of pulse blood pressure (PBP, dashed curve), which is the difference between systolic and diastolic arterial blood pressure.

According to results of performed modeling, the following preliminary conclusions can be made by visual comparison of plots with HR, SBP, DBP and PBP beat-to-beat values:

1. For the given model parameters (and as it was mentioned above, in the absence of feedback), the HR variability has a significant impact on the BP variability.
2. Systolic and diastolic BP are directly proportional to HR.
3. Pulse pressure is inversely related to HR.
4. The effect of HR on systolic and diastolic BP is manifested in varying degrees. In other case, pulse pressure values would remain constant.
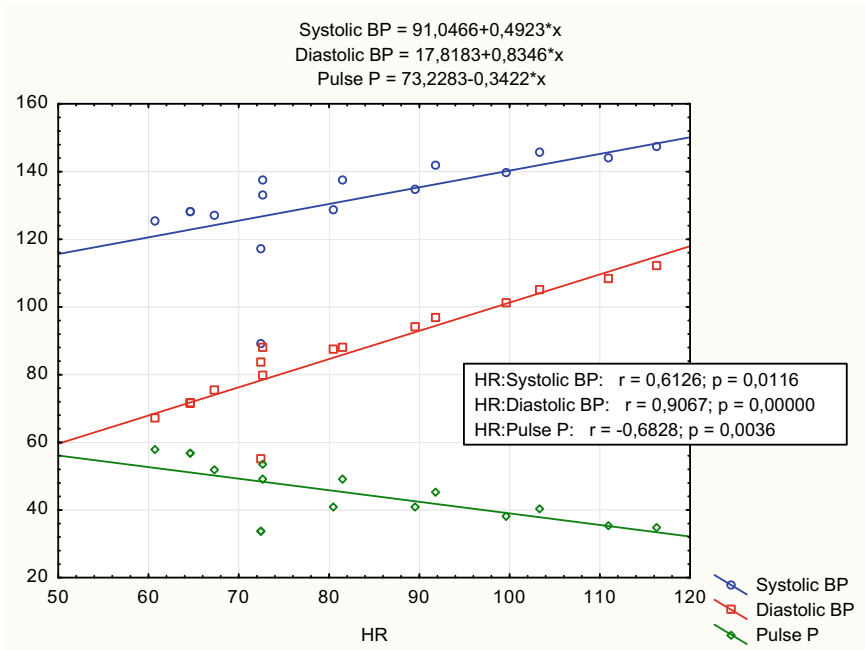5. HR affects diastolic blood pressure more than systolic blood pressure.

Fig. 11.4 Dependences of SBP, DBP and PBP on HR. We also present the parameters of linear approximation (above), correlation coefficient (r) values and p-values

The variants for the signals considered on Fig. 11.3 are shown in Fig. 11.4, these are dependences of beat-to-beat values of SBP, DBP and PBP on HR.

The corresponding parameters of linear approximation (above), correlation coefficient values (r) and p- values (results calculated using Statistica software) are shown in the same figure. The numerical values of the correlation coefficients confirm our preliminary findings: correlation coefficients, r are sufficiently high (r > 0.6) and significant (p < 0.05), with the highest correlation coefficient being the correlation between DBP-HR, and the negative one being the correlation between PBP-HR (Fig. 11.4).

11.4 Conclusion

Thus, the obtained simple model allows studying the dependence of beat-to-beat arterial blood pressure values on HR and other basic parameters of the system hemodynamics only on the basic level. Accordingly, further studies would be connected with the study of the influence of the model main parameters on the dependence of HR-BP: among others are the values of HRP amplitude, total peripheral resistance, elasticity (capacitance) and internal resistance. The next stage of the study should

be the refinement and complication of the model due to consecutive introduction of feedback loops, first of all, arterial baroreflex and regulation of total peripheral resistance.

References

1. D.E. Mohrman, L.J. Heller, in *Cardiovascular Physiology*, 9th edn. (McGraw-Hill Professional, Canada, 2018)
2. M. Albaghdadi, Baroreflex control of long-term arterial pressure. Rev. Bras. Hipertens **14**(4), 212–225 (2007)
3. L.S. Waghmare, T.K. Srivastava, Conceptualizing physiology of arterial blood pressure regulation through the logic model. Adv. Physiol. Educ. **40**(4), 477–479 (2016)
4. N. Singh, K.J. Moneghetti, J.W. Christle et al., Heart rate variability: an old metric with new meaning in the era of using mHealth technologies for health and exercise training guidance. Part One: Physiol. Methods. Arrhythm Electrophysiol Rev. **7**(3), 193–198 (2018)
5. S. Massaro, L. Pecchia, Heart rate variability (HRV) analysis: a methodology for organizational neuroscience. Organ. Res. Methods **22**(1), 354–393 (2019)
6. Y.M. Ishbulatov, A.S. Karavaev, The history and methods of heart rate variability front. Public Health **5**(265), 1–12 (2019)
7. M. Nardelli, N. Vanello, G. Galperti, A. Greco, E.P. Scilingo, Assessing the quality of heart rate variability estimated from wrist and finger PPG: a novel approach based on cross-mapping method. Sensors **20**(11), 1–13 (2020)
8. V.I. Ponomarenko et al., Model of cardiovascular system autonomic regulation with a circuit of baroreflectory control of mean arterial pressure in the form of delayed-feedback oscillator. Izv. Sarat. University Ser. Phys. **15**(2), 32–37 (2015)
9. J.T. Ottensen, Modelling the dynamical baroreflex-feedback control. Math. Comput. Model. **31**, 167 (2000)
10. J.V. Ringwood, S.C. Malpas, Slow oscillations in blood pressure via a nonlinear feedback model. Amer. J. Physiol. **280**, 1105 (2001)
11. N. Westerhof, J.W. Lankhaar, B.E. Westerhof, The arterial Windkessel. Med. Biol. Eng. Compu. **47**(2), 131–141 (2009)
12. N. Stergiopulos, J. Meister, N. Westerhof, Evaluation of methods for estimation of total arterial compliance. Amer. J. Physiol.-Heart Circulatory Physiol. **268**(4), 1540–1548 (1995)
13. Y. Aboelkassem, Z. Virag, A hybrid Windkessel-Womersley model for blood flow in arteries. J. Theor. Biol. **462**, 499–513 (2019)
14. M.A. Bahloul, T.M. Laleg-Kirati, Three-element fractional-order viscoelastic arterial Windkessel model. in *2018 40-th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE, 2018), pp. 5261–5266
15. M. Paisal, A. Sufyan, I. Taib et al., An analysis of blood pressure waveform using windkessel model for normotensive and hypertensive conditions in carotid artery. J. Adv. Res. Fluid Mech. Thermal Sci. **57**(1), 69–85 (2019)

Chapter 12
Analysis of Classification Methods for Wi-Fi Signals Based on the Use of Channel State Information Spatial Features and CNN

Maksim A. Lopatin, Stanislav A. Fyodorov, and Ge Dong

Abstract Due to Wi-Fi technology, signal characteristics can be collected and analyzed, including Channel State Information (CSI), which contains a lot of information about the signal. This paper discusses the structure of CSI packages in terms of how their components correlate to each other, as well as the importance of CSI features for use in machine learning. A simple preprocessing of data before feeding it into the model is described. The paper explores the possibility of classifying the location of a person in an apartment through CSI using a Convolutional Neural Network (CNN). The structure of CNN is described, as well as the results of classification in comparison with other methods of machine learning on the same data.

12.1 Introduction

Nowadays, Wi-Fi technology can be used to collect Channel State Information (CSI), which is able to describe well the physical medium of data transmission. This allows the use of CSI in various tasks of machine learning (ML), which has grown in popularity in recent years [1–4]. In machine learning, when analyzing Wi-Fi signals, any existing methods can be used, including methods of deep learning using neural networks [5].

In this paper, we used a Convolutional Neural Network (CNN) [6, 7] to recognize the location of a person in an apartment. Channel State Information has many features for classification, in our case it is equal to 224 and One of the purposes of this article is to see how they relate to each other. And can we neglect some of them without losing the accuracy of the classification.

M. A. Lopatin (✉) · S. A. Fyodorov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: maksim.lopatin.spb@mail.ru

G. Dong
Tsinghua University, Beijing, China

Our goal is to assess the correlation of the Wi-Fi signal subcarriers, assess their importance for classification, and compare the effectiveness of CNN compared to simpler machine learning methods. In this paper we use a convolutional neural network of a simpler structure with simpler input data than similar articles [9, 10]. Also, here in a simpler form the whole process of the experiment is described, namely, the removal of samples of the Channel State Information and their preprocessing before sending them to the machine learning model. In similar papers [15], more complex sampling processing schemes are presented using information about the phases of subcarriers of signals, which are really sensitive to changes in the environment, and which are difficult to cope with [14].

12.2 CSI Data Correlation in ML

CSI is a matrix of complex values of signal subcarriers. These values are samples of signals of subcarrier that are part of the OFDM modulation and describe the amplitude and phase of the signal subcarrier at a particular time. We use two routers, each with two antennas. Thus, the signal is transmitted from the receiver to the transmitter along four paths (Fig. 12.1). On each of these paths, the signal consists of 56 subcarriers, which are located at adjacent frequencies. Ultimately, $2 \cdot 2 \cdot 56 = 224$ features can be used to train and test machine learning models. Before processing, each data packet is concatenated as an array as follows: $h_{11} + h_{12} + h_{21} + h_{22} +$ «target value», where, for each path, the subcarriers are in the correct order of increasing frequency.

In this paper, only the absolute values of complex amplitudes of subcarriers are used because they provide a better machine learning result than the phase values of the signal subcarriers [15]. We have considered how the subcarriers of a complex signal (hereinafter referred to as their amplitudes) correlate with each other. To do this, we construct a pairwise correlation matrix of signal subcarriers (Fig. 12.2). To avoid small scale, we only use some subcarriers from the combined array of 224 features, including edge subcarriers for each «h» signal path.

Fig. 12.1 Wi Fi signal transmission scheme. h11, h12, h21 and h22 are the signal paths
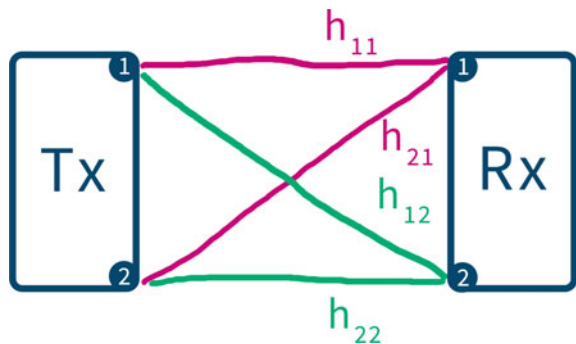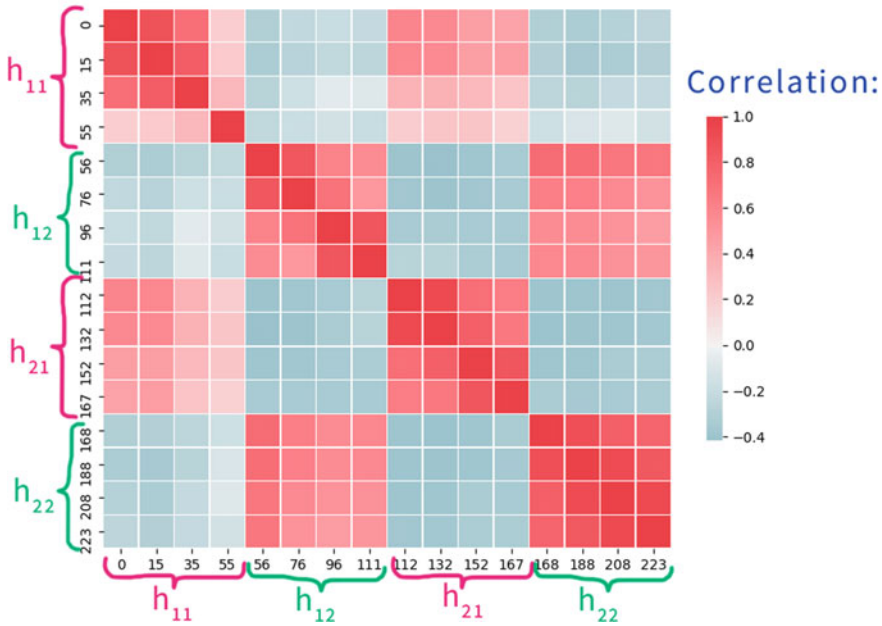
Fig. 12.2   Correlation matrix of subcarriers (amplitude) in combined feature array

For clarity, the correlation matrix in the Fig. 12.2 added boundary subcarriers belonging to different paths. It is noticeable that the paths with a common receiver (Rx) have a good correlation with each other. We can conclude that CSI shape is most influenced by the parameters of the receiving antennas (Fig. 12.1).

In Fig. 12.2 every cell is a correlation between subcarriers (two columns of the corresponding subcarriers in the dataset). Subcarriers not included in the correlation matrix have correlation values close to those of their neighbors from the small square (red or blue) in which they are located in Fig. 12.2.

With such a large number of features for machine learning, it can be assumed that some of these features are redundant. To test this hypothesis, a graph of the cumulative sum of the importance of the features was built (Fig. 12.3) where the stochastic gradient descent is used to assess the feature importance [8].

The cumulative importance in Fig. 12.3 does not go into saturation until absolutely all signs are summed up. Among all the features there are no ones that could be eliminated without deteriorating the classification accuracy. That is, all 224 possible features contain useful unique information for the machine learning method.
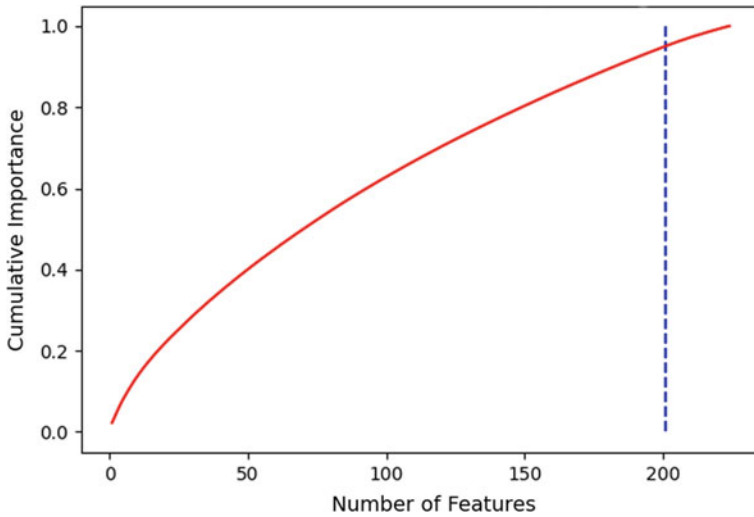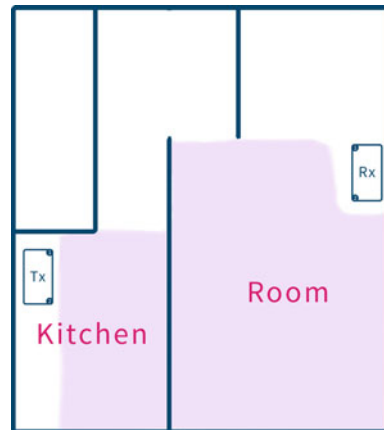
With such a small number of training samples, excellent classification results using machine learning methods cannot be expected. However, getting many CSI samples is a difficult task, because with a large distance between Rx and Tx, not all CSI packets arrive at the Rx. Also, when processing the recorded data, many packets that were sent to the receiver, but are not special packets sent by Tx, are filtered out.

Thus, our task is to classify a person's location using machine learning. Since the classification is binary, the metric of the quality of training models is accuracy (the probability of correct classification). And error is the probability of wrong classification. In this case, binary classification means that only two options are possible—either the person is in the kitchen or in the room.

12.4 CNN Classification

When classifying with CNN, we used the ability to represent one data packet (224 subcarriers) as a two-dimensional array for classification. Recall that we have 4 signal paths with 56 subcarriers each. By default, when sending them to the model for training, we arranged them in a row. But when fed to a convolutional neural network, the 4 paths will be located under each other, so that the subcarriers having the same frequencies are under each other (Fig. 12.5).

The training was carried out using the Keras [11] and TensorFlow [12] libraries in Python programming language. CNN structure is shown in Fig. 12.6. For intermediate layers of the "Dense" type, the "relu" activation function is used. The structure of the convolutional neural network was chosen empirically. Note that dropout layers are often found in the CNN structure. This is necessary for more stable training and repeatability of the classification accuracy.

When comparing the classification results, it is noticeable that CNN compares favorably with other simpler machine learning methods (Table 12.1). This is because CNN studies the data deeper during training, which allows it to see patterns in the data that are hidden from others by the method. In addition, this machine learning method is designed to work with a large number of features.
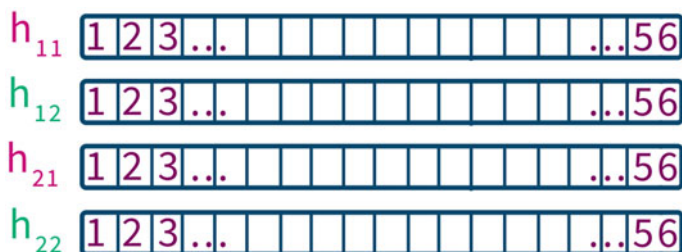

Fig. 12.5 The image formed from every CSI packet (amplitude) for the CNN

```
Layer (type)                    Output Shape              Param #
=================================================================
conv2d (Conv2D)                 (None, 4, 56, 28)            280

dropout (Dropout)               (None, 4, 56, 28)              0

max_pooling2d (MaxPooling2D)    (None, 2, 28, 28)              0

dropout_1 (Dropout)             (None, 2, 28, 28)              0

conv2d_1 (Conv2D)               (None, 2, 28, 12)           3036

dropout_2 (Dropout)             (None, 2, 28, 12)              0

flatten (Flatten)               (None, 672)                    0

dense (Dense)                   (None, 80)                 53840

dropout_3 (Dropout)             (None, 80)                     0

dense_1 (Dense)                 (None, 40)                  3240

dropout_4 (Dropout)             (None, 40)                     0

dense_2 (Dense)                 (None, 2)                     82
=================================================================
```

Fig. 12.6 CNN structure in Keras

Table 12.1 Classification results with different ML methods for human position in apartment

ML method	Accuracy, %	Training time, seconds
CNN	82.5	55.9
Support vector classifier	76.7	8.1
Random forest	76.0	6.6
Adaptive boosting	74.0	8.4
Gaussian naive bayes	68.3	0.1
K-nearest neighbors	64.2	2.6
Logistic regression	62.4	20.8

The fact that CNN showed such good classification results for difficult conditions (little data for training, large distance between routers, different location of a person) compared to other machine learning methods makes it the best choice for this experiment.

12.5 Conclusion

In this article, we looked at the correlation and cumulative importance of features for Channel State Information. The main conclusions are that each feature is important for a successful classification, and the shape of the received signal strongly depends on the individual characteristics of the receiver antennas, which can be seen in Figs. 12.1 and 12.2.

Also, one of the features of the experiment is the use of ordinary home routers operating according to the Wi-Fi 4 standard on the OpenWrt firmware because this set is the most accessible and easy to customize.

With the help of CNN, it was possible to achieve sufficient accuracy in classifying the location of a person in an apartment. In similar results of articles by other authors, the accuracy is higher [5], however, in our article we used a small sample size for training, which led to a decrease in accuracy. It was part of an experiment to be able to do the classification under these constraints. You can find the data and code used to carry out the classification in the GitHub repository of this project [13].

References

1. V.A. Muliukha, L.U. Laboshin, A.A. Lukashin, N.V. Nashivochnikov, Analysis and classification of encrypted network traffic using machine learning. in *2020 XXIII International Conference on Soft Computing and Measurements* (*SCM*) May. (IEEE, 2020), pp. 194–197
2. M. Morshedi, J. Noll, A survey on prediction of PQoS using machine learning on Wi-Fi networks. in *2020 International Conference on Advanced Technologies for Communications* (*ATC*). (IEEE, 2020), October), pp. 5–11. https://doi.org/10.1109/ATC50776.2020.9255457
3. L. Utkin, P. Drobintsev, M. Kovalev, A. Konstantinov, Combining an autoencoder and a variational autoencoder for explaining the machine learning model predictions. in *2021 28th Conference of Open Innovations Association* (*FRUCT*) January (IEEE, 2021), pp. 489–494. https://doi.org/10.23919/FRUCT50888.2021.9347612
4. A. Dziedzic, V. Sathya, M.I. Rochman, M. Ghosh, S. Krishnan, Machine learning enabled spectrum sharing in dense LTE-U/Wi-Fi coexistence scenarios. IEEE Open J. Veh. Technol. **1**, 173–189 (2020)
5. S. Gïney, A. Erdoğan, M. Aktaş, M. Ergün, Wi-Fi based indoor positioning system with using deep neural network. in *2020 43rd International Conference on Telecommunications and Signal Processing* (*TSP*) July, (IEEE, 2020), pp. 225–228. https://doi.org/10.1109/TSP49548.2020.9163548
6. X. Xie, Z. Guo, A CNN approach of activity recognition via channel state information (CSI). in *2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge*

Discovery (CyberC) October. (IEEE, 2019), pp. 329–336. https://doi.org/10.1109/CyberC.2019.00063

7. Q. Duan, X. Wei, J. Fan, L. Yu, Y. Hu, CNN-based intrusion classification for IEEE 802.11 wireless networks. in *2020 IEEE 6th International Conference on Computer and Communications* (*ICCC*). (IEEE, 2020, December), pp. 830–833. https://doi.org/10.1109/ICCC51575.2020.9345293

8. Repository of feature selector (2021). https://github.com/WillKoehrsen/feature-selector. Last Accessed 09 March 2021

9. H. Li, X. Zeng, Y. Li, S. Zhou, J. Wang, Convolutional neural networks based indoor Wi-Fi localization with a novel kind of CSI images. China Commun. **16**(9), 250–260 (2019). https://doi.org/10.23919/JCC.2019.09.019

10. K.M. Chen, R.Y. Chang, S.J. Liu, Interpreting convolutional neural networks for device-free Wi-Fi fingerprinting indoor localization via information visualization. IEEE Access **7**, 172156–172166 (2019). https://doi.org/10.1109/ACCESS.2019.2956187

11. Official Keras website, https://keras.io/. Last Accessed March 09 2021

12. Official TensorFlow website, https://www.tensorflow.org/. March 09 2021

13. GitHub repository of the paper, https://github.com/maksimio/csi_classification. Last Accessed March 09 2021

14. T. Zhang, T. Song, D. Chen, T. Zhang, J. Zhuang, Wigrus: a wifi-based gesture recognition system using software-defined radio. IEEE Access **7**, 131102–131113 (2019). https://doi.org/10.1109/ACCESS.2019.2940386

15. M. Won, S. Sahu, K.J. Park, DeepWiTraffic: low cost WiFi-based traffic monitoring system using deep learning. in *2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems* (*MASS*). (IEEE. 2019), pp. 476–484. https://doi.org/10.1109/MASS.2019.00062.

Chapter 13
Application of Top-Down Deconvolution Tools to Bottom-Up Mass Spectrometry Data: Advantages and Limitations

Kirill Tyshuk and Kira Vyatkina

Abstract High-resolution mass spectrometry is a powerful technique for the analysis of proteins and peptides enhancing the capabilities for de novo sequencing. Prior to interpretation, top-down mass spectra acquired from intact proteins should be deconvoluted; in contrast, this step is often omitted when processing bottom-up spectra acquired from peptides. However, it turned out that application of top-down deconvolution tools to high-resolution bottom-up tandem (MS/MS) mass spectra can be strongly beneficial, since they filter out many unreliable peaks, leaving only the ones particularly likely corresponding to true fragment ions. In this work, we examine behavior of three popular top-down deconvolution tools—MS-Deconv, Xtract, and Hardklör—on high-resolution bottom-up mass spectra acquired from tryptic peptides of carbonic anhydrase 2 (CAH2). We verify how many isotopic clusters present in the input spectra are missed by each tool, and how many peaks in the deconvoluted spectra correspond to fragment ions, compare the obtained results, and indicate advantages and limitations of applying those tools to bottom-up data.

Keywords Computational Proteomics · High-Resolution Mass Spectrometry · Spectral Deconvolution

13.1 Introduction

Mass spectrometry (MS) has established itself as a method of choice for analyzing proteins [1, 2]. The two key strategies applied to this end are bottom-up and top-down MS [3, 4]. The former requires investigated proteins to be proteolytically

K. Tyshuk · K. Vyatkina (✉)
Saint Petersburg State University, 7-9 Universitetskaya Emb, St. Petersburg 199034, Russia
e-mail: vyatkina@spbau.ru

K. Vyatkina
Alferov University, 8/3 Khlopina St, St. Petersburg 194021, Russia

Saint Petersburg Electrotechnical University "LETI", 5 ul. Professora Popova, St. Petersburg 197376, Russia

digested into peptides, while the latter analyses intact proteins. Since the advent of high-resolution orbitrap instruments [5] accessible to many laboratories, top-down proteomics is becoming more and more widespread [6–8]; however, the bottom-up approach can also benefit substantially from high resolution [9–12].

Prior to interpretation, top-down mass spectra need to be deconvoluted; the most popular deconvolution software tools include Xtract [13], MS-Deconv [14], and Hardklör [15, 16]. On the contrary, bottom-up spectra typically are not deconvoluted. Yet it has recently been demonstrated that upon deconvolution with the above-mentioned algorithms, high-resolution bottom-up mass spectra can be effectively de novo sequenced in a top-down like fashion using the Twister method [17–20].

To make this more precise, top-down deconvolution acts as extremely selective filtration, leaving only the most reliable peaks while eliminating all the rest. Typically, a deconvoluted spectrum contains at most thirty peaks, and many spectra lose *all* the peaks. As a consequence, deconvoluted tandem (MS/MS) mass spectra can no longer be interpreted on an individual basis but can be successfully processed in a bunch.

However, the criteria applied by top-down deconvolution algorithms to distinguish between high- and low-quality isotopic envelopes are not perfectly suitable for the bottom-up case, and some "good" isotopic envelopes of fragment ions with small m/z values are discarded just because they differ in shape from the ones expected to occur in a top-down spectrum. Thus, it is desirable to adjust those approaches for processing bottom-up data, still preserving their selectivity.

As a first step, we need to find out how much information contained in the raw spectra in m/z space is lost upon deconvolution and examine the "good" isotopic envelopes that get discarded to determine how the scoring functions used by the algorithms should be adjusted in order to avoid their elimination.

In this work, we examine high-resolution MS/MS spectra acquired from tryptic peptides of carbonic anhydrase 2 (CAH2), along with their deconvoluted counterparts obtained using MS-Deconv, Xtract and Hardklör. We verify how many isotopic clusters are missed by those tools, and how many of the retained peaks correspond to fragment ions, and indicate advantages and limitations of applying them to high-resolution bottom-up data.

13.2 Materials and Methods

13.2.1 Dataset

In our experiments, we used the "tryptic" dataset for CAH2 published in [8], which consists of 47,536 tryptic MS/MS spectra contained in the files 140411_QE_Cah-1.raw (18,707 MS/MS spectra), 140411_QE_Cah-2.raw (21,230 MS/MS spectra), 140411_QE_Cah-3.raw (3377 MS/MS spectra), and 140411_QE_Cah-4.raw (4222 MS/MS spectra). In brief, the respective measurements were performed on a nano LC system (Ultimate 3000; Thermo Fisher Scientific, Gemeringen, Germany) coupled

online to a Q-Exactive Plus Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). MS and MS/MS spectra were acquired at a resolution of 70,000 and 17,500, respectively; MS/MS spectra were obtained by HCD fragmentation. Further details can be found in [20].

13.2.2   Deconvolution

As in [20], the raw bottom-up MS/MS spectra were centroided and converted into mzXML format with ReAdW 4.3.1, and subsequently deconvoluted using MS-Deconv v. 0.8.07370 with the default parameters (maximum charge state: 30; maximum monoisotopic mass of fragment ions: 49,000 Da; signal-to-noise ratio: 1; envelopes of precursor ions were deconvoluted to derive the precursor masses of MS/MS spectra).

Thermo Xtract 3.0 was also applied to the raw data with its default top-down parameters, including Low and High Mass of 56.00 and 60,000.00 and signal-to-noise ratio of 2, except for the maximum charge state, which was set to 10.

Hardklör was run with the default parameters as well, except for 'resolution', 'centroided' and 'ms_level' set to 17,500, 1 = yes and 2, respectively.

13.2.3   Peptide Identification

MS/MS spectra were identified through a search with MS-GF + [21, 22] against a database containing the sequences of CAH2 and the 32 contaminants listed in [20] with the default parameters except for the maximum peptide length and maximum precursor charge, which were set to 50 and 5, respectively.

13.3   Results

13.3.1   Test Datasets

To compare the deconvolution software tools MS-Align, Xtract and Hardklör, we considered separately the four sets of MS/MS spectra from the files 140411_QE_Cah-1.raw, 140411_QE_Cah-2.raw, 140411_QE_Cah-3.raw, and 140411_QE_Cah-4.raw, and selected from each of those the mass spectra identified with the E-value below 10^{-15}. Further, we annotated those spectra with respect to the peptides they were matched to and kept only the ones for which the peaks identified as monoisotopic *b*- or *y*-ions of charge at most 5 provided the peptide sequence coverage of at

least 25%. In this way, we obtained four high-quality datasets, which will be further numbered from 1 to 4.

In Table 13.1, the number of spectra in each dataset is indicated, along with the average number of peaks in those, and the average sequence coverage (in percent) provided by monoisotopic *b*- or *y*-ions of charge at most 5.

Subsequently, we checked whether those spectra were properly deconvoluted by the three tools under consideration. It turned out that while MS-Deconv and Hardklör processed all the spectra so that the deconvoluted ones did contain at least one peak, Xtract failed to process some of those, which led to deconvoluted spectra with no peaks. And namely, Xtract could not handle 4, 9, 18 and 18 spectra from the first, second, third and fourth dataset, respectively. Statistics on those spectra is given in Table 13.2; it indicates that their quality matches quite good the average for the test datasets.

To perform a fair comparison between the deconvolution tools, we excluded from the test datasets the spectra that were not appropriately handled by Xtract. For the sake of completeness, in Table 13.3, we provide statistics on the final datasets used for comparison of MS-Deconv, Xtract and Hardklör; as expected, it is much similar to the one given in Table 13.1.

Table 13.1 Composition of the four test datasets

Dataset	Number of the selected spectra	Average number of peaks	Average sequence coverage with monoisotopic *b*- and *y*-ions of charge at most 5, %
1	62	471	53.56
2	91	462	38.31
3	30	548	46.72
4	39	598	33.36

Table 13.2 Statistics on the spectra that could not be properly deconvoluted with Xtract

Dataset	Number of the mishandled spectra	Average number of peaks	Average sequence coverage with monoisotopic *b*- and *y*-ions of charge at most 5, %
1	4	298	42.25
2	9	484	34.59
3	18	537	50.92
4	18	519	35.90

Table 13.3 Composition of the final datasets used to compare the performance of MS-Deconv, Xtract and Hardklör

Dataset	Number of the mishandled spectra	Average number of peaks	Average sequence coverage with monoisotopic b- and y-ions of charge at most 5, %
1	58	482	54.30
2	82	459	38.85
3	12	561	41.73
4	21	667	31.13

13.3.2   Comparison of Deconvolution Tools

The deconvolution software tools MS-Align, Xtract and Hardklör were tested on the four datasets, statistics on which is provided in Table 13.3. The obtained results are summarized in Table 13.4.

As it can be seen, Xtract eliminates almost all b- and y-ions. Interestingly, both MS-Deconv and Hardklör manage to obtain higher sequence coverage with monoisotopic N- and C-terminal fragment masses, as compared to that with b-and y-ions of charge at most 5. In case of MS-Deconv, this can be explained by the presence of some b-and y-ions of charge more than 5 in the spectra, while the more impressive difference in sequence coverage achieved by Hardklör should be due to thorough analysis of

Table 13.4 Composition of the final datasets used to compare the performance of MS-Deconv, Xtract and Hardklör

Dataset	Deconvolution tool	Average number of peaks	Average sequence coverage with monoisotopic N- and C-terminal fragment masses, %
1	MS-Deconv	28	47.44
	Xtract	16	0.08
	Hardklör	119	75.94
2	MS-Deconv	29	55.82
	Xtract	45	0.46
	Hardklör	120	76.38
3	MS-Deconv	20	40.46
	Xtract	29	1.82
	Hardklör	50	62.07
4	MS-Deconv	24	39.67
	Xtract	34	0.27
	Hardklör	63	62.30

dependencies between the fragment ions, which allows to retrieve the respective monoisotopic masses also from modified ions.

Thus, processing the raw spectra with Hardklör results in deconvoluted spectra with higher sequence coverage with monoisotopic N- and C-terminal fragment masses, while MS-Deconv produces deconvoluted spectra with sequence coverage being approximately the same as with *b*- and *y*-ions in the raw spectra. This suggests that if the spectra should be further handled on an individual basis, it might be beneficial to use Hardklör; however, if they are processed together, as Twister does, MS-Deconv will be likely the choice of preference, as demonstrated in [20]. And this is also consistent with the fact that even though Hardklör is applied for deconvoluting top-down data, it was originally developed for processing bottom-up spectra [15].

13.3.3   Detection of Isotopic Envelopes by MS-Deconv

Despite MS-Deconv sometimes slightly improves sequence coverage as compared to that provided by the raw spectra, it does not recognize several isotopic envelopes, because the scoring function used to compare a candidate envelope to the theoretical one was suggested for the top-down case, in which isotopic envelopes have a different kind of shape, as compared to the bottom-up case. At the same time, experimental isotopic envelopes closely matching theoretical ones are usually properly processed by MS-Deconv. Examples of a high- and modest-quality matches between an experimental and theoretical isotopic envelope are given in Figs. 13.1 and 13.2, respectively; both these experimental envelopes were detected and processed by MS-Deconv. The theoretical envelopes for the respective monoisotopic masses were generated using Emass [23]. Visualization was performed with a software tool we developed to this end.

To estimate the quality of an experimental envelope, a scoring function should be introduced, which would allow comparing it to its theoretical counterpart. Such function employed by MS-Deconv is described in [14]; in particular, it accounts for discrepancies in both the *m/z* values of the corresponding peaks and intensities of the latter. Our preliminary computational experiments suggest that the same concept is applicable to the bottom-up case; however, the deviation of the observed and predicted intensities should be penalized heavier than in the top-down case, as well as the absence of peaks supposed to have large enough intensity.

13.4   Discussion

While deconvolution traditionally is not applied to bottom-up data, it can be strongly beneficial to deconvolute high-resolution bottom-up MS/MS spectra in a top-down fashion. In this way, we can obtain mass spectra with very few yet much reliable
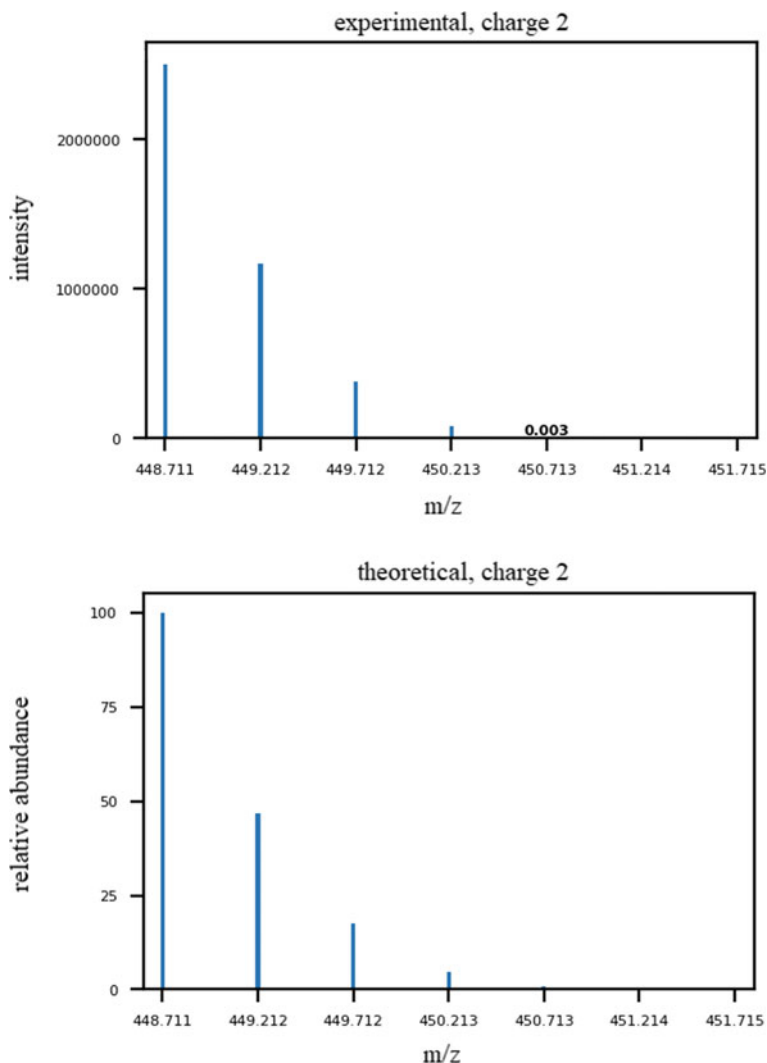
Fig. 13.1 A high-quality match between an experimental and theoretical envelope (charge 2)

peaks. Subsequently, those spectra should be processed all together rather than indi-vidually, and in this way, particularly accurate interpretation of the MS/MS data can be achieved. This is especially important for the task of de novo sequencing.

At the same time, adaptation of the existing top-down deconvolution software tools to the bottom-up case would clearly improve their performance. However, they can be more or less suitable for such adjustment: as our computational exper-iments demonstrated, Xtract—arguably the most popular top-down deconvolution
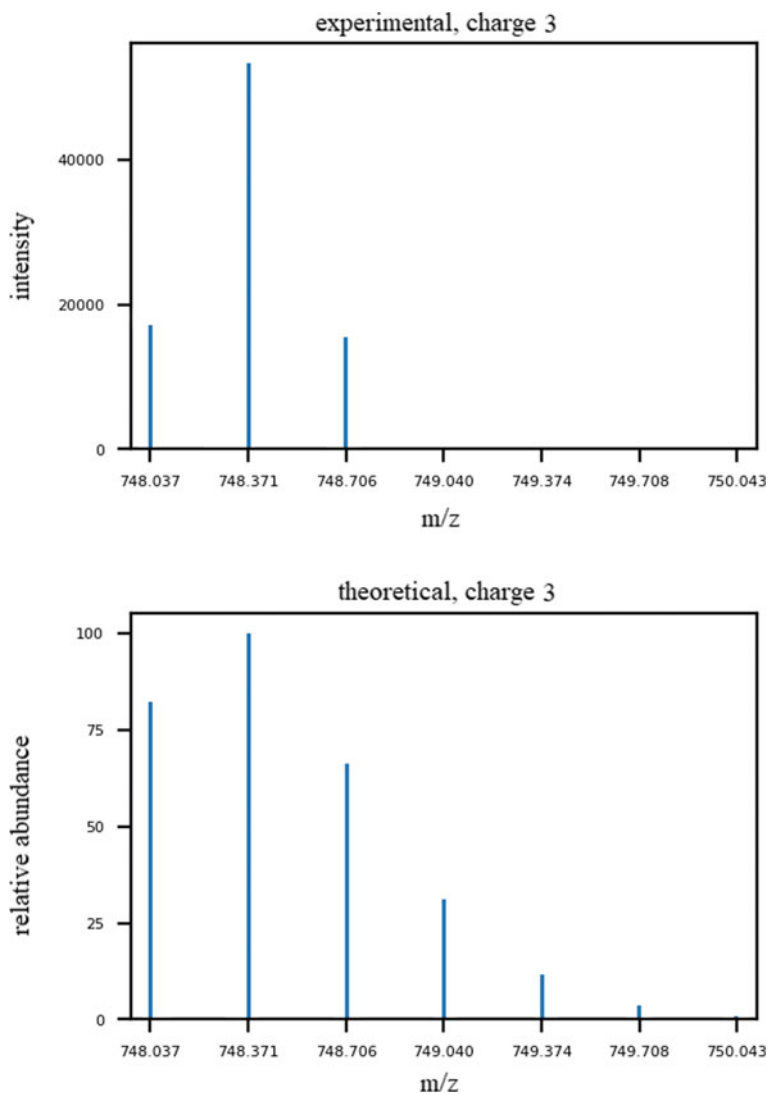
Fig. 13.2 A modest-quality match between an experimental and theoretical envelope (charge 3)

method—performs quite poor on the bottom-up data, while the results of MS-Deconv look very promising and will likely be further improved through appropriate modification of its scoring function.

Having developed efficient and reliable methods for deconvoluting high-resolution bottom-up spectra in a top-down way, we will further come up with novel algorithms for interpretating this kind of data, this gaining new opportunities for analysis of peptides and proteins.

13.5 Conclusions

In this work, we demonstrated that among the most widely used top-down deconvolution tools, MS-Deconv is the most promising candidate for adaptation to the case of high-resolution bottom-up tandem mass spectra. Our future research will be dedicated to adjusting the scoring function it uses to detect isotopic envelopes, so that to improve its performance in this unforeseen particular case.

References

1. T.C. Walther, M. Mann, Mass spectrometry-based proteomics in cell biology. J. Cell Biol. **190**(4), 491–500 (2010)
2. Y. Zhang, B.R. Fonslow, B. Shan, M.-C. Baek, J.R. Yates, Protein analysis by shotgun/bottom-up proteomics. Chem. Rev. **113**(4), 2343–2394 (2013)
3. N.L. Kelleher, H.Y. Lin, G.A. Valaskovic, G.A. Aaserud, E.K. Fridriksson, F.W. McLafferty, Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. J. Am. Chem. Soc. **121**(4), 806–812 (1999)
4. B.T. Chait, Mass spectrometry: Bottom-up or top-down? Science **314**(5796), 65–66 (2006)
5. Q. Hu, H. Li, A. Makarov, M. Hardman, R.G. Cooks, The Orbitrap: a new mass spectrometer. J. Mass Spectrom. **40**(4), 430–433 (2005)
6. T.K. Toby, L. Fornelli, N.L. Kelleher, Progress in top-down proteomics and the analysis of proteoforms. Annu. Rev. Anal. Chem. **9**, 499–519 (2016)
7. D.P. Donnelly, C.M. Rawlins, C.J. DeHart et al., Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. Nat. Methods **16**, 587–594 (2019)
8. K.A. Cupp-Sutton, S. Wu, High-throughput quantitative top-down proteomics. Mol. Omics **16**(2), 91–99 (2020)
9. A.M. Frank, M.M. Savitski, M.L. Nielsen, R.A. Zubarev, P.A. Pevzner, De novo peptide sequencing and identification with precision mass spectrometry. J. Proteome Res. **6**(1), 114–123 (2007)
10. M. Mann, N.L. Kelleher, Precision proteomics: the case for high resolution and high mass accuracy. PNAS **105**(47), 18132–18138 (2008)
11. C. Pan, B. Park, W. McDonald, P. Carey, J. Banfield, N. VerBerkmoes, R. Hettich, N. Samatova, A high-throughput de novo sequencing approach for shotgun proteomics using high-resolution tandem mass spectrometry. BMC Bioinf. **11**, 118 (2010)
12. M.M. Savitski, T. Mathieson, I. Becher, M. Bantscheff, H-Score, a mass accuracy driven rescoring approach for improved peptide identification in modification rich samples. J. Proteome Res. **9**(11), 5511–5516 (2010)
13. D.M. Horn, R.A. Zubarev, F.W. McLafferty, Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. J. Am. Soc. Mass Spectrom. **11**(4), 330–332 (2000)
14. X. Liu, Y. Inbar, P.C. Dorrestein, C. Wynne, N. Edwards, P. Souda, J.P. Whitelegge, V. Bafna, P.A. Pevzner, Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. Mol. Cell. Proteomics **9**(12), 2772–2782 (2010)

15. M.R. Hoopmann, G.L. Finney, M.J. MacCoss, High-speed data reduction, feature detection, and MS/MS spectrum quality assessment of shotgun proteomics data sets using high-resolution mass spectrometry. Anal. Chem. **79**(15), 5620–5632 (2007)
16. M.R. Hoopmann, M.J. MacCoss, R.L. Moritz, Identification of peptide features in precursor spectra using Hardklör and Krönik. Curr. Protoc. Bioinfo. Chap.**13**, Unit13.18 (2012)
17. K. Vyatkina, S. Wu, L.J.M. Dekker, M.M. VanDuijn, X. Liu, N. Tolić, M. Dvorkin, S. Alexandrova, T.M. Luider, L. Paša-Tolić, P.A. Pevzner, De novo sequencing of peptides from top-down tandem mass spectra. J. Proteome Res. **14**(11), 4450–4462 (2015)
18. K. Vyatkina, S. Wu, L.J.M. Dekker, M.M. VanDuijn, X. Liu. N. Tolić, T.M. Luider, L. Paša-Tolić, P.A. Pevzner, Top-down analysis of protein samples by de novo sequencing techniques. Bioinformatics **32**(18), 2753–2759 (2016)
19. K. Vyatkina, De novo sequencing of top-down tandem mass spectra: a next step towards retrieving a complete protein sequence. Proteomes **5**(1), 6 (2017)
20. K. Vyatkina, L.J.M. Dekker, S. Wu, M.M. VanDuijn, X. Liu., N. Tolić, T.M. Luider, L. Paša-Tolić, P.A. Pevzner, De novo sequencing of peptides from high-resolution bottom-up tandem mass spectra using top-down intended methods. Proteomics **17**(23–24) (2017)
21. S. Kim, N. Gupta, P. Pevzner, Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. J. Proteome Res. **7**(8), 3354–3363 (2008)
22. S. Kim, P. Pevzner, MS-GF+ makes progress towards a universal database search tool for proteomics. Nat. Commun. **5**, 5277 (2014)
23. A.L. Rockwood, P. Haimi, Efficient calculation of accurate masses of isotopic peaks. J. Am. Soc. Mass Spectrom. **17**, 415–419 (2006)

Chapter 14
Algorithm for Fetal Activity Passive Monitoring

Check for updates

Yulia O. Bobrova, Olga N. Kapranova, and Kseniya V. Filipenko

Abstract Fetal activity is an important indicator of fetal health. Analysis of the number and duration of fetal movement episodes can provide conclusions about fetal development and its general health. However, short-term observation of these indicators can give a false result due to the changing behavior of fetal activity, depending on the stage of pregnancy. Long-term monitoring will make it possible to timely record the initial stage of pathology and provide timely assistance to the mother and fetus. The purpose of this work is to compare passive recorders: accelerometers and gyroscopes, which can be used as part of a wearable long-term monitoring system. The paper presents an algorithm for identifying episodes of fetal activity to facilitate the analysis of fetal condition and reduce the subjectivity of processing the signal recorded by the system. According to the results of the work, the average sensitivity and average accuracy of the algorithm for identifying episodes of fetal activity based on the gyroscope signal was 90%. At the same time, the average sensitivity of identifying episodes of fetal activity based on the accelerometer signal was 87.50% (calm state) and 79.16% (active state), average accuracy was 87.50% (calm state) and 71.20% (active state).

14.1 Introduction

Today, taking care of the health of the pregnant woman and the fetus is a public health priority in most countries. Technologies in the field of perinatal diagnostics are developing rapidly, and therapeutic approaches are constantly being improved. However, the risk of developing critical fetal conditions and related pathologies, which can subsequently lead to childhood disability or death, is still high [1]. The percentage of children with disabilities aged 0–3 years is 7.59% of the total child disability. In

Y. O. Bobrova (✉) · O. N. Kapranova · K. V. Filipenko
Saint Petersburg Electrotechnical University "LETI", St. Petersburg, Russia
e-mail: jobobrova@etu.ru

2020, this amounted to 53.5 thousand children [2]. Long-term monitoring and, as a result, timely recognition of critical conditions of the fetus will reduce the number of perinatal deaths and ensure the health of the child. The use of remote monitoring systems allows a pregnant woman to reduce the frequency of hospital visits. This allows her to maintain her usual lifestyle and reduce the likelihood of infectious diseases. In addition, in conditions when a woman is not able to visit a medical institution due to its remoteness or the complexity of the epidemiological situation, remote monitoring will allow the doctor to continue monitoring her condition and the condition of the fetus.

Currently, the main method for monitoring the state of the fetus is cardiotocography (CTG). This method is highly accurate, but the complexity of the equipment and the limited time of the study do not allow its use in long-term monitoring systems. The simplest method of controlling fetal movements is maternal perception. However, this method is subjective; its sensitivity varies from 37 to 88% [3]. Alternative passive methods for recording and analyzing the fetal activity signal described in the literature are:

- Fetal electrocardiography (fECG). This method allows for non-invasive, long-term and safe monitoring using compact devices. From the recorded data it is possible to produce the fetal VCG, which makes it possible to obtain information on fetal movements. However, the amplitude of the registered maternal ECG signal is significantly greater than the amplitude of the fetal ECG. Therefore, to isolate the fetal ECG, it is necessary to use complex signal processing algorithms [4].
- Fetal magnetocardiography (MCG) [5]. Compared to fetal ECG, the signal-to-noise ratio in fetal ICG is significantly higher. Therefore, the features required for the analysis can be identified with greater accuracy. However, the high cost, the need for specialized and complex equipment makes this method unsuitable for long-term remote monitoring.
- Acoustic sensors. Such systems are characterized by low cost and availability. Acoustic sensors. Such systems are characterized by low cost and availability. But they are inferior in accuracy to other methods. For example, system [6] is successfully able to detect startle movements, but is not sensitive to breathing or general movements
- The use of various transducers such as accelerometers [7, 8], force-sensitive resistors [9], gyroscopes is a promising technology for monitoring fetal activity. Currently, there are no uniform accepted methods for using certain types of sensors, their number and relative position. There is also the task of developing algorithms for processing the received signals to isolate episodes of fetal activity against the background of non-fetal artifacts. Passive systems based on recording fetal impact on the anterior abdominal wall have the potential to accurately detect fetal movements in real time.

14.2   Proposed System for Fetal Activity Registration

In [10], we proposed a system based on a matrix of passive recorders. This system records the physical activity of the fetus. Fetal activity is a vivid marker of the state of the fetus: its nervous development, as well as the presence or absence of hypoxia [11]. Figure 14.1 shows the layout of the system recorders on the patient.

The system includes abdominal sensors (Pass_Rec_1–Pass_Rec_3). They record the signal of fetal activity and the noise signal associated with the movements of the mother and various artifacts. In addition, the proposed system has a reference sensor (Pass_Rec_0). It registers only the noise signal and is used during processing to

Fig. 14.1 Passive recorders layout
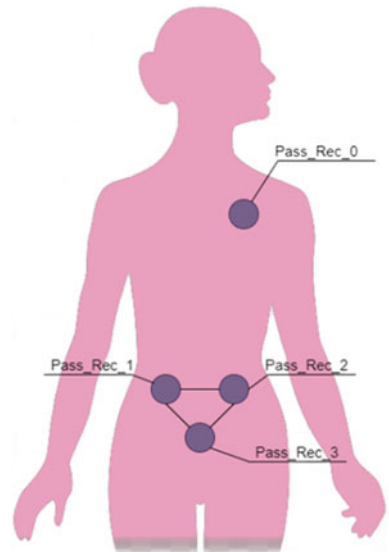
Fig. 14.2 Sensor registration axes
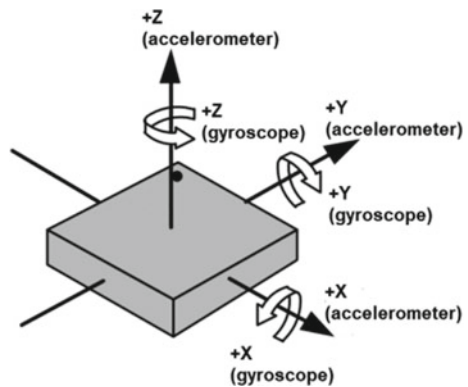
remove noise and artifacts from the abdominal signals. Figure 14.2 shows a diagram of the sensor registration axes [12].

In works [10, 13] we considered the use of accelerometers for recording the signal of fetal activity. The accelerometer records the projection of the apparent acceleration along three axes (X, Y and Z). In this paper, we will take a look at the use of gyroscopes. Gyroscopes measure the change in the angle of orientation of the body (sensor), which also allows the determination of fetal movement.

14.3 Fetal Activity Signal Simulation

14.3.1 Signal Simulation Model

We have developed a model of the fetal activity signal source [10]. The model is necessary to confirm the correctness of the choice of recorders, to analyze their best relative position and, thereby, to increase the efficiency of the system being developed. The layout takes into account the biomechanical features of signal propagation in a real biological environment. For this, it includes viscoelastic structures similar in their mechanical characteristics to real biological tissues. In addition, the model takes into account the morphological shape of the surface, to which recorders will be applied in the future. Figure 14.3 shows a diagram of the structure of the model in section.

Figure 14.3 shows the following components of the layout: 1 is the outer layer of the layout, made of solid silicone, which simulates the outer layers of the skin, from the surface of this layer the signal is recorded by sensors; 2 is viscous inner layer of the model, made of viscous silicone, which simulates a visco-elastic environment through which a real signal passes; 3 points motors that generate a useful signal and noise (motor in the center). The signal amplitude on the model surface depends on the number of motors involved in generating it, their relative position and the duration of the pulse, which controls the operating time of the motors.

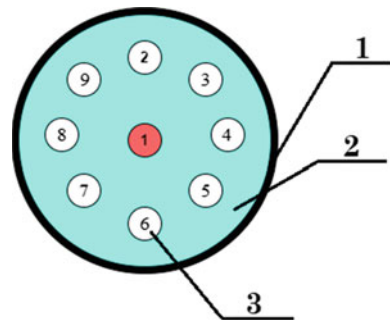**Fig. 14.3** Cross-sectional diagram of the signal source model

Fig. 14.4 Amplitude of the recorded signal vs pulse duration of accelerometer (blue) and gyroscope (red)
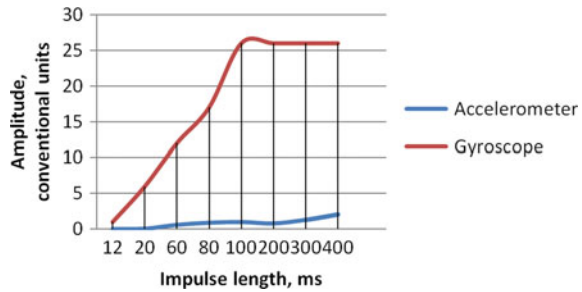
Figure 14.4 shows a graph of the dependences of the amplitude of the signal recorded from the surface of the model by the gyroscope and accelerometer on the duration of the control pulse.

Figure 14.4 shows that there is a strong dependence of the amplitude of the signal recorded from the surface of the model by the gyroscope on the duration of the control pulse. Over a significant interval (until the sensitivity threshold is reached), this dependence is linear. At the same time, the dependence of the signal amplitude on the pulse duration in the case of accelerometer measurements is also close to linear. But unlike a gyroscope, an accelerometer is less sensitive to changes in the input signal level.

Thus, the analysis of the gyroscope signal can make it possible to isolate the episodes of fetal movements more successfully, due to greater sensitivity to changes in the amplitude of the signal under investigation.

14.3.2   Signals Received from the Model

The layout presented in Sect. 3.1 allows generating a signal similar to the fetal activity signal in different amplitude ranges. The duration of the control pulse, which turns on a particular vibration motor, regulates the intensity of its work. This, in turn, affects the amplitude of the signal recorded on the surface of the model. Figure 14.5 shows the result of the study of the sensitivity of the recorders with the minimum duration of the control pulse (12 ms).

The results presented in Fig. 14.5 showed that with the same intensity of the signal generated by the model, the gyroscope is more sensitive than the accelerometer. On the x-axis of the gyroscope, the peaks corresponding to the moments of switching on the vibration motors are visually well distinguishable. At the same time, on the signal recorded by the accelerometer, it is almost impossible to determine the moments when the signal was sent. In addition, the signal recorded by the accelerometer has significantly lower amplitude than the signal recorded by the gyroscope, and is also more prone to baseline drift.

The results obtained during the study of the characteristics of the recorders using the signal source model showed that gyroscopes can also be used for passive recording
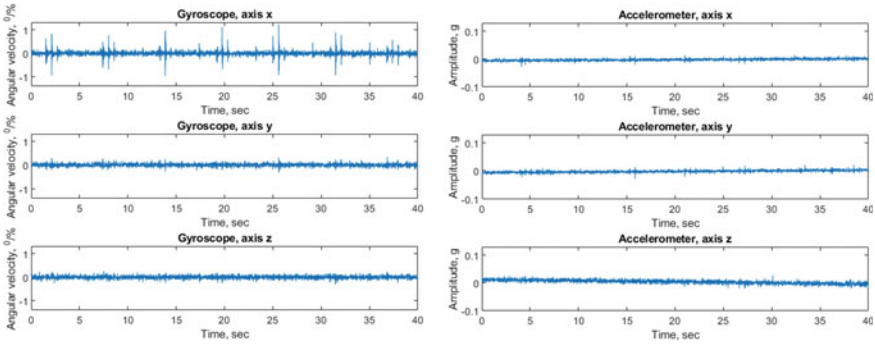
Fig. 14.5 Signals recorded along the axes of the gyroscope (left) and accelerometer (right)

of fetal motor activity, supplementing accelerometric recorders, or even replacing them altogether.

14.4   System Testing in Real Conditions

After confirming the possibility of signal registration by passive sensors using the developed model, we recorded real signals. We have obtained the signals from the accelerometers and gyroscopes. Registration was carried out in a calm state in a sitting position and in a walking condition.
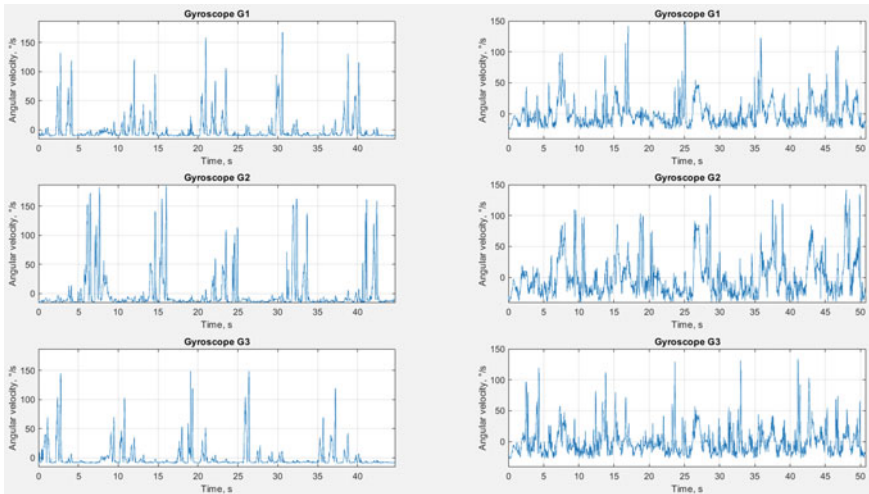
Fig. 14.6 Gyroscopes: signals received while sitting (left) and walking (right)
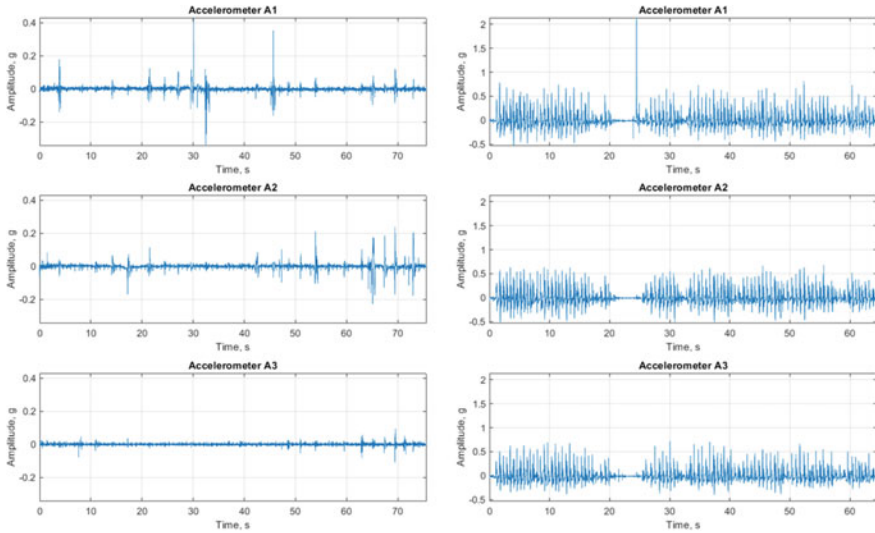
Fig. 14.7 Accelerometers: signals received while sitting (left) and walking (right)

Figure 14.6 shows the results of registration of movements in a sitting position and when walking by gyroscopes located in the abdominal region.

Figure 14.7 shows the signals received from abdominal accelerometers when registering movements in a sitting position and when walking.

As already mentioned, earlier we carried out work on the analysis of signals from accelerometric sensors. The developed signal processing algorithms and the results are given in the work [13]. In this paper, we will consider an algorithm for processing signals received from a gyroscope. They are similar to signals from an accelerometer, since signals from the same source are recorded. However, due to the different principles of sensor operation, the algorithms are slightly different.

Figure 14.8 shows the developed algorithm for processing signals from gyroscopes.

First of all, the received data is read, namely, signals from three three-axis abdominal gyroscopes. The resulting data can then be submitted in an easy to read format. To do so, we convert the measured angular velocity into standard units of measurement, namely in °/s. In our experiment, the maximum scale range was set equal to $\pm\ 125°/s$ A built-in 16-bit ADC was used to obtain data. In this case, the 250°/sec range is described by 2^{16} values. And the 1 /s $= 2^{16}/250 = 65{,}536/250 = 262.144$.

The received signal contains noises caused by the processes of the mother's vital activity. To exclude them from the signal, filtering is performed using a second-order Butterworth bandpass filter. The selected cutoff frequencies correspond to the frequencies of the fetal activity signal. Previous studies have shown that this signal is expected to be in the 0.5–10 Hz range [14].
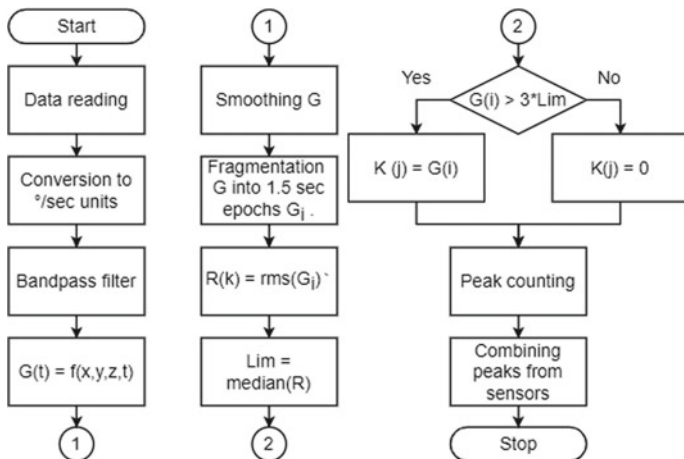
Fig. 14.8 Diagram of the signal processing algorithm

For further analysis, combine the signals from the three axes of each sensor into the resulting signals. For this, we use the Formula (14.1), where Где G_x, G_y and G_z are signals from the gyroscope axes.

$$G_i(t) = \sqrt{G_x^2(t) + G_y^2(t) + G_z^2(t)} \qquad (14.1)$$

The received signal is then smoothed and next we use the threshold method of extracting the peaks. As a threshold, the median of the root mean square values is calculated by epochs with a duration of 1.5 s. The one and a half second interval was chosen as the average duration of continued fetal movement. The counting of peaks on each of the sensors is carried out by comparison with the threshold value set as 3Lim. At the last stage of the algorithm, the results of motion detection for each sensor are combined. This is necessary to eliminate repetitive peaks detected simultaneously by two or more recorders [15, 16].

The result of the algorithm's work on the example of one of the signals is shown in Fig. 14.9.

The results of the algorithm performance for processing the recorded signals from abdominal region are considered in Table 14.1.

The accuracy is calculated using the Formula (14.2):

$$Acc = (TP + TN)/(TP + TN + FP + FN) \qquad (14.2)$$

The sensitivity is calculated using the Formula (14.3):
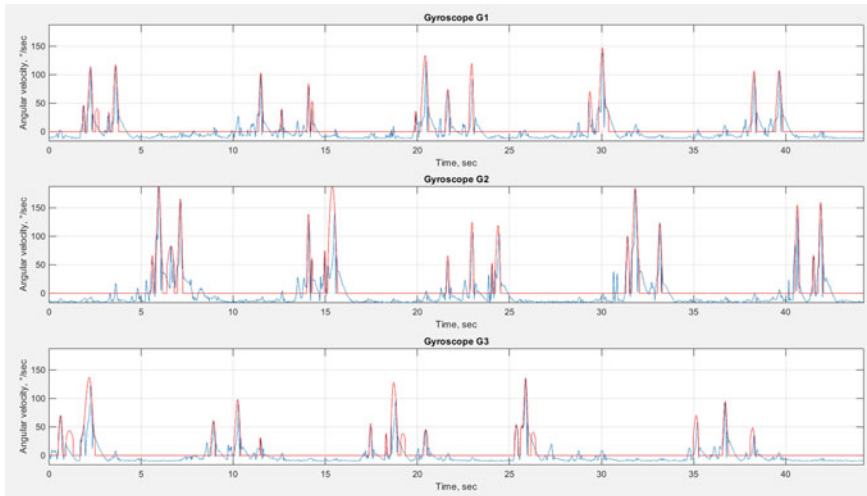
$$Sens = TP/(TP + FN) \qquad (14.3)$$

Fig. 14.9 The result of the algorithm performance. Registered signal (blue) and detected motion (red)

Table 14.1 Signal processing results

Record number	TP	FN	Acc, Sens %	Average Acc, Sens, %
1 (sitting)	29	1	96.97	90
2 (sitting)	26	4	86.67	
3 (walking)	25	5	83.33	
4 (walking)	28	2	93.33	

In our experiment, a true negative (TN) result is always absent, and according to the processing results, the number of false positive (FP) defined movements is also 0. Thus, when this algorithm is implemented on these signals, the sensitivity and accuracy are equal. Average accuracy is 90% and average sensitivity is also 90%.

For accelerometers in the [13], the following values were obtained: when identifying episodes in a calm state (sitting or lying position), the average accuracy is 87.50%, the average sensitivity is 87.50%; when detecting episodes of fetal activity in an active state (walking), the average accuracy is 71.20%, the average sensitivity is 79.16%.

Thus, at this stage of the work, the algorithm for detecting episodes of fetal activity from the signal from the abdominal gyroscopic sensors shows better results than the analysis of signals from accelerometers.

14.5   Conclusion

In this work, a system for passive monitoring of the fetus was proposed, aimed at improving medical support for pregnant women and allowing a timely response to changes in fetal activity. We have compared two types of passive sensors: accelerometer and gyroscope. As a result of testing on a model of the fetal activity signal source, it was found that the gyroscope response to the incoming impact is more pronounced than that of the accelerometer. The gyroscope is more sensitive and less prone to baseline drift. As a result of the analysis of real signals in a calm and active state, the gyroscope-based algorithm provided greater accuracy (90%) than the accelerometer-based algorithm (accuracy 87.50 and 71.20% under various conditions). Thus, gyroscopes can be used both as independent recorders and as part of an accelerometric system to improve its accuracy in the monitoring system.

References

1. Russian statistical yearbook: statistical handbook/ Federal State statistics service (Rosstat). (Moscow, 2021)
2. Federal State Information System Federal Register of Disabled (2021). https://sfri.ru/. Accessed on 06 April 2021
3. A.B. Queyam, S.K. Pahuja, D. Singh, Non-invasive feto-maternal well-being monitoring: a review of methods. J. Eng. Sci. Technol. Rev. **10**(1), 177–190 (2017)
4. M. Altini, P. Mullan, M. Rooijakkers, S. Gradl, J. Penders, N. Geusens, B. Eskofier, Detection of fetal kicks using body-worn accelerometers during pregnancy: trade-offs between sensors number and positioning. in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 290–292 (2016)
5. A. Schmidt, R. Witte, L. Swiderski, J. Zöllkau, U. Schneider, D. Hoyer, Advanced automatic detection of fetal body movements from multichannel magnetocardiographic signals. Physiol. Measurem. **40**(8), (2019)
6. J. Lai, R. Woodward, Y. Alexandrov, Q. ain Munnee, C.C. Lees, R. Vaidyanathan, Nowlan, N.C., Performance of a wearable acoustic system for fetal movement discrimination. PloS One **13**(5), (2018)
7. E. Ryo, K. Nishihara, S. Matsumoto, H. Kamata, A new method for long-term home monitoring of fetal movement by pregnant women themselves. Med. Eng. Phys. **34**(5), 566–572 (2012)
8. X. Zhao, X. Zeng, L. Koehl, G. Tartare, J. De Jonckheere, A wearable system for in-home and long-term assessment of fetal movement. IRBM **41**(4), 205–211 (2020)
9. R.H. Al-Ashwal, A.R. Sazali, Fetal movement detection and characterization based on force sensing. Curr. Sci (2018)
10. Y.O. Bobrova, O.N. Kapranova, K.V. Filipenko, Method of fetal movement registration for remote monitoring systems. in *International Youth Conference on Electronics, Telecommunications and Information Technologies*, pp. 583–591 (2021)
11. J.J. Stanger, D. Horey, L. Hooker, M.J. Jenkins, E. Custovic, Fetal movement measurement and technology: a narrative review. IEEE Access **5**, 16747–16756 (2017)
12. MPU-6050 Product Specification. InvenSense. Document Number: PS-MPU-6000A-00 (2013). https://invensense.tdk.com. Accessed 19 Aug 2013
13. K.V. Filipenko, O.N. Kapranova, Y.O. Bobrova, The algorithm for fetal activity signal processing. in *2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)*, pp. 1739–1743 (2021)

14. M.S. Khlif, B. Boashash, S. Layeghy, T. Ben-Jabeur, M. Mesbah, C. East, P. Colditz, Time-frequency characterization of tri-axial accelerometer data for fetal movement detection. in *2011 IEEE International Symposium on Signal Processing and Information Technology* (2011)
15. N. Yusenas, J. Intaravichai, P. Tirasuwannarat, T. Ouypornkochagorn, Preliminary study to detect fetal movement by using acceleration sensor and MEMS microphone. in *2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pp. 290–292 (2018)
16. A.K. Ghosh, S.F. Burniston, D. Krentzel, A. Roy, A.S. Sheikh, T. Siddiq, R. Vaidyanathan, A novel fetal movement simulator for the performance evaluation of vibration sensors for wearable fetal movement monitors. Sensors **20**(21), 6020 (2021)

Chapter 15
Lossless Genome Data Compression Using V-Gram

Samaneh Navvabi [ORCID]**, Mojtaba Najafi Oshnari** [ORCID]**, and Boris Novikov** [ORCID]

Abstract The high-throughput sequencing technology developed in recent years has enabled the public to obtain a large amount of genomic data, which has caused serious concern about the storage cost as an effective data compression algorithm. Genomic data is still an unresolved problem in genomic data research. This paper proposed a lossless Genome Data Compression Algorithm using V-gram with the help of constructing a dictionary out of sequential data, based on Unsupervised machine learning and predicting next Characters. As a result, we got a 24% Compression rate.

Keywords Data compression · V-Gram · Big data · Unsupervised machine learning

15.1   Introduction

In bioinformatics, the FASTA format is a universal standard for storing nucleotide sequences or amino acid sequences. Fasta is a text-based format that consists of single letters which represent each nucleotide or amino acids. Genome sequences are represented by letter called nucleotides or bases with four different characters, namely guanine (G), thymine (T), adenine (A) and cytosine (C). These sequences are shown in the standard IUB/IUPAC nucleic acid and amino acid codes [1].

Genome research is driving the development of personalized medicine and the discovery of drugs for cancer, rare diseases, and other serious diseases. However, genome data are much more heterogeneous than those in physics, and the generated files are enormous and considered confidential personal information. Thus, they need to maintain in a way that supports continuous discovery, and then need to be held to meet regulatory compliance requirements. Biologists are joining the Big Data Club and having trouble storing, processing, and moving information that was once the

S. Navvabi (✉) · M. Najafi Oshnari · B. Novikov
Saint Petersburg State University, St. Petersburg, Russia
e-mail: borisnov@acm.org

area of astronomers and high-energy physicists. Therefore, we must compress these data, but due to the high entropy rate in genome data, we must try a different approach than traditional [2].

In the prominent position, compression deals with how the information is structured in the data, and weaving this structure leads to the optimization of the data display and, consequently, to the reduction of their volume. If data production is achieved, steps have been taken to understand better and present them more easily. In cases where it is impossible to identify the source correctly, compression algorithms try to discover this structure approximately and use it in compressing data. Conceptual tools developed in coding knowledge that has helped to shorten data have helped determine these structures, which have also been used to analyze biological data.

Formal grammars have been successful models in analyzing age sequences and extracting their structure. The study of DNA from a linguistic point of view, which has led to the extraction of formal grammars from their structure, has a lifespan equal to DNA discovery's lifespan.

In recent decades, storage requirements have grown spectacularly. Storage funding has more problems than sequencing. That is a big problem that the modern scientist has to face. Sequencing has become more problematic, and this problem makes the whole process more difficult. The motivation to sequence and produce new data has diminished [3].

Such data comes in the form of Short strings typically with lengths in the range of 75–150. Each character represents a nucleotide and can accept the values of A (adenine), C (cytosine), G (guanine), T (thymine), or N (error in base call) [4]. All this data needs to be retrieved and processed to keep in I/O traffic because of the slow processing power [5].

The important thing is to compress information without losing the data. For instance, the data created at a specific time may use to refer to the results of other previously created data, which implies the reprocessing of a possibly much larger data set. Thus, it is necessary to develop a robust and lossless Data compression algorithm.

Construction of the V-Grams dictionary depends on the minimum description length principle. This approach allows us to achieve competitive results in data compression tasks like Genome data compression compared to other methods.

This paper proposed a lossless Genome Data Compression Algorithm using V-gram with the help of constructing a dictionary out of a sequential data, based on Unsupervised machine learning and predicting next Characters. The paper structured as follows: Literature review, Material and Methods, Results, and Conclusion and future work.

15.2   Literature Review

A promising approach to genome data storing as big data is using compression methods. Normally compression methods are dividing into two major approaches: traditional compression algorithms and recurrent neural network algorithms. In order to compress noisy data like genome data, the conventional approach unable to reach a high compression rate. Biocompress, GenCompress-1, Gencompress-2, XM, DNA-COMPACT, MFCompress, and DeepZip model are the most frequent methods among the recurrent neural network algorithms.

Currently, universal text compression algorithms, such as gzip [6] and bzip [7], doesn't suit the compression of genomic information since these algorithms were designed for the compression of English text. Besides, the desoxyribonucleic acid sequences carry four bases: 2 bits ought to be adequate to store every base and follow no clear rules like those of text files that can't offer correct compression results.

For instance, bzip2 can compress 9.7 MB of information to 2.8 MB (the compression ratio (CR) is extensively better than 2 bits in keeping with base (bpb)). Nevertheless, that is a far from fine in phrases of compression efficiency. Thus, greater powerful compression strategies for genomic data are required. Recently, numerous lossless compression algorithms were counselled for this project thinking about the capabilities of that genomic information.

15.2.1   Biocompress

Grumbach and Tahi [8] is initially purposed in 1993 for the compression of genomic data. An attempt is formed to seek out the precise and palindromic iterations within the target sequence during a mounted size window. These actual and palindromic iterations are then coded by position and, therefore, the length of their first occurrences.

Biocompress2 [9] encodes the non-repeated regions with an arithmetic coder of order 2 [10] as an improved version.

Based on the detection of the approximate repetitions, the algorithms **Gen-Compress-1** and **Gencompress-2** [11] are proposed. In the first algorithm, Gen Compress-1 uses the Hamming distance (replacement only) for the approximate iterations, while Gencompress-2 in another version uses the Editions distance (erase, insert, replace) to encode the approximate iterations. The context-based compression method is one of the most effective methods for compressing genome sequences based on the n-gram dictionary. It uses the previous n bases to predict the next base probability, followed by the arithmetic coder.

15.2.2   XM

Kaipa et al. [12] XM is a hybrid of expert compression models that calculate the probability of the next nucleotide. The expert consists of three models, Order 1 Markov Model, Order 2 Markov Model, and Copy Expert, which assumes the next nucleotide as part of the repeat sequence. The probabilities of experts are merged and sent to the arithmetic coder.

15.2.3   The DNA COMPACT

Li et al. [13] is a two-pass algorithm. In the first pass, the exact repeats and complementary palindromes are found and encoded, and in the second pass, the remaining sequence is encoded with context models. In this context model, the logistic regression method is used and performs better than the Bayes averaging used by XM.

15.2.4   MFCompress

Pinho and Pratas [14] uses several competing n-gram dictionary models to encode the DNA sequences. Since the size of the n-gram dictionary increases exponentially with the size of n, n has a memory limit, usually less than twelve bases. This method is effective when the genome sequences have local features such as sequence motifs [15], CpG island [16], and tandem repeat [17]. However, the repetition of this local feature is not a complete repetition, an approximate iteration, and in addition to its local relevance, the genome also has global dependencies, such as [18, 19].

15.2.5   DeepZip

Tatwawadi [20] is one of those learning methods suggested for text compression. It predicts the next character/word based on the model of the Recurrent Neural Network (LSTM / GRU) [21]. Although it is attempted to compress the data of the human chromosome 1 genome, the weight of the model is not weighted, and a single pass is made through the data. The RNN, LSTM or GRU models can easily capture long-range correlations.

The structure of the V-Gram dictionary is according to the principle of minimum description length. This allows us to achieve competitive results in data compression tasks. The V-grams Dictionary can be used as terms in the text representation of the term bag. From a practical point of view, V-Gram has at slightest two noteworthy

points of interest over time: fixed dictionary and volume simple text normalization. V-gram dictionary can be much smaller than the word dictionary but demonstrate the same score.

15.3 Materials and Methods

15.3.1 Background

The method discussed in this article is based on the Algorithm introduced by Igor et al. [22], but with a few modifications that are briefly explained below.

Data: Set of sequences X
Result: Optimal A
$A_0 = \mathcal{A}$;
$X_0 =$ the occurrence of alphabet phrases in X;
$t = 0$;
repeat
   $t = t + 1$;
   $X_t = \emptyset$;
   repeat
      Pick x at random from X;
      Split $x \to \{x_k\}$ using terms from A_{t-1} and frequencies from X_{t-1};
      foreach *unigram a and bigram (a, b) from* $\{x_k\}$ **do** put $a \to X_t$ and $(a, b) \to X_t$
   until $StoppingCriterion(X_t)$;
   if t mod $2 = 1$ **then**
      $A_t = Expand(A_{t-1}, X_t)$;
   else
      $A_t = Reduce(A_{t-1}, X_t)$;
   end
until t mod $2 = 0$ *and* $A_t \cap A_{t+1} > N$;

Algorithm 1: General schema of the algorithm

The overall schema of dictionary A construction (Algorithm 1) comes with two iterative phases:

1. **Expansion phase:** Adding the foremost informative component to the dictionary. The most informative component in the equation are estimated using accumulated statistics for pairwise frequency and Kullback-Liebler divergence (15.1).

$$\hat{A} = \arg\min_A H(X, A) = \arg\min_A D_{KL}\left(p(b|a, A, P)||p(b|A, P)\right)$$

$$= \arg\min_A \sum_{x \in X} \sum_{k=1}^{|x|_A} p(x_k|x_{k-1}, A, P) \log \frac{p(x_k|x_{k-1}, A, P)}{p(x_k|A, P)}$$

$$= \arg\min_A \sum_{x \in X} \sum_{k=1}^{|x|_A} p(x_k|x_{k-1}, A, P) \log \frac{p(x_k, x_{k-1}|A, P)}{p(x_k|A, P)p(x_{k-1}|A, P)}$$

$$= \arg\min_A \sum_{(a,b) \in A^2} f(a, b|A, X) p(b|a, A, P) \log \frac{p(a, b|A, P)}{p(a|A, P)p(b|A, P)}.$$

(15.1)

2. **Reduction phase**: Removing the the slightest informative component from the dictionary. For each dictionary component a_j except members of $\mathcal{A}$, removing the dictionary's least descriptive components We can calculate the cost of removing each dictionary component a_j excluding members of $\mathcal{A}$. We format each element a using the rest of the current dictionary $A_k \setminus a$, and we update metrics for coding elements $(a = (b_1, \ldots, b_s), b_j \in A_k \setminus a)$. We append the frequency of a to the coding during the update.

In terms of the variability of element length, entries in the constructed dictionary A are referred to as V-Grams. Iterative conditional optimization is used in the algorithm scheme. A step in the optimal direction is taken in this scheme, and the result is projected to the conditions. All of the algorithm's phases are greedy and optimize the same function: X code length.

15.4 Results

In our experiments, we tested V-Gram on the full Human genome (GRCh38.p12) as well as the human mitochondrial genome (NC_012920.1) which is the entirety of hereditary information contained in human mitochondria and Complete DNA sequence of yeast chromosome XI (NC_001143.9) which includes more than three hundred thousand characters and happens to be very resistant to compression. Analyses are led to look at the normal compression of information X without extra pre-processing (like arranging). For this examination, the data was compacted by various ways (Table 15.1).

A stable correlation trend was found between compression rate and vocabulary size (Fig. 15.1). This fact illustrates one of the approach's main characteristics, namely constructing an optimal feature set. For a feature selection method, rather than expanding the feature set size would increment computational expense.

For this experiment, the data from National Center for Biotechnology Information is used; extracted txt file from GRCh38.p12 and NC_001143.9 and NC_001143.9 compressed with different Dictionary sizes, and direct correlation between Dictionary size and the Compression rate was observed highest compression rate was 24%.

Table 15.1 Genetic data compression

Method	Size, bytes	Compression rate (%)
Original data(GRCh38.p12)	3313061631	–
Biocompress-2	1900999330	57.38
Dictionary size 500	851456840	25.7
Dictionary size 1000	848143781	25.6
Dictionary size 2000	844830716	25.5
Dictionary size 10000	824952345	24.9
Dictionary size 20000	815013170	24.6

Fig. 15.1 The compression rate of genetic data depends on the size of the dictionary
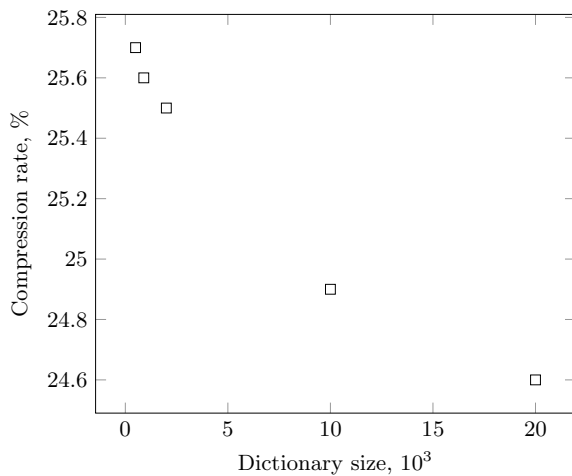
Another essential feature of the proposed method is creating v-grams of any length, as demonstrated by the dictionary elements created for this compression task. Table 15.2 contains examples of nontrivial dictionary elements of long length with frequencies. In contrast to trivial sequences such as homopolymers, these examples are referred to as nontrivial.

Table 15.2 Genetic sub-sequences as Example

V-grams	Frequency
AGCCTCAGAGCAAGAAAAAACTCTGTCTCGGGTGA	101
AAGAACAGCATCATGCTACCTGACTATGGCTGGGTCAAACTA	70
CTTTCATACAGAGTTGAACCTTGAGTTCAGCTTTCAAACACTCTTT	30
AGTAATGGGAAGCTGGAGTCAAATGGTATTTCTAGTTCTTCTAGAT	206

15.5 Conclusion

Our work has shown that it is possible to compress human genome sequences by a recurrent neural network approach. This will help better understand the laws governing DNA sequence data because identifying better data compression is closely related to identifying better data prediction models. Biological analyses explore the more features of genome sequences in the future, and compression methods could provide more information about redundancy and a higher increase in compression performance and The training of the deep-learning model for the human genome using background information like mutations, motifs, and tandem repeats, etc.

References

1. W.R. Pearson, D.J. Lipman, Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. **85**(8), 2444–2448 (1988)
2. D. Greenfield, V. Wittorff, M. Hultner, The importance of data compression in the field of genomics. IEEE Pulse **10**(2), 20–23 (2019)
3. M.S. Akhtar, I.A. Alaraidh, M.K. Swamy, Data measurement, data redundancy, and their biological relevance, in *Essentials of Bioinformatics*, vol III (Springer, Berlin, 2019), pp. 103–107
4. C. Sessegolo, C. Cruaud, C. Da Silva, A. Cologne, M. Dubarry, T. Derrien, V. Lacroix, J.M. Aury, Transcriptome profiling of mouse samples using nanopore sequencing of cdna and rna molecules. Sci. Rep. **9**(1), 1–12 (2019)
5. C. Bischof, D. Wilfinger, Big data-enhanced risk management. Trans. FAMENA **43**(2), 73–84 (2019)
6. Gzip, Available online. http://www.gzip.org/
7. Bzip, Available online. http://www.bzip.org/
8. P. Grumbach, F. Tahi, Compression of DNA sequences, in *Proceedings DCC93: Data Compression Conference* (IEEE, 1993), pp. 340–350
9. S. Grumbach, F. Tahi, A new challenge for compression algorithms: genetic sequences. Information Process. Manage. **30**(6), 875–886 (1994)
10. D. Salomon, G. Motta, *Handbook of Data Compression* Springer Science & Business Media (2010)
11. X. Chen, S. Kwong, M. Li, A compression algorithm for DNA sequences and its applications in genome comparison. Genome Informatics **10**, 51–61 (1999)
12. K.K. Kaipa, A.S. Bopardikar, S. Abhilash, P. Venkataraman, K. Lee, T. Ahn, R. Narayanan, Algorithm for DNA sequence compression based on prediction of mismatch bases and repeat location, in *2010 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW)* (IEEE, 2010), pp. 851–852
13. P. Li, S. Wang, J. Kim, H. Xiong, L. Ohno-Machado, X. Jiang, Dna-compact: Dna compression based on a pattern-aware contextual modeling technique. PloS one **8**(11) (2013)
14. A.J. Pinho, D. Pratas, Mfcompress: a compression tool for fasta and multi-fasta data. Bioinformatics **30**(1), 117–118 (2013)
15. T.L. Bailey, N. Williams, C. Misleh, W.W. Li, Meme: discovering and analyzing dna and protein sequence motifs. Nucleic Acids Res. **34**(suppl_2), W369–W373 (2006)
16. K. Jabbari, G. Bernardi, Cytosine methylation and cpg, tpg (cpa) and tpa frequencies. Gene **333**, 143–149 (2004)
17. J. Jorda, B. Xue, V.N. Uversky, A.V. Kajava, Protein tandem repeats-the more perfect, the less structured. FEBS J. **277**(12), 2673–2682 (2010)

18. A. Sanyal, B.R. Lajoie, G. Jain, J. Dekker, The long-range interaction landscape of gene promoters. Nature **489**(7414), 109 (2012)
19. A. Miele, J. Dekker, Long-range chromosomal interactions and gene regulation. Mol. Biosyst. **4**(11), 1046–1057 (2008)
20. K. Tatwawadi, Deepzip: lossless compression using recurrent networks (2018). https://web. stanford.edu/class/cs224n/reports/2761006.pdf
21. H. Sak, A. Senior, F. Beaufays, Long short-term memory recurrent neural network architectures for large scale acoustic modeling, in *Fifteenth Annual Conference of the International Speech Communication Association* (2014)
22. I. Kuralenok, N. Starikova, A. Khvorov, J. Serdyuk, Construction of efficient v-gram dictionary for sequential data analysis, in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18 (ACM, New York, NY, USA, 2018), pp. 1343–1352. http://doi.acm.org/10.1145/3269206.3271789

Chapter 16
The Method of Generating Random OQPSK Signal Packets Based on Optimal FTN Pulses with Reduced PAPR and Frequency Band

Anna Ovsyannikova ⬤**, Sergey Makarov** ⬤**, and Ge Dong**

Abstract Faster-than-Nyquist (FTN) is a promising technique used to improve spectral efficiency of a communication system. However, FTN signals have high peak-to-average power ratio (PAPR). To eliminate this drawback and keep spectral efficiency at a high level, we propose the method of constructing FTN signals with offset quadrature modulation via scaling up the pulses in quadrature channels and further pulse optimization with the constraint on the constant signal envelope. The values of optimization parameters providing both PAPR reduction and occupied frequency bandwidth reduction compared to the signals based on root-raised-cosine pulses with the roll-off factor 0.3 are found.

Keywords Faster-than-nyquist · Optimization problem · PAPR · OQPSK · RRC · Spectral efficiency

16.1 Introduction

In case of the channels with constant parameters and additive noise the values of spectral efficiency above 2 bps/Hz can be achieved due to application of signals based on Faster-than-Nyquist (FTN) pulses. Such pulses have a long duration T_s which significantly exceeds the transmission time T of one information bit of binary message, and occupy small bandwidth (one-sided) $\Delta F < 0.5/T$ [1]. FTN signals are usually formed in two ways.

According to the first way, FTN signals are formed on the basis of root raised cosine (RRC) pulses via filtering [1–4]. Time and spectral characteristics of such signals depend on the shape of raised cosine frequency response of pulse shaping filter with the roll-off factor $0 \leq \beta \leq 1$ [5]. For $\beta = 0$ the pulse reduces to the shape

A. Ovsyannikova (✉) · S. Makarov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: ovsyannikova_as@spbstu.ru

G. Dong
School of Aerospace Engineering, Tsinghua University, Beijing, China

$\sin(x)/x$. In practice, only pulses of finite duration can be formed, having a fairly compact spectrum, but a low reduction rate of out-of-band emissions. With a long pulse duration $T_s > T$, random intersymbol interference occurs in the signal packet, which leads to energy losses [1]. Signals based on RRC pulses have high peak-to-average power ratio (PAPR). It is a drawback since high PAPR causes degradation of the efficiency of power amplifiers [6–8]. The value of PAPR is influenced by several factors such as pulse duration, the value of the roll-off factor, the symbol rate, and the size of signal constellation [9].

The second way assumes the synthesis of optimal FTN pulses [10–13]. The optimization problem includes the constraints on the compactness of spectrum of signals, the reliability of detection and technical and economic constraints. To increase the spectral efficiency, pulse duration $T_s > T$ is chosen. It leads to intersymbol interference, which is not random, but controllable due to a specific optimization constraint [13, 14].

The disadvantage of both methods is high PAPR of signal packets [13, 15]. In this regard, the possibility of constructing signal packets providing the same spectral efficiency but reduced PAPR becomes especially interesting, since amplifiers can be used more effectively [16]. Due to application of the signals with reduced PAPR, the average power of oscillations can be increased, while the total operating energy costs are kept unchanged. As a result, the quality of detection can be improved. In [13], the problem of PAPR reduction for optimal FTN signals with quadrature modulation was considered. The optimization constraint on PAPR for the case of BPSK (binary phase shift keying) was used to obtain optimal pulses for each quadrature channel. This made it possible to reduce PAPR by 2 dB compared to the signals based on RRC pulses.

Further PAPR reduction in conditions of fixed energy losses and spectral parameters can be done with the use of the optimization constraint on the constancy of instantaneous power, or the constancy of signal envelope [17]. In this work, the method of generating optimal FTN signals with OQPSK (offset quadrature phase shift keying) and with the constraint on the constancy of instantaneous power is proposed.

The purpose of this work is to determine the shapes of optimal FTN pulses providing the gain in spectral efficiency and the energy gain in terms of PAPR in case of OQPSK compared to the signals based on RRC pulses.

16.2   Optimization Problem

The packet $y(t)$ of N signals with OQPSK and amplitude A_0, carrier frequency ω_0, an initial phase ϕ_0 and a pulse shape $a(t)$ can be presented the next way:

$$y(t) = y_1(t) - y_2(t), \tag{16.1}$$

$$y_1(t) = I(t)\cos(\omega_0 t + \phi_0), \ y_2(t) = Q(t)\sin(\omega_0 t + \phi_0), \tag{16.2}$$

$$I(t) = \sum_{k=0}^{(N-1)/2} c_{2k} a(t - 2kT), \quad Q(t) = \sum_{k=0}^{(N-1)/2} c_{2k+1} a(t - 2kT - T). \qquad (16.3)$$

The sequence (16.1) is divided into two quadrature streams $I(t)$ and $Q(t)$ of odd and even symbols $c_k = \pm 1$ transmitted at the rate $R = 1/2T$ at doubled time interval T.

The problem of optimizing the pulse shape according to the criterion of the specified reduction rate of out-of-band emissions looks as follows:

$$\arg\{\min_{a(t)} J\}, \quad J = \int_{-\infty}^{\infty} g(f) \left| \int_{-\infty}^{\infty} a(t) \exp(-j2\pi ft) dt \right|^2 df. \qquad (16.4)$$

Here J is an optimization functional, $a(t)$ is a pulse shape of duration $T_s = LT$ which minimizes the functional J, $g(f) = f^{2n}$ is a weighting function ($L, n \in \mathbb{N}$).

Let us consider the numerical solution to the problem (16.4). The function $a(t)$ can be presented as an expansion into limited Fourier series with m coefficients. Assume that $a(t)$ is an even function at interval $[-T_s/2, T_s/2]$:

$$a(t) = \frac{a_0}{2} + \sum_{k=1}^{m-1} a_k \cos(2\pi kt/T_s). \qquad (16.5)$$

Then the problem of minimizing the optimization functional (16.4) is reduced to the problem of searching for the set of expansion coefficients which minimize the function of many variables:

$$J(\{a_k\}_{k=1}^{m-1}) = \frac{T_s}{2} \sum_{k=1}^{m-1} a_k^2 (2\pi k/T_s)^{2n}. \qquad (16.6)$$

The number of m coefficients of limited Fourier series is determined by the accuracy of $a(t)$ representation and usually is equal to m=10–14 [13] for FTN pulses with duration up to $T_s = 16T - 32T$.

The constraints on the signal energy and boundary conditions which define reduction rate of out-of-band emissions are written the next way:

$$\int_{-T_s/2}^{T_s/2} a^2(t) dt = 1, \quad a^{(k)}(t) \big|_{t=\pm T_s/2} = 0, \quad k = 1 \ldots (n-1). \qquad (16.7)$$

Here $a^{(k)}$ is the k-th derivative of $a(t)$.

The constraint on the cross-correlation coefficient determines the influence of adjacent pulses on each other. This coefficient depends on the symbol rate R, pulse shape $a(t)$, and the pulse duration $T_s = LT$. The constraint on the cross-correlation

coefficient determines bit error rate (BER) performance and can be presented by the following inequality:

$$\max_{k=1\ldots\lfloor(L-1)/2\rfloor} \left| \int_{-LT/2+k/R}^{LT/2} a(t)a(t-k/R)dt \right| < K_0. \tag{16.8}$$

If the symbol rate is equal to $R = 1/2T$ in each quadrature channel, the constraint (16.8) transforms into the next expression:

$$\max_{k=1\ldots\lfloor(L-1)/2\rfloor} \left| \int_{-LT/2+2kT}^{LT/2} a(t)a(t-2kT)dt \right| < K_0. \tag{16.9}$$

For instance, for a pulse $a(t)$ of duration $T_s = 8T$ ($L=8$) this expression can be illustrated by Fig. 16.1. Here vertical dashed lines denote the limits of integration. It can be seen that the highest value of cross-correlation coefficient correspond to $k=1$.

Fig. 16.1 Illustration of the constraint on the cross-correlation coefficient (16.9)
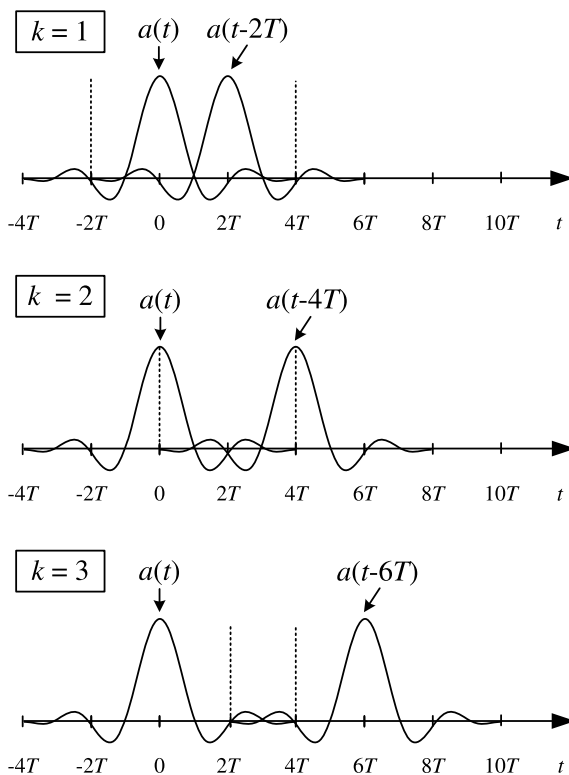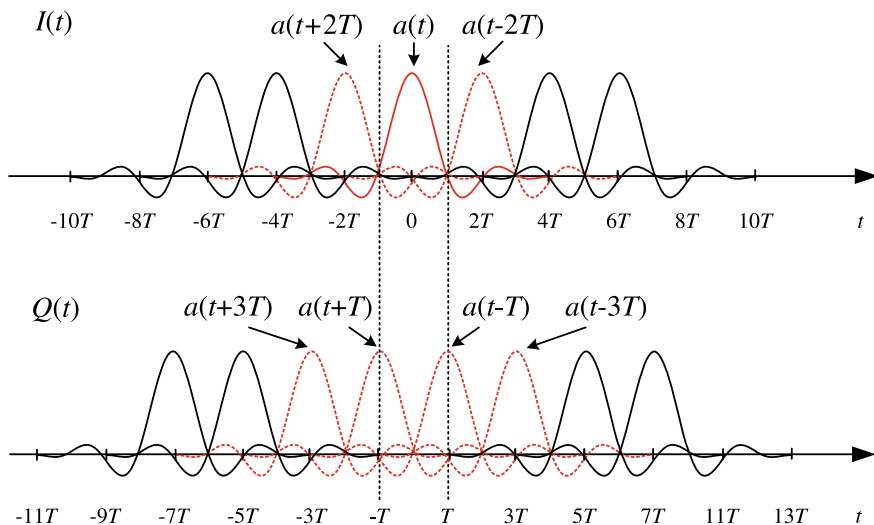
Fig. 16.2 Illustration of the constraint on the constancy of instantaneous power for OQPSK (16.9)

In general, the constraint on the constancy of instantaneous power [17] can be written with the use of notations from (16.1):

$$I^2(t) + Q^2(t) = \text{const.} \tag{16.10}$$

Figure 16.2 shows the sequence of the pulses $a(t)$ of duration $T_s = 8T$ $(L{=}8)$ in in-phase and quadrature components. The symbol rate in each quadrature channel is equal to $R = 1/2T$. Let us consider the time interval $[-T, \ T]$. In the in-phase component, there is an interference between the pulses $a(t + 4T)$, $a(t + 2T)$, $a(t)$, $a(t - 2T)$, $a(t - 4T)$. Similarly, in the quadrature component, there is an interference between the pulses $a(t - 3T)$, $a(t - T)$, $a(t + T)$, $a(t + 3T)$. The point of the constraint on the constancy of instantaneous power is to find the optimal pulse shape $a(t)$ which meets the requirement (16.10). Since $a(t)$ is symmetrical about $t{=}0$, let us consider only the pulses $a(t)$, $a(t - T)$, $a(t - 2T)$, $a(t - 3T)$, $a(t - 4T)$ at the time interval $[0, T]$. Taking into account (16.1) and (16.10), we can write down the following equation:

$$t \in [0, T]: (c_0 a(t) + c_2 a(t - 2T) \ + c_4 a(t - 4T))^2 + $$
$$(c_1 a(t - T) + c_3 a(t - 3T))^2 = \text{const} = a^2(0). \tag{16.11}$$

For the pulses of duration $T_s = 8T$ $(L = 8)$ the constraint on the constancy of instantaneous power in integral form can be defined by the next expression:

$$\int\limits_{0}^{T} \Bigg[\big(c_0 a(t) + c_2 a(t - 2T) + c_4 a(t - 4T) \big)^2 +$$

$$\big(c_1 a(t - T) + c_3 a(t - 3T) \big)^2 - a^2(0) \Bigg]^2 dt = 0, \ \forall \, c_i \ (i = 0 \ldots 4). \quad (16.12)$$

In case of an arbitrary pulse duration under condition of the symmetry about $t=0$, the expression (16.12) is generalized into (16.13):

$$\int\limits_{0}^{T} \Bigg[\bigg(\sum_{k=0}^{L/4} c_{2k} a(t - 2kT) \bigg)^2 + \bigg(\sum_{k=0}^{L/4-1} c_{2k+1} a\big(t - (2k + 1)T \big) \bigg)^2 - a^2(0) \Bigg]^2 dt = 0,$$

$$\forall \, c_i \ (i = 0 \ldots L/2). \quad (16.13)$$

To limit the maximum value of the integral in (16.13) with considering all possible symbol combinations (q is the combination number), we should introduce the parameter ε (16.14):

$$\max_{\{q\}} \Bigg\{ \int\limits_{0}^{T} \Bigg[\bigg(\sum_{k=0}^{L/4} c_{2k}^q a(t - 2kT) \bigg)^2 +$$

$$\bigg(\sum_{k=0}^{L/4-1} c_{2k+1}^q a\big(t - (2k + 1)T \big) \bigg)^2 - a^2(0) \Bigg]^2 dt \Bigg\} < \varepsilon. \quad (16.14)$$

The inequality (16.14) is the constraint on PAPR for the optimization problem (16.4). The usage of this constraint allows not only obtaining the optimal pulse shape $a(t)$ of arbitrary duration, but also taking into account the scaling of $a(t)$ and the change in the transmission rate. Obviously, scaling up $a(t)$ along time axis with the constraints (16.7) leads to frequency band reduction. In the next sections, we are going to consider the method of generating OQPSK signal packets via pulse scaling.

16.3 Results of the Optimization Problem Solving

The main goal of searching for FTN signals with OQPSK is to obtain optimal pulse shape with the considered constraints taken into account via scaling up the pulses in quadrature channels. Hereinafter the found optimal pulse shape is denoted as $a_{opt}(t)$. The procedure of the optimization problem solving and its features are given in [13] in detail. In this section, we are going to present the optimal pulse shape $a_{opt}(t)$ and basic spectral and time parameters of signal packets.
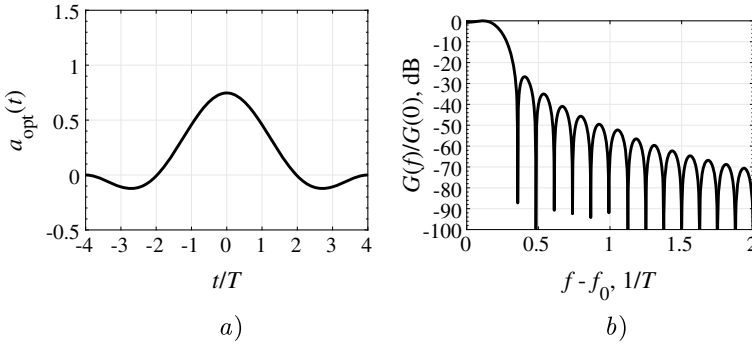
Fig. 16.3 $a_{opt}(t)$ after scaling up twice (**a**) and corresponding energy spectrum (**b**)

Table 16.1 The values of PAPR and $\Delta F_{99\%}$ for signal packets with OQPSK based on scaled pulses

	RRC, $\beta = 0$	RRC, $\beta = 0.3$	$a_{opt}(t)$
$\Delta F_{99\%}$, $1/T$	0.54	0.59	0.56
PAPR, dB	6.72	6.06	6.16

Let us start with the optimal pulse of duration $T_s = 4T$ obtained for the symbol rate $R = 1/T$ with the cross-correlation coefficient $K_0 = 0.1$ and providing reduction rate of out-of-band emissions $1/f^6$ (n=2) [13]. After scaling up twice the optimal pulse $a_{opt}(t)$ provides the same value of cross-correlation coefficient $K_0 = 0.1$ for the rate $R = 1/2T$. Note that the constraint on the constancy of instantaneous power (16.14) is not used at this step of the optimization procedure.

In Fig. 16.3a the optimal pulse shape $a_{opt}(t)$ of duration $T_s = 4T$ scaled up to $T_s = 8T$ is shown. The normalized energy spectrum $G(f)/G(0)$ of a random signal packet based on $a_{opt}(t)$ pulse shape is presented in Fig. 16.3b.

The occupied frequency bandwidth $\Delta F_{99\%}$ (two-sided) and PAPR of a bandpass signal packet consisting of 10000 symbols with $a_{opt}(t)$ pulse shape and OQPSK are given in Table 16.1. This table also includes the values of PAPR and $\Delta F_{99\%}$ for the bandpass signals based on RRC pulses of duration $T_s = 4T$ scaled up to $T_s = 8T$ with the roll-off factor $0 \leq \beta \leq 1$. As it can be seen from Table 16.1, the least PAPR value corresponds to the signals based on RRC pulses with $\beta = 0.3$ while the occupied frequency bandwidth is almost the same for all the cases.

Now we can obtain the optimal pulse shape with the constraint on the constancy of instantaneous power (16.14). The pulse shape $a_{opt}(t)$ from Fig. 16.3 is used as an initial approximation. Calculation of ε according to (16.14) gives the value 0.7. To achieve PAPR reduction, ε needs to be decreased step-by-step during the optimization with the fixed parameters $T_s = 8T$, $K_0 = 0.1$, $n = 2$.

The optimal pulse shape $a_{opt}(t)$ for $\varepsilon = 0.2$ and its energy spectrum are demonstrated in Fig. 16.4. Introducing the constraint on PAPR causes degradation of the side lobes of the pulse shape. The amplitude of the side lobes decreases significantly.
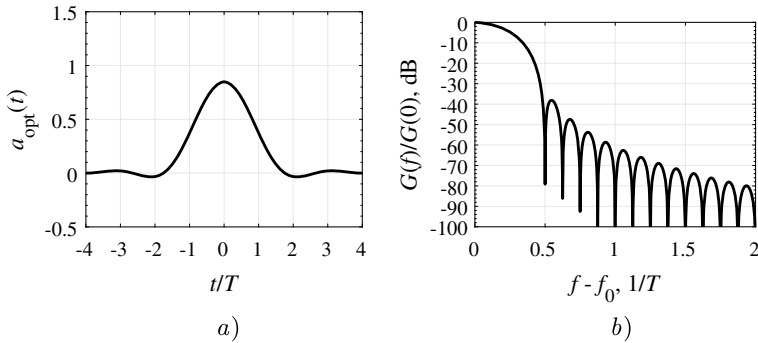
Fig. 16.4 $a_{opt}(t)$ obtained for $\varepsilon = 0.2$ (**a**) and corresponding energy spectrum (**b**)
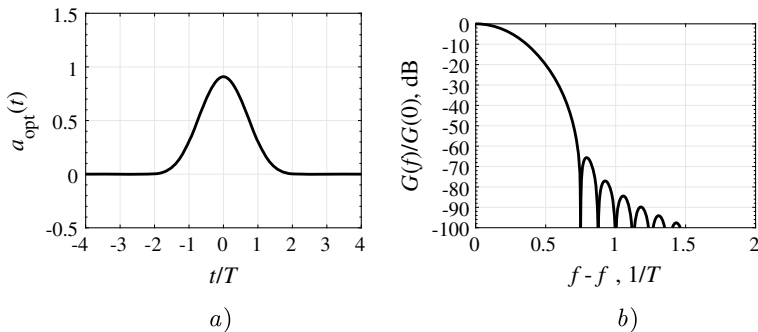
Fig. 16.5 $a_{opt}(t)$ obtained for $\varepsilon = 0.07$ (**a**) and corresponding energy spectrum (**b**)

Application of signal packets with OQPSK and the obtained optimal pulse shape $a_{opt}(t)$ provides PAPR=4.6 dB and occupied frequency bandwidth $\Delta F_{99\%} = 0.73/T$. In other words, the pay for the gain of 1.56 dB in PAPR is the expansion of the occupied frequency bandwidth by 30%.

More stringent requirements for PAPR (i.e., further decrease in the constraint on the constancy of the instantaneous power ε) leads to almost zero level of the side lobes of the optimal pulse shape. It is worth noting that for $\varepsilon < 0.07$ the amplitude of the side lobes of the optimal pulse becomes so small that the value of cross-correlation coefficient K_0 does not exceed 0.05 instead of 0.1. Bandpass signal packets with OQPSK and $a_{opt}(t)$ with $\varepsilon = 0.07$ provide PAPR=3.8 dB and occupy frequency bandwidth $\Delta F_{99\%} = 0.85/T$ (Fig. 16.5). These signals outperform the signals with OQPSK and the optimal pulse shapes with the constraint on PAPR presented in [13]. In that paper, the minimum achieved PAPR was equal to 4 dB while the occupied frequency bandwidth was equal to $\Delta F_{99\%} = 1.17/T$.
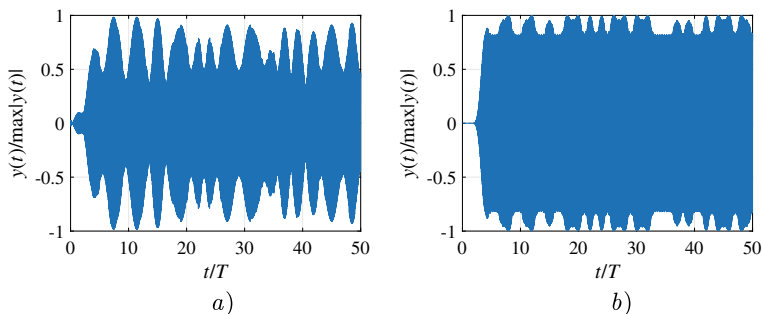
Fig. 16.6 OQPSK signal based on RRC pulses with β=0.3 (**a**) and optimal pulse $a_{opt}(t)$ with $\varepsilon = 0.07$ (**b**)

Figure 16.6 illustrates normalized random FTN signal packets with OQPSK (16.1) based on RRC pulses with $\beta = 0.3$ (**a**) and optimal pulses $a_{opt}(t)$ with $\varepsilon = 0.07$ (**b**). As it is seen from Fig. 16.6, the average power for the signal packet based on optimal pulses $a_{opt}(t)$ with $\varepsilon = 0.07$ is much higher than for the signal packet based on RRC pulses with $\beta = 0.3$.

16.4 Summary of Simulation Results

Let us consider the results of the optimization problem solving with different values of the constraint on the constancy of instantaneous power ε. Figure 16.7 shows the dependencies of occupied frequency bandwidth $\Delta F_{99\%}$ defined by the criterion of containing 99% of signal energy (*a*), occupied frequency bandwidth ΔF_{-30dB} defined by the level of -30 dB of energy spectrum (*b*), and PAPR (*c*) on ε. It can be concluded from Fig. 16.7a), *c*) that for $\varepsilon < 0.07$ optimal pulses $a_{opt}(t)$ have wider bandwidth $\Delta F_{99\%}$ and smaller PAPR than signals based on RRC pulses with $\beta = 0$, or $\sin(x)/x$. At the same time, the value of $\varepsilon = 0.5$ provides the minimum value of ΔF_{-30dB}. Then the gain in spectral efficiency relative to the signals based on the pulses $\sin(x)/x$ reaches 28%.

Let us compare signals based on optimal pulses $a_{opt}(t)$ and signals based on RRC pulses with $\beta = 0.3$. The equality of the occupied frequency bandwidth $\Delta F_{99\%}$ is achieved at $\varepsilon = 0.55$, while PAPR of the optimal signals is 0.4 dB smaller. Besides, there is a gain about 30% in ΔF_{-30dB} provided by optimal signals.

Since the the results of pulse shape optimization with the constraint on PAPR are reasonable to use in packet transmission, it is necessary to consider the dependency of PAPR on the packet length (Fig. 16.8).

When signals based on RRC pulses scaled up twice along time axis and signals based on optimal pulses optimized according to the rate $R = 1/2T$ are applied, the dependency of PAPR on packet length has monotonically decreasing character. To
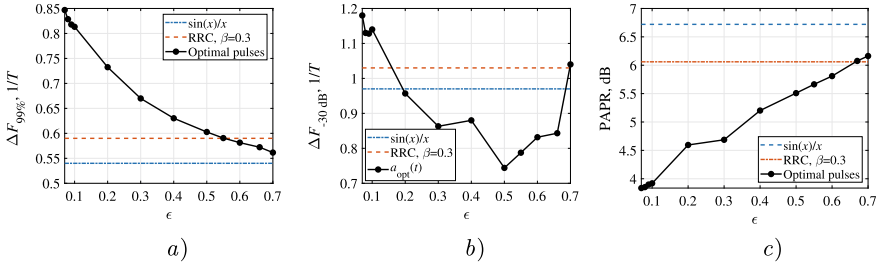
Fig. 16.7 Dependencies of $\Delta F_{99\%}$ (**a**), $\Delta F_{-30\text{dB}}$ (**b**), and PAPR (**c**) on ε
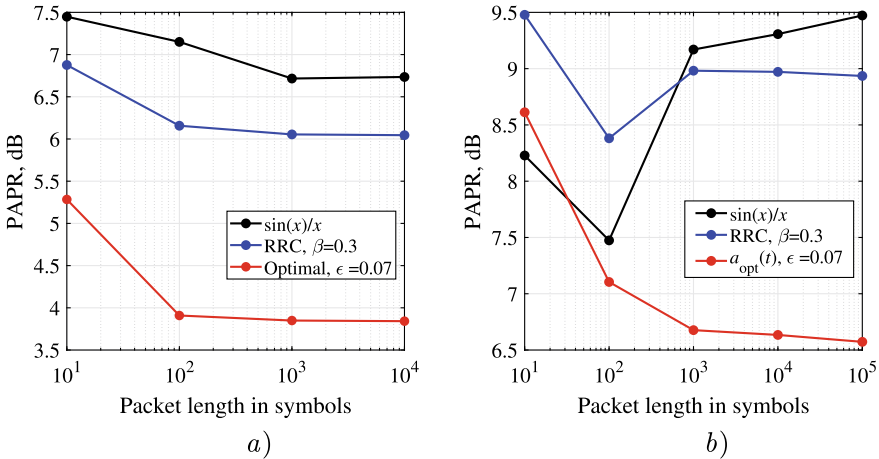
Fig. 16.8 PAPR dependency on the packet length for the symbol rate at each quadrature channel $R = 1/2T$ (**a**) and $R = 1/T$ (**b**)

estimate PAPR of signals with the pulses scaled up twice to $T_s = 8T$ it is enough to construct the packet consisting of 1000 symbols if the symbol rate at each quadrature channel is equal to $R = 1/2T$. Optimal signals based on $a_{opt}(t)$ with $\varepsilon = 0.07$ provide at least 2.2 dB gain in PAPR. Increasing the symbol rate twice leads to an increase in PAPR. However, the gain provided by optimal signals remains at least 2.2 dB if the packet length is at least 1000 symbols.

16.5 Conclusions

In this work the method of generating random signal packets with OQPSK providing reduced PAPR and occupied frequency bandwidth with the use of optimal FTN pulses is proposed. The point of this method is to perform time scaling of the optimal pulses in each quadrature channel along with the introduction of a new constraint on PAPR

into the optimization problem. This method includes the constraint on the required value of cross-correlation coefficient which provides guaranteed level of bit error rate performance in a channel with additive white Gaussian noise. The obtained optimal pulse shapes used for FTN signaling with OQPSK have the following advantages:

1. Compared to signals with OQPSK from [13], both PAPR reduction to 3.8 dB and occupied frequency bandwidth reduction to $\Delta F_{99\%} = 0.85/T$ are provided. The gain in PAPR is about 0.2 dB while the gain in occupied frequency bandwidth reaches 27%.
2. Compared to the signals based on RRC pulses with the roll-off factor $\beta = 0.3$, PAPR is reduced at most by 2.2 dB. As a consequence, the spectral efficiency $R/\Delta F_{99\%}$ decreases by 44%.
3. Compared to the signals based on RRC pulses with the roll-off factor $\beta = 0.3$, optimal signal packets provide both smaller PAPR and narrower occupied frequency bandwidth $\Delta F_{99\%}$ for the optimization parameter $0.55 \leq \varepsilon \leq 0.65$.
4. The occupied frequency bandwidth $\Delta F_{-30\text{dB}}$ determined by the level -30 dB of energy spectrum is much narrower than for signals based on RRC pulses in a wide range of values of the optimization parameter which limits PAPR ($0.2 \leq \varepsilon \leq 0.68$).

It is shown that in case of packet transmission the packet length should be at least 1000 symbols to achieve the maximum benefit in PAPR due to application of signals with OQPSK based on optimal pulses.

References

1. J.B. Anderson, *Bandwidth Efficient Coding*, 1st edn. (Wiley, Hoboken, NJ, USA, 2017)
2. J. Fan, S. Guo, X. Zhou, Y. Ren, G.Y. Li, X. Chen, Faster-Than-Nyquist Signaling: An Overview. IEEE Access **5**, 1925–1940 (2017). https://doi.org/10.1109/ACCESS.2017.2657599
3. E. Bedeer, M.H. Ahmed, H. Yanikomeroglu, A Very Low Complexity Successive Symbol-by-Symbol Sequence Estimator for Faster-Than-Nyquist Signaling. IEEE Access **5**, 7414–7422 (2017). https://doi.org/10.1109/ACCESS.2017.2663762
4. A. Caglan, A. Cicek, E. Cavus, E. Bedeer, H. Yanikomeroglu, Polar coded faster-than-nyquist (FTN) signaling with symbol-by-symbol detection, in *IEEE Wireless Communications and Networking Conference* (IEEE, 2020), pp. 1–5. https://doi.org/10.1109/WCNC45663.2020.9120701
5. I. Lavrenyuk, E. Maksimova, Y. Sadovaya, Research on FBMC/OQAM spectral and energy characteristics for different prototype filters, in *Springer Proceedings in Physics*, vol. 255 (Springer, Cham, 2021), pp. 399–411. https://doi.org/10.1007/978-3-030-58868-7_45
6. A. Pergushev, V. Sorotsky, Signal distortion decreasing in envelope tracking power amplifiers, in *Proceedings of the 2018 IEEE International Conference on Electrical Engineering and Photonics* (IEEE, 2018), pp. 44–47. https://doi.org/10.1109/EExPolytech.2018.8564443

7. Z.N. Ghanim, B.M. Omran, OFDM PAPR reduction for image transmission using improved tone reservation. Int. J. Electr. Comput. Eng. **11**(1), 416–423 (2021). https://doi.org/10.11591/ijece.v11i1

8. A. Pergushev, V. Sorotsky, A. Ulanov, Criteria for selection envelope tracking power supply parameters for high peak-to-average power ratio applications, in *Proceedings of the 2019 IEEE International Conference on Electrical Engineering and Photonics* (IEEE, 2019), pp. 13–16 https://doi.org/10.1109/EExPolytech.2019.8906793

9. D. Puzko, Y. Batov, A. Gelgor, D. Tkachenko, P. Angueira, J. Montalban, Evaluation of finite discrete RRC-pulse parameters to simulate DVB-S2 with LDM, in *Proceedings of the 2019 IEEE International Conference on Electrical Engineering and Photonics* (IEEE, 2019), pp. 140–143. https://doi.org/10.1109/EExPolytech.2019.8906847

10. A. Said, J.B. Anderson, Bandwidth-efficient coded modulation with optimized linear partial-response signals. IEEE Trans. Inf. Theory **44**, 701–713 (1998). https://doi.org/10.1109/18.661514

11. S.B. Makarov, M. Liu, A.S. Ovsyannikova, S.V. Zavjalov, I.I. Lavrenyuk, W. Xue, Optimizing the shape of faster-than-nyquist (FTN) signals with the constraint on energy concentration in the occupied frequency bandwidth. IEEE Access **8**, 130082–130093 (2020). https://doi.org/10.1109/ACCESS.2020.3009213

12. A. Gelgor, T. Gelgor, New pulse shapes for partial response signaling to outperform faster-than-nyquist signaling, in *IProceedings of the 2019 IEEE International Conference on Electrical Engineering and Photonics* (IEEE, 2019), pp. 144–148. https://doi.org/10.1109/EExPolytech.2019.8906884

13. S.B. Makarov, M. Liu, A.S. Ovsyannikova, S.V. Zavjalov, I. Lavrenyuk, W. Xue, Y. Xu, A reduction of peak-to-average power ratio based faster-than-nyquist quadrature signals for satellite communication. Symmetry **13**(2), 346–364 (2021). https://doi.org/10.3390/sym13020346

14. T.H. Phuoc Nguyen, A. Gelgor, Means to enhance the bandwidth gain from applying multicomponent signals in DVB-S2, in *Proceedings of the 2019 IEEE International Conference on Electrical Engineering and Photonics* (IEEE, 2019), pp. 173–176. https://doi.org/10.1109/EExPolytech.2019.8906865

15. Q. Li, Y. Gao, F.-K. Gong, W.-Y. Zhao, H.-Y. Ding, Y. Zhang, PAPR analysis for faster-than-nyquist signaling in satellite communications, in *International Conference on Wireless Communications and Signal Processing* (IEEE, 2020), pp. 708–711. https://doi.org/10.1109/WCSP49889.2020.9299879

16. C. Xu, T. Bai, J. Zhang, R.G. Maunder, S. Sugiura, Z. Wang, L. Hanzo, Constant-envelope space-time shift keying. IEEE J. Selected Top. Sig. Proc. **13**(6), 1387–1402 (2019). https://doi.org/10.1109/JSTSP.2019.2922587

17. W. Shang, Y. Xu, J. Qi, W. Xue, S.B. Makarov, Optimal waveform of the partial-respond signal based on minimum out-of-band radiation criterion. Appl. Sci. **7**(10), 1086–1097 (2017). https://doi.org/10.3390/app7101086

Chapter 17
Iterative Algorithm with Feedback for Coherent Detection of Non-orthogonal Spectrally Efficient Multi-frequency Signals

Dac Cu Nguyen (ID)**, Sergey B. Makarov** (ID)**, Anna S. Ovsyannikova** (ID)**, and Canh Minh Nguyen**

Abstract In this paper, a detection algorithm with feedback decision for Spectrally Efficient Frequency Division Multiplexing (SEFDM) signals is proposed. Simulation modelling of information transmission was held using suggested algorithm. Simulation results have shown that the presented algorithm is quite effective. The proposed algorithm increases computational complexity by at most 1% and provides significant energy gains compared to the detection algorithm which does not take into account intercarrier interference (ICI). For example, the energy gain can reach 7.5 dB for SEFDM signals with bandwidth compression factor 0.6 when error probability is equal to 10^{-3}.

Keywords Coherent · Non-orthogonal · SEFDM · BER performance · Iterative algorithm · Feedback

17.1 Introduction

Spectral efficiency improvement can be obtained by reducing the occupied bandwidth while maintaining the volume of transmitted information. An increase in spectral efficiency of information transmission systems using LTE, WiFi, WiMAX, DVB-T protocols with using multi-frequency signals is possible with a decrease in the spacing between subcarriers while maintaining their total number [1–5].

Such SEFDM (Spectrally Efficient Frequency Division Multiplexing) signals are generated from orthogonal multi-frequency signals by decreasing the frequency separation between subcarriers by a factor of $1/\alpha$, where α is the bandwidth compression factor. In [6–9], algorithms for generating and receiving SEFDM signals based

D. C. Nguyen (✉) · S. B. Makarov · A. S. Ovsyannikova
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

C. M. Nguyen
University of Transport and Communications, Ha Noi, Viet Nam

on DFT/IDFT (Discrete Fourier Transform/Inverse Discrete Fourier Transform) and FFT/IFFT (Fast Fourier Transform/Inverse Fast Fourier Transform) blocks are proposed.

Since the signals transmitted at the subcarriers are not orthogonal, intercarrier interference (ICI) occurs due to the influence of the signals transmitted at adjacent subcarriers on each other. This influence introduces additional interference during the demodulation of SEFDM signals transmitted on each subcarrier frequency.

Algorithms for receiving SEFDM signals in the presence of ICI interference for channels with additive white Gaussian noise (AWGN) with an average power spectral density of $N_0 / 2$ were analyzed in [10–16]. Such algorithms are based on two methods. The first is based on element-by-element coherent detection. This method is the simplest to implement, but the use of such detection algorithms does not allow obtaining high performance of detection. The second method presupposes the use of algorithms maximum likelihood sequence estimation (MLSE), which makes it possible to obtain high energy efficiency. However, the implementation of algorithms according to the second method turns out to be very complicated and does not allow realizing high absolute message transmission rates.

This work proposes a compromise solution to the problem of receiving SEFDM signals under the condition of significant ICI, namely, the use of an iterative algorithm for coherent element-by-element processing with decision feedback.

17.2   Generation of SEFDM Signals

A sequence of SEFDM signals with the number N of used subcarriers ω_n and complex symbols C_n of the channel alphabet on the n-th subcarrier can be represented as:

$$s(t) = \sum_{k=-\infty}^{\infty} \sum_{n=0}^{N-1} a(t - kT) C_{nk} e^{j\omega_n t}, \ t \in (-\infty; \infty), \quad (17.1)$$

or

$$s^{(0)}(t) = \sum_{n=0}^{N-1} a(t) C_n e^{j\omega_n t}, \ t \in [0; T],$$

where $\omega_n = n\omega_0 = 2\pi n \Delta f$; $\omega_0 = 2\pi \Delta f$, $n = 0 \ldots (N-1)$, Δf is the frequency separation between adjacent subcarriers.

The amplitude pulse $a(t)$ can have an arbitrary shape and duration, which are determined by the requirements for the degree of compactness of the oscillation spectrum. In the simplest case, $a(t)$ is rectangular and can be represented by the following Formula:

$$a(t) = \begin{cases} 1, \ t \in [0, \ T] \\ 0, \ \text{otherwise} \end{cases}. \tag{17.2}$$

One of the main characteristics of SEFDM signals is the normalized frequency spacing of adjacent subcarriers $\alpha = \Delta f \cdot T$. For classical OFDM (Orthogonal Frequency Division Multiplexing) signals, $\alpha = 1$. For SEFDM signals, the value $\alpha < 1$, where α plays the role of the frequency division multiplexing factor. With decreasing α, the bandwidth decreases by $(1 - \alpha) \cdot 100\%$ in comparison with OFDM signals. At such values of α, mutual interference appears between signals located at subcarriers.

Generation of the SEFDM signal can be performed based on FFT/IFFT blocks [6]. The discrete k-th signal $s^{(k)}(i)$ in the sequence in (17.1) contains L time samples:

$$s^{(k)}(i) = \sum_{n=0}^{N-1} C(n) e^{j2\pi \frac{ni\Delta f}{F_s}} = \sum_{n=0}^{N-1} C(n) e^{j2\pi \frac{ni}{N}}. \tag{17.3}$$

where $i = 0 \ldots (L - 1); L \leq N, L$ is the number of time samples per duration $T, L = F.T = \alpha \cdot N$, and $F_s = N \cdot \Delta f$ is sampling rate. The number of subcarriers in the SEFDM signal is $N_{used} = N$. Then, to obtain a non-orthogonal subcarrier frequency spacing, the algorithm for generating SEFDM signals using an N-points IFFT block, one can use the method of eliminating (ignoring) a part of discrete values at the output of this block. The block diagram for generating such signals is shown in Fig. 17.1. From the output of the digital module, the complex samples of the signals s_i are fed to the analog signal conversion module for transmission at the carrier frequency ω_c. Then the real (Re) and imaginary (Im) components of digital samples are separated, and the signal spectrum is transferred to the carrier frequency. Finally, bandpass filtering (BPF) and amplification are applied to the formed bandpass SEFDM signal.

By choosing the numbers of the rejected (ignored) samples at the output of the IFFT block, it is possible to control the choice of the value of the normalized frequency separation $\alpha = \Delta f T$ of adjacent subcarriers.
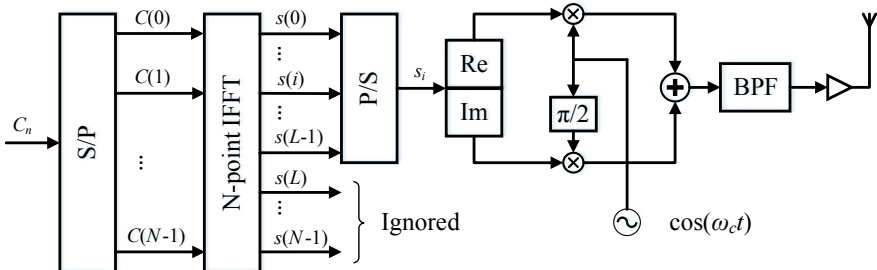
Fig. 17.1 Block diagram of a device for generating SEFDM signals

17.3   Iterative Detection Algorithm with Decision Feedback

Continuous oscillation $r(t)$ at the input of the receiving device at the analysis interval $[0, T]$ has the following form:

$$r(t) = s(t) + w(t), \tag{17.4}$$

where $s(t)$ is transmitted continuous signal, $w(t)$ is additive white Gaussian noise (AWGN). Figure 17.2 shows a block diagram of a receiving device for SEFDM signals. The analyzed process (17.4) enters the input of the receiving device. After amplification with a low-noise amplifier LNA and a band-pass filter BPF, the analyzed process is subjected to a quadrature transformation and translation of $r(t)$ to zero frequency. After low-frequency filtering and analog-to-digital conversion ADC, the sample values of the real Re and imaginary Im part of the analyzed oscillation are fed to the digital module of the N-point Fast Fourier transform FFT. At the output of the digital module, soft decisions of the transmitted symbols $\{C^*(n)\}, n = 0, 1, \ldots(N-1)$ of the channel alphabet are obtained. With a non-orthogonal spacing between subcarriers, intercarrier interference occurs, leading to a decrease in BER (bit error rate) performance of signal detection at each subcarrier frequency.

Soft decisions that appear at the output of the parallel-to-serial converter P/S can be written as:

$$C^*(n) = C(n) + \sum_{m=0, m \neq n}^{N-1} K_{mn} C^*(m), \tag{17.5}$$

where K_{mn} are correlation coefficients between signals transmitted at the m-th and n-th subcarriers. We will assume that ICI is caused only by signals transmitted on adjacent subcarriers. Soft decisions are sent to the Rangekeeper (RK) using feedback block, the block diagram of which is shown in Fig. 17.3. Let us consider the operation of this block in more detail.

First of all, a decision is made on the value of the extreme symbol $C(0)$ from the soft decision $C^*(0)$. Decision is obtained by using Rangekeeper RK (Fig. 17.3). The
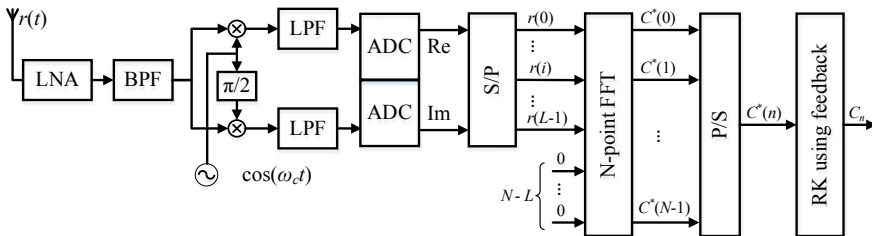
Fig. 17.2 Block diagram of a receiving device for SEFDM signals
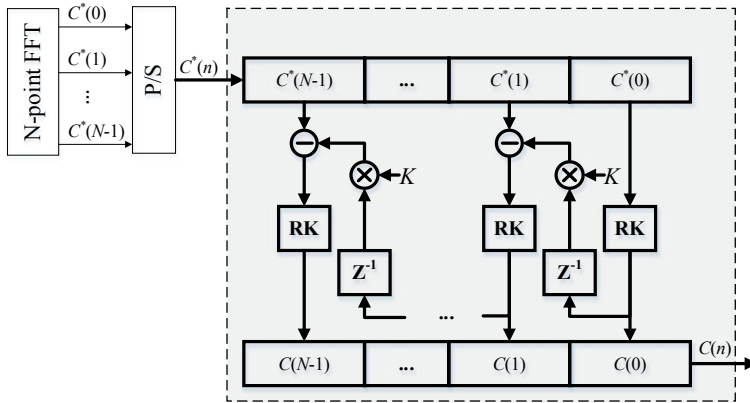
Fig. 17.3 Block diagram of the RK using feedback

decision about this symbol is delayed by the time Δt with the help of the delay line $\mathbf{Z}^{-1}$ to correct the soft decision about the second symbol $C(1)$ considering Formula (17.5). Here Δt is the time interval between the occurrence of adjacent values $C^*(n)$ in the memory register. After deciding on the second symbol, this decision participates in correcting the soft decision on the third received symbol $C(2)$. This decision feedback procedure continues until a final decision on symbol $C(N - 1)$ is obtained.

The coefficient K is calculated in advance and depends on the values of the channel alphabet symbols and the correlation coefficient of signals located at the same time interval and at adjacent subcarriers.

17.4   Analysis of the Computational Complexity of the Detection Algorithm with Feedback

Let us estimate the computational complexity of the detection algorithm with decision feedback by the number of pairs of elementary multiplication-addition or multiplication-subtraction operations. First of all, we will assess the complexity of the algorithm for coherent element-by-element signal detection when ICI is not compensated. With the number of subcarriers $N = 2^n$ computational complexity $Compl(FFT)$ for receiving one N-subcarrier SEFDM-symbol is determined by the expression [6]:

$$Compl(FFT) = N \log_2 N = Nn. \tag{17.6}$$

When using a feedback algorithm to receive signals at each subcarrier frequency, it is necessary to do one extra multiplication and one extra subtraction (Fig. 17.3). Then, the computational complexity $Compl(FFT)_{FB}$ for receiving one N-subcarrier

Fig. 17.4 The dependence
of the computational
complexity of the feedback
detection algorithm versus
number of used subcarriers
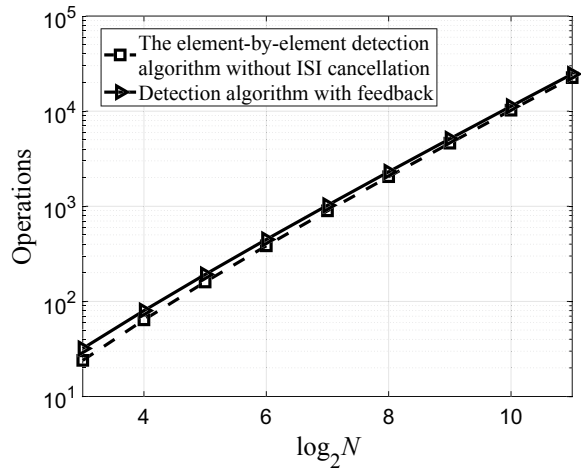
SEFDM via symbol-by-symbol detection algorithm with feedback is calculated using
the following formula:

$$Compl(FFT)_{FB} = Compl(FFT) + N = N(n + 1). \qquad (17.7)$$

The graphs of the dependence of the computational complexity of algorithms for
receiving SEFDM signals with using feedback on the number of used frequencies
N are shown in Fig. 17.4. The same figure shows the results of determining the
computational complexity when using the coherent element-by-element detection
algorithm without ICI cancellation.

From the graphs shown in Fig. 17.4, it can be seen that the use of the detection
algorithm with feedback increases the computational complexity by no more than 1%
compared to the use of the detection algorithm, which does not take into account ICI.
In addition, as can be seen from the graphs, this increase in computational complexity
decreases with an increase in the number of subcarriers used.

17.5   Simulation Results and Discussions

The purpose of the simulation is to evaluate the noise immunity of receiving multi-
frequency signals with a non-orthogonal frequency spacing and using a decision
feedback algorithm. The simulation model is built in the Matlab environment. The
simulation parameters are selected as follows:

- modulation method: BPSK;
- number of subcarriers: $N = 128$;
- bandwidth compression factor $\alpha = 0.6; 0.7; 0.8; 0.9$ and 1.0.
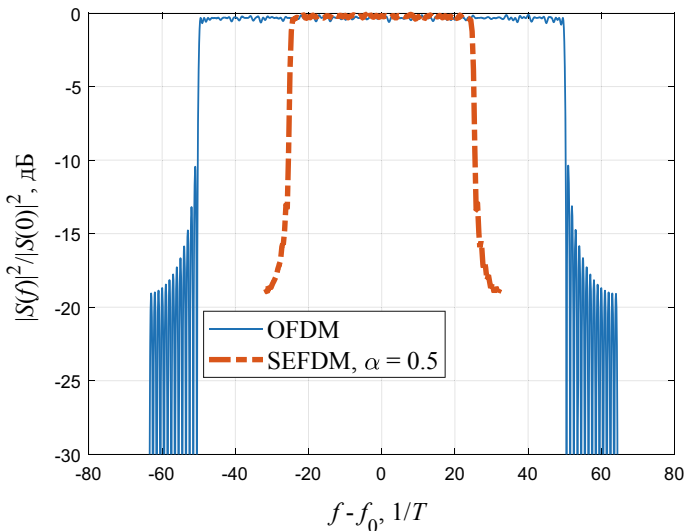
Fig. 17.5 Spectrums of OFDM signals and SEFDM signals at $\alpha = 0.5$

First of all, let us consider the obtained energy spectrum $|S(\omega)|^2$ of the analyzed multi-frequency signals. Figure 17.5 shows the spectrums of OFDM signals and SEFDM signals at $\alpha = 0.5$. It can be seen that the bandwidth of SEFDM signals decreases 2 times when compared with the bandwidth of OFDM signals.

Next, we analyze BER performance of SEFDM signal for the case of using detection algorithm with decision feedback. The dependences of the error probability on the signal-to-noise ratio (E_b/N_0, where E_b is the energy of the useful signal at the subcarrier frequency) are shown in Fig. 17.6a. From the analysis of the dependencies presented in this figure, it follows that in the case of the application of the detection algorithm with feedback on the decision, BER performance of SEFDM signal turns out to be slightly lower than BER performance of OFDM signals. The energy loss is no more than 5 dB in the range of values of the error probability BER $p = 10^{-5}$. However, when comparing these results with the BER performance of SEFDM signals when using a detection algorithm that does not take into account ICI (Fig. 17.6b), there is a significant energy gain. For example, the gain for BER $p = 10^{-3}$ is about 3 dB for SEFDM signals at $\alpha = 0.8$.

The results of simulation modeling of BER performance of receiving SEFDM signals are shown in Table 17.1. This table summarizes the energy gains achieved by using the detection algorithm with decision feedback versus using the detection algorithm, that does not take into account ISI. As can be seen from the table, for the values of the error probability $p = 10^{-2}$ and $p = 10^{-3}$, the energy gains reach values from 0.2 to 7.5 dB.
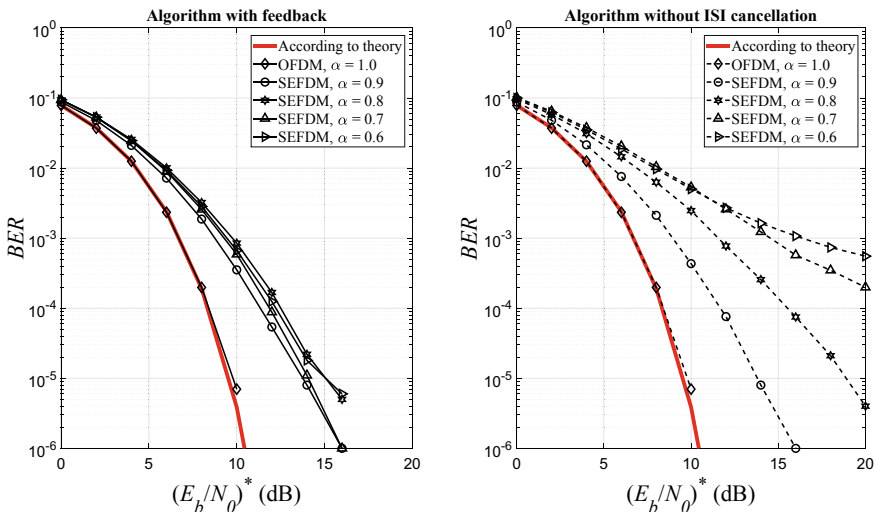
Fig. 17.6 BER performance for different values of α: Detection algorithm with feedback (**a**); Detection algorithm without ICI cancellation (**b**)

Table 17.1 Energy gains

	$\Delta f = 1/T$	$\Delta f = 0.9/T$	$\Delta f = 0.8/T$	$\Delta f = 0.7/T$	$\Delta f = 0.6/T$
$p = 10^{-2}$	0 (dB)	0.2 (dB)	2 (dB)	3 (dB)	3 (dB)
$p = 10^{-3}$	0 (dB)	0.5 (dB)	3 (dB)	5.5 (dB)	7.5 (dB)

17.6 Conclusions

In this paper, a coherent element-by-element detection algorithm with decision feedback is proposed for SEFDM signals with non-orthogonal frequency spacing.

It is shown that the proposed algorithm is quite effective. Thus, the energy gains provided by the application of the detection algorithm with decision feedback in comparison with the use of the detection algorithm, which does not take into account ICI interference, turn out to be very significant. For error probabilities $p = 10^{-2}$ and $p = 10^{-3}$, the energy gains reach values from 0.2 dB to 7.5 dB for various values of the non-orthogonal subcarrier spacing.

In addition, it is shown that the use of the detection algorithm with feedback increases the computational complexity by no more than 1% compared to the use of the conventional detection algorithm, which does not take into account ICI interference. Along with that, this increase in computational complexity decreases with an increase in the number of subcarriers used.

References

1. I. Darwazeh, H. Ghannam, T. Xu, The first 15 Years of SEFDM: a brief survey. in *11th International Symposium on Communication Systems, Networks and Digital Signal Processing* (CSNDSP), (Budapest, Hungary, 2018), pp. 1–7
2. S.V. Zavjalov, S.V. Volvenko, S.B. Makarov, A method for increasing the spectral and energy efficiency SEFDM signal. IEEE Commun. Lett. **20**(12), 2382–2385 (2016)
3. S. Gorbunov, A. Rashich, BER performance of SEFDM signals in LTE fading channels. in *41st International Conference on Telecommunications and Signal Processing* (TSP), (Athens, Greece, 2018), pp. 1–4
4. E.N. Smirnova, A.S. Ovsyannikova, S.V. Zavjalov, G. Dong, On features of implementation of SEFDM-transmitter with optimal shape of envelope. J. Phys.: Conf. Ser. **1236**(1), (2019) IOP Publishing
5. D. Vasilyev, A. Rashich, SEFDM-signals euclidean distance analysis. in *IEEE International Conference on Electrical Engineering and Photonics*, (*EExPolytech*), (St. Petersburg, Russia, 2018), pp. 75–78
6. A.B. Kislitsyn, A.V. Rashich, N.N. Tan, Generation of SEFDM-signals using FFT/IFFT. in *International Conference on Next Generation Wired/Wireless Networking*, (Springer, Cham, 2014), pp. 488–501
7. T. Wu, R. Grammenos, Reduced complexity maximum likelihood detector for DFT-s-SEFDM systems. in *2019 27th European Signal Processing Conference* (*EUSIPCO*) (IEEE, 2019), pp. 1–5
8. S. Isam, I. Darwazeh, Simple DSP-IDFT techniques for generating spectrally efficient FDM signals. in *IET International Symposium on Communication Systems, Networks and Digital Signal Processing* (IEEE, 2010), pp. 20–24
9. S. Isam, I. Darwazeh, IDFT-DFT techniques for generating and receiving spectrally efficient FDM signals. Amer. J. Eng. Appl. Sci. (2010)
10. A. Rashich, A. Kislitsyn, D. Fadeev, T.N. Nguyen, FFT-based trellis receiver for SEFDM signals. in *2016 IEEE Global Communications Conference* (*GLOBECOM*) (IEEE, 2016), pp. 1–6
11. A. Kislitsyn, A. Krylov, A. Rashich, Experimental evaluation of SEFDM trellis based demodulator. in *2020 IEEE International Conference on Electrical Engineering and Photonics* (*EExPolytech*) (St. Petersburg, Russia, 2020), pp. 110–113
12. A. Gelgor, V.P. Nguyen, Outperforming conventional OFDM and SEFDM signals by means of using optimal spectral pulses and the M-BCJR algorithm. in *26th International Conference on Telecommunications* (*ICT*), (Hanoi, Vietnam , 2019), pp. 130–134
13. S.V. Zavjalov, S.B. Makarov, S.V. Volvenko, Nonlinear coherent detection algorithms of nonorthogonal multifrequency signals. in *International Conference on Next Generation Wired/Wireless Networking*, (Springer, Cham, 2014), pp 703–713
14. I. Kanaras, Spectrally efficient multicarrier communication systems: signal detection, mathematical modelling and optimization. University College London (2010)
15. S. Ahmed, I. Darwazeh, Inverse discrete Fourier transform-discrete Fourier transform techniques for generating and receiving spectrally efficient frequency division multiplexing signals. Am. J. Eng. Appl. Sci. **4**, 598–606 (2011)
16. M. Jia, Z. Wu, Z. Yin, Receiver design combining iteration detection and ICI compensation for SEFDM. J. Wireless Com. Netw. **2018**(30), 1–11 (2018)

Chapter 18
Development of a Universal System for Three-Dimensional Data Visualization

Maksim A. Lopatin, Anatolii A. Briushinin, Danila D. Savin, and Irena M. Gureeva

Abstract The necessity of using an information processing system that uses a universal approach to data visualization is justified. An example of such an approach is proposed. The structure of the system and the logic of the selected solution are considered in detail. The input data format is proposed, which brings any information to a single form, which allows using the same algorithm for three-dimensional visualization of various types of information. A software structure has been developed for automatic video generation in 3D/2D format with geometric objects, information about which is transmitted in a JSON file with a special structure. The data visualization approach was implemented as an add-on in the Blender 3D modeling environment. Examples of the application of the developed software are given. Possible limitations and prospects of the Blender software application for three-dimensional data visualization are considered. The assessment of the used approach is carried out.

Keywords 3D visualization · Data processing · Three-dimensional modeling · Data visualization

18.1 Introduction

We receive most of the information coming from outside through visual receptors (about 80%), therefore, the visual perception channel is the most informative [1–9]. Therefore, the desire to visualize any processes, phenomena, or actions of any laws is a logical solution that is actively used in physics, information technology, and other sciences [9–16]. At the beginning, two-dimensional static images were used as visualization methods, whether they were drawings, diagrams, graphs, etc. With

M. A. Lopatin (✉) · A. A. Briushinin · D. D. Savin · I. M. Gureeva
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: maksim.lopatin.spb@mail.ru

I. M. Gureeva
e-mail: gureeva.im@edu.spbstu.ru

the development of technologies and the emergence of other ways of visual representation of information, such as three-dimensional models and videos (including in 3D format), it became possible to create specialized visualization programs [17]. Creating a universal visualization system can significantly speed up and simplify the process of analyzing information [6, 7, 12–14, 18–24].

The construction of a universal data visualization system can be based on the creation of a complex product (a certain program) consisting of many blocks, each of which solves a specific visualization task. This approach, for example, is implemented by AnyLogic. Its libraries are collections of elements created for solving a specific modeling problem or describing an application area [22, 25–29]. However, with this approach, it is impossible to foresee all the possible tasks that the program will have to solve.

The paper suggests a different approach for the formation of a universal visualization system. It is based on the principle of pre-processing information for visualization in order to further use a single program algorithm. The accompanying goals of this project are to create a structure, as well as to implement software that would not require the user to have knowledge of modeling and three-dimensional graphics technologies to obtain the desired result.

18.2 An Approach that Improves the Versatility of Solving the Data Visualization Problem

Data can have a different set of characteristics, and because of this heterogeneity, the problem of finding a universal solution for visualization arises. To solve this problem, it is necessary to bring all the incoming information to a single form.

Unification can be done using a special input data format, which is a file created either with the help of a generator or manually (Fig. 18.1). In any of these cases, the output must be a file built according to a single template. This will ensure the versatility of the input data.
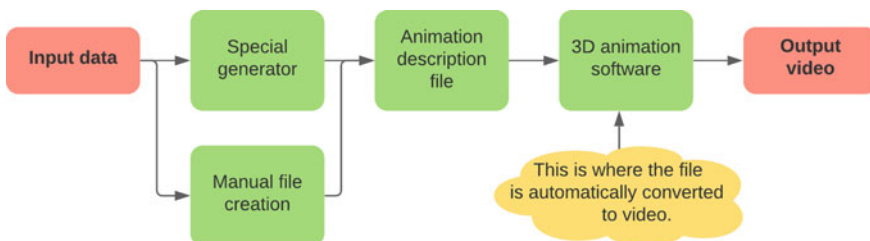
Fig. 18.1 Logical scheme of the solution

In this case, using only one algorithm, we can visualize any action, process, or object that the user needs, from a mathematical graph to a physical map of the area [25–28].

Initially, it was necessary to make the input data available in some universal form that is easy to interpret. For this purpose, we chose the way to arrange the coordinates and characteristics (shape, color, angle of inclination) of various graphic primitives (cube, sphere, plane) in a JSON file.

After reading the contents of such a file, the objects are placed in the scene. Further there should be some dynamics, that is, direct movement of objects and (or) changes in their characteristics (shape, color, angle of inclination).

The dynamics can be placed in the input file, prescribing changes in position, color, and size. Or you can make it so that the file is constantly read again. In this paper, the choice fell on the first option as the easiest to use, the file with the input data can be made manually and stored somewhere.

18.3   Software Package Structure

Now let us take a closer look at the software structure consisting of several modules (blocks) (Fig. 18.2). The module for processing input data taken from the JSON file decrypts the contents of the file and, at the output, tells the next module (the placement module) which primitives to take and where to place them. The placement module interprets this information and places the objects. After placing the elements, some dynamics should occur. Therefore, we need a modification module that is responsible for directly moving objects and changing their characteristics. The visualization module or the video format conversion module is responsible for getting a video showing the data visualization at the output.
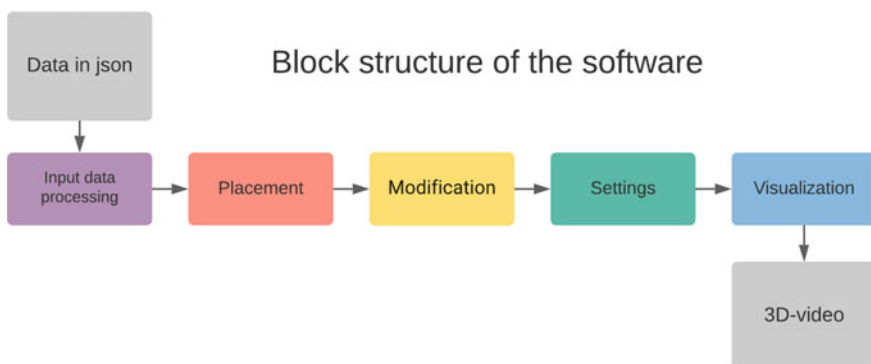
Fig. 18.2   The software block diagram

18.4 Practical Implementation

For the practical implementation of the algorithms described above, we decided to use the Blender software version 2.90.1. A JSON file is used for storing data, and Python 3 was chosen as the programming language.

Modules for processing input data, placement, and modification are implemented in a single file. The program writes data from a file to RAM, creating a list of dictionaries. Each dictionary contains two fields: "cur_frame" (the frame number) and "obj" (the set of objects that are on the scene in this frame) (Fig. 18.3).

Then, based on the names of objects recorded in the list of dictionaries (the "id" field), the program places objects on the scene and, passing through each specified frame, sets changes in the shape, color and position of objects in space. After that, the objects are placed in a special 3D view window, where you can view their placement and external characteristics in detail (Fig. 18.4).

The video format conversion module is already built into Blender, so we don't need to implement it manually. The software we use allows you to create an image in 3D format, in particular, a Side-by-Side. In contrast to 2D visualization, twice as many images (frames) will be generated here, so the speed of visualization is on the first plan. There are several rendering engines available in Blender, but the

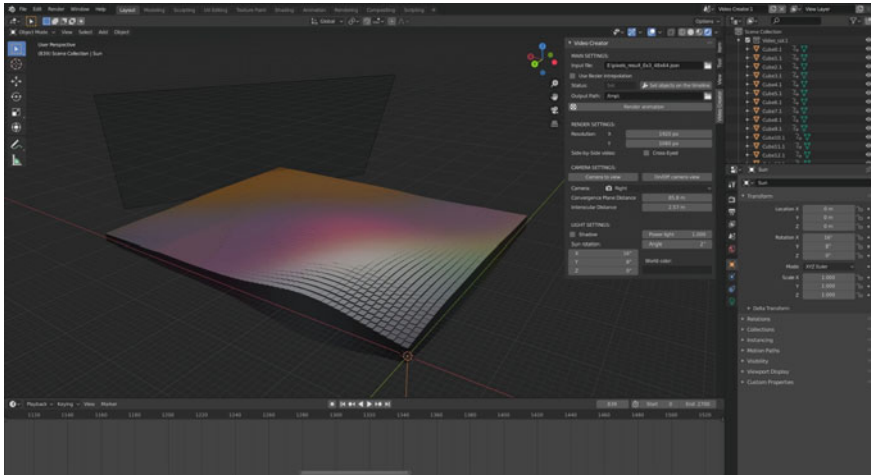**Fig. 18.3** Example of a JSON file describing the behavior of objects

Fig. 18.4 Part of the main working area when interacting with the addon-3D view windows

most balanced one is EEVEE. It is as fast as possible and despite the fact that it is a simulation of the real world (it is not a ray tracer, like Cycles), it retains a decent image quality.

The settings module is implemented in the form of a graphical interface and is divided into two parts: 1-basic and 2-additional (Fig. 18.5). Thanks to this solution, a balance is established between simplicity and functionality—the user only needs to interact with the upper field of the panel. The lower field of the panel number 2, in turn, is used for more flexible settings of both the scene and the output video.

Let us consider an example of using the developed software: there is a set of images (10 pieces), as well as a generator that receives images of the same size (in pixels) as the input and creates a file of the required format. As a result of the work of all modules a three-dimensional heat map is created, where each pixel is a rectangular parallelepiped with the color of the corresponding pixel. Its height depends on the brightness of the color (Fig. 18.6).

18.5   Evaluation of the Using Approach

Having received the finished software product, we faced some limitations. The first is the large weight of the input JSON file in the case of a large number of objects in the scene or with a long animation duration. The second limitation is related to performance—in Blender, reading a file in JSON format with a large number of objects (more than 1000) and their subsequent placement does not occur instantly (that is, unnoticed by the user) and can take minutes on a medium-performance computer.
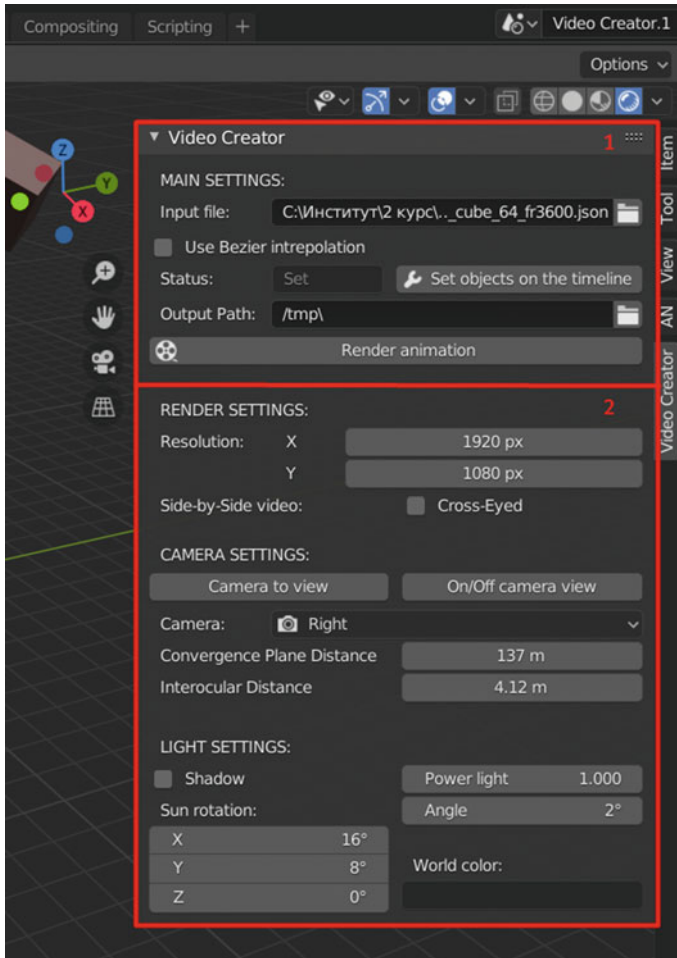
Fig. 18.5 Addon interface panel

As a result, the product we created can handle a small number of objects (up to about 3000) without much discomfort for the user and is convenient for visualization, in which the model for visualization is easy to make up of simple objects (for example, a three-dimensional heat map made up of dynamically changing color and size of rectangular parallelepipeds).

It is worth noting that we used a "static" approach, in which a file for reading in Blender is written in advance, and after opening it in Blender, the final video goes through the rendering procedure for some time; a ready-made video file in mp4 format is created. The "dynamic" approach is to broadcast the data live. Blender is not suitable for this; you can use Unity for this approach. With this approach, you will need to develop a special file for feeding real—time data to the input.
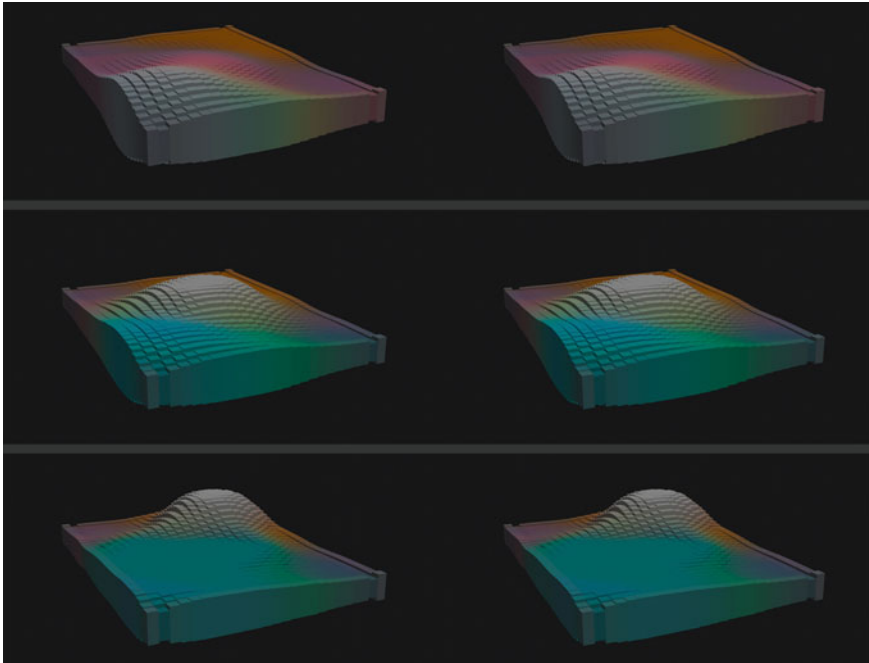
Fig. 18.6 A sequential sample of 3 frames of video in the Side-by-Side format (stereo pair), a JSON file with data was created by a special generator and is a three-dimensional heat map that changes over time

18.6   Conclusion

The task of data universalization is difficult due to the fact that it is necessary to maintain a balance between simplicity and diversity of input data. When combining these concepts, the preponderance always occurs in the direction of one thing. The results showed that the developed system managed to achieve a balance between simplicity and functionality. On the one hand, the creation of a graphical interface provides the convenience of using the developed software. On the other hand, the proposed input data format makes it easy to interpret the data for visualization and compare the characteristics of geometric objects with them. The characteristics of geometric objects are sufficient to ensure high functionality due to a large degree of freedom in transferring the characteristics of the studied information to the properties of geometric objects. It is also important to note that the visualization produced by our program is incremental. That is, the user has the opportunity to observe the exact result of the visualization and adjust the animation before starting the rendering process.

The demonstrated approach to data visualization is viable and can be used in practice. The software and documentation for it prepared as part of the implementation of this work can be found in our repository on GitHub [30].

References

1. M. Liu, W. Xue, P. Jia, S.B. Makarov, B. Li, Research on spectrum optimization technology for a wireless communication system. Symmetry **12**(1), 34 (2020)
2. J. Qi, S.B. Makarov, M. Liu, B. Li, W. Xue, Research on an optimization method for a partially responsive continuous phase modulated (CPM) signal based on an optimal generic function. Symmetry **11**(9), 1114 (2019)
3. N.A. Lukashev, Compact microwave frequency standard on Hg-199 ions for navigation systems. J. Phys.: Conf. Ser. **1236**(1), 012068 (2019)
4. R.V. Davydov, M.S. Mazing, V.V. Yushkova, A.V. Stirmanov, V.Y. Rud, A new method for monitoring the health condition based on nondestructive signals of laser radiation absorption and scattering. J. Phys.: Conf. Ser. **1410**(1), 012067 (2019)
5. J. Jang, S. Seo, W.-G. Ahn, J. Lee, J. Park, Performance analysis of an interference cancellation technique for radio navigation. IET Radar Sonar Navig. **12**(4), 426–432 (2018)
6. A.-O.A.H. Ali, in *Algorithm for positioning in non-line-of-sight conditions using unmanned aerial vehicles.* Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS, vol. 11118, (2018), pp. 496–508
7. A.A. Petrov, Some directions of quantum frequency standard modernization for telecommunication systems. in *Lecture Notes in Computer Science* (*including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*) LNCS, vol. 11118, (2018), pp. 641–648
8. K.J. Smirnov, InP/InGaAs photocathode for hybrid SWIR photodetectors. J. Phys.: Conf. Ser. **1368**(2), 022073 (2019)
9. K.J. Smirnov, S.F. Glagolev, G.V. Tushavin, High speed near-infrared range sensor based on InP/InGaAs heterostructures. J. Phys.: Conf. Ser. **1124**(1), 022014 (2018)
10. A.N. Ermolaev, G.P. Krishpents, M.G. Vysoczkiy, Compensation of chromatic and polarization mode dispersion in fiber-optic communication lines in microwave signals transmittion. J. Phys.: Conf. Ser. **741**(1), 012071 (2016)
11. N.M. Grebenikova, A.V. Moroz, M.S. Bylina, M.S. Kuzmin, Remote control of the quality and safety of the production of liquid products with using fiber-optic communication lines of the internet. IOP Conf. Ser.: Mater. Sci. Eng. **497**, 012109 (2019)
12. Y. Gao, H. Yu, S. Li, C. Yang, Acquisition method of Loran-C signal based on matched filter. in *Joint Conference of the IEEE International Frequency Control Symposium and the European Frequency and Time Forum*, (2015), pp. 265–269
13. Y. Qin, Y. Chen, Signal processing algorithm of ship navigation radar based on azimuth distance monitoring. Int. J. Metrol. Quality Eng. **10**, 12–19 (2019)
14. N.S. Myazin, Y.I. Neronov, V.I. Dudkin, On the need for express control of the quality of consumer goods within the concept 'Internet of things' . IOP Conf. Ser.: Mater. Sci. Eng. **497**(1), 01211 (2019)
15. N.S. Myazin, V.I. Dudkin, N.M. Grebenikova, V.Y. Rud', A.S. Podstrigaev, Fiber–optical system for governance and control of work for nuclear power stations of low power. in *Lecture Notes in Computer Science* (*including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*) LNCS, vol. 11660, (2019), pp. 744–756
16. A.A. Moroz, A new scheme for transmitting heterodyne signals based on a fiber-optical transmission system for receiving antenna devices of radar stations and communication systems. in

Lecture Notes in Computer Science (*including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*) LNCS, vol. 11660, (2019), pp. 710–718

17. Professional 3D visualization systems (2020). http://3d.svega-computer.ru/ Last Accessed 22 Mar 2020

18. E. Savchenko, E. Velichko, New techniques for measuring zeta-potential of colloidal system. in *Saratov Fall Meeting 2018: Optical and Nano-Technologies for Biology and Medicine. International Society for Optics and Photonics*, vol. 11065, (2019), pp. 110651U

19. Savchenko, E.A., A.N. Skvortsov, E.N. Velichko, E.Y. Savchenko, Determination of electrophoretic mobilities by DLS: homodyne versus heterodyne setup. in *2019 IEEE International Conference on Electrical Engineering and Photonics* (*EExPolytech*) (IEEE, 2019), pp. 315–317

20. S.E. Logunov, M.G. Vysoczky, V.V. Davydov, O.A. Titova, On the possibility of using the photometric method for monitoring the position of a mobile marine object. in *Proceedings of the 2018 IEEE International Conference on Electrical Engineering and Photonics*, (*EExPolytech* 2018), vol. 8564402, (2018), pp. 282–284

21. S.E. Logunov, M.G. Vysoczky, O.A. Titova, Peculiarities of registration of magnetic field variations by a quantum sensor based on a ferrofluid cell. J. Phys.: Conf. Ser. **1135**(1), 012069 (2018)

22. A.V. Moroz, K.Y. Malanin, A.A. Krasnov, V.Y. Rud, Features of the construction of the noise compensation circuit of a small-sized active phased antenna array. J. Phys.: Conf. Ser. **1400**(4), 044009 (2019)

23. A.V. Moroz, K.Y. Malanin, A.A. Krasnov, Development of a compensation system based on horn antennas for an active phased antenna array. in *Proceedings of the 2019 Antennas Design and Measurement International Conference* (*ADMInC-2019*) vol. 8969090, (Saint-Petersburg, 2019), pp. 114–116

24. V.V. Davydov, N.S. Myazin, S.E. Logunov, V.B. Fadeenko, A contactless method for testing inner walls of pipelines. Russ. J. Nondestr. Test. **54**(3), 213–221 (2018)

25. K.-M. Pehme, J. Burlakovs, M. Kriipsalu, Urban hydrology research fundamentals for waste management practices. Res. Rural Develop. **1**, 160–167 (2019)

26. A. Kumar, A.P.S. Pharwaha, Dual band minkowski fractal antenna for maritime radio navigation services. Int. J. Innov. Technol. Explor. Eng. **8**(11), 500–505 (2019)

27. S. Van, A. Cheremisin, A. Chusov, F. Switala, R. Davydov, New architectural and planning solutions as a mean of ecological modernization and reconstruction of urban environment (on the example of Shanghai city, China). IOP Conf. Ser.: Earth Environ. Sci. **390**(1), 012011 (2019)

28. S. Van, A. Cheremisin, A. Glinushkin, R. Davydov, V. Yushkova, Application of new architectural and planning solutions to create an ecological city (on the example of Shanghai, China). in *E3S Web of Conferences* vol. 140, (2019), pp. 09008

29. V.V. Davydov, N.S. Myazin, V.B. Fadeenko, S.E. Logunov, Formation of the structure of NMR signal from condensed medium in a weak magnetic field. Tech. Phys. Lett. **44**(2), 153–156 (2018)

30. Official GitHub repository of the paper (MIT license) (2021). https://github.com/maksimio/Video-Creator/ Last Accessed 09 Mar 2021

Chapter 19
Extracting Human Brain Signals from the EEG Records Using LabVIEW and Advanced Signal Processing

Andrei Medvedev ⬛, **Valentina Temkina** ⬛, **Armen Makaryan,**
Eduard Sivolenko ⬛, **Babken Hovhannisyan** ⬛, **and Hamlet Ayvazyan** ⬛

Abstract Human brain signals recognition is increasingly becoming a very popular aspect for disease prediction. The main important features of EEG records are liveliness and the robustness against falsification. A comprehensive study on human brain signals based on EEG records remain the challenge. Nowadays, low-cost wireless EEG recording devices are growing very fast on the market. These devices from different vendors have a huge potential in many applications, and the user might need to test the bunch of the devices to choose the preferable one. To extract human brain signals from the complex record there is a need, at first, to remove noise, breathing, and heart beating components. In this study, we propose a method for human brain signal identification using EEG records obtained from low-cost devices using the LabVIEW library. After that, we suggest using the advanced signal processing of the human brain signals. In this article, the achieved results are very promising. A special antenna and signal processing toolchain were developed. The signal processing based on file record and playback. After playback processing, the results are separated into two main graphs: one is the breathing and heart beating components; another one is the brain signals graph. This method is very promising for predicting future areas of human brain damage.

A. Medvedev · V. Temkina (✉)
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: valenssiya@bk.ru

A. Medvedev
e-mail: medvedev@rphf.spbstu.ru

A. Makaryan · E. Sivolenko · B. Hovhannisyan
Radiophysics and Electronics, Yerevan State University, Yerevan, Armenia

H. Ayvazyan
Engineering and Physics Institute, Russian-Armenian University, Yerevan, Armenia

19.1 Introduction

This article based on independent component analysis (ICA) [1–3]. All human actions such as heartbeats, eye blinking, and breathing generate signals in the human brain. These signals have their effect on the measured brain signal. To pick out the brain signal from the rest of the signals, the ICA method was performed. For example, you can remove the heartbeat signal from the received signals and perform further analysis on the appropriate signals.

Human brain consists of millions of neurons. As a big control center, human brain transmits electromagnetic waves. The electroencephalogram (EEG) is a recording of the electrical activity of the brain from the scalp [Human brain]. The recorded waveforms reflect the cortical electrical activity. Electroencephalography (EEG) helps to acquire brain signals corresponding to various states from the scalp surface area. Human brain transmits several signals with different frequencies. In general, we have 5 active frequency ranges that are responsible for the brain activity Table 19.1.

We designed an antenna and developed software that allows us to acquire the signals coming from the human and remove unnecessary components using ICA method. Independent component analysis (ICA) [4–6] generates a new set of statistically independent multivariate (vector) time series from the original multivariate time series that are statistically dependent on each other.

ICA is a part of LabVIEW Advanced Signal Processing Toolkit, which allows us to process signals visually in time–frequency domain. We use this toolkit because it is unique and very effective. The Advanced Signal Processing Toolkit provides VIs for time frequency analysis, time series analysis, wavelet analysis, and system identification.

The Time Frequency Analysis toolkit allows us to transform received or recorded time domain signal to a linear or quadratic time–frequency domain representation. It is possible, from one hand, to extract useful information from the time–frequency

Table 19.1 The rhythms of the brain

Gamma (25–140 Hz)	Linked to normal visual consciousness and rapid-eye-movement sleep, these might help to decipher multiple sensory signals
Beta (13–32 Hz)	These are associated with normal wakeful consciousness and concentration and are suppressed during movement. Higher frequencies signal irritation, anger, nervousness, or fear. Lower frequencies indicate tiredness or sleepiness. Their activity is higher when the brain applies logical and analytical thinking to solve a task or problem. This part of brainwave spectrum is the key examination factor for this work
Alpha (8–13 Hz)	Arising in the occipital lobe, alpha waves are associated with wakeful rest with eyes closed
Theta (4–8 Hz)	The slowest brainwaves are associated with deep often dreamless sleep. Their presence implies that the body is in a process of deep regeneration
Delta (0.5–4 Hz)	The slowest brainwaves are associated with deep often dreamless sleep. Their presence implies that the body is in a process of deep regeneration

domain representation of a signal, and, from the other hand, to process a time–frequency domain representation and reconstruct a time domain signal.

The Time Series Analysis is widely used to perform preprocessing, statistical analysis, correlation analysis, spectrum estimation, and model estimation on a univariate or multivariate (vector) time series. Also, it is possible to extract useful information from the time series with different analysis methods for different applications. The Wavelet Analysis VIs perform transforms and inverse transforms between a signal and the wavelet coefficients of the signal. The aim to use Feature Extraction Vis in this article is to perform denoising, detrending, probability density function estimation, peak detection, edge detection, and ridge detection on a 1D or 2D signal [7, 8].

Using this powerful toolkit, it is possible to preprocess raw data from a dynamic system like a human brain and to develop human brain model that reflects the behavior of brain damaged parts. Extracting the human brain signal is broadly used for different needs. The aim of this paper is to localize the damaged areas, record the signals, denoise the recorded signal with deleting independent components and predict the diseases like epilepsy.

19.2 Research Methods

In this article the signal processing algorithm is based on "cocktail party effect". It is the phenomenon of the brain ability to focus one's auditory attention on a particular stimulus while filtering out a range of other stimuli, as when a person can focus on a single conversation in a noisy room. In this case as well, the listener is the developed system, which is trying to filter out a specific signal in the noisy environment. The system could identify discrete pieces of information by classifying them by salience. This effect is what allows the system to "tune into" a single frequency and "tune out" all others. This phenomenon is often described in terms of "selective attention" or "selective hearing".

There are two ICA methods: fast ICA [9, 10] and Matrix ICA [11–14]. The first method adjusts the separating matrix to maximize the negentropy. Negentropy measures non-gaussianity of a time series [15].

$$J(y_i) = [E\{G(y_i)\} - E(v)], i = 1 \ldots, n \tag{19.1}$$

where n is number of IC, y_i is independent component, v is a Gaussian variable of zero mean and unit variance, and G is any non-quadratic function specified in method. The methods could be square, cube, tanh, Gaussian, etc.

The second method calculates the auto-correlation matrix of the measurements and the auto-correlation matrix of the measurements filtered by a specified finite impulse response (FIR) filter. The last one sorts the independent components before returning them. The filter must keep some significant frequency band of the time series but cannot be an all-pass filter. You can calculate the autocorrelation with

different delays and then average them to mitigate the effect of noise. The two averaged correlation matrices form the matrix pencil. The following equation describes these steps [3]:

$$X'_t = filter(X_t), \qquad (19.2)$$

where bk is the weight, and k is the delay. bk and k are fixed values. Then, we calculate the general eigenvector of the matrix:

$$R_{xx}E = R_{xx}ED, \qquad (19.3)$$

where E is the matrix of eigenvectors and D is a diagonal matrix. When the elements of D are distinct, E is the separating matrix.

19.3 Results

The results achieved are feasible. It was possible to extract the independent components from the received signal and left only the necessary ones. The received signal is provided below (see Fig. 19.1).

It is very difficult to separate or extract useful information from the received complex signal. Even in the spectrum after FFT, we will receive huge number of frequencies and it will be impossible to handle the useful one. That could be done using advanced signal processing toolkit with bicoherency and ICA functions. The first one, the bicoherency, will separate components which are phase coupled such as the signals from brain [3, 4, 16]. Also, the bicoherency will suppress the components with phase coupling [17–22]. The bicoherency results are shown in the Fig. 19.2.
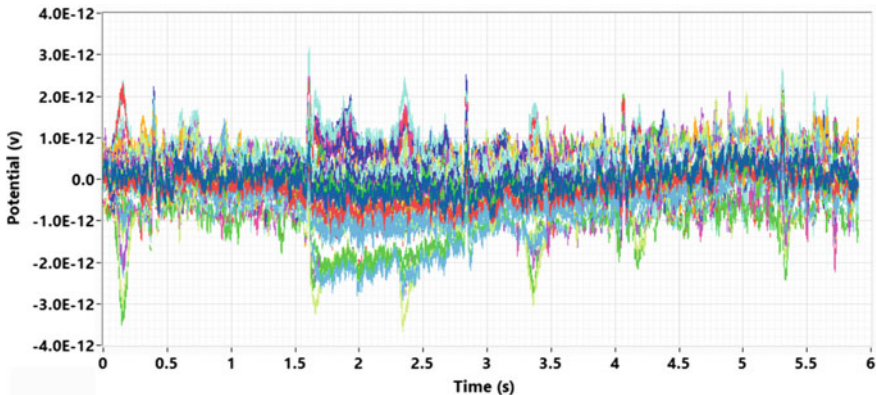
Fig. 19.1 Recorded EEG signal with all components

Fig. 19.2 Phase coupled
information extraction based
on bicoherency processing:
a full graph, **b** Z
(amplitude)-Y (frequency)
view, **c** top (X (frequency)-Y
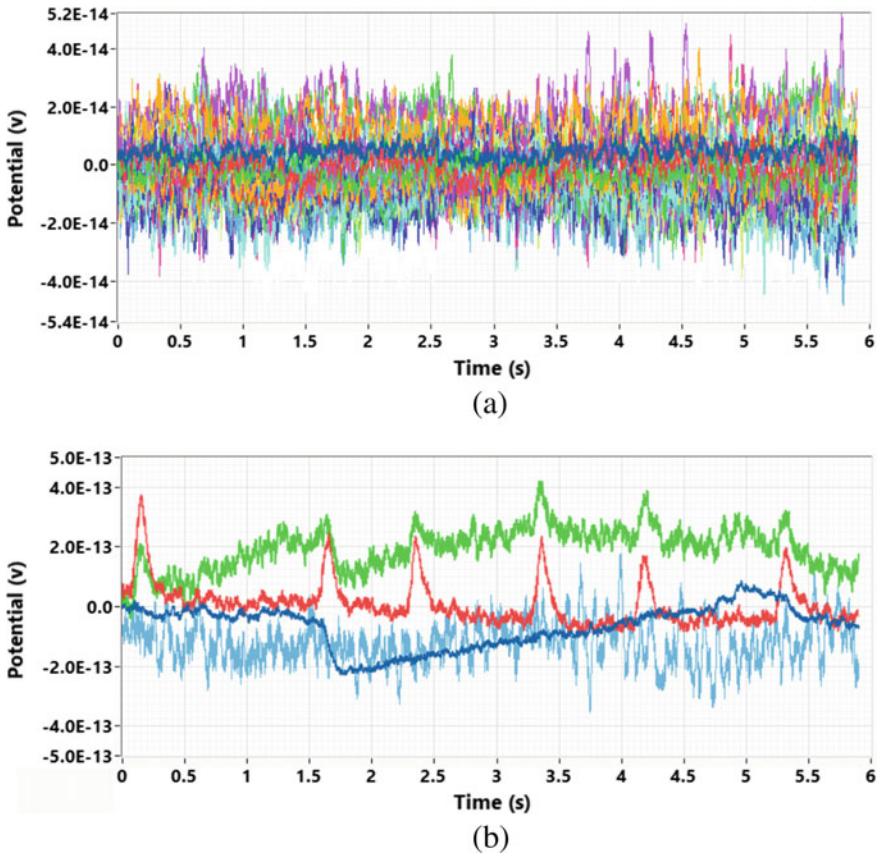(frequency)) view
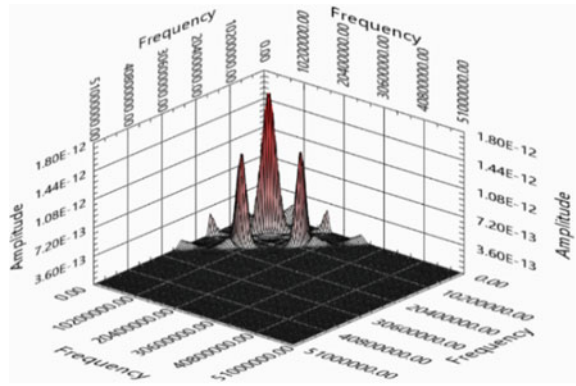
(a)
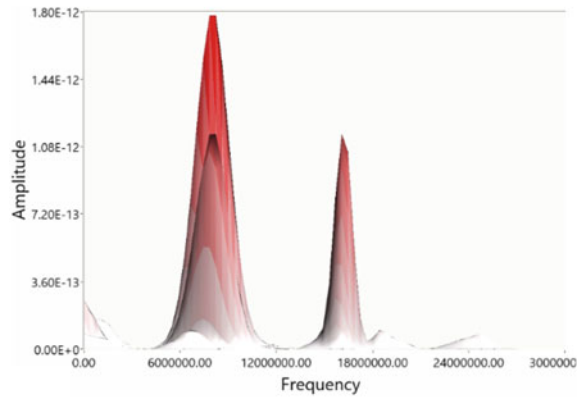
(b)
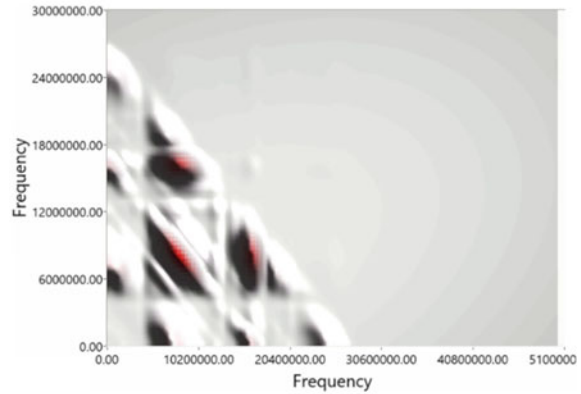
(c)

Fig. 19.3 a Reconstructed signal, **b** removed independent components

The second one will remove independent components from the received signal [23–27]. The useful signal and removed independent components are shown in the Fig. 19.3 on graphs (a) and (b) accordingly.

19.4   Conclusion

In this study, we propose a method for human brain signal identification using EEG records obtained from low-cost devices using the LabVIEW library. The results achieved are very promising. A special antenna and signal processing toolchain were developed. The results are presented above. However, the aim and novelty of this work is not the antenna and the processing toolchain. The value is the necessary form of signal extraction. The breathing and heartbeat components have been successfully removed from the main signal. This method is very promising for the prediction and

localization of the future areas of human brain damage. This method was tested with recorded data. The recorded data was received from hospital. The next step is planned to be identification and real-time signal processing.

References

1. C. Chang, Z. Ding, S.F. Yau, F.H.Y. Chan, A matrix-pencil approach to blind source separation of colored nonstationary signals. IEEE Trans. Signal Process. **48**(3), 900–907 (2000) IEEE
2. A. Hyvarinen, E. Oja, Independent component analysis: algorithms and applications. Neural Netw. **13**, 411–430 (2000)
3. E. Sivolenko, B. Hovhannisyan, A. Medvedev, AM Signal finding in the environment full of micro doppler shifts using bispectral analysis. in *2019 IEEE International Conference on Electrical Engineering and Photonics* (*EExPolytech*) (IEEE, St. Petersburg, 2019), pp. 117–119
4. S.I. Ivanov, L.B. Liokumovich, A.V. Medvedev, Estimation of the parameters of the phase modulated signal in presence of the background noise using complete sufficient statistics. in *2017 XX IEEE International Conference on Soft Computing and Measurements, SCM* (IEEE, St. Petersburg, 2017), pp. 11–13
5. A. Tharwat, Independent component analysis: an introduction. Appl. Comput. Info. 1–15 (2018)
6. M.E. Fouda, E. Neftci, A. Eltawil, F. Kurdahi, Independent component analysis using RRAMs. IEEE Trans. Nanotechnol. **18**, 611–615 (2018)
7. A.S. Barhatte, R. Ghongade, S.V. Tekale, Noise analysis of ECG signal using fast ICA. in *2016 Conference on Advances in Signal Processing*, CASP, Pune (2016), pp. 118–122
8. L. Yuan, Z. Zhou, Y. Yuan, S. Wu, An improved FastICA method for fetal ECG extraction. Comput. Math. Methods Med. **7061456** (2018)
9. J. Miettinen, K. Nordhausen, S. Taskinen, FastICA algorithms and their improved variants. R J. **10**(2), 148–158 (2018)
10. S. Nikam, S. Deosarkar, Fast ICA based technique for non-invasive fetal ECG extraction. in *2016 Conference on Advances in Signal Processing*, CASP, Pune (2016), pp. 60–65
11. K.E. Hild, S.S. Nagarajan, Source localization of EEG/MEG data by correlating columns of ICA solution with lead field matrix. in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing—ICASSP '07* (IEEE, 2007), pp. IV-1177-IV-1180
12. J. Miettinen, S. Taskinen, K. Nordhausen, H. Oja, Fourth moments and independent component analysis. Stat. Sci. **30**(3), 372–390 (2015)
13. J. Miettinen, K. Nordhausen, Oja, S. Taskinen, J. Virta, The squared symmetric FastICA estimator. Signal Process. **131**(402–411), 149–150 (2017)
14. J. Miettinen, K. Nordhausen, H. Oja, S. Taskinen, Asymptotic covariance matrices of some BSS mixing and unmixing matrix estimates (2017)
15. Z. Vahabi, S. Kermani, Desired accuracy estimation of noise function from ECG signal by fuzzy approach. J. Med. Signals Sens. **2**(3), 176–183 (2012)
16. S.I. Ivanov, L.B. Liokumovich, A.V. Medvedev, Synthesis of the demodulation algorithm for the phase modulated signals in presence of the background noise using complete sufficient statistics. in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, LNCS, vol. 11118 (Springer, 2018) pp. 666–674
17. L.B. Liokumovich, A.V. Medvedev, K. Muravyov, P. Skliarov, Signal detection algorithms for interferometric sensors with harmonic phase modulation: distortion analysis and suppression. Appl. Opt. **56**(28), 7960–7968 (2017)
18. V.D. Kuptsov, S.I. Ivanov, A.A. Fedotov, V.L. Badenko, High-precision analytical TDoA positioning algorithmfor eliminating the ambiguity of coordinates determination. in *IOP Conference Series Materials Science Engineering*, vol. 904(012013), (IOP Publishing, 2020)

19. V.D. Kuptsov, S.I. Ivanov, A.A. Fedotov, V.L. Badenko, Multi-target method for small unmanned vehicles parameters remote determination by microwave radars. J. Phys. Conf. Ser. **1515**(032045), (2020) IOP Publishing

20. A.H. Hartog, L.B. Liokumovich, N.A. Ushakov, O.I. Kotov, T. Dean, T. Cuny, A. Constantinou, F.V. Englich, The use of multi-frequency acquisition to significantly improve the quality of fiber-optic-distributed vibration sensing. Geophys. Prospect. **66**, 192–202 (2018)

21. M.E. Sobolev, A.L. Gorelik, O.L. Vlasova, A novel physico-mathematical technique of analyzing the quantitative electroencephalograms: development and application. St.Petersburg State Polytech. Univer. J.: Phys. Math. **13**(3), 119–129 (2020)

22. I. Govorov, E. Dikareva, E. Velichko, V. Kesic, E. Komlichenko, Human-machine alliance against cervical cancer. Int. J. Gynecol. Cancer **31**(2), 299 (2021)

23. A.I. Erofeev, O.A. Zakharova, S.G. Terekhin, P.V. Plotnikova, I.B. Bezprozvanny, O.L. Vlasova, An optogenetic study of the electrophysiological properties of hippocampal neurons in PS1-M146V transgenic mice (a model of Alzheimer's disease). Neurosci. Behav. Physiol. **49**(2), 199–207 (2019)

24. V.M. Mostepanenko, E.N. Velichko, M.A. Baranov, Role of electromagnetic fluctuations in organic electronics. J. Electron. Sci. Technol. **18**(1), 52–58 (2020)

25. E. Nepomnyashchaya, E. Velichko, O. Kotov, Determination of noise components in laser correlation spectroscopic devices for signal-to-noise ratio estimation. in *Proceedings of the 2019 IEEE International Conference on Electrical Engineering and Photonics* (*EExPolytech 2019*) (IEEE, St. Petersburg, 2019), pp. 321–324

26. E.A. Savchenko, A.N. Skvortsov, E.N. Velichko, A.R. Madzhhinov, S.S. Nezhinskikh, Analysis of intensity/time series obtained in homodyne evanescent wave DLS electrophoretic experiments. J. Phys.: Conf. Ser. **1236**(1), 012042 (2019), IOP Publishing

27. E.A. Savchenko, A.N. Skvortsov, E.N. Velichko, S. Elena, Determination of electrophoretic mobilities by DLS: homodyne vs heterodyne setup. in *Proceedings of the 2019 IEEE International Conference on Electrical Engineering and Photonics* (*EExPolytech 2019*) (IEEE, St. Petersburg , 2019), pp. 315–317

Chapter 20
Identification of Mortality Predictors for Elderly Patients with Airway Cancer

Irina A. Kondratyeva , **Irina I. Shpakovskaya** , **Dina V. Trotsyuk** ,
Alexander S. Krasichkov , **Anastasiya A. Polyakova** ,
and Zulfia A. Zaripova

Abstract The article analyzes the data of elderly patients with airway cancer after surgery. To reduce cancer mortality, it is necessary to identify markers that can predict worsening of a patient's condition. The study involved 83 patients, for whom 47 different parameters were recorded. The data contain not only general laboratory parameters, but also indicators of blood gas composition, hemodynamic data and autonomic indices, indicators of metabolism and cardiorespiratory exercise testing at different load levels. Statistical data analysis is carried out using the SPSS program and is based on the following methods: one-side Kolmogorov–Smirnov test, Mann–Whitney test, Chi-square test and the use of logistic regression to build a multivariate model. The target variable is the result of the operation, which can take the following values: no complications, complications after surgery, death. The result of this analysis, namely the identification of markers of mortality, will determine which of the patients need special medical supervision in the postoperative period. Goal of the investigation is to analyze the data of elderly oncology patients who underwent surgery and to identify the parameters that most affect patient mortality. This paper presents a part of the study where predictive analysis is carried out using statistical methods. In the future, the same analysis will be carried out using neural networks to compare the results and increase the reliability.

Keywords Predictive monitoring · Cancer · Statistical analyzes · Statistical model · Chi-square test · Sensitivity and specificity · SPSS

I. A. Kondratyeva (✉) · I. I. Shpakovskaya · A. S. Krasichkov
Saint Petersburg Electrotechnical University 'LETI', Saint Petersburg, Russian Federation

D. V. Trotsyuk
Federal State Autonomous Educational Institution of Higher Education, Belgorod State University, Belgorod, Russia

A. A. Polyakova
Saint Petersburg Pediatric Medical University, Saint Petersburg, Russian Federation

Z. A. Zaripova
The Pavlov First Saint Petersburg State Medical University, Saint Petersburg, Russian Federation

20.1 Introduction

According to the WHO, cancer is one of the leading causes of death in the world. Every year the number of cancer patients grows significantly, however, cancer mortality can be reduced through timely diagnosis and appropriate treatment. There are a large number of methods for predicting postoperative complications, but their share remains high even with elective surgical interventions.

Statistically, the number of surgeries among patients of older age groups is 4 times higher than the number of surgeries in younger patients [1], while the frequency of surgical interventions increases annually [2]. The outcome of surgical treatment depends on a number of factors, including the urgency of the operation, the features of the surgical technique, the changes that develop in the body under the influence of the underlying disease and concomitant pathology [3, 4].

However, many cancer patients, in addition to cancer, often have concomitant diseases [5], which are often excluded from clinical trials, therefore, it is necessary to take into account not only the basic laboratory parameters, but also a number of others: blood gas parameters, hemo-dynamics data and vegetative indices, indicators of metabolism and cardiorespiratory stress testing at different levels of stress [6, 7].

Cardiorespiratory stress testing allows you to assess the body's response to stress and assess the functional reserves of the patient. The results obtained for patients of older age groups with multiple pathologies and age-related changes in the body are of particular value in cardiorespiratory stress testing.

In this paper, methods of statistical analysis of the data of elderly patients after surgery with airway cancer are used in order to determine the markers of mortality, which make it possible to determine which of the patients need special observation by doctors in the postoperative period.

20.2 Methods

The study involved 83 patients, for whom 47 different parameters were recorded after surgery. The data contain not only general laboratory parameters, but also indicators of blood gas composition, hemodynamic data and autonomic indices, indicators of metabolism and cardiorespiratory exercise testing at different load levels.

Statistical analysis to determine potential predictors of mortality was performed using the SPSS program. Initially, the work carried out an exploratory analysis of categorical and quantitative variables of the database for the grouping variable outcome, the construction of a mathematical model and the calculation of the characteristics of the variation series: mathematical expectation, variance, standard deviation, coefficients of variation, skewness, kurtosis and other statistics. The gap analysis showed that all variables have no more than 6% missing values and can be involved in further analysis.

Further, for quantitative variables, the distribution was analyzed according to the one-sided Kolmogorov–Smirnov test [8]. Using the Kolmogorov–Smirnov criterion, the distributions were checked for normality. According to the results of the analysis, variables were selected that didn't pass this test, and variables distributed according to a uniform law were also excluded from further analysis.

For the selected quantitative variables, the establishment of interaction with the target variable was carried out by nonparametric methods according to the Mann–Whitney and Kolmogorov–Smirnov test. After this analysis, 14 quantitative variables were selected: erythrocytes, leukocytes, postoperative creatinine, total bilirubin, ALT, AST, amylase, glucose, potassium, sodium, ionized, calcium, lactate, tCO2, O2sat, which are potentially potential predictors.

For nominal variables, the correlation with the target variable is checked using the chi-square test using contingency tables [9]. As a result, the analysis did not reveal nominal variables that are potential predictors.

A one-dimensional logistic regression model was built on the basis of potential predictors. The impact of each of the potential predictors on predicting death is shown in Fig. 20.1 using the ROC curve [10–12]. Table 20.1 shows the numerical values of the ROC curve.

Figure 20.1 and Table 20.1 show that the variable oxygen saturation has the greatest influence on predicting mortality, as well as the variables amylase and inotropes.
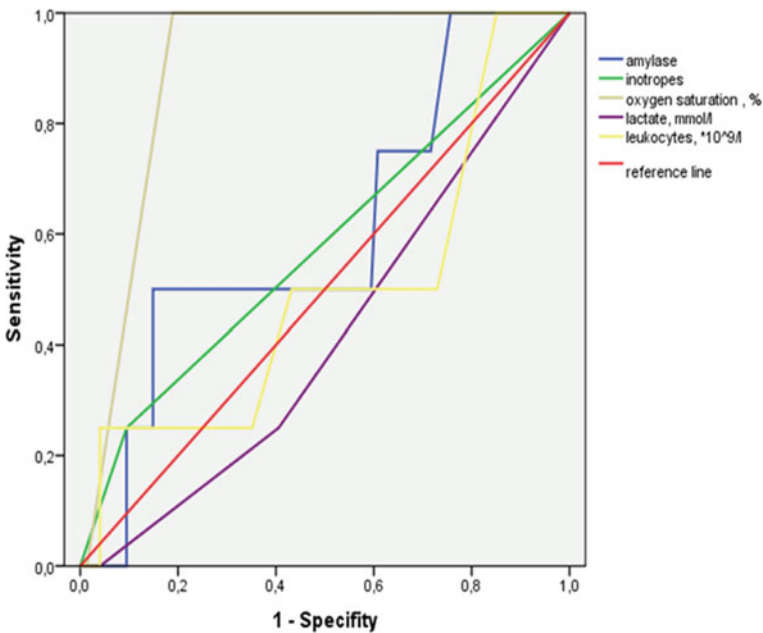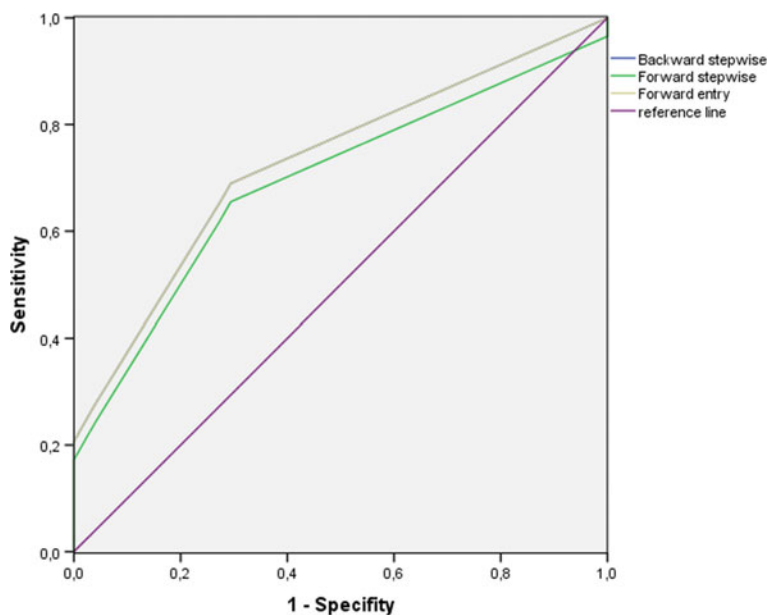

Fig. 20.1 ROC-curve for one-dimensional logistic regression model

Table 20.1 ROC-curve for one-dimensional logistic regression model values

Variable	S	Standard error	Asymptotic Significance	Asymptotic 95% confidence interval	
				Bottom line	Upper line
Amylase	0.605	0.144	0.483	0.322	0.888
Inotropes	0.578	0.160	0.602	0.263	0.892
O_2sat, %	0.899	0.040	0.008	0.821	0.976
Lactate, mmol/l	0.417	0.136	0.579	0.150	0.684
Leukocytes, $*10^9$/l	0.497	0.161	0.982	0.181	0.812

The next step in statistical analysis is to build a multivariate logistic regression model [13]. Four multidimensional models were built and analyzed: the backward stepwise method and the forward stepwise method (two options each: with and without the inclusion of a constant), as a result, the first model showed the highest reliability in terms of the model's diagnostic parameters and the obtained Nigelkirk coefficient. This version of the model was chosen for construction by the forced entry method [14–16].

The impact of each of the model on predicting death is shown in Fig. 20.2 using the ROC curve. Table 20.2 shows the numerical values of the ROC curve.

Figure 20.2 and Table 20.2 show that the variable lactate and inotroes have the greatest influence on predicting mortality.

Fig. 20.2 ROC-curve for multivariate-dimensional logistic regression model

Table 20.2 ROC-curve for multivariative-dimensional logistic regression model values

Variable	S	Standard error	Asymptotic significance	Asymptotic 95% confidence interval	
				Bottom line	Upper line
Backward stepwise	0.553	0.167	0.724	0.226	0.880
Forward stepwise	0.539	0.167	0.791	0.212	0.867
Forced entry	0.553	0.167	0.724	0.226	0.880

20.3   Conclusion

Thus, as a result of the work, a logistic regression model has been built that predicts the fatal outcome of surgery for elderly patients. Exploratory analysis and logistic regression model identified two predictors positively affecting the target outcome variable: lactate and inotropes.

The statistical approach used to achieve the stated goal is based on such analyses methods as the one-side Kolmogorov–Smirnov test, Mann–Whitney test, Chi-square test and the use of logistic regression to build a multivariate model. Based on the analysis of modern literature [1–7, 17], in which statistical studies of the parameters of patients with cancer were carried out, the results obtained in the framework of this study, taking into account the methods used, can be considered reliable.

The predictors identified in the study will help oncologists to determine whether an elderly cancer patient needs the special medical supervision after surgery. Statistical modeling is a part of a study that includes data analysis using statistical methods (the result of which is presented in this paper) and data analysis using neural networks. The use of several forecasting methods will make sure that the results obtained are reliable.

References

1. C. Naughton, R.O. Feneck, The impact of age on 6-month survival in patients with cardiovascular risk factors undergoing elective non-cardiac surgery. **61**(5), 768–76 (2007). https://doi.org/10.1111/j.1742-1241.2007.01304.x
2. T.G. Weiser, A.B. Haynes, G. Molina et al., Estimate of the global volume of surgery in 2012: an assessment supporting improved health outcomes. The Lancet **385**(S11) (2015). https://doi.org/10.1016/S0140-6736(15)60806-6
3. D.L. Vetrano, A. Calderón-Larrañaga, A. Marengoni et al., An international perspective on chronic multimorbidity: approaching the elephant in the room. J. Gerontol. A Biol. Sci. Med. Sci. **73**(10), 1350–1356 (2017). https://doi.org/10.1093/gerona/glx178
4. E. Fabbri, Y. An, M. Zoli et al., Aging and the burden of multimorbidity: associations with inflammatory and anabolic hormonal biomarkers. J. Gerontol. A Biol. Sci. Med. Sci. **70**(1), 63–70 (2015). https://doi.org/10.1093/gerona/glu127
5. D. Ha, A. Malhotra, A.L. Ries, W.T. O'Neal, M.M. Fuster, Heart rate variability and heart rate recovery in lung cancer survivors eligible for long-term cure. Respiratory Physiol. Neurobiol. **269** (2019). https://doi.org/10.1016/j.resp.2019.103264

6. K. Matsunami, K. Tomita, H. Touge, H. Sakai, A. Yamasaki, E. Shimizu, Physical signs and clinical findings before death in Ill elderly patients. Amer. J. Hospice Palliative Med.® **35**(4), 712–717 (2019). https://doi.org/10.1177/1049909117733661

7. C.C. Van't Klooster, P.M. Ridker, N.R. Cook, J.G.J.V. Aerts, J. Westerink, F.W. Asselbergs , F.L.J. Visseren, Prediction of lifetime and 10-Year risk of cancer in individual patients with established cardiovascular disease. JACC: CardioOncol (2020). https://doi.org/10.1016/j.jaccao.2020.07.001

8. D. Mizrahi, J.E. Fardell, R.J. Cohn, R.E. Partin, C.R. Howell, M.M. Hudson, L.L. Robison, K.K. Ness, J. McBride, P. Field, C.E. Wakefield, D. Simar, The 6-minute walk test is a good predictor of cardiorespiratory fitness in childhood cancer survivors when access to comprehensive testing is limited. Int. J. Cancer **147**, 847–855 (2020). https://doi.org/10.1016/j.ahj.2017.03.023

9. M.S. Anker, N. Ebner, B. Hildebrandt, J. Springer, M. Sinn, H. Riess, S.D. Anker, U. Landmesser, W. Haverkamp, S. Haehling, Resting heart rate is an independent predictor of death in patients with colorectal, pancreatic, and non-small cell lung cancer: results of a prospective cardiovascular long-term study. Eur. J. Heart Fail. **18**, 1524–1534 (2016). https://doi.org/10.1002/ejhf.670

10. A.D. Crosswell, K.G. Lockwood, P.A. Ganz, J.E. Bower, Low heart rate variability and cancer-related fatigue in breast cancer survivors. Psychoneuroendocrinology **45**, 58–66 (2014). https://doi.org/10.1016/j.psyneuen.2014.03.011

11. C. Chen, J. Zhou, H. Yu, Q. Zhang, L. Gao, X. Yin, Y. Xia, Identification of important risk factors for all-cause mortality of acquired long QT syndrome patients using random survival forests and non-negative matrix factorization. Heart Rhythm (2020). https://doi.org/10.1016/j.hrthm.2020.10.022

12. J.V. Carter, J. Pan, S.N. Rai, S. Galandiuk, ROC-ing along: evaluation and interpretation of receiver operating characteristic curves. Surgery **159**(6), 1638–1645 (2016). https://doi.org/10.1016/j.surg.2015.12.029

13. S.G. Baker, The central role of receiver operating characteristic (ROC) curves in evaluating tests for the early detection of cancer. JNCI J. National Cancer Inst. **95**(7), 511–515 (2003). https://doi.org/10.1093/jnci/95.7.511

14. Z.H. Zhang, Y.W. Han, H. Liang, L.M. Wang, Prognostic value of serum CYFRA21–1 and CEA for non-small-cell lung cancer. Cancer Med. **4**(11), 1633–1638 (2015)

15. Q. Yang, P. Zhang , R. Wu, K. Lu, H. Zhou, Identifying the best marker combination in CEA, CA125, CY211, NSE, and SCC for lung cancer screening by combining ROC curve and logistic regression analyses: is it feasible? Disease Mark. 1–12 (2018). https://doi.org/10.1155/2018/2082840

16. L. Chen, D. Lu, K. Sun, Y. Xu, P. Hu, X. Li, F. Xu, Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. Gene (2019). https://doi.org/10.1016/j.gene.2019.01.001

17. D. Trivanović, M. Petkovic, D. Stimac, New Prognostic index to predict survival in patients with cancer of unknown primary site with unfavourable prognosis. Clin. Oncol. **21**(1), 43–48 (2009). https://doi.org/10.1016/j.clon.2008.09.007

Chapter 21
FPGA Implementation of Outer Decoders for Sequential Decoder of Polar Codes

Nikita Makarov, Aleksei Krylov, Andrey Rashich, and Nguyen Ngoc Tan

Abstract The block sequential decoding of polar codes greatly decreases the latency of decoder, keeping the advantages of basic sequential decoding algorithm: low complexity and good BER performance. The decoder throughput increases due to block processing of multiple bits on each iteration using decoders for outer codes. These outer codes come from generalized Plotkin decomposition and are relatively simple for implementation. In this work we propose the decoders structures for typical outer codes and provide their implementation in FPGA. We show that the total number of look-up tables and flip-flops required by these decoders is relatively small, compared to the total resources consumption.

Keywords Sequential decoder · Plotkin decomposition · Polar codes

21.1 Introduction

Polar codes achieve the throughput of a wide class of information transmission channels, while having simple construction, encoding, and decoding procedures. Polar codes with CRC [1] and polar subcodes are superior to modern LDPC and turbo codes [2]. Therefore, the polar codes were chosen to encode the control channels in the 5-th generation mobile communication (5G).

Polar code performance highly depends on selected decoding algorithm, there are several main types of decoders: Successive Cancellation (SC) [3], Successive Cancellation List (SCL) [1], Successive Cancellation Stack (SCS) and Sequential decoder [4].

N. Makarov (✉) · A. Krylov · A. Rashich · N. N. Tan
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

A. Krylov
e-mail: krylov_ae@spbstu.ru

A. Rashich
e-mail: rashich@cee.spbstu.ru

Successive Cancellation decoder shows poor performance, but at the same time has the lowest latency and computational complexity. This decoder is applicable with limited computational complexity and low performance requirements.

Successive Cancellation List decoder shows better performance at the same latency compared to the SC decoder, but has high computational complexity and extremely inefficient use of resources for list sizes of interest. This is due to the fact that the SCL decoder consists of several parallel SC-decoders and the block which choses a number of best paths. Note that the latency of SC and SCL decoders are equal in terms of iterations number, but in general the latter decoder has longer iterations due to the sorting requirement.

SCS decoder was proposed to solve the complexity problems of SCL decoder. The main idea of SCS decoder is that the stack (or priority queue) stores information about the L most likely paths, but only one, most probable, path is processed on each iteration. This approach reduces the decoder computational complexity compared to SCL decoder, but at the same time increases the memory consumption and maximum decoding delay, because now the decoder can switch between paths of different lengths.

Sequential decoder is an improved SCS decoder, which is equipped with heuristic function [5]. Such approach dramatically decreases decoder latency without BER performance degradation. Therefore, the sequential decoder keeps the BER performance benefits of SCL decoder but has much lower complexity [5].

Given basic SC, SCL and sequential decoders one of the main goals is to increase their throughput. Here two main approaches are possible: intensive and extensive. The former includes various types of architecture unrolling and deep pipelining [6]. This increases resource consumption, while throughput grows. The latter utilizes various advanced techniques. The most popular are the following:

- GLAH [7]—G-node lookahead, allows to reduce the number of clock cycles during decoding, but requires much more memory compared to basic architecture.
- DTS [6]—double thresholding scheme, the method allows to achieve much lower latency, while maintaining low complexity, but there BER degradation, compared to the basic architecture.
- MBD [8, 9]—multi-bit decoding, this method is accompanied by decoding delays reduction, but at the same time negatively affects the resources consumption.
- Selective expansion (SE) [6]—the method is based on the properties of polar codes, each source word bit of the polar code corresponds to a synthetic channel, and different synthetic channels have different reliability. The less reliable synthetic channels are decoded using the SCL decoder, while the more reliable bits are decoded by the SC decoder. This method reduces decoding latency, but at the same time increases BER.
- Block decoding—utilizes the decomposition of polar code to a number of outer codes and decoding the codewords for one such code at each iteration [10].

In this paper we consider the outer decoders implementation to use with sequential decoder, which introduces the block sequential decoder (BSD). It is shown in [10] that the throughput of BSD outperforms the one of the SCL decoder with even small

list size (2), thus outperforming it in terms of complexity and BER performance. This article also discusses the results of the outer decoders implementation using Xilinx Kintex UltraScale + FPGA.

This paper is organized as follows. Background on block sequential decoding for polar codes is presented in Sect. 21.2. General structures of outer decoders are proposed in Sect. 21.3. Performance and resource consumption estimates are provided in Sect. 21.4. Final conclusions are drawn in Sect. 21.5.

21.2   Background on Block Sequential Decoding for Polar Codes

The sequential decoder operates on multiple paths, like an SCL decoder, but processes only one path per iteration. An information vector comes to the input of the decoder and the decoder builds a number of partial information vectors-candidates $v_0^{\varphi-1}$. One of the main blocks of the sequential decoder is the priority queue (PQ), which can store up to D partial paths $v_0^{\varphi-1}$, where φ is the current phase of the path. The paths in the priority queue are stored as tuples $\left(M, \ v_0^{\varphi-1}\right)$, where M is the path score, and sorted in descending order, i.e. the higher the score, the more likely it is that the path corresponds to the original code word.

Figure 21.1 shows LLR processing tree for $N = 8$. The input of the LLR processing tree is a vector $S_{0, \ j}$, which is a vector of modified LLR bits obtained through a classical SC decoder and vector Ps_0^{N-1}—partial sums vector, which is an paths score. Processing unit (PU)—is a universal processing unit capable of performing both P-and Q-operations on all layers of LLR processing tree [11, 12]. At each

Fig. 21.1 LLR processing tree for $N = 8$
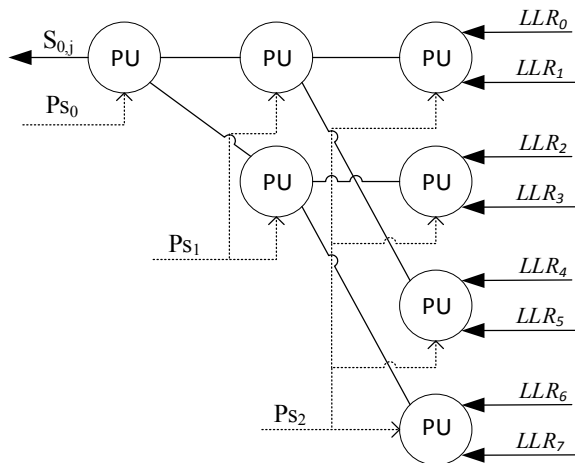
decoding phase, a new bit is evaluated, and we perform all the calculations along the decoding tree.

The main idea of the basic sequential decoding algorithm is that the polar code is recursively decomposed until codes of length 1 are obtained, i.e. decoding along the tree of calculators reaches the end, to the bit of this phase. In [10], it was shown that it is possible to reduce the complexity of sequential decoding using generalized Plotkin decomposition and outer codes decoders. The main idea of this approach is that the generalized Plotkin decomposition is recursively applied to the polar code until one obtains codes, which admit efficient decoding. I.e. when decoding through the calculation tree, we reach a certain decoding layer, get the output of the N-LLRs, and use decoders of outer codes for them.

Figure 21.2 shows an example of code decomposition and use of outer decoders. Output port B of decoders are decoded data bits and port M_{out} is ellipsoidal weight value, which is transformed path score [13, 14]. Typical codes for block decoding: $(N, 0)$, $(N, 1)$, $(N, 2)$, (N, N), SPC, first order Reed-Muller and second order Reed-Muller codes. It is important that the use of outer codes in the decoding of polar codes gives a gain in reducing the resources consumed.

In [10], presented the results of analysis using block decoding. The results of the analysis show a gain in the amount of resources consumed compared to the basic architecture of the sequential decoder.

21.3   General Structures of Outer Decoders

21.3.1   Decoder of (N, 0) Code

Decoding of $(N, 0)$ codes is performed by calculating the ellipsoidal weight (also known as the correlation discrepancy) $E\left(c^{(i)}, S_0^{n-1}\right)$, where S_0^{n-1} is the input *LLRs* vector and i—the decoding phase. The ellipsoidal weight is a converted path estimation function, given into a form suitable for use with a sequential decoder of outer codes.

The result of decoding the $(N, 0)$ code is the value of the ellipsoidal weight M_{out} calculated during decoding.

The decoder input $(N, 0)$ receives N-dimensional *LLRs* vector and ellipsoidal weight M from previous phase. Figure 21.3 shows the decoder structure for $(N, 0)$ code. The input *LLRs* are processed in N similar subblocks, which implement the following equation:

$$\text{Out } 0(x) = \begin{bmatrix} 0, & x \leq 0 \\ |x|, & x > 0 \end{bmatrix} \tag{21.1}$$

Finally, the output M_{out} of the decoder is the sum of all *LLRs* passed through *Out0* subblocks and the ellipsoidal weight from previous phase.
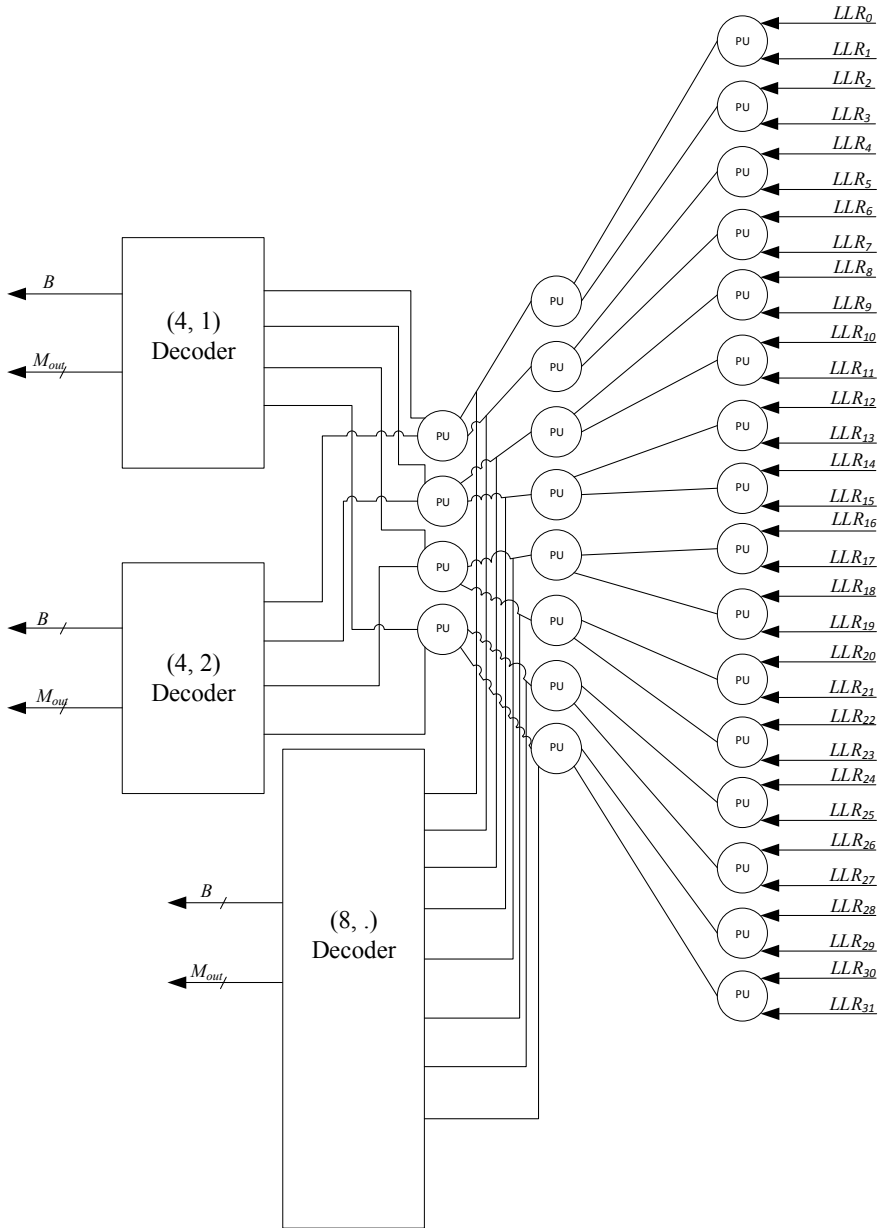
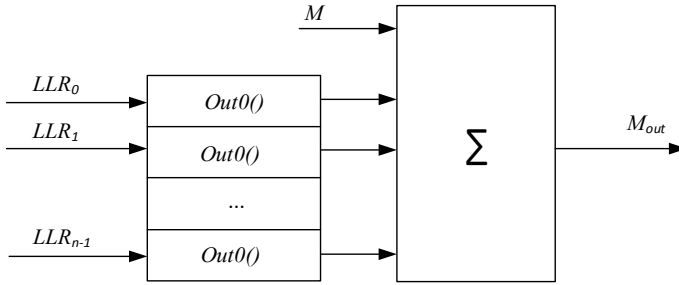**Fig. 21.2** Example of code decomposition and use of outer decoders

Fig. 21.3 Decoder structure for $(N, 0)$ code

21.3.2 Decoder of (N, 1) Code

Decoding of $(N, 1)$ code is performed by calculating the corresponding ellipsoidal weight $E\left(c^{(i)}, S_0^{n-1}\right)$ for each codeword and then sorting in ascending order $E\left(c^{(i)}, S_0^{n-1}\right)$ where S_0^{n-1} is the *LLRs* vector.

The result of decoding the code $(N, 1)$ is two single-digit code words B_0, B_1 and an ellipsoidal weight value M_{out0}, M_{out1}.

The decoder input $(N, 1)$ receives N-dimensional *LLRs* vector and ellipsoidal weight M from previous phase. Figure 21.4 shows the decoder structure for $(N, 1)$ code. The input *LLRs* are processed in two blocks, the first block consists of N similar subblocks, which implement equation in Eq. 21.1, second block consists of N similar subblocks, which implement the following equation:
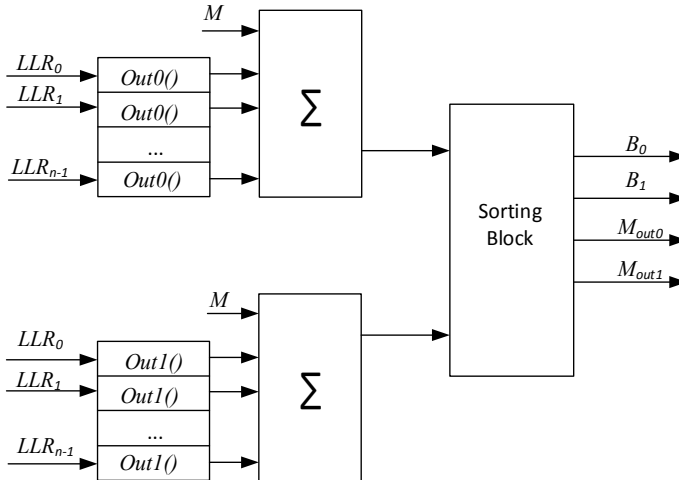
Fig. 21.4 Decoder structure for $(N, 1)$ code

$$\text{Out } 1(x) = \begin{cases} 0, & x > 0 \\ |x|, & x \le 0 \end{cases} \qquad (21.2)$$

The all *LLRs* passed through *Out0* subblocks are added together and all *LLRs* passed through *Out1* subblocks are added too. In the next step, the resulting values are sorted in order of decreasing likelihood, i.e. the larger value, the higher priority, and added to value of ellipsoidal weight from the previous phase. Finally, the output of the decoder is two single-bit codewords, sorted in order of decreasing likelihood, and their corresponding ellipsoidal weight.

21.3.3   Decoder of (N, 2) Code

As in the case of decoding $(N, 1)$ code, decoding $(N, 2)$ code is performed by calculating the corresponding ellipsoidal weight $E\left(c^{(i)}, S_0^{n-1}\right)$ for each code word and then sorting in ascending order $E\left(c^{(i)}, S_0^{n-1}\right)$.

The result of decoding the code $(N, 2)$ is four two-bit code words B_0, B_1, B_2, B_3 and an ellipsoidal weight value M_{out0}, M_{out1}, M_{out2}, M_{out3}.

The decoder input $(N, 2)$ receives N-dimensional *LLRs* vector and ellipsoidal weight M from previous phase. Figure 21.5 shows the decoder structure for $(N, 2)$ code. The input *LLRs* are processed in four blocks, the first two blocks perform the same functions as the code decoder blocks $(N, 1)$, the third block consists of alternating $N/2$ sub-blocks *Out1* in Eq. 21.2 and *Out0* in Eq. 21.1, the fourth of alternating $N/2$ sub-blocks *Out0* and *Out1*. The *LLRs* passed through each block are added together. In the next step, the resulting values are sorted in order of decreasing likelihood and added to the value of the ellipsoidal weight from the previous iteration. Finally, the output of the decoder is four two-bit code words, sorted in order of decreasing likelihood, and their corresponding ellipsoidal weight.

21.3.4   Decoder of (N, N) Code

Decoding of (N, N) code is performed by calculating the corresponding ellipsoidal weight $E\left(c^{(i)}, S_0^{n-1}\right)$ for each code word.

The decoder input (N, N) receives N-dimensional *LLRs* vector and ellipsoidal weight M from previous phase. Figure 21.6 shows the decoder structure for (N, N) code. The result of decoding the code (N, N) is L n-dimensional code words B_0, B_1, … B_{L-1}, where $B_0 = \{B_0, B_1, … B_{n-1}\}$ that are sorted in order of decreasing likelihood and the corresponding ellipsoidal weights of these code words M_{out0}, M_{out1}, … $M_{out(L-1)}$.
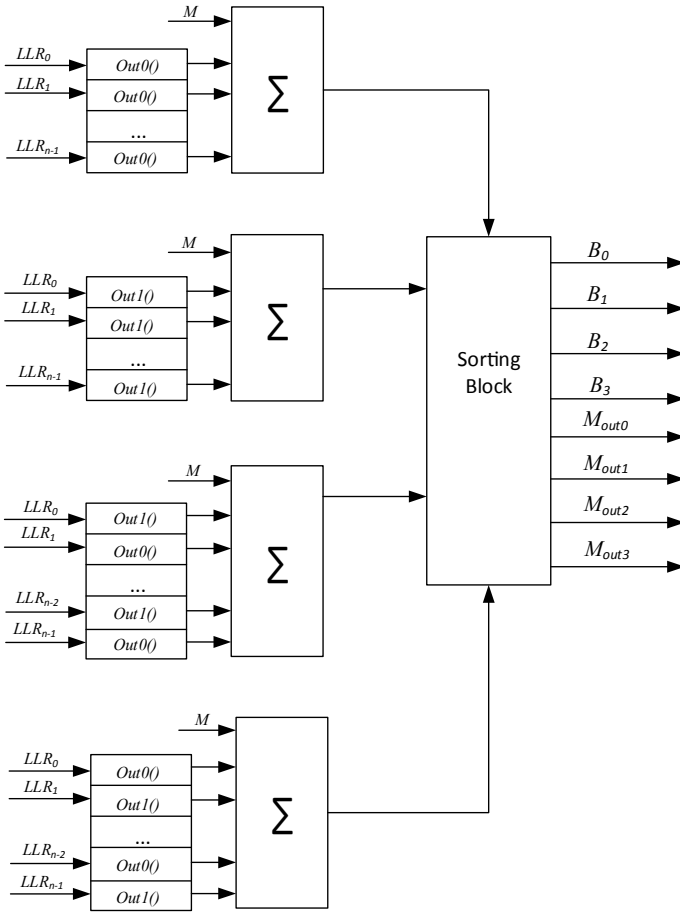
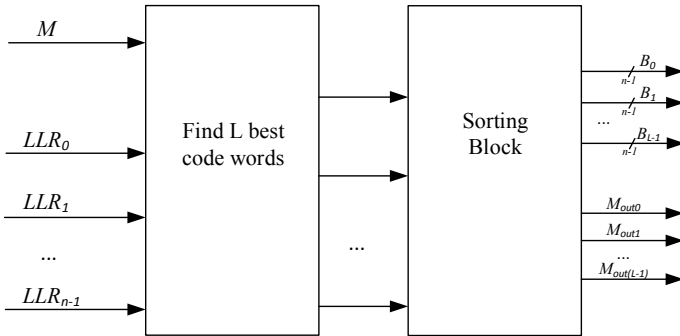**Fig. 21.5** Decoder structure for $(N, 2)$ code

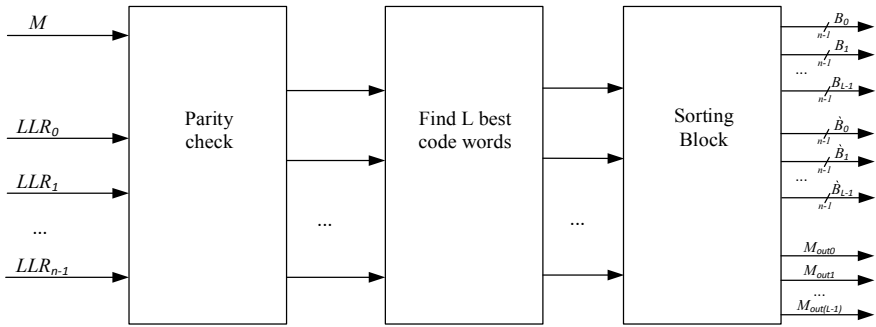**Fig. 21.6** Decoder structure for (N, N) code

Fig. 21.7 Decoder structure for SPC code

21.3.5 Decoder of Single Parity Check Code

Decoding the SPC code is similar to decoding the (N, N) code, except that when decoding the SPC code, a parity check is performed.

The decoder input SPC receives N-dimensional *LLRs* vector and ellipsoidal weight M from previous phase. Figure 21.7 shows the decoder structure for SPC code. The result of decoding the code SPC is L n-dimensional code words and depending on the result of the parity check it is $B_0, B_1, \ldots B_{L-1}$, where $B_0 = \{B_0, B_1, \ldots B_{n-1}\}$ or $\dot{B}_0, \dot{B}_1, \ldots \dot{B}_{L-1}$, where $\dot{B}_0 = \{\dot{B}_0, \dot{B}_1, \ldots \dot{B}_{n-1}\}$, that are sorted in order of decreasing likelihood and the corresponding ellipsoidal weights of these code words M_{out0}, $M_{out1}, \ldots M_{out(L-1)}$.

21.3.6 Decoder of First Order Reed-Muller Code

The decoder input first order Reed-Muller receives N-dimensional *LLRs* vector and ellipsoidal weight M from previous phase. Figure 21.8 shows the decoder structure for first order Reed-Muller code. The input *LLRs* are processed in block of Hadamard transform [15], the result of the Hadamard transformation is $2 \bullet n$ L-dimensional codewords. Then the resulting code words are sent to the sorting block. Finally, the output of the decoder is $2 \bullet n$ L-dimensional codewords, sorted in order of decreasing likelihood, and their corresponding ellipsoidal weight.

21.3.7 Decoder of Second Order Reed-Muller Code

Decoding second order Reed-Muller code is identical to decoding the first order Reed-Muller code, difference is that when decoding the second order Reed-Muller code, the repetition encoding method is used, i.e. the code word is written to the
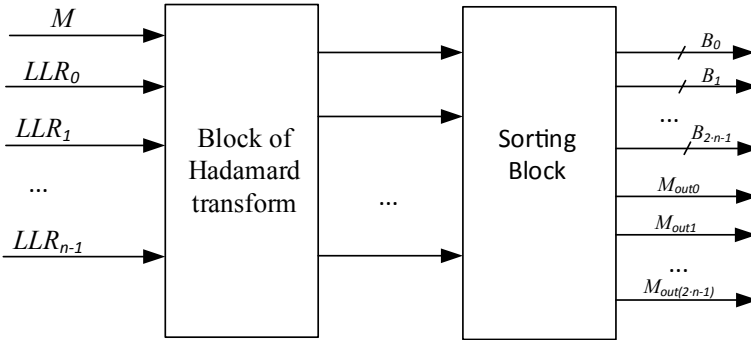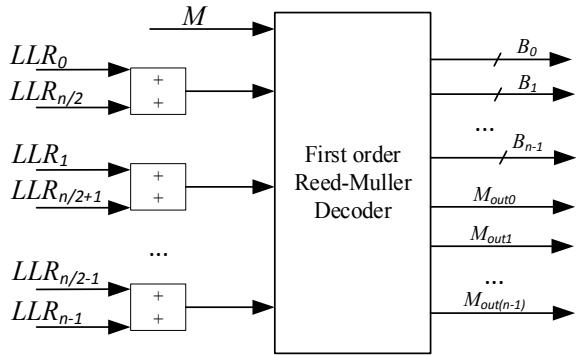
Fig. 21.8 Decoder structure for first order Reed-Muller code

Fig. 21.9 Decoder structure for second order Reed-Muller code

output stream twice. The same *LLRs* are added together and received to the input of the first order Reed-Muller decoder.

Figure 21.9 shows the decoder structure for second order Reed-Muller code. The output of the decoder is $2 \bullet n$ *L*-dimensional codewords, sorted in order of decreasing likelihood, and their corresponding ellipsoidal weight.

21.4   Implementation Results

Outer code decoders were implemented in RTL using the SystemVerilog language. The decoders architecture was developed for ASIC. Thus, the registers were not used for pipelining, as is done in FPGAs, so as not to reduce the decoder throughput. Also, for verification and testing, a pre-synthesis decoder was implemented in Kintex UltraScale + . The following outer decoders were implemented: $(N, 0)$, $(N, 1)$, $(N, 2)$, (N, N), SPC, first order Reed-Muller and second order Reed-Muller decoder.

The inputs of the decoders are defined as an 8-dimensional *LLRs* vector, the bit depth of the *LLR* is set to 38 bits, and the output value of the ellipsoidal weight

Table 21.1 Resources decoders of outer codes

Block	LUTs	FFs
$(N, 0)$ Decoder	311	42
$(N, 1)$ Decoder	829	88
$(N, 2)$ Decoder	1765	138
(N, N) Decoder	433	99
SPC Decoder	3357	100
First order Reed-Muller	3775	105
Second order Reed-Muller	2129	104

of decoders is 41 bits. The resource consumption by outers decoders and decoders subbank is presented in Table 21.1.

21.5 Conclusion

In this work we proposed the decoders structures for typical outer codes in block sequential decoder of polar codes, which arise from generalized Plotkin decomposition, and constitute the heart of block sequential decoding algorithm. We provided their implementation in FPGA. The decoders of outer codes can be arranged in a single bank, in which the decoders are switched according to a certain algorithm during sequential decoding process. The proposed decoders structures require 300–3000 LUTs and 40–140 FFs. In total, the decoder bank takes about 13,000 LUTs and 700 FFs. The decoding latency is 1–2 clock cycles. The results obtained can be used in the implementation of the decoder architecture presented in [16].

References

1. I. Tal, A. Vardy, List decoding of polar codes. IEEE Trans. Inf. Theory **61**(5), 2213–2226 (2015)
2. P. Trifonov, G. Trofimiuk, A randomized construction of polar subcodes. in *2017 IEEE International Symposium on Information Theory* (ISIT), (2017), pp. 1863–1867
3. E. Arikan, Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. IEEE Trans. Inf. Theory **55**(7), 3051–3073 (2009)
4. V. Miloslavskaya, P. Trifonov, Sequential decoding of polar codes. IEEE Commun. Lett. **18**(7), 1127–1130 (2014)
5. P. Trifonov, A score function for sequential decoding of polar codes. in *2018 IEEE International Symposium on Information Theory* (ISIT) (2018), pp. 1470–1474
6. Y. Fan, C. Xia, J. Chen, C. Tsui, J. Jin, H. Shen, B. Li, A low-latency list successive-cancellation decoding implementation for polar codes. IEEE Areas Commun. **34**(2), 303–317 (2016)
7. C. Zhang, B. Yuan, K. Parhi, Reduced-latency sc polar decoder architectures. IEEE Int. Conf. Commun, (ICC), 3471–3475 (2012)
8. B. Yuan, K. Parhi, Low-latency successive-cancellation list decoders for polar codes with multibit decision. IEEE Trans. VLSI Syst **23**(10), 2268–2280 (2015)

9. C. Xiong, J. Lin, Z. Yan, Symbol-based successive cancellation list decoder for polar codes. in *IEEE Workshop Signal Process. Systems* (SiPS), (2014), pp. 1–6
10. G. Trofimiuk, N. Iakuba, S. Rets, K. Ivanov, P. Trifonov, Block sequential decoding techniques for polar subcodes. In: CoRR. Available: http://arxiv.org/abs/1808.07268. Last Accessed 19 Feb 2021
11. C. Xia, J. Chen, Y. Fan, C. Tsui, J. Jin, H. Shen, B. Li, A high-throughput architecture of list successive cancellation polar codes decoder with large list size. IEEE Trans. Signal Process. **66**(14), 3859–3874 (2018)
12. C. Leroux, A.J. Raymond, G. Sarkis, W. J.. Gross,: A semi-parallel successive-cancellation decoder for polar codes. IEEE Trans. Signal Process. **61**(2), 289–299 (2013)
13. A. Valembois, M. Fossorier, Box and match techniques applied to soft-decision decoding. IEEE Trans. Inf. Theory **50**(5), 796–810 (2004)
14. H. Moorthy, T. Lin, S.T. Kasami, Soft-decision decoding of binary linear block codes based on an iterative search algorithm. IEEE Trans. Info. Theory **43**(3), 1030–1040 (1997)
15. J. Shanks, Computation of the fast walsh-fourier transform. IEEE Trans. Comput. 457–459 (1969)
16. A. Krylov, A. Rashich, A. Gelgor, D. Fadeev, polar codes sequential decoder hardware architecture. in *2019 42nd International Conference on Telecommunications and Signal Processing* (TSP*)*, (Budapest, Hungary, 2019), pp. 261–264

Part II
Electronics and Nanotechnologies

Chapter 22
Integrated Digital Voltage Regulator with Phase-Frequency Control for Wireless Passive Microdevices

Alexander S. Sinyukin⬤**, Boris G. Konoplev**⬤**, and Andrey V. Kovalev**⬤

Abstract A low dropout voltage regulator for low power wireless applications is presented. For implementation low voltage operation of the device and for compatibility with manufacturing processes of digital integrated circuits, all analog circuits in the regulator were replaced by digital ones. A digital phase-frequency detector with shift direction control based on XOR logic cell which compares signal phases determined by reference voltage and output voltage is developed. A digital bi-directional k-bit serial-in parallel-out shift register divided into two parts for decreasing output voltage settling time is described. An auto-activated clock generator capable to lower power consumption is implemented.

Keywords Low Dropout · Digital Control · Low Voltage Operation · Nanoscale CMOS · Internet of Things

22.1 Introduction

In recent times wireless small-sized radio-frequency devices are often used in different realms of human activities like industry, commerce or medicine. Passive microdevices are of particular interest among them since they don't contain internal power cell, e.g. storage battery, so such devices can operate much longer and their footprint is less compared to active analogues. Some of the most widespread applications of the wireless passive devices are Internet of Things (IoT) [1, 2], wireless sensor networks (WSN) [3, 4] and radio-frequency identification (RFID) [5, 6].

A. S. Sinyukin (✉) · B. G. Konoplev · A. V. Kovalev
Southern Federal University, 44, Nekrasovsky lane, Taganrog 347922, Russia
e-mail: sinyukin@sfedu.ru

B. G. Konoplev
e-mail: kbg@sfedu.ru

A. V. Kovalev
e-mail: avkovalev@sfedu.ru

Wireless passive microdevices include power supply module which provide supply voltage for entire integrated circuit of the device. One of the most important components of the power supply module is a voltage rectifier with voltage multiplying also known as RF-DC converter [7, 8]. This circuit converts the radio-frequency energy obtained from coupled base station or harvested from the environment into direct voltage and then rises rectified voltage level in order to meet the requirements of utilized technology. Another important component of power supply module is voltage regulator [9, 10]. Its purpose is to regulate and conclusive define the supply voltage based on direct voltage received from RF-DC converter. In portable stand-alone low-power devices, low dropout (LDO) voltage regulators are usually used [11–13]. They represent DC linear voltage regulators which are able to regulate output voltage (the supply voltage in our case) even if input voltage (the voltage from multiplier) is very close to this output voltage.

Generally LDO voltage regulator consists of an operational amplifier or error amplifier (EA), an analog controlled power MOS transistor and negative feedback loop as well as reference voltage source. The MOS gate is coupled to the amplifier output, the input voltage is applied to the MOS source, and its drain is connected to the regulator output, which is closed by the feedback loop with one of the EA inputs. The reference voltage is applied to another EA input and when regulated voltage varies from the reference, the EA output voltage is increasing (in case of PMOS and the rise of regulated voltage) or decreasing (in case of voltage drop). As a result channel thickness of the control MOS transistor reduces or grows respectively which leads to decreasing or increasing of the output voltage thereby the regulation is carried out.

The operation of traditional analog voltage regulators is complicated or unrealizable at low operating voltage (lower than 0.5 V) which is determined by difficulty of operational amplifier work at such voltage levels [14, 15]. For this reason nowadays voltage regulators based entirely or partially on digital components have became widespread in micropower integrated applications [16–18]. Most common disadvantages of the present LDO voltage regulators [9–20] are reliance on analog components, long output voltage settling time and in some instances high power consumption and structural complexity.

The purpose of this paper is to propose novel LDO voltage regulator design based on modified digital cells with decreased output voltage settling time and reduced power consumption, which is able to work at low voltages.

22.2 Methods

For elimination or diminishing the influence of the present voltage regulators disadvantages, the development of LDO voltage regulator based on digital components is proposed. Two voltage controlled oscillators (VCO) are used as signal shapers in the proposed development. The reference voltage is applied to the first oscillator and the output voltage which should to be regulated is applied to the second oscillator. The

phase difference of signals produced by these oscillators indicates about necessity or absence of necessity to regulate (to decrease or to increase) the output voltage.

The rectangular pulse sequences are shaped at outputs of the oscillators. In the case of the equality of the reference voltage level and the regulator output voltage level, the frequencies of these pulse sequences will be equal to each other and the phase difference will be equal to zero. If this is not the case, non-zero phase difference applied to inputs of a phase-frequency detector (PFD) with shift direction control will be a sign for changing the output voltage of the regulator.

The proposed phase-frequency detector with shift direction control is fully digital device which consists of RS-latch, OR2 logic cell and XOR (exclusive OR) logic cell. The XOR cell function is in its turn implemented by means of nine NOR2 (negative OR) logic cells. The detector has two inputs (A1, A2) and two outputs (B1, B2) and it transmits logical "1" or logical "0" to one of its outputs (B2) depending on the input signals phase difference. This phase difference is in turn determined by positive or negative difference between the reference and the output voltages. To the other output of the phase-frequency detector (B1) the clock generator activation signal is transmitted (in case of phase difference presence). Then control signal from the detector output B2 and clock signal are transmitted to a digital controller.

More simple XOR logic cell structures based on fewer NOR logic cells are known, however such structures have only one output. Since two outputs have been required for the proposed structure (for clock signal activation and for shift direction control), additional NOR logic cells have been appended. Figure 22.1 shows functional diagram of the proposed phase-frequency detector with shift direction control.
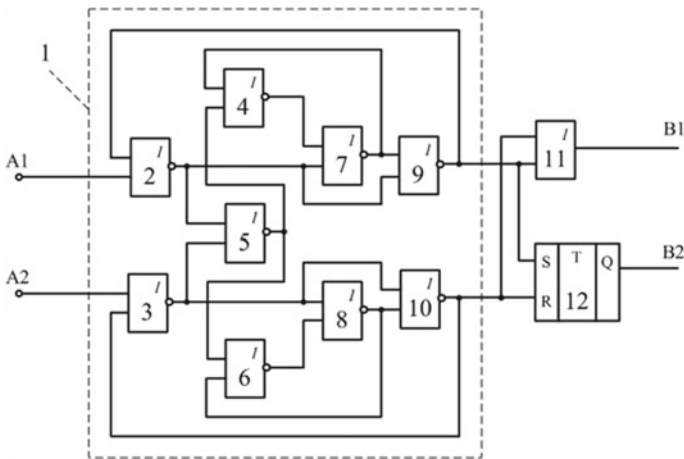
Fig. 22.1 Functional block diagram of the phase-frequency detector: A1—signal from the first VCO output, A2—signal from the second VCO output, 1—XOR logic cell, 2–10—NOR2 logic cells, 11—OR2 logic cell, 12—RS-latch, B1—signal transmitted to the clock generator input, B2—signal transmitted to the control inputs of the digital controller and carrying information about shift direction of registers
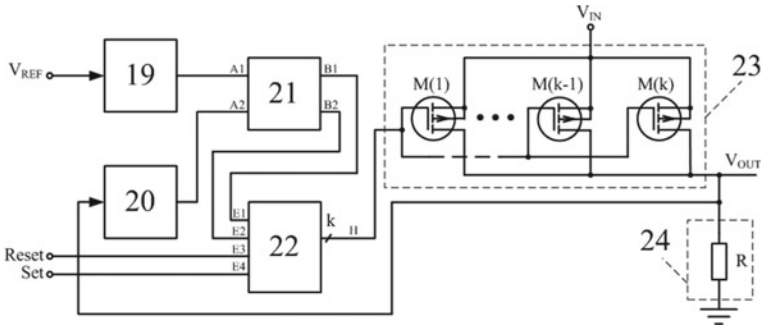
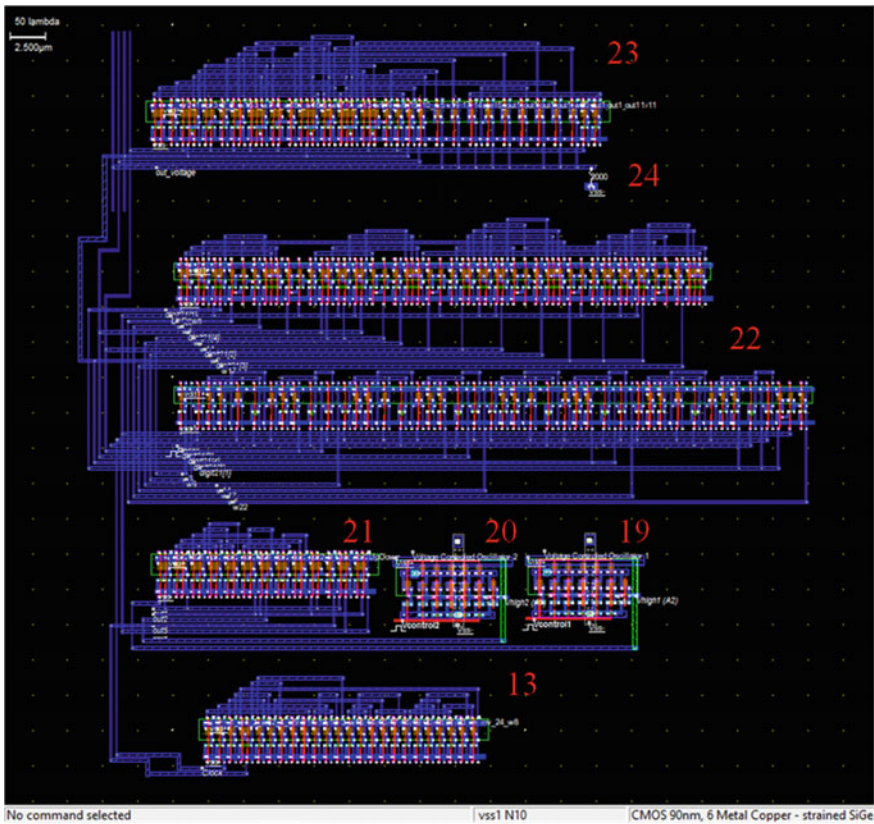**Fig. 22.3** Functional block diagram of the LDO voltage regulator

Fig. 22.4 Layout of the digital voltage regulator with phase-frequency control (CMOS 90 nm technology)

Alongside the transistors array (23) at the layout top, the decoder is placed which transforms the digital controller binary output to the thermometer code appropriate for controlling the PMOS transistors in array. In the proposed implementation, k is equal to 16.

22.3   Results

The proposed LDO voltage regulator operates as follows. Firstly preset adjustment of the shift registers and thus of the array of k PMOS transistors is performed. For this the reset signal is applied to the reset input R of n-bit shift register and the initial set signal is applied to the set input S of m-bit shift register. As a result at the outputs of the register 17, logical "0" settle which are applied to the coupled with them PMOS gates and thereby n transistors turn on. At the outputs of the register 18, logical "1" settle which are transmitted to the coupled with them transistors gates, hence m PMOS transistors turn off. This way the initial output voltage settles at the regulator output and this voltage level is a certain fraction of the input voltage, determined by m and n parameters.

Major task of the voltage regulator is to maintain the output voltage at the constant specified level equal to the reference voltage in this case. If the regulator output voltage begins to differ from reference voltage, e.g. in the case of the load current or the input voltage amplitude varying, then the phase of signal produced by the second VCO (20) deviate from the phase of signal produced by the first VCO (19). Further, signals from the outputs of the oscillators arrive to the PFD inputs.

At the input of the phase-frequency detector, XOR logic cell is placed which is based on sequential logic in the proposed implementation and formed by nine NOR2 logic cells. One can consider two cases of the output voltage deviation regarding the reference level.

In the first instance, the output voltage level rises, i.e. becomes higher than the reference voltage level, resulting in the frequency of the second oscillator (20) increases and the phase of its output signal begins to outstrip the phase of signal produced by the first VCO (19)—the phase difference emerges. The signal from the first oscillator output arrives to the first input (A1) of the detector and the signal from the second oscillator output arrives to the second input (A2) of the detector. The second input of the detector is logical "1" during the time corresponding to the phase difference of the signals from the VCOs outputs. The first input of the phase-frequency detector is logical "0" the same amount of time. Such input signals combination leads to the output of NOR2 10 logic cell is logical "1" which is transmitted to the input of OR2 11 logic cell resulting in the clock generator enables. Besides that, logical "1" from the output of NOR2 10 logic cell is applied to reset input R of the RS-latch. The logical "0" signal is produced at the RS-latch output which is applied to control inputs C of the shift registers (17, 18). Therefore the registers shift so that the number of logical "1" at their outputs increases which leads to

growth of turned-off PMOSs and eventually to the output voltage level lowering—the negative feedback loop closes.

In the second instance, the output voltage reduces regarding the reference voltage. With that the frequency of the second oscillator decreases and its signal phase start to lag behind the phase of signal produced by the first oscillator. For this reason during time interval corresponding to these signals phase difference, the first input of PHD is logical "1" while the second input of the detector is logical "0". According to the internal logic of the XOR logic cell, at such combination of input signals logical "1" signal is produced at the output of NOR2 9 logic cell which through OR2 logic cell activates clock generator and also enters to the set input S of the RS-latch. As a result, the latch output is logical "1" and it is transmitted to the control inputs of the registers. Logical "1" at the control inputs of the registers lead to such shift that number of logical "0" at registers outputs increases. Consequently the number of turned-on PMOSs rises and the output voltage grows which indicates the closing of the negative feedback loop too.

It is seen from the digital voltage regulator behaviour that clock generator is activating only when the phase difference emerges between the signals from VCO outputs to the detector inputs. In that case one of the OR2 11 logic cell inputs is logical "1" resulting in, according to truth table, logical "1" is transmitted from the output of this cell to the clock generator input thereby activating it.

Figure 22.5 shows the simulation results of the proposed digital regulator with phase-frequency control carried out using Microwind.
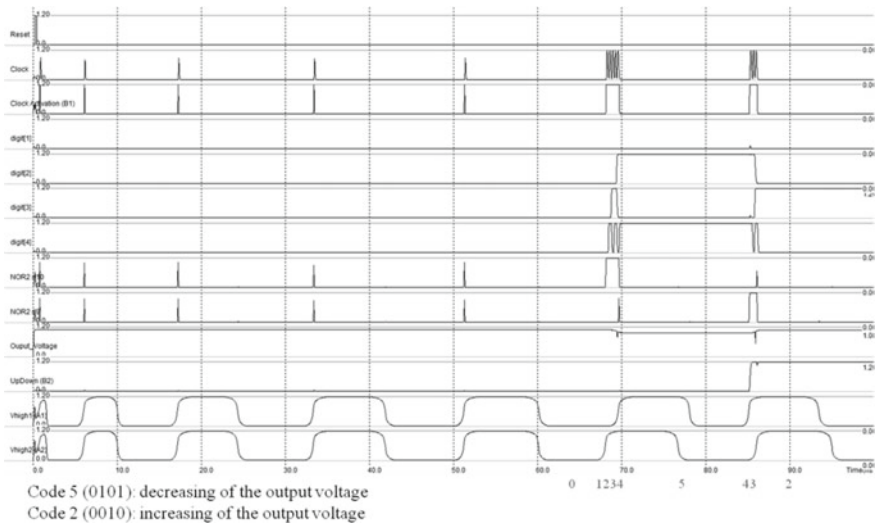

Code 5 (0101): decreasing of the output voltage
Code 2 (0010): increasing of the output voltage

Fig. 22.5 The results of the simulation of the proposed digital LDO voltage regulator

22.4 Discussion

As it is seen from Fig. 22.5, at the time instant approximately equal to 67 ns the output voltage begins to exceed the reference voltage therefore phase of the signal (Vhigh2) produced by the voltage controlled oscillator connected with the regulator output begins to outstrip phase of the signal (Vhigh1) produced by oscillator controlled by reference voltage. As a consequence of this, clock generator activates (Clock Activation) and the registers begin to shift towards increasing of number of logical "1" at their outputs (shift from zero to five). With shift of logical "1" corresponding to the code 5 in decimal system, the number of turned off transistors becomes sufficient for reducing the Output Voltage level. At moment of time corresponding 85 ns, phase of the signal produced by reference oscillator begins to outstrip phase of the signal related to the output voltage. The registers begin to shift towards the opposite direction (from five to two) and at shift corresponding to the code 2 the rise of the output voltage occurs.

In the proposed LDO voltage regulator analog components have been replaced by digital components (phase-frequency detector, digital controller based on shift registers) with the use of nanoscale MOS transistors capable to operate in subthreshold region. As a result of this, power consumption, noise sensitivity, sensitivity to temperature and process variations have been reduced. The possibility to scale the voltage regulator based on digital cells has been implemented too which in turn allows achieve compatibility with digital design flows. Besides that, transition to digital structure of the regulator has allowed reduce output voltage ripples and has permitted to operate at low voltage levels, which is essential for wireless micropower devices especially for those which obtain the power supply through harvesting radio-frequency energy from ambient space. The use of the auto-activated clock generator has also allowed decrease the power consumption.

As well in the proposed voltage regulator the separation of bi-directional k-bit serial-in parallel-out shift register into n-bit shift register and m-bit shift register is implemented. Such technique allows without delays settle the output voltage at specified level which is determined by ratio of m and n, since m of paralleled PMOS in array of k transistors turn off and others turn on immediately upon supplying power. Eventually it allows reduce the time (approximately to 3 ns) required to settle the output voltage at desired level compared to counterparts.

22.5 Conclusion

This paper presents the design of an LDO voltage regulator based on digital components. The simulation and analysis of the proposed regulator were carried out. The separation of the k-bit shift register into two parts led to reducing the output voltage settling time to 3 ns. The usage of digital cells, a PMOS array and implementation of auto-activated clock generator allowed the low value of the regulator power

consumption to be attained. The digital components as a base for the regulator allowed lowering the sensitivity to external influence, achieve compatibility with digital design flows and make possible the device operation at low voltage levels (less than 0.5 V).

References

1. A. Takacs et al., Recent advances in electromagnetic energy harvesting and wireless power transfer for IoT and SHM applications. in *2017 IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics* (IEEE, Donostia, Spain, 2017), pp. 1–4
2. M. Taghadosi et al., High efficiency energy harvesters in 65nm CMOS process for autonomous IoT sensor applications. IEEE Access **6**, 2397–2409 (2018)
3. J. Nicot et al., Remote RF powering of ambient sensors. in *2016 IEEE International Conference on Electronics, Circuits and Systems*, (IEEE, Monte Carlo, Monaco , 2016), pp. 760–763
4. M. Stoopman et al., An RF-Powered DLL-Based 2.4-GHz transmitter for autonomous wireless sensor nodes. IEEE Trans. Microwave Theory Techniques **65**(7), 2399–2408 (2017)
5. L. Shangguan et al., The design and implementation of a mobile RFID tag sorting robot. in *14th Annual International Conference on Mobile Systems, Applications, and Services*, (ACM, Singapore, 2016), pp. 31-42
6. G. Zhang et al., Architecture characteristics and technical trends of uhf rfid temperature sensor chip. Act. Passive Electron. Compon. **2018**(9343241), 1–8 (2018)
7. A. Collado et al., Rectifier design challenges for RF wireless power transfer and energy harvesting systems. Radioengineering **26**(2), 411–417 (2017)
8. N. Swangpattaraphon et al., Design consideration of 433.92MHz ISM band RF rectifier for wireless IoT cloud application. in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference*, (IEEE, Las Vegas, NV, USA, 2019), pp. 685–689
9. G. De Vita et al., Ultra-low-power series voltage regulator for passive rfid transponders with subthreshold logic. Electron. Lett. **42**(23), 1350–1352 (2006)
10. S. Fan et al., A high-efficiency radio frequency rectifier-booster regulator for ambient WLAN energy harvesting applications. in *2018 IEEE MTT-S International Wireless Symposium,* (IEEE, Chengdu, China, 2018), pp. 1–3
11. M. Lüders et al., Fully-integrated LDO voltage regulator for digital circuits. Adv. Radio Sci. **9**, 263–267 (2011)
12. J.M. Hinojo et al., FVF-based low-dropout voltage regulator with fast charging/discharging paths for fast line and load regulation. ETRI J. **39**(3), 373–382 (2017)
13. X. Cheng et al., An ultra-low power output capacitor-less low-dropout regulator with slew-rate-enhanced circuit. J. Semicond. **39**(3), 1–6 (2018)
14. Y. Okuma et al., 0.5-V Input digital LDO with 98.7% current efficiency and 2.7-μA quiescent current in 65nm CMOS. in *IEEE Custom Integrated Circuits Conference 2010*, (IEEE, San Jose, CA, USA, 2010), pp. 1–4
15. D.D. Bernardo et al., 0.5 V output digital low dropout (DLDO) voltage regulator with VCO-based digital feedback loop. in *2017 IEEE Region 10 Conference*, (IEEE, Penang, Malaysia, 2017), pp. 505–509

16. G. Zamora-Mejía et al., A digitally enhanced LDO voltage regulator for UHF RFID passive tags. IEICE Electron. Expr. **13**(12), 1–10 (2016)
17. S.B. Nasir et al., All-digital low-dropout regulator with adaptive control and reduced dynamic stability for digital load circuits. IEEE Trans. Power Electron. **31**(12), 8293–8302 (2016)
18. M. Huang et al., An output-capacitor-free analog-assisted digital low-dropout regulator with tri-loop control. in *2017 IEEE International Solid-State Circuits Conference*, (IEEE, San Francisco, CA, USA ,2017), pp. 342–343
19. J.-J. Chen et al., Sub-1V capacitor-free low-power-consumption LDO with digital controlled loop. in *2008 IEEE Asia Pacific Conference on Circuits and Systems*, (IEEE, Macao, China , 2008), pp. 526–529
20. A. Raychowdhury et al., Digitally phase locked low dropout regulator. Patent US 0,241,890 (2015)

Chapter 23
Aerosol Transport Modeling
for Evaporator of CVD Reactor

Roman Kleimanov ⬤**, Andrey Korshunov** ⬤**, Ivan Komarevtsev** ⬤**,
and Sergey Alexandrov** ⬤

Abstract In this paper authors propose a new method for modeling a process of droplet evaporation inside of chemical vapor deposition (CVD) reactor. This method allows to exclude so complicated part of modeling as deforming meshes and calculation of heat transfer surfaces between surfaces of droplets and transport gases. This method also allows to minimize time of modeling. The method based on modeling of transport gas flow and calculation of concentration of dispersed phase (aerosol droplets). This method has some limitations and allows to model transport of diluted species with disperse phase concentration lower than 10%. This method was used to predict the heater power for decamethylcyclopentasiloxane evaporator for SiO_2 deposition process. Experimental study shows good matching of numerical study data and measured evaporation heater power.

Keywords Chemical vapor deposition · Computational fluid dynamics · Modelling · Heat transfer · Aerosol transport

23.1 Introduction

The problem of supplying viscous reagents with a high boiling point can be solved using aerodynamic nebulizers. Such atomizers are a nozzle that accelerates the transport gas flow into which a liquid (reagent) is injected. The interaction of the high-velocity gas flow and the liquid droplets causes the droplets break up. However, in the case of using a carrier gas with a low density, its consumption increases significantly [1, 2]. The transport gas supply with high volume flow may be unacceptable from the point of view of the low-pressure chemical vapor deposition (CVD) process. In such cases, it is necessary to feed the reactant in the form of vapors carried by the

R. Kleimanov (✉) · A. Korshunov · I. Komarevtsev
Alferov University, St. Petersburg 194021, Russia
e-mail: kleimanovrv@mail.ru

S. Alexandrov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

gas stream into the reactor at a low flow rate [3]. The evaporation process should take place at the lowest possible temperature, since very often the temperatures of evaporation and thermal decomposition for complex compounds are very close. Even the presence of the reagent for a long time at high temperatures can lead to the occurrence of undesirable chemical reactions in it. Thus, another criterion for the correct operation of the evaporator is a sufficiently fast evaporation process. To increase the heat exchange surface and accordingly the evaporation rate the reagent must first be sprayed. Further evaporation occurs in a hot gas flow. To minimize the change in the composition of the reagent, contact of the reagent with the hot metal is not allowed; this requires the manufacture of the evaporator from inert materials such as quartz. From the point of view of intensifying the heat transfer process, it is preferable to use turbulators on the evaporator walls for more efficient heat transfer from the hot walls to the gas flow but this complicates the design and manufacturing of non-metal evaporator. The simplest is the coaxial pipe design, in which the flow is organized with several turns through 180 degrees. This allows to reduce the linear dimensions of the evaporator while maintaining sufficient heat transfer surface area. To build an evaporator capable of supplying a reagent to a CVD reactor it is necessary to determine the power of the heaters which will be enough for complete evaporation of the reagent without overheating it. When calculating a two-phase flow taking into account heat transfer, it is required to build an interface between droplets of a liquid and a gas. In the case of a finely dispersed aerosol, this requires the use of moving meshes, which causes sharp increasing of degrees of freedom solving for. In this work, an approach is used that is applicable in the case of a low (less than 10%) concentration of the dispersed phase (aerosol) in the gas flow. In the case when the nature of the primary spraying of the liquid is known in advance (the droplet size distribution), it is possible to set the averaged droplet-gas heat transfer parameters. Then, to calculate heat transfer, it is required to determine the temperature field and make its correction taking into account the concentration of the dispersed phase and its evaporation.

23.2   Modelling Methods

The essence of the method is to model the viscous transfer of an atomized liquid droplets by a gas flow. In the case of fine atomization, when the size of the droplets makes it possible to consider them hydraulically small, the flow of the dispersed phase will be carried away by the gas flow with practically no reverse effect on the streamlines of the gas flow. The criterion for the smallness of the droplet size is the Stokes number or Stk. The Stokes number is defined as the ratio of the characteristic time of a particle (or droplet) to a characteristic time of the flow. A particle with a low Stokes number follows fluid streamlines (perfect advection), while a particle with a large Stokes number is dominated by its inertia and continues along its initial trajectory.

Smaller Stokes numbers represent better tracing accuracy; for Stk»1, particles will detach from a flow especially where the flow decelerates abruptly. For Stk<<1, particles follow fluid streamlines closely. If Stk < 0.1, tracing accuracy errors are below 1% [4].

Thus, it is possible to construct a system of transport equations based on the characteristics of the sprayed droplets and the transport gas flow. At the first stage, the concentration field of the dispersed phase is constructed. This is followed by the stage of correction of the temperature field taking into account the evaporation of the dispersed phase. In this case, the temperature field also changes. Due to the density change at the last stage, the velocity (pressure) field is also corrected. After several iterations the convergence of the construction of the flow pattern is achieved taking into account the evaporation of aerosol. As a computational method, we used a numerical flow model built on the basis of the finite element analysis (FEA) made in the COMSOL Multiphysics software package. The finite element method allows modeling both laminar and turbulent flows within the Knudsen numbers corresponding to the viscous flow regime. For high vacuum flows molecular dynamics methods should be used. However, high-vacuum systems imply the supply of reagents in the form of vapor not droplets. In this case evaporation from the liquid surface at reduced pressure should be considered and not heat exchange of droplets in the gas stream [5, 6]. The system under consideration implies operation at pressures close to atmospheric and liquid spraying directly into the evaporator flow path.

23.3   Results

The result of applying this method is illustrated in Figs. 23.1, 23.2, 23.3. Figure 23.1 shows the temperature field in the evaporator using helium as transport gas. Decamethylcyclopentasiloxane acts as an evaporated liquid with a mass flow rate of 5 ml / min. Transport gas-helium with a flow rate of 6.7 l/min. The heater is located on the outer wall of the tube and consists of three sections along the evaporator. The highest power is supplied in the middle section (in the area of the hemispherical fairing inside the evaporator flow path). The power of the heaters was varied so that the wall temperature in different areas of the evaporator did not exceed 160 degrees Celsius.

The proposed method allows you to quickly change the type of gas to quickly assess the performance of the system with various combinations of evaporated liquid and transport gas. Replacing helium with argon (Fig. 23.2) with the same flow rate, a significant deterioration in heat transfer is observed. Condensation of the reagent on the domed fairing causes it cooling. An increase in the power of the heaters will lead to local overheating of the walls and will not allow the reagent to completely evaporate in the flow. Figure 23.3 also shows the eddies from the ascending flows of the evaporating reagent, which are also reflected in the temperature field. Figure 23.2 shows the velocity field and streamlines for calculations with argon. Intense vortex formation in this case leads to overheating of the reagent due to an increase in its
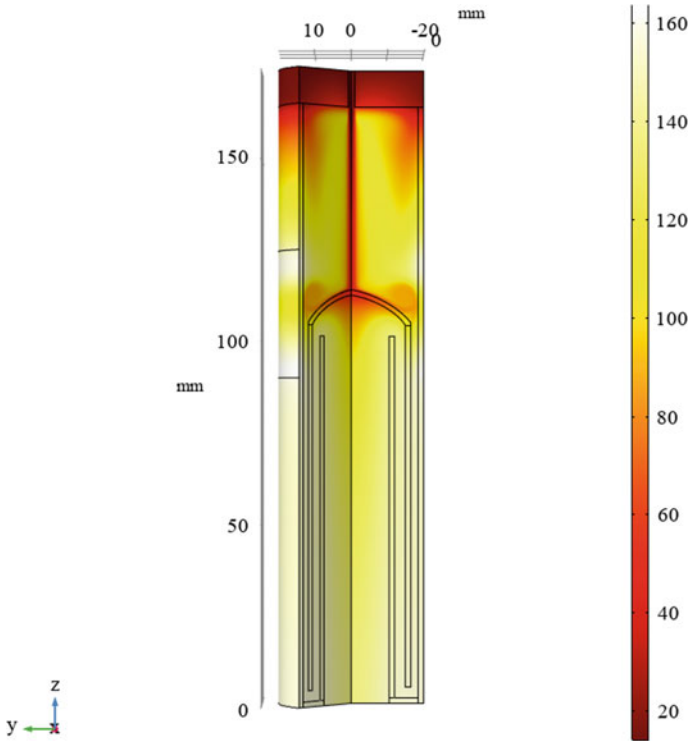
**Fig. 23.1** Evaporator temperature field, transport gas—helium, volume flow rate 6.7 l/min

residence time in the hot zone. Deterioration of heat transfer at the walls in this case leads to local overheating of the reagent, while there are areas where the wall temperature is insufficient for evaporation of the reagent and condensation occurs.

Experimental verification of the correctness of the calculations showed a fairly good agreement in terms of calculating the power required for complete evaporation of the reagent in the gas flow. It should also be noted that the change in flow conditions is quite correct, taking into account the evaporation of liquid in the gas flow. Comparison of the calculated and experimental data is shown in Table 23.1.

23.4 Conclusions

The proposed method allows in the presence of some restrictions to simplify the adjustment procedure of a CVD process. This makes it possible to carry out more variant calculations and select the gas for the required reagent consumption and the conditions for its evaporation. Obviously in the example shown argon is less preferable than helium for given reactor feeding conditions. It is not possible to determine

Fig. 23.2 Evaporator temperature field, transport gas—argon, volume flow rate 6.7 l/min
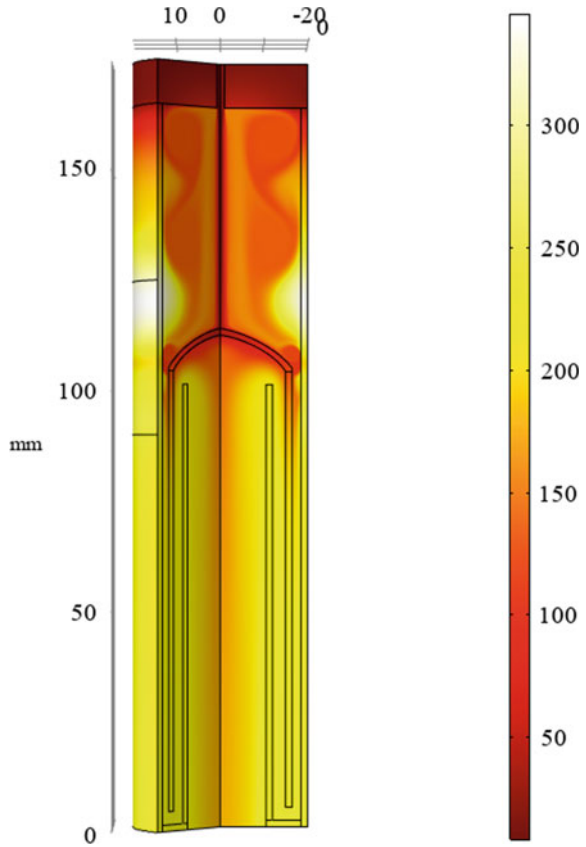
all these aspects at the stage of preliminary calculations without experiment [7, 8]. Compared to full-scale direct modeling of two-phase flow using movable meshes this method is much simpler and faster. As limitations should note the errors in determining the power caused by the use of the averaged parameters of the sprayed droplets as well as the impossibility of simulating high droplet concentrations.
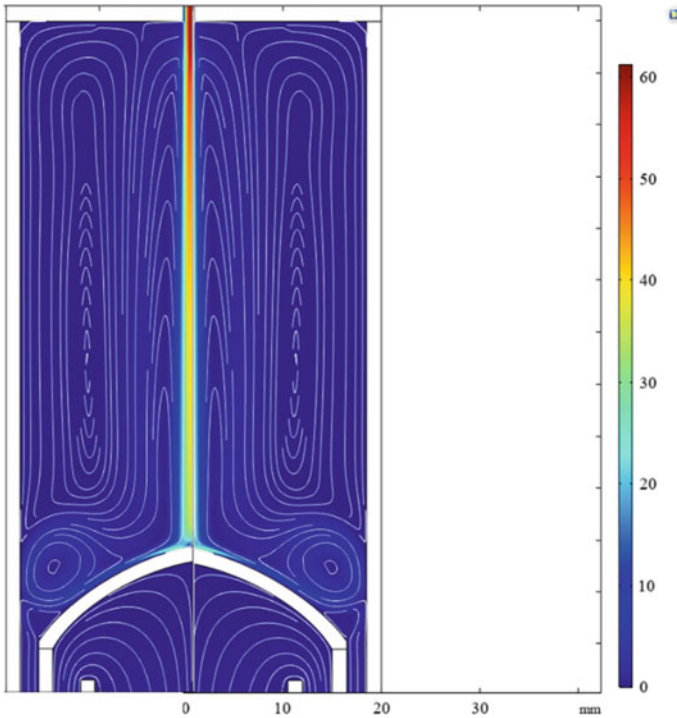
Fig. 23.3 Vortex formation at the top of the evaporator (argon)

Table 23.1 Comparison of the calculated and experimental values of temperatures at the control points on the walls of the evaporator and the power of the heaters

	Heater 1	Δ, %	Heater 2	Δ, %	Heater 3	Δ, %
t °C, FEA	140	3.5	130	3.7	160	9.3
t °C, Exp	135		135		145	
P, W, FEA	47	7.8	35	7.7	70	8.5
P, W, Exp	51		32,5		76.5	

References

1. Low Pressure Chemical Vapor Deposition—Technology and Equipment (2021). https://www.crystec.com/klllpcvde.htm. Last Accessed 03 Jan 2021
2. R. Bunshah, *Handbook of Deposition Technologies for Films and Coatings* (Noyes Publications, New York, 1993)
3. I.A. Shareef, G.W. Rubloff, M. Anderle et al., Subatmospheric chemical vapor deposition ozone/TEOS process for SiO_2 trench filling. J. Vacuum Sci. Technol. B: Microelectron. Nanometer Struct. Process. Measurem. Phenomena. **13**(4), 1888–1892 (1995)
4. C. Tropea, A. Alexander Yarin, J. Foss, in *Springer Handbook of Experimental Fluid Mechanics.* (Springer, 2007)

5. S.P. Belyaev, L.M. Levin, Techniques for collection of representative aerosol samples. Aerosol Science **5**(4), 325–338 (1974)
6. R. Israel, D.E. Rosner, Use of a generalized stokes number to determine the aerodynamic capture efficiency of non-stokesian particles from a compressible gas flow. Aerosol Sci. Technol. **2**(1), 45–51 (1982)
7. A. Vorobyev et al., Controlled formation of iron oxide nanoparticles by pulse-modulated RF discharge at atmospheric pressure. Mater. Today: Proceed. **30**, 417–421 (2020)
8. P. Bespalova et al., Degradation of NiO-GNPs electrochromic coating. in *IEEE International Conference on Electrical Engineering and Photonics* (EExPolytech), (IEEE, 2019), pp. 244–246

Chapter 24
Numerical Simulation of CVD Reactor for Oxide Semiconductor Layer Deposition

Roman Kleimanov , **Andrey Korshunov** , **Anastasia Kondrateva** ,
Platon Karaseov , **Maxim Mishin** , **Yakov Enns** ,
and Ivan Komarevtsev

Abstract We present a numerical model of a chemical vapor deposition (CVD) reactor for producing oxide semiconductor layers containing gold nanoparticles. To implement this technological process, it is required to ensure the stability of the substrate temperature and the uniformity of layer deposition. The reactor housing is made from a quartz tube, inside which there is a graphite stage with a substrate holder. The substrate holder is heated by an inductor located outside the quartz tube. It is required to determine the heating time of the substrate and the uniformity of its heating. Since it is not allowed to place measuring devices in the reactor to control the temperature homogeneity during the process, a numerical analysis is carried out at various operating modes (temperature, flow and heating intensity, etc.). The numerical model is built in the COMSOL Multiphysics software package. The model includes the calculation of the induction heating of the substrate holder when the reactor is purged with an inert gas and a mixture of gas with a reagent, as well as the calculation of the configuration of the electromagnetic field for further evaluation of the field effect on the forming of gold nanoparticles on a nickel oxide substrate. It is required to ensure the most uniform deposition of the oxide layer over the entire area of the substrate without using the rotation of the substrate holder during the process.

Keywords Chemical vapor deposition · Computational fluid dynamics · Modelling · Heat transfer · Aerosol transport

24.1 Introduction

Finite elements analysis (FEA) is widely used for the numerical simulation of various processes including chemical vapor deposition (CVD). The main difficulty is the construction of a reaction scheme with an accurate description of all forward and

R. Kleimanov (✉) · A. Korshunov · A. Kondrateva · P. Karaseov · M. Mishin · Y. Enns ·
I. Komarevtsev
Alferov University, St. Petersburg 194021, Russia
e-mail: kleimanovrv@mail.ru

reverse reactions occurring in the reactor volume and directly on the substrate where the deposition takes place. However, in addition to chemical reactions there is a wide range of CVD problems that can be solved with computational fluid mechanics. In particular this is the organization of the flow with the correct spatial distribution of gases in the reactor flow path. The trivial geometry of the flow path (cylindrical or flat structures) does not guarantee that the flow will be organized in an aerodynamically optimal way [1].

Measurements such as particle image velocimetry (PIV), placement of aerodynamic probes in the flow path, etc. devices are often impossible for technological reasons [2]. In a high-vacuum deposition process the flow is organized in a molecular mode (when the mean free path of gas molecules is comparable to the scale of the flow path of the reactor) the geometry does not have such a strong effect on the flow as at pressures close to atmospheric. The range of problems under consideration implies the deposition of a layer of oxide semiconductors the subsequent deposition of a thin gold film and its dewetting under the action of heating to create an ensemble of nanoparticles on a semiconductor substrate [3]. This requires precise control of the substrate temperature both during the deposition of the oxide semiconductor sublayer and during subsequent operations. A horizontal tubular reactor with induction heating of a graphite substrate holder for the deposition of layers of nickel and titanium oxides from organometallic precursors on quartz substrates is considered [4].

The peculiarity of the induction heater is the complexity of measuring the temperatures of the substrate and substrate holder by contact methods. To ensure uniform heating a numerical simulation of the heating process is carried out when the reactor is purged with an inert gas. The purpose of the simulation is to determine the uniformity of heating the substrate holder as well as the heating time and response to changes in the gas flow regime in the reactor [5].

24.2   Modelling Methods

To build a numerical model, the finite element method implemented in the COMSOL Multiphysics software package is used. The reactor is a horizontal quartz tube. Outside is placed the coil of the inductor made from a copper tube, inside of which circulates water to cool the inductor. Heating the substrate holder eliminates the need to heat the entire volume of gas to the temperatures of the onset of the deposition reaction. It is advisable to feed the reagent at temperatures above the condensation temperature of the organometallic compound. Heating to the start temperature of the reaction by the temperature of the hot substrate holder also reduces unwanted deposition of reaction products outside the substrate. Inside the quartz tube there is a graphite table-substrate holder, on which thin quartz substrates are located. A solid-state reactor model without end caps is shown in Fig. 24.1.

Multiphysics simulation allows for conjugate calculations such as the substrate holder cooling by gas flow. The joint solution of the equations describing the heating
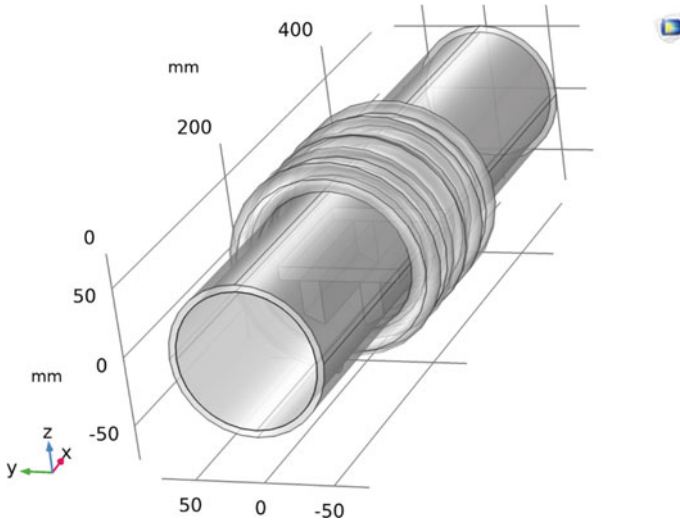
Fig. 24.1 General view of the reactor with an induction heater

of the substrate holder by eddy currents and its cooling by the gas flow makes it possible to obtain a correct distribution of the temperature depending on the gas flow rate and the operating mode of the inductor (for example, the frequency or the waveform from the generator). A feature of induction heating is the formation of a skin layer with maximum heat release on the surface of the workpiece. In this case, cooling also occurs due to the extraction of heat from the surface of the stage, but also depends on the flow conditions. Since a thermal boundary layer is formed on the surfaces of the substrate holder the heat transfer conditions are not the same over the area of the substrate holder. The difference can be partially offset by a slight (several degrees) slope of the surface on which the substrate is located towards the flow to reduce the thickness of the boundary layer. The Prandtl number (defined as the ratio of momentum diffusivity to thermal diffusivity) of argon is less than 1, which means that the thickness of the temperature boundary layer increases more slowly than the dynamic boundary layer thickness [6, 7]. This means that the heat transfer across the plate does not change as intensively as the boundary layer thickness increases; nevertheless, this change should be taken into account when evaluating the temperature uniformity. The inclination of the substrate holder changes the nature of the flow due to less displacement of the flow by the boundary layer from the surface by the viscous boundary layer; this can also have a positive effect on the uniformity of the deposition of oxide layers on the substrate.
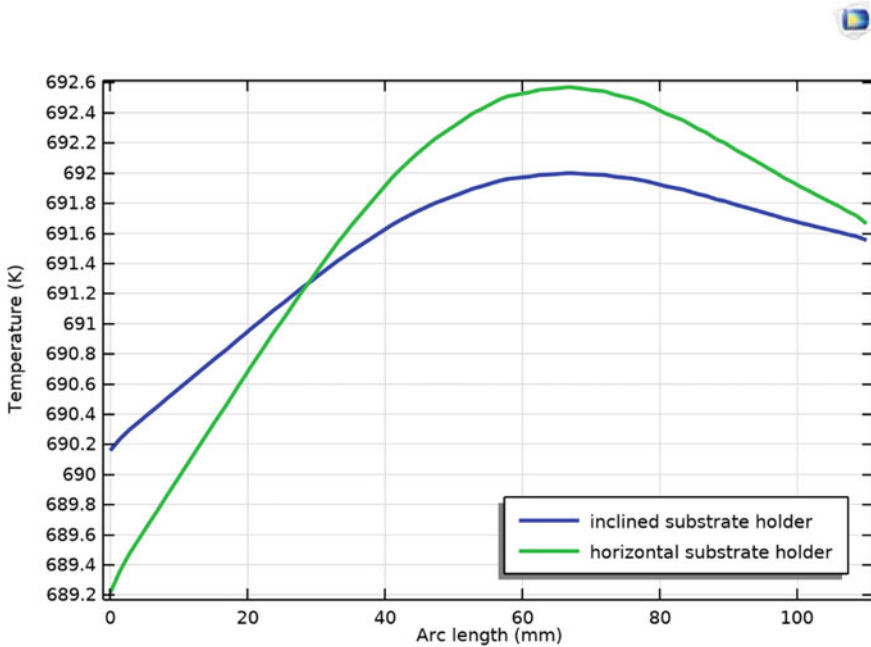
Fig. 24.2 Distribution of temperatures along the length of the substrate holder at different slopes

24.3 Results

Analysis of the heating of the substrate holder showed a good temperature uniformity over the entire area. Figure 24.2 shows the distribution of temperatures along the length of the working surface of the substrate holder at the operating temperature. Sufficient uniformity of temperatures over the entire area can be noted, with the exception of the edge sections. Tilt of the plate by 3 degrees affects the uniformity of temperatures without significant distortion of the flow in the reactor. The temperature field around the substrate holder is distorted as the gas warms up and heat is transferred to the walls of the reactor; therefore, simulation was carried out with a significant holding time to achieve thermal equilibrium. The obtained temperature inhomogeneities within two degrees can be provided over the entire operating temperature range (up to 700 K) (Fig. 24.3).

24.4 Conclusions

The performed numerical simulation of the heating of the tooling in the CVD reactor revealed the need to tilt the substrate holder to reduce the temperature difference. It was found that a longer exposure is required to establish thermal equilibrium
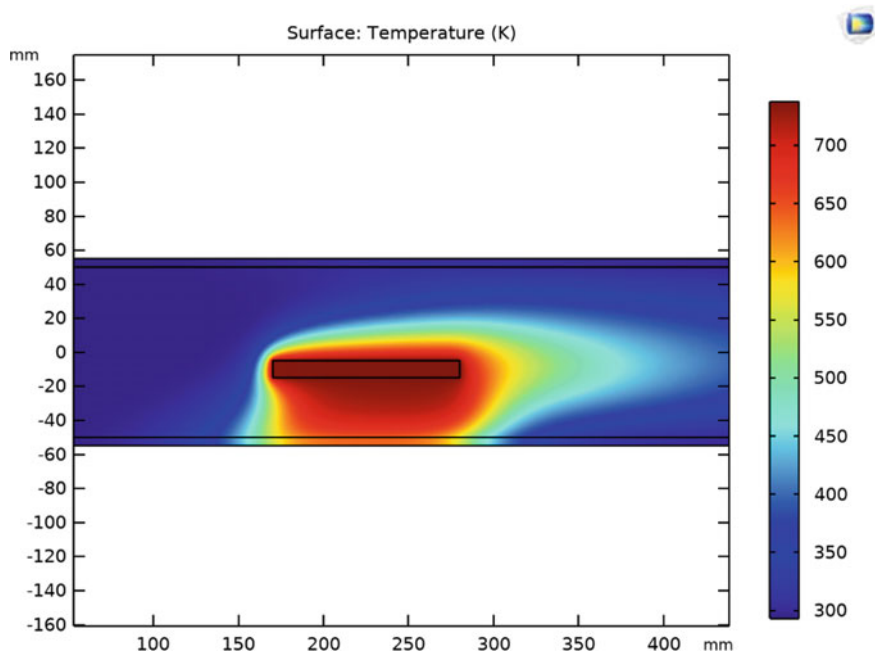
Fig. 24.3 Temperature distribution around a substrate holder

around the substrate holder when the reagent is supplied. The data obtained make it possible to reduce the amount of experimental verification of the reactor operability and to shorten the period of its debugging. An increase in the homogeneity of the temperature and concentration fields in the future will make it possible to achieve greater uniformity in the deposition of coatings on various substrates.

References

1. B.J. Kirby, in *Micro- and Nanoscale Fluid Mechanics Transport in Microfluidic Devices.* (Cambridge University Press, 2010)
2. A. Melling, Tracer particles and seeding for particle image velocimetry. Meas. Sci. Technol. **8**(12), 1406–1416 (1997)
3. C.N. Rao, Kulkarni et. al., Metal nanoparticles and their assemblies. Chem. Soc. Rev. **29**(1), 27–35 (2000)
4. G.B. Stringfellow, in *Organometallic Vapor-Phase Epitaxy: Theory and Practice*, (Elsevier Science , 2012), pp. 3
5. N. Tay, F. Bruno, M. Belusko, Experimental validation of a CFD model for tubes in a phase change thermal energy storage system. Int. J. Heat Mass Transf. **55**, 574–585 (2012)

6. D.M. McEligot, M.F. Taylor, The turbulent Prandtl number in the near-wall region for low-Prandtl-number gas mixtures. Int. J. Heat Mass Transfer **39**, 1287–1295 (1996)
7. E.M. Sparrow, S.V. Patankar, Relationships among boundary conditions and nusselt numbers for thermally developed duct flows. J. Heat Transf. **99**, 483 (1977)

Chapter 25
Design of GaAs Flip-Flop Using Depletion-Mode-Only Transistors

Gleb I. Stesev ⓘ**, Dmitry O. Budanov** ⓘ**, and Evgenii V. Balashov** ⓘ

Abstract In this paper, the design of logic gates based on GaAs 0.5 μm pHEMT technology is proposed. It uses Depletion-mode-only transistors, which is applicable to technological processes without Enhancement-mode transistors available. An attempt has been made to reduce power consumption by using high-value resistors. Basic logic gates (inverter, NOR, NAND) have been implemented and the topology of flip-flop has been designed, alongside with the NAND gate. The results of this design can be used for any desirable logic circuit, combinational or sequential.

Keywords GaAs · Logic circuits · pHEMT · Depletion mode · Flip-flop

25.1 Introduction

Gallium arsenide is a widely used material in the semiconductor industry although a big part of it belongs to silicon and germanium. However, GaAs has its own niche in high-frequency applications due to higher electron mobility (under normal conditions around 5000 cm^2/V·s, which is 5–7 times higher than in Si). Besides, GaAs outperforms Si in terms of noise due to having a wider and direct band energy gap [1].

One of the major drawbacks of GaAs is relatively low hole mobility (around 250 cm^2/V·s, even lower than in Si). Hence, the usage of p-channel GaAs devices could negate previously described potential. Owing to that, the complementary logic approach cannot be applied, leading to higher power consumption. As a result, GaAs logic circuits are less commonly used [2, 3].

But in some cases GaAs circuits are preferable, e.g. high speed logic or digital control circuits for phase shifters, attenuators, amplifiers, etc. Designing such circuits in GaAs technology allows to manufacture them with the analog part on the same die. This increases the degree of integration and removes the necessity of external digital control chip [4–10].

G. I. Stesev (✉) · D. O. Budanov · E. V. Balashov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: stesev.gi@edu.spbstu.ru

Every digital circuit can be described as either time-independent (combinational) or time-dependent (sequential). Output signals of a combinational circuit depend only on the input signals at any given time. The examples of such circuits are inverter, OR, AND and other Boolean operations. Sequential circuits, on the other hand, have a special characteristic called internal state, which is used as an input to define the very next internal state. Such devices include latches, flip-flops, registers, etc. [11–16]

Basic logic gates, such as NOR or NAND, can be used to design any desirable logic circuit, including a sequential one. The main part of a digital control circuit is a serial-to-parallel code converter, which can be implemented by connecting multiple flip-flops in series. The basic design of such a converter was proposed in [17]. In this paper, the improvements have been provided in terms of power consumption and overall logic levels stability.

25.2 Simulation Details

All simulations have been carried out using AWR Microwave Office. The following elements from the given Process Design Kit (PDK) were used:

- 1-finger Depletion-mode FET: gate width 15…50 μm;
- 1-finger diode: gate width 20…50 μm;
- Thin film resistor: width 6…300 μm; length 8…300 μm; resistance 1700 Ω/sq.

To reduce occupied area and signal delay, as well as the power consumption, the widths of transistors and diodes have been set to their minimal values –15 μm and 20 μm respectively. As for the resistors, it is not recommended to set their dimensions lower than 10–20 μm. It would not reduce the area of the chip significantly but could provide tolerance issues.

The supply voltages are equal to $\pm$ 5 V and the logic levels are –3 V (low), 0 V (high).

The signal delay is measured as the time interval from the input occurrence to the output response. The decision level is 50% of the amplitude. Out of all possible switching processes, only the greatest signal delay is considered.

Strip lines are omitted from the presented schematics for better readability. But their stray parameters were taken into account during the simulation. All proposed elements were simulated at 25 °C.

25.3 Logic Gate Structure

In previous design [8], the following logic gate structure was used (Fig. 25.1). It has two diode-connected transistors as current sources and a high-value resistor (HVR) for level shifting. Shifter part is necessary to provide compatibility between

Fig. 25.1 Buffered FET
logic gate from the previous
design
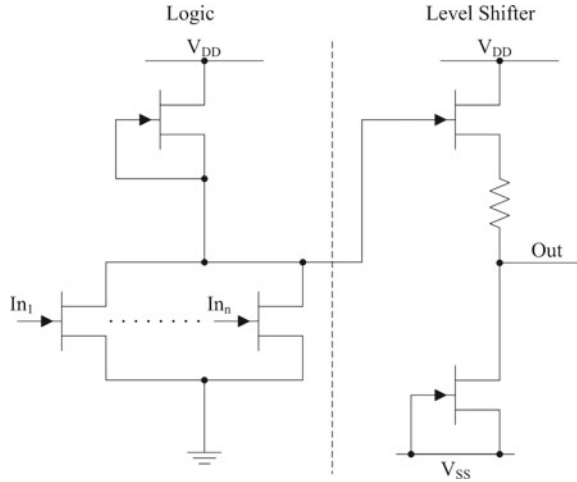
logic gates, since Depletion-mode transistors have negative threshold voltage. They
operate with logic levels shifted towards negative values as a result [1, 18, 19].

Improved gate structure has a series of diodes for level shifting and HVRs to
reduce the power consumption (Fig. 25.2). With these improvements the current of
the supply sources has been reduced approximately from 19 to 3 mA. On the top
of that, the previous structure was sensitive to the HVR value mismatch, since the
voltage shift is linearly dependent on the resistance according to the Ohm's law. In
the new structure the voltage shift is occurred owing to the height of the Shottky
barrier, which is constant in the given technology. HVRs serve the role of current
limiters and could be mismatched to some extent, as long as their resistances remain

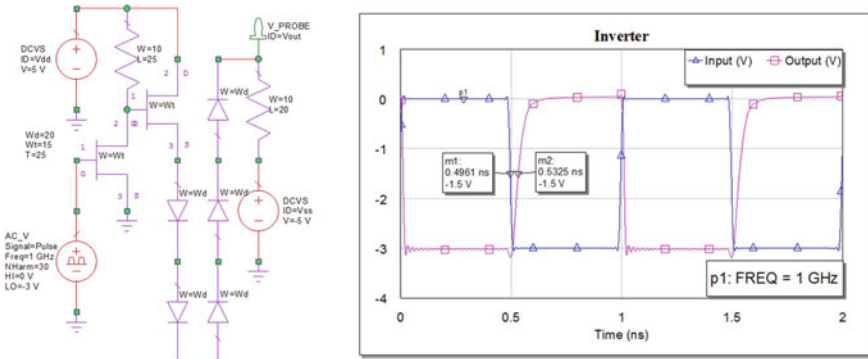**Fig. 25.2** Improved logic
gate structure
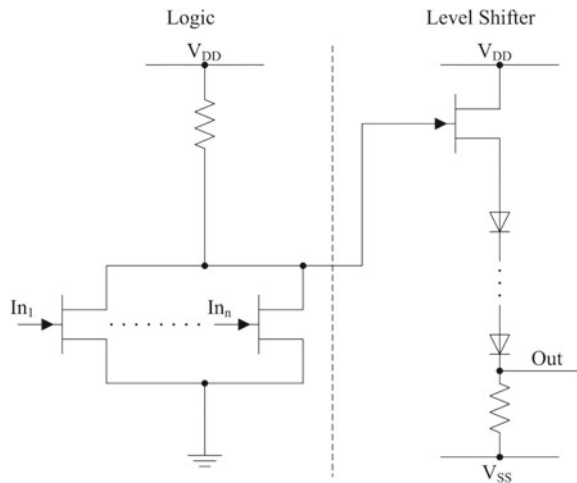
Fig. 25.3 Inverter schematic and simulation results

relatively high. This makes the output logic levels more consistent and all digital circuits more stable overall.

25.4 Logic Circuits Simulation

25.4.1 *Inverter*

The inverter structure and its simulation results are shown in Fig. 25.3. Pulse signal with frequency 1 GHz is passed to the circuit input.

The signal delay is 36.4 ps. The average supply current is 2.66 mA.

25.4.2 *Series Connected Inverters*

Since the signal delay is measured for an unloaded circuit, it must be compared with the loaded one for validation of the results. Simulation results for three inverters connected in series are shown in Fig. 25.4. The total signal delay and supply current both should be approximately three times higher than for a single inverter. The measurements for this comparison are shown in Table 25.1.

25.4.3 *NAND*

The NAND gate structure and its simulation results are shown in Fig. 25.5. Two pulse 1 GHz signals with the phase difference of 90° passed to the circuit inputs.

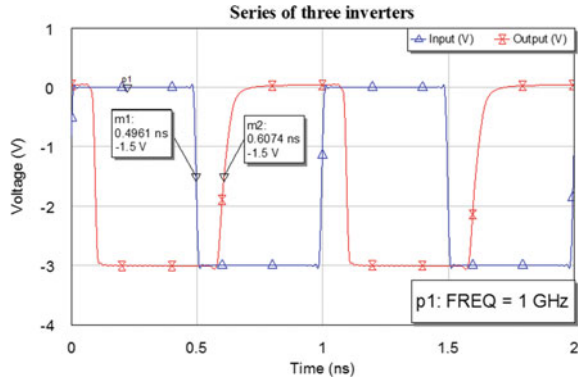**Fig. 25.4** Simulation results for series of inverters

Table 25.1 Average delay and supply current (*per inverter*) for two reviewed schematics

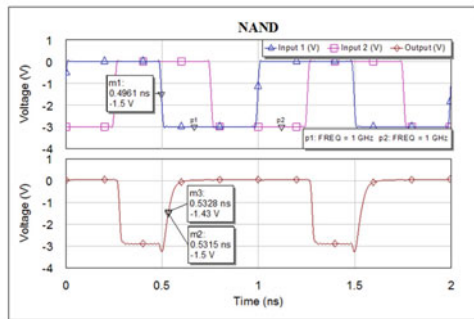| Schematic	Delay, ps	Supply current, mA
Unloaded inverter	36.4	2.66
Series of 3 inverters	37.1	2.72
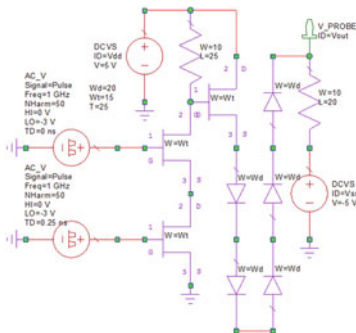
Fig. 25.5 NAND schematic and simulation results

Since including of the second transistor in series slightly shifts output low logic level (approximately from –3 V to –2.9 V), it might be considered to shift the 50% threshold level for the output accordingly. Nevertheless, the two signal delay values are 3.5% apart, so the threshold will remain the same for the generality.

The signal delay is 35.4 ps. The average supply current is 2.86 mA.

25.4.4   NOR

The NOR gate structure and its simulation results are shown in Fig. 25.6. The inputs are the same as for the NAND gate.
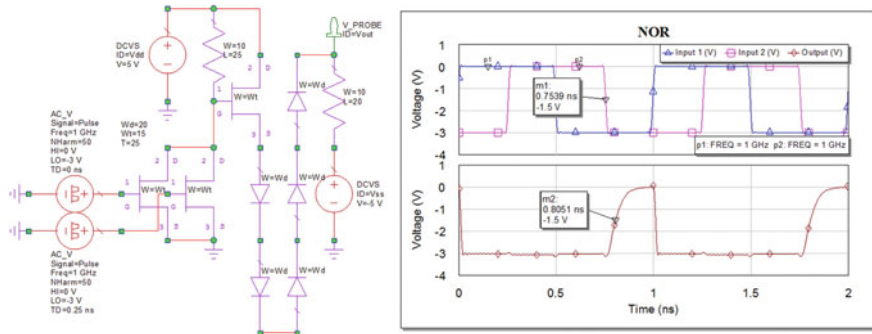
Fig. 25.6 NOR schematic and simulation results

The signal delay is 51.2 ps. The average supply current is 2.46 mA.

25.4.5 Flip-Flop

A flip-flop structure can be implemented using NAND or NOR basic logic gates (Fig. 25.7). A NAND-based flip-flop would be positive-edge-triggered, whereas a NOR-based one would be negative-edge-triggered. The NAND-based structure was chosen because of the lower delay time [18, 20].

The simulation results are shown in Fig. 25.8. Pulse signal with frequency 0.3 GHz is passed to the Data input. Sequence of short pulses with frequency 0.6 GHz is passed to the Clock input. The signal delay for switching to either of logic levels is 138 ps. The average supply current is 16.7 mA.

Fig. 25.7 Flip-flop designation **a** and its implementation using NAND **b** and NOR **c** gates
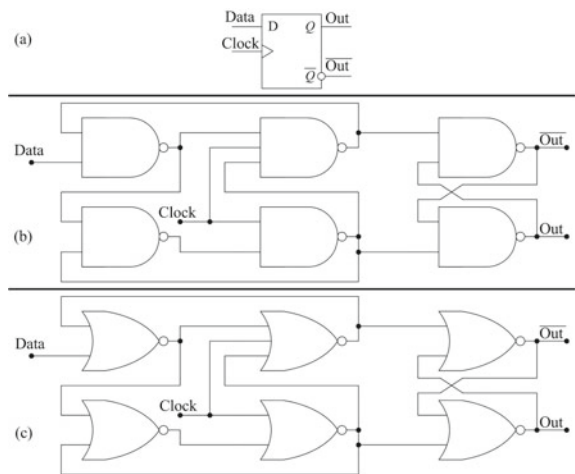
Fig. 25.8 Simulation results
for the proposed
NAND-based flip-flop
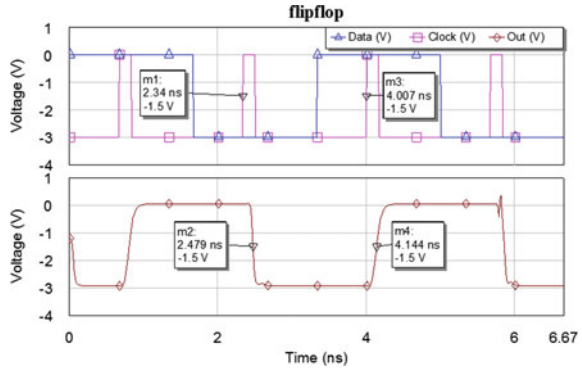
25.5 Flip-Flop Topology

The topologies given are presented to evaluate the occupied area of the circuits. There
is still room for improvements before the chip manufacturing. Because of this and
the assumption that the GaAs logic circuits are not meant to be stand-alone chips, the
supply and ground buses are not considered. They can be connected with additional
metallization layer and vias, or with the same metallization as the data buses using
bridges on the intersections.

The topology for the NAND gate is shown in Fig. 25.9. The dimensions are 176
$\times$ 129 μm and the occupied area is 22700 μm^2. Traditionally, placing two supply

Fig. 25.9 NAND topology
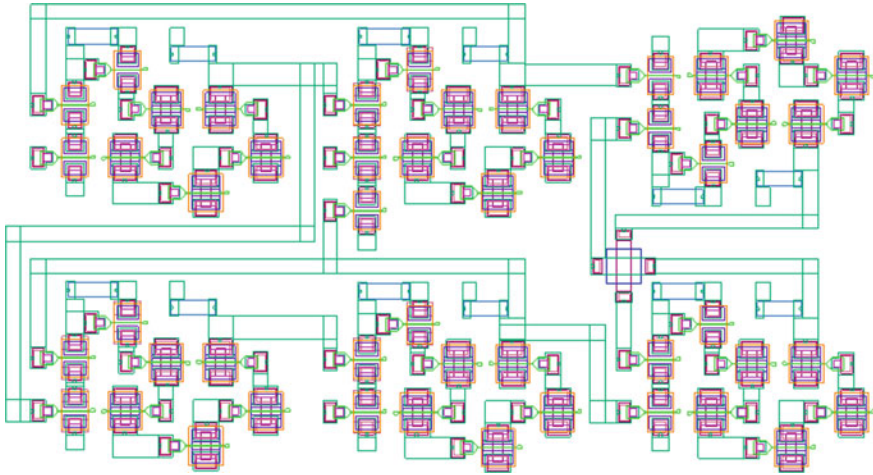
Fig. 25.10 Flip-flop topology

buses at the top and the bottom simplifies their connection. In this case, allocating them to one side allowed to fold the topology structure, reducing the occupied area.

The topology for the flip-flop is shown in Fig. 25.10. The dimensions are 590 × 318 μm and the occupied area is 187620 μm^2.

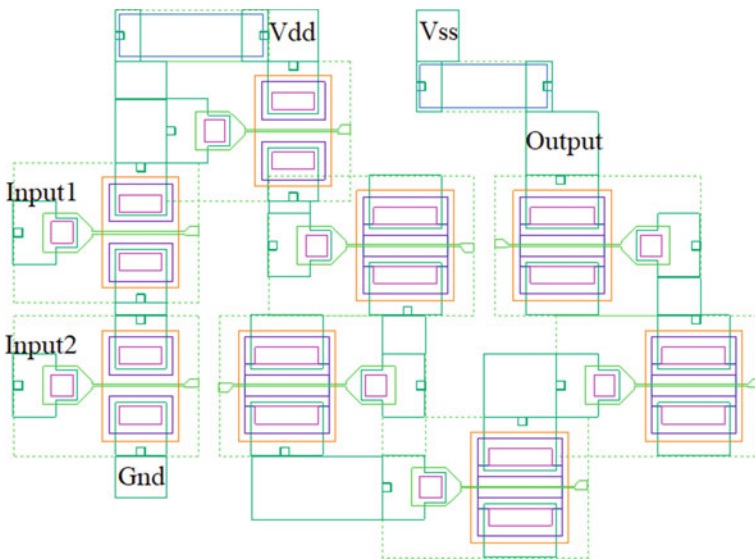## 25.6 Analysis of the Obtained Results

Basic logic gates (inverter, NOR, NAND) and flip-flop were simulated. During the simulation supply voltages ± 5 V were used. The simulation results of main parameters such as signal delays, supply current and power consumption are shown in Table 25.2. A common way to describe circuit performance is using Power-Delay-Product (PDP) characteristic. It describes energy efficiency of a logic circuit and is the product of power consumption and input–output delay time [20, 21].

Table 25.2 Signal delay and power consumption of the proposed schematics

Circuit	Delay, ps	Supply current, mA	Power consumption, mW	PDP, pJ
Inverter	36.4	2.66	26.6	0.97
NAND	35.4	2.86	28.6	1.01
NOR	51.2	2.46	24.6	1.26
Flip-flop	138	16.7	167	23.05

Fig. 25.11 Supply current
of the proposed flip-flop
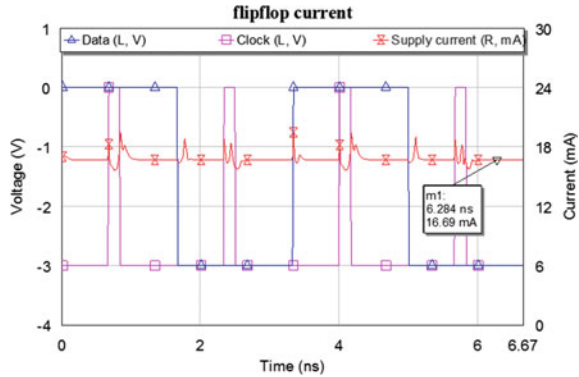
Table 25.3 Flip-flop designs
comparison

Flip-flop design	Delay, ps	Power consumption, mW	PDP, pJ
Previous [17]	330	350	115.5
Proposed	138	167	23.05
MOSFET [20]	232	$292 \cdot 10^{-6}$	$67.8 \cdot 10^{-6}$

Because of constant flowing currents the obtained power consumption and PDP are significantly higher than for CMOS logic. Supply current for the proposed flip-flop is shown in Fig. 25.11. Nevertheless, the previous proposed design was successfully improved in terms of all given parameters (Table 25.3). It's also important to keep in mind the potential to trade-off delay and power consumption by altering HVRs values.

The future improvements may include further topology modernizations, such as combining logic transistors' connected terminals and optimization of series connected diodes.

25.7 Conclusions

This paper has demonstrated the design of digital circuits in GaAs technology and allowed to evaluate the possibilities of the given technology. The proposed schematics and topologies can be used as a verification circuit for various element libraries.

It is important to note about the restrictions of the PDK applied in this work. More complex technological processes may open access to Enhancement-mode transistors. This would significantly improve the performance in terms of consumed power and occupied area. Channel length in the given PDK is 500 nm whereas modern MOSFETs achieve single-digit nm values of this parameter. From the other hand, the advantage of relatively bigger elements is greater reliability in harsh environments.

The future line of this work includes research and developing new logic gate structures to design circuits with better performance. The further optimization of the proposed topologies would also improve their efficiency.

References

1. T. Nilsson, C. Samuelsson, *Design of MMIC Serial to Parallel Converter in Gallium Arsenide* (Linköping University, Linköping, The Department of Electrical Engineering Division of Electronic Devices, 2001)
2. A. Sedra, K. Smith, *Microelectronic Circuits* (Oxford University Press, Oxford, 2010)
3. E. Bushehri, in *Critical Design Issues for GaAs VLSI Circuits.* (Microelectronics Centre, Middlesex Polytechnic, London, 1992)
4. L. Li, Y. Gao, R. Qian, X. Sun, W-band four-channel receiver MMIC in 0.1um GaAs PHEMT technology, in *IEEE Asia-Pacific Microwave Conference (APMC)*, (IEEE, Singapore, 2019), pp. 1512–1514. https://doi.org/10.1109/APMC46564.2019.9038426
5. E. Balashov, I. Rumyancev, An unbalanced transformerless vector-sum phase shifter architecture, in *IEEE NW Russia Young Researchers in Electrical and Electronic Engineering Conference (EIConRusNW)*, (IEEE, St. Petersburg , 2016), pp. 491–494. https://doi.org/10.1109/EIConRusNW.2016.7448229
6. I. Rumyancev, A. Korotkov, Survey on beamforming techniques and integrated circuits for 5G systems, in *IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*, (IEEE, St. Petersburg, 2019), pp. 76–80, https://doi.org/10.1109/EExPolytech.2019.8906842
7. K. Razheva, I. Rumyancev, Study of quadrature couplers for reflection-Type phase shifters in CMOS technology, in *IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, (IEEE, St. Petersburg and Moscow, 2020), pp. 158–161. https://doi.org/10.1109/EIConRus49466.2020.9038913
8. D. Akhmetov, A. Korotkov, D. Morozov, M. Pilipko, I. Rumyancev, Radio frequency identification system of internet of things based on CMOS integrated circuits, in *Proceedings of 2017 IEEE East-West Design and Test Symposium (EWDTS)*, (IEEE, 2017), pp. 1–3. https://doi.org/10.1109/EWDTS.2017.8110155
9. E. Balashov, A. Korotkov, I. Rumyancev, Calibration of phase shifters on basis of vector-sum signals. Radioelectron. Commun. Syst. **61**(11), 515–521 (2018). https://doi.org/10.3103/S0735272711811002X
10. D. Akhmetov, A. Korotkov, I. Rumyancev, 2.4–2.5 GHz fractional-N frequency synthesizer with integrated VCO in 0.18 um CMOS for RFID systems, in *IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*, (IEEE, St. Petersburg, 2018), pp. 64–68. https://doi.org/10.1109/EExPolytech.2018.8564373
11. M. Yenuchenko, A. Korotkov, D. Morozov, M. Pilipko, A switching sequence for unary digital-to-analog converters based on a knight's tour. IEEE Trans. Circuits Syst. I Regul. Pap. **66**(6), 2230–2239 (2019). https://doi.org/10.1109/TCSI.2018.2890412
12. G. Alekseev, I. Mukhin, V. Repin, D. Morozov, The development of a high speed 8-bit pipelined ADC on 0.25um BiCMOS, in *2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, (IEEE, St. Petersburg and Moscow, 2019), pp. 1616–1619. https://doi.org/10.1109/EIConRus.2019.8657101
13. D. Budanov, D. Morozov, M. Pilipko, An 8-bit analog-to-digital converter with a novel encoder using 90 nm CMOS, in *IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*, (IEEE, St. Petersburg, 2018), pp. 56–59. https://doi.org/10.1109/EExPolytech.2018.8564441

14. M. Manokhin, M. Pilipko, Design of a CMOS radio frequency to direct current power converter, in *2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, (IEEE, St. Petersburg and Moscow, 2019), pp. 116–119. https://doi.org/10.1109/EIConRus.2019.8657295

15. A. Kozlov, M. Pilipko, A second-order sigma-delta modulator with a hybrid topology in 180nm CMOS, in *2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, (IEEE, St. Petersburg and Moscow, 2020), pp. 144–146. https://doi.org/10.1109/EIConRus49466.2020.9039246

16. M. Yenuchenko, M. Pilipko, D. Morozov, A 10-bit segmented M-string DAC. In: 2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), (IEEE, St. Petersburg and Moscow , 2018), pp. 265–268, doi: 10.1109/EIConRus.2018.8317081.

17. G. Stesev, D. Budanov, E. Balashov, Serial-to-parallel converter using GaAs D-mode transistors, in *IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*, (IEEE, St. Petersburg, 2020), pp. 67–70. https://doi.org/10.1109/EExPolytech50912.2020.9243862

18. A. Bentini, Design and realization of GaAs digital circuit for mixed signal MMIC implementation in AESA applications. Int. J. Microwave Sci. Technol. **11** (2011). https://doi.org/10.1155/2011/387137

19. G. Ghione, M. Pirola, R. Quaglia, W. Ciccognani, E. Limiti, T. Cavanna, Compact GaAs HEMT D flip-flop for the integration of a SAR MMIC core-chip digital control logic, in *Workshop on Integrated Nonlinear Microwave and Millimeter-Wave Circuits*, (Goteborg, 2010), pp. 62–65. https://doi.org/10.1109/INMMIC.2010.5480137

20. G. Ganesan, K. Shaji, Design of positive and negative edge triggered D-flip flop using AlGaAs/GaAs MODFET technology, in *2nd International Conference on Devices, Circuits and Systems (ICDCS)*, (Combiatore, 2014), pp. 1–5. https://doi.org/10.1109/ICDCSyst.2014.6926217

21. S. Maheshwari, J. Patel, S. Nirmalkar, A. Gupta, Logical effort based power-delay-product optimization, in *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, (Delhi, 2014), pp. 565–569. https://doi.org/10.1109/ICACCI.2014.6968530

Chapter 26
Comparative Study of Ion-Induced Damage Formation in GaN and *beta*-Ga$_2$O$_3$

Konstantin V. Karabeshkin, Platon A. Karaseov, Andrei I. Struchkov, Andrei I. Titov, Alexander Azarov, and Daniela S. Gogova

Abstract Accumulation of ion-beam-induced structural disorder in wide bandgap semiconductors β-Ga$_2$O$_3$ and GaN irradiated at room temperature by 40 keV P ions has been studied by Rutherford backscattering/channeling spectrometry. Results show that despite the damage versus depth profiles are bimodal in both β-Ga$_2$O$_3$ and GaN, the kinetics of radiation damage accumulation is drastically different for these two materials. It is also demonstrated that β-Ga$_2$O$_3$ is more than one order less radiation resistant as compared to GaN for the irradiation conditions under consideration.

Keywords Wide band-gap semiconductors · Gallium nitride · Gallium oxide · Radiation defects · Ion implantation

26.1 Introduction

Gallium nitride (GaN) and beta gallium oxide (β-Ga$_2$O$_3$) are wide bandgap semiconductors with bandgaps of approximately 3.4 and 4.85 eV, respectively. Research interest in these materials has been stimulated by their unique properties [1–8] and number of promising applications. Indeed, GaN has become the third major semiconductor in technology after Si and GaAs and still expands its presence on the market primarily for optoelectronic devices able to work in the blue and UV spectral ranges. In contrast to GaN, research on β-Ga$_2$O$_3$ is still on the initial stage. Due to its ultrawide bandgap and extremely high break-down field value (~8 MV/cm) which is larger than those of III-nitrides (3.3 MV/cm for GaN) and silicon carbide (2.5 MV/cm for 4H-SiC), gallium oxide is considered as the most promising candidate for the

K. V. Karabeshkin · P. A. Karaseov · A. I. Struchkov (✉) · A. I. Titov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

A. Azarov
Centre for Materials Science and Nanotechnology, University of Oslo, Oslo, Norway

D. S. Gogova
Central Lab of Solar Energy At the Bulg. Acad. Sci, 1784 Sofia, Bulgaria

next generation power electronics as well as for solar blind UV photodetectors and sensors.

Ion implantation is a powerful technique for materials modification and, in particular, for selective area doping in device processing. However, the ion irradiation of semiconductors is always accompanied by formation of radiation defects, which can dramatically affect the irradiated layers properties. Therefore, a deep understanding of the processes of defects formation and their evolution is urgently needed for electronic device fabrication, development of new devices and expanding their functionality.

A large number of studies has already been devoted to the accumulation of defects during the implantation of accelerated heavy and medium-mass ions in GaN [6–13]. In its turn, the data on defect formation in gallium oxide during irradiation with similar ions are limited and all the main results concern only the most stable beta phase [14–17]. At the same time, such data are urgently needed.

The aim of the present report is to compare the damage accumulation in the selected semiconductors for room temperature irradiation by 40 keV P ions.

26.2 Experimental

Wurtzite GaN epilayers, grown by MOCVD on *c*-plane sapphire substrates and monoclinic β-Ga$_2$O$_3$ single crystals grown by HVPE (Tamura Corporation), have been irradiated using 500 kV HVEE implanter by 40 keV P$^+$ ions in a wide dose range. All the implants were carried out at room temperature (RT) and at 7° off the channeling directions. The ion fluxes were close to each other and equal to 1.9×10^{12} and 1.51×10^{12} cm^{-2} s^{-1} for GaN and β-Ga$_2$O$_3$ targets, respectively. In order to compare the irradiation effects in two materials the ion doses were normalized to the average number of displacements per atom (DPA). So, the normalized values of ion fluxes were 3.6×10^{-3} and 2.4×10^{-3} DPA s^{-1} for GaN and β-Ga$_2$O$_3$, respectively. Values of DPA were calculated using TRIM code (version SRIM 2013) [18] with effective threshold energies for atomic displacements of 25 eV for all Ga, O and N sub-lattices. DPA values are the concentrations of ion-beam-generated lattice vacancies at the maximum of the nuclear energy loss profile n_v normalized to the atomic concentration: DPA = $n_v \times \Phi/n_{at}$, where for GaN $n_{at} = 8.85 \times 10^{22}$ atoms $\times$ cm^{-3} and for β-Ga$_2$O$_3$ $n_{at} = 9.45 \times 10^{22}$ atoms $\times$ cm^{-3}. So, 1 DPA is equal to 5×10^{14} cm^{-2} for gallium nitride and 6.3×10^{14} cm^{-2} for *beta* – gallium oxide.

Implantation-produced disorder was measured by Rutherford backscattering/channeling (RBS/C) spectrometry with 0.7 meV ^{4}He^{2+} ions incident along the channeling direction and backscattered to 103°. The effective number of scattering centers (referred to below as "relative disorder") were deduced from RBS/C spectra using one of the conventional algorithms [19].

26.3   Results and Discussion

Figure 26.1 shows the depth distributions of relative disorder produced in GaN and β-Ga$_2$O$_3$ by 40 keV P ions implanted with the doses resulting in roughly the same values of disorder maxima in the target bulk for both materials. These doses were low enough to prevent formation of high defects concentration into the materials and, accordingly, affect the disorder saturation. The TRIM calculated vacancy generation function in β-Ga$_2$O$_3$ is also plotted in this figure. It should be noted that according to the TRIM simulations the generation functions in GaN and β-Ga$_2$O$_3$ are nearly identical with the difference being within 5%.

It is clearly seen that the main features of both distributions are quite similar for both materials. Really, the both disorder versus depth profiles are bimodal with the distinct bulk defect peaks (BDP) and surface disordered layers (SDL). However, the shape of defect profiles as well as the kinetics of disorder accumulation with increasing dose are significantly different in GaN and β-Ga$_2$O$_3$. The first thing that attracts attention is more than an order of magnitude difference in the rate of accumulation of stable radiation damage for these two semiconductors as clearly illustrated by Fig. 26.2 showing the dose dependencies of the relative disorder in the BDP. It is easy to see that, for low concentration of defects, radiation disorder accumulates in BDP approximately 20 times faster in the case of ion irradiation of gallium oxide than for gallium nitride.

Moreover, Fig. 26.2 also demonstrates that for these two gallium compounds BDP amplitude saturates with increasing dose and does not reach complete amorphization level. However, for GaN, the saturation level is usually about $0.3-0.4$ [6–10], but for β-Ga$_2$O$_3$—0.8 (this work) or 0.9 [15]. Actually, the last data are equal each other taking in account experimental errors.

Fig. 26.1 Depth profiles of relative disorder in GaN (dose 3.9 DPA) and β-Ga$_2$O$_3$ (dose 0.147 DPA) implanted at room temperature with 40 keV P ions. A profile of P-ion-generated lattice vacancies (with an arbitrary vertical scale) predicted by TRIM simulations is also shown by dashed line
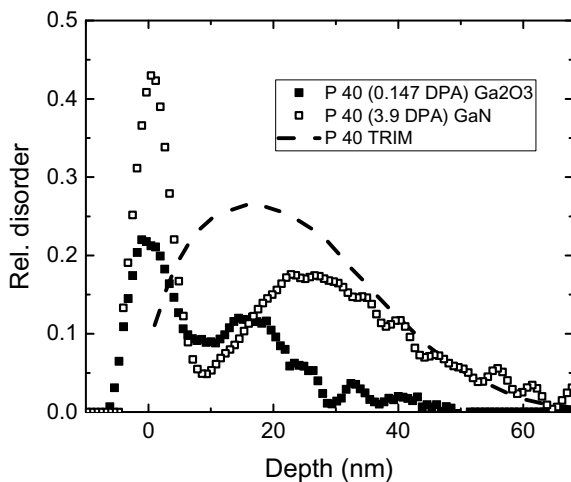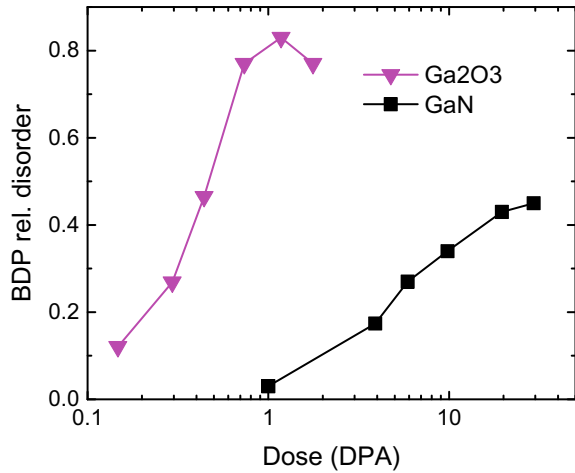
Figure 26.3 shows the dose dependences of the integral disorder in the SDL peak
for both semiconductors. It is seen that the damage in the SDL accumulates more
efficiently in β-Ga$_2$O$_3$ as compared to that of GaN. Thus, according to the results
shown in Figs. 26.2 and 26.3, it can be concluded that the radiation resistance of
GaN is more than an order of magnitude higher than that for β-Ga$_2$O$_3$.

An additional fact is that the position of BDP moves deeper into the crystal
bulk for GaN with ion dose accumulation (see also [6, 12]), while, BDP position
practically does not change in β-Ga$_2$O$_3$. As illustrated by Fig. 26.4 showing the dose
dependencies of BDP position for both semiconductors. Indeed, it should be noted
that the dose range for β-Ga$_2$O$_3$ is approximately an order of magnitude lower than
that for GaN and in such a narrow range, position of stable defect maximum for
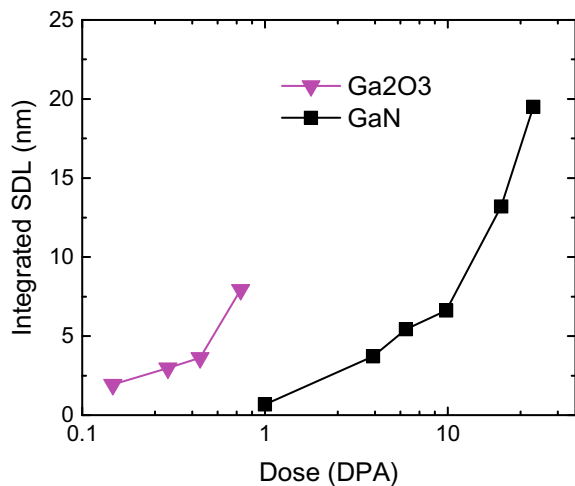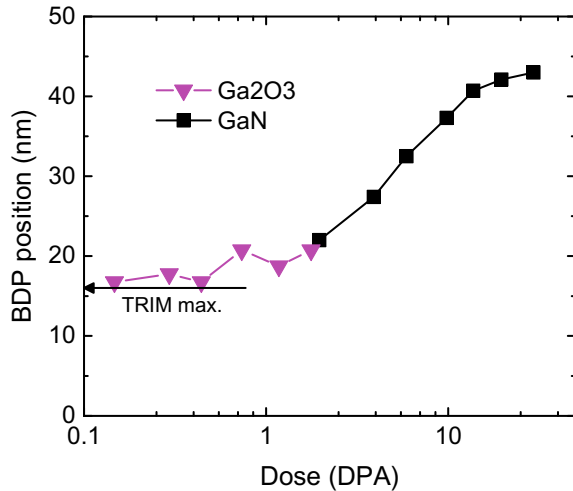
Fig. 26.4 Bulk peak position vs ion dose in GaN and β-Ga$_2$O$_3$ bombarded at room temperature by 40 keV P ions. The arrow indicates the position of generation function maximum calculated by TRIM

the last material also does not change very much. Such a difference in the behavior of the BDP with increasing ion dose, can, at least partly, explain the difference in the saturation values of the BDP for both semiconductors. Really, it was shown in [20] that one of the reasons for the BDP saturation, is caused by the shift of this defect maximum inward from the region where the point defects are generated—the building material for the formation of stable damage.

Therefore, this mechanism can be excluded for Ga$_2$O$_3$ since there is no BDP shift position with increasing ion dose in this material. However, other reasons causing the damage growth saturation can play a role. For example, as was shown in [16] an implantation-induced phase transition from β to κ-Ga$_2$O$_3$ can take place during ion irradiation and dramatically change both the damage accumulation kinetic as well as channeling conditions during RBS/C analysis. Thus, the mechanisms of the damage saturation in β-Ga$_2$O$_3$ and GaN should be different, and more detailed studies are required to better understand the physical nature of the discussed effects.

26.4 Conclusions

The kinetics of damage accumulation during P ion irradiation at room temperature were compared for two gallium compounds: β-Ga$_2$O$_3$ and GaN. The following main results have been achieved:

(1) Damage accumulation rates of both the BDP and SDL in β-Ga$_2$O$_3$ is about an order of magnitude higher than those for GaN indicating that β-Ga$_2$O$_3$ is more susceptible to ion irradiation.

(2) The BDP position in GaN shifts towards the crystal bulk with increasing ion dose, while that in β-Ga$_2$O$_3$ practically does not change.

(3)   The BDP value saturates below full amorphization in both semiconductors; however, the saturation level for β-Ga$_2$O$_3$ is approximately twice higher than that for GaN.

References

1. J. Zhang, J. Shi, D. Qi, L. Chen, K. Zhang, Recent progress on the electronic structure, defect, and doping properties of Ga2O3. APL Mater. **8**, 020906 (2020)
2. S.J. Pearton, F. Ren, M. Tadjer, J. Kim, Perspective: Ga$_2$O$_3$ for ultra-high power rectifiers and MOSFETS. J. Appl. Phys. **124**, 220901 (2018)
3. S.J. Pearton, F. Ren, E. Patrick, M.E. Law, A.Y. Polyakov, Review-Ionizing radiation damage effects on GaN devices. ECS J. Solid State Sci. Technol. **5**, Q35 (2016)
4. A.Y. Polyakov, N.B. Smirnov, I.V. Shchemerov, D. Gogova, S.A. Tarelkin, I.-H. Lee, S.J. Pearton, Electrical properties of bulk, non-polar, semi-insulating M-GaN grown by the ammonothermal method. ECS J. Solid State Sci. Technol. **7**, 260 (2018)
5. D.A. Zatsepin, D.W. Boukhvalov, A.F. Zatsepin, Y.A. Kuznetsova, D. Gogova, Y. V. Shur, A.A. Esin, Atomic structure, electronic states, and optical properties of epitaxially grown β-Ga2O3 layers. Superlattices Microstruct. **120**, 90 (2018)
6. S.O. Kucheyev, A.Y. Azarov, A. I. Titov, P.A. Karaseov, T.M. Kuchumova, Energy spike effects in ion-bombarded GaN. J. Phys. D **42**, 085309 (2009)
7. S.O. Kucheyev, J.S. Williams, C. Jagadish, J. Zou, G. Li, A.I. Titov, Effect of ion species on the accumulation of ion-beam damage in GaN. Phys. Rev. B **64,** 035202 (2001)
8. S.O. Kucheyev, J.S. Williams, S.J. Pearton, Ion implantation into GaN. Mater. Sci. Eng. R **33**, 51 (2001)
9. E. Wendler, A. Kamarou, E. Alves, K. Gaertner, W. Wesch, Three-step amorphisation process in ion-implanted GaN at 15 K. Nucl. Instrum. Methods Phys. Res. B **206**, 1028 (2003)
10. F. Gloux, T. Wojtowicz, , P. Ruterana, K. Lorenz, E. Alves, Transmission electron microscopy investigation of the structural damage formed in GaN by medium range energy rare earth ion implantation. J. Appl. Phys. **100,** 073520 (2006)
11. A.Y. Azarov, A.I. Titov, S.O. Kucheyev, Effect of pre-existing disorder on surface amorphization in GaN. J. Appl. Phys. **108**, 033505 (2010)
12. A.I. Titov, P.A. Karaseov, A.Y. Kataev, A.Y. Azarov, S.O. Kucheyev, Model for radiation damage buildup in GaN. Nucl. Instr. Meth. Phys. Res. Sec. B, B **277**, 80 (2012)
13. P.A. Karaseov, K.V. Karabeshkin, A.I. Titov, et al., Single and molecular ion irradiation-induced effects in GaN: experiment and cumulative MD simulations. J. Phys. D: Appl. Phys. **50**(50), 505110 (2017)
14. K. Lorenz, M. Peres, M. Felizardo et al., Doping of Ga2O3 bulk crystals and NWs by ion implantation. Proc. of SPIE. **8987**, 89870M-M89871 (2014)
15. E. Wendler, E. Treiber, J. Baldauf, S. Wolf, C. Ronning, High-level damage saturation below amorphisation in ion implanted β-Ga$_2$O$_3$. Nucl. Instr. Meth. Phys. Res. Sec. B, B379, **85** (2016)
16. E.A. Anber, D. Foley, A.C. Lang, et al., Structural transition and recovery of Ge implanted β-Ga2O3. Appl. Phys. Lett. **117**, 152101 (2020)
17. J. Kim, S.J. Pearton, C. Fares et al., Radiation damage effects in Ga$_2$O$_3$ materials and devices. J. Mater. Chem. C **7**, 10 (2019)
18. J.F. Ziegler, J.P. Biersack, U. Littmark, in *The Stopping and Range of Ions in Solids* (Pergamon Press, NY, 1985). Ziegler, J. F. SRIM-2013 software package. http://www.srim.org

19. K. Schmid, Some new aspects for the evaluation of disorder profiles in silicon by backscattering. Radiat. Eff. **17**, 201 (1973)
20. A.I. Titov, K.V. Karabeshkin, P.A. Karaseov, A.I. Struchkov, The formation of radiation damage in GaN during successive bombardment by light ions of various energies. Vacuum **173**, 109149 (2020)

Chapter 27
Molecular Effect in Damage Formation in β-Ga$_2$O$_3$

Konstantin V. Karabeshkin, Andrei I. Struchkov, Andrei I. Titov, Alexander Azarov, Daniela S. Gogova, and Platon A. Karaseov

Abstract Density of collision cascades is one of the most important parameters determining defect formation in materials under ion bombardment. Here we study the role of the cascade density on the damage buildup in β-Ga$_2$O$_3$ using "molecular effect", i.e. compare the damage formation under irradiations with monatomic (F, P) and small molecular (PF$_2$) ions. Results show that a strong collision cascade density effect takes place in β-Ga$_2$O$_3$, so that molecular ions produce more damage both near the surface and in the region between surface and bulk defect peaks.

Keywords Gallium oxide · Ion implantation · Radiation damage formation · Collision cascade · Molecular effect

27.1 Introduction

Developments in various fields of electronics are generating a huge demand for a new generation of power switching and control devices with improved performance and new materials suitable for their manufacture [1, 2]. Wide bandgap materials, such as GaN, InGaN, and SiC have been successfully industrialized in the last decade with a large impact on our daily life. Nowadays, the research is directed towards ultra-wide bandgap semiconducting materials with an energy gap exceeding 4 eV. Beta

K. V. Karabeshkin · A. I. Struchkov (✉) · A. I. Titov · P. A. Karaseov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: andrei.struchkov@yandex.ru

P. A. Karaseov
e-mail: platon.karaseov@spbstu.ru

K. V. Karabeshkin
CJSC "ELAR", St.-Petersburg, Russia

A. Azarov
Centre for Materials Science and Nanotechnology, University of Oslo, Oslo, Norway

D. S. Gogova
Central Lab of Solar Energy at the Bulgarian Academy of Sciences, Sofia, Bulgaria

gallium oxide (β-Ga$_2$O$_3$), having a wide bandgap of approximately 4.85 eV, and especially the high break-down field value (~8 MV $\times$ cm^{-1}) is the most attractive material for the next generation power electronics [2–6] as well as for solar blind UV photodetectors [7] and sensors [8]. It is obtainable directly from the melt by scalable growth methods [2, 3, 7, 8] with further scale-up capabilities that would enable low production costs as compared with other ultra-wide bandgap materials. The electrical conductivity of β-Ga$_2$O$_3$ seems to have a multiple origin, such as native defects resulting from growth environment (growth method, temperature, and atmosphere), residual impurities and intentional doping. Most of β-Ga$_2$O$_3$ samples are electrically insulating when grown under oxidizing conditions and semiconducting under reducing conditions [9]. Despite a possible realization of hole conductivity in gallium oxide grown under specific conditions have been recently reported [10, 11], p-type doping is still challenging.

Ion implantation is a widely used technique in semiconductor technology for selective area properties alteration, like doping, electrical isolation, etc. However, irradiation with accelerated ions is always accompanied by defect production and further formation of stable radiation damage. Kind and rate of damage production is generally affected by irradiation parameters such as ion energy, ion mass, ion-beam current, substrate temperature, etc. [12]. Stopping ions colliding with the target atoms kick-out them from regular lattice positions forming so-called collision cascade consisting of vacancies and self-interstitial atoms. Volumetric density of these individual cascades can dramatically affect the resulting damage, as have been experimentally shown for such materials as Si [13–15], SiC [16, 17], ZnO [18] and GaN [19–23]. Corresponding effects can be conveniently studied using irradiation with molecular ions [13–21]. Information on ion-beam induced damage formed in β-Ga$_2$O$_3$ during irradiation with accelerated ions is very sparse. In this contribution, we study the structural disorder in β-Ga$_2$O$_3$ single crystals irradiated at room temperature with monatomic P$^+$ and F$^+$ and molecular PF$_2^+$ ions.

27.2   Experimental

Monoclinic β-Ga$_2$O$_3$ single crystals grown by HVPE method (Tamura Corporation equipment), have been irradiated by F$^+$, P$^+$, and PF$_2^+$ ions using 500 kV HVEE implanter. All the implants were carried out at room temperature and at an angle 7° off the channeling direction.

Special care of the experimental conditions has to be taken as irradiation parameters strongly influence the resulting damage appearing in the target. All experiments were carried out in compliance with the correct irradiation conditions that allow performing meaningful analysis of the experimental data. As it has been established earlier [14, 21, 24], that have to be kept constant are ion energy per atomic mass unit (amu), dose expressed in displacements per target atom (*DPA*), beam flux in *DPA/s*. In this case, the only difference between all irradiations would be the density of individual collision cascades formed by stopping monatomic and molecular ions.

Table 27.1 List of irradiation parameters used in the experiments

Ion	Energy		Fluence per 1 *DPA*	Flux		Dose rate
	keV	keV/amu	10^{14} cm^{-2}	10^{11} cm^{-2} s^{-1}	μA/cm^2	10^{-3}*DPA*/s
F	25	1.3	11.4	27.6	0.442	2.41
P	40	1.3	6.28	15.1	0.242	2.41
PF$_2$	100	1.45	3.00	7.25	0.116	2.41

DPA values are the concentrations of ion-beam-generated lattice vacancies at the maximum of the nuclear energy loss profile n_v normalized to the atomic concentration: $DPA = n_v \times \Phi/n_{at}$, where $n_{at} = 9.45 \times 10^{22}$ atoms $\times$ cm^{-3} in β-Ga$_2$O$_3$. In the case of molecular ion bombardment, dose was estimated as $DPA(\text{PF}_2) = DPA(\text{P}) + 2 \times DPA(\text{F})$. In order to normalize ion fluences, we calculated values of DPA using TRIM code (version SRIM 2013) [25] with effective displacement threshold energy equal to 25 eV for both Ga and O sublattices.

Implantation-produced disorder was measured by Rutherford backscattering spectrometry in channeling mode (RBS/C). The effective number of scattering centers (which is proportional to the number of displaced target atoms and hence referred to below as "relative disorder") was deduced from RBS/C spectra using one of the conventional algorithms [26].

Table 27.1 presents the list of all irradiation parameters used in the present study.

27.3 Experimental Results

Typical RBS/C spectra collected after irradiation of β-Ga$_2$O$_3$ samples with 40 keV P and 100 keV PF$_2$ ions are shown in Figs. 27.1a and b, respectively. It is clearly seen from this figure that, at low doses the damage-depth distribution created by monatomic ion (P) species is bimodal consisting of two peaks, one located at the sample surface and second in the crystal bulk (see Fig. 27.1a). Amount of damage grows up with the dose increase and two peaks merge into one that grows up and enhances deeper into the bulk.

Molecular ions produce more damage than monatomic ones (see Fig. 27.1b). It reveals significant and wide surface peak and slightly pronounced bulk maximum. Consistently with previous reports [27, 28], damage does not reach full amorphization level saturating at ~ 80–90%. Note that, according to the binary collision approximation, amount of point defects generated during irradiation was kept constant for all ions used. Hence, the difference seen in the formation of stable disorder is caused by nonlinearities appearing due to different densities of individual collision cascades created by monatomic P and molecular PF$_2$ ions.

The last statement is further illustrated by Fig. 27.2, where the relative disorder created in the β-Ga$_2$O$_3$ sample by irradiation with different ions to the same dose

Fig. 27.1 RBS/C spectra collected from β-Ga$_2$O$_3$ samples irradiated with 40 keV P (**a**) and 100 keV PF$_2$ (**b**) ions. Doses (in *DPA*) are indicated in the legend. Random refers to the spectra taken at non-channeling direction
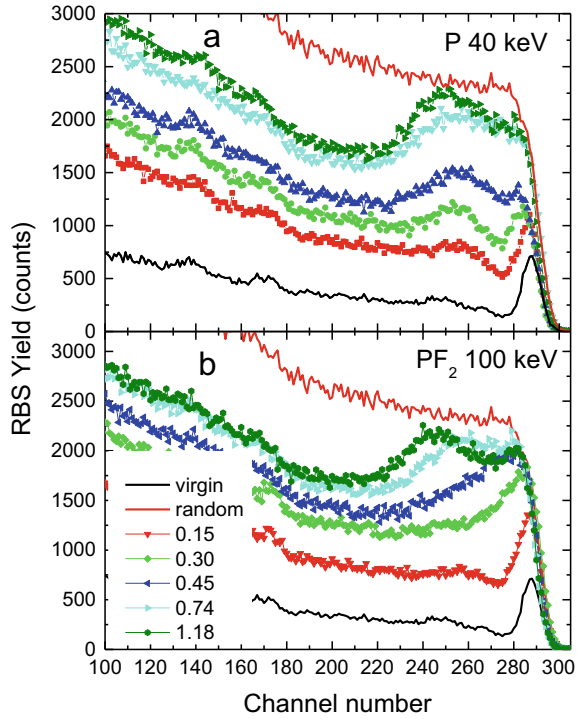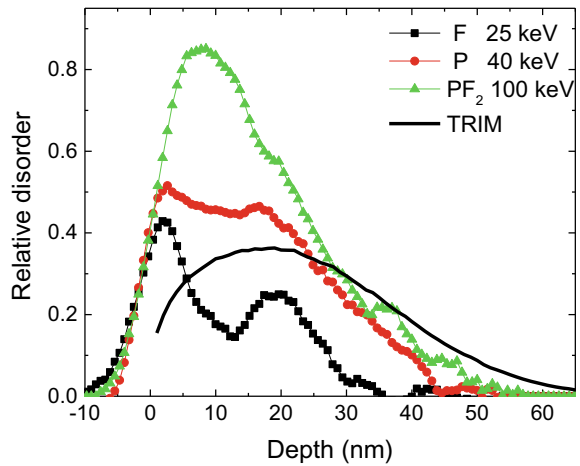

Fig. 27.2 Relative disorder created after irradiation to the same dose of 0.44 *DPA* by different ions as indicated in the legend, as extracted from RBS/C spectra (symbols) and displacement generation function simulated by TRIM [23] (black line)

of 0.44 *DPA* is plotted as a function of target depth. It could be seen in Fig. 27.2 that depth distribution of the damage created by relatively light monatomic ions is bimodal split into two peaks located at the sample surface and in the bulk. At small doses the heavier P ions generate higher level of damage than F ions over the all depths. Bulk defect peak position coincides well with the maximum of atomic displacement generation (compare positions of maxima of bulk peaks and vacancy generation function in Fig. 27.2). Molecular ions produce more damage at the depths between surface and bulk peaks, thus enhancing the surface defect peak. Thus, the so-called "*molecular effect*" is experimentally found. Below we will discuss a possible mechanism of the effects observed.

27.4 Discussion

First, let us briefly point out the main features of the processes that take place in the target during our experiments. When solids are bombarded with accelerated ions, two major mechanisms underlie energy transfer from ion to the target [12]. Energy could go to an electronic subsystem (thus creating ionization and excitation of atoms) and to target nuclei. Ions are able to shift nuclei from equilibrium positions in the crystal lattice in elastic collisions creating point defects, like vacancies and interstitial atoms. The latter process is frequently referred to as collision cascade formation. Volumetric density of collision cascade (f_{av}) is one of the important parameters that affect radiation-induced effects [12, 17, 21]. Recently in [21] we have developed a simple way to estimate f_{av} using binary collision approximation data generated by the TRIM code [25]. The average density of individual collision cascade formed in the β-Ga$_2$O$_3$ target by F, P, and PF$_2$ ions and the fraction of ions η that form at least one sub-cascade in a 5 nm thick layer parallel to the sample surface in the vicinity of a given depth are shown in the Fig. 27.3. It should be noted that our experiments were designed in such a way to make the cascades density the only difference between irradiation cases while all other parameters were kept the same (see Table 27.1). Thus, we ascribe higher amount of damage formed by P ions to two main reasons: (i) heavier P ions produce denser cascades than those for F ions (see Fig. 27.3a), and (ii) smaller fraction of light F ions produce atomic displacements.

When solids are bombarded with cluster ions, collision cascades induced by individual constituent atoms of the cluster overlap each other. It is seen from Fig. 27.3a that the collision cascades produced by P and F atoms, which form a cluster ion, effectively overlap only in the near-surface region (in the first 15 nm), giving rise to formation of collision cascade with increased density. Indeed, dense cascade generated by a cluster ion deposits higher energy density as compared to that for monatomic ions at the same velocity. Subsequently the formation of structure damage in the target material is enhanced under the impact of PF$_2$ molecules. This enhancement is called "molecular effect". It explains the larger amount of structural damage created by molecular ions in the region close to the target surface.

Fig. 27.3 **a** Average density of individual collision cascade formed in the target by the same ions. **b** Fraction of ions η that form at least one sub-cascade in a 5 nm thick layer in the vicinity of a given depth
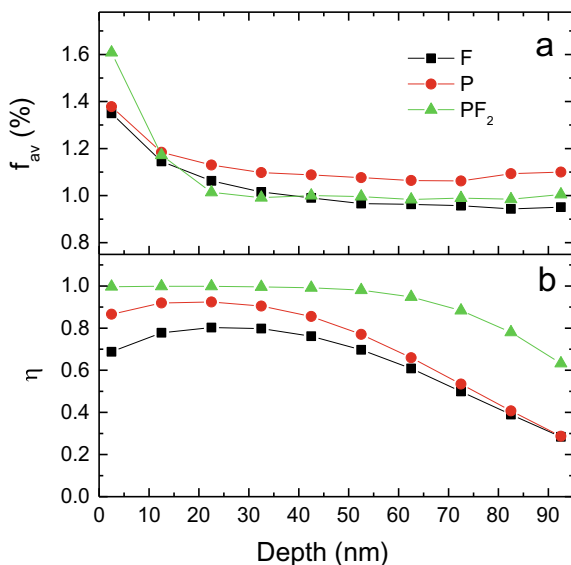
27.5 Conclusions

Damage formation in β-Ga_2O_3 under monatomic F^+ and P^+ as well as molecular PF_2^+ ions was studied. Damage created by relatively light monatomic ions is bimodal consisting of two distinct peaks located at the sample surface and in the bulk, close to the maximum of nuclear energy loss. In contrast, molecular ions produce more damage both near the surface and in the region between surface and bulk defect peaks, so the apparent damage profiles induced by these ions become unimodal with a single broad defect peak. The observed damage enhancement is attributed to the strong collision cascade density effect in ion-bombarded β-Ga_2O_3. Furthermore, we demonstrate that, in contrast to many other materials, damage accumulation in β-Ga_2O_3 saturates at ~ 80–90% of amorphization level with increasing ion dose in all cases studied. The dose needed to reach this heavily damaged state is estimated to be about 1 *DPA*.

References

1. P.J. Wellmann, Z. Anorg, Power electronic semiconductor materials for automotive and energy saving applications–SiC, GaN, Ga_2O_3 and diamond. Allg. Chem. **643**, 1312 (2017)
2. S.I. Stepanov, V.I. Nikolaev, V.E. Bougrov, A.E. Romanov, Gallium oxide: properties and applications—a review. Rev. Adv. Mater. Sci. **44**, 63 (2016)
3. A. Kuramata, K. Koshi, S. Watanabe, Y. Yamaoka, T. Masui, S. Yamakoshi, High-quality β-Ga_2O_3 single crystals grown by edge-defined film-fed growth. Japan. J. Appl. Phys. **55**, 1202A2 (2016)

4. A.Y. Polyakov, N.B. Smirnov, I.V. Shchemerov, D. Gogova, S.A. Tarelkin, S.J. Pearton, Compensation and persistent photocapacitance in homoepitaxial Sn-doped β-Ga$_2$O$_3$. J. Appl. Phys. **123**(11), 115702 (2018)

5. S.J. Pearton, J. Yang, P.H. Cary IV, F. Ren, J. Kim, M.J. Tadjer, M.A. Mastro, A review of Ga$_2$O$_3$ materials, processing, and devices. Appl. Phys. Rev. **5**, 011301 (2018)

6. Y. An, X. Chu, Y. Huang, Y. Zhi, D. Guo, P. Li, Z. Wu, W. Tang, Au plasmon enhanced high performance β-Ga$_2$O$_3$ solar-blind photo-detector. Prog. Nat. Sci.: Mater. Int. **26**, 65 (2016)

7. K. Shimamura, E.G. Víllora, T. Ujiie, K. Aoki, Excitation and photoluminescence of pure and Si-doped β-Ga$_2$O$_3$ single crystals. Appl. Phys. Lett. **92**, 201914 (2008)

8. A.Y. Polyakov, N.B. Smirnov, I.V. Shchemerov, D. Gogova, S.A. Tarelkin, I.-H. Lee, S.J. Pearton, Electrical properties of bulk, non-polar, semi-insulating M-GaN Grown by the ammonothermal method. ECS J. Solid State Sci. Technol. **7**, 260 (2018)

9. D.A. Zatsepin, D.W. Boukhvalov, A.F. Zatsepin, Y.A. Kuznetsova, D. Gogova, Y. V. Shur, A.A. Esin, Atomic structure, electronic states, and optical properties of epitaxially grown β-Ga$_2$O$_3$ layers. Superlattices Microstruct. **120**, 90 (2018)

10. M.R. Lorenz, J.F. Woods, R.J. Gambino, Some electrical properties of the semiconductor β-Ga$_2$O$_3$. J. Phys. Chem. Solids **28**, 403–404 (1967)

11. E. Chikoidze et al., Enhancing the intrinsic p-type conductivity of the ultrawide bandgap Ga$_2$O$_3$ semiconductor. J. Mater. Chem. C **7**, 10231 (2019)

12. W. Wesch, E. Wendler (eds.) Ion beam modification of solids: ion-solid interaction and radiation damage. (Springer, 2016)

13. A.I. Titov, S.O. Kucheyev, V.S. Belyakov, AYu. Azarov, Damage buildup in Si under bombardment with MeV heavy atomic and molecular ions. J. Appl. Phys. **90**, 3867 (2001)

14. A.I. Titov, AYu. Azarov, L.M. Nikulina, S.O. Kucheyev, Damage buildup and the molecular effect in Si bombarded with PFn cluster ions. Nucl. Instrum. Methods Phys. Res. B **256**, 207 (2007)

15. A.I. Titov, P.A. Karaseov, AYu. Azarov, S.O. Kucheyev, Damage formation in Si under irradiation with PF +n ions of different energies. Nucl. Instrum. Methods Phys. Res. B **267**, 2701 (2009)

16. AYu. Azarov, A.I. Titov, P.A. Karaseov, A. Hallen, Effect of collision cascade density on radiation damage in SiC. Nucl. Instrum. Methods Phys. Res. B **267**, 1247 (2009)

17. L.B. Bayu Aji, J.B. Wallace, S.O. Kucheyev, Effects of collision cascade density on radiation defect dynamics in 3C-SiC. Sci. Rep. **7**, 4703 (2017)

18. A.Y. Azarov, S.O. Kucheyev, A.I. Titov, P.A. Karaseov, Effect of the density of collision cascades on ion implantation damage in ZnO. J. Appl. Phys. **102**, 083547 (2007)

19. A.I. Titov, P.A. Karaseov, V.S. Belyakov, K.V. Karabeshkin, A.V. Arkhipov, S.O. Kucheyev, AYu. Azarov, Molecular effect on surface topography of GaN bombarded with PF4 ions. Vacuum **86**, 1638 (2012)

20. P.A. Karaseov, K.V. Karabeshkin, E.E. Mongo, A.I. Titov, M.W. Ullah, A. Kuronen, F. Djurabekova, K. Nordlund, Experimental study and MD simulation of damage formation in GaN under atomic and molecular ion irradiation. Vacuum **129**, 166 (2016)

21. S.O. Kucheyev, A.Y. Azarov, A.I. Titov, P.A. Karaseov, T.M. Kuchumova, Energy spike effects in ion-bombarded GaN. J. Phys. D: Appl. Phys. **42**, 085309 (2009)

22. V.A. Shalygin et al., Selective terahertz emission due to electrically excited 2D plasmons in AlGaN/GaN heterostructure. J. Appl. Phys. **126**(18), 183104 (2019)

23. M.Y. Vinnichenko, et al., Near-infrared optical absorption in GaN/AlN quantum wells grown by molecular-beam epitaxy. J. Phys.: Conf. Ser. **1482**, 012021 (2020)

24. AYu. Azarov, A.I. Titov, Accumulation of structural defects in silicon irradiated with PF +n cluster ions with medium energies. Semiconductors **41**, 5 (2007)

25. J.F. Ziegler, J.P. Biersack, U. Littmark, in *The Stopping and Range of Ions in Solids* (Pergamon Press, NY, 1985). Ziegler, J. F. SRIM-2013 software package. http://www.srim.org

26. K. Schmid, Some new aspects for the evaluation of disorder profiles in silicon by backscattering. Radiat. Eff. **17**, 201 (1973)

27. E. Wendler, E. Treiber, J. Baldauf, S. Wolf, C. Ronning, High-level damage saturation below amorphisation in ion implanted β-Ga2O3. Nucl. Instr. Meth. Phys. Res. Sec. B, B379, **85** (2016)
28. K. Lorenz, M. Peres, M. Felizardo et al., Doping of Ga2O3 bulk crystals and NWs by ion implantation. Proc. of SPIE. **8987**, 89870M-M89871 (2014)

Chapter 28
Investigation of Self-organized Protein Films Formed Under an External Electric Field

Maksim Baranov ⓘ**, Sergey Rozov** ⓘ**, and Dmitry Dicky**

Abstract Biomolecular electronics technology is currently developing rapidly. Because of this, there is a strong demand for new materials and development of new principles of biomolecular devices construction, and protein films can become such materials. In this paper we study self-organized albumin protein films. Optical microscopy and computer simulation have been used to analyze self-organization processes and to detect complex macrostructures formed under external electric field. Computational and experimental results are presented which can be used in biomolecular electronics.

Keywords Self-organization · Protein films · Biomolecular electronics

28.1 Introduction

Thin films are important objects, which attract the attention of many researchers [1–4]. Special attention is paid to biomolecular films, including protein ones [5–7]. They have a great variety of applications in medicine, drugs and food packaging, and biomolecular electronics [8–10].

Biological molecules have the property of assembling into ordered macrostructures, i.e. the ability to self-organise. Many characteristics of biomolecular films are important for the development of biomolecular electronics elements: dielectric, optical, mechanical, etc. [11, 12]. Many of these characteristics depend on the self-organization parameters. An important parameter is the stability of supramolecular biomolecular structures [13–17].

Previous work has shown that the parameters of self-organization can be controlled by varying the parameters of the experiment [18]. These are parameters such as a protein and salt concentration in the solution, dehydration temperature, and surface material.

M. Baranov (✉) · S. Rozov · D. Dicky
Peter the Great St. Petersburg Polytechnic University, St. Petersburg 195251, Russia

To develop elements of biomolecular electronics, it is necessary to investigate the effects of electric fields on the formation of biomolecular films [2, 19]. Integrated circuits are known to have small value fields. These fields can have a certain impact on the functionality of devices. Therefore, the purpose of this work is to study the effect of the electric field on the process of self-organization in protein films.

In this work the processes of self-organisation of protein films under the influence of external electric fields are investigated. Molecular dynamics and optical microscopy methods were used to analyse film structures.

28.2   Methods of Investigation

Self-organization processes in proteins start in the synthesis of proteins at the ribosome. In this case, self-organization is described by kinetic equations when the conformation of protein molecules changes. In the case of protein film formation, the laws of non-linear thermodynamics, which describe phase transitions within dried protein solutions, start to play an important role at the macro level. Depending on the experimental parameters, different types of structures are formed in protein films. Figure 28.1 shows the examples of the obtained structures.

Figure 28.1 was obtained and described in our previous work [20]. It presents albumin protein films obtained from an aqueous solution (Fig. 28.1a) and a salt solution (Fig. 28.1b). In the absence of salts in the initial protein solutions, spiral structures or radial cracks appear, but in the presence of salts, dendrites form. Both types of structures may have different geometric parameters depending on the experimental parameters. The mathematical analysis of such structures is described in [21, 22]. In addition, it was shown that in the case of dendritic structures, a phase separation of protein and salt is observed in the films.
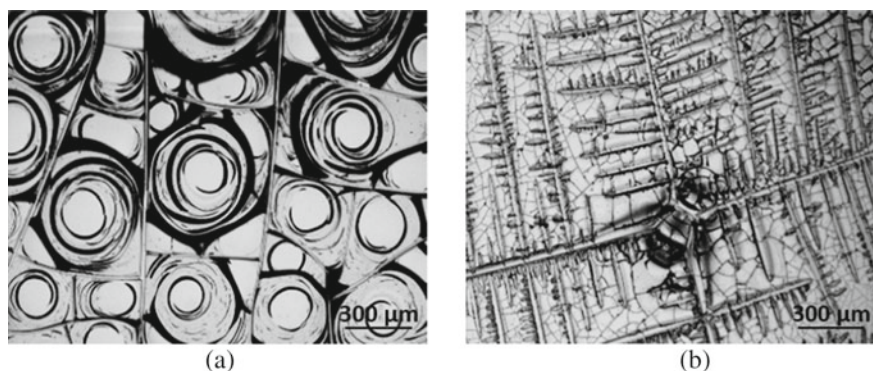

Fig. 28.1 Examples of typical structures observed in biomolecular films: spiral structure domains, or cells (**a**); dendrite cluster domains (**b**)

28.2.1 Optical Microscopy

The processes of self-organization of structures in protein films under the influence of an external electric field were investigated using the method of optical microscopy. An albumin protein solution with a concentration of 20% was chosen to prepare the solutions. Solutions with concentrations of 2 and 5% were prepared with an initial volume of 2 ml and placed in Petri dishes of 30 mm diameter. The drying process took place in a thermostat at normal pressure and temperature of 308 K. Flat electrodes connected to a direct current source were used to generate an external electric field. The area of the electrodes was several times larger that of the Petri dishes, so edge effects at the electrode boundaries can be neglected. Fields of 1, 3 and 5 V/cm were applied to the samples. Figure 28.2 shows the flowcharts of the experiments.

Optical microscope Olympus CX 33 and Altami USB camera (model UCMOS10000KPA) were used to obtain protein film images. These images have following properties: an area about 0.1 cm^2, a resolution of 1280 × 960 pixels, a depth of 24 bits.
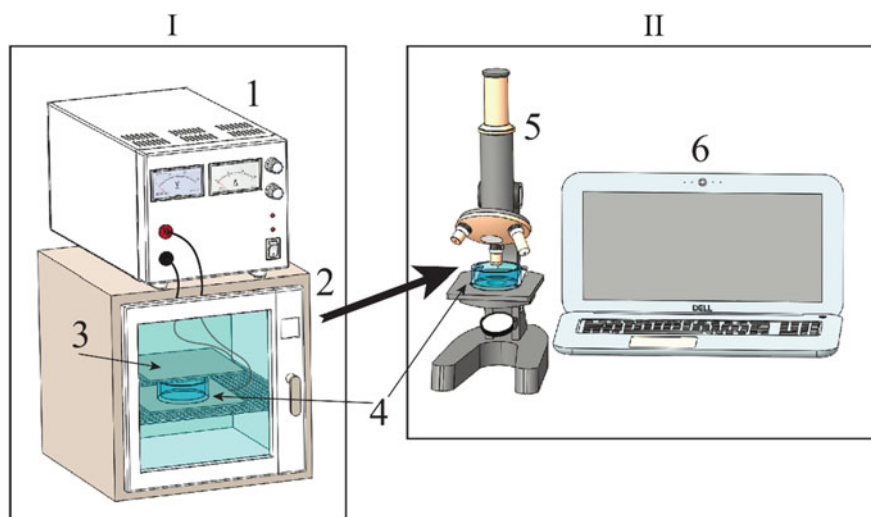

Fig. 28.2 Schematic representation of an experimental setup for the study of the spatial structure of a protein film during its dehydration in an electric field: 1—the generator of the electric field; 2—thermostat; 3—fixed condenser plates; 4—Petri dish with aqueous solution; 5—optical microscope and camera; 6—computer

28.2.2 Molecular Simulation

The processes of protein self-organization are characterized by the free energy value of protein molecules. It is known that the protein energy of a densely packed globule E is minimal and increases with a small expansion of the globule [24]. This paper attempts to explore the free energies of a molecule to analyze the self-organization processes of cluster proteins under the influence of an external electric field.

Molecular simulation is a promising method for investigation of the properties of molecular systems [21, 23–25]. A Visual Molecular Dynamics program was used to create molecular systems. The estimated systems consist of four albumin molecules, surrounded by water molecules. The distance from the boundaries of the molecules to the edge of the cube was taken equal to 10 Å.

For greater reliability, each system contains four albumin protein molecules. Na^+ and Cl^- ions were added to the solvation cube to neutralize the electric charge of the system. Nanoscale Molecular Dynamics (NAMD) allows the application of an external electric field to molecular systems expressed in units kcal/(mol·Å·e) where 1 kcal/(mol·Å·e) $\approx 4.35 \times 10^8$ V/m.

Simulations of biomolecular systems were performed in the NAMD program using CHARMM27 force fields. Before simulations of each system, minimization algorithm was used to stabilize molecular clusters energy. Minimization of energy was carried out for 5000 time steps, which corresponds to 1 ps. In the next step, the energy of the molecular clusters was calculated for 1 ns at temperature of 308 K. The time dependences of the energy values were constructed on the basis of these data.

28.3 Results

Various images have been obtained: the most characteristic ones look like quasiperiodic structures consisting of specific cells with the size of 200–500 nm (Fig. 28.3).

As illustrated in Fig. 28.3, the density of radial cracks in films increases with an increase of the protein concentration in the initial solution., As the applied electric field increases, the number of structures first increases, and then decreases in the case of low protein concentration, and the structures lose their spiral form in the case of greater concentrations. It can be concluded that an external parameter in the form of an electric field with sufficient influence produces chaos in the studied system. It is worth noting that the films formed under the influence of the electric field are unstable and have weak adhesion to the substrate.

As a result of computer simulation, the time dependences of the energies of molecular clusters were obtained from the values of the applied electric fields. Figure 28.4 shows the corresponding graphs.

Examination of these dependencies indicates that external electric fields weakly affect the energy of molecular clusters. The differences in the influence of electric fields with values of a few V/cm units are practically unobservable. It is likely that to
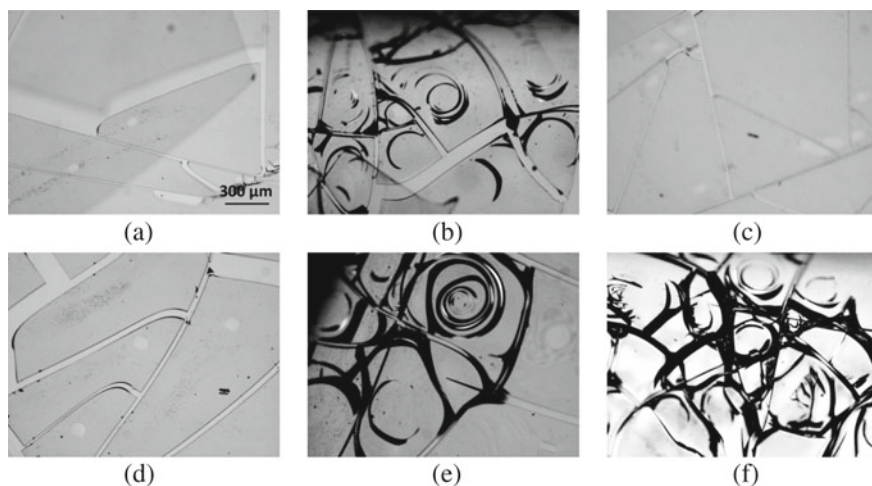
Fig. 28.3 Images of albumin protein films, formed from 2% (**a**, **b**, **c**) and 5% (**d**, **e**, **f**) solution under the influence of the external electric fields of 1 V/cm (**a**, **d**); 3 V/cm (**b**, **e**); 5 V/cm (**c**, **f**)
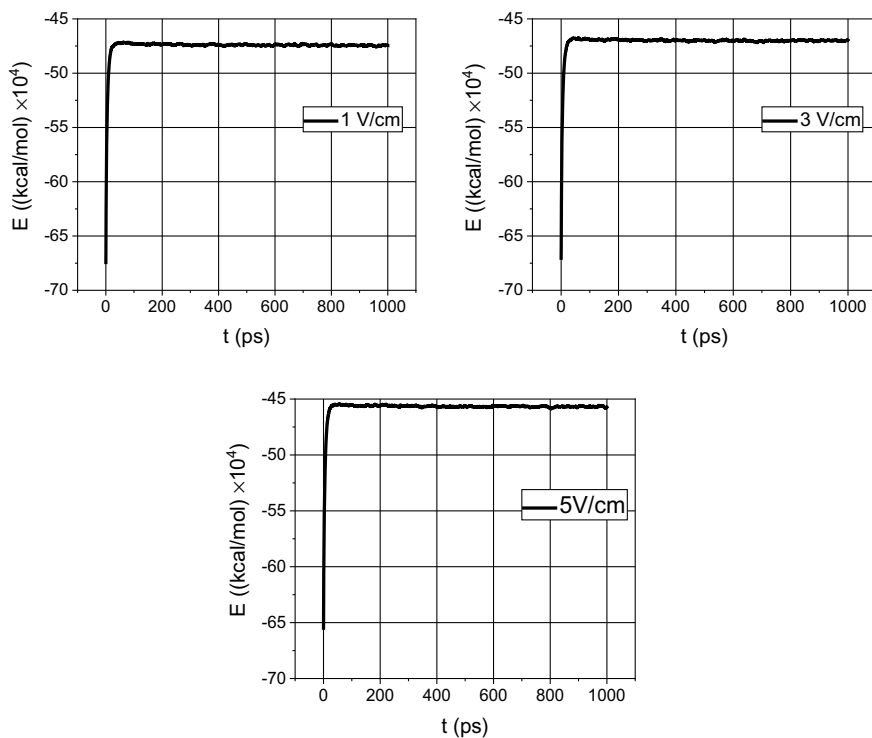

Fig. 28.4 Energies of molecular cluster with intensity of external electric field of 1, 3 and 5 V/cm versus time

achieve a greater difference in the effect of weak fields on the energy of the molecular clusters, it is necessary to increase the modelling time.

Comparing the results obtained by optical microscopy and molecular dynamics methods, we can conclude that weak constant electrical fields applied to biomolecular films have a weak influence on the self-organization processes in thin biomolecular films produced by isothermal dehydration. This result could be useful in the development of biomolecular electronic devices.

28.4 Conclusion

This paper discusses the study of self-organization processes in thin protein films, formed under influence of the external electric field. Optical microscopy and molecular modeling methods show that weak electrical fields of about several V/cm have slight effect on the energy of molecular clusters. A change in shape disorders and reduction of spiral structures with an increase in the magnitude of external electric field is observed. In the future, it is planned to model molecular clusters on inorganic substrates under strong external electric fields.

References

1. A. Pavlov, A. Ustinov, V. Petrov, Spin relaxation length for medium energy electrons in Pd and LiF ultrathin films. J. Electron Spectros. Relat. Phenom. **223**, 62–66 (2018)
2. A. Tretiakov, V. Kapralova, N. Sudar, O. Gryshkov, B. Glasmacher, Dielectric properties of PVDF based thin films and electrospun mats. J. Phys.: Conf. Ser. **1236**, 012009 (2019)
3. M.A. Elistratova, I.B. Zakharova, G.V. Li, R.M. Dubrovin, O.M. Sreseli, The effect of crystallization conditions on the spectral characteristics of tetraphenylporphyrin thin films. Semiconductors **53**, 51–54 (2019)
4. O. Tsybin, Nano-device with an embedded molecular film: mechanisms of excitation. Internet of Things. Smart Spaces, and Next Generation Networks and Systems. 772–777 (2015)
5. A.V. Redkov, A.V. Osipov, S.A. Kukushkin, Molecular dynamics simulation of the indentation of nanoscale films on a substrate. Tech. Phys. Lett. **42**, 639–643 (2016). https://doi.org/10.1134/S1063785016060274
6. K.G. Gareev, E.K. Nepomnyashchaya, Obtaining and characterizing a water-based magnetic fluid. Bull. Russ. Acad. Sci. Phys. **83**(7), 904–905 (2019)
7. E. Velichko, E. Savchenko, E. Nepomnyashchaya, D. Dyubo, O. Tsybin, Nanocommunication system with a laser activated molecular film. Internet of Things, Smart Spaces, and Next Generation Networks and Systems. 649–655 (2018)
8. N. Amdursky, E.D. Głowacki, P. Meredith, N. Amdursky, E.D. Głowacki, P. Meredith, Macroscale biomolecular electronics and ionics. Wiley Online Libr. **31**, 1802221 (2018). https://doi.org/10.1002/adma.201802221
9. M.C. Petty, T. Nagase, H. Suzuki, H. Naito, Molecular electronics. Springer Handbooks. 1 (2017)

10. E.N. Velichko, E.K. Nepomnyashchaya, K.G. Gareev, J. Martínez, M.C. Maicas, Characterization of Magnetite-Silica Magnetic Fluids by Laser Scattering. Appl. Sci. **11**(1), 183 (2021)
11. E.N. Velichko, E.K. Nepomnyashchaya, A.V. Sokolov, T.Y. Kudryashova, Laser correlation spectrometer for assessing the size and dynamics of changes in the size of structures in biological fluids. Opt. Spectrosc. **128**(7), 959–963 (2020)
12. E.A. Savchenko, E.N. Velichko, the use of speckle correlation analysis to determine blood flow velocity. Opt. Spectrosc. **128**(7), 998–1003 (2020)
13. E.N. Velichko, M.A. Baranov, V.M. Mostepanenko, Change of sign in the Casimir interaction of peptide films deposited on a dielectric substrate. Mod. Phys. Lett. A.**35**, 1–6 (2020)
14. E.N. Velichko, G.L. Klimchitskaya, E.N. Nepomnyashchaya, Casimir effect in optoelectronic devices using ferrofluids. J. Electr. Sci. Technol. **18**(1), 100024 (2020)
15. G.L. Klimchitskaya, V.M. Mostepanenko, E.K. Nepomnyashchaya, E.N. Velichko, Impact of magnetic nanoparticles on the Casimir pressure in three-layer systems. Phys. Rev. B **99**(4), 045433 (2019)
16. M.A. Baranov, G.L. Klimchitskaya, V.M. Mostepanenko, E.N. Velichko, Fluctuation-induced free energy of thin peptide films. Phys. Rev. E **99**(2), 022410 (2019)
17. E.N. Velichko, V.M. Mostepanenko, G.L. Klimchitskaya, Effect of increased stability of peptide-based coatings in the Casimir regime via nanoparticle doping. Phys. Rev. B **102**(16), 161405 (2020)
18. E. Savchenko, I. Kolokolnikov, E. Velichko, V. Osovskikh, L. Kiseleva, Z. Musakulova, Design of liver functional reserve estimation technique based on optical densitometry. Diagnostics **10**(8), 599 (2020)
19. N.V. Kapralov, E.S. Babich, A.V. Redkov, Molecular dynamics simulation of metal nanoislands growth. J. Phys.: Conf. Ser. **929** (2017).
20. E. Velichko, T. Zezina, M. Baranov, E. Nepomnyashchaya, O. Tsybin, Dynamics of polypeptide cluster dipole moment for nano communication applications. Internet of Things, Smart Spaces, and Next Generation Networks and Systems 675–682 (2018)
21. W. Inthavong, A. Kharlamova, C. Chassenieux, Matter, Structure and flow of dense suspensions of protein fractal aggregates in comparison with microgels. Soft matter **12**, 2785–2793 (2012). https://doi.org/10.1039/C5SM02893K
22. M. Baranov, E. Velichko, K. Greshnevikov, Analysis of fractal structures in dehydrated films of protein solutions. Symmetry **13**(1), 123 (2021)
23. A.V. Finkelstein, O. Ptitsyn. Protein Physics: a Course of Lectures **252** (Elsevier, 2016)
24. P. Grigorev, A. Zinovev, D. Terentyev, G. Bonny, E.E. Zhurkin, G. Van Oost, J.M. Noterdaeme, Molecular dynamics simulation of hydrogen and helium trapping in Tungsten. J. Nucl. Mater. **508**, 451–458 (2018). https://doi.org/10.1016/j.jnucmat.2018.05.052
25. F. Ganazzoli, G. Raffaini, Dendrimer dynamics: a review of analytical theories and molecular simulation methods. Polymers. **12**, 1387 (2020)

Chapter 29
Optimizing of ALPHIE Grid System with Particle-In-Cell Simulations

Dmitry Dyubo, Jorge Gonzalez, Oleg Tsybin, and Luis Conde

Abstract The Alternative Low Power Ion Engine (ALPHIE) currently in development by UPM Plasmalab at Universidad Politécnica de Madrid (UPM) is a promising spacecraft ion engine. The counter flow of charged particles through the grid system is a unique characteristic of this thruster. In its grid system, the ionizing electrons coming from an external cathode which also provides the neutralizing electrons in the plasma plume travel inwards the ionization chamber. These electrons are accelerated by the potential jump between the two grids (usually 500–600 V). The generated ions are extracted by the same grid system and accelerated to supersonic speeds, usually in the order of 40–50 km/s. A fully particle-in-cell (PIC) approach is used to study the effect of relevant parameters (thickness and distance between grids) in the grid system on the ion and electron dynamics. The new code fpakc (Finite element PArticle Kinetic Code) is employed for such a task. The Polytechnic University has successfully used the CST (Computer Simulation Technology) licence code installed on a supercomputer for this purpose. A fpakc 2D axial-symmetric beam-type domain is used to study one hole in two-grid system. The use of a larger time step for ions to reduce computational time is checked. The thickness and distance between grids are chosen as relevant parameters. The results will be used in the next iteration of ALPHIE to improve the ionizing electrons inflow and the extracted ion current. In addition, the new calculations will compare the advantages and disadvantages of two different approaches, CST and fpakc.

D. Dyubo (✉) · O. Tsybin
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: dyubo_db@spbstu.ru

O. Tsybin
e-mail: otsybin@rphf.spbstu.ru

J. Gonzalez
DIFFER-Dutch Institute for Fundamental Energy Research, De Zaale 20, 5612 AJ Eindhoven, The Netherlands

L. Conde
Department of Applied Physics, ETS Ingeniería Aeronáutica Y del Espacio, Universidad Politécnica de Madrid, 28040 Madrid, Spain

29.1 Introduction

Computer modeling of electric engines [1–5] is of appropriate relevance to characterize the performance and improve the design of the system. Ion Engines [6, 7] for space propulsion present a high specific impulse and relatively low thrust, which make them especially useful for long-term accelerated motion. Thus, a lot of effort is made to develop and use numerical codes to provide support for the design and testing of new plasma engines.

The UPM Plasmalab at the Universidad Politécnica de Madrid (UPM) is currently developing ALPHIE (Alternative Low Power Hybrid Ion Engine), a new concept of thruster that aims to provide high specific impulses with low power consumption, suitable for middle-size satellites in LEO.

ALPHIE has a unique grid system in which a counter-flow of charged particles (ions and electrons) appear. Electrons are generated by a source outside the engine with a double purpose: to neutralize the ion beam and to create an inflow of electrons towards the ionization chamber. These electrons are accelerated by the grid system up to energies around 500 eV, meaning that the ionization of neutrals is done with high energy electrons.

In this work, a brief review of previous work in St. Petersburg Polytechnical University with the CST Studio Suite is done. Then, the operation of CST is compared with fpakc, a new particle-in-cell finite element code developed by the UPM. Further, the problem of the unique grid system in ALPHIE is presented. Finally, performance for a range of geometrical parameters of the grid system calculated with fpakc are presented and some hints on how to improve ALPHIE design for its next iteration are given.

29.2 CST vs fpakc

CST Studio Suite is a high-performance 3D electromagnetic analysis software package for the design, analysis and optimization of electromagnetic components and systems.

CST Particle Studio (PS) is a tool for fast and accurate analysis of charged particle dynamics in 3D electromagnetic fields. CST PS is fully integrated into CST design environment allowing for its intuitive modeling capabilities and powerful import interfaces.

The particle tracking solver can be used to compute the trajectories of charged particles in an electrostatic and magneto static field. The implemented gun iteration function allows calculating the self-consistent electrostatic field, which takes into account the reaction of the particle motion to the electrostatic potential distribution.

Arbitrary solid surfaces can be defined as particle sources. These surfaces emit particles based on a predefined emission pattern. The main task of the solver is to calculate particle trajectories, self-consistent electrostatic field, space charge distribution and particle current.

The tracking solver is primarily capable of tracking charged particles through electrical and magnetic static fields. The governing equations of the particle tracking are the momentum and position update:

- momentum update (analytic and discrete form of Newton-Lorentz equation):

$$\frac{d}{dt}(m\overline{v}) = q\left(\overline{E} + \overline{v} \times \overline{B}\right) \tag{29.1}$$

$$m^{n+1}\overline{v}^{n+1} = m^n\overline{v}^n + q\Delta t\left(\overline{E}^{n+1/2} + \overline{v}^{n+1/2} \times \overline{B}^{n+1/2}\right) \tag{29.2}$$

- position update (analytic and discrete form):

$$\frac{d\overline{r}}{dt} = \overline{v} \tag{29.3}$$

$$\overline{r}^{n+3/2} = \overline{r}^{n+1/2} + \Delta t\overline{v}^{n+1} \tag{29.4}$$

The CST particle studio offers multiple emission models for particles leaving a cathode. The particle tracking simply computes the trajectory of the particles through a precalculated electromagnetic field.

The gun-iteration consists of an iterative application of electrostatic solver and particle tracker. After calculating the electrostatic field the particles are emitted and tracked and their space charge is monitored. This space charge is used to modify the right hand side of the electrostatic solver (the charge-vector). Afterwards a new electrostatic field is calculated which incorporates some knowledge of the particles' flight path. This procedure is repeated until the relative difference of the emitted current and space charge meets a specific convergence criteria.

The disadvantage of such software is the lack of a module for calculating the mechanical parameters of ion and ion-plasma accelerators. Like the Monte Carlo method and some other similar software packages, the CST package cannot reveal the relationship between the mechanical forces acting in ion accelerators and internal microscopic processes. Using the CST method the analysis of combined ion and electron trajectories has been performed for the first time in the previous studies [3].

The Finite Element PArticle Code (fpakc) the UPM Plasmalab at the Technical University of Madrid (UPM) can be employed to perform time-dependent simulations that can complement some of the missing capabilities of CST. This code models species in plasma [8–11] (ions, electrons and neutrals) following the trajectories of macro-particles as they move and interact between them and the boundaries of the domain.

The fpakc uses macro-particles to simulate the dynamics of different plasma species [12–15] (mainly ions, electrons and neutrals). These macro-particles represent a large amount of real particles. In the evolution of these particles, external forces (as the electromagnetic field), interaction between particles (as collisions) and interaction with the boundaries of the domain are included.

This procedure differs from the iterative method of CST and it is possible to provide a higher accuracy, although more computational time may be required.

29.3   ALPHIE Grid System

The plasma engine ALPHIE (Alternative Low Power Hybrid Ion Engine) [16–18] is a new concept aimed to provide high thrust (~1mN) at low power consumption (<600 W) and low mass flux (<1 sccm). This new thruster has an unique grid system in which a counter-flow of ions and electrons. A sole source of electrons outside the ionization chamber is responsible of both, neutralizing the plasma beam and generating the inflow of electrons for ionization.

Thus, this grid system is not charge space limited and previous studies used for classical gridded ion engines are not applicable. Previous studies [5] have been carried out to analyze the counter-flow of charged species by means of a fully particle-in-cell code [19–23].

The fpakc code introduced above is now used to study relevant parameters from the ALPHIE grid system to improve its behavior. The distance between grids (dg) as well as their thickness (tg) is modified and their effect in the ion and electron distribution through the grids is studied. This sensitivity analysis will have a huge impact in the next iteration of ALPHIE, currently in design phase (Fig. 29.1).

A single hole of the grid system is studied in the axial-symmetric domain presented in Fig. 29.2. The Upstream region refers to the ionization chamber while Downstream is the vacuum outside the engine from which the ionizing electrons come.

In the current configuration of ALPHIE, $dg = 2.5$ mm and $tg = 0.4$ mm. This case will be used as a reference case to study the impact of modifying these two geometric parameters.

Electrons are introduced from the downstream boundary simulating the electrons coming from the external cathode. At the Upstream condition, ions and electrons coming from the ionization process are introduced. Grids are set to -150 and -600 V respect to the Upstream boundary condition which is keep at 0 V. Particles are simulated until they reach a quasi-steady situation

29.4   Results

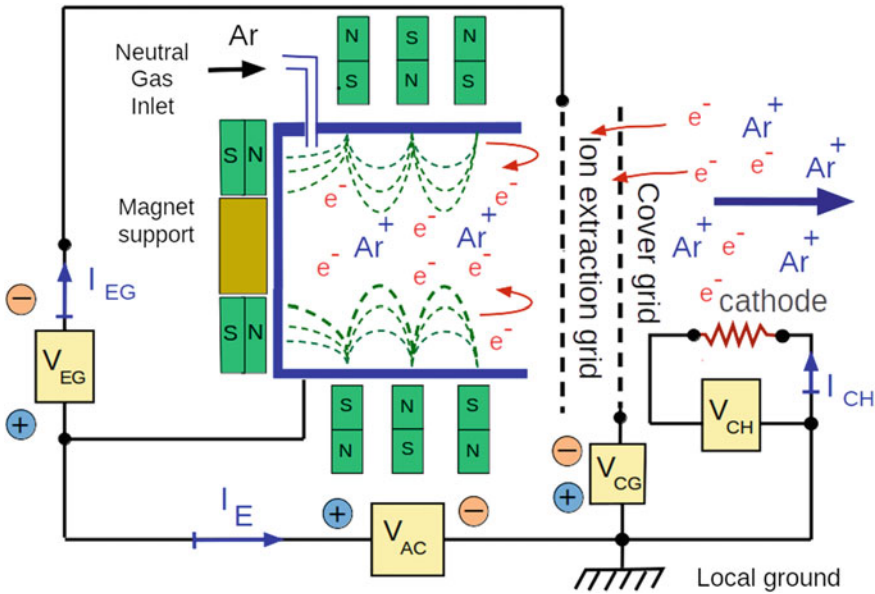This section presents the main results obtained from the CST and fpakc simulations.

Fig. 29.1 ALPHIE's electrical scheme. The external cathode has the double function of ionizing and neutralizing the plasma beam [4]
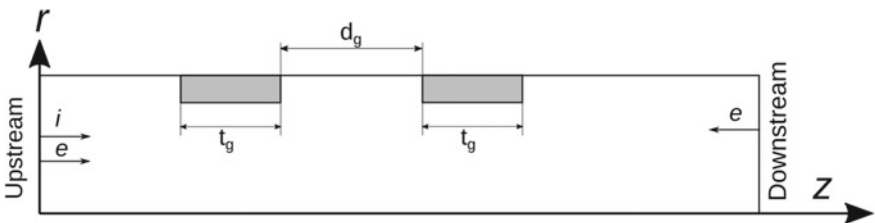

Fig. 29.2 Representation of a one-hole of the ALPHIE grid system simulated in this work. The relevant parameters for the sensitivity study are the distance between grids d_g and the thickness of the grids t_g

Initially, the CST simulation was used to obtain a similar 3D model of the device and perform the trajectory analysis. Figure 29.3 shows the acceleration performed by two grids system. The top grid is shown in the figure in the form of a solid disk.

Then, the capability of fpakc to advance different species with different time steps will be tested. This allows to push the heavy ions only at certain number of steps, reducing the computational requirements for each case.

Figure 29.4 shows two simulations of the same case, one in which particles evolve with the same time step (τ) and another one in which the time step for ions is 100 times larger.
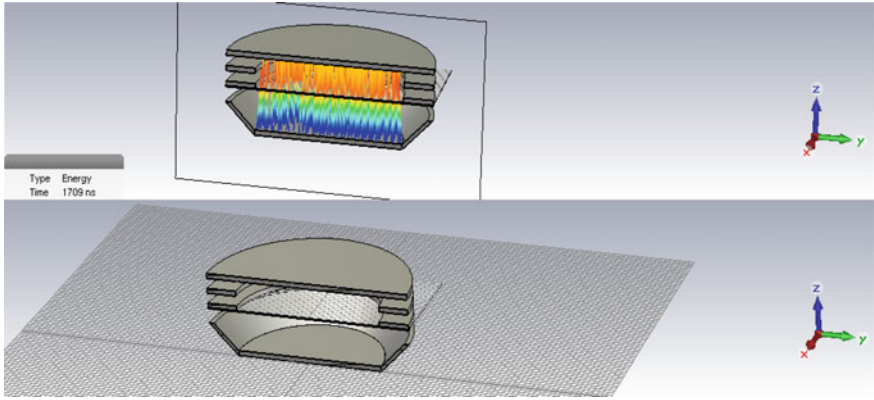
Fig. 29.3 3D 2-grids device model and particle trajectories in CST studio
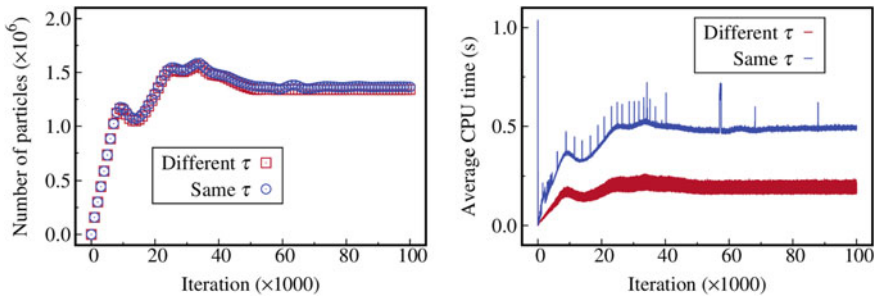

Fig. 29.4 Comparison between using the same time step or a larger time step for ions than electrons. Left: total number of macro-particles in the domain. Right: average time per iteration

The number of particles present in the numerical domain for both cases is exactly the same, the average computation time per iteration is reduced to almost half is different time steps are used per each species.

Moreover, currents for ions and electrons are similarly enough, as Fig. 29.5 show. This results in the capability to use a larger time step for heavy species without losing too much accuracy but gaining an incredible speed up, important for the sensibility analysis performed next.

Now, the results of varying the two geometric parameters of ALPHIE grid system (distance between grids and thickness of the grids) are presented now. Firstly, the results at the exhaust plane of the ion density and axial velocity are presented in Figs. 29.6 and 29.7. Simulations presented in Fig. 29.6 indicate there is no significant improvement in decreasing the grid thickness but increasing the grid thickness as ion density in the exhaust plane is greatly reduced near the axis. This means that grid thickness is only constrained by design limitations.

On the other hand, the grid system seems to be very dependent on the distance between grids. Figure 29.7 show that a variation of just 0.5 mm in the distance
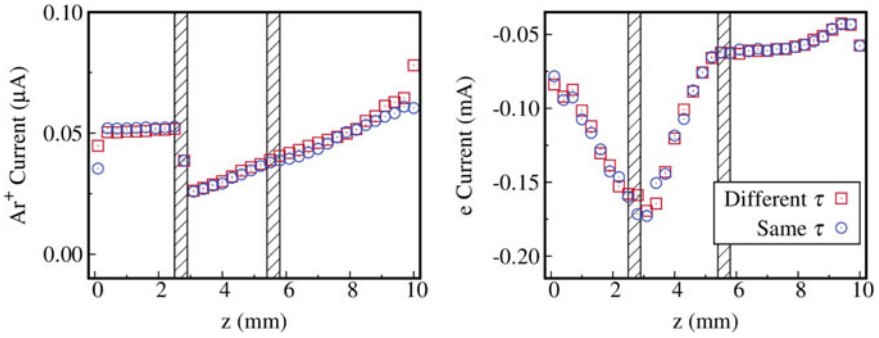
Fig. 29.5 Axial current for ions (left) and electrons (right) for cases with different and the same time steps. Results are indistinguishable
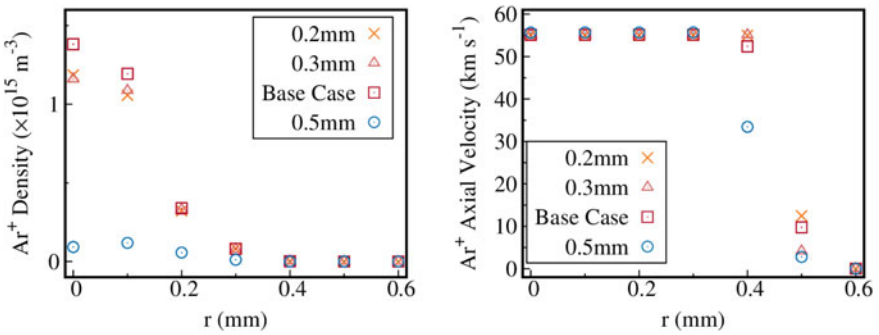

Fig. 29.6 Ion density (left) and axial velocity (right) for variations in the grid thickness (t_g). Base Case refers to $t_g = 0.4$ mm. No significant effect arises except for very thick grids
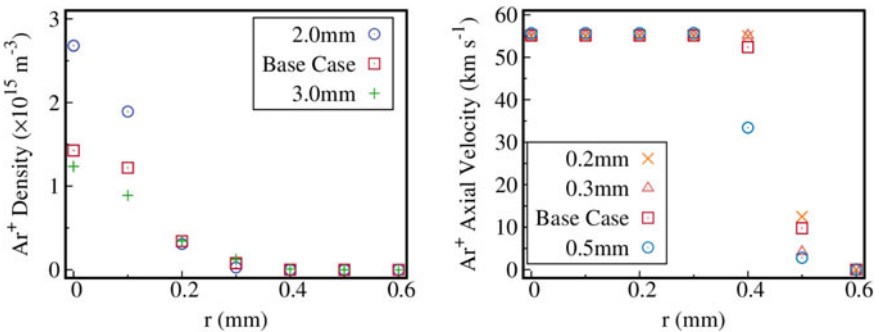

Fig. 29.7 Ion density (left) and axial velocity (right) for variations in the distance between grids (d_g). Base Case refers to $d_g = 2.5$ mm. Huge improvement of density near the axis is found for $dg = 2.0$ mm
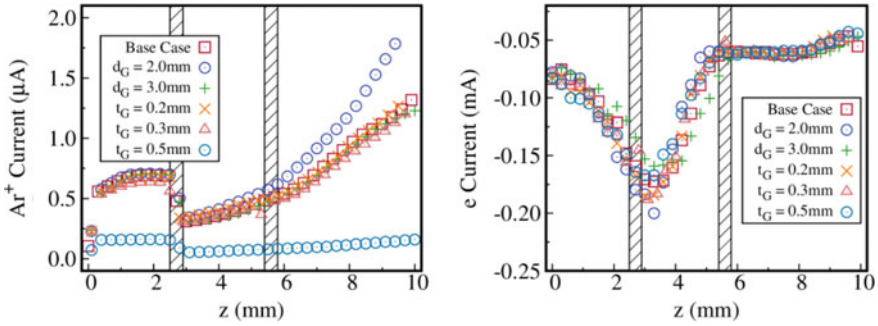
Fig. 29.8 Current of ions (left) and electrons (right) through the ALPHIE's grid system. The main improvement respect to the Base Case comes from decreasing the separation between grids

between grids modifies the ion density obtained. Specially, if distance between grids is reduced, ion density almost doubles. This behavior is similar to classical one-species ion extraction grids, in which the separation of grids highly influences the output current. Moreover, reducing the distance between grids results in a more collimated beam, as density increases near the axis and reduces far from it.

An important result of these simulations is that axial velocity of ions seems unaltered by the grid system. This could be the result of the counter-flow of ions and electrons through the grid system dominating the total acceleration of ions while the grid geometry affects its distribution and current.

The current of ions and electrons through the grid system is analyzed now. Figure 29.8 pictures the total current for ions and electrons along the axial coordinate. As previous analysis indicates, reducing the separation between grids has a huge impact in the extracted ion current.

However, electrons, due to their high speed and low mass, are unaffected by the modifications to the grid geometry. This is a positive note on ALPHIE grid system as electrons are fundamental for the ionization of the neutral gas, meaning that the geometry can be optimized for increasing the ion current without perturbing the influx of electrons required for the operation of ALPHIE. However, this also means that there is no means to improve electron inflow by geometrical factors and that the only option is to increase the emitted electrons by the external source.

29.5 Conclusions and Future Work

The newly developed code fpakc has proven to be an excellent tool to perform a sensibility analysis to improve the grid system of the ALPHIE plasma thruster, currently in development. The capability of the code to use different time steps for each species has been tested not finding any significant differences but saving a

reasonable amount of computational resources. The different approaches of fpakc and a previously used suite for electromagnetic devices have been analyzed.

Moreover, a sensitivity analysis of some geometrical parameters has been performed in order to improve the grid system extraction of ions. The main modifications respect to the current configuration of ALPHIE suggested by the simulations is the reduction in the distance between grids and to not exceed 0.5 mm of grid thickness. These changes will be taken into account in the next iteration design of ALPHIE.

In comparison with fpakc, CST can provide the trajectory analysis in the common tracking area of ion and electrons faster and simpler. However, the fpakc performs a more detailed and more precise modeling considering different physical processes. Combining both of these techniques seems to be promising in the field of ion thruster development.

References

1. S. Mazouffre, Electric propulsion for satellites and spacecraft: established technologies and novel approaches. Plasma Sour. Sci. Technol. **25**, 033002 (2016)
2. D.M. Goebel, I. Katz, *Fundamentals of Electric Propulsion Ion and Hall Thrusters.* (John Wiley & Sons, Hoboken, New Jersey, USA Ch. 1, 6 and 7, 1, 2008), pp. 243–389
3. D. Dyubo, OYu. Tsybin, A computer model of the contact ionization ion accelerator for the electrically powered spacecraft propulsion. St. Petersburg State Polytech. Univer. J. Phys. Math. **13**(1), 78–91 (2020)
4. D. Dyubo, J. González, O. Tsybin, L. Conde, Plasma thrusters for in-space propulsion; new trends and physical limitations, in *International Youth Conference on Electronics, Telecommunications and Information Technologies. Springer Proceedings in Physics*, eds. E. Velichko, M. Vinnichenko, V. Kapralova, Y. Koucheryavy, vol. 255 (2021)
5. J. Gonzalez, L. Conde, Particle-in-cell simulations of the ion extraction and acceleration processes in the alternative low power hybrid ion engine (ALPHIE). Phys. Plasmas **26**(4), 043505 (2019)
6. R.G. Jahn, in *The Physics of Electric Propulsion*, vol. 12 (Dover Publications, Mineola, 2006)
7. E. Ahedo, Plasmas for space propulsion. Plasma Phys. Control. Fusion **53**(12), 124037 (2011)
8. E. Vekshina, V. Rozhansky, E. Kaveeva, I. Senichenkov, N. Khromov, Modeling of Globus-M connected double-null discharge. Plasma Phys. Control. Fusion **61**, 125009 (2019)
9. P. Pankratiev, Y. Barsukov, A. Vinogradov, V. Volynets, A. Kobelev, A.S. Smirnov, Selective SiN/SiO2 etching by SF6/H2/Ar/He plasma, in *AIP Conference Proceedings* (2019), p. 2179
10. O.A. Emelyanov, E.G. Feklistov, N.V. Smirnova, K.A. Kolbe, E.V. Zinoviev, M.S. Asadulaev, A.A. Popov, A.S. Shabunin, K.F. Osmanov, Corona discharge plasma application for in vitro modulation of fibroblast proliferation and wound healing, in *AIP Conference Proceedings* (2019), p. 2179
11. A.Y. Yashin, V.V. Bulanin, A.V. Petrov, E.A. Zhivulin, V.K. Gusev, G.S. Kurskiev, V.B. Minaev, M.I. Patrov, Y.V. Petrov, Study of turbulence in the Globus-M tokamak plasma during the transition to the ELM-free H-mode, in *AIP Conference Proceedings* (2019), p. 2179
12. V. Bulanin, E. Gusakov, V. Gusev, S. Heuraux, C. Lechte, V. Minaev, A. Petrov, Y. Petrov, A. Yashin, N. Sakharov, N. Teplova, G. Zadvitskiy, Simulation of Doppler backscattering off filaments in the Globus-M spherical tokamak, in *46th EPS Conference on Plasma Physics, EPS* (2019)
13. A.S. Kukushkin, V.Y. Sergeev, B.V. Kuteev, Analysis of pumping conditions in DEMO-FNS, in *46th EPS Conference on Plasma Physics* (2019)

14. D.S. Sorokina, I.Y. Senichenkov, V.A. Rozhansky, E.O. Vekshina, Modeling of Globus-M2 spherical tokamak edge with nitrogen seeding. Phys. Plasmas **25**, 122514 (2018)
15. A.Y. Dnestrovskiy, A.S. Kukushkin, B.V. Kuteev, V.Y. Sergeev, Integrated modelling of core and divertor plasmas for the DEMO Fusion Neutron Source hybrid facility. Nucl. Fusion **59**, 096053 (2019)
16. L. Conde, J.L. Domenech-Garret, J.M. Donoso, J. Damba, S.P. Tierno, E. Alamillo-Gamboa, M.A. Castillo, Supersonic plasma beams with controlled speed generated by the alternative low power hybrid ion engine (ALPHIE) for space propulsion. Phys. Plasmas **24**(12), 123514 (2017)
17. J. Damba, P. Argente, P.E. Maldonado, A. Cervone, J.L. Domenech-Garret, L. Conde, J. Phys.: Multiprobe characterization of plasma flows for space propulsion Journal of Physics: Conference Series, Vol. 958; 12th International Workshop on Electric Probes in Magnetized Plasmas 2017 (IWEP 2017), Naklo, Slovenia, 4–7 Sept 2017
18. L. Conde, J.L. Domenech-Garret, J.M. Donoso, R.E. Del, M.A. Castillo, Plasma accelerator with modulated thrust. U.S. Patent 10,172,227 B2 (2019)
19. S. Cho, H. Watanabe, K. Kubota, S. Iihara, K. Fuchigami, K. Uematsu, I. Funaki, Study of electron transport in a Hall thruster by axial–radial fully kinetic particle simulation. Phys. Plasmas **22**(10), 103523 (2015)
20. S.H. Jenab, F. Spanier, Phys. Rev. E **95**, 053201 (2017)
21. V. Kolobov, R. Arslanbekov, J. Phys.: Conf. Ser. **719**, 012020 (2016)
22. D. Tskhakaya, K. Matyash, R. Schneider, F. Taccogna, The Particle-In-Cell method. Contrib. Plasma Phys. **47**(8–9), 563–594 (2007)
23. L. Liu, W. Zou, H. Wang, F. Guo, D. Liu, PIC simulation of the anode plasma in a high-power hollow cathode diode. Phys. Plasmas **25**(2), 022502 (2018)

Chapter 30
Design and Optimization of a Shunt RF MEMS Switch with a Hybrid Contact Type

Alexey Tkachenko 📧, **Igor Lysenko** 📧, **Mark Denisenko** 📧, **and Olga Ezhova** 📧

Abstract This paper presents the results of the design, optimization and simulation of a radio-frequency (RF) switch made using microelectromechanical systems (MEMS) technology. The device is a capacitive shunt switch with a hybrid contact type and a high capacitance ratio and a small air gap. To increase the capacitance ratio in the developed design of the RF MEMS switch, a fixed capacitor with metal–insulator-metal (MIM) plates is used. As the material of the insulator layer, a material with a high permittivity—titanium oxide—is used. To reduce the value of the activation voltage and increase the speed of the switch in the presented design of the RF MEMS switch, a zig-zag type of elastic suspension element is used. According to the results of the simulation of the optimized design of the RF MEMS switch, the activation voltage is no more than 3.5 V with a closing time of no more than 6.5 us. In addition, the RF MEMS switch presented by the results of electromagnetic modeling is characterized by low insertion loss -0.06 dB @ 3.6 GHz in the open position and a high isolation value -43.5 dB @ 3.6 GHz in the closed position. The effective frequency spectrum of the presented RF MEMS switch with a hybrid contact type is the S-band, which includes various types of ground and satellite radio communications.

30.1 Introduction

In the last decade, devices manufactured using microelectromechanical systems (MEMS) technology have undergone tremendous development in various fields of information and communication technologies. In particular, recent trends in the interaction of integrated circuit components (ICs) with the external environment have

A. Tkachenko (📧) · I. Lysenko · M. Denisenko · O. Ezhova
Southern Federal University, Taganrog, Rostov-on-Don Region 347922, Russia

led to the development of innovative devices and technologies that can combine electrical properties with several different physical domains. In particular, MEMS devices using a miniature movable structure, the movement or position of which can be controlled by means of an electrostatic, thermal, magnetic, liquid and electromagnetic activation mechanism. Reliability, compactness and a high degree of integration were among the key factors to achieve successful products in the technology market, such as ST Microelectronics accelerometers for car airbags and Nintendo Wii controllers, Apple iPhone gyroscopes, inkjet devices, Texas Instruments digital micro-mirrors for projectors.

One of the most promising areas of application of MEMS technology is associated with circuits and devices of ultra-high frequency (microwave) and radio-frequency (RF) communication. In recent years, the introduction and dissemination of new wireless communication standards, in particular the fifth generation of mobile radio (5G), has set new challenges in the development of hardware for transceivers. At the level of a single electronic device, the most important characteristics to consider are: low power consumption, high linearity and high throughput, which should characterize each component to ensure high performance of wireless radio systems and 5G mobile networks. At the system level, highly technological and reconfigurable circuits and devices allow modern wireless radio systems to be multiplatform, that is to integrate all the most complex communication standards used (such as UMTS, GSM, WiFi, Bluetooth, WiMAX) without losing the compactness and quality of radio signal transmission. RF MEMS technology allows you to create devices and circuits that can meet the requirements and capabilities mentioned above.

Key RF subsystems in 5G RF transceivers include antennas, tunable filters, power amplifiers, and MIMO (multiple-input multiple-output). At the same time, it can be noted that in all these RF systems or subsystems of 5G, from a hardware point of view, the RF switch is one of the most fundamental and important components that is used to route signals along RF signal transmission paths with a high degree of efficiency; its RF characteristics, switching time, RF signal power and their reliability can directly affect the corresponding properties and performance of 5G applications. At the same time, in accordance with the concept presented in [1] the key RF systems and subsystems of 5G transceivers can be replaced by corresponding RF MEMS devices.

Currently, among all classes of switching RF devices, PIN-diode switches are currently the most popular. At frequencies up to 10 GHz, they have virtually no equal in speed, simplicity of circuit solutions and cheapness. However, starting as early as 8 GHz, PIN-diode switches are not able to provide high isolation in the closed switch position, even when cascading. This is due to the increasing influence of the barrier capacitance, which is on the order of 0.2–0.8 pF for commercially available PIN diodes. In addition, PIN diodes cause noticeably greater losses in the open position and at lower frequencies. This problem is partially solved by the use of field-effect transistors, but at frequencies up to 40 GHz, coaxial switches are mainly used, which have good RF parameters, but are extremely expensive. At frequencies up to 1 GHz, high-frequency relays are sometimes used, but their disadvantages include low switching speed, at least 5 ms, and high-power consumption, as well
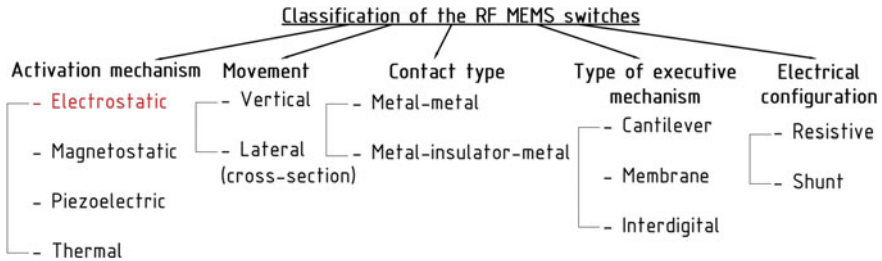
Fig. 30.1 Classification of RF MEMS switches

as high weight and size characteristics. RF MEMS switches can effectively replace PIN-diode analogs at frequencies up to 10 GHz and compete with transistor and coaxial switches in the range up to 40 GHz or more [2, 3].

Figure 30.1 shows the classification of RF MEMS switches according to the activation mechanism, the movement of the suspended parts, the type of contact, the type of mechanism used, and the electrical configuration.

Electrostatic RF MEMS switches are the most common among micromechanical switches today. This is due to the almost zero energy consumption in the switched-on state, the small size of the element, the compatibility of the device manufacturing process with the technological processes of silicon technology and AIIIBV-based technology, the relative ease of manufacture and the short switching time. Electrostatic RF MEMS switches are divided into two types—cantilever and membrane switches with metal–metal and metal-dielectric-metal contacts, respectively. Resistive RF MEMS switches have some advantages over capacitive RF MEMS switches because they are characterized by a lower value of the tripping voltage and a higher switching speed. On the other hand, capacitive RF MEMS switches are characterized by greater reliability and are subject to fewer failure mechanisms. In addition, for the design of RF MEMS switches of medium and low power, switches with a capacitive contact type have an important advantage—the ability to design switch designs with a low value of the actuation voltage.

In the last decade, a number of studies have been conducted on methods for reducing the value of the tripping voltage and increasing the switching speed of RF MEMS switches. The switching speed can be increased by using materials that reduce the effective mass of the moving parts of the structure [4], while another study suggests that the switching speed can be increased either by using materials that allow achieving a low stiffness coefficient of the elastic suspensions of the structure [5] or by using elastic suspension elements with a low stiffness coefficient, having a serpentine or zig-zag shape [6].

Currently, capacitive RF MEMS switches with metal-dielectric-metal contact are the most common in the technology in RF MEMS devices. In capacitive RF MEMS switches, the movable electrode of the structure is a metal membrane suspended from the anchor areas by elastic suspensions. The dielectric layer is applied to the signal line of the coplanar RF transmission line. Thus, a capacitor with metal-dielectric

layer-metal plates is formed. The main problem of this approach is the imperfect roughness of the contacting layers, which leads to a decrease in the capacitance ratio of this type of RF MEMS switches. A decrease in the capacitance ratio, in turn, leads to a shift in the resonant frequency from the required one, and a decrease of the isolation value in the close-state of the RF MEMS switch.

To date, there are a number of studies that have been conducted to achieve a high value of the capacitance ratio of capacitive RF MEMS switches and a low value of the actuation voltage. In [7], a design of a capacitive RF MEMS switch using a ceramic dielectric layer with a high permittivity is proposed. In [8], a high capacitance ratio was achieved by using a curved design of the movable electrode of a capacitive RF MEMS switch. Another method used to achieve a high value of the capacitance coefficient is to increase the air gap between the movable electrode-membrane and the dielectric layer [9, 10]. However, there are some obvious disadvantages of these methods, which are that the problem of the charge of the dielectric layer becomes more significant the smaller the thickness of the applied dielectric layer, and the electromechanical parameters of the switch change when the air gap changes.

Meanwhile, a number of studies have proposed some approaches for obtaining a high value of the capacitance ratio, which consist in the use of dielectric materials with high permittivity. Such dielectric materials include: HfO_2 [11], STO [7, 12], Ta_2O_5 [13], (Ba, Sr) TiO_3 [14], PZT [15] and metal oxide dielectrics with high dielectric characteristics. As a result, the value of the capacitance ratio of shunt RF MEMS switches is more than 100 [14, 15]. However, the value of the capacitance ratio is limited by the minimum thickness of the dielectric layer, the maximum value of the dielectric constant, and the maximum value of the air gap between the movable switch electrode and the RF transmission line. In this regard, the methods used in [8–10] are not the most appropriate.

To overcome the described disadvantages, an additional fixed metal–dielectric layer–metal (MIM) capacitor on the substrate is developed and a switch is used to turn it on or off from the circuit. This results in a capacitance ratio that is independent of the roughness of the lower electrode, and is therefore ideal for dielectrics with high permittivity and roughness. This method is also applicable to the design of RF MEMS switches with large capacitances in the closed position based on nitride dielectrics. The disadvantage of this method of increasing the capacitance ratio is the high geometric dimensions of the movable switch electrode, as well as the additional contact resistance between the metal of the movable membrane electrode and the upper metal layer of the MIM capacitor.

The object of this work is to develop a design of an RF MEMS switch with a high value of the capacitance ratio, a small value of the air gap using a combined approach. This approach of increasing the value of the capacitance ratio consists in using the design of a floating metal movable electrode without restrictions of the minimum thickness of the dielectric layer and the minimum size of the air gap, as well as using the material of the dielectric layer with a high permittivity. To reduce the value of the actuation voltage and increase the speed of the switch in the presented design of the RF MEMS switch, an inhomogeneous zig-zag elastic suspension is used. In

addition, the developed design of the RF MEMS switch should be characterized by a small form factor for use in RF devices and systems of 5G mobile networks.

30.2   Description of the Approach

A typical design of a shunt RF MEMS switch contains a CPW, a movable metal membrane, a thin dielectric layer, and fixed control electrodes, as shown in Fig. 30.2. The metal membrane is suspended above the CPW with an air gap of g_0. When a constant actuation voltage is applied to the fixed down actuation electrodes relative to the metal membrane, the electrostatic force attracts the metal membrane down to the dielectric layer, and as a result, the RF signal is shunted to the ground lines of the CPW [16].

A simple and effective approach to obtaining a high value of the capacitance ratio of a shunt RF MEMS switch is to connect an additional fixed capacitor with metal–insulator-metal (MIM) plates to the main shunt capacitor, as shown in Fig. 30.3.

In this design, an additional fixed MIM capacitor is connected to a shunt capacitor with metal-air-metal (MAM) plates formed by the upper metal layer of the MIM capacitor, a movable metal membrane and an air space between them. The MIM capacitor is connected to the MAM capacitor in series, in the case when the movable metal membrane is in the upper position. In the case when the movable metal membrane is in the lower position, the MAM capacitor changes to the resistance in
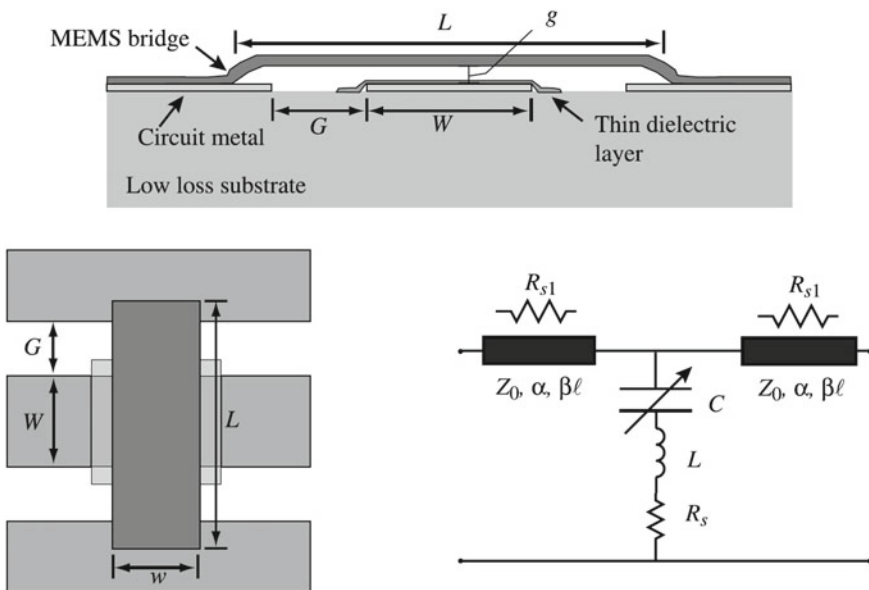

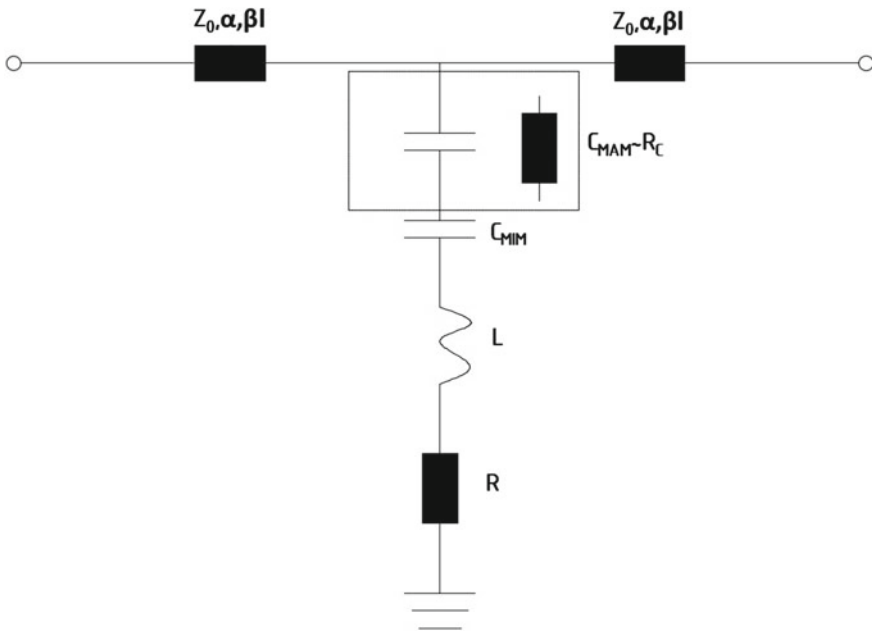**Fig. 30.2** Typical design of a capacitive shunt RF MEMS switch [16]

Fig. 30.3 Electrical circuit diagram of the developed RF MEMS switch

the electrical circuit. The characteristic resistance of a coplanar waveguide is equal Z_0. $Z_{bl}(\alpha, \beta l)$ denotes the characteristic resistance of the transmission line between the wave port and the edge of the movable metal membrane, α—the transmission attenuation constant, and β—the electrical length of the transmission line. In the case when the movable metal membrane is in the lower position, the MAM capacitor changes to the resistance in the electrical circuit.

The 3D topology of the proposed RF MEMS switch is shown in Fig. 30.4. The upper metal layer of the fixed MIM capacitor and the relatively narrow transmission line of the CPW are located under the movable electrode-membrane of the switch in order to obtain a low capacitance value in the upper position. Changing the size of the dielectric layer ensures that the wave resistance (50 Ω) is matched between the RF input of the CPW transmission line and the connected load.

30.3 Design Optimization and Flexibility

30.3.1 Material Selection

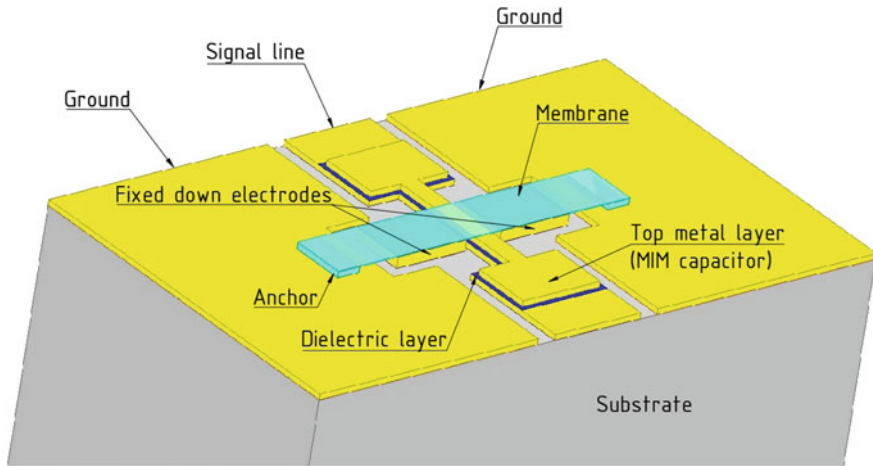Based on the material selection diagrams [17–19], it was observed that:

Fig. 30.4 3D topology of the proposed RF MEMS switch with a high capacitance factor

- as the material of the micromechanical elements of the RF MEMS switch design, gold and aluminum are the appropriate material to obtain the desired parameters that provide the best performance;
- as the substrate material of RF MEMS switches, the best candidate is sapphire and high-purity aluminum oxide Al_2O_3, which have the highest thermal conductivity, the lowest dielectric loss tangent, and the highest relative permittivity. At the same time, the roughness of the surface of sapphire is better, and the relative permittivity of these materials is the same. This material has exceptional electrophysical properties over a wide frequency range, high wear resistance and corrosion resistance at very high temperatures, as well as high mechanical strength;
- as the material of the dielectric layers used in the designs of RF MEMS shunt switches with capacitive contact type, the best candidate is Al_2O_3, followed by SiN, AlN, TiO_2. TiO_2 is the most suitable dielectric layer material used due to their low-temperature deposition process, which greatly improves the service life of switches with a reduced electric charge effect of the dielectric layer for RF applications.

30.3.2   Coplanar Waveguide

In this paper, a coplanar waveguide was chosen for the design of the RF MEMS switch design, since in this case it is possible to use thick dielectric substrates (for example, sapphire—0.5 mm). Since the grounding lines are located on the front side of the substrate, they can be used as control electrodes. Thus, the use of coplanar waveguides in the design of RF MEMS switch designs is technologically more accessible and makes it possible to manufacture structures of various configurations and complexity.

Table 30.1 Specification of coplanar waveguide

Component	Length (μm)	Width (μm)	Depth (μm)	Material
Substrate	650	400	500	Sapphire
CPW (G S G)	20	100	20	Copper

A coplanar waveguide for transmitting a high-frequency RF signal is formed from a symmetrically arranged signal line with a width W relative to the ground lines and an equal longitudinal gap G between them.

The dimensions of the developed coplanar waveguide are shown in Table 30.1.

Figure 30.5 shows the CPW for the developed RF MEMS switch.

If C is the capacitance per unit length of the line, and C_0 is the capacitance per unit length of the line in the absence of a dielectric layer, then the effective permittivity for the final CPW is determined by Eq. (30.1) [16]:

$$\varepsilon_{eff} = \frac{C}{C_0} = 1 + \frac{1}{2}(\varepsilon_r - 1)\frac{K(k)}{K(k')} \times \frac{K(k'_s)}{K(k_s)} \tag{30.1}$$

where K is the complete elliptic integral of the first kind and the values of k and k' are determined by the geometry of the line using the Eq. (30.2):

$$k = \frac{c}{b}\sqrt{\frac{b^2 - a^2}{c^2 - a^2}}$$

$$k' = \sqrt{1 - k^2} = \frac{a}{b}\sqrt{\frac{c^2 - b^2}{c^2 - a^2}} \tag{30.2}$$

The thickness of the substrate is 500 μm >> the thickness of the CPW line (6 um). Hence, for an infinitely thick substrate $k_s = k$. The equation for the effective permittivity is reduced to the Eq. (30.3):

$$\varepsilon_{eff} = \frac{1}{2}(\varepsilon_r + 1) \tag{30.3}$$

where ε_r is the relative permittivity of the substrate. Next, the phase velocity is given by the Eq. (30.4):

$$V_p h = \frac{c'}{\sqrt{\varepsilon_{eff}}} \tag{30.4}$$

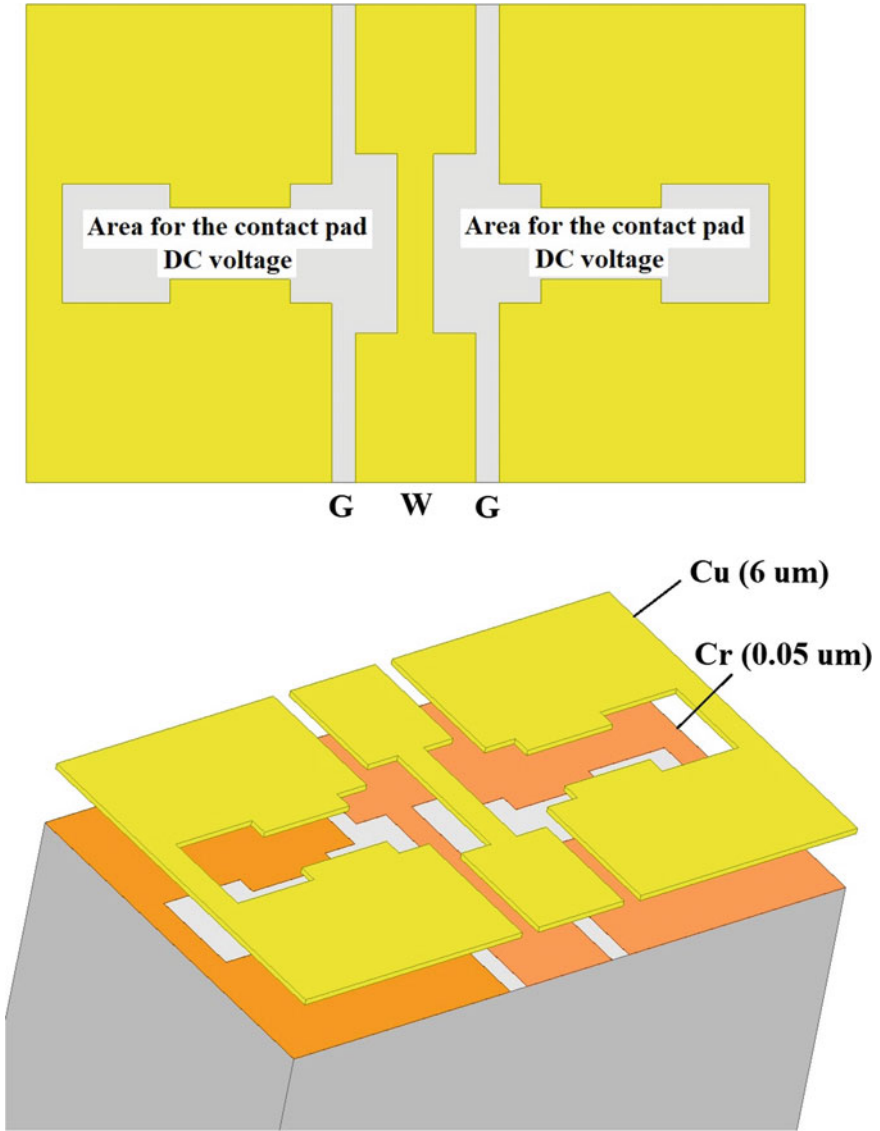where c' is the speed of light in free space. Thus, the characteristic resistance Z_0 is given by the Eq. (30.5):

Fig. 30.5 CPW for the developed RF MEMS switch

$$Z_0 = \frac{30\pi}{\sqrt{\varepsilon_{eff}}} \times \frac{K(k)}{K(k')} \qquad (30.5)$$

Shown in Fig. 30.5 the CPW switch designed for the RF MEMS design with a characteristic resistance of 50 Ω meets the requirement of matching the RF input and RF output.

30.3.3 Membrane Design

Figure 30.6 shows the optimized design of the movable electrode—membrane of the developed RF MEMS switch. Figure 30.6 shows the optimized design of the movable electrode-membrane of the developed RF MEMS switch. The passage of a high-frequency RF signal along the CPW signal line depends on the position of the switch membrane. A constant control bias voltage is applied to the fixed down electrodes relative to the membrane. A positive electric potential is applied to the fixed down actuation electrodes, and a negative electric potential is applied to the membrane.
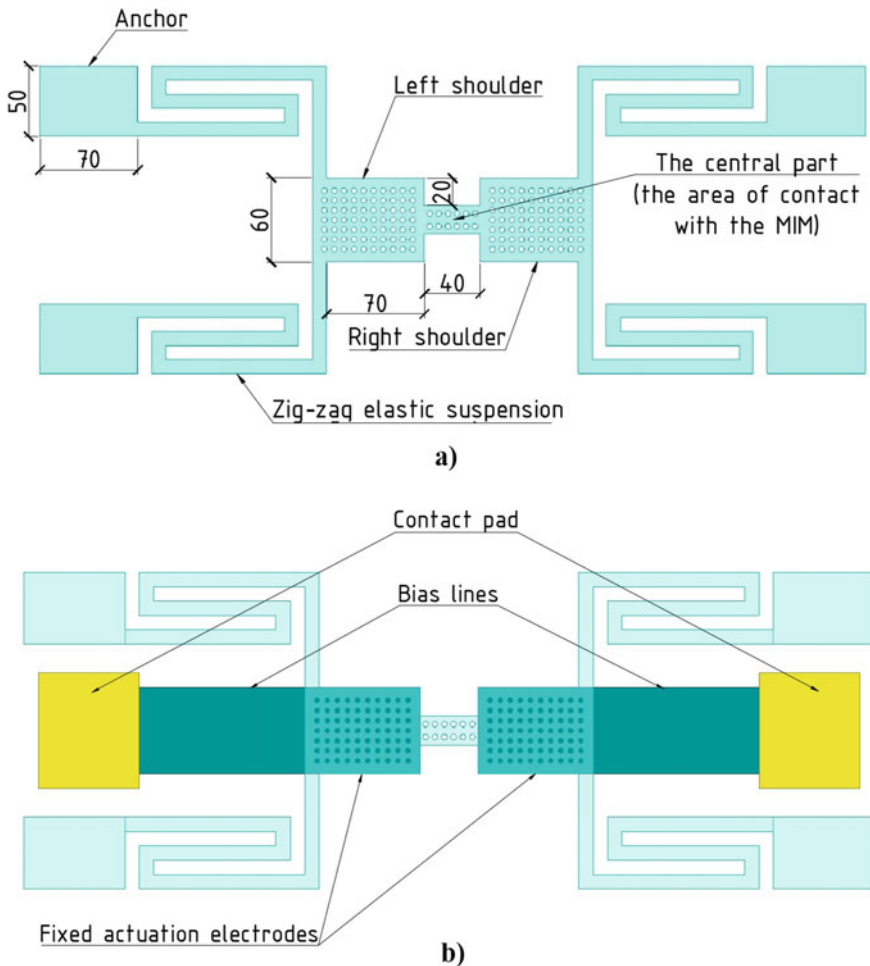
Fig. 30.6 Optimized membrane design of the developed RF MEMS switch: **a** construction of the membrane, anchor areas and zig-zag elastic suspensions; **b** construction of fixed down actuation electrodes, bias lines and contact pads

Table 30.2 Dimensions of the developed RF MEMS switch

Parameter	Dimensions (μm)	Material
Membrane	$70 \times 60 \times 1$	Aluminum
Hole diameter	4	
Anchors	$70 \times 50 \times 8$	Aluminum
Fixed down actuation electrode	$80 \times 60 \times 12$	Copper
Air gap, g_0	1	–
Dielectric film (MIM)	$300 \times 100 \times 0.2$	TiO_2
Dielectric film (on fixed down actuation electrode)	$80 \times 60 \times 0.1$	SiO_2

As a result, the membrane is attracted to the fixed down actuation electrodes by an electrostatic force. When the membrane comes into contact with the fixed down actuation electrodes, it also comes into contact with the upper metal layer of the additional fixed capacitor of the MIM structure at the point of their intersection.

The variable capacitor of the MAM structure is replaced in this case by an equivalent contact resistance. In this case, the value of the contact resistance negatively affects the amount of insertion loss. Therefore, the contact area in this case should be as small as possible.

Through holes are also provided in the membrane to reduce the stiffness coefficient of the movable electrode, reduce the value of the control bias voltage, and also to reduce the isothermal damping coefficient, which positively affects the switching speed. In addition, through holes in the membrane make it easier to remove the sacrificial layer of photoresist during the manufacturing process, as well as release some of the residual stresses in the membrane structure after the manufacturing process and reduce its mass, which leads to a higher mechanical resonant frequency [16].
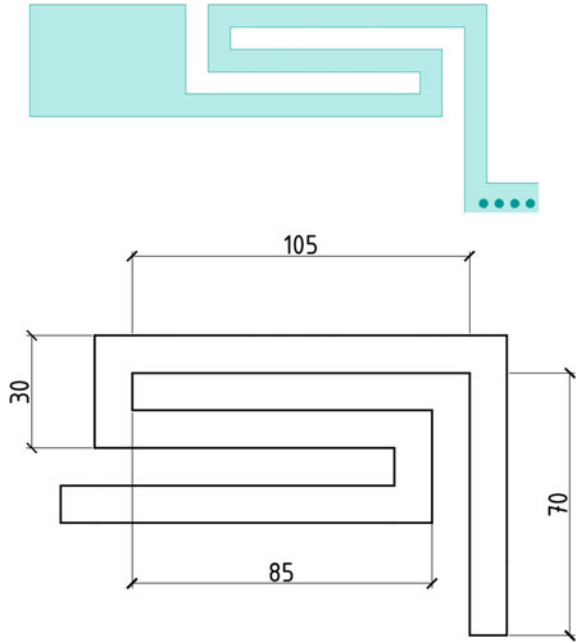
Table 30.2 shows the geometric dimensions of the main parts—the membrane, the fixed down actuation electrodes, the bias lines and the contact pads.

30.3.4 Zig-Zag Elastic Suspensions

Figure 30.7 schematically shows the developed zig-zag elastic suspension. Each elastic suspension consists of five successively connected elastic suspensions with two-fold symmetry, forming the shape of a meander, as well as one connecting elastic suspension having the shape of an elastic beam. Therefore, it is enough only to analyze one elastic suspension, and the resulting stiffness coefficient will be equal to one-fourth of the stiffness coefficient of one elastic suspension.

The bending symmetry is important for the translation of the boundary condition, so the guiding final boundary condition was determined at the ends of the elastic suspension. The displacement and rotation of the ends of the elastic suspension are assumed to be zero, except for the direction of the applied force.

The resulting stiffness coefficient of the presented zig-zag elastic suspension is determined by the Eq. (30.6):

$$K = k_s + k_b \tag{30.6}$$

where k_s is the stiffness coefficient of the elastic suspension in the form of a meander and k_b is the stiffness coefficient of the connecting elastic suspension in the form of a cantilever beam, respectively.

The effective stiffness coefficient of all four elastic suspensions is determined by the Eq. (30.7):

$$K_{eff} = \frac{K}{4} \tag{30.7}$$

The stiffness coefficient of an elastic suspension in the form of a meander is determined by the Eq. (30.8) [16]:

$$k_s = \frac{48}{l_a^2 \left(\frac{GJ}{EI_x} l_a + l_b \right) n^3}, n \gg \frac{3l_b}{\frac{GJ}{EI_x} l_a + l_b} \tag{30.8}$$

where E is the Young's modulus; G is the shift modulus defined by the following expression: $G = E/(2(1 + \upsilon))$; υ is the Poisson's ratio; J is the torsion constant,

defined by the following expression: $J = 0.413 I_p$; I_p is the polar moment of inertia, defined by the expression: $I_p = I_x + I_z$; I_x is the moment of inertia along the x axis, defined by the expression: $I_x = w t^3 / 12$; I_z is the moment of inertia along the z axis, defined by the expression: $I_z = t w^3 / 12$; n is the number of meanders.

The stiffness coefficient of an elastic suspension in the form of a beam is determined by the Eq. (30.9) [16]:

$$k_b = E w \left(\frac{t}{l} \right)^3 \tag{30.9}$$

30.4   Modeling

30.4.1   Electromechanical Modeling

In numerical modeling of continuum mechanics problems, the quality of the calculated finite element grid is an important parameter. The question of grid quality is of critical importance in contact problems, since, on the one hand, small elements reflect the behavior of the structure with higher accuracy when loads are applied to it, and on the other hand, such problems, due to their non-linearity, require significant computational resources. A geometric model of the movable electrode and zig-zag elastic suspensions of the developed RF MEMS switch design and a fragment of a finite element grid with a given characteristic cell size of 2 μm are shown in Fig. 30.8.
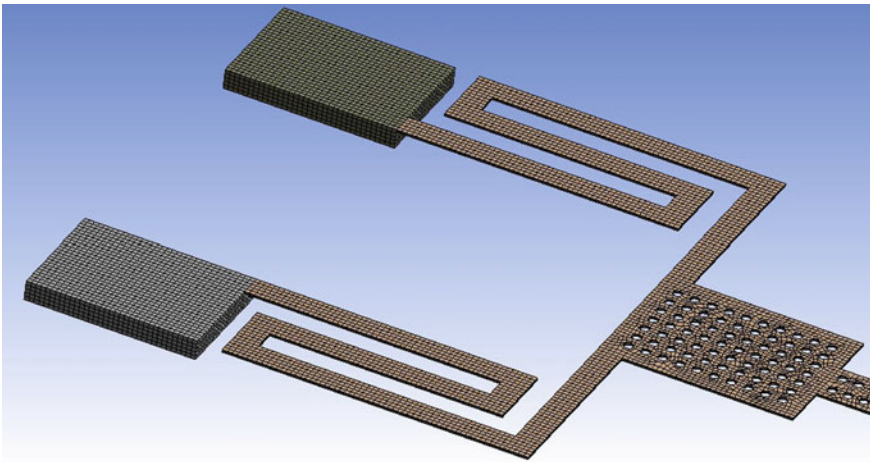
Fig. 30.8 Finite element model of the membrane and zig-zag elastic suspensions

The problems of numerical modeling that allow the occurrence of contacts are characterized by non-linearity, which greatly complicates their solution. The nonlinearity in this case is related to the dependence of internal forces on displacements. That is, when the movement of the nodes of the finite element grid reaches a certain value, forces appear that should prevent the further movement of the nodes. Thus, contact problems are characterized by an abrupt dependence of internal forces on the movements of nodes, a high degree of nonlinearity, and require significant computational resources to solve.

Figures 30.9, 30.10 and 30.11 illustrates the results of numerical simulation of the electromechanical parameters of the developed RF MEMS switch design.
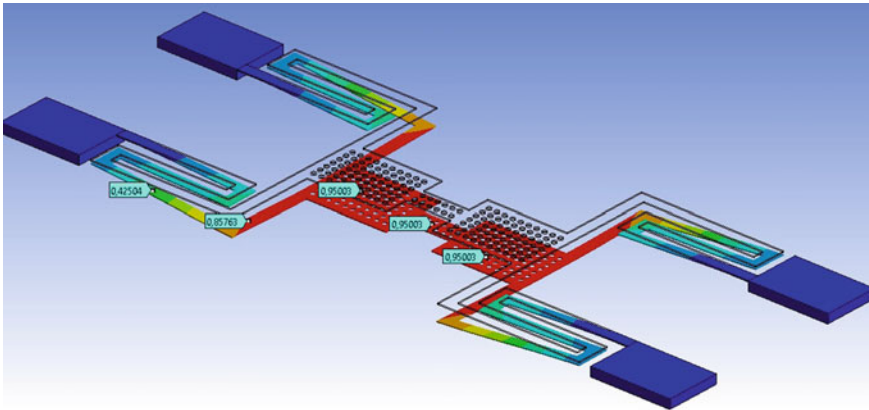
Fig. 30.9 Distribution of the displacement of the membrane and elastic suspensions during electrostatic activation (the value of the applied voltage $V_p = 3.5$ V)
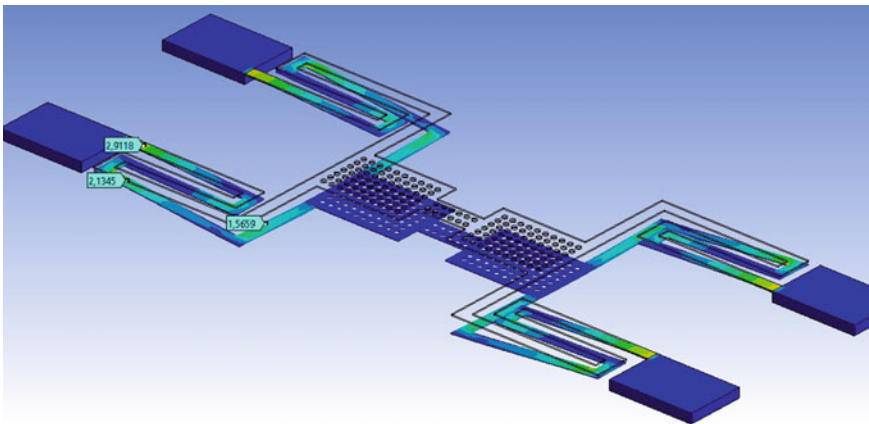
Fig. 30.10 Distribution of mechanical stresses of the membrane structure and elastic suspensions under electrostatic activation
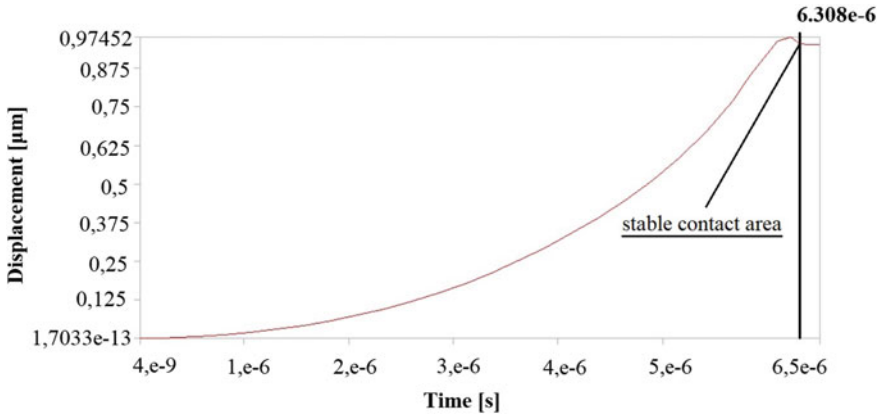
Fig. 30.11 Switching time of the developed RF MEMS switch design during electrostatic activation

When calculating the switching time, the value of the actuation voltage was taken from the expression $t_s = 1.4V_p$.

According to the results of the numerical simulation of the electromechanical model of the developed RF MEMS switch design, it follows that the value of the actuation voltage is 3.5 V, and the switching time to the down-state of the movable electrode is 6.3 us.

Also, according to the results of the simulation, it follows that the complete closure of the membrane of the developed RF MEMS switch occurs when a force of 0.175 uN is applied.

30.4.2 Electromagnetic Modeling

Figure 30.12 and Fig. 30.13 show the results of numerical simulation of the S-parameters (scattering parameters) in the ANSYS HFSS software package of the developed RF MEMS switch design in the open position (up-state of the membrane), namely, S_{11}—reflection loss, S_{12}—insertion loss, expressed in dB.

Figure 30.14 shows the results of numerical simulation of S-parameters, namely, S_{12}—isolation of the developed design of the RF-MEMS switch in the closed position (the down-state of the membrane).

Based on the results of the electromagnetic simulation, it can be concluded that the developed design of the RF MEMS switch is suitable for use in the S-frequency range. In the open state of the switch, the insertion loss does not exceed −0.06 dB, in the closed state, the isolation is −43.8 dB at the central resonant frequency of 3.6 GHz.
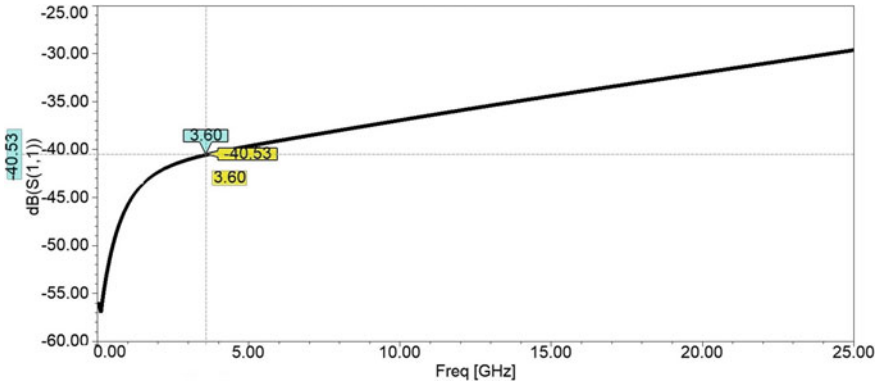
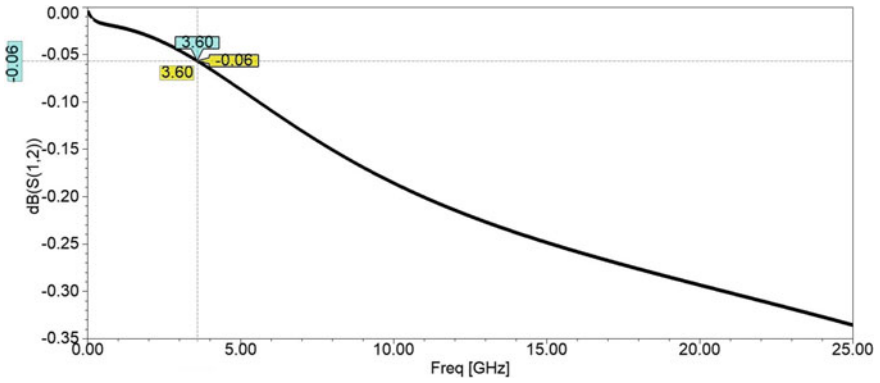**Fig. 30.12** Modeling the S-parameters (S_{11}) in the up-state of the membrane

Fig. 30.13 Modeling the S-parameters (S_{12}) in the up-state of the membrane
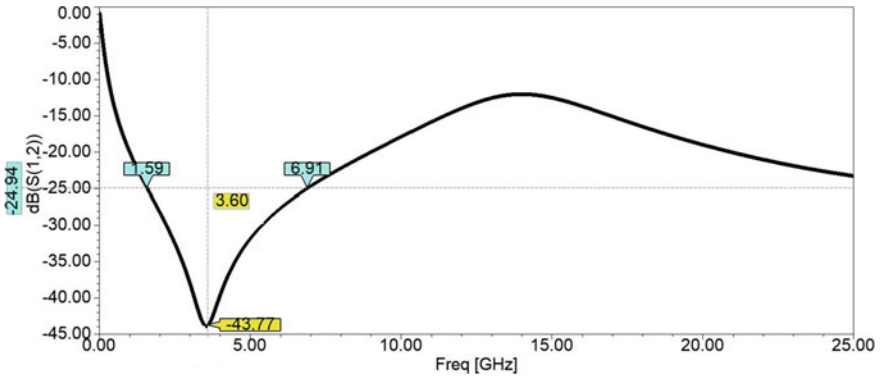

Fig. 30.14 Modeling the S-parameters (S_{12}) in the down-state of the membrane
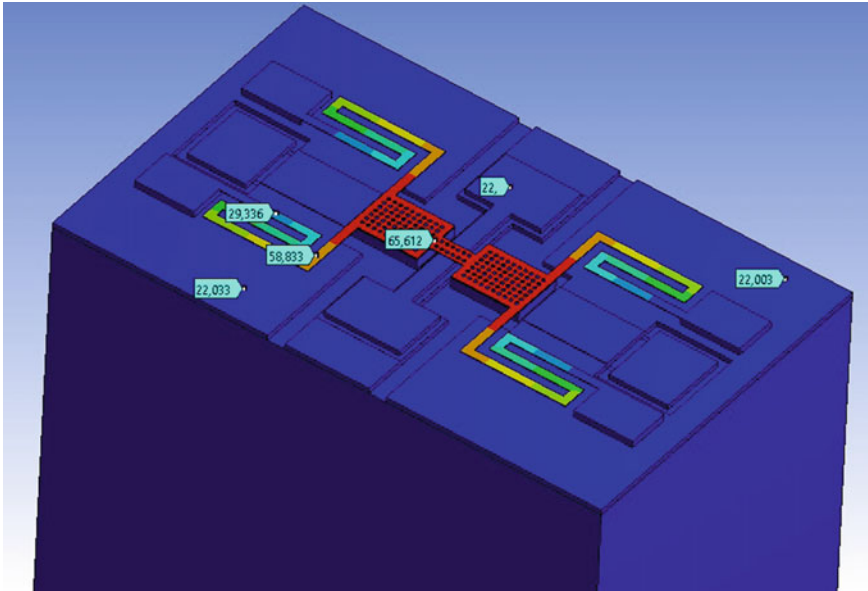
Fig. 30.15 Results of numerical simulation of the temperature distribution in the developed design of the RF MEMS switch in the open state

30.4.3   Thermal Modeling

The material of the coplanar waveguide is an adhesive layer of chromium with a thickness of 50 nm and a layer of copper with a thickness of 6 μm. Cooling is provided by convection from the surface (h = 20 pW/μm^2 ×°C). It was also assumed that the crystal is attached to the base with a constant temperature of 22 °C. As an example, the switched power of the 3.6 GHz RF signal is 1 W. The temperature distribution in this case for the developed RF MEMS switch design is shown in Fig. 30.15.

The thermal analysis shows that the maximum heating in the developed design of the RF MEMS switch occurs in the contact area-the area of the movable electrode. The movable electrode also acts as a heat sink from the contact area. The heating temperature in both cases is not critical and is not close to the softening temperature of the metal of the movable electrode, and therefore there is no sticking of the contacts.

30.5   Conclusion

This paper presents the developed design of an RF MEMS switch with a hybrid contact type and a high capacitance ratio. The proposed switch provides high RF performance with isolation of more than −43.5 dB at 3.6 GHz in the closed position and very low insertion loss of −0.06 dB at 3.6 GHz in the open position. In addition,

Table 30.3 Performance comparisons of the proposed RF MEMS switch with generic designs

References	[7]	[10]	[20]	[21]	[22]	[23]	This paper
Insertion loss (dB)	0.08	0.15	0.35	0.2	0.7	1.5	0.06
Isolation (dB)	42	40	37	38.5	35	20	43.5
Actuation voltage, (V)	8	4–6	~118	12	30	21	3.5
Air gap (μm)	2.5–3.5	~15	2	~2	3–5	2	1
Switching time (μs)	N/A	~20	N/A	~15	7	<10	6.3
Capacitance ratio	600	250	22	64.5	85.5	100	5701

the presented switch is characterized by a low activation voltage of 3.5 V, a switching time of 6.3 μs and small weight and size characteristics, which makes it suitable for use in devices and systems of 5G mobile networks, many wireless communication systems, space systems and satellites.

The performance comparisons of the proposed RF MEMS switch with a hybrid contact type and high capacitance ratio with generic designs are shown in Table 30.3.

The results obtained can be used as a foundation for the design of complex RF MEMS networks, such as multi-position attenuators (for example, 5–8 bits), the development of which would also reduce hardware redundancy in radio-frequency interface modules (RFFE) designed for mobile 5G networks. More signal shaping functions, such as attenuation and phase shift, can be combined in the same RF MEMS device, which makes such a technical solution even more attractive for 5G applications. In general, the development and development of an element component base based on passive RF MEMS elements, such as antenna switches, intermediate frequency filters, LC filters and resonators, will lead to the replacement of traditional semiconductor analogues, while increasing the performance of mobile 5G networks, while causing a reconsideration of the architecture of transceivers.

References

1. J. Bouchaud, B. Knoblich, RF MEMS switches deliver on early promise. Sens. Trans. J. **86**, 1802–1808 (2007)
2. A. Berezniak, A. Korotkov, Solid-state microwave switches: circuitry, manufacturing technologies and development trends. Review (Part 1). Radioelectronics Commun. Syst. **56**(4), 159–177 (2013)
3. A. Berezniak, A. Korotkov, Solid-state microwave switches: circuitry, manufacturing technologies and development trends. Review (Part 2). Radioelectron. Commun. Syst. **56**(5), 213–226 (2013)

4. K. Guha et al., A new analytical model for switching time of a perforated MEMS switch. Microsyst. Technolog. **26**(3), 3143–3152 (2018)
5. R. Kumar, Design and simulation of MEMS shunt capacitive switch for lower switching time. 3C Tecnología. Glosas de innovación aplicadas a la pyme. Spec. Issue Recent Trends Comput. Sci. Electron. 167–178 (2019)
6. A.K. Ravirala et al., Design and performance analysis of uniform meander structured RF MEMS capacitive shunt switch along with perforations. Microsyst. Technolog. **24**(2), 901–908 (2018)
7. J.Y. Park et al., Fully integrated micromachined capacitive switches for RF applications, in *IEEE MTT-S International Microwave Symposium Digest*, vol. 1 (IEEE, Boston, MA, USA, 2000), pp. 283–286
8. R. Al-Dahleh, R.R. Mansour, High-capacitance-ratio warped-beam capacitive MEMS switch designs. J. Microelectromech. Syst. **19**, 538–547 (2010)
9. C. Chang, P. Chang, Innovative micromachined microwave switch with very low insertion loss. Sens. Actuators, A **79**, 71–75 (2010)
10. C. Bozler et al., MEMS microswitch arrays for reconfigurable distributed microwave components, in *Proceedings of the IEEE Antennas and Propagation Society International Symposium* (IEEE, Salt Lake City, UT, USA, 2000), pp. 587–591
11. M. Angira et al., On the investigation of an interdigitated, high capacitance ratio shunt RF-MEMS switch for X-band applications, in *Proceedings of NSTI Nanotechnology*, vol. 2. (TechConnect Briefs, Washington DC, United States, 2013), pp. 189–192
12. A. Persano et al., Ta$_2$O$_5$ thin films for capacitive RF MEMS switches. J. Sens. **2010**, 1–5 (2010)
13. Y. Liu et al., High-isolation BST-MEMS switches, Microwave Symposium Digest, vol. 1. (IEEE, Seattle, WA, USA, 2002), pp. 227–230
14. A. Ziaei et al., Lifetime characterization of capacitive power RF MEMS switches, in *Proceedings of 35th European Microwave Conference*, vol. 3. (IEEE, Paris, France, 2005), pp. 227–230
15. G. Wang et al., Novel reliable RF capacitive MEMS switches with photodefinable metal–oxide dielectrics. J. Microelectromech. Syst. **16**(3), 550–555 (2007)
16. G.M. Rebeiz, *RF MEMS: Theory, Design, and Technology* (Wiley, New York, 2004)
17. I.E. Lysenko et al., Research of the microelectromechanical switch with different materials of metal membrane, in *Proceedings of SPIE 11022, International Conference on Micro- and Nano-Electronics 2018*, vol. 10226, (SPIE, Zvenigorod, Moscow Region, Russia 2019), pp. 1–12
18. I.E. Lysenko et al., The mechanical effects influencing on the design of RF MEMS Switches. Electronics **9**(2), 1–26 (2020)
19. I.E. Lysenko et al., Analytical approach in the development of RF MEMS switches electronics. Electronics **7**(12), 1–23 (2018)
20. K. Demirel et al., A new temperature-tolerant RF MEMS switch structure design and fabrication for Ka-Band application. J. Microelectromech. Syst. **25**, 60–68 (2016)
21. M. Fernandez-Bolanos et al.: Reliability of RF MEMS capacitive switches and distributed MEMS phase shifters using AlN dielectric, in *Proceedings of the IEEE 22nd International Conference on Micro Electro Mechanical Systems*, (IEEE, Sorrento, Italy 2009), pp. 638–641
22. C.L. Goldsmith et al., Performance of low-loss RF MEMS capacitive switches. IEEE Microwave and Guided Wave Lett **8**(8), 269–271 (1999)
23. P. Blondy et al., Dielectric less capacitive MEMS switches, in *Proceedings of the 2004 IEEE MTT-S International Microwave Symposium Digest* (IEEE, Fort Worth, TX, USA 2004), pp. 573–576

Chapter 31
Destruction of Unresectable Brain Tumors: Simulation of Thermal Spread and Tissue Damage During MRI-Guided Laser Ablation

Gesine Hentschel, Johannes Johansson, Christina Winkler, Birgit Glasmacher, and Karin Wårdell

Abstract Laser induced thermal therapy (LITT) emerged in recent years as a minimal invasive treatment method for otherwise oftentimes inoperable brain tumors, such as glioblastomas. During the thermal ablation process, the procedure carries the risk of destroying healthy brain tissue adjacent to the tumor. Limitations in the spatial distribution of the real-time monitoring MR thermography system currently allow only a rough representation of the damage zone during surgery. For this reason, improved pre-operative simulations of tissue heating and the resulting tissue damage could be valuable to optimize clinical treatment protocols while minimizing the risks of the procedure. In this study a method is presented, to simulate the ablation process of the LITT with the finite elements method (FEM) using the simulation software *COMSOL MULTIPHYSICS*. Thereby the temperature distribution is displayed and, based on this, the extent of the tissue damage during the process is simulated. Therefore, relevant parameters for the ablation process, such as optical and thermal properties, blood per-fusion, and the interface between healthy and tumor tissue were investigated and their influence on temperature distribution and extent of tissue damage was described.

Keywords LITT · Minimal invasive surgery · Laser in medicine · Finite elements method

31.1 Introduction

Glioblastomas are the most common and aggressive malignant brain tumors in adult humans. The Robert-Koch-Institute registered 6102 new cases in in 2015. More than

G. Hentschel (✉) · C. Winkler · B. Glasmacher
Institute for Multiphase Processes, Leibniz University Hannover, An der Universität 2, 30823 Garbsen, Germany

G. Hentschel · J. Johansson · K. Wårdell
Department for Biomedical Engineering, Linköping University, 581 85 Linköping, Sweden

90% of patients diagnosed with this grade of brain tumor die within five years after the initial diagnosis [1]. The treatment of gliomas has shown that the extent of resection performed during surgery correlates directly with the life expectancy of treated patients. A larger resection of low-grade gliomas can therefore lead to more favorable results and alter their malignant transformation. However, there are associated disadvantages. A larger resection often carries the risk of damage to adjacent tissue which may subsequently lead to neurological morbidity [2].

In recent years, Laser induced thermal therapy (LITT) has become an option for surgeons to address this problem. By destroying the tumor while minimizing damage to adjacent brain regions. The tumor is destroyed by hyperthermic coagulation due to photon diffusion and heat transfer within the tissue in a minimal invasive intervention [3, 4]. The first clinical application of LITT as a treatment method for brain tumors dates back to 1990 [3].

To apply this method safely, a precise knowledge of the temperature distribution and extent of tissue damage is essential. Therefore, the temperature distribution is monitored by means of real-time MR thermography. Currently, this is still a rather superficial method and it is limited in spatial resolution. This shortcoming and the associated risk of injuring adjacent healthy tissue structures can be addressed by performing simulations of the entire ablation process. They display the temperature distribution with a high spatial resolution and can be implemented in the necessary planning process before the operation. Clinical treatment protocols can be effectively optimized [3].

This requires adequate governing equations that can model the temperature dependence of tissue properties as realistic as possible [5]. However, most simulation models of LITT outlined in the literature are based on simplifications of the real problem, both in geometry and in the governing equations [4, 6–10]. Finite element simulations on brain tissue are only described in a few publications so far [6, 11, 12].

In this study, the temperature distribution and associated extent of tissue damage during the ablation process of a glioblastoma were simulated through dynamic changes in optical and thermal tissue properties. The results were compared with simulations in which these effects were neglected. Furthermore, to test the influence of the model geometry, the thermal spread of the process was modelled for a glioma surrounded by healthy brain tissue.

31.2   Materials and Methods

To investigate the influence of dynamic optical and thermal parameters, three simulation models were compared to each other. The basic FEM model, described the ablation process as time–dependent and with fixed optical and thermal parameters. The model was refined by first modelling the optical parameters and then the blood flow dynamically. Subsequently, the influence of these changes on the ablation process were investigated. Lastly, the model's geometry was adapted by

implementing healthy brain tissue surrounding the tumor into the model. All models were generated with *Comsol Multiphysics* (Version 5.5 Comsol AB, Sweden).

31.2.1 Geometry

To simulate the ablation process, a 3D model of the LITT catheter with the surrounding tumor tissue was developed in *Comsol Multiphysics*. The template for the catheter was the *Visualase* LITT system from the company Medtronic (USA).

The catheter has a total diameter of 1.65 mm and consists of a two-channel cannula.

According to Carpentier et al., a diode laser with a wavelength of 980 nm, a power of 10 W and a length-adjustable diffusor tip at the end of an optical fiber, is integrated in the *Visualase* system to deliver laser light to the tissue [13]. The optical fiber is positioned in the inner channel, which has a diameter of 0.778 mm. Medtronic uses polycarbonate (PC) as the outer covering material. The applied material and tissue data are summarized in Table 31.1.

During the entire ablation process, a sodium chloride solution (NaCl) flows through the outer channel, which is used to cool the catheter. To simplify the model, the flow properties of the NaCl solution inside the catheter were neglected during the simulation and replaced by a temperature boundary condition. It was assumed that the solution maintains a temperature of 20 °C during the entire process.

The tumor tissue, is represented in the model by a cylindrical volume body with a diameter of 20 mm and a height of 40 mm. The catheter penetrates a total of 22 mm into the tumor tissue, while the length of the light emitting area was set to 3 mm (Fig. 1a).

31.2.2 Mathematical Methods

The core of the simulation is a mathematical model which is composed of three sections: heat generation, heat transport and heat effects [18]. For all three models certain input parameters must be defined in the simulation. Their embedding and representation in *Comsol Multiphysics* are described below.

Heat generation depends on the laser properties of the optical fiber inside the catheter and on the optical tissue properties. Since no experimental data for the tissue parameters of the simulated glioblastoma were available in the context of this study, they were taken from various scientific studies instead (Table 31.1). The Light Distribution Equation describes the light distribution of the laser light emitted by the catheter within the tissue.

$$\nabla \cdot (-D\nabla\phi(r)) + \mu_a\phi(r) = S(r) \tag{31.1}$$

Table 31.1 Input parameters used in this study to simulate the LITT

Property	Symbol	Value	Unit	References
Density (polycarbonate)	ρ_{pc}	1.2	kg m^{-3}	[14]
Heat capacity at constant pressure (polycarbonate)	c_{pc}	1.2×10^{-3}	J kg^{-1} K^{-1}	[14]
Thermal conductivity (polycarbonate)	k_{pc}	0.22	W K^{-1} m^{-1}	[14]
Absorption coefficient glioblastoma	μ_a	101	m^{-1}	[13]
Reduced scattering coefficient	μ'_s	2110	m^{-1}	[13]
Absorption coefficient coagulated tissue	$\mu_{a,\,coagulated}$	189	m^{-1}	[15]
Reduced scattering coefficient coagulated tissue	$\mu'_{s,\,coagulated}$	2632.6	m^{-1}	[15]
Absorption coefficient healthy white brain matter	$\mu_{a,w}$	122	m^{-1}	[15]
Reduced scattering coefficient healthy white brain matter	$\mu'_{s,w}$	3698	m^{-1}	[15]
Mass density of tumor tissue	ρ	1020	kg m^{-3}	[6]
Mass density of white matter tissue	ρ_w	1041	kg m^{-3}	[16]
Heat capacity at a constant pressure tumor tissue	c_p	3500	J kg^{-1} K^{-1}	[6]
Heat capacity at a constant pressure white brain matter	$c_{p,w}$	3583	J kg^{-1} K^{-1}	[16]
Thermal conductivity tumor tissue	k	0.6	W m^{-1} K^{-1}	[6]
Thermal conductivity white brain matter	k_w	0.48	W m^{-1} K^{-1}	[16]
Arterial temperature of blood	K	310.15	K	[17]
Specific heat capacity of blood	c_b	4180	J kg-1 K-1	[17]
Blood perfusion rate	ω_b	0.0046	s^{-1}	[17]
Mass density of blood	ρ_b	1000	kg m^{-3}	[17]
Arrhenius equation activation energy	ΔE	2.577×10^5	J mol^{-1}	[16]
Arrhenius equation frequency factor	A	7.39×10^{39}	s^{-1}	[16]

where, $\phi(r)$ describes the light fluence rate, D the diffusion coefficient, μ_a the absorption coefficient of the tumor tissue and $S(r)$ the source term. The source term is defined as the absorbed energy per unit volume and is expressed for the light distribution equation as

$$S(r) = -\mu_a \cdot \phi(r). \qquad (31.2)$$

The model simulates the light transfer where the light is only dispersed within the tumor tissue domain. At the diffuse light emitting area of the catheter, a Flux/Source boundary condition is introduced to adjust the boundary flux between catheter and
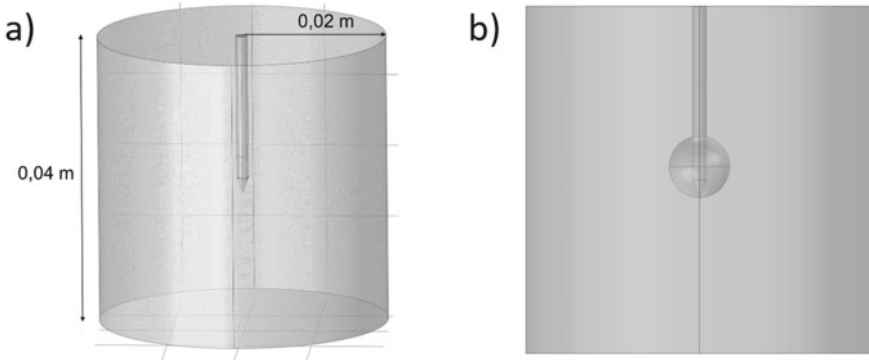
Fig. 31.1 Geometry of the simulation model. **a** Geometry of model were only the temperature distribution inside tumor tissue was calculated. **b** Geometry of the model after an interface between tumor tissue and healthy white brain matter was added

tumor tissue. The Flux/Source boundary condition is defined as

$$-n \cdot (-D\nabla\phi(r)) = g - q\phi(r). \tag{31.3}$$

where **n** is the outward unit vector perpendicular to the surface and g is the boundary flux, q describes the boundary absorption and is set to zero. This defines that all light reflected back to the catheter from the tissue is completely reflected at the catheter surface. The size of the boundary flux is defined as

$$S_P = \frac{10 \ W}{A_l} \tag{31.4}$$

where A_l is the area of light emittance. The Flux/Source boundary absorption is defined only for the domain of the light emitting area.

The second section of the mathematical model is the simulation of the heat transport within the tumor tissue. This process was modelled using the *Pennes*-Heat Distribution Equation. The Heat Distribution Equation is stored in *Comsol Multiphysics* through the Heat Transfer Module and is displayed as following

$$\rho c_p \frac{\delta T}{\delta t} - \nabla \cdot (k\nabla T) = Q_e + \rho_b c_b \omega_b (T_b - T) + Q_{met} \tag{31.5}$$

where ρ is the density of the tumor tissue (kg m^{-3}), c_p describes the heat capacity at a constant pressure (J kg^{-1} K^{-1}) and k the thermal conductivity of the tissue (W m^{-1} K^{-1}). T_b, c_b, ω_b, and ρ_b, are the arterial temperature of the blood (°C), the specific heat of the blood (J kg^{-1} K^{-1}), the blood perfusion rate (s^{-1}) and the mass density of the blood (kg m^{-3}), respectively. T is the temperature (°C) at a given time t (s). The temperature of the tumor tissue at the beginning of the simulation corresponds to T_b.

Q_{met} is the metabolic and Q_e is the external heat source, both measured in W m^{-3}. In this model the Q$_{met}$ is neglected. The external heat source corresponds to the source term and is referred to as

$$Q_e = S(r) = \mu_a \cdot \phi(r), \tag{31.6}$$

where μ_a is again the absorption coefficient of the tumor tissue and $\phi(r)$ the light fluence rate. Through $S(r)$, the Photon Distribution Equation and Heat Distribution Equation are connected via the light energy to heat source coupling [19].

The cooling generated inside the catheter by the flowing NaCl solution. To take this into account a temperature boundary condition is defined on the outside of the NaCl channel. The boundary condition defines the temperature of the saline cooling T_0 with

$$T_0 = 293.15 \ K \tag{31.7}$$

The last section of the mathematical model, the heat effects, were calculated according to the *Arrhenius-Damage-Equation*. They can be determined in *Comsol Multiphysics* within the Heat-Transfer Module. The program computes a damage integral which is given in the equation form

$$\Omega(r, \tau) = ln\left(\frac{C(r, 0)}{C(r, \tau)}\right) = \int_0^\tau A \cdot e^{-\frac{E_A}{RT(r,t)}} dt, \tag{31.8}$$

where A is the frequency factor, R the universal gas constant, ΔE the activation energy and T the temperature. $C(r,0)$ is the concentration of undamaged molecules at the beginning of the process and $C(\tau,0)$ at time τ. The values for the parameters of the *Arrhenius Equation* are shown in Table 31.1. Ω the *Damage Index*, is dimensionless, linearly dependent of exposure time and exponentially dependent on temperature.

A *Damage Index* of 1 corresponds to an irreversible tissue damage of 100%. Based on the results of Mohammed et al., complete tissue destruction is assumed in this model already at a θ of $>= 0.6$. A θ of 0.6 corresponds to a 50% reduction of healthy tissue cells [19].

Dynamic Optical Properties. To investigate the influence of optical tissue properties on the temperature distribution and the extent of tissue damage, the described mathematical model had to be extended. Mohammed et al. implemented dynamic optical properties in a way that they were directly dependent on the *Damage Index* [19]. Therefore, the optical tissue properties μ_a and μ'_s, were recalculated from the Damage Value in every new time step. The equations were integrated as follows:

$$\mu_a = \mu_{a, \ native} + \frac{\Omega}{\ln(2) + \Omega} \left(\mu_{a, \ coagulated} - \mu_{a, \ native}\right) \tag{31.9}$$

$$\mu'_s = \mu'_{s,native} + \frac{\Omega}{\ln(2) + \Omega} \left(\mu'_{s, \ coagulated} + \mu'_{s, \ native}\right) \tag{31.10}$$

Here, Ω stands for the *Damage Index* from 0 to infinity for complete tissue destruction, while a value of 1 corresponds with 37% living cells left in the tissue area, which, as stated before, coincides with an irreversible damage of 100% [19]. $\mu_{a,native}$ and $\mu'_{s,native}$ stand for the native values of absorption and reduced scattering coefficients respectively and $\mu_{a,\,coagulated}$ and $\mu_{s,\,coagulated}$ for the values of absorption and scattering coefficients of the coagulated tissue.

Values for the optical tissue parameters of coagulated brain tumor tissue were not apparent in literature. For this reason, values were taken from a study by Yaroslavsky et al., which investigated optical tissue properties of coagulated white matter. The exact values are shown in Table 31.1.

Temperature Dependent Blood Perfusion. Additional to dynamic optical properties, the influence of a temperature-dependent blood perfusion on temperature distribution and tissue damage was investigated. It was assumed that the blood flow rate increases linearly, starting from an initial blood flow rate until the *Damage Index* reaches a value of 0.6. When a mesh point reaches this value, the blood vessels were then assumed to be coagulated and the blood flow was set to zero.

To reproduce this behavior in the simulation, the blood perfusion rate within the model had to be made dependent on temperature and *Damage Index*. The previously described blood flow rate from Eq. 31.5 with a fixed value was instead dynamically designed. The dynamic blood perfusion rate was represented as follows:

$$\omega_d = \begin{cases} (\omega_b + bT), & \Omega < 0.6 \\ 0 & , \quad \Omega \geq 0.6 \end{cases} \tag{31.11}$$

where ω_d (s^{-1}) is the dynamic blood perfusion rate, ω_b (s^{-1}) is the initial blood perfusion rate, b (s^{-1}) is the mean blood perfusion, T (K) is the temperature and Ω is the *Damage Index* at this time step. Since no values for increasing blood flow rates of brain tumors were available in the literature, the value for the maximum blood perfusion rate was taken from the study by Kim et al. [17]. However, it should be noted that the values used by them were the blood perfusion rates of human skin. Since the skin is also a well perfused human tissue, the blood perfusion rates are assumed to be sufficiently accurate for the following simulations. The maximum and initial blood flow rate were then used to calculate the mean blood flow rate b using linear interpolation.

Nonlinear Algorithm for the FEM calculation. The algorithm presented in this paragraph was calculated at each mesh point for each time step in the FEM model to take the nonlinear calculations into account.

The total exposure time was 180 s and the initial body temperature was set to 37 °C. A small mesh size was selected for important areas of the geometry, such as the light emitting area, to gain a high accuracy within the simulation of the temperature distribution. In addition to the mesh density, the accuracy of the model can also be controlled by the internal time step size. Small time step size increases the accuracy, but also the computation time of the simulation. Here, the internal time step size was
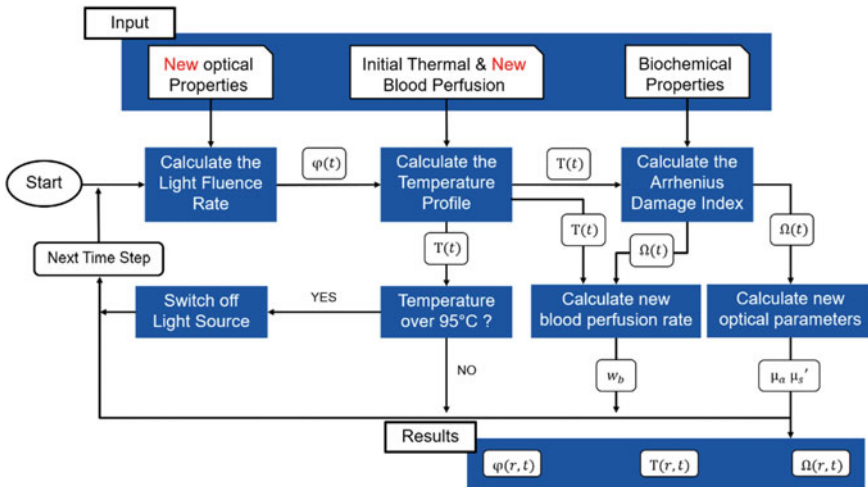
Fig. 31.2 Simulation loop for the completely dynamic simulation model. Simultaneously to the calculation of the optical tissue properties, the new blood flow rate was calculated from the values of temperature and damage index of the current simulation loop

set to 0.01 s, which means the program calculates values every 0.01 s. The solution plotting was set to save the values calculated per time step only once per second.

The simulation loop for these nonlinear calculations of the fully dynamic model is shown in Fig. 31.2. To avoid boiling of the tissue, a termination criterion was introduced parallel to the loop. If a temperature of more than 95 °C is calculated for a time step, the model switches off the light source and moves on to the next time step. The optical fiber is kept switched off until the calculated temperature falls below 95 °C again.

Interface between tumor tissue and physiological brain tissue. In all previously described simulation models, the entire tissue surrounding the catheter was simulated as tumor tissue. In vivo, tumor tissue of glioblastoma naturally borders onto healthy white matter of the brain. To investigate the influence of a more complex simulation geometry, the behavior of temperature distribution and tissue damage at the interface between tumor and healthy physiological tissue has been simulated. The fully dynamic model was used in the simulations. The geometry of the existing model had to be adapted. For this purpose, a new spherical domain was added, corresponding as the tumor (Fig. 31.1). A cylindrical volume surrounding the tumor was simulated as physiological white brain matter tissue for this simulation (Table 31.1).

In order to be able to test the temperature distribution and tissue damage at different tumor sizes, several simulations with various tumor radii were carried out.

31.3   Results

The simulations of the three models show substantial differences in their temperature distribution and volume of the coagulated tissue based on their specified properties. The basic model was simulated with static optical and thermal parameters. In the second model, dynamic optical parameters were added to the simulation. The third fully dynamic model additionally implemented the dynamic blood perfusion rate.

Figure 31.3 displayed the temperature distribution after 180 s ablation time for all three simulations. The black line described the area within which a tissue damage of $\Omega = 0.6$ or higher was calculated. This means that at least 50% of the cells within this area were necrotic. Tissue damage propagated itself in a ring-shaped and clearly defined area around the catheter.

For the basic simulation model with static optical and thermal parameters (Fig. 3a), a uniform circular temperature distribution around the catheter can be seen. Individual areas of different temperatures can be clearly distinguished from each other. This was especially the case for high temperatures. For lower temperatures, the heating appeared to be more diffuse. The temperature distribution in the second simulation model (Fig. 3b) was similar in structure to the temperature distribution of the basic simulation model, but the overall area of heated and necrotic tissue was reduced. The temperature distribution of the fully dynamic model (Fig. 3c) differed from the other simulations. It can be seen that the total amount of tissue heated and/or coagulated by the ablation process was greatly reduced by implementing the dynamic blood perfusion rate. The diffuse heating of tissue at lower temperatures was no longer detectable. Instead, a clear separation between tissue areas heated to over 50 °C and tissue areas which are hardly heated was evident.

The volumes of coagulated tissue that were estimated with the three simulation models are shown in Fig. 31.4. The blue curves show the simulated volume of tissue heated to temperatures of 50 °C and higher respectively for every model. The green curves represent the volume of tissue that reached an ablation index of 0.6 or higher.
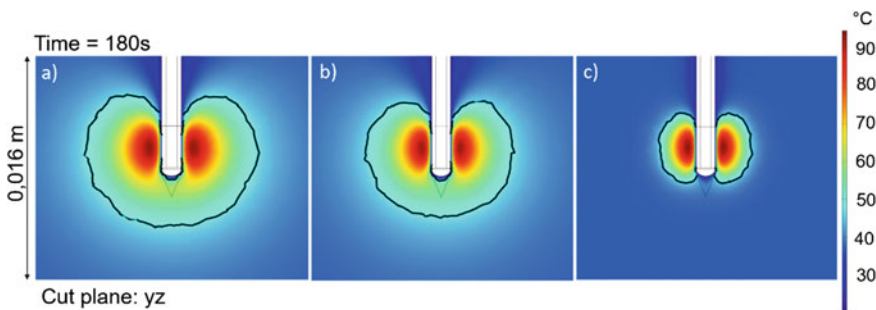

Fig. 31.3 Temperature distribution of the three simulation models. **a** Simulation result for the basic simulation model with fixed optical and thermal properties. **b** Simulation result for the model with dynamic optical properties. **c** Simulation result for the model with dynamic optical properties and a temperature dependent blood perfusion
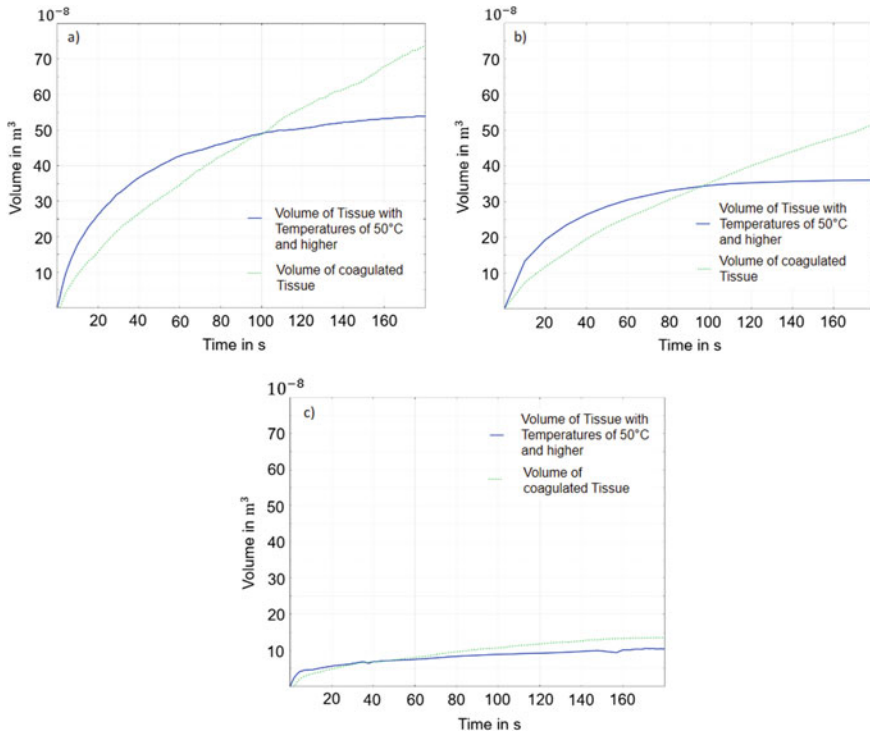
Fig. 31.4 Volume of coagulated and heated tumor tissue plotted against ablation time for all three simulation models. **a** Simulation result for the basic simulation model with fixed optical and thermal properties. **b** Simulation result for the model with dynamic optical properties. **c** Simulation result for the model with dynamic optical properties and a temperature dependent blood perfusion

For the basic simulation model (Fig. 4a), the tissue volume heated to temperatures of 50 °C and above increased in the first half of the ablation process, while the process slowed down in the second half. After the implementation of dynamic optical properties (Fig. 4b) the progression of heated and coagulated tissue is altered. After 80 s, the volume fraction reached a maximum value and then remained stable for the rest of the ablation time. The volume reduction of both parameter progressions was visible. Figure 4c shows that the dynamically designed blood perfusion has an influence on temperature distribution and its associated tissue damage. The volume of both heated and coagulated tissue was considerably reduced in this simulation model. The volume fraction of heated tissue increases strongly in the first five seconds of the ablation model, and after that strove towards a maximum value. In contrast to the previous simulations, the increase in the volume of necrotic tissue was similar to the course of tissue heating. At the beginning of the process, the volume increase was more rapidly and then tended towards a maximum limit value.

To test the influence of the geometry of the tumor on temperature distribution, simulations with different tumor radii and additional healthy brain tissue were carried
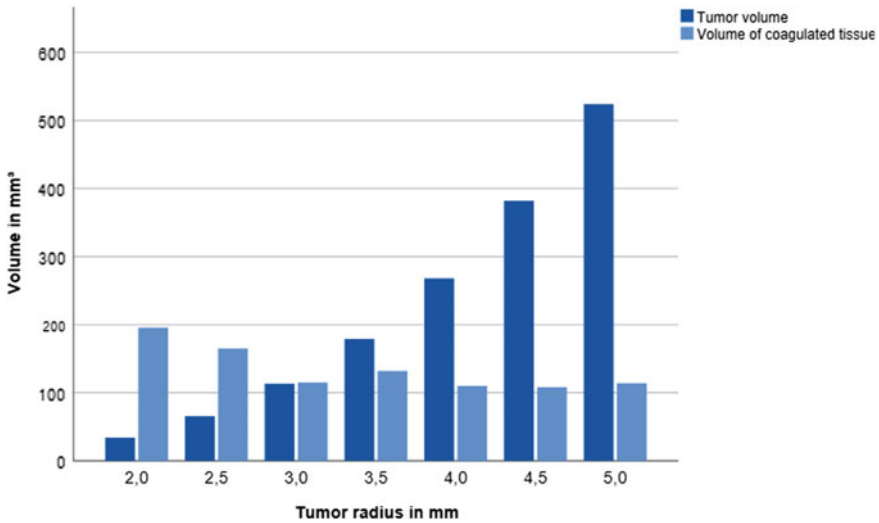
Fig. 31.5 Comparison between the volume of coagulated tissue and the tumor volume. As the tumor radii increased, the volume of coagulated tissue did not grow as well, but remained at a constant level

out. Figure 31.5 compared the volume of coagulated tissue and the tumor volume, simulated by the fully dynamic model. It can be seen that for the two smallest tumor radii, the volume of necrotic tissue exceeded the tumor volume.

Thus, more tissue was heated to a lethal thermal range than the tumor volume occupied. In the simulation, the 3 mm radius tumor could be completely destroyed by the ablation process without destroying surrounding physiological white matter tissue. As the tumor radii increased, the volume of coagulated tissue did not grow as well, but remained at a constant level.

31.4   Discussion

The presented study focused on the simulation of the ablation process of a glioblastoma using the LITT method. The aim was to investigate the influence of model geometry and complexity of the mathematical calculations on temperature distribution and the extent of tissue damage. For this purpose three simulation models were compared to each other.

The first simulated ablation process was time-dependent but with fixed thermal and optical parameters. Subsequently, in the second model, first dynamic optical parameters and then, in the third model, a dynamic blood perfusion was implemented in the simulation.

Additionally, it should be taken into account that, when the accuracy of the simulated ablation process is being discussed, there are limitations due to not accessible

tissue data. In the context of this study, not all values for the required input parameters were available in literature. These alterations could have influenced the simulation results. Accordingly, it can be assumed that the results of the simulations in this study will deviate from in vitro or even in vivo investigations.

Despite these limitations, the presented study was able to show the influence of the underlying mathematical model and tissue geometry on the simulation results of temperature distribution and extent of tissue damage.

Comparing the simulations (a), (b) and (c), the uniform layer-by-layer temperature distribution can be seen in all three simulations. Individual layers of the same temperature level can be distinguished. This temperature distribution is typical for a linear light source and can be explained by the decrease of light deposition in deeper tissue regions due to an increasing number of absorption events [18]. An increased distance from the light source leads to an exponential decrease of the light fluence rate [17].

Furthermore, simulations (3a) and (3b) showed that the volume of coagulated tissue at the end of the process exceeds the volume of tissue that is heated to 50 °C. This leads to the conclusion that the tumor tissue partially reaches the value of lethal tissue damage of $\Omega = 0.6$ already at temperatures below 50 °C. These observations coincide with the process of hyperthermia at a temperature range between 42 and 50 °C, in which molecular changes are initiated which, if irradiated for a sufficiently long time, can already cause coagulation of the tissue [18].

The simulation models show, that with dynamic optical parameters implemented, the temperature increase reached a steady state at some point. The non-linear behaviour of the temperature profile due to the increased scattering coefficient in the coagulated tissue, leads to a higher light energy deposition near the optical fiber. This effect was further enhanced by the implementation of dynamic blood perfusion, since the cooling effect of blood perfusion in deeper tissue layers slowed down the spread of the temperature profile [17].

It was found, that the influence of dynamic optical tissue parameters significantly reduces the volume of both heated and coagulated tissue. Simulation 2b calculated a 30% lower volume of tissue heated to 50 °C and this reduction is also visible in the temperature distribution. The volume of coagulated tissue was also reduced by about 30% compared to simulation 3a by implementing the dynamic optical parameters. The influence of the optical parameters seems to slow down the propagation of the temperature and coagulation front significantly from the beginning of the process. An increase of the scattering coefficient of the coagulated tissue results in a reduction of the penetration depth of the temperature and coagulation fronts [19]. Accordingly, the influence of the scattering coefficient has a great impact, reducing the volumes of coagulated tissue and the penetrating depth of heated tissue. It is noted, that authors cannot verify whether the increase of the scattering coefficient is representative since data on coagulated glioblastoma tissue was not available in the literature.

Influence of dynamic blood perfusion on the temperature distribution and volume of coagulated tissue is more evident. The penetration extent of coagulated tissue was reduced, as the increased blood flow in the hyperemic ring removes the heat from the coagulation front [17]. Furthermore, it is visible that the diffuse tissue heating

to increased temperatures is no longer present in simulation (2c). The heating of the tissue outside the coagulation front is reduced. This is due to the presence of two heat sinks, one of which is the constant basic perfusion in this tissue area and the other is an even higher perfusion zone due to dilated blood vessel which is located between the coagulation front and the heated one.

Dynamic thermal and optical parameters changed the behavior of temperature distribution and extent of tissue damage strongly. The volume of coagulated tissue of simulation (3a) was simulated 30% higher compared to simulation (3b) and 82.9% higher compared to simulation (3c). Dynamic parameters must therefore be inserted into the simulation in order to be able to make a sufficient prediction of the occurring tissue damage.

Another aspect of interest was the behaviour of temperature propagation and extent of tissue damage at the interface between tumor tissue and physiological white brain matter, in order to investigate the influence of model geometry. The transition of light deposition into another tissue with different optical properties has a strong influence on temperature propagation and tissue damage and therefore needs to be considered in simulations. Thus, the physiological tissue seems to heat up more quickly and therefore shows more pronounced tissue damage.

As the tumor radius increases, the volume of coagulated tissue reaches a limit value is therefore in good agreement with the previous simulations in which solely tumor tissue is simulated. The tissue damage reaches a steady state, when the radius of the tumor is so large that the interface between physiological tissue and tumor tissue no longer has an influence on temperature propagation and tissue damage. These simulations show, that an exact geometry of the tumor, but also of the tissue adjacent to the tumor, is necessary to be implemented in order to be able to make precise prognoses about the extent of the tissue damage induced by the treatment.

31.5 Conclusion

The aim of this study was the implementation of a FEM model, which simulated the ablation process of a glioblastoma using the LITT method. For this purpose the temperature distribution and, based on this, the extent of the tissue damage were simulated as precisely as possible. In this work, a time-dependent model with static optical and thermal parameters was compared to simulations that contained dynamic optical parameters and one with a dynamic blood perfusion. The influence of the optical parameters and the dynamic blood perfusion on the thermal distribution and tissue damage were investigated. The study shows that a fully dynamic design of the simulation model influences the temperature distribution and the extent of tissue damage. Mainly due to the increased scattering coefficient and the cooling effect of dynamic blood perfusion, the penetration depths of the temperature and coagulation front were reduced. Compared to the dynamic model, the static simulation model overestimated the extent of tissue damage by 82.9%. The extension of the simulation model shows that the interface between healthy and tumor tissue contributed to the

extent of tissue damage by the corresponding changes in the optical parameters. For example, larger tissue volumes were destroyed with smaller tumor radii than with larger tumor radii.

Future experimental work, for example the establishment of the actual perfusion change with temperature and its impact on the ablation process, will be necessary to confirm the clinical significance of these results.

References

1. Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. Gemeinsame Publikation des Zentrums für Krebsregisterdaten und der Gesellschaft der epidemiologischenKrebsregister in Deutschland e.V.: Krebs in Deutschland (2013/2014)
2. D.A. Hardesty, N. Sanai, The value of glioma extent of resection in the modern neurosurgical era. Front. Neurol. **3** (2012)
3. O. Ashraf, N.V. Patel, S. Hanft, S. Danish, Laser-induced thermal therapy in Neuro-oncology: a review. World Neurosurg. **112**, 166–177 (2018)
4. S.C. Jiang, X.X. Zhang, Effects of dynamic changes of tissue properties during laser-induced interstitial thermotherapy (LITT). Lasers Med. Sci. **19**, 197–202 (2005)
5. C. Rossmann, D. Haemmerich, Review of temperature dependence of thermal properties, dielectric properties, and perfusion of biological tissues at hyperthermic and ablation temperatures. Crit. Rev. Biomed. Eng. **42**, 467–492 (2014)
6. T. Uzuka, R. Tanaka, H. Takahashi, K. Kakinuma, J. Matsuda, K. Kato, Planning of hyperthermic treatment for malignant glioma using computer simulation. Int. J. Hyperth. Offic J. Eur. Soc. Hyperth. Oncol. North American Hyperth. Group **17**, 114–122 (2001)
7. D.J. Schutt, D. Haemmerich, Effects of variation in perfusion rates and of perfusion models in computational models of radio frequency tumor ablation. Med. Phys. **35**, 3462–3470 (2008)
8. M. Ahmed, Z. Liu, S. Humphries, S. Nahum Goldberg, Computer modeling of the combined effects of perfusion, electrical conductivity, and thermal conductivity on tissue heating patterns in radiofrequency tumor ablation. Int. J. Hyperth. **24**, 577–588 (2008)
9. M. Iizuka, A. Vitkin, M. Kolios, M. Sherar, The effects of dynamic optical properties during interstitial laser photocoagulation. Phys. Med. Biol. **45**, 1335–1357 (2000)
10. H.J. Schwarzmaier, I.V. Yaroslavsky, A.N. Yaroslavsky, V. Fiedler, F. Ulrich, T. Kahn, Treatment planning for MRI-guided laser-induced interstitial thermotherapy of brain tumors–the role of blood perfusion. J. Magnet. Reson. Imaging: JMRI **8**, 121–127 (1998)
11. J. D. Johansson, Impact of tissue characteristics on radio-frequency lesioning and navigation in the Brain: Simulation, experimental and clinical studies, in *Linköping Studies in Science and Technology. Dissertations*, vol. 1230 (2009)
12. K.K. Mendis, R.L. Stalnaker, S.H. Advani, A constitutive relationship for large deformation finite element modeling of brain tissue. J. Biomech. Eng. **117**, 279–285 (1995)
13. A. Carpentier, R.J. McNichols, R. Stafford, J.-P. Guichard, D. Reizine, S. Delaloge, E. Vicaut, D. Payen, A. Gowda, B. George, Laser thermal therapy: Real–time MRI–guided and computer–controlled procedures for metastatic brain tumors. Lasers Surg. Med. **43**, 943–950 (2011)
14. Kern GmbH: Polycarbonate (PC). https://www.kern.de/de/technisches-datenblatt/polycarbonat-pc?n=2301_1. Last accessed 25 July 2020
15. A.N. Yaroslavsky, P.C. Schulze, I.V. Yaroslavsky, R. Schober, F. Ulrich, H.-J. Schwarzmaier, Optical properties of selected native and coagulated human brain tissues in vitro in the visible and near infrared spectral range. Phys. Med. Biol. **47**, 2059–2073 (2002)

16. It is foundation. https://itis.swiss/virtual-population/tissue-properties/database/. Last accessed 25 July 2020
17. B.-M. Kim, S.L. Jacques, S. Rastegar, S. Thomsen, M. Motamedi, Nonlinear finite element analysis of the role of dynamic changes in blood perfusion and optical properties in laser coagulation of tissue. IEEE J. Sel. Top. Quantum Electron. **2**, 922–933 (1996)
18. M. Niemz, *Laser-Tissue Interactions: Fundamentals and Applications*, 4th edn. (Springer International Publishing, Heidelberg, 2019)
19. Y. Mohammed, J. Verhey, A finite element method model to simulate laser interstitial thermo therapy in anatomical inhomogeneous regions. Biomed. Eng. Online **4**, 1–16 (2005)

Chapter 32
Design of CMOS Operational Amplifiers with Dynamic Offset Cancellation

Check for updates

Ara Abdulsatar Assim and **Evgenii Balashov**

Abstract In this paper, the design of CMOS Operational Amplifiers based on 50 nm CMOS technology is presented. Operational amplifiers use dynamic offset cancellation techniques to decrease the offset voltage and flicker noise. Several methods to dynamically reduce offset are compared, commonly known as chopping and auto-zeroing, and a combination of both methods may also be used. It was shown that the dynamic offset cancellation techniques based on chopping are more effective, since the flicker noise was reduced by nearly 4800% (from 102 to 2 uV/$\sqrt{}$Hz). The flicker noise reduction percentage for auto-zeroing amplifiers strongly depends on the used configuration, a simple discrete-time auto-zeroing amplifier with only one operational amplifier is not the most optimal configuration to use because it reduces flicker noise by 366.6% only (i.e. from 28.6 to 7.8 uV/$\sqrt{}$Hz), a continuous-time auto-zeroing amplifier reduces noise by 328% (from 31.6 to 0.09 uV/$\sqrt{}$Hz) while a discrete-time auto-zeroing operational amplifier with a supplementary amplifier reduces flicker noise by 4563% (from 28.7 to 0.63 uV/$\sqrt{}$Hz), its performance is slightly worse than the operational amplifier with chopping compensation.

Keywords Dynamic offset cancellation · Operational amplifier · Chopping · Auto-zeroing · CMOS

32.1 Introduction

The offset voltage problem rises due to random uncertainties that are unavoidable during the transistor manufacturing process. The effect of the offset voltage degrades the characteristics of the differential amplifiers' input stages in operational and instrumental amplifiers (even in high frequency amplifiers) [1–18]. MOSFET transistors

A. A. Assim (✉) · E. Balashov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: assim.a@edu.spbstu.ru

A. A. Assim
Salahaddin University, Erbil 44001, Iraq

have inconsistent threshold voltages, the threshold voltage depends on the transistor's doping level, and doping levels are not the same for all transistors in a circuit. Another widespread problem is incompatibility between the MOS transistors' dimensions (lengths and widths) [8–11]. This deteriorates an amplifier's performance. Another problem is the flicker noise [7–16]. Generally, flicker noise is higher in low frequencies, it is commonly referred to as (flicker noise) [7, 11–13], the point at which flicker noise lessens to a value equal to the thermal noise is known as corner frequency [9]. The value of the offset voltage heavily depends on the MOS technology [7] and is less for transistors with longer channel lengths. Commonly available op-amps have an offset voltage of 1 to 10 mV maximum [6, 10–12], this value isn't very notable but nevertheless, due to op-amp's large gain nature, this value is increased by tens or hundreds, this causes distortion at the output (the output signal is clipped from above or below), another issue rises due to offset, that is the limitation of the op- amp's largest allowable input voltage, an example of that is the amplifier designed in [5, 13], it can be optimized by adding one of the offset voltage cancellation methods to it, thus, its crucial to get rid of offset voltage [4–14]. Dynamically compensated op-amps are widely used for many applications, including but not limited to signal mixers, analog to digital converters, instrumentation amplifiers [5, 14–17], etc. The proposed circuits can be realized in practice and achieve similar results experimentally [16].

32.2 Offset Cancellation Techniques

32.2.1 A Basic Auto-Zeroing Configuration (BAZ)

A simple auto-zeroing compensated amplifier is given in Fig. 32.2, its operation is based on the periodical autocalibration, in which the common mode voltage (VCM) is used as a reference. Clock signals C1 and C2 are out of phase by 180 degrees. If C1 is high and C2 is low, then inputs Vn and Vp and the output (OUT) of the op-amp are shorted. As a result, the input voltage on the input of the amplifier is the offset voltage. The amplifier's output is equal to the input offset voltage (10 mV). The capacitor is charged with the help of the feedback loop. In the second cycle C2 is high and C1 is low. The amplifier works as an inverting op-amp, but with the offset removed, since the voltage appearing on the capacitor cancels the offset voltage. This results in nulled offset voltage at the input. The circuit diagram of the used op-amp is provided in Fig. 32.1, the offset voltage is modeled as a DC voltage source (VOS), connected to the positive input (Vp), in this op-amp, VDD is 1 V, Vbias is 353 mV and GND is 0 V.

Figure 32.3 shows the PSS (Periodic Steady State) noise analysis results for the basic auto-zeroing amplifier. It can be seen how flicker noise is reduced from 28.69 uV/$\sqrt{}$Hz to 7.826 uV/$\sqrt{}$Hz. Moreover, the circuits use real transistors, thus the clock signals leak to the output and increase output noise, that issue is not resolved in this work.
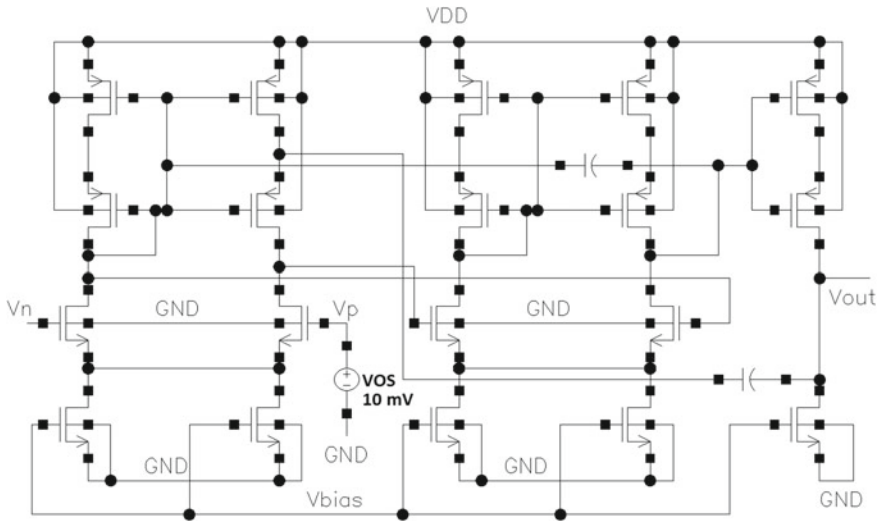
Fig. 32.1 Schematic of the operational amplifier
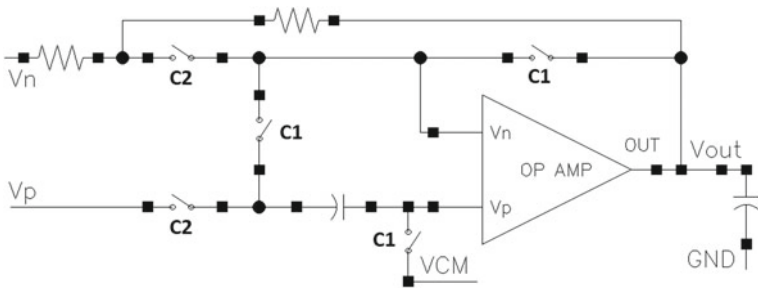

Fig. 32.2 A basic auto-zeroing configuration

Time-domain analysis shows an offset error of 10.19 mV has been corrected, i.e., offset voltage has been reduced from 10.18 mV to −6.79 uV. It is worth to mention that the compensation process does not happen immediately, it takes about 150 us before it starts to stabilize.

32.2.2 Auto-Zeroing with an Auxiliary Amplifier

A better configuration for auto-zeroing can be made by adding an auxiliary amplifier, this greatly reduces the flicker noise (see Fig. 32.4). If C1 is on, the input signals drive the feedforward amplifier and are amplified normally, in the following cycle, the feedforward amplifier is driven by the output signal (Vout) and the common mode
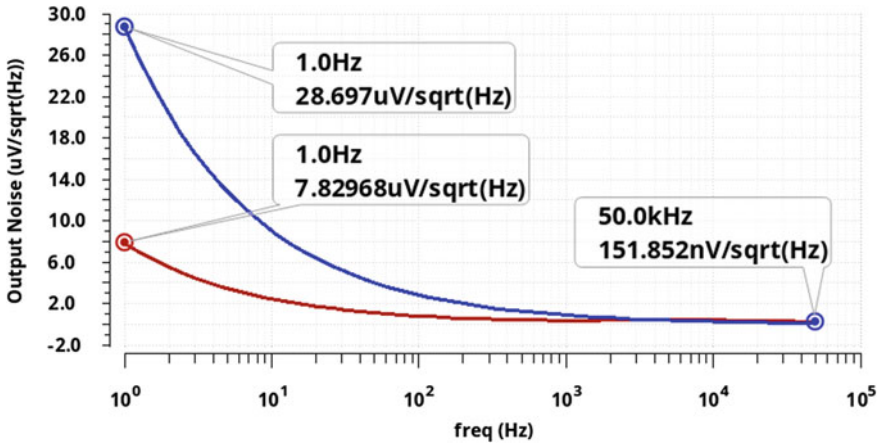
Fig. 32.3 Dependence of the noise versus frequency for a basic auto-zeroing amplifier with (blue) and without (red) compensation
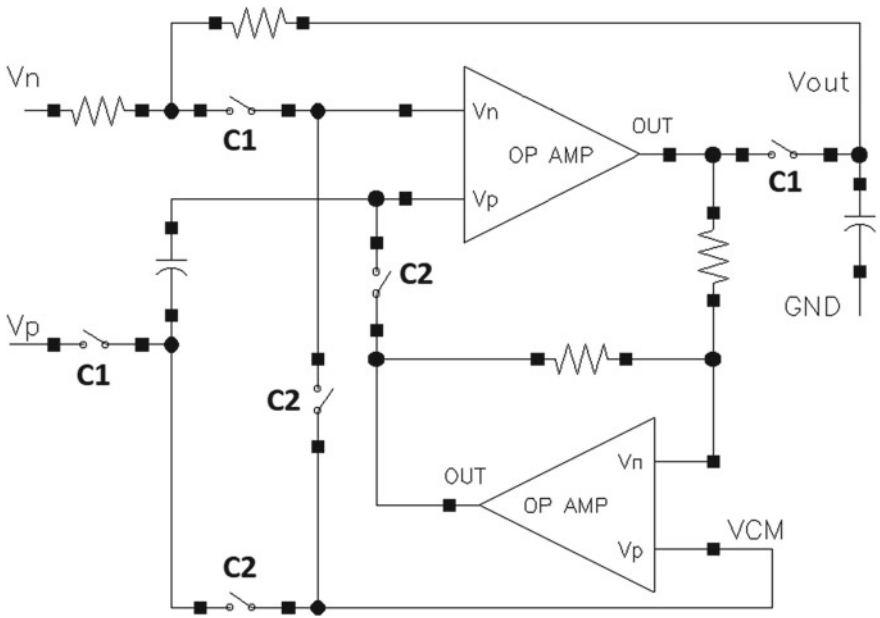
Fig. 32.4 Auto-zeroing amplifier with an auxiliary amplifier

voltage (VCM). The capacitor is fed by Vp if C1 is on and by VCM if C2 is off, it measures the offset voltage and compensates it. The simulation results PSS (Periodic Steady State) analysis for the auto-zeroing amplifier with an auxiliary amplifier in Fig. 32.4 is provided in Fig. 32.5. Noise analysis results in Fig. 32.5 prove that adding
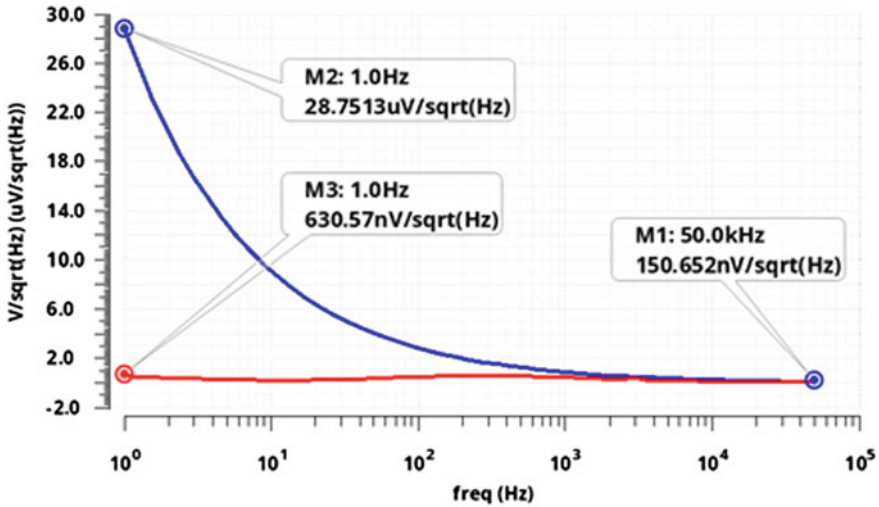
Fig. 32.5 Dependence of the noise versus frequency for the amplifier in Fig. 32.4 with (red) and without (blue) compensation

an auxiliary amplifier in the feedback path drastically enhances the amplifier's noise performance, this configuration is far more superior than the previous auto-zeroing configuration. Noise is reduced from 28.75 to 0.630 uV/$\sqrt{}$Hz.

Additionally, the transient analysis depicts that an offset voltage error of 9.249 mV has been corrected, note that the offset voltage compensation process happens much faster compared to the previous case (BAZ amplifier).

32.2.3   Continuous-Time Auto-Zeroing (CTAZ)

The two auto-zeroing amplifiers previously proposed were designed for use in discrete-time applications. A different scheme is designed that is suitable to be used in continuous-time applications, commonly known as continuous-time auto-zeroing amplifier. The amplifier consists of the two amplifiers (Fig. 32.6). In the first moment, the first amplifier amplifies the input signal, and the second amplifier is calibrated. In the second moment the first amplifier is calibrated, and the second amplifier amplifies the signal. At first cycle when C2 is on (C1 is off), the amplifier above is driven by Vn and Vp, they are amplified and appear at (Vout), at the next cycle, C2 is off (C1 is on), the amplifier below works as a typical amplifier but the amplifier above nulls offset, and output is taken from it, that will lead to a continuous signal at both cycles with removed offset voltage.

The results of PSS (Periodic Steady State) noise analysis of the CTAZ amplifier is given in Fig. 32.7, the flicker noise has been reduced from 31.60 to 0.096 uV/$\sqrt{}$Hz and an error offset voltage of 9.999 mV has been corrected (-10.019 mV to 19.75 uV).
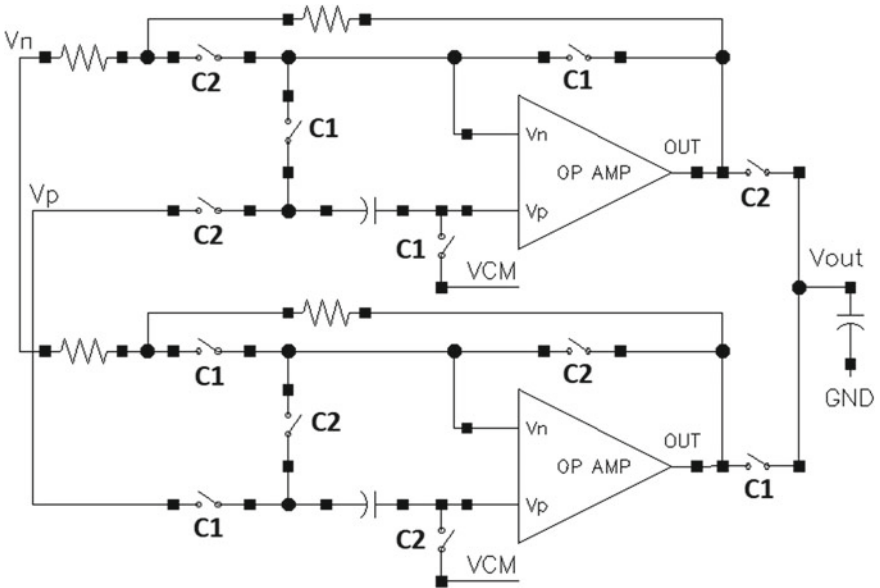
Fig. 32.6 Continuous-time auto-zeroing amplifier (CTAZ)
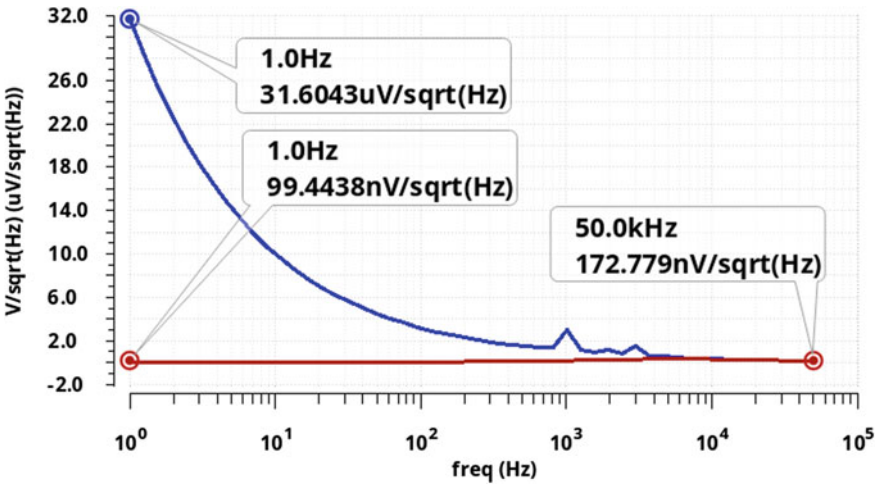
Fig. 32.7 Dependence of the noise versus frequency for CTAZ amplifier in Fig. 32.5 with (blue) and without (red) offset compensation

32.2.4   Chopping

Chopping is different from auto-zeroing, it is based on modulation instead of charge compensation. A basic chopper amplifier is provided in Fig. 32.8.

In this method a fully differential operational amplifier should be used. The Butter-worth low-pass filter is used to get rid of the voltage spikes in the output signal. The differential input signals $V+$ and $V-$ modulates the clock signal by the first chopper. As a result, the spectrum of the signal is transferred to a higher frequency that is equal to the chopping frequency. Then the signals along with the offset are amplified by the fully differential amplifier. The second chopper demodulates the amplified signal and changes its frequency to its original value. But the offset is shifted to the frequency of the clock harmonics. A major advantage of chopping is that no noise folding is produced. The flicker noise is reduced from 102.7 to 2.132 uV/$\sqrt{\text{Hz}}$ as shown in Fig. 32.9, in addition to that, an offset voltage error of 9.92 mV has been corrected.
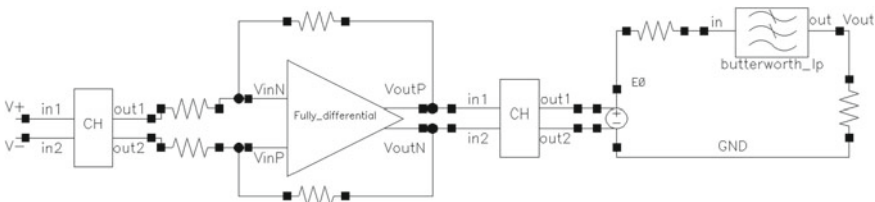
Fig. 32.8 A basic chopper amplifier
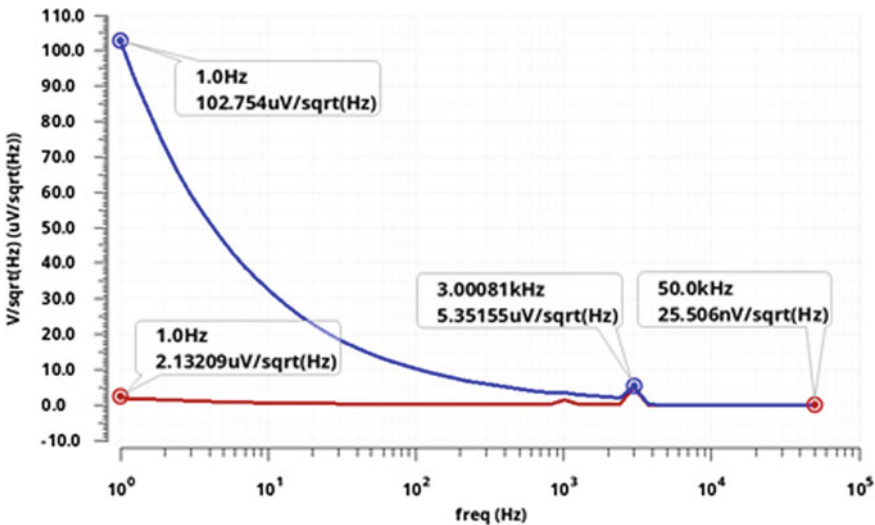
Fig. 32.9 Dependence of the noise versus frequency of chopper amplifier's with (red) and without (blue) compensation

Table 32.1 Noise comparison of the proposed offset cancellation methods

Amplifier type	Noise before/after compensation (uV/$\sqrt{}$Hz)	Noise reduction (%)	Offset voltage before/after compensation (uV)	Voltage reduction (%)
Basic auto-zeroing (AZ)	28.7/7.8	367	10,000/41.15	24,301
AZ with an auxiliary amplifier	28.8/0.6	4564	10,000/11	90,909
Continuous-time auto-zeroing	31.6/0.1	329	10,000/46	21,739
Chopping	102.8/2.1	4820	10,000/10.42	95,969

32.3   Discussion

The noise performance of the presented dynamically compensated op-amps are summarized in Table 32.1 (based on simulation using Cadence). The result shows that chopping and AZ with an auxiliary amplifier are reducing the flicker noise more effectively (by approximately 12.4–14.6 times more), however the choice of an amplifier for a certain application does not only depend on the flicker noise reduction rate. For continuous-time applications such as analog-to-digital conversion, CTAZ should be used despite its inferior characteristics in comparison to the other amplifiers.

32.4   Conclusion

In this paper, four operational amplifier configurations for compensating input offset voltage, which were designed in 50 nm CMOS technology, were presented. In auto-zeroing amplifiers, the value of the capacitor must be chosen carefully, because it plays a major role in offset voltage compensation process. While for chopper amplifiers, the chopping frequency is the most important parameter, in this work a frequency of 10 kHz was chosen. The result shows that chopping and AZ with an auxiliary amplifier reduce flicker noise more effectively (by approximately 12.4–14.6 times more).

References

1. E.V. Balashov, N.V. Ivanov, A.S. Korotkov, SOI instrumentation amplifier for high-temperature applications, in *2020 IEEE East-West Design & Test Symposium (EWDTS)* (IEEE, Bulgaria, 2020), pp. 1–4
2. A. Korotkov, D. Morozov, M. Pilipko, M. Yenuchenko, Sigma-Delta ADC on SOI technology for working at high temperatures. Radioelectron. Commun. Syst. **63**(11), 586–595 (2020)

3. I.A. Rumyancev, A.S. Korotkov, Survey on Beamforming Techniques and Integrated Circuits for 5G Systems, in *2019 IEEE International Conference on Electrical Engineering and Photonics*. (IEEE, St. Petersburg, Russia, 2019), pp. 76–80

4. R.J. Baker, *CMOS Circuit Design, Layout, and Simulation*, 3rd edn. (Wiley-IEEE Press, New Jersey, 2010)

5. R. Wu, J. Huijsing, K. Makinwa, *Precision Instrumentation Amplifiers and Read-Out Integrated Circuits,* 1st edn. (Springer, Netherlands, 2013)

6. J. Witte, K. Makinwa, J. Huijsing, *Dynamic Offset Compensated CMOS Amplifiers,* 1st edn. (Springer, Netherlands, 2009)

7. V. Raghuveer, K. Balasubramanian, S Sudhakar, A $2\mu V$ low offset, 130 dB high gain continuous auto zero operational amplifier, in *International Conference on Communication and Signal Processing (ICCSP)*, (IEEE, Chennai, 2017), pp. 1715–1718

8. X. Kuang, T. Wang, F. Fan, The design of low noise chopper operational amplifier with inverter, in *2015 IEEE 16th International Conference on Communication Technology (ICCT),* (IEEE, Hangzhou, 2015), pp. 568–571

9. Y. Kusuda, A 60 V auto-zero and chopper operational amplifier with 800 kHz interleaved clocks and input bias current trimming. IEEE J. Solid-State Circuits **50**(12), 2804–2813 (2015)

10. Q. Fan, J. Huijsing, A. Makinwa, A multi-path chopper stabilized capacitively coupled operational amplifier with 20V-Input-Common-Mode range and 3 V offset, in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, (IEEE, San Francisco, 2013), pp. 176–177

11. Y. Kusuda, A 5.6 nV/$\sqrt{}$ Hz chopper operational amplifier achieving a 0.5 μV maximum offset over rail-to-rail input range with adaptive clock boosting technique. IEEE J. Solid-State Circuits **51**(9), 2119–2128 (2016)

12. A. Grasso, D. Marano, G. Palumbo, S. Pennisi, Design methodology of subthreshold three-stage CMOS OTAs suitable for ultra-low-power low-area and high driving capability. IEEE Trans. Circuits Syst. I: Regul. Papers **62**(6), 1453–1462 (2015)

13. L. Magnelli, F. Amoroso, F. Crupi, G. Cappuccino, G. Iannaccone, Design of a 75-nW, .05-V subthreshold complementary metal-oxide-semiconductor operational amplifier. Int. J. Circuit Theory Appl. **42**(9), 967–977 (2014)

14. G. Shukla, N. Srivastava, A. Shadab, DC offset voltage reduction in cascaded instrumentation amplifier, in *2013 Students Conference on Engineering and Systems (SCES),* (IEEE, Allahabad 2013), pp. 1–5

15. N. Bortun, M. Stan, G. Brezeanu, High precision bidirectional chopper amplifier with extended common mode input voltage range, in *International Semiconductor Conference (CAS)* (Sinaia, 2015), pp. 297–300

16. M. Duwe, T. Chen, Offset correction of low power, high precision op amp using digital assist for biomedical applications, in *IEEE International Symposium on Circuits and Systems (ISCAS)* (IEEE, Seoul, 2012), pp. 850–853

17. V. Ivanov, M. Shaik, A 10MHz-bandwidth 4μs-large-signal-settling 6.5nV/$\sqrt{}$Hz-noise 2μV-offset chopper operational amplifier, in*2016 IEEE International Solid-State Circuits Conference (ISSCC)* (IEEE, San Francisco, 2016), pp. 88–89

18. T. Mai, K. Schmid, J. Rober, A. Hagelauer, R. Weigel, A fully differential Operational Amplifier using a new Chopping Technique and Low-Voltage Input Devices, in *24th IEEE International Conference on Electronics, Circuits and Systems (ICECS)* (IEEE, Batumi, 2017), pp. 74–77

Chapter 33
Research of the Microelectromechanical Device Cross-Shaped Element

Igor E. Lysenko, Nikolai F. Kidyaev, and Olga A. Ezhova

Abstract This article examines a cross-shaped design element of a linear acceleration and angular velocity sensor with three sensitivity axes. Variants of changes in the suspension design of the linear acceleration and angular velocity sensor are presented. The parameterized geometry of the gyroscope-accelerometer is modeled based on the ANSYS software package. The dependence of inertial mass movements on changes in the parameters of the cross-shaped gyroscope-accelerometer element is obtained.

Keywords MEMS · Gyroscope · Accelerometer · ANSYS

33.1 Introduction

The development of MEMS sensors has significantly influenced the implementation of fundamentally new applications and methods of human interaction with them. Sensors of small accelerations and linear speeds are firmly established in the market of smartphones, game consoles, tablets. Currently, the improvement of microelectromechanical sensors is one of the most urgent problems.

Thus, the article solves the problem of improving the model of a three-axis micromechanical sensor of linear accelerations and angular velocities with two inertial masses. This will improve the characteristics of the MEMS sensor model and increase the competitiveness of MEMS devices in the microelectronic industry market [1–15].

I. E. Lysenko · N. F. Kidyaev (✉)
Southern Federal University, Shevchenko str., 2, Taganrog, Russia 347922
e-mail: nkidyaev@sfesu.ru

O. A. Ezhova
LLC «MEMS Design», Iniciativnaya str., 60, Taganrog, Russia 347936
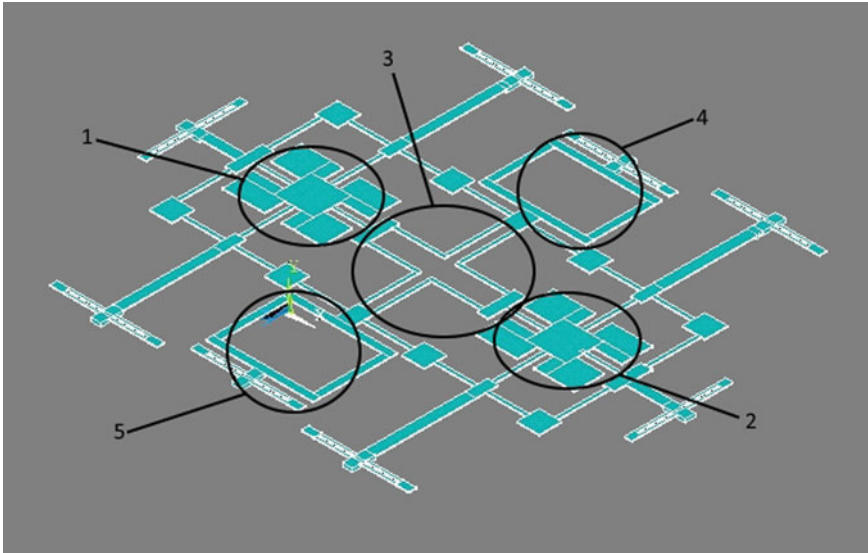
Fig. 33.1 MEMS Gyroscope-Accelerometer design

33.2   MEMS Gyroscope-Accelerometer Designs

In this paper, the design of the linear acceleration and angular velocity sensor with three sensitivity axes is shown in Fig. 33.1. In the proposed suspension design, the cruciform element of rigidity (3), consisting of series–parallel connections of elastic beams, is a distinctive element from its analogues. It is located in the center of the structure for connecting the electrostatic actuators (4, 5) to the inertial masses (1, 2). Within the framework of this article, three additional models of the micromechanical gyroscope-accelerometer suspension with modified parameters of the beams around the cruciform element are considered. Figures 33.2 and 33.3 show the versions of the cross-shaped element [3, 14].

The width of the standard beams around the cruciform stiffener is 5 microns, and the width of the wide beams around the cruciform stiffener is 10 microns.

The width of thin beams around the cross-shaped stiffener is 2.5 microns.

33.3   Simulation of MEMS Gyroscope-Accelerometer

For a comparative analysis of the designs of the micromechanical gyroscope-accelerometer presented in the article, parametrizable elastic suspension model by the ANSYS software environment was developed [13, 15].
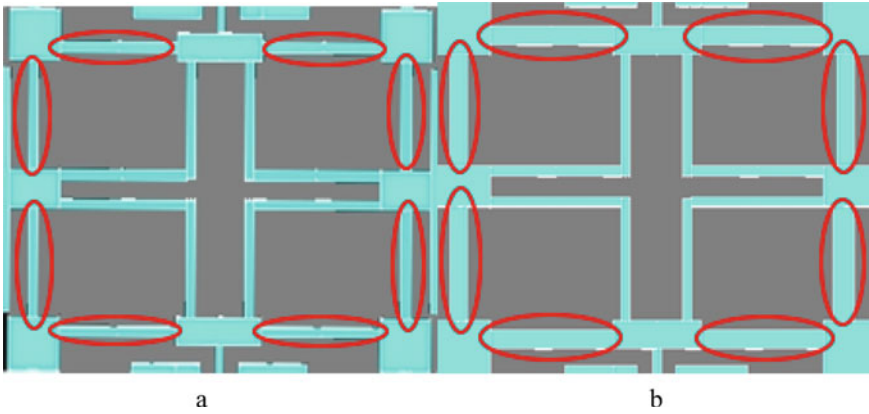
Fig. 33.2 Construction of a cross-shaped element with standard (**a**) and wide (**b**) beams
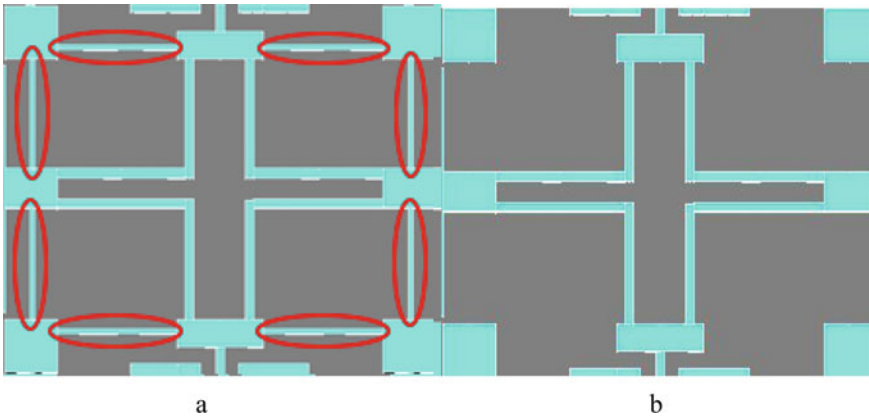
Fig. 33.3 Cruciform element design with thin beams (**a**) and without beams (**b**)

For static analysis, a force of 10^{-9} N was applied to all types of suspension from the side of the electrostatic actuators. These forces are applied in the direction of each other. The results of the simulation are shown in Figs. 33.4, 33.5, 33.6 and 33.7.

The simulation result shows that the total displacement vector of the inertial masses is $2.5*10^{-13}$ m.

The simulation result shows the total displacement vector of the inertial masses is $1.36*10^{-14}$ m. Due to the increased rigidity of the cross-shaped element of the micromechanical gyroscope-accelerometer, the movement of the inertial masses is so small that the ANSYS does not form a color scale of deformation.

The simulation result shows that the total displacement vector of the inertial masses is $9.18*10^{-13}$ m. The natural result is achieved by reducing the resistance of the cruciform element and the electrostatic drives driven.
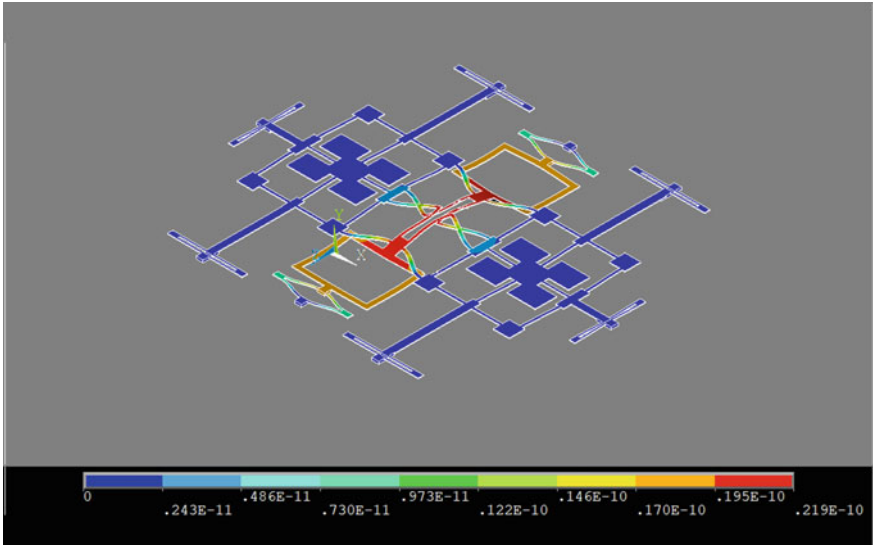
Fig. 33.4 The result of simulation of a suspension with a cross-shaped element beam width of 5 microns
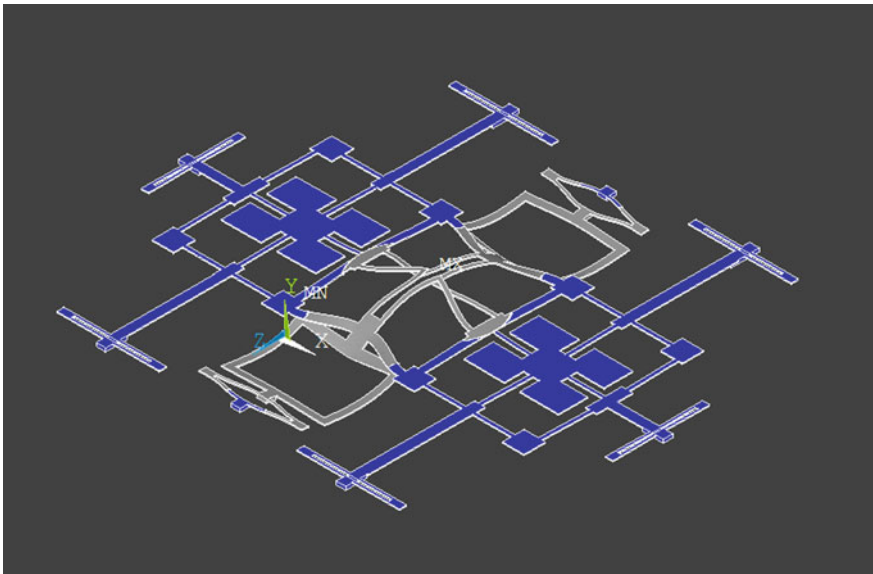
Fig. 33.5 The result of simulation of a suspension with a cross-shaped element beam width of 10 microns
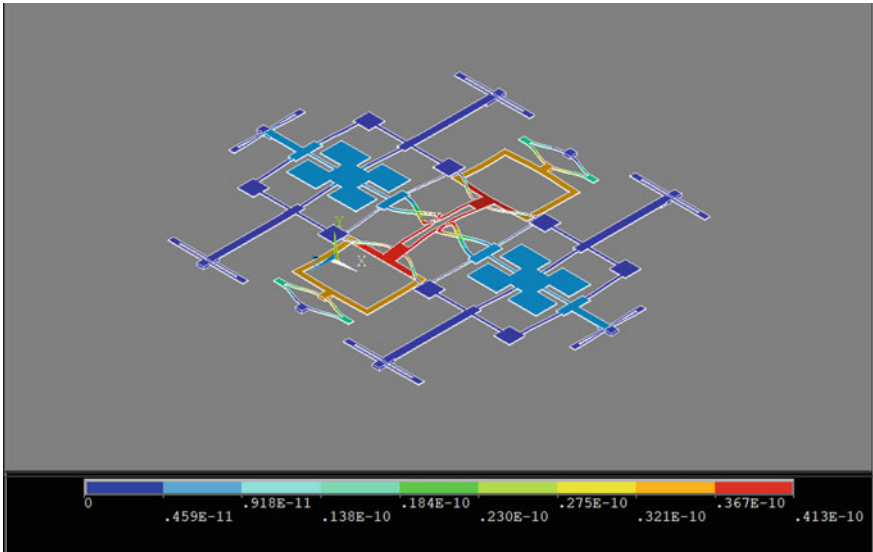
Fig. 33.6 The result of a simulation of a suspension with a cross-shaped element beam width of 2.5 microns
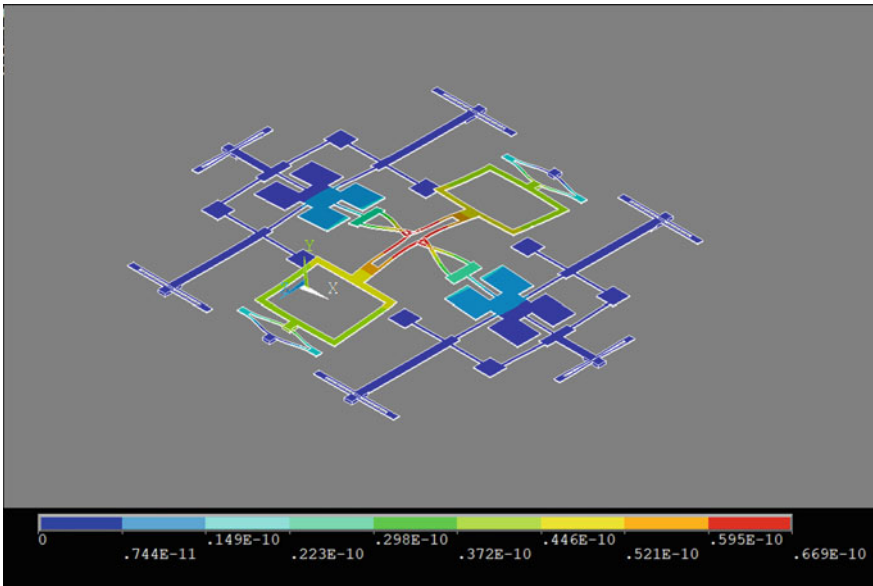

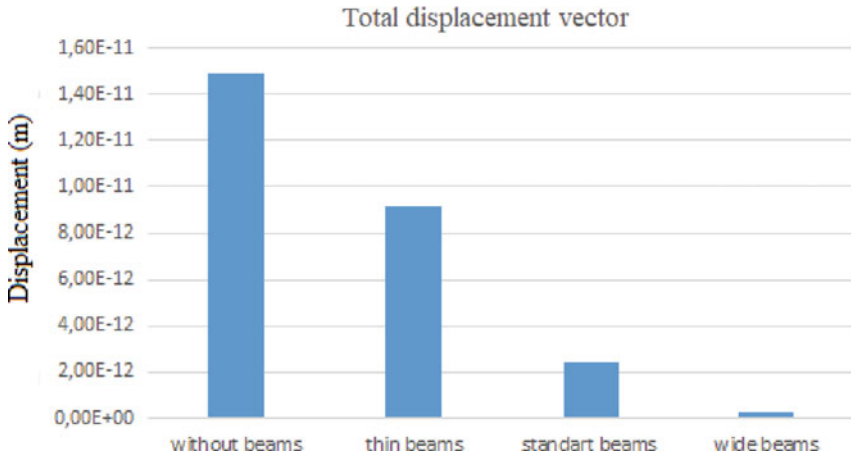**Fig. 33.7** The result of modeling a suspension without beams surrounding a cross-shaped element

Fig. 33.8 Results of modeling the suspensions of a micromechanical gyroscope-accelerometer

Also, for clarity, the suspension was modeled without beams, around the cross-shaped structure of a micromechanical gyroscope-accelerometer. Figure 33.7 shows the result of the simulation.

The simulation result shows the total displacement vector of the inertial masses is $1.49*10^{-12}$ m. This type of suspension showed the greatest result of the total displacement vector of the inertial masses which can be seen in Fig. 33.8.

The graph shows that a suspension without beams is able to move inertial masses further than a classic suspension by 5 times and 30% further than a suspension with thin beams.

From the simulation results, it can be seen that the suspension without beams has a strong deformation along the Y-axis. This can greatly affect the stable operation of the micromechanical gyroscope-accelerometer. The main vector of movement of inertial masses under the action of electrostatic drives in the given design is the vector along the X-axis. The movements of the inertial masses only along the X axis are shown in Fig. 33.9.

The graph shows that the suspension with beam thickness of 2.5 microns has the greatest movement of inertial masses relative to the rest.

33.4 Conclusion

In this article, four types of elastic suspension of a micromechanical gyroscope-accelerometer with three axis of sensitivity were presented. To obtain the results of modeling the suspensions presented, 3D models of a micromechanical gyroscope-accelerometer with different types of cross-shaped element were developed using the ANSYS software. The simulation results show that the greatest movement of
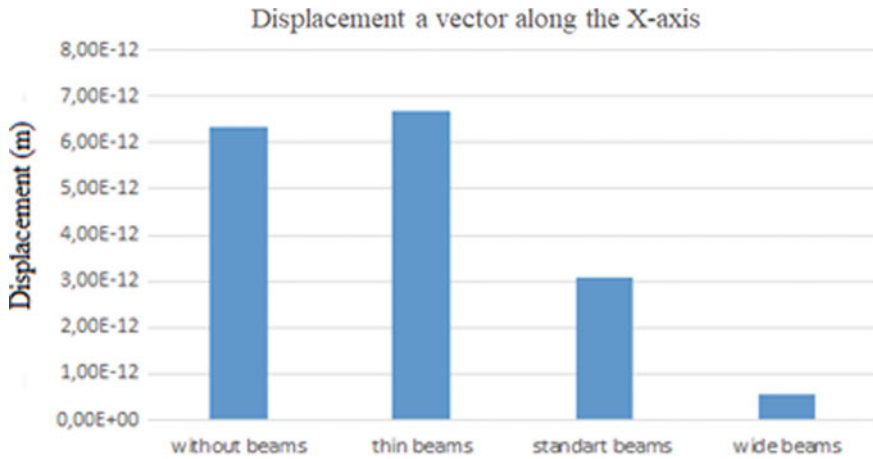
Fig. 33.9 The suspension movement along the X-axis vs beam width

inertial masses along all three axes has a suspension, whose cruciform element is not surrounded by beams. But we cannot use it to improve the design of the MEMS gyroscope-accelerometer since it has less stability in the Y-axis. The best option is a suspension with 2.5 μm thick beams surrounding the cross-shaped element of the MEMS gyroscope-accelerometer model. The advantage of this suspension is the movement of the inertial masses along the X-axis and sufficient stability of the movement along the Y-axis.

References

1. I.E. Lysenko, O.A. Ezhova, A.V. Tkachenko, D.V. Naumenko, K. Guha, K.S. Rao, The results of modelling of a micromechanical accelerometer, in *Journal of Physics: Conference Series* (2019)
2. V.. Ya.. Raspopov, *Micromechanical Devices* (Tula state University, Tula, 2007)
3. I.E. Lysenko, O.A. Ezhova, A.V. Tkachenko, D.V. Naumenko, Analysis of the micromechanical three-axis accelerometer, in *Book of Abstracts of 7th International School and Conference «Saint Petersburg OPEN 2020» on Optoelectronics, Photonics, Engineering and Nanostructures* (St. Petersburg Academic University, 2020)
4. S. Beeby, G. Ensell, M. Kraft, N. White, *MEMS Mechanical Sensors* (London Artech House, 2004)
5. I.E. Lysenko, A.V. Tkachenko, E.V. Sherova, A.V. Nikitin, Analytical approach in the development of RF MEMS switches. Electronics **7**(12), 1–23 (2018)

6. I.E. Lysenko, A.V. Tkachenko, O.A. Ezhova, Research of the microelectromechanical switch with different materials of metal membrane, in *Proceeding of SPIE, Proceeding of the International Conference Micro- and Nanoelectronics—2018 (ICMNE-2018)*, vol. 10226 (2019), pp 1–12
7. V. Rochus, Finite element modeling of strong electro-mechanical coupling in MEMS (University de Liege, 2006)
8. J. Taye, K. Guha, S. Baishya, Design and analysis of RF MEMS shunt capacitive switch for low actuation voltage & high capacitance ratio, in *Physics of Semiconductor Devices* (2014), pp. 445–448
9. M. Gyimesi, D. Ostergaard, Electro-Mechanical Transducer for MEMS Analysis in ANSYS, in *International Conference Modeling/Simulation of Microsystems (MSM99)*, vol. 7, (1999), pp. 270–273
10. P. Osterberg, H. Yie, X. Cai, J. White, S. Senturia, Self-consistent simulation and modeling of electrostatically deformed diaphragms, in *17th IEEE International Conference on Microelectromechanical System* (1994), pp. 28–32
11. R. Osiander, D. Garrison, J. Champion, MEMS and microstructures in aerospace applications (Taylor & Francis Group, 2006), p. 369
12. J. Korvink, O. Paul, in *MEMS: A Practical Guide to Design, Analysis, and Applications* (William Andrew Publishing, 2006)
13. Nanotechnology in microelectronics, ed. by O.A. Ageev, B.G. Konoplev, M.: NAUKA2019–511p ISBN 978-5-02-040201-0 (2019)
14. I.E. Lysenko, B.G. Konoplev, N.F. Kidyaev, S.I. Shafrostova, Patent for invention No. 2683810 "Integrated micro-mechanical gyroscope-accelerometer"
15. I.E. Lysenko, D.V. Naumenko, O.A. Ezhova, Design and simulation high aspect ratio torsion suspension of MEMS z-axis accelerometer, in *Book of Abstracts of 7th International School and Conference «Saint Petersburg OPEN 2020» on Optoelectronics, Photonics, Engineering and Nanostructures* (St. Petersburg Academic University, St. Petersburg, 2020)

Chapter 34
Research of the Influence of Elastic Suspensions on MEMS Gyroscope-Accelerometer Dynamics

Igor E. Lysenko and **Dmitry Y. Sevostyanov**

Abstract An original angular velocity and linear acceleration sensor design with three sensing axes is presented. To study the dynamic characteristics of the gyroscope-accelerometer, a parametrizable geometric model was created in the Ansys CAD system, and a finite element model was obtained. Three different designs of elastic suspensions were investigated. The gyroscope-accelerometer was modeled with a change in geometric parameters of each suspension. As a result of the simulation, the dependences of the natural vibration frequencies on the parameters of elastic suspensions were obtained. The results show that the gyroscope-accelerometer's natural frequencies along the X, Y and Z axes decrease as the length of the elastic suspension beams increases.

Keywords MEMS gyroscope · MEMS accelerometer · Finite element model · Gyroscope-accelerometer

34.1 Introduction

Currently, the market for microelectromechanical systems (MEMS) is growing, and applications are expanding [1, 2]. MEMS sensors have become an integral part of many systems, such as smartphones, unmanned aerial vehicles (drones), bionic prostheses, and control and navigation systems in aerospace vehicles [3, 4]. A significant part of MEMS sensors are micromechanical gyroscopes-linear acceleration sensors and accelerometers-angular velocity sensors. Today, the market is widely represented by one-and two-axis linear acceleration and angular velocity sensors, which are two separate devices in different cases [3, 4].

To reduce the cost and size of the final product, it is advisable to use gyroscopes-accelerometers, which are a single design with one or two sensitive elements. An important part of the sensing element is the elastic suspension that connects the inertial mass to the base of the micromechanical device. The geometric parameters of

I. E. Lysenko · D. Y. Sevostyanov (✉)
Southern Federal University, Taganrog 347900, Russia

the elastic suspension significantly affect the dynamic properties of the gyroscope-accelerometer. The aim of this work is to study the influence of the geometric parameters of elastic suspensions on the dynamic characteristics of the developed gyroscope-accelerometer.

34.2 Methods

The developed design of the micromechanical gyroscope-accelerometer has three types of elastic suspensions shown in Fig. 34.1. To connect the inertial mass to the electrostatic displacement converter, a suspension of five beams is used, shown in Fig. 34.1a. The suspension shown in Fig. 34.1b consists of two beams and connects the electrostatic displacement converter to the anchor. The suspension shown in Fig. 34.1c consists of three beams and also connects the electrostatic displacement converter to the anchor.

The ANSYS 19 CAD system was used for the research was used to conduct modal analysis and to obtain the dynamic characteristics of the developed gyroscope-accelerometer-natural oscillation frequencies. In it, a geometric model of the gyroscope-accelerometer was constructed and a finite element grid was superimposed, as shown in Fig. 34.2.

The obtained simulation results were approximated by the least squares method in MATLAB.
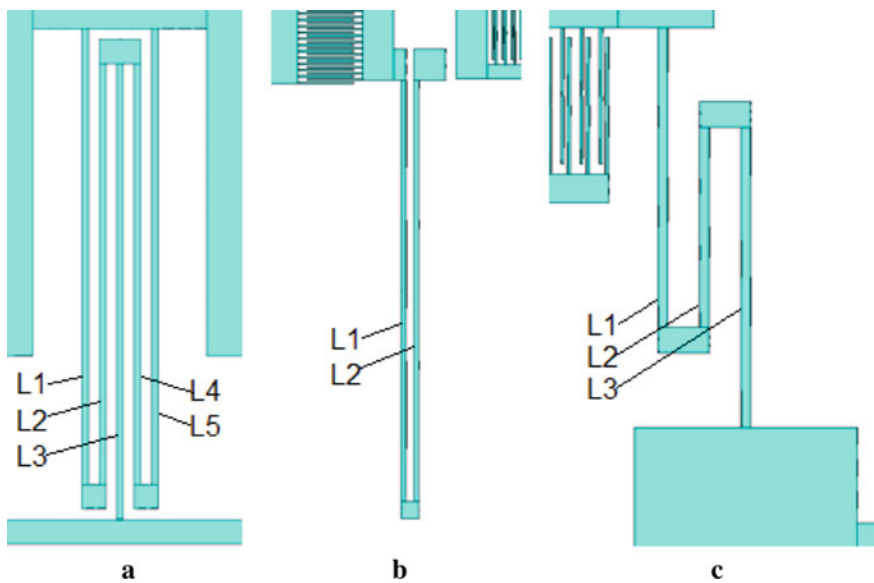
Fig. 34.1 Types of elastic suspensions: suspension of five (**a**), two (**b**), and three (**c**) beams
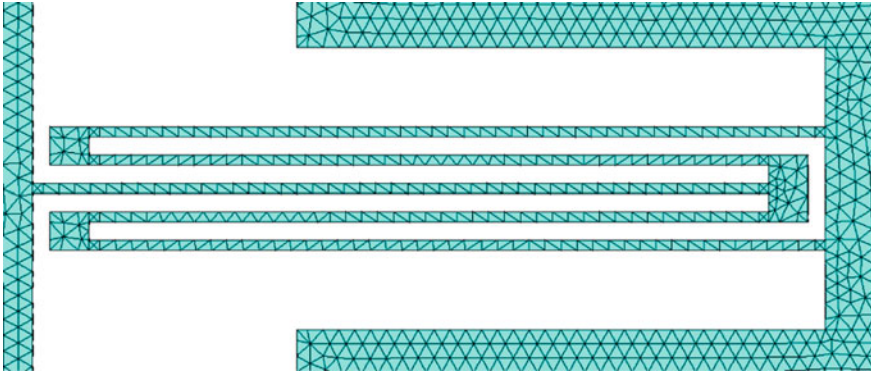
Fig. 34.2 Finite element model of a gyroscope-accelerometer

34.3   Results

The results of modeling the dynamic characteristics of a micromechanical gyroscope-accelerometer with a change in the geometric characteristics of a suspension consisting of five beams are shown in Fig. 34.3. Figure 34.3a shows the dependence of the natural vibration frequencies of the sensor element on the length of the beams *L2*, *L4* and *L3* of the elastic suspension along the X-, Y- and Z-axis. Figure 34.3b shows the dependence of the natural vibration frequencies on the length of the elastic suspension beams *L1*, *L5*, *L2*, *L4* and *L3* along the X-, Y- and Z-axis. Figure 34.3b shows the dependence of the natural vibration frequencies on the length of the elastic suspension beams *L1*, *L5*, and *L2*, *L4* along the X-, Y- and Z-axis. Figure 34.3 shows the dependence of the natural vibration frequencies on the thickness of the elastic suspension beams along the X-, Y- and Z-axis.

The results of modeling the dynamic characteristics of a micromechanical gyroscope-accelerometer with a change in the geometric characteristics of a suspension consisting of two beams are shown in Fig. 34.4. Figure 34.4a shows the dependence of the natural vibration frequencies on the length of the beams *L1* and *L2* of the elastic suspension along the X-, Y- and Z-axis. Figure 34.4b shows the dependence of the natural vibration frequencies on the thickness of the elastic suspension beams along the X-, Y- and Z-axis.

The results of modeling the dynamic characteristics of a micromechanical gyroscope-accelerometer with a change in the geometric characteristics of a suspension consisting of two beams are shown in Fig. 34.5. Figure 34.5a shows the dependence of the natural vibration frequencies on the length of the elastic suspension beams *L1*, *L3*, and *L2* along the X-, Y- and Z-axis. Figure 34.5b shows the dependence of the natural vibration frequencies on the length of the elastic suspension beams *L2* and *L3* along the X-, Y- and Z-axis. Figure 34.5b shows the dependence of the natural vibration frequencies on the thickness of the elastic suspension beams along the X-, Y- and Z-axis.
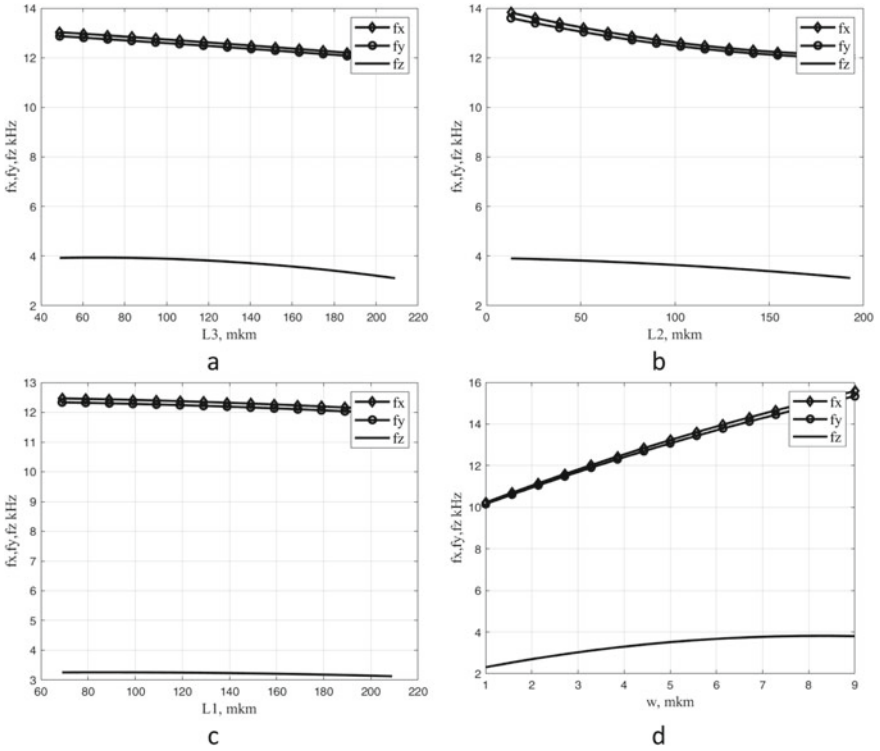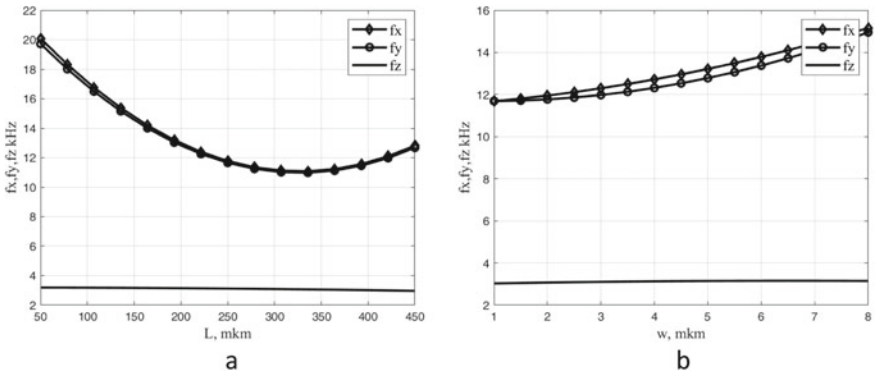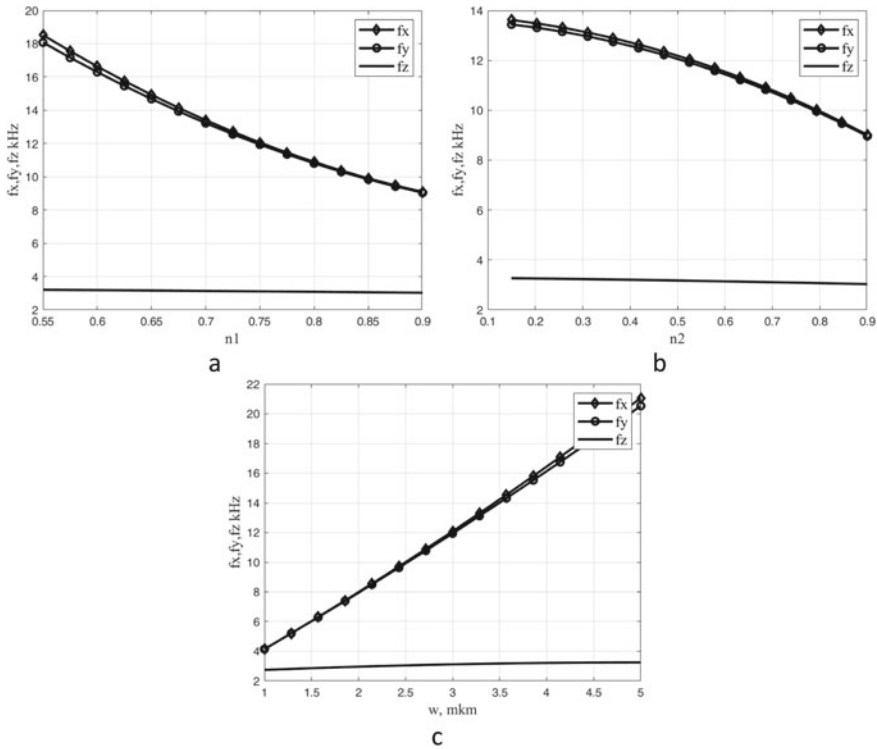
Fig. 34.3 The dependence of the natural vibration frequencies of the sensor element on changes in the geometric parameters of the suspension of five beams: length of the beam *L3* (**a**), length of the beam *L2* (**b**), length of the beam *L1* (**c**), and structural layer thickness *w* (**d**)

Fig. 34.4 The dependence of the natural vibration frequencies of the sensing element on the change in the geometric parameters of the suspension of two beams: the length of the beam *L* (**a**) and the structural layer thickness *w* (**b**)

Fig. 34.5 The dependence of the natural vibration frequencies of the sensor element on changes in the geometric parameters of the suspension of three beams: the beam length coefficient *n1* (**a**), the beam length coefficient *n2* (**b**), and from the structural layer thickness w (**c**)

Analyzing the results obtained, it can be concluded that an increase in the length of the beams in all types of suspensions leads to a decrease in the natural vibration frequencies, especially along the X- and Y- axis. An increase in the thickness of the structural layer of suspensions leads to an increase in natural frequencies, especially along the X- and Y- axis.

34.4   Conclusion

In this work, the design and simulation of a micromechanical gyroscope-accelerometer were carried out. The influence of the geometric parameters of elastic suspensions on the dynamic characteristics of the gyroscope-accelerometer is investigated. The simulation results showed that an increase in the width of the elastic suspension beams leads to a decrease in the stiffness of the elastic suspensions and,

consequently, to a decrease in the natural vibration frequencies. Changing the thickness of the structural layer of the suspension also has a significant impact on the dynamic characteristics of the suspension. The proposed model is suitable for testing various designs of elastic suspensions to obtain optimal dynamic characteristics. The results obtained can be useful for developers of micromechanical devices.

References

1. S. Sysoeva, Automobile gyroscopes. Compon. Technol. **1**, 58–66 (2007)
2. G. Krishnan, U. Kshirsagar, Micromachined high–resolution accelerometers. J. Indian Inst. Sci. **87**, 333–361 (2007)
3. S. Timoshenkov, Application of MEMS–sensors in navigation and orientation systems of mobile objects. Izvestiya vuzov. Electron. **6**, 51–56 (2012)
4. D. Pavlov, Development of a mathematical model of a MEMS accelerometer. Bull. Novgorod State Univer. **8**(91), 22–25 (2015)

Chapter 35
Adjustable Current Source for Pulse Electrochemical Deposition Installation

Wei Cao

Abstract This paper presents a printed circuit board with a controlled current source for pulse electrochemical deposition of thin films. The developed current source consists of an analog-to-digital converter, a digital-to-analog converter, a buffer amplifier, a transistor, and is controlled by Arduino Mega board through the designed LabVIEW program. The measurement results of the designed current source are presented. The source provides an output current of up to 70 mA. The duration of time intervals is regulated with 1 ms increments and provides the required operating modes.

Keywords Pulse electrochemical deposition · Current source · Arduino · LabVIEW

35.1 Introduction

Today's economy is developing rapidly. Due to people's inadequate anticipation of the negative effects of highly developed industries and unfavorable prevention, there are three major global crises: energy shortage, environmental pollution, and ecological destruction. Energy is the material basis of human activities. The development of human society is inseparable from the emergence of high-quality energy and the use of advanced energy technologies. Therefore, the most urgent task is to vigorously develop green energy to alleviate these three crises. It is urgent to develop green energy [1].

In the past few decades, energy harvesting technology, which converts natural dissipating energy into usable energy, has been studied to help mitigate energy depletion around the globe, and thus various methods to harvest this energy have been suggested. In particular, solid-state conversion of heat to electricity is expected to be a key energy harvesting technology in the future. Thermoelectric energy harvesting

W. Cao (✉)
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: tsao.v@edu.spbstu.ru

can replace and does not require batteries in wireless networks for different applications [2]. Thermoelectric power generation technology is an energy harvesting technology developed based on the Seebeck effect of thermoelectric materials.

There are various ways of producing materials for thermoelectric generators. One of them is electrochemical deposition, which requires an external current source. The development of such cheap current source is the main goal of this work.

35.2 Production of Thermoelectric Materials

A thermoelectric generator (TEG) is a semiconductor device that converts the temperature difference into an electrical energy due to the [3–5]. Thermoelectric materials used to generate thermoelectric power are mainly semiconductor materials such as Bi_2Te_3, SiGe, and MnTe.

The thermoelectric power generation equipment made of these thermoelectric materials is small in size, light in weight, no vibration, no noise, long service life, and the advantage of long-term operation in extremely harsh environment is very suitable for a variety of uncontrolled sensors, satellite power supplies, beacons and navigation signs, as well as for areas of medical and physiological research [6]. TEGs are classified into two types: volumetric and thin-film devices. Thin-film TEGs are manufactured using solid-state semiconductor technologies compatible with the technology of manufacturing microelectromechanical systems (MEMS) [2].

In [5], we learned that due to its technical compatibility with silicon and high thermal stability, the application of semiconductor silicide may be one of the most important stages to improve the performance of thermoelectric generators. Therefore, the TEG of Uni-Leg structure is designed. The premise of the TEG structure is that the thermoelectric elements of two n-type and p-type thermoelectrically active semiconductor materials are interchangeable (TE), and they can be thermally connected in parallel and electrically connected in series (π structure), Or TE is made of only thermoelectrically active material semiconductor material-single thermocouple (Uni-Leg), as shown in Fig. 35.1. The Uni-Leg structure simplifies the structure of the thermoelectric generator; therefore, it reduces the cost of off-the-shelf products [7]. Therefore, it can be used in daily life.

Among thermoelectric materials, N-type bismuth telluride (Bi_2Te_3) and P-type antimony telluride (Sb_2Te_3) are attractive because of their high characteristics for thermoelectric generation applications near room temperature [8–12]. There are many methods for the synthesis of these materials. Pulsed electrochemical deposition is one of the widely used methods for producing thick films [13].
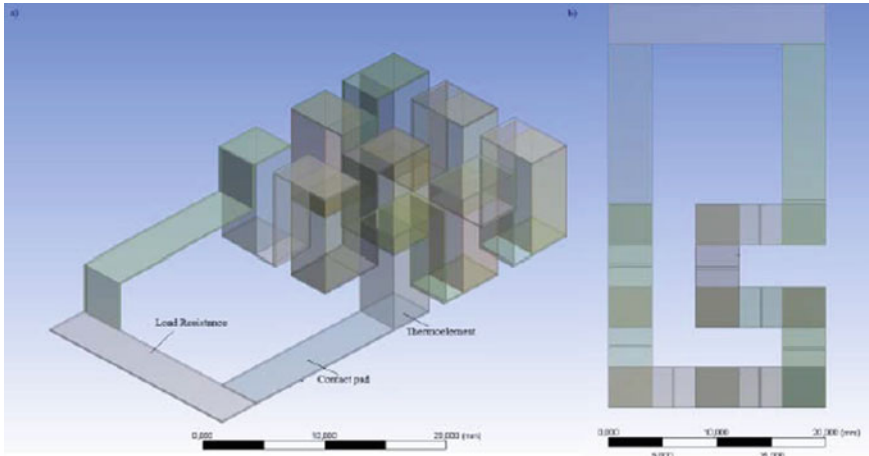
Fig. 35.1 3D geometry model of a Uni-Leg TEG [7]

35.3 System hardware

This design uses Arduino Mega 2560 from Arduino to build the main controller. Arduino Mega 2560 is a micro control board based on ATmega2560. It has 54 digital input/output ports (15 of which can be used as PWM outputs), 16 analog input ports, and 4 UART interface, 16 MHz crystal oscillator, USB port, battery port, ICSP header and reset button [14]. Arduino Mega 2560 sends and receives data through MAX5533 and the serial port [15].
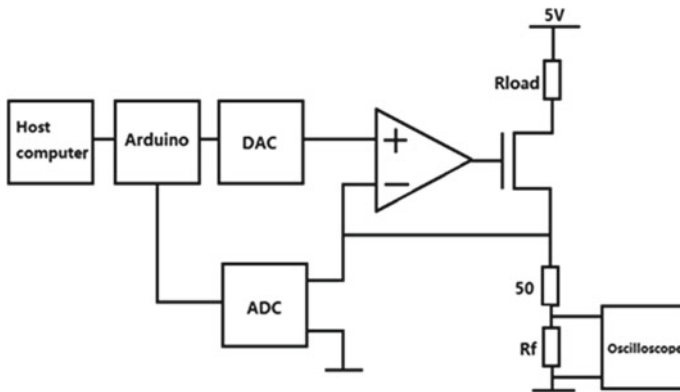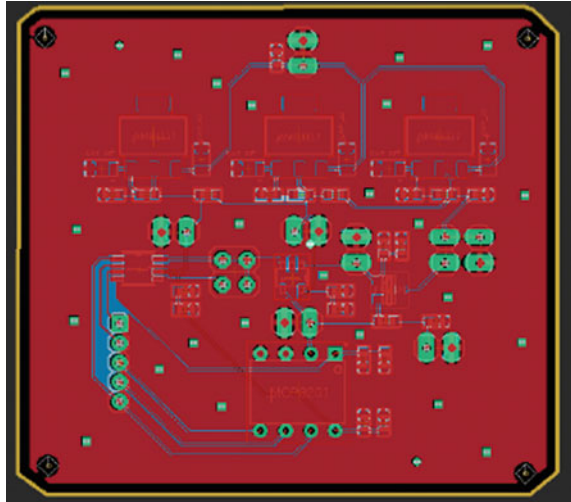

Fig. 35.2 The block diagram of the developed current source

Fig. 35.3 Final PCB

The block diagram of the developed current source is shown in Fig. 35.2 and consists of a digital-to-analog converter, an operational amplifier, a field-effect transistor, an analog-to-digital converter and several resistors. The transistor plays the role of a current source controlled through a voltage follower on the operational amplifier using a digital-to-analog converter, the output voltage of which is set by the Arduino board. The voltage across the resistors connected to the drain of the transistor, proportional to the generator output current, controlled by a high-precision analog-to-digital converter.

The resistance Rf is the feedback resistance, and Vin+ and Vin− are the operational amplifiers. In this design integrated circuits MAX5533, MCP3201 and MCP606 are used. The current source is controlled in the developed program in the LabVIEW environment through the Arduino Mega 2560 (Fig. 35.3).

35.4 System Software and Measurement Results

In [13], which describes the synthesis and study of thick films of electro-chemically deposited thermoelectric materials Bi_2Te_3 and Sb_2Te_3, some information is presented on the required parameters of the source for a pulsed electro-chemical deposition unit, among which the range of generated currents up to 35 mA and applied voltages up to 1.5 V. Synthesis of a thick film in this method is performed using an alternating current. The applied potentials used in pulsed deposition alternate between ON and OFF intervals. In the ON interval, potential Eon is applied to the sample and the flowing current is controlled for film growth, as in conventional electrochemical deposition. Meanwhile, the current is maintained at 0 mA during the OFF interval.

In this case, it should be possible to set the pulse duration equal to 100 ms for the ON state and 200 ms for the OFF state.

The user interface of the designed LabVIEW program is shown in Fig. 35.4. Though this interface user can control time intervals and maximum output current of the source due to the internal 2-bit controlled reference of the digital-to-analog converter. Thus, the maximum current is 23.8 mA.

The designed controlled current source has been tested at room temperature (Fig. 35.5). Figure 35.6 shows the oscillogram of the developed current source output signal. X-axis scale: 100 ms/div, Y-axis scale: 5 mV/div.

Figure 35.6 was obtained in high resolution mode. In peak detect mode we can see some digital noise that will be eliminated during future work.
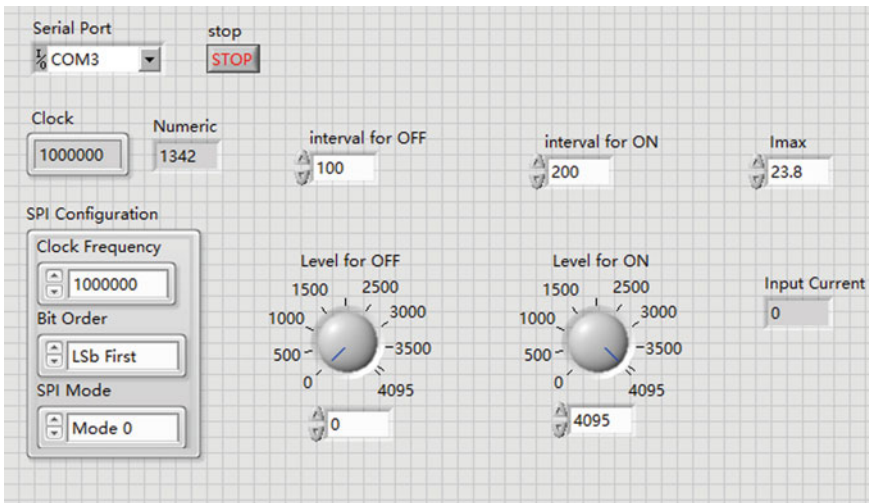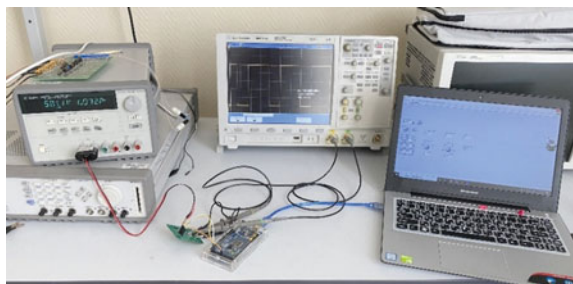

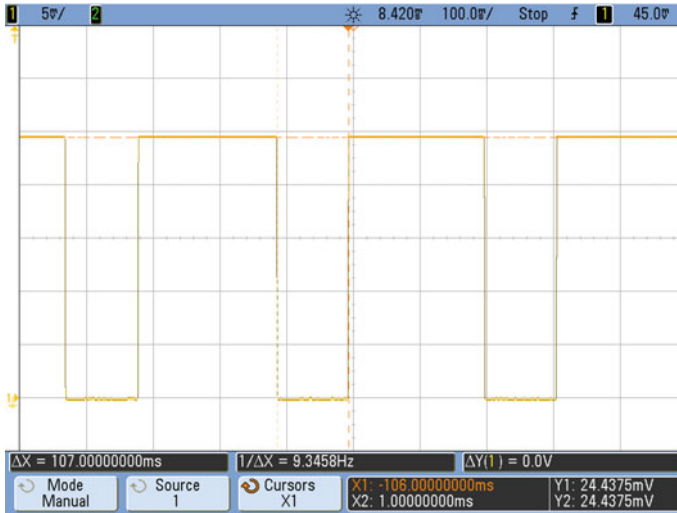**Fig. 35.4**   Front panel layout

Fig. 35.5   Measurement setup

Fig. 35.6 Output signal of controlled current source

35.5 Conclusion

The paper presents designed adjustable current source for a pulsed electrochemical deposition unit with the following parameters:

- the output current of the source is up to 24 mA when the DAC reference source is operating with a minimum voltage and up to 70 mA with a maximum voltage;
- durations of time intervals are adjustable in 1 ms steps and provide the required operating mode.

References

1. D. Ahuja, M. Tatsutani, Sustainable energy for developing countries. SAPIENS Surv. Persp. Integrating Environ. Soc. **2**(1) (2009)
2. A.S. Korotkov, V.V. Loboda, S.V. Dzyubanenko, E.M. Bakulin, Design of a thin-film thermoelectric generator for low-power applications. Russ. Microlectron. **48**(5), 326–334 (2019)
3. R. Buslaev, A. Galitskaya, V. Loboda, Simulation of flexible thermoelectric generators based on Bi$_2$Te$_3$/Sb$_2$Te$_3$ synthesized by electrochemical deposition method, in *2019 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)* (IEEE, Piscataway, NJ, 2019), pp. 54–57
4. A. Korotkov, V. Loboda, S. Dzyubanenko, E. Bakulin, Fabrication and testing of MEMS technology based thermoelectric generator, in *2018 7th Electronic System-Integration Technology Conference (ESTC)* (IEEE, Piscataway, NJ, 2018), pp. 1–4

5. A.S. Korotkov, V.V. Loboda, Simulation and design of thin-film thermoelectric generators, in *2018 International Symposium on Fundamentals of Electrical Engineering (ISFEE)* (IEEE, Piscataway, NJ, 2018), pp. 1–4

6. S.V. Volvenko, D. Ge, S.V. Zavjalov, A.S. Gruzdev, A.V. Rashich, E.L. Svechnikov, Experimental wireless ultra wideband sensor network for data collection, in *2017 Progress in Electromagnetics Research Symposium-Spring (PIERS)* (IEEE, Piscataway, NJ, 2017), pp. 965–970

7. R. Buslaev, V. Loboda, Simulation of Uni-Leg thermoelectric generator, in *2018 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)* (IEEE, Piscataway, NJ, 2017), pp. 27–31

8. L.M. Goncalves, P. Alpuim, G. Min, D.M. Rowe, C. Couto, J.H. Correia, Optimization of Bi_2Te_3 and Sb_2Te_3 thin films deposited by co-evaporation on polyimide for thermoelectric applications. Vacuum **82**(12), 1499–1502 (2008)

9. K. Kaspar, U. Pelz, H. Hillebrecht, Polyol synthesis of nano-Bi_2Te_3. J. Electron. Mater. **43**(4), 1200–1206 (2014)

10. C. Boulanger, Thermoelectric material electroplating: a historical review. J. Electron. Mater. **39**(9), 1818–1827 (2010)

11. H. You, S.H. Baek, K.C. Kim, O.J. Kwon, J.S. Kim, C. Park, Growth and thermoelectric properties of Bi_2Te_3 films deposited by modified MOCVD. J. Cryst. Growth **346**(1), 17–21 (2012)

12. D. Huo, J. Liang, H. Li, S. Xie, Y. Yu, Application of electro-chemical analysis technology in the research of metal electrodeposition. J. Casting Tech. **12**, 153–161 (2016)

13. N.H. Trung, K. Sakamoto, N.V. Toan, T. Ono, Synthesis and evaluation of thick films of electrochemically deposited Bi_2Te_3 and Sb_2Te_3 thermoelectric materials. Materials **10**(2), 154 (2017)

14. S.F. Barrett, Arduino microcontroller processing for everyone! Syn. Lect. Dig. Circ. Syst. **8**(4), 1–513 (2013)

15. C.C. Zhih, S.K.V. Ragavan, M. Shanmugavel, Development of a simple, low-cost autopilot system for multi-rotor UAVs, in *2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS)* (IEEE, Piscataway, NJ, 2015), pp. 285–289

Chapter 36
Influence of Contact Material on Photoinduced Current Flow in Si/TiO$_2$/GNPs/TiO$_2$ Nanocomposites

Alexandr Vorobyev, Yakov Enns, Anastasia Kondrateva, Polina Bespalova, Ivan Komarevtsev, and Maxim Mishin

Abstract The aim of this work was to obtain Si/TiO$_2$/GNPs/TiO$_2$ nanocomposite coatings photoactive in the visible wavelength range. A method to increase the optical response of the oxide nanostructures is proposed. It combines a magnetron deposition of oxide matrix with the thermal deposition and aggregation of metallic nanoparticles, allowing the preparation of wide band gap oxide heterostructures with buried plasmonic Au nanoparticles. Thin 2D layers of TiO$_2$ was deposited by reactive magnetron sputtering, followed by 2.5 nm gold layer thermal deposition and N$_2$-atmosphere thermal annealing at 550 °C to induce coalescence of Au nanoparticles. Then Si/TiO$_2$/GNPs was covered with TiO$_2$ by reactive magnetron sputtering. For the electrical properties of the formed nanocomposites study, special samples were prepared on silicon substrates. Various metals were used as electrical contacts: Ti, V, Cr, Al, Zr, Au alloy Ti70-V30. The I–V characteristics of different ohmic and Shottky contacts were obtained in the presence and absence of VIS irradiation.

Keywords Heterojunction · Gold nanoparticles · Photo-generation · Nanocomposites · Transition metal oxides

36.1 Introduction

In recent years, oxygen-containing semiconductor materials has been paid an increasing attention [1]. This interest is due to their wide distribution and the relative simplicity and breadth of the spectrum of methods for their formation. In optical applications, titanium dioxide plays a special role [2]. However, its widespread usage

A. Vorobyev (✉) · Y. Enns · A. Kondrateva · P. Bespalova · I. Komarevtsev · M. Mishin
Alferov St. Petersburg National Research Academic University, St. Petersburg 194021, Russia

Y. Enns · A. Kondrateva · P. Bespalova
Peter the Great St. Petersburg Polytechnic University, St. Petersburg 195251, Russia

I. Komarevtsev
Ioffe Physical-Technical Institute of the Russian Academy of Sciences, St. Petersburg 194021, Russia

is limited by the fact that the intrinsic absorption peak of the material is in the UV region [3, 4].

The introduction of plasmon activators, for example, gold nanoparticles (GNPs), into the matrix of wide-gap material allows us to translate the material fundamental absorption edge into the visible region [5, 6] and control the type and magnitude of the conductivity of the material [7].

Transition metal oxides are actively used as catalysts in gas chemistry and in fuel cells, sensing elements of sensor systems, etc. They are also potentially interesting for widespread use in various electronic devices. Including not only as traditional insulating and passivating layers and transparent electrodes, but also as active regions with substantially nonlinear properties. In particular, composite thin-film coatings based on transition metal oxides TiO_2 [5, 8], Fe_2O_3 [9, 10], Cu_2O, NiO [7, 11] are of interest for practical application. The potential use of such materials is limited by the fact that most of them are wide-gap n-type semiconductors, the optical absorption spectrum of which lies outside the visible range.

Controlling optical transparency, electrical conductivity, and carrier lifetimes is challenging. Growth technique, process parameters and reagent purity have the most significant influence on the result. Additional opportunities for controlled control of properties, in addition to doping with impurities, is the introduction of metal nanoparticles into the matrix material, that is, the creation of composite materials. In ensembles of nanoparticles, resonant excitation of surface plasmon oscillations at discrete frequencies of the visible region of the spectrum can manifest itself, which ensures strong absorption of light in this range. The characteristics of surface plasmon resonance are determined both by the geometric parameters of nanoinclusions (shape, diameter, relative position, array topology) and by the physicochemical properties (stoichiometry, crystallinity, surface morphology, elemental composition) of the matrix.

36.2   Methods

The object of investigation was a nanocomposite containing a matrix of titanium dioxide (TiO_2) and filler—gold nanoparticles (GNPs), formed on silicon substrates.

The nanocomposite was formed in several stages. During the first stage, a TiO_2 oxide coating 40 nm thick, obtained by reactive magnetron sputtering according to the method [12], was deposited on a single-crystal silicon substrate.

Further, a gold film was formed on the surface of the obtained oxides by the method of vacuum thermal spraying [5]. The thickness of the gold layer for all samples was 2 nm.

Then the plates were annealed for an hour at a temperature of 550 °C in a nitrogen atmosphere. This led to the formation on the surface of oxide coatings of a two-dimensional array of gold nanoparticles with a characteristic size of 10 nm. At the final stage, the particles were overgrown with a TiO_2 matrix 150 nm thick by the

Fig. 36.1 Schematic
representation of the sample
under study and the scheme
for measuring I–V
characteristic
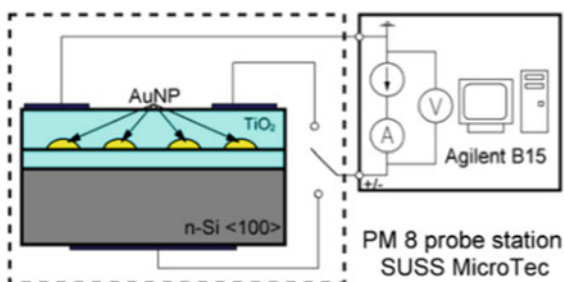
method of reactive magnetron sputtering. A schematic representation of the obtained nanocomposite is shown in Fig. 36.1.

36.3   Result and Discussion

To study the flow of the photoinduced current, the I–V characteristics of various samples were obtained in the presence and absence of irradiation.

Metallic contacts made of (Ti, V, Cr, Al, Zr, Ti70-V30) alloy with an area of 0.5 $\times$ 2.5 mm^2 were applied to the surface of the obtained nanocomposite by magnetron sputtering, as well as Au contacts by vacuum thermal sputtering at a distance of 6 mm from each other. from friend. An aluminum contact with an area of 8 mm$_2$ $\times$ 8 mm$_2$ was deposited on the back side of each sample (on the back side of the Si substrate). A schematic representation of the resulting sample is shown in Fig. 36.1.

The influence of various metals used as a contact to the material of the nanocomposite matrix was experimentally studied by measuring the current-voltage characteristics (VAC) under irradiation and in the absence of irradiation. The study was carried out using a PM 8 probe station (SUSS MicroTec) and an Agilent B1500A semiconductor analyzer. The I–V characteristics were measured in two modes: a bias was applied either in the plane of the nanocomposite (symmetric contact to the matrix material) or between the nanocomposite (matrix material) and the silicon substrate.

At positive bias, there is practically no response to radiation. Such behavior can be explained using the band diagram of the structure shown in Fig. 36.2, considering the values of the band gap and the electronic affinity of the materials included in the studied samples [13].

The TiO$_2$-Si contact is a heterojunction (Fig. 36.2a), at the boundary of which a barrier is formed. It can be assumed that the value of the barrier at the transition is very small due to the proximity of the location of the bottom of the conduction band of TiO$_2$ ($E_c \sim -4.21$ eV) and Si ($E_c \sim -4.51$ eV). The discontinuity of the conduction bands in the case of Si is of the order of $\Delta E_c \sim 0.3$ eV [13].

At a positive bias, electrons from silicon are injected into TiO$_2$. Even at small displacements, the barrier at the Si-TiO$_2$ interface disappears, and the entire applied voltage falls mainly on TiO$_2$. The current through the sample is determined by
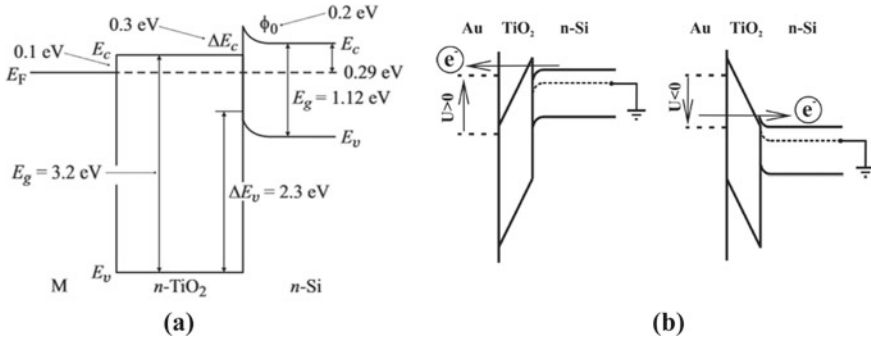
Fig. 36.2 Zone diagram of the silicon-titanium dioxide contact with no external bias **a** [13] and upon positive and negative bias **b**, applied [5]

processes in TiO_2 [5]. In the absence of effective charge separation at the Si-TiO_2 interface, there is practically no effect on irradiation.

For all used contact materials, the dependencies are the same. Minor differences in current and threshold voltages are due to features occurring at the contact-TiO_2 interface.

Let us consider the features of the I–V characteristic when applying reverse bias. Figure 36.3 shows typical current-voltage characteristics of samples with different types of contacts at reverse bias.

Based on the obtained I–V characteristics, it can be concluded that different materials of the contacts have a different effect on the photoinduced current generated in the samples under the action of irradiation.

However, the nature of the dependencies for all contact materials is similar. Under irradiation, a sharper increase in the dependence of current on voltage is presumably

Fig. 36.3 Comparative I–V characteristics of samples with different contact materials at negative bias
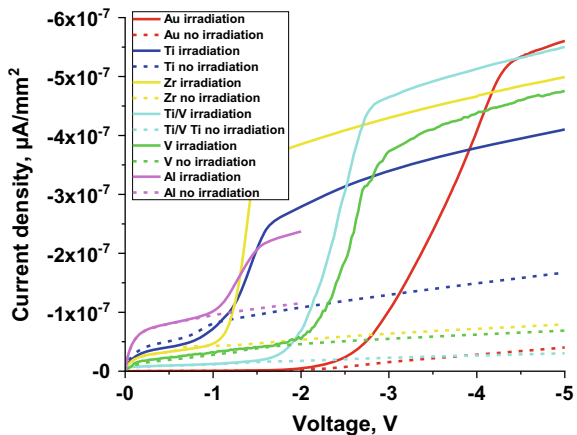
associated with the plasmon absorption of light by gold particles and the transition of electrons from GNPs to TiO$_2$ [8].

To describe the processes of photoinduced current flow, a semi-empirical model was developed that describes the processes of charge generation under the action of light, as well as charge transfer in the nanocomposite [5].

With a negative bias (in the absence of irradiation), the current flow corresponding to the reverse branches of the I–V characteristic is due to the injection of electrons from the contact material into TiO$_2$ and then into Si. The low concentration of intrinsic carriers in TiO$_2$ and the reverse displacement of the Schottky barrier between the supply contact and TiO$_2$ cause weak reverse currents in the absence of irradiation [5]. In this case, a space charge region (SCR) is formed in Si, and the voltage drop is distributed between TiO$_2$ and Si. The reverse current through the structure is mainly determined by generation—recombination processes near the SCR [14, 15].

Upon irradiation, a sharp increase in the current is observed at reverse bias. This is due to the plasmon absorption of light by gold particles and the transition of electrons from gold particles to titanium dioxide [2].

A characteristic feature of all the obtained I–V characteristics is the presence of two regions (inflections), which are two different barriers at the contact-TiO$_2$ and TiO$_2$-Si interfaces. Since the differences in the samples are determined only by the contact material, it is possible to draw conclusions about the use of the contact material, which makes it possible to achieve the best photoelectric properties of the nanocomposite, namely, the highest degree of sensitivity under irradiation.

Based on Fig. 36.3, it is possible to draw a conclusion about the effectiveness of various materials used as contacts to the nanocomposite under study. The results are shown in Fig. 36.4 and in Table 36.1.

The fact that the process of photogeneration of charge carriers and the increase in current under the action of irradiation are caused by plasmon absorption of light by gold particles can be confirmed by the practically identical level of current saturation when using different materials of contacts.

The position of the first inflection on the I–V characteristic, presumably, corresponds to different heights of the Schottky barrier at the contact-TiO$_2$ interface.

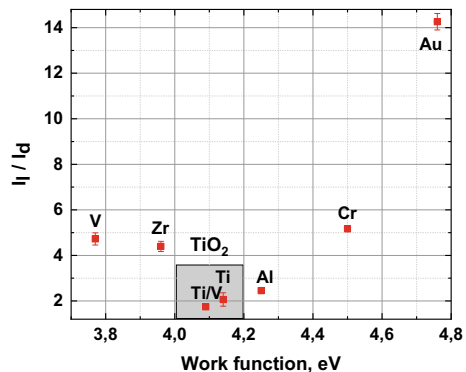**Fig. 36.4** Ratio of photocurrent to shadow current vs the work function of the upper contact material

Table 36.1 The influence of
the contact material on the
response magnitude

Material	Work function, eV	I_I/I_d
V	3.77	4.7 ± 0.3
Zr	3.96	4.4 ± 0.2
Ti70-V30 alloy	4.09	1.7 ± 0.1
Ti	4.17	2.1 ± 0.3
Al	4.3	2.5 ± 0.1
Cr	4.5	5.2 ± 0.1
Au	4.76	14.26 ± 0.36

Fig. 36.5 Opening voltage
(first bend) and voltage of the
current output to saturation
(second bend) vs difference
between the work functions
of contact material and TiO_2
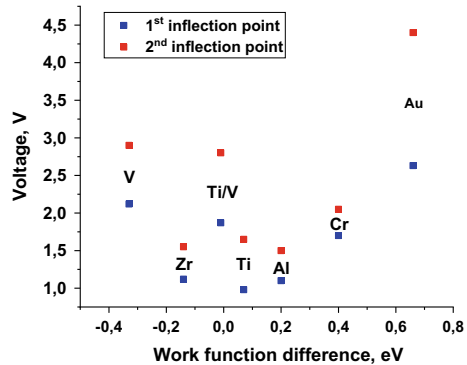
Figure 36.5 shows the dependences of the opening voltage (the first inflection point) and the saturation current output voltage (the second inflection point) on the difference between the work functions of the contact material and TiO_2.

It can be concluded that the higher the barrier at the contact-TiO_2 interface, the greater the displacement is required to activate the photogeneration process.

36.4   Conclusions

It was shown that the heterostructures fabricated exhibit the high optical response under light irradiation in visible range. The structures formed were characterized by sufficient conductivity, the ratio of light and dark currents reached 5/1 for Si/TiO_2/GNPs/TiO_2. Various metals (Ti, V, Cr, Al, Zr, Au, Ti70-V30 alloy) were used as electrical contacts. The I–V characteristics of different ohmic and Shottky contacts were obtained in the presence and absence of VIS irradiation.

References

1. R. Pampin, et al.. Electrodes-oxide-semiconductor device for biosensing: renewed conformal analysis and multilayer algorithm. J. Electroanal. Chem. **856**, 113651 (2020)
2. F. Li, et al., Electrospun TiO_2//SnO_2 Janus nanofibers and its application in ethanol sensing. Mater. Lett. **262**, 127070 (2020)
3. S. Shablaev, R. Pisarev, Giant nonlinear absorption in the NiO antiferromagnet. Phys. Solid State **45**(9), 1742–1745 (2003)
4. P. Tikhonov et al., Investigation into the sensory properties of fine-grain polycrystalline films based on cobalt, nickel, and praseodymium oxides. Glass Phys. Chem. **31**(5), 700–708 (2005)
5. M. Mishin, et al., The mechanism of charge carrier generation at the TiO_2-n-Si heterojunction activated by gold nanoparticles. Semicond. Sci. Technol. **33**(7), 075014 (2018)
6. I. Fasaki et al., Optimized hydrogen sensing properties of nanocomposite NiO: Au thin films grown by dual pulsed laser deposition. Sens. Actuators B Chem. **176**, 103–109 (2013)
7. C. Etrich et al., Effective optical properties of plasmonic nanocomposites. Materials **7**(2), 727–741 (2014)
8. Z. Zhao et al., Three-dimensional plasmonic Ag/TiO_2 nanocomposite architectures on flexible substrates for visible-light photocatalytic activity. Sci. Rep. **7**(1), 1–11 (2017)
9. P. Bespalova et al., Characterization of iron oxide coatings prepared by MOCVD method from Fe $(CO)_5$. Mater. Today Proc. **30**, 434–438 (2020)
10. A. Vorobyev et al., Controlled formation of iron oxide nanoparticles by pulse-modulated RF discharge at atmospheric pressure. Mater. Today Proc. **30**, 417–421 (2020)
11. K. Nakamura et al., Properties of plasmon-induced photoelectric conversion on a TiO_2/NiO p–n junction with Au nanoparticles. J. Phys. Chem. Lett. **7**(6), 1004–1009 (2016)
12. A. Kondrateva et al., Electrooptical properties of TiO_2 doped with gold nanoparticles. Semiconductors **54**(14), 1885–1888 (2020)
13. V. Kalygina et al., Mechanism of conductivity in titanium oxide thin films and structures metal-TiO_2-Si. Semiconductors **50**(6), 1036–1040 (2016)
14. V. Gaman, Physics of semiconductor devices. NTL, Tomsk (2000) [in Russian]
15. E. Gol'Dman, A. Zhdan, Generation of electron-hole pairs at the surface of a semiconductor in the fringing fields of macroscopic charged clusters: effects of electrical inhomogeneities of various length scales. Semiconductors **29**(3), 219–224 (1995)

Chapter 37
Interband Absorption in Asymmetric Biconvex Lens-Shaped Quantum Dot

Mher A. Mkrtchyan

Abstract In this manuscript, an analytical model has been considered, describing the behavior of particles localized in an asymmetric biconvex lens-shaped quantum dot. The interband absorption has been investigated in strong size quantization regime. The dependencies of the interband absorption and photoluminescence coefficients on incident light energy have been plotted.

Keywords Lens-shaped quantum dot · Adiabatic approximation · Interband absorption · Photoluminescence

37.1 Introduction

Quantum dots (QDs) are objects with nontrivial optical properties and on the other hand, these structures can play important role for many applied fields such as optical sensing, nonlinear optics, fiber optics, laser manufacturing [1–3].

The most QDs geometries lead to calculation problems and description of energy spectra and wavefunctions in these systems. Therefore, the most studied and applied ones are spherical [4, 5] and cylindrical [6, 7] geometries. For such geometries, in most cases we can give analytical solutions of Schrödinger equation. On the other hand, QDs with more complex geometries have been theoretically considered and experimentally investigated. These include conical [8, 9], ellipsoidal [10, 11], and lens-shaped [12–15] QDs geometries. For such geometries we need to use various numerical methods [16–18]. QDs with lens-shaped geometry allow the use of adiabatic approximation.

In this paper, asymmetric biconvex lens-shaped QD with impenetrable walls is investigated. The interband absorption and photolumencence (PL) coefficients dependencies on incident light energy are plotted.

M. A. Mkrtchyan (✉)
Russian–Armenian University, Yerevan, Armenia
e-mail: mher.mkrtchyan@rau.am

37.2 Theory

Let us consider an electron localized in asymmetric biconvex lens-shaped QD. For confining potential we can write

$$\widehat{V}_{conf}(\rho, z) = \begin{cases} 0, & inside\ ABLQD \\ \infty, & outside\ ABLQD \end{cases}. \tag{37.1}$$

In the size quantization regime, the Coulomb interaction between particles can be neglected. Thus, the problem is reduced to the one-particle particle motion. Now we can present system Hamiltonian in cylindrical coordinates as the sum of radial and axial ones

$$\begin{cases} H_{\rho} = -\frac{\partial^2}{\partial\rho^2} - \frac{1}{\rho}\frac{\partial}{\partial\rho} - \frac{1}{\rho^2}\frac{\partial^2}{\partial\varphi^2} \\ H_z = -\frac{\partial^2}{\partial z^2} + \widehat{V}_{conf}(\rho, z) \end{cases}. \tag{37.2}$$

Along axial direction the particle is localized in the one-dimensional quantum well with following boundaries

$$\begin{cases} z_1 = \sqrt{R_2^2 - \rho^2} + R_2 - h_2 \\ z_2 = \sqrt{R_1^2 - \rho^2} + h_1 - R_1 \end{cases}, \tag{37.3}$$

where h_1 and h_2 are semi-axes for each convex part of asymmetric biconvex lens-shaped QD, R_1 and R_2 are the curvature radii of two spheres intersection.

This problem has been discussed in [12] and for axial wavefunction and energy spectrum, we can write

$$\begin{cases} \chi(z, \rho) = \sqrt{\frac{2}{a(\rho)}} \sin\left(\frac{\pi n_z}{a(\rho)}\right) z \\ E_{n_z}^z = \frac{\pi^2 \hbar^2}{2\mu a^2(\rho)} n_z^2 \end{cases}, \tag{37.4}$$

where $a(\rho) = \sqrt{R_2^2 - \rho^2} + \sqrt{R_1^2 - \rho^2} + (h_1 + h_2) - (R_1 + R_2)$ is one-dimensional quantum well length, $n_z = 1; 2; 3; \ldots$ is axial quantum number. For wavefunction and energy spectrum one we have

$$\begin{cases} f(\rho) = C\rho^{|m|} e_1^{-\frac{\lambda\rho^2}{2}} F_1\left(-n_\rho, |m| + 1; \lambda\rho^2\right) \\ E_{n_\rho, m}^\rho = \hbar\Omega\left(2n_\rho + |m| + 1\right) \end{cases} \tag{37.5}$$

where $\lambda = \frac{\mu \Omega}{\hbar}$, $\Omega^2 = \frac{R_1 + R_2}{2\mu(h_1 + h_2)^3 R_1 R_2} \pi^2 \hbar^2 n_z^2$ $n_\rho = 0; 1; 2; \ldots$ is the radial quantum number, m is magnetic quantum number C is the wavefunction normalization constant, $_1F_1(a, b; c)$ is the confluent hypergeometric function.

37.3 Results

Asymmetric biconvex lens-shaped QD made of *InAs* have been considered. For the absorption coefficient, we have the following expression

$$K(\hbar\omega) = K_0 \cdot \omega \left(\int \Psi_e(\vec{r}_e) \Psi_h(\vec{r}_h) dV \right)^2 \sum_{\nu_e, \nu_h} \delta \left(\hbar\omega_{e,h} - E_g - E^e_{\nu_e} - E^h_{\nu_h} \right), \quad (37.6)$$

where K_0 is the quantity proportional to the square of the modulus of the matrix element of the dipole moment, taken over the Bloch functions, $\Psi_e(\vec{r}_e)$, $\Psi_h(\vec{r}_h)$ are electron and hole wavefunctions, E_g is band gap, $E^e_{\nu_e} \left(E^h_{\nu_h} \right)$ is electron (hole) energy, $\nu_e(\nu_h)$ is set of quantum numbers for electron (hole), $\hbar\omega_{e,h}$ is incident light energy.

Figures 37.1 and 37.2 show the dependences of absorption coefficient on the energy of incident light for the diagonal $\left| m, n_\rho, n_z \right\rangle \to \left| m', n'_\rho, n'_z \right\rangle$ transitions, in the regime of strong size quantization, when an electron and a hole are in identical states. As it can be seen from the figures, intensity of the absorption coefficient decreases with increasing the axial quantum numbers for electron and light hole. Also, with the temperature decrease, absorption peaks linearly decreases for both $\left| 0, 0, 1 \right\rangle \to \left| 0, 0, 1 \right\rangle$ and $\left| 0, 0, 2 \right\rangle \to \left| 0, 0, 2 \right\rangle$ cases. It should be noted that we see broadening of peaks on all dependences. This happens because $\Gamma(T)$ is directly proportional to temperature.

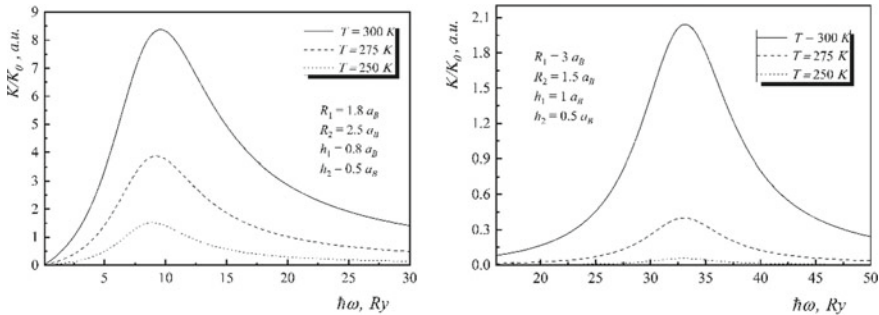Note that we have following selection rules for quantum numbers

Fig. 37.1 Dependence of the interband absorption coefficient on the frequency of incident light for the light hole to electron diagonal $\left| 0, 0, 1 \right\rangle \to \left| 0, 0, 1 \right\rangle$ (left) and $\left| 0, 0, 2 \right\rangle \to \left| 0, 0, 2 \right\rangle$ (right) transitions

Fig. 37.2 Dependence of the interband absorption coefficient on the frequency of incident light for the heavy hole to electron diagonal $|0, 0, 1\rangle \rightarrow |0, 0, 1\rangle$ (left) and $|0, 0, 2\rangle \rightarrow |0, 0, 2\rangle$ (right) transitions
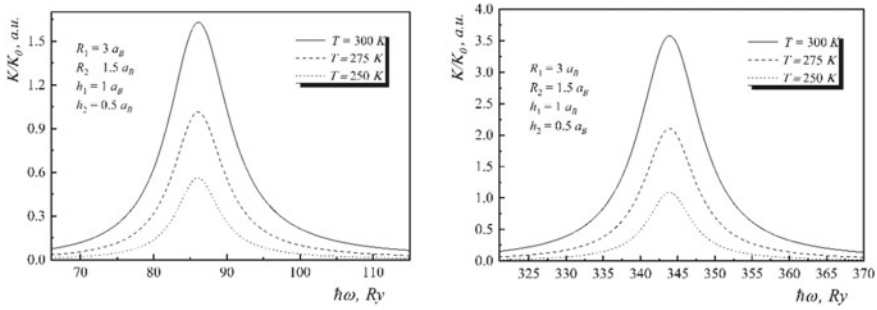
$$
\begin{cases}
n_r \rightarrow n'_r, \\
m \rightarrow -m', \\
n_z \rightarrow n'_z,
\end{cases}
\tag{37.7}
$$

for both light hole to electron and heavy hole to electron transitions.

The PL coefficient are calculated using the relation

$$
R(\hbar\omega) = R_0 \cdot \frac{1}{kT} \cdot \hbar\omega \cdot K(\hbar\omega) \cdot \frac{f_c(1 - f_v)}{f_c - f_v},
\tag{37.837.8}
$$

where R_0 is the quantity proportional to the square of the modulus of the matrix element of the dipole moment, taken over the Bloch functions, k is Boltzmann constant, f_c and $(1 - f_v)$ are the probabilities that the state of the conduction band is filled and the state of the valence band is empty, respectively.

Figures 37.3 and 37.4 show the dependences of PL coefficient on the energy of incident light in the regime of strong size quantization. As it can be seen from the figures, the intensity of the PL increases for both light hole to electron and heavy hole to electron transitions with temperature. It should be noted that we see broadening of peaks on all graphs. This happens because $\Gamma(T)$ in absorption coefficient is directly proportional to temperature.

37.4 Conclusion

In present article, behavior of electrons localized in an asymmetric biconvex lens-shaped quantum dot in external electric field has been studied. The confinement

Fig. 37.3 Dependence of the PL coefficient on the frequency of incident light for the light hole to electron transitions
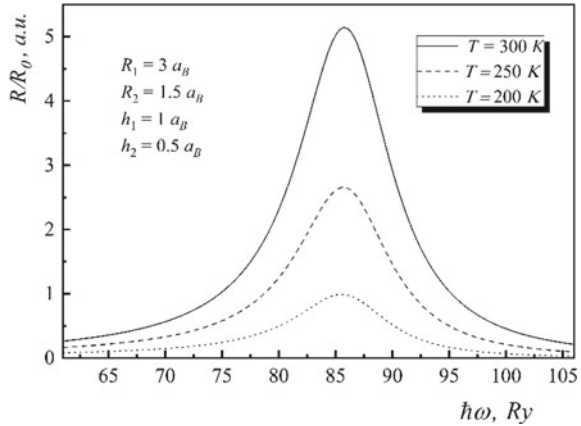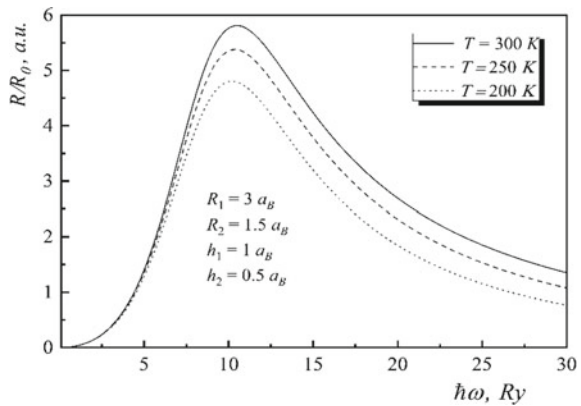

Fig. 37.4 Dependence of the PL coefficient on the frequency of incident light for the heavy hole to electron transitions

potential is considered as infinite. Interband absorption and PL coefficients dependencies on incident light frequency are plotted for the strong size quantization regime.

References

1. D. Bimberg, M. Grundmann, N.N. Ledentsov, *Quantum Dot Heterostructures* (Wiley, 1998)
2. P. Harrison, A. Valavanis, *Quantum Wells, Wires and Dots: Theoretical and Computational Physics of Semiconductor Nanostructures* (Wiley, 2016)
3. T. Chakraborty, *Quantum Dots: A Survey of the Properties of Artificial Atoms* (Elsevier, 1996)
4. V.M. Fomin, V.N. Gladilin, J.T. Devreese, E.P. Pokatilov, S.N. Balaban, S.N. Klimin, Photoluminescence of spherical quantum dots. Phys. Rev. B **57**(4), 2415 (1998)
5. I. Karabulut, S. Baskoutas, Linear and nonlinear optical absorption coefficients and refractive index changes in spherical quantum dots: effects of impurities, electric field, size, and optical intensity. J. Appl. Phys. **103**(7), 073512 (2008)

6. W.S. Li, C.Y. Chen, Electron-phonon interaction in a cylindrical quantum dot. Physica B **229**(3–4), 375–382 (1997)
7. C.H. Liu, B.R. Xu, Theoretical study of the optical absorption and refraction index change in a cylindrical quantum dot. Phys. Lett. A **372**(6), 888–892 (2008)
8. D.B. Hayrapetyan, E.M. Kazaryan, H.A. Sarkisyan, Magneto-absorption in conical quantum dot ensemble: possible applications for QD LED. Opt. Commun. **371**, 138–143 (2016)
9. D.B. Hayrapetyan, A.V. Chalyan, E.M. Kazaryan, H.A. Sarkisyan, Direct interband light absorption in conical quantum dot. J. Nanomater. **2015** (2015)
10. E. Sadeghi, A. Avazpour, Binding energy of an off-center donor impurity in ellipsoidal quantum dot with parabolic confinement potential. Physica B **406**(2), 241–244 (2011)
11. T. Chen, W. Xie, S. Liang, The nonlinear optical rectification of an ellipsoidal quantum dot with impurity in the presence of an electric field. Physica E **44**(4), 786–790 (2012)
12. L.S. Petrosyan, Electron states in a biconvex thin quantum lens in the presence of an external homogeneous magnetic field. J. Contemp. Phys. **40**, 28 (2005)
13. M.R.K. Vahdani, G. Rezaei, Linear and nonlinear optical properties of a hydrogenic donor in lens-shaped quantum dots. Phys. Lett. A **373**(34), 3079–3084 (2009)
14. H.A. Sarkisyan, D.B. Hayrapetyan, L.S. Petrosyan, E.M. Kazaryan, A.N. Sofronov, R.M. Balagula, D.A. Firsov, L.E. Vorobjev, A.A. Tonkikh, Realization of the Kohn's theorem in Ge/Si quantum dots with hole gas: theory and experiment. Nanomaterials **9**(1), 56 (2019)
15. A.H. Rodriguez, C. Trallero-Giner, S.E. Ulloa, M.-J. Antuña, Electronic states in a quantum lens. Phys. Rev. B **63**(12), 125319 (2001)
16. E.K. Gross, R.M. Dreizler, *Density Functional Theory* (Springer Science & Business Media, 2013)
17. I. Babuška, T. Stroubouli, *The Finite Element Method and Its Reliability* (Oxford University Press, Oxford, 2001)
18. K.G. Dvoyan, E.M. Kazaryan, L.S. Petrosyan, Electronic states in quantum dots with ellipsoidal symmetry. Physica E **28**(4), 333–338 (2005)

Chapter 38
Binding Energy of Magnetobiexciton in Ellipsoidal Quantum Dot

Yuri Y. Bleyan

Abstract The theoretical investigation of biexciton in ellipsoidal quantum dot in the presence of the magnetic field is done in the scope of Heisenberg's uncertainty principle. The dependency of the energy for the magnetobiexciton on the magnetic field is calculated. The magnetobiexciton energy increases with the increase of the magnetic field magnitude. The dependence of the magnetobiexciton binding energy on the magnetic field is computed and the binding energy decreases with the increase of magnetic field value. Finally, the magnetic field threshold values, depending on the geometrical parameters of the quantum dot, after of which the biexciton stability is disrupted, are obtained.

Keywords Magnetobiexciton · Ellipsoidal quantum dot · Binding energy · Heisenberg's uncertainty principle

38.1   Introduction

Reduction in size with the strong electronic and optical properties make the semiconductor quantum dots (QD) as promising candidates for optoelectronic devices. The QDs have potential applications in the research areas of high-density data storage, quantum information processing, quantum computing and quantum cryptography [1]. Since the motion of charge carriers is localized in three directions in QDs, the electronic (energy, binding energy, recombination energy (difference between the total energies of the initial and final states after recombination) etc.) and optical (oscillator strength, absorption coefficient, photoluminescence etc.) properties can be manipulated and controlled by changing its size and the barrier material. Especially interesting problem is the investigation of four-particle systems in QDs.

The biexciton is a neutral compound quasiparticle, consisting of two electrons and two holes. The conversion of excitons to biexciton complexes leads to various

Y. Y. Bleyan (✉)
Russian–Armenian University, Yerevan, Armenia
e-mail: yuri.bleyan@rau.am

optical and other interesting effects. Some of that effects are photoluminescence emission peaks shift, sharp increase in two-photon absorption, change in the nonlinear susceptibility [2–7].

Due to the size quantization effect the binding energy of the biexciton is increased, as a result of which low-dimensional biexcitons have much higher stability [8]. There are several ways for biexciton's observation: two-photon absorption of light from the ground state to the biexciton state, pump-probe experiments etc.

In former times the magnetic field effects on excitonic complexes have been neglected as a magnetic field is technologically less significant than an electric field in the device applications. However, a magnetic field has two important advantages. The first advantage is that the magnetic field increases the exciton-biexciton binding energy by squeezing the electrons and holes together. Therefore, a magnetic field can cause improved nonlinear optical properties. Secondly, it offers far richer insight into the physics of excitonic complexes, especially for low-dimensional systems [9].

Spherical, spheroidal and ellipsoidal QDs are in the spotlight of the investigators as the ensembles of these structures can be easily synthesized and controlled by the help of technological parameters [10–13]. In spheroidal and ellipsoidal QDs the generalized Kohn theorem can be implemented [14, 15]. As the ellipsoidal QD has two geometrical parameters (small and large semiaxes) it gives an opportunity to control the energy spectra in a wide range [8].

Therefore, the theoretical investigation of biexciton in ellipsoidal QD in the presence of magnetic field is an actual problem.

38.2 Theory

Let's consider the biexciton states in strongly oblate ellipsoidal quantum dot (SOEQD) with impenetrable walls in the presence of magnetic field (magnetobiexciton). The potential energy of the particle in the SOEQD in cylindrical coordinates can be written as:

$$U_{conf}(\rho, \varphi, Z) = \begin{cases} 0, & \frac{\rho^2}{a^2} + \frac{Z^2}{c^2} \leq 1 \\ \infty, & \frac{\rho^2}{a^2} + \frac{Z^2}{c^2} > 1 \end{cases}, \quad a^2 \gg c^2 \tag{38.1}$$

where c and a, respectively, the small and large semiaxes of the SOEQD [12]. All dimensionless lengths and energies in the paper are measured in the Bohr effective radius of the electron $a_e = \hbar^2 \varepsilon / m_e^* e^2$, the Rydberg effective energy $E_R = \hbar^2 / 2 m_e^* a_e^2$ of the electron, m_e^* is effective mass of the electron and ε is the dielectric constant.

The Hamiltonian of the four interacting particles (two electrons and two holes) in the presence of the magnetic field in cylindrical coordinates has the following form:

$$\widehat{H}(\vec{r}_1, \vec{r}_2, \vec{r}_a, \vec{r}_b) = \sum_j \frac{1}{2m_j^*}\left(\vec{P}_j + \frac{e}{s}\vec{A}_j\right)^2 + \sum_{\substack{i,j \\ i \neq j}} \frac{q_i q_j}{\varepsilon|\vec{r}_i - \vec{r}_j|} + \sum_j U_{conf}(\vec{\rho}_j, z_j)$$

$$(38.2)$$

where $j = \{1, 2, a, b\}$, $q_{1,2} = -|e|$, $q_{a,b} = |e|$, $\vec{r}_1$ and $\vec{r}_2$ are coordinates of electrons, $\vec{r}_a$ and $\vec{r}_b$ are coordinates of holes. Here $\vec{P}_j$ is the particle impulse operator, $\vec{A}_j$ — vector potential of the magnetic field, s is the light speed in vacuum, e is the electron charge. The calibration $A_\rho = 0$, $A_\varphi = \frac{1}{2}H\rho$, $A_z = 0$ ensures the parallelism of the magnetic field to the axial direction.

As a first step, let's estimate the two-dimensional magnetobiexciton energy for the ground level in the scope of the Heisenberg's uncertainty principle. We will assume that the electrons and holes are located on the vertices of the square.

This means that $\rho_1 = \rho_2 = \rho_a = \rho_b \equiv \rho$, $p_1 = p_2 = p_a = p_b \equiv p$ and for the distances between all the particles one can write:

$$|\rho_1 - \rho_a| = |\rho_2 - \rho_a| = |\rho_1 - \rho_b| = |\rho_2 - \rho_b| = \sqrt{2}\rho,$$
$$|\rho_1 - \rho_2| = |\rho_a - \rho_b| = 2\rho. \qquad (38.3)$$

By using Heisenberg's uncertainty principle $p \sim \hbar/\rho$ and making some simplifications for the energy of the system, one can present the energy in dimensionless magnitudes as

$$E(\rho) \sim \frac{2}{\rho^2}(1 + \sigma) + (1 + \sigma)\left(\frac{\pi^2}{2c^2}\right) + (1 + \sigma)\left(\frac{\pi^2}{2a^2c^2}\right)\rho^2$$
$$- \frac{(4\sqrt{2} - 2)}{\rho} + \frac{1}{2}\gamma^2\left(1 + \frac{1}{\sigma}\right)\rho^2 \qquad (38.4)$$

where $\sigma = \frac{m_e^*}{m_h^*}$, $\gamma = \frac{\hbar\omega_H}{2E_R}$, $\omega_H = \frac{eH}{m_e^* s}$. The condition for minimizing the energy of the system has the form:

$$\frac{dE(\rho)}{d\rho} = 0, \qquad (38.5)$$

After some transformations one can get the following equation for $\rho_{\min}$:

$$(1 + \sigma)\left(\frac{\pi^2}{a^2c^2}\right)\rho_{\min}^4 + \gamma^2\left(1 + \frac{1}{\sigma}\right)\rho_{\min}^4 + \left(4\sqrt{2} - 2\right)\rho_{\min} - 4(1 + \sigma) = 0$$

$$(38.6)$$

If solve the Eq. (38.6) numerically, one can find the dependence of the minimum radius $\rho_{\min}$ on the geometric parameters of the SOEQD. After that, by substituting

$\rho_{\min}$ value in Eq. (38.4), one can calculate the magnetobiexciton's minimum energy for the ground state.

38.3 Results and Discussion

Let us now move to the discussion of the obtained results. Figure 38.1 shows the dependence of the magnetobiexciton energy on the value of the magnetic field for different values of small semiaxis, when the value of the large semiaxis is fixed.

As can be seen from the figure the magnetobiexciton energy increases with the increase of magnetic field magnitude. Note, that value $\gamma = 1$ corresponds to the value of the magnetic field $H \approx 0.16T$.

It is worth to note that the dependence on the small semiaxis is stronger than on the large semiaxis. This is due to the fact that the size quantization in the axial direction prevails over the radial. That is why there are represented different values for the small semiaxis in Fig. 38.1.

Fig. 38.1 The dependence of the magnetobiexciton minimum energy of the ground state on the magnetic field for the fixed value of the large semiaxis
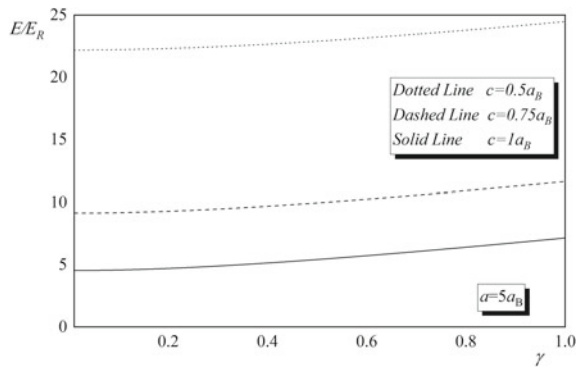
Fig. 38.2 The dependence of the magnetobiexciton binding energy of the ground state on the magnetic field for the fixed value of the large semiaxis
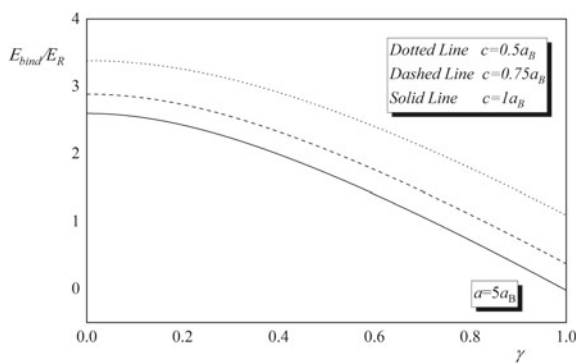
Table 38.1 The magnetic
field threshold values,
depending on the geometrical
parameters of the quantum
dot, after of which the
biexciton stability is broken
(i.e. the binding energy
becomes negative)

Large semiaxis	Small semiaxis	H, T
		XX
$a = 5a_B$	$c = 0.5a_B$	0.32214
	$c = 0.75a_B$	0.274272
	$c = 1a_B$	0.248522

In Fig. 38.2 the dependence of the magnetobiexciton binding energy on the
magnetic field for the different values of the small geometrical parameter of the
ellipsoid, when the large geometrical parameter is fixed, is obtained.

It is clear from the figure that the magnetobiexciton binding energy decreases with
the increase of the magnetic field. Also, it is evident that the lowest binding energy
has the magnetobiexciton, which has the highest value of the small semiaxis.

Finally, in the Table 38.1 the magnetic field threshold values, depending on the
geometrical parameters of the quantum dot, after of which the biexciton stability is
disrupted, have been shown.

38.4   Conclusion

In the present paper calculations were made for the biexciton in the ellipsoidal QD in
the presence of the magnetic field in the scope of Heisenberg's uncertainty principle.
The dependence of the magnetobiexciton's energy on the small semiaxis of the ellip-
soidal QD and the ratio of the effective masses of electron and hole is computed. The
magnetobiexciton binding energy is calculated as a function on the small geometrical
parameter (small semiaxis) and the magnetic field threshold values, depending on the
small and large semiaxes of the quantum dot, after of which the biexciton stability
is disrupted, have been shown.

References

1. D. Lalitha, A.J. Peter, C.W. Lee, Magneto-biexciton in a group III–V diluted magnetic quantum
   dot. Int. J. Mod. Phys. B **28**(27), 1450188 (2014)
2. L. Bányai, I. Galbraith, C. Ell, H. Haug, Excitons and biexcitons in semiconductor quantum
   wires. Phys. Rev. B **36**(11), 6099 (1987)
3. H. Wang, J. Shah, T.C. Damen, L.N. Pfeiffer, Polarization-dependent coherent nonlinear optical
   response in GaAs quantum wells: dominant effects of two-photon coherence between ground
   and biexciton states. Solid State Commun. **91**(11), 869–874 (1994)

4. T. Takagahara, Biexciton states in semiconductor quantum dots and their nonlinear optical properties. Phys. Rev. B **39**(14), 10206 (1989)
5. S. Baskoutas, A.F. Terzis, Biexciton Luminescence of PbS and PbSe colloidal quantum dots in E-MAA and phosphate glasses. J. Comput. Theor. Nanosci. **5**(1), 88–92 (2008)
6. S. Baskoutas, A.F. Terzis, Biexciton luminescence in InAs nanorods. J. Appl. Phys. **98**(4), 044309 (2005)
7. D.B. Hayrapetyan, Y.Y. Bleyan, D.A. Baghdasaryan, H.A. Sarkisyan, S. Baskoutas, E.M. Kazaryan, Biexciton, negative and positive trions in strongly oblate ellipsoidal quantum dot. Phys. E. **105**, 47–55 (2019)
8. Y.Y. Bleyan, D.B. Hayrapetyan, H.A. Sarkisyan, E.M. Kazaryan, Optical properties of biexcitons in ellipsoidal quantum dot, in *Quantum Technologies*, vol. 10674. (International Society for Optics and Photonics, 2018), p. 106741Q
9. A. Balandin, S. Bandyopadhyay, Magnetobiexcitonic states in a quantum wire. Phys. Rev. B **54**(8), 5712 (1996)
10. E. Sadeghi, A. Avazpour, Binding energy of an off-center donor impurity in ellipsoidal quantum dot with parabolic confinement potential. Phys. B **406**(2), 241–244 (2011)
11. A. Bagga, S. Ghosh, P.K. Chattopadhyay, Energy levels in spheroidal quantum dots with finite barrier heights. Nanotechnology **16**(11), 2726 (2005)
12. D.A. Baghdasaryan, D.B. Hayrapetyan, E.M. Kazaryan, Oblate spheroidal quantum dot: electronic states, direct interband light absorption and pressure dependence. Eur. Phys. J. B **88**(9), 223 (2015)
13. D.A. Baghdasaryan, D.A. Hayrapetyan, E.M. Kazaryan, Optical properties of narrow band prolate ellipsoidal quantum layers ensemble. J. Nanophotonics **10**(3), 033508 (2016)
14. D.B. Hayrapetyan, E.M. Kazaryan, H.A. Sarkisyan, Implementation of Kohn's theorem for the ellipsoidal quantum dot in the presence of external magnetic field. Phys. E. **75**, 353–357 (2016)
15. D.B. Hayrapetyan, E.M. Kazaryan, H.A. Sarkisyan, On the possibility of implementation of Kohn's theorem in the case of ellipsoidal quantum dots. J. Contemp. Phys. **48**(1), 32–36 (2013)

Chapter 39
Nonlinear Optical Absorption in Strongly Prolate Conical Quantum Dot in an External Electric Field

Khachik S. Khachatryan

Abstract Linear and nonlinear absorption coefficients in a strongly prolate conical quantum dot in the presence of external axial electric field have been considered. Infinite confinement potential has been chosen. The selection rules for intraband transitions have been obtained. The generation coefficients of the second and third harmonics depending on the photon energy were investigated both in the absence and in the presence of external electric field.

39.1 Introduction

Quantum dots (QDs) are an interesting class in semiconductor nanostructures because in these structures the motion of charge carriers is restricted in all three directions [1–3]. This makes it possible to manipulate the energy levels of charge carriers by purposefully growing QDs. Therefore, QDs are promising materials in various fields of science: medicine, biology, optoelectronics and etc.

Some complexity in geometries of the QDs leads to the number of problems in the theoretical description. For the theoretical description of QDs, various approximate and numerical methods were used: variational method, perturbation theory, adiabatic method, DFT, finite element method [4–7].

The conical quantum dots (CQD) are nanostructures with nontrivial geometry, the analytic description of which is a difficult problem. In the case of strongly prolate or oblate CQDs the most suitable method for the description of such systems is the adiabatic approximation [8, 9].Theoretical study of physical processes in the CQD has been implemented in the various papers [10–14].

K. S. Khachatryan (✉)
Russian–Armenian University, Yerevan, Armenia
e-mail: khachik.khachatryan@rau.am

Semiconductor quantum dots have a very high nonlinear optical susceptibility compared to bulk semiconductors. Therefore, the theoretical study of nonlinear optical effects in QDs gains importance [15–18].

In this article nonlinear optical absorption has been considered for the case of strongly prolate CDQ. Influence of the external electric field on the optical properties CQD was taken into account. The second harmonic generation (SHG) and third harmonic generation (THG) dependencies from incident light energy have been plotted.

39.2 Theory

Let us consider a strongly prolate CQD in the presence of the external axial electric field. The confining potential of QD one can choose in the following form [9, 14]:

$$V_{conf}(x, y, z) = \begin{cases} 0, \, particle \in CQD, \\ \infty, \, particle \notin CQD. \end{cases} \tag{39.1}$$

The Hamiltonian of the problem has the following form

$$\widehat{H}(x, y, z) = -\frac{\hat{p}^2}{2\mu} + V_{conf}(x, y, z) - eFz, \tag{39.2}$$

where F is electric field strength.

This problem of strongly prolate CQD in external electric field has been considered in [14]. For the radial energy spectra have been obtained

$$E_{n_\rho,m}(z) = \frac{\hbar^2 \lambda_{n_\rho,m}^2}{2\mu R^2(z)}. \tag{39.3}$$

The total wave function of the system was sought in the following form

$$\Psi(\rho, \varphi, z) = \frac{e^{im\varphi}}{\sqrt{2\pi}} f(\rho; z) \chi(z), \tag{39.4}$$

where $f(\rho; z) = \frac{\sqrt{2}}{R(z)} \frac{J_m(\lambda_{n_\rho,m} \rho/R(z))}{\{J_{m+1}(\lambda_{n_\rho,m})\}}$ is the radial part of wavefunction, $\chi(z)$ is axial one. In the frame of adiabatic approximation axial energy spectrum have been calculated.

$$E_{n_z} = \frac{\hbar^2 \lambda_{1,0}^2}{2\mu R_0^2} \left\{ \frac{1}{\left(1 - \frac{z_0}{H}\right)^2} - \frac{2\frac{z_0}{H}}{\left(1 - \frac{z_0}{H}\right)^3} \right\} - \left(\frac{\hbar^2}{2\mu} \left\{ \frac{\hbar^2 \lambda_{1,0}^2}{\mu R_0^2 H \left(1 - \frac{z_0}{H}\right)^3} - eF \right\}^2 \right)^{\frac{1}{3}} \cdot \text{æ}_{n+1},$$

(39.5)

where $\lambda_{n_\rho,m}$ are zeros of Bessel function of the first kind, R_0 is base radius of CQD, H is CQD height, μ is particle effective mass, $\text{æ}_{n+1}(n = 0; 1; \ldots)$ are zeros of the Airy function, z_0 is determined numerically.

The wavefunction for the axial part can be written as

$$\chi(z) = C A_i \left(\frac{1}{\xi_0} \left(z - \frac{\vartheta}{D} \right) \right),$$

(39.6)

where C is the normalization constant, $A_i(t) = \frac{1}{\sqrt{\pi}} \int_0^\infty \cos\left(Ut - \frac{1}{3}U^3\right) dU$, $\xi_0 = \left(\frac{\hbar^2}{2\mu D} \right)^{\frac{1}{3}}$, $D = \beta - e\vartheta$, $\vartheta = E_{n_z} - \alpha$, $\alpha = U_0 \left\{ \frac{1}{\left(1 - \frac{z_0}{H}\right)^2} - \frac{2\frac{z_0}{H}}{\left(1 - \frac{z_0}{H}\right)^3} \right\}$, $\beta = \frac{2U_0}{H\left(1 - \frac{z_0}{H}\right)^3}$, $U_0 = \frac{\hbar^2 \lambda_{1,0}^2}{2\mu R_0^2}$.

39.3 Intraband Light Absorption

Now let us consider the direct intraband light absorption in the CQD. Analytical expressions for the linear and nonlinear optical absorption coefficient (AC) are obtained [17, 18]

$$\alpha^{(1)}(\omega) = \omega \cdot \sqrt{\frac{\eta}{\varepsilon_r}} \frac{\sigma \cdot \hbar\Gamma_{fi}}{\left(E_{fi} - \hbar\omega\right)^2 + (\hbar\Gamma_{fi})^2} \left|M_{fi}\right|^2,$$

(39.7)

$$\alpha^{(3)}(\omega, I) = -\omega \sqrt{\frac{\eta}{\varepsilon_r}} \left(\frac{I}{\varepsilon_0 n_r c} \right) \frac{\left|M_{fi}\right|^4 2\sigma \cdot \hbar\Gamma_{fi}}{\left[\left(E_{fi} - \hbar\omega\right)^2 + (\hbar\Gamma_{fi})^2\right]^2}$$

$$\times \left\{ 1 - \frac{\left|M_{ii} - M_{ff}\right|^2}{4\left|M_{fi}\right|^2} \left(\frac{\left(E_{fi} - \hbar\omega\right)^2 - \left(\hbar\Gamma_{fi}\right)^2 + 2E_{fi}\left(E_{fi} - \hbar\omega\right)}{\left(E_{fi}\right)^2 + \left(\hbar\Gamma_{fi}\right)^2} \right) \right\},$$

(39.8)

where η is the permeability of the system, σ is the electron density in a QD, $E_{fi} = E_f - E_i$ is energy difference between the final and initial states (f and i, respectively), I is the incident optical intensity, n_r is the refractive index, $\Gamma_{fi} = 1/\tau_{fi}$ is the relaxation rate for states f and i (corresponds to the full width at half maximum), $M_{fi} = \langle \Psi_f | ez | \Psi_i \rangle$ is the matrix element of dipole moment. The total absorption coefficient is as follows:

$$\alpha(\omega, I) = \alpha^{(1)}(\omega) + \alpha^{(3)}(\omega, I). \tag{39.9}$$

The expression of susceptibility to SHG in a three-level system is given by the expression [16–18]:

$$\chi^{(2)}(2\omega) = \frac{e^3 \sigma}{\varepsilon_0} \frac{M_{in_1} M_{n_1 f} M_{fi}}{\left(2\hbar\omega - E_{fi} - i\hbar\Gamma_{fi}\right)\left(\hbar\omega - E_{n_1 i} - i\hbar\Gamma_{n_1 i}\right)}. \tag{39.10}$$

The THG susceptibility in a four-level system is [16–18]

$$\chi^{(3)}(3\omega) = \frac{e^4 \sigma}{\varepsilon_0} \frac{M_{in_1} M_{n_1 n_2} M_{n_2 f} M_{fi}}{\left(3\hbar\omega - E_{fi} - i\hbar\Gamma_{fi}\right)\left(2\hbar\omega - E_{n_2 i} - i\hbar\Gamma_{n_2 i}\right)\left(\hbar\omega - E_{n_1 i} - i\hbar\Gamma_{n_1 i}\right)}. \tag{39.11}$$

Here n_1 and n_2 are the quantum numbers of intermediate levels.

39.4   Results and Discussion

The case of CQD made of *InAs* will be considered. The material parameters which have been used in the calculations are as follows:$\mu = 0.023\, m_0$, where m_0 is the free electron mass, $\varepsilon_r = 15.15$. Note that dependencies are plotted for $I = 2.5 \cdot 10^6\, \text{W/m}^2$ and $\sigma = 3 \cdot 10^{22}\, \text{m}^{-3}$ values. The broadening linewidth is taken to be 0.5 meV at 77 K and 5 meV at 100 K for *InAs* QD [18].

Figures 39.1 and 39.2 present the dependencies of the linear, nonlinear and total absorption coefficient on the energy of incident light in the external field, for different $|n_r, m, n_z|\rangle \rightarrow |n'_r, m', n'_z|\rangle$ transitions. Since the nonlinear absorption coefficient has the opposite sign to the linear one, the total absorption will be reduced. The calculations were performed for two different temperatures: $T = 77\,\text{K}$ (left) and
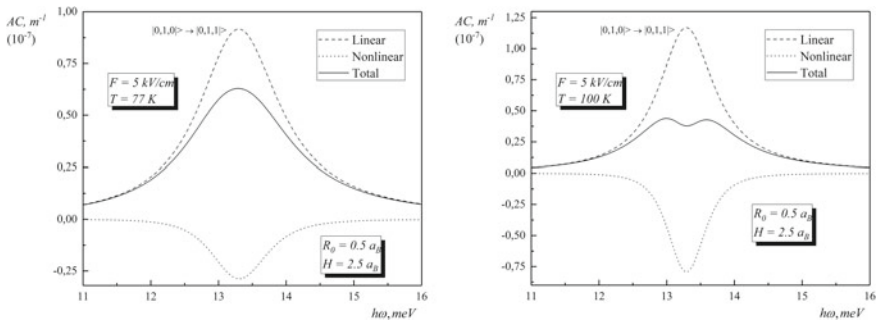
Fig. 39.1 Absorption coefficient vs incident light photon energy for $|0, 1, 0|\rangle \rightarrow |0, 1, 1|\rangle$ transition
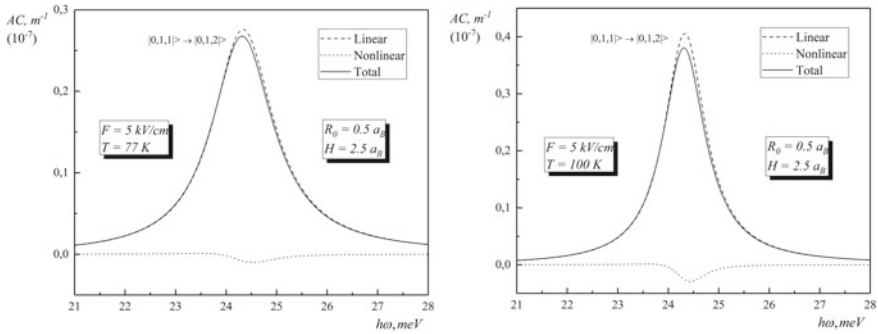
Fig. 39.2 Absorption coefficient vs incident light photon energy for $|0, 1, 1|\rangle \rightarrow |0, 1, 2|\rangle$ transition

$T = 100\,$K (right). As can be seen from the figures, with increasing temperature, the value of total absorption also increases. Note that as the temperature raises, the internal chaos of the system increases, therefore, linear and nonlinear optical absorption increases.

Figure 39.3 shows the dependencies of SHG on the energy of incident light for different electric fields. The calculations were performed for two different temperatures:$T = 77\,$K (left) and $T = 100\,$K (right). As can be seen from the figures, the SHG peaks increase with an increase in the eclectic field. Note that in all cases the intensity of the first peak is greater than that of the second.

Figure 39.4 presents the dependencies of THG on the energy of incident light for different electric fields. Calculations have been done for two different temperatures: $T = 77\,$K and $T = 100\,$K. It follows from the figure that with an increase in the electric field, the THG peaks increase. Note that in all cases the intensity of the first two peaks is greater than that of the third.
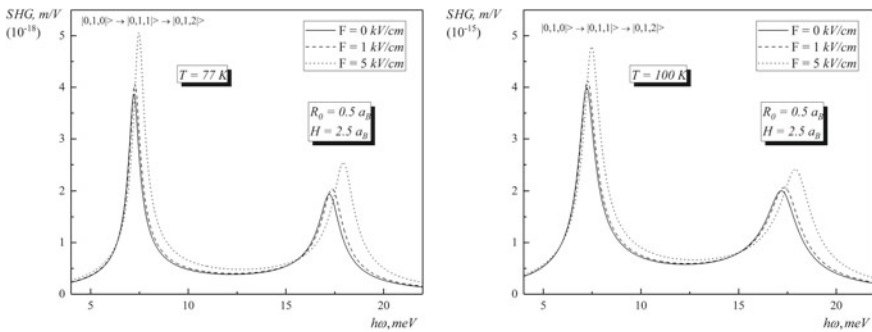
Fig. 39.3 SHG for transition for $|0, 1, 0|\rangle \rightarrow |0, 1, 1|\rangle \rightarrow |0, 1, 2|\rangle$ transition
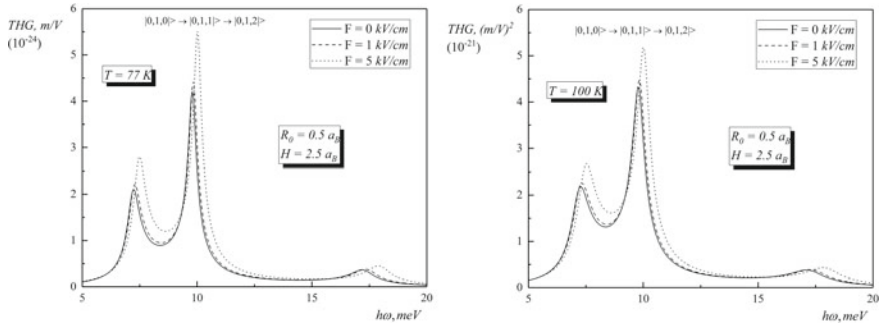
Fig. 39.4 THGfor $|0, 1, 0|\rangle \rightarrow |0, 1, 1|\rangle \rightarrow |0, 1, 2|\rangle$ transition

39.5   Conclusion

The intraband absorption in strongly prolate CQD made of *InAs* is investigated. The dependences of linear and nonlinear absorption coefficientson the incident light quantum energyon strongly prolate CDQ are studied. The influence of the external electric field strength on the nonlinear characteristics of CQD was taken into account. SHG and THG dependencies on the incident light energy, external electric field and temperatures are shown.

References

1. P. Harrison, A. Valavanis, *Quantum Wells, Wires and Dots: Theoretical and Computational Physics of Semiconductor Nanostructures.* (Wiley, 2016)
2. Zh. Wang, *Quantum Dots* (Springer Science Business Media LLC, New York, 2008)
3. T. Chakraborty, *Quantum Dots: A Survey of the Properties of Artificial Atoms* (Elsevier, Amsterdam, 1999)
4. E.K. Gross, R.M. Dreizler, *Density Functional Theory* (Springer Science & Business Media, 2013)
5. S. Wu, L. Wan, Electronic structures in a CdSe spherical quantum dot in a magnetic field: diagonalization method and variational method. J. Appl. Phys. **111**(6) (2012)
6. R. Khordad, H. Bahramiyan, S.A. Mohammadi, Influence of impurity on binding energy and optical properties of lens shaped quantum dots: finite element method and Arnoldi algorithm. Chin. J. Phys. **54**(1), 20–32 (2016)
7. W. Liu, Y. Li, F. Liu, W. Jiang, D. Zhang, J. Liang, Visible-light-driven photocatalytic degradation of diclofenac by carbon quantum dots modified porous g-C_3N_4: mechanisms, degradation pathway and DFT calculation. Water Res. **151**, 8–19 (2019)
8. V. Galitski, B. Karnakov, V. Kogan, *Exploring Quantum Mechanics: A Collection of 700þ Solved Problems for Students, Lecturers and Researchers* (Oxford University Press, Oxford, 2013)
9. A.A. Gusev, O. Chuluunbaatar, S.I. Vinitsky, E.M. Kazaryan, H.A. Sarkisyan, The application of adiabatic method for the description of impurity states in quantum nanostructures. J. Phys. Conf. Ser, **248** (2010)

10. D.B. Hayrapetyan, E.M. Kazaryan, H.A. Sarkisyan, Magneto-absorption in conical quantum dot ensemble: possible applications for QD LED. Opt. Commun. **371**, 138–143 (2016)
11. D.B. Hayrapetyan, A.V. Chalyan, E.M. Kazaryan, H.A. Sarkisyan, Direct interband light absorption in conical quantum dot. J. Nanomater. (2015)
12. J.H. Huh, C. Hermannstädter, K. Akahane, H. Sasakura, N.A. Jahan, M. Sasaki, I. Suemene, Fabrication of metal embedded nano-cones for single quantum dot emission. Jpn. J. Appl. Phys. **50**(6S) (2011)
13. A. El Aouami, E. Feddi, M. El-Yadri, N. Aghoutane, F. Dujardin, C.A. Duque, H.V. Phuc, Electronic states and optical properties of single donor in GaN conical quantum dot with spherical edge. Superlattices Microstruct. **114**, 214–224 (2018)
14. K.S. Khachatryan, D.B. Hayrapetyan, E.M. Kazaryan, H.A. Sarkisyan, Strongly prolate conical quantum dot in an external electric field, in *International Youth Conference on Electronics, Telecommunications and Information Technologies* (Springer, Cham, 2021), pp. 185–192
15. S. Sauvage, P. Boucaud, F. Glotin, R. Prazeres, J.M. Ortega, A. Lemaitre, J.M. Gérard, V. Thierry-Mieg, Third-harmonic generation in InAs/GaAs self-assembled quantum dots. Phys. Rev. B**59**(15) 9830 (1999)
16. R. Khordad, Third-harmonic generation in a double ring-shaped quantum dot under electron-phonon interaction. Opt. Commun. **391**, 121–127 (2017)
17. I. Karabulut, S. Baskoutas, Second and third harmonic generation susceptibilities of spherical quantum dots: effects of impurities, electric field and size. J. Comput. Theor. Nanosci. **6**(1), 153–156 (2009)
18. T.A. Sargsian, M.A. Mkrtchyan, H.A. Sarkisyan, D.B. Hayrapetyan, Effects of external electric and magnetic fields on the linear and nonlinear optical properties of InAs cylindrical quantum dot with modified Pöschl-Teller and Morse confinement potentials. Physica E Low-dimens. Syst. Nanostruct. **126**, 114440 (2021)

Chapter 40
External Electric Field Effect on Interband Optical Absorption and Photoluminescence in Vertically Coupled Cylindrical Quantum Dots with Modified Pöschl-Teller Potential

Tigran Sargsian

Abstract A theoretical study of external electric field effect on the interband optical absorption and photoluminescence spectra of vertically coupled cylindrical quantum dots with modified Pöschl-Teller potential has been carried out. Interband optical absorption and photoluminescence plots have been obtained at different temperatures and external electric field strengths. It is shown that the application of an external electric field separates the charge carriers spatially, which in turn causes a blue shift of the absorption and photoluminescence peaks and a reduction of their peak values.

Keywords Vertically coupled quantum dots · Modified Pöschl-Teller potential · Interband absorption · Photoluminescence · Electric field

40.1 Introduction

The huge amount of theoretical, experimental and technological knowledge accumulated in the field of semiconductor nanostructures makes it reasonable to use these structures not only in the field of classical computation but also in the quantum computation. Quantum dots (QDs) or artificial atoms are one of the most noteworthy candidates for realizing qubits and quantum gates for quantum information processing [1, 2]. Electronic and optical properties of QDs are highly dependent on the geometrical shape and sizes, which, in turn, can be precisely controlled during the growth of the QD [3–7]. The application of additional external electric field also has a significant influence on the physical properties of QDs [8–11].

Two QDs or artificial atoms placed on each other form a structure named vertically coupled QDs or artificial molecule. Coupling between two QDs modifies their electronic structure and one can control this modification by controlling the interdot distance [12–14].

T. Sargsian (✉)
Russian-Armenian University, 123 Hovsep Emin Str, 0051 Yerevan, Armenia
e-mail: tigran.sargsian@rau.am

The coupled QDs structure can be theoretically modeled by the confinement potential, which will appear during the growth process of coupled QDs. There are a number of theoretical confinement potential models that can be used for modeling and simulating coupled QDs [15–18].

In this work, vertically coupled InAs quantum dots with modified Pöschl-Teller potential (MPTP) have been considered in the presence of an external electric field. Interband optical absorption and photoluminescence coefficients have been obtained for these structures in a strong quantization regime.

40.2 Theory

Consider the motion of a particle in vertically coupled cylindrical QDs, which are confined by MPTP in z axis and by the parabolic potential in the radial direction. In the regime of strong size quantization, the energy of Coulomb interaction between an electron and a hole is much smaller than the energy caused by the walls of the QDs and, thus, can be neglected. Then the problem reduces to finding the energy states of electrons and holes separately. In this case, the Hamiltonian of a particle in cylindrical coordinates can be written as:

$$\widehat{H} = \frac{\widehat{\vec{p}}_p^2}{2m_p^*} + U_{conf}(\vec{r}) \pm e\,\overrightarrow{F}\,\vec{r}, \tag{40.1}$$

in which m_p^* is the effective mass of the particle, $\widehat{\vec{p}}_p$ is the particle momentum operator, $U_{conf}(\vec{r})$ is the confinement potential, e is the electron charge magnitude, $\overrightarrow{F}(0, 0, F)$ is the electrical field intensity, $\vec{r}$ is the radius-vector and the signs - or + before the last term depend on the type of charge carrier (electron or hole).

In dimensionless variables the system Hamiltonian can be presented as the sum of Hamiltonians of radial and axial subsystems:

$$\widehat{H} = \widehat{H}_{radial} + \widehat{H}_{axial}, \tag{40.2}$$

where:

$$\widehat{H}_{radial} = -\frac{\partial^2}{\partial\rho^2} - \frac{1}{\rho}\frac{\partial}{\partial\rho} - \frac{1}{\rho^2}\frac{\partial^2}{\partial\varphi^2} + U_{conf}(\rho, \varphi), \tag{40.3}$$

$$\widehat{H}_{axial} = -\frac{\partial^2}{\partial z^2} \pm fz + U_{conf}(z). \tag{40.4}$$
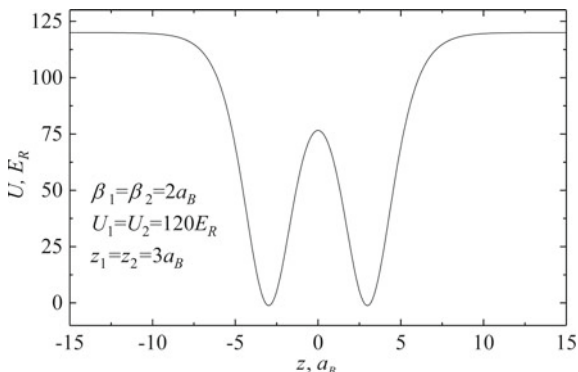
Fig. 40.1 The form of MPTP

Here all lengths are given in effective Bohr radius $a_B = \frac{\varepsilon_r \hbar^2}{e^2 m^*}$, and all energies in effective Rydberg energy $E_R = \frac{\hbar^2}{2m^* a_B^2}$. In addition, $f = \frac{eFa_B}{E_R}$ denotation is made.

As it is mentioned above, for the radial direction the parabolic confinement potential and for the axial direction MPTP has been chosen:

$$U_{conf}(\rho, \varphi) = \gamma_0 \rho^2/4, \tag{40.5}$$

$$U_{conf}(z) = \max\left[\tilde{U}_1, \tilde{U}_2\right] - \frac{\tilde{U}_1}{ch^2\left(\frac{z+z_1}{\tilde{\beta}_1}\right)} - \frac{\tilde{U}_2}{ch^2\left(\frac{z-z_2}{\tilde{\beta}_2}\right)}, \tag{40.6}$$

where $\gamma_0 = \frac{2m^* \omega a_B^2}{\hbar}$, ω is the frequency of parabolic confinement potential, $\tilde{U}_{1,2}$ are the depth, $\tilde{\beta}_{1,2}$ the half-width and $z_{1,2}$ the offsets of first and second QDs, respectively, and $\max\left[\tilde{U}_1, \tilde{U}_2\right]$ is the greater value from the QD depths. The form of the potential in the axial direction is shown in the Fig. 40.1.

For one QD the total wave function of the system is searched in the following form:

$$\psi(r, \varphi, z) = Ce^{im\varphi} R(r)\chi(z), \tag{40.7}$$

where C is the normalization coefficient and m is the magnetic quantum number.

In the case of the absence of external fields the problem has been solved analytically [19] and in the strong size quantization regime for the dimensionless wave functions and energy levels we can write:

$$R(\rho) = e^{-\frac{\gamma_0 \rho^2}{2}} (\gamma_0 \rho^2)^{|m|/2} {}_1F_1 \left[-n_\rho, |m| + 1, \gamma_0 \rho^2 \right],$$

$$\chi_{n_z}(z) = \left(1 - th^2 \left(\frac{z}{\beta} \right) \right)^{\frac{-(1+2n_z) + \sqrt{1+4\beta^2 U_0}}{4}} \cdot$$

$$\cdot {}_2F_1 \left(-n_z, \sqrt{1 + 4\beta^2 U_0} - n_z, \frac{1}{2} \left(1 + \sqrt{1 + 4\beta^2 U_0} \right) - n_z, \frac{1 - th(z/\beta)}{2} \right),$$

$$(40.8)$$

$$E_\rho = 2\gamma_0 (N + 1),$$

$$E_z = U_0 - \frac{1}{4\beta^2} \left[-(1 + 2n_z) + \sqrt{1 + 4\beta^2 U_0} \right]^2. \qquad (40.9)$$

Here $N = 2n_\rho + |m|$ is the oscillatory quantum number, n_ρ and n_z are radial and axial quantum numbers, respectively. In the case of the presence of external electric field, the problem has been solved in the framework of finite element method for axial direction. For the coupled QDs the problem has been solved with the help of numerical discretization method.

40.3 Results and Discussion

Now let us consider the interband light absorption and photoluminescence in vertically coupled cylindrical QDs with MPTP confining potential in the presence of external electric field. As it was mentioned above, QDs made of InAs will be considered. The material parameters which have been used in the calculations are the following: $m_e^* = 0.023 m_0$, $m_{hh}^* = 0.41 m_0$, where m_0 is the free electron mass, $E_g = 0.35$ eV.

For the absorption coefficient one can write the following equation:

$$\alpha(\omega) = \alpha_0 \omega \left(\int \Psi_e(\vec{r}_e) \Psi_h(\vec{r}_h) dV \right)^2 \sum_{\nu_e, \nu_h} \delta \left(\hbar \omega_{e,h} - E_g - E_e - E_h \right), \qquad (40.10)$$

where α_0 is a coefficient proportional to the square of the modulus of the matrix element of the dipole moment taken on the Bloch functions, $\Psi_e(\vec{r}_e)$ and $\Psi_h(\vec{r}_h)$ are electron and heavy hole wave functions, respectively, E_e is electron energy, E_h is heavy hole energy.

The PL spectra are calculated using the relation:

$$R(\hbar\omega) = R_0 \cdot \hbar\omega \cdot \alpha(\omega) \cdot e^{\frac{\hbar\omega - E_g}{kT}} \cdot e^{\frac{0.5 - E_g}{kT}}, \qquad (40.11)$$
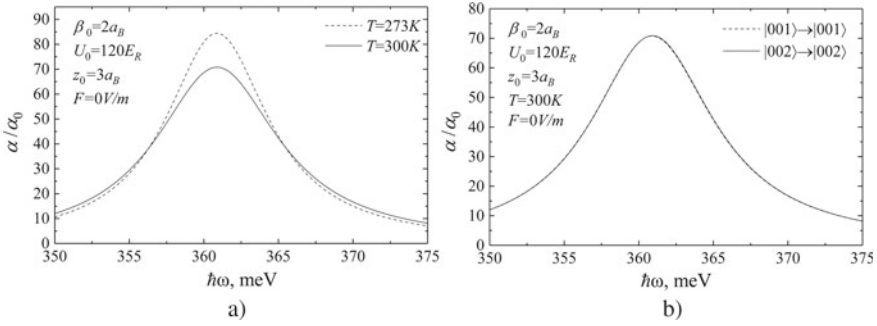
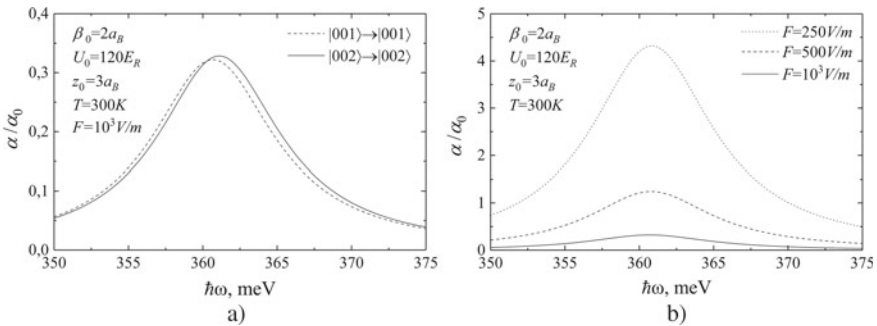**Fig. 40.2** Dependence of absorption coefficient on the incident light energy, **a** for $|001\rangle \rightarrow |001\rangle$ transition, when external electric field is absent at 273 and 300 K, **b** for $|001\rangle \rightarrow |001\rangle$ and $|002\rangle \rightarrow |002\rangle$ transition at 300 K

where R_0 is a coefficient proportional to the square of the modulus of the matrix element of the dipole moment taken on the Bloch functions.

It should be noted, that coupled QDs discussed here are symmetrical, e.g. $U_0 = \widetilde{U}_1 = \widetilde{U}_2 = 120\,E_R$, $\beta_0 = \tilde{\beta}_1 = \tilde{\beta}_2 = 2a_B$ and $z_0 = z_1 = z_2 = 3a_B$.

In Fig. 40.2a the dependence of absorption coefficient on the incident light photon energy is plotted for $|001\rangle \rightarrow |001\rangle$ transition at 273 and 300 K temperatures when external electric field is absent. Here the $|n_\rho, m, n_z\rangle$ notation is used. Figure 40.2b presents the dependence of absorption coefficient on the incident light photon energy at 300 K but for both transitions $|001\rangle \rightarrow |001\rangle$ and $|002\rangle \rightarrow |002\rangle$, and one can see that due to the symmetry of the coupled QDs these levels are degenerated and absorption curves coincide. One can notice that curves in Fig. 40.2 are broadened. The broadening mechanisms and parameters are discussed in details in [20]. When external electric field is present ($F = 10^3$ V/m) one can see that the corresponding curves are separated (see Fig. 40.3a). Thus, the presence of external electric field

Fig. 40.3 Dependence of absorption coefficient on the incident light photon energy **a** for $|001\rangle \rightarrow |001\rangle$ and $|002\rangle \rightarrow |002\rangle$ transitions, at 300 K temperature, **b** for different values of electric field
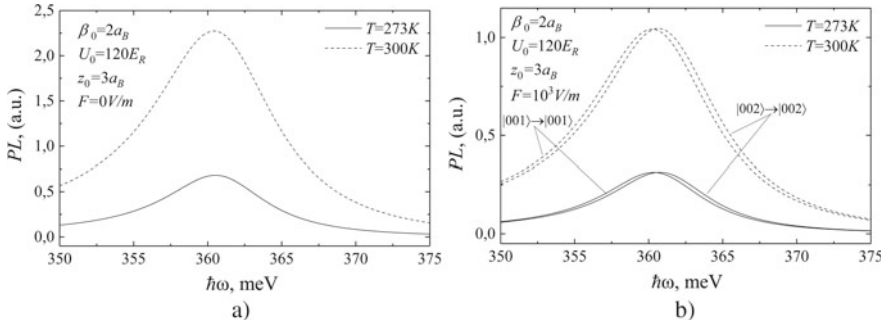
Fig. 40.4 Dependence of PL on the light energy for $|001\rangle \rightarrow |001\rangle$ transition at 273 and 300 K (**a**), and for transitions $|001\rangle \rightarrow |001\rangle$ and $|002\rangle \rightarrow |002\rangle$ at 273 and 300 K (**b**)

removes the degenerate states. In Fig. 40.3b one can see the behavior of the absorption coefficient curves at different values of external electric field intensity.

Figure 40.4a presents the dependence of PL spectra on the light energy for the $|001\rangle \rightarrow |001\rangle$ transition at 273 and 300 K when external electric field is absent. Figure 40.4b presents the influence of external electric field on the PL spectra for both transitions $|001\rangle \rightarrow |001\rangle$ and $|002\rangle \rightarrow |002\rangle$.

40.4 Conclusion

In the work presented, the theoretical study of external electric field effect on the interband optical absorption and photoluminescence spectra of vertically coupled cylindrical quantum dots with modified Pöschl-Teller potential has been carried out. Coupling between two quantum dots affects their electronic structure, which can be controlled by changing the interdot distance. Such structures are great candidates for physical realization of qubits and quantum gates, which are the foundation of quantum computers. It is shown that the position of the absorption curve peak is independent of temperature, while the intensity is dependent. It is shown that the application of external electric field separates charge carriers spatially, which in turn causes a blue shift of the absorption and photoluminescence peaks and a reduction of their peak values. It should be noted, that the absorption curves for transitions $|001\rangle \rightarrow |001\rangle$ and $|002\rangle \rightarrow |002\rangle$ at 300 K merge, while the presence of external electric field separates these curves, in other words, the degeneration of levels is moved.

References

1. X. Zhang, H. Li, K. Wang, G. Cao, M. Xiao, G. Guo, Qubits based on semiconductor quantum dots. Chin. Phys. B **27**, 020305 (2018). https://doi.org/10.1088/1674-1056/27/2/020305

2. L. Fedichkin, M. Yanchenko, K. Valiev, Coherent charge qubits based on GaAs quantum dots with a built-in barrier. Nanotechnology **11**, 387–391 (2000). https://doi.org/10.1088/0957-4484/11/4/339

3. C. Cornet, A. Schliwa, J. Even, F. Doré, C. Celebi, A. Létoublon, E. Macé, C. Paranthoën, A. Simon, P. Koenraad, N. Bertru, D. Bimberg, S. Loualiche, Electronic and optical properties of InAs/InP quantum dots on InP(100) and InP(311)B substrates: theory and experiment. Phys. Rev. B. **74** (2006). https://doi.org/10.1103/PhysRevB.74.035312

4. Z. Zeng, C. Garoufalis, S. Baskoutas, G. Bester, Electronic and optical properties of ZnO quantum dots under hydrostatic pressure. Phys. Rev. B. **87** (2013). https://doi.org/10.1103/PhysRevB.87.125302

5. D. Hayrapetyan, E. Kazaryan, H. Tevosyan, Optical properties of spherical quantum dot with modified Pöschl-Teller potential. Superlattices Microstruct. **64**, 204–212 (2013). https://doi.org/10.1016/j.spmi.2013.09.002

6. R. Singh, G. Bester, Effects of atomic ordering on the electronic and optical properties of self-assembled InxGa1−xAs/GaAs semiconductor quantum dots. Phys. Rev. B. **84** (2011). https://doi.org/10.1103/PhysRevB.84.241402

7. L. Li, D. Moldovan, W. Xu, F. Peeters, Electric-field and magnetic-field dependence of the electronic and optical properties of phosphorene quantum dots. Nanotechnology **28**, 085702 (2017). https://doi.org/10.1088/1361-6528/aa55e8

8. İ Karabulut, S. Baskoutas, Linear and nonlinear optical absorption coefficients and refractive index changes in spherical quantum dots: effects of impurities, electric field, size, and optical intensity. J. Appl. Phys. **103**, 073512 (2008). https://doi.org/10.1063/1.2904860

9. D. Baghdasaryan, D. Hayrapetyan, V. Harutyunyan, Optical transitions in semiconductor nanospherical core/shell/shell heterostructure in the presence of radial electrostatic field. Physica B **510**, 33–37 (2017). https://doi.org/10.1016/j.physb.2017.01.017

10. S. Baskoutas, E. Paspalakis, A. Terzis, Electronic structure and nonlinear optical rectification in a quantum dot: effects of impurities and external electric field. J. Phys. Condens. Matter **19**, 395024 (2007). https://doi.org/10.1088/0953-8984/19/39/395024

11. E. Kadantsev, P. Hawrylak, Theory of exciton fine structure in semiconductor quantum dots: quantum dot anisotropy and lateral electric field. Phys. Rev. B. **81**. https://doi.org/10.1103/PhysRevB.81.045311

12. M. Bayer, Coupling and entangling of quantum states in quantum dot molecules. Science **291**, 451–453 (2001). https://doi.org/10.1126/science.291.5503.451

13. C. Kagan, C. Murray, Charge transport in strongly coupled quantum dot solids. Nat. Nanotechnol. **10**, 1013–1026 (2015). https://doi.org/10.1038/nnano.2015.247

14. M. Frederick, V. Amin, E. Weiss, Optical properties of strongly coupled quantum dot-ligand systems. J. Phys. Chem. Lett. **4**, 634–640 (2013). https://doi.org/10.1021/jz301905n

15. D. Hayrapetyan, S. Amirkhanyan, E. Kazaryan, H. Sarkisyan, Effect of hydrostatic pressure on diamagnetic susceptibility of hydrogenic donor impurity in core/shell/shell spherical quantum dot with Kratzer confining potential. Physica E **84**, 367–371 (2016). https://doi.org/10.1016/j.physe.2016.07.028

16. D. Hayrapetyan, E. Kazaryan, T. Kotanjyan, H. Tevosyan, Exciton states and interband absorption of cylindrical quantum dot with Morse confining potential. Superlattices Microstruct. **78**, 40–49 (2015). https://doi.org/10.1016/j.spmi.2014.11.025

17. D. Hayrapetyan, E. Kazaryan, H. Tevosyan, Impurity states in a cylindrical quantum dot with the modified Pöschl-Teller potential. J. Contemp. Phys. (Armenian Academy of Sciences) **49**, 119–122 (2014). https://doi.org/10.3103/S1068337214030062

18. D. Hayrapetyan, A. Achoyan, E. Kazaryan, H. Tevosyan, Electronic states in a cylindrical quantum dot with the modified Pöschl-Teller potential in the presence of external magnetic

field. J. Contemp. Phys. (Armenian Academy of Sciences) **48**, 285–290 (2013). https://doi.org/10.3103/S1068337213060054

19. D. Hayrapetyan, E. Kazaryan, H. Tevosyan, Direct interband light absorption in the cylindrical quantum dot with modified Pöschl-Teller potential. Physica E **46**, 274–278 (2012). https://doi.org/10.1016/j.physe.2012.10.005

20. T. Sargsian, M. Mkrtchyan, H. Sarkisyan, D. Hayrapetyan, Effects of external electric and magnetic fields on the linear and nonlinear optical properties of InAs cylindrical quantum dot with modified Pöschl-Teller and Morse confinement potentials. Physica E **126**, 114440 (2021). https://doi.org/10.1016/j.physe.2020.114440

Chapter 41
Features of the Conductive Carbon Coatings Formation on Titanium Electrodes Using C_{60} Ion Beams

Vladimir Pukha, Andrey Belmesov, Alexander Glukhov, Igor Khodos, Mahdi Khadem, Dae-Eun Kim, Kirill Krainov, Alexander Shakhmin[ID], and Platon Karaseov[ID]

Abstract The regularities of superhard carbon nanocomposite (CNCs) growth on a titanium substrate using accelerated C_{60} ion beam are investigated. Substrate temperature T_s and ion energy E are the main parameters varied during film growth. The formation of CNC coatings with a specific electrical resistance below $1\ \Omega \times m$ is observed at a substrate temperature T_s above 300 ºC and ion energies from 5 to 8 keV. CNC coatings obtained at T_s in the range of 300–400 ºC consist of graphite crystals with a size of 1–2 nm enclosed in a diamond-like matrix. The sp^2/sp^3 bond ratio weakly depends on T_s in this temperature range and slightly decreases with ion energy increase (from 0.33 to 0.26 at 5–8 keV). Diamond-like coating grows on a substrate at T_s lower than 300 ºC by 5 and 7 keV ions. At an ion energy of 8 keV and above sputtering of the substrate was found. In all cases a TiC layer is formed at the substrate-coating interface due to the ion-beam mixing. Formation of this interlayer provides good adhesion of the coating to the substrate. The CNC coating on Ti exhibits high corrosion resistance and good protective properties. Thus, composite carbon films could be used as a conductive electrode in various chemical and bio applications.

Keywords Carbon coatings · Carbon nanocomposite · Ti · Cluster ion growth · C_{60} · Ion bombardment · Conductive coatings · Corrosion resistance

V. Pukha (✉) · A. Belmesov · A. Glukhov
Institute of Problems of Chemical Physics (IPCP RAS), Chernogolovka, Russia

I. Khodos · M. Khadem
Institute of Microelectronic Technology and High Purity Materials, Chernogolovka, Russia

D.-E. Kim
Department of Mechanical Engineering, Yonsei University, Seoul, South Korea

K. Krainov
PAO Severstal, Cherepovets, Vologda Region, Russia

A. Shakhmin · P. Karaseov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: platon.karaseov@spbstu.ru

41.1   Introduction

Titanium and its alloys are currently the main materials for medical implants in orthopedics, traumatology, and dentistry [1] as titanium has a unique combination of properties required for biomaterials, namely, bioinertness, hypoallergenicity, lack of toxicity, etc. These properties appear primarily due to the thin layer of titanium oxide always present on the surface. The disadvantage of titanium is poor tribological properties (high coefficient of friction and wear) [2] which make it practically impossible to use it in rubbing joint pairs. Moreover, titanium oxide is not stable in the corrosive environment of the organism which complicates its use as electrode biosystems.

To date, various technologies have been proposed for modifying the surface of titanium alloys, aimed both to solve problems with wear resistance and to improve the electrical properties of the surface. One of the candidates for hardening the surface of titanium and its alloys are carbon based coatings, including diamond and diamond-like carbon films [3, 4]. Superhard carbon-based coatings, including hydrogenated ones, are collectively referred to as diamond-like carbon (DLC). DLC is a non-crystalline carbon with a high proportion of diamond-like (sp3) bonds. DLC possess excellent physical properties (high hardness, high modulus of elasticity), and also are chemically inert to any acids, alkaline solutions and organic solvents. DLC can be effectively used to coat implants in contact with bone and soft tissues of an organism. Metal implants with a diamond-like carbon coating exhibit high biocompatibility. Carbon-based coating does not affect blood coagulation; it serves as an effective barrier preventing the diffusion of metal ions into surrounding body. The validity of the use of carbon materials as a coating of endoprostheses and implants is based on the ability of carbon materials to stimulate osteogenesis. In addition, carbon coatings have a biostimulating effect, improving the regeneration of the tissues surrounding the implant. Products of wear and tear or destruction do not have a harmful effect on the surrounding tissues, lymph nodes and the body as a whole [5]. The combination of a low coefficient of friction (COF) and high wear resistance makes it possible to significantly increase the durability of precision friction pairs, for example artificial joints. In addition, DLC is an attractive electrode material due to its antifouling properties. For example, DLC exhibited high static resistance to the adhesion of *Staphylococcus aureus* and *Staphylococcus epidermidis* compared to titanium, tantalum and chromium [6–9].

Another positive development for the use of hard carbon coatings is their electrochemical properties. The rate of electrochemical reactions significantly depends on the electrode surface properties. The kinetics of oxygen and hydrogen evolution on a diamond-like carbon film is much slower than on the most commonly used metal electrodes [10]. Doping with boron atoms is used in order to make the films conductive [11]. Ion beam irradiation can be used to improve its properties [12, 13]. Such coatings are an excellent electrode material due to the large potential window in an aqueous solution, low background current, fast electron kinetics, and chemical stability [14, 15]. These characteristics of carbon-based electrodes have been

used in many applications. Unfortunately, the chemical vapor deposition method used to grow conductive diamond-like carbon films is carried out at high temperatures (>800 °C), which limits the choice of substrate materials. Superhard carbon coatings, which can be obtained at low substrate temperatures (T_s), usually have high values of electrical resistance and also require doping to increase conductivity. Doping with nitrogen and boron is not effective for DLC films. Nitrogen doped DLC is n-type semiconductor where nitrogen acts as a weak compensative donor. Boron is a weak p-type impurity for non-hydrogenated DLCs. Doping with metals is much more effective, but in this case the diffusion of metal ions into surrounding tissues is highly expected [10].

Recently, we have demonstrated pure carbon nanocomposite (CNC) coatings synthesized using accelerated C_{60} ion deposition [16 17 18]. CNCs are formed by a few keV energetic ions at relatively high substrate temperatures (300–400 °C), and consist of 1–2 nm in size graphite nanocrystals enclosed in an amorphous diamond-like matrix [17]. In contrast to DLC coatings, CNCs have a high ratio of hardness to Young's modulus (0.13–0.14) along with high hardness (H ~ 40–60 GPa), which reduces the likelihood of brittle fracture of the coating under load [18]. In addition, they have excellent biocompatibility [16], which allows them to be used as medical antifriction coatings (for example, artificial joint parts coating). Such films have a high conductivity comparable to that of graphite the necessary sp3-bonded carbon structure, which usually exhibits a wide potential window. In addition, sp2 carbon is very important for obtaining the electrode activity required for the oxidation of target molecules.

Thus, modification of the surface of titanium and its alloys with CNC coatings can not only protect the titanium surface from corrosion and improve its tribological properties, but also make it possible to use it as an electrode material for detecting a number of biomolecules, for measuring biopotentials and controlling bionic prostheses with their help. In this work, we investigated the regularities of the deposition of coatings on a Ti surface under irradiation with 5–8 keV C_{60} ions. Sputtering of a target by fullerene ions has a significant effect on the formation of coatings plays important role in this energy range [19]. Special attention is paid to the revealing relation between CNC coating structure and properties and deposition conditions.

41.2 Experimental

The installation and the method of deposition of carbon nanocomposites are described in detail in [17]. Polished VT1-0 titanium plates 10 mm^3 × 10 mm^3 × 0.5 mm^3 in size were used as substrates. The substrate was heated using halogen lamps and kept at temperatures (T_s) 100, 200, 300, 350, and 400 °C. Fullerene C_{60} powder of 99.5% purity (NeoTechProduct, St. Petersburg, Russia) was used as a starting material. Fullerene powder was purified by vacuum distillation before use. The ion beam passed through a system of electrostatic lenses, and a magnetic mass separator onto a substrate covered a ~1 cm diameter spot. The ion energies were 5, 7, and

8 keV. Base pressure in the chamber during deposition was ~10^{-4} Pa. In order to distinguish samples, we have developed the following notation to encode all the important parameters. The first two characters represent the substrate material (Ti), one digit after the dash indicates the energy of the fullerene ion, and the last three digits after the second dash indicate the deposition temperature. For example, Ti-5-300 denotes a film deposited with a 5 keV beam at 300 °C.

After C_{60} ion bombardment, samples were examined using X-ray photoelectron spectroscopy (XPS), atomic force microscopy (AFM), Rutherford backscattering (RBS), and transmission electron microscopy (TEM). NanoFab-25 (NT-MDT) was used to study samples by XPS. An Au film less than a nanometer thick was deposited on a carbon film to check a possible charge-related shift in the position of the Cls peak [20]. The crystal structure of the coatings was determined by TEM using a JEM-2100 microscope. Titanium substrate was dissolved in a mixture of HNO_3 and HF acids (10:1), and then the carbon film was washed in deionized water and placed onto a copper electron microscopic grid.

Electrical measurements were carried out using the four-probe technique. Corrosion tests were carried out in a three-electrode liquid cell (0.9% NaCl in water) with a platinum auxiliary electrode, and a silver-chloride reference electrode. Test sample was used as a working electrode. The measurements were carried out using Elins P40x potentiostat at 25 °C, in argon ambient to avoid oxygenation. The potential was scanned in the range from −0.5 to 1 V at a rate of 0.3 mV/s. Corrosion current and potential were calculated from the data obtained using the standard procedure [21].

41.3   Results and Discussion

41.3.1   Carbon Coating Deposition

Carbon coating with thickness proportional to the beam current and deposition time was observed on the titanium surface irradiated using 5 and 7 keV C_{60} ions at all T_s. No coating is observed on samples irradiated by 8 keV ions at T_s < 200 °C. Carbon film appears on the substrate surface if T_s exceeds 300 °C. The absence of a continuous coating at low temperatures was confirmed by the RBS method and survey XPS spectra (see Fig. 41.1). Figure 41.1a reveals the presence of significant amount of titanium atoms right on the surface of the Ti-8-200 sample, whereas it is strongly shifted deep for Ti-5-400 due to the presence of carbon film. Figure 41.1b shows the presence of O1s and C1s peaks in the XPS survey spectra in both cases. In addition, the Ti2p peak is clearly seen in the XPS spectrum of the sample grown at 100 °C, which is absent in the spectrum of the sample obtained at Ts = 400 °C. Similar phenomenon was earlier observed on silicon target at the same temperatures and C_{60} ion energies [19, 22], as well as at lower energies for substrates coated with In_2O_3-SnO_2 [23] and was explained by the strong effect of sputtering process.
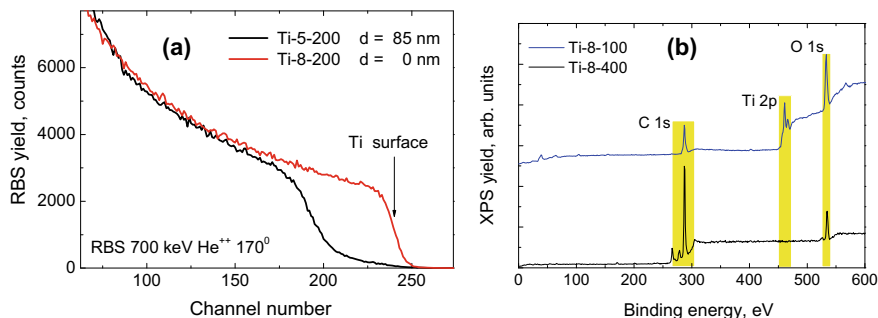
Fig. 41.1 RBS spectra of selected samples as indicated **a**, and XPS spectra of carbon films deposited on Ti substrates **b**

41.3.2 Carbon Bond Composition

It is known that the properties of amorphous carbon coatings are significantly affected by the hybridization of electrons of atoms. In particular, the hardness and Young's modulus of the coating are mainly determined by the sp3 bond concentration. X-ray photoelectron spectroscopy was used to determine the sp2/sp3 bond ratio in the obtained coatings. Figure 41.1b illustrates typical survey spectra. Typical C1s peaks are shown in the Fig. 41.2 together with their deconvolution. It can be easily seen that C1s peak consists of two main peaks at 285 and 284 eV appear from sp3 and sp2 bonded atoms respectively [20]. Side peaks seen at about 286.4 and 288.7 eV are due to C-O and C=O bonds respectively. An additional peak at an energy of 281.6 eV corresponding to C-Ti bonds is present in the spectra of samples irradiated with 8 keV C_{60} ions at $T_s = 100$ and 200 °C (see Fig. 41.2b). Similar effect was found in the case of Si target irradiation [22]. Titanium carbide also appears at a lower C_{60} ion energy at the initial stages of coating growth, regardless of the substrate temperature.
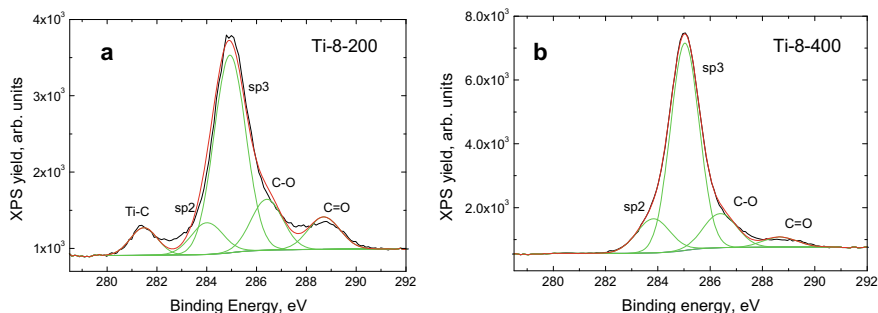
Fig. 41.2 Typical XPS spectra obtained from samples irradiated with 8 keV C_{60} ions at low **a** and high **b** substrate temperature

Approximation with Gaussian function was used to deconvolute C1s spectra. Identical parameters were used in the deconvolution analysis of spectra. The composition of the film was determined as the ratio of the areas under the C1s peak components (see Fig. 41.2). In all cases studied, the sp3/sp2 ratio exhibits very slight dependence on the substrate temperature within the values corresponding to the nanocomposite coatings formation. It decreases from 3.9 to 3.5 at 5 keV, and from 2.8 to 2.2 at 8 keV, with T_s increase from 100 to 400 °C, and 300 to 400 °C respectively.

41.3.3 Coating Structure

TEM studies have shown that coatings grown at temperatures of 100 and 200 °C are amorphous. A characteristic halo is observed in the microdiffraction patterns (not shown here). An increase in T_s leads to the formation of graphite nanocrystals. The reflection in the form of a blurred ring corresponding to the interplanar spacing of 0.33 nm (which is close to 0.338 nm of (002) interplanar spacing of graphite) is well distinguished in the microdiffraction pattern taken from the sample grown by 5 keV ions at $T_s = 300$ °C (See Fig. 41.3a). This reflection becomes narrower and more intense with T_s increase, while the intensity of the halo from amorphous carbon decreases. Graphite (004) reflection appears at $T_s = 350$ °C and becomes better manifested at 400 °C. Thus, the perfection and size of graphite nanocrystals increases with an increase in the deposition temperature, although the amorphous halo is still present in the microdiffraction pattern. Ion energy change from 5 to 8 keV has a little effect on the coating structure. Formation of graphite nanocrystals is slightly enhanced at $T_s = 350$ and 400 °C. Thus, it can be concluded that CNC coatings are formed on titanium in the temperature range from 300 to 400 °C. CNC consists of 1–2 nm in size graphite nanocrystals enclosed by an amorphous carbon matrix. Similar effect was earlier observed on silicon substrates [17].
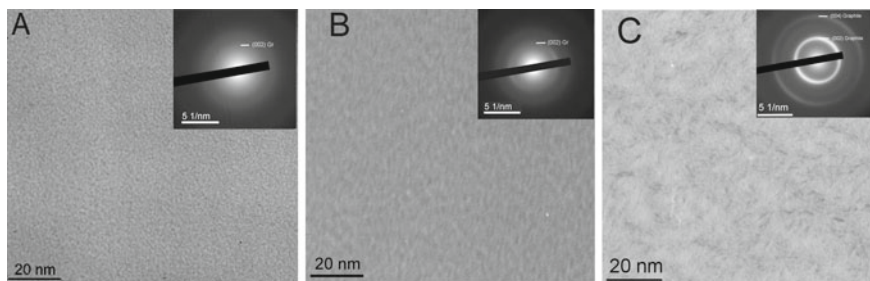
Fig. 41.3 Bright field TEM images of nanocomposite carbon coatings obtained by 5 keV C$_{60}$ ion beam irradiation of a substrate at a temperatures of **a** 300 °C, **b** 350 °C, **c** 400 °C. The inset shows microdiffraction patterns from these regions. Dark spots in the image (C) are from graphite nanocrystals

It should be noted that a significant change in the crystal structure of coatings in the deposition temperature range of 300–400 °C (see Fig. 41.3) occurs with an almost unchanged sp3/sp2 bond ratio (see Fig. 41.2).

41.3.4   Potentiometry

Corrosion resistance in a medium simulating biological fluids was studied in the potentiodynamic mode. According to the measurement results (see Fig. 41.4), it can be seen that the carbon coating protects titanium well, since there is a significant shift in the corrosion potential and a decrease in the corrosion current. Also, by the absence of additional peaks on the polarization curve of a sample protected by a carbon film, one can judge the continuity of the carbon coating. Table 41.1 presents corrosion parameters obtained for bare and CNC coated Ti substrate. It is seen that corrosion potential is almost twice higher whereas corrosion current in almost order of magnitude lower for the CNC coated sample.

Fig. 41.4 Polarization curves measured on bare and CNC coated Ti plates
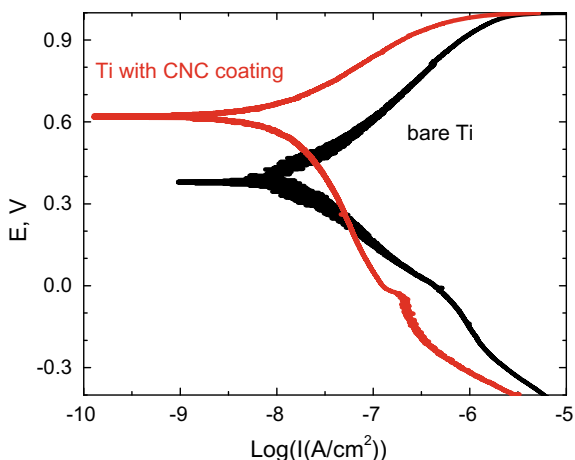
Table 41.1 Corrosion characteristics

	Corrosion potential, V	Corrosion current, A/cm^2
Bare Ti	0.387	1.4×10^{-9}
CNC coated Ti	0.625	3.3×10^{-10}

41.3.5 Electric Conductivity

In order to investigate the electrical conductivity of CNC coatings special sample geometry has been developed. Four silver strips 1 mm^2 × 5 mm^2 in size spaced by 4 mm distance between them were deposited onto a polished sapphire substrate. Then, a thin titanium layer was deposited above these strips through a 4 mm wide mask by electron beam evaporation until electric conduction appeared between them and the resistance decreased to 100 Ω. The resulting structure was used as a substrate for the deposition of CNC coatings. The layer thicknesses were 10 and 200 nm for Ti and CNC, respectively as determined by AFM measurements at a step formed by a silicon mask. The resistance decreased to 90 Ω after CNC coating grown by 8 keV ion beam at T_s = 400 °C. If we consider Ti and CNC coating as two independent conductors connected in parallel, then the specific resistance of the CNC coating is 1.7×10^{-4} Ω × m. A decrease in the ion energy to 5 keV at the same T_s leads to a decrease in the resistance to 98 Ω and a specific resistance to ~1×10^{-3} Ω × m.

41.4 Conclusions

In conclusion, we report formation and properties of carbon-based nanocomposite that can be used as hard conductive coating for electro-chemical, biological and medical applications. Amount of sp^3 bonded carbon atoms in the film slightly decreases with temperature raise from 100 to 400 °C. Nanocomposite film consisting of graphite nanocrystals (~1.5 nm in size) embedded in diamond-like matrix forms at T_s above 300 °C. Nanocomposite is formed due to the spatial separation of sp^3 and sp^2 components, and not due to the conversion of sp^3 to sp^2. Bombardment with 8 keV C_{60} ions sputters Ti substrate at the temperatures up to 200 ºC. Growth of nanocomposite carbon film starts at T_s = 300 ºC. Amount of sp^3 bonds in the films is slightly lower in this case and formation of graphite nanocrystals is less pronounced. Formation of Ti-C interlayer at the interface between the composite and the substrate provides excellent adhesion. Low sheet resistance provides possibility to use the composite films as a highly conductive protective electrode in various applications. Titanium electrodes covered with a conductive chemically inert coating are very suitable to measure biopotentials and control bionic prostheses.

References

1. M. Kaur, K. Singh, Review on titanium and titanium based alloys as biomaterials for orthopaedic applications. Mater. Sci. Eng. C Mater. Biol. Appl. **102**, 844–862 (2019)
2. G. Wang, P. Zhang, Zhao, Endotoxin contributes to artificial loosening of prostheses induced by titanium particles. Med Sci. Monit. **24**, 7001–7006 (2018)
3. M.R. Derakhshandeh, M.J. Eshraghi, M. Javaheri, S. Khamseh, M.G. Sari, P. Zarrintaj, M. Mozafari, Diamond-like carbon-deposited films: a new class of biocorrosion protective coatings. Surface Innov. **6**(4–5), 266–276 (2018)
4. Y. Lu, G. Huang, S. Wang, C. Mi, S. Wei, F. Tian, H. Cao, Y. Cheng, A review on diamond-like carbon films grown by pulsed laser deposition. Appl. Surf. Sci. **541**, 148573 (1–12) (2021).
5. R. Narayan (ed.), *Diamond-Based Materials for Biomedical Applications.* (Elsevier, 2013)
6. A. Matthew, M. Ben, R. Neil, *In vitro* and *in vivo* investigations into the biocompatibility of diamond-like carbon (DLC) coatings for orthopaedic applications. J. Biomed. Mater. Res. **58**(3), 319–328 (2001)
7. G. Dearnaley, J.H. Arps, Biomedical applications of diamond-like carbon (DLC) coatings: a review. Surf. Coat. Tech. **200**, 2518–2524 (2005)
8. E. Kaivosoja, S. Sainio, J. Lyytinen, T. Palomäki, T. Laurila, S.I. Kim, J.G. Han, J. Koskinen, Carbon thin films as electrode material in neural sensing. Surf. Coat. Tech. **259**, 33–38 (2014)
9. E. Kaivosoja, S. Myllymaa, V. Kouri, K. Myllymaa, R. Lappalainen, Y.T. Konttinen, Enhancement of silicon using micro-patterned surfaces of thin films. Eur. Cell. Mater. **19**, 147–157 (2010)
10. A. Zeng, V.F. Neto, J.J. Gracio, Q.H. Fan, Diamond-like carbon (DLC) films as electrochemical electrodes. Diam. Rel. Mater. **43**, 12–22 (2014)
11. E.N. Shubina, P.A. Karasev, A.I. Titov, N.N. Karasev, A.V. Pozdnyakov, The effect of Diborane additive on the plasma-chemical properties of deposited carbon films. Tech. Phys. Lett. **43**(1), 81–84 (2017)
12. P.A. Karaseov, O.A. Podsvirov, K.V. Karabeshkin, A.Ya. Vinogradov, A.Yu. Azarov, N.N. Karasev, A.I. Titov, A.S. Smirnov, Influence of ion irradiation on internal residual stress in DLC films. Nucl. Instrum. Meth. Phys. Res. B **268**, 3107–3110 (2010)
13. P.A. Karaseov, V.S. Protopopova, K.V. Karabeshkin, E.N. Shubina, M.V. Mishin, J. Koskinen, S. Mohapatra, A. Tripathi, D.K. Avasthi, A.I. Titov, Swift heavy ion irradiation of metal containing tetrahedral amorphous carbon films. Nucl. Instr. Meth. Phys. Res. B **379**, 162–166 (2016)
14. M.A. Elistratova, I.B. Zakharova, O.E. Kvyatkovskii, Self-organization features of tetraphenyl-porphyrins according to quantum chemical calculations. Macroheterocycles **12**(4), 370–374 (2019)
15. I.B. Zakharova, D.I. Dolzhenko, V.F. Borodzyulya, N.T. Sudar', The electroforming effect in polycrystalline fullerene C_{60} films. Tech. Phys. Lett. **45(2)**, 142–144 (2019)
16. O.V. Penkov, V.E. Pukha, S.L. Starikova, M. Khadem, V.V. Starikov, M.V. Maleev, D.-E. Kim, Highly wear-resistant and biocompatible carbon nanocomposite coatings for dental implants. Biomaterials **102**, 130–136 (2016)
17. V.E. Pukha, E.N. Zubarev, A.N. Drozdov, A.T. Pugachov, S.H. Jeong, S.C. Nam, Growth of nanocomposite films from accelerated C_{60} ions. J. Phys. D Appl. Phys. **45**(33), 335302 (2012)
18. O.V. Penkov, V.E. Pukha, E.N. Zubarev, S.S. Yoo, D.E. Kim, Tribological properties of nanostructured DLC coatings deposited by C60 ion beam. Tribol. Int. **60**, 127–135 (2013)
19. V. Pukha, J. Popova, M. Khadem, D.-E. Kim, I. Khodos, A. Shakhmin, M. Mishin, K. Krainov, A. Titov, P. Karaseov, Formation of functional conductive carbon coating on Si by C_{60} ion beam, in *Proceedings of the YETI 2020, St. Petersburg, Russia, Springer Proceedings in Physics*, vol. 255 (2021), pp. 131–139
20. J. Diaz, G. Paolicelli, S. Ferrer, F. Comin, Separation of the sp3 and sp2 components in the C1s photoemission spectra of amorphous carbon films. Phys. Rev. B **54**(11), 8064–8069 (1996)
21. W.H. Kao, Y.L. Su, Y.T. Hsieh, Effects of duplex nitriding and TiN coating treatment on wear resistance, corrosion resistance and biocompatibility of Ti6Al4V Alloy. J. Mater. Eng. Perform. **26**, 3686–3697 (2017)

22. P.A. Karaseov, O.A. Podsvirov, A.I. Titov, K.V. Karabeshkin, A.Ya. Vinogradov, V.S. Belyakov, A.V. Arkhipov, L.M. Nikulina, A.L. Shakhmin, E.N. Shubina, N.N. Karasev, Effect of ion bombardment on the phase composition and mechanical properties of diamond-like carbon films. J. Surf. Invest. **8**, 45–49 (2014)
23. M.V. Maleev, E.N. Zubarev, V.E. Pukha, A.N. Drozdov, A.S. Vus, A.V. Penkov, Features of interaction of accelerated C_{60} ions with the surface of ITO target. Phys. Surf. Eng. **13**(2), 184–195 (2015). ((In Russian))

Chapter 42
Development of a Data Acquisition System for Monitoring Environmental Parameters of Laboratories

Vladislav A. Kruglov, Vladislav S. Reznik, Konstantin S. Sychev, and Van Sinian

Abstract Currently, industrial sensors employing RS-485 communication channel are widely used. A large variety of such sensors allows a wide range of tasks to be solved. The RS-485 communication link allows the simultaneous use of up to 128 sensors, which in turn leads to the need for a data acquisition system, as well as the output of these sensors in a convenient form for the end-user. In accordance with this, this paper is devoted to the development of a sensor data acquisition system using an RS-485 communication channel. This data acquisition system contains an RS-485 communication link, an Ethernet link for data output to the end user, and a USB link used for debugging and controlling the board firmware. The control board uses a single-board computer Raspberry PI compute Module 3. This article discusses the main protection schemes of the channels used, the main schemes necessary for the correct operation of Raspberry Pi Compute Module 3, as well as the basic principles of the data acquisition system.

Keywords Data acquisition system · Protection circuits · RS-485 · Ethernet · USB

42.1 Introduction

Currently, in almost any industry sector, there is a need to control the processes being carried out [1–9]. Nowadays many devices have been developed to implement various measurements [2, 5, 6, 9–18]. Different methods of measuring physical parameters are used depending on the tasks to be solved [10, 11, 13–17, 19–25]. Optical sensors have the greatest potential [14–17, 26–32]. In some cases, they allow transmitting information over fiber-optic communication lines over long distances

V. A. Kruglov (✉) · V. S. Reznik · K. S. Sychev
Institute of Analytical Instrumentation of the Russian Academy of Sciences, St. Petersburg 190103, Russia

V. Sinian
Xi'an Shiyou University, Xi'an 257517, China

(remote control) [6, 33, 34]. When monitoring complex technological processes, it is necessary to use many sensors.

In this case, systems that perform centralized control based on information received from sensors have a great advantage in monitoring.

To ensure centralized management, it is necessary to obtain synchronized information, which can be provided by a system for collecting data from sensors that can be installed directly at the point of observation and a communication channel with external access. Similar problems are solved when processing signals in radar stations from many receiving antennas [35–40].

42.2 Structure of the Data Acquisition System

Due to the wide possible functionality of sensors that communicate via the RS-485 communication channel, this implies a wide application of the data acquisition system itself. Therefore, one of the important parameters of the system itself will be low cost, small size, the ability to place it in various places, as well as convenience and ease of use.

Based on these parameters, the system under development is made in the form of several printed circuit boards, the dimensions of which allow them to be placed in the D6MG standard case. The advantage of this choice is several factors at once: the possibility of placing this housing on a DIN-rail, which in turn allows you to use the data acquisition system as a ready-made housing device and place it in various places, as well as the possibility to use terminal blocks as connectors, which in turn simplifies the operation of this device.

In addition, for ease of operation, the developed system has several LEDs on the front panel, each of which displays the status of the system, a reset button, as well as a USB connector necessary for configuring the system.

The block diagram of the developed data acquisition system is shown in Fig. 42.1.

The main part of the system is a single-Board microcomputer Raspberry Pi Compute Module 3. This microcomputer exchanges data with communication channels, as well as provides data exchange between them.

Due to the fact that Compute Module 3 does not have the ability to communicate using RS-485 and Ethernet communication channels, it was decided to use different interface converters.

The RS-485 interface was converted to UART by using the MAX485 chip. In addition, the implementation scheme of the RS-485 channel includes several different protection schemes, such as protection against overvoltage of the communication channel or against incorrect connection.

Also, to protect the rest of the circuit and to ensure the independence of the signal circuits of the system and the connected sensors, a galvanic isolation of the part of the circuit responsible for the RS-485 communication channel and the rest of the circuit was implemented.
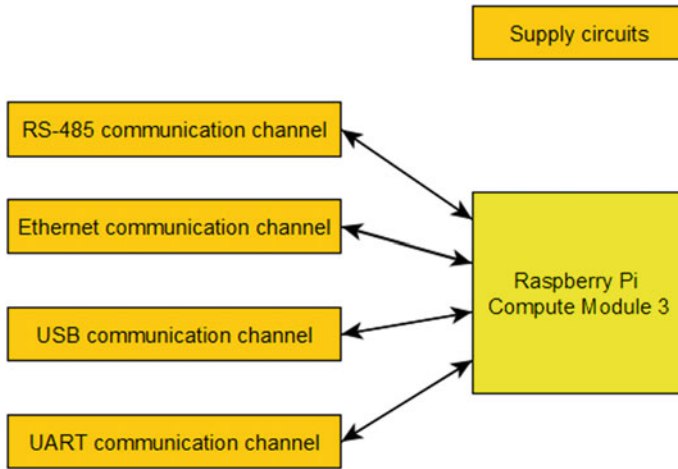
Fig. 42.1 Block-diagram of data acquisition system

The Ethernet interface was converted to SPI by using the ENC28J60 chip. Inter-ference protection schemes were implemented for the correct operation of the interface.

To ensure the initial configuration of the system, as well as the configuration of the Compute Module interfaces, it was decided to use the existing UART interface. For the convenience of debugging and configuration, it was decided to use the USB communication channel in order to debug from the computer. The CP2102 chip is used to convert the UART interface to USB.

Considering that debugging can be performed from another device, similar to the RS-485 channel, galvanic isolation was implemented.

To configure the system for working with sensors, a second USB communication channel is used, implemented as a connector to which various USB drives can be connected.

In order to ensure ease of operation, it was decided to use + 12 V as the supply voltage. The power supply circuits of the various components of the circuit also include protection circuits against incorrect connection, as well as against overvoltage.

42.3   The Principle of Operation of a Data Acquisition System

Since the data acquisition system must work with different sensors using the RS-485 communication channel, it is necessary to configure the system to work with different sensors.

Fig. 42.2 Block diagram of working process of the data acquisition system

To implement this configuration, a USB communication channel is used. Before working, the user needs to write a configuration file to a USB drive that would contain data about the sensors used. This file must be written using YAML. The Python programming language is used to process this data.

The block diagram of the data acquisition system after the configuration is shown in Fig. 42.2.

After the correct configuration, the data acquisition system starts exchanging data with the sensors. In this case, the survey is performed sequentially. The number of sensors is limited by the capabilities of the RS-485 communication channel and cannot exceed 128 sensors.

Next, the data is processed by Compute Module 3 and a future shipment is generated, which is then sent to the user's server using an Ethernet communication channel.

42.4 Hardware Realization

After the basic principles of building the described data acquisition system have been implemented, it is necessary to implement it at the hardware level. To do this, you need to develop a circuit and a printed circuit board of the system.

An example of circuitry solutions used in the development of the hardware level of the data acquisition system is shown in Fig. 42.3. This schematic diagram shows
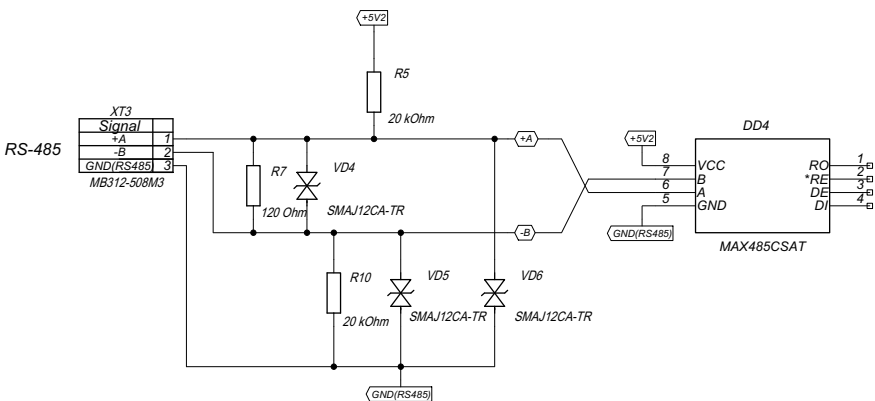
Fig. 42.3 Schematic of RS-485 communication channel

the communication channel with RS-485 sensors, as well as its protection circuits, the scheme for converting the RS-485 interface to UART.

As one of the protection circuits, TVS-diodes are used, which prevent the influence of transients, such as, for example, electrostatic discharge, overvoltage or switching transitions. To protect USB-based communication channels, an IP4234 protection circuit is used, which consists of several diodes that prevent the communication channels, as well as the power and ground lines from closing to each other.

For this device, printed circuit boards were developed, which are shown in Fig. 42.4. For a convenient design of printed circuit boards, two were developed: the first, which houses the main elements of the data acquisition system communication channels with sensors, a debug channel based on the UART protocol, and an Ethernet-based channel. It also houses the Compute Module 3 itself in the SODIMM-200 connector.

The second printed circuit board allows 4 LEDs to be displayed on the front panel of the device housing to display the system status (communication with sensors, communication with the server, connection via one of the channels based on UART or USB, as well as the presence of system power), as well as a USB connector for external storage.
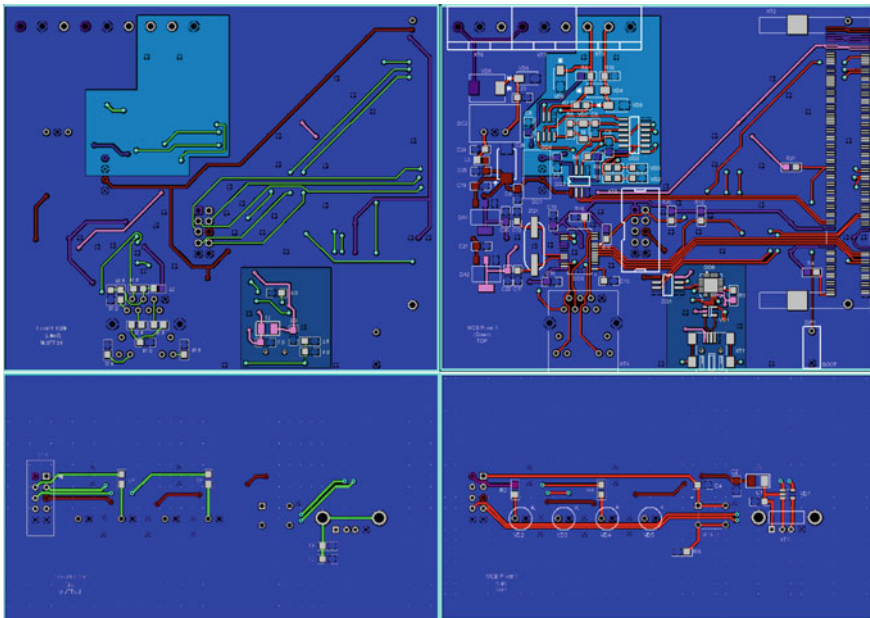
Fig. 42.4 Printed circuit boards of data acquisition system
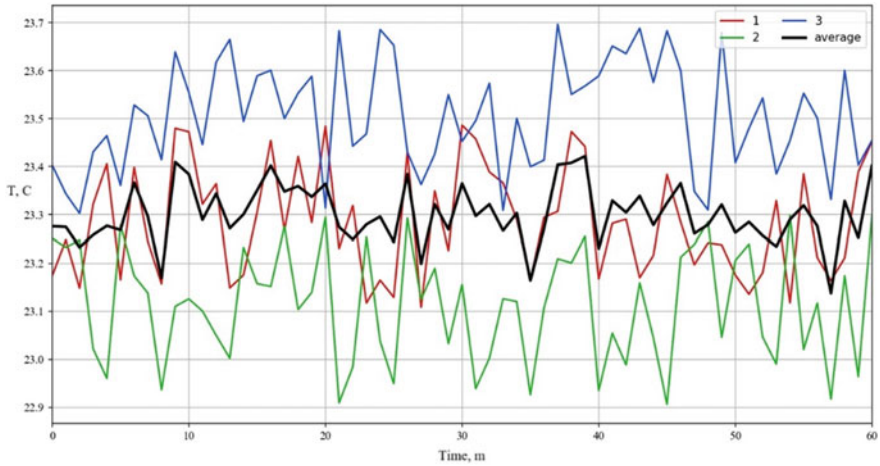
Fig. 42.5 Results of measurements

42.5 Results

As a result, the developed system was tested using 3 temperature sensors. These sensors were installed in various parts of the laboratory, and the obtained data were displayed on the server screen in real time. In addition, the obtained parameters were averaged. The results are shown in Fig. 42.5.

At the moment, there are several successful analogues of such systems on the market based on the Compute Module from Raspberry, which can be performed both in the form of simple expansion cards or ready-made devices. However, the significant disadvantages of such solutions are their high price and the impossibility of their convenient use in laboratories, for example, the lack of a convenient case for placing the system on a DIN rail. The developed system is designed to solve some of the existing shortcomings.

42.6 Conclusion

Thus, the developed data acquisition system allows communication with various sensors using the RS-485 communication channel. Due to the use of Compute Module 3, the system is able to work with various sensors. In addition, the implemented protection circuits allow the unit to be protected against various contingencies.

References

1. O.S. Antonova, G.E. Rudnidskya, A.N. Tupik, A.L. Bylianitsa, V.E. Evstrapov, Polymerase chain reaction: devices and methods. A systematic review of literature. Sci. Instrum **21**(4), 5–21 (2011)
2. V.V. Davydov, N.S. Myazin, A.V. Kiryukhin, Nuclear-Magnetic Flowmeter—Relaxometers for monitoring coolant and feedwater flow and status in Npp. At. Energ. **127**(5), 274–279 (2020)
3. V.V. Davydov, N.S. Myazin, V.N. Dudkin, R.V. Davydov, Peculiarities of monitoring the state of a flowing medium by the method of nuclear magnetic resonance. Tech. Phys. Lett. **46**(1), 55–58 (2020)
4. N.S. Myazin, V.V. Davydov, V.V. Yushkova, N.I. Taranda, VYu. Rud, On the need to control the state of the flowing media by the values of relaxation constants. J. Phys.: Conf. Ser. **1410**(1), 012130 (2019)
5. V.V. Davydov, N.S. Myazin, E.N. Velichko, Characteristics of spectrum registration of condensed medium by the method of nuclear-magnetic resonance in a weak field. Tech. Phys. Lett. **43**(7), 607–610 (2019)
6. N.S. Myazin, V.I. Dudkin, N.M. Grebenikova, V.V. Davydov, V.Y. Rud', A.S. Podstrigaev, Fiber—Optical System for Governance and Control of Work for Nuclear Power Stations of Low Power, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11660 (LNCS, 2019), pp. 744–756
7. V.V. Davydov, N.S. Myazin, S.E. Logunov, V.B. Fadeenko, A contactless method for testing inner walls of pipelines. Russ. J. Nondestr. Test. **54**(3), 213–221 (2018)
8. G.L. Klimchitskaya, V.M. Mostepanenko, E.K. Nepomnyashchaya, E.N. Velichko, Impact of magnetic nanoparticles on the Casimir pressure in three-layer systems. Phys. Rev. B **99**(4), 045433 (2019)
9. V.V. Davydov, N.S. Myazin, V.I. Dudkin, E.N. Velichko, Investigation of condensed media in weak fields by the method of nuclear magnetic resonance. Russ. Phys. J. **61**(1), 162–168 (2018)
10. N. Myazin, Y. Neronov, V. Dudkin, V. Davydov, V. Yushkova, Environmental express monitoring of territory and water bodies at various stages of construction and improvement. MATEC Web Conf. **245**, 11013 (2018)
11. B. Gizatullin, M. Gafurov, A. Rodionov, G. Mamin, C. Mattea, S. Stapf, S. Orlinskii, Proton-radical interaction in crude oil—a combined NMR and EPR study. Energy Fuels **32**(11), 11261–11268 (2018)
12. S.V. Dyachenko, M.A. Vaseshenkova, K.D. Martinson, I.A. Cherepkova, A.I. Zhernovoi, Synthesis and properties of magnetic fluids produced on the basis of magnetite particles. Russ. J. Appl. Chem. **89**(5), 690–696 (2016)
13. F. Shariaty, M. Baranov, E. Velichko, M. Galeeva, V. Pavlov, Radiomics: extracting more features using endoscopic imaging, in *2019 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)* (2019), pp. 181–194
14. E.A. Savchenko, A.N. Skvortsov, E.N. Velichko, E.Yu. Savchenko, Determination of electrophoretic mobilities by DLS: Homodyne vs Heterodyne setup, in *2019 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)* (IEEE, 2019), pp. 315–317
15. M.A. Baranov, E.N. Velichko, A.A. Andryakov, Image processing for analysis of bio-liquid films. Opt. Mem. Neural Netw. (Inform. Opt.) **29**(1), 26–32 (2020)
16. E. Savchenko, E. Velichko, New techniques for measuring zeta-potential of colloidal system, in *Saratov Fall Meeting 2018: Optical and Nano-Technologies for Biology and Medicine. International Society for Optics and Photonics*, vol. 11065 (2019), p. 110651U
17. E.A. Savchenko, E.N. Velichko, E.T. Aksenov, E.K. Nepomnyashchaya, Combined method for laser selection, positioning and analysis of micron and submicron cells and particles, in *2018 International Conference Laser Optics* (IEEE, 2019), pp. 539–539
18. A.A. Petrov, V.V. Davydov, N.M. Grebennikova, On the potential application of direct digital synthesis in the development of frequency synthesizers for quantum frequency standards. J. Commun. Technol. Electron. **63**(11), 1281–1285 (2018)

19. N.S. Myazin, V.V. Davydov, VYu. Rud, V.V. Yushkova, V.I. Dudkin, New method for determining the composition of liquid media during the express control of their state using the nuclear magnetic resonance phenomena. J. Phys.: Conf. Ser. **1400**(6), 066008 (2019)

20. V.V. Davydov, Nuclear magnetic spectrometer for studying flows of liquid media. Meas. Tech. **59**(1), 1202–1209 (2017)

21. R.V. Davydov, M.S. Mazing, V.V. Yushkova, A.V. Stirmanov, VYu. Rud, A new method for monitoring the health condition based on nondestructive signals of laser radiation absorption and scattering. J. Phys.: Conf. Ser. **1410**(1), 012067 (2019)

22. A.S. Grevtseva, K.J. Smirnov, V.V. Davydov, VYu. Rud, Development of methods for results reliability raise during the diagnosis of a person's condition by pulse oximeter. J. Phys.: Conf. Ser. **1135**(1), 012056 (2018)

23. V.V. Davydov, E.N. Velichko, N.S. Myazin, V.Y. Rud', A method for studying the magnetic susceptibility of colloidal solutions in ferrofluidic cells. Instrum. Exp. Tech. **61**(1), 116–122 (2018)

24. V.V. Davydov, V.I. Dudkin, N.S. Myazin, V.Y. Rud', On the possibility of studying condensed media in the express mode using the nuclear-magnetic-resonance method. Instrum. Exp. Tech. **61**(1), 140–147 (2018)

25. A.I. Zhernovoi, A.A. Komlev, S.V. D'yachenko, Magnetic characteristics of MgFe2O4 nanoparticles obtained by glycine–nitrate synthesis. Tech. Phys. **61**(2), 302–305 (2016)

26. S.E. Logunov, V.V. Davydov, M.G. Vysoczky, M.S. Mazing, New method of researches of the magnetic fields force lines structure. J. Phys.: Conf. Ser. **1038**(1), 012093 (2018)

27. K.J. Smirnov, V.V. Davydov, S.F. Glagolev, N.S. Rodygina, N.V. Ivanova, Photocathodes for near infrared range devices based on InP/InGaAs heterostructures. J. Phys.: Conf. Ser. **1038**(1), 012102 (2018)

28. S.E. Logunov, VYu. Rud, R.V. Davydov, A.V. Moroz, K.J. Smirnov, Optical method for studying the magnetic field structure. J. Phys.: Conf. Ser. **1326**(1), 012024 (2019)

29. S.E. Logunov, V.V. Davydov, M.G. Vysoczky, O.A. Titova, Peculiarities of registration of magnetic field variations by a quantum sensor based on a ferrofluid cell. J. Phys.: Conf. Ser. **1135**(1), 012069 (2018)

30. S.E. Logunov, M.G. Vysoczky, V.V. Davydov, O.A. Titova, On the possibility of using the photometric method for monitoring the position of a mobile marine object, in *2018 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech 2018)*, vol 8564402 (2018), pp. 282–284

31. K.J. Smirnov, V.V. Davydov, Y.V. Batov, InP/InGaAs photocathode for hybrid SWIR photodetectors. J. Phys.: Conf. Ser. **1368**(2), 022073 (2019)

32. S.E. Logunov, AYu. Koshkin, V.V. Davydov, Quantum autonomous magnetic field sensor. J. Phys.: Conf. Ser. **1124**(1), 041025 (2018)

33. N.S. Myazin, Yu.I. Neronov, V.I. Dudkin, V.V. Davydov, A.A. Petrov, On the need for express control of the quality of consumer goods within the concept 'Internet of things'. IOP Conf. Ser.: Mater. Sci. Eng. **497**, 012111 (2019)

34. N. Grebenikova, A. Moroz, M. Bylina, M. Kuzmin, Remote control of the quality and safety of the production of liquid products with using fiber-optic communication lines of the Internet. IOP Conf. Ser.: Mater. Sci. Eng. **497**(1), 012109 (2019)

35. A.A. Moroz, R.V. Davydov, V.V. Davydov, A new scheme for transmitting heterodyne signals based on a fiber-optical transmission system for receiving antenna devices of radar stations and communication systems, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11660 (LNCS, 2019), pp. 710–718

36. A.V. Moroz, K.Y. Malanin, A.A. Krasnov, VYu. Rud, Features of the construction of the noise compensation circuit of a small-sized active phased antenna array. J. Phys.: Conf. Ser. **1400**(4), 044009 (2019)

37. A.V. Moroz, V.V. Davydov, Fiber-optical system for transmitting heterodyne signals in active phased antenna arrays of radar stations. J. Phys.: Conf. Ser. **1368**(2), 022024 (2019)

38. A.V. Moroz, K.Y. Malanin, A.A. Krasnov, Development of a compensation system based on Horn Antennas for an Active Phased Antenna Array, in *Proceedings of the 2019 Antennas Design and Measurement International Conference (ADMInC-2019)*, vol. 8969090 (Saint-Petersburg, 2019), pp. 114–116
40. G.A. Pchelkin, V.B. Fadeenko, Features of the transmission of microwave signals at offshore facilities. J. Phys.: Conf. Ser. **1368**(2), 022045 (2019)
40. V.B. Fadeenko, V.A. Kuts, D.A. Vasiliev, V.V. Davydov, New design of fiber-optic communication line for the transmission of microwave signals in the X-band. J. Phys.: Conf. Ser. **1135**(1), 012053 (2018)

Part III
Photonics and Optical Information

Chapter 43
Determination of Electrophoretic Mobility of Magnetic Fluids by Electrophoretic Light Scattering Method

Ekaterina A. Savchenko [ID]**, Olga I. Tkach, Alexey N. Skvortsov** [ID]**, and Peter A. Nekrasov**

Abstract The electrophoretic mobility of magnetic fluids was determined by electrophoretic light scattering approaches. Parameters of magnetic liquids such as particle size, electrophoretic mobility and zeta potential have been studied, which are particularly important for the study of binding capacity, solution stability and the tendency to coagulation. The experimental setup for measuring electrophoretic mobility is presented. The results obtained with our setup were compared with data measured with a commercially available Zetasizer Nano instrument. The results of an experimental study of the parameters of magnetic fluids are discussed.

Keywords Magnetic fluids · Dynamic light scattering · Electrophoresis · Electrophoretic light scattering

43.1 Introduction

Surface functionalized magnetic nanoparticles have been widely used in a range of biological applications [1–3]. Magnetite Fe_3O_4 is an easily degradable material, and it is therefore useful in bioseparation and catalytic processes. Magnetite nanoparticles have also been extensively studied in biomedicine [4] due to their superparamagnetic properties, high biocompatibility, and lack of toxicity to humans. Raw magnetite nanoparticles possess high surface energy and thus tend to aggregate quickly. Such strong aggregation degrades their adsorption properties and magnetic efficiency; so the nanoparticles are frequently coated with organic or inorganic layers to prevent aggregation. Such coatings not only stabilize the magnetite nanoparticles but can be easily used for further functionalization.

E. A. Savchenko (✉) · O. I. Tkach · A. N. Skvortsov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

P. A. Nekrasov
Smorodintsev Research Institute of Influenza, Ministry of Healthcare of the Russian Federation, St. Petersburg, Russia

In this case, it is of great importance to study the mechanisms of charge transfer in magnetic fluids and features of the flow of electrokinetic phenomena. There are several methods to study electrokinetic phenomena. The method that is used in the present work is electrophoretic light scattering, which combines the phenomena of dynamic light scattering and electrophoresis [2]. This method was chosen due to short time of measurements, minimal sample preparation; the method is also non-destructive and does not damage the sample.

43.2 Techniques

Electrophoretic light scattering (ELS) is based on measuring the fluctuations of scattering intensity from colloidal nanoparticles [5]. The total motion of the dispersed particles is combined from random Brownian motion, directed motion of colloidal particles in electric field (electrophoresis), and electroosmotic flow of the liquid. In the electric field the particles of the sample gain a certain velocity named electrophoretic velocity, which is proportional to the applied electric field. The proportionality coefficient is electrophoretic mobility, which is an important characteristic property of the particles.

Measurements were made by recording of autocorrelation characteristics (ACF) of the light scattered by moving particles [5]. Uncharged particles do not contribute in this case. The measured signal results from two types of motion from all the particles in the light scattering volume. The ACF is a product of cosine Function (43.1) exponential decay function

$$G(\tau) = Ae^{-Dq^2\tau} \cdot e^{iq\mu E \cdot cos\theta/2},
\tag{43.1}$$

where the first factor is related to random motion, D is diffusion coefficient. The second factor corresponds to directed motion in the electric field, E is electric field, μ is electrophoretic mobility, and q is scattering vector.

As ACF was a known function, it was possible to define the oscillation period Δt, which yielded the value of electrophoretic mobility μ through Eq. (43.2)

$$\Delta t = \frac{2\pi}{\mu E q cos\theta/2}
\tag{43.2}$$

where q is scattering vector, θ is the scattering angle.

If all of parameters are known, we can calculate zeta potential using formula

$$\zeta = 3\frac{\mu\eta}{2f(ka)\varepsilon\varepsilon_0}
\tag{43.3}$$

where η is dynamic viscosity of sample, $f(ka)$ is Henry function, k is thickness of double layer (Stern layer), a is radius of the particle, ε_0 vacuum permittivity, ε absolute permittivity. We used Smoluchowski theory [6, 7], which considers $f(ka)$ $\rightarrow 1$.

Two measurement modes exist: homodyne and heterodyne. In heterodyne mode the interference pattern between the scattered light and reference beam with predefined frequency shift is analyzed. However, the influence of the electric field on colloidal solution produces some difficulties for the heterodyne mode. Parasitic effects like electroosmotic flow and thermal convection due to Joule heating make it very difficult to match the phases of the two beams work [8–10].

Therefore, the homodyne mode was used in this work for simplicity because there is no need for additional items in scheme forming a direct beam. Also, it is not necessary to reconcile the wavefront of the direct and scattered beams, which is a difficult task and requires careful control. A high-sensitivity photodetector and high sampling frequency was used for noise resistance.

In the future, we plan to develop a method for matching wavefronts [9, 11–13] and create an installation for the heterodyne method.

43.3   Experimental Setup

The experimental setup was developed in accordance with the selected method. Its main structural elements are presented in Fig. 43.1.

The beam of a 10 mW distributed Bragg reflector laser with wavelength of 633 nm was passed through the pinhole (to eliminate the side modes of the laser) and was focused using a short-focus lens on the sample sell. The diameter of the irradiation
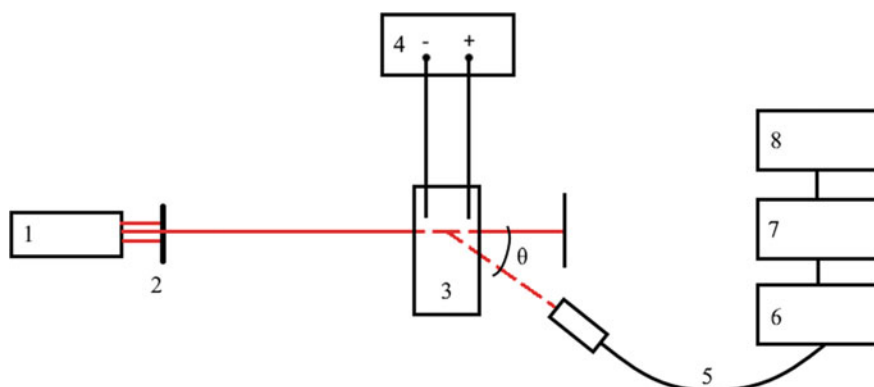
Fig. 43.1 Setup for measuring the electrophoretic light scattering: 1—laser, 2—diaphragm (pinhole), 3—electrodes and sample, 4—power supply, 5—optical fiber, 6—photomultiplier, 7—ADC, 8—computer

spot behind the lens was estimated by the following formula:

$$D_3 = \frac{1.22 \ast f}{D_2} \tag{43.4}$$

where f is the focal length of the lens, D_2 is spot size before the lens. In the present case the beam diameter at the focus of the lens comprised 80 μm. The radiation was focused inside the cell with the investigated liquid, where it was partially scattered. The scattered light was captured by optical fiber [14–17] at the angle θ = 15° (larger scattering angles were not used because of a stronger influence of diffusion on light scattering) and detected by a photomultiplier. The signal from the photomultiplier was digitized by an A/D converter board [15] with sampling frequency of 100 kHz. One continuous series comprised 3 s, which made it possible to detect fluctuations resulting from the electrokinetic motion. These fluctuations had larger oscillation period than the fluctuations related to Brownian motion (Doppler frequency is less than 100 Hz).

43.4 Results

The objects of our studies were samples of magnetite Fe_3O_4 suspended in water. The average size of the particles was about 100 nm and typical concentrations comprised 0.02%. Concentration of solution was selected to provide the necessary amount of 500–1000 particles in the scattering volume. The ferrofluids samples were thoroughly mixed prior the measurements. The dynamic viscosity was measured for temperature 25 °C. The autocorrelation functions measured from ferrofluids sample with and without external electric field are presented in Fig. 43.2.

We can see the periodic ACF under the influence of the electric field [18, 19], while without electric field ACF is decaying exponential. For following calculation electrophoretic mobility by formula (43.2) we use an average period of ACF about 0,03 s shown at Fig. 43.2. Next, we compared the results obtained using the developed experimental setup and the measurements by commercial device Zetasizer Nano (Malvern Panalytical, United Kingdom). The obtained values of electrophoretic mobility and zeta-potential are presented in Table 43.1. Also, we need to match values of zeta potential with a well-known line between stable and unstable suspensions which taken as + 30 or −30 mV [20, 21].

The data obtained by the experimental setup developed in the present work match the data from the commercial device Zetasizer Nano within the acceptance error. In homodyne mode we cannot determine the sign of electrophoretic mobility [11] and therefore we can compare values only by absolute value.

The difference in the obtained results is due to different experimental schemes for the homodyne and heterodyne modes, and insufficient control of the dynamic

Fig. 43.2 Autocorrelation functions of ferrofluid samples without external electric field (**a**), and in external electric field of 25 V/cm (**b**)

a)

b)
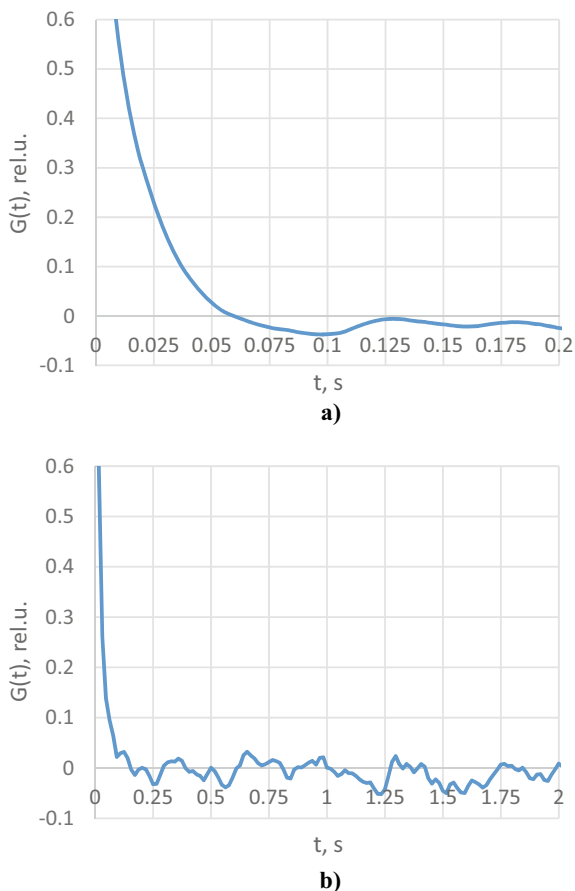
Table 43.1 Electrophoretic mobility and zeta-potential obtained in various measurement setups

Parameters	Developed experimental setup	Commercial device Zetasizer Nano
Electrophoretic mobility, μm $\cdot$ cm/s $\cdot$ V	2.13 ± 0.50	-2.11 ± 0.30
Zeta-potential, mV	-26.40 ± 1.28	-26.90 ± 3.89

viscosity of the sample and its temperature dependence. In general, the expected zeta-potential values [22–25] of ferrofluids have been successfully received.

43.5   Conclusion

In this work we developed an experimental setup for measuring electrophoretic light scattering in homodyne mode. We were able to determine electrophoretic mobility for a solution of magnetite nanoparticles (Fe_3O_4) and the zeta potential, which is important for the solution stability. The data obtained by the developed experimental setup match the data from the commercial device Zetasizer Nano within the acceptance error. According to resulting values, ferrofluid of Fe_3O_4 in aqueous medium is acceptable for using as a medical agent for bloodstream, and it is also a potential transport or diagnostic agent.

References

1. E.K. Nepomnyashchaya, A.V. Prokofiev, E.N. Velichko, I.V. Pleshakov, Y.I. Kuzmin, Investigation of magneto-optical properties of ferrofluids by laser light scattering techniques. J. Magn. Magn. Mater. **431**, 24–26 (2017)
2. E. Mirkin, E. Savchenko, E. Savchenko, Use of the method of configuring the interval target in the problem of synthesis of the neural network classifier for diagnosing cardiovascular diseases, in *2020 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)* (Saint Petersburg, 2020), pp. 140–144
3. V.V. Tuchin, Tissue optics: light scattering methods and instruments for medical diagnosis, in *SPIE Tutorial Texts in Optical Engineering TT38* (Bellingham, WA, USA, 2007), p. 882
4. L. Vékás, Ferrofluids and Magnetorheological Fluids. Adv. Sci. Technol. **54**, 127–136 (2008)
5. M. Medrano, A.T. Perez, L. Lobry, F. Peters, Electrophoretic mobility of silica particles in a mixture of toluene and ethanol at different particle concentrations. Langmuir **25**(20), 12034 (2009)
6. I.N. Serdyuk, N.R. Zaccai, *Methods in Molecular Biophysics. Structure, Dynamics, Function*, 1st edn. (Cambridge University Press, 2007)
7. A. Sikora, D. Bartczak, D. Geißler, V. Kestens, G. Roebben, Y. Ramaye, Z. Varga, M. Palmai, A. Shard, C. Minelli, A systematic comparison of different techniques to determine the zeta potential of silica nanoparticles in biological medium Analytical Methods. Anal. Methods **7**, 9835–9843 (2015)
8. S.A. Tereshchenko, P.V. Shalaev, Yu.P. Masloboev, S.A. Dolgushin, V.A. Deshabo, I.K. Yudin, Electrokinetic potential of nanorods and cells in liquid dispersions. Biomed. Eng. **50**(5), 333–338 (2017)
9. B. Robertson, J. Hubbard, Determining mobility from homodyne AC electrophoretic light scattering. J. Colloid Interface Sci. **153**(2), 459–460 (1992)
10. R. Xu, Particle characterization: light scattering methods. Part. Technol. Ser. **13**, 289–343 (2002)
11. K. Makino, H. Ohshima, Electrophoretic mobility of a colloidal particle with constant surface charge density. Langmuir **26**(23), 18016–18019 (2010)
12. V. Uskoković, Dynamic light scattering based microelectrophoresis: main prospects and limitations. J. Dispers. Sci. Technol. **33**(12), 1762–1786 (2012)

13. P. Carvalho, M. Felício, N. Santos, S. Gonçalves, M. Domingues, Application of light scattering techniques to nanoparticle characterization and development. Front. Chem. **6**(237), 1–17 (2018)
14. A. Markvart, L. Liokumovich, N. Ushakov, Fiber-Optic Fabry-Perot interferometric sensor for pulse wave registration, in *2019 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)* (Saint Petersburg, 2019), pp. 341–343
15. N. Grebenikova, V. Davydov, A. Moroz, M. Bylina, M. Kuzmin, Remote control of the quality and safety of the production of liquid products with using fiber-optic communication lines of the Internet. IOP Conf. Ser. Mater. Sci. Eng **497**(1), 012109 (2019)
16. V. Temkina, A. Medvedev, A. Mayzel, Computer simulation of the fiber optic electric field sensor. J. Phys. Conf. Ser. **1236**, 012031 (2019)
17. N.A. Ushakov, A.A. Markvart, L.B. Liokumovich, Pulse wave velocity measurement with multiplexed fiber optic Fabry-Pérot interferometric sensors. IEEE Sens. J. **20**(19), 11302–11312 (2020)
18. F.G. Halaka, Dielectrophoretic dynamic light-scattering (DDLS) spectroscopy. Proc. Nat. Ac. Sci. **100**(18), 10164–10169 (2011)
19. N. Bakir, V. Pavlov, S. Zavjalov, S. Volvenko, A. Gumenyuk, M. Rethmeier, Development of a novel optical measurement technique to investigate the hot cracking susceptibility during laser beam welding. Weld World **63**, 435–441 (2019)
20. E. Kokufuta, Light scattering and electrophoretic light scattering of biopolymers. *Encyclopedia of Biocolloid and Biointerface Science*, vol. 2, 1st edn. (John Wiley & Sons, Inc., 2016)
21. M. Larsson, A. Hill, J. Duffy, Suspension stability: Why particle size, zeta potential and rheology are important. Annu. Trans. Nord. Rheol. Soc. **20**, 209–214 (2012)
22. E.N. Velichko, E.K. Nepomnyashchaya, K.G. Gareev, J. Martínez, M.C. Maicas, Characterization of magnetite-silica magnetic fluids by laser scattering. Appl. Sci. **11**(1), 183 (2021)
23. G.L. Klimchitskaya, V.M. Mostepanenko, E.K. Nepomnyashchaya, E.N. Velichko, Impact of magnetic nanoparticles on the Casimir pressure in three-layer systems. Phys. Rev. B **99**(4), 045433 (2019)
24. J. Matusiak, E. Grządka, Stability of colloidal systems—a review of the stability measurements methods. Annales universitatis mariae curie-skłodowska Lublin **73**(1), 33–45 (2017)
25. E.N. Velichko, M.A. Baranov, V.M. Mostepanenko, Change of sign in the Casimir interaction of peptide films deposited on a dielectric substrate. Mod. Phys. Lett. A **35**(03), 2040020 (2020)

Chapter 44
Infrared Finger-Piece Sensor for ICG Concentration Measurements

Ilya Kolokolnikov [ID]**, Elena Velichko** [ID]**, Lyubov Kiseleva** [ID]**, and Victor Osovskikh** [ID]

Abstract Indocyanine green (ICG) is widely used in medicine as contrast media. It has a high absorbance peak at 805 nm, which is known to be an isosbestic point of hemoglobin. Therefore, the measurement of indocyanine green absorbance is independent of the oxyhemoglobin to deoxyhemoglobin ratio, i.e., blood oxygen saturation. Indocyanine green is unable to penetrate through cell and vascular walls and can be completely excreted by the liver in several hours, which makes ICG a safe and well localized compound. Human body tissues have a high transmittance in the infrared, and, hence, the absorbance peak in the near infrared makes ICG a very convenient solution for qualitative and quantitative optical in vivo analysis. For instance, the technologies for liver function assessment using the ICG concentration dynamics have been known for a long time and PDR (Plasma Disappearance Rate) is a predictive parameter that has proven to be highly reliable. ICG is also used to measure a cardiac output, and for ophthalmic angiography, perfusion diagnostics of tissues, and selective tissue overheating. In this paper we present a finger-piece device for measuring the ICG concentration dynamics in a patient's blood and algorithms of signal processing.

Keywords Indocyanine green · Optical sensor · Attenuation analysis · Near infrared · Colorant concentration

44.1 Introduction

Indocyanine green has been used as a contrast media since the late 1950s [1], and research on the use of its properties in diagnostic medicine is still going on [2, 3]. Indocyanine green is used for diagnostics in many areas of medicine. It is applied to reliably measure the patient's liver function [4], blood plasma volume, cardiac

I. Kolokolnikov (✉) · E. Velichko
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

L. Kiseleva · V. Osovskikh
Russian Research Center for Radiology and Surgical Technologies, St. Petersburg, Russia

output [5]. ICG is used in fluorescent imaging, e.g., angiography of blood flow in the choroid of the eye [6]. In laser surgery a contrast colorant can be used to selectively heat tissues [7].

Measurements of the ICG concentration dynamics in the patient's blood are primarily used to measure the liver function. Since ICG is almost completely eliminated by the liver, the rate of its elimination is direct diagnostic parameter indicating the ability of the liver to remove extraneous substances.

The basic method for liver function assessment using ICG is a fully invasive method of sequential blood samples that requires collection of several samples of the patient's stained blood several times after a colorant intravenous injection [8]. With further development of optical technics for colorant concentration measurements an improved method appeared. The semi non-invasive method of course involves the injection of the colorant into the patient's blood, but concentration measurements are made by a finger-piece optical sensor. Although the in vivo optical measurements have less accuracy than measurements on the blood sample in a cuvette, they can be carried out with a much higher frequency. Optical measurement methods [9, 10] and especially fluorescent analysis [11–13] have a variety of applications in diagnostics and medical practice.

Some devices that can be used in medical practice for in vivo ICG concentration measurements exist nowadays. They are, for instance, LiMON (Pulsion Medical Systems, Germany) and DDG 2001 (Nihon Kohden, Japan). But the only device available in Russian Federation—LiMON—does not provide the full data set of measurements. An incorrect measurement of the required parameter (PDR) negates the advantage in rapidity of the non-invasive test—a repeat can be performed only after a few hours, when the colorant is completely removed from the blood. A certain amount of incorrect measurements by LiMON and significant cost of a single measurement necessitates the development of the device that provides a full data set to a doctor.

44.2 Materials and Methods

Indocyanine green is a tricarbocyanine colorant—a sodium salt which is normally available in the powder form and can be dissolved in various solvents, 5% sodium iodide is usually added to ensure a better solubility. In several first seconds after injection ICG becomes completely bounded to blood plasma proteins (globulin, $\alpha1$-lipoproteins) and distributes within blood volume in 2–3 min [14]. ICG is almost exclusively captured by the liver and then excreted to the bile without entering the entero-hepatic circulation [15]. Being bounded to plasma proteins ICG has a sharp absorption peak at 805 nm—in a shortest wavelength infrared range, where water and, accordingly, most human body tissues, have a high transmittance.

The colorant concentration measurement technique is based on absorption analysis using the Beer-Lambert-Bouguer law for light attenuation in an absorbing medium [16]. The intensity of light transmitted through an absorbing medium is
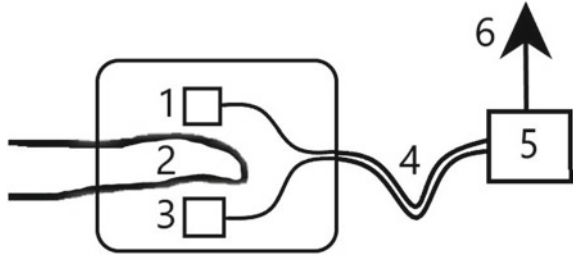
described by an exponential dependence on the optical path and on the concentration
of the absorbing substance if the medium is a solution.

$$I = I_0 \cdot 10^{-A} \tag{44.1}$$

where I is the intensity of the light transmitted through the absorbing medium, I_0
is the reference point—intensity of the light transmitted through the medium before
colorant deposition, and A is the absorbance.

$$A = \varepsilon c z \tag{44.2}$$

where ε is the absorptivity, coefficient describing interaction of light and solvent for
direct transmission, c is the solute concentration, and z is the optical path length.

After the selection of elements and measurements on a laboratory model, which
were described in earlier reports [17, 18] we created a finger-piece device (Fig. 44.1)
to measure the ICG concentration dynamics in vivo.

A photodetector provides information about the transmitted light intensity. The
intensity measured before colorant injection defines reference point I_0. After cutting
off the first-pass peaks, we leave only the elimination area, in which we recalculate
the absorbed light intensity to absorbance which is proportional to concentration.
PDR is measured in percent per minute and can be found as a slope factor of the
logarithmized absorbance curve.

44.3   Results

Our device was placed on the next finger to a LiMON sensor, for parallel measure-
ments, during a usual ICG clearance test. The data obtained are presented in
Fig. 44.2.

The data acquisition was performed with a period of 400 ms. The signal
was distorted by high frequency component created by scattering in the dynamic
inhomogeneous medium and the variable quantity of the contrast medium in the
capillaries.
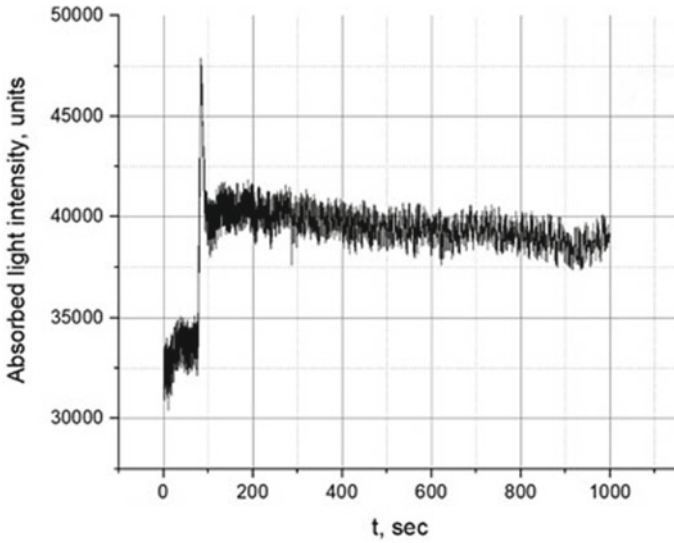
Fig. 44.2 Obtained signal

On the plot of the signal, we can see the first-pass peak, after the moment of injection, when absorbed light intensity grows sharply. Then the first recirculation peak can be identified, it represents the wave of plasma with a high ICG concentration and shows while colorant is still not distributed throughout the blood volume.

After recalculating the absorbed light intensity data to absorbance units the elimination area was logarithmized once again to find PDR.

$$\frac{A}{A_1} = \frac{c}{c_1} = \exp(-0.01 \cdot PDR \cdot t), \tag{44.3}$$

where A_1 and c_1 are the absorbance and concentration in the region of stabilized level of absorbed light intensity after the recirculation peak—90 to 150 s from the beginning of the measurement, PDR is the plasma disappearance rate—the indicator of liver function, t is the time.

PDR can be found as a slope of the straight line:

$$PDR \cdot t = -100 \ln\left(\frac{A}{A_1}\right) \tag{44.4}$$

By using the least squares linearization we got PDR = 2.34%/min.

44.4   Conclusion

A finger-piece device for in vivo measurements of the indocyanine green (ICG) concentration dynamics has been developed. Dynamic measurements of the ICG concentration in a patient's blood allow one to assess the liver function expressed as a referent parameter PDR (Plasma Disappearance Rate), which has already proved to be a reliable prognostic parameter for resection and transplantation procedures.

The colorant concentration was determined by using the absorption analysis method. Being bounded with plasma proteins indocyanine green has the absorption peak at 805 nm, which makes measurements independent of the saturation level.

The first measurements on a patient were performed. Further research will be devoted to reducing noise in determining the required parameter and collecting statistics. Research for the further method progress is related to the practical experience in fluorescence measurements [19, 20] and data matching of ICG concentration dynamics and blood flow velocity [21].

References

1. I.J. Fox, L.G. Brooker, D.W. Heseltine, H.E. Essex, E.H. Wood, A tricarbocyanine dye for continuous recording of dilution curves in whole blood independent of variations in blood oxygen saturation. Proc Mayo Clim **32**, 478–484 (1957)
2. A.J. Ruiz, M. Wu, E.P.M. LaRochelle, D. Gorpas, V. Ntziachristos, T.J. Pfefer, B.W. Pogue, Indocyanine green matching phantom for fluorescence-guided surgery imaging system characterization and performance assessment. J. Biomed. Opt. **25**(5), 056003 (2020)
3. M. Abdelaziz, I. Krejci, D. Fried, Enhancing the detection of proximal cavities on near infrared transillumination images with Indocyanine Green (ICG) as a contrast medium: in vitro proof of concept studies. J. Dentistry **91**, 103222 (2019)
4. S.G. Sakka, A. Meier-Hellmann, K. Reinhart, Prognostic value of indocyanine green-plasma disappearance rate in critically ill patients, in *Critical Care*, vol. 6(1). (BioMed Central, 2002), pp. 1–1
5. T. Iijima, T. Aoyagi, Y. Iwao, J. Masuda, M. Fuse, N. Kobayashi, H. Sankawa, Cardiac output and circulating blood volume analysis by pulse dye-densitometry. J. Clin. Monit. **13**(2), 81–89 (1997)
6. P.E. Stanga, J.I. Lim, P. Hamilton, Indocyanine green angiography in chorioretinal diseases: indications and interpretation: an evidence-based update. Ophthalmology **110**(1), 15–21 (2003)
7. W.R. Chen, R.L. Adams, A.K. Higgins, K.E. Bartels, R.E. Nordquist, Photothermal effects on murine mammary tumors using indocyanine green and an 808 nm diode laser: an in vivo efficacy study. Cancer Lett. **98**(2), 169–173 (1996)
8. A.W. Hemming, C.H. Scudamore, C.R. Shackleton, M. Pudek, S.R. Erb, Indocyanine green clearance as a predictor of successful hepatic resection in cirrhotic patients. Am. J. Surg. **163**(5), 515–518 (1992)

9. E.K. Nepomnyashchaya, E.N. Velichko, E.T. Aksenov, T.A. Bogomaz, Laser correlation spectroscopy for immune testing, in *International Conference Laser Optics (ICLO)*. (IEEE, St. Petersburg, Russia, 4–8 June 2018), p. 561

10. R.V. Davydov, V.I. Antonov, V.V. Yushkova, V.V. Davydov, Y. Rud', K.J. Smirnov, A new method of processing a pulse wave in rapid diagnosis of the human health. J. Phys. Conf. Ser. **6**, 066037 (2019)

11. M.L. Kuranova, T.A. Ledashcheva, E.K. Tulush, D.L. Beliaev, S.V. Zherebtsov, N.M. Pleskach, I.M. Spivak, Diagnostics of ataxia-telangiectasia by the express-test found on the method of indirect immunofluorescence. Tsitologiia **55**(8), 560–565 (2013)

12. A.I. Sulatskaya, A.V. Lavysh, A.A. Maskevich, I.M. Kuznetsova, K.K. Turoverov, Thioflavin T fluoresces as excimer in highly concentrated aqueous solutions and as monomer being incorporated in amyloid fibrils. Sci. Rep. **7**(1), 1–11 (2017)

13. A.G. Pershina, O.Y. Brikunova, A.M. Demin, M.A. Abakumov, A.N. Vaneev, V.A. Naumenko, L.M. Ogorodova, Variation in tumor pH affects pH-triggered delivery of peptide-modified magnetic nanoparticles. Nanomed. Nanotechnol. Biol. Med. **32**, 102317 (2021)

14. A. De Gasperi, E. Mazza, M. Prosperi, Indocyanine green kinetics to assess liver function: ready for a clinical dynamic assessment in major liver surgery? World J. Hepatol. **8**(7), 355 (2016)

15. E. Levesque, E. Martin, D. Dudau, C. Lim, G. Dhonneur, D. Azoulay, Current use and perspective of indocyanine green clearance in liver diseases. Anaesthesia Critical Care Pain Med. **35**(1), 49–57 (2016)

16. T.G. Mayerhöfer, J. Popp, Beer's law derived from electromagnetic theory. Spectrochimica Acta Part A Molecular Biomolecular Spectroscopy **215** (2019-05-15) (in German)

17. E. Savchenko, I. Kolokolnikov, E. Velichko, V. Osovskikh, L. Kiseleva, Z. Musakulova, Design of liver functional reserve estimation technique based on optical densitometry. Diagnostics **10**(8), 599 (2020)

18. I. Kolokolnikov, E. Savchenko, E. Velichko, Spectrophotometry setup for solute concentration determination, in *2020 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*. IEEE (2020, October), pp. 258–260

19. E.A. Savchenko, E.K. Nepomnyashchaya, D.B. Dyubo, E.N. Velichko, O.Y. Tsybin, Registration of fluorescence in biomolecular solutions using dynamic pin photodiode. J. Phys. Conf. Ser. **917**(4), 042010 (2017, November). IOP Publishing

20. E.A. Savchenko, A.A. Andryakov, E.N. Velichko, Molecular dynamics of rhodamine 6G solutions as revealed by the computer processing of fluorescence microscopy images. Opt. Memory Neural Netw. **29**(3), 157–164 (2020)

21. E.A. Savchenko, E.N. Velichko, The use of speckle correlation analysis to determine blood flow velocity. Opt. Spectrosc. **128**(7), 998–1003 (2020)

Chapter 45
Ring-Shaped Contamination Detection System

Yakov Kovach, Artem Petrenko, Stanislav Rochas, Daniil Shiryaev, Alexey Borodkin, and Evgeni Kolodeznyi

Abstract The paper describes designed and fabricated ring-shaped contamination detection system based on the effect of total internal reflection. The design of the detection system was determined by optoelectronic and thermal phenomena computer simulations. The performance of the assembled system was confirmed experimentally and the input range for setting the system sensitivity was ~10%.

45.1 Introduction

Optical navigation, control and monitoring systems efficiency may be reduced during outdoors operation due to contamination of their components [1–6]. Global informatization and automation are bound with a growing demand of new types of sensors such as lidars [7, 8]. For complete sensor autonomy it is necessary to detect and remove contamination from sensor's surface due to a contamination's negative affection [1–6]. The contamination of transmitting and receiving optical devices can be controlled with use of additional integrated optical detectors. The use of such small dimension optical detectors in bound with cleaning unit control systems forms a fully automated cleaning device which can be used in combination with different types of navigation systems.

An automotive transport, the number of which will increase significantly on public roads according to the various news agencies estimates [9], is a promising area for the described detectors application. In recent years, autonomous cars started to develop rapidly, and such aspect as the disruption of their navigation systems requires attention as it poses a danger to road users [10]. Optical detectors utilization allows one to register lidar and video cameras, which are commonly used in autonomous cars navigation, contamination and take timely measures to clean up contamination

Y. Kovach (✉) · A. Petrenko · S. Rochas · D. Shiryaev · A. Borodkin · E. Kolodeznyi
ITMO University, St. Petersburg 197101, Russia
e-mail: j-n-kovach@itmo.ru

using automatic cleaning systems. Additionally, such sensors can be used as a rain sensor for mobile vehicles and for stationary observation posts.

There are several types of detectors that respond to rain drops on the windshield of the car had been developed in the past to increase vehicles safety and reliability. The performance of such rain sensors can be based on changes in electrical circuit capacitance [11, 12] or resistivity [13]. In capacitance sensors contamination interferes with the fringe field, which causes a change in the sensor capacitance. Such sensors can detect the position and type of an interference material. Resistive sensors act as a potentiometer the change in resistance of which depends on the amount of liquid contamination on the surface. There are also optical detectors based on detecting the radiation intensity changing due to the total internal reflection phenomena [14–16]. The optical system consists of an LED and a photodiode for a radiation and a signal detection, which are positioned that the light signal of the light source falls on a glass surface between the critical angle for the transition of radiation from glass into air and a critical angle for the transition of radiation from glass to water. If there is contamination on the glass surface, the light is partially scattered into the external environment, while the signal at the photodiode decreases.

45.2   System Description

The paper describes the scalable cylindrical-shaped optical contamination detector. The detection system consists of two printed circuit boards (PCB) for an optical signal emission and detection. The emitting PCB places 16 GaAs-based 50 mW infrared (IR) light-emitting diodes (LED). The detection PCB places 16 infrared phototransistors with resistors and operational amplifiers. Both PCBs are placed between an optical part in the form of a ring with a complex profile. The optical part designed as a detection system's casing with a 100 mm diameter, which corresponds to the diameter of the most common commercial versions of lidars on the market.

The operating principle of the detection system is based on the effect of total internal reflection of infrared LEDs radiation [15, 16]. Contamination registration is carried out according to the level of the LEDs optical signal detecting by phototransistors. LEDs and phototransistors are divided into 4 working autonomous sectors of the same area. The signal inside of each sector is transmitted to an operational amplifier, which acts as an adder, and to the AVR microcontroller analog inputs for subsequent signal transmission to cleaning systems.

45.3   Experimental Results

An optical part of the detection system was simulated in Zemax software. Five million rays were considered in every simulation. The following optical part heights: 20, 30 and 40 mm were studied. For each height wall thicknesses were modeled from 5 to

2 mm with a 1 mm step. Due to the estimation of the influence of the optical system dimensions on the maximum illumination in the photodetector plane the following parameters were selected: height of 40 mm, wall thickness of 2 mm.

Three different profiles of the optical part were proposed and studied: a profile with a prism, a profile with curved prism edges, and a profile with prism and microlenses (Fig. 45.1a). Although the profile with prism and microlenses is preferable in terms of maximum illumination at the point, the profile with curved prism edges is preferable due to manufacturing resource intensity and a larger area of illumination (Fig. 45.1b). The integrated optical power is slightly different between both profiles and twice as high as for the profile with a prism.

Figure 45.2 shows the dependence of the photodetector mean illumination on the IR LED optical power for a clean and contaminated outer surface of the optical part, where 1, 2 are the dependences for the profiles with curved prism edges with and without water layer, respectively, 3, 4 are the dependences for the profiles with a prism with and without water layer, respectively. The mean illumination change with contamination for the profile with curved prism edges is 29% and for the profile with a prism is 12%.

The simulation of the profile with curved prism edges without the water layer on the surface (Fig. 45.3a) and with the water layer (Fig. 45.3b) with use of 50 mW optical power radiation source demonstrates that the presence of an additional layer shifts the radiation beam, thus the maximum illumination in the sensitive area of the photodetector decreases from 165 to 140 mW cm^{-2}, which corresponds to the 15% of losses.

The influence of water droplets location was also investigated. An array of one hundred hemispheres with a diameter of 1 mm was set on the optical part outer surface. Herein 2×10^7 emanating rays were considered while 1.56×10^7 rays reach the photodetector's plane, which corresponds to the 22% of losses. Thus, we
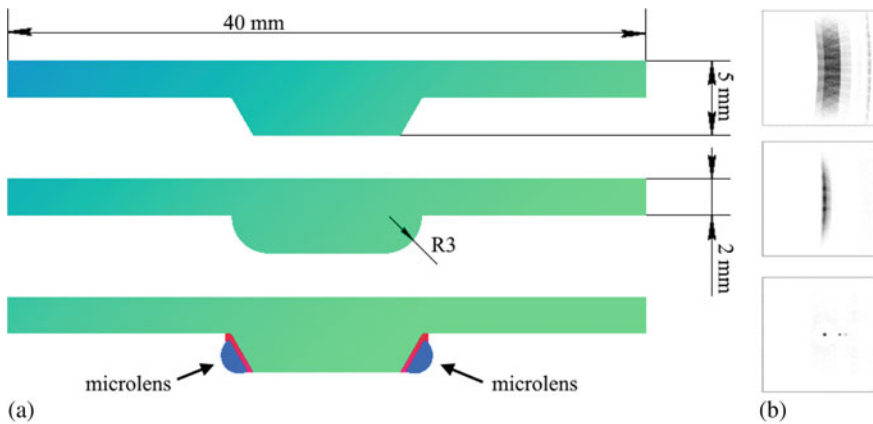
Fig. 45.1 a Types of simulated profiles: with a prism, with curved prism edges, with prism and microlenses; **b** corresponding illumination in the photodetector plane

Fig. 45.2 The dependence
of the photodetector mean
illumination on the infrared
light-emitting diode optical
power

(a)

(b)
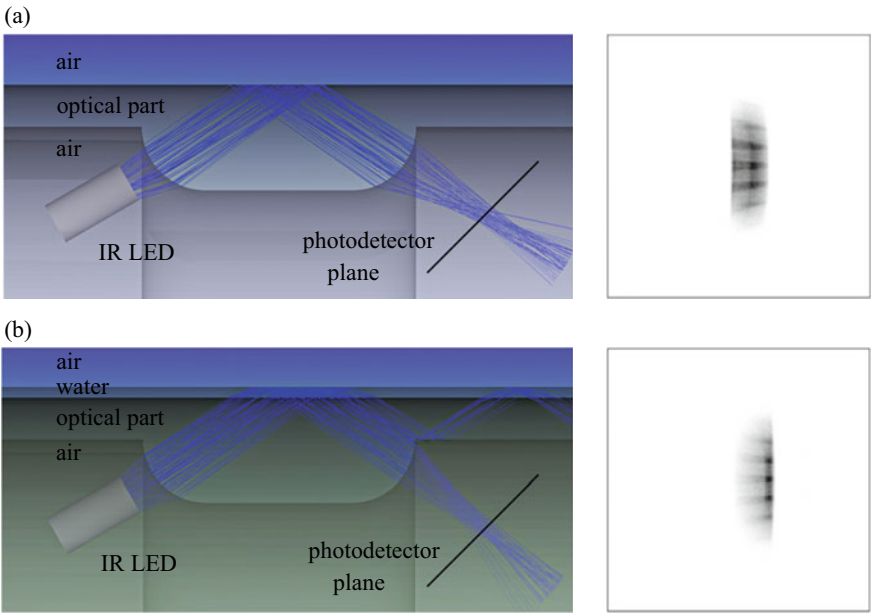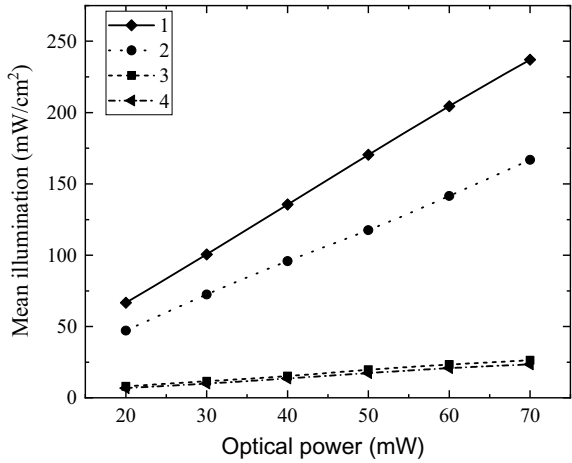
Fig. 45.3 **a** Ray diagram and illumination in the photodetector plane for the profile with curved
prism edges without additional water layer; **b** with a water layer on the outer surface

can assume that the optical part designed from the profile with curved prism edges
can be used as a rain sensor.

The contamination detection of the optical part outer surface implies the installa-
tion of several phototransistors and LEDs inside of the ring-shaped casing. For the
contamination detection not less than 50% of the outer surface of the optical part

sixteen LEDs were taken. For a 5 mm GaAs LED the illuminated spot diameter on the optical part is 11.34 mm, which corresponds to 13°1' of the ring-shaped casing. Thus, the use of sixteen LEDs overlaps 208°16' of the optical part which allows to detect 58% of the outer surface contamination.

To simulate thermal processes of the detection system using the finite element method, COMSOL Multiphysics was used. The quarter of the computational domain was considered, and a 40 mW heat source was specified within each GaAs LED. Convective heat exchange was set on surfaces adjacent to an air. The constructed computational grid is shown in Fig. 45.4a. The maximum temperature, equal to 333.8 K, was reached on the GaAs LED chip (Fig. 45.4b).

The estimation of the ring-shaped detection system (Fig. 45.5a) operability was carried out according to the Fig. 45.5b at standard conditions. The following devices and elements were used: power supply "KEYSIGHT N6705C", voltmeter "GDM-78342", the detection system.
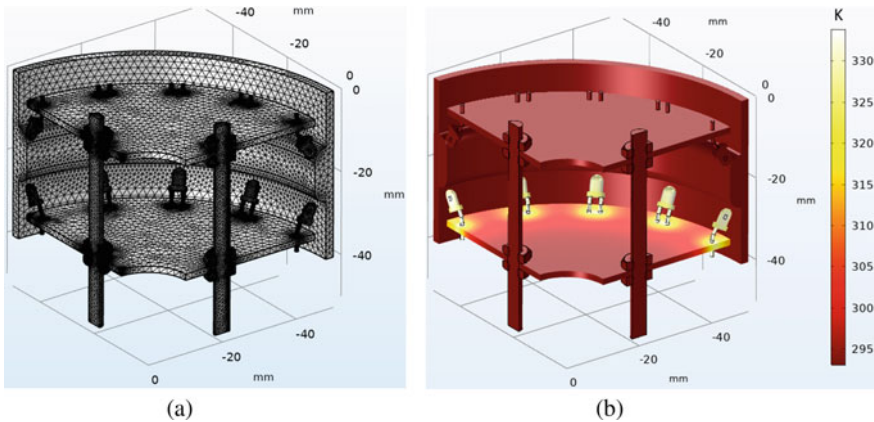
Fig. 45.4 **a** Constructed computational grid of the detection system; **b** temperature distribution of the detection system
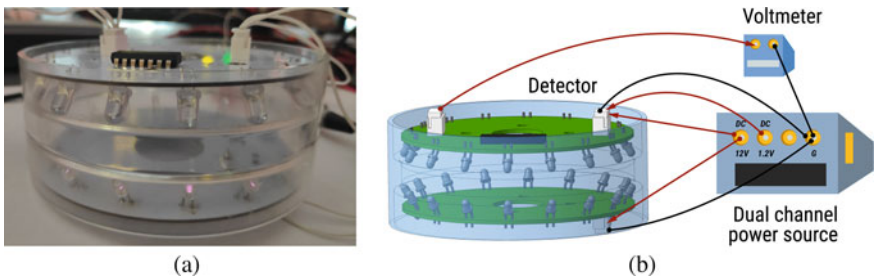
Fig. 45.5 **a** Photo of the ring-shaped detection system; **b** detection system measurement circuit

Table 45.1 Photodetector mean output voltage

	Sector 1	Sector 2	Sector 3	Sector 4
Output voltage with a clean optical part, V	3.37	3.41	3.36	3.39
Output voltage with contaminated optical part, V	3.05	3.1	2.95	3.05
Relative voltage difference, %	9.5	9.1	12.2	10.0

The voltage at the terminals of the operational amplifiers was measured for a clean and contaminated outer surface of the optical part, measurement results are presented in Table 45.1.

The PCB based on AVR microcontroller was designed and fabricated. The input signal from the detecting system arrives to the analog-digital converter (ADC) of AVR microcontroller. ADC of used AVR controller has a resolution of 0.005 V, which allows detecting a voltage drop at the outputs of the operational amplifier. From the Table 45.1 mean relative voltage difference equal 10.2% or 0.35 V, thus the AVR microcontroller can detect voltage drop from the optical part outer surface contamination. The algorithm of signal processing turns on digital outputs of the AVR controller as the instantaneous value of an input signal differs from the initial value, the average value of the input signal in first 10 s after the detection system is turned, by 5%.

The output signal of AVR microcontroller activates relays scheme based on relays and bipolar transistors which start the cleaning cycle using the electric drive and using the water pump. The duration of a complete cleaning cycle is 7 s, after that microcontroller turns off the relays scheme. For the supply voltage of the circuit (12 V) and the supply voltage of the phototransistors (1.2 V) the step-down DC-DC converter TPS5622 is used.

45.4 Conclusion

The study of the optoelectronic and thermal phenomena allowed us to define the preferable parameters for the detection system design in terms of application with the cleaning system. The use of PCB with 16 IR GaAs LEDs with an optical output power of 50 mW and PCB with 16 phototransistors provides contamination detection of the 58% optical part outer surface area. Due to the fabricated PCB with an AVR microcontroller the automation detection of 4 detection system separate sectors contamination is provided.

The detection system operability is confirmed experimentally at standard conditions and the photodetectors relative voltage difference ~10% provides a range to adjust the sensitivity of the detection system. The obtained detection system can be used with the majority of light-transmitting cylinder-shaped devices, such as autonomous vehicle lidar due to its dimensions and scalability.

References

1. J. Rivero, I. Tahiraj, O. Schubert, C. Glassl, B. Buschardt, M. Berk, J.. Chen, Characterization and simulation of the effect of road dirt on the performance of a laser scanner, in *IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)* (IEEE, 2017), pp 1–6

2. S. Abdul, A. Ali, K. Follivi, K. Sousso, A. Kodjo, Z. Nadjet, The perception system of intelligent ground vehicles in all weather conditions: a systematic literature review. Sensors **20**(22), 6532 (2020)

3. S. Zang, M. Ding, D. Smith, P. Tyler, T. Rakotoarivelo, K. Ali, The impact of adverse weather conditions on autonomous vehicles. IEEE Veh. Technol. Mag. **14**(2), 103–111 (2019)

4. M. Byeon, S. Yoon, Analysis of automotive lidar sensor model considering scattering effects in regional rain environments. IEEE Access **8**, 102669–102679 (2020)

5. D. Vriesman, B. Thoresz, D. Steinhauser, A. Zimmer, T. Brandmeier, An experimental analysis of rain interference on detection and ranging sensors, in *IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)* (IEEE, 2020), pp 1–5

6. N. Goberville, M. El-Yabroudi, M. Omwanas, J. Rojas, et al., Analysis of LiDAR and camera data in real-world weather conditions for autonomous vehicle operations. SAE Int. J. Adv. Curr. Prac. Mob. **2**(2020-01-0093), 2428–2434 (2020)

7. L. Maleki, Advances in LiDAR: the autonomous vehicle perspective, in *OSA Laser Congress (ASSL, LAC, LS&C)* (OSA, 2019), pp. LTu4B-1

8. Yole Developpment, Woodside Capital Partners: The Automotive LiDAR Market Report (2018). https://d19j0qt0x55bap.cloudfront.net/production/onboardings/5e5421415aaa397b55 2399b4/documents/file/Yole_WCP-LiDAR-Report_April-2018-FINAL.pdf. Last accessed 2021/04/16

9. X. Mosquet, T. Dauner, N. Lang, N. Rubmann, A. Mei-Pochtler, R. Agrawal, F. Schmieg, Revolution in the driver's seat: the road to autonomous vehicles. Boston Consult. Group **11** (2015)

10. A. Skarbek-Żabkin, M. Szczepanek, Autonomous vehicles and their impact on road infrastructure and user safety, in *XI International Science-Technical Conference Automotive Safety* (IEEE, 2018), pp 1–4

11. A. Vasile, I. Vasile, A. Nistor, L. Vladareanu, M. Pantazica, F. Caldararu et al., Rain sensor for automatic systems on vehicles. Adv. Topics Optoelectron. Microelectron. Nanotechnol. **V 7821**, 78211W (2010)

12. I. Bord, P. Tardy, F. Menil, Influence of the electrodes configuration on a differential capacitive rain sensor performances. Sens. Actuators B Chem. **114**(2), 640–645 (2013)

13. J. Mukul, J. Kaustubh, D. Sonawane, S. Vinayak, M. Joshi, A novel and cost effective resistive rain sensor for automatic wiper control: circuit modelling and implementation, in *2013 Seventh International Conference on Sensing Technology (ICST)* (IEEE, 2013), pp. 40–45

14. N. Kyoo, Omni-directional rain sensor utilizing scattered light reflection by water particle on automotive windshield glass. Sensors 1728–1731 (2011)

15. V. Matusevich, F. Wolf, E. Tolstik, R. Kowarschik, A transparent optical sensor for moisture detection integrated in a PQ-PMMA medium. IEEE Photonics Technol. Lett. **25**(10), 969–972 (2013)

16. M. Trierweiler, T. Peterseim, C. Neumann, Automotive LiDAR pollution detection system based on total internal reflection techniques, in J.K. Kim, M.R. Krames, M. Strassburg (eds.) *Light-Emitting Devices, Materials, and Applications XXIV*, vol. 11302 (SPIE, 2020), pp. 1130216–1

Chapter 46
Pulse Wave Measurement Using Fiber-Optic Intermodal Interferometric Sensor

Aleksandr Markvart, Daria Kulik, Alexander Petrov, Leonid Liokumovich, and Nikolai Ushakov

Abstract The paper presents a fiber-optic pulse wave sensor based on an intermodal interferometer formed by a singlemode-multimode-singlemode fiber structure. The sensor was interrogated by means of spectral interferometry approach. Several signal processing methods aimed at suppressing the nonlinear distortions caused by complex mechanical coupling between the pulse wave and the sensor are considered. Pulse wave signal with signal-to-noise ratio about 40 dB was acquired. The proposed sensor can be used for medical diagnostics and vital signs monitoring and can be further modified to be implemented in a portable form-factor.

Keywords Singlemode-multimode-singlemode · Fiber optic sensor · Pulse wave · Spectral interferometry · Biomedical sensor

46.1 Introduction

Recently, an increasing number of such chronic diseases as diseases of the cardiovascular system, connective tissue dysplasia and others have become widespread [1–3]. At the same time, timely diagnosis is the most important stage of treatment and prevention of complications. One of the widely used methods for diagnosing these diseases is pulse wave analysis [1, 4, 5]. The advantages of this method include non-invasiveness, high diagnostic confidence, and potential ease of implementation. Typically, high-precision setups for recording pulse wave signals are stationary and are available only in medical institutions. The examples of such devices include SphygmoCor (Atcor Medical, Sydney, Australia), Complior (Artech Medical, Pantin, France), and others [6, 7]. The main factors limiting their implementation in personal and portable devices is their high cost and sensitivity to body movements, which can

A. Markvart (✉) · D. Kulik · A. Petrov · L. Liokumovich · N. Ushakov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: markvart_aa@spbstu.ru

N. Ushakov
e-mail: n.ushakoff@spbstu.ru

significantly distort the recorded signals. On the other hand, the development and implementation of portable and personal medical devices based on the registration and processing of pulse wave signals will make it possible to diagnose the above diseases at the earliest stage, significantly facilitating and increasing the effectiveness of their treatment.

Optical sensing techniques are widely used in biomedical inspection, including measurement of blood flow parameters [8–10]. Among other approaches, fiber optic sensors are great candidates to measure pulse wave signals due to their noninvasiveness, bio-compatibility, immunity to electromagnetic interference, small footprint of the sensing element, extremely high achievable accuracy and resolution. Several recent papers report the approaches for cardiac monitoring [11–13] and pulse wave measurement [14–18] by means of fiber Bragg gratings (FBG) and Fabry-Perot interferometric (FPI) sensors. However, when attached to the skin, such sensors are subject to bending and uneven deformations, which will distort the shape of their spectral transfer function and reduce the accuracy of the measurements. In order to overcome this limitation, a sensor capable of distinguishing between different perturbations, such as strain and bending. An example of such sensor is a fiber-optic intermodal interferometer, based on singlemode-multimode-singlemode structure (SMS) [19–22]. In addition to ability to multi-quantity sensing, SMS sensors also offer higher strain sensitivities than FBG sensors [21].

In the current work, we report on a use of SMS sensor for pulse wave monitoring and compare the obtained signals with those obtained using an FBG sensor, interrogated by the same hardware. Two algorithms for processing signals of SMS sensor are considered, their linearity and attained signal-to-noise ratio is compared with the signal obtained using FBG sensor.

46.2   Experimental Setup

The schematic drawing of the interrogation hardware and the sensor is presented in Fig. 46.1a, b. Since both the used sensors (SMS and FBG) require measurement of their spectra in order to demodulate the sensor signals, we were able to use the same interrogation hardware for both sensors and therefore compare their performance. Superluminescent light emitting diode (SLED) model Exalos EXS210066-01 SLED (output power up to 5 mW, central wavelength 1.55 μm, -6 dB spectral width 160 nm, flat-top spectrum shape), installed on Exalos EBD5000 driver board was used as a light source. The optical spectra of the sensors were measured with an Ibsen I-MON USB512 spectrometer (spectrum measurement interval was [1.51; 1.595] μm, variable integration time from 10 μs to 100 μs, spectra acquisition rate up to 3 kHz). In case of an SMS sensor the light was transmitted directly from SLED to spectrometer through the sensor, while in case of FBG reflection spectrum was measured, which was done by connecting FBG to the SLED and the spectrometer via optical circulator, as shown in Fig. 46.1b.
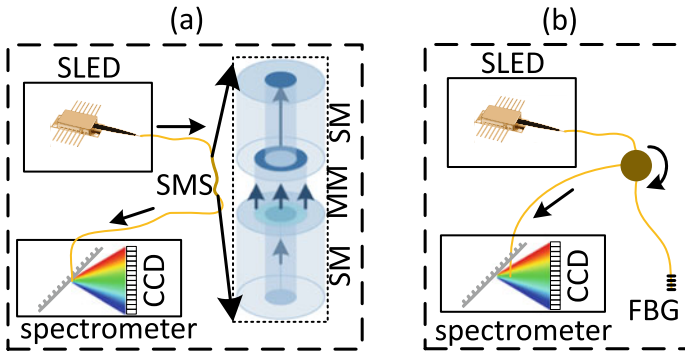
Fig. 46.1 Schemes of interrogation setups for SMS sensor (**a**) and FBG sensor (**b**)
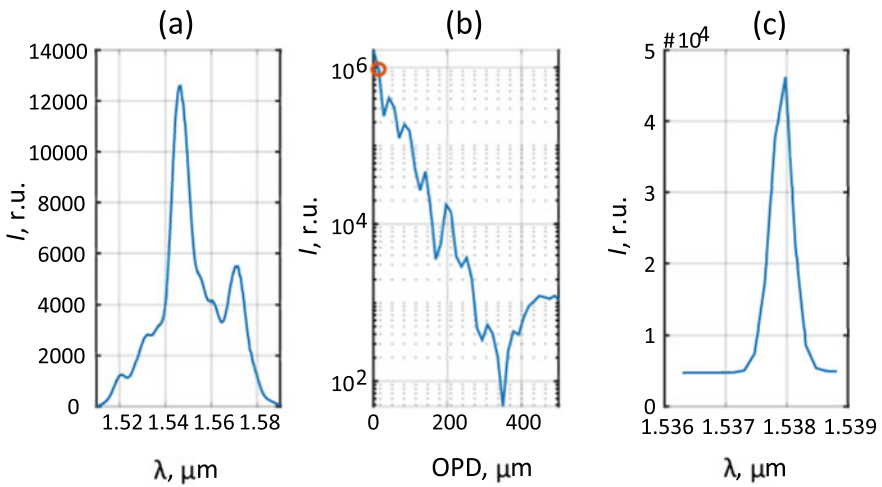
Fig. 46.2 Examples of optical spectrum of SMS sensor (**a**) and its FFT, circle shows the first sample (**b**); FBG sensor (**c**)

The SMS sensor was fabricated using standard fiber splicer Fujicura FCM 45M. The multimode (MM) section consisted of a 5 cm long step-index Thorlabs FG050LGA fiber. At both ends, it was spliced to singlemode (SM) SMF-28 patch-cords. Its optical spectrum and its FFT are shown in Fig. 46.2a, b, respectively. The FBG sensor was a commercially acquired grating with about 1540 nm resonance wavelength, its reflection spectrum is shown in Fig. 46.2c. The fabricated sensors were placed on the skin over the arteries and fixed by a flexible adhesive tape as shown in Fig. 46.3a, b.

(a)   (b)
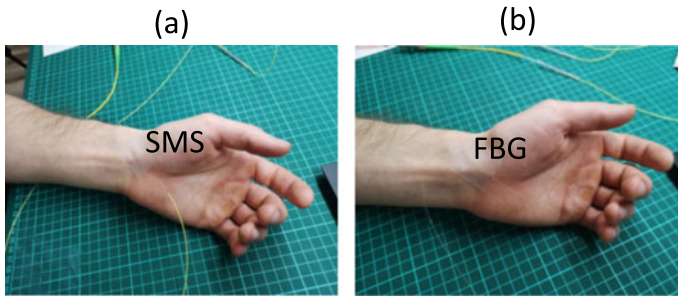
Fig. 46.3 Photos of SMS (**a**) and FBG (**b**) sensors, installed on Subject's radial artery

46.3 Signal Processing

The measured optical spectra were stored on the personal computer, to which the spectrometer was connected via USB protocol and further processed in Matlab. FBG signal was demodulated by least-squares fitting of the measured spectra with gaussian functions as was done in [23, 24] with its position, width and amplitude being fit parameters. It was shown in [24] that such FBG demodulation approach is nearly optimal in terms of demodulated signal noise.

For SMS sensor, two signal processing approaches were utilized: one based on spectra correlation [25] and other one, based on FFT-based phase calculation, as it is often done in case of spectral interferometry [22, 26]. Prior to calculating FFT, the measured optical spectrum was interpolated so that spectra samples correspond to uniform spacing of optical frequency, not the wavelength. In this case interferometer OPD directly corresponds to fringe oscillation frequency, hence resulting in minimal broadening of FFT peaks. Also, the optical spectrum was multiplied at a Blackman window before FFT calculation in order to mitigate the cross-talk between different interference components. Phase unwrapping [27] was applied to the calculated FFT argument in order to remove the signal discontinuities.

As can be seen in Fig. 46.2b, optical spectrum of the fabricated SMS sensor contains a great number of interference components, which was expected for SMS sensor with a step-index multimode section. Obviously, different harmonic components of the SMS optical spectrum correspond to interference of MM fiber modes of different orders. This may potentially lead to different responses of these components' phases to the PW signal.

SMS spectrum components with lower optical path difference (OPDs) are of interest for our task, which is also advantageous due to their higher magnitudes, as can be seen in Fig. 46.2b. The phase of the first sample of SMS spectra FFT (the 0-th sample corresponds to the constant component) was used as the main demodulated signal for SMS sensor with FFT demodulation. The phases of other FFT samples were compared with it to prove the above reasoning.

46.4   Pulse Wave Measurement

The pullse wave measurements were performed for one subject, who was in a sitting position, hand put on a table. Pulse wave signals were measured at the radial artery, at the wrist. Although the subject didn't move deliberately, there still were some quasi-static motion artefacts in the demodulated signals, which, however, can easily be removed by either high-pass or band-pass filter with the lower cut-off frequency about 0.6 Hz (lower than the typical heart rate).

The following parameters of the demodulated signals were compared: signal to noise ratio (SNR), estimated as a ratio of standard deviation of high-pass filtered signal and white noise level, evaluated from high-pass filtered signal according to [28], signal to quasi-static ratio (SQSR), estimated as ratio of standard deviations of high-pass filtered and non-filtered signals and subjective score, indicating correspondence to typically observed pulse wave signals at the radial artery. The comparison results are shown in Table 46.1. Fragments of demodulated signals of FBG and SMS sensors with different processing algorithms are shown in Fig. 46.4.

Table 46.1 Comparison of pulse wave signals measured with different sensors and signal processing

	FBG	SMS, corr	SMS, FFT#1	SMS, FFT#3
SNR, dB	30	23	38	40
SQSR, dB	−2	−21	−11	−21
Linearity	+	−	+	−
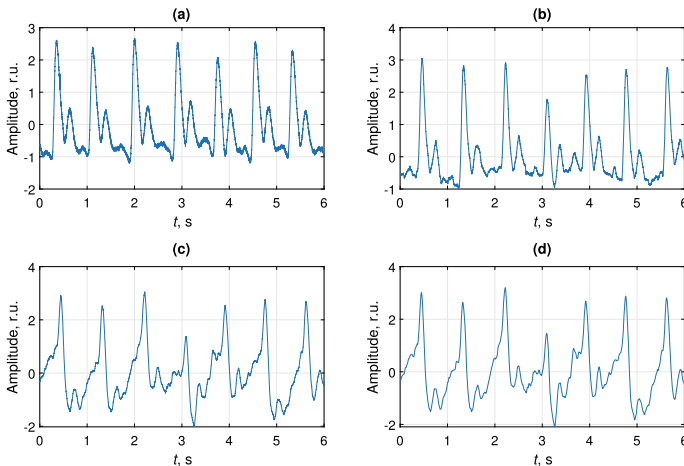
Fig. 46.4 Comparison of pulse wave signals obtained with FBG sensor (**a**); SMS sensor, FFT processing, sample #1 (**b**); SMS sensor, FFT processing, sample #3 (**c**); SMS sensor, correlation processing (**d**)

As can be seen in Table 46.1, FFT-based processing of SMS sensor signal results in the highest SNR value. Choosing the proper FFT component also leads to an improvement in quasistatic component suppression. As can be seen in Fig. 46.4, linear sensor response was obtained with the both FBG and SMS sensors, since the signals in Fig. 46.4a, b correspond to the typically observed pulse wave signal very well. However, SMS turns out to be preferable due to lower noise and quasi-static component levels.

46.5 Conclusion

A singlemode-multimode-singlemode fiber-optic interference sensor was studied in terms of its efficiency for pulse wave signal measurement. It was shown that interference of lower-order modes of multimode fiber can provide a demodulated signal of high quality, both in terms of signal to noise ratio and linearity. This finding is quite important for technology transfer and development of portable pulse wave sensors, which will require spectrometers with very mild requirements to spectral resolution. Moreover, a close analogue of SMS sensors, a multimode-coreless-multimode fiber sensors, which in addition do not require a spatially coherent light source, can be also studied in terms of their potential to such tasks. In contrast, FBG sensor requires both spatially coherent light and high spectrl resolution of optical spectrometer, hence, its realization in portable form-factor is hardly feasible. Therefore, our work might pave the way to non-invasive, highly accurate, low-cost and portable pulse wave and other biomedical sensors, which can revolutionize the health care system.

References

1. S. Laurent, J. Cockcroft, L. Van Bortel, P. Boutouyrie, C. Giannattasio, D. Hayoz, B. Pannier, C. Vlachopoulos, I. Wilkinson, H. Struijker-Boudier, Expert consensus document on arterial stiffness: methodological issues and clinical applications. Eur. Heart J. **27**(21), 2588–2605 (2006). https://doi.org/10.1093/eurheartj/ehl254
2. A. Kiselev, R. Vaz, A. Knyazeva, A. Khudiakov, S. Tarnovskaya, J. Liu, A. Sergushichev, S. Kazakov, D. Frishman, N. Smolina, T. Pervunina, J. Jorholt, G. Sjoberg, T. Vershinina, D. Rudenko, A. Arner, T. Sejersen, A. Lindstrand, A. Kostareva, De novo mutations, in *FLNC leading to early-onset restrictive cardiomyopathy and congenital myopathy*. Human Mutation, vol. 39, issue 9, pp. 1161–1172 (2018). https://doi.org/10.1002/humu.23559
3. A. Emelyanov, T. Shtam, R. Kamyshinsky, L. Garaeva, N. Verlov, I. Miliukhina, A. Kudrevatykh, G. Gavrilov, Y. Zabrodskaya, S. Pchelina, A. Konevega, Cryo-electron microscopy of extracellular vesicles from cerebrospinal fluid. PLOS ONE **15**(1), e0227, 949 (2020). https://doi.org/10.1371/journal.pone.0227949
4. M. Peltokangas, A. Vehkaoja, J. Verho, V.M. Mattila, P. Romsi, J. Lekkala, N. Oksala, Age dependence of arterial pulse wave parameters extracted from dynamic blood pressure and blood volume pulse waves. IEEE J. Biomed. Health Informatics **21**(1), 142–149 (2017). https://doi.org/10.1109/JBHI.2015.2503889

5. R.R. Townsend, Arterial stiffness: recommendations and standardization. Pulse **4**(1), 3–7 (2016). https://doi.org/10.1159/000448454

6. T. Pereira, C. Correia, J. Cardoso, Novel methods for pulse wave velocity measurement. J. Med. Biol. Eng. **35**(5), 555–565 (2015)

7. P. Salvi, *Pulse Waves How Vascular Hemodynamics Affects Blood Pressure*, 2nd edn. (Springer, New York, 2017)

8. E.A. Savchenko, E.N. Velichko, The use of speckle correlation analysis to determine blood flow velocity. Optics Spectrosc. **128**(7), 998–1003 (2020). https://doi.org/10.1134/S0030400X2007019X

9. E. Savchenko, I. Kolokolnikov, E. Velichko, V. Osovskikh, L. Kiseleva, Z. Musakulova, Design of liver functional reserve estimation technique based on optical densitometry. Diagnostics **10**(8), 599 (2020). https://doi.org/10.3390/diagnostics10080599

10. E.N. Velichko, E.K. Nepomnyashchaya, A.V. Sokolov, T.Y. Kudryashova, Laser correlation spectrometer for assessing the size and dynamics of changes in the size of structures in biological fluids. Optics Spectrosc. **128**(7), 959–963 (2020). https://doi.org/10.1134/S0030400X20070255

11. K. Katayama, S. Chino, S. Kurasawa, S. Koyama, H. Ishizawa, K. Fujimoto, Classification of pulse wave signal measured by FBG sensor for vascular age and arteriosclerosis estimation. IEEE Sensors J. **20**(5), 2485–2491 (2020). https://doi.org/10.1109/JSEN.2019.2952833

12. D. Lo Presti, C. Massaroni, J. D'Abbraccio, L. Massari, M. Caponero, U.G. Longo, D. Formica, C.M. Oddo, E. Schena, Wearable system based on flexible FBG for respiratory and cardiac monitoring. IEEE Sensors J. **19**(17), 7391–7398 (2019). https://doi.org/10.1109/JSEN.2019.2916320

13. L. Dziuda, F.W. Skibniewski, M. Krej, J. Lewandowski, Monitoring respiration and cardiac activity using fiber Bragg grating-based sensor. IEEE Trans. Biomed. Eng. **59**(7), 1934–1942 (2012). https://doi.org/10.1109/TBME.2012.2194145

14. S. Kurasawa, S. Koyama, H. Ishizawa, K. Fujimoto, S. Chino, Verification of non-invasive blood glucose measurement method based on pulse wave signal detected by FBG sensor system. Sensors (Switzerland) **17**(12), 2702 (2017). https://doi.org/10.3390/s17122702

15. M.F. Domingues, C. Tavares, N. Alberto, A. Radwan, P. André, P. Antunes, High rate dynamic monitoring with Fabry-Perot interferometric sensors: an alternative interrogation technique targeting biomedical applications. Sensors **19**(21), 4744 (2019). https://doi.org/10.3390/s19214744

16. Y. Haseda, J. Bonefacino, H.Y. Tam, S. Chino, S. Koyama, H. Ishizawa, Measurement of pulse wave signals and blood pressure by a plastic optical fiber FBG sensor. Sensors **19**(23), 5088 (2019). https://doi.org/10.3390/s19235088

17. C. Leitão, P. Antunes, P. André, J.L. Pinto, J.M. Bastos, Central arterial pulse waveform acquisition with a portable pen-like optical fiber sensor. Blood Pressure Monit. **20**(1), 43–46 (2015). https://doi.org/10.1097/MBP.0000000000000079

18. N.A. Ushakov, A.A. Markvart, D.D. Kulik, L.B. Liokumovich, Comparison of pulse wave signal monitoring techniques with different fiber-optic interferometric sensing elements. Photonics **8**(5), 142 (2021). https://doi.org/10.3390/PHOTONICS8050142

19. S, Silva, O. Frazão, J. Viegas, L.A. Ferreira, F.M. Araújo, F.X. Malcata, J.L. Santos, Temperature and strain-independent curvature sensor based on a singlemode/multimode fiber optic structure. Meas. Sci. Technol. **22**(8), 085, 201 (2011). https://doi.org/10.1088/0957-0233/22/8/085201

20. Q. Wu, A.M. Hatta, P. Wang, Y. Semenova, G. Farrell, Use of a bent single SMS fiber structure for simultaneous measurement of displacement and temperature sensing. IEEE Photonics Technol. Lett. **23**(2), 130–132 (2011). https://doi.org/10.1109/LPT.2010.2093515

21. C. Lu, J. Su, X. Dong, T. Sun, K.T.V. Grattan, Simultaneous measurement of strain and temperature with a few-mode fiber-based sensor. J. Lightwave Technol. **36**(13), 2796–2802 (2018). https://doi.org/10.1109/JLT.2018.2825294

22. Y. Cardona-Maya, A.B. Socorro, I. Del Villar, J.L. Cruz, J.M. Corres, J.F. Botero-Cadavid, Label-free wavelength and phase detection based SMS fiber immunosensors optimized with

cladding etching. Sensors Actuators B Chem. **265**, 10–19 (2018). https://doi.org/10.1016/j.snb.2018.03.002

23. V. Mizragi, J.E. Sipe, Optical properties of photosensitive fiber phase gratings. J. Lightwave Technol. **11**(10), 1513–1517 (1993). https://doi.org/10.1109/50.249888
24. A. Markvart, L.B. Liokumovich, I. Medvedev, N. Ushakov, Smartphone-based interrogation of a chirped FBG strain sensor inscribed in a multimode fiber. J. Lightwave Technol. **39**(1), 282–289 (2021). https://doi.org/10.1109/jlt.2020.3024713
25. A.V. Petrov, I.E. Chapalo, M.A. Bisyarin, O.I. Kotov, Intermodal fiber interferometer with frequency scanning laser for sensor application. Appl. Optics **59**(33), 10, 422 (2020). https://doi.org/10.1364/ao.400345
26. N.A. Ushakov, A.A. Markvart, L.B. Liokumovich, Pulse wave velocity measurement with multiplexed fiber optic Fabry-Perot interferometric sensors. IEEE Sensors J. **20**(19), 11302–11312 (2020). https://doi.org/10.1109/JSEN.2020.2997465
27. K. Itoh, Analysis of the phase unwrapping algorithm. Appl. Optics **21**(14), 2470 (1982)
28. N.A. Ushakov, L.B. Liokumovich, Signal processing approach for spectral interferometry immune to λ/2 errors. IEEE Photonics Technol. Lett. **31**(18), 1483–1486 (2019). https://doi.org/10.1109/LPT.2019.2934093

Chapter 47
Quarter Wave Plate for Fiber Optic Current Sensor: Comparison of Modeling and Experimental Study

Valentina Temkina , **Andrei Medvedev** , **Alexey Mayzel** , **and Eduard Sivolenko**

Abstract Fiber optic current sensor is the high-precision device for measuring large current values in high-voltage grids. A quarter wave plate is an important part of this sensor, as its quality affects the accuracy of the entire system. Firstly, this plate must be fiber so as not to violate the integrity of the device. Secondly, its length must be equal to the quarter of fiber beat length. It is just a few millimeters. Finally, the optical length of quarter wave plate depends on its temperature. Due to all this facts, a perfect quarter wave plate preparation is a very difficult task, and it is often not ideal in a real device. In this paper, we studied the plate imperfections effect on the performance of the fiber optic current sensor. The mathematical model of the sensor was developed using the Jones matrices formalism. The non-ideality of a quarter-wave plate was considered by introducing a small phase shift between two orthogonal linear modes, which led to the formation of an elliptical polarization mode instead of a circular one at the plate output. The similar experimental study was conducted for the real prototype of a fiber optic current sensor and compared with the modeling results. Thus, it was found that the quarter wave plate affects not only the sensor measurement error, but also the contrast of the interference signal.

Keywords Fiber optic current sensor · Current measuring · Quarter wave plate · Jones matrices formalism

V. Temkina (✉) · A. Medvedev · A. Mayzel
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: valenssiya@bk.ru

A. Medvedev
e-mail: medvedev@rphf.spbstu.ru

E. Sivolenko
Radiophysics and Electronics, Yerevan State University, Yerevan, Armenia

47.1   Introduction

The fiber optic current sensor (FOCS) operation principle is based on the Faraday Effect, which occurs in a sensitive element (spun optical fiber) due to the presence of a magnetic field. The specificity of the spun fiber is maintaining of the circular polarization of the light propagating through it. In the electric systems the magnetic field is generated by an electric current propagating through a conductor. The spun fiber is wrapped around a conductor. Then, due to the Faraday Effect, a phase shift proportional to the current is formed between the light modes with the right and left circular polarizations [1–9].

To ensure the introduction of circularly polarized light into the spun fiber, a quarter-wave (QW) plate is used in the FOCS scheme. It converts linearly polarized light modes into modes with circular polarization, and vice versa. It is known that the measurement accuracy of the FOCS depends on the accuracy of this transformation [10–12]. Therefore, firstly, the QW plate must be fiber to maintain the high accuracy of the fiber-optic measuring system. It allows not violating the integrity of the sensor. Secondly, a QW plate is a segment of a birefringent fiber, the length of which should be equal to

$$L = L_{\mathrm{b}} \cdot \left(\frac{1}{4} + \frac{n}{2} \right), \tag{47.1}$$

where L_{b} is the beat length of fiber, n is an integer. For example, the beat length of the commercial PM fiber HB1250 is equal to 3.28 mm. Then the length of the QW plate should be equal to $0.82 + 1.64n$ mm. To manufacture a fiber QW plates, optical component suppliers also offer a special fiber with an increased beat length, e.g. Fibercore company offers to use the fiber HB1500(8.9/125)QW with $L_{\mathrm{b}} = 8.84$ mm for our application. In this case, the length of the QW plate should be equal to $2.21 + 4.42n$ mm. However, even using of a special fiber, it is very difficult task to make a QW plate with an accuracy of hundredths of a millimeter. In addition, the QW plate temperature will affect the conversion accuracy of the polarization of light by the plate. The temperature deviation leads to the optical path length deviation of orthogonal light modes in the plate. As a result, the phase shift between them at the QW plate output becomes different from $\pi/2$, and an elliptical polarization is obtained instead of circular one.

Thus, the QW plate in the FOCS scheme is a particularly important component with difficult-to-achieve accuracy and strong influence of external factors. Therefore, it is investigated in this paper how the fiber QW plate imperfections will affect the FOCS performance characteristics.

47.2 Mathematical Modeling of the Fiber Optic Current Sensor with Quarter Wave Plate Imperfections

There is an optical fiber coupler, a polarizer, an electro-optical modulator (EOM), a delay line, a fiber QW plate, a sensitive spun fiber, a mirror and a photodetector in the FOCS scheme (see Fig. 47.1).

We use the additional phase modulation of linearly polarized light by a harmonic signal

$$\Delta \psi_M(t) = \psi_m \cos \omega_m t, \tag{47.2}$$

where ψ_m is modulation amplitude, ω_m is modulation frequency. Phase detection is based on the method of the harmonic ratio of the modulation frequency.

Previously, we have developed the FOCS model based on the Jones matrices formalism [13]. According to it, each element of a fiber-optic system can be represented by a 2×2 matrix describing the transformation of the polarization state of the light transmitted through this element. Then the total matrix of the system $[T]$ is obtained by successive multiplication of the matrices of each element in the order opposite to the direction of light propagation [14]. When all FOCS elements are ideal, $[T]$ is equal

$$[T] = \begin{bmatrix} j \cos(\varphi_{mod} + \varphi_F) & 0 \\ 0 & 0 \end{bmatrix}, \tag{47.3}$$

where φ_F is the phase difference formed in spun fiber between light modes of circular polarization (due to the Faraday Effect), φ_{mod} is the phase difference formed by electro-optical modulator between light modes of linear polarization.

The matrix of an ideal QW plate is represented as

$$[F_{\lambda/4}] = \begin{bmatrix} e^{-\frac{j\pi}{4}} & 0 \\ 0 & e^{\frac{j\pi}{4}} \end{bmatrix}. \tag{47.4}$$
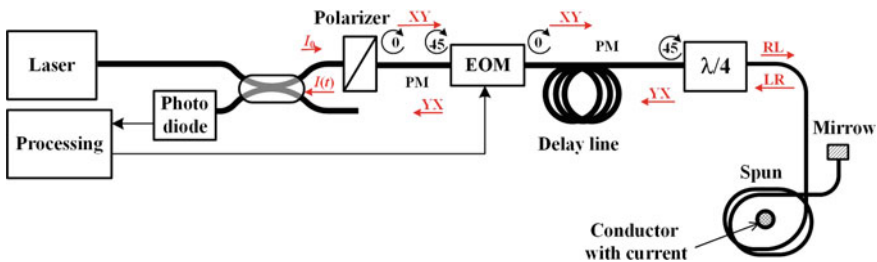
Fig. 47.1 FOCS scheme

When two linearly polarized light modes pass through the phase plate, a phase shift between them is equal to π/2 and circular polarization of light is formed at the plate output. However, as described above, it is technologically very difficult to manufacture an ideal fiber QW plate. Therefore, in a real FOCS device, the phase shift introduced by the QW plate will have an error $\Delta\varphi$. In addition, the QW plate temperature in the actual operating conditions will be unstable, which will also introduce an error in the phase shift.

We considered the QW plate imperfections in the FOCS developed model. Then the phase shift between two orthogonal linearly polarized modes introduced by the plate is equal to $\pi/2 + \Delta\varphi$, and the imperfect QW plate matrix is given as

$$\left[F_{\lambda/4}\right]^{imp} = \begin{bmatrix} e^{-j\left(\frac{\pi}{4}+\frac{\Delta\varphi}{2}\right)} & 0 \\ 0 & e^{j\left(\frac{\pi}{4}+\frac{\Delta\varphi}{2}\right)} \end{bmatrix}. \tag{47.5}$$

We substituted this matrix in the common matrix of the system [T], described in detail in [13], instead of the ideal QW plate matrix $[F_{\lambda/4}]$. Multiplying all the matrices in the system, we get the analytical expression for the common matrix of the system [T] with the QW plate imperfection

$$[T] = \begin{bmatrix} -\cos\varphi_F \cdot \sin\Delta\varphi + j(\cos\varphi_F \cdot \cos\Delta\varphi \cdot \cos\varphi_{mod} - \sin\varphi_F \cdot \sin\varphi_{mod}) & 0 \\ 0 & 0 \end{bmatrix}. \tag{47.6}$$

If the light at the FOCS input is linearly polarized, then the input Jones vector is equal to

$$[D_{\text{in}}] = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{47.7}$$

Then the output Jones vector $[D_{\text{out}}]$ and the intensity I of the light beam are defined as

$$[D_{\text{out}}] = [T] \cdot [D_{\text{in}}]$$
$$= \begin{bmatrix} -\cos\varphi_F \cdot \sin\Delta\varphi + j(\cos\varphi_F \cdot \cos\Delta\varphi \cdot \cos\varphi_{mod} - \sin\varphi_F \cdot \sin\varphi_{mod}) \\ 0 \end{bmatrix}, \tag{47.8}$$

$$I = \left[D_{out}^*\right]^T \cdot [D_{out}] = \frac{1}{2} + \frac{1}{4}\sin^2\Delta\varphi + \left(\frac{1}{4}\cos^2\Delta\varphi - \frac{1}{4}\right)\cos 2\varphi_{mod}$$
$$+ \frac{1}{2}\left(\frac{1}{2}\cos\Delta\varphi - \frac{1}{2}\right)^2 \cos(2\varphi_{mod} - 2\varphi_F)$$
$$+ \frac{1}{2}\left(\frac{1}{2}\cos\Delta\varphi + \frac{1}{2}\right)^2 \cos(2\varphi_{mod} + 2\varphi_F). \tag{47.9}$$

According to Eq. (47.9), if the QW plate is ideal and the error $\Delta\varphi = 0$, then the intensity of the output optical signal of the sensor looks like the intensity of an interferometer

$$I(\Delta\varphi = 0) = \frac{1}{2} + \frac{1}{2}\cos(2\varphi_{\text{mod}} + 2\varphi_F). \tag{47.10}$$

Taking into account the phase modulation of light by a harmonic signal (47.2), Eq. (47.9) can be rewritten in the following form

$$\begin{aligned}
I = {} & \frac{1}{2} + \frac{1}{4}\sin^2\Delta\varphi + \left(\frac{1}{4}\cos^2\Delta\varphi - \frac{1}{4}\right)\cos(2\psi_m \sin\omega_m t) \\
& + \frac{1}{2}\left(\frac{1}{2}\cos\Delta\varphi - \frac{1}{2}\right)^2 \cdot \cos(2\psi_m \sin\omega_m t - 2\varphi_F) \\
& + \frac{1}{2}\left(\frac{1}{2}\cos\Delta\varphi + \frac{1}{2}\right)^2 \cos(2\psi_m \sin\omega_m t + 2\varphi_F).
\end{aligned} \tag{47.11}$$

It can be seen that the equation for intensity I consists of one constant and three variables components, proportional to $\cos(z\sin t)$. It is a widely accepted method of using an expansion in terms of Bessel functions for similar components according to expressions

$$\cos(z\sin t) = J_0(z) + 2\sum_{k=1}^{\infty} J_{2k}(z)\cos 2kt, \tag{47.12}$$

$$\sin(z\sin t) = 2\sum_{k=1}^{\infty} J_{2k-1}(z)\sin(2k-1)t, \tag{47.13}$$

where $J_k(z)$ are Bessel functions of the first kind of the k-th order.

We applied the sum and difference formulas for cosine to the Eq. (47.11) and took an expansion in terms of Bessel functions. At the same time, in the expression (47.12), we limited the value of k to 2, and in the expression (47.13) to 1. This is due to the FOCS signal processing method [13]. It relies on synchronous extraction of the first, second and fourth harmonics of the modulation frequency and the computation of their relationship.

Thus, the equation for intensity I can be rewritten in the following form

$$\begin{aligned}
I = {} & \frac{1}{2} + \frac{1}{4}\sin^2\Delta\varphi + \left(\frac{1}{4}\cos^2\Delta\varphi - \frac{1}{4}\right)[J_0(2\psi_m) + 2J_2(2\psi_m)\cos 2\omega_m t \\
& + 2\,J_4(2\psi_m)\cos 4\omega_m t] + \frac{1}{2}\left(\frac{1}{2}\cos\Delta\varphi - \frac{1}{2}\right)^2 \\
& \cdot \{[J_0(2\psi_m) + 2J_2(2\psi_m)\cos 2\omega_m t + 2\,J_4(2\psi_m)\cos 4\omega_m t]\cos 2\varphi_F
\end{aligned}$$

$$+ 2J_1(2\psi_m) \sin \omega_m t \sin 2\varphi_F\} + \frac{1}{2}\left(\frac{1}{2}\cos\Delta\varphi + \frac{1}{2}\right)^2$$

$$\cdot \{[J_0(2\psi_m) + 2J_2(2\psi_m)\cos 2\omega_m t + 2\,J_4(2\psi_m)\cos 4\omega_m t]\cos 2\varphi_F$$

$$- 2J_1(2\psi_m)\sin\omega_m t \sin 2\varphi_F\}. \tag{47.14}$$

Based on Eq. (47.14), analytical expressions for first, second and forth harmonics were obtained

$$U_1 \sim -J_1(2\psi_m)\sin 2\varphi_F \cos\Delta\varphi, \tag{47.15}$$

$$U_2 \sim \left(\frac{1}{2}\cos^2\Delta\varphi - \frac{1}{2}\right)\cdot J_2(2\psi_m)$$

$$+ J_2(2\psi_m)\cos 2\varphi_F\left(\left[\frac{1}{2}\cos\Delta\varphi - \frac{1}{2}\right]^2 + \left[\frac{1}{2}\cos\Delta\varphi + \frac{1}{2}\right]^2\right), \tag{47.16}$$

$$U_4 \sim \left(\frac{1}{2}\cos^2\Delta\varphi - \frac{1}{2}\right)\cdot J_4(2\psi_m)$$

$$+ J_4(2\psi_m)\cos 2\varphi_F\left(\left[\frac{1}{2}\cos\Delta\varphi - \frac{1}{2}\right]^2 + \left[\frac{1}{2}\cos\Delta\varphi + \frac{1}{2}\right]^2\right). \tag{47.17}$$

It should be noted that the second and fourth harmonics of the modulation frequency are the same up to the values of the Bessel functions $J_2(2\psi_m)$ and $J_4(2\psi_m)$. Therefore, we use the ratio of these harmonics in the FOCS signal processing algorithm to compensate the error that occurs due to an uncontrolled deviation of the modulation amplitude from the optimal value. This compensation algorithm is described in detail in [13]. Thus, the FOCS signal with the compensation of measurement dependence on modulation amplitude and with an imperfect QW plate is defined by the following equation

$$\varphi_F^r = \text{arctg}\left[\frac{J_1(2\psi_m)}{J_2(2\psi_m)}\cdot g(2\psi_m)\right.$$

$$\left.\cdot\frac{\cos\Delta\varphi\sin 2\varphi_F}{\left(\frac{1}{2}\cos^2\Delta\varphi - \frac{1}{2}\right) + \cos 2\varphi_F\left(\left[\frac{1}{2}\cos\Delta\varphi - \frac{1}{2}\right]^2 + \left[\frac{1}{2}\cos\Delta\varphi + \frac{1}{2}\right]^2\right)}\right], \tag{47.18}$$

where φ_F and φ_F^r are measured and calculated Faraday phase shifts, $g(2\psi_m)$ is the error correction function [13].

In addition, Eq. (47.14) shows that the imperfection of the QW plate affects not only the current measurement error, but also the constant component in the interference signal I, and hence its contrast.

47.3 Experimental Results of QW Plate Temperature Impact on FOCS Signal Processing

Our goal was to measure the impact of the QW plate temperature on the first, second and fourth harmonics of the modulation frequency. We modified our FOCS laboratory prototype, which we described in [13], by isolating a fiber QW plate in a small chamber controlled with a Peltier cooler, driven by a triangular signal generator with a period of 500 s [15]. As a result, the QW plate is heated from 23 to 58 °C, and then cooled. Temperature of the QW plate is measured by thermocouple. It should also be noted that the fiber QW plate has made of a special fiber HB1500 (8.9/125)QW with a beat length $L_b = 8.84$ mm by the method described in [16]. The sensing element is a coil of 159 turns of spun fiber. Three turns of copper wire are passed through this fiber coil. The industrial frequency (50 Hz) current flows through the copper wire.

The experimental results of QW plate temperature impact on the first, second and fourth harmonics of the modulation frequency are shown in Figs. 47.2, 47.3, 47.4 and 47.5.

It can be seen that the amplitudes of the second and fourth harmonics depend parabolically on the QW plate temperature, which will lead to an error in calculating
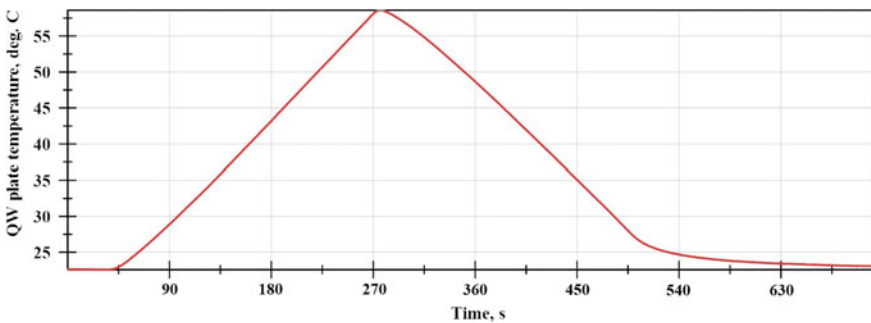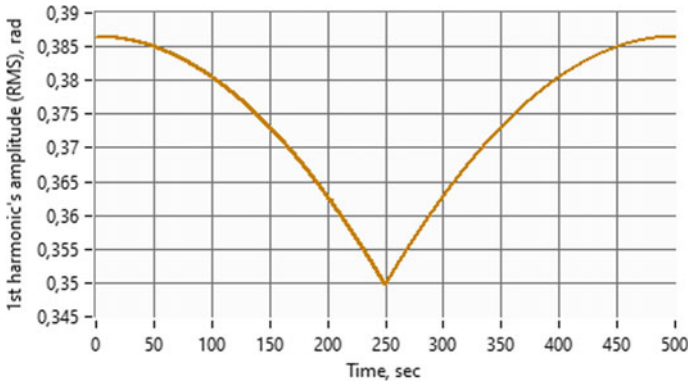

Fig. 47.2 QW plate temperature versus time
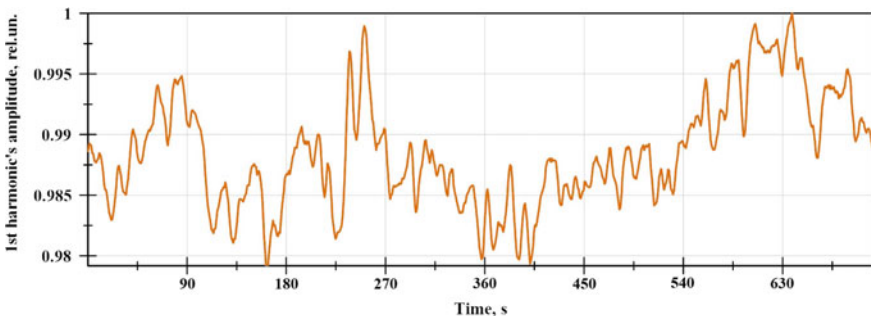
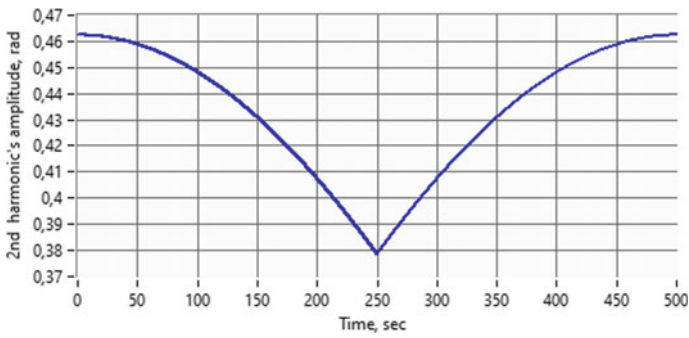**Fig. 47.3** Amplitude of the first harmonic of the modulation frequency versus time
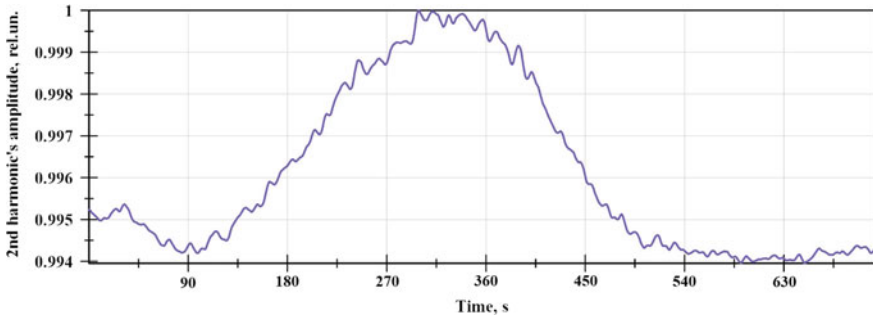
Fig. 47.4 Amplitude of the second harmonic of the modulation frequency versus time
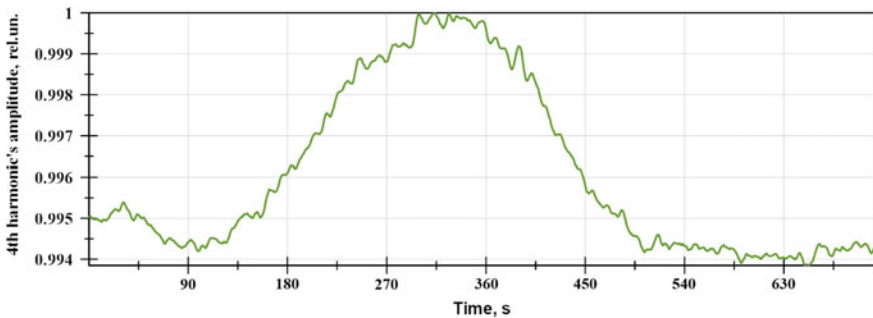
Fig. 47.5 Amplitude of the fourth harmonic of the modulation frequency versus time

the amplitude of the measured current signal. When the plate temperature increased from 23 to 58 °C, the amplitudes of these harmonics are increased by 0.6%. In addition, the response of the system to changes in the QW plate temperature is not instantaneous. As for the first harmonic of the modulation frequency, in this case, the dependence on the plate temperature is not observed.

47.4   Comparison of Modeling and Experimental Study

To compare the results of our modeling and experiment, we have to find a match between the QW plate temperature change in experiment and the corresponding change in the phase shift introduced by the plate between two linearly polarized modes in modeling. It is known that a temperature change of the fiber plate leads both to a change of its length and the birefringence coefficient of the fiber [11]. Then the erroneous phase shift introduced by the QW plate can be represented as

$$\Delta\varphi = \varphi_{\lambda/4}(C_L + C_B)\Delta T = \varphi_{\lambda/4}C_\varphi\Delta T, \tag{47.19}$$

where $\varphi_{\lambda/4} = \pi/2$, C_L is thermal coefficient of linear expansion, C_B is the thermal coefficient of the birefringence coefficient change, ΔT is the temperature deviation of QW plate. Since the thermal coefficient of linear expansion of quartz glass is very small and equal to $C_L = 5 \cdot 10^{-7} \text{K}^{-1}$ [17], the main contribution to the erroneous phase shift is made by the birefringence coefficient change [11].

We obtained the experimental results for the QW plate temperature effect on the ellipticity of the state of light polarization at the QW plate output in our previous work [18]. Using these results, we determined the value of $C_\varphi = -0.008 \text{K}^{-1}$ for our fiber QW plate.

To compare the results of experiment and modeling, we substituted the value of the modulation amplitude $\psi_m = 1.31$ rad from the experiment into Eqs. (47.11), (47.12) and (47.13). In addition, we set the Faraday phase shift φ_F as a sinusoidal signal with a frequency of 50 Hz and $\Delta\varphi$ according to Eq. (47.19), where the value of temperature T increases linearly from 23 to 58 °C and then decreases with a period of 500 s (see Fig. 47.6). Using the Eqs. (47.11), (47.12) and (47.13), we plotted the modeled dependences of the first, second, and fourth harmonic amplitudes versus time (see Figs. 47.7, 47.8 and 47.9).

Our modeling shows that when the QW plate temperature changed from 23 to 58 °C, the first harmonic amplitude decreased by 9%, the second and fourth harmonic amplitudes decreased by 18%.

Comparing modeling results with the experimental data, one can see that:

(1)   The experimental dependences of the second and fourth harmonic amplitudes on the QW plate temperature have the same shape as the modeled dependences. However, the direction of amplitudes change is opposite and there is a smoothed extremum. Probably, the smoothing of the extremum is due to the fact that the fiber of the QW plate does not heat up instantly.
(2)   The impact of the QW plate temperature on the first harmonic amplitude was not observed in the experiment, although such dependence exists in modeling results.
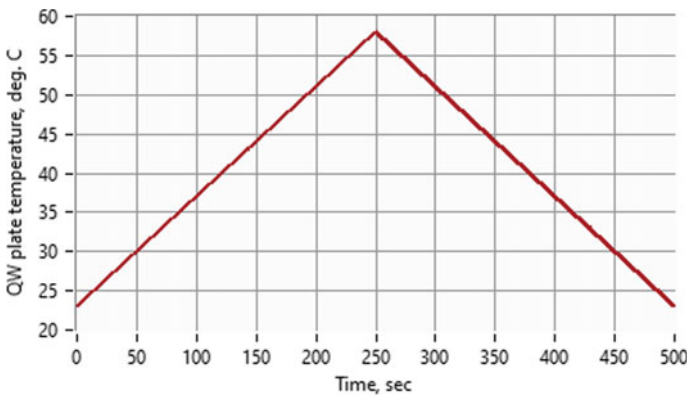
Fig. 47.6   QW plate temperature versus time (modeled dependence)

Fig. 47.7 Amplitude of the first harmonic of the modulation frequency versus time (modeled dependence)

Fig. 47.8 Amplitude of the second harmonic of the modulation frequency versus time (modeled dependence)
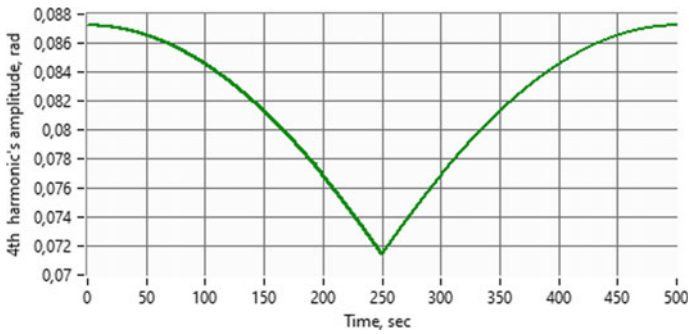
Fig. 47.9 Amplitude of the fourth harmonic of the modulation frequency versus time (modeled dependence)

47.5   Conclusion

The fiber QW plate is the weakest point in the FOCS optical circuit. It is very laborious task to make the ideal fiber QW plate with the length of a few millimeters. Moreover, the plate is strongly influenced by external factors, especially temperature. These lead to a change in the amplitudes of the detected harmonics of the modulation frequency and, as a result, to errors in FOCS signal processing.

In this paper, we present the comparison of the mathematical modeling and the experimental study of the FOCS with the temperature dependent QW plate. Our model is based on the Jones matrices formalism and demonstrates that the plate imperfections caused by changes in its temperature introduce an error both in the detected harmonics of the modulation frequency and in the contrast of the interference signal at the optical circuit output. As a result of modeling, the first harmonic amplitude should decrease by 9% and the second and fourth harmonic amplitudes by 18%, while the QW plate temperature increases from 23 to 58 °C. The experimental curves look similar to modeled dependences but differ in following. There is no influence of the QW plate temperature on the first harmonic amplitude. Besides, the second and fourth harmonic amplitudes increase only by 6% in the experiments instead of predicted 18% and there is a smoothed extremum in the experiments. We consider the smoothing extremum can be due to the fact that QW plate fiber has thermal inertia. The reason for the direction change in the deviation of the second and forth harmonic amplitudes has yet to be investigated.

References

1. R.M. Silva, H. Martins, I. Nascimento, J.M. Baptista, A.L. Ribeiro, J.L. Santos, P. Jorge, O. Frazão, Optical current sensors for high power systems: a review. Appl. Sci. **2**, 602–628 (2012)
2. M. Lenner, A. Frank, L. Yang, T. Roininen, K. Bohnert, Long-term reliability of fiber-optic current sensors. IEEE Sens. J. **20**(2), 823–832 (2019)
3. W. Wang, X. Wang, J. Xia, The nonreciprocal errors in fiber optic current sensors. Opt. Laser Technol. **43**(8), 1470–1474 (2011)
4. A. Boev, V. Gubin, S. Morshnev, J. Przhijalkovskij, M. Rjabko, A. Sazonov, N. Starostin, J. Chamorovskij, Fibre-optic current sensor. Patent RU 2437106 C2, G01R 15/24, G01R 19/2. Date of publication: 20.12.2011 Bull. 35
5. D. Kiesewetter, V. Malyugin, A. Reznik, N. Zhuravleva, Spectral-correlation method of investigation of highvoltage electrical insulation components, in *2018 IEEE XXVII Interna-tional Scientific Conference Electronics—ET*. (IEEE, Sozopol, 2018), pp. 1–3
6. R. Decosta, B. Altschul, Mode analysis for energetics of a moving charge in Lorentz and CPT-violating electrodynamics. Phys. Rev. D **97**(5), 055029 (2018)
7. P. Agruzov, I. Pleshakov, E. Bibik, S. Stepanov, A. Shamray, Fast magnetooptical mod-ulation in optofluidic devices based on ferrofluidfilled microstructured optical fibers, in *2015 European*

Conference on Lasers and Electro-Optics - European Quantum Electronics Conference, paper CK_P_13. OSA (2015)

8. M. Shlyagin, P. Agruzov, I. Pleshakov, A. Prokofiev, E. Bibik, Incident-power-dependent refractive index of ferrofluid in magnetic field measured with a fiber optic probe. Optik **186**, 418–422 (2019)

9. N.V. Larionov, I.M. Sokolov, Influence of electric and magnetic fields on interference effects upon multiple light scattering in cold atomic ensembles. J. Exp. Theor. Phys. **127**(2), 264–273 (2018)

10. Y.V. Przhiyalkovsky, S.K. Morshnev, N.I. Starostin, V.P. Gubin, Modified sensing element of a fibre-optic current sensor based on a low-eigenellipticity spun fibre. Quantum Electron. **44**, 957–964 (2014)

11. A.S. Petukhov, A.S. Smirnov, V.V. Burdin, The temperature properties of fiber quarter-wave plate of minimum length. Prikladnaia fotonika **2**(1), 80–87 (2015)

12. A. Vlasov, Y. Plotnikov, A. Ashirov, A. Aleynik, A. Varlamov, A. Stam, The method for protection of sensitive fiber optic components from environmental noise and vibration impacts, *2019 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech).* (IEEE, St. Petersburg, 2019), pp. 305–307

13. V. Temkina, A. Medvedev, A. Mayzel, Research on the methods and algorithms improving the measurements precision and market competitive advantages of fiber optic current sensors. Sensors (Switzerland)**20**(21), 1–22 (2020). Article № 5995

14. E. Collett, *Polarized Light in Fiber Optics.* (PolaWave Group, Lincroft, NJ, USA, 2003)

15. E. Bakulin, S. Dzyubanenko, S. Konakov, A. Korotkov, V. Loboda, Y. Yugay, Thermoelectric Peltier micromodules processed by thin-film technology. J. Phys. Conf. Ser. **1124**. (2018)

16. V. Temkina, A. Medvedev, A. Mayzel, A. Mokeev, A. Kirpichenko, Manufacturing method and stability research of the fiber quarter-wave plate for fiber optic current sensor, in *Proceedings of the 2019 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech).* (Saint Petersburg, Russia, 2019), pp. 278–281

17. S.M. Tripathi, A. Kumar, R.K. Varshney, Y.B.P. Kumar, E. Marin, J.-P. Meunier, Strain and temperature sensing characteristics of single-mode–multimode–single-mode structures. J. Lightwave Technol. **27**(13), 2348–2356 (2009)

18. A.V. Kirpichenko, V.S. Temkina, A.V. Medvedev, A.V. Mayzel, A.S. Mokeev, Metod izgotovleniya i issledovanie temperaturnyh svoistv volokonnoj chetvert'volnovoj plastinki [Method of manufacturing and investigation of the temperature properties of a fiber quarter-wave plate], in *Nedelya nauki SPbPU: materialy nauchnoj konferencii c mezhdunarodnym uchastiem [Week of Science SPbPU: materials of the scientific conference with international participation].* (Polytech-Press, St. Petersburg, 2019), pp. 87–90

Chapter 48
Non-invasive Blood Glucose Estimation Using Two Multiplexed Fiber-Optic Fabry-Perot Interferometric Sensors and Pulse Wave Signal Features Analysis

Daria Kulik, Dmitry Zubko, Aleksandr Markvart, Leonid Liokumovich, and Nikolai Ushakov

Abstract The paper describes and verifies a non-invasive blood glucose estimation method, which uses two multiplexed fiber-optic Fabry-Perot interferometric sensors interrogated by means of spectral interferometry. The sensors were installed at the wrist and ankle and the corresponding pulse wave velocity was measured. A mathematical model for estimating the blood glucose concentration from pulse wave signal features was developed. The model accuracy was proven by estimating blood glucose values from pulse wave signals measured in an additional validation experiment.

48.1 Introduction

Diabetes mellitus is one of the most common chronic diseases both in Russia and in most of developed countries of the world. Regular measurement of blood glucose concentration (BGC) is an important means of diagnosis and control of diabetes [1–3]. Traditionally, this procedure is carried out on a sample of the patient's blood, which makes it difficult to constantly monitor (patients with type 1 diabetes are required to make up to ten measurements a day), and it is objectionable from the point of view of hygiene psychological comfort. Thus, the task of non-invasive control of blood glucose concentration is of great interest from the point of view of science and medicine and attracts the interest of researchers from around the world for more than 30 years[4, 5]. Currently, there are many areas of development for non-invasive control of blood glucose concentration: bioimpedance analysis [6], spectroscopy [7], optical coherent tomography [8], registration and analysis of the pulse wave signal [9], photoacous-

D. Kulik · D. Zubko · A. Markvart (✉) · L. Liokumovich · N. Ushakov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: markvart_aa@spbstu.ru

tic techniques [10] and other methods. Various optical methods are commonly use for estimation of blood properties [11–13], which can be used for blood glucose estimation as well. Among all the above-mentioned methods, the approach of BGC estimation based on measurement and analysis of pulse wave signals appears to be the safest, as it doesn't require any interaction of electromagnetic radiation or ultrasound waves with biological tissues, yet, it appears to be quite promising since glucose concentration changes biochemical and mechanical properties of blood [14], hence, influencing pulse wave propagation conditions.

Fiber optic sensors are good candidates to measure pulse wave signals due to their bio-compatibility, immunity to electromagnetic interference, small footprint of the sensing element, extremely high achievable accuracy and resolution. Recent paper reports the approaches for non-invasive BGC measurement with a fiber Bragg grating [9]. However, when attached to the skin, the Bragg grating is subject to bending and uneven deformations, which will distort the shape of its spectral transfer function and reduce the accuracy of measurements. Moreover, the implementation of high-precision measurements using Bragg gratings requires expensive and quite bulky equipment.

In our previous work [15] we present an approach to noninvasively measure pulse wave velocity by means of two multiplexed fiber-optic Fabry-Perot interferometric sensors interrogated by means of spectral interferometry. Here we utilize this approach for BGC estimation.

48.2   Experimental Setup

The schematic drawing of proposed glucose estimation system is presented in Fig. 48.1a. We used two fiber optic sensors two measure the pulse wave at different locations. Broad spectrum light from a SLED (Exalos EXS210066-01) was transferred to the sensors via the optical circulator, denoted as ⌒ in Fig. 48.1a and divided between the sensors by a 50:50 coupler. The light waves, reflected from the sensors, went back through the same fibers, were combined at the coupler and directed to the spectrometer (Ibsen I-MON USB512) via the optical circulator. Both sensors were extrinsic Fabry-Perot interferometers (EFPI), formed by two fiber facets, mated in a glass capillary, as shown in the inset in Fig. 48.1a. Cavity lengths of fabricated EFPIs were set to about 56 and 143 μm which are sufficiently different in order to separate the signals by means of frequency division multiplexing and avoid cross-talk, as described in [16]. The fabricated sensors were placed on the skin over the radial artery of the wrist and tibialis posterior artery of the medial ankle and fixed by a medical tape as it presented in Fig. 48.1b. The blood pressure induced the capillary deformation that finally led to changing of the gaps between fiber ends. Thus, the pulse waveforms were reconstructed by the registration of the gaps in the Fabry-Perot sensors. To determine the gaps we used spectral interferometry as a well-established method for interferometric sensor interrogation [17–19]. More details about the setup can be found in our previous work [15].
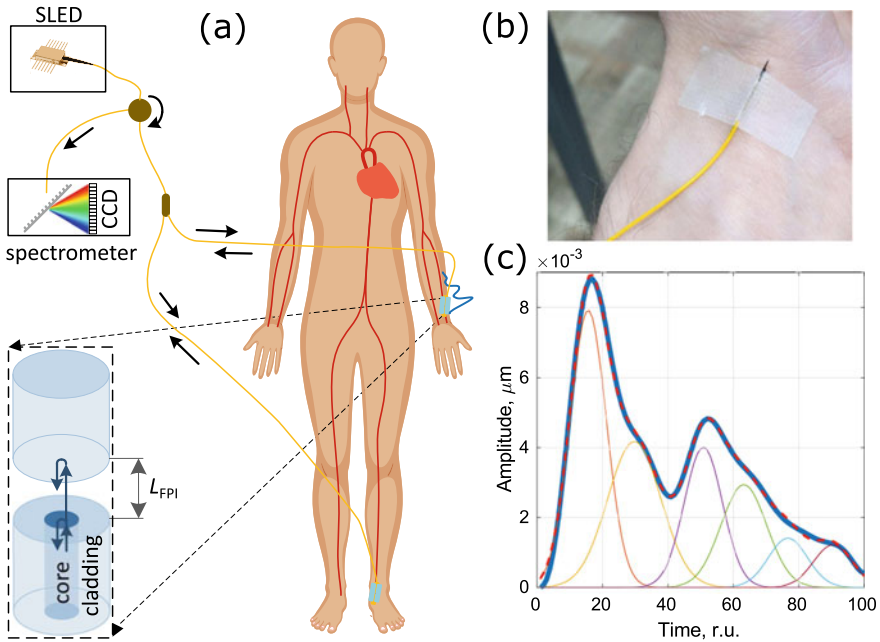
Fig. 48.1 a Concept of the proposed glucose estimation setup. **b** Photo of fabricated sensor installed on radial artery. **c** An example of a pulse wave segment measured at the radial artery on the wrist (solid bold blue line) and its decomposition into direct and reflected waves. The sum of six Gaussian functions is shown with a dashed bold red line. The first Gaussian peak (solid thin red line) corresponds to the direct wave, the rest—to reflected waves

48.3   Blood Glucose Estimation

Measurement of the initial data for the development of a mathematical model describing the blood glucose concentration was performed similarly to the standard glucose tolerance test. Measurements were taken under fasting condition, the subject (healthy male, 32 years old) was in a supine position, pulse wave signals were measured at the wrist and medial ankle. 5 min after the start of measurement, subject drank a glass of warm water with dissolved 75 g of glucose, after which, for an hour, the pulse wave signals were recorded and every 2 min the blood glucose level was invasively measured. The results of invasive glucose concentration measurements were used as reference values for developing a mathematical model. At the same time, the arterial pressure was monitored, which revealed that it remained constant throughout the measurements.

The measured pulse wave signals were used to determine the pulse wave velocity, the signals themselves were divided into segments corresponding to single pulse waves. The starting points were chosen as the boundaries of the pulse wave slopes. To eliminate the influence of the heart rate on the measurement results, for each

segment, an interpolation procedure was carried out, in a result of which each segment contained the same number of samples, and the sampling rate was normalized to the pulse period. For each segment, signal features were found by approximating a segment by a superposition of six Gaussian functions similarly to [20–22], and as a result, each segment was described by 18 parameters (amplitude, position and width of the Gaussian peak). Initial approximations and restrictions on the values of the approximation parameters were chosen in such a way that the Gaussian functions corresponded to direct and reflected waves in the pulse wave signal. Pulse wave segment example measured on the radial artery at the wrist and its approximation by six Gaussian functions is shown in Fig. 48.1c.

The set of the initially estimated parameters (pulse wave velocity and signal features) corresponds to separate pulse waves. Due to the effect of pulse wave velocity variability, the values of these parameters were averaged over an interval of 120 s length (about 110–120 pulse waves). In this case, the boundaries of the averaging intervals were chosen in such a way that the midpoints of these intervals correspond to the time moments of reference glucose concentration measurements.

After preparing the data, the pulse waves parameters that were significant for describing the blood glucose concentration were found. Signal features were combined into a matrix X, the first column of which contained the reference glucose concentration values, the second column - values of the pulse wave velocity, and the remaining columns – pulse wave signal features. It was revealed that the best results are obtained when the analysis included only the features of the signal measured at the wrist. The analysis included the coordinates of the approximated Gaussian functions corresponding to the reflected waves, from which the coordinates of the direct wave peaks were subtracted, the widths of all approximated Gaussian functions and the amplitudes of the approximated Gaussian functions, corresponding to the reflected waves, normalized on the direct wave amplitude (16 signal features: 5 values of the relative positions and amplitudes of the reflected waves and 6 values of the widths of the peaks of the reflected waves, total matrix contained 18 columns). The principal component analysis (PCA) decomposition was applied to this matrix for dimension reduction and the search for correlations between the features and glucose concentration. As a result of PCA decomposition, the feature matrix is represented in a form

$$X = V \cdot C^T, \tag{48.1}$$

where V is principal components matrix (columns contain principal components) and C is a matrix of expansion coefficients (rows correspond to the original signals, columns – to principal components). The normalization was carried out in such a way that the standard deviations of the components were equal, and the expansion coefficients varied from -1 to 1, so that the absolute value of expansion coefficients characterizes the degree of importance of the corresponding principal component to describe the columns of the original matrix. A graphical representation of the matrix of the coefficients for the first 10 components is shown in Fig. 48.2a, the coefficients corresponding to the significant components and parameters of the pulse wave are highlighted with rectangles.

(a)

(b)
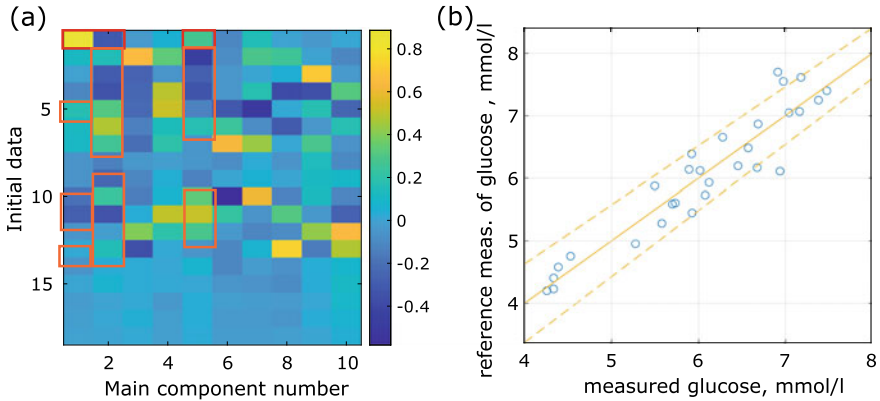
Fig. 48.2   a The matrix of the decomposition coefficients into the main components of the matrix containing the reference values of the glucose concentration (the corresponding coefficients in the first row). The pulse wave velocity (the second row), the delay of the reflected waves relative to the direct wave (rows 3–7), the width of the peaks corresponding to the straight wave (8-th line) and reflected waves (lines 9 through 13). Relative amplitudes of reflected waves (lines 14 through 18). The rectangles mark the coefficients of expansion into significant components, the modulus of which exceeds 0.15. **b** Correspondence between the measured blood glucose value from the validation experiment and the corresponding reference measurements of blood glucose concentration. The dots show the experimental data, the solid line corresponds to the function y = x, the dashed lines show the confidence interval of 95%

During the analysis of the decomposition coefficients, it was found that the column of blood glucose concentration is described by the three main components, the values of the expansion coefficients for which were equal to 0.88, −0.31 and 0.27, and its expansion coefficients for other components did not exceed 0.15. Further, by analyzing the rest of the decomposition coefficients, it was determined which pulse wave features are significant for describing the blood glucose concentration—those parameters for which the values of the decomposition coefficients for three components, describing the blood glucose concentration exceeded 0.15, were selected as significant, the rest were excluded from the analysis. Based on the results of this analysis, it was revealed that the largest values of the decomposition coefficients were observed for the values of the pulse wave velocity (decomposition coefficients in two of the significant components were equal to 0.44 and 0.63), the relative positions and widths of all approximated Gaussian functions corresponding to reflected waves, while the amplitudes of the approximated Gaussian functions do not provide useful information.

Further, a linear regression model was built using the significant parameters of the pulse wave signals, describing the blood glucose concentration. The obtained value of the coefficient of determination of 0.88 testifies to the high accuracy of the developed model. The correctness of the model was also verified during a repeated validating experiment similar to the first one. The validation experiment was conducted at the same time around 9 a.m., 8 d after the original. Blood pressure values were also

monitored during the validation experiment and differed from the values measured during the initial experiment by no more than 4 mm Hg. Pulse wave signal features obtained during the validating experiment were substituted into the developed model, and the resulting estimated blood glucose concentration was compared with the invasively measured reference values. The correspondence between the reference and measured values is shown in Fig. 48.2b. Solid line depicts a straight-line y = x, dashed lines—95% confidence intervals of the values predicted by the developed model.

Thus, the developed mathematical model accurately describes blood glucose concentration, which is demonstrated by the repeatability of estimated blood glucose values in a validating experiment.

48.4   Conclusion

A non-invasive method for estimating blood glucose level is proposed. The method is based on the measurement and further processing of pulse wave signals, which were acquired using two multiplexed fiber-optic sensors. The pulse wave signal features, most relevant for blood glucose estimation, were identified. The performed analysis of pulse wave signal features might shed some light on biomechanics of pulse wave propagation and its relation to various health issues. A mathematical model, able to predict BGC from those signal features was developed and experimentally verified through an independent measurement. Further studies will include more thorough verification of the proposed approach, involving a greater number of participants and various experimental conditions.

References

1. M.J. Davies, D.A. D'Alessio, J. Fradkin, W.N. Kernan, C. Mathieu, G., P. Rossing, A. Tsapas, D.J. Wexler, J.B. Buse, Management of hyperglycemia in type 2 diabetes, 2018. *A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD) (2018)*. https://doi.org/10.2337/dci18-0033
2. G. Llauradó, V. Ceperuelo-Mallafré, C. Vilardell, R. Simó, N. Freixenet, J. Vendrell, J.M. González-Clemente, Arterial stiffness is increased in patients with type 1 diabetes without cardiovascular disease: a potential role of low-grade inflammation. Diabetes Care **35**(5), 1083–1089 (2012). https://doi.org/10.2337/dc11-1475
3. O. Glotov, E. Serebryakova, M. Turkunova, O. Efimova, A. Glotov, Y. Barbitoff, Y. Nasykhova, A. Predeus, D. Polev, M. Fedyakov, I. Polyakova, T. Ivashchenko, N. Shved, E. Shabanova, A. Tiselko, O. Romanova, A. Sarana, A. Pendina, S. Scherbak, E. Musina, A. Petrovskaia-Kaminskaia, L. Lonishin, L. Ditkovskaya, L. Zhelenina, L. Tyrtova, O. Berseneva, R. Skitchenko, E. Suspitsin, E. Bashnina, V. Baranov, Whole-exome sequencing in Russian

children with non-type 1 diabetes mellitus reveals a wide spectrum of genetic variants in mody-related and unrelated genes. Molecul. Med. Rep. **20**(6), 4905–4914 (2019). https://doi.org/10.3892/mmr.2019.10751

4. W. Villena Gonzales, A. Mobashsher, A. Abbosh, The progress of glucose monitoring-a review of invasive to minimally and non-invasive techniques. Dev. Sens. **19**(4), 800 (2019). https://doi.org/10.3390/s19040800

5. D. Bruen, C. Delaney, L. Florea, D. Diamond, D. Bruen, C. Delaney, L. Florea, D. Diamond, Glucose sensing for diabetes monitoring: recent developments. Sensors **17**(8), 1866 (2017). https://doi.org/10.3390/s17081866

6. T. Yilmaz, R. Foster, Y. Hao, Radio-frequency and microwave techniques for non-invasive measurement of blood glucose levels. Diagnostics **9**(1), 1–34 (2019). https://doi.org/10.3390/diagnostics9010006

7. A. Lambrecht, T. Beyer, K. Hebestreit, R. Mischler, W, Petrich, Continuous Glucose Monitoring by Means of Fiber-Based, Mid-Infrared Laser Spectroscopy. Appl. Spectrosc. **60**(7), 729–736 (2006). https://doi.org/10.1366/000370206777887071

8. R.O. Esenaliev, K.V. Larin, I.V. Larina, M. Motamedi, Noninvasive monitoring of glucose concentration with optical coherence tomography. Opt. Lett. **26**(13), 992–994 (2001). https://doi.org/10.1364/ol.26.000992

9. S. Kurasawa, S. Koyama, H. Ishizawa, K. Fujimoto, S. Chino, Verification of non-invasive blood glucose measurement method based on pulse wave signal detected by FBG sensor system. Sens. (Switzerland) **17**(12), 2702 (2017). https://doi.org/10.3390/s17122702

10. J. Kottmann, J. Rey, M. Sigrist, Mid-infrared photoacoustic detection of glucose in human skin: towards non-invasive diagnostics. Sensors **16**(10), 1663 (2016). https://doi.org/10.3390/s16101663

11. E.A. Savchenko, E.N. Velichko, The use of speckle correlation analysis to determine blood flow velocity. Opt. Spectrosc. **128**(7), 998–1003 (2020). https://doi.org/10.1134/S0030400X2007019X

12. E. Savchenko, I. Kolokolnikov, E. Velichko, V. Osovskikh, L. Kiseleva, Z. Musakulova, Design of liver functional reserve estimation technique based on optical densitometry. Diagnostics **10**(8), 599 (2020). https://doi.org/10.3390/diagnostics10080599

13. E.N. Velichko, E.K. Nepomnyashchaya, A.V. Sokolov, T.Y. Kudryashova, Laser correlation spectrometer for assessing the size and dynamics of changes in the size of structures in biological fluids. Opt. Spectrosc. **128**(7), 959–963 (2020). https://doi.org/10.1134/S0030400X20070255

14. E. Topsakal, T. Karacolak, E.C. Moreland, Glucose-dependent dielectric properties of blood plasma, in *2011 30th URSI General Assembly and Scientific Symposium, URSIGASS 2011* (2011). https://doi.org/10.1109/URSIGASS.2011.6051324

15. N.A. Ushakov, A.A. Markvart, L.B. Liokumovich, Pulse wave velocity measurement with multiplexed fiber optic fabry-perot interferometric sensors. IEEE Sens. J. **20**(19), 11302–11312 (2020). https://doi.org/10.1109/JSEN.2020.2997465

16. N.A. Ushakov, L.B. Liokumovich, Multiplexed extrinsic fiber fabry-perot interferometric sensors: resolution limits. J. Lightwave Technol. **33**(9), 1683–1690 (2015). https://doi.org/10.1109/JLT.2015.2396201

17. C. Li, S. Chen, Y. Zhu, Maximum likelihood estimation of optical path length in spectral interferometry. J. Lightw. Technol. **35**(22), 4880–4887 (2017). https://doi.org/10.1109/JLT.2017.2743214

18. N.A. Ushakov, L.B. Liokumovich, Signal processing approach for spectral interferometry immune to $\lambda/2$ errors. IEEE Photon. Technol. Lett. **31**(18), 1483–1486 (2019). https://doi.org/10.1109/LPT.2019.2934093

19. N.A. Ushakov, A. Markvart, L.B. Liokumovich, Enhancing the resolution limits of spectral interferometric measurements with swept- wavelength interrogation by means of a reference interferometer. Appl. Opt. **54**(19), 6029–6036 (2015). https://doi.org/10.1364/AO.54.006029

20. U. Rubins, Finger and ear photoplethysmogram waveform analysis by fitting with Gaussians. Med. Biolog. Eng. Comput. **46**(12), 1271–1276 (2008). https://doi.org/10.1007/s11517-008-0406-z

21. R. Couceiro, P. Carvalho, R.P. Paiva, J. Henriques, I. Quintal, M. Antunes, J. Muehlsteff, C. Eickholt, C. Brinkmeyer, M. Kelm, C. Meyer, Assessment of cardiovascular function from multi-Gaussian fitting of a finger photoplethysmogram. Physiolog.l Measurem. **36**(9), 1801–1825 (2015). https://doi.org/10.1088/0967-3334/36/9/1801
22. N.A. Ushakov, A.A. Markvart, D.D. Kulik, L.B. Liokumovich, Comparison of pulse wave signal monitoring techniques with different fiber-optic interferometric sensing elements. Photonics **8**(5), 142 (2021). https://doi.org/10.3390/PHOTONICS8050142

Chapter 49
Kerosene- and Water-Based Magnetic Fluids Stability Studied by Spectral Analysis

Arseniy Alekseev, Elina Nepomnyashchaya (ORCID)**, Elena Velichko** (ORCID)**, and Ivan Pleshakov**

Abstract The paper is aimed at studying the magnetic fluids by analyzing their transmission spectra. We compared stability of kerosene- and water-based samples with equal concentration of magnetic nanoparticles. The transmission spectra were detected for different concentrations of kerosene-based and water-based specimens, their stability was analyzed after dilution and sonification. We concluded that kerosine-based fluid is more stable under the dilution than water-based fluid. The data obtained allow better understanding of the stability limits of these materials.

Keywords Magnetic fluids · Transmission spectra · Absorption · Spectroscopy

49.1 Introduction

Magnetic fluids (MF) are colloidal systems consisting of magnetic particles suspended in a carrier fluid. Due to unique properties, they are widely used in modern technology, for instance in optical systems for recording and transmitting information controlled by a magnetic field [1]. In such application both water- and kerosene-based magnetic fluids can be used [2–4]. In particular, MF can be used in the design of fiber-optic modulators [2–10]. The application of these substances in such systems is associated with their optical properties, such as refraction, scattering and absorption [11–14]. That is why the study of the optical properties of magnetic fluids is an important task for modern studies in optoelectronics [15–19].

Our research focused on the transmission of light through MF with different dilutions. We have previously investigated the dependence of the transmission on the type of surfactant and on the concentration of magnetite phase in the solution [19, 20]. This work is devoted to assessing the sedimentation and aggregate stability

A. Alekseev (✉) · E. Nepomnyashchaya · E. Velichko
Peter the Great St. Petersburg Polytechnic University, Polytechnicheskaya 29, St. Petersburg, Russia

I. Pleshakov
Ioffe Institute, Polytechnicheskaya 26, St. Petersburg, Russia

[21] of MF depending on the solvent. In the study, we analyzed transmission spectra of magnetite, dispersed in distilled water and kerosene.

49.2 Mathematical Analysis

Our research is based on the concept of the light attenuation cross section σ, which shows the attenuation of light caused by a single particle. It is equal to the sum of the scattering cross section σ_{sca} and the absorption cross section σ_{abs}

$$\sigma = \sigma_{sca} + \sigma_{abs} \tag{49.1}$$

For the scattering by small particles ($R < \lambda$, R stands for the particle radius, λ is the light wavelength) the Rayleigh approximation is applicable and the expressions for the cross sections can be written as

$$\sigma_{sca} = \frac{8}{3}\left(\frac{2\pi R}{\lambda}\right)^4 \left|\frac{(\tilde{m}^2 - 1)}{(\tilde{m}^2 + 2)}\right| \tag{49.2}$$

$$\sigma_{abs} = \frac{8\pi R}{\lambda}\, \mathrm{Im}\left(\frac{(\tilde{m}^2 - 1)}{(\tilde{m}^2 + 2)}\right), \tag{49.3}$$

where $\tilde{m}$ is the complex refractive index of the particles with respect to the carrier liquid

$$\tilde{m} = m + ik$$

$$k = \frac{\varepsilon''}{2m}$$

$$\tilde{\varepsilon} = \varepsilon' + i\varepsilon'',$$

where $\tilde{\varepsilon}$ is an effective dielectric constant.

Equations (49.2) and (49.3) show that the attenuation of light is proportional to the particle size. As the particle size increases the attenuation of light increases proportionally. In other words, solutions that are more prone to aggregation will transmit less light than solutions with good aggregate stability.

At the same time, large aggregates will settle. This will lead to a decrease in the concentration of particles suspended in the carrier liquid [14] and, as a consequence, to a decrease in optical density and to increase in transmission. This dependence of the particle concentration on optical density is described by Beer's law

$$\lg \frac{I_0}{I} = kC,$$

where I_0 is initial light intensity, I is light intensity after passing through the cuvette, k is aspect ratio, C is concentration.

49.3   Materials and Methods

To carry out the experiments, we used a setup for detecting transmission spectra of fluids shown in Fig. 49.1.

The broadband illuminator L10290 (Hamamatsu) consisting of deuterium and halogen lamps was chosen as the illumination source. The spectrum of its radiation lies in the range 200–1100 nm, which makes it possible to study spectral characteristics in a wide range of wavelengths. The optical fiber was used for transmitting light to the sample.

The spectrometer C10083MD (Hamamatsu) was used to detect the spectrum of the transmitted light. It allowed high-precision spectroscopic measurements in the 330–850 nm range. All experiments were carried out in dark room to avoid additional exposure.

The following samples were chosen in the research. The particles in each MF had size in the range of 5–40 nm (in dry).

- Kerosene-based ferrofluid (Fe_3O_4) with concentration of the magnetic phase in the initial liquid equal to 18% (according to the manufacturer's passport data), synthetic oil was used for stabilization.

Fig. 49.1 Setup for detecting transmission spectra of magnetic fluids
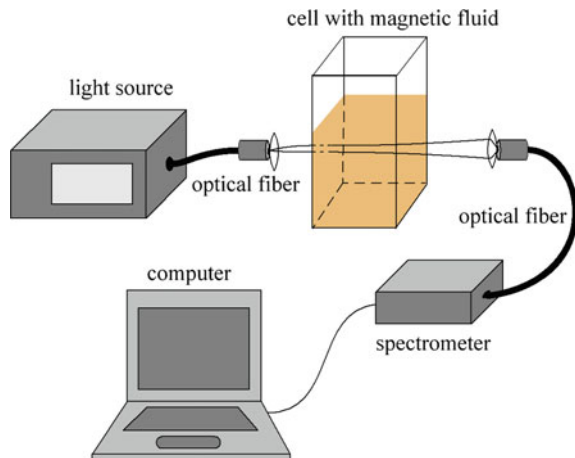
- Water-based ferrofluid (Fe_3O_4) with concentration of the magnetic phase in the initial liquid equal to 27% (according to the manufacturer's passport data), organic oil was used for stabilization.

To achieve the same transparencies for both samples and study the aggregate stability we diluted MF in water or kerosene. In this work, the following dilutions of ferromagnetic fluids were used.

- Kerosene-based ferrofluid was diluted 40,000, 20,000 and 13,000 times.
- Water-based ferrofluid was diluted 8000, 4000 and 2700 times.

Magnetic fluids based on water and kerosene were chosen due to the fact that they are the most used nowadays.

49.4   Results and Discussion

The transmission spectra of the samples were detected immediately after dilution, after dispersion them in an ultrasonic chamber for 10 min (sonification), and after 10 days since dilution of MF to assess their stability. For dilution, a solvent was used, in our case distilled water and kerosene, without the additional stabilizers and surfactants. The results are shown in Figs. 49.2 and 49.3.
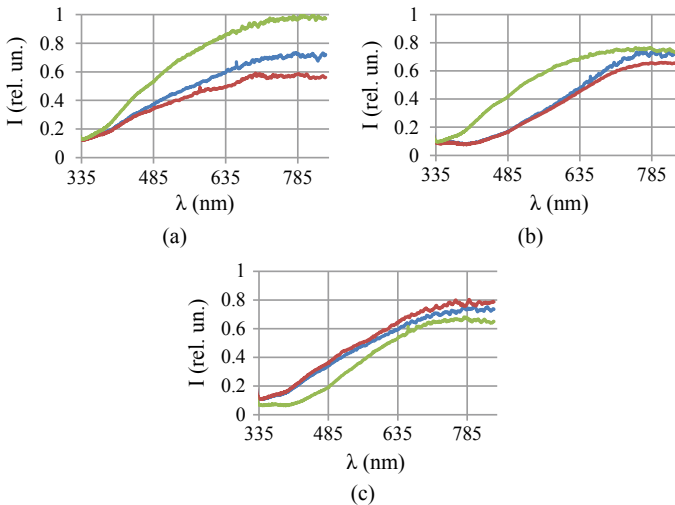
Fig. 49.2 Transmission of a magnetic fluid based on kerosene with dilution of 40,000 **a**, 20,000 **b**, 13,000 **c** times immediately after dilution (blue line), after sonification (red line), and 10 days after dilution (green line)
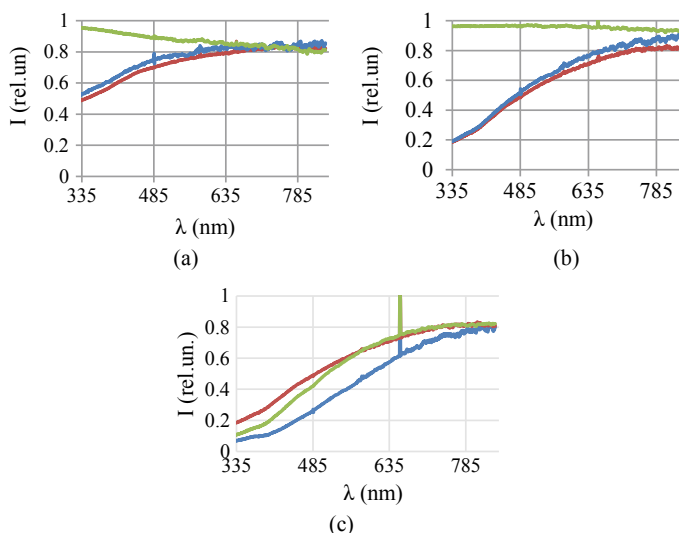
Fig. 49.3 Transmission of magnetic fluid based on water with dilution of 8000 **a**, 4000 **b**, 2700 **c** times immediately after dilution (blue line), after sonification (red line), and 10 days after dilution (green line)

As seen from the figures, the transmission of a magnetic fluid in the ultraviolet region is minimal. We believe that this is due to the absorption effects, which coincide for the studied MFs (in water and kerosene). With increasing concentration, transmission decreases in the entire spectral range.

The kerosene-based MF sample studies after sonification show lower transmission. We believe that this is due to ultrasonic-induced aggregation, which was noted by some researchers before. This effect is less noticeable for high concentrations of MF in kerosene. At the same time, the effect of ultrasound on the water system at high concentrations has the predicted opposite effect.

The transmission obtained 10 days after the dilution of the kerosene-based MF is higher for low concentrations of the magnetic phase in the sample. We believe that this is due to partial loss of aggregate stability and following sedimentation [14].

The transmission obtained for a higher concentration of the magnetic phase in kerosene demonstrates a decrease over time. This may be explained by formation of clusters, which reduced transmission, but did not become large enough to precipitate (larger density and viscosity of the liquid makes sedimentation difficult).

Water-based MF samples with low magnetic phase concentrations ten days after dilution became almost transparent, which indicates a stability violation. The presence of sediment in the fluid was also visually observed. At the same time, the non-diluted fluid did not precipitate. The 2700 times diluted water-based MF sample still shows partial stability after 10 days.

Thus, according to the results of the experiment, it can be concluded that the kerosene-based MF is more stable, while the water-based MF completely lost its properties when diluted by 4000 times or more.

49.5   Conclusion

In this work, the sedimentation and aggregate stability of ferromagnetic fluids were investigated depending on the solvent used. Transmission spectra were obtained for magnetic fluids based on kerosene and water.

Comparing the experimental results, we can conclude that all the studied MF samples have the strongest attenuation of light in the UV region of the spectrum. The work also indicated differences in the stability of fluids based on kerosene and water. We established the effect of ultrasonic waves and dilution with a solvent without the use of additional stabilizers on studied MFs. It was shown that the water-based ferrofluid is less stable with respect to dilution.

References

1. S.E. Logunov, V.V. Davydov, M.G. Vysoczky, O.A. Titova, Peculiarities of registration of magnetic field variations by a quantum sensor based on a ferrofluid cell. J. Phys. Conf. Ser. **1135**(1), 012069. (2018)
2. G.L. Klimchitskaya, V.M. Mostepanenko, E.K. Nepomnyashchaya, E.N. Velichko, Impact of magnetic nanoparticles on the Casimir pressure in three-layer systems. Phys. Rev. B **99**(4), 045433 (2019)
3. E.N. Velichko, G.L. Klimchitskaya, E.K. Nepomnyashchaya, Casimir repulsion though a water-based ferrofluid. Mod. Phys. Lett. A **35**(03), 2040016 (2020)
4. G.L. Klimchitskaya, V.M. Mostepanenko, E.N. Velichko, Effect of agglomeration of magnetic nanoparticles on the Casimir pressure through a ferrofluid. Phys. Rev. B **100**(3), 035422 (2019)
5. H.E. Horng, J.J. Chieh, Y.H. Chao, S.Y. Yang, C.-Y. Hong, H.C. Yang, Designing optical-fiber modulators by using magnetic fluids. Opt. Lett. **30**, 543 (2005). https://doi.org/10.1364/ol.30.000543
6. V.V. Davydov, V.I. Dudkin, N.S. Myazin, R.V. Davydov, On the possibility of studying ferrofluids by a nuclear magnetic magnetometer with a flowing sample. J. Commun. Technol. Electron. **65**, 558–564 (2020)
7. P.M. Agruzov, I.V. Pleshakov, E.E. Bibik, A.V. Shamray, Magneto-optic effects in silica core microstructured fibers with a ferrofluidic cladding. Appl. Phys. Lett. **104**, 071108 (2014). https://doi.org/10.1063/1.4866165
8. P.M. Agruzov, I.V. Pleshakov, E.E. Bibik, S.I. Stepanov, A.V. Shamrai, Transient magneto-optic effects in ferrofluid-filled microstructured fibers in pulsed magnetic field. EPL (Europhysics Letters) **111**, 57003 (2015). https://doi.org/10.1209/0295-5075/111/57003

9. A. Prokofiev, E. Nepomnyashchaya, I. Pleshakov, Y. Kuzmin, E. Velichko, E. Aksenov, Study of specific features of laser radiation scattering by aggregates of nanoparticles in ferrofluids used for optoelectronic communication systems, in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems* (Springer, Cham, 2016), pp. 680–689. https://doi.org/10.1007/978-3-319-46301-8_59

10. X. Yang, D. Li, Q. Li, X. Meng, Spatial self-phase modulation of a Gaussian beam transmitted through a ferrofluid. Appl. Opt. **59**(32), 10069–10075 (2020)

11. A.V. Prokofiev, I.V. Pleshakov, M. Shlyagin, P.M. Agruzov, E.E. Bibik, Y.I. Kuz'min, Noise characteristics of the optical response of ferrofluids to a magnetic field. Tech. Phys. Lett. **45**, 743–745 (2019). https://doi.org/10.1134/S1063785019080121

12. D. Jing, L. Sun, J. Jin, M. Thangamuthu, J. Tang, Magneto-optical transmission of magnetic nanoparticle suspensions for different optical applications: a review. J. Phys. D Appl. Phys. **54**, 013001 (2020)

13. A. Sanz-Felipe, J.C. Martín, Ferrofluids with high dynamic ranges of optical transmission, in *Third International Conference on Applications of Optics and Photonics,* vol. 10453. (International Society for Optics and Photonics. 2017), p. 104530P

14. M.G. Shlyagin, P.M. Agruzov, I.V. Pleshakov, A.V. Prokofiev, E.E. Bibik, Incident-power-dependent refractive index of ferrofluid in magnetic field measured with a fiber optic probe. Optik **186**, 418–422 (2019). https://doi.org/10.1016/j.ijleo.2019.04.130

15. S.N. Lysenko et al., Novel parameter predicting stability of magnetic fluids for possible application in nanocomposite preparation. Appl. Surf. Sci. **463**, 217–226 (2019). https://doi.org/10.1016/j.apsusc.2018.08.236

16. X. Zhou, et al., Magnetic field sensing based on SPR optical fiber sensor interacting with magnetic fluid. IEEE Trans. Instrum. Meas. **68**(1), 234–239 (2018). https://doi.org/10.1109/TIM.2018.2834222

17. V. Temkina, A. Medvedev, A. Mayzel, Computer modeling of fiber optic current sensor, in *International Youth Conference on Electronics, Telecommunications and Information Technologies* (Springer, Cham, 2021) , pp. 263–276

18. L.Y. Xiong, B. Zhang, H.Y. Ji, W. Wang, X. Liu, S.L. He, J.L. Shen, Active optically controlled broadband terahertz modulator based on Fe_3O_4 nanoparticles. IEEE Trans. Terahertz Sci. Technol. **8**(5), 535–540 (2018)

19. E.K. Nepomnyashchaya, A.V. Prokofiev, E.N. Velichko, I.V. Pleshakov, Y.I. Kuzmin, Investigation of magneto-optical properties of ferrofluids by laser light scattering techniques. J. Magn. Magn. Mater. **431**, 24–26 (2017)

20. E.N. Velichko, E.K. Nepomnyashchaya, K.G. Gareev, J. Martínez, M.C. Maicas, Characterization of magnetite-silica magnetic fluids by laser scattering. Appl. Sci. **11**(1), 183 (2021)

21. J.L. Viota, F. González-Caballero, J.D.G. Durán, A.V. Delgado, Study of the colloidal stability of concentrated bimodal magnetic fluids. J. Colloid Interface Sci. **309**(1), 135–139 (2007). https://doi.org/10.1016/j.jcis.2007.01.066

Chapter 50
Data Analysis in Polarized Multispectral Imaging

Valeriia Osochenko, Elina Nepomnyashchaya , **Marina Galeeva** ,
and Nikita Kirpanev

Abstract Non-invasive skin diagnostic methods are already being used in medical practice, but doctors are still in need with new ones. In this paper we suggest a new realization of a polarized multispectral imaging method and an experimental setup. We take snapshots of skin with scattered light in it. The processing includes object-motion compensation and computation of the residual polarization degree that gives us information about skin morphology. Description of the processing algorithm and received images analysis are presented in this paper. The result of this work can be of use in dermatology and trichology.

Keywords Polarization visualization · Biological tissue · Non-invasive methods · Skin · Dermatoscopy

50.1 Introduction

It is a well-known fact that it is much easier to prevent a disease than to treat it, especially when it comes to cancer. Nowadays a surgical removal of the cancerous tumors is the most common treatment [1]. A preoperational demarcation can be quite useful. But before it is necessary to detect and diagnose the disease [2]. The statistics says the doctors overlook about 50% of melanoma cases [3]. Non-invasive methods for diagnosing diseases [4], in particular the skin condition especially in combination with computational system are a promising field because it helps therapists to identify the malignant tumors and when using them the biological tissues are not damaged. Some of these methods are already being widely used in medicine. For example, dermatoscopy where the white light, magnification and cross-polarization effect are used to diagnose melanoma [5]. It is based on polarized visualization [6],

V. Osochenko (✉) · E. Nepomnyashchaya · M. Galeeva · N. Kirpanev
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

M. Galeeva
e-mail: zanina_ma@spbstu.ru

465

which is a powerful method for skin analysis [7]. The other method, spectrophoto-metric intracutaneous analysis (SIAscopy), uses multiple light souses with different wavelengths of light. Dermatoscopy and SIAscopy also can be used to investigate blood vessels or hematomas [8–11].

The main problem is that the dermatoscopy, used nowadays in medicine, still has low specificity for skin diseases, especially non-melanoma cancer [12, 13]. Modern non-invasive methods, based on optical coherent tomography, are usually expensive and cannot be widely used in common clinics and dermatology centers [14]. In addition, the analysis of images, received using different dermatoscopy techniques is usually provided only based on doctor's experience and is not realized using machine methods. Meanwhile processing methods development is an actual and important task [15–17].

In this study we decided to unite benefits of polarization visualization methods, used in regular dermatoscopy, and multispectral imaging, used in SIAscopy. Joint use of these two methods can sufficiently increase the sensitivity and specificity of non-invasive dermatological analysis [18].

The other benefit of our approach is that our setup is not expensive comparing to other multispectral scanning devices. Moreover, a multispectral light source which we use provides us with different light penetration so we can examine different skin layers [19]. The absorbance spectra differ between 440 and 650 nm because of melanin and hemoglobin influence. The light of visible range can be used for oxyhe-moglobin, deoxyhemoglobin and melanin chromophore examination [20]. To detect these differences, we used narrow-band linearly polarized light for skin lighting.

50.2   Description of the Experimental Setup

The method of polarizing visualization is used to obtain high-contrast images containing information about the surface and subsurface structures of biological tissues. In our experiment, we used linearly polarized light to illuminate an object (in our case—birthmarks and scars). Then through imaging system we registered backscattered light what gave information about the skin morphology.

The experimental setup is presented in Fig. 50.1. As a light source we used SMD diodes with power 3 W, amperage 0.7 A and voltage 3.0–3.8 V. The distances and angle between objects are not that crucial. After passing through the polar-izer, we have a linearly polarized light. In the tissue the light scatters. We shot two images: 1 when the axes of polarizer and analyzer are parallel, 2 when the axes are perpendicular.

In the images there are polarized and depolarized components. In the first case we get an image of the surface layer with reflected light, so we examine the surface structure of the tissue. By changing the orientation of the filter axes to the orthogonal one, depolarized light enters the imaging system. The images from the camera are transmitted to the computer. After that we get information about the internal structures

Fig. 50.1 The experimental setup: 1—light source (SMD diode with different wavelength: 440, 530, 575 and 640 nm), 2—lens, 3—polarizer, 4—object, 5—analyzer, 6—color filter, 7—camera
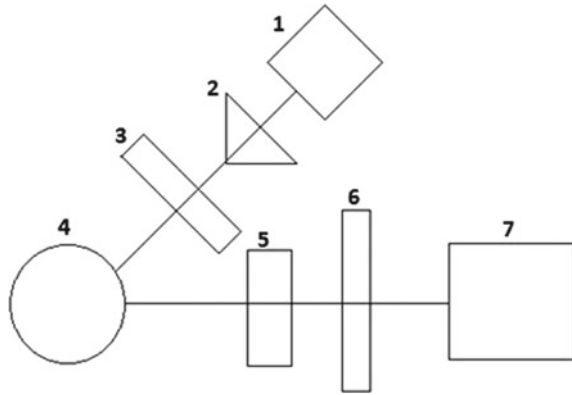
of the tissue on which it was scattered and depolarized. The resulting image shows subsurface layers and no surface structural features [21].

50.3   Data Processing

After shooting two images with reflected and depolarized light we process them in Wolfram Mathematica according to Eq. (50.1) to get a high-contrast image of the investigated area.

$$POL\ DEG = (PAR - PER)/(PAR + PER) \qquad (50.1)$$

PAR stands for an image with parallel axes, PER stands for an image with perpendicular axes, POL DEG is a resultant image that shows polarization degree.

On that step of our experiment, we faced an obstacle. When changing the filters axis, the object moved slightly. As a result, there was an image shift. To correct it we used a reference mark so that the program could distinguish it and process the image considering that mark.

To correct the rotation angle of images with different polarizations, a reference mark search algorithm was used. During the operation of the algorithm, one image was fixed, and the position/angle of rotation of objects in the second image changed relative to it.

Image processing performed as follows. First, we look for key points on both images with SURF [22] detector. After that, we match images key points with k-nearest neighbours algorithm [23] and filter extra ones. As a result, we have control points for further processing (see Fig. 50.2).
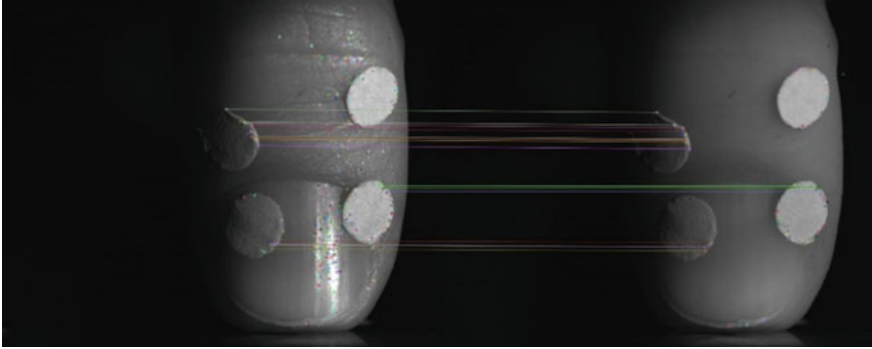
Fig. 50.2 A scar: in the first image the axes are parallel, in the second—perpendicular. By lines mapped marks after filtering are shown

On the next step for each pair of points (the point from the first image and the corresponding point in the second image) the coordinate difference (x, y) is calculated. Based on the average value of the coordinate difference and its sign one of the images shifts up or down, left or right.

After shifting, the image rotation angle is additionally adjusted. Once again, we define key points for the shifted image and match it with second image, which remains unchanged. This process is similar to the explained above. The next step is to define a homography matrix [24] for two images—an image with a shift and an image relatively which the transformation takes place. Homography matrix represents the transformation of one image into another if they are taken from one camera with different angle. Using this matrix, the perspective is restored.

The perspective restoration algorithm provides an accurate overlay of the reference marks while the boundaries of the object do not necessarily coincide (see Fig. 50.3).
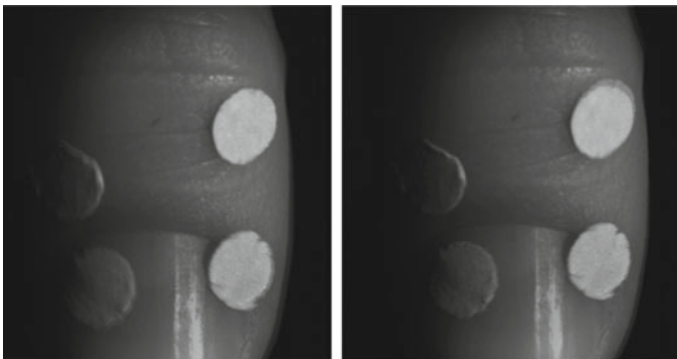
Fig. 50.3 Overlaying a processed image on a fixed one: shift only and shift and restoration perspective

Fig. 50.4   A scar under the red light: parallel axes (left), orthogonal axes (middle), processed image (right)

It is obvious that exact overlap is provided at the reference marks. At the same time the borders of the object (finger) are not superimposed perfectly; in addition, the area of one of the stickers has enlarged (Fig. 50.3). Thus, because of restoring perspective, we can expect an accurate overlay in the area of existing key points, but distortions appear outside of them.

After the shift compensation the image showing the degree of residual polarization was processed according to formula 50.1. An example of the resulting images is shown in Fig. 50.4.

50.4   Results and Discussion

The results of one mole imaging in different light ranges are presented in Figs. 50.5, 50.6, 50.7 and 50.8.

In the resulting picture we see a skin defect area framed with the reference marks. The processed image has a high contrast of the defect and shows the boundaries of it clearer. The contrast ratio in the processed image is formed of the light scattered from different structures in tissue, so it shows the difference of skin layer structures. The resultant image also depends on the pigmentation, as one can see the studied mole is more visible in 440 and 530 nm spectral ranges, while in 575 and 640 nm the mole
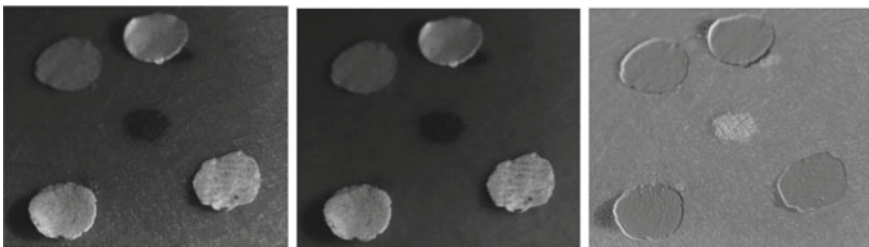
Fig. 50.5   440 nm radiation: parallel axes (left), orthogonal axes (middle), processed image (right)
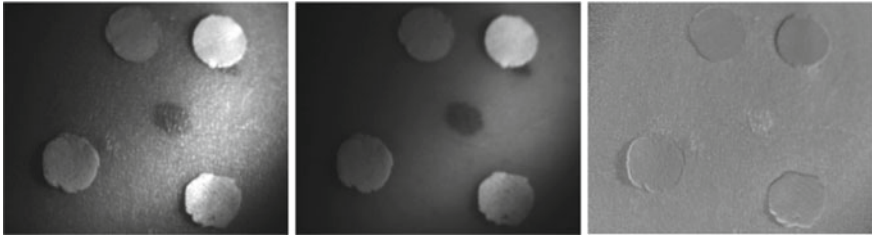
Fig. 50.6 575 nm radiation: parallel axes (left), orthogonal axes (middle), processed image (right)

Fig. 50.7 530 nm radiation: parallel axes (left), orthogonal axes (middle), processed image (right)

Fig. 50.8 640 nm radiation: parallel axes (left), orthogonal axes (middle), processed image (right)

image lost some contrast. We can expect, that different types of skin abnormalities, for example sin cancer, will be more visible in different spectral ranges.

50.5 Conclusion

In this study we suggested a simple experimental setup for joint application of dermatoscopy and SIAscopy methods. We suggested an algorithm for processing detected images in different polarizations. Results on polarized multispectral imaging of benignant mole processed by suggested algorithm are presented. It is visible that some images are darker than the others while diaphragm is not changed, different color filters cause this. Different skin structures absorb light in different ways, this is

the reason why the images of the same birthmark under different light is not the same. The presented setup and processing algorithm can be useful for doctors to distinguish noncancerous and malignant tumors on early stages of cancer. Such images show the boundaries of object, melanin concentration and skin morphology.

References

1. H.A. Shain, B.C. Bastian, From melanocytes to melanomas, in *Nature Reviews Cancer 2016*, vol. 16 (Macmillan Publishers Limited, 2016), pp. 345–358
2. M.A. Galeeva, On specific features of the endoscopic image processing. J. Phys. Conf. Ser. **1**(1236), 012036 (2019)
3. O.A. Melsitov, V.P. Sherendak, S.G. Konovalov, O.O. Myakinin, Automatic malignant melanoma recognition using a dermatoscopy imaging tool, in *International Conference on Information Technology and Nanotechnology (ITNT)* (IEEE, Samara, 2018)
4. E. Savchenko, I. Kolokolnikov, E. Velichko, V. Osovskikh, L. Kiseleva, Z. Musakulova, Design of liver functional reserve estimation technique based on optical densitometry. Diagnostics **10**(8), 599 (2020)
5. L.V. Demidov, D.V. Sokolov, I.V. Bulycheva, B.V. Shashkov, A.N. Makhson, G.N. Vorozhtsov, S.G. Kuzmin, V.V. Sokolov, The progress in the diagnosis of cutaneous melanoma. J. Clin. Oncol. Res. Inst. N. N. Blokhin RCRC RAMS **1**(18), 36–41 (2007)
6. E.N. Velichko, E.K. Nepomnyashchaya, K.G. Gareev, J. Martínez, M.C. Maicas, Characterization of magnetite–silica magnetic fluids by laser scattering. Appl. Sci. **11**(1), 183 (2021)
7. P. Trubin, A. Murashov, D. Suntsov, E. Velichko, Improvement of a dactyloscopic authentication security using polarimetric technique, in *Springer Internet of Things, Smart Spaces, and Next Generation Networks and Systems*, LNCS (Springer, 2020), pp. 67–73
8. Y.Y. Sergeev, V.Y. Sergeev, Dermoscopy use in daily dermatological practice. Kremlin Med. J. **18** (2018)
9. A.G. Ushenko, P.M. Grigorishyn, Diagnosis of hemangioma and blood tissues using laser polarimetry (Repository of Grondo State Medical University, 2019)
10. O.O. Myakinin, A.G. Khramov, D.S. Raupov, S.G. Konovalov, S.V. Kozlov, A.A. Moryatov, Texture analysis in skin cancer tumor imaging, in *Springer Multimodal Optical Diagnostics of Cancer* (Springer, Cham, 2020), pp. 465–504
11. M.B. Aksenenko, T.G. Ruksha, N.B. Tolstikhina, S.R. Kuznetsov, Using SIAscopy to assess the intensity of pathologic changes in skin when melanocytic neoplasms develop. Vestn Dermatol Venereol **4**, 31–36 (2011)
12. H.D. Heibel, L. Hooey, C.J. Cockerell, A review of noninvasive techniques for skin cancer detection in dermatology. Am. J. Clin. Dermatol. **21**, 513–524 (2020)
13. J. Malvehy, G. Pellacani, Dermoscopy, confocal microscopy and other non-invasive tools for the diagnosis of non-melanoma skin cancers and other skin conditions. Acta Derm Venereol **218**, 22–30 (2017)
14. J. Lademann, N. Otberg, H. Richter, L. Meyer, H. Audring, A. Teichmann et al., Application of optical non-invasive methods in skinphysiology: a comparison of laser scanning microscopyand optical coherent tomography with histological analysis. Skin Res. Tech. **13**, 119–132 (2007)
15. M. Baranov, E. Velichko, K. Greshnevikov, Analysis of fractal structures in dehydrated films of protein solutions. Symmetry **13**(1), 123 (2021)

16. I. Govorov, E. Dikareva, E. Velichko, V. Kesic, E. Komlichenko, Human-machine alliance against cervical cancer. Int. J. Gynecol. Cancer (2020)
17. E. Velichko, E. Nepomnyashchaya, M. Baranov, M.A. Galeeva, V.A. Pavlov, S.V. Zavjalov, E. Komlichenko, A concept of smart medical autonomous distributed system for diagnostics based on machine learning technology, in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems* (Springer, Cham, 2020), pp. 515–524
18. K.A. Tiwari, R. Raišutis, J. Liutkus, S. Valiukevičienė, Diagnostics of melanocytic skin tumours by a combination of ultrasonic, dermatoscopic and spectrophotometric image parameters. Diagnostics **10**(9), 632 (2020)
19. I.A. Bratchenko, M.V. Alonova, O.O. Myakinin, A.A. Moriatov, S.V. Kozlov, V.P. Zakharov, Hyperspectral visualization of skin pathologies in the visible spectrum. Comput. Opt. **2**(40), 240–248 (2016)
20. A.N. Yaroslavsky, X. Feng, S.H. Yu, P.R. Jermain, T.W. Iorizzo, V.A. Neel, Dual-wavelength optical polarization imaging for detecting skin cancer margins. J. Invest. Dermatol. **140**, 1994–2000 (2020)
21. V.V. Tuchin, Optics of biological tissues. Fizmatlit (2012)
22. H. Bay, A. Ess, T. Tuytelaars, L.V. Gool, Speeded up robust features computer vision and image understanding. **110** (3), 346–359 (2008)
23. M. Cao, W. Jia, Z. Lv, W. Xie, L. Zheng, X. Liu, Two-pass K Nearest Neighbor search for feature tracking. IEEE Access **6**, 72939–72951 (2018)
24. R. Szeliski, *Computer Vision: Algorithms and Applications*, 1st edn. (Springer, Berlin, Heidelberg, 2010)

Chapter 51
Features of Degradation of Silicon-Based Solar Photovoltaic Cells

Linda Boudjemila , **Vadim V. Davydov** , **Vladislav G. Malyshkin** , **and Vasiliy Yu. Rud'**

Abstract The article substantiates the need for the development of solar power plants in the energy system of various countries. Currently known solar cells have limited band gaps which makes their efficiency not very high. It has been observed that in long-term operation, degradation is the most important factor limiting the efficiency of photovoltaic cells, which is already low. Various factors caused by nature that reduce the efficiency of solar energy systems are mentioned. So, this work intends to present a method for efficient analysis of the photovoltaic cells degradation data. The necessity of developing methods for modeling the efficiency of a solar power plant for the photovoltaic cells degradation study has been proved. The calculation results for the rates of Si photovoltaic cells degradation are presented for different countries. It was noted that the main factor in the degradation process formation is manufacturing technology of photovoltaic cells.

Keywords Solar power plant · Degradation · Silicon · Photovoltaic cell · Simulation · Solar energy conversion efficiency

51.1 Introduction

The development of scientific and technological progress requires more and more electrical energy for a stable functioning of various spheres of human activity every

L. Boudjemila · V. V. Davydov (✉)
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

L. Boudjemila
Department of Mechanical Engineering, University of Science and Technology, Houari Boumedien Bab Ezzouar 16111, Algiers, Algeria

V. V. Davydov · V. Yu. Rud'
All-Russian Research Institute of Phytopathology, Moscow Region 143050, Russia

V. G. Malyshkin · V. Yu. Rud'
Ioffe Institute, St. Petersburg 195256, Russia

year [1–8]. The deterioration of the environmental situation in the world [3, 9–13] creates stable preconditions for the development and introduction of alternative sources of electrical energy into the energy system. One of these is solar energy [11–15]. The rapid development of solar energy is due not only to the solution of environmental problems [9–11, 16–18]. There are also a number of other factors. The main ones are:

- A constant decrease of solar cell price from $3.88/W in 2008 to $0.80/W in 2013 [19] and then under $0.23/W in 2020 [20];
- Increasing the efficiency of solar cells. Currently, the best laboratory photovoltaic converters based on silicon have an efficiency of above 28%, and commercial cells of 22% and above [21];
- State financing of solar energy development programs and tariff regulation [14, 15].

The solar cells are typically classified in four generations based on time of development and categories of materials used [22–24]. The most common solar cells on the market are the first-generation solar cells which comprise mono-crystalline and poly-crystalline silicon. In 2012–2021 silicon wafer prices have undergone more than 10 times decline [24, 25] which made mono-crystalline and polycrystalline silicon technology most attractive and put thin-film solar cells at a disadvantage. Many companies have been forced to update their technology or go bankrupt [26]. This has required extensive research, both in academia and industry (e.g. the operation of solar panels in power plants under various conditions). One of the important elements in these studies is the modeling of the operation of a solar power plant, taking into account the climatic conditions of operation of photovoltaic converters, the processes of their degradation, etc.

51.2   Features of the Degradation of Solar Photovoltaic Cells

It should be noted that the processes of degradation of solar photovoltaic cells are the main reason that reduces the amount of power generated by a solar power plant during its long-term operation [14, 15, 19–26]. The climatic factors affect the acceleration of the degradation processes of photovoltaic cells. Therefore, it is extremely important to have these data when simulating the operation of a solar power plant. In the Figs. 51.1 and 51.2 we present NASA data on climatic factors (illumination of a horizontal surface in energy units per day, temperature and wind speed; these last two parameters are measured at 2 m altitude. In Fig. 51.1 the data on climatic factors in the territory of the city of Algeria, Algeria, and in Fig. 51.2 the city of St. Petersburg, Russian Federation, is presented for 10 years. A period of 10 years is the effective life of a solar power plant. The type of PV cells chosen is mono-crystalline silicon in both locations as this type is very common within easy reach. An analysis of the
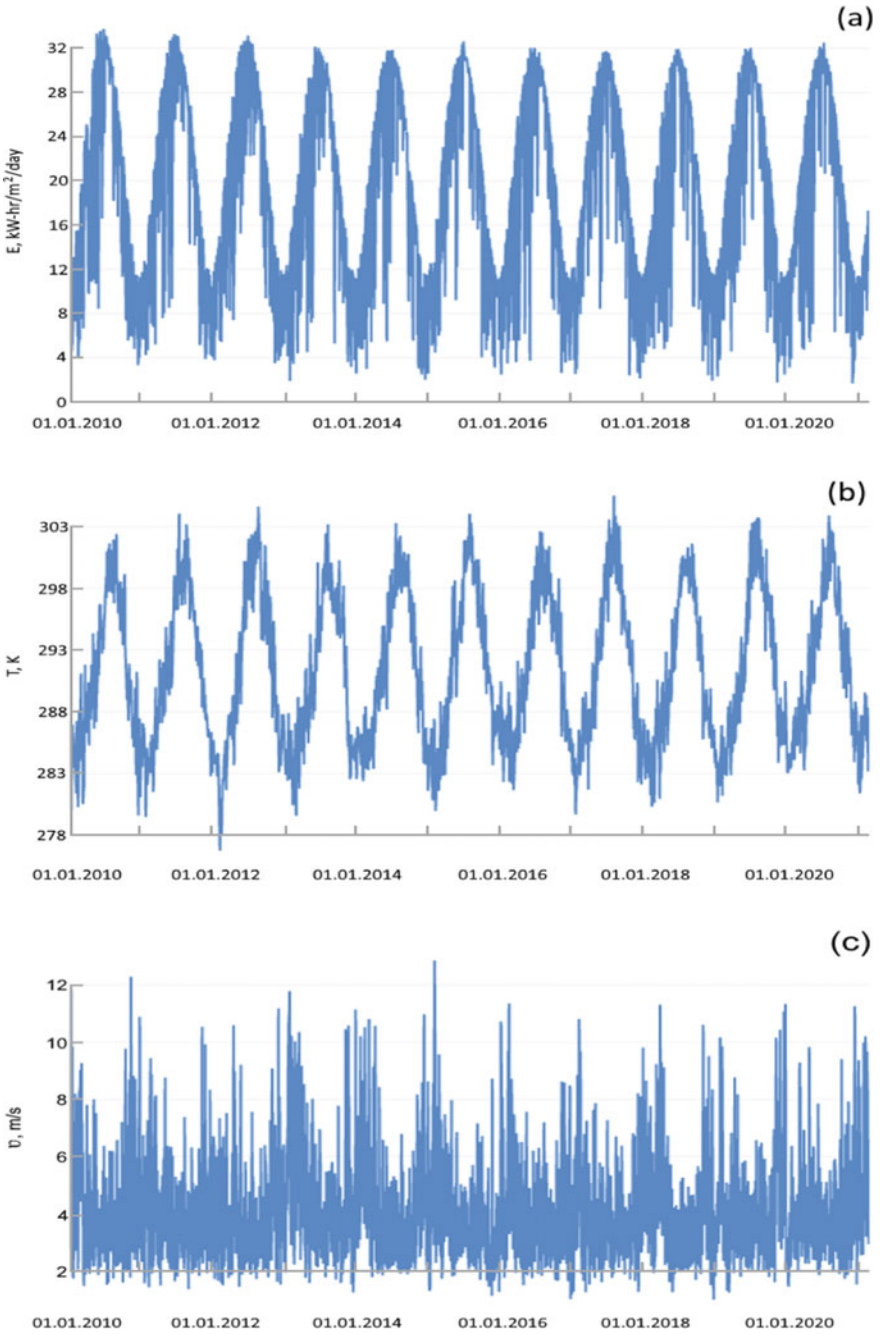
Fig. 51.1 The climatic data over 10 years in Algeria: incoming solar irradiation **a**, temperature T **b**, wind speed **c**
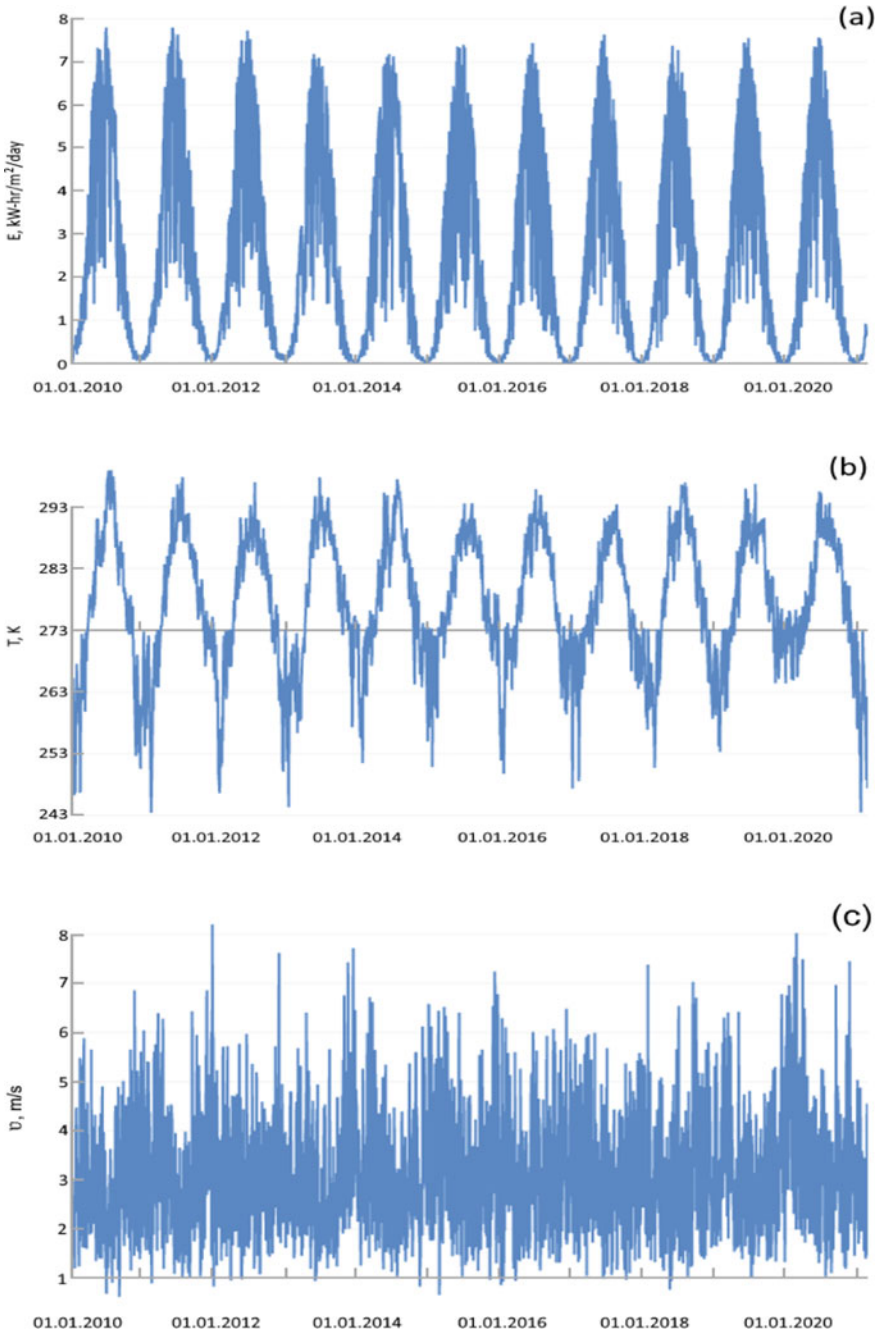
Fig. 51.2 The climatic data over 10 years in St. Petersburg: incoming solar irradiation **a**, temperature T **b**, wind speed **c**

climatic conditions between the two cities shows large differences in illumination and temperature and little difference in wind speed. The first two factors have a significant impact on the power generated by a solar power plant throughout the year. Wind speed has a different effect on the rate of degradation of photovoltaic converters (warm or cold wind and humidity are of great importance).

In case of low temperatures and strong winds, the process of degradation of contacts on photovoltaic cells is faster. This fact must be taken into account when modeling degradation processes in St. Petersburg. In Algeria, the impact of these processes is less significant. In Algeria, the degradation process is affected by the heating of photovoltaic cells (their expansion occurs causing displacement of layers and damage to contacts). The various heterogeneities in solar panels associated with production technology are beginning to have a more significant impact on the work-flow. The preferred method of degradation estimation in such a situation is the annual (YOY) approach [21]. The algorithm includes the following: normalizing the measured PV power, by site irradiance and temperature; filtering the data (the most common filter is a low irradiance cutoff of about 200 W/m^2); data analysis. There are actually two major categories of degradation calculation methods: with an independent measurement of solar irradiation and without it. Advances in algorithm processing [21, 22] allows to obtain sufficiently accurate value of the degradation rate even without independent irradiation measurements. Data availability is one of the major issues. NREL maintains an array of PV systems for the US, representing many PV module technologies with an operating history in exceed of 10 years. Global data is much more scarce and require substantial effort to obtain. Figures 51.3 and 51.4 show the results of calculations of the rate of distribution of the generated power by a solar power plant and the confidence interval for the degree of degradation of photovoltaic stations for Algeria and St. Petersburg.

The obtained results of modeling the operation of a solar power plant show that degradation processes occur more actively at solar power plants located in the region of St. Petersburg. For the research, data was taken on the operation of SES, in which photovoltaic converters of the same company are used.
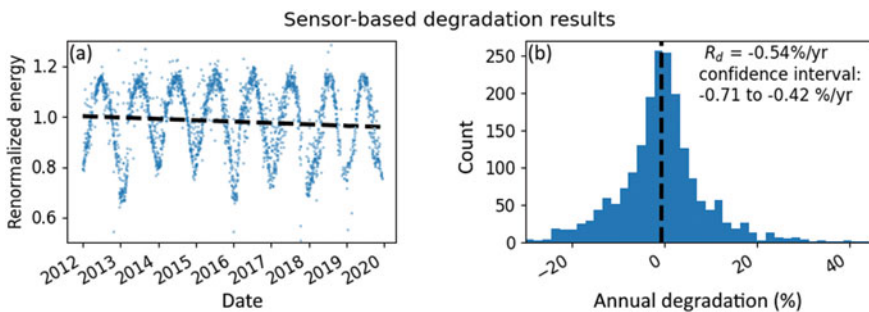
Fig. 51.3 Work of solar power plants in Algeria over the past 8 years
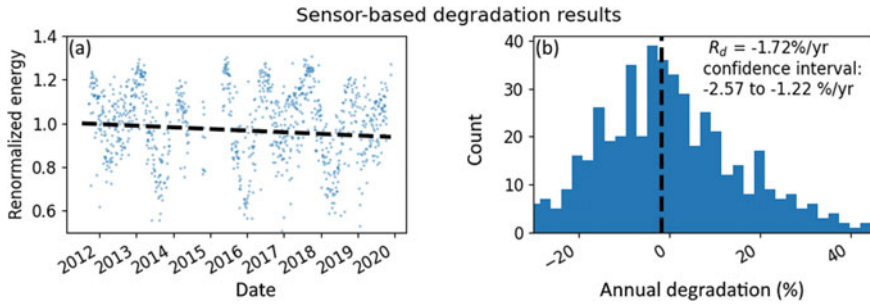
Fig. 51.4 Work of solar power plants in St. Petersburg over the past 8 years

51.3 Conclusion

The results obtained show that the location of the solar power plant has a major role in the degradation. The location of Saint Petersburg provides more complicated climatic conditions than the city of Algeria which is located in North Africa. The rate of degradation in Saint Petersburg (with rainy, windy and snowy weather) is quite high because of the air humidity which is much higher comparing to the city of Algeria which has humid but very hot climate. Another important factor in increasing the rate of degradation of photovoltaic cells is the presence of violations or minor deviations during their transportation which can cause some invisible cracks leading to fast degradation. Especially in case of high humidity and temperature, the defects become visible to the naked eye. The damage in one photovoltaic cell, which is installed on a solar panel, leads to increase in the rate of degradation of other elements located on this panel even if they do not have defects.

The results obtained show that the climatic conditions of operation of photovoltaic cells do not have a major effect on the rate of their degradation during the first years of use. The amount of degradation defects becomes more important and affects the efficiency after a long period of functioning in a harsh weather.

References

1. N.A. Lukashev, A.P. Glinushkin, V.S. Lukyantsev, Microwave low mass-dimensional frequency standard on Hg-199 ions. J. Phys. Conf. Ser. **1410**(1), 012211 (2019)
2. G. Fokin, Passive geolocation with unmanned aerial vehicles using TDOA-AOA measurement processing, in *Proceedings—21st International Conference on Advanced Communication Technology (ICACT)* (PyeongChang Kwangwoon_Do, South Korea, 2019), pp. 360–365
3. A.H.A. Al-odhari, G. Fokin, A. Kireev, Positioning of the radio source based on time difference of arrival method using unmanned aerial vehicles, in *2018 Systems of Signals Generating and Processing in the Field of on Board Communications* (Moscow, 2018), pp. 1–5
4. M. Makolkina, V.D. Pham, R. Kirichek, A. Gogol, A. Koucheryavy, Interaction of AR and IoT applications on the basis of hierarchical cloud services, in *Lecture Notes in Computer*

Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11118 LNCS (2018), pp. 547–559

5. G.A. Pchelkin, V.B. Fadeenko, Features of the transmission of microwave signals at offshore facilities. J. Phys. Conf. Ser. **1368**(2), 022045 (2019)

6. V.B. Fadeenko, G.A. Pchelkin, O.O. Beloshapkina, Features of construction of the scheme of fiber-optic communication system for transmission of analog signals in the frequency range from 0.135 to 40 GHz. J. Phys. Conf. Ser. **1410**(1), 012238 (2019)

7. A.A. Petrov, Some directions of quantum frequency standard modernization for telecommunication systems, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11118 LNCS (2018), pp. 641–648

8. A.A. Petrov, On the potential application of direct digital synthesis in the development of frequency synthesizers for quantum frequency standards. J. Commun. Technol. Electron. **63**(11), 1281–1285 (2018)

9. N.S. Myazin, V.V. Yushkova, N.I. Taranda, V.Yu. Rud, On the need to control the state of the flowing media by the values of relaxation constants. J. Phys. Conf. Ser. **1410**(1), 012130 (2019)

10. N.S. Myazin, V.V. Yushkova, V.I. Dudkin, New method for determining the composition of liquid media during the express control of their state using the nuclear magnetic resonance phenomena. J. Phys. Conf. Ser. **1400**(6), 066008 (2019)

11. A.A. Petrov, Rubidium atomic clock with improved metrological characteristics for satellite communication system, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10531 LNCS (2017), pp. 561–568

12. K.J. Smirnov, InP/InGaAs photocathode for hybrid SWIR photodetectors. J. Phys. Conf. Ser. **1368**(2), 022073 (2019)

13. K.J. Smirnov, S.F. Glagolev, G.V. Tushavin, High speed near-infrared range sensor based on InP/InGaAs heterostructures. J. Phys. Conf. Ser. **1124**(1), 022014 (2018)

14. A.V. Bobyl, A.G. Zabrodskii, S.A. Kudryashov, V.M. Makarov, E.E. Terukova, A.F. Erk, Renormalized model for solar power plants economic efficiency evaluation. Energy **6**, 46–58 (2017)

15. A.V. Sachenko, Yu.V. Kryuchenko, V.P. Kostylyov, A.V. Bobyl, E.I. Terukov, S.N. Abolmasov, A.S. Abramov, D.A. Andronikov, M.Z. Shvarts, I.O. Sokolovskyi, M.S. Evstigneev, Temperature dependence of photoconversion efficiency in silicon heterojunction solar cells: theory vs experiment. J. Appl. Phys. **119**(22), 225702 (2016)

16. S. Van, A. Cheremisin, A. Chusov, F. Switala, R. Davydov, New architectural and planning solutions as a mean of ecological modernization and reconstruction of urban environment (on the example of Shanghai city, China). IOP Conf. Ser. Earth Environ. Sci. **390**(1), 012011 (2019)

17. S. Van, A. Cheremisin, A. Glinushkin, R. Davydov, V. Yushkova, Application of new architectural and planning solutions to create an ecological city (on the example of Shanghai, China). E3S Web Conf. **140**, 09008 (2019)

18. N.S. Myazin, V.V. Davydov, V.V. Yushkova, V.Y. Rud', A new method of determining the state of water and agricultural areas in real time. Environ. Res. Eng. Manag. **75**(2), 28–35 (2019)

19. M. Bazilian, I. Onyeji, M. Liebreich, I. MacGill, J. Chase, J. Shah, S. Zhengrong, Reconsidering the economics of photovoltaic power. Renew. Energ. **53**, 329–338 (2013)

20. M. Fischer, ITRPV 11 th edition. Report presentation and key findings. Web-Seminar June 9 (2020)

21. D. Jordan, C. Deline, S. Kurtz, G. Kimball, M. Anderson, Robust PV degradation methodology and application. IEEE J. Photovoltaics **8**(2), 525–531 (2018)

22. J.J. John, A. Alnuaimi, A. Elnosh, M. Stefancich, P. Banda, Estimating degradation rates from 27 different PV modules installed in desert conditions using the NREL/Rdtools, in 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (IEEE, 2018), pp. 712–714

23. M. Dhimish, A. Alrashidi, Photovoltaic degradation rate affected by different weather conditions: a case study based on PV systems in the UK and Australia. Electronics **9**(4), 650–653 (2020)

24. L. Lima, C. Vasconcelos, J. Simão, H. de Mendonça, The quantitative easing effect on the stock market of the USA, the UK and Japan. J. Econ. Stud. **10**, 0081 (2016)
25. T. Miyakoshi, J. Shimada, K. Li, The dynamic effects of quantitative easing on stock price: evidence from Asian emerging markets, 2001–2016. Int. Rev. Econ. Finan. **49**, 548–567 (2017)
26. M. Taguchi, E. Maruyama, M. Tanaka, Temperature dependence of amorphous/crystalline silicon heterojunction solar cells. Jpn. J. Appl. Phys. **47**(2), 814–818 (2008)

Chapter 52
Simultaneous Measurement of Strain and Curvature with a Fiber Optic Singlemode-Multimode-Singlemode Structure

Aleksandr Markvart, Leonid Liokumovich, and Nikolai Ushakov

Abstract Current paper reports a novel approach for simultaneous measurement of strain and curvature by a single mode-multimode-single mode (SMS) sensor based on different dependencies of mode propagation constants on the perturbations. For that, optical transmission spectra of SMS sensor are to be measured and processed with a fast Fourier transform (FFT). FFT analysis allows to decompose the complex interference signal onto components corresponding to individual mode groups. This, in turn, allowed us to carry out simultaneous measurement of strain and curvature, which, to the best of our knowledge, was done for the first time with a SMS sensor.

Keywords Optical fiber sensors · Intermode interference · Single-mode-multimode-single-mode · Sms sensor · Spectral interferometry · Strain · Curvature · Fourier transform

52.1 Introduction

Fiber-optic sensors are a rapidly developing area of interest for both academia and industry due to their high accuracy, immunity to electromagnetic interference, easy integration into complex constructions, small footprint, ability to multiplex several sensors and perform remote measurements. Measurement of different physical quantities like temperature, strain, curvature and so on is required for various applications, including, but not limited to structural health monitoring and industrial inspection, medical diagnostics for health monitoring and during surgery. One the most widely used fiber optics sensors is a fiber Bragg grating (FBG). It's main limitation in practical tasks is cross-sensitivity and impossibility of simultaneous registration of different impacts (for example strain and temperature) by a single sensor. The using of multiple sensors with different sensitivities is needed in this case, which complicates the scheme.

A. Markvart (✉) · L. Liokumovich · N. Ushakov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: markvart_aa@spbstu.ru

One of the most promising proposed types of fiber optic sensors is sensor based on singlemode-multimode-singlemode (SMS) structure. In simplest case it consists of three sections of optical fibers, where the multimode fiber (MMF) is spliced between two singlemode fibers (SMF), is straightforward to fabricate and low-cost [1]. The MMF section is a sensing element here, which typical length ranges from several to tens of centimeters, depending on the type of MMF. The first SMF launches several modes in MMF, which are then coupled to the main mode of the second SMF. Due to different propagation constants of MMF's modes, the intensity of output SMF mode depends on their optical path differences (OPD), hence, demonstrating interference behaviour. In fact, SMS sensor is a variant of mode-mode interferometer [2, 3]. The sensor interrogation is usually carried by spectral interferometry [4, 5]. It was demonstrated that it is possible to eliminate the cross-sensitivity between different physical quantities and measure them simultaneously with a single SMS sensor [6–8]. It was also shown that SMS sensors have higher temperature, strain sensitivities compared with FBGs [4, 8, 9]. Silva et al. [6] demonstrated the curvature sensor without cross-sensitivity to temperature and strain. Wu et al. [7] demonstrated simultaneous temperature and displacement measurements. Lu et al. [8] demonstrated simultaneous temperature and strain measurement.

However, the main factor, limiting the further development of state of the art SMS sensors is the use of simple signal processing algorithms, such as peak/dip wavelength tracking and peak/dip width estimation. It limits the dynamic range and the resolution of such sensors. Moreover, a detailed analysis of spectral characteristics of SMS structures may result in better understanding of their performance and fundamental limitations. In other areas of spectral interferometry applications, such as extrinsic Fabry-Perot interferometers (EFPI) and FBGs, a great attention is paid to development of advanced signal processing algorithms [10–13].

Here we present a novel approach for SMS interference signal processing for efficient simultaneous measurement of different impacts. Simultaneous measurement of strain and curvature is, to the best of our knowledge, reported for the first time.

52.2 Description of Singlemode-Multimode-Singlemode Structure

The investigated SMS structure was fabricated using standard fiber splicer Fujicura FCM 45M. As a multimode section, we have used a 32 cm long step-index Thorlabs FG050LGA fiber. At both ends, it was spliced to SMF-28 patchcords. The interrogation of SMS sensor was performed by NI PXIe 4844 optical sensor interrogator. Since it is generally used for interrogation of FBG and EFPI sensors, for which reflective spectral characteristics are generally analyzed, we had to use two channels of interrogator, using channel#1 as a source and channel#0 as a receiver, for that it was connected to SMS output via optical circulator as shown on scheme in left part of Fig. 52.1.
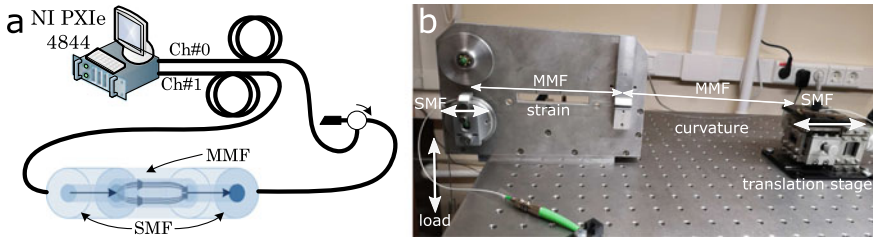
Fig. 52.1 The scheme of SMS sensor and interrogation setup (left); experimental arrangement for controlling strain and curvature of MMF section

The main goal of this work was to simultaneously measure strain and curvature of MMF section of SMS. Since simultaneous and precisely controlled bending and straining of fiber requires very sophisticated setup, we have decided to apply strain and bending to two distinct parts of MMF, which was done by a special experimental arrangement, shown in the right part of Fig. 52.1. The left part of MMF section was subject to strain, while the right part was bent. The middle of multimode fiber was fixed by a stationary clamp; SMF on the left side was attached to a roller, to which different loads were applied, causing strain of left part of MMF; SMF on the left side was attached to a translation stage, which was shifted in left direction to increase the bend, which can be assumed close to circular in case of small slack of the fiber. The bending was done analogously to [6]. This relatively simple setup allowed to simultaneously and independently control strain and curvature of two parts of MMF section.

Transmission spectra of SMS sensor and FFT of one of the spectra are shown in Fig. 52.2, a, and b, respectively. FFT of presented spectra are nearly indistinguishable, therefore, only one of them is illustrated. Complex structure of spectra and their FFT indicate that there is a large number of modes that are excited in multimode fiber and interfere when are coupled to the main mode of output SMF [5]. Spectra shown in Fig. 52.2, a were obtained in three cases: no strain, curvature $2.75\,\text{m}^{-1}$; slight strain $0.32\,\text{m}\epsilon$, curvature $2.75\,\text{m}^{-1}$; no strain, curvature $2.85\,\text{m}^{-1}$. It can be seen in enlarged fragments of SMS spectra that the behaviour of spectra dips due to various perturbations of MMF section is quite complex and unpredictable. Therefore, instead of empirically investigating how different dips (or peaks) behave in case of different perturbations, we propose to use integral parameters of SMS spectra, for example, phases of FFT components, as for example, was done in [14] for high-birefringence fiber loop sensors.

FFT of SMS spectrum can be represented as a decomposition of complex spectrum, resulting from OPDs. Since propagation constants for different modes of step-index fiber are not equidistant [15], the spacing between the FFT peaks increases as the OPD of interference component becomes greater. Since resolution of spectra FFT in OPD domain in presented case is relatively low ($\sim$30 μm, which is limited by the width of the spectral range, in which the spectra are measured), interference components with smaller OPDs overlap in FFT domain and it is hard to extract infor-
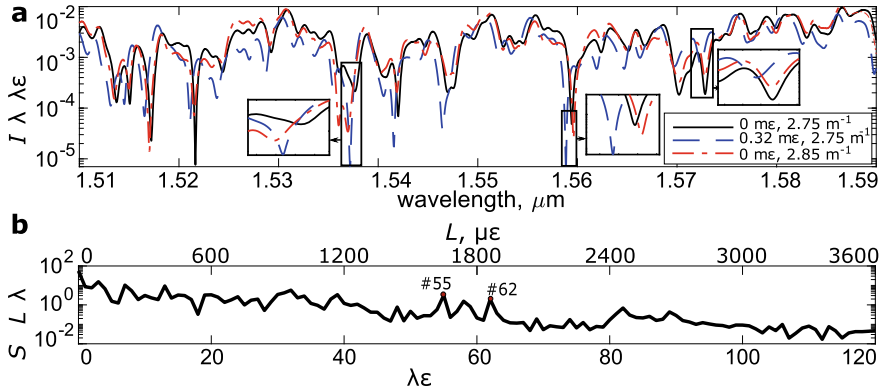
Fig. 52.2 Experimental spectra of SMS sensor (**a**) and FFT of typical SMS spectrum (**b**)

mation from them independently. On contrary, in case of larger OPDs interference components become unambiguously distinguishable, and for instance, FFT components #55 and #62 have relatively large amplitude and are presented by clear peaks, and will be used in the following analysis.

52.3 Signal Analysis

An experiment for calibrating sensor sensitivities to strain and curvature was performed. For that, we have varied strain in limits from $0\,\mathrm{m}\varepsilon$ to $0.32\,\mathrm{m}\varepsilon$ with step $0.08\,\mathrm{m}\varepsilon$ (5 steps overall) and curvature in limits from 2.75 to $2.85\,\mathrm{m}^{-1}$ with step $0.02\,\mathrm{m}^{-1}$ (6 steps overall). During each step of curvature variation, the whole range of strain values was scanned, as a result, 30 points with different pairs of strain and curvature values were measured. For each point, 100 SMS spectra were measured in order to increase the accuracy of sensor sensitivities' estimation. All spectra were saved and further processed in Matlab.

For each measured spectra, phases of 55-th and 62-nd FFT samples, denoted as φ_{62} and φ_{55} were calculated, their dependencies on strain ε and curvature ρ were fitted by plane equations, which are presented in Eq. (52.1). Cross-sections of φ_{62} and φ_{55} dependencies on strain and curvature are shown in Fig. 52.3a and b, respectively.

$$
\begin{aligned}
\varphi_{62} &= 6.64 \cdot 10^{-3} \times \Delta\varepsilon + 6.04 \times \Delta\rho, \\
\varphi_{55} &= 5.40 \cdot 10^{-3} \times \Delta\varepsilon + 1.07 \times \Delta\rho,
\end{aligned}
\tag{52.1}
$$

Adjusted R^2 values were calculated for each fit, comprising $R^2_{adj} = 0.9997$ for φ_{62} and $R^2_{adj} = 0.9992$ for φ_{55}, which indicates excellent linearity of the sensor. Also, from fit coefficients, it can be seen that the system of equations, relating φ_{62} and φ_{55}
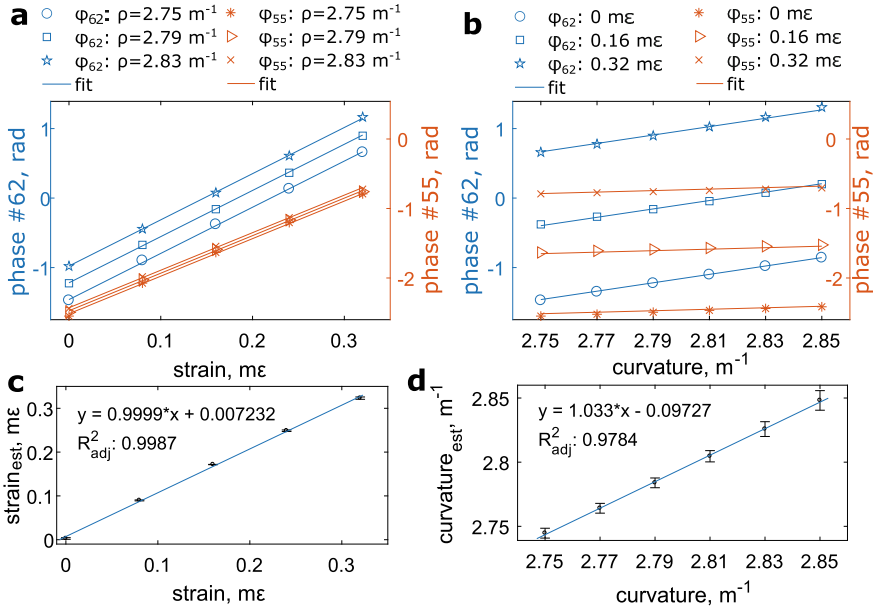
Fig. 52.3 Dependencies of phases of 55-th and 62-nd FFT samples on strain (**a**) and curvature (**b**); dependencies of estimated strain on induced strain (**c**) and estimated curvature on induced curvature (**d**)

with strain and curvature, is non-degenerate, allowing one to simultaneously obtain the sensor readings for strain and curvature from experimentally measured φ_{62} and φ_{55}.

The possibility to correctly estimate strain and curvature values was validated by finding the solution of system of Eq. (52.1) and calculating ε and ρ for all measured SMS spectra. Since the values of induced strain and curvature were known, the dependencies of estimated values on the true ones were analyzed and linearly fitted, corresponding plots can be seen in Fig. 52.3, c and d. Corresponding fitting equations and adjusted R^2 values are also presented in Fig. 52.3. Again, exceptional linearity with no cross-talk was obtained, with larger variation of sensor readings in case of curvature measurement, which may be due to nonideal conditions of experiment.

52.4   Conclusion

Simultaneous measurement of strain and curvature by singlemode-multimode-singlemode is reported for the first time. Proposed SMS sensor consisted of conventional step-index multimode fiber, spliced between two SMF-28 fiber. Novel FFT-based algorithm for processing SMS signals was proposed, which demonstrated

excellent linearity and no cross-talk during simultaneous measurement of strain and curvature of multimode fiber section.

References

1. Q. Wu, Y. Qu, J. Liu, J. Yuan, S.P. Wan, T. Wu, X.D. He, B. Liub, D. Liuc, Y. Ma, Y. Semenova, P. Wang, X. Xin, G. Farrell, Singlemode-multimode-singlemode fiber structures for sensing applications—a review. IEEE Sens. J. 1–1 (2020). https://doi.org/10.1109/JSEN.2020.3039912
2. O.I. Kotov, M.A. Bisyarin, I.E. Chapalo, A.V. Petrov, Simulation of a multimode fiber interferometer using averaged characteristics approach. Journal of the Optical Society of America B **35**(8), 1990–1999 (2018). https://doi.org/10.1364/josab.35.001990
3. A.V. Petrov, I.E. Chapalo, M.A. Bisyarin, O.I. Kotov, Intermodal fiber interferometer with frequency scanning laser for sensor application. Appl. Opt. **59**(33), 10,422–10,431 (2020). https://doi.org/10.1364/AO.400345. http://ao.osa.org/abstract.cfm?URI=ao-59-33-10422
4. S.M. Tripathi, A. Kumar, R.K. Varshney, Y.P. Kumar, E. Marin, J.P. Meunier, Strain and temperature sensing characteristics of single-mode–multimode–single-mode structures. J. Lightwave Technol. **27**(13), 2348–2356 (2009). http://jlt.osa.org/abstract.cfm?URI=jlt-27-13-2348
5. M. Kumar, A. Kumar, S.M. Tripathi, A comparison of temperature sensing characteristics of sms structures using step and graded index multimode fibers. Opt. Commun. **312**, 222–226 (2014). https://doi.org/10.1016/j.optcom.2013.09.034. http://www.sciencedirect.com/science/article/pii/S0030401813008559
6. S. Silva, O. Frazão, J. Viegas, L.A. Ferreira, F.M. Araújo, F.X. Malcata, J.L. Santos, Temperature and strain-independent curvature sensor based on a singlemode/multimode fiber optic structure. Measurement Sci. Technol. **22**(8), 085,201 (2011). https://doi.org/10.1088/0957-0233/22/8/085201
7. Q. Wu, A.M. Hatta, P. Wang, Y. Semenova, G. Farrell, Use of a bent single SMS fiber structure for simultaneous measurement of displacement and temperature sensing. IEEE Photon. Technol. Lett. **23**(2), 130–132 (2011). https://doi.org/10.1109/LPT.2010.2093515
8. C. Lu, J. Su, X. Dong, T. Sun, K.T.V. Grattan, Simultaneous measurement of strain and temperature with a few-mode fiber-based sensor. J. Lightwave Technol. **36**(13), 2796–2802 (2018). https://doi.org/10.1109/JLT.2018.2825294
9. Q. Wu, Y. Semenova, A.M. Hatta, P. Wang, G. Farrell, Bent sms fibre structure for temperature measurement. Electron. lett. **46**(16), 1129–1130 (2010)
10. N.A. Ushakov, L.B. Liokumovich, Signal processing approach for spectral interferometry immune to λ/2 errors. IEEE Photon. Technol. Lett. **31**(18), 1483–1486 (2019). https://doi.org/10.1109/LPT.2019.2934093
11. C. Li, S. Chen, Y. Zhu, Maximum likelihood estimation of optical path length in spectral interferometry. J. Lightwave Technol. **35**(22), 4880–4887 (2017). https://doi.org/10.1109/JLT.2017.2743214
12. Z. Yu, A. Wang, Fast white light interferometry demodulation algorithm for low-finesse fabry-perot sensors. IEEE Photon. Technol. Lett. **27**(8), 817–820 (2015). https://doi.org/10.1109/LPT.2015.2391912
13. D. Tosi, Advanced interrogation of fiber-optic bragg grating and fabry-perot sensors with KLT analysis. Sensors **15**(11), 27,470–27,492 (2015). https://doi.org/10.3390/s151127470
14. D. Leandro, M. Bravo, A. Ortigosa, M. Lopez-Amo, Real-time FFT analysis for interferometric sensors multiplexing. J. Lightwave Technol. **33**(2), 354–360 (2015). https://doi.org/10.1109/JLT.2014.2388134
15. H.G. Unger, H. Unger, *Planar Optical Waveguides and Fibres* (Clarendon Press, Oxford, 1977)

Chapter 53
A Broadband Analog Fiber-Optic Line with Recirculating Memory Loop for Variable Microwave Signal Delay

Dmitrii V. Kondakov 🆔**, Sergei I. Ivanov** 🆔**, and Alexander P. Lavrov** 🆔

Abstract An active device for variable time delay of broadband radio signals is described. The device is made on the base of analogue fiber-optic link included in software-reconfigurable recirculating memory loop. An analogue fiber-optic link is an intensity modulated direct detection type link with an external integrated-optical Mach–Zehnder modulator. The time delay is changed in steps: multiples of the minimum delay time. The minimum time was 552 ns in our case. It was determined by the length of fiber-optic link, which was about 100 m of SMF-28 fiber. To change the time delay, recirculation was organized using high-speed radio frequency switches. Switches control was carried out using FPGA. The developed device works with radio signals in 0.5…14 GHz frequency band. Experiments have been carried out, which have shown the device operability in a wide range of radio signal delay times. The number of cycles of passage of the radio signal in the memory loop varied in the range from 1 to 30. The error in setting the radio signal delay time was no more than 2 ns. The influence of the loop gain in the recirculating memory loop on the output signal level and noise level in the developed delay line is estimated. The obtained measurement results were used to calculate the dynamic range of the delay line for different numbers of recirculation cycles.

Keywords Broadband analogue fiber-optic link · Recirculating memory loop · Microwave signal variable delay

53.1 Introduction

Microwave photonic systems have many advantages over traditional only electronic systems: low propagation loss, high instantaneous bandwidth, and insensitivity to electromagnetic interference [1, 2]. So that systems included fiber-optic line were used for variety of applications: in diverse signal processing and generation, reference clock distribution in radioastronomy, phased array antenna beamforming, radar range

D. V. Kondakov · S. I. Ivanov (✉) · A. P. Lavrov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: ivanov_si@spbstu.ru

testing, counter measurement tasks and many others [1–11]. The important utilization of a fiber-optic delay line in modern electronic systems is a 'simple' wideband signal memory. Other possible variants of wideband signal delay line realization one can find in [12, 13].

In some tasks like radar range testing or electronic countermeasure applications an RF memory is necessary to store complex-waveform microwave pulses for retransmission them or detailed explore them. A RF memory within fiber-optic delay approaches is much more wideband variant then one within electronic approaches due to A/D conversion. When RF memory has to be changed in wide range of time delay than recirculating type RF memory is an attractive one [14–17].

In recirculating type RF delay lines, the RF signal is converted to the optical domain, and at the end of the delay line, to obtain the RF signal again, conversion from the optical domain is performed. The signal delay time is determined by the length of the optical path (the length of the optical fiber). The recirculation loop itself is realized in the optical domain or in the radio frequency domain. The delay time can be changed in multiples by changing the number of repeated passes through the loop. The input and output of an RF signal to the delay line and control of the number of signal passes through the loop is performed by using high-speed switches installed at the input and output of the delay line. A delay line of this type was investigated in [18].

We have been studied the functioning of this type analog fiber-optic recirculating delay line with switchable time delay focused on more thorough measurements of signal amplitude when it reticulates in loop. Also, in our laboratory setup of a fibre optic recirculating delay line, we use RF optics and opto-RF converters with a wider radio frequency band than in [18].

53.2 Analogue Fiber-Optic Switchable Recirculating Delay Line

A block diagram of the fiber-optic switchable recirculating delay line (DL) is presented in Fig. 53.1. The base of DL is a wideband analogue fiber-optic link. In the laboratory set-up we used Optiva OTS-2 (Emcore) link. This link is intensity modulation direct detection type (IM-DD) one [19, 20].

An input RF signal comes through switch SW1, amplifier Amp, and attenuator ATT to input of Rf-optic converter Tr Emcore. This link transmitter module realizes intensity modulation (IM) of DFB laser radiation (wavelength in 1.5 mcm band, ITU grid) by MZM modulator driven by input RF signal with frequencies in 0.05…18 GHz band. We can comment: analogue optical transmitters with external intensity modulation by MZM modulators have significantly wider RF bandwidth than optical transmitters with direct intensity modulation by laser injection current modulation [1, 20].
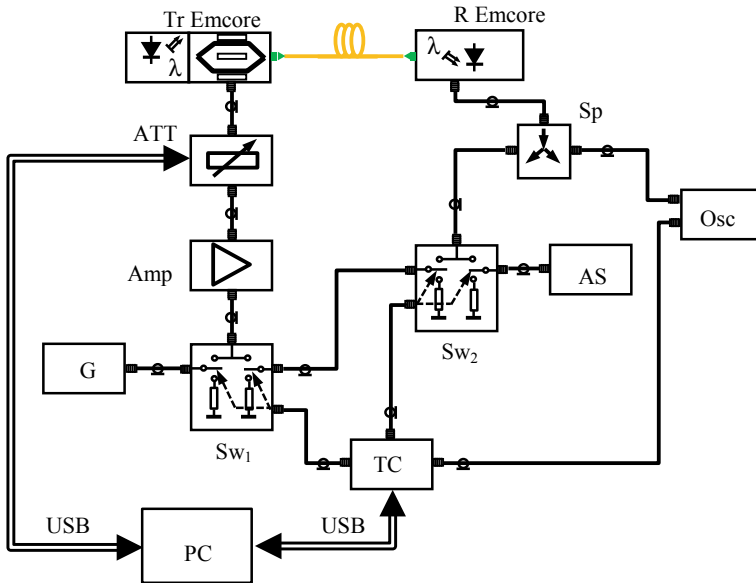
Fig. 53.1 The block diagram of the fiber-optic switchable recirculating delay line

Intensity modulated optical signal is delayed in the fiber optic coil, and then the delayed signal is converted back to RF frequency domain by this link receiver module R Emcore (direct detection). In our lab experimental setup of DL the receiver output RF signal comes through splitter Sp in two ways. The one way—back to link input (to switch SW1), so forming loop for recirculating functioning, the second way—for RF signal registration on oscilloscope OCS. At the output of recirculating DL there is the switch for snatch (pull out) delayed RF signal. In our lab setup it is switch SW2 installed after the splitter. The snatched RF signal we can exam by spectrum analyzer SA. The switches SW1 and SW2 work under control from time control (TC) unit.

The switching time of an RF switches may be in nanosecond range: 10 ns as in [21]. TC unit also gives synchro pulse to 'trigger' input of OSC. Some details concerning our exam of the Optiva OTS-2 18 GHz link: link gain, 1 dB compression point, output noise level, own intrinsic time delay in the 'transmitter + receiver' modules pair one can find in [22].

The time delays Td implemented in our DL depend on the fiber segment length L as well as the number (N) of recirculation assigned to the DL from TC unit: Td = T1·(N + 1). The minimum delay time (when N = 0) is T1 = (L·n)/c, where n is the effective refractive index of fiber mode (single mode fiber SMF-28 type), c is the light velocity in free space. In the DL lab setup we use 100 m fiber segment (fiber spoon), so T1 = 552 ns. The length L of fiber segment determines time discrete (resolution = step) when one switches total time delay in DL. This time discrete depends on the DL application (Radar, Radio intelligence, Radio countermeasures means).
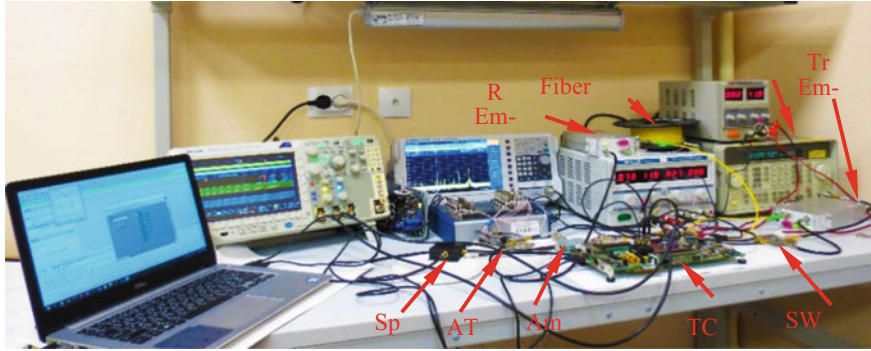
Fig. 53.2 General view of the experimental setup

We can measure time delay in analogue fiber-optic links with high precision using vector network analyzer (VNA) [23]. A VNA measures the S-parameters (S11, S12, S21, and S22) of microwave two-port networks. To measure network time delay the parameter S21 is important: |S21(f)| is the amplitude response of a two-port device, and arg[S21(f)] is its phase response, $\varphi(f)$. From the measured phase response $\varphi(f)$ in an analogue FO link specified frequency band Δf, the group delay time Tdel is determined: Tdel $= \Delta\varphi(f)/(\Delta f \cdot 360)$, where $\Delta\varphi(f)$ − phase change, in degrees.

For phase unwrapping in $\Delta\varphi(f)$ a VNA has to measure S21(f) with large amounts of points in measurement trace. We use the VNA S5085 (Planar) with number of points in S(f) traces up to 200,001 [24].

We comment also that the VNA S5085 software allows selecting superimposed RF signals (case of multipath RF signal propagation through a two-port networks under test), and these possibilities was used in our research of delay line based on LiNbO$_3$ crystal [25].

The general view (a photo) of the lab experimental setup is shown in Fig. 53.2.

53.3   Experimental Results

In experiments we changed some of the DL parameters: time duration Tsw1 and Tsw2 when RF switches SW1 and SW2 close the recirculating loop and their mutual time shift Tsh also, the number of recirculation N, the ATT setting, the input RF signal level Pin and some other parameters. In presented results simple one tone (sinusoidal) signal was used as the DL RF input from generator 8684C (Agilent). The ATT used in our DL lab setup was ZX-76-31R5-P-S + type (MiniCurcuits), it allows controlling one pass gain G$_{RL}$ in the recirculating loop (open loop gain) with small step 0.5 dB. When one uses the DL with large recirculation number N than the open loop gain G$_{RL}$ is adjusted to be close to 1 across the DL operating RF bandwidth. If the loop gain G$_{RL}$ > 1 than a recirculating RF signal will increase

gradually in amplitude (as $(G_{RL})^N$) until its saturation. It is necessary to note: when number N of recirculation is large even in case $G_{RL} < 1$ the DL saturation will be reached by fiber-optic link own noise because a fiber-optic link noise factor is large [1]. We adjusted open loop gain G_{RL} in selected bandwidth to different values: $G_{RL} \approx 1$, $G_{RL} < 1$, $G_{RL} > 1$. The DL functioning is monitored on an OSC and a SA (see Fig. 53.2).

The DL works in repetitive mode under control from the TC unit. In our experimental setup the TC unit was realized on the Xilinx KC705 evaluation board with the XC7K325T-2FFG900C FPGA of the Kintex 7 family located on it [26]. On the base of the FPGA some counters with user-defined operating modes are implemented. Modes are set from PC via the UART interface. The TC unit generates synchro pulse (like meander) with period Trep and two pulses Tsw1 and Tsw2 to control the RF switches SW1 and SW2. Durations of Tsw1 and Tsw2 and their mutual position into period Trep are set independently. We set Trep = T1·M, M—integer number, so number of recirculation N = M − 1 and RF pulse time delay in the DL under test will be Td = (M − 1)·T1.

We test the DL with a Tsw1-wide pulse at a different RF input signal carrier frequencies. A Tsw1 is duration of control pulse when the SW1 is connected to the RF signal source (generator 8684C in our lab setup). A Tsw1 was slightly less than T1: Tsw1 = 548 ns, T1 = 552 ns. So we can see signal zero level for a short time at the end of each signal pass in the loop. In Fig. 53.3 one can see the repetitive RF pulse train in the DL—as the bottom trace; the upper trace is the TC unit synchro pulse with period Trep = T1·10, so N = 9—number of recirculation. The 9-th RF pulse was snatched from the DL by the SW2, Tsw2 = 556 ns, so in Fig. 53.3 one sees zero level instead of 9-th pulse. In this experiment $G_{RL} < 1$: −2 dB.
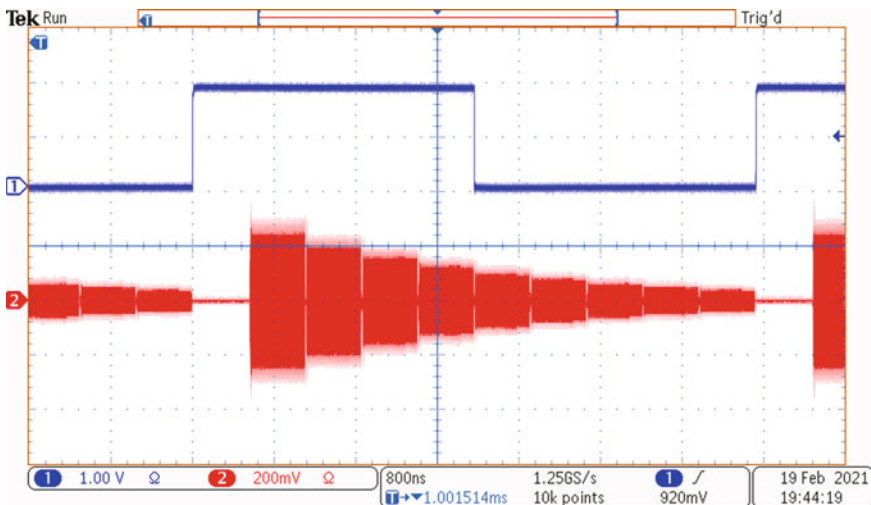
Fig. 53.3 Receiver output RF signal in the recirculating DL. N = 9 recirculation, loop gain G_{RL} = −2 dB

Figure 53.4 shows the similar experiment, but $G_{RL} \approx 1$: 0 dB, and Fig. 53.5 corresponds to the case when $G_{RL} > 1$: + 3.5 dB, M = 10 again, so the delay time for the snatched 9-th RF pulse is (Trep–T1) = 5468 ns.

Figures 53.6, 53.7 and 53.8 show the DL repetitive work with different period Trep = T1·M and $G_{RL} < 1$. One can see the snatched RF pulses for cases M = 5,
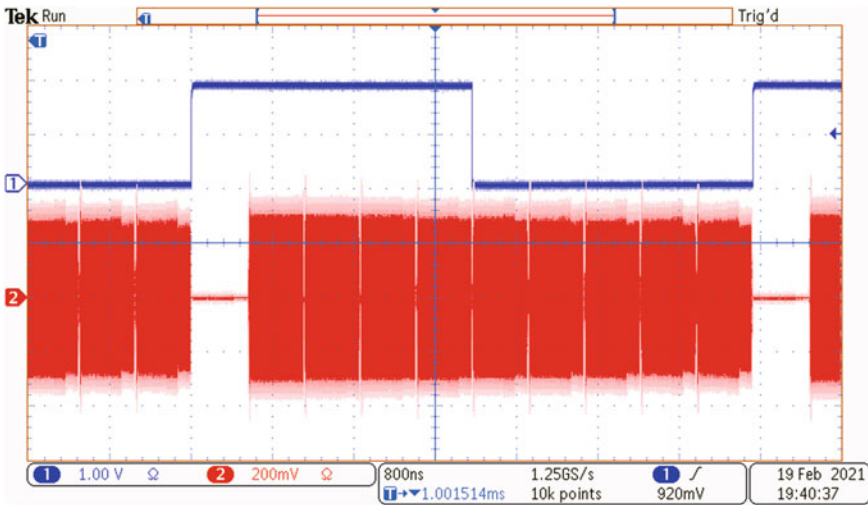
Fig. 53.4 Receiver output RF signal in the recirculating DL N = 9 recirculation, loop gain $G_{RL} \approx 0$ dB
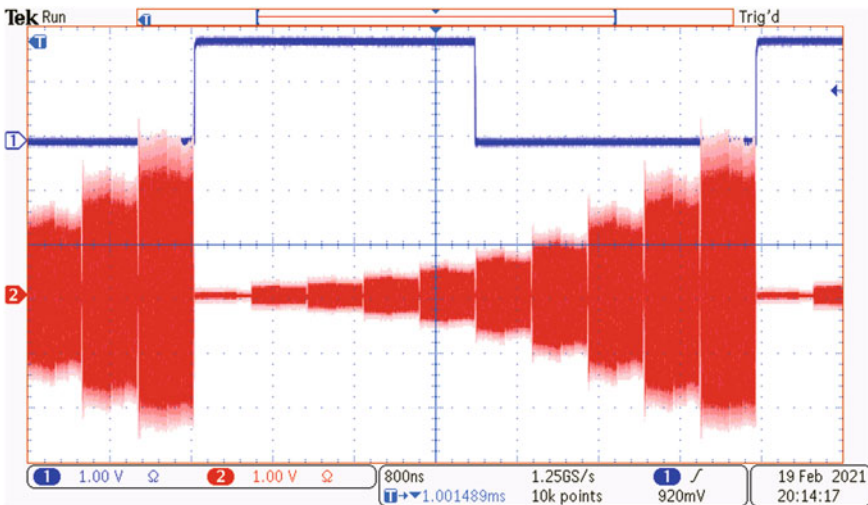
Fig. 53.5 Receiver output RF signal in the recirculating DL N = 9 recirculation, loop gain $G_{RL} = 3.5$ dB
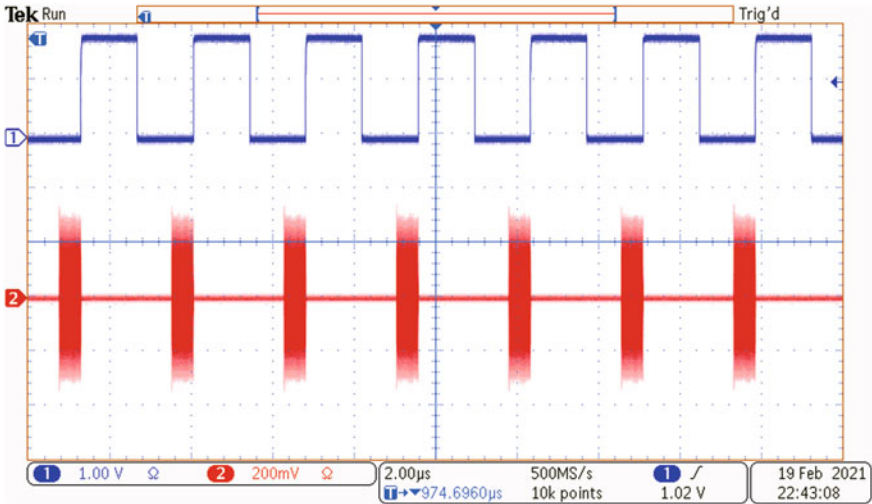
Fig. 53.6   Delay line output RF signal. Time delay is 2208 ns, N = 4 recirculation
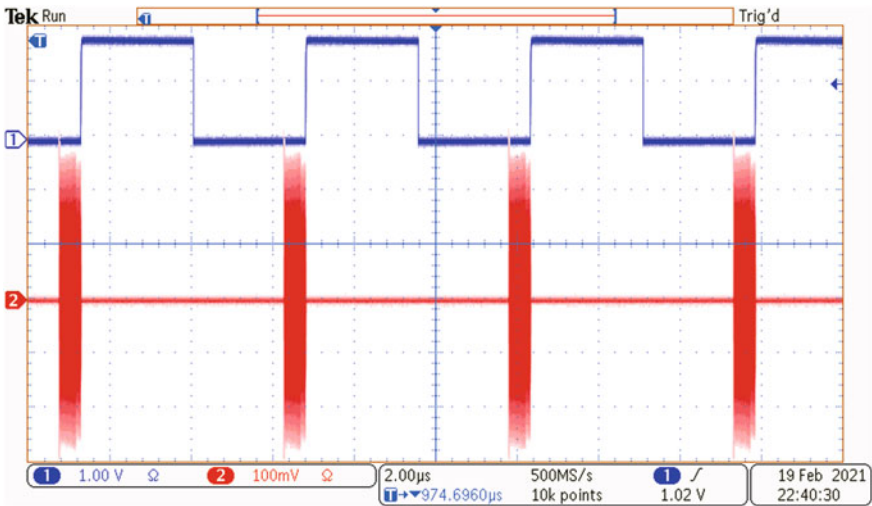
Fig. 53.7   Delay line output RF signal. Time delay is 5468 ns, N = 9 recirculation

10 and 20 accordingly, so time delays Td were 2.208, 5.468 and 10.488 mcs. In all these cases a RF pulses have duration 548 ns.
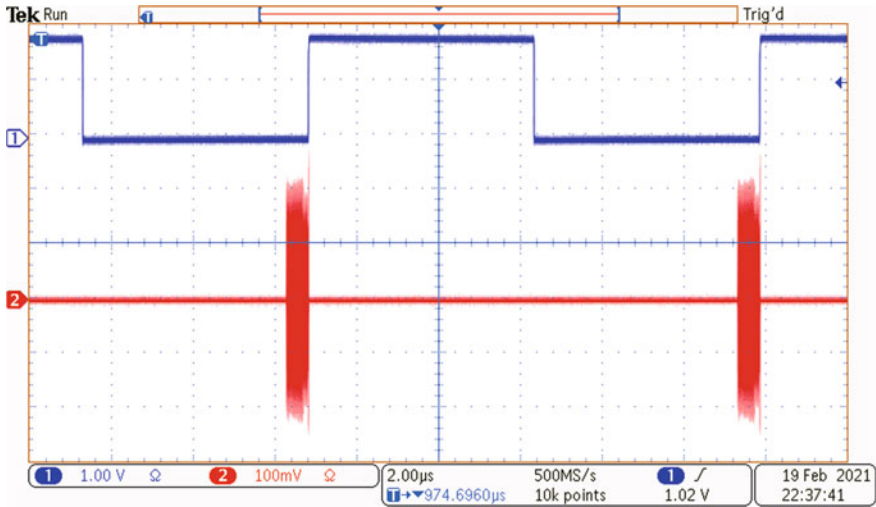
Fig. 53.8 Delay line output RF signal. Time delay is 10488 ns, N = 19 recirculation

53.4 Discussion and Conclusion

Experimental studies of the developed device with switchable time delay of RF signals have shown the promise of using it in systems where a wide range of insertion time delay and wide bandwidth for delayed RF signals are required. The changes of insertion time delay are realized by using a recirculating loop. The technical implementation of this delay line based on an analogue fiber-optic link with an external electro-optic Mach–Zehnder modulator and software-reconfigurable recirculating loop allowed in our experimental studies to change the number of passes through recirculating loop in the range from 1 to 30. Maximum number of the RF signal delay cycles is determined by the delay line operating bandwidth, the unevenness of the total RF amplitude-frequency response, discrete in settings times Tsw1, Tsw2, their mutual position and also period Trep under control the RF switches from the TC unit—approx. 2 ns (when using the Xilinx KC705 evaluation board), the transient time of the RF switches itself (for HMC347ALP3E no more than 2 ns [21]) and the gain control accuracy in recirculating loop (for the software-controlled attenuator ZX-76-31R5-P −0.5 dB). Our preliminary analysis shows that the maximum number of RF signal delay cycles can reach several hundred. The maximum delay time is determined by the length of the fiber-optic link and can be (depending on the used operating bandwidth) several tens of microseconds. The maximum operating bandwidth of this type delay line is currently determined by the operating bandwidth of the RF switches used and in our experiment it was 14 GHz [21]. The minimum delay time of RF signal is determined by the minimum length of the analogue fiber-optic link, as well as the latency of the RF switches time control unit and the transient time of the RF switches themselves, and amounts to several tens of nanoseconds for the used

software-controlled Xilinx evaluation board. One additional question concerning of fiber-optic line functioning is possible nonlinear effect in fiber. Typical transmitter optical power in analogue fiber-optic line is approx. 10 mW, so fiber nonlinear effect is negligible quantity. Another question is the fiber dispersion D(λ) influence [27]. We exploited just the fiber dispersion D(λ) in developing photonic beamformer for linear phased array antenna [9]. In this investigation of fiber DL we have small frequency bandwidth in optical domain of only 2Fmax, Fmax is maximal frequency of input RF signal (20…40 GHz), so we consider fiber dispersion D(λ) influence also negligible. In the future, it is planned to carry out a theoretical and experimental analysis of a noise-like RF signal transformation in the investigated switchable time delay line. We believe that a delay line of this type could be of interest not only in radar and countermeasures, but also in other applications, for example in meteorological radars and lidar systems [28] to imitate return pulses in them.

References

1. V.J. Urick, J.D. Mckinney, K.J. Williams, *Fundamentals of Microwave Photonics* (Wiley, Hoboken, New Jersey, 2015)
2. J. Yao, Microwave photonics. IEEE J. Lightwave Technol. **27**(3), 314–335 (2009)
3. I.L. Newberg, C.M. Gee, G.D. Thurmond, H.W. Yen, Long microwave delay fiber-optic link for radar testing. IEEE Trans. MTT **38**(5), 664–666 (1990)
4. R.A. Minasian, E.H.W. Chan, X. Yi, Microwave photonic signal processing. Opt. Exp. **21**(19), 22918–22936 (2013)
5. J.F. Diehl, J.M. Singley, C.E. Sunderman, V.J. Urick, Microwave photonic delay line signal processing. App. Opt. **54**(31), F35–F41 (2015)
6. J. Capmany, J. Mora, I. Gasulla, J. Sancho, J. Lloret, S. Sales, Microwave photonic signal processing. J. Lightwave Technol. **31**, 571–586 (2013)
7. B. Shillue, W. Grammer, C. Jacques, R. Brito, J. Medows, J. Castro, Y. Masui, R. Treacy, J.F. Cliche, The ALMA photonic local oscillator system. Proc. SPIE **8452**, 845216 (2012)
8. T.N. Bakhvalova, M.E. Belkin, I.V. Gladyshev, S.A. Kudzh, A.S. Sigov, Photonics technology as a way for upgrading key technical features of radio delay devices. Russ. Technol. J. (In Russ.) **5**(3), 4–21 (2017)
9. V.A. Volkov, D.A. Gordeev, S.I. Ivanov, A.P. Lavrov, I.I. Saenko, Photonic beamformer model based on analog fiber-optic links' components. J. Phys. Conf. Ser. **737**, 012002 (2016). https://doi.org/10.1088/1742-6596/737/1/012002
10. A. Petrov, E. Velichko, V. Lebedev, I. Ilichev, P. Agruzov, M. Parfenov, A. Varlamov, A. Shamrai, Broad-band fiber optic link with a stand-alone remote external modulator for antenna remoting and 5G wireless network application, in *NEW2AN 2019, ruSMART 2019, LNCS*, vol. 11660, ed. by O. Galinina, S. Andreev, S. Balandin, Y. Koucheryavy (Springer, Cham, 2019), pp. 727–733
11. V.V. Davydov, N.V. Sharova, E.V. Fedorova, E.P. Gilshteyn, V.A. Vologdin, A.Y. Karseev, K.Y. Malanin, I.V. Fedotov, Fiber-optics system for the radar station work control, in *NEW2AN 2015, ruSMART 2015, LNCS*, vol. 9247, ed. by O. Galinina, S. Andreev, S. Balandin, Y. Koucheryavy (Springer, Cham, 2015), pp. 712–721
12. Y. Shahoei, J. Yao, Delay lines, in *Wiley Encyclopedia of Electrical and Electronics Engineering*, ed. by J. Webster. (Wiley, Hoboken, 2014)

13. A.A. Markvart, L.B. Liokumovich, N.A. Ushakov, Tunable optical delay lines based on a system of coupled whispering gallery mode resonators. J. Phys. Conf. Ser. **1326**, 012017 (2019)
14. I. Koffman, P. Herczfeld, A. Daryoush, B. Even-Or, R. Markowitz, A fiber optic recirculating memory loop operating at microwave frequencies, in *1988 IEEE Military Communications Conference.* https://doi.org/10.1109/MILCOM.1988.13363 (1988)
15. T.A. Nguyen, E.H.W. Chan, R.A. Minasian, Photonic multiple frequency measurement using a frequency shifting recirculating delay line structure. J. Lightwave Technol. **32**(20), 3831–3838 (2014)
16. B. Vizoso, C. Vfizquez, R. Civera, M. Lopez-Amo, Amplified fiber-optic recirculating delay lines. J. Lightwave Technol. **12**(2), 294–305 (1994)
17. B. Even-Or, S. Lipsky, R. Markowitz, Recirculating fiber-optical RF memory loop in countermeasure systems. Proc. SPIE **1371**, 161–169 (1990)
18. A.S. Podstrigaev, A.S. Lukiyanov, A.A. Galichina, A.P. Lavrov, M.V. Parfenov, Wideband tunable delay line for microwave signals based on RF photonic components, in *NEW2AN 2020, ruSMART 2020, LNCS*, vol. 12525, ed. by O. Galinina, S. Andreev, S. Balandin, Y. Koucheryavy. (Springer, Cham, 2020), pp. 424–431. https://doi.org/10.1007/978-3-030-657 260_38
19. Optiva OTS-2 18 GHz Unamplified Microwave Band Fiber Optic Links, https://emc ore.com/wp-content/uploads/2016/03/Optiva-OTS-2-18GHz-Unamplified.pdf. Last accessed 2021/03/13
20. S.I. Ivanov, A.P. Lavrov, I.I. Saenko, Main characteristics study of analog fiber-optic links with direct and external modulation in transmitter modules, in *2018 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)* (2018), pp. 264–267. https://doi.org/10.1109/EExPolytech.2018.8564391
21. Switch HMC347ALP3E datasheet, https://www.analog.com/en/products/hmc347alp3e.html. Last accessed 2021/04/01
22. S.I. Ivanov, A.P. Lavrov, I.I. Saenko, Investigation of key components of photonic beamforming system for receiving antenna array, in *NEW2AN 2015, ruSMART 2015, LNCS*, vol. 9247, ed. by S. Balandin, S. Andreev, Y. Koucheryavy. (Springer, Cham, 2015), p. 679688. https://doi.org/10.1007/978-3-319-23126-6_61
23. S.I. Ivanov, A.P. Lavrov, I.I. Saenko, Measurements and stabilization of the radio signals time delay when their transmitting over long wideband analog fiber optics links, in *2019 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)* (2019), pp. 50–53. https://doi.org/10.1109/EExPolytech.2019.8906807
24. S5085 2-Port 8.5 GHz Analyzer. Extended Specification Sheets, https://coppermountaintech.com, https://online.fliphtml5.com/pbaab/jjxa/#p=1. Last accessed 2021/03/13
25. A.V. Varlamov, S.I. Ivanov, A.P. Lavrov, A.V. Shamray, Time-frequency analysis of acoustic waves in an integrated acousto-optical modulator based on $LiNb_{O3}$ crystal. J. Phys. Conf. Ser. **1697**, 012174 (2020). https://doi.org/10.1088/1742-6596/1697/1/01217
26. Xilinx Kintex-7 FPGA KC705 Evaluation Kit, https://www.xilinx.com/products/boards-and-kits/ek-k7-kc705-g.html. Last accessed 2021/04/01
27. U. Gliese, S. Norskov, T.N. Nielsen, Chromatic dispersion in fiber-optic microwave and millimeter-wave links. IEEE Trans. Microwave Theor. Tech. **44**(10), 1716–1724 (1996). https://doi.org/10.1109/22.538964
28. V.E. Privalov, V.G. Shemanin, Lidar system for monitoring radioactive contamination of atmospheric air. J. Opt. Technol. **84**(5), 289–293 (2017). https://doi.org/10.1364/JOT.84.000289

Part IV
Telecommunications and Navigation Systems

Chapter 54
Research of an X-band Oscillator Based on a SiGe Amplifier with Dielectric Resonator

Check for updates

Egor V. Egorov , **Sergey B. Makarov** , **and Victor M. Malyshev**

Abstract This paper presents the model of 9.93 GHz low-phase noise reference oscillator and its experimental sample with electrical and mechanical tuning. The oscillator is built on a two-stage transistor amplifier with external feedback. In the feedback, the resonant system with electrical and mechanical frequency tuning is based on DR with unloaded Q factor of 20,000. The loaded Q factor of such a system exceeds 7,300 over the entire electrical tuning range. The phase noise level of oscillator prototype is less than -120 dBc/Hz at 10 kHz offset from the 9.93 GHz carrier with output power over 0 dBm. A comparison of calculations and measurements shows that the noise model of the transistor is quite suitable for calculations at the analysis of frequencies F above 10 kHz. At frequencies F below 1 kHz, there is a significant discrepancy between the phase noise of the oscillator obtained experimentally and from the AWR design based on the transistor spice model. Possible ways to overcome this discrepancy are presented.

Keywords Reference oscillator · Microwave · Low noise amplifier · Phase noise · Oscillatory system · Frequency tuning · Dielectric resonator · Q factor

54.1 Introduction

Modern reference oscillators (RO) are widely used in many radio electronic systems for frequency synthesis and signals [1–5]. The frequency stability is a vital parameter in conditions of rapid technologies development [6, 7]. The phase noise (PN) is a fundamental qualitative indicator of any reference oscillator. Phase noise characteristics are important for digital, as well as analog, communication. Nowadays, you have to use complex processing to extract useful information from a noisy and distorted signal. For example, in radar systems, PN degrades the ability to process Doppler information in radar. And, in digitally modulated communication systems, phase noise degrades error vector magnitude (EVM) [8]. In the SHF range, low-noise

E. V. Egorov (✉) · S. B. Makarov · V. M. Malyshev
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: egorkin.e@list.ru

SiGe amplifiers and high-Q resonant systems based on dielectric resonator (DR) are currently used to achieve low PN level of oscillator. With a low production cost of such RO, it is possible to achieve a PN level less than -120 dBc/Hz at 10 kHz offset.

Currently, CAD is widely used in the development of various types of microwave devices and, in particular, oscillators [9–11]. They allow you to speed up the development process. This mainly refers to the calculation of energy and frequency characteristics. To a lesser extent, this applies to the noise characteristics of the RO. This is due to the lack of a developed noise model of individual active elements [12]. The paper presents the results of modeling and measuring the phase noise of reference oscillator with an oscillatory system (OS) based on DR with varactor frequency tuning (DRO).

54.2 Oscillator Description

The reference oscillator consists of a low-noise two-stage amplifier with external feedback. As a feedback, a modified design of a resonant system with a DR located in a metal cavity is used [13]. The DR is a cylindrical disk resonator with an absolute dielectric constant of $\varepsilon = 30$. DR dimensions are 6 mm in diameter and 2.7 mm in height. The unloaded Q factor (Q_{un}) of the resonator is 20,000. The oscillating system has a resonant frequency f_0 equal to 9.93 GHz, the transmission loss L at f_0 are equal to 8.5 dB, and the loaded quality factor Q_L is not less than 7300 in the entire range of electrical frequency tuning. Electrical tuning of 1.1 MHz is carried out using MACOM MA46H120 varactors [14]. Mechanical tuning of such oscillator is more than 25 MHz and it is obtained by the screw.

The common emitter low noise amplifier was designed using SiGe BFP843 transistor [15]. The parameters of the LNA: gain G at the operating frequency f_0 near 10 GHz is 10 dB, the noise figure NF = 2.6 dB, output power for 1 dB compression P1dB = 1.7 dBm, Output Third Order Intercept IP3 is 9.5 dBm. Oscillator structure is shown in Fig. 54.1. The reflective phase shifter was used as the phase shifter φ, and the Minicircuits ZX10-2-126-S splitter was used as the directional coupler [16].

The S-parameters of DRO's individual linear parts were measured and simulated in the AWR environment before design. It turned out that the losses in the connecting cables and the phase shifter do not allow achieving self-excitation conditions when using the LNA on a single transistor. Therefore, the second stage was added to the amplifier, connected to the first stage through the 2.5 dB attenuator. As a result, the LNA gain reaches 17.5 dB. This made it possible to provide conditions for the self-excitation of DRO. Moreover, self-excitation of the oscillator was observed only in a narrow range of angles φ. By choosing the angle φ, it is possible to achieve a minimum PN level. For all varactor's bias voltages U_V (0-15 V) the PN level S_φ is less than -120 dBc/Hz at 10 kHz frequency offset from the carrier 9.93 GHz. Figure 54.2 shows the spectral power density of oscillator's phase noise for $U_V = 12$ V, measured with the Rohde & Schwarz FSWP26 phase noise analyzer.

Fig. 54.1 Oscillator structure
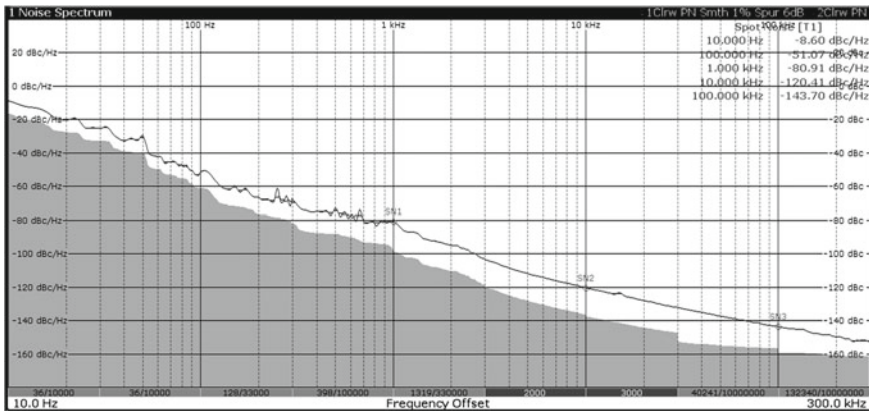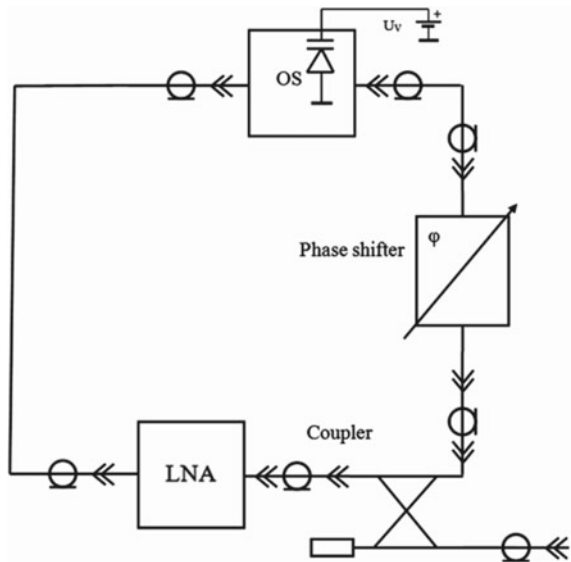
Fig. 54.2 Phase noise spectrum

54.3   Simulation

Before creating the DRO experimental sample, oscillator was simulated in the AWR Microwave Office. Experimentally measured characteristics were used as parameters of individual parts of the model (Fig. 54.1). Both the simulated DRO and the amplifier has frequency and energy parameters close to those of the experimental models.

The spice model of the BFP843 transistor offered by the manufacturer, in which a collector current source of flicker noise is introduced [17], was used in LNA design [15]. The spectral density of flicker noise S (F) is described by the formula (54.1),
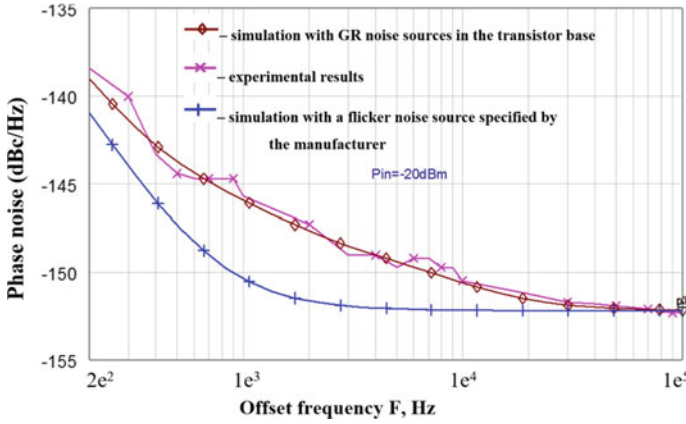
Fig. 54.3 PN spectral density S_φ (F) of the amplifier

where I is the current through the device, AF = 1.529, КF = 3.12×10^{-16}.

$$S\ (F) = \text{KF} \cdot I^{\text{AF}}/F \tag{54.1}$$

The parameters of the flicker source in the spice model can be adjusted. Figure 54.3 shows the results of measuring the PN spectral density S_φ (F) of the experimental amplifier and AWR model for input power level $P_{\text{in}} = -20$ dBm. It can be seen that the coincidence of the experimental and simulated results is observed at frequencies offset more than 10 kHz (white noise of the spectrum presumably caused by thermal noise sources). At frequencies less than 10 kHz, there is a significant discrepancy. The phase noise level of the experimental LNA significantly exceeds the PN of the AWR model. The PN of the amplifier model can be approximated to the experimental results by increasing the KF to 6.24×10^{-11}. It is clear that such a difference in the KF estimates, determined from the PN of the experimental amplifier and determined by the transistor manufacturer based on modern methods of extracting noise parameters [18], indicates that at frequencies F less than 10 kHz the effect of another noise source appears in the PN of LNA sample. It is known that such a noise source can be a base current noise source caused by low-frequency fluctuations of the base current [17]. To create these fluctuations without changing the spice model of the transistor, a low-frequency noise source (I_Burst) was located at the input of the transistor base. I_Burst parameters were determined based on the coincidence of experimental PN level of the amplifier and its model. As can be seen from Fig. 54.3, the frequency dependence S_φ (F) of the experimental amplifier is close to the contribution of three generation-recombination noise (g-r noise) sources. The spectral density of this current noise is described by the formula (54.2), where FB is the cutoff frequency [17, 19].

$$S\ (F) = \text{KB} \cdot I^{\text{AB}}/\left(1 + (F/\text{FB})^2\right) \tag{54.2}$$

FB3=1.2257598937965e-7 KB3=7.91863625245272e-5   KB1=5.15802339651711e-6   FB1=8.10492387962611e-6   KB2=1.26314393355579e-5   FB2=1.28824236084596e-6

L_BURST
ID=NS2
KB=KB3
AB=2
FB=FB3 GHz

L_BURST
ID=NS1
KB=KB1
AB=2
FB=FB1 GHz

L_BURST
ID=NS3
KB=KB2
AB=2
FB=FB2 GHz

Rist=0.01

PORT
P=1
Z=50 Ohm

RES
ID=R1
R=Rist Ohm

RES
ID=R2
R=Rist Ohm

RES
ID=R3
R=Rist Ohm

PORT
P=2
Z=50 Ohm
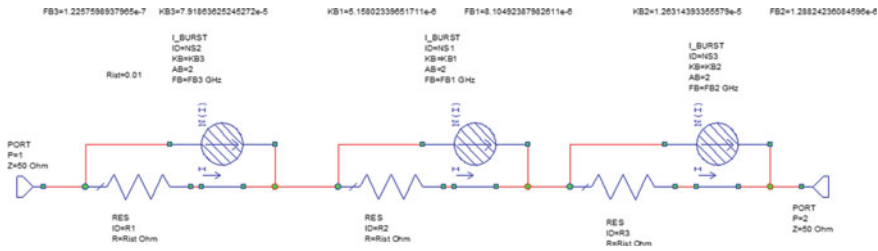
Fig. 54.4 Low-frequency noise *Source* I_Burst

Figure 54.4 shows the AWR base noise model. As can be seen from Fig. 54.3, the g-r noise parameters, found using parametric optimization (KB1 $= 5.2 \times 10^{-6}$; FB1 $= 8.1 \times 10^{-6}$; KB2 $= 1.3 \times 10^{-5}$; FB2 $= 1.3 \times 10^{-6}$; KB3 $= 7.9 \times 10^{-5}$; FB3 $= 1.2 \times 10^{-7}$; AB $= 2$), allow us to get PN level quite close to the experimental dependence.

Comparison of PN level of DRO in three different cases is shown in Fig. 54.5.

As can be seen from the graph, the most accurate coincidence of the spectrum is also observed at offset frequencies above10 kHz (the main contribution to the noise comes from white noise of the spectrum source). At lower frequencies offset, the DRO PN level, using an amplifier with a single flicker collector noise, has a much lower value with respect to the experimental one. A more accurate match there is in the DRO model, which uses I_Burst noise source in the base circuit of the transistor's amplifier. Since the low-frequency phase noise may be caused by fluctuations in the barrier capacitances of the transistor [12], the contribution of which depends on the level of the variable signal, this can be attributed to the difference in the experimental PN level and the DRO model with the noise source in the base circuit. However,
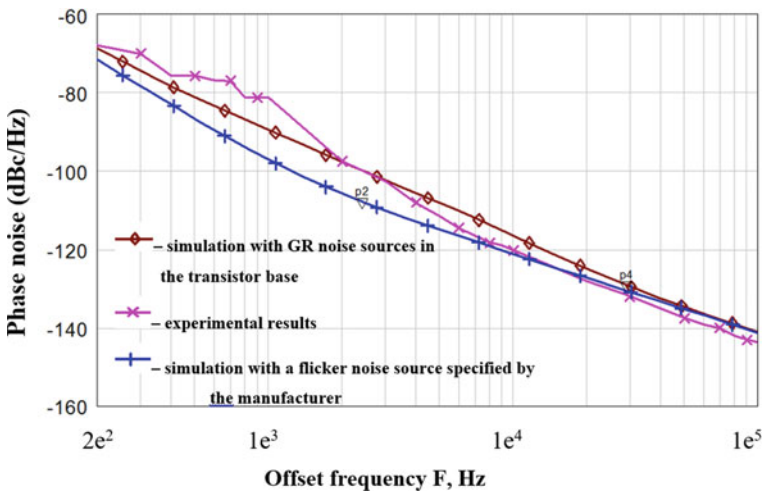
Fig. 54.5 PN spectral density S_φ (F) of reference oscillators

the use of an additional noise source allows us to give a more accurate phase noise estimate of the DRO over the entire frequency offset. In the presence of the measured phase noise of the amplifier, the PN level of the oscillator at low frequencies offset can be estimated more accurately at the design stage.

54.4 Conclusion

The paper presents the results of a comparison an experimental oscillator, a LNA and their models designed in the AWR MWO to obtain phase noise level estimates. The results show the coincidence of the PN for the white noise part of the spectrum. For flicker part of the spectrum, the measured phase noise of the amplifier should be used to estimate the phase noise level of the reference oscillator. The 9.93 GHz dielectric resonator oscillator with electrical tuning more than 1 MHz and phase noise level less than -120 dBc/Hz at 10 kHz offset was developed.

References

1. A. Chenakin, Frequency synthesis: Current status and future projections. Microwave J, **60**(4), 22–36 (2017)
2. A.B. Nikitin, E.I. Khabitueva, A 6–12 GHz wideband hybrid vco. In 2018 IEEE International Conference on Electrical Engineering and Photonics, EExPolytech 2018, (St. Petersburg, Russia, 2018), pp. 37–39
3. A.B. Nikitin, E.I. Khabitueva, Microwave Ultra-Wideband VCO Design. In International Conference on Actual Problems of Electron Devices Engineering APEDE 2018, (Saratov, Russia, 2018), pp. 108–112
4. E.I. Khabitueva, A.B. Nikitin, The PLL synthesizer based on 6–12 GHz wideband hybrid VCO. J. Phys. Conf. Series **1326** (2019)
5. D.B. Akhmetov, A.S. Korotkov, I.A. Rumyancev, 2.4–2.5 GHz fractional-n frequency synthesizer with integrated VCO in 0.18 um CMOS for RFID systems. In: 2018 IEEE International Conference on Electrical Engineering and Photonics, EExPolytech 2018, (St. Petersburg, Russia, 2018), pp. 64–68
6. A.A. Petrov, V.E. Shabanov, D.V. Zalyotov, V.V Davydov, A.L. Bulyanitsa, D.V. Shapovalov, Modernization of the frequency synthesizer of cesium atomic clock. In 2018 IEEE International Conference on Electrical Engineering and Photonics, EExPolytech 2018, (St. Petersburg, Russia, 2018), pp. 52–55
7. D.B. Akhmetov, A.S. Korotkov, The reference spur reduction technique for frequency synthesizers. In 2018 IEEE International Conference on Actual Problems of Electron Devices Engineering, ElConRus 2018, (Russia, 2018), pp. 164–166
8. How to Optimize Local Oscillator Phase Noise for EVM Measurements, https://blogs.keysight.com/blogs/tech/rfmw.entry.html/2020/05/08/optimize_phase_noise-fYkg.html. Last Accessed 14 Mar 2021

9. E. Khabitueva, A. Nikitin, D. Okulov, Comparison of various em simulators in the design of a wideband microwave VCO. In 2020 IEEE International Conference on Electrical Engineering and Photonics, EExPolytech 2020, (St. Petersburg, Russia, 2020), pp.26–29
10. V. Karpova, N.A. Ivanov, Symmetrical Design of a Microstrip Tunable Bandpass Filter. Part of the Springer Proceedings in Physics book series **255**, 159–167 (2021)
11. E.V. Egorov, A.V. Ivanova, S.B. Makarov, V.M. Malyshev, Ultra-Low-Noise Reference Oscillator Based on a Dielectric Resonator with Mechanical and Electrical Frequency Tuning. Part of the Springer Proceedings in Physics book series **255**, 621–628 (2021)
12. A.V. Yakimov, A.V. Klyuev, M.A. Krevskii, The Nature of Introduced Phase 1/f Noise in Microwave Oscillators. J. Commun. Technol. Electron. **65**(1), 84–89 (2020)
13. E.V. Egorov, V.M. Malyshev, Oscillating system of a reference microwave generator with screened dielectric resonator excited from an end face. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems, vol. 10, No. 2, pp. 45–57 (2017)
14. MA46H120 Series. GaAs Constant Gamma Flip-Chip Varactor Diode, http://www.macom.com. Last Accessed 09 Apr 2019
15. Gummel-Punn model BFP843, Infineon spice model library for RF transistors, https://www.infineon.com/cms/en/product/rf-wireless-control/rf-transistor/ultra-low-noise-sigec-transistorsfor-use-up-to-12-ghz/bfp843/#!simulation. Last Accessed 29 Oct 2020
16. Mini-Circuits ZX10–2–126-S, Datasheet, https://www.minicircuits.com/WebStore/dashboard.html?model=ZX10-2-126-S%2B. Last Accessed 14 Mar 2021
17. C. Mukherjee, T. Jacquet, A. Chakravorty, T. Zimmer, J. Böck, K. Aufinger, C. Maneu, Low-frequency noise in advanced SiGe: C HBTs—Part I: Analysis. IEEE Trans. Electron Devices **63**(9), 3649–3655 (2016)
18. M. Rudolph, C. Fager, D.E. Root (eds.), *Nonlinear Transistor Model Parameter Extraction Techniques* (Cambridge University Press, Cambridge, 2011)
19. AWR Microwave Office Element Catalog 2020, Cadence Design Systems, https://awrcorp.com/download/faq/english/docs/Elements/Elements.htm. Last Accessed 03 Dec 2021

Chapter 55
Geometry and Radiating Patterns of Segmented Long-Ring Travelling Wave Resonator Antennas

Dmitry Dicky

Abstract The results of the development and study of long-ring ($L \gg \lambda$, L is ring length) low-profile antennas based on the traveling wave resonator (TWR) idea are presented. Depending on the configuration of the radiating ring, different radiating patterns can be obtained. It opens comprehensive capability to use these antennas in many applications, especially in satellite and mobile telecommunication as a vehicle antenna. Dual orthogonal polarized radiation mode and multi-frequency capability remain almost in all configurations of the segmented long-ring TWR antennas. The theoretical analysis, design, and calculations are performed with special computer programs using the matrix methods of circuit theory and vector potential calculations.

Keywords Travelling Wave Resonator Antenna · Antenna Radiation Pattern · Antenna Polarization

55.1 The General Idea of a Non-segmented TWR Antenna

The TWR antenna (Fig. 55.1) includes the radiating ring, situated at the height *h* above the ground plane, and the strip-line directional coupler (DC), one port of which is an exciting one. The general idea and some fundamental relations between the incident and reflected waves are given in [1, 2].

Several works are devoted to the theory of thin loop antennas. As shown in [1], under certain circumstances, when radiating power is equal to the energy pumped into the ring with a directional coupler, the reflections from the supply port is zero. Wu [3] and Whiteside [4] have shown that when the radiating ring is situated over a conductive surface, the ring's current distribution varies a little. This fact allows setting the amplitude of the ring's current to some constant. Lately, numerical results for the ring's current distribution for the TWR antenna with different supply systems are obtained in [5].

D. Dicky (✉)
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Fig. 55.1 The geometry of
strip-line TWR antenna
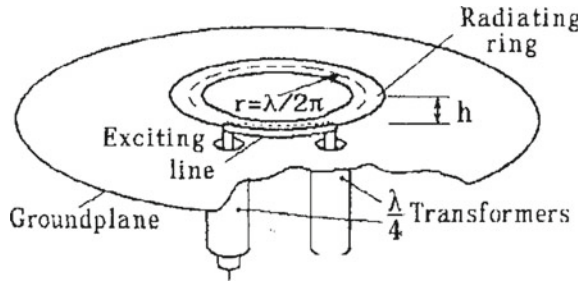
In the case of the circular ring under a resonant wavelength $\lambda_{res} = \frac{L}{n}$, $n = 1, 2, 3, \ldots$ the following equations for the orthogonal components E_θ and E_φ are evaluated in [6, 7]. Practical equations for the TWR antenna pattern for the ring over the conductive surface are obtained in [5, 7]:

$$E_\varphi = \frac{nI_0}{2r}\left[J_{n-1}(n\sin\theta) - J_{n+1}(n\sin\theta)\right]\sin\left(2\pi\frac{h}{\lambda}\cos\theta\right)$$

$$E_\theta = \frac{nI_0}{2r}\left[J_{n-1}(n\sin\theta) + J_{n+1}(n\sin\theta)\right]\sin\left(2\pi\frac{h}{\lambda}\cos\theta\right)\cos\theta$$

where, I_0—resonant current of the ring, h—ring's height, J_n—the cylinder Bessel function. Axis ($\theta = 0$) radiation takes place only when $n = 1$, i.e. $L = \lambda$ (for the circular ring). As calculations and experimental data show, the axis gain of the strip-line TWR antenna is about 9.5 dB.

55.2 The Segmented Long-Ring TWR Antenna

Further increase of directivity can be achieved by increasing the radiating area, i.e. the ring diameter. Although the formulas above show zero radiation in the axis direction for $L = n\lambda$ ($n \neq 1$), it is valid only for a circular ring. Other geometry of the radiating element (not a circular ring) does not allow to write simple expressions for components of the radiating field. Therefore, in the assumption of the constant amplitude of the ring's current, the radiating pattern can be obtained numerically.

Figure 55.2 shows the radiating line's configuration, which provides high directivity for $L = n\lambda$ ($n > 1$). Here all outer segments A, B, C,... radiate the right circular-polarised (CP) wave.

The inner segments a, b, c,... radiate the left CP field. It may be shown that if the full length of the ring is equal to $(m + l)\lambda$, where m-number of segment pairs, co-polarised waves of the outer segments sum up in-phase in the axis direction, whereas cross-polar waves of the inner elements sum up into zero in the axis direction.
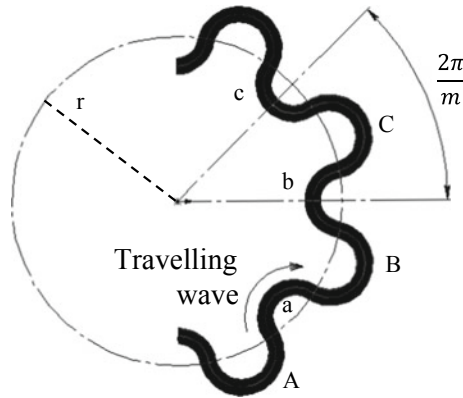
Fig. 55.2 Segmented ring configuration

Similarly, axis radiation of the segments **a, b, c**,… is amplified, and the one of segments **A, B, C**…. is depressed when $n = m - l$. Two examples of the possible configurations of the segmented TWR antenna for $n = m + l$ are shown in Fig. 55.3. The calculated maximum of the axis gain is 16.6 dB for $m = 4$ and 14.5 dB for $m = 3$.

Figure 55.4 shows the radiating patterns under the different arc lengths of useful segments for $m = 4$ (Fig. 55.3a). Calculations were performed with the Vector Potential method's help by integrating the travelling wave current distribution with its mirror image over the ground plane.

In practice, it is possible to depress the radiation of the useless segments (to decrease the sidelobes) even more by placing them on the lower height. Figure 55.5 shows the experimental model of the segmented TWR antenna with $n = 5, \ m = 4$. The strip-line configuration is performed as a patch on a thin, flexible substrate curved by lines C1–C2, as shown in Fig. 55.5.
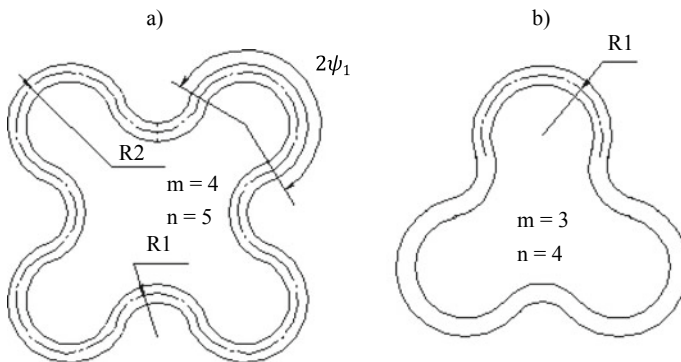
Fig. 55.3 Possible configurations of the segmented rings
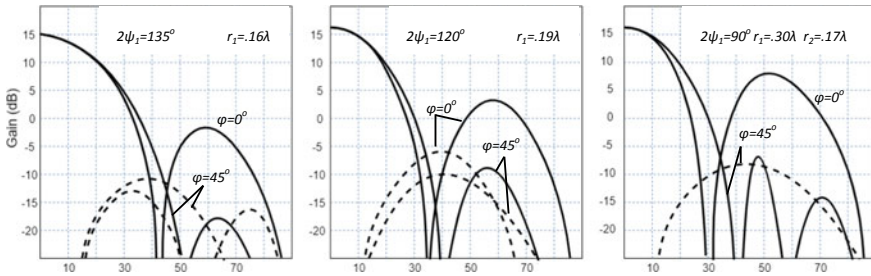
Fig. 55.4 Field patterns of segmented TWR antenna ($m = 4$, $n = 5$) for the principal ($\varphi = 0^o$) and diagonal ($\varphi = 45^o$) planes. Solid line is co-polarisation, dashed line is cross-polarization
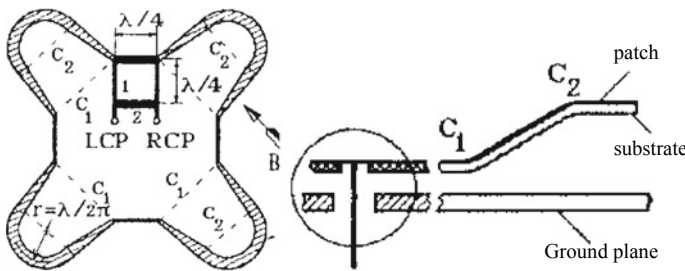
Fig. 55.5 The geometry of the segmented TWR experimental antenna with $n = 5$, $m = 4$

It is possible to obtain the radiation of two orthogonal polarised fields with this antenna, with the help of a branch-line hybrid junction, as shown in Fig. 55.5. If only one circular polarization is required, branches 1, 2 (Fig. 55.5) may be removed, and the exciting system becomes simple.

Besides, the availability of a low-height strip-line structure in the centre provides many possibilities to supply the ring in several points to increase the bandwidth.

Figure 55.6 shows the dual-band segmented antenna ($\lambda = 18$ cm and 30 cm) produced for passive radio-thermal earth mapping. This antenna is installed on the small aircraft wing.

This antenna has a gain of 11.9 dB, and the sidelobe level is -11.5 dB. The bandwidth of each channel is about 5%, with the VSWR less than 2, the axial ratio is better than 0.85. The radiating pattern at $\lambda = 18$ cm in principal ($\varphi = 0$) plane is shown in Fig. 55.7.

Taking into account the zero axis radiation of circular TWR antennas with $n \neq 1$ and non zero radiation of the segmented TWR antenna, some omnidirectional antenna can be approached by the intermediate segmentation of TWR antennas with $n > 1$ (as shown in Fig. 55.3b).

Fig. 55.6 Dual-band
segmented TWR antenna
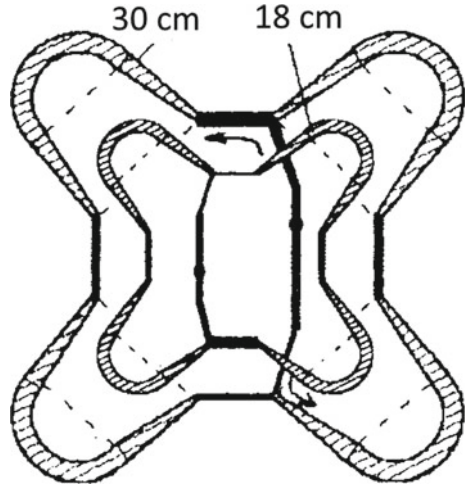
Fig. 55.7 The radiation
pattern of the experimental
segmented long-ring TWR
antenna
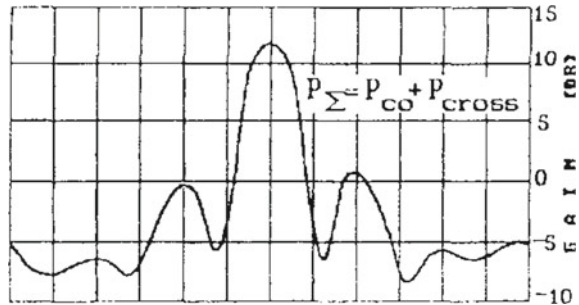
55.3   The Bandwidth of the Long-Ring TWR Antenna

The bandwidth of the long-ring ($n > l$) TWR antenna is less (approximately by n
times) than the bandwidth of the ring TRW antenna. There are two possible ways
to increase the bandwidth: (a) feed the radiating element in several points, as shown
in Fig. 55.6, to decrease the electrical length between the exciting points, (b) multi-
stage construction of TWRs (like was used in [9, 10] for $n = 1$) to implement the
multi-peak frequency response pattern. An equivalent scheme of multi-stage TWR
antenna is shown in Fig. 55.8 (left).

In this construction, only the upper ring (actually segmented, but for simplicity
depicted as circular) is the radiating one, and all other ones are intermediate
resonators. Intermediate resonators do not have to be segmented and can be
constructed as convenient. Rings are linked by the system of non-directional
elements, which in sum provide the directional coupling effect. Figure 55.8 (right)
shows the frequency response (mutual coupling between RCP and LCP inputs) for
the dual-stage TWR.
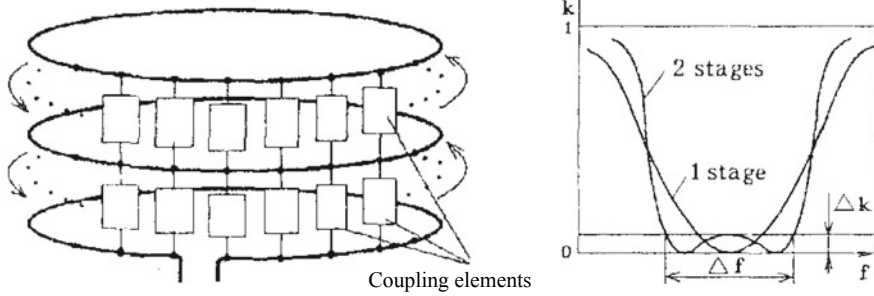
Fig. 55.8 Two-stage TWR antenna and its frequency response

Calculations and experimental data show that the use of both of these methods (a and b) gives the bandwidth of more than 10–12% at the level of mutual coupling $\Delta k = 2 - 5\%$ even for $n = 5$.

55.4 Discussion

There are many designs of the wide-band low-profile antennas [11–13]. These designs show suitable parameters for the particular area of application—broadband communication, near field source localization, short-range radars, etc. The proposed design is suitable for a wide range of application, from radars to passive radio location. Namely, the radiating line configuration allows for varying the radiating pattern, while changing the supply system allows for varying the frequency response. A single segmented TWR antenna has a bandwidth up to 10–12%, while a stacked TWR antenna can cover up to several octaves in frequency.

The mutual coupling between ports defines the noise figure of the segmented TWR antenna. For example, if $\Delta k < -10$ dB in the desired frequency range, the noise figure of the antenna is below 30 K. VSWR is generally below 1.2.

Another advantage of the proposed antenna is that it has a stable field pattern in the desired frequency range while other antennas like in [14–16] have a frequency-dependent field pattern and high VSWR (about 2).

55.5 Conclusion

Multi-stage segmented long-ring TWR antennas are low-profile, super-directive antennas with a gain above 11 dB in the axis direction and low sidelobes level. They can be used in a wide range of applications where low-profile and high directivity is a must. The multi-frequency capability of such antennas makes them suitable for mobile and satellite communication systems.

References

1. J.L. Altman, Microwave circuits. D. Van Nostrand Company, 1st edn. (1964)
2. A.L. Feldshtein, L.N. Yavich, Synthesys of multi-ports networks at UHF. Svyaz, M (1971)
3. T.T. Wu, Theory of thin loop antenna. J. Math. Phys., Nov-Dec (1962)
4. H. Whiteside, W.P. King, The loop antenna as a probe. IEEE Trans. Antennas Propagat. Oct (1963)
5. D.V. Dicky, Radiating structures base on travelling wave resonator. Ph.D. thesis, SPb, (2004)
6. A.Z. Fradin, Antennas and feeders. M., Svyaz (1977)
7. A.L. Drabkin, V.L. Zuzenko, Antennas. M., Svyaz (1977)
8. V.M. Bogod, V.N. Dikiy, et al., Multiwave Feed based on running wave resonator with a combined phase centre for radioastronomy. Astro fiz. issledovania (Izvestia SAO), **17,** pp. 124–130 (Russian) (1983)
9. V.N. Dikiy, O.A. Kuzmin, C.G.M. van't Klooster, The multi-frequency band capability of the Radioastron antenna. In *Proceedings of JINA 92, Nice*, pp.305–308 (1992)
10. V.N. Dikiy, D.V. Dikiy, V.K. Nuzhin, Multifrequency antenna of circular polarization with nearly continuous bands overlapping. Proceedings of JINA 94, Nice, p.548–549. (1994).
11. Herscovici N., A wide-band single-layer patch antenna. IEEE Trans. Antennas Propag. Apr (1998)
12. Z.N. Chen, M.Y.W. Chia, Broad-band suspended probe-fed plate antenna with low cross-polarization levels. IEEE Trans. Antennas Propag. (2003)
13. Z.N Chen, M.Y.W. Chia, C.L. Lim, A stacked suspended plate antenna. Microwave and optical technology letters (2003)
14. R. Pokuls, J. Uher, D.M. Pozar, Dual-frequency and dual-polarization microstrip antennas for SAR applications. IEEE Trans. Antennas Propagat. (1998)
15. S.D. Targonski, D.M. Pozar, Dual-band dual polarized printed antenna element. Electron. Lett. (1998)
16. W.S. Chen, C.K. Wu, K.L. Wong, Novel compact circularly polarized square microstrip antenna. IEEE Trans. Antennas Propagat. (2001)

Chapter 56
Modeling the Effect of Geomagnetic Field Anomalies on the Frequency of Satellite Onboard Rubidium Clock in the GPS Navigation System

Daniil Borisov , **Roman Lozov** , **Maxim Kuznetsov** ,
Anastasiya Nitkina , **and Aleksandra Chekireva**

Abstract This paper presents the results of modeling the influence of the Earth's magnetic anomalies on the positioning accuracy of the GPS navigation system. We estimate the orientational frequency shift influence of a quantum standard with optical pumping on the positioning accuracy of the GPS navigation system. Calculations were carried out using the British Geological Survey (BGS) API, which is based on the World Magnetic Model (WMM). Comparison of calculated and experimental data on the time dependences of positioning errors by calculating the values of the correlation coefficients is presented.

Keywords Navigation satellites systems · Atomic clock · GPS · WMM · BGS · Geomagnetic field · Light shift

56.1 Introduction

The modern world cannot be imagined without satellite navigation systems. After all, it solves a rather important navigation-time problem, the result of which is the determination of the location of the object in space, the rate of change of its coordinates, as well as the establishment of the exact time [1–5].

However, the solution of the above problem is achieved if the time scales of special on-board computing equipment are bound, the deterioration of accuracy due to various external sources of errors of which can lower the quality of the desired result. Such onboard equipment can be an optically pumped rubidium atomic clock, and one of the sources of error is the geomagnetic field.

The operation of the atomic clock on Rb^{87} vapor is based on the method of double radio-optical resonance. With this method, two quantum transitions are simultaneously carried out: optical and microwave ranges [6]. The first is responsible for the redistribution of the populations of the hyperfine sublevels F of the ground state $^2S_{1/2}$,

D. Borisov (✉) · R. Lozov · M. Kuznetsov · A. Nitkina · A. Chekireva
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

and the second causes the corresponding induced transitions. Each F level from the hyperfine structure is split into 2F + 1 sublevels with different values of the magnetic number m_F [7].

A microwave transition between F = 2, m_F = 0 and F = 1, m_F = 0 of the main energy level $^2S_{1/2}$ (0–0 transition) is used as a reference. Its choice as a standard is explained by the fact that the sublevels m_F = 0 are least of all exposed to the influence of parasitic magnetic fields that can arise from various sources. But this does not at all negate the fact that the effect of magnetic fields makes a certain contribution to the energy hyperfine structure, changing the length of the quantum transition line, which leads to the so-called light shift [8, 9].

The magnitude of the light shift depends on many factors, such as the nuclear spin of the alkaline atom, the intensity and polarization of the radiation from the pumping source, the composition and pressure of the inert gas in the working absorption cell, as well as the angle between the direction of the pumping light and the magnetic field vector. The latter affects the movement of a communications satellite in orbit, when the magnitude and direction of the geomagnetic field, including due to magnetic Earth anomalies, have different variations, which has a direct effect on the measurement error of modern rubidium atomic clocks with optical pumping used in the on-board apparatus-tour, and this, in turn, leads to an error in the solution carried out by various navigation systems (GLONASS, Galileo and GPS), the navigation-time problem. To weaken the effect of an external magnetic field on the frequency stability of rubidium atomic clocks, multilayer magnetic shields, including cylindrical ones, made of materials with high magnetic permeability, are used. Practice shows that the calculated shielding coefficients for a constant magnetic field (~ 10^5–10^6) can differ significantly from the experimental values (~ 10^2–10^3) [10], which is due to the presence of technological holes connecting electronic components in screen designs. Therefore, the screening coefficients of the magnetic screen are different for different directions of the magnetic field, which, due to this fact, will certainly affect the vapors of the working substance of the atomic clock. The foregoing determines the relevance of a detailed study of the effect of light frequency shift on the metrological characteristics of navigation satellite systems.

Thus, the purpose of this work is to establish the influence of the Earth's magnetic field on the satellite navigation systems accuracy by means of computational-theoretical modeling for a GPS satellite taking into account data from a real map of the planet's magnetic field (the map includes various magnetic anomalies), which results in a light shift of the microwave resonance frequencies of the onboard optically pumped rubidium atomic clock.

56.2   Computational Modeling

Earlier in [11], an algorithm was already presented that made it possible to determine the effect of a magnetic field on the accuracy of a rubidium atomic clock, and then calculate the positioning error of a GPS communication satellite. However, such an

algorithm (presented in [11]) did not consider the real magnetic field, which includes various anomalies, but only the theoretical one, which is a uniform field over the entire radius of the calculation. The algorithm of the authors of this article considers the real situation when calculating the positioning error.

This algorithm has the following stages of work.

1. The calculation of the Earth's magnetic field (considering all anomalies of the magnetic field) is carried out at the distance of the GPS communication satellite. This calculation is performed using a special API developed by the British Geological Survey (BGS), which is based on the World Magnetic Model (WMM).
2. The calculation of the light frequency shift of the rubidium atomic clock on board the GPS communication satellite is carried out.
3. The calculation of the positioning error of the GPS communication satellite is carried out (taking into account only the influence of the geomagnetic field, and hence the light shift of the frequency of the onboard rubidium atomic clock).
4. Comparison of the results of GPS positioning (considering the Earth's magnetic field) with real data presented in open literature is carried out [12, 13].

The input parameters of the program are [14, 15]:

1. The angle of inclination of the satellite orbit to the plane of the Earth's equator;
2. The major semi axis of the elliptical orbit of the communications satellite under study;
3. Eccentricity of the elliptical orbit of the communications satellite under study;
4. Working magnetic field inside the magnetic shield of the investigated atomic clock;
5. Location of the atomic clock on board the communications satellite;
6. The number of desired calculated points for one rotation period of the communication satellite;
7. The number of periods of rotation of the communication satellite.

The screening of the magnetic shields of the rubidium atomic clock in the program is carried out the considering data presented in article [10].

56.3 Results

The calculation of the positioning error of the GPS communication satellite, which considers only the effect of the real geomagnetic field with magnetic anomalies (which is described by WMM), is presented in Fig. 56.1 by red dash-dotted line. The black solid line represents the experimental dependence, which is described in [13].

The distribution (taken from the Internet) of the WMM magnetic fields is shown in Fig. 56.2.
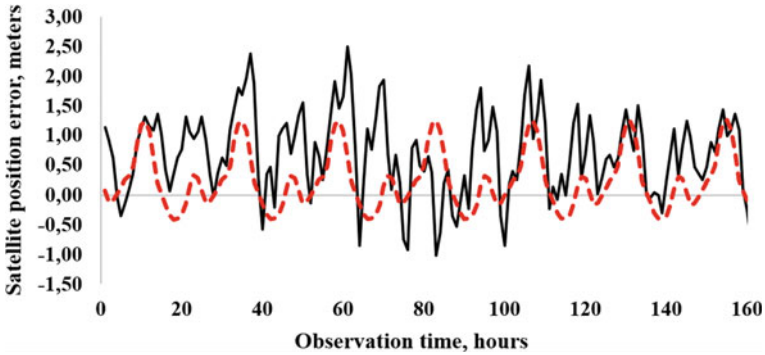
Fig. 56.1 Experimental (solid black line) and calculated (dotted red line) time dependencies GPS satellite position error
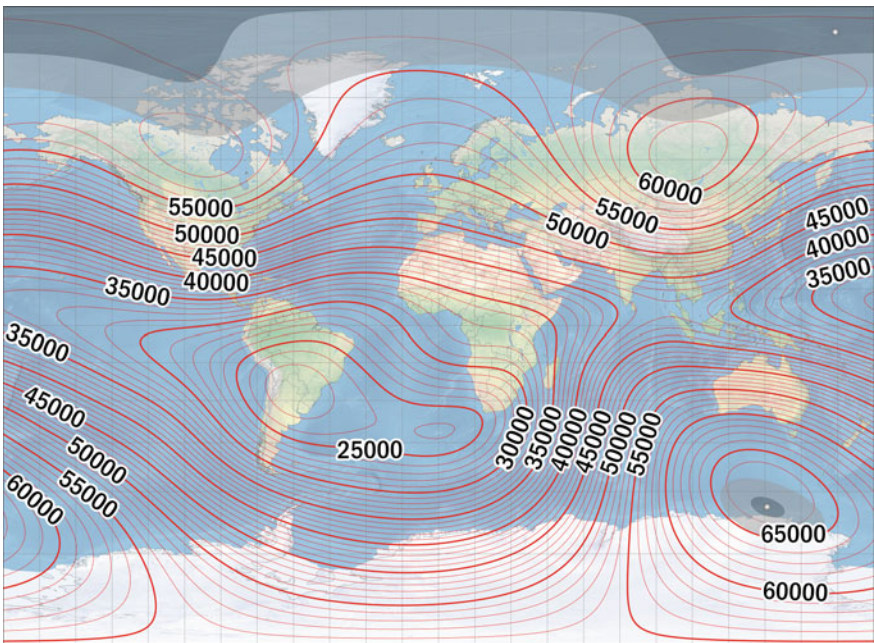
Fig. 56.2 Distribution of WMM magnetic fields (values are given in nT)

A comparison of experimental and calculated by the developed algorithm dependencies of the positioning error of the GPS communication satellite on the observation time is presented in the form of correlation coefficients in Table 56.1.

Table 56.1 Correlation coefficients values to dependencies in Fig. 56.1 at different starting times

Δt	−6 T	−4 T	−2 T	0	2 T	4 T	6 T
R_{GPS}	0.42	0.43	0.42	0.42	0.43	0.42	0.42

Δt is starting time, T is the satellite rotation period, R_{GPS} is correlation coefficient value (Fig. 56.1).

56.4 Conclusion

As a result of the research carried out, an algorithm was implemented that calculates the positioning error of a GPS communication satellite. This algorithm considers the effect of the real magnetic field of the Earth with all anomalies on the frequency of the rubidium atomic clock. The real geomagnetic field was considered using the BGS API.

The positioning error versus the observation time calculated by the developed algorithm was compared with the experimental data.

It can be concluded that the effect of a real magnetic field, considering magnetic anomalies, contributes to the positioning error of a GPS communication satellite due to the appearance of a light shift. The greater the magnetic anomaly, the greater the light shift, and hence the greater the positioning error of the GPS communication satellite (large peaks in Fig. 56.1).

The similarity of the correlation coefficients in Table 56.1 is explained by the rotation period of the GPS communication satellite, which is half the rotation period of the Earth.

References

1. A. Rachitskaya, I. Tsikin, Gnss integrity monitoring in case of a priori uncertainty about user's coordinates, in *2018 IEEE International Conference on Electrical Engineering and Photonics, EExPolytech*, vol. 1, no. 1 (2018), pp. 83–87
2. A. Melikhova, I. Tsikin, Optimum array processing with unknown attitude parameters for GNSS anti-spoofing integrity monitoring, in *41st International Conference on Telecommunications and Signal Processing* vol. 3, no. 1 (2018), pp. 1–4
3. I. Tsikin, E. Shcherbinina, Algorithms of GNSS signal processing based on the generalized maximum likelihood criterion for attitude determination, in *25th Saint Petersburg International Conference on Integrated Navigation Systems* vol. 3, no. 1 (2018), pp. 1–4
4. R. Lozov, A. Baranov, S. Ermak, V. Semenov, Comparison of orientational error of an optically pumped quantum sensor in on-board equipment of Galileo and GPS satellite systems. J. Phys. Conf. Ser. **1236**(1)
5. A. Melikhova, I. Tsikin, Decision-making algorithms based on generalized likelihood ratio test for angle-of-arrival GNSS integrity monitoring, in *25th Saint Petersburg International Conference on Integrated Navigation Systems* vol. 3 no. 1 (2018), pp. 4–5
6. W. Happer, Optical pumping. Rev. Mod. Phys. **44**(2), 169–170 (1972)

7. F. Riehle, *Frequency Standards: Basics and Applications* (Wiley, New York, 2006)
8. A. Davydenko, E. Popov, I. Tsikin, Radio direction finding of highly dynamic objects based on the MUSIC method using antenna arrays with a small number of elements. Radioengineering **84**(12), 5–16 (2020)
9. W. Happer, Effective operator formalism in optical pumping. Phys. Rev. **163**(1), 12–25 (1967)
10. R. Lozov, S. Ermak, A. Baranov, V. Semenov, The influence of the magnetic field direction variations on the frequency stability of a gas cell atomic clock, in *Proceedings of the 2019 IEEE International Conference on Electrical Engineering and Photonics* vol. 1 no. 1 (2019), pp. 46–49
11. E. Borisevich, A. Korolev, R. Lozov, The orbits shape influence of the navigation satellite systems on positioning accuracy. Int. Youth Conf. Electron. Telecommun. Inf. Technol. **2**(1), 761–775 (2021)
12. D. Svehla, *Geometrical Theory of Satellite Orbits and Gravity Field*, vol. 1 no. 1 (Springer Theses), pp. 1–28
13. A.S. Bandura, Evaluation of the accuracy of the data of navigation messages of spacecraft of satellite radio navigation systems. Sci. Tech. J. Inf. Technol. Mech. Opt. **2**(1), 2–5 (2006)
14. G. Heald, Revised GPS satellite frequency corrections (2021)
15. X. Yin, Modelling and assessment of Galileo and Galileo/GPS velocity determination with stand-alone receiver. J. Navig. **2**(1), 1–13 (2021)

Chapter 57
Features of a Fiber-Optics Transmission System Using Dispersion-Managed Optical Solitons

Artur A. Ermolaev [ORCID]**, Maria A. Shevchenko** [ORCID]**, Elena I. Andreeva** [ORCID]**, and Dmitry P. Andreev** [ORCID]

Abstract The analysis of the parameters of a high-speed information transmission system using dispersion-managed solitons is carried out. It is shown that the quality of information transmission can be increased by choosing the input–output point of the symbol sequence on the dispersion map. Fiber-optics soliton information systems are distinguished by high stability of data transmission. However, in real fibers, it is possible to realize the soliton regime at certain approximations. For high-speed WDM systems, it is advisable to use the dispersion-managed soliton. The feature of this mode is the low threshold level forming soliton in an optical fiber, which ensures a low level of inter-channel interference. The dispersion map parameters largely determine the bit rate. However, with the given parameters of the dispersion map, it is possible to select the parameters of the symbol pulse, providing a higher bit rate and higher signal-to-noise ratio in the system. A computer simulation of a fiber-optic data transmission system using optical solitons has been carried out. It is shown that when constructing a system with a dispersion map, which assumes input–output of spectrally limited pulses, the quality of data transmission increases. The method for the optimization of the parameters of the soliton data system is presented.

Keywords Fiber optics transmission system · Optical soliton

57.1 Introduction

Soliton information transmission systems are distinguished by high stability and high information capacity as product of the bit rate on the transmission distance [1–15]. However, in the real fibers, it is possible to realize the soliton regime at certain approximations. Various types of the soliton-like regime are used to solve

A. A. Ermolaev · M. A. Shevchenko · E. I. Andreeva
The Bonch-Bruevich St. Petersburg State University of Telecommunications, St. Petersburg 193232, Russia

D. P. Andreev (✉)
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: andreev.dp@edu.spbstu.ru

the specific problem. The classical regime of the fundamental soliton provides a high speed in the transmission channel but requires a relatively high peak power of symbol pulses. The use of wavelength division multiplexing (WDM) is hindered by the effect of the soliton's interaction. The pass-average soliton regime is suitable for long-haul systems and has symbol width and bit rate limitations. For high-speed WDM systems, it is advisable to use the dispersion-managed soliton mode. The distinguishing feature of this mode is the low threshold level forming soliton in an optical fiber, which ensures a low level of inter-channel interference.

Dispersion management is actively used in the high-speed information transmission systems. It turns out that if fibers with varying dispersion are used or if fibers with different dispersions are alternated in a special way, the parameters of soliton systems can be significantly improved. Periodic dispersion (dispersion-managed) circuits consist of alternating fiber lengths with opposite sign dispersion. Dispersion-managed systems are attractive because of the following.

– The average variance of the entire system can be made very small, which improves the characteristics of the systems as a whole.
– The dispersion of each individual segment (β_2) remains large enough, which reduces the influence of the effects of four-wave mixing and third-order dispersion.

57.2   The Main Characteristics of the Dispersion Map

Dispersion scheme (Fig. 57.1) consist of the two fibers with the anomalous (negative) dispersion β2a and length la and positive dispersion β2n and length ln

$$L_{map} = l_a + l_n$$

where L_{map} is the period of the dispersion map.

Path-average dispersion:

Fig. 57.1 Typical dispersion management map
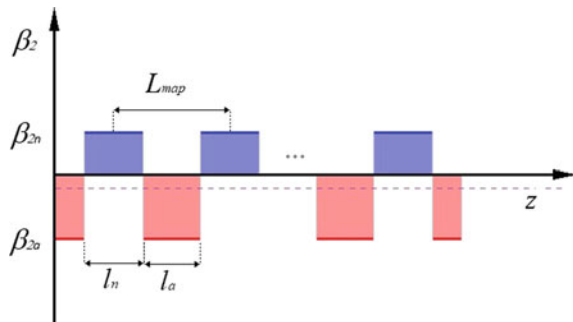
$$\overline{\beta_2} = \frac{\beta_{2n}l_n + \beta_{2a}l_a}{l_n + l_a}$$

The scheme map is selected in such a way that the average variance turns out to be small, although it remains anomalous (path-average dispersion parameter of the map, $\overline{D} > 0$, and value $\overline{\beta_2} < 0$).

The input pulse width T_o is determined depending on the initial chirping C_o and the parameter T_{map} is dispersion map parameter [4].

$$T_o = T_{map}\sqrt{\frac{1 + C_o^2}{|C_o|}}$$

where T_o is pulse width and C_o is pulse chirp between the segments with different dispersion, and.

$$T_{map} = \sqrt{\frac{|\beta_{2n}\beta_{2a}l_nl_a|}{\beta_{2n}l_n - \beta_{2a}l_a}}$$

In the segment middle value $C(z) \cong 0$ and pulse width became T_{min} [4]

$$T_{min} = \frac{T_o}{\sqrt{\left(1 + C_o^2\right)}} = \frac{T_{map}}{\sqrt{|C_o|}}$$

The value of T_{min} can be used to estimate the allowable bit rate.
Initial soliton pulse energy E_o [4]

$$E_0 = 2\sqrt{2\pi}\frac{\beta_{2a}\varepsilon_a(l_a) + \beta_{2n}\varepsilon_2(l_n)}{\left(\frac{\gamma_{0a}}{\sqrt{c_a}}\right)lnr_a + \left(\frac{\gamma_{0n}}{\sqrt{c_n}}\right)lnr_n}$$

while $(i = a, n)$

$$\varepsilon_i(z) \equiv \frac{1}{2}\frac{c_i - a_ib_i}{\sqrt{c_i^3 b_i}}\left[\tan^{-1}\left(\sqrt{\frac{c_i}{b_i}}(l_i)\right) - \tan^{-1}\left(\sqrt{\frac{c_i}{b_i}}(l_i - 2z)\right)\right]$$

$$r_i = \frac{T_0 - l_i\sqrt{c_i}}{T_0 - l_i\sqrt{c_i}}$$

$$a_i \equiv \frac{C_0^2}{l_i^2}$$

$$b_i \equiv \frac{1}{2}\beta_{21}C_0l_i + T_0^2$$

$$c_i \equiv \mp\frac{\beta_{2i}C_0}{2l_i}$$

Fig. 57.2 Spatial dynamics of changes in the pulse width $T(z)$ (1–2) of a soliton and a chirp $C(z)$ (3–4) at $C_o = 3$ (1, 3) and $C_o = 1$ (2, 4) (**a**) on the period of the dispersion map $l_a = l_n = 48$ km, $|\beta_2| \cong 3$ ps^2/km (**b**)
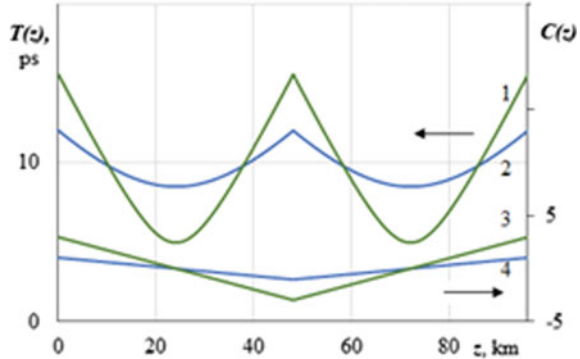
The soliton is formed as averaged over a long nonlinearity manifestation. For pulse chirp [4]:

$$
C(z) = \begin{cases}
C_o\left(1 - \frac{2z}{l_a}\right), 0 \le z \le l_a \\
-C_o\left(1 - \frac{2z}{l_n}\right)(z - l_a), l_a \le z \le L_{map}
\end{cases}
$$

For pulse width [4]:

$$
T^2(z) = \begin{cases}
T_o^2 + 2\beta_{2a}C_o\left(1 - \frac{z}{l_a}\right)z, \ 0 \le z \le l_a \\
T_o^2 + 2\beta_{2n}C_o\left(1 - \frac{1}{l}(z - l_a)\right)(z - l_a), \ l_a \le z \le L_{map}
\end{cases}
$$

Along the segment length, the parameters of the soliton pulse, such as the pulse width $T(z)$, the frequency chirp $C(z)$, experience periodic changes, keeping their values T_o and C_o at the segment boundary (Fig. 57.2).

57.3 Computer Simulation

The fiber-optic system with a period $l_a = l_n = 48$ km was modeled in OptiSystem. Input soliton pulse width, corresponding to $T_o = 12$ ps and $C_o = 1$, so that the $T_{map} = 8.5$ ps realized the bit rate of 10 Gbit/s.

The dispersion map parameters largely determine the bit rate. However, with the given parameters of the dispersion map, it is possible to select the parameters of the symbol pulse (Figs. 57.3 and 57.4), providing a higher bit rate and higher SN-ratio in the system. For this, a differential dispersion map is used, which makes it possible to input/output data at the midpoint of the segment, the so-called point of zero chirp, where the pulse width takes on a minimum value.
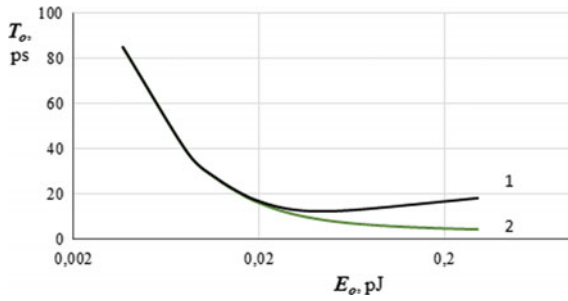
Fig. 57.3 Pulse width T_o, **a** at the boundary of segments with different dispersion (1) and the pulse width T_{min} (2) in the middle of the segment vs initial energy E_o of the soliton pulse and the pulse power P_o, **b** at the boundary of segments with different dispersion (1) in the middle of the segment (2)

Fig. 57.4 Peak power P_o in the middle of a segment vs chirp C_o at the boundary of segments with different dispersion
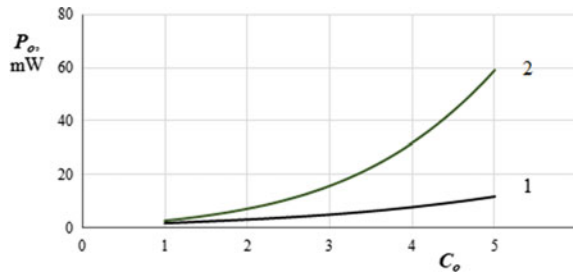
As shown by computer simulation with the OptiSystem-program, with an initial value of $C_o = 1$ and entering the middle of a segment with anomalous dispersion, the value of the Q- parameter increases by 3 times. An increase in the initial energy E_o of the pulse and, accordingly, an increase in $C_o = 3$ additionally increases the Q- parameter up to 5.6 times (Fig. 57.5). However, a further increase in these
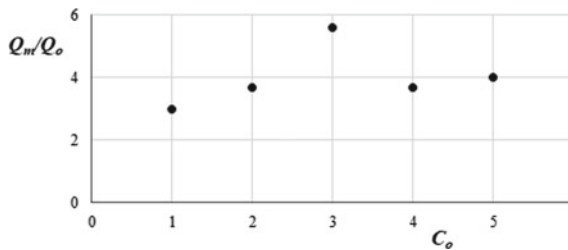
Fig. 57.5 Q-parameter at the output of the system Q_m when entering the middle of a segment with anomalous dispersion (at the point of zero chirp) vs the value of C_o chirp at the boundary of segments with variance of different signs, normalized to the value of Q_o (the value of the Q-parameter at $C_o = 1$) and data input at segment boundaries

values C_o already leads to a decrease in the value of the Q-parameter since the pulse width at the segment boundary and the interaction of solitons significantly increase.

57.4 Conclusion

A computer simulation of a fiber-optic data transmission system using optical solitons has been carried out. It is shown that when constructing a system with a dispersion map, which assumes input–output of spectrally limited pulses, the quality of data transmission increases. The optimal parameters of such a system can be chosen.

References

1. G. Agrawal, Nonlinear fiber optics. 5th Academic Press, (San Diego, USA, 2013)
2. A. Shcherbakov, E. Andreeva, Performance data of lengthy-span soliton transmission system. Opt. Fiber Technol. **2**, 127–133 (1996)
3. Y. Kivshar, G. Agrawal, Optical solitons from fibers to photonic crystals. The Institute of Optics University of Rochester (New York, USA, 2003)
4. E. Poutrina, G. Agrawal, J. Lightwave Technol **21**(4), 990–996 (2003)
5. E. Andreeva, M. Bylina, S. Glagolev, P. Chaimardanov, Proc. Telecommun. Univ. **4**(1), 5–12 (2018)
6. A. Hasegawa, *Massive WDM and TDM Soliton Transmission Systems* (Academic Press, New York, USA, 2002)
7. E. Andreeva, M. Bylina, S. Glagolev, P. Chaimardanov, Proc. Telecommun. Univ. **4**(2), 26–35 (2018)
8. M. Tarasenko, V. Lenets, K. Malanin, N. Akulich, V. Davydov, J. Phys. Conf. Series **1038**(1), 012035 (2018)
9. I. Potapov, YETI-Proc. pp. 215–219 (2020)
10. V. Fadeenko, V. Kuts, D. Vasiliev, V. Davydov, New design of fiber-optic communication line for the transmission of microwave signals in the X-band. J. Phys. Conf. Series **1135**(1), 012053 (2018)
11. M. Ferreira, *Nonlinear Effects in Optical Fibers* (Wiley, Hoboken, New Jersey, USA, 2011)
12. N. Grebenikova, V. Davydov, A. Moroz, M. Bylina, M. Kuzmin, IOP Conf. Series: Mater. Sci. Eng. **497**, 012109 (2019)
13. L. Mollenauer, G. Gordon, *Solitons in Optical Fibers* (Academic Press, San Diego, USA, 2006)
14. N. Myazin, V. Dudkin, N. Grebenikova, V. Rud, A. Podstrigaev, Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics) 11660 LNCS, pp. 744–756 (2019)
15. G. Agrawal, Nonlinear fiber optics: its history and recent progress. Opt. Soc. Am. B J. **28**(12), 1–10 (2011)

Chapter 58
Application of Nonlinear Algorithms with Decision Feedback for FTN Signals Coherent Detection

Check for updates

Ilya Lavrenyuk , **Sergey Makarov** , **and Wei Xue**

Abstract An approach to the problem of increasing the bandwidth efficiency of communication systems, which has recently received considerable attention, is the introduction of so-called controlled inter-symbol interference in the time and/or frequency domain of the signal. However, in this case, in order to minimize energy loss, more computationally complex detection algorithms should be applied to eliminate the negative effects of inter-symbol interference in the signal. One of the barriers to implementing the above approach in existing communication standards is the high complexity of implementing detection algorithms. This paper presents a study on application of coherent detection algorithms with decision feedback for the faster-than-Nyquist signaling to decrease detection complexity.

Keywords Faster-than-Nyquist · Detection algorithm · Decision feedback · Intersymbol interference

58.1   Introduction

Random sequences of binary FTN (Faster-Than-Nyquist signaling) signals *s(t)* with the increased duration exceeding the transmission time T of one bit of messages without coding can provide a channel rate R above the "Nyquist barrier" [1, 2]. High spectral efficiency is achieved using signals of duration $T_s = LT$ (L = 2, 3, …) and energy E_s, which can be formed using a filter with an extremely narrow band of occupied frequencies ΔF. The message transmission occurs in the conditions of intersignal (intersymbol) interference ISI, caused by the superposition of adjacent signals on each other, which significantly affects the decrease in the bit error rate performance at high channel message transmission rates. The properties of such

I. Lavrenyuk (✉) · S. Makarov
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

W. Xue
College of Information and Communication Engineering, Harbin Engineering University, Harbin, China

527

FTN signals are considered in [3–5] and estimates of their detection performance in channels with additive white Gaussian noise (AWGN) with an average power spectral density $N_0/2$ are given.

To receive FTN signals, in most cases, algorithms of coherent detection are used, including maximum likelihood sequence estimation (MLSE) algorithms for processing message packets. These algorithms are the most efficient and in ISI conditions they carry out a weighted enumeration of all possible combinations of received signals. However, the hardware complexity of the implementation of such algorithms is not realizable for high absolute transmission rates. Reducing the complexity of processing algorithms is achieved by using the MLSE detection methods based on the Viterbi algorithm and its modifications [3, 6, 10]. However, it is not possible to achieve any significant gains in reducing the number of computational operations when demodulating signals. At the same time, when using the classical algorithm of element wise detection without taking into account ISI, which ensures the maximum speed of the signal demodulator, energy losses due to the presence of significant ISI reach 10–20 dB at specific bit rates R > 1/T. Similar difficulties arise when detection signals with interference in the frequency domain, such as in SEFDM approach [11–15]. The approach of joint application of error-correction codes and optimized signal shapes considered in papers [7–9]. Such approach demonstrates promising results in bandwidth efficiency gain without extremely complex detection and energy efficiency degradation.

It is reasonable to formulate the problem of reducing energy costs due to the transition to the application of algorithms for receiving FTN signals, which occupy an intermediate place between easily implemented element wise processing, which does not take into account ISI, and algorithms for MLSE detection taking into account ISI, but having a complex. These algorithms include nonlinear iterative algorithms for coherent reception with decision feedback. The main idea of using such algorithms is that in the presence of intersymbol interference, the signals that follow before and after the analyzed signal fall into the observation interval of the useful signal. Under these conditions, it is possible to use decisions about the previous (following before the analyzed) received symbols as known when receiving this symbol. The so-called ideal feedback on decisions about the previous received symbols is introduced into the reception algorithm. This makes it possible to compensate for intersymbol interference from previous received symbols [16, 17].

The purpose of this work is to determine the possibility of increasing the performance of FTN signal sequences detection when using coherent processing algorithms with decision feedback at increased bit rates R > 1/T.

58.2 Definition of Random FTN Sequences

Let us consider an FTN signal with an arbitrary amplitude pulse shape *a(t)*, having a maximum value A_0, a carrier frequency and a pulse duration $T_s = LT$ (L > 1). We represent it in the following form:

$$s_r(t) = A_0 a(t) d_r^{(0)} cos(2\pi f_0 t) \tag{58.1}$$

where $d_r^{(0)}$ is the modulation symbol in the time interval from 0 to T. For BPSK $d_r^{(0)}$ = ± 1 and r = 1, 2.

A random sequence consisting of N signals (58.1) for an arbitrary symbol rate can be written as:

$$y(t) = A_0 \sum_{n=-N/2}^{N/2} a(t - \xi n T) d_r^{(n)} cos(2\pi f_0 t) \tag{58.2}$$

Sequence (58.2) provides binary data transmission at symbols rate R = 1 / ξT (0 < ξ ≤ 1) above the "Nyquist barrier" (at $\Delta F = 1/2\ T$). In the general case, for FTN signals with the channel alphabet base M, the values in (58.1) are determined by the following expression:

$$d_j^{(n)} = \frac{M - 2j + 1}{M - 1}, j = 1...M$$

For example, for pulse amplitude modulation and M = 4 we have j = 1, 2, 3, 4 and $d_1^{(n)} = 1$; $d_2^{(n)} = 0.33$; $d_3^{(n)} = -0.33$; $d_4^{(n)} = -1$. The values $d_j^{(n)}$ are evenly spaced from + 1 to −1. For FTN signals, the values $d_j^{(n)}$ have equal occurrence probabilities for each n.

Random sequences (58.2) of FTN signals can be based on square root raised cosine (RRC) pulses. Figure 58.1 shows the normalized RRC functions of the pulse $a(t)$ for the roll-off factor of the frequency characteristic $|H(f)|^2$ of the filter $\beta = 0.0$,
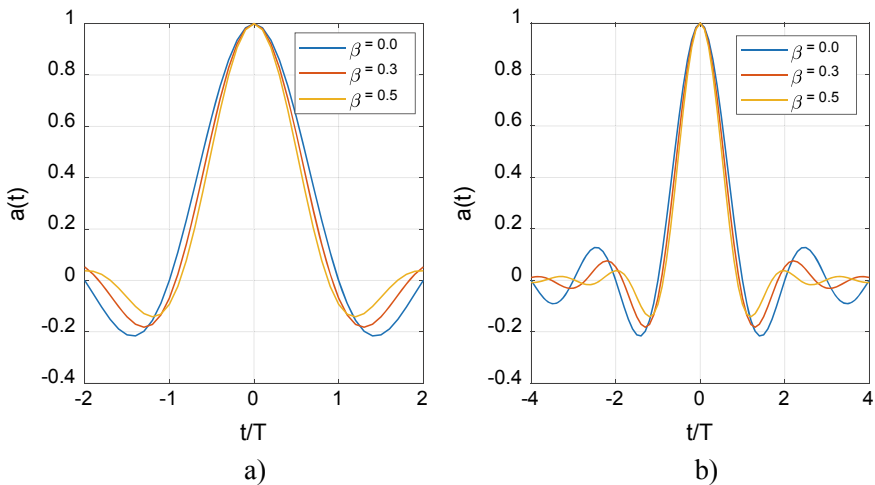
Fig. 58.1   RRC impulse type with different roll-off factors: a) $T_s = 4\ T$; b) $T_s = 8\ T$

0.3, 0.5 and duration $T_s = 4\,T$ (Fig. 58.1a) and $T_s = 8\,T$ (Fig. 58.1b). Note that for the value of the parameter $\beta = 0$, the form of the amplitude pulse $a(t)$ coincides with the form of the function *sin(x)/x*. This amplitude pulse has the maximum amplitude side lobes in time domain and, accordingly, the highest level of intersymbol interference.

In general, a random signal sequence generated from truncated RRC pulses, even at a transmission rate of R = $1/T$, will have a high enough level of intersymbol interference. Consider a nonlinear algorithm for coherent detection with decision feedback under such conditions.

58.3   Algorithm for Coherent Detection with Decision Feedback.

When analyzing coherent signal processing, we will assume that the measurement of the initial phase of a high-frequency oscillation with a carrier frequency is determined using a phase-locked loop (PLL) device according to the preamble before the information sequence (58.2). Clock synchronization is also determined from a periodic sequence of symbols using a clock synchronization system (CSS). Under these conditions, the analyzed process at the input of the receiving device has the form:

$$r(t) = \mu s_l(t) + \mu y_-(t, i) + \mu y_+(t, q) + n(t). \tag{58.3}$$

Here, *n(t)* it is additive white Gaussian noise (AWGN) with an average power spectral density of $N_0/2$, which is the transmission coefficient over the communication channel, which is determined by the level of fading μ. In (58.3) for R = $1/T$, the signal sequence (58.2) is converted into two preceding $y_-(t, i)$ and subsequent $y_+(t, q)$ signal sequences. The indices i and q correspond to the i-th preceding and q-th subsequent combinations of characters.

$$y_-(t, i) = \sum_{k=-1}^{-(L-1)} s_{ri}^{(p)}(t - k\xi T) = \sum_{k=-1}^{-(L-1)} d_{ri}^{(p)} a(t - k\xi T) cos(2\pi f_0 t); \tag{58.4}$$

$$y_+(t, q) = \sum_{k=1}^{L-1} s_{rq}^{(p)}(t - k\xi T) = \sum_{k=1}^{L-1} d_{rq}^{(p)} a(t - k\xi T) cos(2\pi f_0 t). \tag{58.5}$$

When writing the likelihood ratio functional, we will take into account the the decisions on the previously transmitted symbols. At the same time, we will keep the averaging over the subsequent unknown symbols. Then for the M-ary channel alphabet we get:

$$\sum_{q=1}^{M^{L-1}} \exp\left(\frac{1}{N_0}\left(\begin{array}{l} 2\mu \int_0^{LT} r(t)(s_l(t) + y_-(t,i) + y_+(t,q))dt \\ -\mu^2 \int_0^{LT} (s_l(t) + y_-(t,i) + y_+(t,q))^2 dt \end{array}\right)\right) >$$
$$> \sum_{q=1}^{M^{L-1}} \exp\left(\frac{1}{N_0}\left(\begin{array}{l} 2\mu \int_0^{LT} r(t)(s_r(t) + y_-(t,i) + y_+(t,q))dt \\ -\mu^2 \int_0^{LT} (s_r(t) + y_-(t,i) + y_+(t,q))^2 dt \end{array}\right)\right). \tag{58.6}$$

Let us transform (58.6), assuming that all signals following the analyzed one on the current symbol interval are additional noise and a priori information about the form of implementation of this noise is not taken into account in processing:

$$\exp\left(\frac{1}{N_0}\left(2\mu \int_0^{LT} r(t)(s_l(t) + y_-(t,i))dt - \mu^2 \int_0^{LT} (s_l(t) + y_-(t,i))^2 dt\right)\right) >$$
$$> \exp\left(\frac{1}{N_0}\left(2\mu \int_0^{LT} r(t)(s_r(t) + y_-(t,i))dt - \mu^2 \int_0^{LT} (s_r(t) + y_-(t,i))^2 dt\right)\right). \tag{58.7}$$

Then expanding (58.7) in a power series and limiting ourselves only to the first terms of the expansion, we get: the *l*-th symbol is registered if the inequality holds:

$$\int_0^{LT} r(t)(s_l(t) - s_r(t))dt$$
$$> \frac{1}{2}\mu\left(\int_0^{LT} (s_l(t) + y_-(t,i))^2 dt - \int_0^{LT} (s_r(t) + y_-(t,i))^2 dt\right); \tag{58.8}$$

Consider the implementation of the solution feedback algorithm for FTN signals with binary modulation. Taking into account the fact that for BPSK $s_1(t) = -s_2(t)$, and simplifying (58.8), we obtain the following rule for making a decision:

$$\int_0^{LT} r(t)s_l(t)dt \gtrless \mu \int_0^{LT} s_l(t)y_-(t,i)dt \tag{58.9}$$

Consider representation (58.9) through baseband quadrature components. We will write $r(t)$ and $s_l(t)$ in the form:

$$\begin{aligned} r(t) &= A_{rc}(t)\cos(2\pi f_0 t) - A_{rs}(t)sin(2\pi f_0 t) \\ s_l(t) &= A_{lc}(t)\cos(2\pi f_0 t) - A_{ls}(t)sin(2\pi f_0 t) \end{aligned} \tag{58.10}$$

Taking into account (58.2) and (58.10), we rewrite (58.9) in the form:

$$d_l^{(0)} \int_0^{LT} A_{rc}(t)a(t)dt \underset{<}{\overset{>}{\gtrless}} \mu d_l^{(0)} \int_0^{LT} a(t) \sum_{k=-1}^{-(L-1)} d_{ri}^{(p)} a(t - k\xi T)dt \qquad (58.11)$$

As can be seen from (58.11), the algorithm for coherent detection with decision feedback involves comparing the result of correlation processing of a signal over the interval of its existence and comparing this result with a threshold, the value of which is determined by the interference value of the previous signal combinations. In addition, the level of the threshold depends on the transmission ratio μ of the communication channel, which is determined by the level of fading. Algorithm (58.11) provides feedback to the depth L of symbols for the transmission rate R = 1/T. At a transmission rate higher than this value, more than L number of previous symbols falls into the observation interval, which can be taken into account when receiving. Of course, the depth of feedback can also vary depending on the requirements for the hardware and software implementation (58.11). Algorithms (58.9) and (58.11) can be performed based on correlators or matched filters.

Let us consider the block diagram of the implementation based on correlators, which is shown in Fig. 58.2. From the analyzed process at the input of the receiving device with the help of phase (PLL) and clock synchronization (CSS), the reference waveform and clock synchronization frequency are defined from the message preamble. To implement algorithm (11), the channel transmission coefficient μ is estimated in the "Estimation μ" block. After the transfer of the analyzed process $r(t)$ to zero frequency and low-pass filtering from the components at the doubled carrier frequency, the correlation processing of the process $A_{rc}(t)$ takes place. At the output of the "symbol decision", estimates of the received symbols $\hat{d}_j^{(n)}$ of the
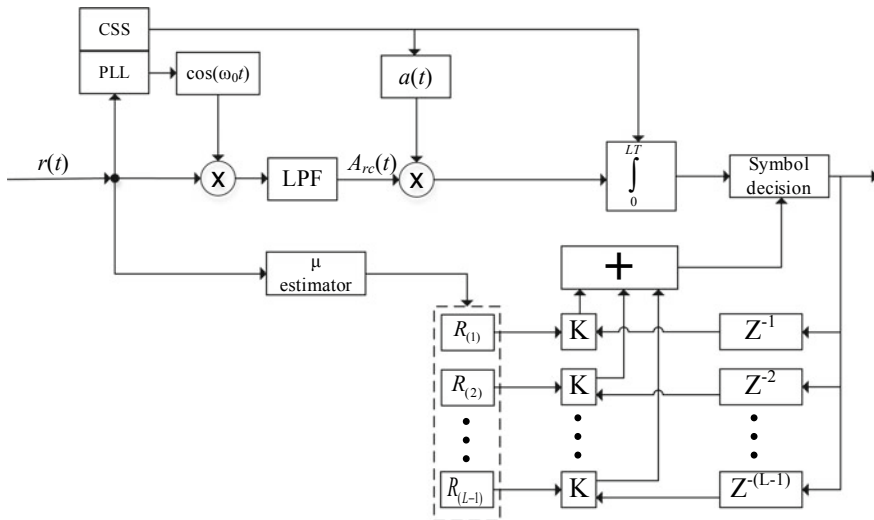
Fig. 58.2 Schematic flow-chart for implementation of algorithm (58.11)

channel alphabet appear. These estimates are taken into account when choosing the response threshold of the feedback loop. This loop includes time T delay elements $Z^{-k} (k = 1, 2...L - 1)$ for a transmission rate R = 1/T. As R increases, the number of delay elements increases accordingly. When constructing a feedback loop, it is necessary to have estimates of the cross-correlation coefficients of the transmitted signals included in (11). For this, the coefficients are calculated $R_{(i)}$ (i = 1, 2,…L-1). They are equal:

$$R_{(1)} = \int_0^{LT} a(t)a(t - T)dt\,L;$$

$$R_{(2)} = \int_0^{LT} a(t)a(t - 2T)dt;$$

$$\vdots$$

$$R_{(L-1)} = \int_0^{LT} a(t)a(t - (L - 1)T)dt$$

The coefficients $R_{(i)}$, taking into account the estimate μ, are added to the value of the response threshold of the decision block. This is done using the switching elements K and the adder.

Thus, in the symbol decision block the results of correlation processing (see (58.11)) are compared with the threshold value, the value of which changes and depends on the received previous symbols. Of course, parts of the signals following the analyzed signal will also fall into the integration interval. They considered as hindrance. To reduce the complexity of the implementation of the reception algorithm (58.11), it is possible to adjust the depth of the feedback on the solution, taking into account not all the previous signals that affect the current analyzed process, but only a part of them. Coefficient storage units $R_{(i)}$, switching elements K and delay elements Z^{-k} can be implemented on the basis of, for example, an FPGA. Of particular importance is the calculation of the μ parameter which is transmission coefficient over the communication channel. This is especially important when receiving signals under fading conditions. The simplest device of the μ *estimator* block can be performed with the classical automatic gain control scheme.

58.4   Simulation Results

Let us consider the results of simulation of a nonlinear algorithm with decision feedback when receiving FTN signals formed on the basis of RRC pulses of different duration. The purpose of the simulation is to estimate the bit error rate (BER) performance of coherent detection of FTN signals in the presence of additive white Gaussian noise (AWGN) in the transmission channel with an average power spectral density $N_0/2$ at various data rates.
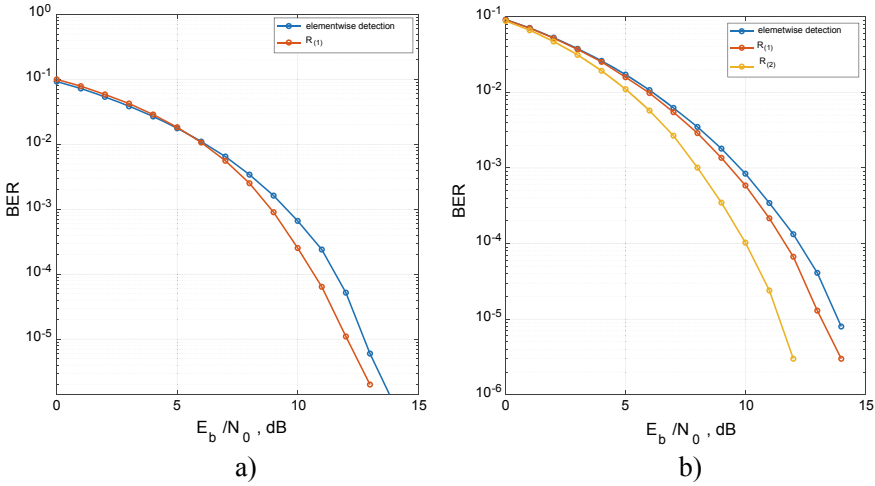
Fig. 58.3 BER performance for an RRC pulse with $\beta = 0$ and R = 1/T at different pulse durations: **a** $T_s = 2\,T$, **b** $T_s = 4\,T$

When performing simulation, binary phase shift keying with RRC shaping was used, the forms corresponding baseband pulses are shown in Fig. 58.1. The number of signals in sequence (58.2) N = 10^6. Figure 58.3 shows the dependences of the error probability (BER) on the signal-to-noise ratio (E_b/N_0, where E_b is the signal $s_r(t)$ energy in (58.1)) for the data rate R = 1/T. As a modulation RRC pulse, a pulse with a duration of $T_s = 2\,T$ (Fig. 58.3, a)) and $T_s = 4\,T$ (Fig. 58.3, b)) was taken when the roll-off factor of the frequency response $|H(f)|^2$ of the filter $\beta = 0$ (pulse type $sin(x)/x$).

For comparison, Fig. 58.4 shows the dependences of the error probability for the case of using the coherent element wise detection algorithm without compensation for intersymbol interference. Hereinafter, we will understand the notation of the feedback depth $R_{(1)}$, as consisting of one feedback pass; designation $R_{(2)}$—consisting of two feedback passes, namely $R_{(1)} + R_{(2)}$; designation $R_{(3)}$—consisting of three passes $R_{(1)} + R_{(2)} + R_{(3)}$ and so on. From the analysis of the dependences shown in Fig. 58.3, the following conclusions can be drawn. First, the use of the algorithm with feedback on the decision allows to increase the BER performance of detection compared to using the algorithm without ISI compensation even at a relatively low transmission rate R = 1/T for the case when signals are partial responsed. The energy gain is about 1 dB at $T_s = 2\,T$ (Fig. 58.3a) and about 2.5 dB at $T_s = 4\,T$ (Fig. 58.3b) in the range of error probabilities p = $10^{-4} - 10^{-5}$. Secondly, an increase in the feedback depth (Fig. 58.3, b) to two ($R_{(2)}$) gives a significant improvement in performance amounting to up to 2 dB in comparison with the case of using only one feedback pass ($R_{(1)}$).

Consider the BER performance when the message transmission rate is increased to R = 1.25/T for $T_s = 4\,T$ (Fig. 58.4, a) and $T_s = 8\,T$ (Fig. 58.4, a), for different
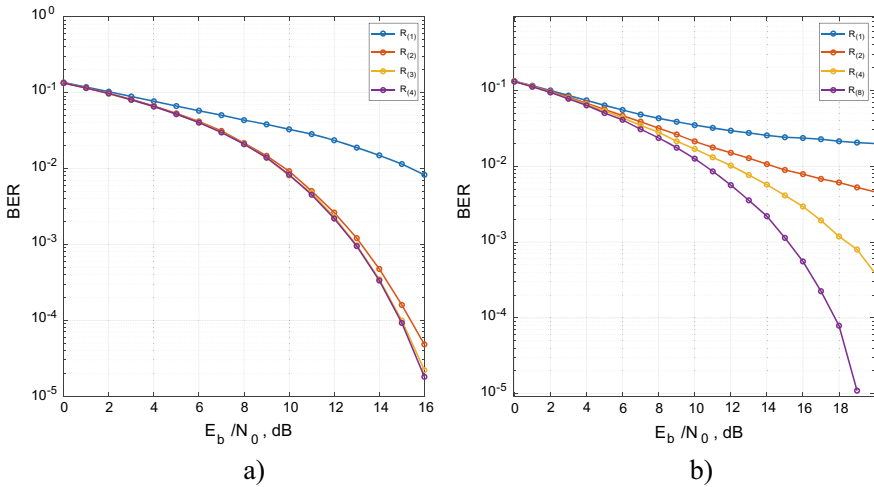
Fig. 58.4 BER performance for an RRC pulse with $\beta = 0$ and R $= 1.25/T$ at different pulse durations: **a** $T_s = 4\,T$, **b** $T_s = 8\,T$

values of the feedback depth. It can be seen that with an increase in R and, naturally, with an increase in ISI, the BER performance decreases. If we compare the obtained error probability values with similar BER values for the transmission rate R $= 1/T$ (Fig. 58.3), it is easy to see that the energy loss will be more than 4 dB at the feedback depth $R_{(4)}$ for the signal duration $T_s = 4\,T$ and more 8 dB at feedback depth $R_{(8)}$ for signal duration $T_s = 8\,T$. However, one should not conclude from this comparison that algorithms with decision feedback are ineffective. With an increase in the message transmission rate and an increase in ISI, the number of signals falling into the analysis interval increases. Therefore, to reduce the BER, it is necessary to increase the depth of the feedback.

Let us analyze the efficiency of the feedback algorithm with a further increase in the transmission rate. For RRC modulation, we use a pulse with a rounding parameter of the amplitude-frequency characteristic $|H(f)|^2$ filters $\beta = 0.3$ with a signal duration $T_s = 8\,T$. Figure 58.5 shows the detection results for the message transmission rate up to R $= 1.33/T$ and the feedback depth $R_{(4)}$.

As can be seen from the comparison of the curve in Fig. 58.5 for $R_{(4)}$ and the curve for $R_{(8)}$ in Fig. 58.4 b) with an increase in the transmission rate from R $=$ 1.25/T–R $= 1.33/T$, but with a decrease in the ISI level (see Fig. 58.1 for $\beta = 0.3$) it is even possible to improve the BER performance. The energy gain is about 2 dB in the range of error probabilities p $= 10^{-4}-10^{-5}$.

With a further increase in the transmission rate to the value R $= 1.66/T$, but with an increase in the parameter $\beta = 0.5$ (decrease in the ISI level), it becomes possible to obtain an increase in the detection quality when using the algorithm with feedback on the decision for $R_{(3)}$. In this case, it is possible to obtain an energy gain in relation to

Fig. 58.5 BER performance for an RRC pulse with $T_s = 8\,T$, $\beta = 0.3$ and $R = 1.33/T$
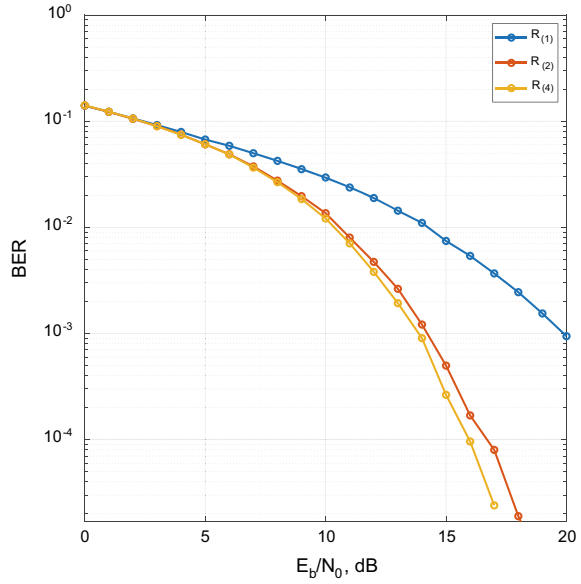
Fig. 58.6 BER performance for an RRC pulse with $T_s = 8\,T$, $\beta = 0.5$ and $R = 1.66/T$
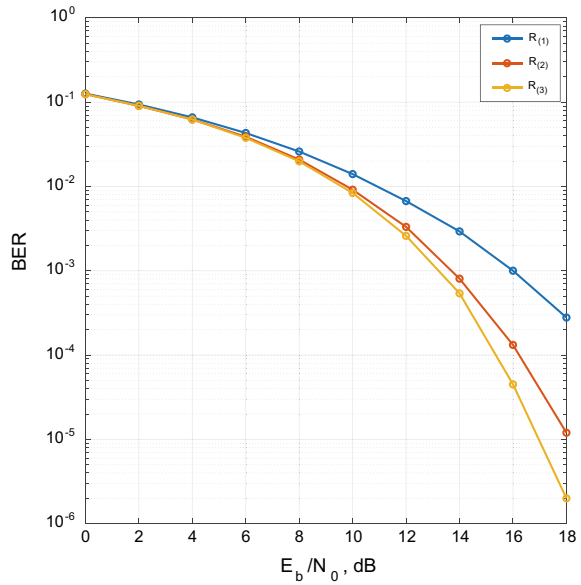
the previous figure (Fig. 58.5 for $R_{(4)}$) of about 1 dB in the same range of error probabilities. Thus, the region of applicability of the algorithm can be designated. Firstly, this detection algorithm can efficiently be applied in non-strong-ISI scenario (for example, corresponding to high ξ values in the cases considered in the simulation).

Secondly, the algorithm does not lead to a significant increase in the complexity of the detection for large pulse length L values (the complexity increases linearly with the L growing) and can efficiently implemented in computational devices. Thirdly, the given algorithm can be the core for further modification, for example, related to taking into account subsequent signals.

58.5 Conclusions

As a result of this work, the following is shown.

1. The application of the algorithm with decision feedback makes it possible to increase the BER performance of reception in comparison with the use of the algorithm without ISI compensation even at a relatively low transmission rate $R = 1/T$ for partially responded signals. The energy gain is from 1 dB to 2.5 dB in the range of error probabilities $p = 10^{-4} - 10^{-5}$.
2. An increase in the feedback depth gives a significant improvement in the BER performance, providing energy gains of up to 4–5 dB, in comparison with the case of using the feedback algorithm, but with one pass.
3. An Increase of More Than 60% in the Bit Rate When Using FTN Signals Based on RRC Pulses Does not Increase the Bit Error Rate When Using Algorithms with Decision Feedback.
4. With an increase in the symbol rate of message transmission, the application of algorithms with decision feedback provides energy gains, in comparison with the elementwise detection, which does not take into account the ISI, more than 10–12 dB in the range of error probabilities $p = 10^{-4} - 10^{-5}$.

References

1. J. Mazo, Faster-than-nyquist signaling. The Bell System Technical Journal **54**(8), 1451–1462 (1975)
2. J. Anderson, F. Rusek, V. Öwall, Faster-than-Nyquist signaling. Proc. IEEE **101**(8), 1817–1830 (2013)
3. A. Liveris, C. Georghiades, Exploiting faster-than-Nyquist signaling. IEEE Trans. Commun. **51**(9), 1502–1511 (2003)
4. J. Anderson, F. Rusek, Constrained capacities for faster-than-Nyquist signaling. IEEE Trans. Inf. Theory **55**(2), 764–775 (2009)
5. I. Lavrenyuk, A. Ovsyannikova, S. Zavjalov, S. Volvenko, S. Makarov, Improving energy efficiency of finite time FTN pulses detection by choosing optimal envelope shape, in *26th International Conference on Telecommunications 2019* (IEEE, Vietnam, 2019), pp. 289–294

6. J. Anderson, F. Rusek, A. Prlja, Receivers for faster-than-Nyquist signaling with and without turbo equalization, in *IEEE International Symposium on Information Theory 2008* (IEEE, Canada, 2008), pp. 464–468

7. A. Ovsyannikova, S. Zavjalov, S. Volvenko, On the joint use of turbo codes and optimal signals with increased symbol rate, in *10th International Congress on Ultra-Modern Telecommunications and Control Systems and Workshops 2018* (IEEE, Russia, 2018), pp. 1–4

8. A. Ovsyannikova, S. Zavjalov, S. Volvenko, Influence of correlation coefficient on spectral and energy efficiency of optimal signals, in *10th International Congress on Ultra-Modern Telecommunications and Control Systems and Workshops 2018* (IEEE, Russia, 2018), pp. 1–4

9. A. Ovsyannikova, S. Zavjalov, S. Volvenko, On the joint use of optimal faster than Nyquist signals with polar and turbo codes, in *International Conference on Electrical Engineering and Photonics 2018* (IEEE, Russia, 2018), pp. 79–82

10. A. Gelgor, T. Gelgor, New pulse shapes for partial response signaling to outperform faster-than-Nyquist signaling, in *International Conference on Electrical Engineering and Photonics 2019* (IEEE, Russia, 2019), pp. 144–148

11. A. Kislitsyn, A. Krylov, A. Rashich, Experimental evaluation of SEFDM Trellis based demodulator, in *International Conference on Electrical Engineering and Photonics 2020* (IEEE, Russia, 2020), pp. 110–113

12. A. Rashich, S. Gorbunov, Computational complexity analysis of SEFDM time and frequency domain equalizers, in *International Conference on Electrical Engineering and Photonics 2019* (IEEE, Russia, 2019), pp. 94–97

13. A. Rashich, S. Gorbunov, ZF equalizer and trellis demodulator receiver for SEFDM in fading channels, in *26th International Conference on Telecommunications 2019* (IEEE, Vietnam, 2019), pp. 300–303

14. A. Rashich, S. Gorbunov, Spatial receive diversity for SEFDM based system, in *42nd International Conference on Telecommunications and Signal Processing 2019* (IEEE, Hungary, 2019), pp. 265–268

15. A. Gelgor, V. Nguyen, Outperforming conventional OFDM and SEFDM signals by means of using optimal spectral pulses and the M-BCJR algorithm, in *26th International Conference on Telecommunications 2019* (IEEE, Vietnam, 2019), pp. 130–134

16. E. Bedeer, M. Ahmed, H. Yanikomeroglu, A very low complexity successive symbol-by-symbol sequence estimator for faster-than-Nyquist signaling. IEEE Access **5**, 7414–7422 (2017)

17. E. Bedeer, M. Ahmed, H. Yanikomeroglu, Reduced complexity optimal detection of binary faster-than-Nyquist signaling, in *International Conference on Communications 2017* (France, 2017), pp. 1–6

Chapter 59
Selection Features of Different Standard Optical Fiber in CCTV Fiber-Optics Systems

Yury E. Krivenko ⓘ **, Semen E. Logunov** ⓘ **, Maxim N. Davydov** ⓘ **, Elena I. Andreeva** ⓘ **, and Dmitriy P. Andreev** ⓘ

Abstract The high-quality video signal has a wide spectrum typically. Therefore, in order to transmit such a signal over long distances through an optical cable, low-loss, low-dispersion optical fiber is required. In fiber-optic video systems, as well as in optical communication systems, standard single mode optical fibers (SSMF, standard G 652) are usually used. One of the advantages of these fibers is the ability to use CWDM in a wide spectrum (from 1270 to 1610 nm). At the same time, non-zero dispersion fiber (NZDSF, standard G 655) is more optimal near the wavelength of 1550 nm. However, as studies have shown, these optical fibers have an increased sensitivity to bending. This fact can be used to traffic interception. In WDM-system, for example, CWDM, due to nonlinear effects in fiber the part of the optical power from the working channel can enter the adjacent channels. Computer simulation of the four-wave multiplexing (FWM) effect in CWDM-system has been carried out. It is shown that fiber-optics systems with SSMF have more protection from traffic interception than systems with NZDSF. To transmit a high-confidential video signal, special techniques, such as frequency modulation, can be used, or additional noise signals can be added.

Keywords Video systems · Security systems · Fiber-optic cables

59.1 Introduction

The use of the optical cable in the communication systems provides such advantages as information capacity, reliability, durability and protection [1–15]. To achieve these advantages, different types and standards of fibers are used. However, this does not exclude the interception of the transmitted information.

Y. E. Krivenko · E. I. Andreeva
The Bonch-Bruevich Saint-Petersburg State University of Telecommunications, St. Petersburg, Russian Federation

S. E. Logunov · M. N. Davydov · D. P. Andreev (✉)
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation
e-mail: andreev.dp@edu.spbstu.ru

The optimization of some system parameter is associated with difficulties in maintaining the proper level of other system parameters. One of the key issues in the design of a video surveillance system is the choice of the type of optical fiber as a direct signal transmission medium. The characteristics of the optical fiber must ensure the maximum of the transmission distance on the one hand, and the protection from the interception of the transmitted information on the other. To provide the transmission a high-quality video signal, it is necessary to transmit a wide- spectrum optical signal. Therefore, for long distance optical signal transmission, fibers with a low loss and low dispersion can be used, such as NZDSF—nonzero dispersion shifted fibers (standard G.655). In fiber-optic video surveillance systems, as well as in optical communication lines, standard optical fibers (SSMF, Standard Single Mode Fiber, standard G.652) are most often used. One of the advantages of these fibers is the ability to use the wavelength division multiplexing in a wide spectrum range (from 1270 to 1610 nm, CWDM). At the same time, the optimal parameters for loss and dispersion in the spectral window near the wavelength of 1550 nm are provided by NZDSF fibers. However, as studies have shown, these fibers have an increased sensitivity to bending. This can lead to additional losses when laying optical cables and an increased risk of unauthorized access. To mitigate these risks, the Bending Loss Insensitive Fiber (BLIF—fibers, standard G.657) can be used.

The purpose of this work was a comparative study of the fiber-optic system security against the interception of the transmitted information by such parameters as the fiber bend sensitivity and nonlinear effects in fibers.

59.2 Bend Loss Sensitivity of the Fibers.

The main methods of interception are connected with the formation of optical informative signals by outputting a part of the optical information signal, e. g., by removing a part of the radiation when the fiber is bended.

To study the effect of bending of an optical fiber on its characteristics, a stand was assembled, the block diagram of which is shown in Fig. 59.1.
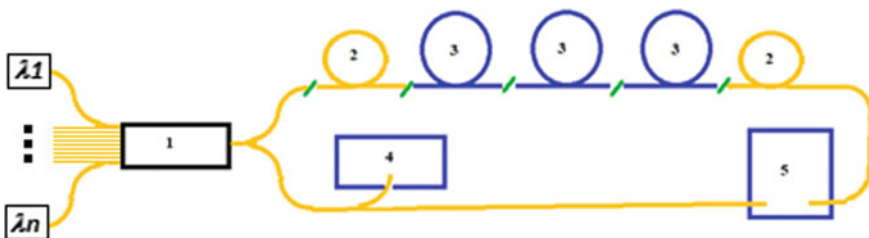
Fig. 59.1 Block diagram of the measuring setup: 1—WD-multiplexor, 2—measuring optical fiber, 3—tested optical fibers, 4—optical spectrum analyzer, 5—optical power meter
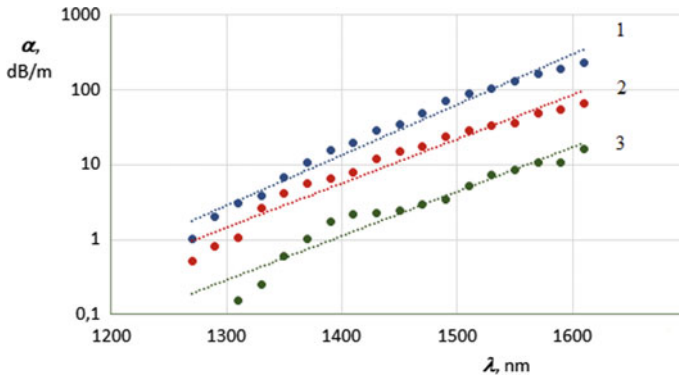
Fig. 59.2 The bend loss α in dependence from wavelength λ for NZDSF (G.655) (1), for SSMF (G.652) (2) and BLIF (G.657A2) (3) is presented. (The bend diameter is 12.5 mm.)

An experimental study of the level of bending losses was carried out by the substitution method. Testing was carried out in 2 stages. At the first stage, the spectral dependence of the level of introduced losses was recorded in a wide range of wavelengths: from 1270 to 1610 nm. CWDM SFP modules (Small Form-Factor Pluggable) were used as radiation sources. Registration of the optical signal was carried out by the differential method taking into account the spectral sensitivity of the photodetector [2]. Optical power measurements were carried out using a Rubin-300 stationary optical power meter. The source parameters were monitored using a Yokogawa AQ6370C optical spectrum analyzer.

Experimental investigation of the bend sensitivity of the different fiber is presented in Fig. 59.2. NZDSF has maximal bend sensitivity within studied fibers. So, this fiber has big risk of the information interception.

59.3 Nonlinear Effects in the Fibers

In WDM systems, part of the transmitted signal can fall into the adjacent spectral channel due to FWM (Four Wave Mixing). The simulation was carried out in the OptiSystem program. The optical signal was fed into the spectral channels of the CWDM grid: 1470 and 1490, 1530 and 1550 nm. After passing the effective loss manifestation length $L_{eff} = 20$ km, signals due to the original ones were observed in the neighboring spectral channels. The level of these signals is 53 … 55 dB lower than that transmitted in NZDSF and 55 … 60 dB lower than that transmitted in SSMF. This fact reflects the difference in the nonlinearity coefficient of these optical fibers (Table 59.1). Standard optical fibers are characterized by greater protection from unauthorized access then NZDSF.

Table 59.1 Nonlinear fiber coefficient

Fiber	A_{eff}, μm^2	n_2, 10^{-20} m^2/W$_T$	n_2/A_{eff}, 10^{-10} W^{-1}	$\gamma = (2\pi n_2)/(\lambda A_{eff})$, W^{-1} km^{-1}
SSMF	80	2.35	2.94	1.2
NZDSF	55	2.6	8.7	3.5

A_{eff}—the effective core area.
n_2—nonlinear refractive index.
Γ—nonlinearity coefficient (Fig. 59.3).

a)   b)
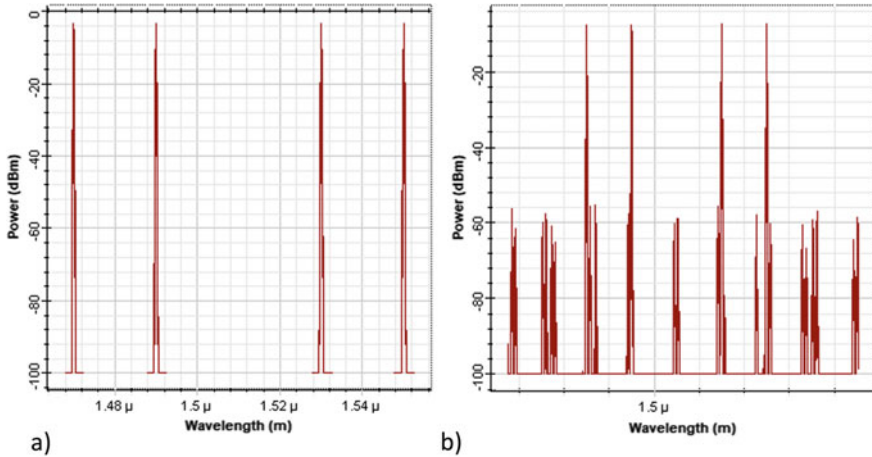
Fig. 59.3 Optical signal at input (**a**), and output (**b**) optical fiber with dispersion $D = 2$ ps/nm/km. Signal wavelength: $\lambda = 1470$, 1490, 1530 and 1550нм. Distance 20 km

59.4 Conclusion

It is experimentally confirmed that NZDS-fibers (with low dispersion) of the G.655-standard are characterized by increased sensitivity to bending, as result of which they require greater security to avoid interception of the transmitted signal traffic.

To transmit a high-privacy video signal, special modulation methods, such as frequency modulation [4], can be used, or additional noise signals can be added.

References

1. D. Andreev, E. Andreeva, A. Sergeev, V. Sumkin, Influence of optical cable bending in fiber optic systems video surveillance and subscriber access with spectral multiplexing. ICAT-2020 Proc., pp. 57–61 (2020)
2. E. Andreeva, V. Valyukhov, V. Kuptsov, Hight-quality video transmission over fiber-optic network with CWDM-technique. ICAT-2019 Proc., pp. 56–60 (2019)

3. V. Grishachev, Traffic interception in optical network: informative parasitic electromagnetic radiation. Photonics **13**(3), 280–294 (2019)
4. E. Andreeva, V. Valyukhov, V. Kuptsov, Fibre optics CCTV systems and security alarms. ICAT-2018 Proc., pp. 57–61 (2018)
5. V. Kuptsov, V. Valyukhov, K. Kopalin, V. Sumkin, Optical cable bending in fiber optic systems video surveillance and subscriber access with spectral multiplexing: A technique. ICAIT – Proc., pp. 68–72 (2020)
6. V.D. Kuptsov, V.P. Valjukhov, Sensitivity of photoreception devices fiber-optic communication lines. Comput. Telecommun. Control **11**(6), 31–36 (2010)
7. M. Tarasenko, V. Lenets, K. Malanin, N. Akulich, V. Davydov, Features of use direct and external modulation in fiber optical simulators of a false target for testing radar station. J. Phy. Conf. Series **1038**(1), 012035 (2018)
8. V. Fadeenko, V. Kuts, D. Vasiliev, V. Davydov, New design of fiber-optic communication line for the transmission of microwave signals in the X-band. J. Phys.: Conf. Series **1135**(1), 012053 (2018)
9. V. Grishachev, Photonics for security systems and data security. Photonics **5**(6), 58–63 (2011)
10. N. Grebenikova, V. Davydov, A. Moroz, M. Bylina, M. Kuzmin, Remote control of the quality and safety of the production of liquid products with using fiber-optic communication lines of the Internet. IOP Conf. Series: Mater. Sci. Eng. **497**, 012109 (2019)
11. V. Grishachev, Traffic intercept in optical network: method of optical tunneling. Photonics **14**(8), 280–294 (2020)
12. A.V. Moroz, R.V. Davydov, V.V. Davydov, A new scheme for transmitting heterodyne signals based on a fiber-optical transmission system for receiving antenna devices of radar stations and communication systems. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11660 LNCS, pp. 710–718 (2019)
13. A. Me'Ndez, T. Morze, Specialty optical fibers handbook. Academic Press, USA (2007)
14. M. Johnson, Optical fibre, cables and systems. ITU-T Manual (2010)
15. G. Agrawal, *Nonlinear Fiber Optics, 5th Academic Press* (San Diego, USA, 2013)

Chapter 60
Evaluation of the Transmission Efficiency of Multi-frequency Signals with an Unknown Initial Phase in the Meteor Burst Communication System

Ekaterina Mashkova ⓘ **, Sergey Zavjalov** ⓘ **, Sergey Volvenko** ⓘ **, and Wei Xue** ⓘ

Abstract Due to the intermittent nature of the meteor channel, the transmission is conducted at a fairly low speed. Usually meteor burst communication uses single-frequency signals, but multi-frequency signals can be used to increase the date rate. However, such signals have a disadvantage in the form of a high value of the peak-to-average power ratio of the emitted oscillations. To reduce this parameter, one can apply an amplitude limit on the transmitting module, which will increase the average signal power while maintaining the peak instantaneous power. It is interesting to consider a situation that is close to reality: the lack of information about the initial phase of the signal on the receiving module. This article discusses a variant of receiving OFDM and SEFDM signals with an unknown initial phase when transmitting in a meteor radio channel. It is proposed to introduce an initial phase estimation unit on the receiving module and make a corresponding change in the reference oscillation based on the records of the received signals. The obtained estimates of the system characteristics show that it is possible to reduce the PAPR by 4–6 dB when using multi-frequency signals without large losses in system performance. For the OFDM case, the measured delivery time and error probability are less than for the SEFDM signal transmission under the same conditions.

Keywords Meteor burst communications · Meteor channel · Multi-frequency signals · Initial phase · OFDM · PAPR

E. Mashkova (✉) · S. Zavjalov · S. Volvenko
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
e-mail: mashkova.ed@edu.spbstu.ru

W. Xue
College of Information and Communication Engineering, Harbin Engineering University, Harbin, China

60.1   Introduction

Meteor burst communication (MBC) has recently been actively developed, especially for prospective use on the Northern Sea Route. This type of radio communication has features associated with the intermittent nature of the appearance of meteors in the Earth's atmosphere. Falling into the Earth's atmosphere at a speed of several tens of kilometers per second, meteor particles heat up. As a result of thermal ionization, a trace of ionized gas is formed. This occurs at altitudes from 70 to 120 km [1]. The size of the resulting ionized trace depends on the mass and speed of the meteor. The trail can be tens of kilometers long. The possible radius is about a meter and it increases over time due to diffusion.

One of the disadvantages of using MBC is a rather low data transfer rate. Typically, MBC systems use single-frequency signals [2–5], but in order to increase the transmission rate, it makes sense to consider the transmission of multi-frequency signals. This article describes a research on working with OFDM and SEFDM signals [6–11]. But they, like everything else in this world, have a disadvantage: a high value of the peak-to-average power ratio (PAPR) [12, 13]. PAPR is the ratio of the maximum power of a signal P_{max} to its average power P_{avg}:

$$\text{PAPR} = \frac{Pmax}{Pavg} \tag{60.1}$$

High values of the PAPR can lead to the fact that the amplifiers aren't used to the maximum power, therefore, the efficiency of the amplifier stages is low [12–16]. In order to use them more effectively, the implemented model includes an amplitude limiter working together with a power amplifier [10, 17–20]. This method of reducing the PAPR is called clipping. In this case, we have two effects. On the one hand, the average signal strength increases, and the probability of error at a given distance at a fixed noise level decreases. On the other hand, at a certain value of the amplitude limit, the system becomes more subject to distortion, and the probability of error increases. If the PAPR control scheme is not applied, clipping may occur unintentionally in the system due to exceeding the saturation level of the power amplifier. Thus, there are some optimal values of the reduction value of PAPR, at which it is possible to minimize the probability of error. The clipping method is described in [12, 21–23], where the advantages and disadvantages of its use for various multi-frequency signals are shown in more detail.

MBC is used in radiolocation, radio navigation, communications and various tracking systems. In each of these systems, there is a problem of synchronization of processes. For coherent reception, it is necessary to compensate for the phase mismatch of the received signal and the signal of the reference generator used in the receiver. The study to which [24] is devoted, and on which this study is based, assumes that the initial phase of the received signal is precisely known in the receiving module. However, in reality, this information in noise conditions can't be accurately known and is subject to evaluation. The purpose of this work is to study the efficiency of

transmitting multi-frequency signals with an unknown initial phase by implementing in the receiving module a block for evaluating the initial phase from the records of the received signal. After that, the phase adjustment of the mixer of the reference generator in accordance with the estimated phase is assumed.

60.2   Simulation Model

The simulation is based on the use of a broadcast protocol that does not assume that the transmitting module has information about the success of data transmission. However, it is quite simple to implement, since the transmission is carried out continuously, regardless of the fact of the formation of the channel, so the stations operate in the prescribed modes without wasting time and energy on changing modes—from transmission to receiving and back. The simulation model is implemented in the Matlab R2018b software package and is presented in Fig. 60.1.

A sequence of bits of a given length is formed in the information source block. In the transmitting module modulation symbols are formed based on the received bits, and BPSK manipulation is used. $N = 8$ subcarriers are used for data transmission. The frequency difference between the subcarriers is given as α/T, where the parameter α determines the type of multi-frequency signal: for OFDM $\alpha = 1$, for SEFDM $\alpha < 1$. Further actions to limit and amplify occur with the signal in the time domain, so the IFFT is implemented. In the case of OFDM, the received signals from the IFFT output have the property of orthogonality. Multi-frequency signals with orthogonal frequency compaction provide high BER performance in the conditions of intercarrier interference due to the distribution of information over frequency channels in comparison with single-frequency signal systems. In order to increase the average signal power, the symbols go to the input of the pre-amplifier, and then they are fed to the input of the limiter. It sets the PR value by which we want to reduce the PAPR. Then the data is transmitted in the meteor channel, after which it is assumed to return to the analysis in the frequency domain.

After parallelizing the data streams, the input of the initial phase estimation unit receives a signal of the subcarrier frequency allocated specifically for phase adjustment. The frequency of this subcarrier is reported to the receiving module and is considered known. In the initial phase evaluation block, the frequency components of the received signal are determined. After that, the angle by which the phase of the element corresponding to the specified frequency of the subcarrier under consideration is rotated relative to the zero value is calculated. The phase of the generated reference oscillation is rotated by this angle. The resulting oscillation is fed to the receiver mixer. Further, the received information symbols are demodulated, which are transmitted using the remaining $(N-1) = 7$ subcarriers. After that, the received bits are sent to the error probability calculation block. If the information was transmitted without errors, the delivery time is calculated.

Specified simulation parameters for the signal and meteor trail:
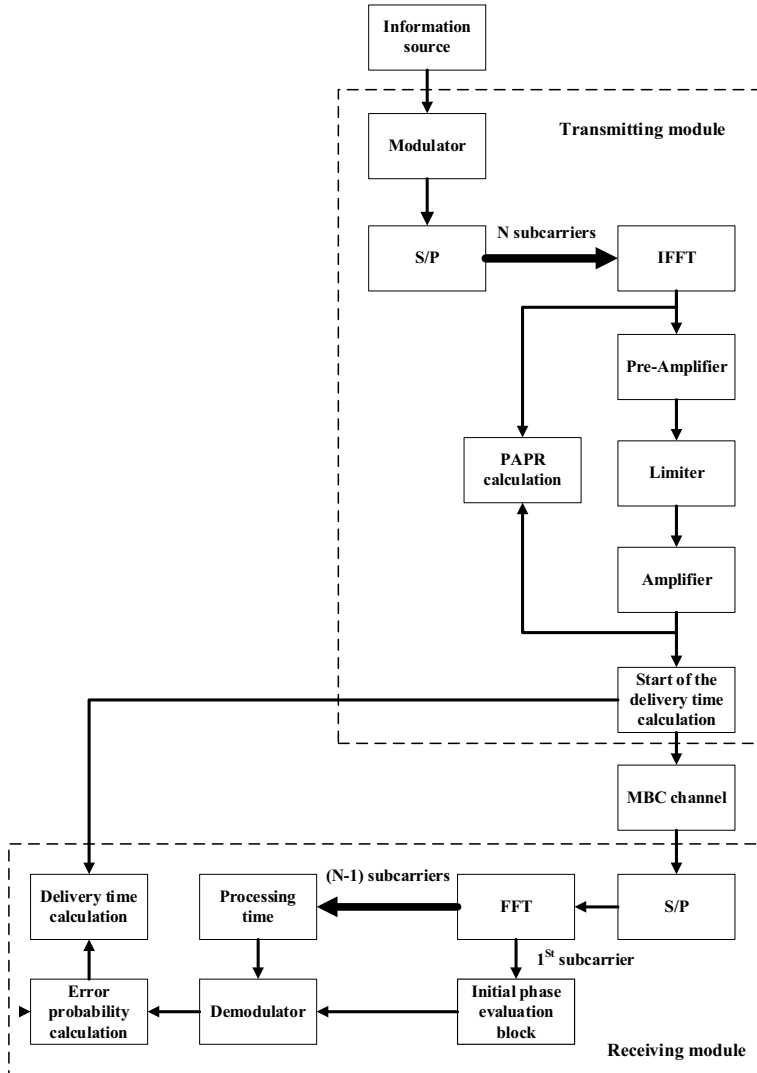
Fig. 60.1 Block diagram of the implemented model of the MBC system

- OFDM symbol duration: $T_{OFDM} = 10^{-4}$ s;
- number of symbols in one packet: $N_{OFDM} = 100$ symbols;
- accordingly, the duration of the package: $T = 0.01$ s;
- sampling frequency: $F_s = 10$ MHz;
- mean duration of a meteor trail: $T_c = 330$ ms;
- mean duration between two meteor trails: $\Delta T_c = 800$ ms.

In this case, we assume that the receiver knows the time of receipt of OFDM symbols.

60.3   Results

To check the adequacy of the model and the correct operation of the initial phase evaluation block, the BER performance was investigated (Fig. 60.2). Variants of multi-frequency OFDM ($\alpha = 1$) and SEFDM ($\alpha = 0.9$) signals without limiting the amplitude on the transmitting module are considered. It can be seen that the experimental data correspond to the behavior of the theoretical curve, however, in the case of SEFDM, the BER performance is worse, which is associated with the appearance of additional interference of signals from neighboring subcarrier frequencies.

Figure 60.3 shows the obtained dependences of the error probability on PR. It can be concluded that when transmitting multi-frequency signals with an unknown phase, it is possible to reduce the PAPR up to 4 dB without BER performance degradation. Further restriction will lead to an increase in the error probability, as the level of interference in the frequency domain between signals located at different subcarrier frequencies begins to increase, which in turn worsens the BER performance. Also a similar effect of increasing interference is observed with a decrease in the value of α.

Figure 60.4 shows the resulting dependence of the delivery time on PR. We see that for OFDM signals with an unknown initial phase, it is possible to reduce the PAPR by 6 dB and keep the delivery time almost unchanged. In the case of SEFDM signals,
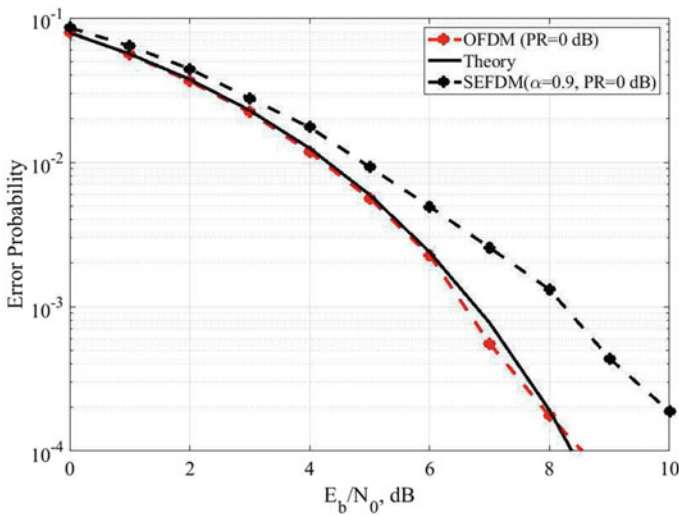
Fig. 60.2   BER performance in the AWGN channel
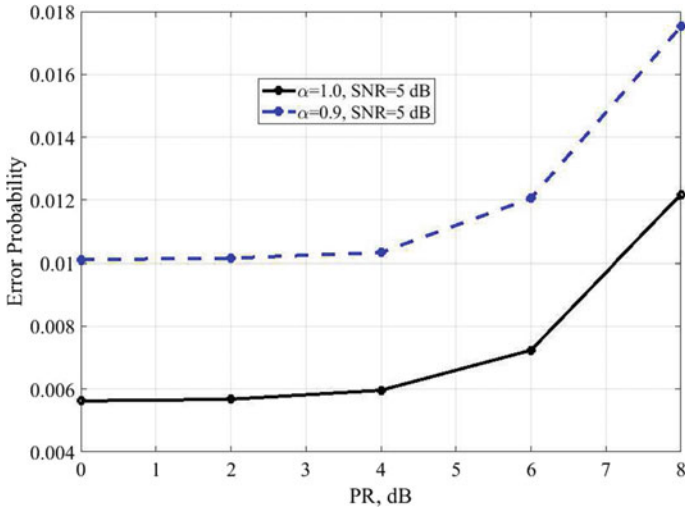
Fig. 60.3 Dependence of the error probability on the decrease in the peak-to-average power ratio
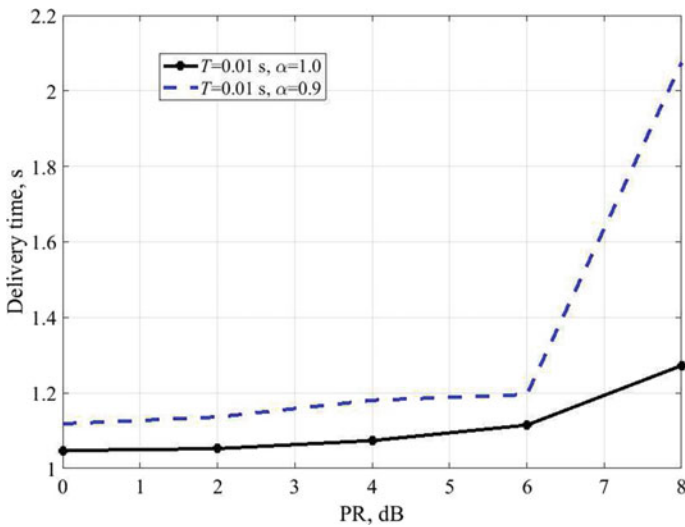
Fig. 60.4 Dependence of delivery time on the decrease in the peak-to-average power ratio when transmitting via the meteor channel

the delivery time is slightly longer, but the trend of the curve behavior remains, and it is also clear that a decrease in the PAPR by 4–6 dB won't lead to a significant increase in the delivery time. And so the existing interference between the subcarrier frequencies for SEFDM signals, coupled with additional distortions introduced in the power amplifier or during clipping, can significantly complicate the reception

of such signals, which affects the results both for the error probability and for the delivery time.

The article [26] on which this paper is based cites the results of such a study, but when received it is assumed that the initial stage is accurately known. The results obtained in [26] suggest that in the case of OFDM signals, when the PAPR decreases by 6 dB, the delivery time almost does not change, which is consistent with the results obtained. Also [26] suggests that it is possible to minimize the probability of an erroneous reception if you select the effective value of the PAPR limit. In the case of adding the initial phase evaluation block Fig. 60.3 doesn't have a specific minimum, but there is a "shelf" that allows you to make an estimate of possible values by which you can reduce the PAPR without significant losses in the accuracy of reception.

60.4   Conclusion

In this paper, we consider a model of a meteor communication system that allows us to obtain estimates of such parameters as the delivery time and BER when transmitting multi-frequency signals under conditions of an unknown initial phase. During the simulation, estimates of the delivery time for the cases of OFDM and SEFDM signals were obtained, which allow us to conclude that the transmission of OFDM signals under equal conditions takes less time. At the same time, it is noted that without large losses in system performance, it is possible to reduce the PAPR of the used multi-frequency signals by 4–6 dB, which can allow using the power amplifiers used in the system more efficient.

Further development of the model involves the introduction of clock and symbol synchronization at the receiver. Also, the implementation of the start of transmission at a random time after the appearance of the meteor trail will be close to real conditions. In addition, it is necessary to add control over the integrity and authenticity of the transmitted data.

References

1. A.I. Sulimov, A.V. Karpov, Simulation of frequency-selective properties of meteor scatter radio links, in *Russian Open Conference on Radio Wave Propagation (RWP)* vol. 1, pp. 488–491. (IEEE, July 2019), pp. 488–491. https://doi.org/10.1109/RWP.2019.8810184
2. P.A. Forsyth, E.L. Vogan, D.R. Hansen, C.O. Hines, The principles of JANET—a meteor-burst communication system. Proc. of the IRE. **45**(12), 1642–1657 (1957)

3. P. Bartholome, I. Vogt, COMET—a new meteor-burst system incorporating ARQ and diversity reception. IEEE Trans. Commun. Technol. **16**(2), 268–278 (1968)

4. G.L. Schaefer, SNOTEL: the world's first and largest data collection system using meteor burst technology, in *Proceeding of the Symposium on the Hydrological Basis for Water Resources Management* (Beijing, 1990), pp. 229–238

5. J. Weitzen, J. Larsen, R. Mawrey, Design of a meteor scatter communication-network for vehicle tracking, in *Vehicular Technology Society 42nd VTS Conference—Frontiers of Technology: From Pioneers to the 21st Century.* vol. 1(2). pp. 75–78 (1992)

6. V.I. Slyusar, K.A. Vasil'ev, Potential limits of frequency division multiplexing of N-OFDM signals based on Hartley's basis functions. Radioelectron. Commun. Syst. **51**(3), 129–133 (2008). https://doi.org/10.3103/S0735272708030035

7. V.I. Slyusar, V.G. Smolyar, Communication channels frequency multiplexing on the basis of superrayleigh signals resolution. Izvestiya Vysshikh Uchebnykh Zavedenij. Radioelektronika **46**(7), 30–39 (2003)

8. I. Lavrenyuk, E. Maksimova, Y. Sadovaya, Research on FBMC/OQAM spectral and energy characteristics for different prototype filters, in *2021 Springer Proceedings in Physics*, vol. 255 (2021), pp. 399–411. https://doi.org/10.1007/978-3-030-58868-7_45

9. I. Darwazeh, H. Ghannam, T. Xu, The first 15 years of SEFDM: a brief survey, in *2018 11th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP)* (Budapest, 2018), pp. 1–7. https://doi.org/10.1109/CSNDSP.2018.8471886

10. D.C. Nguyen, S.V. Zavjalov, A.S. Ovsyannikova, S.V. Volvenko, E. Smirnova, C.M. Nguyen, On application of hard-clipping and soft-clipping for SEFDM signal with optimal envelope, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12526 LNCS* (2020), pp. 174–184. https://doi.org/10.1007/978-3-030-65729-1_16

11. H. Ghannam, I. Darwazeh, SEFDM: spectral efficiency upper bound and interference distribution, in *2018 11th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP)* (18–20 July 2018/Budapest, Hungary, 2018), pp. 1–6

12. E.O. Antonov, A.V. Rashich, D.K. Fadeev, N. Tan, Reduced complexity tone reservation peak-to-average power ratio reduction algorithm for SEFDM signals, in *Proceedings of the 39th International Conference on Telecommunications and Signal Processing* (2016), pp. 445–448

13. L. Wang, C. Tellambura, An overview of peak-to-average power ratio reduction techniques for OFDM systems, in *2006 IEEE International Symposium on Signal Processing and Information Technology* (Aug 2006), pp. 840–845

14. V. Sorotsky, R. Zudov, Application of neural networks to reduce distortion of RF signals in switch mode power amplifiers, in *2020 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)* (St. Petersburg, Russia, 2020), pp. 76–79. https://doi.org/10.1109/EExPolytech50912.2020.9243973

15. A. Pergushev, V. Sorotsky, A. Ulanov, Criteria for selection envelope tracking power supply parameters for high peak-to-average power ratio applications, in *2019 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)* (St. Petersburg, Russia, 2019), pp. 13–16. https://doi.org/10.1109/EExPolytech.2019.8906793

16. A. Pergushev, V. Sorotsky, Signal distortion decreasing in envelope tracking power amplifiers in *2018 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)* (St. Petersburg, 2018), pp. 44–47. https://doi.org/10.1109/EExPolytech.2018.8564443

17. K.T. Nguyen et al., Oscillation characteristics in waveguide-based TWT amplifiers, in *2015 IEEE International Vacuum Electronics Conference (IVEC)*, pp. 1–2 (2015). https://doi.org/10.1109/IVEC.2015.7223810

18. E.N. Smirnova, S.B. Makarov, S.V. Zavjalov, B. Polozhintsev, Influence of the amplitude limitation of signals with the sin(x)/x envelope on the spectral and energy characteristics, in *Proceedings of the 2020 IEEE International Conference on Electrical Engineering and Photonics, EExPolytech N 9243855*, pp. 164–167 (2020). https://doi.org/10.1109/EExPolytech50912.2020.9243855

19. A. Chini, Y. Wu, M. El-Tanany, S. Mahmoud, Hardware nonlinearities in digital TV broadcasting using OFDM modulation. IEEE Trans. Broadcasting **44**(1), 12–21, (Mar 1998). https://doi.org/10.1109/11.713052

20. D.C. Nguyen, S.V. Zavjalov, A.S. Ovsyannikova, The Effectiveness of application of multi-frequency signals under conditions of amplitude limitation, in *lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), 11660 LNCS* (2019), pp. 681–687. https://doi.org/10.1007/978-3-030-30859-9_59

21. S.C. Thompson, J.G. Proakis, J.R. Zeidler, The effectiveness of signal clipping for PAPR and total degradation reduction in OFDM systems, in *Global Telecommunications Conference, GLOBECOM 2005*, vol. 5 (IEEE, 2005), pp. 2811

22. S.I.A. Ahmed, Spectrally efficient FDM communication signals and transceivers: design, mathematical modelling and system optimization, in *A Thesis Submitted for the Degree of Doctor of Philosophy*, (2011), pp 182–187

23. D.K. Fadeev, A.V. Rashich, Optimal input power backoff of a nonlinear power amplifier for SEFDM system, in *Proceedings of the NEW2AN 2015 and 8th Conference* (Aug 2015), pp. 669–678

24. D.C. Nguyen, S.V. Zavyalov, S.V. Volvenko, Transmission efficiency of multi-frequency signals in MBC using amplitude limitation on the transmitting module. Comput. Telecommun. Control **13**(1), 42–52 (2020). https://doi.org/10.18721/JCSTCS.13104

Chapter 61
Hardware Architectures of Multipliers in Binary Galois Fields with Characteristic Less Than 1024

Valerii Perhun, Andrey Rashich, and Tan Ngoc Nguyen

Abstract The multiplication in Galois field is actively used on various steps of algebraic decoders of BCH codes. The short codes are usually used as the component codes of more complex structures, like Open FEC. The decoding of the overall construction is usually iterative and thus requires very fast decoders of component codes. The fast hardware architectures of multipliers for short code decoders are considered in this paper and carefully compared in terms of gates consumption and critical path length for various field sizes. The best architecture candidates are proposed depending on the field size and requirements.

Keywords Multiplier · Binary Galois field · Polynomial · Karatsuba · Hybrid · Architecture · Implementation · Gates · Critical path

61.1 Introduction

The algebraic decoding, which is used to decode cyclic block codes, especially BCH codes, involves multiplication in Galois field operation on various steps of decoding procedure [8]. The short codes (of length less than 1024) are usually used not by themselves but in the complex coding structures, like Open FEC. In such cases very fast decoder of inner codes architectures are needed, which in turn arise the problem of fast and simple (in terms of gates number) blocks of basic operations. The multiplication is one of most demanding. The main ideas of this paper are to find the best algorithms of multiplication in terms of gates number and critical path length or to develop new architectures of multipliers like hybrid multiplication architecture based on the Karatsuba algorithm and direct multiplication algorithm in binary Galois field with characteristic less than 1024. The important note here is that the architectures are to be parallel (providing maximum throughput), because they

V. Perhun (✉) · A. Rashich · T. N. Nguyen
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

A. Rashich
e-mail: rashich@cee.spbstu.ru

are to be used in component codes decoders, which appear in the iterative decoding of Open FEC.

Multipliers in binary Galois Field can be implemented in different ways: using logarithm and anti-logarithm tables, full multiplier approach or using various ingenious algorithms, which can reduce the number of gates or critical path or both (e.g.: Schönhage-Strassen, Toom-Cook, Karatsuba, Montgomery, Massey-Omura multipliers [1–9]). Also, one can combine different algorithms and get hybrid multiplication schemes in $GF(2^m)$, where m is the extension of the original field.

The paper is organized as follows. In the Sect. 61.2 the architecture of the full multiplier is described, its complexity and critical path length are considered. Section 61.3 presents the Karatsuba algorithm and hybrid multiplier architecture. Section 61.4 describes the multiplexer-based multiplier architecture on logarithm and anti-logarithm tables. Section 61.5 dedicated to normal basis multiplier, its complexity and critical path length are calculated. Section 61.6 presents a comparison of multiplier architectures in terms of gates number and critical path length.

61.2   Full Multiplier

Consider two elements in binary Galois field $GF(2^m)$ represented by a polynomial basis $\{1, \alpha, \alpha^2, \ldots, \alpha^{m-1}\}$, which are presented in general form:

$$b_{m-1}\alpha^{m-1} + b_{m-2}\alpha^{m-2} + \cdots + b_1\alpha + b_0 \tag{61.1}$$

$$c_{m-1}\alpha^{m-1} + c_{m-2}\alpha^{m-2} + \cdots + c_1\alpha + c_0 \tag{61.2}$$

Three steps are needed to get the multiplication result using full multiplier approach. The first step is the pairwise multiplication of all coefficients of the first polynomial by all coefficients of the second. Next step is the addition of those pairwise multiplication results, which correspond to the same power of x. The result of such operation is the following polynomial (not simplified):

$$(b_{m-1}c_{m-1})\alpha^{2m-2} + (b_{m-1}c_{m-2})\alpha^{2m-3} + (b_{m-2}c_{m-1})\alpha^{2m-3} + \cdots + b_0c_0 \tag{61.3}$$

The final step is the polynomial modulo reduction using the field irreducible polynomial. This operation is done using Euclidean algorithm for polynomials division [10].

All calculations are performed in binary Galois fields, therefore all coefficients at each step take values 0 or 1, multiplications are logical AND and addition is done modulo 2 (XOR).

In general, the polynomial in (61.3) has m^2 terms. It is then simplified to $2m - 1$ terms at the second step.
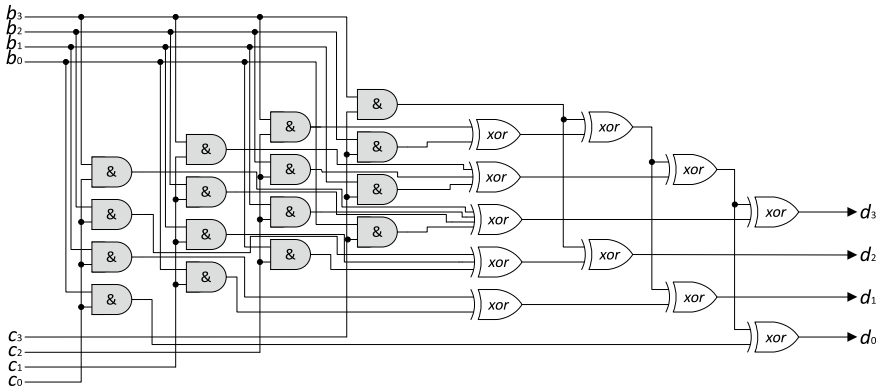
Fig. 61.1 Full multiplier in $GF(2^4)$ with primitive polynomial $f(x) = x^4 + x^3 + 1$

The example architecture of such full multiplier for $GF(2^4)$ with primitive polynomial $f(x) = x^4 + x^3 + 1$ is shown in Fig. 61.1.

The multiplication of each coefficient in $GF(2^m)$ is done by m^2 logical ANDs. The coefficients additions at the second step require $(m-1)^2$ XORs. Euclid algorithm requires $(m-1)(k-1)$ XOR gates, where k is the number of non-zero coefficients in primitive polynomial. In total full multiplier requires $m^2 + (m-1)^2 + (m-1)(k-1)$ logic gates.

The critical path length consists of the following terms: 1 is the parallel multiplying of coefficients, $\lceil \log_2 m \rceil$ is the modulo two addition at the second step, $(m-1)$ is a reduction.

Note that multi-input XORs are considered as trees of two-input XORs.

61.3   Hybrid Karatsuba Multiplier

The Karatsuba algorithm of multiplying two elements in $GF(2^m)$ is based on multipliers of half bit width numbers.

The Karatsuba algorithm is as follows [7]. Let us consider the multiplication of two elements (61.1) and (61.2). They can be represented as m bit-width vectors:

$$A = a_{m-1} a_{m-2} \ldots a_1 a_0 \tag{61.4}$$

$$B = b_{m-1} b_{m-2} \ldots b_1 b_0 \tag{61.5}$$

Each vector can be represented as a sum of two terms, each has $n = m/2$ digits:

$$A_0 = a_{n-1} a_{n-2} \ldots a_0$$

$$A_1 = a_{m-1}a_{m-2}\ldots a_n$$
$$A = A_0 + A_1 * S^n \tag{61.6}$$

In (61.6) S is the number base, for $GF(2^m)$ it is 2:

$$B_0 = b_{n-1}b_{n-2}\ldots b_0$$
$$B_1 = b_{m-1}b_{m-2}\ldots b_n$$
$$B = B_0 + B_1 * 2^n \tag{61.7}$$

Note that multiplication by 2^n in (61.6) and (61.7) is just a shift and requires no resources. Denote this shift operation as *.

If m is odd or vectors (61.4) and (61.5) have different numbers of digits in their representation, zeros are added at the most significant positions of the input vectors in such way, that the number of digits is the same.

Let multiply AB:

$$\begin{aligned}
AB &= (A_0 + A_1 * 2^n)(B_0 + B_1 * 2^n) \\
&= A_0B_0 + A_0B_1 * 2^n + A_1B_0 * 2^n + A_1B_1 * 2^{2n} \\
&= A_0B_0 + (A_0B_1 + A_1B_0) * 2^n + A_1B_1 * 2^{2n}
\end{aligned} \tag{61.8}$$

The multiplication in (61.8) requires 4 multipliers of n-digit numbers. Karatsuba proposed the approach, which requires only 3 multipliers. Consider the following representation of $(A_0B_1 + A_1B_0)$:

$$(A_0B_1 + A_1B_0) = (A_0 + A_1)(B_0 + B_1) - A_0B_0 - A_1B_1 \tag{61.9}$$

All operations are calculated in binary Galois field; therefore, the subtraction is equivalent to addition. Note that addition here is just the XOR operation. The result is the following:

$$AB = A_0B_0 + ((A_0 + A_1)(B_0 + B_1) - A_0B_0 - A_1B_1) * 2^n + A_1B_1 * 2^{2n} \tag{61.10}$$

At this point one can recursively reapply Karatsuba algorithm for each multiplication in (61.10) so that finally all operations are performed on single-bit elements. But for ASIC or FPGA implementation and for fields with $m \leq 10$ it is better to use the following hybrid approach. At first apply the Karatsuba algorithm on the input operands. It will turn out to 3 multipliers of $m/2$-digit numbers. Then perform full multiplication for them.

The multiplication result is to be reduced modulo the field irreducible polynomial as for the full multiplier architecture.

The example of such multiplier for $GF(2^4)$ without Euclidean algorithm is shown in Fig. 61.2.
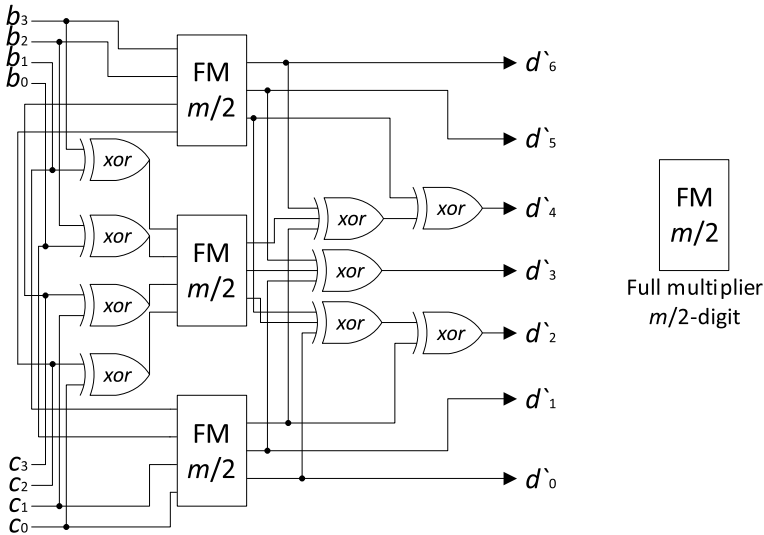
Fig. 61.2 The architecture of hybrid Karatsuba multiplier for $GF(2^4)$

The number of gates for this scheme is defined by the following: three full multipliers of $m/2$-digit, which require $3((m/2)^2 + ((m/2) - 1)^2)$ gates, two bitwise XORs of $m/2$-bit numbers, m gates, three bitwise XORs of $(m - 1)$-bit numbers, $3(m - 1)$ gates. Note that during the last operation the shifted versions of input vectors are XORed, thus, in one position there is no XOR operation, the input bit is just copied into the result. And the last step is Euclidean algorithm. Thus hybrid scheme requires $1.5m^2 + m - 1 + (m - 1)(k - 1)$ gates.

The critical path length of proposed hybrid architecture is $5 + \lceil \log_2(m/2) \rceil + (m - 1)$. The critical path length consists of the full multiplier of $m/2$-bit inputs, the XORs in Karatsuba algorithm (the term $(A_0 + A_1)(B_0 + B_1)$ in (61.10)) and Euclidean algorithm as in the full multiplier architecture.

61.4 Logarithm, Anti-Logarithm Table-Based Multiplier

Each finite field has a primitive element—α [11]. The powers of primitive element generate all non-zero elements of this field. Thus, the multiplication of two input elements is the addition of powers modulo $2^m - 1$. The conversion from polynomial representation to the powers of primitive element and back can be done using precomputed tables.

These tables may be implemented using ROMs. The drawback of such approach is that it takes at least one clock cycle to extract an element from memory. The better way is to implement tables using multiplexers. In this case the multiplexer address
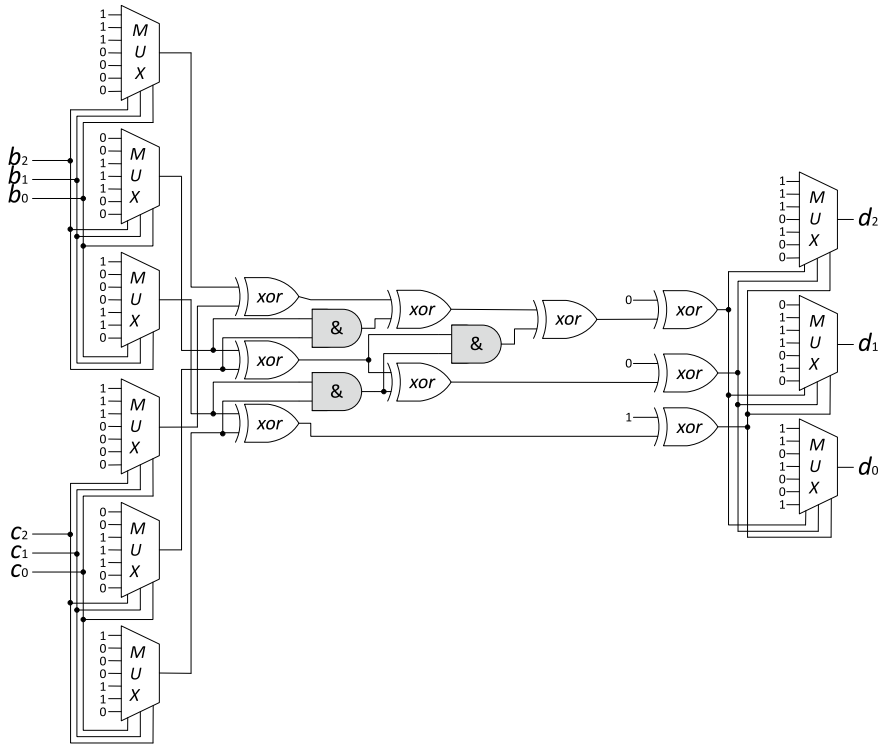
Fig. 61.3 Scheme of table-based multiplier in $GF(2^3)$, $f(x) = x^3 + x + 1$

input receives input element in polynomial representation, data inputs are connected to constants from table. The reverse conversion is done in the same way.

The example of such multiplier for $GF(2^3)$ with primitive polynomial $f(x) = x^3 + x + 1$ is shown in Fig. 61.3.

The multiplexer occupies $3(2^m - 2)$ gates, but this value is an upper bound estimate, because during synthesis multiplexer will be simplified. This is due to the fact that data inputs are connected to constants.

The addition and modulo reduction operations require $6m - 6$ gates. There are $3m (2^m - 1):1$ multiplexers in scheme. The total number of gates is $9m(2^m - 2) + 6m - 6$.

The critical path length is $2m + (m + 1) + 2m$. The first and the last terms correspond to conversion between polynomial and power representations. The second term is addition modulo $2^m - 1$.

61.5   Normal Basis Multiplier

Consider two elements in binary Galois field $GF(2^m)$ represented by a normal basis $\{\beta, \beta^2, \beta^4, \ldots, \beta^{2^{m-1}}\}$, which are presented in general form:

$$b_{m-1}\beta^{2^{m-1}} + b_{m-2}\beta^{2^{m-2}} + \cdots + b_1\beta^2 + b_0\beta \tag{61.11}$$

$$c_{m-1}\beta^{2^{m-1}} + c_{m-2}\beta^{2^{m-2}} + \cdots + c_1\beta^2 + c_0\beta \tag{61.12}$$

The multiplication of two elements is done by parallel Massey-Omura multiplier [12–14]. The product of (61.11) and (61.12) is

$$d_{m-1}\beta^{2^{m-1}} + d_{m-2}\beta^{2^{m-2}} + \cdots + d_1\beta^2 + d_0\beta \tag{61.13}$$

From (61.13) and the fact that β is the root of the irreducible polynomial, one obtains the result of multiplication for each d_i.

The multiplication is done in the following steps. At the first step all the coefficients of the input polynomials are multiplied pairwise as in the full multiplier. Further the resulting coefficients, which correspond to the same power, are XORed. Then due to the fact that β is the root of the irreducible polynomial the result of multiplication in normal basis is obtained.

The example architecture of such multiplier for $GF(2^4)$ with primitive polynomial $f(x) = x^4 + x^3 + 1$ is shown in Fig. 61.4.

The normal basis multiplier requires $m^2 + 0.5(m^2 - m) + m2^{m-2}$ gates, where m^2 gates are required for the pairwise multiplication of all input coefficients and $0.5(m^2 - m)$ gates correspond to XORs of coefficients with the same power and
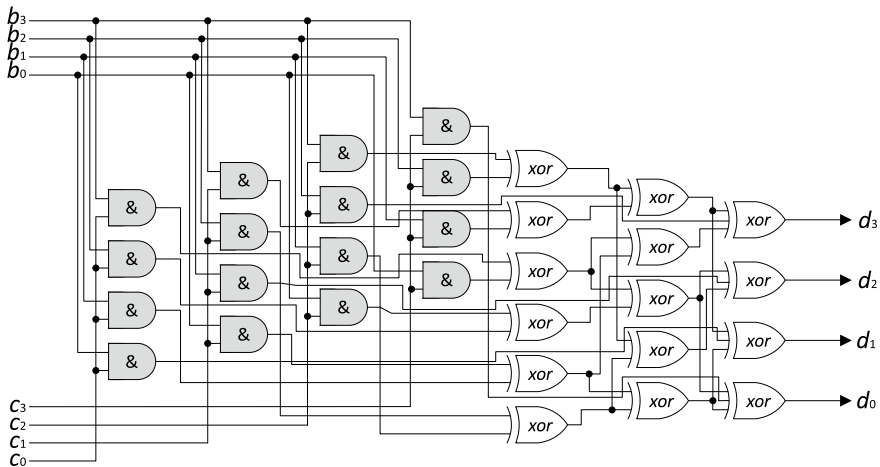
Fig. 61.4   Normal basis multiplier in $GF(2^4)$ with primitive polynomial $f(x) = x^4 + x^3 + 1$

2^{m-2} gates are required be the XORs for one digit of the output vector, there are m digits in total.

The critical path length is $1 + 1 + (m - 1)$. The first term is parallel multiplication of each coefficients and the second term is the modulo two addition of the coefficients with same power and the last term is XORs.

Note, that the number of gates and critical path can be reduced. The architecture improvement is based on a reformulation of the multiplication as in [13] or on other ways of multiplication of two elements in normal basis [15].

61.6   Comparison of Multiplier Architectures

Table 61.1 shows how many gates requires each algorithm of multiplication for $m < 10$. The Table 61.2 represent the length of critical path for each multiplier for $m < 10$.

The critical path of normal basis multiplier is always less than of other schemes for $m > 2$. But hybrid scheme requires less gates compared to full multiplier starting from $m = 6$. Scheme with tables always require much more gates but this is an upper bound estimate. Such scheme better suits for implementation in signal processor rather than in ASIC/FPGA.

Table 61.1 Gates number for various types of multipliers and m values

m	Full	Hybrid	Table	Normal
2	7	9	42	7
4	31	33	522	38
6	71	69	3378	147
8	141	131	18,330	604
10	199	177	92,034	2705

Table 61.2 Critical path length for various types of multipliers and m values

m	Full	Hybrid	Table	Normal
2	3	6	11	3
4	6	9	21	5
6	9	12	31	7
8	11	14	41	9
10	14	17	51	11

61.7 Conclusions

In this paper a number of architectures for multiplier in binary Galois fields of size not greater than 1024 were discussed. They are direct multiplication scheme, Karatsuba and hybrid Karatsuba algorithm, normal basis multiplier and logarithm and anti-logarithm table-based scheme. These schemes were compared in terms of gates consumption and critical path length for various field sizes. The hybrid approach requires fewer gates than the full multiplier scheme for $m > 6$ but has greater critical path. The normal basis multiplier has the shortest critical path among other approaches for $m > 2$ considered field sizes. The table-based approach loses to all others in each term, although the before-synthesis-optimization estimates were shown. This algorithm suits better for software implementation rather than hardware.

References

1. T. Steinke, R. Sainuddin, A rigorous extension of the Schönhage–Strassen integer multiplication algorithm using complex interval arithmetic, in *Computability and Complexity in Analysis*, ed. by X. Zheng, N. Zhong. (CCA 2010), EPTCS 24 (2010), pp. 151–159
2. K. Millar, M. Lukowiak, S. Radziszowski, Design of a flexible Schönhage–Strassen FFT polynomial multiplier with high-level synthesis to accelerate HE in the cloud, in 2019 International Conference on ReConFigurable Computing and FPGAs (ReConFig), (Cancun, Mexico, 2019), pp. 1–5
3. M. Bodrato, Toward optimal Toom-Cook multiplication for univariate and multivariate polynomials in characteristic 2 and 0, in *WAIFI*, ed. by C. Carlet, B. Sunar, LNCS 4547 (Springer, Heidelberg, 2007), pp. 116–133
4. J. Ding, S. Li, Z. Gu, High-speed ECC processor over NIST prime fields applied with toom-cook multiplication. IEEE Trans. Circuits Syst. I Regul. Pap. **66**(3), 1003–1016 (2019)
5. S. Vladimirov, *Matematicheskie Osnovy Teorii Pomekhoustojchivogo Kodirovaniya* (The Bonch-Bruevich Saint-Petersburg State University of Telecommunications, Saint-Petersburg, Russia, 2016)
6. N. El Mrabet, A. Guillevic, S. Ionica, Efficient multiplication in finite field extensions of degree 5, in *Progress in Cryptology—AFRICACRYPT 2011* ed. by A. Nitaj, D. Pointcheval. AFRICACRYPT 2011. Lecture Notes in Computer Science, vol 6737 (Springer, Berlin, 2011)
7. Z. Gu, S. Li, A division-free Toom-Cook multiplication-based montgomery modular multiplication. IEEE Trans. Circuits Syst. II Express Briefs **66**(8), 1401–1405 (2019)
8. K. Okamoto, N. Homma, T. Aoki, Formal design of arithmetic circuits over galois field based on normal basis representation. IEICE Trans. Information Syst. **97**, 2270–2277 (Maruzen Co., Ltd.) (2014)
9. C. Kizikale, O. Egecioglu, C.K. Koc, A matrix decomposition method for optimal normal basis multiplication. IEEE Trans. Comput. **65**(11), 3239–3250 (2016)
10. R. Gallager, Information Theory and Relaible Communication (Wiley, USA 1968)
11. R. Morelos-Zaragoza, *The Art of Error Correcting Coding* (Wiley, USA, 2002)
12. Wang, Troung, Shao, Deutsch, Omura, Reed, VLSI architecture for computing multiplications and inverses, in $GF(2^m)$. IEEE Trans. Comput. **C-34**(8), 709–717 (1985)
13. A. Reyhani-Masoleh, M.A. Hasan, A new construction of Massey-Omura parallel multiplier over $GF(2^m)$. IEEE Trans. Comput. **51**(5), 511–520 (2002)
14. B. Rashidi, S.M. Sayedi, R.R. Farashhi, An efficient and high-speed VLSI implementation of optimal normal basis multiplication over $GF(2^m)$. VLSI J. **55**, 138–154 (2016)

15. O. Eğecioğlu, Ç.K. Koç, Reducing the complexity of normal basis multiplication, in *Arithmetic of Finite Fields* ed. by Ç. Koç, S. Mesnager, E. Savaş. WAIFI 2014. Lecture Notes in Computer Science, vol 9061 (Springer, Cham, 2015)

Chapter 62
On the Possibility of Application of Optimal Signals with Restriction on Energy Concentration on SDR HackRFOne

Aleksandr V. Zhila, Anna S. Ovsyannikova, and Wei Xue

Abstract The article presents a method for optimizing single carrier signals with the criterion of a given energy concentration and restrictions on energy and peak-to-average power ratio (PAPR). The method of numerical solution of the optimization problem is shown. The influence of constraints on the solution of the optimization problem is considered. An experimental study was conducted on the basis of the HackRFOne SDR platform.

62.1 Introduction

Modern communication systems require the development of information transfer methods. One of the promising directions for systems of the "Beyond 5G" and "Pre-6G" classes is the application of optimal signals, which can be obtained as a result of solving an optimization problem with various optimality criteria: the reduction rate of out-of-band emissions, energy concentration in bandwidth, signal-to-noise ratio at the receiving point and many others. The choice of optimality criterion depends on the specific scenario of applicability [1–5].

In this paper, we will consider a method for optimizing single carrier signals according to the criterion of a given energy concentration in a frequency band. The relevance of such a criterion is explained by the peculiarities of the construction of information transmission systems. Since information transmission system contains filters, it can accordingly work with signals in a limited frequency band. Thus, an

A. V. Zhila (✉) · A. S. Ovsyannikova
Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

W. Xue
College of Information and Communication Engineering, Harbin Engineering University, Harbin, China

increase in the concentration of energy in a given frequency band can lead to an increase in the quality of information processing.

Further in the article, the optimization method is complicated by adding a limitation on the peak to average power ratio (PAPR) of the radiated oscillations. The value of PAPR is calculated as the ratio of the peak instantaneous power P_p to the average power P_{av} of the radiated oscillations. The problem of a high value of PAPR leads to underutilization of power amplifiers and an increase in the level of nonlinear distortions in the signal, which, in turn, leads to an increase in the level of out-of-band emissions [6–9]. Because of this, the average signal power is reduced and the BER performance degrades [10, 11]. Thus, a decrease in the PAPR is undoubtedly in demand, especially in the promising systems "Beyond 5G" and "Pre-6G" [12–14].

There are various methods of reducing the PAPR value, such as Selective Mapping (SLM), Tone Reservation (TR), Tone Injection (TI) [15]. All of these methods have drawbacks such as signal distortion, lower transmission rates, and high computational complexity. The method for solving the optimization problem proposed in this article allows avoiding the above-mentioned drawbacks, since the limitation of the PAPR of the radiated oscillations is proposed even at the stage of signal waveform synthesis.

The purpose of this article is to analyze the possibility of obtaining the forms of optimal signals with the criterion of optimality of a given energy concentration in a fixed frequency band and in the presence of restrictions on the PAPR of the radiated oscillations.

When synthesizing optimal signals, a search is performed for the function $a(t)$, which describes the envelope of signals with duration T_s and provides the minimum of the following functional:

$$
\arg\left\{\min_{a(t)}(J)\right\}, J = BEC - BEC_d, BEC
$$
$$
= \int_{-\Delta F/2}^{\Delta F/2} \left| \int_{-T_s/2}^{T_s/2} a(t)\exp(-j2\pi ft)dt \right|^2 df \qquad (62.1)
$$

where BEC is the initial energy concentration in the ΔF band, BEC_d is the desired concentration of energy in the ΔF band, set as a certain percentage (90%, 95%, 99%).

Note that analytical solutions to problems of this class can be found only for a small set of input parameters without additional restrictions. The issues of reducing the value of the PAPR lead to the complication of the nature of the optimization problem. For this reason, we will choose a variant of the numerical solution, in which we use the expansion of the function $a(t)$ in a bounded Fourier series.

It is convenient to consider the limiting condition on the PAPR of the radiated oscillations in the form of the following inequality:

$$
PAPR < (PAPR^* - \Delta PAPR), \qquad (62.2)
$$
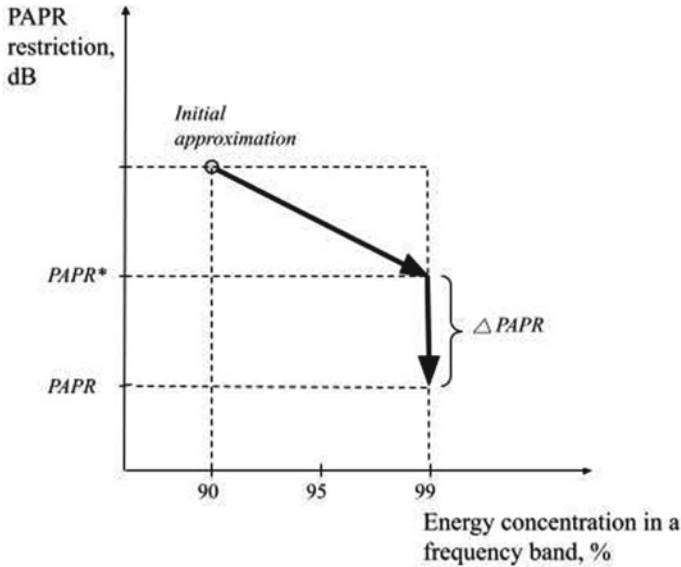
Fig. 62.1 A graphical representation of the optimization procedure

where PAPR* is the initial value of the PAPR of the radiated oscillations. As a result of solving the optimization problem with such a limitation, an envelope shape will be obtained that will reduce the PAPR of the radiated oscillations by ΔPAPR dB relative to the original signal. Also, the PAPR of the pulse sequence will be equal to the PAPR of one pulse, since there is no interference and the modulation used is BPSK.

Figure 62.1 shows a graphical view of the procedure for solving the optimization problem. On it, along the axes of a two-dimensional graph, the selected values of the energy concentration in a fixed band and the limitation on the PAPR are plotted.

The initial approximation for solving the optimization problem is the shape of the envelope, which was obtained in the course of optimization with the criterion for ensuring a given decay rate of the level of out-of-band emissions. At the first step of the optimization problem, restrictions are introduced on the energy concentration in a given frequency band. This changes the shape of the envelope $a(t)$. With an increase in the energy concentration, the value of the PAPR of the radiated oscillations decreases. In the second step, a limitation on the PAPR is introduced. The result of solving the optimization problem by a numerical method will be m coefficients of the expansion of the envelope $a(t)$ in the Fourier series $\{a_k\}$.

The sequence of solutions to the optimization problem is shown in Fig. 62.2. Initially, the required solution accuracy is set and the initial approximation is set. It should be noted that the choice of the initial approximation can significantly affect the final result of the algorithm. The required solution accuracy is determined by the choice of the number of expansion coefficients m of the envelope $a(t)$ in a
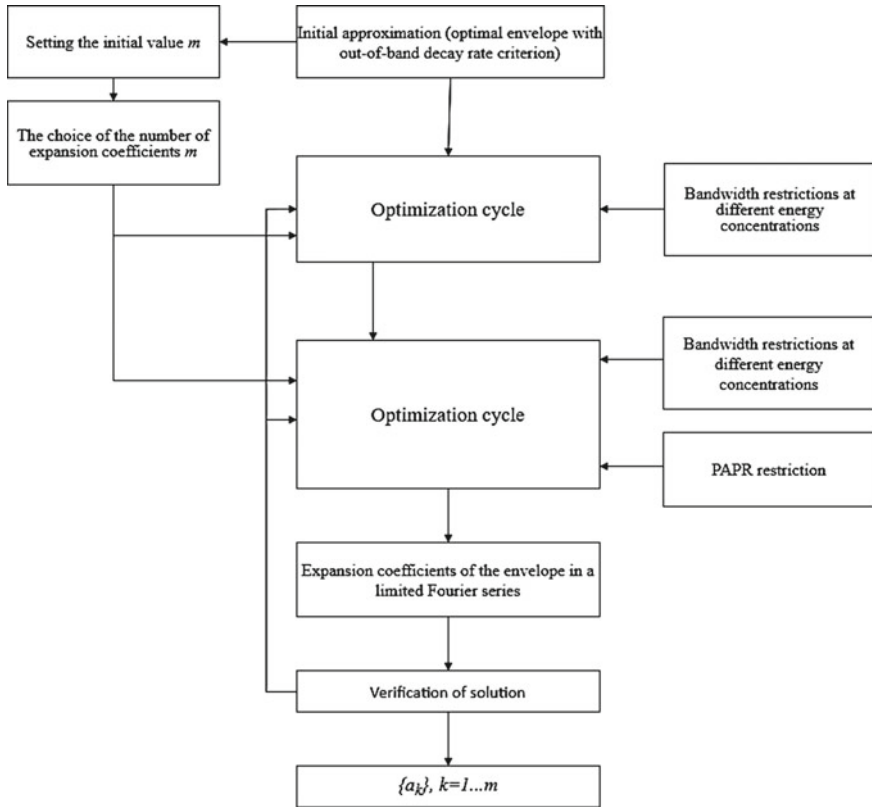
Fig. 62.2 Block diagram of the optimization problem

bounded Fourier series. The value m, the initial approximation and the constraints of the optimization problem are the input parameters of the optimization algorithm (Fig. 62.2).

At the output of the optimization block, the coefficients of the expansion of the envelope $a(t)$ into a limited Fourier series are calculated. The resulting solution accuracy is assessed. If necessary, the value of m is increased and the optimization algorithm is restarted.

Also, one of the main stages of work was an experiment based on the HackRF One SDR platform [16]. This technology allows us to change its settings very easily and to use it for research of different forms of optimal signals. HackRF One has a wide frequency range from 1 MHz to 6 GHz, a bandwidth of 20 MHz, 8 RX ADC bits (input bits), 8 TX DAC bits (output bits). In addition, this platform has a cost several times less than NI USRP 2920. The block diagram of the experiment is shown in Fig. 62.3.

To begin with, various envelopes were recorded in Matlab environment, obtained in the course of solving the optimization problem. Then a complex envelope was

Fig. 62.3 Block diagram of the experiment

formed in the Visual Studio program and fed to the transmitter, the role of which is played by the HackRF One device. In the hardware part of the SDR, it is transferred to the carrier and transmitted to the Agilent DSO9104A oscilloscope and the Agilent Technologies N9342C spectrum analyzer. Experiments were carried out with the modem parameters: carrier frequency—402 MHz, sampling frequency—10 MHz, number of samples per bit—100, modulation scheme DBPSK. Also, in the experiment, we consider the case of a fixed peak power.

62.2 Optimization Problem

62.2.1 Examples of Numerical Solutions to the Optimization Problem

Let us consider the results of solutions when choosing a limitation on the concentration of energy with a further limitation of the PAPR of the radiated oscillations.

The case without limiting the PAPR is shown in Fig. 62.4. In this case, the
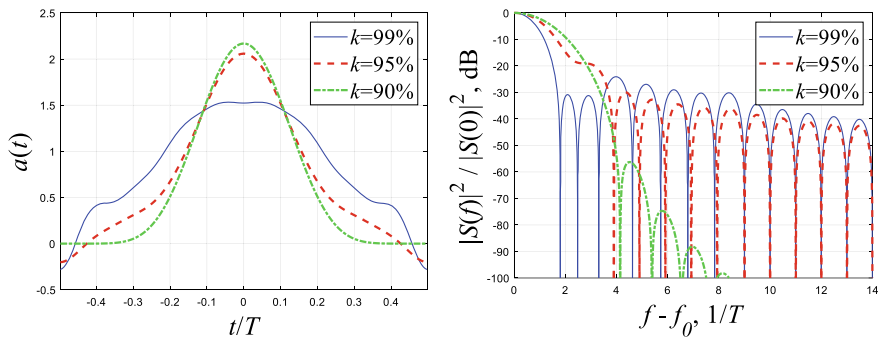
Fig. 62.4 Solution results for the case without limitation on the PAPR values and the corresponding energy spectra

Table 62.1 Results of solving an optimization problem with a constraint on the energy concentration and then adding a constraint on the PAPR

m	n	BEC_d (%)	Initial PAPR (dB)	PAPR with limitation (dB)	Occupied frequency bandwidth, $1/T$
9	8	90	6.7	6.5	3.17
		95	6.3	6	3.75
		99	3.7	3.6	4.83

frequency band was fixed at an energy concentration of 90%, then the BEC_d parameter was increased at a given band equal to 3.2/T (Table 62.1). The PAPR value without increasing the energy concentration is 6.7 dB. After adding the limitation on the BEC_d parameter in the fixed band (Fig. 62.3), the PAPR values decreased from 6.7 dB for $BEC_d = 90\%$ and to 3.7 dB for $BEC_d = 99\%$. Note that the envelope shape tends to a rectangular shape while energy concentration increases.

Also, according to the data from Table 62.1, it can be seen that 90% of the energy of the signal under consideration is contained in the frequency band equal to 3.17/T, after the introduction of a restriction on the energy concentration, the frequency band for 99% became equal to 4.83/T. In the frequency domain, this limitation influenced the redistribution of energy from the main lobe to out-of-band emissions. Let us move on to the case with the addition of a limitation on the value of the PAPR.

Figure 62.5 shows the optimal envelopes when limiting the PAPR to 6.5, 6, 3.7 dB. In this case, it was possible to synthesize signals with a relatively small limitation on the PAPR (0.1–0.2 dB), since this order of limitations complicates the optimization problem and complicates the search for solutions. Also, the optimization problem is solved in such a way that the frequency band does not change due to the introduction of a limitation on the PAPR.

Let us consider a technique for optimizing a PAPR limited signal with subsequent limitation on energy concentration in a given frequency band (Fig. 62.6). The initial value of the PAPR limitation is 6 dB (Table 62.2). After increasing the $BECd$ param-
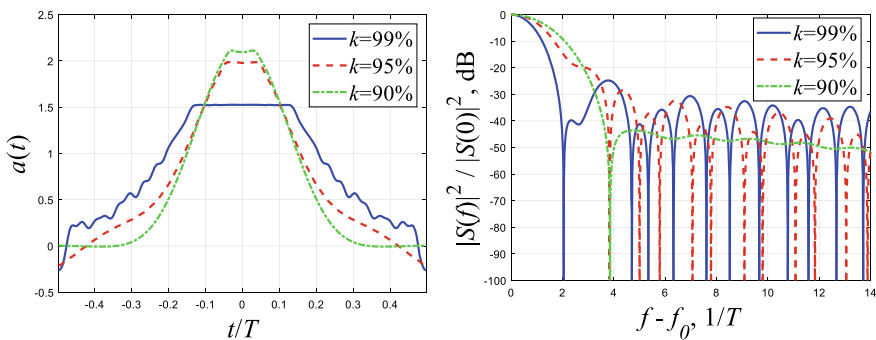
Fig. 62.5 Solution results for the case with a limitation on the PAPR value and the corresponding energy spectra
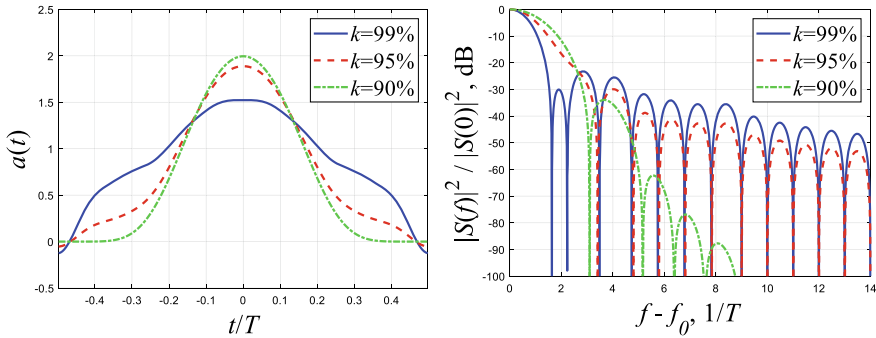
Fig. 62.6 Solution results for the case of a PAPR limited signal with subsequent limitation on the energy concentration and the corresponding energy spectra

Table 62.2 Results of solving an optimization problem with a signal limited by the PAPR and a subsequent restriction on the energy concentration

m	n	BEC_d (%)	Initial PAPR (dB)	PAPR with limitation (dB)	Occupied frequency bandwidth, $1/T$
9	8	90		6	2.82
		95	6	5.5	3.32
		99		3.5	4.22

eter by 9% in the 2.82/T band, the PAPR reached 3.5 dB. Lower values of the PAPR and a smaller occupied frequency band relative to the above considered case allow us to conclude that this technique is more efficient.

62.2.2 Experimental Study Based on the HackRF One SDR Platform

Let us move on to examining the results obtained during the experiment on HackRF One. To analyze the PAPR, signals with different envelopes were recorded with subsequent transfer to Matlab. The obtained experimental values of the PAPR are given in Table 62.3. It can be seen that the experimentally obtained PAPR values are greater than the simulated PAPR values by about 1 dB. This fact is explained by the presence of transient processes when the signal phase changes and parasitic amplitude modulation due to the nonlinearity of the amplifiers.

Figure 62.7 shows the oscillograms of the optimal signal at the concentration of energy in a given frequency band without limiting the PAPR of the radiated oscillations. The difference in the form of the envelopes on the oscillograms from the envelopes in Figs. 62.4, 62.5 and 62.6 is explained by the fact that the experiment recorded the peak power.

Table 62.3 Simulation results and experimental values at different signal envelopes

Optimization technique used	BEC_d (%)	Simulation results (dB)	Experimental PAPR (dB)
A case without restriction on PAPR	90	9.6	9.6
	95	9.2	9.1
	99	6.7	7.3
A case with a restriction on a PAPR	90	9.5	10
	95	9	9.6
	99	6.7	7.5
A case of a PAPR limited signal with subsequent limitation on the energy concentration	90	8.9	10.3
	95	8.5	9.8
	99	6.7	7.4
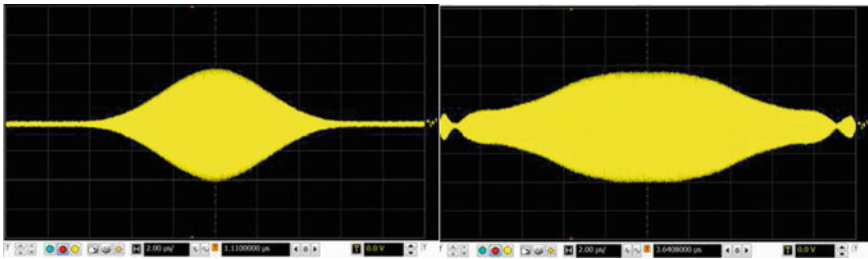
Fig. 62.7 Optimal envelope shapes ($BEC_d = 90$ and 99%) without PAPR limitation at 2 μs

Figure 62.8 shows oscillograms of the optimal signal at energy concentration in a given frequency band with a limitation on the PAPR of the radiated oscillations.

Consider a technique for optimizing a PAPR limited signal with subsequent limitation on the energy concentration in a given frequency band (Fig. 62.9).

Let us consider the characteristics of the experimentally obtained energy spectra: without limitation on the PAPR (Fig. 62.10a), with limitation on the PAPR
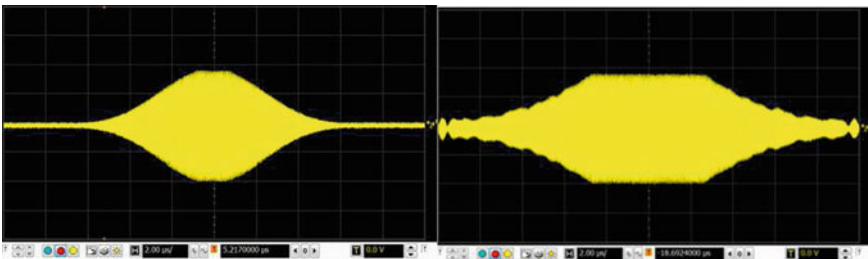
Fig. 62.8 Optimal envelope shapes ($BEC_d = 90$ and 99%) with PAPR limitation at 2 μs
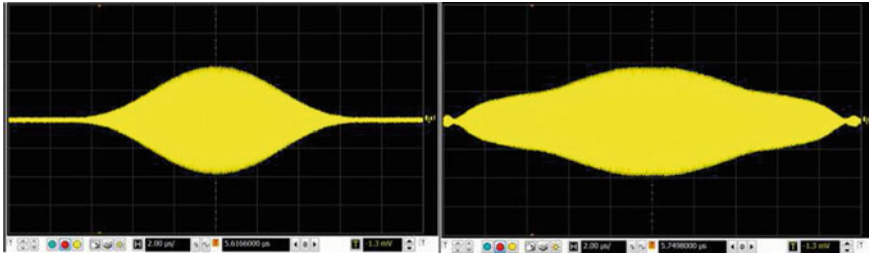
Fig. 62.9 Optimal envelope shapes ($BEC_d = 90$ and 99%) of a PAPR limited signal with subsequent limitation on the energy concentration at 2 μs
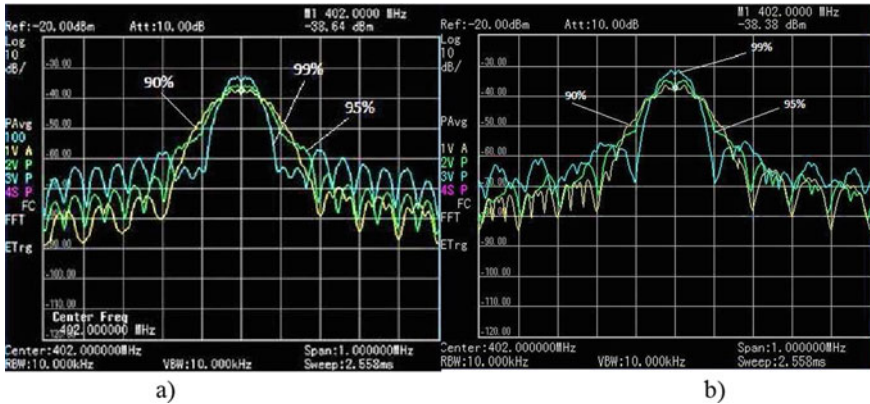
Fig. 62.10 The energy spectrum of the optimal signal with and without PAPR limitation

(Fig. 62.10b), limited in the signal PAPR with subsequent limitation on the energy concentration (Fig. 62.11).

It can be seen that an increase in the energy concentration in a given frequency band leads to narrowing of the spectrum near zero, which corresponds to the results obtained as a result of modeling.

62.3 Conclusion

The problem of synthesis of optimal spectrally effective signals for the criterion of a given energy concentration in the frequency band with the addition of a limitation on the PAPR is considered, as well as the numerical solution of this optimization problem using the expansion of the signal envelope in a limited Fourier series. Two methods for solving the problem were shown and compared.

The influence of restrictions on the solution of the optimization problem is shown. In particular, it was shown that with an increase in the value of the BEC_d parameter,

Fig. 62.11 The energy
spectrum of a PAPR limited
signal with subsequent
limitation on the energy
concentration at 2 μs
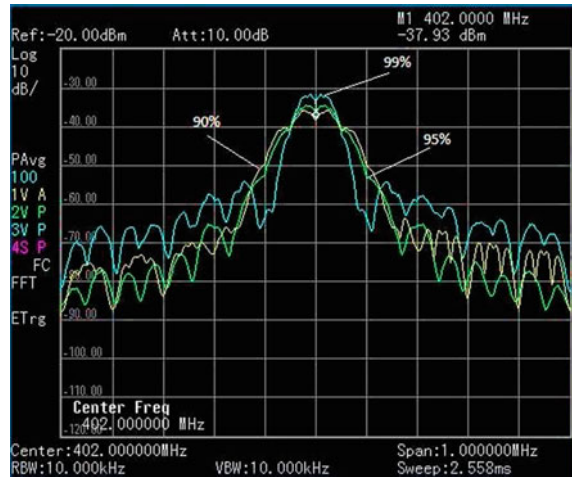
which is responsible for the energy concentration in a given frequency band, by 9
percent, the occupied frequency band increases by 65% for the first technique, and
when optimizing the PAPR-limited signal, the occupied frequency band increases by
67%. Also, with an increase in the energy concentration in a fixed frequency band,
the PAPR decreases by about 1.7 times. It has been shown that the second technique
provides a lower PAPR and occupied bandwidth relative to the first.

An experimental study was carried out on the basis of the HackRF One SDR
platform with different signal envelopes. The simulation results are consistent with
the experimental data.

References

1. M. Liu, W. Xue, P. Jia, S.B. Makarov, B. Li, Research on spectrum optimization technology
   for a wireless communication system. Symmetry **12**(1), art. № 34, 1–12 (2020). https://doi.
   org/10.3390/SYM12010034
2. S.B. Makarov, M. Liu, A.S. Ovsyannikova, S.V. Zavjalov, I.I. Lavrenyuk, W. Xue, J. Qi,
   Optimizing the shape of Faster-Than-Nyquist (FTN) signals with the constraint on energy
   concentration in the occupied frequency bandwidth, in *IEEE Access*, 8, art. № 9139940 (2020),
   pp. 130082–130093. https://doi.org/10.1109/ACCESS.2020.3009213
3. A. Gelgor, T. Gelgor, New pulse shapes for partial response signaling to outperform faster-
   than-Nyquist signaling, in *2019 IEEE International Conference on Electrical Engineering and*

Photonics (EExPolytech) (2019), pp. 144–148. https://doi.org/10.1109/EExPolytech.2019.890 6884

4. M. Liu, W. Xue, S.B. Makarov, J. Qi, B. Li, Comparative study of analytic solution and numerical solution of baseband symbol signal based on optimal generic function, in *Mathematical Problems in Engineering*, art. № 8045217, pp. 1–12 (2019). https://doi.org/10.1155/2019/804 5217

5. P. Nguyen Tan Hoang, A. Gelgor, Optimization of shaping pulse by spectral mask to enhance DVB-S2, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11660 LNCS (2019), pp. 649–660. https://doi.org/10.1007/978-3-030-30859-9_56

6. P.P. Vizarreta, G. Montoro, P.L. Gilabert, Hybrid envelope amplifier for envelope tracking power amplifier transmitters, in *2012 42nd European Microwave Conference* (IEEE, 2012) pp. 128–131

7. V. Sorotsky, R. Zudov, Application of neural networks to reduce distortion of RF signals in switch mode power amplifiers, in *2020 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)* (IEEE, 2020), pp. 76–79

8. A. Pergushev, V. Sorotsky, A. Ulanov, Criteria for selection envelope tracking power supply parameters for high peak-to-average power ratio applications, in *Proceedings of the 2019 IEEE International Conference on Electrical Engineering and Photonics*, EExPolytech 2019 (2019), pp 13–16. https://doi.org/10.1109/EExPolytech.2019.8906793

9. A.V. Zhila, A. S. Ovsyannikova, S.V. Zavjalov, Estimation of the possibility of PAPR reduction for optimal signals, in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems* (Springer, Cham, 2020), pp. 402–414

10. E.N. Smirnova, S.B. Makarov, S.V. Zavjalov, B. Polozhintsev, Influence of the amplitude limitation of signals with the sin(x)/x envelope on the spectral and energy characteristics, in *Proceedings of the 2020 IEEE International Conference on Electrical Engineering and Photonics*, EExPolytech 2020, art. № 9243855 (2020), pp. 164–167. https://doi.org/10.1109/EExPolytech50912.2020.9243855

11. D.C. Nguyen, S.V. Zavjalov, A.S. Ovsyannikova, The effectiveness of application of multi-frequency signals under conditions of amplitude limitation, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11660 LNCS (2019), pp. 681–687. https://doi.org/10.1007/978-3-030-30859-9_59

12. T. Wild, V. Braun, H. Viswanathan, Joint design of communication and sensing for beyond 5G and 6G systems, in *IEEE Access* (2021), pp. 1–1. https://doi.org/10.1109/ACCESS.2021.305 9488

13. A. Papazafeiropoulos, H.Q. Ngo, P. Kourtessis, S. Chatzinotas, J.M. Senior, Towards optimal energy efficiency in Cell-Free Massive MIMO systems, in *IEEE Transactions on Green Communications and Networking* (2021). https://doi.org/10.1109/TGCN.2021.3059206

14. A. Ometov, D. Moltchanov, M. Komarov, S.V. Volvenko, Y. Koucheryavy, Packet level performance assessment of mmWave backhauling technology for 3GPP NR Systems. IEEE Access **7**, 9860–9871 (2019). https://doi.org/10.1109/ACCESS.2018.2890558

15. K. Mhatre, U.P. Khot, Efficient selective mapping PAPR reduction technique. Proc. Comput. Sci. **45**, 620–627 (2015). https://doi.org/10.1016/j.procs.2015.03.117

16. HackRF One Homepage. https://greatscottgadgets.com/hackrf/one/. Last accessed 2021/02/19

Chapter 63
Design of Compact Shielded Multilayer Directional Coupler

Check for updates

Alexey V. Andropov, Sergey V. Kuzmin, and Konstantin O. Korovin

Abstract Although the general trend in HF technology is to reduce the analogue part of the feeder line, directional couplers are still in demand. The development of directional coupler designs has been towards increasing the operating frequency bandwidth and reducing weight and dimensions. The use of directional couplers in high-density equipment introduces additional boundary conditions that significantly affect the performance, which have been taken into account in this paper. The paper discusses the implementation of directional couplers in a microstrip line designed for communication links in the 2–3 GHz frequency range, on a substrate of the widely available and commonly used RO4350 material, with a coupling factor of 3–7 dB. These directional couplers have been designed with additional boundary conditions to reduce the overall size of the unit. In the numerical modelling, the correction factors used for the preliminary analytical calculations were refined.

Keywords Directional coupler · Microstrip line

63.1 Introduction

Despite the general trend to decrease the analog part of the feeder line, directional couplers are still widely used in microwave technique.

There are many microstrip coupler designs but, due to the technological and series production limitations, just a few of them are widely used in industry. These are mainly structures with a lateral and facial coupling [1].

Printed design directional couplers with side coupling, are usually employed used when it is necessary to provide transient attenuation of 10–20 dB for measuring power in a feeder line. At the same time, they strive to increase the coupling factor, as a rule, making use of the structures described in [2–4]. Since the distance between the strip lines is big and practicable by printed circuit board technology for such

A. V. Andropov (✉) · S. V. Kuzmin · K. O. Korovin
Bonch-Bruevich Saint-Petersburg State University of Telecommunications, St. Petersburg, Russia

coupling factor, directional couplers of this type have reproducible characteristics and are successfully used.

In addition, such devices withstand high levels of average and impulse power, especially after coating with varnish.

Directional couplers with a transient attenuation of 3–8 dB are mainly necessary when creating beamforming network (BFN), in particular, of matrix type [5]. With coupling factor mentioned above, the gap between the strip lines in the side-coupled structure gets narrow and practically can't be realized with the PCB (printed circuit board) technology. Even using other technologies (vacuum deposition, etc.), it is not possible to ensure repeatability, high operating capacity, manufacturability and low cost. This was the reason for transition to the structures with facial coupling.

Commercially available samples of multilayer printed directional couplers with a face coupling through a thin dielectric layer have a number of disadvantages.

The most significant of them are the following.

- During installation on the printed circuit board, parasitic reactivity occurs, which, as a result, degrades the characteristics of directional coupler in the product.
- The most popular transitional attenuation values, which are in series production, are not always suitable for building a beam forming network to form the necessary amplitude-phase distribution.
- The production of directional couplers using the technology of multilayer printed circuit boards makes prototyping difficult due to the high cost of single samples.

These disadvantages have led to the need to search for a more technologically advanced structure of directional couplers. Now, directional couplers with front coupling through a slot in a screen are actively developing. It was noted in [6] that the degree of coupling is determined by the width of the gap. A similar structure is described in [7]. There, in the field of communications, strip lines have wider than regular lines and beveled at 45° to reduce the ratio reflections. In [8], the shape of the stripes in the bond region is offered in the form of rhombuses, and the slot is in the form of hexagon. In [9] the stripes and the slit are made in ellipsoid form, which, with the correct selection dimensions, can significantly improve performance.

63.2   Problem Statement

The directional coupler described in [9] has an open structure. The strips in the coupling area are unshielded and can be affected by external boundaries such as enclosure covers or adjacent elements onboard. If, for example, a BFN has a dense layout [10], it is necessary to reduce the dimensions of the elements included. In this regard, it became necessary to create a methodology for designing such directional couplers, taking into account the effect of shields, in order to reduce the overall dimensions. As an example of utilization of such technique, we consider directional couplers built for communication lines in the frequency range of 2–3 GHz, on a

substrate made of generally used and readily available material RO4350, with a coupling factor of 3–7 dB.

63.3   Experimental Technique

At the moment, there is no analytical model that allows one to synthesize the structure of directional coupler under study with sufficient accuracy. Well-known analytic expressions help to get the preliminary dimensions of the structure. Therefore, the design of directional coupler includes the stage of preliminary analytical calculations and the stage of numerical experiments in CAD. Since numerical experiments in CAD require sufficient computational power, the finite difference time domain (FDTD) technique was chosen for the calculations. At the same time, it is possible to use computing systems based on graphic processor units (GPUs), which leads to manifold acceleration of calculations, and the utilization of optimization becomes possible.

63.4   Analytical Calculations

Let us present analytical expressions from [9, 11, 12] for calculating structures shown in Fig. 63.1.
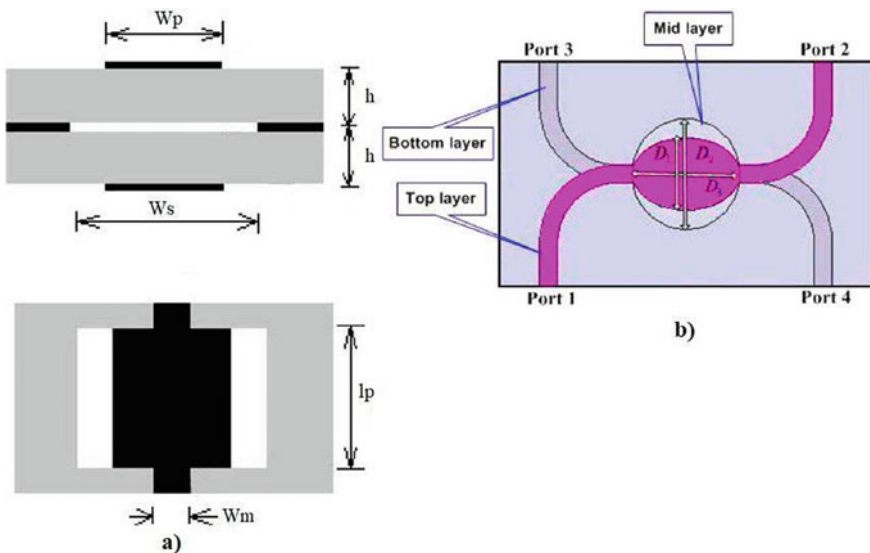
Fig. 63.1   Geometry of the problem, **a** Initial structure, **b** structure under study

The design shown in Fig. 63.1b consists of two dielectric layers and three metal layers. Microstrip lines, turning into ellipses are formed on the outer layers of metal. Some of the metal in the inner layer is removed and an elliptical window is formed. The centers and semi-major axes of all ellipses coincide. The minor semiaxis of the inner elliptical window is larger than the identical minor semiaxes of the outer metallized ellipses.

The required wave impedances for even and odd wave modes are found from the equations:

$$Z_{oe} = Z_0 \sqrt{\frac{1 + 10^{\frac{-C_{dB}}{20}}}{1 - 10^{\frac{-C_{dB}}{20}}}},$$

$$Z_{oo} = Z_0 \sqrt{\frac{1 - 10^{\frac{-C_{dB}}{20}}}{1 + 10^{\frac{-C_{dB}}{20}}}}, \tag{63.1}$$

where C_{dB} is the coupling factor of the coupler in dB, Z_0 is the characteristic impedance of the supply line.

For the structure shown in Fig. 63.1a the indicated wave impedances will be as follows:

$$Z_{oe} = \frac{60\pi}{\sqrt{\varepsilon_r}} \frac{K(k_1)}{K'(k_1)},$$

$$Z_{oo} = \frac{60\pi}{\sqrt{\varepsilon_r}} \frac{K'(k_2)}{K(k_2)}, \tag{63.2}$$

where

$$k_1 = \sqrt{\frac{\sinh^2\left(\frac{\pi w_s}{4h}\right)}{\sinh^2\left(\frac{\pi w_s}{4h}\right) + \cosh^2\left(\frac{\pi w_p}{4h}\right)}},$$

$$k_2 = \tanh\left(\frac{\pi w_p}{4h}\right),$$

$$\frac{K(k)}{K'(k)} = \begin{cases} \frac{2}{\pi} \ln\left(2\sqrt{\frac{1+k}{1-k}}\right), & \text{for } 0.707 < k \leq 1 \\ \frac{\pi}{2\ln\left(2\sqrt{\frac{1+\sqrt{1-k^2}}{1-\sqrt{1-k^2}}}\right)}, & \text{for } 0 \leq k \leq 0.707. \end{cases}$$

By equalizing the corresponding wave impedances in Eqs. (63.1) and (63.2), we obtain the desired line dimensions. In the simplest case, the solution can be found graphically.

Let's introduce auxiliary designations: Rx = 0.5D1, Ry = 0.5D3, Rm = 0.5D2.

To find the semiaxes of the ellipses, correction factors are introduced:

$$D_3 = k_y\left(\sqrt{l_p^2 + w_p^2} + l_p\right)$$

$$D_1 = k_x\left(\frac{w_p l_p}{D_3}\right),$$ (63.3)

$$D_2 = k_m\left(\frac{w_s l_p}{D_3}\right),$$

where $l_p = \Lambda/4$ in microstrip line.
From [1] $k_x = 1.273$, $k_y = 0.5$, $k_m = 1.273$.

63.5   Numerical Experiment

63.5.1   Preparing a Model for a Numerical Experiment

Initially, for preliminary calculations and optimization of characteristics, a simplified model is formed, shown in Fig. 63.2.

Since the directional coupler will be used in the future in the electronic devices with high packing density [12] the boundary conditions simulate the situation when metal parts of the structure are located on all sides of the coupler. Optimization of characteristics is carried out with allowance for these boundary conditions. If they change, the characteristics of the directional coupler can change significantly. The distance from the strip to the case cover is chosen equal to triple substrate height H, which will allow the use of other components, for example, SMD resistors, without significantly affecting their performance.

As can be seen from Fig. 63.2, the preliminary model is simplified. To make an experimental sample, one needs to separate the supply lines on the sides to make the installation of radio frequency connectors possible. This simplified model is required to accelerate the optimization of the directional coupler characteristics when

Fig. 63.2 Preliminary
model of directional coupler
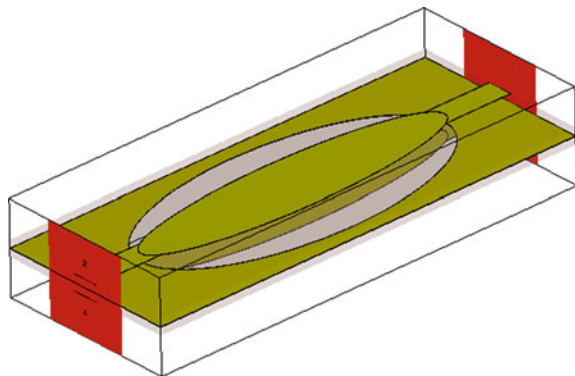
changing the semiaxes of the ellipses at the expense of reducing the size of the structure.

Another reason for the construction of such initial model is the ability to correctly shift the reference planes of the ports to the beginning of the ellipses to exclude the get rid of supply lines, which is necessary to speed up the calculations.

63.5.2 Results of Model Analysis

Let us substitute in the model the length of the semiaxes calculated analytically. As a result of numerical calculations in CAD, using the finite difference method in the time domain, we obtain the results shown in Figs. 63.3 and 63.4

We optimized the performance, which took quite a long time. The results can be seen in Figs. 63.5 and 63.6.

The VSWR is decreased by an impedance transformer between the regular microstrip line and outer ellipses. The dimensions of the transformer can be found as a result of numerical experiments, but it takes a long time. Therefore, we will apply a different technique.

We shift the reference planes ports to the beginning of the outer ellipses and obtain scattering matrix of directed coupler. Then we export this matrix in CAD based on microwave circuit theory, add impedance transformers to the new model
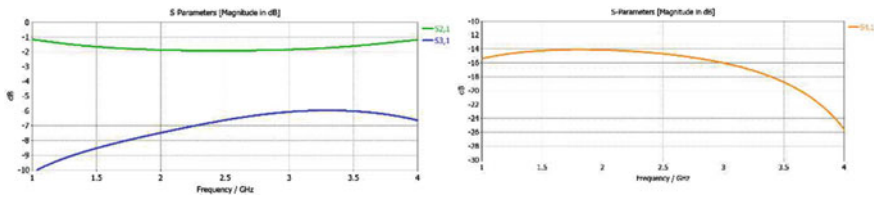
Fig. 63.3 Transmission coefficients (Rx = 2.75 mm, Ry = 9.03 mm, Rm = 2.93 mm)
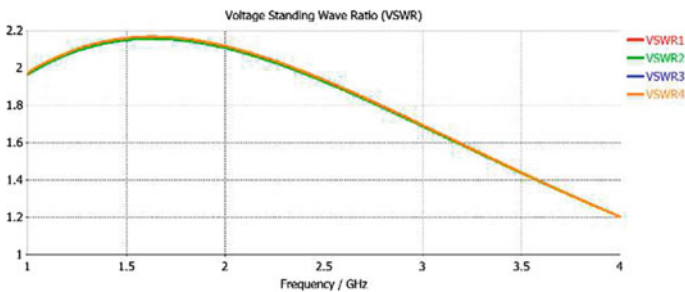
Fig. 63.4 VSWR (Rx = 2.75 mm, Ry = 9.03 mm, Rm = 2.93 mm)
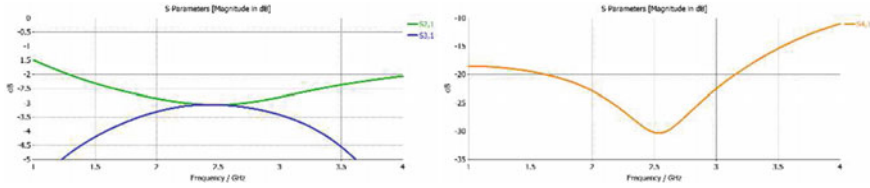
Fig. 63.5 Transmission coefficients after optimization (Rx = 2.42 mm, Ry = 11.0 mm, Rm = 5.0 mm)
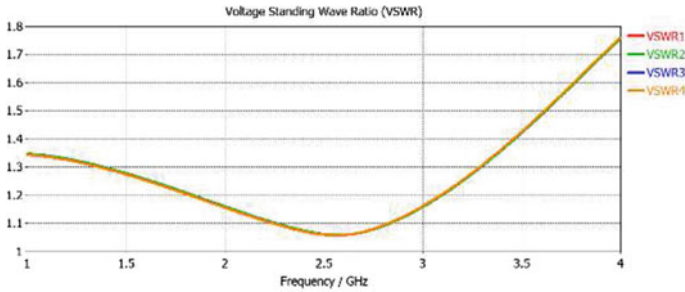
Fig. 63.6 VSWR after optimization (Rx = 2.42 mm, Ry = 11.0 mm, Rm = 5.0 mm)

and carry out optimization of performance. The results are shown in Figs. 63.7 and 63.8. Transformer dimensions Ltr = 3.92 mm, Wtr = 1.67 mm.

The approach to calculating the size of the transformer, described above, is hybrid and combines analytical calculations by the method of microwave circuit theory and numerical optimization methods. By doing this, we increase the speed of calculation manifold, compared with CAD calculations using numerical methods of electrodynamics.

The new refined model of directed coupler is shown in Fig. 63.9. Outputs are spaced apart, since on the basis of this model a model for field experiments is made.

Fig. 63.7 Transmission coefficients after optimization in CAD based on circuit theory
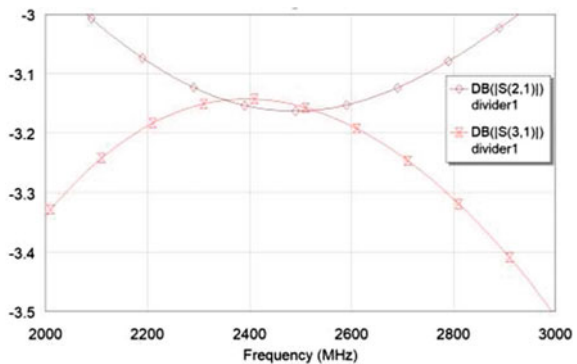
Fig. 63.8 VSWR after optimization in CAD based on circuit theory
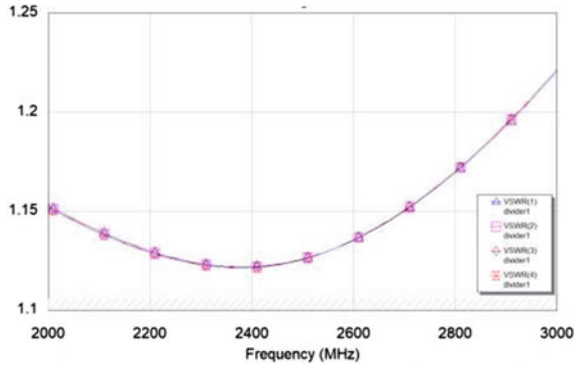
Fig. 63.9 Refined directed coupler model with transformers
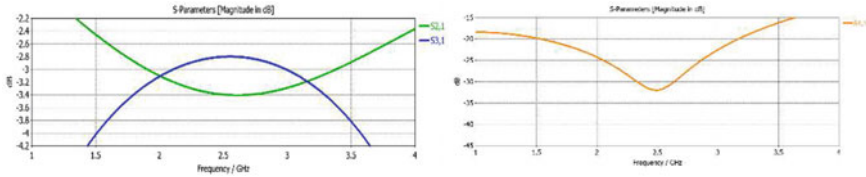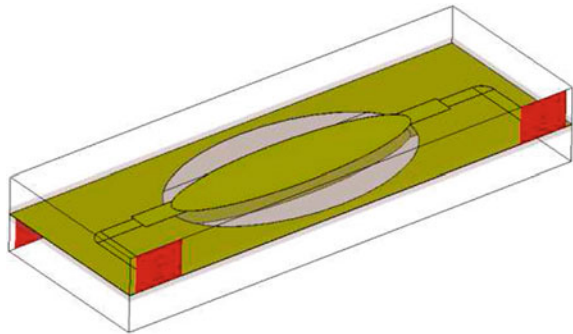
Fig. 63.10 Transmission coefficients (Rx = 2.64 mm, Ry = 10.5 mm, Rm = 5.86 mm, Ltr = 3.11 mm, Wtr = 1.7 mm)

The results of the characteristics optimization after the next correction of geometric dimensions are shown in Figs. 63.10 and 63.11.

63.5.3 Refining Analytic Expressions

The geometric dimensions obtained as a result of optimization in the course of numerical experiments allow adjust the coefficients in (63.3). The new coefficients
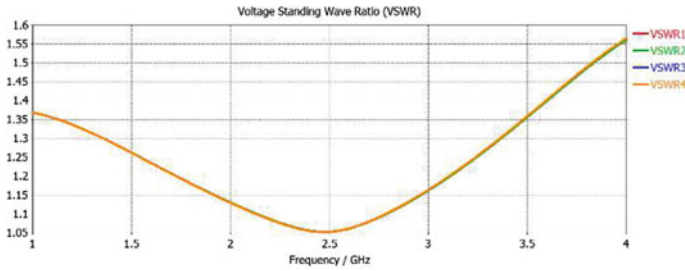
Fig. 63.11 VSWR (Rx = 2.64 mm, Ry = 10.5 mm, Rm = 5.86 mm, Ltr = 3.11 mm, Wtr = 1.7 mm)

will be as follows: kx = 1.363, ky = 0.609, km = 2.640. We use them to obtain the dimensions of the directional coupler, the characteristics of which are upshifted by 300 MHz.

The results of numerical simulation are shown in Figs. 63.12 and 63.13.

As a result of numerical simulation, one can also obtain the coefficients in (63.3) for different values of attenuation. The results are shown in Table 63.1.

Thus, expressions (63.3), with allowance for the coefficients from Table 63.1, make synthesis of the required dimensions of the directional coupler possible, without lengthy numerical experiments, decreasing the time to obtain the results.
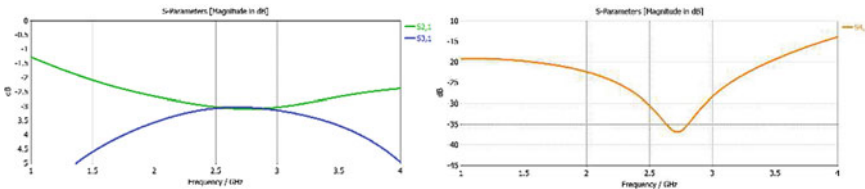
Fig. 63.12 Transmission coefficients (Rx = 2.41 mm, Ry = 9.86 mm, Rm = 4.98 mm)
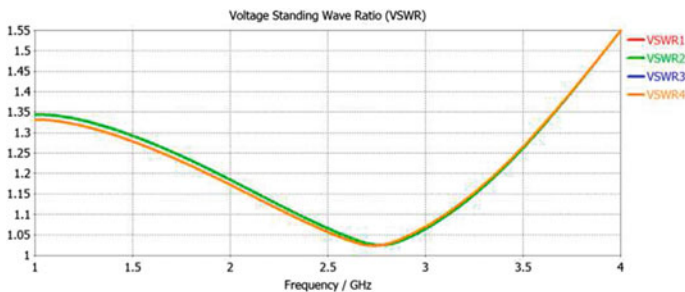
Fig. 63.13 VSWR (Rx = 2.41 mm, Ry = 9.86 mm, Rm = 4.98 mm)

Table 63.1 Coefficients that were obtained as a result of numerical simulation

k	C (dB)			
	3.0	6.0	6.6	10.5
kx	1.363	1.339	1.315	1.252
ky	0.609	0.613	0.614	0.615
km	2.640	1.823	1.840	1.696

63.6 Conclusion

The presented technique allows one to accelerate the design of directional coupler with the face coupling of elliptical conductors through an elliptical slit in the screen, with account for external closely spaced screens. The coefficients obtained in numerical experiments can be used to synthesize a directional coupler design with a transient attenuation of 3–11 dB without lengthy numerical calculations. As a result of insertion of shields into the structure of the directional coupler, the total height of the structure is reduced by more than by factor of 2.

Further development of the subject of miniature and face-to-face shielded microwave devices through the slit can be directed to increase operating frequency bands, as suggested, for example, in [13].

Based on the structures considered in [10], broadband phase-shifting circuits can also be created [14], which are especially relevant in modern antenna arrays. Introduction of adjustment elements in the design can allow the creation of a new class of phase shifters with manual and electromechanical control [15]. Shielded broadband directional taps and phase shifters can be the basis for the construction of Butler matrices in two-dimensional multi-beam antenna systems.

References

1. V.A. Neganov, G.P. Yarovoy, Theory and application of microwave devices, in *Radio and Communication*, Samara (2006)
2. High-directivity Microstrip Coupler Having Periodically Indented Conductors. Patent US, no. 3629733A, 21.12.1971
3. S. Uysal, C.W. Turner, J. Watkins, Nonuniform transmission line codirectional couplers for hybrid MIMIC and superconductive applications. Trans. Microwave Theor. Techniques **42**(3), 407–414 (1994)
4. S. Uysal, H. Aghvami, Synthesis, design, and construction of ultra-wide-band nonuniform quadrature directional couplers in inhomogeneous media. Trans. Microwave Theor. Techniques **37**(6), 969–976 (1989)
5. S.Z. Ibrahim, M.E. Bialkowski, Wideband butler matrix in microstrip slot technology, in *Asia Pacific Microwave Conference 2009* (IEEE, Singapore, 2009), pp. 2104–2107
6. G.I. Veselov, *Microelectronic Microwave Devices* (Vysshaia shkola, Moscow, 1988)
7. V.I. Gvozdev, E.I. Nefedov, *Volumetric Microwave Integrated Circuits* (Nauka, Moscow, 1985)

8. L. Abdelghani, T.A. Denidni, M. Nedil, Design of a broadband multilayer coupler for UWB beamforming applications, in *Proceedings of the 41st European Microwave Conference 2011*. IEEE, pp. 810–813

9. A.M. Abbosh, M.E. Bialkowski, Design of compact directional couplers for UWB applications. Trans. Microwave Theor. Techniques **55**(2), 189–194 (2007)

10. A.V. Moroz, V.V Davydov, K.Y. Malanin, A.A. Krasnov, V.Yu. Rud, Features of the construction of the noise compensation circuit of a small-sized active phased antenna array. J. Phys.: Conf. Ser. **1400**(4) (2019)

11. Yu.V. Pimenov, V.I. Volman, A.D. Muravtsov, *Technical Electrodynamics* (Radio i sviaz, Moscow, 2000)

12. A.M. Abbosh, Ultra-wideband phase shifters. Trans. Microwave Theor. Techniques **55**(9) (2007)

13. N.S. Binti Muklas, S.K.A. Rahim, N. Seman, A design of compact ultra wideband coupler for butler matrix. Wireless Personal Commun. **70**, 915–926 (2013)

14. W.S. Chang, C.Y. Chang, A High slow-wave factor Microstrip structure with simple design formulas and its application to microwave circuit design. IEEE Trans. Microw. Theor. Tech. **60**(11), 3376–3383 (2012)

15. V. Kochemasov, L. Belov, A. Maistrenko, Phase shifters with manual and electromechanical control. Components Technol. **6**(179), 57–68 (2016)

Chapter 64
Development of the Controllable C-field Current Stabilizer for the Atomic Beam Tube of the Cesium Atom Clock

Kseniya Arinushkina, Anna Grevsteva, Anton Valov, and Nikita Lukashev

Abstract The feasibility of using quantum frequency standards (QFS)—atomic clocks as reference measures in satellite navigation systems (SNS) compared to other reference generators models is justified. It is noted that among all atomic clocks, cesium QFS occupy a special place. These standards are the primary reference frequency on which the international time scale is based. Given the importance of atomic clocks in the operation of satellite systems, the development of new and modernized QFS models is an important and urgent task. The main goal of new developments or upgrades is to improve the metrological characteristics of QFS. In case of its application for mobile objects, its dimensions, weight and power consumption also become important characteristics. One of the factors causing instability in the operation of QFS is considered. The development of a device for stabilizing the magnetic field at a given level in the atomic-beam tube of the QFS and the results of its operation are presented. The short-term frequency stability of the output signal of the frequency standard is improved by 10% and its temperature sensitivity is reduced twice.

Keywords Automatic Frequency Adjustment · Frequency Stabiliser · Cesium Frequency Standard · Stable STI Provision · Operational Amplifier · Digital-to-Analog Converter · Atomic Beam Tube

64.1 Introduction

The development of modern systems of communication, navigation and exploration of outer space is impossible without the time determination with high accuracy [1–7]. To solve this problem, various devices are used in navigation systems, electronic

---

K. Arinushkina (✉) · A. Grevsteva
Peter the Great St. Petersburg Polytechnic University, St. Petersburg 195251, Russia

A. Valov · N. Lukashev
Bonch-Bruevich St. Petersburg State University of Telecommunications, St. Petersburg 193232, Russia

589

equipment, telecommunication equipment, and computer and cellular networks [1, 6, 8–12]. These devices have different accuracy and stability of frequency determination, which is recalculated into units of time [9, 13–17]. Among them, quantum frequency standards (QFS), which are one of the most high-precision frequency measures [3, 5, 10, 12, 15, 17–21], have the best parameters. Therefore, only QFS are used as reference sources of high accuracy signals in satellite systems GLONASS, GALLEO and GPS [2, 3, 9, 10, 18, 22–25].

Cesium QFS occupy a special place among all quantum standards, since they give the primary reference frequency, which is the basis of international time scale [26–29]. Operation of the cesium QFS is based on the principle of tuning the frequency of the high-stability quartz oscillator (QG) relative to the cesium Cs-133 quantum transition frequency in the atomic beam tube (ABT), equal to 9,192,631,770 Hz. [26–29].

The relevance of the topic of this paper is determined by the fact that at present the frequency error of cesium standards has nearly reached its minimum. To improve the characteristics further, it is necessary to eliminate the influence of a number of factors, such as instabilities of magnetic field (for example, long-term drift of the current source, temperature dependence, the influence of an external magnetic field), of microwave power level, etc.

The solution to this problem is extremely important in the context of new requirements for the accuracy of determining coordinates. The work presents a developed scheme that allows stabilizing the magnetic field at a given level in the ALT.

The practical significance of the work lies in the development of a new device and substantiation of new methods for improving the metrological characteristics of the midrange.

64.2   Current Stabilizer

It has previously been found that the wavelengths of the spectral lines and the periods of the respective radiations can be used to determine the unit of length and the unit of time, respectively [11, 18, 20, 24]. The splitting of energy levels when a magnetic field is applied to atoms is called the Zeeman effect. Level splitting results in the splitting of spectral lines into several components (Fig. 64.1).

The transition $|F = 3, mf = 0| \leftrightarrow |F = 4, mf = 0|, mf = 0|$ is used for the operation of QFS on cesium-133 atoms, since it has a minimum dependence on the magnetic field that creates it. Atoms in the state of $F = 4, mF = 0$ or in the state of $F = 3, mF = 0$ are selected in the cesium ABT using a magnetic polarizer. The cesium beam is affected by three magnetic fields as it travels from the furnace to the detector. In order to split a beam of atoms into several beams, an inhomogeneous magnetic field, also called the A-field, acts. A magnetic field called the B-field is used to finally focus on the detector only atoms belonging to the same pair of energy levels used, which also serves to deflect the other atoms.

Fig. 64.1 Energy levels of the cesium atom in the main and first excited state
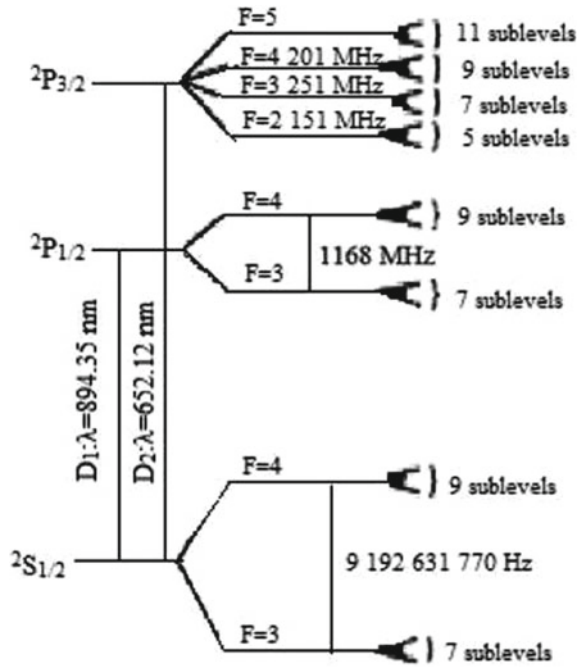
In the interval between these magnetic fields, the atoms fly through a volume resonator with conducting walls, in which electromagnetic oscillations of a certain frequency are excited, that is, a so-called C-field is created. If, under the influence of these vibrations, the Cs atom with the energy E1 goes over to the energy state E2, then the field of the second magnet will throw it away from the detector, because for an atom that has passed into the E2 state, the field of the second magnet will no longer be focusing and this atom will bypass the detector. Thus, the current through the detector will be reduced by an amount proportional to the number of atoms that have made energy transitions under the influence of the electromagnetic resonator. Transitions of Cs atoms from the E2 state to the E1 state will be recorded in the same way. Atoms in this state are detected and allow determining the frequency of the interrogation field, since the probability of transition has a maximum at the frequency of microwave resonance. Atoms in this state are detected and allow determining the frequency of the interrogation field, since the probability of transition has a maximum at the frequency of microwave resonance.

Scanning the atomic resonance frequency f results in a detector current similar to that represented in Fig. 64.2.

The central frequency of this resonance circuit (Zeeman line) depends linearly on the magnetic field inside the atomic beam tube. By measuring the resonance frequency, it is possible in some cases to stabilize the magnetic field inside the tube. Under the new operating conditions of the QFS, this becomes insufficient.

$$f_0 = 9.192.631.770\,\text{Hz}$$
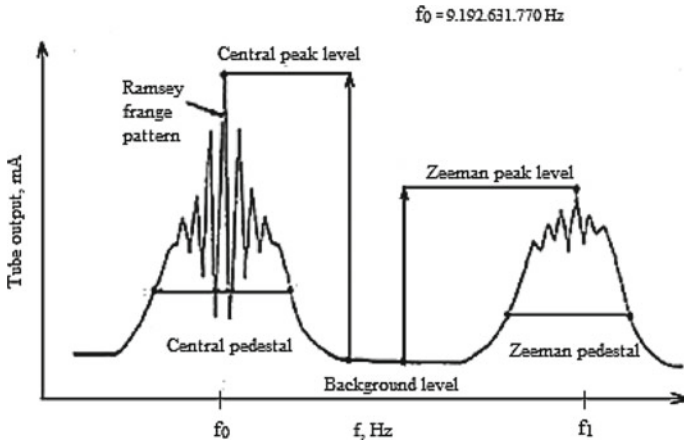
Fig. 64.2 Cesium ABT output vs microwave frequency

To stabilize the current, it is proposed to use a different dependence of different transitions of the ultra-thin structure of cesium on the magnetic field. The dependence of the main transition on the magnetic field is expressed by:

$$\Delta f_0 = 427 \cdot H_0^2 \tag{64.1}$$

Dependence of the first lateral transition on the field relative to the main one:

$$\Delta f_1 = f_1 - f_0 = 700{,}841 \cdot H_0 \tag{64.2}$$

Current stabilizer is used for generation of DC current on load supplied to solenoid of ABT medium field. The structural diagram of the developed current stabilizer is shown in Fig. 64.3.
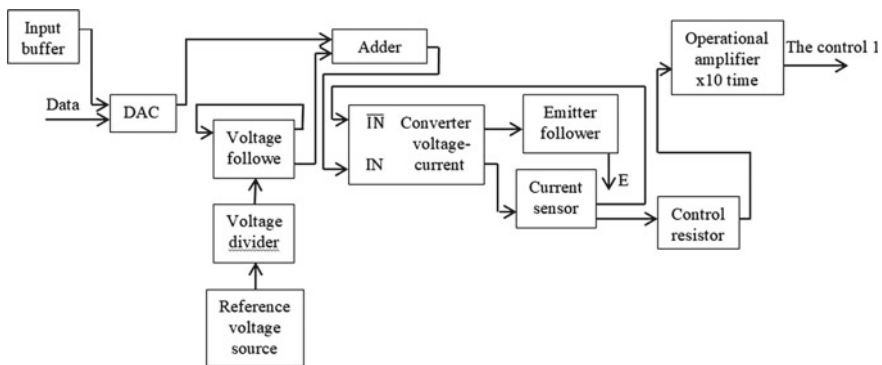
Fig. 64.3 Structural diagram of current stabilizer

The voltage generated at the current sensor is fed to the inverse input of the operational amplifier (Op-Amp), thus providing negative feedback. The transistor is connected according to the emitter follower circuit and amplifies the output current of the Op-Amp.

The voltage generated when the output current flows through the control resistor is fed to the Op-Amp, which amplifies this voltage 10 times and sends it to the "The control 1" output. In the case of a displacement of the magnetic field value, an unbalance voltage occurs, which allows monitoring the output current of the CS.

In the technical conditions for the Zener diode 2S198K OSM, we find that the temporary instability of the stabilization voltage is $\delta_{\text{sourse}} = \pm 0.02\% = 2 \times 10^{-4}$ per 1000 h. The service life of the equipment is 100,000 h.

Thus, the current drift over the service life will be:

$$\delta_{\text{sourse}} = 2 \times 10^{-4} \times 100 = 2 \times 10^{-2},$$

where δ_{sourse} is the voltage drift of the Zener diode for 1000 h; 100 is the number of intervals of 1000 h during the service life of 100,000 h.

Then the possible drift of the output current over the service life will be:

$$\Delta \text{Ist} = \text{Inom} \times 2 \times 10^{-2} = 20 \text{ mA} \times 2 \times 10^{-2} = 0.4 \text{ mA},$$

where ΔIst is the possible drift of the output current over the service life from the voltage drift of the Zener diode; Inom is the rated value of the CT output current.

64.3 Results of Experimental Studies of MHF Operation with New Current Stabilizer

In the previous version of the mean field current stabilizer, the signal from the voltage follower went directly to the voltage-current converter. The development of a new branch, consisting of an input buffer, a DAC and an adder, made it possible to avoid a current drift of 400 μA over 10 years of the stabilizer's service life. This branch provides an atomic transition adjustment.

In the temperature range from −20 °C to 20 °C, in accordance with the technical conditions, the current drift is not more than 1 μA/°. Tuning will take the current back.

The prototype of the current stabilizer was included in the design of the QFS, the output characteristics of which have been measured. After processing the obtained data, graphs of the dependence of Allan variance $\sigma(\tau)$ [°/h] on time τ were built for the modernized (Fig. 64.4) and the previous design of the QFS (Fig. 64.5).

The results obtained show an improvement in Allan variance $\sigma(\tau)$ by 12%. Studies of the operation of the QFS were carried out for 12 days in a temperature chamber. As a result, it was found that the temperature coefficient of the standard frequency
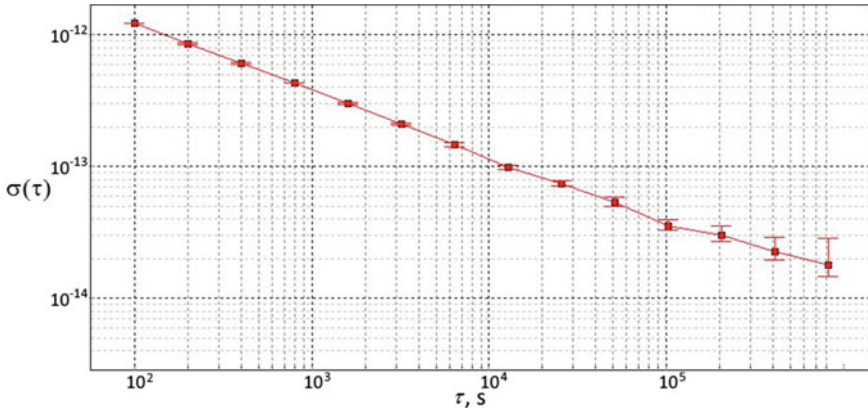
Fig. 64.4 A plot of Allan variance versus time using a new magnetic field stabilization system
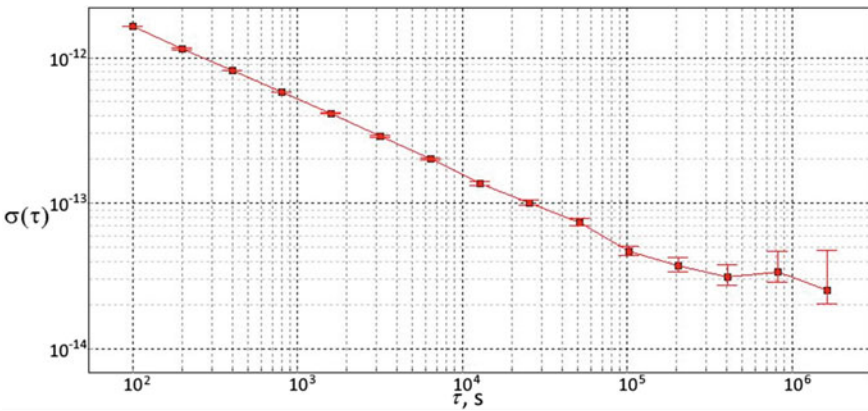
Fig. 64.5 A plot of Allan variance values vs time for the QFS of previous design

decreased by 2 times when the current adjustment ring was closed. Further tests were carried out in a chamber in which a spaceflight is simulated. From these two tests of the standard performance, conclusions were drawn regarding the improvement in Allan variance.

64.4   Conclusion

The results obtained show the validity of the proposed solutions to improve the stabilization of the magnetic field in ABT, which were implemented in the modernized

design of the quantum frequency standard based on cesium-133 atoms. The experiments carried out made it possible to determine new directions of development for the modernization of the design of the magnetic field stabilizer in the ABT.

References

1. N. Grebenikova, A. Moroz, M. Bylina, M. Kuzmin, Remote control of the quality and safety of the production of liquid products with using fiber-optic communication lines of the Internet, in *IOP Conference Series: Materials Science and Engineering* vol 497, 012109 (2019)
2. E. Batori, N. Almat, C. Affolderbach, G. Mileti, GNSS-grade space atomic frequency standards: current status and ongoing developments, in *Advances in Space Research* (2020)
3. E. Batori, N. Almat, C. Affolderbach, F. Gruet, G. Mileti, High-performance pulsed laser-pumped Rb clock for GNSS, in *Proceedings of European Navigation Conference, ENC 2020*, Dresden, Germany (Institute of Electrical and Electronics Engineers Inc), 9317388 (2020)
4. N. Lukashev, A. Glinushkin, V. Lukyantsev, Microwave low mass-dimensional frequency standard on Hg- 199 ions. J. Phys. Conf. Ser. **1410**(1), 012211 (2019)
5. A.H.A. Al-Odhari, Positioning of the radio source based on time difference of arrival method using unmanned aerial vehicles, in *2018 Systems of Signals Generating and Processing in the Field of on Board Communications* (Russian Federation, Moscow, 2018), pp. 1–5
6. A.-O.A.H. Ali, *Algorithm for Positioning in Non-line-of-Sight Conditions Using Unmanned Aerial Vehicles.* Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11118 LNCS, pp. 496–508 (2018)
7. P. Forman, Atomichron: the atomic clock from concert to commercial product. Proc. IEEE **73**(7), 1181–1204 (1985)
8. J. Vanier, Atomic frequency standards: basic physics and impact on metrology, in *Recent Advantages in Metrology and Fundamental Constants.* Course CXLVI (2001), pp. 397–452
9. B. Hofmann-Wellenhof, *GNSS—Global Navigation Satellite Systems.* GPS, GLONASS, GALILEO and more (SpringerWien, N.Y., 2008)
10. F. Vannicola, GPS Block IIF atomic frequency standard analysis, in *Proceeding of the 42th Ann. Precise Time and Time Interval Meeting* (2010), pp.181–195
11. M. Takamoto, An optical lattice clock. Nature **435**(7040), 321–324 (2005)
12. Y. Chan, W. Johnson, S. Karuza, A. Young, J. Camparo, Self-monitoring and self-assessing atomic clocks. IEEE Trans. Instrum. Meas. **59**(2), 330–334 (2010)
13. B. Jaduszliwer, J. Camparo, Past, present and future of atomic clocks for GNSS. GPS Solutions **25**(1), 27 (2021)
14. A. Makdissi, Evaluation of the accuracy of the optically pumped caesium beam primary frequency standard of the BNM-LPTF. Metrologia **38**, 409–425 (2001)
15. A. Petrov, *Rubidium Atomic Clock with Improved Metrological Characteristics for Satellite Communication System.* Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 10531 LNCS, pp. 561–568 (2017)
16. V. Formichella, J. Camparo, P. Tavella, Influence of the ac-Stark shift on GPS atomic clock timekeeping. Appl. Phys. Lett. **110**(4), 043506 (2017)
17. A. Petrov, V. Davydov, A. Vologdin, V. Zalyotov, Dependence of microwave-excitation signal parameters on frequency stability of caesium atomic clock. J. Phys. Conf. Ser.**643**(1), 012087 (2015)
18. A. Hudson, J. Camparo, Quantum system dynamics and the quasistatic approximation. Phys. Rev. Appl. **13**(6), 064007 (2020)
19. A. Huber, Hydrogen-deuterium 1S–2S isotope shift and the structure of the deuteron. Phys. Rev. Lett. **80**, 468–471 (1998)

20. A. Petrov, V. Shabanov, D. Zalyotov, A.L.A. Bulyanitsa, D. Shapovalov, Modernization of the frequency synthesizer of cesium atomic clock, in *IEEE International Conferenceon Electrical Engineering and Photonics* (EExPolytech-2018), 8564389 (Saint-Petersburg, 2018), pp. 52–55
21. F. Martinez, *Performance of New GNSS Satellite Clocks* (Karlsruher Institut fur Technologie) Scientific Publishing (2014)
22. A. Bauch, Caesium atomic clocks: function, performance and applications. Sci. Technol. **14**, 1159–1173 (2003)
23. A. Petrov, D. Shapovalov, About the microwave excitation signal formation in the quantum frequency standard on cesium atoms-133. J. Phys. Conf. Ser. **1400**(4), 044008 (2019)
24. J. Kuster, Long-term experience with caesium beam frequency standards, in *Proceeding of the 1999 Joint Meeting of the European Frequency and Time Forum and the IEEE International Frequency control Symposium*, vol 87, issue 5 (1999), pp. 156–168
25. K. Hagimoto, Accuracy evaluation of the optically pumped Cs frequency standard at NRLM. IEEE Trans. Instrum. Meas. **48**, 496–499 (2009)
26. A. Hudson, J. Camparo, Mesoscopic physics in vapor-phase atomic systems: collision-shift gradients and the 0–0 hyperfine transition. Phys. Rev. A**98**(4), 042510 (2018)
27. A. Petrov, Digital frequency synthesizer for ^{133}Cs-vapor atomic clock. J. Commun. Technol. Electron. **62**(3), 289–293 (2017)
28. A. Petrov, On the potential application of direct digital synthesis in the development of frequency synthesizers for quantum frequency standards. J. Commun. Technol. Electron. **63**(11), 1281–1285 (2018)
29. V. Formichella, J. Camparo, P. Tavella, Atomic clocks and the continuous-time random-walk. European Physical Journal B **90**(11), 206 (2017)

Author Index