proceedings of the

27th Small Systems Simulation Symposium 2018

February 12-14
Faculty of Electronic Engineering
Niš, Serbia

Organized by

Faculty of Electronic Engineering Niš

Innovation Centre of Advanced Technologies

Proceedings of the 7th Small Systems Simulation Symposium 2018

Publisher:

Faculty of Electronic Engineering, University of Niš
Aleksandra Medvedeva 14,
18000 Niš
http://elfak.ni.ac.rs

Editor:

Marko Dimitrijević

Printed by:

UNIGRAF X-COPY, Niš

Number of copies printed: 40
Year of publication: 2018

CIP - Каталогизација у публикацији - Народна библиотека Србије, Београд

519.876.5(082)
004.942(082)

SMALL Systems Simulation Symposium (7th ; 2018 ; Niš)
Proceedings of the 7th Small Systems Simulation Symposium 2018, February
12-14, Niš, Serbia / organized by Innovation Center of Advanced Technologies and Faculty of Electronic Engeneering, Niš ; [editor Marko
Dimitrijević]. - Niš : Faculty of Electronic Engineering, 2018 (Niš :
Unigraf X-Copy). - 106 str. : ilustr. ; 30 cm

Tekst štampan dvostubačno. - Tiraž 40. - Bibliografija uz svaki rad. -
Registar.

ISBN 978-86-6125-199-3

a) Симулација - Зборници
COBISS.SR-ID 257014540

--

Preface

To the Proceedings of the 7th Small Systems Simulation Symposium

Dear colleagues,

This year we have the 7th Small Systems Simulation Symposium, 18 years after the first Symposium took place here at the Faculty of Electronic Engineering of the University of Niš. The idea come from professor Vančo Litovski. At the beginning, researchers from University of Southampton, Middlesex University, University from Besancon, Technical University of Vienna, Delft University and others helped SSSS become international symposium. In the following years, support from Balkan countries, as well Germany, Armenia and Spain was invaluable. Thanks to the determination of the researchers from Laboratory for Electronic Design Automation (LEDA) and contribution of our foreign friends, we succeeded to endure an international meeting that lasts all these years. From the beginning, we were trying to promote and review our bi-annual research results and plan future endeavors.

SSSS is dedicated to simulation, which will always impose new and more complex tasks. Past years, functional verification – closely related to simulation – becomes unavoidable task in electronic design automation. This year papers are categorized in three regular sessions dedicated to modeling, circuit and system design and simulations. The authors addressed number of topics, including internet of things, functional verification, electronic component modeling, RF electronic simulation, co-simulation techniques and educational applications. The plenary session will be an opportunity to hear presentations from professor Dimitar Trajanov, Faculty of Computer Science and Engineering, Skopje addressing dark data and IoT, and Goran Panić from IHP Institute, Frankfurt Oder, addressing verification of an embedded sensor node SoC. Symposium also hosts round table which will address cooperation between academic institutions and industry.

I want to express my gratitude to professor Vančo Litovski, who established the symposium, for advice and help. This Symposium would not happen without broad support from Faculty of Electronic Engineering, all members of LEDA laboratory and Innovation Centre of Advanced Technologies, Niš.

Respectfully,

Marko Dimitrijević

Assistant-professor

The 7th Small Systems Simulation Symposium

was supported by

Ministry of Education, Science, and Technological Development of Serbia,

Office for Local Economic Development and Projects, Niš, Serbia

Iritel a.d. Beograd, Serbia

Tagor Electronic d.o.o. Niš, Serbia

Istraživačko-razvojni institut NIRI d.o.o. Niš, Serbia

Photon Optronics d.o.o. Niš, Serbia

Pogled telekomunikacije d.o.o. Niš, Serbia

Klaster Ni-KAT Niš, Serbia

STEERING COMMITTEE

M. Andrejević-Stošović, University of Niš (Serbia)
B. Blanuša, University of Banja Luka (Bosnia and Herzegovina)
S. Bojanić, Universidad politécnica de Madrid (Spain)
I. Bushehri, Lime Microsystems, (United Kingdom)
B. Dokić, University of Banja Luka (Bosnia and Herzegovina)
G. S. Djordjević, University of Niš (Serbia)
N. Janković, University of Niš (Serbia)
M. Jevtić, University of Niš (Serbia)
T. Kazmierski, University od Southampton
V. Litovski, University of Bath (United Kingdom)
V. Melikyan, Director of Educational Department, Synopsys Armenia CJSC
S. Milenković, Lime microsystems (United Kingdom)
D. Milovanović, University of Niš (Serbia)
O. Nieto, Universidad politécnica de Madrid (Spain)
D. Pantić, University of Niš (Serbia)
P. Petković, University of Niš (Serbia)
B. Reljin, University of Belgrade (Serbia)
M. Simić, School of Aerospace, Mechanical and Manufacturing Eng., (Australia)
M. Savić, Lime Microsystems (Serbia)
D. Trajanov, FCSE (FINKI) Skopje (Macedonia)
V. Zerbe, Technical University of Erfurt (Germany)
M. Živanov, University of Novi Sad (Serbia)
M. Zwolinski, University of Southampton (United Kingdom)

ORGANIZING COMMITEE

M. Dimitrijević, University of Niš (Serbia)
M. Andrejević-Stošović, University of Niš (Serbia)
S. Đorđević, University of Niš (Serbia)
B. Jovanović, University of Niš (Serbia)
V. Litovski, University of Bath (United Kingdom)
M. Milić, University of Niš (Serbia)
J. Milojković, ICAT (Serbia)
D. Mirković, University of Niš (Serbia)
P. Petković, University of Niš (Serbia)
Z. Petković, University of Niš (Serbia)
D. Stevanović, ICAT, Niš (Serbia)

SYMPOSIUM SECRETARY

Miljana Milić

Faculty of Electronic Engineering, University of Niš

Aleksandra Medvedeva 14, 18000 Niš, Serbia

Tel: +381 18 529321

miljana.milic@elfak.ni.ac.rs

Proceedings of

The 7th Small Systems Simulation Symposium

Faculty of Electronic Engineering, Niš, Serbia, 12-14 February, 2018

Contents

S3 System Design and simulation

Dark Data in Internet of Things (IoT): Challenges and Opportunities

Dimitar Trajanov, Vladimir Zdraveski, Riste Stojanov and Ljupco Kocarev

Abstract - Nowadays we are witnessing the establishment of the data-driven science as a new scientific paradigm, that is opening a waste amount of new opportunities for scientific and technological advances. The data is becoming the main asset in today's science and technology. Unfortunately, a significant amount of available and stored data is not used today. This data is known as a dark data. Starting from this point, the primary goal of this paper is to raise the awareness of the opportunities that are explored with the dark data utilization in companies and organizations, by giving an overview of the underlining technologies, proposing a methodology and showing example projects that utilize the dark data in the IoT domain.

Keywords – Dark data, Internet of Things (IoT), Machine Learning, Data Science.

I. INTRODUCTION

If we look back in the history of science and technology, four scientific paradigms can be identified and distinguished [1]. Each new paradigm is a revolutionary improvement over the previous one, and it is based on invention and employment of entirely new and different set of tools and methods.

The first paradigm is the science based on empirical observations, and this is the era of Thales, Archimedes, Pythagoras, and Aristotle that continue with the invention of scientific methods which has been employed in the Middle Ages by Ibn al-Haytham and Roger Bacon.

The second paradigm of science and technology started in the mid-17th century with the invention of calculus, by Isaac Newton and Gottfried Leibniz, which forms the core of theoretical models and generalizations. This completely new tool makes a basis for scientific revolution and many new innovations and inventions including the results of James Clerk Maxwell and Albert Einstein.

Many scientific problems and the theoretical models became too complex with the time, so the analytical solution was no longer feasible [2]. With the advancement of computers in 1950, the third paradigm of computational science was born. This has allowed simulations of complex real-world phenomena based on the theoretical models of the second paradigm. Many products that we use today are developed using computer models and simulations like microprocessors, planes, rockets and many more.

The fast development of digital and computer systems in the last few decades, and especially with advancement in the data storage systems resulted in the creation and permanent storage of an enormous amount of data. This data stored in computer readable format recently has given rise to the fourth paradigm, which is known as data exploration or e-Science [1], and it unifies the first three paradigms of empirical observations, theory, and computation and simulation. The term e-Science is often defined as a coupling between accurate information tools, science, and scientists, whereas many of the practical scientific processes are guided and directed by the available data in each domain.

The data-driven science for the first time in the scientific history gives a chance to someone other than human to create new models and programs. With the progress in machine learning algorithms, now it is possible computers to learn the dependencies between parameters and to enable models' creation in areas that were impossible before. Some remarkable examples are models for personalized medicine, human speech recognition, playing Chess or Go, or even driving a car. This fast expansion of the machine learning is tightly connected to the availability of a considerable number of open source and free or commercial libraries and tools that simplify the process of using machine learning techniques [3]. Regarding the programming language popularity in the field of machine learning, Python is a leading one and is followed by R, Java, JavaScript, C and C++ [4].

Internet of Things (IoT) is another important trend that is happening in the last decade, and it is heavily contributing to a faster transition of science through its fourth paradigm. The IoT describes the connection of devices to the internet using embedded software and sensors to communicate, collect and exchange data with each other. IoT combines connectivity with sensors, devices, and people, enabling a form of a free-flowing conversation between human and machine, software and hardware [5]. Furthermore, forecasters predict there will be approximately 30 billion connected things by 2020 [5] These connected things include wearable sensors; smartphones; sensor-embedded gaming systems, such as music players; and in-vehicle sensor devices. Users carrying mobile sensing devices are becoming the source of

Dimitar Trajanov, Vladimir Zdraveski, Riste Stojanov and Ljupco Kocarev are with the Faculty of Computer Science and Engineering – ss. Cyril and Methodius University - Skopje, Ruger Boskovik 16 Skopje Macedonia, E-mail: {dimitar.trajanov, vladimir.zdraveski, riste.stojanov, ljupco.kocarev}@finki.ukim.mk

Ljupco Kocarev is also with Macedonian Academy of Science and Art, Bul. Krste Misirkov 2, P.O. Box 428, Skopje, Macedonia, E-mail: lkocarev@manu.edu.mk

a vast amount of varied data. At the same time, inherent user and device movement and ubiquitous connectivity are creating opportunities for dense spatial and temporal sensing.

Fig. 1. Multiple applications that create big data and percentage of data that will be transmitted and used (Source: Cisco Global Cloud Index, 2015–2020 [7])

This large number of IoT devices by 2020 will generate 600 ZB per year by 2020, 275 times higher than projected traffic going from data centers to end users/devices (2.2 ZB); 39 times higher than total projected data center traffic (15.3 ZB) [7]. Unfortunately, in today's IoT applications only less than one percent of the data is examined, and more than 99 percent of collected data is lost before reaching operational decision makers [8][7]. Most of the data that is actually used, for example, in manufacturing automation systems are used only for real-time control or anomaly detection. A great deal of additional value remains to be acquired, by using more of the data, linking this data to other data sets, as well as deploying more sophisticated IoT applications, such as using performance data for predictive maintenance or analyzing workflows to optimize operating efficiency. So, IoT can be a key source of big data that can be analyzed to capture value, and open data, which can be used by more than one entity [8]. This unused and hidden data that have a potential of creating new values is known as Dark Data.

Based on all previous findings and trends, the main goal of this paper is to raise awareness of opportunities and challenges that come from utilizing IoT created Dark Data for new applications that have a potential of scientific improvement of our lives or to increase productivities and effectiveness in the companies.

II. Dark Data in IoT: Concepts and Supporting Technologies

Gartner defines dark data as "the information assets organizations collect, process and store during regular business activities, but generally, fail to use for other purposes." It includes all data objects and types that have yet to be analyzed for any business or competitive intelligence or aid in business decision making [9]. Organizations often retain dark data for compliance purposes only.

Although this Dark Data may initially appear irrelevant, this information represents a large portion of available opportunities that many companies ignore. There is a substantial risk that comes with not addressing this data. If a company decides not to invest in the analysis and processing of dark data while its competitors do, the company will fall behind. Investing in dark data is an opportunity that greatly outweighs the cost [10].

A. Sources of data and data variety

Traditionally, most organizations primary collect transactional data used to run their business like orders, inventory tracking, customers interactions, operations or financial transaction. The internet services have created some entirely new sets of data that includes social networks data, pictures, videos, and a lot of textual information. But the real explosion of data variety is happening with the establishment and massive development of IoT [11]. The IoT devices are embedded into many objects that we are using today, with the expectation that they will be a part of every object in the near future.

In contrast with computers, the IoT devices are equipped with a broad spectrum of different sensors that gives the opportunity to sense many parameters of the physical environment. Some of the more common sensor types that are part of today's IoT are sensors for: Sound, audio, and acoustics; Pressure and force; Optic, light and imaging; Temperature and thermal; Motion and velocity; Flow, liquid, chemical and gas; Magnetic; Air, water and land pollution; and Proximity, position and presence. Because all those sensors are spatially distributed, the spatial information can be associated with the measurements.

The real-time data from sensors usually is used for alarms or real-time control. To be used in more applications, like optimization and prediction, data from sensors need to be stored at discrete time intervals and then used in new advanced applications. A typical industrial practice involves acquiring one measurement per second from each IoT device [12]. With this resolution of measurements, there will be 60 measurements per minute, 3,600 measurements per hour, leading to 31.5 million measurements per year. If we have an industrial facility with 3,000 parameters that are measured by IoT devices, then 94.6 billion measurements are generated per year. The net result is huge volumes of both structured and unstructured data that need to be analyzed to reveal patterns and associations.

Another type of data that is a real candidate for dark data are audio, video and image files. Contents of these media files, such as the people in the recording or dialogue within a video, will remain locked within the file itself if it is not processed.

Combining operational business data with sensor-generated information has a potential to deliver innovative ways to drive high performance and intelligent decision-making in every industry. So, the real power comes from

the data integrations and building more complex models of the real world.

B. Dark Data and IoT Intersecting Technologies

From the technological point of view, several other emerging technologies and trends are closely related to the Dark Data in IoT. The three concepts are worth to be mentioned here and have or will have a significant impact on improving the data usage in the IoT domain. These technologies are Industry 4.0, Digital Twins and Data Lakes.

In recent years, there have been great advances in Industrial Internet of Things (IIoT) and its related domains, such as industrial wireless networks (IWNs), big data, and cloud computing [14]. These emerging technologies are introducing advanced digitalization within factories, and in combination with future-oriented technologies in the field of "smart" objects (machines and products) have resulted in a new fundamental paradigm shift in industrial production. The future production will be based on modular and efficient manufacturing systems in which products control their own manufacturing process. That is supposed to accomplish the manufacturing of individual products in a batch size of one while maintaining the economic conditions of mass production. Motivated by this future expectation, the phrase "Industry 4.0" was established for a planned "4th industrial revolution" [15]. While in Industry 3.0 the focus was on the automation of single machines and processes, in Industry 4.0 the central goal is the digitization of all physical assets and integration into digital ecosystems with value chain partners.

One of the Gartner top 10 strategic technology trends for 2018, closely connected to the IoT and Dark Data, are Digital twins [16]. A digital twin is a digital representation of a real-world entity or system. The "twin" is the "digital" transformation of a real-world object or system which can be visualized and controlled by an analyst or manager in a manner that is location agnostic [17]. In the context of IoT, digital twins are linked to real-world objects and offer information on the state of the counterparts, respond to changes, improve operations and add value.

From a simulation point of view, the Digital Twin approach is the next wave in modeling, simulation and optimization technology. In the last decades, simulation become a state of the art technology, but it is restricted to a computer and numeric experts that use standard tools to answer specific design and engineering questions [18]. In this direction, digital twin opens two new challenges. The first one is a design of new algorithms and methods for real-time simulations of a variety of different physical objects and processes. The second challenge is how to utilize the collected data from digital twins to improve business and production processes in a company or our lives.

Digital twin in an IoT platform contains two categories of data. The metadata which does not change and includes the details that describe the device such as serial number, firmware version, device location, type of sensors attached, sensors precision, and some other parameters like model or year of manufacturing. The second category of data is real-time sensor and status data of the device.

If we look at the sources for dark data, we can note that it can be represented as any data types: structured, semi-structured or unstructured data. In order to store and manage this data, we need a new storage concept that will support all data types and additionally will support storage of streaming data that come from IoT devices. The methodology that supports all these requirements is the "Data Lake."

A "Data Lake" is a methodology enabled by a massive data repository, based on low-cost technologies, which improves the capture, refinement, archival, and exploration of raw data within an enterprise. A data lake contains raw unstructured or multi-structured data that for the most part has unrecognized value for the firm [19]. The data lakes are typically built to handle large and fast arriving volumes of unstructured data (in contrast to data warehouses' highly structured data) from which further insights are derived. Thus, the lakes use dynamic (not pre-build static like in data warehouses) analytical applications. The data in the lake becomes accessible as soon as it is created (again in contrast to data warehouses designed for slowly changing data). The data lake strategies can combine SQL and NoSQL database approaches and online analytic processing (OLAP) and online transaction processing (OLTP) capabilities [20]. The data lake has emerged as the recognized mechanism to enable organizations to define, manage and govern the use of various big data technologies. That represents an evolution of big data towards the mainstream use in an enterprise and the associated focus on management of such assets [24].

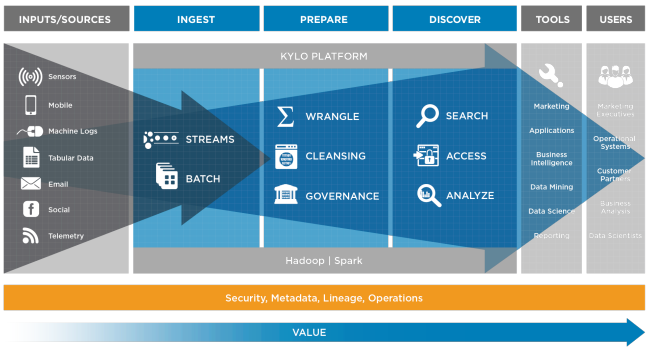


Fig. 2. Kylo Data Lake Architecture

To dynamically handle the structure of stored data, the Data lakes need to include metadata or semantic model of the data that adds a layer of context over the data defining the meaning and interrelationships with other data. Usually, the technologies of the Semantic web that are standardized by W3C are used, enabling easy integration of data stored

in the Data lake with Linked Data Cloud or other semantic web sources [21].

All major cloud providers like Microsoft [22] and Google [23] are offering a Data Lake storage. There is also high-quality open source enterprise-ready data lake management platform like Kylo[1], that enables self-service data ingest and data preparation with integrated metadata management, governance, security and best practices inspired by Think Big's 150+ big data implementation projects. The architecture of the Kylo platform is shown on Fig 2.

C. Security issues

Storing and securing data sometimes can gain more expense (and sometimes greater risk) than value. When both the location and the contents of files are unknown, they can be easily mined by hackers and used against a company. This can result in legal action, high payouts, and lost businesses [13]. Dark data can be a security risk and a barrier to operations when companies are unaware that the data even exists.

From the other side, IoT devices generate vast amounts of data on a daily basis that remain trapped in private infrastructures, due to the inability to control access to them and to connect this data with the rest of the world. By nature, this data is heterogeneous and sensitive, revealing the context, habits, and behavior of the owners. Hence, the publication of such data requires flexible security policies to control the access and interaction rights. However, in emergency scenarios, targeted disclosure of the private information is often needed in order to provide timely response [26].

Traditionally, the protection of the data generated by IoT devices is mainly based on securing the communication channel, most often through Transport Layer Security (TLS) or its Datagram TLS (DTLS) variant [28]. Additionally, the OAuth protocol is often used to limit the access to the available data and services, but it is not possible to include a context that plays a key role in this domain [29]. The IoT devices generate streams of data, and the work in [30] presents how these streams can be protected with centralized policies.

Even though there are models for different aspects of the IoT authorization, such as stream protection, context awareness, information flow control and identity providing (with certificates or OAuth), there is no complete solution that provides policies that cover all these features together. Moreover, the attacks that take over the IoT are becoming more often, introducing the need for device discoverability protection and configuration control. None of the analyzed solutions provides overcoming the heterogeneity in the IoT domain in the process of data protection. Among these challenges, a complete policy model should also cover all the features from the traditional enterprise (API based) systems, since the IoT devices are coordinated and consumed by this kind of applications, and the policies should provide distributed and complete protection of the whole infrastructure

III. Dark Data Collection and Usage Methodology

For any business, data is vital, because it holds the key to successfully manage the company, to attract new customers and increase growth. That is why the big data is big business. Dark data is not just a small portion of big data. It is the biggest slice of the pie and holds a massive amount of potential for those who can control it [13]. But, the central point to realize about dark data is that it does not have to stay dark. At the moment when dark data is used to gain insights, the data becomes actionable and is no longer dark.

D. How to start and build-up on current dark data.

In many cases, the organizations are just not aware of the dark data existence. So, in the beginning, there is a need to raise the awareness of existence and opportunities that can come from the dark data. Afterwards, the infrastructure that will support dark data analytics needs to be put in place. Creating a Data Lake infrastructure is the preferred solution, where gigabytes of data will be moved from multiple locations. This new storage will keep all data in one integrated system, where it will be easy to access and not to be forgotten again.

Based on our previous experience in many data-oriented projects [21][26][27][32][36] the following methodology is proposed:

1. *Get access*. Getting administrative access to everything, including all servers, hard drives and any other storage facilities used
2. *Search for data*. Search and identify all available data sources. Look at the applications, devices, peoples, and processes.
3. *Catalog data*. Analyze and categorize all data that is used by identified data sources, including the data stored in relational databases, logs, text data, multimedia data, IoT streams, IoT metadata, auditing data, and any other data that is stored.
4. *Security and privacy*. In this step, all legality issues need to be identified, and for all datasets, the assessment of security and privacy issues need to be conducted.
5. *Determine the value*. Based on the business needs determine which questions are the most important to be answered first. Identify datasets that will support answers to these questions.
6. *Move the data*. Store all or most of the data in the centralized Data Lake.

[1] https://kylo.io/

7. *Expand the data*. In this step, the goal is to find if there is additional important data that is sensed or collected but not stored. Examples include: some sensor data, intermediate data, additional more detailed log data, or data that is present but is not digitalized. These actions will require additional effort, so some estimate of the value of this data related to the price of getting it will be needed.
8. *Interlink the data*. Data that was collected come from different applications and sources and usually is not interlinked. We need to keep in mind that not only data but also relations carry information. In many cases, this information can be crucial for the business processes and models because it connects two or more different parts of the business.
9. *Link to external data*. Link the data with the external data sources like weather conditions, geolocations, stock exchange, news, large public and open data sets like DBpedia or Wikipedia.
10. *Create new data-driven applications*. Based on the business needs create new data-driven applications. In this process, usually, statistics and machine learning can be used to analyze the data (clustering, PCA, anomaly detection, novelty detection) or to create new models that will be used for predictions. Special emphasis needs to be given to data visualization in order to most effectively communicate the results with the users.

E. How to support future data based services

When you are designing a new service or application, there are two important issues associated with data that need to be followed. First, the collected data should be fine-grained as possible, meaning that you need to collect all possible details that are measured or are available. You can always get summaries and aggregate data, but there is no possibility to go back and derive more detailed data. The second principle is that data need to be available in real time. There is always a possibility to batch data or to slow down, but it cannot be speeded up.

The new developments in IoT like Industry 4.0 and Digital Twins are creating a lot of digital data. Only small portions of this data are available in factories or in our homes, so currently seems that for a big part of this data there are no ideas how to use it. But this does not mean that all these unusable details called “digital exhaust” do not need to be stored [11]. The idea behind the usage of Data Lake as a storage option is to allow an easy way to store many different types of data. This approach significantly reduces the effort that is required to store all available details. The Data Lake allows collection of data for future needs before it’s possible to know what those needs are, so it has tremendous potential. Data is not limited by the scope of thinking present when the data is captured but is free to answer questions we do not know how to ask yet: “Data itself is no longer restrained by initial schema decisions, and can be exploited more freely by the enterprise [24].

Having a lot of details about your business explores the opportunities to create new models and algorithms that would improve the business outcomes. In many cases, there is a general perception that models and algorithms are extremely complicated, but in many cases, there can be a very simple connection between business variables [11].

It is clear that there are real costs associated with the collection, transmission, processing, and storage of data. Some of the new technologies like Data Lake simplify the process and thus lower the costs, but still, there is a need for thoughtful rationalization of how much data is enough.

Processing a large data without a specific purpose in mind will possibly lead to failure. Indeed, dark analytics efforts that are surgically precise in both intent and scope often deliver the greatest value. Like every analytics journey, successful efforts begin with a series of specific questions. What problem are you solving? What would we do differently if we could solve that problem? Answering these questions makes it possible for dark analytics initiatives to illuminate specific insights that are relevant and valuable [25].

IV. DARK DATA UTILIZATION EXAMPLES

In the past couple of years, we have worked on several IoT connected projects where some forms of dark data were used. Three of our projects will be presented in more details, each of them covering different aspects of using dark data. (1) In the first project is a Data Lake platform for Smart City that was created to support storage and analytics of variety of different data types including industrial and personal IoT devices. (2) In the second presented project is about Power Grid Analytics and the concept of interlinking the data with external data sets are shown. (3) The last project shows the creation of metadata in the form of Ontologies and how analysis of dark data that is stored in semi-structured text files can be utilized to automate the process of System of chip synthesis.

A. Smart City Platform

We have developed a concept of a platform that complements the slow low-resolution (annual, quarterly or monthly) city's ISO indicators with high resolution, sensing, social media extracted knowledge and person-centric indicators, which can provide real-time input into models to infer or forecast future smart city indicators. The storage of the data is organized like Data Lake allowing variety of data types to be stored. Therefore, the software solution provides methods for observing and measuring phenomena of common interest (e.g. traffic conditions, air pollution, noise in urban areas), over large geographic areas, which exploits the inherent mobility of sensing devices. By combing with the data from social networks, news, blogs and other data sources this creates a generic

city's footprint that could guide solutions and technologies towards smart city's transition [31].

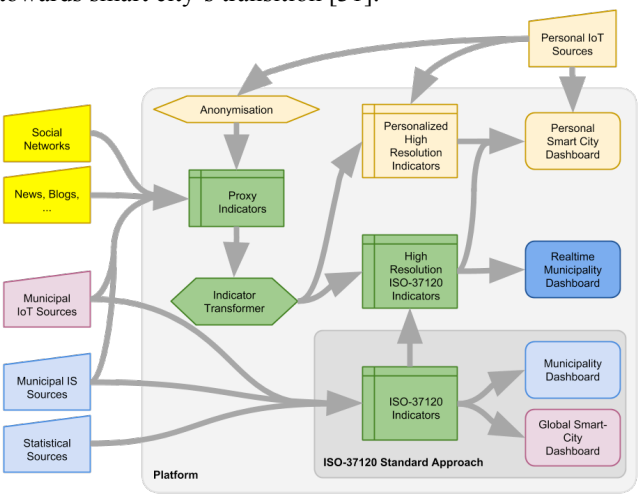


Fig. 3. Smart City Platform Architecture with Data Lake Storage

The core of the architecture, shown in Fig 3., is based on the ISO-37120 indicators (or ISO4City, ISO4C) and extended to a more granular (temporal, spatial) set of high resolution ISO-37120 indicators that provide transformation of the core indicators towards the real-time personalized dashboard. To tune the data, additional customization and contextualization to a viewpoint of a citizen is performed with the personalized high resolution indicators' filter, which is directly affected by the personal IoT sources. The data coming from users passes through anonymization filter, then the whole dataset is reduced through the proxy indicators filter and transformed towards the high-resolution ISO-37120 indicators, using the indicator transformer. The block Personal IoT sources represents integration with the personal (citizen's owned smart devices) IoT sensors. The specific Personal IoT block is designed to support the integration of personal IoT devices to the platform. The Personal IoT module architecture is component-based and enables easy development of new cartridges to support the integration of different classes of IoT devices. Integration of personal IoT raises the questions of privacy and security of the personal data.

B. Power grid analytics

The power grid ontology (PGO), shown in Fig 4, introduces a data model for power distribution system's data annotation. PGO is developed on top of schema.org and reuses and inherits many entities and properties. However, for the most specific domain requirements new entities and properties are introduced. The core entity is Node that represents a node in the power network, such as generator, substation, pillar of a transmission line and a power meter. Generators could be renewable such as wind turbine, solar, biomass, geothermal and hydro turbine and nonrenewable such as nuclear, coal, natural gas, crude oil and petroleum. After a sufficiently large data set is annotated with the PGO, it could be used to generate (periodically and continuously) a set of reports, such as to find the most over-loaded node or transmission line or to find the nodes with variable frequency or most frequent voltage drops. Several power grid data sets have been published recently, such as SciGRID[2] and GridKit[3] , that we have used to evaluate our concept and the ontology itself.

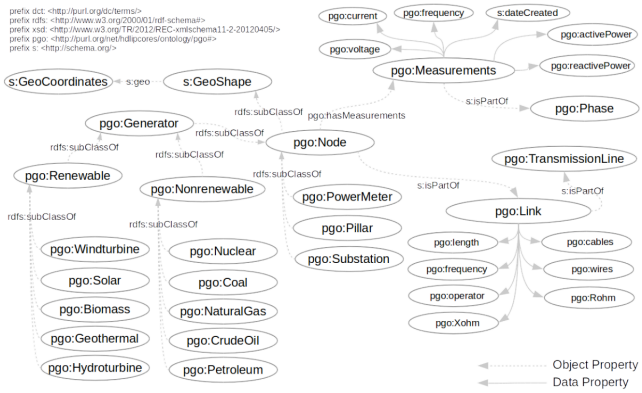


Fig. 4. Power grid ontology provides a data model for power distribution system's data (such as network topology and continuous and periodic measurements) annotation

As an example of linking data form the Power Grid with external data sources (DBPedia[4]), and introducing more granular indicators, we could find the for-example cities placed in radius of 20 km (or any other distance) from a power plant and calculate the total population living around the power plants. The results from this interlinking is shown on Table 1.

Table 1 - Cities in radius of 20 km from a power plants

Power Plant	Distance (m)	City	Population
Gemeinschaftskraftwerk Kiel	3021.53	Kiel	240832
Koepchenwerk	5607.79	Hagen	191241
Koepchenwerk	11579.8	Dortmund	575944
Koepchenwerk	18027.1	Bochum	361876
Kraftwerk Scholven	10144.6	Bottrop	117450
Kraftwerk Scholven	11287.4	Gelsenkirchen	260900
Kraftwerk Scholven	15564.4	Herne, N. R.-W.	166187
Kraftwerk Scholven	15747.0	Oberhausen	214990
Kraftwerk Scholven	16571.7	Essen	589075
Kraftwerk Scholven	19545.8	Bochum	361876
Statkraft Kraftwerk Knapsack II	11083.7	Cologne	1057327
TOTAL			4137698

C. HDL IP Cores system

In the HDL IP cores system first we have created meta data in the form of Ontologies, and we designed the system for synthesis of System on Chip that utilize the dark data that is extracted and analyzed from semi-structured HDL source text files.

[2] SciGrid, http://scigrid.de/
[3] GridKit, https://github.com/bdw/GridKit
[4] DBPedia SPARQL endpoint, http://live.dbpedia.org/sparql

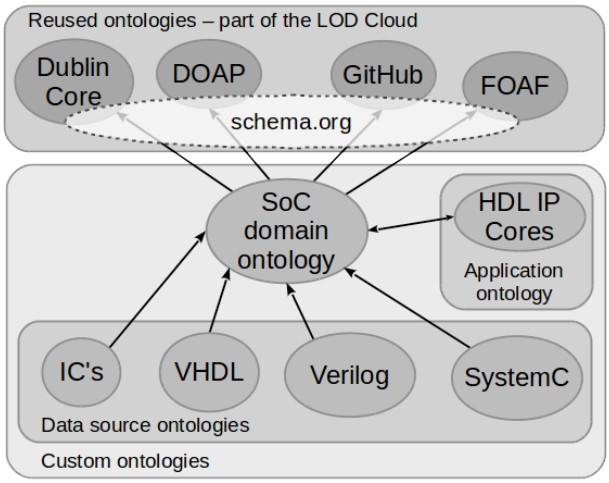


Fig. 5. SoC design ontologies

The Linked Data best practices [32] and the concept of semantic systems development, explored in [33], states that for any semantic system there must be a domain ontology, describing the knowledge in the given domain, a data-source ontology, strongly related to the type of the data source, and an application ontology, which describes the application entities and relations. Following this paradigm, we developed a set of ontologies, shown in Fig 5.

The ontologies on the top of the cloud in Fig 5. have already been mentioned and the ontology in the middle (SoC) is our domain ontology. It serves as a basic hardware architecture description schema, sufficient to annotate an existing system on a chip. At the bottom, there are ICs, VHDL, Verilog and SystemC data source ontologies, each extending the main hardware concepts from SoC, but providing additional language specific schema, suitable for its specific HDL. At the right-hand side of Fig. 5. SoC design ontologies, the HDL IP Cores application ontology (hipc.owl) is shown. It covers all classes and relations required to deploy a functional application and with that offers client-side features which provide a novel concept in the storage and retrieval of HDL IP Cores [34][35][36].

V. Conclusion

In this paper, the current trends and technologies related dark data in the IoT domain were presented. In order to utilize the benefits, of data driven science and machine learning, we need to get access and utilize all available data. To apply machine learning to your domain, you need a lot of data. There are two solutions for that. First is to create new applications that collect all data that is required to train machine learning based models, then to deploy the application, collect data for some time, and at the end use the collected data as an input to a machine learning algorithm. The second approach is to identify all white and dark data that is already in the system, to collect them and then to try to create machine learning algorithms based on the currently collected data. The advantage of the second approach is that it is cheaper, faster and what is more important it can provide access to a very large number of historical data.

Based on current trends in science and technology, companies first need to manage to store as much data as possible, and then they find a way to use this data creatively and innovatively that will give them a key completive advantage.

References

[1] T. Hey, S. Tansley, K. M. Tolle, et al., The fourth paradigm: data-intensive scientific discovery, vol. 1. Microsoft research Redmond, WA, 2009.

[2] Agrawal, Ankit, and Alok Choudhary. "Perspective: materials informatics and big data: realization of the "fourth paradigm" of science in materials science." APL Materials 4, no. 5 (2016): 053208.

[3] Joseph Misiti, A curated list of awesome Machine Learning frameworks, libraries and software, Available online at: https://github.com/josephmisiti/awesome-machine-learning, Accessed: 10 Jan 2018.

[4] Jean Francois Puget,"The Most Popular Language For Machine Learning Is", IBM Developer Works Blog, https://www.ibm.com/developerworks/community/blogs/jfp/entry/What_Language_Is_Best_For_Machine_Learning_And_Data_Science?lang=en. (2016)

[5] Ernst & Young Report, "Internet of Things: Human-machine interactions that unlock possibilities", Ernst & Young Global Limited, (2016).

[6] A. Nordrum, "Popular Internet of Things Forecast of 50 Billion Devices by 2020 Is Outdated," IEEE Spectrum, 18 Aug. 2016; http://spectrum.ieee.org/tech-talk/telecom/internet/popular-internet-of-things-forecastof-50-billion-devices-by-2020-is-outdated.

[7] Cisco White Paper, "Cisco Global Cloud Index: Forecast and Methodology, 2015–2020", Cisco Public, 2016

[8] Bughin, J., J. Manyika, J. Woetzel, M. Chui, P. Bisson, and R. Dobbs. "The Internet of Things: mapping the value beyond the hype.", McKinsey Global Institute (2015).

[9] Sony Shetty, How to Tackle Dark Data, Gartner, (2017), https://www.gartner.com/smarterwithgartner/how-to-tackle-dark-data/

[10] Frank Moreno, "Dark data discovery: Improve marketing insights to increase ROI", IBM Business analytics blog, Sep 20, 2016, https://www.ibm.com/blogs/business-analytics/dark-data-discovery-improve-marketing-insights-to-increase-roi/

[11] Rossman, John. "The Amazon Way on IoT: 10 Principles for Every Leader from the World's Leading Internet of Things Strategies", Clyde Hill Publishing, 2016

[12] Fuhr, Peter L., Marissa E. Morales Rodriguez, Sterling Rooke, and Penny Chen. "Convergence and

Commercial Momentum-Industrial Internet of Things Evolution." InTech 2017, no. 2 (2017).
[13] Stephen Mackey, "The Rise of Dark Data and How It Can Be Harnessed", KDnuggets: Opinions, Interviews, Reports, https://www.kdnuggets.com/2016/03/rise-dark-data-how-harnessed.html (2016)
[14] Wan, Jiafu, Shenglong Tang, Zhaogang Shu, Di Li, Shiyong Wang, Muhammad Imran, and Athanasios V. Vasilakos. "Software-defined industrial internet of things in the context of industry 4.0." IEEE Sensors Journal 16, no. 20 (2016): 7373-7380.
[15] Lasi, Heiner, Peter Fettke, Hans-Georg Kemper, Thomas Feld, and Michael Hoffmann. "Industry 4.0." Business & Information Systems Engineering 6, no. 4 (2014): 239-242.
[16] Kasey Panetta, "Gartner Top 10 Strategic Technology Trends for 2018", Gartner, (2017)
[17] Datta, Shoumen Palit Austin. "Emergence of Digital Twins-Is this the march of reason?." Journal of Innovation Management 5, no. 3 (2017): 14-33.
[18] Rosen, Roland, Georg von Wichert, George Lo, and Kurt D. Bettenhausen. "About the importance of autonomy and digital twins for the future of manufacturing." IFAC-PapersOnLine 48, no. 3 (2015): 567-572.
[19] Fang, Huang. "Managing data lakes in big data era: What's a data lake and why has it became popular in data management ecosystem." In Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), 2015 IEEE International Conference on, pp. 820-824. IEEE, 2015.
[20] Miloslavskaya, Natalia, and Alexander Tolstoy. "Big Data, Fast Data and Data Lake Concepts." Procedia Computer Science 88 (2016): 300-305.
[21] Trajanov, Dimitar, Riste Stojanov, Milos Jovanovik, Vladimir Zdraveski, Petar Ristoski, Marjan Georgiev, and Sonja Filiposka. "Semantic sky: a platform for cloud service integration based on semantic web technologies." In Proceedings of the 8th International Conference on Semantic Systems, pp. 109-116. ACM, 2012.
[22] Ramakrishnan, Raghu, Baskar Sridharan, John R. Douceur, Pavan Kasturi, Balaji Krishnamachari-Sampath, Karthick Krishnamoorthy, Peng Li et al. "Azure Data Lake Store: A Hyperscale Distributed File Service for Big Data Analytics." In Proceedings of the 2017 ACM International Conference on Management of Data, pp. 51-63. ACM, 2017.
[23] Halevy, Alon Y., Flip Korn, Natalya Fridman Noy, Christopher Olston, Neoklis Polyzotis, Sudip Roy, and Steven Euijong Whang. "Managing Google's data lake: an overview of the Goods system." IEEE Data Eng. Bull. 39, no. 3 (2016): 5-14.
[24] Amber Lee Dennis, "Data Lakes 101: An Overview", Dataversity, August 25 2016, Available online at: http://www.dataversity.net/data-lakes-101-overview/
[25] Tracie Kambies, Paul Roma, Nitin Mittal, Sandeep Kumar Sharma, "Dark analytics: Illuminating opportunities hidden within unstructured data", Deloitte Insights Tech Trends 2017, (2017),
[26] Stojanov, Riste, Sasho Gramatikov, Igor Mishkovski, and Dimitar Trajanov. "Linked Data Authorization platform." IEEE Access (2017).
[27] Jovanovik, Milos, and Dimitar Trajanov. "Consolidating drug data on a global scale using Linked Data." Journal of biomedical semantics 8, no. 1 (2017): 3.
[28] A. Niruntasukrat, C. Issariyapat, P. Pongpaibool, K. Meesublak, P. Aiumsupucgul, and A. Panya, "Authorization mechanism for mqtt-based internet of things," in Communications Workshops (ICC), 2016 IEEE International Conference on, pp. 290–295, IEEE, 2016
[29] Fremantle, Paul, and Benjamin Aziz. "OAuthing: privacy-enhancing federation for the internet of things." In Cloudification of the Internet of Things (CIoT), pp. 1-6. IEEE, 2016.
[30] B. Carminati, E. Ferrari, J. Cao, and K. L. Tan, "A framework to enforce access control over data streams," ACM Transactions on Information and System Security (TISSEC), vol. 13, no. 3, p. 28, 2010.
[31] Zdraveski, Vladimir, Kostadin Mishev, Dimitar Trajanov, and Ljupco Kocarev. "ISO-Standardized Smart City Platform Architecture and Dashboard." IEEE Pervasive Computing 16, no. 2 (2017): 35-43.
[32] Schmachtenberg, Max, Christian Bizer, and Heiko Paulheim. "Adoption of the linked data best practices in different topical domains." In International Semantic Web Conference, pp. 245-260. Springer, Cham, 2014.
[33] Hebeler, John, Matthew Fisher, Ryan Blace, and Andrew Perez-Lopez. Semantic web programming. John Wiley & Sons, 2011.
[34] V. Zdraveski, M. Jovanovik, R. Stojanov, and D. Trajanov, "Hdl ip cores search engine based on semantic web technologies," ICT Innovations 2010, Communications in Computer and Information Science, vol. 83, no. 2, pp. 306 – 315, 2011.
[35] V. Zdraveski, A. Dimitrovski, D. Trajanov. "HDL IP Cores System as an Online Testbench Provider". Small Systems Simulation Symposium, Volume: 5th, Nis, Serbia, February 2014.
[36] V. Zdraveski, M. Jovanovik, R. Stojanov, D. Trajanov. "HDL IP Cores Searh Engine Based on Semantic Web Technologies".

Verification of an Embedded Sensor Node System-on-Chip

Goran Panić

Abstract - In this paper the verification methodology for an embedded low power sensor node system-on-chip design has been presented. A mixed-signal, power-gated, processor-based sensor node microcontroller has been implemented and verified. The chip implements a number of peripherals, several analog components and a flash memory for program storage. The paper describes applied verification methodology including simulation steps, power analysis and chip measurements.

Keywords – sensor node, low power, system-on-chip, system verification.

I. Introduction

Nowadays, advanced system-on-chips (SoCs) implement not only digital components, SRAMs and I/Os, but also analog components, sensors, passives and non-volatile memories integrated on chip [1]. As most modern SoCs are developed for use in battery-powered devices, the power is considered to be one of the most important design constraints. The power reduction in SoC is accomplished by implementation of advanced low power techniques, which introduce new challenges to design and verification methodology [2]. Low power SoCs support complex power management schemes that allow alteration of supply voltage or complete shut-off of inactive parts of the chip. Usually, a low power system is designed to support several power modes controlled by a power control unit. In a specific power mode, selected chip components are powered down or they are set to low voltage or retention state. Consequently, the verification of such systems is a challenging task, since the verification methodology must prove correct chip functionality with respect to all implemented power modes [3].

Additional complexity to SoC verification is introduced by the integration of analog components. The functionality of an analog circuit cannot be efficiently verified in a digital simulation. Even though, the behavior of analog circuits can be described in a mixed-signal language such as Verilog-AMS [4], which is supported by most verification tools, it is still required to perform full SPICE level simulations of an analog circuit in order to create a golden simulation model to be used as a reference to the behavioral model. A behavioral Verilog-AMS model might be applied in a digital-only simulation for specific tests, but at the cost of increased simulation time. However, designers often choose to test only the interface to an analog block by using a simplified verification model of the analog-to-digital interface.

The verification of a processor-based SoC requires thorough testing of processor's instruction set as well as the functional testing of peripherals and system bus. If a program memory, such as an embedded Flash, is integrated on chip, the verification methodology must prove correct functionality of memory erase, write and read tasks initiated from memory controller or a dedicated port. All system components, memory access and communication between processor and peripherals must be functionally verified.

Finally, a verification methodology must be created in a way that it provides all required inputs for efficient chip testing after the fabrication. During post-production measurements, the stimuli from a functional simulation are applied to the chip and the chip response is being recorded. The measurements results are then compared to the results of functional simulation.

This paper describes the verification methodology applied to the design of an embedded sensor node microcontroller. The designed sensor node SoC integrates a 16-bit RISC processor, multiple I/O peripherals, crypto cores, embedded Flash, SRAMs and analog components. It also implements advanced low power techniques such as power gating and frequency islands. The chip was fabricated and tested.

The rest of the paper is organized as follows: The Section II discusses basic tasks of design verification. The Section III gives details on the target design and Section IV describes applied verification methodology. Section V concludes the paper.

II. Design Verification

Design verification applies set of actions in order to prove physical, electrical and functional correctness of design. The design is verified for power, timing, physical correctness and functional behavior. Functional verification is a fundamental step in design verification, and it is tightly coupled with timing verification. Functional verification is typically done in a functional simulation. Most verification tools support direct RTL (register transfer level) simulations based on common hardware description languages (VHDL or Verilog) as well as more complex methodologies such as universal verification methodology (UVM) defined for simulations based on SystemVerilog, a hardware description and verification language that allows complex, assertion-based simulation scenarios [5]. The

Goran Panić is with IHP GmbH, Im Technologiepark 25, 15236 Frankfurt Oder, Germany, E-mail: panic@ihp-microelectronics.com

UVM provides a SystemVerilog base class library helping users to create well-structured testbenches with clear separation between code structures related to the verification environment and those related to the creation of stimuli [6]. It also allows portability of verification data, reusability of verification tests and enables good interoperability between multiple verification engineers and hardware designers [7].

Per definition, a SoC integrates a number of components such as processor or advanced peripherals on a single chip. Many integrated SoC components are pre-qualified IP-Cores (intellectual property cores) already proven in an in-depth stand-alone verification process. Therefore, it is common to divide the SoC verification problem to the verification of IPs and top-level verification. The IP-level verification requires full-coverage verification of an IP, based on a random simulation technique that includes coverage monitors and scoreboard checkers integrated in the testbench environment. The IP-testbench can be reused in the top-level simulation if desired. The top-level verification requires intensive verification of certain system-level aspects such as communication between components, power management control and bus performance. The top-level simulation usually includes a number of direct tests for peripherals as well.

The back-annotation includes functional verification of design after it has been synthesized or layouted. The post-synthesis verification proves design functionality with respect to timing, which is defined by signal delays in synthesized cell gates. The delays in wires are approximated by so called wire-load models that give statistical approximation of wire delays. In the post-layout verification the wire delays are calculated by RC extraction performed on place- and-routed design. The timing simulation is performed for multiple corners with respect to different power modes that are associated to specific supply voltages and clock frequencies. The quality of clock tree insertion is also verified in the post-layout simulation.

Power verification includes analysis of leakage and dynamic power, analysis of electromigration and signal integrity, checking of voltage drop, and analysis of rush currents and turn-on time. Early power analysis is performed early in the design flow and full analysis is done in the back-end flow. Physical verification searches for violations in design geometry and connectivity, and it is performed during the sign-off checks.

Finally, fabricated SoCs are tested and measured on wafer. The wafer tests include functional tests as well as structural tests that identify physical failures in the chip such as scan chain tests. The chip measurements also prove the compliance to electrical specifications for the chip.

III. SENSOR NODE DESIGN

In focus of this paper is verification of a complex sensor node system on chip designed for sensor applications with strong security demands [8]. To better understand the complexity of the design to be verified, the architecture of the target system, as well as, the steps related to system design and implementation are briefly described.

A. System Architecture

A sensor node microcontroller, which architecture is shown in Fig. 1, has been designed to cope with the security problems in wireless sensor networks (WSN). The security challenges in WSN have been addressed by hardware implementation of several advanced crypto algorithms. The system integrates hardware accelerators for AES (advanced encryption standard), ECC (elliptic curve cryptography), and SHA-1 (secure hash algorithm). The core of the system is a 16-bit RISC processor compatible to Texas Instruments MSP430X architecture. Additionally, the chip integrates a baseband processor core supporting direct-sequence spread spectrum, a spread spectrum modulation technique used to reduce overall signal interference. The system implements 16 kB of RAM and 64 kB of non-volatile sector-erasable Flash enhanced with error correction capability (EDAC). The designed SoC integrates a number of peripherals, e.g. I/O digital ports, timers, serial ports and controllers for Flash and analog-to-digital converter (ADC). The clock distribution is provided by an integrated DCO and two external sources, one for slow and one for fast clock input. The integrated clock controller allows individual clock setup for each peripheral. The ADC is interfaced to an integrated preamplifier circuit that provides the connection to an external biomedical sensor (BMS). The power management control is maintained by the power controller unit. The system implements five power-gated islands allowing power shut-down of selected crypto cores, the baseband core and UART blocks. The system also supports global clock gating of all peripherals. The debug capability is provided through a dedicated I2C debug port.

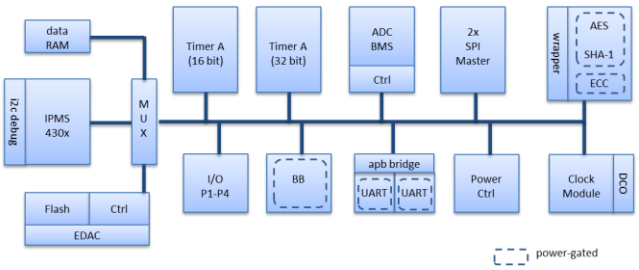


Fig. 1. Sensor node microcontroller

B. System Design and Implementation

The first step in the design of a sensor node microcontroller was system planning that included the exploration of architectures, algorithms and protocols related to the system application. The planning phase resulted in the decision on hardware/software partitioning

and the definition of system architecture and power saving strategy for design.

The RTL design process included HDL (hardware description language) design and verification of system components and system itself. Some system components have been taken from existing IP libraries and some have been designed from scratch. The power control logic for selected power saving scheme was designed and verified in RTL as well.

The implementation steps followed the completion of RTL design for the system. The implementation flow started with the creation of power intent description file for the design. The power intent file contains a set of TCL commands, which describe power domains and supply voltages in the design. It also contains the information on low-power specific cells and libraries to be used in the implementation flow. The design power intent was described in CPF format (common power format) supported by the Cadence power-aware implementation tools. The CPF was also used in power-aware RTL simulations and back-annotations to verify correct behavior of the implemented power controller and to test the state switching between different system power modes.

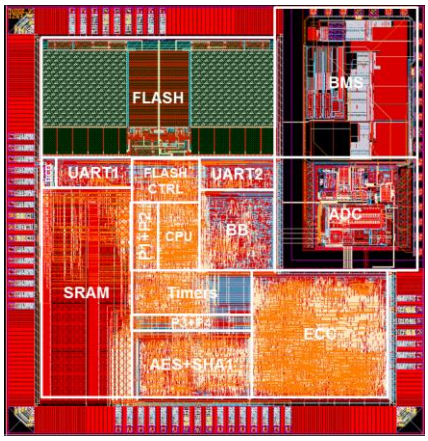


Fig. 2. Sensor node chip layout

The implementation steps included synthesis of the RTL design to physical gates and the backend implementation steps (floorplanning, clock tree insertion, place and route). The creation of power islands was performed during the floorplanning phase. The resulting layout of the system is shown in Fig. 2. The chip was implemented in IHP 0.25 um BiCMOS technology, and it runs with maximum frequency of 11.4 MHz. The estimated and measured power consumption of the chip was around 10 mW at 1 MHz.

IV. Verification Methodology

The verification of the designed SoC included functional and power verification during design process and the testing of fabricated chip. The functional verification included simulations of digital and analog parts and system level simulations. Before the system integration had started, each IP peripheral was thoroughly tested in a dedicated simulation environment. Only UARTs were tested by a UVM-enabled verification process using SystemVerilog verification models. All other IP-cores had their dedicated Verilog/VHDL simulation environment. The peripheral tests included exhaustive tests of IP functionality performed on the RTL-level. The analog components were tested in pure analog and mixed-signal simulations, including schematic simulations and SPICE simulations of the extracted layout netlists. The analog cores were also characterized for power and checked for design rule violations. The system-level tests were created to prove the overall system functionality with respect to timing and power constraints.

A. System Level Tests

The top-level simulation suite included a set of more than 30 test programs that execute from Flash. The test programs implement program routines designed to verify basic peripheral functionality and the functionality of whole system. In a simulation, the testbench loads the image of a test program to the Flash functional model and starts the execution. The results of test operations are then written to external ports and evaluated in the testbench. Additionally, the traces of signal changes during the simulation were captured in an EVCD file (extended value change dump), a standard industry-enabled file format for capturing simulation data, which was then used as an input for post-production tests. Since the Flash programmability was provided by the I2C debug port, a number of I2C-based tests were created to test the basic Flash operations, erase, write, read, and sector-based operations. For that purpose, a simulation model of I2C debug device was created and implemented in the test environment. The I2C-based simulations were used to simulate the instruction set of the processor and to create the program loading sequences for post-production tests. A simplified illustration of the applied verification environment is shown in Fig. 3.

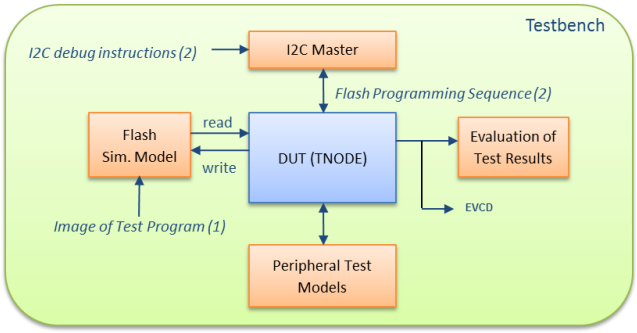


Fig. 3. Testbench environment: (1) tests executed from Flash; (2) tests executed via I2C debug port.

The behavior of power control logic was tested using dedicated test scenarios that forced controller to switch between different power modes of the chip. Those simulations were performed using power-aware capability of the Cadence simulation tool (Innovus) able to interpret design power intent information specified in a CPF file. The CPF-based simulations were designed to check for correct functionality of the power controller unit and to report eventual power state violations. The signal activity from the power-related simulations was captured in a full-scope EVCD file that is used when performing dynamic power simulations of the routed design. The results of power estimations for different case scenarios were compared in post-production to the power measurement results of fabricated chips.

B. Post-Production Tests

The post-production tests included functional wafer testing of fabricated microcontroller chips and power measurements. The tests were performed with an industrial tester device (Agilent Verigy 9300) capable for both wafer and package testing. The tester is constructed to operate with input stimuli extracted from EVCD files generated in simulations. A standard test procedure applies input stimuli to the device under test and samples the response on the chip output ports. The recorded output values are then compared to the expected values extracted from the same EVCD file. The measured sensor-node chip could not be stimulated externally, since it was designed to read the program data from an integrated Flash memory. Therefore, the test stimuli data had to be written to the Flash memory and executed from it. For that purpose a set of initial programming sequences that program the Flash via I2C debug port was created. The I2C programming sequence was erasing the Flash, writing the test program to it and reading the code back to check if it was correctly written. Following this approach, a complete set of test programs was sequentially stored and executed from Flash in a fully automatized test procedure.

The basic functionality of ADC was tested separately by applying predefined DC values to ADC inputs and reading back the converted data from the ADC internal registers. The extracted values were compared to the expected values obtained from analog simulations.

The chip power consumption was measured for representative test cases and compared to the power estimation results. Finally, the functioning chips were selected for packaging and further utilization.

V. CONCLUSION

The verification of a complex SoC is a challenging task requiring detailed planning of verification methodology with respect to the design process and post-production chip measurements. This paper described a verification approach in the design of a complex sensor node microcontroller SoC. The paper discussed main aspects of SoC verification and gave the details of the applied verification methodology. A special attention was given to the verification challenges introduced by implementation of low power techniques and integration of analog components in the system. The practical approach to the verification of a Flash-based system is described in detail.

ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Commission's Horizon 2020 under grant agreement from project myAirCoach – No. 643607.

REFERENCES

[1] Arden, W., Brillouët, M., Cogez, P., Graef, M., Huizing, B., & Mahnkopf, R., "*More-than-Moore white paper. Version 2"*, ITRS, 2010.

[2] Flynn, D., Aitken, R., Gibbons, A., & Shi, K., "*Low power methodology manual: for system-on-chip design"*, Springer Science & Business Media, 2007.

[3] Srikanth, J., Janick, B., Yoshio, I., & Flynn, D., "*Verification Methodology Manual for Low Power*," Synopsys, 2009.

[4] Kundert, K., & Zinke, O., "*The designer's guide to Verilog-AMS*," Springer Science & Business Media, 2006.

[5] Mintz, M., & Ekendahl, R., "*Hardware Verification with System Verilog: An Object-Oriented Framework (Vol. 230)*," Springer Science & Business Media, 2007.

[6] Accellera, "*UVM 1.0 Reference Manual*", 2011

[7] Yun, Y. N., Kim, J. B., Kim, N. D., & Min, B., "*Beyond UVM for practical SoC verification*", In SoC Design Conference (ISOCC), 2011 International (pp. 158-162). IEEE, 2011.

[8] Panić, G., Stecklina, O., & Stamenković, Z., "*An embedded sensor node microcontroller with crypto-processors*," Sensors, 16(5), 607, 2016

Low-Power and Low-Cost Dedicated Bit-Serial Hardware Neural Networks for Epileptic Seizure Prediction

Si Mon Kueh and Tom J Kazmierski

Abstract— **This paper investigates the feasibility of using bit-serial architecture as a method of designing an extremely low-power and low-cost neural network processor for epilepsy seizure prediction. The concept of a novel bit-serial data processing unit (DPU) is presented which implements the functionality of a complete neuron and uses bit-serial arithmetic. An array of these DPUs are controlled by a simple finite state machine. We show that epilepsy detection through such low-cost and low-energy dedicated neural hardware is feasible. The proposed processor extracts epileptic seizure characteristics from electroencephalogram (EEG) waveforms. In order to facilitate the classification of EEG waveforms we develop a dedicated feature extraction hardware that provides inputs to the neural network. This approach has been tested using various network configurations and has been compared with related work. A sample complete system which can predict epileptic seizures with high accuracy has been implemented on an ALTERA Cyclone V FPGA and the hardware uses 3088 ALMs which constitutes about 5% of the Cyclone V A7 capacity.**

I. Introduction

In recent years, the World Health Organization (WHO) have found that 50 million of the world's population are affected by a hidden disability known as epilepsy [1]. Approximately 80% of the reported epileptic cases occur in developing countries where readily available treatment facilities and medications are not generally accessible. Currently, epilepsy is commonly treated with the use of anti-epileptic drugs (AEDs). Early and accurate seizure prediction is essential in preventing seizures by the timely administration of such drugs. Existing state-of-the art seizure prediction systems rely on complex software methods and require significant C PU p ower. T hese m ethods u se elaborate mathematical models of non-linear dynamic systems which are solved using time-domain or frequency-domain analysis [2]. Artificial N eural N etworks (ANNs) h ave also been shown to predict epileptic seizures reliably with an accuracy over 90% [3]. ANNs are an efficient classifier used commonly in conjunction with linear numerical methods of feature extraction to facilitate epilepsy detection. However, as of today, there is still no reliable, home-based and low-cost seizure prediction system which could be used as an aid for timely administration of AEDs and used by an individual epileptic patient. In this paper we propose to consider simple, dedicated hardware neural networks that are optimised for seizure prediction from electroencephalogram (EEG) waveforms and can be personalised to reflect the characteristics of an individual patient. Such systems can be implemented in the form of affordable, wearable equipment without the need to resort to complex software and powerful computers. Firstly, we briefly review state-of-the-art seizure detection methods. We focus on linear seizure prediction models [4] which have the advantage of simplicity and versatility compared with non-linear ones that are capable of addressing the non-stationary nature of the EEG signals. Secondly we present a dedicated hardware implementation of an artificial neuron, based on a bit-serial Data Processing Unit (DPU) which is extremely small and can be used in vector arrangements where a single sequential controller drives an array of such DPUs. We demonstrate that the proposed DPU has has the capability of simulating a biological neuron and can be expanded into a neural network that successfully differentiates between epileptic seizure and non-seizure EEG waveforms. The EEG waveforms used in our investigation are taken from real patients and available online [5] in public domain.

Faculty of Physical Sciences and Engineering, University of Southampton, U.K. {[1]smk1g10, [2]tjk}@ecs.soton.ac.uk

II. Conventional Classification Methods for Epilepsy Detection

Here we summarise briefly the main conventional classification techniques for machine learning which are applicable to medical diagnosis including epilepsy. These methods are the Naive Bayes Classifier, Decision Tree Classifier, k-Nearest-Neighbours (k-NNs) Classifier, support vector machines (SVM) and classifiers based on neural networks. They are briefly reviewed in the following subsections.

A. Naive Bayes (NB) Classifier

NB is a simple probabilistic classifier utilising the Bayes Theorem. It can also be considered as a conditional probability model. This classifier is often used in data mining and it is also applicable to automated medical diagnosis thus making it suitable for epilepsy detection. The Naive Bayes classifier uses the independence assumption that focuses on each feature independently of each other while ignoring any possible correlation between the features [6]. One of the main advantages of using the Naive Bayes classifier is the limited use of training data for classification.

B. Decision Tree Classifier (DTC)

Decision trees are an efficient way to classify sets of data. As a sample is only tested against a subset of the classes, traversing a decision tree does not require complex computations. It has been suggested recently [7] to use neural networks are used in the design of a DTC. There are a few disadvantages when using a decision tree. It will

not be as accurate as the other classifiers. Furthermore, the performance of the DTC will heavily depend on the effectiveness of the particular design [7]. DTCs tend to be less robust than other methods as a very small change in the training datasets might result in a huge change in the output prediction.

C. k-Nearest-Neighbours (k-NN) Classifier

A k-NN classifier is a non-parametric, non-linear yet relatively simple classifier. This classifier is effective when dealing with large data sets. It relies on class assignment based on a nearby data set where similarities between the samples used are measured with a distance function. A recent work [8] points out that k-NN is applicable to medical classification problems. The basic algorithm for a k-NN classifier is relatively similar to that of a neural network classifier. Both have a training stage and a prediction stage. The training stage of the k-NN classifier involves all the different samples which are stored in a form of memory. A neural network, on the other hand, uses the training stage to calculate the weights with the highest accuracy to predict a target output.

D. Support Vector Machine (SVM) Contribution to Epilepsy Detection

SVMs have been used to analyse EEG signals which contain a great deal of detail about the brain activity. A smart sensor IC was proposed [9] with a CMOS chip that has an area of $0.35u$m for scalp EEG acquisition. This chip is integrated with the local processing of the sensor node. Feature vectors of the signal are extracted and classified through machine learning. In order to produce a functional system for epilepsy detection, a number of sensors would have to be worn to achieve spatial correlation. Each individual output of the classifier could then be combined to detect the onset of an epileptic seizure.

Support Vector Machines have also been used in lung cancer diagnosis in conjunction with image processing techniques [10]. SVMs are suitable for such applications as they possesses the advantage of high generalisation and an assurance of global optimisation. They have been successfully used in many other fields that require classification.

E. Automatic Epilepsy Detection Using Artificial Neural Networks (ANNs)

It is possible that the prediction of the onset of a seizure occurrence can be achieved with the assumption that the EEG generated is a very complex but linear system. However, the brain is non-linear. By analysing the power spectrum, it is also possible to continue the analysis through a linear approach [11]. Back propagation neural networks include two stages, a forward propagation stage and a back propagation stage. The normal neural operation uses the forward propagation to pass along the EEG sample provided along the input layer to the hidden layer where calculations are being made which in turn is passed to the output layer to produce the output sample of the neural network which can determine if a seizure occurrence will appear with the input EEG sample. The back propagation stage includes a learning process which reduces the error between the calculated output sample and the target output (possibility of seizure occurrence). This process is performed by adjusting the weights of the neural network in real time [11]. Spiking Neural Networks (SNNs) are a third generation ANNs that have been researched in recent years [12]. SNNs are different from other forms of ANNs as each individual spiking neuron propagates information by the timing of the neuron, rather than using the rate of the spikes. It was also found that SNNs are effective in brain modeling [13], [14]. This is useful as methods can be sought to detect epilepsy through the process of modelling the brain of an epileptic patient. Hardware implementations of SNNs were performed using NVIDIA CUDA [12]and the SpiNNaker [15]. The latter has the capability to simulate and implement the SNN which is used in brain modelling mentioned above. There are a few advantages and disadvantages when using hardware implementation on NVIDIA GPUs. The constant read-only memory is proved to have higher access speed than global memory. However, there is a requirement for more graphics processing unit (GPU) memory. Accessing the parameters of an individual neuron is also slow [12].

III. Bit-Serial Architecture with Relation to Neural Network Processors

Bit-serial architectures which process data bit by bit during each clock cycle are largely historic. Most modern processors use bit-parallel data processing for performance. However, when high performance is not a priority but instead the emphasis is on very low-power and low-cost bit-serial computing has its advantages. In modern applications bit-serial processing is still used sometimes in digital filters where input samples are processed in a bit-serial manner, although. Usually, however, the overall samples included in the filter's window frame are processed in parallel.

IV. A bit-serial hardware neural network

A novel approach is proposed to implement a low-cost hardware neural network which is primarily intended for use in portable equipment to predict epilepsy seizures. We consider the classical model of a perceptron that receives a vector input pattern x_i where $i = 1, \ldots, I$ and I the size of the vector. These inputs are weighted by the weight vector of a given perceptron $(w_1, w_2, \ldots, I)$ which is obtained in the off-line learning process. The neuron is a summation unit that performs the sum of products to calculate its output u. The output u is then processed by the activation function used in the output neuron. In our case the activation function is a simple threshold operation converting u into a logic signal y which has the value of '0' or '1'.

$$u = \sum_{i=1}^{I} w_i x_i \tag{1a}$$

$$y = \Phi(u) \tag{1b}$$

The conventional bit-serial architecture can model this behaviour with ease and complex feed forward neural networks (FNNs) based on such neurons can be created using simple, regular hardware structures controlled by simple state machines. The learning process of such designs can be accomplished off-line by using simulation software.

Each DPU in a given FNN layer performs the same operations and receives control signals issued by the layer control FSM to carry out the bit-serial additions and multiplications. This way, an FNN layer becomes an SIMD machine controlled by a single FSM. The development of an FPGA implementation of the neural processor is fast straightforward. We have used an FPGA implementation to carry out a number of test and study the potential of the proposed processor to classify epileptic EEG patterns. Table I shows that an 8-bit DPU requires only 24 Logic Elements (LEs) on an inexpensive Altera Cyclone V FPGA, out of over 300,000 LEs available on a Cyclone V chip. The control path of for a network with three layers requires 103 LEs (Central Control FSM: 3 LEs, 2 layer FSMs: 18 LEs each and 2 counters: 32 LEs each). This compares favourably with the size of the datapaths of typical bit-serial processors mentioned in the Table. Bearing in mind that the control logic of the proposed approach requires only simple state machines, rather than fully-fledged program control paths used in general-purpose processors, expected overall benefits of an ASIC implementation will include faster operation and lower power consumption.

Hardware	Development Chip	LE Count
Bit Array [16]	ASIC	56 Altera Equivalent LEs
Cellular Processor [17] (Data Path)	Virtex 5	26 Altera equivalent LEs
Proposed Neural Processor	Cyclone V	24 LEs

TABLE I: Cost comparison between three different processors from previous work [18].

V. Proposed Feature Extraction Hardware - Slope Calculator

In order to complete the wearable seizure detection system, it is imperative to include a novel and simple feature extraction hardware to provide the inputs to the BSNN. The proposed hardware will use picoMips as the basis of the design.

The data path consists of two synchronous RAM and a simple subtractor in the form of a ALU module. The data path is controlled by a simple FSM module. The hardware cost requires only 13 ALMs when synthesised on a Altera Cyclone V chip. This hardware will serve as a mean of extracting slope of the EEG waveform from two adjacent points on the sample.

VI. EEG waveform classification

The input data used in the evaluation of the proposed FNN was obtained from an on-line open source [5] which provides sets of EEG waveforms for both seizure free instances and EEG waveforms during seizures taken from the brain (epileptogenic zone) of the same patient. Figure 2 shows an samples of an epileptic and a normal EEG. Our results we obtained from a number of implementations of the proposed FNN and were evaluated using standard metrics [19] in seizure detection, namely: the sensitivity (TPR), specificity (TNR), positive predictive value (PPV) and negative predictive value (NPV). The hardware implementations were trained offline in MATLAB and then tested with two sets of 100 EEG waveforms. As part of the validation process, the same input data used for training was used to test the n-1-1 network (i.e. n neurons in the input layer, one neuron in the hidden layer and one output neuron).

It was found that the n-1-1 network configuration exhibits very bad recognition rates. From the results it can be concluded that a multi-input single neuron in the hidden layer is not sufficient to detect epilepsy accurately. Therefore, other configurations have been tested, for example a 40-n-1 network with n hidden neurons. The DPUs used in these tests had a 12-bit precision to provide high accuracy. In summary, the network configuration 40-30-1 provides promising results in terms of detecting epileptic waveforms. Further tests have been conducted using a larger number of inputs and more hidden layers to further validate and optimise the network.

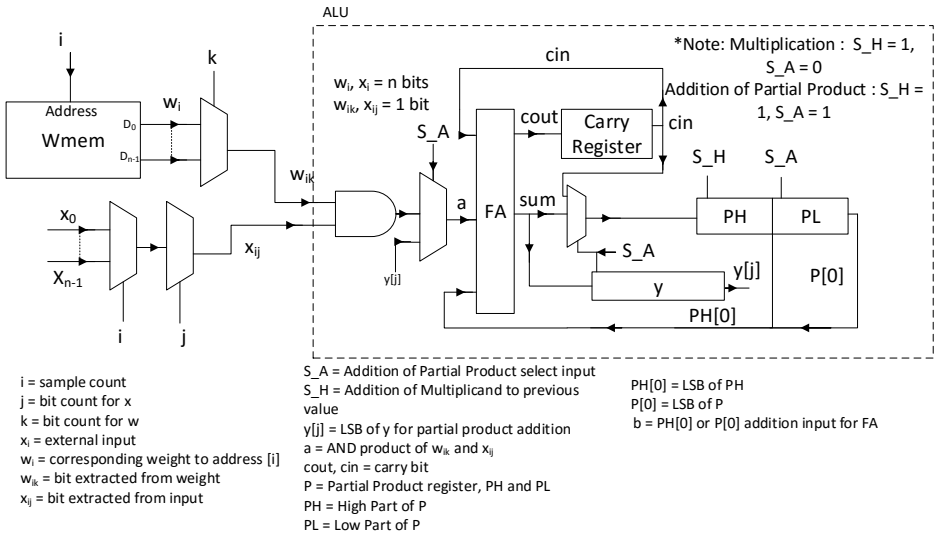
A. Hardware Network Validation and Testing

As the main purpose of this work is to distinguish seizure-free waveforms in epileptic patients from seizure waveforms, healthy patient brain waveforms are not included in the design testing. Results of the tests carried out at the validation stage have been compared with those of various software methods used in epilepsy detection [3].

Using the training datasets, the 11-7-1 hardware neural network with a 12 bit architecture has a specificity and sensitivity of 60%. It could recognise 30 out of 50 waveform used to train the network in a MATLAB model. The feature vector values consist of the same metrics as those provided in related work [3]. These values contain mean (X_{Mean}), median (X_{Median}), mode (X_{Mode}), standard deviation (X_{StdDev}), first quartile (X_{Q1}), third quartile (X_{Q3}), inter-quartile range (X_{IQR}), skewness (X_{skew}), kurtosis ($X_{kurtosis}$), minimum (X_{Min}), and maximum (X_{Max}). Ten other network configuration have been designed and tested. These configurations analysed using MATLAB in order to determine the mean square error (mse) in each case. From Table II, it can be seen that a single hidden layer with 100 neurons has a similar performance to that of a double layer network with 10 neurons in each layer.

B. EEG Waveform Slope Used as Feature Vector

In the previous subsection, a feature vector consisting of various statistic metrics is used. The maximum accuracy was 80% when tested using additional data. However, disparities



bit	LE Cost	
	Cyclone V	Stratix IV
6	20	20
8	24	28
12	28	33

Fig. 1: DPU Design with logic element counts included in table.

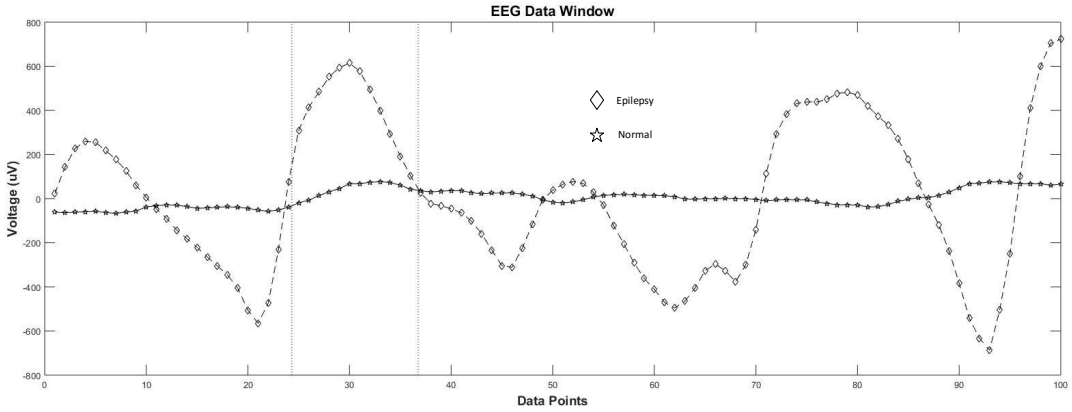


Fig. 2: Sample EEG input data.

Network Configuration	Correct recognition against training data	Correct recognition against additional tests
11-25-1	52%	60%
11-40-1	56%	50%
11-65-1	60%	30%
11-100-1	66%	55%
11-10-10-1	62%	60%
11-20-20-1	56%	80%
11-30-30-1	58%	60&
11-40-40-1	64%	45%
11-10-10-10-1	54%	50%
11-5-5-5-1	56%	30%

TABLE II: Correct recognition rates of different hardware ANN configurations.

when testing the same network configuration against the training data should be noted. In this respect, the 11-20-20-1 network shows some promising results. In this section, some experiments have been conducted to obtain better accuracy by using the slope of the EEG waveform at different points as a feature vector. The tested network configurations are 11-10-10-1, 11-20-20-1, 11-30-30-1 and 11-40-40-1.

Network configuration	TPR	TNR	PPV
11-10-10-1	57%	100%	80%
11-20-20-1	52%	44%	42%
11-30-30-1	66%	64%	58%
11-40-40-1	63%	100%	100%

TABLE III: Statistics for network configuration evaluation against training data.

VII. Conclusion

In conclusion, experiments with bit-serial neurons confirm that an extremely small logic system can successfully implement effective epileptic seizure detection. The

Network Configuration	TPR	TNR	PPV
11-10-10-1	75%	33%	43%
11-20-20-1	50%	50%	40%
11-30-30-1	25%	44%	10%
11-40-40-1	53%	33%	80%

TABLE IV: Statistics for network configuration evaluation against additional data.

key benefit of a dedicated neural processor compared to known, equivalent general-purpose processors, is that very small control logic and a low bit-precision are sufficient to obtain correct operation. Multiple tests have been conducted with various network configuration to test the feasibility of detecting epilepsy when using the proposed approach. Future work involves further investigation into suitable sizes and accuracies of bit-serial FNNs which will be followed by a development of a low-power ASIC.

REFERENCES

[1] "Who epilepsy fact sheet," http://www.who.int/mediacentre/factsheets/fs999/en/, accessed: 2017-09-01.

[2] B. Litt and J. Echauz, "Prediction of epileptic seizures," *The Lancet Neurology*, vol. 1, no. 1, pp. 22 – 30, 2002.

[3] Wang L., e.a., "Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis," *Entropy*, vol. 19, no. 6, 2017.

[4] G. Giannakakis, V. Sakkalis, M. Pediaditis, and M. Tsiknakis, *Methods for Seizure Detection and Prediction: An Overview*. Springer, NY, 2015, pp. 131–157.

[5] Andrzejak, Ralph G., e.a., "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E*, vol. 64, p. 061907, Nov 2001.

[6] A. Sharmila and P. Geethanjali, "Dwt based detection of epileptic seizure from eeg signals using naive bayes and k-nn classifiers," *IEEE Access*, vol. 4, pp. 7716–7727, 2016.

[7] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 660–674, May 1991.

[8] L. Arbach, J. M. Reinhardt, D. L. Bennett, and G. Fallouh, "Mammographic masses classification: comparison between backpropagation neural network (bnn), k nearest neighbors (knn), and human readers," in *CCECE 2003 - Canadian Conference on Electrical and Computer Engineering. Toward a Caring and Humane Technology (Cat. No.03CH37436)*, vol. 3, May 2003, pp. 1441–1444 vol.3.

[9] D. Sukumaran, Y. Enyi, S. Shuo, A. Basu, D. Zhao, and J. Dauwels, "A low-power, reconfigurable smart sensor system for eeg acquisition and classification," in *Circuits and Systems (APCCAS), 2012 IEEE Asia Pacific Conference on*, Dec 2012, pp. 9–12.

[10] K. Xia, G. Xu, and N. Xu, "Lung cancer diagnosis system based on support vector machines and image processing technique," in *Intelligent Information Hiding and Multimedia Signal Processing, 2006. IIH-MSP '06. International Conference on*, Dec 2006, pp. 143–146.

[11] Kiranmayi, G.R. and Udayashankara, V., "Neural network classifier for the detection of epilepsy," in *Circuits, Controls and Communications (CCUBE), 2013 International conference on*, Dec 2013, pp. 1–4.

[12] A. Arista-Jalife and R. Vazquez, "Implementation of configurable and multipurpose spiking neural networks on gpus," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, June 2012, pp. 1–8.

[13] Izhikevich, E.M., "Simple model of spiking neurons," *Neural Networks, IEEE Transactions on*, vol. 14, no. 6, pp. 1569–1572, Nov 2003.

[14] Cheng-Wen Ko, Hsiao-Wen Chung , "Automatic spike detection via an artificial neural network using raw EEG data: effects of data preparation and implications in the limitations of online recognition," *Clinical Neurophysiology - 1*, vol. 111, no. 3, pp. 477–481, March 2000.

[15] X. Jin, M. Lujan, L. Plana, S. Davies, S. Temple, and S. Furber, "Modeling spiking neural networks on spinnaker," *Computing in Science Engineering*, vol. 12, no. 5, pp. 91–97, Sept 2010.

[16] B. Svensson and T. Nordstrom, "Execution of neural network algorithms on an array of bit-serial processors," in *Proc.10th Int. Conf. on Pattern Recognition*, vol. ii, June 1990, pp. 501–505 vol.2.

[17] D. Walsh and P. Dudek, "A compact fpga implementation of a bit-serial simd cellular processor array," in *2012 13th International Workshop on Cellular Nanoscale Networks and their Applications*, Aug 2012, pp. 1–6.

[18] S. M. Kueh and T. Kazmierski, "A dedicated bit-serial hardware neuron for massively-parallel neural networks in fast epilepsy diagnosis," in *2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT)*, Nov 2017, pp. 105–108.

[19] H. J. Carey, M. Manic, and P. Arsenovic, "Epileptic spike detection with eeg using artificial neural networks," in *2016 9th International Conference on Human System Interactions (HSI)*, July 2016, pp. 89–95.

Estimation of Power Amplifier Package Model from Frequency Sweep Measurements

Dušan N. Grujić, Milan Savić and Pavle Jovanović

Abstract - In this paper we present a method for estimating power amplifier package model from spectrum analyser frequency sweep measurements. Assumed power amplifier package model is physical and it includes parasitic capacitances and bondwire inductance. Packaged model parameters can be estimated from frequency sweep measurements. Estimated package model was used to design the matching network for 2.14 GHz band, and the correctness of the proposed method and assumed model was confirmed by comparing expected results obtained by simulation with measurements.

Keywords – Power amplifier, package model, optimization.

I. Introduction

Software defined radio is disrupting RF system design in the same way as FPGA did for digital system design. The need to cover many standards and reconfigurability is driving the expansion of software defined radio even in the markets which have been dominated by few big companies, such as telecommunication infrastructure.

Modern software defined radios operate in very wide frequency range spanning from few MHz well into GHz range. For maximum flexibility external matching circuit, usually consisting of inductors, capacitors and baluns is used to enable operation in such extreme frequency span. Matching network is designed to optimize the performance for a given application, e.g. wideband frequency response, maximum output power etc.

Designing a matching network for a given application is not a trivial task since it consists of fixed on-chip part, i.e. parasitic capacitance, bond-wire inductance etc., and external components. On-chip parasitics must be taken into account, especially in the GHz range, but it is difficult to simulate or measure them.

One of the ways to characterize on-chip parasitics is to run an electromagnetic simulation. However, the accuracy of results depends on the accuracy of electromagnetic models. It is difficult to construct an accurate model since the knowledge of packaging geometry and material properties are not known or incomplete. For example, exact bonding profile (geometry) is not always known so the accuracy of simulated bond-wire inductance is questionable. Material properties, such as molding plastic permittivity and loss tangent, are either completely unknown or specified at low frequency. All of these issues complicate the design of electromagnetic models and lower the confidence in simulation results. Analytic models, such as [1-5] can aso be used, but unknown bondwire geometry and material properties limit the accuracy of results.

Package parasitics can in principle be measured by a vector network analyser. Success of determining package parasitics from packaged chip measurements depends on the impedance of on-chip circuit. In the case of on-chip power amplifier the output impedance is almost purely capacitive, and the chip is a highly reflective load which is difficult to accurately measure. Two-port measurements, where one port is on-chip and the other on PCB, would solve the problem of bondwire measurement, but would require on-wafer probing and fabrication of test chip, which is very expensive.

In this paper we present a method for estimating the package model of software defined radio power amplifier by using only a spectrum analyser frequency sweep measurements. Method and its assumptions are described in Sec. II. Simulation and experimental validation is presented in Sec. III, while concluding remarks are given in Sec. IV.

D. Grujić is with School of Electrical Engineering, University of Belgrade and Lime Microsystems. P. Jovanović and M. Savić are with Lime Microsystems. E-mail: {d.grujic, p.jovanovic, m.savic}@limemicro.com

II. Packaged Power Amplifier Model

Software defined radio should operate in very wide frequency range and is usually designed as direct conversion transceiver to avoid the use of tunable filters. Block diagram of software defined radio direct conversion transmit chain – Fig. 1 – consists of I and Q baseband paths, followed by quadrature mixer and a power amplifier. This architecture does not require an image rejection filter, but must have very good amplitude and phase balance to provide sufficient image rejection and preserve RF signal fidelity. Depending on the chosen modulation, standard CMOS processes may provide sufficient amplitude and phase matching to satisfy required image rejection, while in other cases some kind of calibration must be performed to satisfy system requirements.

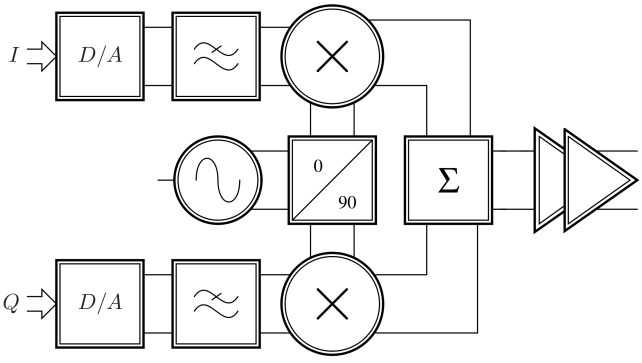


Fig. 1. Software defined radio RF transmit chain block diagram

For a single sideband excitation, generated by a proper choice of I and Q digital baseband signals, the whole

transmit chain can be simplified as power amplifier driven by a voltage source, as shown in Fig. 2. This simplification has an underlying assumption that the quadrature mixer-power amplifier interface is frequency independent, which is true for a well designed chip. Furthermore, CMOS power amplifier can be represented as a voltage constrained current source [6-8] with a shunt capacitance. Current source voltage constraints originate from physical limitations of MOS transistors – entering a triode region where they do not operate as current sources and breakdown voltage. Shunt capacitance is the sum of cascade transistor drain junction, drain-gate and drain-source parasitic capacitances.

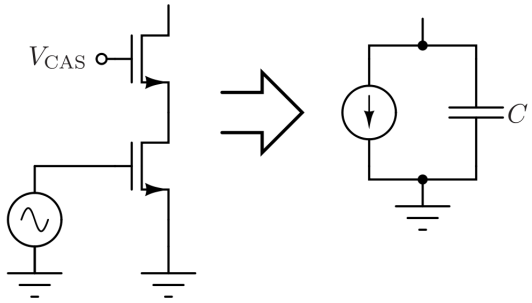


Fig. 2. Power amplifier equivalent model

Integrated power amplifiers are usually designed as differential circuits to boost the output power and to suppress the second order harmonic. Packaged differential amplifier equivalent model, shown in Fig. 3, consists of equivalent voltage constrained current source, parasitic transistor shunt capacitance, series bondwire inductance and resistance, and package parasitic capacitance. In order to design a matching circuit, the values of these parasitic elements should be estimated.

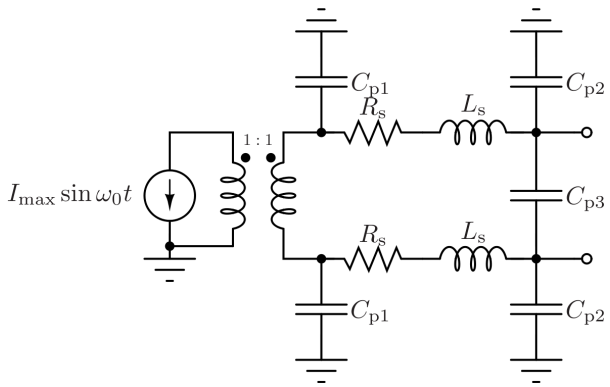


Fig. 3. Packaged differential power amplifier model

Assumed packaged differential amplifier model has six unknowns – driving current amplitude $I_{\max}$, parasitic capacitances C_{p1}, C_{p2} and C_{p3}, series resistance R_{s} and inductance L_{s}. Determination of assumed power amplifier model parameters is discussed in the following section.

III. Model Parameter Extraction and Experimental Validation

Assumed equivalent power amplifier model has six independent parameters, and it requires a system of at least six equations to solve. Parameters of assumed equivalent power amplifier model can be obtained from frequency sweep measurements, where the number of frequencies is at least six – preferably more, and by optimizing the parameters to fit the measurement data. However, the packaged power amplifier is mounted on printed circuit board, has bias inductors, DC block capacitors and a balun to facilitate spectrum analyser measurements, as shown in Fig. 4. Since the printed circuit board geometry and element values are known, it is possible to determine the values of equivalent packaged power amplifier model from measured output power at different frequencies by performing a frequency sweep and measuring the output power.

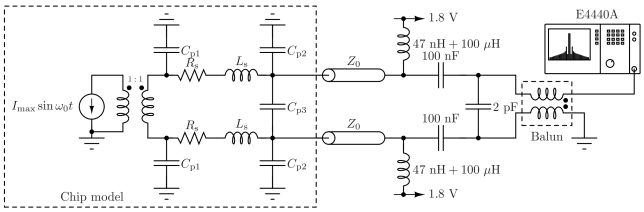


Fig. 4. Setup for frequency sweep measurement

The results of frequency sweep measurements and optimized packaged power amplifier model simulation are shown in Fig. 5. Assumed packaged power amplifier model is in good agreement with measurement results, so it is expected that model parameters correspond to physical reality.

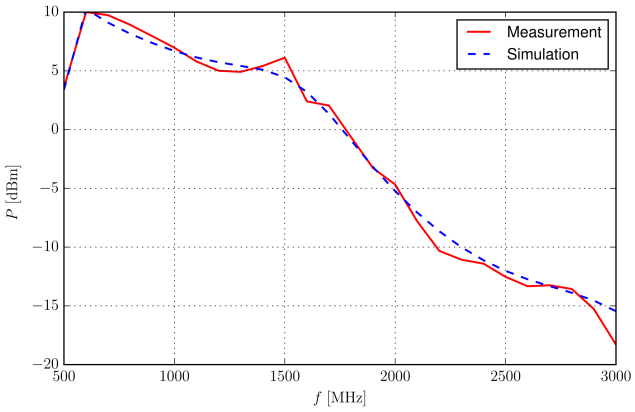


Fig. 5. Measured and simulated frequency sweep. Simulation was performed with optimized package model.

Optimized model parameters can be used to predict performance in the desired frequency band. In our case, the desired frequency band was centered around 2.14 GHz, and

the goal was to design a matching circuit which would optimize the output power. Matching circuit was designed by using the optimized packaged power amplifer model and changing the printed circuit board components. Model obtained by optimization is shown in Fig. 6.

If the assumed packaged power amplifier model is correct, then it is expected that the optimized matching network simulation results should be in good agreement with measurement results. Experimental validation was performed by changing the printed circuit board components to the ones obtained by optimization and frequency sweep power measurements. Comparison of simulation and measurement results are shown in Fig. 7.

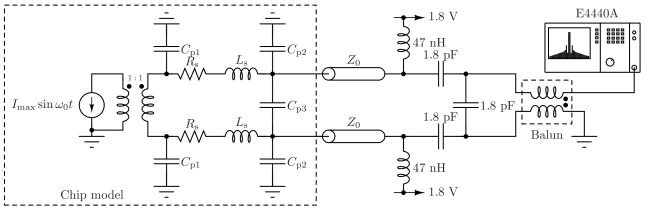


Fig. 6. Setup for measuring the optimized matching circuit for 2.14 GHz band

From Fig. 7 it can be seen that simulation and measurement results are in reasonable agreement, and that the optimized component values can be used as a good starting point for fine tuning of matching network. It also shows that the assumed package model and estimated model parameters are correct.

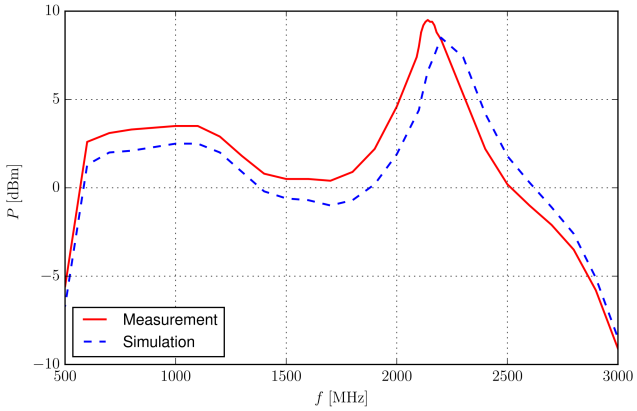


Fig. 7. Comparison of expected and measured output power with matching circuit for 2.14 GHz band

IV. Conclusion

In this paper we have shown a method of estimating power amplifier package model from frequency sweep measurements. Measurement of frequency dependent software defined radio output power was used to fit the parameters of physical power amplifier package model. The model includes parasitic capacitance and bondwire inductance. Estimated model was used to design a matching network for 2.14 GHz band, and was verified by measurement. The proposed method may be used to estimate the package model without the need for network analyser measurements or test chip if assumptions are satisfied.

References

[1] A. L. Nazarian, L. F. Tiemeijer, D. L. John, J. A. van Steenwijk, M. de Langen, R. M. T. Pijper, "*A Physics-Based Causal Bond-Wire Model for RF Applications*", IEEE Transactions on Microwave Technology and Techniques, Vol. 60, No. 12, Dec. 2012, pp. 3683-3692

[2] S. Hauptmann, M. Hellfeld, C. Knochenhauer and F. Ellinger, "*Modeling bond wires for millimeter wave RFIC design*," 2009 IEEE International Conference on Microwaves, Communications, Antennas and Electronics Systems, Tel Aviv, 2009, pp. 1-4.

[3] M. Sigalov, D. Regev, E. Kabatsky, R. Shavit, "*Quad Flat Non-lead Package Characterization and Circuit Modeling*", PIERS Proceedings, Moscow, Russia, August 18–21, 2009, pp. 642-646

[4] J. Lim, D. Kwon, J.-S. Rieh, S.-W. Kim, S. H., "*RF characterization and modeling of various wire bond transitions*," in IEEE Transactions on Advanced Packaging, vol. 28, no. 4, pp. 772-778, Nov. 2005.

[5] T. P. Wang and Y. F. Lu, "*Fast and Accurate Frequency-Dependent Behavioral Model of Bonding Wires*," in IEEE Transactions on Industrial Informatics, vol. 13, no. 5, pp. 2389-2396, Oct. 2017.

[6] H. Darabi, "*Radio Frequency Integrated Circuits and Systems*", Cambridge University Press, 2015

[7] S. Cripps, "*RF Power Amplifiers for Wireless Communications*", Artech House, 2015

[8] D. N. Grujić, L. Saranovac, "*Broadband Power Amplifier Limitations due to Package Parasitics*", Serbian Journal of Electrical Engineering, Oct. 2015, Vol. 12, Issue 3, pp. 275-291

LMS8001 Companion Board – A Highly Configurable 4-Channel Frequency Shifter Platform Utilising the LMS8001A RFIC

Milan Savić, Miloš Božić, Mihajlo Božović, Pavle Jovanović, Dušan N. Grujić

Abstract – This paper presents the LMS8001-Companion board, which is a highly configurable 4-channel frequency shifter platform utilising the LMS8001A integrated circuit.

Keywords – Frequency shifter, RFIC, LMS8001, MIMO, 5G.

I. Introduction

The LMS8001-Companion board provides a highly integrated, highly configurable, four-channel frequency shifter platform, utilising the LMS8001A integrated circuit. One of the typical applications is extending Lime Micro transceiver family RF frequency range up to 10 GHz.

Massive MIMO (Multiple Input Multiple Output) technology, which is supported by the latest 3GPP specifications for LTE, will be supported in 5G from the first deployments. Generally, massive MIMO provides both coverage and capacity gains at sub-6GHz frequency and coverage gains at mmWave frequencies for 5G [1].

Highly integrated 4-channel frequency shifter RF IC [2], featured on the presented board, enables compact massive MIMO solutions at any sub-6GHz band, and is paving the way for the above-6GHz applications.

II. LMS8001-Companion Board

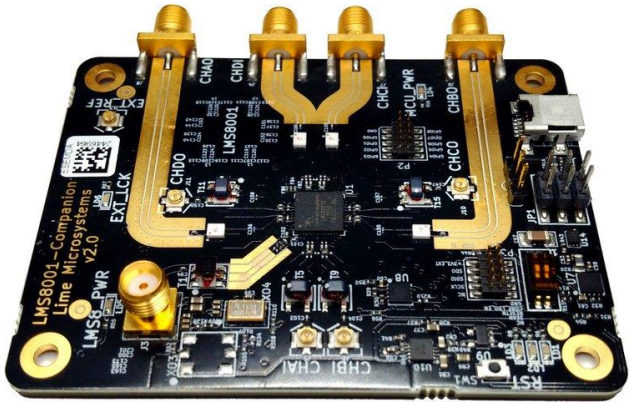


Fig. 1. LMS8001-Companion Board

M. Savić, M. Božić, M. Božović, and P. Jovanović are with Lime Microsystems. D. Grujić is with School of Electrical Engineering, University of Belgrade and Lime Microsystems.

E-mail: {m.savic, m.bozic, m.bozovic, p.jovanovic, d.grujic} @limemicro.com

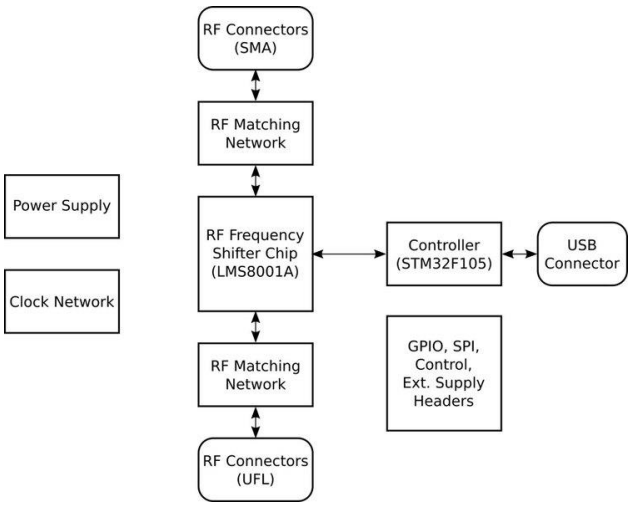


Fig. 2. LMS8001-Companion Block Diagram

LMS8001-Companion board is provided as an open-source hardware, and is available through the Myriad-RF community [3].

Board information, setup and installation instructions, control software basics, as well as basic measurement results are provided in the documentation provided.

A. LMS8001A Integrated Circuit

Core functionality of the LMS8001-Companion board is provided by the featured LMS8001A integrated circuit [4].

LMS8001A Features:

- Single chip up/down RF frequency shifter with continuous coverage up to 10 GHz RF output range
- Four independent highly reconfigurable RF paths all driven by the same LO
- Fully differential signals
- Few external components
- Low voltage operation, 1.2 and 1.8V. Integrated LDOs to run on a single 1.8 V supply
- 56-pin QFN package
- Serial Port Interface
- Power down control available via ENABLE pins and/or equivalent SPI registers
- Synchronous loading of pre-set operation profiles by

GPIO pins. More options are also available using corresponding SPI registers

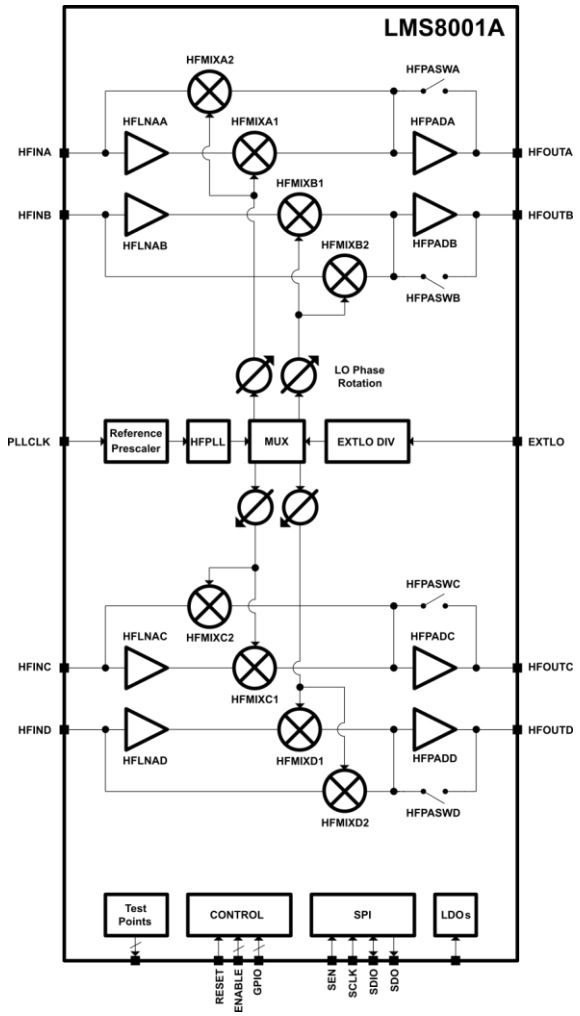


Fig. 3. Structure of LMS8001A up/down RF frequency shifter

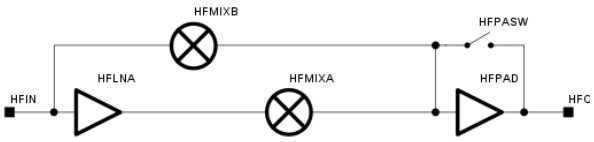


Fig. 4. Channel block diagram

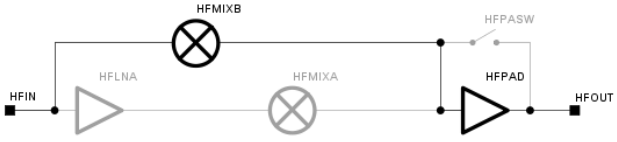


Fig. 5. Example of up-conversion channel configuration

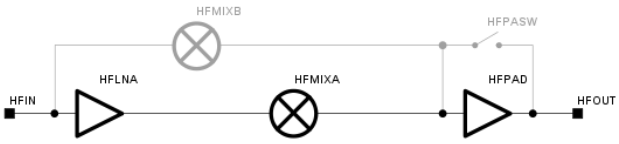


Fig. 6. Example of down-conversion channel configuration

B. Board Features

Connections:

- 4 x SMA connectors - RF
- 4 x UFL connectors – RF
- 1 x SMA connector – External LO
- 1 x UFL connector – External clock reference

RF Matching:

- Channel A Input (UFL) – 1–3 GHz broadband
- Channel B Input (UFL) – 1–3 GHz broadband
- Channel C Input (SMA) – 10 GHz band
- Channel D Input (SMA) – 5 GHz band
- Channel A Output (SMA) – 10 GHz band
- Channel B Output (SMA) – 5 GHz band
- Channel C Output (UFL) – 1–3 GHz broadband
- Channel D Output (UFL) – 1–3 GHz broadband

USB Interface:

- USB - mini B
  (STM32 controller STM32F105RBT6)

Clock System:

- 40MHz on board VCTCXO
- Possibility to lock VCTCXO to external clock

Board Size:

- 60mm x 100mm (2.36'' x 3.94'')

C. Control Software

Utilities for use with hardware based upon LMS8001 RFICs from Lime Microsystems are provided as open-source software [5].

LMS8Suite software provides full control of the LMS8001A integrated circuit.

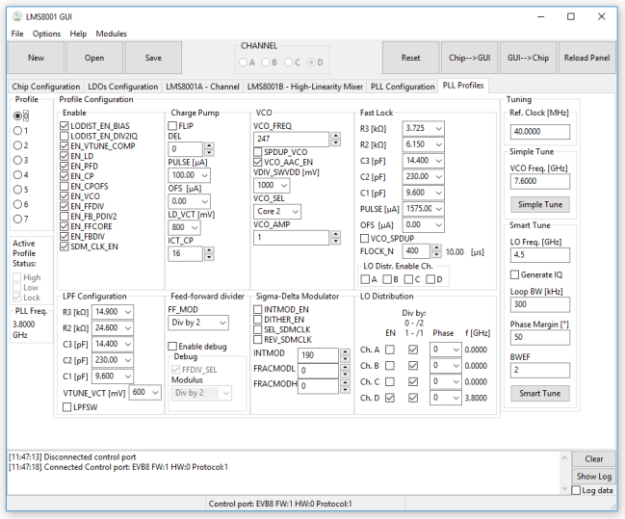


Fig. 7. LMS8Suite Control Software – PLL Profiles Tab

D. LMS8001 PLL-Sim

Lime-Micro LMS8001 PLL-Sim software is intended to provide the user deeper insight into the operation of PLL frequency synthesizer inside the LMS8001 frequency conversion IC. There are many ways to confugure LMS8001 PLL Core. In order to somewhat facilitate getting optimal configuration of PLL core for targeted input parameters such as VCO (or LO) frequency, loop crossover frequency, phase margin etc, many functions are implemented to automate the whole optimization process.

This software is also available as the open-source [6].

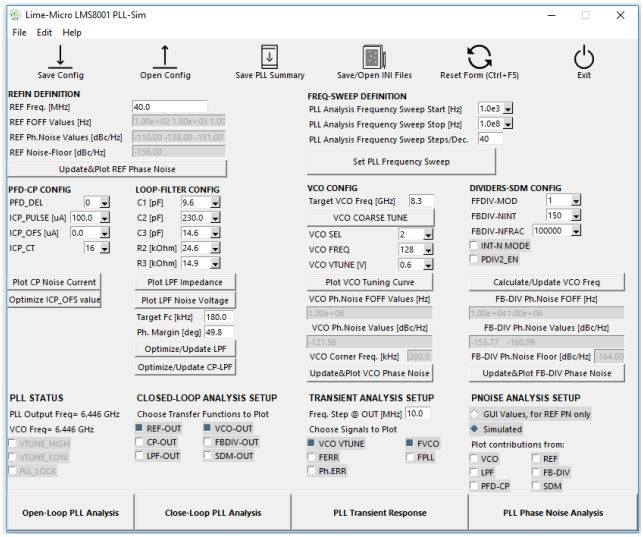


Fig. 8. LMS8001 PLL-Sim – Main Window

III. Measurement Result Example

In the following the measurement results of the channel D configured for downconversion (Fig. 6) are presented. The input frequency was swept between 5 and 6 GHz, and output frequency was set to 1.2 GHz, unless otherwise stated.

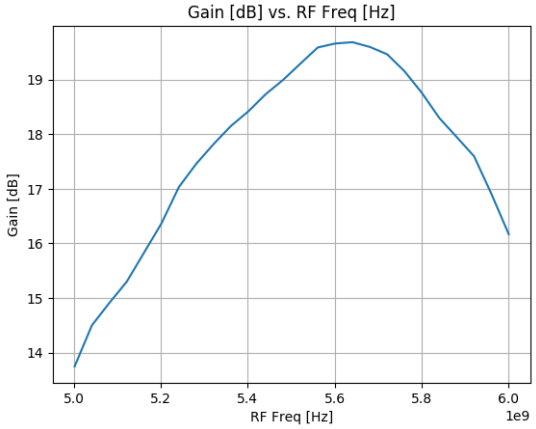


Fig. 9. Conversion Gain

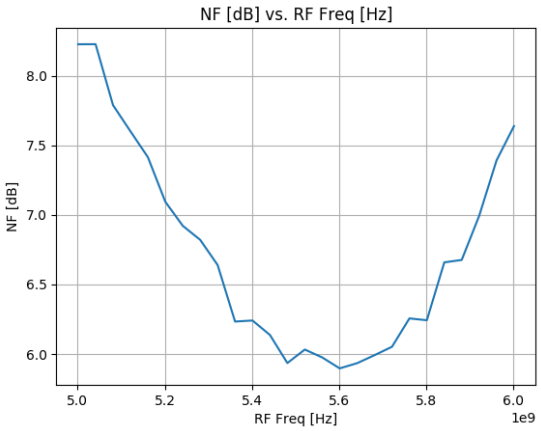


Fig. 10. Noise Figure

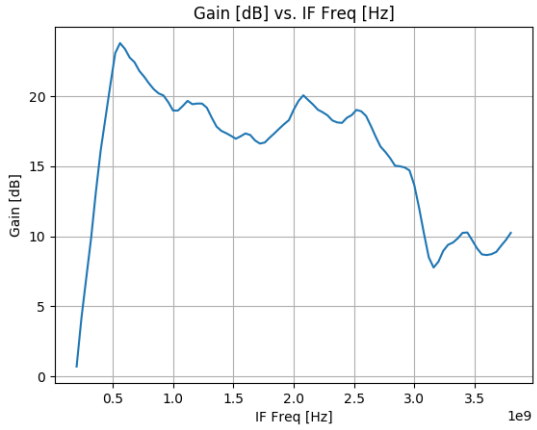


Fig. 11. Gain vs. IF Frequency, RF Frequency = 5.5 GHz

IV. Conclusion

This paper presenteed the LMS8001-Companion board, which is a highly configurable 4-channel frequency shifter platform utilising the LMS8001A integrated circuit from Lime Microsystems.

References

[1] Nokia, *"Beamforming for 4.9G/5G Networks – Exploiting Massive MIMO and Active Antenna Technologies"*, White Paper, Available at: https://onestore.nokia.com/asset/201377/Nokia_5G_Beamforming_mMIMO_White_Paper_EN.pdf

[2] Litovski, V., Zwolinski, M., "*VLSI Circuit Simulation and Optimization*", Chapman and Hall, London, 1997.

[3] Lime Microsystems, "*LMS8001-Companion*", available at: https://github.com/myriadrf/LMS8001-Companion

[4] Lime Microsystems, "*LMS8001 Documentation*", available at: https://github.com/myriadrf/LMS8001-docs

[5] Lime Microsystems, "*LMS8Suite*", available at: https://github.com/myriadrf/LMS8Suite

[6] Lime Microsystems, "*LMS8001 PLL-Sim*", available at: https://github.com/myriadrf/LMS8001-PLL-Sim

IGBT versus VDMOS switches in DC-to-AC inverters, some basic comparisons

Jelena Milojković, Simon Le Blond, Marko Dimitrijević and Vančo Litovski

Abstract - Renewable energy sources invariably require interfaces for power conversion, and new switches based on modern semiconductors may offer converters with improved efficiency. This study investigated properties of two contemporary voltage-controlled power transistors, and their suitability for use in DC-to-AC inverters. Namely Si IGBT and Si VDMOS were considered. These devices were implemented in a Low-Voltage Single-Phase H-Bridge based Inverter (LVSPHBI) and simulated with LTSPICE, with simulations for different levels of power delivered to a linear resistive load while keeping the DC voltage source constant. The efficiency and total harmonic distortion (*THD*) were used as measures of merit. The study includes frequency dependence of the efficiency and the *THD* for various technologies. It was found that each of the transistors has advantages from proper point of view.

Keywords – DC-to-AC inverters, IGBT, VDMOS

I. Introduction

Due to the high impact of fossil fuels on the environment, renewable energy sources have become the most important consideration in the development of electrical power systems [1]. The output of primary renewable sources, however, are strongly dependent on microclimate conditions (wind speed, insolation), which as a consequence, lead also to variable (DC or AC) output voltage.

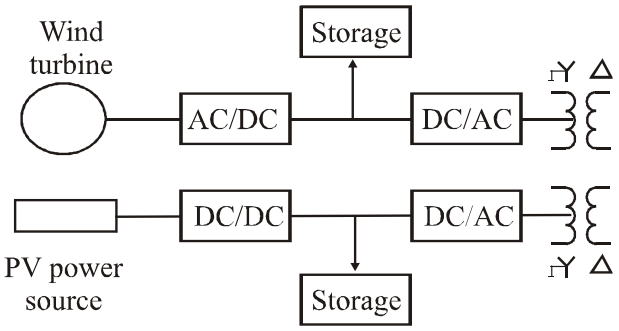


Fig. 1. Power conversion suits in two alternative-energy-source chains. Top wind energy and bottom photovoltaic energy

Jelena Milojković is with Innovation Centre of Advanced Technologies, Niš, Serbia, E-mail: jelena.milojkovic@icnt.rs

Simon Le Blond is with University of Bath, United Kingdom, E-mail: s.p.leblond@bath.ac.uk

Marko Dimitrijević is with Faculty of Electronic Engineering, University of Niš, Sebia, E-mail: marko.dimitrijevic@elfak.ni.ac.rs

Vančo Litovski is independent researcher, Serbia, E-mail: vanco@elfak.ni.ac.rs

Fig. 1 depicts two "processing" chains which start with a renewable source and end with a source of stabilized (amplitude and frequency) AC voltage ready to be connected to the grid. It consists of several stages, performing conversion and inversion of the original unstabilized AC or DC voltage.

As can be seen from Fig. 1, the DC-to-AC inverter represents an unavoidable subsystem in the connection of a renewable energy source to the grid, be it micro-, smart-, or ordinary [2], [3]. This is why much attention was given to the development and design of various types of DC-to-AC inverter [4].

Within the extensive variety of converter implementations, we consider here the most common, namely, the Low-Voltage Single-Phase H-Bridge based Inverter (LVSPHBI). In particular, the device is built from voltage controlled switches, which in modern times, have replaced thyristors [5], [6].

Our intention was to collect and publicize comparative information about the behavior of LVSPHBIs using two technologies: the silicon VDMOS and the silicon IGBT. As a measure for comparison, two main metrics were used: the efficiency and the Total Harmonic Distortion (THD).

In the next section, we begin with the properties of the devices from the implementation perspective for DC-to-AC inverters. Then, results of a systematic search with responses obtained by simulation will be reported, enabling comparisons between the devices. Finally, some conclusions from this study will be drawn.

II. Main properties of the components implemented

In the search for fast switching devices able to handle large voltages and currents existing silicon devices the N-VDMOS (Vertical Double Diffused MOS) and the IGBT (Insulated Gate Bipolar Transistor) are frequently used, due to the fact they are voltage controlled and alike thyristors, consume low energy at the input terminal which, in addition, needs no current source.

Application of these components in low-voltage solid-state circuit breakers and comparisons of the circuit properties was reported in [7], while similar studies are reported for the implementations in single-phase low-power low-voltage T-type DC-to-AC inverters [8] where no VDMOS was taken into account.

Table 1. Power transistor's maximum voltages at the output terminal, maximum temperatures, and capacitances

Type	V_{max} (V)	T_{max} (°C)	C_{ISS} (pF)	C_{RSS} (pF)	C_{OSS} (pF)
Si MOS (V_{DS}=25 V) R5021ANX	600	150	2300	70	1000
Si IGBT (V_{AK}=20 V) NGTB30N135IHRWG	1350	150	5290	100	124

Fig. 2 shows the structure of a modern high-voltage and high-power device, specifically the cross-section of a VDMOS. Its practical implementation is in fact a complex integrated circuit containing hundreds of cells (like the one depicted) connected in parallel. To allow for high voltages a depletion region is inserted between the drain and the channel while high currents are achieved by parallelization.

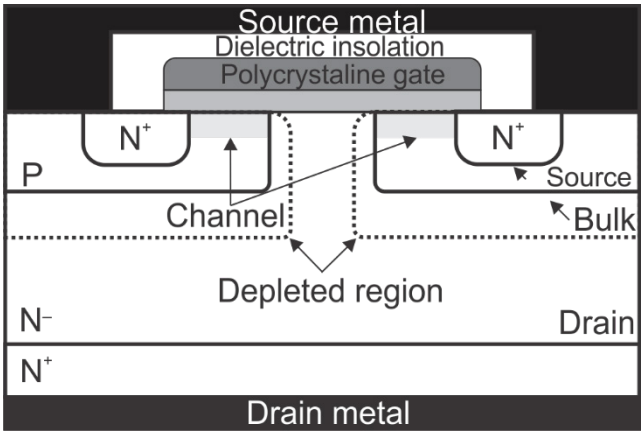


Figure 2. Cross section of an enhancement type N-channel VDMOS

In this paper the following components will be used for creation of the inverter: the R5021ANX VDMOS [9]; the NGTB30N135IHRWG IGBT [10].

Table 1 depicts some of the properties of the components listed above (taken from original datasheets). Among them the maximum voltage at the output terminal is of prime importance. The components selected both have maximum voltage above 400 V which makes them well suited to the subsequent experiment.

For high-power implementation point of view (having in mind the expenses for cooling) the electronic maximum operating temperature (T_{max}) is of prime importance.

The rest of Table 1 is related to device capacitances. The influence of capacitance the transistor's responses to a pulsed input signal will be discussed briefly later on.

To build a realistic a picture of the behavior of the above components we will first consider their static characteristics. To that end, the simple circuit of Fig. 3 was created. It allows for graphic analysis of the output characteristic and determination of the minimum voltage for different values of the circuit parameters e.g. V_{DD} and R_L.

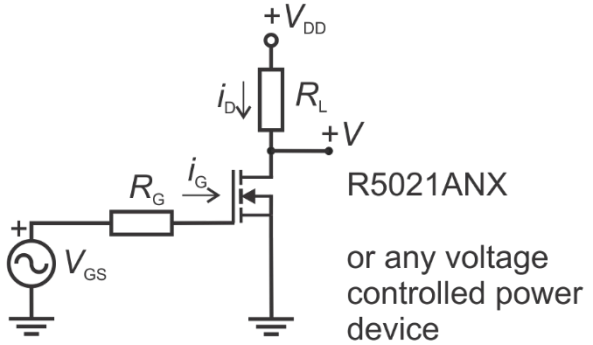


Figure 3. The test circuit

Some of the results obtained by LTSPICE [11] simulation are depicted in Fig. 4. In Fig. 4a we used V_{DD}=50V and R_L=5/8 Ohm. (It would be more realistic to use 8 times greater V_{DD} and R_L but the figure would not be so readable, while the conclusions would be the same.)

For a given gate voltage, (say, equal to the amplitude of the input pulse) the minimum value of the output voltage is restricted by the intersection between the load line and the proper output characteristic of the transistor, here denoted by the letters A and B. As can be seen from Fig. 4 the difference between the minimum voltages is in favor of the IGBT (denoted as Si_IGBT) as compared to the VDMOS (denoted as Si_MOS). Note, if 8 times larger V_{DD} and R_L were used, the upper-left end of the load line would not change, while the lower-right would move 8 times further to the right, resulting in even larger minimum voltage for the VDMOS.

The interception points A-B define the maximum swings of the output voltage and current and consequently the efficiency. Since alternating currents (AC) are considered, the larger these swings are the better.

The same effect may be observed from the pulse responses of the circuits containing the same set of transistors as listed above.

The simulation results for a 1 μs long input impulse are depicted in Fig. 5. Fig. 5a represents the output voltage of the circuit of Fig. 3 for V_{DD}=400 V and R_L=$310^2/(2\cdot5500)$ Ω which will later be used for the case where required inverter output power is P_{out}=5.5 kW. While the bottom of the output

pulse is lower making its amplitude larger the large input capacitance of the IGBT maps to a larger delay of the rising edge of the response as compared with the VDMOS.

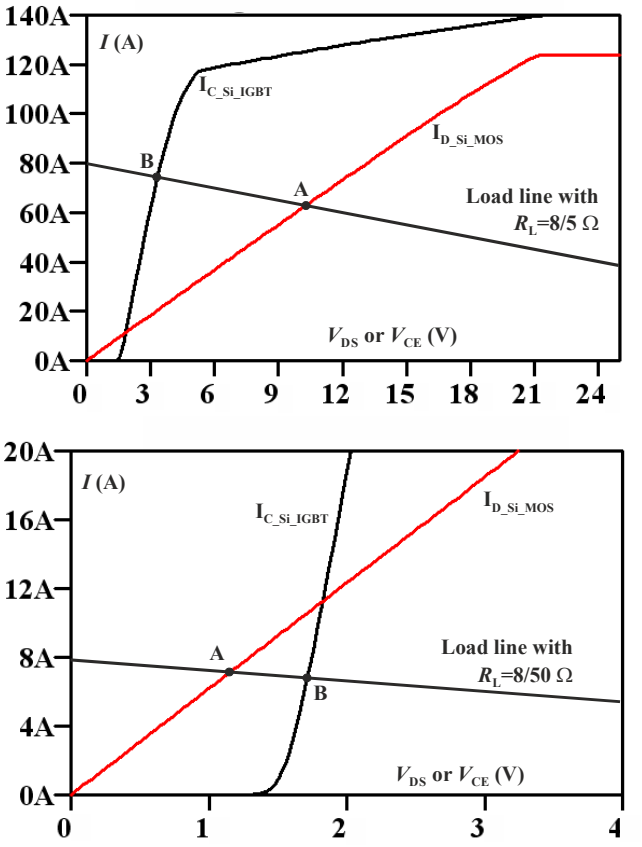


Fig. 4. Illustration of the value of the minimum output voltage (points A and B) for different technologies and different output powers.

In the case when larger power is required (P_{out}=10 kW) the value of the load resistor was changed to $R_L=310^2/(2\cdot10^4)$ Ω. The simulation results for this case are depicted in Fig. 5b. Here only the floor of the output pulse is shown in order to emphasize the differences introduced by different technologies. For large powers the amplitude of the output pulse is reduced and so is the efficiency. This reduction, however, strongly varies across the different technologies starting with approximately 2.5 V for the case of the IGBT and ending with 13 V for the Si_MOS.

III. The DC-to-AC Inverter

The structure of the single-phase H-bridge inverter is dual to the full-wave rectifier in the sense that the DC and AC ports are interchanged. Both, however, need a low-pass filter to eliminate the unwanted components of the output signal. Fig. 6 depicts the schematic used in this paper for the inverter implemented with IGBTs. Here in all cases E=200V.

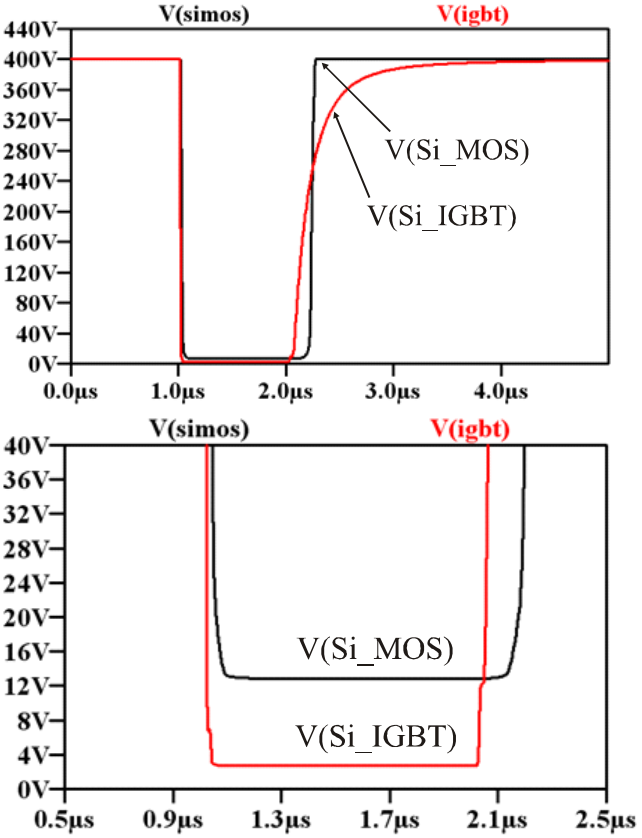


Fig. 5. Time domain responses of the circuit of Fig. 3 for different types of transistors. a) The complete pulse responses. $R_L=310^2/(2\cdot5500)$ Ω as (later used in the inverter simulation) and b) the floor of the output pulses highlighted. $R_L=310^2/(2\cdot10^4)$ Ω as (later used in the inverter simulation). The value of r_G corresponds to the value used for simulation later

The *pvm* signal is obtained from a subsystem (not shown) which generates a pulse-width-modulated signal [12], [13]. It is characterized by two quantities: the frequency, f_s, of the triangular waveform set against the 50 Hz sinusoid comparator to produce a pulse train, and the index representing the quotient of the amplitudes of the sinusoid and the triangular waveform, M_a. The latter is the parameter that determines the pulse train and thus is the main controlling parameter allowing adjustments of the amplitude of the output voltage. In reality, the connection of the *pwm* signal and its complement to the input terminals of the switching devices is quite complex, but for the purpose of this study, it may be satisfactorily approximated with a simple resistor, r_G.

IV. Simulation Results

The input (battery) power was calculated as follows

$$P_{bat} = 2 \cdot E \cdot I_{bat} \quad (1)$$

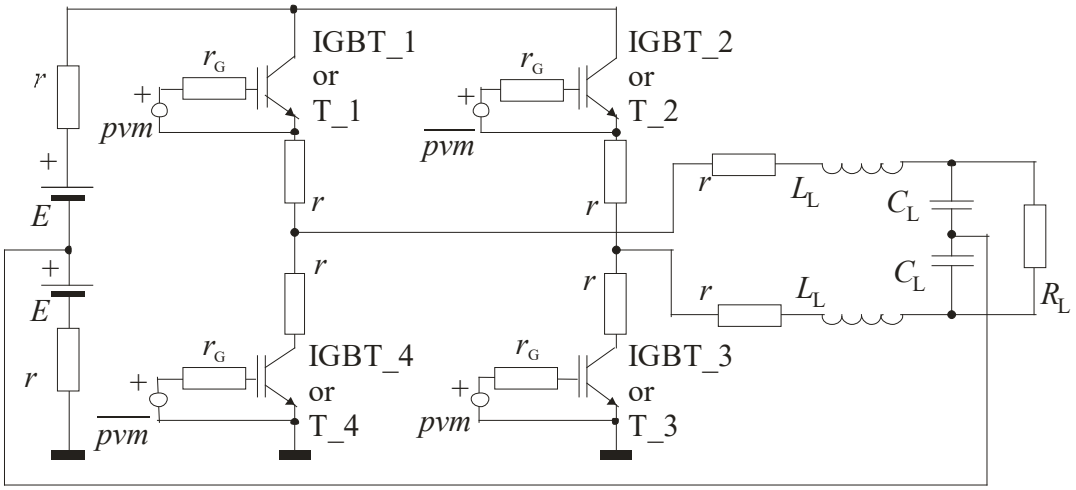


Fig. 6. The schematic used as a AC-to-DC inverter.

where I_{bat} is the DC component of the battery current produced (after transient simulation of the inverter) by the [.four] command in LTSPICE. On the other side, the AC power delivered to the load was calculated by

$$P_{load} = V_{load} \cdot I_{load} / 2 \quad (2)$$

where V_{load} and I_{load} represent the amplitudes of the load voltage and current, respectively, which were produced (after transient simulation of the inverter) by the [.four] command of LTSPICE.

The efficiency was obtained by the following expression

$$\eta = 100 \cdot P_{load} / P_{bat} \,[\%] \quad (3)$$

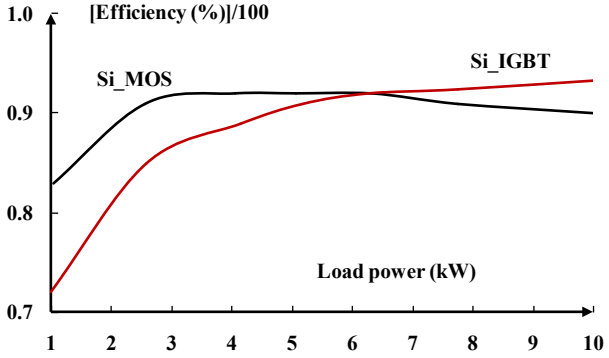
(To make results clearer, in the next section, the efficiency as a decimal is used, i.e. η/100).

To produce the dependences of the efficiency and the *THD* on the power delivered to the load, repetitive simulations of the circuit of Fig. 6 were performed. Two quantities were changed to obtain the drawings for a given value of the load power. Namely, the load resistance was calculated first to be $R_L=310^2/(2\cdot P_{load})$, and then the circuit was repeatedly simulated with various values of M_a until the output voltage became almost equal (but slightly larger) to 310 V.

The results for efficiency are depicted in Fig. 7. When considering Fig. 7 and Fig. 5b simultaneously, one must also recall that during the inverter's on state, two transistors are connected in series thus reducing the amplitude of the output voltage by a value twice as large as that depicted in Fig. 5b (or Fig. 4).

From the perspective of efficiency, based on Fig. 7, at low powers (as depicted in Fig. 4b) the VDMOS is advantageous as compared with the IGBT based inverter. Whilst its efficiency suffers at low powers, its low value of the minimum voltage at high currents, as discussed above, allows for high efficiency when high power is required.

As described in detail in [14] the harmonic distortions are one of the most important indices of the quality of the delivered electrical power. Here we performed Fourier analysis, using LTSPICE, up to the 250th harmonic of the power frequency i.e. up to 12.5 kHz. Since linear resistive load was considered only, the results given below for the load voltage are the same for the load current.



Fig, 7. Efficiency as a function of the power frequency power delivered to the load for different technologies

The results are depicted in Fig. 8. The distortions introduced by the inverter are very small. For example, for the case of IGBT based inverter at P_{load}=2.5 kW the THD is equal to 0.35% (almost the largest value). The VDMOS based inverter introduces even smaller distortions which decrease dramatically at very low powers.

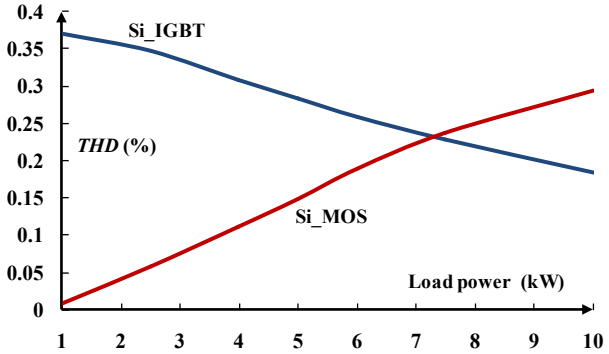


Fig. 8. *THD* as a function of the power frequency power delivered to the load for different technologies

Finally, the dependence of the inverter's properties as a function of the frequency f_s used in the *pwm* signal generator is of interest. At low frequencies, irritating audible noise is frequently produced by this type of equipment due to the nonlinearities that generate harmonics. For this reason, higher frequencies are preferred. The efficiency and the distortions at higher frequencies are therefore of interest.

Fig. 9 depicts the dependence of the efficiency of both types of inverter as a function of the frequency used in the PWM signal generator. The general conclusion is that the efficiency decreases with the rise of f_s. Given the trend suggested by the curve, it could be argued the Si_MOS based inverter is preferable at high frequencies. From the perspective of distortion, as can be seen from Fig. 10, at high frequencies both the IGBT based and VDMOS based inverters perform very well.

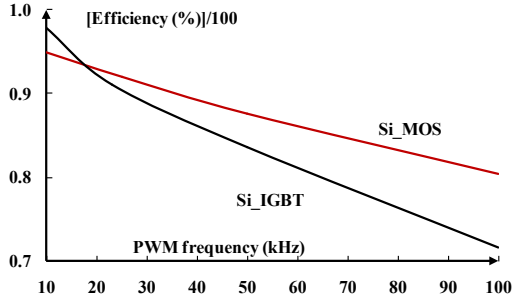


Fig. 9. Efficiency as a function of the frequency f_s used in the PWM subsystem for different technologies

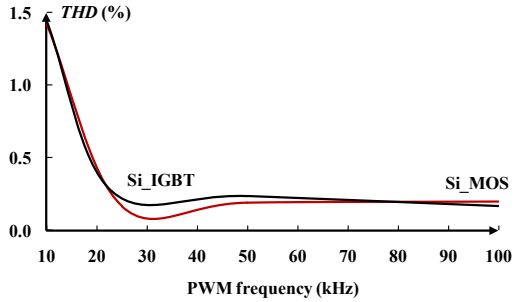


Fig. 10. *THD* as a function of the frequency f_s used in the PWM subsystem for different technologies

V. Conclusion

The DC-to-AC inverter is a ubiquitous subsystem of modern renewable alternative energy sources. Its properties map themselves directly into the characteristics of the waveforms delivered to the load (or the grid), thus the device deserves much attention in the search for improvement. This study investigated the advantages and disadvantages of using existing Si based voltage controlled switching devices in a conventional single-phase single-triangle PWM H-bridge inverter. The efficiency and the *THD* were used as measures of merit. LTSPICE simulation was used for all results reported here. Based on the results obtained, we may conclude that the VDMOS and the IGBT have application niches of their own in which they perform better.

Acknowledgement

This research was partly funded by The Ministry of Education and Science of Republic of Serbia under contract No. TR32004.

References

[1] Michaelides, E., "Alternative energy sources", Springer Verlag, Berlin, Germany, 2012.

[2] Keyhani, A., "Design of Smart Power Grid Renewable Energy Systems", John Wiley and Sons, Inc, Hoboken, N.J., 2011

[3] Li, F., Qiao, W., Sun, H., Wan, H., Wang, J., Xia, Y., Xu, Z., and Zhang, P., "Smart Transmission Grid: Vision and Framework", IEEE Trans. on Smart Grid, Vol. 1, No. 2, 2010, pp. 168-177.

[4] Wang, X, Guerrero, J. M., Blaabjerg, F, and Chen, Z, "A Review of Power Electronics Based Microgrids", Int. J. of Power Electronics, Vol. 12, No.1, 2012, pp.181-192.

[5] B.Burger and D. Kranzer, "Extreme high efficiency PV-power converters," in Proc. Power Electron. Appl. Conf., 2009, pp. 1–13.

[6] Dapkus II, D.A., "Using MOS-Gated Power Transistor in AC Switch Applications", International Rectifier, Technical paper DT 94-5, www.irf.com, last visited December 18, 2015.

[7] Milojković, J., Litovski, V., and Le Blond, S., "Low-Voltage Circuit Breakers Based on WBG Solid-State Devices",J. of Circuits, Systems and Computers,Vol. 27, No. 2, February 2018 [19 pages]https://doi.org/10.1142/S0218126618500202.

[8] Gurpinar, E. and Castellazzi, A., "Single phase T-type inverter performance benchmark using Si IGBTs, SiC MOSFETs and GaN HEMTs", IEEE Trans. on Power Electronics, Vol. 31, No. 10, 2016, pp. 7148-7160.

[9] http://www.rohm.com/web/eu/search/parametric/-/search/MOSFETs?gclid=CO3M5abH5cgCFQSVGwodJqUJ8w, last visited November 2017.

[10] http://www.onsemi.com/pub/Collateral/NGTB30N135IHR-D.PDF, last visited November 2017.

[11] http://www.linear.com/solutions/ltspice, last visited November 2017.

[12] Lai, *Z.,* and Smedley, K. M., "A General Constant-Frequency Pulsewidth Modulator and Its Applications", IEEE Trans. on Circuits and Systems-I: Fundamental Theory and Applications, Vol. 45, No. 4, April 1998, pp. 386-396.

[13] Hobraiche, J., Vilain, J.-P., Macret, P., and Patin, N., "A New PWM Strategy to Reduce the Inverter Input Current Ripples", IEEE Trans. on Power Electronics, Vol. 24, No. 1, January 2009, pp. 172-180.

[14] Heydt, G. T., "Electric power quality", Stars and Circle Publications, West Lafayette, Indiana, USA, 1991.*and Optimization*", Chapman and Hall, London, 1997.

LSM Polyphase g_m-C Filter

Srdjan Milenković, Vančo Litovski, Slobodan Bojanić and Octavio Nieto Taladriz Garcia

Abstract – In modern telecommunication systems, receiver sensitivity is of paramount importance. On the other hand, receiver sensitivity is directly related to the receiver noise figure (NF). The lower NF the better receiver sensitivity. For portable devices, power consumption is also very important. This paper presents active polyphase g_m-C filter design approach which can be used as an integrated solution taking care of the above mentioned issues. The filer occupies low space when integrated while consuming low power. At the same time, due to its polyphase nature it suppresses received signal image at the negative frequencies when the receiver runs in low IF mode hence improving SNR i.e. NF. As an example, Bluetooth is used as the target communication standard.

Keywords – Polyphase filters, Bluetooth, LSM, IF, gyrator, transconductance.

I. Introduction

Integrated intermediate frequency (IF) filters are very attractive from a cost, size, and manufacturing point of view. The major drawbacks, compared to passive filters, are higher power consumption and a lower dynamic range. To minimize power consumption, the lowest possible centre (IF) frequency f_0 should be chosen. With traditional band-pass filters this will cause filter warping because of the non-linear low-pass-to-band-pass frequency transformation $H_{lp}(j\omega) \rightarrow H_{bp}(j\omega + 1/j\omega)$.

Furthermore, a low IF filter will suffer from an in-band image signal. This image is, with a traditional high IF, sufficiently suppressed by filters in front of the mixer, but such filtering is not practical with a low IF [4].

One solution to both frequency warping and in-band image is to use complex (polyphase) band-pass filters. Such filters have a linear frequency transformation and a high image rejection, thus supporting low IF architectures. For example, the specifications for the Bluetooth short-range radio require some 20 dB image rejection, which is easily achievable with an integrated design. Typical cellular systems like UMTS/FDD, requiring 50 dB of in-band blocking, are also well within the reach of the filter described here. The overall behaviour of the filter designed and presented in this paper fulfils the targeted Bluetooth specifications.

CMOS transconductance-C filter design techniques were reported in [2] [4]. In this paper we will design a polyphase filter based on a low pass prototype exhibiting 30 dB stop-band rejection while having excellent passband amplitude characteristic. Similar designs (based on elliptic prototypes) were reported in [4] [1].

Srdjan Milenković and Vančo Litovski are with Faculty of Electronic Engineering, University of Niš, Serbia, E-mail: srdjan@milenkovici.net, vanco@elfak.ni.ac.rs.

Slobodan Bojanić and Octavio Nieto Taladriy Garcia are with Universidad Politécnica de Madrid, E-mail: slobodan@ die.upm.es; octavio.nieto@die.upm.es

II. The LSM g_m-C filter

LSM polynomial filters [6] [8] are known to have critical monotonic amplitude characteristic in the passband and minimum area under the attenuation characteristic which makes them favourable from many design aspects as compared to their monotonic and non-monotonic counterparts. After the introduction of the rational version of the LSM filters [7] improvements were obtained both in the pass-band (by further reducing the area under the attenuation characteristic in the pass-band) and in the stop-band (by further improving their selectivity).

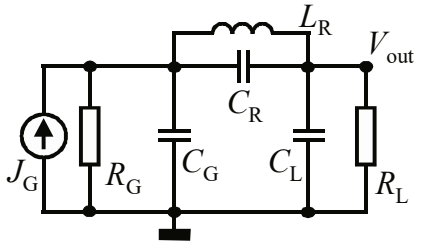


Figure 1. Third order LSM prototype filter

Table 1. Normalized element values of the prototype filter

Label	Normalized value
R_G	1
C_G	1.3133
C_R	0.1318
L_R	1.4964
C_L	0.8857
R_L	1

Having that in mind we used a third order LSM filter [9] exhibiting maximal stop-band attenuation of 30 dB as a prototype. Its transfer function is normalized so that it exhibits 3 dB at the cut-off frequency ω_c. Its schematic is depicted in Fig. 1 while Table 1 contains the normalized values of the circuit elements for ω_c =1 rad/s.

A g_m-C or Transconductance-C [2] based filter is a structure reminiscent to the active RC filters with the main difference being the absence of resistors; and integrated voltage amplifiers being substituted by operational transconductance amplifiers (OTA) [3]. The benefit is mainly in reduction of the silicon area since in application related to signal processing the OTA does not need large transistors.

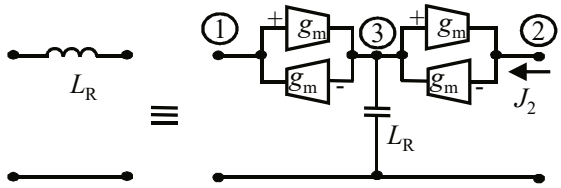


Figure 2. The equivalence used to create a simulated inductance.

One of the implementation of OTA in filtering circuits is to use gyrators [5] in order to simulate inductors. The analogy depicted in Fig. 2 [10] was used in order to simulate the floating inductor. It comes from the following development (based on the assumption of ideal OTA).

$$\begin{aligned} &j\omega L_R V_3 + g_m V_1 - g_m V_2 = 0 \\ &J_2 = -g_m V_3 \end{aligned} \tag{1}$$

After substitution one obtains

$$g_m(V_1 - V_2) = -j\omega L_R J_2 / g_m \tag{2}$$

and

$$(V_1 - V_2)/(-J_2) = j\omega L_R / g_m^2 \tag{3a}$$

For g_m=1 S, one gets

$$(V_1 - V_2)/(-J_2) = j\omega L_R \,. \tag{3b}$$

After substitution of the simulated inductor of Fig. 2 into the circuit of Fig. 1, simulation was performed using LTSPICE [11]. The resulting frequency characteristic is depicted in Fig. 3.

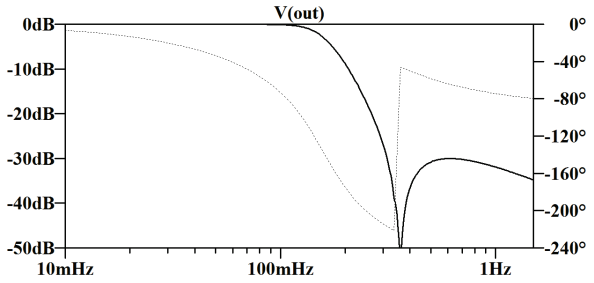


Figure 3. Frequency response of the LSM filter with L_R substituted according to Fig. 2.

III. THE POLYPHASE SOLUTION

The polyphase filter that is presented here has two inputs in-phase (I) and quadrature (Q) and two quadrature outputs (I and Q). Two transfer functions therefore characterize the filter. A low-IF receiver requires a polyphase filter with a passband from positive to positive frequencies, with an attenuation from negative to negative frequencies and with no signal transfer from positive to negative frequencies and vice versa. The transfer functions and the circuit synthesis of such a filter can be found by performing a linear frequency transformation on a low-pass filter characteristic. Equ. 4 gives this for a first order low-pass filter.

$$\begin{aligned} H_{lp}(j\omega) &= \frac{1}{1 + j\omega/\omega_c} \Rightarrow \\ H_{bp}(j\omega) &= \frac{1}{1 + j(\omega - \omega_0)/\omega_c} \end{aligned} \tag{4}$$

where ω_c is the cut-off frequency of the low-pass prototype filter (here considered normalized to unity) while ω_0 is the central frequency of the newly created band-pass filter.

It is usual the transformation given by (4) to be implemented directly to the circuit schematic by creating an equivalent circuit of the capacitor. Namely, i-mplementation of the transformation will lead to the following

$$\begin{aligned} I_{lp}(j\omega) &= j\omega C \cdot V_{lp} \Rightarrow \\ I_{bp}(j\omega) &= j\omega C \cdot V_{bp} - j\omega_0 C \cdot V_{bp} \end{aligned} \tag{5}$$

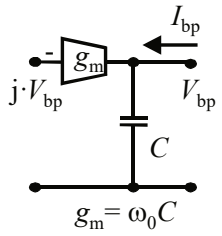


Figure 4. Implementation of the frequency transformation

Note that this transformation is shifting the poles and zeros of the prototype transfer function upwards for a value ω_0 which is usually larger than their imaginary part. In that way the resulting transfer function will have poles and zeros in the top half of the complex frequency plane. That makes $B_{bp}(j\omega)$ different for positive and negative frequencies.

The transformation depicted in Fig. 4 is to be implemented three times in order to create a polyphase filter based on the third order prototype: Twice for the original capacitors and once for the capacitor used within the simulated inductor circuit. The resulting schematic is depicted in Fig. 5. The currents J_{gi} and J_{gq} should be shifted in phase by 90°. In the subsequent simulation their amplitudes are set to 2 A.

Fig. 6 depicts the amplitude characteristic of the circuit of Fig. 5 where the amplitude (in dB) of V_{outi} is shown for positive and negative frequencies. ω_c=1 rad/s was used.

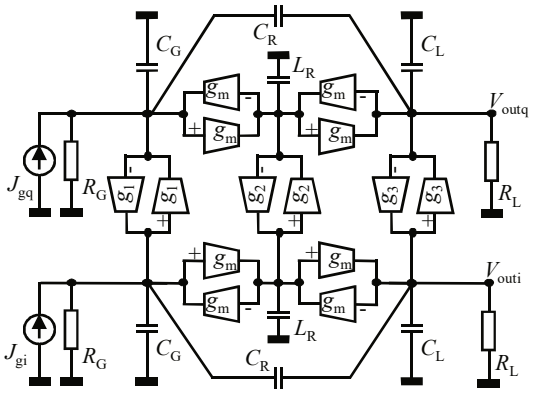


Figure 5. Third order LSM polyphase filter

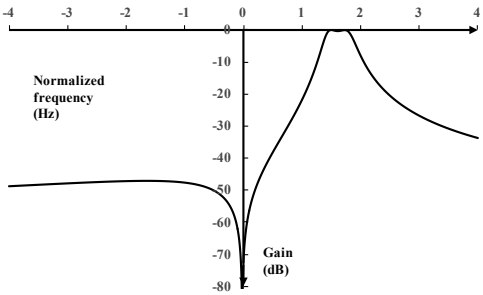


Figure 6. Amplitude characteristic of the LSM polyphase filter ($20 \cdot \log(|V_{outi}|)$ is depicted. ω_0=10·ω_c was used

IV. Some additional considerations

One may be interested on the migration of the poles and zeros due to the transformation given by (4) and (5).

Table 2. Normalized poles and zeros of the prototype LSM filter

Poles	-0.382025	0.917745
	-0.93057	0
	-0.382025	-0.917745
Zeros	0	2.250931
	0	-2.250931

To illustrate Table 2 contains the poles and zeros of the prototype LSM filter. As usual these are complex conjugate and located symmetrically to the real axis. If ω_0= 10 rad/s is used for the transformation the poles and zeros are to be shifted upwards as depicted in Fig. 7. Since no conjugates are present any more, the resulting transfer function has complex coefficients.

For example, the numerator polynomial of the third order LSM transfer function after transformation will become:

$$s^2 - 2js\omega_0 - (\omega_0^2 - \omega_p^2), \tag{6}$$

where ω_p is the frequency of the attenuation pole (here $\omega_p = 2.250931$.

This makes no use of all existing procedures [12] for filter synthesis so that, to our best knowledge, the only possible synthesis approach is the one described above.

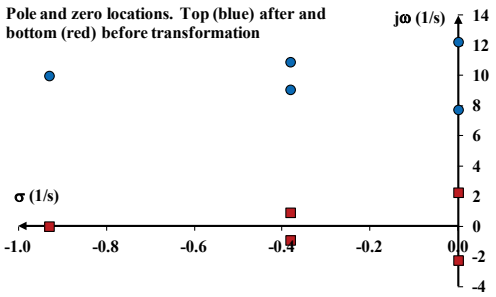


Figure 7. Pole and zero locations before (bottom) and after (top) implementation of the transformation

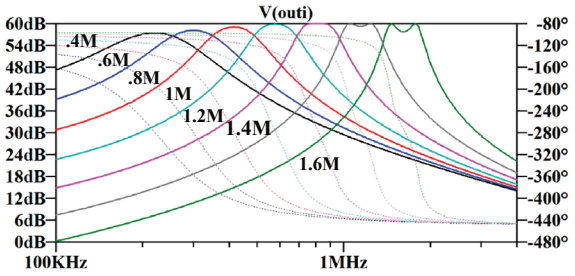


Figure 8. Influence of the ω_0/ω_c ratio to the shape of the passband gain characteristic.

It was noticed during the development of the simulation software that the transformation used here is not to be accepted as completely linear. Namely, due to the non-linear dependence of the amplitude characteristics on the circuit parameters (or transfer function's poles and zeros) distortion may be observed in the passband of the resulting polyphase filter. This phenomenon is illustrated in Fig. 8 where denormalized LSM filter was transformed several times with rising f_0 from 0.4MHz to 1.6MHz while f_c was kept to be 100 kHz. As it can be seen, at low f_0 the passband gain is distorted so that is more like Butterworth's response while at high frequencies with the rise of f_0 a dip grows in the middle of the passband. In the extreme case when $f_0 >> f_c$ one gets a transmission zero in the middle of the passband.

V. Conclusion

Polyphase g_m-C filter has been successfully designed by transforming low pass prototype into polyphase g_m-C filter architecture. The solution offers 30dB stop band attenuation and even better image rejection of 50dB. It has been observed that the performance of the polyphase filter after components denormalization depends on f_0/f_c ratio. However, there is an optimum f_0/f_c ratio around 10 where the resulting polyphase filter preserves frequency response within acceptable tolerances.

References

[1] Behbahani, F., Firouzkouhi, H., Chokkalingam, R., Delshadpour, S., Kheirkhahi, A., Nariman, M., Conta, M., and Bhatia, S., "*A Fully Integrated Low-IF CMOS GPS Radio with On-Chip Analog Image Rejection*", IEEE J. of Solid-State Circuits, Vol. 37, No. 12, December 2002, pp. 1721-1727.

[2] Nauta, B., "*A CMOS Transconductance-C Filter Technique for Very High Frequencies*", IEEE J. of Solid-State Circuits, Vol. 27, No. 2, February 1992, pp. 142-153.

[3] Glozić, D., Litovski, V., Bayford, R., "*ASCOTA3 ADIL: A New Reconfigurable CMOS Analogue VLSI Design Framework*", Microelectronics Journal, Vol. 25, No. 5, Aug., 1994, pp. 335-351.

[4] Andreani, P., Mattisson, S., and Essink, B., "*A CMOS gm-C Polyphase Filter with High Image Band Rejection*", Proc. IEEE ESSCIRC 2000, Sept. 2000, pp. 244–247.

[5] Tellegen, B.D.H., *"The gyrator, a new electric network element"*, Philips Res. Rep. 3, April 1948, pp. 81–101.

[6] Raković, B.D., Litovski, V., *"Least-Squares Monotonic Low pass Filters with Sharp Cutoff"*, Electronic Letters, Vol. 9, No. 4, 1973, pp. 75-76.

[7] Raković, B.D., Litovski, V., "*Monotonic Passband Low-Pass Filter With Chebyshev Stopband Attenu--ation*", IEEE Trans. on Acoustics, speech, and signal processing, Vol. ASSP-22, 1974, pp. 39-44.

[8] Topisirović, D., Litovski, V., and Andrejević Stošović, M., "*Unified theory and state-variable implementation of critical-monotonic all-pole filters*", Int. J. of Circuit Theory and Applications, Article first published online: 22 Oct. 2013; Vol. 43, No. 4, April 2015, pp. 502–515.

[9] Litovski, V., Raković, B.D., "*Attenuation Characteristics and Element Values of Least-Squares Monotonic Passband Filters With an Arbitrary Number of Transmission Zeros*", Publications of Electrical Engineering Faculty, Uni. Belgrade, Series Electronics, Telecommunications and Automation (ETA), No. 97-101, 1974, pp. 67-101.

[10] Uzunov, I.S., "*Theoretical Model of Ungrounded Inductance Realized With Two Gyrators*", IEEE Trans. on Circuits and Systems—II: Express Briefs, Vol. 55, No. 10, October 2008, pp. 981-985.

[11] http://www.linear.com/solutions/ltspice

[12] Zverev, A.I., "*Handbook of Filter Synthesis*", Wiley-Blackwell, Revised edition, ISBN-10: 0471749427, 2005.

Peak Windowing for Peak to Average Power Reduction

Borisav Jovanović, Srđan Milenković

Abstract - Linearization improves power amplifier (PA) efficiency and reduces running cost of the wireless infrastructure. The out-of-band radiation and bit error rate performance degradation are caused by PA nonlinearity. The operation of PA can be restricted to PA linear region by reduction of peak to average power ratio. The peak windowing method for peak to average power reduction is presented in the paper. The results are presented for Quadrature Phase Shift Keying (QPSK) and Wideband Code Division Multiple Access (WCDMA) waveforms.

Keywords - Crest factor reduction, Peak to Average Power Ratio, Peak Windowing method.

I. INTRODUCTION

The drawback of state-of-the-art modulation schemes is the high peak-to-average power ratio (PAPR), which results in intercarrier interference, high out-of-band radiation and bit error rate performance degradation [1]. To overcome this, large peaks at power amplifier (PA) input must stay within linear region of transfer function (Fig. 1). This requires that average PA output power must be much less than maximum saturated power (Fig. 1). PA power is reduced from its optimum operating point for the amount of output backed-off power (OBO). PA becomes energy inefficient. One solution which solves this problem and consequently reduces running cost of wireless infrastructure is dealing with signals which have reduced peak-to-average power ratio. In this case it is possible to increase signal average power without the risk of having the PA operating in non-linear region [1].

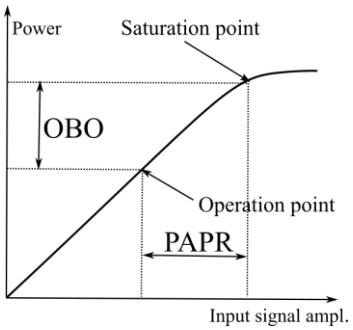


Fig. 1. The nonlinear and linear regions of PA transfer function

Borisav Jovanović is with the Faculty of Electronic Engineering Niš, University of Niš, Aleksandra Medvedeva Street 14, E-mail: borisav.jovanovic @elfak.ni.ac.rs

Srđan Milenković is with Lime Microsystems, Surrey Technology Centre, Occam Road, Guildford GU2 7YG, United Kingdom, E-mail: s.milenkovic@limemicro.com.

The occurrence of large peaks in waveform can be generally avoided:

1) by increasing complexity of the transmitter, including the transmission of some additional data needed at receiver side for the reliable reconstruction of useful data,

2) by modifying the shape of the signals to limit the magnitude of the large peaks at the cost of increased signal distortion [2].

II. CREST FACTOR REDUCTION

A. Crest Factor Measurement

We have already explained how signal characteristics have influence on PA power efficiency. One of these characteristics is the Crest Factor (CF). CF of a signal *s(n)* is defined as the ratio between the magnitudes associated to the largest $s(n)_{\max}$ and the average values $s(n)_{\text{rms}}$ of a signal:

$$CF = \frac{\|s(n)\|_{\max}}{s(n)_{rms}} \tag{1}$$

In literature one more parameter can be found - the Peak to Average Power Ratio (PAPR), which is the squared value of CF:

$$PAPR_{dB} = 10\log_{10}\frac{\|s(n)\|^2_{\max}}{s^2_{rms}} \tag{2}$$

When CF value of a signal is reduced, the PA can operate in the linear region. The primary goal when implementing the Crest Factor Reduction (CFR) is to reduce the CF value without introducing significant in-band and out-of-band distortions. Unfortunately, signal distortion cannot be completely avoided. To quantify the performance of CF reduction operations, the distortion is measured by Error Vector Magnitude (EVM) and Adjacent Channel Power Ratio (ACPR). [2]

B. Signal Quality Measurement

The EVM measures in-band signal distortion. EVM is defined as displacement of the received symbols in I/Q plane compared to the referent symbol positions. To calculate EVM, the average value of symbol power P_{ref} is computed, as well as the mean value of error vector power

P_{error}. The error vector represents the difference vector in I/Q plane that connects the received symbol vector to the referent one. The EVM is defined as a square rooted value of the ratio of mean error vector power P_{error} and average value of referent signal power P_{ref}: For WCDMA the EVM should be less than 12.5%.[3]

$$EVM(\%) = \sqrt{\frac{P_{error}}{P_{ref}}} \cdot 100 \quad (3)$$

The ACPR measures out-of-band signal distortion and it is defined as the ratio of power leaked to the adjacent channels P_{adj} and main channel power P_{main}.

$$ACPR = \frac{P_{adj}}{P_{main}} \quad (4)$$

The most important reason to keep the ACPR low is to prevent unwanted power to be transmitted outside of the frequency band of interest. For WCDMA signal the technical specifications define a minimum allowable ACPR limits for first and the second adjacent channels. These limits are 45 dBc for a first adjacent channel (5 MHz channel offset) and 50 dBc for a second adjacent channel (10 MHz channel offset). [3]

C. Crest Factor Reduction Techniques

Different techniques can be used for CF reduction. They can be divided into following groups: probabilistic (scrambling), coding, adaptive predistortion, clipping techniques, etc. [1] Some of the techniques don't distort the signal at all, at the price of a greater complexity, the others inject some distortion. Scrambling and coding technique requires special coding/decoding of the signal at the receiver side, which is not possible for implementation in our case.

The clipping techniques do not require special signal processing at receiver side, but have disadvantages in introducing both in-band and out-of-band signal distortion. The clipping techniques include Clipping and Filtering Technique (CAF), Block-scaling technique, Peak Windowing technique (PW), Peak Cancellation technique (PC) [3].

We have adopted PW algorithm for CF reduction. To evaluate the performance of PW method, we have created a PW CFR model in SystemC. The designed model has several options. The model implements a FIR filter operation, which is described in the next section in detail. Beside options dedicated for FIR filter configuration, we have option to set desired PAPR value of the output signal. The programmability is another goal of the future final implementation of CFR module in ASIC. When input waveform is changed, ASIC implementation should support adaptation to new parameters (new FIR filter order and coefficients).

The implemented module is evaluated using following waveforms: Quadrature Phase Shift Keying (QPSK) and Wideband Code Division Multiple Access (WCDMA). The results of evaluation are clearly presented.

III. Peak Windowing Method for Crest Factor Reduction

A. Method Description

Clipping and filtering is the conventional method. It includes hard clipping and low-pass filtering. In Peak Windowing method the original signal in the region of the peak is multiplied with a specific windowing function. The Kaiser, Hamming or Hanning functions can be used for this purpose [3].

The operation of clipping is described by Eq. (5):

$$y(n) = c(n)x(n), \quad (5)$$

where *x(n)* is the input signal, *y(n)* is the signal obtained after clipping operation is performed. The signal *c(n)* represents the clipping function:

$$c(n) = \begin{cases} 1, |x(n)| \le A \\ \dfrac{A}{|x(n)|}, |x(n)| > A \end{cases}, \quad (6)$$

where the parameter *A* is the clipping amplitude threshold. The process of clipping limits the *y(n)* to the level of *A*. Note that the signals *x(n)* and *y(n)* are complex signals consisting of I and Q signal components.

The process of clipping causes sharp edges in an output signal waveform, which gives unwanted out-of-band distortion. To reduce this distortion, the windowing function *w(n)* is applied. PW method intends to smooth sharp edges of a shortened signal. This not only improves the ACPR of the resulting signal but also preserves peak amplitude at the selected threshold value.

The windowing operation replaces the clipping coefficients *c(n)* with new ones *b(n):*

$$b(n) = 1 - \sum_{k=-\infty}^{k=\infty} a_k w(n-k), \quad (7)$$

where a_k are the weighting coefficients. Because the input signal is multiplied with windowing function, the output signal spectrum can be considered as convolution of the original signal spectrum and the spectrum of the used window [3]. This convolution is implemented by FIR filter. Also, to ensure that the value of *y(n)* is less than the threshold level A, the condition given by Eq. (8) must be satisfied:

$$b(n) \le c(n) \quad (8)$$

To minimize EVM the last inequality must be near the equality as much as possible. [3] This implies that window length should be narrow. If clipping rate and window length are too large, the adjacent windows overlap. Then, the convolution is generally larger, causing lower values of *b(n)*, more attenuation of *y(n)* and higher EVM.

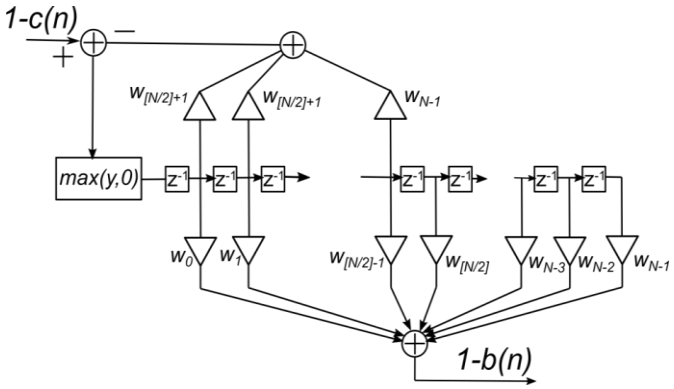


Fig. 2. The Peak Windowing CFR filter architecture

The PW filter structure is shown in the Fig. 2. The FIR filter takes at input the signal 1-*c(n)* (the *c(n)* is defined by Eq. (6)) and produces at output the signal 1-*b(n)*. A feedback structure adjusts the input values of the FIR filter. The delay of the FIR filter is equal to the time required for an input signal to reach the centre tap. Without a feedback, the resulting filter output will be larger than input value due to contributions of adjacent filter taps when clipping extends interval of several samples.

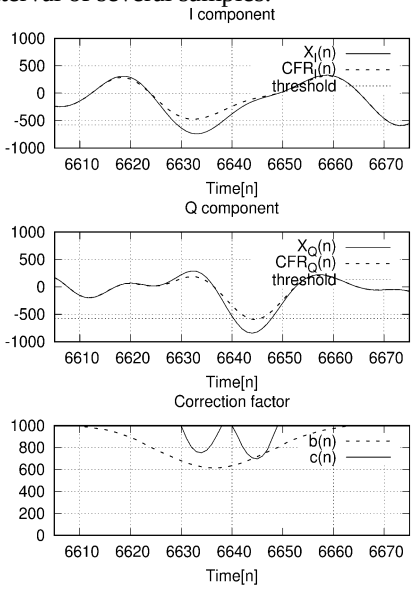


Fig. 3. Top and middle panels present I and Q components respectively of the signals at CFR block inputs X(n) and output CFR(n); the bottom panel presents the signals *b(n)* and *c(n)*

The feedback path scales filter input values. Looking forward to when clipped input value reaches the centre tap, the contribution of all previous input values (between first and centre tap) are calculated and used for correction of the next input value. [3] When incoming clipped signal reaches the centre tap, the contributions from all previous values have already been compensated. Then, the filter output *b(n)* becomes equal to the *c(n)*.

B. The Modelling of PW CFR block in SystemC

The PW algorithm is simulated using SystemC. Several modules are created to support SystemC simulations: the Crest Factor Reduction (CFR) block, PAPR and EVM calculation modules.

The CFR module takes at inputs I and Q quadrature signal components of signal *X(n)* and modifies them by clipping their magnitude to threshold level. Beside operation of clipping, the peak windowing is realized by the same module. For the realization of *c(n)* (Eq. (6)), the magnitude of input signal *X(n)* is calculated and after that, divided by threshold value. The peak windowing implementation is based on the FIR filter architecture given in the Fig. 2. The filter output produces signal *b(n)* (defined by Eq. (6)) which is later used as gain correction of delayed version of input signal. After these signals are multiplied the output signal *CFR(n)* (given in Fig. 3) is derived. Note that *CFR(n)* is complex signal consisting of I and Q components. The CFR module has following parameters: the order of embedded PW CFR filter, the clipping threshold level and arithmetic precision (number of bits of input signals).

For calculation of the PAPR value the new PAPR module is implemented in SystemC. PAPR value can be calculated for any input waveform and its operations are based on Eq. (2). In simulations the PAPR value is calculated for both input and output signals of CFR block.

The EVM module is created to find Error Vector Magnitude of QPSK signal. We have applied the Root Raised Cosine filter for QPSK demodulation. For obtaining ACPR we use already implemented SystemC modules.

III. SIMULATION RESULTS

In simulations we evaluated different clipping thresholds and filter orders. The clipping threshold is changed from 1.0, down to 0.56, with the 0.02 step. When threshold is changed, the PAPR value is also changed. It is calculated using PAPR module. The filter orders are: N=9, 19, 29 and 39. For each combination of selected threshold and filter order we have calculated PAPR, ACPR and EVM for the signal at the output of CFR block.

The goal was to find optimum filter order and clipping threshold which give the best performance in terms of ACPR and EVM. The utilized different waveforms are - QPSK and WCDMA (Test Model 1). The sample rate of these waveforms is 30.72 MS/s.

The results for QPSK and WCDMA are presented in diagrams below. It is known that reduction of PAPR will increase EVM and ACPR. The diagrams clearly show the

effects of varying the filter order.

A. Case 1: The QPSK signal

Without CFR, the PAPR of unclipped QPSK waveform is 6.29dB. The EVM is equal to 0.5% and ACPR is -90dBc.

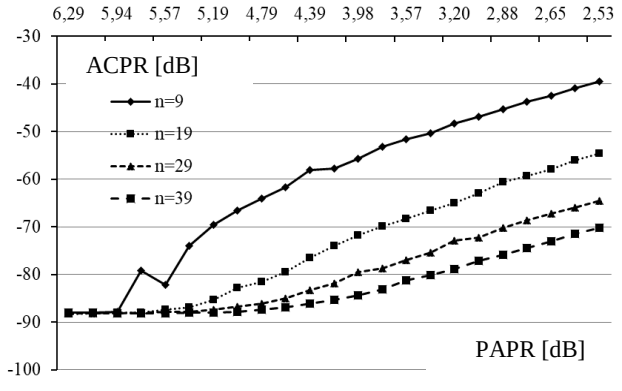


Fig. 4. The ACPR for QPSK signal at the PW filter output as a function of PAPR. The filter order values are: N=9, 19, 29 and 39

When CFR is used, the filter order is N=39 and the PAPR is reduced by 3dB, the ACPR is increased to nearly -80 dBc. EVM is worsened to 3.7%. In the case when PAPR is reduced by 3dB and N=9, the EVM = 1.8% and ACPR = -50 dBc. The Figs. 4 and 5. show trade-off in ACPR and EVM when selecting different filter order. Lower filter orders produce better EVM, but ACPR values get worsened.

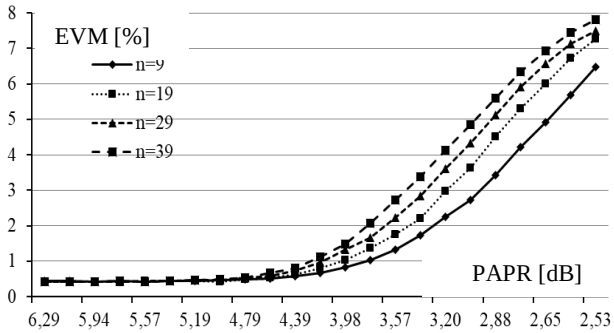


Fig. 5. The EVM of QPSK signal at the PW filter output as a function of PAPR. The filter order values are: N=9, 19, 29 and 39

B. Case 2: The WCDMA signal

The PAPR of input unclipped WCDMA waveform is 10.59 dB. The EVM = 0.3 % and ACPR = -90 dBc. When filter order N=39 is selected and PAPR is reduced by 3dB, the ACPR is increased to nearly -70 dBc and EVM is degraded to 3%. When N=9, the EVM=1.8% and ACPR = -45 dBc The Figs. 6 and 7. show ACPR and EVM plots versus PAPR for WCDMA signal.

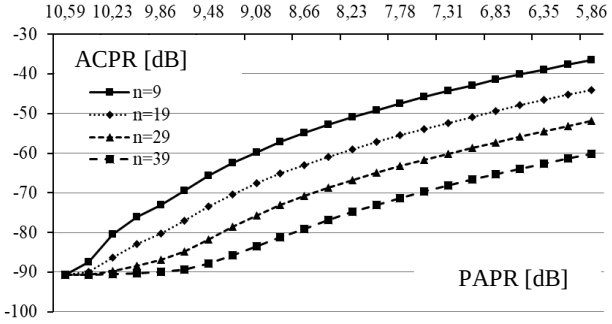


Fig. 6. The ACPR of WCDMA signal at the PW filter output as a function of PAPR. The filter order values are: N=9, 19, 29, 39

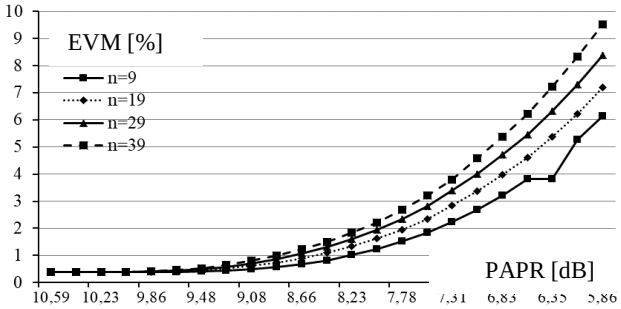


Fig. 7. The EVM of WCDMA signal at the PW filter output as a function of PAPR. The filter order values are: N=9, 19, 29, 39

IV. Conclusion

The signals with large peak to average power ratio require expensive wireless infrastructure and increase the running cost of equipment exploitation. This paper presents the peak windowing method for peak to average power reduction. The method was verified by simulation results. The various windowing lengths and clipping levels are changed for best performance, which is determined by measuring EVM and ACPR. Shorter window length minimizes EVM but degrades the ACPR value. The best performance is obtained for filter order value equal to 39.

References

[1] Lim, D.W., Heo, S.J., No J.S., "*An Overview of Peak-to-Average Power Ratio Reduction Schemes for OFDM Signals*", Journal of Communications and Networks, Vol. 11, No. 3, June 2009, pp. 229-239.

[2] Jiang, T., Wu, Y., "*An Overview: Peak-to-Average Power Ratio Reduction Techniques for OFDM signals*", IEEE Trans. Broadcasting, Vol. 54, No. 2, June 2008, pp. 257-268.

[3] Mistry, H., "*Implementation of a peak windowing algorithm for crest factor reduction in WCDMA*", Master of Engineering Thesis, Simon Fraser University, Canada, 2006

14nm Educational Design Kit: Capabilities, Deployment and Future

Vazgen Melikyan, Meruzhan Martirosyan, Anush Melikyan, Gor Piliposyan

Abstract— An open Educational Design Kit (EDK) which supports a 14nm FinFET design with all the necessary design rules, models, technology files, verification and extraction command decks, scripts, symbol libraries, and PyCells. It also includes a Digital Standard Cell Library (DSCL); an I/O Standard Cell Library (IOSCL); a set of memories (SOM) with different word and data depths; and a phase-locked loop (PLL). These components of the EDK augment any type of design for educational and research purposes. Though the EDK does not contain any foundry information, it allows 14nm FinFET technology with high accuracy to be implemented in the designs.

Keywords - design kit; low power; pycell.

I. Introduction

In the age of nanometer technologies, universities strive to provide the most modern and high quality studies in IC design. In addition to Electronic Design Automation (EDA) tools from leading companies, Educational Design Kits (EDKs), which include Digital and I/O Standard Cell Libraries for different IC fabrication technologies, are also necessary. But creation of such EDKs is challenged by numerous difficulties such as labor-intensive development and considerable complexity of verification. However, the most important of the challenges are the intellectual property (IP) restrictions imposed by IC fabrication foundries which do not allow universities to copy their technology into EDKs. That is why it became necessary for Synopsys to create an EDK which on one hand did not contain confidential information from foundries, and on the other hand, had the characteristics very close to the real design kits of the foundries.

II. Overview Of The Libraries

Synopsys has created 14nm FinFET Educational Design Kit (EDK) which is free from intellectual property restrictions and is targeted for educational and research purposes. It is aimed for programs training highly qualified specialists in the sphere of microelectronics at different universities, training facilities, and research centers. The EDK is intended to support the trainees so they can better master today's advanced design methodologies and the capabilities of Synopsys' state-of-the art IC design tools. It allows students to design different ICs using 14nm technology and Synopsys' EDA tools.

The Synopsys EDK contains the following: a technology kit (TK), a Digital Standard Cell Library (DSCL), an I/O Standard Cell Library (IOSCL), an I/O Special Cell Library (IOSpCL), a set of memories (SOM) and a phase-locked loop (PLL).

For the EDK's development, an abstract 14nm FinFET technology was used. While the EDK does not contain actual foundry data, which is confidential information from foundries, it is very close to the real 14nm technology. Using the abstract 14nm technology allowed Synopsys to create an EDK which can be used for study and research of real 14nm design characteristics.

III. Description Of The Technology Kit

The technology kit (TK) is a set of technology files needed to implement the physical aspects of a design. The generic TK for education contains:

A. Design Rules

These rules were created by using the MOSIS Scalable CMOS (SCMOS) design rules [6]. They provide greater portability of designs than if 14nm rules were developed because the sizes in 14nm rules can be larger by 5-20% than those in real foundry processes. An example design rule is illustrated in Figure1.

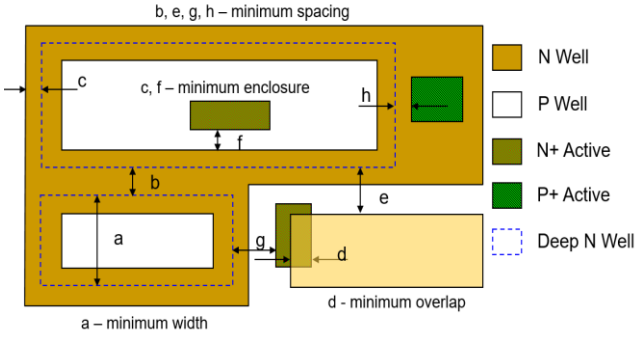


Fig. 1. Example Design Rule

Synopsys Armenia CJSC, Yerevan, Armenia,
E-mail: vazgenm@synopsys.com, meruzha@synopsys.com.
National Polytechnic University of Armenia, Yerevan, Armenia,
E-mail: anush-melikyan@yandex.ru
Synopsys Armenia CJSC, Yerevan, Armenia,
E-mail: piliposy@synopsys.com

B. *Device Formation*

This portion of the TK contains the description of available devices and their layout formation rules. It represents all the devices offered in the 14nm 0.8V/1.5V/1.8V generic process.

Figure 2 illustrates examples of device formation.

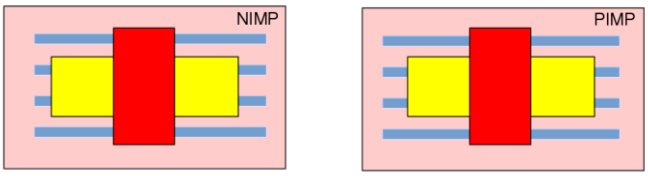


Fig. 2. Examples of Device Formation

C. GDSII Layer Map

This part of the TK contains layer names and GDSII numbers used in the 14nm process. Some layers such as dummy, marking, and text, have been added to the layer map. Any layer numbers may be chosen to form a generic process. A sample of the layer map is shown in Figure 3.

Layer #	Data type	Tape Out Layer	Drawing or Composite Layer	Layer name in Tech/Map File	Layer Name in DRC	Layer Name in LVS	Layer usage description
1	0	YES	Drawing	FIN	FINi	FINi	FIN
2	0	YES	Drawing	FINCUT	FINCUTi	FINCUTi	FIN cut for printing fine spaces
3	0	YES	Drawing	NWELL	NWELLi	NWELLi	NWELL
4	0	YES	Drawing	DNW	DNWi	DNWi	Deep NWELL
5	0	YES	Drawing	DIFF	DIFFi	DIFFi	Active area, thin oxide for device or interconnection
5	0	YES	Drawing	DDMY	DDMYi	DDMYi	Dummy DIFF layer; must be added if there's DIFF density rule violation
6	0	YES	Drawing	PIMP	PIMPi	PIMPi	P+ source/drain ion implantation
7	0	YES	Drawing	NIMP	NIMPi	NIMPi	N+ source/drain ion implantation
8	0	YES	Drawing	DIFF_15	DIFF_15i	DIFF_15i	1.5v thick oxide (second gate oxide)
9	0	YES	Drawing	DIFF_18	DIFF_15i	DIFF_15i	1.8v thick oxide (second gate oxide)

Fig. 3. Sample of Layer map

D. *Process Description*

This section of the TK provides approximate values of dielectric and metal thicknesses.

E. *Generic SPICE model library*

These are based on the Predictive Technology Model [1]. The SPICE model library contains the following devices:
- transistors,
a) 1.8V devices: thick oxide FinFETs,
b) 1.5V devices: thick oxide FinFETs,
c) 0.8V devices: thin oxide FinFETs with typical, high, and low threshold voltages. Each of these devices have five corner models: TT - both typical; FF - both fast; SS - both slow; SF – slow nfet/fast pfet; FS – slow pfet/fast nfet.
- diode
- rpoly and rmet resistors

In order to estimate the accuracy of the SPICE models, the models' parameters were scaled to 32nm technology to compare them with the characteristics of open 32nm models (Figure 4).

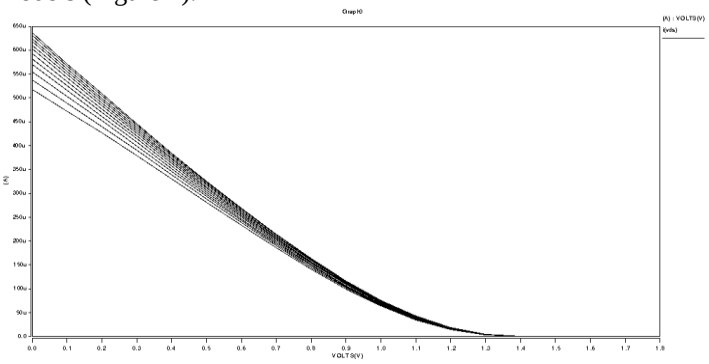


Fig. 4. Bunch of transfer curves for PMOS models

A set of DC transfer curves was obtained and the middle curve from the set was chosen as a typical corner for 1.8V devices, thereby assuring that it is close to the real foundry process. FF, SS, SF and FS corner models were formed by changing the threshold voltage (vth0) and oxide thickness (tox) in the range of +/-5%. Figure 5 shows the transfer curves for TT, FF and SS corners of a thin oxide NFET model.

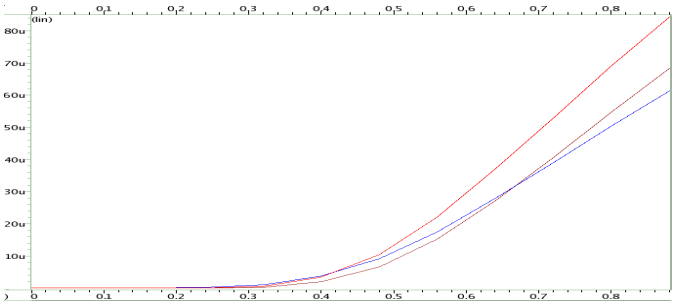


Fig. 5. TT, FF and SS corners of 0.8V Thin oxide High Vth NFET

F. *Milkyway technology file*

This file contains rules used by Synopsys' EDA tools.

G. *OpenAccess[3] symbol library and PyCells*

The OpenAccess symbols and PyCells in this library are: FinFET transistors, resistors, BJTs, and diodes. The PyCells were developed using the Python scripting language complying to PIL 1.1.7 standard to work in any compatible OpenAccess tool, i.e. Synopsys Galaxy™ Custom Compiler.

H. *DRC and LVS rule decks*

These are the design rules needed for Synopsys' IC Validator tools to perform design rule checks and layout vs. schematic.

I. Extraction files

These are files used by the Synopsys StarRC tool for parasitic extraction: ITF, TLU+, mapping, and command files.

J. Support scripts

A variety of additional scripts are required to support the design flow. For example, a script for controlling user input during schematic capture and PyCell setup scripts.

IV. Digital Standard Cell Libraries

The Digital Standard Cell Library (DSCL) is used for designing different ICs in 14nm FinFET technology with Synopsys' EDA tools. The DSCL builds using 1P9M 0.8V/1.5V/1.8V design rules and is aimed at optimizing the main characteristics of an IC design.

The DSCL contains a total of about 1200 cells. The library includes typical combinational logic cells with different drive strengths (inverter, inverting buffer, non-inverting buffer; 2-4 input AND, NAND, OR, NOR cells; 2-3 input XOR, XNOR cells; 2/1-2/2/2 AND-OR, AND-OR-Invert, OR-AND, OR-AND-Invert cells; multiplexers (2 to 1, 3 to 1, 4 to 1); half adder 1-bit and sequential, full adder 1-bit, pos edge DFF (w/ async low-active set, set & reset); neg edge DFF (w/ async active-low set, set & reset, only Q out, only QN out); scan pos edge DFF (w/ async active-low set, reset, set & reset, Q, QN; scan neg edge DFF(w/ async active-low set, reset), multibit DFFs, multibit scan DFFs, TAP cells, Synchronizers, ECO cells and Miscellaneous Cells.

The library also contains all the cells which are required for different styles of low power designs [2]. These cells enable the design of ICs with different core voltages to minimize dynamic and leakage power (clock gating cell, non-inverting delay line; pass gate; hold 0 isolation cell - logic AND; hold 1 isolation cell - logic OR and NOR type of isolation cell); high-to-low level shifter; pos edge retention DFF w/ async active-low clear; always on inverter; always on non-inverting buffer; always on high-to-low level shivter).

Composite Current Source (CCS) modeling technology is used for cell characterization to meet the requirements of contemporary low power design methods. CCS provides timing, noise, and power analyses s while considering the relevant nanometer dependencies. CCS allows the requirements of variation-aware analysis to be met.

In order to fully meet the requirements of low power design techniques, the DSCL characterizes for the 18 process/voltage/temperature (PVT) conditions (Table 1) as well as additional PVT conditions is used to characterize level-shifter cells.

TABLE I
Characterization Conditions

Normal Voltage Operating Condition	
Corner	PVT
tt0p8v25c	Typical/0.8/25^{0}c
tt0p8v125c	Typical/0.8/125^{0}c
tt0p8vm40c	Typical/0.8/-40^{0}c
ss0p72v25c	Slow/0.72/25^{0}c
ss0p72v125c	Slow/0.72/125^{0}c
ss0p72vm40c	Slow/0.72/-40^{0}c
ff0p88v25c	Fast/0.88/25^{0}c
ff0p88v125c	Fast/0.88/125^{0}c
ff0p88vm40c	Fast/0.88/-40^{0}c
Low Voltage Operating Condition	
tt0p6v25c	Typical/0.6/25^{0}c
tt0p6v25c	Typical/0.6/125^{0}c
tt0p6v25c	Typical/0.6/-40^{0}c
ss0p6v25c	Slow/0.6/25^{0}c
ss0p6v125c	Slow/0.6/125^{0}c
ss0p6vm40c	Slow/0.6/-40^{0}c
ff0p7v25c	Fast/0.7/25^{0}c
ff0p7v125c	Fast/0.7/125^{0}c
ff0p7vm40c	Fast/0.7/-40^{0}c

The DSCL has all the necessary deliverables: databook/user guide; layer usage file (.doc, .txt); symbols (.sdb, .slib); synthesis files (.db, .lib); Verilog simulation models (.v); HSPICE netlists (.sp); extracted RC netlists for (.spf); GDSII layout views (.gds); report files (.drc, .lvs); LEF files (.lef); FRAM views, layout views and runset files (.fram, .cel).

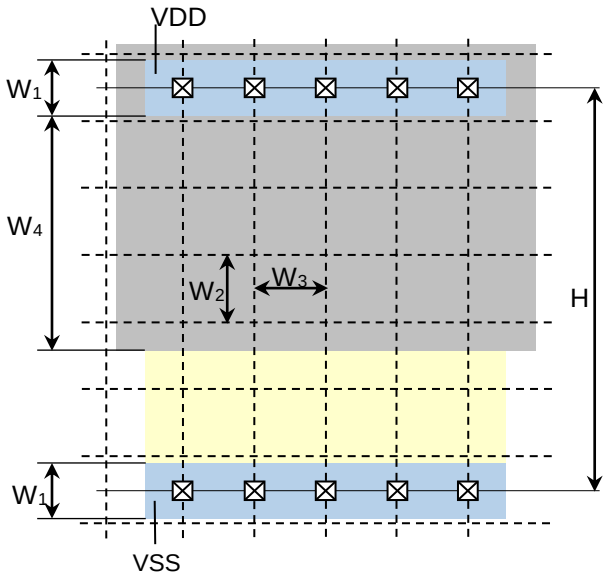


Fig. 6. Physical structure of single height cells

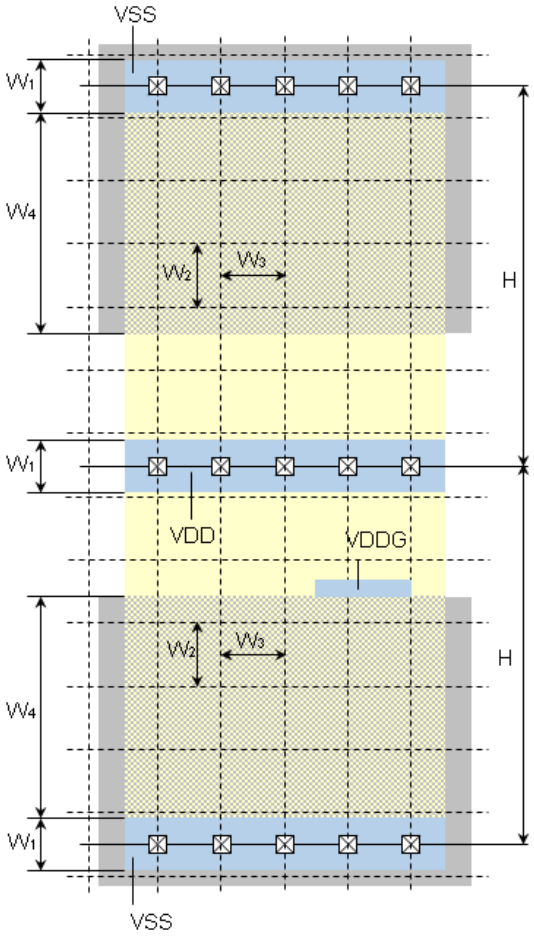


Fig. 7. Physical structure of double height cells

The selection of the physical structures (Figure 6, 7) of the digital cells was made to provide maximum cell density in digital designs as well as to take into consideration the requirements of low power design techniques.

Parameter	Symbol	Value
Cell height	H	0.6
Power rail width	W_1	0.094
Vertical grid	W_2	0.074
Horizontal grid	W_3	0.074
NWell height	W_4	0.3

V. I/O Cell Library

The I/O Standard Cell Library (IOSCL) is used for designing different integrated circuits (ICs) in 14nm technology using Synopsys' EDA tools. It was built using 14nm 1P9M 0.8V/1.5V/1.8V design rules.

Providing a complete set of standard functions, the IOSCL contains 50 cells (including FinFET non-inverting input buffer; FinFET non-inverting bi-directional cell; 4/8/12/16 mA tri-state driver with pull-up and pull-down; analog, non-inverting bidirectional without resistor pad with ESD protection; core power; I/O power; core ground; I/O ground pads; crosscoupling diode; IOVSS to VSS; decoupling capacitors VDD to VSS and IOVDD to IOVSS; break cell; corner pad; filler cell; bonding pad). CCS modeling technology was used for characterization of the IOSCL. All cells are available in wirebond and flip-chip variants with 25um x 200um in size, and all cells are available in EW (East-West) and NS(North-South) variants. for Flip-chip IC design library contains also bump cell for creating flip-chip array(Figure 8).

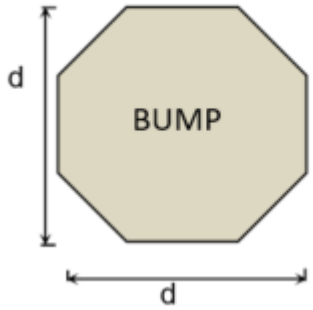


Fig. 8. Physical structure of bump cells for flip-chip array.

The I/O Special Cell Library (IOSpCL) is built using 14nm 1P9M 0.8V/1.5V/1.8V design rules with the same sized as standard I/O cells. The library includes I/Os complying to HSTL[4] and SSTL[5] standards.

CONCLUSION

An Educational Design Kit (EDK) was created and tested by Synopsys. It can be used for educational and research purposes, is free from intellectual property restrictions, and is representative of industrial design kits. It can be used in a wide range of design flows for digital, analog and mixed-signal designs using Synopsys' EDA tools.

REFERENCES

[1] Steine J., Castellanos I., Wood M., Henson J., Love F. Free PDK: An Open-Source Variation-Aware Design Kit.- Proceedings of the International Conference on Microelectronic Systems Education, San Diego, USA, 3-4 June, 2007, pp.173-174

[2] Keating M., Flynn D., Aitken R., Gibbons A., Shi K. Low Power Methodology Manual For System-on-Chip Design: Springer.-New York, 2007.-300p.

[3] Hahn M. OpenAccess: Standard and Practices, Design Syst., San Jose, CA, USA Volume: 29, Issue: 2 April 2012, 44-52pp.

[4] Shaoquan G., Jihua C., Jianwu M., Ting L., Design and verification of CMOS HSTL buffers - YF/spl I.bar/HSTL018, 6th International Conference On ASIC, Oct. 2005, 621-624pp.

[5] Jun-Hyun B, et. al, A crosstalk-and-ISI equalizing receiver in 2-drop single-ended SSTL memory channel, Custom Integrated CircuitsConference (CICC), 2010 IEEE, Sept. 2010, 1-4pp.

[6] R.Godman, K. Bartleson, T. Wood, K. Kranen , V. Melikyan, E. Babayan, 32/28nm Educational Design Kit: Capabilities, deployment and future - 2013 IEEE Asia Pacific Conference on Postgraduate Research in Microelectronics and Electronics (PrimeAsia)

Extended Character Segmentation Approach for Compressed Machine-Typed Documents

Vladan Vučković, Boban Arizanović, and Simon Le Blond

Abstract - This paper presents an analysis of threshold parameters used in a new character segmentation approach for machine-typed documents, together with efficient new image compression and decompression methods used in the real time OCR system. Provided results show that the character segmentation technique is robust to machine-typed documents from different typewriters, giving far superior results than state-of-the-art approaches.

Keywords - Image processing, OCR, Character segmentation, Machine-typed documents, Image compression.

I. Introduction

Character segmentation is a very important pre-processing stage in Optical Character Recognition (OCR) systems [1,2], and together with character recognition [3,4] has been an important subject of research for many years [5]. It should be emphasized that the difficulty of character segmentation is usually underestimated compared to the process of character recognition [6,7]. Previous work that deals with character segmentation in document images can be divided into machine-printed documents [6,8,9], where the document structure and the shape of its elements is regular, and handwritten documents where character segmentation is challenged due to irregular document structure [4,7,10]. Old machine-typed documents are of particular significance because important historical documents are often in this form [4,11,12].

Recent research of character segmentation includes all levels of this process. Analyses on image binarization parameters, used in document image pre-processing, showed that the Otsu method and other Otsu-based methods give the best results on average [12]. As a pre-processing stage of the character segmentation system, image compression and decompression are required to efficiently store the document images. Genetic algorithm based on discrete wavelet transformation information for fractal image compression was presented in [13]. A lossy image compression technique which uses singular value decomposition (SVD) and wavelet difference reduction (WDR) was proposed in [14]. Many methods for character segmentation have been proposed. A technique based on searching for connected regions in the spatial domain performed on a binary image was proposed in [15]. Another process uses the Bayes theorem for segmentation by exploiting prior knowledge, and is adapted for real time tasks [16]. Diverse methods for segmentation of handwritten documents are proposed. One technique exploits clustering in the process of segmentation [17]. To solve the problem of touching characters in handwritten documents, self-organizing maps, SVM classifiers, and Multi-Layer Perceptron are used [10,18].

This paper presents further improvements and analyses of the author's character segmentation approach, which forms part of a real time OCR system for the needs of the "Nikola Tesla Museum" in Belgrade [19-22]. The first contribution presented in this paper is evaluating pre-processing methods for document image compression and decompression, which take place after the image binarization, while the second contribution is a detailed analysis of the threshold parameters used in the character segmentation system. The results show that proposed image compression and decompression methods are better than JPEG and JPEG2000 image compression standards, giving up to 4-fold improved compression ratio. Results also show that the extended character segmentation presented is robust to variation in the typewriter, outperforming other methods in this respect also.

This paper is organized as follows: Section II offers the complete description of the proposed image compression and decompression methods, as well as analysis of threshold parameters. In Section III, a set of experimental results for image compression methods and segmentation accuracy are provided. Finally, discussion of the extended real time character segmentation method, results, and future work is given in Section IV.

II. Extended character segmentation approach

This section proposes new image compression and decompression methods used in the pre-processing stage of the character segmentation system and provides detailed analyses of threshold parameters used in the segmentation stage to control the segmentation process.

A. Image compression and decompression

Image compression/decompression allows decoupling of the character segmentation system into two independent stages. The first is a pre-processing stage with the document image compression and decompression as a final process, and the second is the document image segmentation. The most evident gain here is the ability to execute two independent system parts at different times. In this way, the document image compression and document image segmentation can be executed on different machines. Also, the previously compressed/decompressed document

Vladan Vučković and Boban Arizanović are with the Faculty of Electronic Engineering, University of Nis, Aleksandra Medvedeva 14, 18000 Nis, Serbia, E-mails: vladanvuckovic24gmail.com, bobanarizanovichotmail.com.

Simon Le Blond is with University of Bath, Department of Electronic & Electrical Engineering, Bath, United Kingdom, E-mail: S.P.leBlond@bath.ac.uk.

images can be processed using different versions of the segmentation engine. This is very important since it allows efficient testing of the segmentation engine. Finally, image compression allows efficient storing of document images which can save significant memory space.

A.I Image compression and decompression using RLE

The first proposed image compression and decompression methods employ the RLE algorithm for data compression. This approach is general and can be used for all types of binarized images, but RLE algorithm gives better compression results in the case of document images than other classes e.g. natural images. Illustration of the RLE algorithm including the coding format is given in Fig. 1.

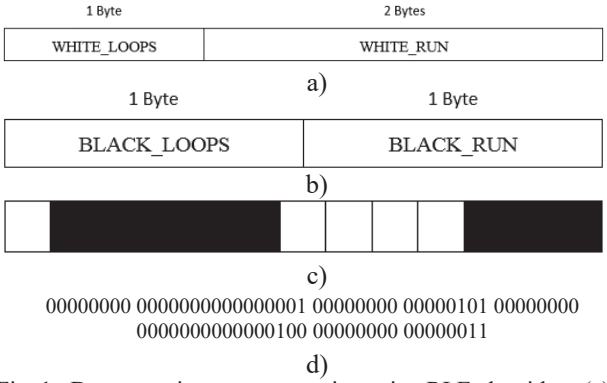


Fig. 1. Document image compression using RLE algorithm: (a) Compression format for white pixel runs, (b) Compression format for black pixel runs, (c) Example of pixel scanline, (d) Compressed pixel scanline

The RLE algorithm counts white and black pixels and stores the information about pixel runs in a compressed file. Storing is achieved using 3 bytes for all information about the white pixels and 2 bytes for information about the black pixels. In both cases the pixels are counted until the maximal value is reached. Since 2 bytes are used for white pixels (WHITE_RUN), this value is $2^{16} = 65536$, while in case of black pixels (BLACK_RUN) this value is $2^8 = 256$. When these values are reached, the WHITE_LOOPS or BLACK_LOOPS byte is incremented and WHITE_RUN and BLACK_RUN values are set to 0. The whole process of counting is then repeated. Since white pixels are a part of the background and dominate in document images, it is expected that 1 byte is not enough for storing the information about the number of consecutive white pixels. On the other hand, black pixel runs are expected to be short since they represent document characters and some spaces between characters are expected, thus only 1 byte is used for storing this information. Document image decompression is straightforward. The first byte is always multiplied by 256 for black pixels or 65536 for white pixels. This value is then incremented by the value of the next byte or 2 bytes. The obtained value represents the number of consecutive pixels of the same color in the current run of pixels. This process is repeated until the end of the compressed file.

A.II Image compression and decompression using document character contour extraction and scanline fill algorithm

The second proposed method employs the combination of the algorithm based on document character contour extraction and the scanline fill algorithm. The document image is processed in the horizontal and vertical direction and distances between the black pixels which represent starting and ending pixels of the black runs are stored in the compressed file. For this purpose, 2 bytes can be used to store the distance between two black pixels. It should be mentioned that in both compression methods the number of bytes used for storing the information about the white and black pixel runs is dependent primarily on the image dimensions. Small images are expected to have short runs, while large images are expected to have long runs of pixels of the same color. Therefore, 1 byte can be used for both white and black pixels in the case of small images, while in case of large images 2 bytes are necessary. Another important factor is a structure of a document image. If textual content dominates in a document image, background areas are less significant and thus even with large images, 1 byte can be used for storing the information about pixel runs. On the other side, if the background area dominates, even with medium images, 2 bytes are not enough to store the information about pixel runs.

After obtaining the offsets of black pixels which represent the contours of the characters, in the first step of the decompression method contours are drawn to the output image. The second step uses the iterative scanline fill algorithm. The main idea here is to scan the whole output image and fill the contours which represent the background with background color, while character contours will be filled with black color. This is achieved by repeating execution of the scanline fill algorithm. The background color of a document image is then replaced with white color and the original binarized document image is obtained.

B. Analyses of threshold parameters

This subsection provides detailed analyses of threshold parameters used in the character segmentation system. Analyses of document image binarization and line, word, and character segmentation are taken into consideration.

B.I Binarization

The binary image is obtained using a thresholding function T which is a gray-level transformation function of the form:

$$T(r) = \begin{cases} 0, & 0 \le r \le T_{hb} \\ 1, & T_{hb} < r \le r_{\max} \end{cases} \quad (1)$$

In the concrete case, value r_{max} is equal to 255. The threshold value T_{hb} that gives the best results is influenced primarily by the quality of the document image. If the document image is of low quality, a higher threshold value is required. Experimental results showed that in case of machine-typed documents the best results are achieved when T_{hb} takes values in the range 160-190. This threshold parameter affects line and word segmentation in some aspects, but character segmentation the most, since the spaces between the characters are small and the modified projection profiles technique can fail to separate the words into characters properly. In the case of images of high quality, T_{hb} can take values closer to 160, while in case of images of low quality this value must be closer to 190. The graphs which show a dependency of segmentation accuracy from the binarization threshold parameter T_{hb} are shown in Fig. 2.

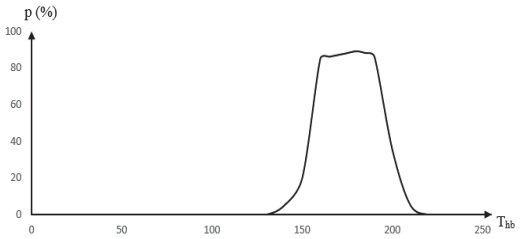


Fig. 2. Dependency of the segmentation accuracy from the binarization threshold parameter

The character segmentation system is intended for processing of document classes which contain documents typed on the same typewriter. Although it does not guarantee that documents which belong to the same class will be of the same quality, this is nevertheless likely. This means that a constant value for binarization threshold parameter can be used for different documents classes.

B.II Segmentation logic

Segmentation logic is based on a modified projection profiles technique, which uses histogram processing. A sliding window is used for calculation of the concentration of black pixels in the area of interest. Document line segmentation exploits the sliding window based method with vertical sliding. Suppose that the sliding window is size of N x H, where N is the width of the binary image f and H is the height of the sliding window. The concentration of black pixels in the sliding window is calculated as:

$$s_n = \sum_{s=0}^{H-1} \sum_{y=0}^{N-1} f(x+s, y) \tag{2}$$

Values s_n represent the values in the array of all sliding window concentrations of black pixels, S. The next condition is used for making a decision about which offsets will be taken as potential delimiters between lines:

$$d_x = \begin{cases} x + \left\lfloor \frac{H}{2} \right\rfloor, & s_n < T_{hswl} \\ -1, & otherwise \end{cases} \tag{3}$$

where d_x represents the offset relative to the top of the image and $\lfloor H/2 \rfloor$ represents integer division. Values d_x represent offsets of the middle lines of the windows which slide vertically along the area of interest. Only the offsets which belong to the sliding window with the concentration of black pixels lower than the chosen threshold value are considered. Using the offset analysis and considering the fact that the document line height is greater than d pixels, the closest offsets on distance greater than d are taken. Value H is the height of the sliding window which slides from top to bottom of the image. In general, the goal is to find local minima which represent spaces between the document lines. If value greater than 1 is chosen, it will lead to worse results, since multiple pixel scanlines are considered. In that case, the bigger value for threshold parameter T_{hswl} is necessary due to higher average concentration of black pixels in the sliding window. This would lead to unpredictable behavior and would be impossible to separate document lines that are very close to each other. Therefore, the sliding window height equal to 1 is the best choice and makes the process of document line segmentation easier to control. The value of the threshold parameter T_{hswl} depends on the document image structure and document image quality. The best choice is the smallest possible value which will separate correctly and completely the document image in lines. Even if this is achieved, it is possible and expected that other separators which cut the document lines will be found, too. For this reason, the previously mentioned value d is used to eliminate the wrong separators. This value is constant for documents inside the same document class and approximately equal to the document image character height.

The word segmentation process uses the sliding window based method with horizontal sliding. Suppose that a sliding window size of W x H is taken, where W is the width of the sliding window and $H = L_{cl} - L_{cu}$ represents the height of the current line. The concentration of black pixels in the sliding window is calculated as:

$$h_n = \sum_{x=L_{cu}}^{L_{cl}} \sum_{t=0}^{W-1} f(x, y+t) \tag{4}$$

where L_{cl} is the offset of the lower border and L_{cu} is the offset of the upper border of the given document line. The histogram is computed for each document line using the values h_n. The essence of the histogram analysis is in determination of the histogram valleys which represent the minima of the concentration of black pixels in the given line. The local minima less than the chosen threshold value are taken and these values represent delimiters between the words. Taking the word segmentation into consideration,

the sliding window width W and the threshold value T_{hsww} should be analyzed. Since the sliding window height is fixed and determined by the previously obtained document line height, the sliding window width controls the width of the spaces that are searched inside the document line. The value used for sliding window width ideally should be around the average width of spaces between the words. An alternative solution is to choose the value higher than average space between the characters inside the word. This will raise the possibility that only the spaces between the words are detected. This value also depends on document image structure, quality, and dimensions. The threshold parameter T_{hsww} has a role to eliminate the potentially wrongly detected spaces between the characters. Choosing the low value for T_{hsww} will ensure only the spaces between the words are detected, thus the latter choice is more suitable. The graph in Fig. 3 shows the dependency of the word segmentation accuracy from the sliding window width.

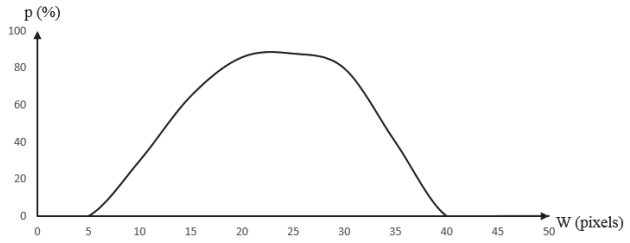


Fig. 3. Dependency of the word segmentation accuracy from the sliding window width

The approach used for character segmentation consists of the word alignment process, the already described histogram based method, and decision-making logic. The word alignment process represents the correction of dislocated words that appear due to the use of old typing machines. Sliding window height is again equal to the previously obtained document line height, while sliding window width depends on various factors. In case that a document image is of good quality, it is possible to perform quality image filtering and remove undesirable noise. But, in case of document images of low quality, even the best possible choice of the value for the binarization threshold parameter would not be helpful. The reason for this lies in sensitivity of the projection profiles technique used in the first part of the character segmentation process. Document images of low quality will have missing pixels in some part of the word and this area will be detected as a space between the characters. Also, the additional pixels in the areas which represent spaces between the characters will prevent the algorithm to detect them as spaces. A sliding window of generally smaller width than that used for word segmentation should be chosen for the first part of character segmentation. The value for sliding window width should approximately correspond to the average space between the characters. It depends primarily on document image dimensions. The threshold value T_{hswc} used to choose the potential delimiters within a word should be low enough to avoid detecting the wrong delimiters, and is strongly dependent on document image structure, quality, and dimensions.

The second part of the character segmentation is more specific and represents a decision-making logic. After determination of the local minima, which represent potential delimiters between characters, the threshold value for document character width T_{hcw} is used for determining the word length. The average character width is calculated as follows:

$$C_n = \left\lfloor \frac{W_w}{T_{hcw}} \right\rfloor \quad (5)$$

$$C_{wavg} = \left\lfloor \frac{W_w}{C_n} \right\rfloor \quad (6)$$

where C_n is the assumed number of characters in a word calculated using the threshold value T_{hcw}, W_w is the width of the given word in pixels, and C_{wavg} represents the average character width used in further processing. The word length is used to determine the number of delimiters in a word. Afterwards, the average character width can be calculated using the assumed word length, which is calculated using the threshold value T_{hcw}. The threshold value T_{hcw} is dependent on document image dimensions. This value ideally should be equal to the average character width, but even the slightly lower or higher value would be suitable. This deviation also depends on document image dimensions. The larger document images allow the bigger deviation and smaller document images allow smaller deviation values.

The segmentation of the given word starts from the left border, and delimiters at distance equal to the determined average character width are taken as the referent delimiters. The crucial part in character segmentation is the choice of the correct delimiter from the potential delimiters. The algorithm goal is to find the potential delimiter that is closest to the referent delimiter, where the maximal distance between the referent and potential delimiter is defined by the maximal offset allowed. In the case where there are no potential delimiters within the allowed distance, the referent delimiter will be chosen to be the real delimiter and the next referent delimiter will be set on a distance equal to the average character width from the chosen delimiter. Suppose that d_1, d_2,…,d_n is the sequence of potential delimiter offsets, where each of them is calculated as:

$$d_i = h_v + \left\lfloor \frac{W}{2} \right\rfloor \quad (7)$$

where h_v represents the given histogram valley and W is the width of the sliding window. The offset of the chosen delimiter is determined as follows:

$$j = \underset{i}{argmin}\left(\left|d_i - d_{ref}\right|\right) \quad (8)$$

$$d = \begin{cases} d_j, & \left|d_j - d_{ref}\right| \le T_{hoffset} \\ d_{ref}, & \left|d_j - d_{ref}\right| > T_{hoffset} \end{cases} \quad (9)$$

where j is the index of the chosen delimiter in the set of all potential delimiters, d_{ref} is the referent delimiter, and $T_{hoffset}$ is the threshold value for the maximal allowed distance between the closest potential delimiter and the referent delimiter. The threshold value $T_{hoffset}$ also primarily depends on document image dimensions. This value represents a deviation from the assumed delimiter position d_{ref}. In any case, this deviation is smaller than determined average character width C_{wavg}.

IV. Experiments

The proposed image compression and decompression methods, as a part of a character segmentation system, are tested on several PC machines. These methods are evaluated from the perspective of the image compression ratio and time complexity, to the perspective of the segmentation accuracy when specific compression methods are used. The second batch of experiments show the adaptability of the character segmentation approach for different classes of machine-typed documents. Documents typed on different typewriters require different threshold values, and by choosing the correct values for the thresholds, it is possible to achieve the high segmentation accuracy for all classes of documents. In order to obtain comparative results, JPEG and JPEG2000 image compression standards are used. Image compression using JPEG and JPEG2000 is performed using low, medium, and high quality compression, including lossless compression in the case of JPEG2000 compression.

The most important metric for evaluating the compression methods is compression ratio. Since the proposed image compression and decompression methods will be used in character segmentation system, their performances specifically on document images should be analyzed. To perform this analysis, image compression methods are tested using two document images. These document images are machine-printed documents since the second method is limited to machine-printed documents which have regular structure, thus character contour extraction is easier. Compression ratio results for these document images and different image compression methods are shown in Table I.

TABLE I

COMPARISON OF THE IMAGE COMPRESSION RATIO FOR MACHINE-PRINTED DOCUMENT IMAGES FOR DIFFERENT IMAGE COMPRESSION METHODS

Image Dimensions	Image File Size (KB)	Compression Ratio								
		JPEG			JPEG2000				**RLE**	**Contour Extraction/Scanline Fill**
		Low	Medium	High	Low	Medium	High	Lossless		
719x328	692	19.771	15.727	9.479	24.714	16.878	11.533	14.417	**67.184**	**46.443**
1266x924	3429	14.169	9.605	5.164	36.095	13.138	6.671	9.741	**27.878**	**19.373**

The proposed methods perform very well on document images based on results from Table I. This conclusion is expected since the presented algorithms perform better when the image contains huge areas of the same color, as in the case of document images. The RLE based method gives the best results, while the contour extraction method gives the second best results on average. To justify the usage of the second image compression method, the segmentation results for document images previously compressed and decompressed using different methods are given in Table II.

TABLE II

COMPARISON OF THE SEGMENTATION ACCURACY RESULTS FOR DIFFERENT IMAGE COMPRESSION METHODS USED IN THE PRE-PROCESSING STAGE

	Segmentation Accuracy (%)			
	JPEG	JPEG2000	**RLE**	**Contour Extraction/Scanline Fill**
Line Segmentation	81.54	81.54	**81.54**	**80.32**
Word Segmentation	78.28	78.28	**78.28**	**78.14**
Character Segmentation	87.08	87.08	**87.08**	**86.92**

The medium or high quality JPEG and JPEG2000 compression is shown since it behaves the same way as lossless compression after the additional binarization. Therefore, the results for JPEG, JPEG2000, and RLE based compression are identical. The most important conclusion here is that contour extraction based compression in combination with scanline fill decompression gives slightly worse results than previous compression methods. The reason for this lies in sensitivity of the evaluation metrics and also in the specificity of the character segmentation technique. In general, this technique is not sensitive to small changes in document image structure and therefore the segmentation accuracy results are similar to those obtained using the lossless compression methods.

Finally, a very important aspect of the image compression methods is time complexity since they are intended for a real time character segmentation system. In order to provide reliable results, the proposed image compression and decompression methods are tested on several PC machines and results are shown in Tables IV and V.

TABLE IV

PROCESSING TIME FOR RLE BASED COMPRESSION AND DECOMPRESSION METHOD (AMD ATHLON™ X4 840 QUAD CORE PROCESSOR 3.1 GHZ)

Image dimensions (pixels)	White Pixels/Black Pixels (%)	Processing Time (ms)			
		RLE Compression		RLE Decompression	
719x328	93.09:6.91	0.26639	0.30149	0.23231	0.27278
1266x924	83.51:16.49	2.04907	2.59600	1.94450	3.05122
2632x3575	98.06:1.94	10.61898	15.07239	13.35769	21.82402
2640x3612	98.69:1.31	10.79404	16.99354	13.31302	21.95595

TABLE V

PROCESSING TIME FOR CONTOUR EXTRACTION BASED COMPRESSION METHOD AND SCANLINE FILL DECOMPRESSION METHOD (DOCUMENT IMAGE SIZE OF 719X328)

PC Machine Specification	Processing Time (ms)			
	Contour Compression		Scanline Fill Decompression	
AMD Athlon™ X4 840 Quad Core Processor 3.1 GHz	0.26672	0.29824	1.59406	2.06575
Intel® Core™ i3-4150 CPU @ 3.50GHz	0.50409	0.52009	1.16008	1.18575
Intel® Core™ i5-750 CPU @ 2.67GHz	1.05348	1.06509	1.94235	1.95962

Time complexity results are obtained after 10000 executions of algorithms implementations. The first method also achieves excellent results in the case of document images with more black pixels which is a characteristic of documents with greater textual content. The second method processing time is primarily affected by the number of closed contours which represent character borders and must be filled using the scanline fill algorithm. Each time the closed contour must be filled, the scanline fill algorithm for region filling must be executed. Since the document image used for obtaining the results in Table V has huge areas of the same color and a small number of characters, the second method proved to be very efficient.

The second part of the experimental section is focused on the adaptability of the character segmentation system. In order to evidence this property, documents are divided into 12 classes, where each class represents a set of documents typed on the same typewriter. Therefore, 12 combinations of threshold values that give the best results for corresponding classes are chosen, and segmentation accuracy results when these combinations are used for processing are provided. The results for all levels of segmentation are shown in Table VI. Results rarely go below 80%. Text line segmentation results are above 90% in cases when spaces between document lines are bigger and when it is easy to determine the position of the lines. It is expected that text line segmentation results are good for all document classes, where spaces between lines are clear,

otherwise threshold values should be chosen carefully. Another important feature are threshold values for word segmentation and the determination of the average character width.

TABLE VI

SEGMENTATION RESULTS FOR CHARACTER SEGMENTATION APPROACH FOR DIFFERENT CLASSES OF MACHINE-TYPED DOCUMENTS

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
1	**88.62**	85.23	85.84	85.76	82.85	86.61	86.63	88.42	85.98	88.27	87.92	86.59
	87.16	87.34	86.97	84.67	84.62	85.43	87.24	85.21	85.14	87.32	86.42	87.43
	85.43	84.28	82.34	81.35	83.22	81.74	84.38	82.77	81.93	82.46	82.74	81.22
2	83.47	**87.51**	85.38	83.21	84.39	85.10	86.12	89.86	86.60	87.64	86.24	88.50
	85.36	**88.35**	84.69	84.18	85.13	86.42	85.46	85.39	84.38	87.05	85.30	87.39
	83.82	**86.79**	83.17	80.54	82.68	82.38	81.18	81.74	80.27	83.41	83.97	85.21
3	84.12	85.19	**87.15**	85.63	81.18	84.17	87.79	87.31	85.23	86.36	88.73	88.18
	85.31	83.25	**88.68**	82.78	82.44	85.68	86.54	86.12	85.83	87.39	88.65	86.35
	82.08	83.10	**85.27**	80.34	80.97	83.71	83.68	84.67	82.59	81.59	85.27	81.46
4	87.14	86.13	85.19	**88.30**	85.53	83.36	85.91	88.55	87.16	85.58	87.34	87.23
	86.29	85.67	85.23	**89.17**	85.49	82.54	86.33	86.34	86.57	86.33	88.28	86.54
	84.05	83.76	84.66	**86.14**	82.69	79.66	83.14	83.46	83.24	83.42	82.73	84.42
5	82.25	83.16	84.38	83.36	**88.59**	83.23	86.71	88.33	86.79	86.18	88.14	86.55
	85.63	84.80	84.57	80.46	**86.72**	85.74	86.93	87.24	85.47	88.27	86.79	85.04
	81.18	82.22	82.31	78.27	**86.71**	80.48	85.42	82.63	82.39	82.75	83.20	80.63
6	85.15	86.88	86.20	82.42	82.83	**92.68**	86.57	87.94	87.76	89.29	86.47	87.64
	81.31	83.07	84.62	84.79	82.57	**89.73**	85.12	85.24	87.52	87.66	85.32	86.32
	82.45	82.67	81.18	79.03	78.46	**87.64**	81.14	81.72	82.40	83.63	80.95	84.45
7	86.65	85.14	85.67	84.87	82.34	84.78	**91.05**	86.70	87.11	87.75	88.74	88.24
	86.25	85.31	85.34	84.49	80.65	85.57	**92.23**	87.23	86.25	86.49	87.51	87.93
	84.12	83.45	82.85	81.43	79.38	82.41	**88.42**	82.11	83.09	81.78	84.73	85.13
8	87.54	85.12	85.49	83.22	82.84	85.64	85.25	**93.18**	86.59	86.46	87.12	87.74
	82.37	84.35	83.06	84.53	83.76	85.72	87.64	**89.62**	85.17	86.23	85.44	86.53
	81.68	82.38	82.40	79.61	80.45	83.59	81.37	**88.14**	81.26	84.62	82.56	82.26
9	86.13	86.63	84.49	83.38	84.41	84.06	87.82	86.27	**92.43**	87.33	85.27	86.79
	84.34	85.54	85.16	82.76	82.70	86.37	86.52	85.37	**89.53**	86.29	87.41	87.43
	82.17	81.33	83.76	80.92	83.27	85.21	84.13	83.28	**88.77**	84.37	80.76	84.33
10	87.63	85.22	84.73	84.24	83.31	85.32	88.40	87.26	88.61	**92.75**	88.53	87.38
	84.41	86.15	83.35	84.55	84.75	87.45	87.25	85.34	87.69	**90.55**	86.22	86.42
	78.62	80.27	81.36	81.34	81.67	86.89	84.61	82.46	84.55	**85.64**	81.94	83.77
11	85.28	84.52	83.86	82.74	82.69	87.36	87.72	88.29	85.69	89.64	**89.11**	89.34
	83.43	83.37	82.93	80.46	84.35	85.28	88.25	87.33	86.23	86.76	**90.19**	87.29
	80.29	81.20	80.32	78.17	83.91	83.77	83.64	85.75	84.36	81.23	**87.49**	84.71
12	87.52	86.91	86.48	85.44	81.11	88.52	86.57	88.54	88.18	88.14	86.31	**91.48**
	84.93	85.13	84.37	83.68	83.67	86.44	87.26	86.30	86.22	86.67	86.47	**90.45**
	81.35	82.78	82.67	81.74	79.43	85.39	83.61	82.68	82.38	82.53	82.74	**88.92**

In both cases, for determining larger spaces between words and characters, higher threshold values are required. Also, for document images with greater character width, it is necessary to use higher threshold values in the process of determination of the average character width. Experiments also showed that the threshold value used for word segmentation could even be fixed, since it is usually clear which spaces represent the spaces between words. In contrast, selection of threshold value for character width is crucial, since it is the entry point to the decision-making algorithm. The threshold value for the maximal allowed offset between the referent and the potential delimiter is also important. Based on experiments, this value is usually between 5 and 10 pixels.

IV. Conclusion

This paper presents an image compression/decompression stage of the author's existing character segmentation approach, together with analysis of the threshold parameters used in the character segmentation stage. In Section II the image compression and decompression methods are presented and analysis of threshold parameters is provided. The presented methods use the RLE data compression algorithm and document character contour extraction for image compression, and the scanline fill algorithm for document image decompression. The character segmentation method is adaptive, since it can be used for character segmentation of the documents with different characteristics. In Section III a set of experimental results is provided for image compression methods and adaptability of the character segmentation system is proved using documents from different classes. The proposed methods perform up to 4 times better than JPEG and JPEG2000 image compression standards. Results show that choosing the correct threshold values leads to high segmentation accuracy for documents of different classes. Future work will be focused on algorithm improvement and its integration into the

complete OCR system.

Acknowledgement

This paper is supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Project III44006-10), Mathematical Institute of Serbian Academy of Science and Arts (SANU), and The "Nikola Tesla Museum" (providing original typewritten documents of Nikola Tesla).

REFERENCES

[1] Bourbakis, N., Pereira, N., Mertoguno, S., *"Hardware design of a letter-driven OCR and document processing system"*, Journal of Network and Computer Applications, Vol. 19, No. 3, 1996, pp. 275-294.

[2] Mao, J., Mohiuddin, K. M., *"Improving OCR performance using character degradation models and boosting algorithm"*, Pattern Recognition Letters, Vol. 18, No. 11-13, 1997, pp. 1415-1419.

[3] Namane, A., Guessoum, A., Soubari, E. H., Meyrueis, P., *"CSM neural network for degraded printed character optical recognition"*, Journal of Visual Communication and Image Representation, Vol. 25, No. 5, 2014, pp. 1171-1186.

[4] Razzak, M. I., Anwar, F., Husain, S. A., Belaid, A., Sher, M., *"HMM and fuzzy logic: A hybrid approach for online Urdu script-based languages' character recognition"*, Knowledge-Based Systems, Vol. 23, No. 8, 2010, pp. 914-923.

[5] Fujisawa, H., *"Forty years of research in character and document recognition-and industrial perspective"*, Pattern Recognition, Vol. 41, No. 8, 2008, pp. 2435-2446.

[6] Lu, Y., *"Machine Printed Character Segmentation - An Overview"*, Pattern Recognition, Vol. 28, No. 1, 1995, pp. 67-80.

[7] Lu, Y., Shridhar, M., *"Character segmentation in handwritten words - An overview"*, Pattern Recognition, Vol. 29, No. 1, 1996, pp. 77-96.

[8] Min-Chul, J., Yong-Chul, S., Srihari, S. N., *"Machine Printed Character Segmentation Method Using Side Profiles"*, Proceedings of IEEE SMC '99 Conference on Systems, Man and Cybernetics, 1999.

[9] Park, H. C., Ok, S. Y., Yu, Y. J., Cho, H. G., *"A word extraction algorithm for machine-printed documents using a 3D neighborhood graph model"*, International Journal on Document Analysis and Recognition, Vol. 4, No. 2, 2001, pp. 115-130.

[10] Lacerda, E. B., Mello, C. A. B., *"Segmentation of connected handwritten digits using Self-Organizing Maps"*, Expert Systems with Applications, Vol. 40, No. 15, 2013, pp. 5867-5877.

[11] Younes, M., Abdellah, Y., *"Segmentation of Arabic Handwritten Text to Lines"*, Procedia Computer Science, International Conference on Advanced Wireless Information and Communication Technologies (AWICT 2015), Vol. 73, 2015, pp. 115-121.

[12] Gupta, M. R., Jacobson, N. P., Garcia, E. K., *"OCR binarization and image pre-processing for searching historical documents"*, Pattern Recognition, Vol. 40, No. 2, 2007, pp. 389-397.

[13] Wu, M., *"Genetic algorithm based on discrete wavelet transformation for fractal image compression"*, Journal of Visual Communication and Image Representation, Vol. 25, No. 8, 2014, pp. 1835-1841.

[14] Rufai, A. M., Anbarjafari, G., Demirel, H., *"Lossy image compression using singular value decomposition and wavelet difference reduction"*, Digital Signal Processing, Vol. 24, 2014, pp. 117-123.

[15] Zheng, Z., Zhao, J., Guo, H., Yang, L., Yu, X., Fang, W., *"Character Segmentation System Based on C# Design and Implementation"*, Procedia Engineering, International Workshop on Information and Electronics Engineering, Vol. 29, 2012, pp. 4073-4078.

[16] Grafmüller, M., Beyerer, J., *"Performance improvement of character recognition in industrial applications using prior knowledge for more reliable segmentation"*, Expert Systems with Applications, Vol. 40, No. 17, 2013, pp. 6955-6963.

[17] Venkateswarlu, N. B., Boyle, R. D., *"New segmentation techniques for document image analysis"*, Image and Vision Computing, Vol. 13, No. 7, 1995, pp. 573-583.

[18] Bae, J. H., Jung, K. C., Kim, J. W., Kim, H. J., *"Segmentation of touching characters using an MLP"*, Pattern Recognition Letters, Vol. 19, No. 8, 1998, pp. 701-709.

[19] Vučković, V., Arizanović, B., *"Efficient character segmentation approach for machine-typed documents"*, Expert Systems with Applications, Vol. 80, 2017a, pp. 210-231.

[20] Vučković, V., Arizanović, B., *"Automatic document skew pre-processor for character segmentation algorithm"*, Facta Universitatis: Electronics and Energetics, Vol. 30, No. 4, 2017b, pp. 611-625.

[21] Vučković, V., Arizanović, B., Le Blond, S., *"Ultra-fast basic geometrical transformations on linear image data structure"*, Expert Systems with Applications, Vol. 91, 2018a, pp. 322-346.

[22] Vučković, V., Arizanović, B., Le Blond, S., *" Generalized N-way iterative scanline fill algorithm for real-time applications"*, Journal of Real-Time Image Processing, Vol. 13, No. 4, 2018b, pp. 1-19.

General One-Dimensional Model of a New Composite Ultrasonic Transducer

Igor Jovanović, Uglješa Jovanović and Dragan Mančić

Abstract - This paper presents an original general one-dimensional model of a new high-power composite ultrasonic transducer. Modelling of the composite transducer is performed using a one-dimensional analysis, which includes only thickness oscillation modes while radial oscillations are ignored. Unlike the most one-dimensional models that do not consider impact of the prestressing bolt or include only a part of it, the realized one-dimensional model includes all transducer components, as well as a central bolt with its head. Additionally, in order to prove the correctness of the transducer modelling process using the proposed model, a transducer model that ignores the central bolt impact will also be presented. Verification of the proposed one-dimensional models is performed by comparing the modelled dependencies of input electrical impedance vs. frequency with the experimental results.

Keywords - High-power ultrasound, Composite ultrasonic transducer, One-dimensional modeling.

I. Introduction

At the very beginning of ultrasound development its application was in sonars. In nowadays, ultrasound is in applied on solid bodies, liquids and gases with the desired effect. The main applications of high-power ultrasound are ultrasonic cleaning and machining of materials. The high-power ultrasonic system consists of an electromechanical transducer and a power source operating in a predefined frequency range. Since ultrasonic energy is transferred to working environment through coupling elements, acoustic power emitted by the transducer depends on acoustic impedance of working environment and achieved adjustment. Design of efficient ultrasonic transducer includes a detailed analysis of mechanical and electrical transducer characteristics during different operating conditions. If the transducer emits ultrasonic waves to a complex acoustic load, the resonant frequency will change due to the change of boundary conditions on the transducer working surface [1].

The most widely used modelling approach for ultrasonic transducers, found in literature, is application of one-dimensional theory using equivalent electromechanical circuits. Electromechanical circuits use Mason's theory to model piezoelectric ceramics and symmetric T quadruples to model passive transducer elements [2], [3]. Therefore, the new composite transducer, analysed in this paper, is presented in the simplest form as a network with two electrical and two mechanical approaches. When model includes a bolt and various electrical connections of the transducer, then the number of electrical and mechanical approaches in the electromechanical equivalent circuit increases. Presence of the bolt becomes significant when determining resonant frequency of transducers with small axial dimensions. On the other hand, the bolt impact can be ignored in case of transducers with longer metal attachments. In literatures [4], [5], [6], [7] and [8], the bolt impact is not taken into the account since it is considered to be negligible, while in literature [8] the T networks of particular passive elements are even more simplified. In literatures [9] and [10], impact of the bolt part passing through ceramics is analysed in the form of parallel connection of the corresponding T quadruples. The bolt part passing through ceramics is presented along with the quadrupole impedances in the mechanical part of the equivalent electromechanical circuit in literatures [11] and [12], or directly via wave equations in literature [13].

In this paper, modelling of the realized composite transducer with new structure, which represents a special unidirectional composite ultrasonic transducer [14], is performed. Prestressing of the structure is achieved using the central bolt which is not in the contact with the central mass. Consequently, the central mass performs compressions and expansions in cycles, simultaneously with the changes of axial dimensions of the entire transducer due to mutually opposite polarization of piezoceramic rings in active blocks. Most of the problems related to impedance and frequency adjustment, as well as to mechanical load couplings, can be avoided by employing such composite transducer.

The paper also presents the general one-dimensional model of the high-power composite ultrasonic transducer with the new structure, which takes into the account all transducer components including a central bolt with the bolt head. The model is represented as a passive electromechanical equivalent circuit, the application of this equivalent circuit is based on the idea that propagation speed of ultrasonic waves is equivalent to electric current, while mechanical force is equivalent to electric voltage. In order to prove the correctness of the described transducer modelling process by the proposed general model, the paper also presents the one-dimensional model that does not include the central bolt impact.

Igor Jovanović, Uglješa Jovanović and Dragan Mančić are with theUniversity of Niš, Faculty of Electronic Engineering, 14 Aleksandra Medvedeva, 18000 Niš, Serbia (e-mail: igor.jovanovic@elfak.ni.ac.rs, ugljesa.jovanovic@elfak.ni.ac.rs and dragan.mancic@elfak.ni.ac.rs).

II. Analytical One-Dimensional Modelling of High-Power Ultrasonic Composite Transducer

The complete general one-dimensional model of the piezoelectric composite transducer that takes into the account the dimensions of the bolt and its head as well as the piezoelectric properties of the exciting ceramic is shown in Fig. 1. The model represented as the equivalent circuit is developed with the one-dimensional theory [2], [3] and it is modified based on the new structure of the composite transducer presented in literature [14]. In the proposed one-dimensional model, input electrical impedance depends on dimensions, material characteristics and resonant frequencies of the each transducer part included in the model.

Elements of the circuit shown in Fig. 1 corresponding to the isotropic metal transducer parts made of different materials are calculated as:

$$Z_{i1} = jZ_{ci}tg\frac{k_i l_i}{2},\ Z_{i2} = \frac{-jZ_{ci}}{\sin(k_i l_i)} \qquad (1)$$

wherein $Z_{ci}=\rho_i v_i P_i$ and $k_i=\omega/v_i$ (for i=1, 2, 3, ..., 6) are characteristic impedances and the corresponding wave numbers. ρ_i are densities, l_i and P_i are lengths and surface areas of the cross-sections, v_i are the velocities of longitudinal ultrasonic waves propagation through the corresponding elements.

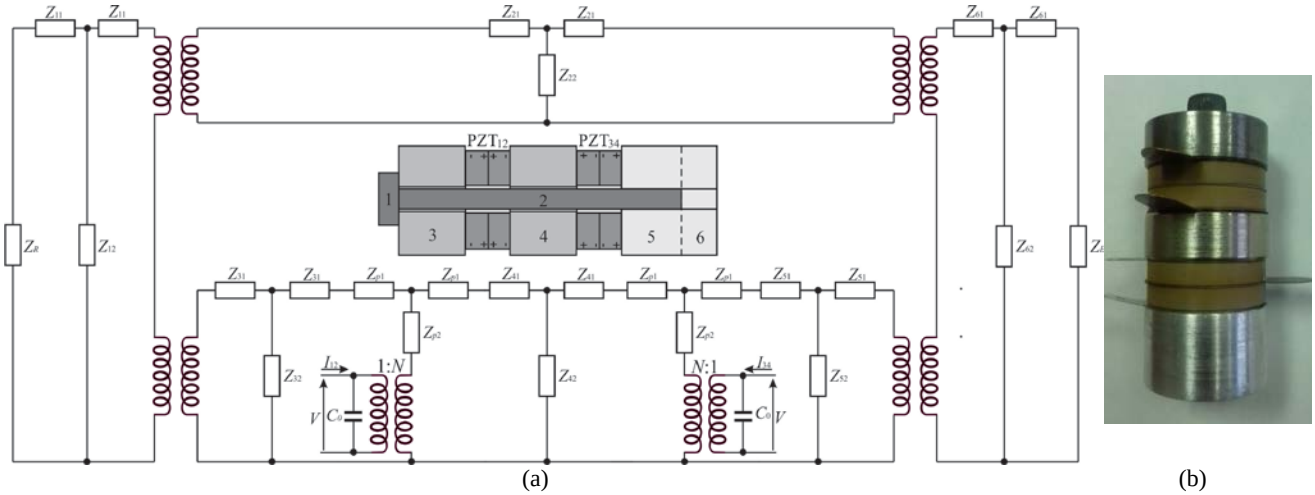
Elements of the circuit shown in Fig. 1 corresponding to the piezoceramic rings in the upper active layer (PZT_{12}) and the piezoceramic rings in the lower active layer (PZT_{34}) are determined as:

$$Z_{p1} = jZ_{cp}tg\frac{nk_p l_p}{2},\ Z_{p2} = \frac{-jZ_{cp}}{\sin(nk_p l_p)} \qquad (2)$$

wherein $Z_{cp}=\rho_p v_p P_p$ and $k_p=\omega/v_p$ are characteristic impedances and corresponding wave numbers, respectively. ρ_p, l_p, P_p are densities, lengths and surface areas of the piezoceramic cross-sections, v_p are velocities of longitudinal ultrasonic waves propagation, respectively. The input electric voltages and currents are marked as V, I_{12} and I_{34}. The piezoceramic models consist of capacitance $C_0=n\varepsilon_{33}^S P_p/l_p$, and ideal transformers with transmission ratios (electromechanical coefficients of the coupling) $N=h_{33}C_0/n$, wherein n is the number of piezoceramic rings per the active layer (in the case of the particular composite transducer n is 2). The piezoelectric properties of the transducer active layers are represented by piezoelectric constant h_{33} and relative dielectric constant of the pressed ceramic ε_{33}^S.

Piezoceramic rings (between which is the central mass situated) are mechanically connected in series with emitter and reflector attachments. Emitter and reflector attachments are closed with acoustic impedances Z_E and Z_R, which are in this case negligible because experimental measurements were conducted with unloaded transducers oscillating in the air. The metal bolt extends along the entire structure and is therefore connected mechanically in parallel to the remaining elements in the scheme. As already noted, most one-dimensional models do not include the bolt impact, or include only a part of it, hence the accuracy of the model in predicting the transducer behaviour is reduced, even in the case of transducers with longer metal attachments.

Due to the complexity of the expression for the composite transducer input impedance, observed on electrical approaches, as well as to prove the correctness of the transducer modelling by the proposed general one-dimensional model, which includes the impact of the prestressing bolt, the analytical model of the composite transducer, which ignores the central bolt impact is presented in the following text.



(a) (b)

Fig. 1. (a) General one-dimensional model; (b) Realized composite transducer

If the central bolt impact is neglected, based on the equivalent circuit scheme shown in Fig. 1 ($Z_{11}=Z_{12}=Z_{21}=Z_{22}=0$) transducer input electrical impedance can be calculated as:

$$Z_e = \frac{Z_{12}Z_{34}}{N^2\left(Z_{12}+Z_{34}\right)+ j2\omega C_0 Z_{12}Z_{34}} \quad (3)$$

wherein $\omega=2\pi f$ is angular frequency, $Z_{12}=V/I_{12}$ and $Z_{34}=V/I_{34}$ are input electrical impedances of the corresponding active layers obtained by following expressions:

$$Z_{12} = \frac{Z_{e5}-\frac{Z_{e2}}{Z_{e3}}Z_{e6}}{Z_{e4}-\frac{Z_{e1}}{Z_{e3}}Z_{e6}},\ Z_{34} = \frac{Z_{e3}-\frac{Z_{e6}}{Z_{e5}}Z_{e2}}{Z_{e1}-\frac{Z_{e4}}{Z_{e5}}Z_{e2}} \quad (4)$$

In Eq. 4 the newly introduced equivalent impedances are obtained by following expressions:

$$Z_{e1} = 1+\frac{Z_{p1}+Z_{41}+Z_{42}}{Z_{e7}}+\frac{Z_{42}}{Z_{e8}},$$

$$Z_{e2} = Z_{42}+Z_{p2}+Z_{p1}+Z_{41}+\frac{Z_{p2}\left(Z_{42}+Z_{p1}+Z_{41}\right)}{Z_{e7}}$$

$$Z_{e3} = Z_{42}+\frac{Z_{p2}Z_{42}}{Z_{e8}}, \quad (5)$$

$$Z_{e4} = 1+\frac{Z_{p1}+Z_{41}+Z_{42}}{Z_{e8}}+\frac{Z_{42}}{Z_{e7}},$$

$$Z_{e5} = Z_{42}+\frac{Z_{p2}Z_{42}}{Z_{e7}},$$

$$Z_{e6} = Z_{42}+Z_{p2}+Z_{p1}+Z_{41}+\frac{Z_{p2}\left(Z_{42}+Z_{p1}+Z_{41}\right)}{Z_{e8}}$$

that is:

$$Z_{e7} = \frac{\left(Z_{rf}+Z_{31}\right)Z_{32}}{Z_{rf}+Z_{31}+Z_{32}}+Z_{31}+Z_{p1},$$

$$Z_{e8} = \frac{\left(Z_{ef}+Z_{(5+6)1}\right)Z_{(5+6)2}}{Z_{ef}+Z_{(5+6)1}+Z_{(5+6)2}}+Z_{(5+6)1}+Z_{p1} \quad (6)$$

Based on Eq. (3), expressions for resonant frequencies can be derived as:

$$Z_{12}Z_{34} = 0 \quad (7)$$

Expressions for antiresonant frequencies can be derived as:

$$N^2\left(Z_{12}+Z_{34}\right)+ j2\omega C_0 Z_{12}Z_{34} = 0 \quad (8)$$

Based on Eqs. (7) and (8) it is obvious that the transducer frequency response depends on the material characteristics of its constituting parts and their geometric dimensions.

In the proposed transducer models, it is assumed that the circuit elements are ideal, i.e. they do not have losses. Losses can be included if piezoelectric constants and constants of elasticity of the transducer metal parts are in the form of complex numbers, in which the imaginary parts represent losses. These models allow only the thickness resonant modes to be predicted and, therefore, do not take into the account the inevitable radial resonant modes.

III. Simulation and Experimental Results of the Composite Transducers

By connecting acoustic impedances at the outer surfaces, the input electrical impedance, $Z_{ul}=U/I$ can be easily obtained. In order to compare results obtained by the models with experimental measurements, the modulus of transducer input impedance was determined, wherein due to the large range of impedance changes the decay function in the decibels was analyzed ($z_{ul}=20\log|Z_{ul}[\Omega]|$ [dB]). It is assumed that the surrounding environment is only air.

Table 1 shows dimensions of the individual transducer with following dimensions of the exciting piezoceramic rings Ø38/Ø13/6.35 mm, which are made of PZT4 piezoceramic equivalent material [15]. The emitter is made of dural while the reflector and the central mass are made of steel with the standard material properties.

TABLE I
Dimensions of composite transducer used in experimental analysis

Dimension [mm]	Composite transducer
$2L_1$	8
$2L_2$	58
$2L_3=2L_4$	11
$2L_5$	16
$2L_6$	21
$2a_1$	13
$2a_2$	8
$2a_3=2a_4=2a_5=2a_6$	40
$2b_3=2b_4$	9
$2b_5=2b_6$	8

Impedance/gain-phase analyzer HP4194A was used to record the experimental characteristic of input electrical impedance vs. the frequency dependency which was compared to the analogue characteristic obtained by the proposed models.

Verification of the proposed one-dimensional models was performed by comparing the modeled characteristics of input electrical impedance vs. frequency dependency with the experimental measurements for the realized unloaded ultrasonic composite transducer. Modeled dependences were obtained firstly by using the proposed one-dimensional model that does not take into the account the central bolt impact (Eq. 3), and then by applying the proposed general one-dimensional model that takes into the account all transducer parts including the central bolt with its head (see Fig. 1).

Fig. 2 shows experimental and modeled dependencies of transducer input electrical impedance in case of the transducer with dimensions given in Table 1. As it can be seen from Fig. 2, there is a great similarity between modeled and experimental dependencies. Both one-dimensional models have satisfactory results when analyzing the transducer in the case of the basic resonant mode. Measured resonant frequency of the first resonant mode is 25.62 kHz. Resonant frequency obtained by the general one-dimensional model is 25.7 kHz and the error it makes in determining this resonant frequency is 0.31%. Resonant frequency obtained by the one-dimensional model that does not take into the account the bolt impact is 25.8 kHz and the error is 0.7%. The proposed model can predict the general shape of the fourth resonant mode (with the measured resonant frequency of 52.23 kHz) but with large deviations (the detailed analysis of determination of the higher resonant modes nature is presented in literature [14]).

In the presented case from the practical aspect, when designing a transducer with a one-dimensional method, a small difference between the precision of prediction of the resonant frequencies of the basic mode is negligible. It is more important to predict all the thickness modes with the appropriate model, and especially the modes close to the operating mode that can imact the transducer behavior, which was in this case achieved by the proposed general one-dimensional model.

Based on the comparison between two characteristics modeled with the one-dimensional theory, it can be assumed that the second resonant mode, obtained by the proposed general one-dimensional model, originates from the bolt impact.

As already stated, it is possible to model the thickness resonant modes using the one-dimensional theory. In this case, the modeled resonant frequencies of the first and the second resonant modes are greater than the ones measured, while for the fourth resonant (third thickness) mode they are lower than the measured resonant frequencies. This is probably due to the presence of the modes, which are not included by the model, that is, primarily radial resonant modes and other possible types of vibration modes.

Resonant modes depend on the coupling between multiple individual modes. A detailed analysis of the individual impact of the transducer parts on the resonant modes is given in literature [14], and it was carried out with the approximatively three-dimensional matrix model [16], which predicts the thickness and the radial oscillation modes, as well as their mutual couplings. In this case, the third resonant mode, which occurs at 44.21 kHz, is a radial resonant mode in conjunction with the second and fourth resonant modes. Impact on the third resonant mode frequencies have only radial changes in the characteristics of the emitter and piezoceramic rings in the lower active layer (PZT_{34}). Coupling of the third resonant mode with the second mode arises from the fact that radial changes of the piezoceramic rings characteristics in the lower active layer (PZT_{34}) also have impact on the second mode. Coupling of the third resonant mode with the fourth mode was created based on the impact of changes of the emitter characteristics in the radial direction. In addition, the greatest impact on frequencies of the fourth resonant mode have changes in characteristics of the central mass, the emitter and all piezoceramic rings in the axial direction.

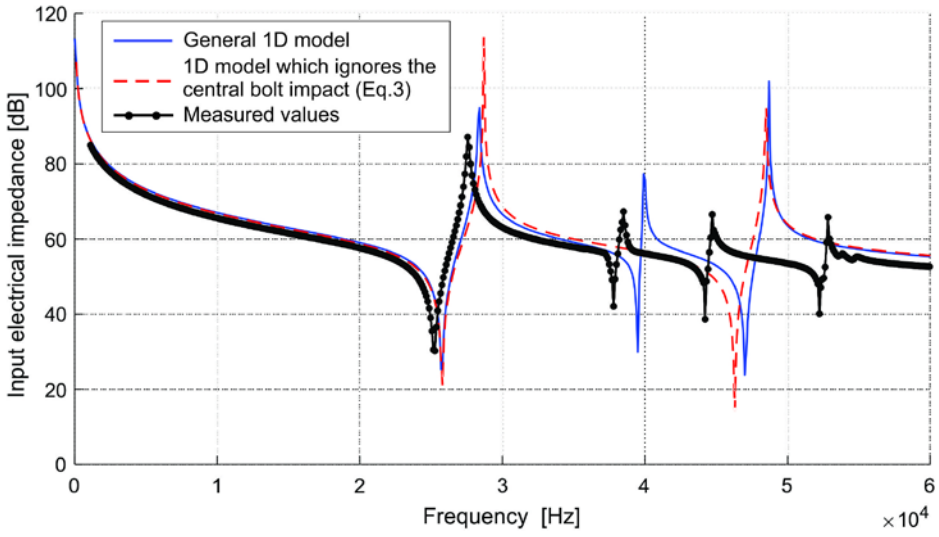


Fig. 2. Input electrical impedance vs. frequency for the realized composite transducer

Due to the strong coupling between these thickness modes with the radial mode, the proposed one-dimensional theory cannot make accurate prediction of the aforementioned resonant modes. Since the model does not take into the account mechanical and dielectric losses, which can be easily added using the characteristics of materials in the form of complex numbers, the minimal and the maximal impedance values are more pronounced on modelled dependencies compared to the measured dependencie.

IV. Conclusion

To sum up the above analysis, the following conclusions may be drawn:

- Because the structure of the composite transducer is complex, its vibrational modes are more complex compared to the traditional sandwich longitudinal transducer. Therefore, more vibrational modes were created, and the frequency characteristics becomes complex.
- Presence of the resonant modes not included by the one-dimensional models is best seen from the dependency shown in Fig. 2. The third measured resonant mode does not represent a thickness resonance, hence the modelled dependencies of the third thickness mode are at lower frequencies than the measured dependencies. Similarly, the dependency of the second thickness mode modelled with the general one-dimensional model is at higher frequencies than the measured dependency.
- The theoretical analysis of the presented research is one-dimensional. It requires the radial dimension to be much lower than its longitudinal dimension. When the radial transducer dimension increases, the radial vibration becomes intense. In this case, the three-dimensional coupled vibration should be considered. In addition, with transducers with small axial dimensions compared to radial dimensions, presence of the bolt becomes important in determining the resonant frequency.

Acknowledgement

This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia under the project TR33035.

References

[1] Prokić, M., Radmanović, M., Hedrih (Stevanović), K., "*The Change of Electrical and Mechanical Resonant Characteristics under Conditions of Various Transducers Loads*", GAMM, Dubrovnik, 1985, pp. 1-24.

[2] Mason, W.P., "*Electromechanical Transducers and Wave Filters*", Van Nostrand, Princeton, NJ, 1948.

[3] Mason, W.P., "*Piezoelectric Crystals and their Applications to Ultrasonic*", Van Nostrand-Reinhold, Princeton, NJ, 1950.

[4] Chong-Mao, L., Li-Qi, H., Chung-Fu, Y., "*Analysis of Broad-band Piezoelectric Sandwich Transducer with Perforated Structure*", Arch. Acoust., vol. 9, no. 3, 1984, pp. 349-354.

[5] Shuyu, L., "*Study on the Multifrequency Langevin Ultrasonic Transducer*", Ultrasonics, vol. 33, no. 6, 1995, pp. 445-448.

[6] Mančić, D., Radmanović, M., "*Projektovanje snažnih ultrazvučnih sendvič pretvarača*", Elektronika-ETF Banjaluka, vol. 1, no. 1, decembar 1997, pp. 66-69.

[7] Hornišová, K., Billik, P., "*Some properties of horn equation model of ultrasonic system vibration and of transfer matrix and equivalent circuit methods of its solution*", Ultrasonic, vol. 54, no. 1, 2014, pp. 330-342.

[8] Shuyu, L., Long, X., Wenxu, H., "*A new type of high power composite ultrasonic transducer*", Journal of Sound and Vibration, vol. 330, no. 7, 2011, pp. 1419-1431.

[9] Neppiras, E.A., "*The Pre-stressed Piezoelectric Sandwich Transducer*", Ultrasonic International 1973 Conference Proceedings, London, England, March 1973, pp. 295-302.

[10] Sherrit, S., Dolgin, B.P., Bar-Cohen, Y., "*Modeling of Horns for Sonic/Ultrasonic Applications*", IEEE 1999 Ultrasonics Symposium Proceedings, Caesars Tahoe, USA, Oct., 1999, pp. 647-651.

[11] Li, T., Ma, J., Low, A.F., "*Horn-Type Piezoelectric Ultrasonic Transducer: Modelling and Applications*", Advances in Piezoelectric Transducers, InTech, London, 2011.

[12] Lin, S.Y., "*Ultrasonic Transducer Principle and Design*", Science Press, Beijing, 2004.

[13] Abdullah, A., Shahini, M., Pak, A., "*An approach to design a high power piezoelectric ultrasonic transducer*", J Electroceram, vol. 22, 2009., pp369-382.

[14] Jovanović, I., Mančić, D., Jovanović, U., Prokić, M., "*A 3D model of new composite ultrasonic transducer*", Journal of Computational Electronics, vol. 16, no. 3, Sep. 2017., pp. 977-986.

[15] "*Properties of Piezoelectricity Ceramics*" Technical Publication TP-226, Morgan Electro Ceramics.

[16] Mančić, D., Stančić, G., "*New Three-dimensional Matrix Models of the Ultrasonic Sandwich Transducers*", Journal of Sandwich Structures and Materials, vol. 12, 2010, pp. 63-80.

Visualization of finite potential wells

M. Jeremić, M. Gocić, S. Trajković, and M. Milić

Abstract - The necessity to describe micro particles through probability is the most important feature of quantum theory. On the other hand, during their lectures, professors of physics face many problems related to students' understanding of quantum mechanics. They appear for any teaching style, book, school or the level of students' previous knowledge. In this article we will emphasize the need for physics teachers of quantum theory to visually illustrate some quantum theory phenomena in order to overcome understanding problems. Applications are developed in order to help students in improving and deepening the knowledge of this important subject.

An example of visualization for a rectangular potential barrier is described in this paper, for which it is necessary to calculate the probability of tunnelling, i.e. the probability that a particle can jump over the edge of the barrier if its energy is less than the potential barrier. This is one of the more complex forms of potential barriers that can have an analytical solution of the Schrödinger equation. Two versions of Computer applications were developed: PC and PocketPC application using Borland Delphi 7 and Visual Studio 2010.

Keywords - Potential wells, Schrödinger equation, Wave function.

I. INTRODUCTION

An old-fashioned tools and methods for presenting the quantum theory lectures cause many understanding problems to students that are attending them. Because of that teachers, professors and scientists have begun to investigate these problems [1], [2], [3], [4]. An interesting result is that most of the students' difficulties are universal [4], [5], and patterns of errors of quantum mechanics appear at all levels of lecturing.

In order to face this problem, we will first give a brief overview of the quantum theory basics.

Due to its wavy nature, as well as the uncertainty relations that arise out of it, it is impossible to observe and analyse microparticles (quantum particles) in the classical sense. Namely, in the case of a classical particle, the equation of motion for a given particle can completely describe its position and impulse at a given time instant. Regardless of the fact that quantum particles do not allow this, there was a need to introduce a function, which would contain wavelike properties of a micro-particle; it would be a part of an equation that can be used for calculation of the probability of finding a micro-particle, as well as its energy. The state of the system in quantum mechanics describes the wave functions $\Psi^a(x, y, z, t)$ that can be determined by solving the Schrödinger equation. However, it turned out that only the square of its module has physical meaning, and that is proportional to the probability of finding the particle W.

In general, the Schrödinger equation describes the change of the wave function over time, which depends on time and the coordinates (positions) of the particle $\Psi(r, t)$, which is shown in Eq. 1, while its form for the case of stationary state is given by Eq. 2 [6].

$$-\frac{\hbar^2}{2m}\Delta\Psi + U(x,y,z,t)\Psi = i\hbar\frac{\partial\Psi}{\partial t} \quad (1)$$

It is well known from the basic of mathematics that the change of this function over time is determined by the partial derivative of the function with respect to time as the variable. Since Ψ is function of more variables therefore we need to use its partial derivatives

$$\Delta\Psi + \frac{2m}{\hbar^2}(E-U)\Psi = 0 \quad . \quad (2)$$

The equation will be solved for a few cases that may be of interest for further applications, and allow obtaining the solution in a closed form. The solution of the differential equation is a function that, when replaced into this equation, gives an identity. It should be emphasized that for the solution of a differential equation we assume any continuous function that has defined corresponding derivatives and which translates equation into an identity. The general expression for the normalization in this case is given by the following form:

$$\int |\Phi(x)|^2 dx = 1 \quad (3)$$

The integration is done along the areas where the particle can move. First, we will observe a particle with $V(x) = 0$.

There are several forms of wave functions that are used in practice, which contain trigonometric functions sinus and cosines. This function can represent a stationary wave that moves from left to right or vice versa. Their linear combination can also be a solution:

$$\Phi(x) = C \cdot \sin k \cdot x + D \cdot \cos k \cdot x \quad (4)$$

The choice of constants depends on the boundary conditions, that is, the conditions that the wave function should fulfil for the physical reasons. Depending on the direction of motion, some of the constants can have a zero value, while the free particle can have an arbitrary value of

M. Jeremić, M. Gocić, and S. Trajković are with the Faculty of Civil Engineering and Architecture, University of Niš, Aleksandra Medvedeva 14, Niš, miljan.jeremic@gmail.com, milan.gocic@gaf.ni.ac.rs, slavisa@gaf.ni.ac.rs.

M. Milić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, Niš, miljana.milic@elfak.ni.ac.rs.

positive energy.

The normalization problem can arise if the summation is done over the entire space, and consequently an infinite value can be expected. After this, we can observe a particle in a well with an infinite potential [7]. The extreme case of the well is shown in the following Figure:

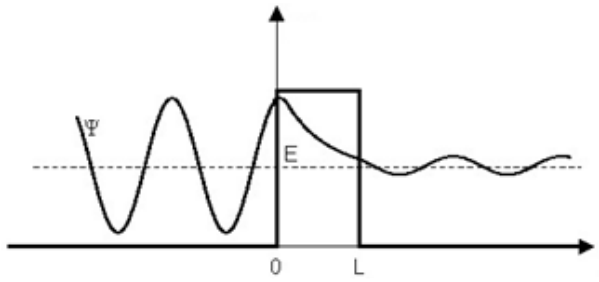


Fig. 1. Particle in a potential well

This well describes the situation when a particle cannot leave a certain space. Then, the probability of finding a particle outside the well, as well as at its borders is zero, and therefore, the wave function must also be zero [8]. Then, the wave number of a particle in an infinitely deep potential well can only have few exact values that correspond to the energy levels. The wave function can now be normalized. The calculation of these functions is illustrated in Figure 2.:

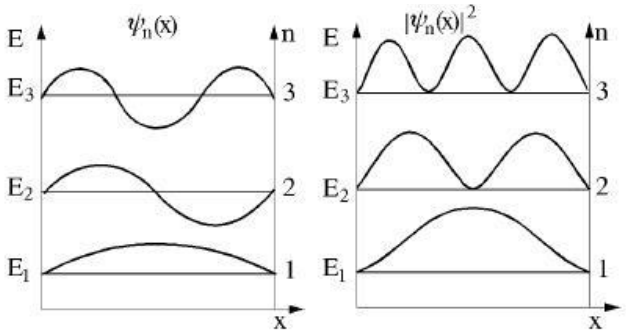


Fig. 2. Wave function

If we analyse the case of a finite potential barrier, there are two situations. In the first situation, the energy of the particle is greater than the energy of the barrier $E > V_0$. More interesting case is when the energy of the particle is smaller than the energy of the barrier. There is a finite, non-zero probability for the particle to appear in the well in Fig. 1, and it decreases with the distance from the boundary of that area.

This situation explains the tunnelling effect. On the left and right side of the barrier, the particles behave freely. As the probability is inversely proportional to the width of the barrier, it is clear that all this makes sense only if the barriers are narrow.

II. Methods and Tools

Borland Delphi 7 and Visual Studio 2010, were used as the environment for the application development [9].

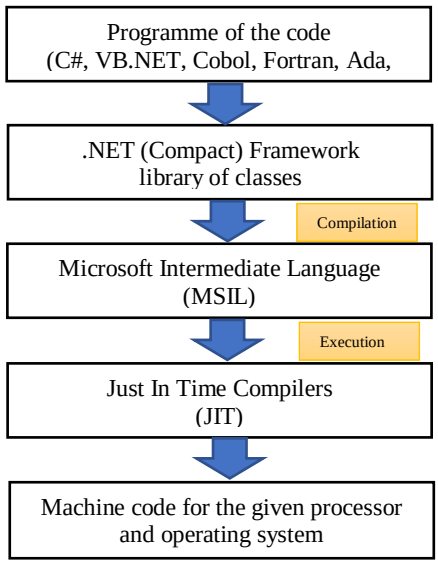


Fig. 3. Architecture of .NET application

Programming in Visual Studio is done with the support of .NET Framework. If, for example, both .NET application and .NET JIT compiler for Mac OS, are available, application developed for Windows OS would also work correctly for Mac OS. The architecture of the .NET application is shown in Fig. 5 [10]. For mobile versions of Windows OP, there are some "shorter" versions of .NET Framework, called .NET Compact Framework. It can be used for the development of PocketPC version of the application.

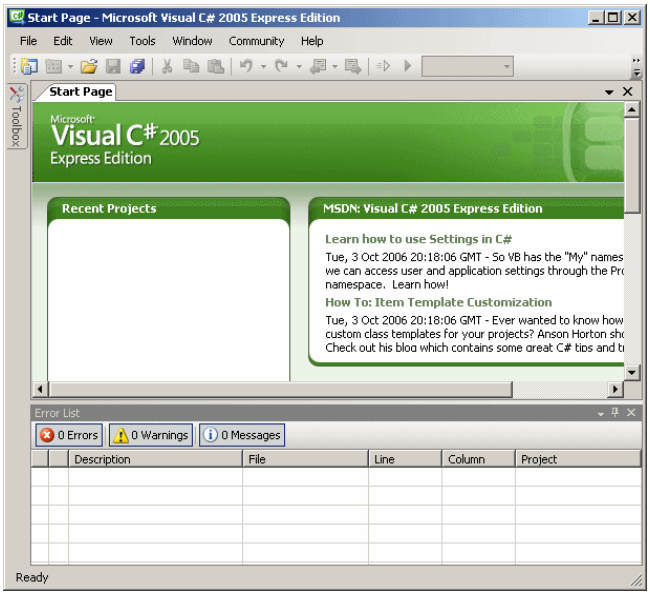


Fig. 4. Project in Visual Studio

Working in Visual Studio (for classical applications development) implies placing the objects/controls in the form/window and defining their visual characteristics and

properties during the program execution. After that, program coding, application testing and debugging need to be done.

Development of the program consisted of two parts: programming the PC application and programming the PocketPC application, using .NET Framework-a and .NET Compact Framework, respectively. Development of a PocketPC application version is illustrated in Fig. 2 [9], [11].

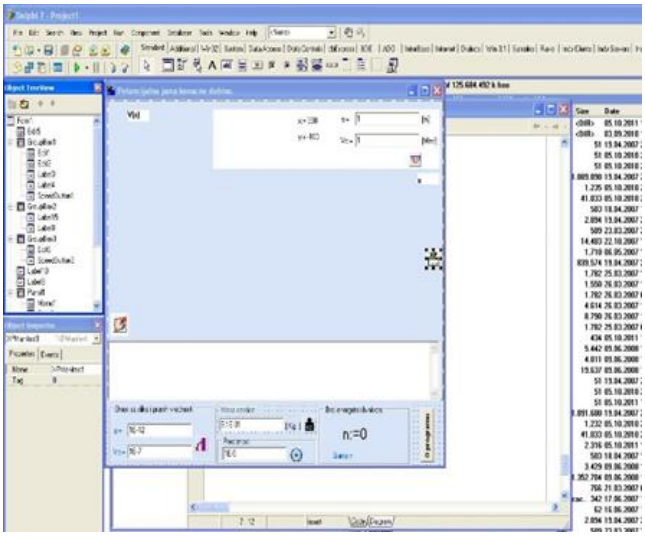


Fig. 5. Project in Delphi

For the PocketPC application it was enough to use textual INI basis for data storing due to fast and easy access. The alternative to these bases were Access and SQL bases, but their advanced functions for data manipulations were far too complex for this application. Every INI „base“ consist of Sections, while they consist of Keys that can store a certain value.

In order to verity the PocketPC application, both emulator and a real device were used. Microsoft ActiveSync was used to establish the communication with a device. This is also required during the execution of the program due to the requirements for a calendar and the synchronization between the files on the PocketPC with those on the computer. During the execution of the program, it stores a certain file in the PocketPC memory, which represents a sort of database placed in My Documents folder of the device, while ActiveSync defines, during the installation, the folder in a PC where the synchronized versions of all necessary files from My Documents folder in a PocketPC were stored. When a device is connected to the PC over the USB cable, ActiveSync compares the corresponding files from both devices, and automatically updates their versions, including the Program.ini file.

Install Creator was used to create installations for both devices. Also during the development of the application MWSnap was used for creation of the screenshots, as well as some high-performance program for image processing (like Adobe Photoshop) [10].

III. Interface and the Application of the Program

With the appearance of a new Visual Studio a new and redesigned graphical controls arrived, that resemble to those used in Office 2007 and 2010 versions. These changes are the most noticeable in menu bars, toolbars and command buttons.

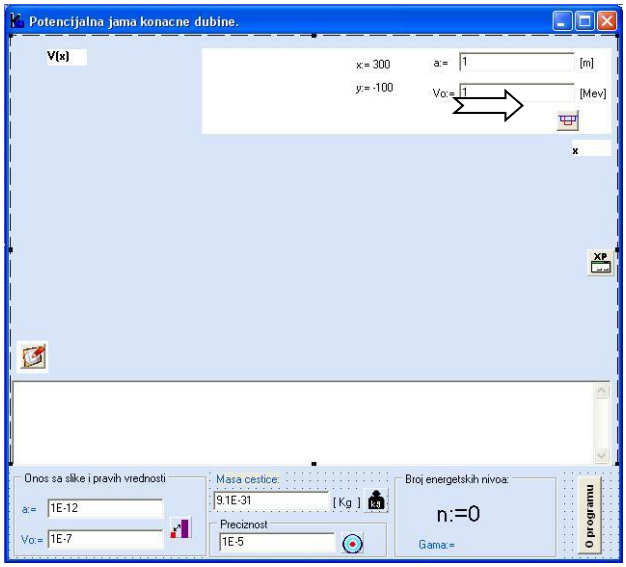


Fig. 6. Interface of the main program

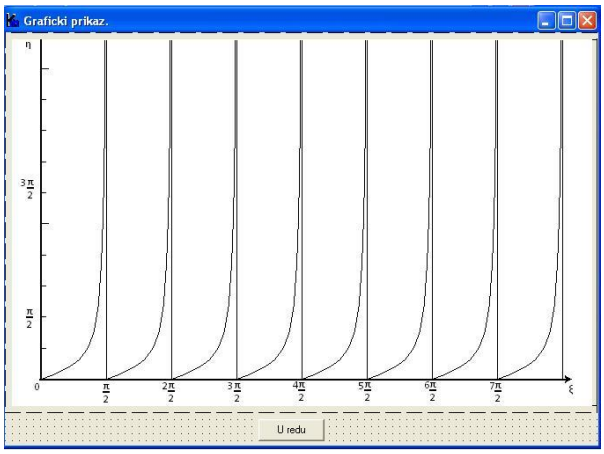


Fig. 7. Graphical illustration

The interface of the main program can be divided into several sections:

1) The main – initial line – where the program name is written;
2) Icons of the components – that are used for displaying graphics in a Dekart coordinate system, and for the visualization of the increase and the decrease of the width and depth of the well (shown in Fig. 10);

3) Toolbar – contains shortcuts toward some menu command;

For command tasters and buttons on the toolbar, tooltip function is implemented. It is a kind of help to the user of the program.

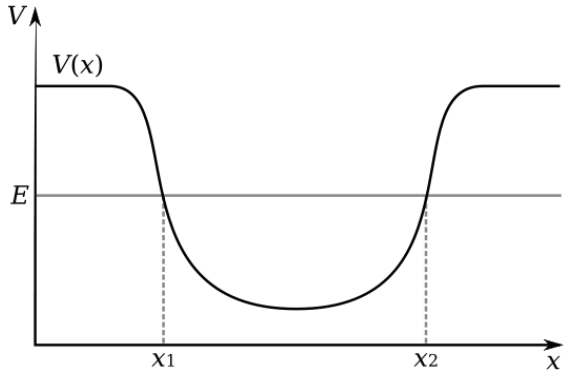


Fig. 8. The visualization of the increase and the decrease of the width and depth of the well

IV. Conclusion

According to the presented results, it is obvious that a complicated mathematical apparatus is required and used to describe phenomena in quantum-mechanics. These methods must be correctly adjusted to meet the mathematical knowledge and understanding of a typical secondary school student. Because of the students' knowledge limitation, the obtained visual results are very close to the correct value and are based on a classical theory, that are sometimes not applicable on a quantum mechanics cases. The correct mathematical calculations require additional mathematical models. In this paper we have tried to bridge the gap between the abstract quantitative formalism of quantum mechanics and the qualitative understanding necessary to clarify and predict different physical phenomena [12]. At the end we can conclude that during the physics education a special attention should be paid to cognitive issues since visualization tools can help students build qualitative representation about different and difficult quantum phenomena.

Acknowledgement

Authors would like to thank to the professors of the Knjazevac gymnasium and its director for the support in preparing the paper, and to the professors of the Faculty of Electronic Engineering at the University of Niš, for teaching us how to think, solve problems and be creative in applying new IT technologies.

References

[1] Am. J. Phys. Vol. 70, No.3, March 2002, available at http://perg.phys.ksu.edu/papers

[2] Fischler, H., Lichtfeldt, M., "*Modern physics and students' conceptions*", Int. J. Sci. Educ. Vol. 14, No. 2, 1992, pp. 181-190.

[3] Jolly, P., et al., "*Visualizing motion in potential wells*", Am. J. Phys. Vol. 66, No. 1, 1998, pp. 57-63.

[4] C. Singh, "*Student understanding of quantum mechanics*", Am. J. Phys. Vol. 69, No. 8, pp. 885-896, 2001.

[5] Styer, D., "*Common misconceptions regarding quantum mechanics*", Am. J. Phys. Vol. 64, pp. 31-34, 1996.

[6] Pejović, M., *Opšti kurs fizike-III deo,* Elektronski fakultet u Nišu, 1995. *Niš.*

[7] Raspopović, M., Kapor, D., Škrinjar, M.., *Fizika za IV razred gimnazije., ZUNS* , 1994. Beograd.

[8] Pejović, M., *Fizička elektronika-kvantna mehanika,* Elektronski fakultet u Nišu, 2005. Niš.

[9] Cantu, M., *DELPHI 2005 bez tajni, Mikro knjiga,* 2005. Beograd.

[10] McClelland D., *Photoshop 5 za Windows za neupućene,* IDG Books Worldwide i Mikro knjiga, 1998. Beograd.

[11] Janković D., *Programski jezik C#,* Elektronski fakultet u Nišu, 2010. Niš.

[12] Guangtian, Z., Singh, C., "Improving students' understanding of quantum measurement. II. Development of research-based learning tools." *Phys. Rev. ST Phys. Educ. Res.* Vol. 8, No. 1, https://www.compadre.org/Repository/document/ServeFile.cfm?ID=11910&DocID=2810 (accessed 10 January 2018).

Contribution to the electrostatics of bi-isotropic materials for the case of charged ring placed near to a sphere of a bi-isotropic material

Žaklina J. Mančić, Zlata Ž. Cvetković and Saša S. Ilić

Abstract –In this paper, calculation of electric and magnetic scalar potential of charged ring in the air, which is placed near to the bi-isotropic sphere, so that its axis coincides with the radial direction of the sphere, is performed. The problem is solved using the image theorem in the bi-isotropic spherical mirror.

Keywords: electrostatics, bi-isotropic sphere, ring of charge.

I. INTRODUCTION

The circular lineic ring appears very often as an element of various systems in electromagnetics [1], such as Helmholtz coils [2], systems for generating homogeneous fields [3], or systems for space acquisition [4]. In such systems it is necessary to determine the potential, the field of the system, and if the body is placed in such system, redistribution of the field, which occurs due to the influence of that body, should be found [5].

A. Foundations of electrostatic analysis of bi-isotropic materials

In the last decades, bi-isotropic materials have been very often encountered in various problems of electromagnetics [6-15]. Bi-isoptropic materials are described using the following constitutive relations [6,7]

$$\mathbf{D} = \varepsilon\mathbf{E} + \xi\mathbf{H}\,, \quad \mathbf{B} = \mu\mathbf{H} + \xi\mathbf{E}\,. \tag{1}$$

According to Telegen, such materials consist of elements that have permanent electrical and magnetic dipoles, parallel or antiparallel with others, so that the electric field in such material simultaneously regulates both electrical and magnetic dipoles. Similarly, the magnetic field in such material regulates both electric and magnetic dipoles at the same time.

Starting from the constituent relation (1) and Maxwell's equations[2], the Laplace's equation for the electric scalar potential is obtained:

Žaklina J. Mančić, Zlata Z. Cvetković and Saša S. Ilić are with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, Niš
E-mail:{zaklina.mancic, zlata.cvetkovic, sasa.ilic}@elfak.ni.ac.rs

$$\Delta\varphi = -\rho_s / \varepsilon_e,\ \varepsilon_e = \varepsilon(1 - \xi^2 / \varepsilon\mu), \tag{2}$$

as well Laplace's equation for the magnetic scalar potential of free electric charges in the bi-isotropic environment:

$$\Delta\varphi_m = \xi\rho_s /(\varepsilon_e\mu), \tag{3}$$

where $\varepsilon\mu \neq \xi^2$ [8]. When determining the electric field in such environments, firstly it is necessary to integrate the Poisson's equations for the electric and magnetic scalar potential, with respecting the existing boundary conditions, and then using the defining relations determine field components:

$$E = -\operatorname{grad}\varphi \text{ and } H = -\operatorname{grad}\varphi_m\,. \tag{4}$$

In this paper, a theoretical calculation of the electric and magnetic scalar potential for the case of a charged ring placed beside the bi-isotropic sphere, is presented as a contribution to the electrostatics of bi-isotropic materials. It should be noted that the axis of the ring coincides with one radial direction of the sphere.

II. The ring of charge beside the bi-isotropic sphere

Let's observe lineic ring of charge described with radius a, charged with q, which is located beside bi-isotropic sphere of radius b, placed so that the axis of the ring coincidents with the z axis of the Cartesian coordinate system, as it is shown in Fig. 1.1. The system is placed in the air/vacuum (ε_0, μ_0).

Starting from the constituent relations for bi-isotropic environments (1), outside the bi-isotropic sphere the charges produce electrical and magnetic scalar potential, that are calculated according to (2) and (3), respectively.

The solutions for the electric and magnetic scalar potential can be assumed using an expansion of the axial potential distribution into a Chebyshev series [1]:

$$\varphi=\begin{cases}\displaystyle\sum_{n=0}^{\infty}\left(A_n r^n+\frac{B_n}{r^{n+1}}\right)P_n(\cos\theta), \text{ for } r\le b\\ \displaystyle\sum_{n=0}^{\infty}\left(C_n r^n+\frac{D_n}{r^{n+1}}\right)P_n(\cos\theta), \text{for } b\le r\le\sqrt{a^2+b^2}\\ \displaystyle\sum_{n=0}^{\infty}\left(\left(C_n-\frac{qP_n(\cos\theta)}{4\pi\varepsilon_0 r_0^{n+1}}\right)r^n+\left(D_n+\frac{qP_n(\cos\theta)r_0^{n1}}{4\pi\varepsilon_0}\right),\frac{1}{r^{n+1}}\right)P_n(\cos\theta), \text{ for } \sqrt{a^2+b^2}\le r,\end{cases} \tag{5}$$

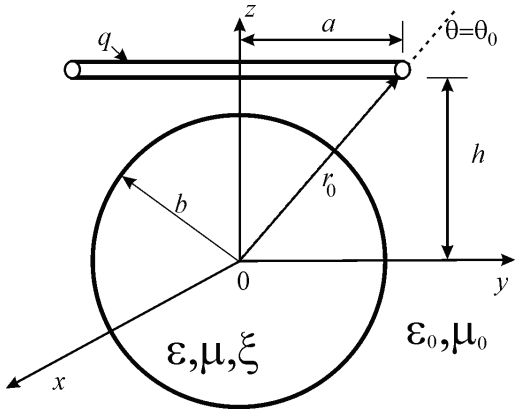


Fig. 1. Lineic ring of charge beside bi-isotropic sphere

as well as

$$\varphi_m=\begin{cases}\displaystyle\sum_{n=0}^{\infty}\left(A_{n1}r^n+\frac{B_{n1}}{r^{n+1}}\right)P_n(\cos\theta), r\le b\\ \displaystyle\sum_{n=0}^{\infty}\left(C_{n1}r^n+\frac{D_{n1}}{r^{n+1}}\right)P_n(\cos\theta), b\le r\end{cases}, \tag{6}$$

where: $A_n, B_n, C_n, D_n, A_{n_1}, B_{n_1}, C_{n1}, D_{n1}$ are the unknown constants determined using boundary conditions. Bearing in mind that the electrical and magnetic scalar potential in the center of the system ($r=0$) must have a finite value, wherease at infinitely large distances ($r\to 0$) is equal to zero, it follows that:

$$B_n=0,\ C_n=\frac{qP_n(\cos\theta)}{4\pi\varepsilon_0 r_0^{n+1}} \tag{7}$$

$$B_{n1}=0,\ C_{n1}=0. \tag{8}$$

Using the condition of the potential continuity on the boundary surface between two mediums (bi-isotropic sphere and air, $r=b$) as well as using the condition of the continuity of vectors $\mathbf{D}_n$ and $\mathbf{B}_n$ for $r=b$, a system of equations is obtained from which the other unknown constants are determined:

$$A_n b^n=C_n b^n+\frac{D_n}{b^{n+1}}, \tag{9}$$

$$D_{n1}=A_{n1}b^{2n+1}, \tag{10}$$

$$\varepsilon A_n n r^{n-1}\Big|_{r=b}+\xi A_{n1} n r^{n-1}\Big|_{r=b}=\varepsilon_0\left(C_n n r^{n-1}-(n+1)\frac{D_n}{r^{n+2}}\right), \tag{11}$$

$$\mu A_{n_1} n r^{n-1}\Big|_{r=b}+\xi A_n n r^{n-1}\Big|_{r=b}=-\mu_0\frac{n+1}{r^{n+2}}D_{n1}. \tag{12}$$

From (7) and (9) it is obtained:

$$D_n=A_n b^{2n+1}-\frac{qP_n(\cos\theta_0)}{4\pi\varepsilon_0 r_0^{n+1}}b^{2n+1}. \tag{13}$$

Next, from (11) and (12):

$$n\varepsilon A_n n b^{n-1}+\xi n A_{n1}b^{n-1}=\varepsilon_0\frac{qP_n(\cos\theta_0)}{4\pi\varepsilon_0 r_0^{n+1}}nb^{n-1}-\frac{n+1}{b^{n+2}}\varepsilon_0 D_n \tag{14}$$

and

$$\mu n A_{n1}b^{n-1}+\xi n A_n b^{n-1}=-\frac{\mu_0(n+1)}{b^{n+2}}D_{n1}. \tag{15}$$

Finally, it is obtained:

$$A_n=\frac{q(2n+1)P_n(\cos\theta_0)\left[\mu n+\mu_0(n+1)\right]}{4\pi r_0^{n+1}\left[\left[n\varepsilon+(n+1)\varepsilon_0\right]\left[\mu n+\mu_0(n+1)\right]-\xi^2 n^2\right]} \tag{16}$$

$$A_{n1}=\frac{-\xi nq(2n+1)P_n(\cos\theta_0)}{4\pi r_0^{n+1}\left[\left[n\varepsilon+(n+1)\varepsilon_0\right]\left[\mu n+\mu_0(n+1)\right]-\xi^2 n^2\right]} \tag{17}$$

$$D_n = q\frac{\left[\xi^2 n^2 + n(\varepsilon_0-\varepsilon)\left[\mu n+\mu_0(n+1)\right]\right]b^{2n+1}P_n(\cos\theta_0)}{4\pi\varepsilon_0 r_0^{\,n+1}\left[\left[n\varepsilon+(n+1)\varepsilon_0\right]\left[\mu n+\mu_0(n+1)\right]-\xi^2 n^2\right]} \quad (18)$$

$$D_{n1} = \frac{-\xi n(2n+1)qb^{2n+1}P_n(\cos\theta_0)}{4\pi r_0^{\,n+1}\left[\left[n\varepsilon+(n+1)\varepsilon_0\right]\left[\mu n+\mu_0(n+1)\right]-\xi^2 n^2\right]}. \quad (19)$$

In the absence of bi-isotropy, for $\xi = 0$, it becomes:

$$A_{n1} = D_{n1} = 0 \quad (20)$$

and the magnetic field does not exist, while constants A_n and D_n become equal to A_{n0} and D_{n0}, respectively, which are calculated in [1], in the case of a dielectric spehere (ε, μ):

$$A_{n0} = \frac{q(2n+1)P_n(\cos\theta_0)}{4\pi r_0^{\,n+1}\left[n\varepsilon+(n+1)\varepsilon_0\right]} \quad (21)$$

and

$$D_{n0} = q\frac{n(\varepsilon_0-\varepsilon)b^{2n+1}P_n(\cos\theta_0)}{4\pi\varepsilon_0 r_0^{\,n+1}\left[n\varepsilon+(n+1)\varepsilon_0\right]}. \quad (22)$$

Let's denote as:

$$\delta_1 = \left[n\varepsilon+(n+1)\varepsilon_0\right] \text{ and } \delta_2 = \left[\mu n+\mu_0(n+1)\right]. \quad (23)$$

Now, it is valid:

$$\frac{A_n}{A_{n0}} = \frac{1}{1-\frac{\xi^2 n^2}{\delta_1\delta_2}} = \frac{1}{1-\chi} \quad (24)$$

where

$$\chi = \frac{\xi^2 n^2}{\delta_1\delta_2} \quad (25)$$

and

$$\frac{D_n}{D_{n0}} = \frac{\left[\xi^2 n^2 + n(\varepsilon_0-\varepsilon)\delta_2\right]\delta_1}{n(\varepsilon_0-\varepsilon)\left[\delta_1\delta_2-\xi^2 n^2\right]} = \frac{1+\gamma}{1-\gamma}, \quad (26)$$

where is

$$\gamma = \frac{\xi^2 n}{(\varepsilon_0-\varepsilon)\delta_1\delta_2}. \quad (27)$$

The presence of bi-isotropy changes the values of the constants A_n and D_n and thus changes the distribution of the potential in the vicinity of the ring in the presence of a bi-isotropic sphere. In order to illustrate the dependence of the constants which apper in the expression of electric scalar potential from intensity of sphere bi-isotropy, the dependence of $\frac{A_n}{A_{n0}}$ and $\frac{D_n}{D_{n0}}$ upon $\frac{\xi^2}{\varepsilon_0\mu_0}$. Let's adopt: $n=1,\ \varepsilon=\varepsilon_r\varepsilon_0,\ \varepsilon_r=2,\ \mu=\mu_r\mu_0,\ \mu_r=1$.

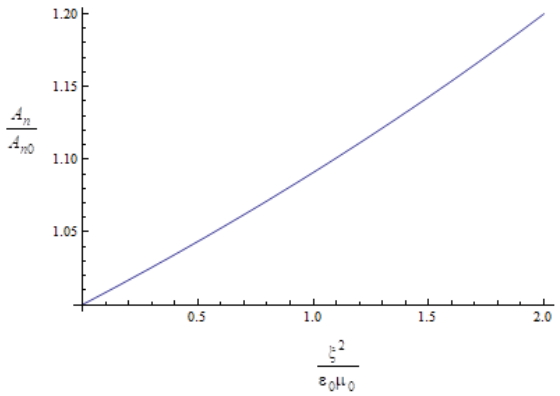


Fig. 2. Dependency $\frac{A_n}{A_{n0}}$ from $\frac{\xi^2}{\varepsilon_0\mu_0}$, for $\varepsilon = 2\varepsilon_0, \mu = \mu_0, n = 1$

In Fig.2, where is shown dependancy $\frac{A_n}{A_{n0}}$ from $\frac{\xi^2}{\varepsilon_0\mu_0}$, it can be observed increasing funciton. Thereby, it has to be fullfilled $0 \le \frac{\xi^2}{\varepsilon\mu} < 1$ [8].

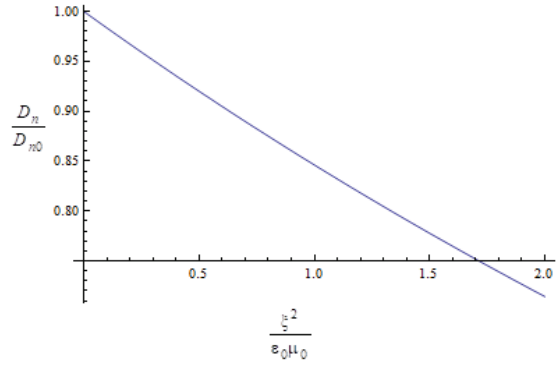


Fig. 3. Dependency $\frac{D_n}{D_{n0}}$ from $\frac{\xi^2}{\varepsilon_0\mu_0}$, for $\varepsilon = 2\varepsilon_0, \mu = \mu_0, n = 1$

From Fig. 3, where the dependency of $\frac{D_n}{D_{n0}}$ upon the parameter $\frac{\xi^2}{\varepsilon_0\mu_0}$, it can be observed that this is a decreasing function that changes the sign for high values of bi-isotropy.

IV. CONCLUSION

In the paper, we have shown the results of electric and magnetic scalar potential calculation for the case when lineic ring of charge is placed near bi-isotropic sphere. It is shown that for bi-isotropic mediums, Poisson's equation can be solved, i.e. Laplace's equation in spherical coordinate system with satisfaction of boundary condition for continuity of potential and normal component of vectors **D** and **B** on the surface of bi-isotropic sphere. Bi-isotropic materials produce magnetic field when they are located in the electric field and the influence of parameter of bi-isotropy ξ is visible in intensity of the field. It is concluded that constants in expressions for electric and magnetic scalar potential, depend on the parameter which describes bi-istropic medium, ξ . For the case $\xi = 0$, the expressions for electric scalar potential become equal to those in [1] and magnetic field does not exist.

ACKNOWLEDGEMENT

This work was supported by the Serbian Ministry of Education, Science, and Technological Development under grant TR-32052.

REFERENCES

[1] D. Veličković, *Calculation methods for electrostatic fields,* Stil, Podvis, 1982. Niš.

[2] J.A.Kong, Electromagnetic Wave Theory, John Wiley and Sons, 2-nd edition, ISBN 0-471-52214-7, 1990.

[3] Zlata Z. Cvetković, Mirjana T. Perić, Ana N. Vučković, "Systems for homogeneous electrostatic field generation using Shebyshev polynomials**",** FACTA UNIVERSITATIS Series: Automatic Control and Robotics Vol. 10, No 1, 2011, pp. 105 – 114.

[4] Z. Cvetković, *Homogeneous stacionary electric and magnetic field generation*, PhD thesis, Faculty of Electronic Engeneering of Nis, 2001.

[5] Zlata Z. Cvetković, Mirjana T. Perić, Ana N. Vučković, "An External Body in Protected Area", FACTA UNIVERSITATIS, Series: Elec. Energ. Vol. 22, no. 2, August 2009, 209-215.

[6] A. H. Sihvola, I.V. Lindell, "Bi-isotropic constitutive relations", Microwave and Optical Technology letters, Vol.4, No.8, July, 1991.

[7] W. Ren, "Bi-isotropic layered mixtures", *Progress in Electromagnetics research*, PIER, 9, pp.103-116, 1994.

[8] D. Veličković, " Wave equations and its solutions in bianisotropic media", Scientific meeting New sintered materials and their applications, Nis, pp. 40-41, 1990.

[9] I.V. Lindeli, "Quasi-static image theory for Bi-isotropic Sphere", *IEEE Trans. Antennas and Propagat*., Vol. 40, No.2, pp. 228-233, Feb.1992.

[10] I. V. Lindell, J. J. Hänninen, "Static image principle for the sphere in isotropic or bi-isotropic space", *Radio Science*, Vol. 35, Issue 3, 2000. pp: 653 – 660, DOI: 10.1029/1999RS002286.

[11] Ž. J. Mančić, S. Ilić, S. Aleksić: One type of strip lines on the surface of biisotropic cylindrical dielectric, *PES 2007*, Niš, September 3-5 2007, pp.59-60

[12] Mančić, J. Ž., Ilić, S. S., "Contribution to the Electrostatic Analysis of Bi-isotropic Mediums of Tellegen Type", *Proceedings of 13th International Conference on Telecommunication in Modern Satellite, Cable and Broadcasting Services (TELSIKS 2017)*, Niš, Serbia, October 18−20, 2017, pp. 359−362.

[13] N.Raičević, S.Ilić. "One method for chiral and metamaterial strip lines determination", *International conference in European Electromagnetics*, EUROEM 2008, Lausanne, Switzerland, p 44, 21-25 july 2008.

[14] D. Velickovic, B. Milovanovic, Z. Mancic, "New approach for equalising phase velocities in coupled microstrip lines", *Electronic letters*, Vol. 33, Issue 17, pp. 1444 -1445, 1997.

[15] D. Veličković, Ž. Mančić, "Image theorem invarement in biisotropic media", PES 96, Niš, 22-23 May, 1996, pp. 59-63.

System-Level Anomaly Detection using Hardware Performance Counters

Mark Zwolinski, Lai Leng Woo, and Basel Halak

***Abstract*—In computer-based systems, anomalous behaviour can result from physical effects, such as variations in temperature and voltage, single event effects and component degradation, as well as from various security attacks such as control hijacking, malware, reverse engineering, eavesdropping and many others. In this paper, we will present a detection technique to detect a change in the system before the system encounters a failure, by using data from Hardware Performance Counters (HPCs). We show how HPC data can be used to create an execution profile of a system based on measured events and any deviation from this profile indicates an anomaly has occurred in the system. The first step in developing a detector is to analyse the HPC data and extract features from the collected data to build a forecasting model. Anomalies are assumed to happen if the observed values fall outside given confidence intervals, which are calculated based on the forecast values and prediction confidence. A detector should provide a warning to the user if anomalies occur consecutively for a certain number of times. We evaluate our detection algorithm on benchmarks that are affected by single bit flip faults. Our initial results show that the detection algorithm is suitable for use for this kind of univariate time series data and is able to correctly identify anomalous data from normal data.**

I. Introduction

Improvements in transistor size and integrated circuit performance have allowed an increase in the number of affordable embedded sensors. With the emergence of the Internet of Things (IoT), these sensors are now being connected together in networks where huge amounts of time series data are streamed, collected and shared. The sensors used in IoT are considered inexpensive and replaceable, however, there is increasing expectation that these sensors function safely, securely and reliably. The concerns have been studied for many years. Safety in embedded systems means reducing the frequency of failures whereas reliability means ensuring the system completes the task without experiencing any failure [1]. Security in the context of an IoT application is to ensure that malicious attackers do not gain control of any of the embedded devices or systems that could lead to disastrous consequences.

Although care has been taken to ensure these systems and sensors function in a safe, secure and reliable manner, they are still exposed to various environmental conditions which may cause problems for the systems and sensors. For example, the sensors may be imperfect, a bit error may appear, or the nature of the physical processes may have some variations. Security attacks on IoT applications, such eavesdropping, control hijacking, malware and others also cause problems to IoT applications.

Mark Zwolinski, Lai Leng Woo and Basel Halak are with Department Electronics and Computer Science, University of Southampton, Southampton, United Kingdom SO17 1BJ
Email: mz@ecs.soton.ac.uk

The impact of these problems is anomalous behaviour in the system, which could lead to device failure. Very often, users are aware of the anomalous behaviour only after a failure has occurred. One practical approach is to detect anomalies from streaming real-time data. Here, we have used *Hardware Performance Counters* (HPCs) to monitor the behaviour of a system. HPCs are sets of special-purpose counters built into processors to record events precisely and accurately in real-time. A system that behaves normally (no error is detected in the system) exhibits a particular profile, and any deviations from this profile indicate an anomaly in the system. The research on anomaly detection in real-time streaming data is not something new, however, but we have yet to find research attempting to detect a change in the behaviour of the system using HPCs. This paper is the first attempt that focuses on early detection of anomalies (deviation from the normal patterns in the system) by utilising the real-time streaming HPCs that is available in the processor itself, and thus, no modification is required to the physical system. By creating a system that has some self-awareness capability and that is able to provide a warning to the user before a failure occurs, we aim to minimise or even avoid potential risk to the user. Overall, the main contributions of our work are as follows:

- We develop an algorithm for early detection of system-level anomalous behaviour using HPCs;
- We explore several anomaly detection methods in a case study;
- We develop a new attribute called the detection time that evaluates the effectiveness of the early detection algorithm; and
- Our results show that the algorithm can be used for early detection of system-level anomalous behaviour.

This paper is organised as follows. Section II looks at anomaly detection in the context of real-time time-series data. Our experiments with HPCs are presented in Section III. Our proposed detection algorithm and experiment based on hardware performance counter are presented in Section IV. In Section V, we discuss the data we obtained from our experiment. Finally, in Section VI, we conclude the paper and make suggestions for future research.

II. Anomaly Detection

Anomalous behaviour, or in short, anomalies, is behaviour that does not conform to a normal, expected pattern and can also be identified as outliers, exceptions, peculiarities, contaminants or other terms according to the domain [2] and anomaly detection refers to finding patterns in data that do not conform to expected behaviour [3]. The science of detecting anomalies is typically applied in applications like fraud detection in credit card applications, loan facilities applications, state benefits, fraudulent usage of credit cards and mobile telecommunication [2], network intrusion detection [4], network performance detection [5] as well as activity monitoring [6].

Anomaly detection is not an easy problem to solve due to various factors such as the nature of the data itself, the availability of labelled data, the types of anomalies to be detected, the application domain and many more. There are numerous existing anomaly detection techniques such as classification-based, clustering-based, nearest-neighbour-based, statistical, information theoretic and spectral [3], however, most of these techniques are often used for detecting anomalies in batches of data and are unsuitable for real-time streaming applications. Techniques that require data labelling such as supervised learning are also not suitable for real-time anomaly detection.

Most of the anomaly detection methods used in real-time streaming time series data are statistical techniques that are computationally lightweight, as one of the main requirements is the ability of the algorithm to learn continuously without storing the whole stream of data. These techniques include sliding windows [7], [8], ARIMA [9], Exponential Smoothing [10], Hierarchical Temporal Memory (HTM) [11] and change detection [12]. For real-time streaming time series data, early detection of anomalies is valuable as it allows actions to be taken, possibly even preventing system failure [11]. As evaluating all the methods is beyond the scope of this paper, we will focus on applying sliding window, Exponential Smoothing and ARIMA techniques to detect anomalies in the streaming data from HPCs.

III. Methodology and Experiment

Similar to what was done elsewhere, [13], [14], we use HPCs to create fault-free execution profiles for several different type of benchmarks.

There are several different fault models used in digital circuits. In our experiment, we focus on the single stuck-fault (SSF) model because this fault model is applicable to many different physical fault regardless of whatever technology is applied.

We have chosen to use the following counters to profile the executions: the number of committed instructions (instructions that were executed); number of function calls; number of integer instructions; and number of load instructions.

Once we have identified the fault model and specific hardware counters for monitoring, we created fault-free executables from various benchmarks and performed the simulation using Gem5 and GemFI, described below. We obtain initial execution profiles using the counters. The next step is to inject faults and observe the anomalies recorded using the counters.

This experiment was conducted on a Microsoft Azure virtual machine, [15]. We created a Linux virtual machine with 16 central processing units (CPUs), 32 GBs of memory and 1TBs of data storage. We used Ubuntu version 14.04 for compatibility with Gem5 and GemFI.

A. Architectural Simulator

Gem5 [16] simulator is an instruction set simulator, widely used in computer architecture research. It supports various Instruction Set Architectures (ISAs) such as X86, ARM, Alpha, Sparc, Mips and Power. Gem5 can operate in two modes: *System Call Emulation (SE)* and *Full System (FS)*. SE mode allows users to emulate most common system calls, thus avoiding the need to model devices or even an operating system (OS). In FS mode, Gem5 models complete system including the OS and devices, executing both user-level and kernel-level instructions.

GemFI [17], is a cycle accurate fault injection tool developed based on Gem5 with the primary objective of enabling fault injection. GemFI supports the Alpha and Intel X86 ISAs. There are two intrinsic functions provided by GemFI API:

- **void FI_init()** initialises the fault injection module.
- **void FI_activate (int id, int command)** is a pseudo-assembly instruction to toggle a fault on a specific thread. The thread is given a numerical identification number.

A set of faults is generated using the fault generator that comes together with GemFI. Each fault contains four attributes: *Location*; *Thread*; *Time*; and *Behaviour*.

B. Benchmarks

The benchmarks used in this experiment are from MiBench [18], which is a set of 35 embedded applications divided into six suites with each suite targeting a specific area of the embedded market. We have chosen the *basicmath*, *bitcount* and *qsort* benchmarks from the Automotive and Industrial Control suite, as well as *Dijkstra* from the Network suite. The basicmath program performs simple mathematical functions. The bitcount algorithm tests the bit manipulation ability of a processor by counting the number of bits in an array of integers and *qsort* uses the popular qsort algorithm to sort a large array of strings into ascending order. *Dijkstra* calculates the shortest path between every pair of nodes in a graph.

C. Experimental Setup

To extract the HPCs features that will be used to monitor system reliability, there are several steps:

1) Set up the benchmarks required for testing.
   Each benchmark was compiled dynamically in two versions – one in the original form and another with GemFI intrinsic functions added. Both versions were compiled for the X86 ISA. For *basicmath*, no input data was required, whereas for *bitcount*, the input data is an array of integers and for *qsort*, the input data contains a

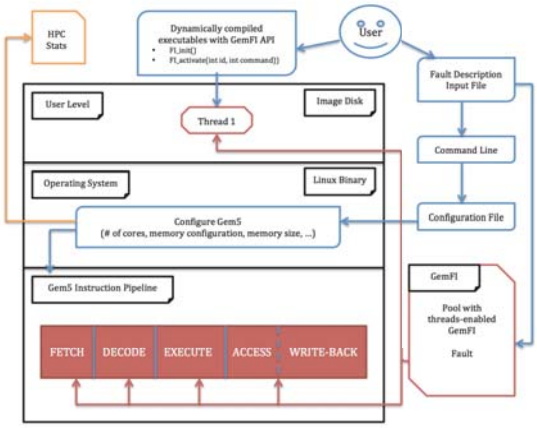


Fig. 1. Overview of the GemFI API, after [17]. The red components show possible fault injection locations; the red octagon is where the executables run.

list of words. The input data for *Dijkstra* is a large graph in the form of an adjacency matrix. The executable files are then placed in the disk image serving as the virtual disk for GemFI.

2) Perform the simulation.
Simulation of the benchmarks was performed in both the Gem5 and GemFI simulators in FS mode. FS mode simulates the execution of the benchmarks in an OS-based simulation environment. A script file is created to assist in the execution of the benchmarks. After fault injection has been initialised and enabled, a set of faults is then created using the fault generator in GemFI. A fault configuration file describing the fault to be injected is provided for GemFI. This file is parsed at startup and each fault is injected into one of the four internal queues, which correspond to a pipeline stage. The simulation continues as normal until it is time for the fault to be injected. Figure 1 provides a general overview of how the simulation works using GemFI API. The blue lines indicates that the tasks belong to the user, the red lines indicate the responsibility of GemFI, and the orange line denotes the HPC values as outputs from the OS.
Each experiment was executed six times: (i) initial run; (ii) with fault activation; (iii) fault injected in the Fetch pipeline; (iv) fault injected in the Decode pipeline; (v) fault injected in the Execute pipeline; and (vi) fault injected in the Load/Store pipeline. The fault model applied in this experiment is a stuck-at-1 fault model, and it is applied at every level in the pipeline.

3) Trace and record the required HPC values.
Two different tracing methods were tried to log the HPCs values obtained. The first method was to obtain the HPCs after the operating system (OS) has booted and another set of HPCs at the end of the execution of the benchmark. However, this method can only provide an indication that an error has occurred which causes the program to either hang, crash or provide incorrect output, but is unable to determine when the fault occurred. The second method was to log the HPCs values at certain intervals. Using this method, we can demonstrate that we are able to create an execution profile for each benchmark, and to detect the instance when an error has occurred. We found that tracing the HPC data at time intervals of 10ms (which is equivalent to 200,000 cycles) is sufficient to create a profile for each benchmark. The HPC data presented below are only for the benchmarks and do not include the OS.

TABLE I
TOTAL INSTRUCTIONS FOR EACH BENCHMARK

Total Instructions	Benchmarks				
	Basicmath	Bitcount	QSort	QSort (2)	Dijkstra
GemFi	590984224	40508678	83324366	82063568	52370245
GemFI w/ Fault Activated	590989240	40511222	83339395	82065794	52373445
Injected Fault - Fetch	13718386802	27945500	59913002	421330792	50701576
Injected Fault - Decode	590989240	40511222	83244369	82065794	52373445
Injected Fault - Execute	586196426	20397641	21025988	42442442	302299133
Injected Fault - Memory	1410304814	40511222	83339395	82065792	52373446

D. Results and Analysis

Figure 2 compares the execution profiles obtained from four different benchmarks using the number of committed instructions (axis Y) plotted against the time interval (axis X). By inspection, the profiles for each benchmark differ from one another, which suggests that HPCs can be used to identify the normal behaviour of the system. Profiles using the number of integer instructions, the number of function calls as well as the number of load instructions were plotted as well (but not displayed here due to space) and these profiles also show distinct differences, suggesting that it is sufficient to monitor the reliability of the system based on one or two counters.

In our experiments, as we injected a stuck-at-1 fault in every pipeline, we discovered that this stuck-at-1 (or 0) fault led to errors such as *segmentation faults*, *invalid opcodes*, *kernel panics* and *invalid instruction pointers*. These errors will then cause the program to either *crash* or *hang*. Table I shows the total instructions executed for each benchmark. In particular, "GemFI" and "GemFI w/ Fault Activated" represent the baseline data for the program being executed successfully. The value in "GemFI w/ Fault Activated" will always be slightly higher compared to "GemFI" due to the additional intrinsic functions that are added. Table I also displays the total instructions for *QSort* and *QSort2* with two different sets of input data.

Table I lists the total instructions executed after a fault is injected in every pipeline. Values that differ from the baseline "GemFI w/ Fault Activated" number indicate that an anomaly has occurred in the program. A value that is below the baseline indicates that the program terminates early, whereas a value that exceeds the baseline indicates that the program hangs. As we discussed earlier, obtaining the data at the end of the execution is only able to tell us whether the program has terminated successfully but cannot determine when an error has occurred. For example, consider the total instructions

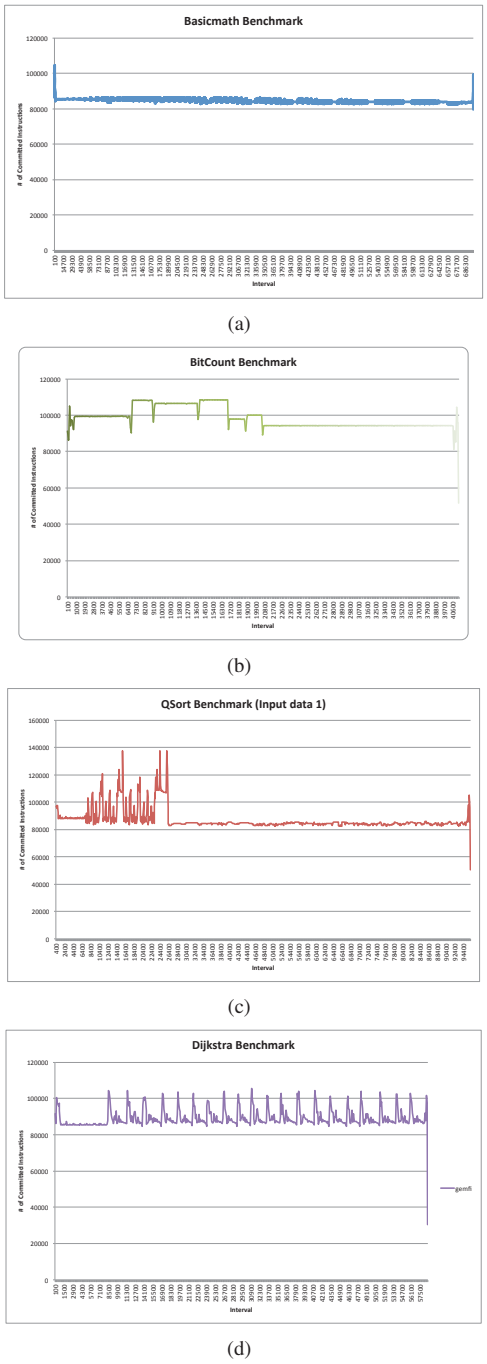


Fig. 2. Execution profiles based on the number of committed instructions for different benchmarks - (a) Basicmath, (b) Bitcount, (c) QSort and (d) Dijkstra

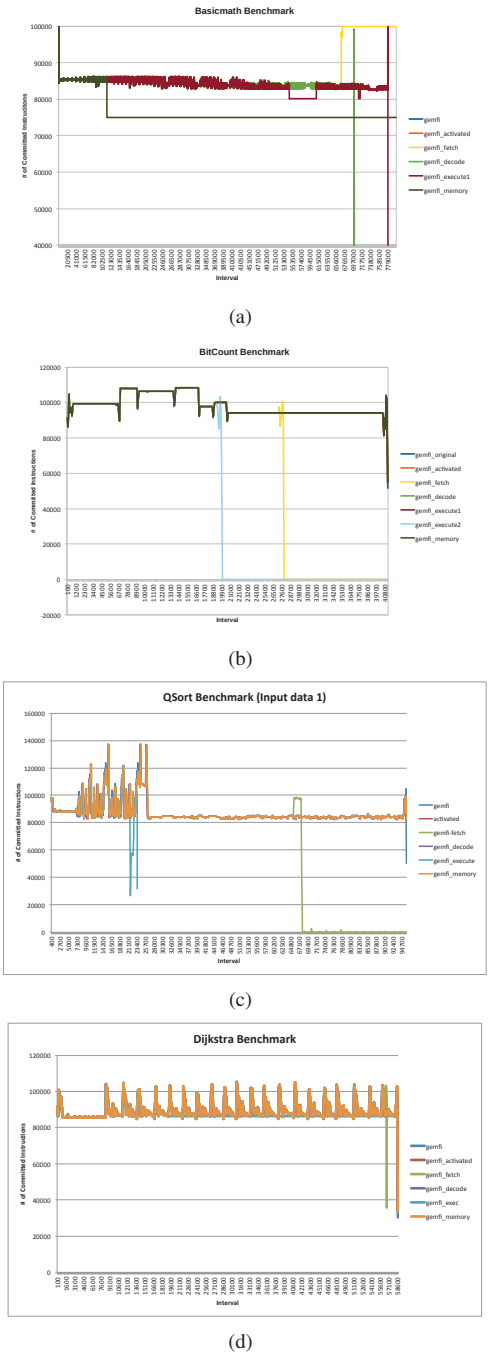


Fig. 3. Execution profiles that shows how a failure that occurred can be detected using only counters for different benchmarks - (a) Basicmath, (b) Bitcount, (c) QSort and (d) Dijkstra

recorded for the *basicmath* benchmark when a fault is injected in the Execute pipeline. If we compare the value of 586196426 against the baseline 590989240, the initial conclusion would be the program had not terminated successfully. However, if we perform instructions tracing in a time interval, we can see that an error has occurred but that the program did terminate successfully. This is illustrated better in Figure 3 (a) where the profile shown by the maroon line indicates a drop at interval 553500 to value 80000 compared to the expected value between 84000 and 85000. The value 80000 stays for a number of cycles until the program managed to recover from the fault and resumed the execution at interval 615000 until completion. We have shown that by tracing the data at intervals and creating a profile using the HPCs, it is possible to capture even a slight change in the program.

In Figure 3 (a) for the Basicmath benchmark, the profile was able to capture the error occurring in the *fetch* pipeline (yellow line) and the *memory* pipeline (dark green line). As noted, it is also possible to detect an error that occurs due to a fault, but is recovered, as shown by the maroon line in the same figure. This figure shows how an error can be visible if we performed data tracing. Other benchmarks, Figure 3 (b), (c) and (d), also show that errors are detectable using the counters. In 3 (b), the HPCs were able to detect a system hang in the *fetch* pipeline (yellow line) and *execute* pipeline (blue line), whereas in 3 (c), the *execute* pipeline (blue line) shows that the program has crashed (hence the line stopped in the midst of execution), and the *fetch* pipeline (green line) shows the program hangs.

These findings provide justification for our hypothesis that tracing HPCs in an interval can be used for anomaly detection in embedded systems.

IV. Early Detection Algorithm

Using GemFI, a single bit-flip fault model was injected into the Dijkstra benchmark. A single hardware counter, the *number of committed instructions*, was sampled at every 0.1μs and gathered at an equally spaced time interval of 100μs; this data is used to monitor the system's behaviour. We then developed an early detection method with a univariate type of time-series data.

The HPC data is a univariate type of time series data that can be represented as $X = \{x(t)|1 \leq t \leq m\}$ where $x\{t\}$ is a vector of continuous streaming data input at time t and can be represented as $x\{t\} = x_1, x_2, x_3, ..., x_t$. Briefly, in our algorithm, we propose early detection of system-level anomalous behaviour using a sliding window where q is the number of hardware counters used to predict the next sequential data x_{t+1}. Data will be classified as anomalous if it falls outside the upper and lower ranges of the predicted value calculated using D^t as the confidence interval, p. If five anomalies are detected consecutively, the system sends a warning to the user, else, the sliding window moves a step forward by removing x_{t-q+1} from the back of the window and adding the current data x_{t+1} to the front of the window to create D^{t+1}. The detail of the algorithm is explained further in following subsections.

A. One-Step-Ahead Prediction

The first step of this algorithm is to determine the size of the sliding window, q. Generally, there are two types of data used to detect anomalies – historical data and real-time data. The difference between using historical data compared to real-time data is that historical data uses previous and subsequent data in the window as input parameters to determine if the current data is anomalous, whereas real-time data uses only the previous data in the window to determine if the next data is anomalous. When the system experiences some anomalies, it is observed that the counter data starts to deviate from the point at which the fault was manifested. Based on this observation, the sliding window, q, will use previous data to predict the next data and this can be written as:

$$D^t = \{x_{t-q+1}, x_{t-q+2}, x_{t-q+3}, ..., x_{t-q+q}\} \quad (1)$$

Once the size of the sliding window has been determined, a univariate autoregressive model of the hardware counter data is used to predict the next counter value using the previous data as input. According to [7], [8], univariate autoregressive models are models that are used to predict future data in a sensor data stream by using specified set of previous measurements from the same sensor. This model is suitable for use in this experiment as it uses only one variable and data is being sampled at the same frequency. D^t is used as the input into the autoregressive model, M, to predict the next data which can be written as:

$$\overline{x}_{t+1} = M(D^t) \quad (2)$$

This work compares three methods for creating a one-step-ahead prediction model, namely Single Exponential Smoothing (SES), Single-Layer Linear Network Predictor (LN) and Autoregressive Moving Average (ARMA).

1) Single Exponential Smoothing (SES): SES is a type of exponential smoothing that weighs past observations with exponentially decreasing weights to forecast future values. The predicted value, $\overline{x}_{t+1}$ is calculated using:

$$\overline{x}_{t+1} = \alpha x_t + (1-\alpha)(D^t) \quad \text{where} \quad 0 < \alpha \leq 1 \quad (3)$$

The constant parameter α is a smoothing constant, used to smooth or damp older observations. The value of α was determined using a trial-and-error approach and the α value that provides the model with the best performance was selected. SES requires at least 3 previous data points for initialisation before a prediction of the next data can be made.

2) Single-Layer Linear Network Predictor (LN): The LN method predicts the next data $\overline{x}_{t+1}$ as a linear combination of the q previous data points using the following equation:

$$\overline{x}_{t+1} = \frac{\sum_{i=0}^{q-1} w_i x_{t-i}}{\sum_{i=0}^{q-1} w_i} \quad (4)$$

where b and $w_0, w_1, ..., w_{q-1}$ is a set of constant weights used to predict the next expected data based on the sliding window D^t. For simplicity, we have assigned the weight vectors to $1, 2, ..., q$ with the weight vector is inversely proportional to the distance of the points in the sliding window, that is, the further the point x_t from $\overline{x}_{t+1}$, the smaller the weight vector will be.

3) Autoregressive Moving Average (ARMA): The ARMA method (also known as the Box-Jenkins method) consists of an autoregressive (AR) part and a moving average (MA) part where the AR part regresses the variable on its past values, while the MA part models the error term. There is no *Integrated* term, as the time series data is found to be stationary based on the *Augmented Dickey-Fuller (ADF)* test. Therefore, the ARMA (p, q) model is used, defined as follows:

$$\overline{x}_{t+1} = c + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - ... - \theta_q \epsilon_{t-q} \quad (5)$$

In our work, we tested the value (p, q) between 0 and 5 in increments of 1, based on an autocorrelation plot (ACF) and a partial autocorrelation plot (PACF); the (p, q) values that provides the model with the best performance are selected.

B. Anomaly Classification

Once the next data has been predicted using either the SES, LN or ARMA prediction methods, the next step is to determine the upper and lower bounds of the predicted value using the confidence interval, p, where the upper and lower bounds determine the acceptable range of values the future hardware counter data can take. We assumed the models' residuals have zero-mean normal distributions and the standard deviation is calculated based on the window size. Therefore, the range of acceptable values is calculated using the following equation:

$$\overline{x}_{t+1} \pm (t_{\alpha/2,q-1} * \frac{\sqrt{\sum(x_t - \bar{x_t})^2}}{q-1}) \quad (6)$$

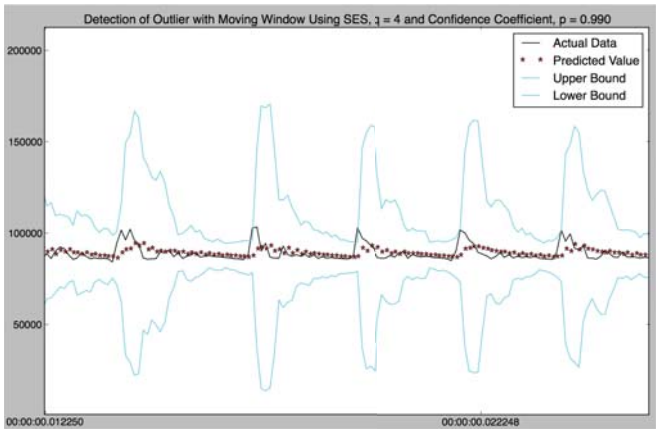
where $\overline{x}_{t+1}$ is the predicted data, $t_{\alpha/2}$ is pth percentile of the t-distribution with $q-1$ degrees of freedom, $\sqrt{\sum(x_t - \bar{x_t})^2}$ is the standard deviation of the model residual and q is the size of the sliding window. If the next actual hardware counter data falls within the range, the data is considered non-anomalous, else if it falls outside the range, the actual data is marked anomalous. As we had observed, the hardware counter data does not have a fixed value, therefore, using a confidence interval is more relevant and beneficial compared to using some random threshold as confidence interval maintain the width of the upper and lower range according to the actual data.

C. Early Detection

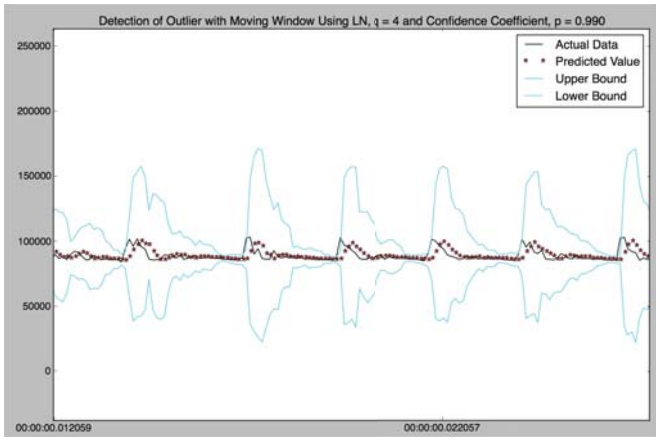
The main objective of the proposed detector is to be able to detect the anomalous behaviour as quickly as possible but at the same time, to avoid the detector being overly sensitive. From our observation, when the system experiences a failure, the hardware counter data begins to deviate from its normal pattern. The number of anomalies detected consecutively is determined using a trial-and-error approach, where the aim is to choose the value that gives the quickest detection time.

V. Discussion and Analysis

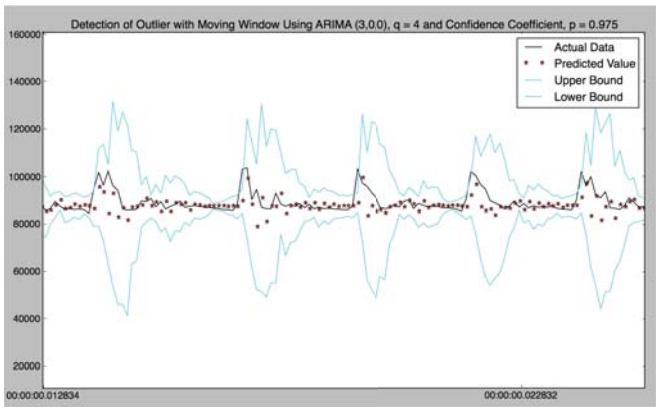
Fig 4 shows the result of one-step-ahead prediction using the three different methods outlined earlier. The upper and lower ranges calculated from Equation 6 are also shown in the same figures. It can be seen that the actual hardware counter data lies very near to the predicted data in all three methods. However, the SES method has provided a rather smoother predicted value, hence there is a bigger residual difference between actual data and predicted data compared to the LN and ARMA methods. Another observation is that both the SES and LN methods have also produced a bigger range of acceptable values compared to the ARMA method. However, as we will show in our analysis, a bigger range of acceptable values will make it harder to detect anomalies.



(a)



(b)



(c)

Fig. 4. One-step-ahead prediction methods - (a) SES Method, (b) LN Method, and (c) ARMA Method

In order to evaluate the effectiveness of the early detection algorithm, we look at how well the detector has performed in classifying the anomalies using the sensitivity, specificity and accuracy statistical attributes. Sensitivity (also known as true positive rate) evaluates how well the detection algorithm correctly identifies the anomalies and specificity (also known as true negative rate), measures how well the detection algorithm

correctly identifies the non-anomalies. Accuracy measures how well the detection algorithm detect both anomalies and non-anomalies. True Positive (TP) and True Negative (TN) are the ideal situation where data points are detected and identified correctly, while False Positive (FP) and False Negative (FN) are undesirable cases which are impossible to eliminate but need to be kept to a minimum. The formulae to calculate sensitivity, specificity and accuracy are defined as follows:

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{7}$$

$$Specificity = \frac{TN}{(TN + FP)} \tag{8}$$

$$Accuracy = \frac{TP + TN}{(TP + FN + TN + FP)} \tag{9}$$

Another attribute that is important in evaluating our detection algorithm is detection time. This is an important attribute as it will determine which method is quickest in identifying the anomalous behaviour in the system. We have developed a formula to calculate detection time as shown below:

$$\text{Detection time} = (TP + FN) * \text{Logging Interval} \tag{10}$$

The goal of the detector is to detect the anomalous behaviour at the earliest time, therefore, the key attribute is the lowest detection time that can be attained from the proposed methods. A low (or early) detection time means low values of TP and FN are required. Lower values of TP and FN means a smaller value of sensitivity will be attained. The window size, k, and confidence interval, p, are the two parameters that are optimised in order to improve the detection algorithm. The parameter k determines how many previous points are used in the calculation to predict $\overline{x}_{t+1}$ and the value k takes in from 4 to 12 in increments of one. Parameter p calculates the acceptable range for a data point to be classify as non-anomalous and the value p varies from 85% to 99% with the increment of 5%. Larger value of p means bigger range of values are acceptable.

Table II shows the statistical analysis of using the ARMA method for predicting the next data with the table showing the top three results where detection time, TP, FN and FP are the lowest. The logging interval is set at 100 μs, and the earliest detection time was found to be 1700 μs. The result shows that the value of specificity where the detector was able to detect and identify non-anomalous points is at least 95% and above in most cases (except where (k, p) is $(4, 90\%)$, the specificity recorded was 92.6%). The best performance of the detector is achieved with $k = 4$ and $p = 99\%$ where actual data points that anomalous, TP is 2 and data points that are normal and identified as normal, TN is 264. The accuracy achieved with this setup was 93.0%.

Table III shows the statistical results for all three methods. For simplicity, only the top three results from each method are presented. From the results, it is clear that all three methods shows high accuracy of at least 89% achieved using LN method and highest being 93.0% which was achieved using the ARMA method. The ARMA method has also been shown to be superior compared to SES and LN, where the anomalous behaviour was detected earliest at 1700μs compared to 1900μs and 2700μs using SES and LN, respectively. The specificity attribute using ARMA method is 93%, which is the highest of all. This means that the one-step-ahead prediction model using ARMA was more accurate as it is able to predict values that lie very close to the actual values.

VI. Conclusion

In this study, we have developed an early detection algorithm that detects anomalies by detecting a change in the hardware performance counter data. This detection algorithm is performed on a univariate time series data using a one-step-ahead prediction applied to a sliding window, k, and the range of acceptable data is calculated from the predicted data using a confidence interval, p. The first step is to slice the data into small windows before applying a prediction model to predict the next counter data. Anomalies are assumed to have occurred if the value falls outside the acceptable range. The algorithm goes a step further to detect if there are 5 anomalies that have occurred consecutively.

The detection algorithm was tested on the Dijkstra benchmark that was affected with single bit flip fault. From the results, we conclude that the early detection algorithm developed in this study is useful in early detection of system-level anomalous behaviour. The detection algorithm is a simple algorithm which is suitable for application in IoT devices as it only requires the time series data of the hardware performance counter and does not require any initial classification of the data. In summary:

- We have developed an algorithm for early detection of system-level anomalous behaviour using hardware performance counter data;
- We explored several methods for one-step-ahead prediction which can be applied in our case study;
- We also developed a new attribute called the detection time that evaluates the effectiveness of the early detection algorithm; and
- Our results show that the algorithm can be used for early detection of system-level anomalous behaviour.

Our study will be further expanded by looking at other plausible techniques which can optimise our detection algorithm to reduce the detection time. One of the ways to reduce the detection time is by minimising the false negatives where the detector identified anomalous data points as normal. We also plan to develop and test the detector in a real IoT device and evaluate the detector in terms of its complexity, computational power, area and cost.

VII. Acknowledgement

This work has been partly supported by Microsoft Azure Research Award number CRM: 0518905.

TABLE II
STATISTICAL ANALYSIS OF ARMA METHOD WITH VARIOUS PARAMETER SETUP

Parameters (k, p)	**(4, 90%)**	**(4, 95%)**	**(4, 99%)**	**(5, 95%)**	**(5, 99%)**	**(6, 95%)**	**(6, 99%)**
TP	3	3	2	3	2	3	2
FN	14	14	15	14	15	14	15
FP	20	11	5	11	8	14	9
TN	249	258	264	258	261	255	260
Sensitivity	0.176	0.176	0.118	0.176	0.118	0.176	0.118
Specificity	0.926	0.959	0.981	0.959	0.970	0.948	0.967
Accuracy	88.11%	91.25%	93.00%	91.26%	91.96%	90.21%	91.61%
Detection Time (μs)	1700	1700	1700	1700	1700	1700	1700

TABLE III
STATISTICAL ANALYSIS USING ARMA, SES AND LN METHODS

	ARMA			**SES (α = 0.3)**			**LN**		
Parameters (k, p)	**(4, 99%)**	**(5, 99%)**	**(6, 99%)**	**(4, 99%)**	**(5, 99%)**	**(6, 99%)**	**(4, 99%)**	**(5, 99%)**	**(12, 99%)**
TP	2	2	2	4	4	4	9	9	9
FN	15	15	15	15	15	15	18	18	18
FP	5	8	9	6	8	9	12	14	13
TN	264	261	260	263	261	260	257	255	256
Sensitivity	0.118	0.118	0.118	0.211	0.211	0.211	0.333	0.333	0.333
Specificity	0.981	0.970	0.967	0.978	0.970	0.967	0.955	0.948	0.952
Accuracy	93.00%	91.96%	91.61%	92.71%	92.01%	91.67%	89.86%	89.19%	89.53%
Detection Time (μs)	1700	1700	1700	1900	1900	1900	2700	2700	2700

REFERENCES

[1] P. Koopman, "Reliability, safety and security in everyday embedded systems," *Dependable Computing, Lecture Notes in Computer Science*, vol. 4746/2007, 2007.

[2] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artificial Intelligence Review*, vol. 22, no. 2, pp. 85 – 126, 2004.

[3] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computer Survey*, vol. 41, no. 3, pp. 15: 1 –15: 58, July 2009.

[4] J. Lin, Q. Zhang, H. Bannazadeh, and A. Leon-Garcia, "Automated anomaly detection and root cause analysis in virtualized cloud infrastructures," in *NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium*, April 2016, pp. 550–556.

[5] Y. Zhang, S. Debroy, and P. Calyam, "Network-wide anomaly event detection and diagnosis with perfsonar," *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, pp. 666–680, Sept 2016.

[6] P. Fiadino, A. D'Alconzo, M. Schiavone, and P. Casas, "Towards automatic detection and diagnosis of internet service anomalies via dns traffic analysis," in *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*, Aug 2015, pp. 373–378.

[7] D. J. Hill and B. S. Minsker, "Anomaly detection in streaming environmental sensor data: A data-driven modeling approach," *Environmental Modeling and Software*, vol. 25, no. 9, pp. 1014–1022, September 2010. [Online]. Available: http://dx.doi.org/10.1016/j.envsoft.2009.08.010

[8] Y. Yu, Y. Zhu, S. Li, and D. Wan, "Time series outlier detection based on sliding window prediction," *Mathematical Problems in Engineering*, no. 879736, 2014.

[9] Z. Hasani, "Robust anomaly detection algorithms for real-time big data: Comparison of algorithms," in *2017 6th Mediterranean Conference on Embedded Computing (MECO)*, June 2017, pp. 1–6.

[10] A. Kumar, A. Srivastava, N. Bansal, and A. Goel, "Real time data anomaly detection in operating engines by statistical smoothing technique," in *2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, April 2012, pp. 1–5.

[11] S. Ahmad, A. Lavin, S. Purdy, and Z. Agha, "Unsupervised real-time anomaly detection for streaming data," *Neurocomputing*, vol. 262, no. Supplement C, pp. 134 – 147, 2017, online Real-Time Learning Strategies for Data Streams. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231217309864

[12] N. Chen, Z. Yang, Y. Chen, and A. Polunchenko, "Online anomalous vehicle detection at the edge using multidimensional ssa," in *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, May 2017, pp. 851–856.

[13] V. Jyothi, X. Wang, S. K. Addepalli, and R. Karri, "Brain: Behavior based adaptive intrusion detection in networks: Using hardware performance counters to detect ddos attacks," in *2016 29th International Conference on VLSI Design and 2016 15th International Conference on Embedded Systems (VLSID)*, Jan 2016, pp. 587–588.

[14] C. Malone, M. Zahran, and R. Karri, "Are hardware performance counters a cost effective way for integrity checking of programs," in *Proceedings of the sixth ACM workshop on Scalable trusted computing*. ACM, 2011, pp. 71–76.

[15] Microsoft, "Get started with azure," World Wide Web, retrieved 2017-02-15. [Online]. Available: https://azure.microsoft.com/en-gb/get-started/

[16] N. Binkert, B. Beckmann, G. Black, S. K. Reinhardt, A. Saidi, A. Basu, J. Hestness, D. R. Hower, T. Krishna, S. Sardashti, R. Sen, K. Sewell, M. Shoaib, N. Vaish, M. D. Hill, and D. A.Wood, "The gem5 simulator," *SIGARCH Comput. Archit. News*, vol. 39, no. 2, pp. 1 – 7, Aug 2011. [Online]. Available: http://doi.acm.org/10.1145/2024716.2024718

[17] K. Parasyris, G. Tziantzoulis, C. D. Antonopoulos, and N. Bellas, "Gemfi: A fault injection tool for studying the behavior of applications on unreliable substrates," in *2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, June 2014, pp. 622–629.

[18] M. R. Guthaus, J. S. Ringenberg, D. Ernst, T. M. Austin, T. Mudge, and R. B. Brown, "Mibench: A free, commercially representative embedded benchmark suite," in *2001 IEEE International Workshop, Proceedings of the Workload Characterization, 2001. WWC-4*, ser. WWC '01. Washington, DC, USA: IEEE Computer Society, 2001, pp. 3–14. [Online]. Available: http://dx.doi.org/10.1109/WWC.2001.15

Verification of Active Front End Converter Using Co-Simulation Technique in the LabVIEW

Bojan Banković, Filip Filipović, Milutin Petronijević, Nebojša Mitrović, Vojkan Kostić

Abstract - This paper provides an overview of a co-simulation methodology for the converter Active Front End control algorithm implementation using a National Instrument's (NI) LabVIEW development environment for visual programming language. It allows a rapid prototyping, design, development, debugging and verification of control algorithm for real hardware module implementation, such as a Field-Programmable Gate Array (FPGA). The complete system was verified and optimised before being implemented in the hardware using the NI Multisim simulation plugin for the LabVIEW graphical programming environment. A case study is presented to demonstrate the co-simulation between a Voltage Source Inverter (VSI), LC filter and load side models in the Multisim and standalone control algorithm in the LabVIEW.

Keywords – Active Front End, control algorithm, co-simulation, Field Programmable Gate Array, LabVIEW, Multisim.

I. INTRODUCTION

The approach of design, prototype and deploy embedded control and monitoring systems is significantly changed in the past decade. It became necessary due to increased complexity in design to take extensive analysis and simulations prior to the first prototype build. The advancements in power electronic lead to the emergence of advanced hardware devices with possibilities for the development of new control algorithms [1]. The time needed to select the most optimal solution in such situations can be long and the choice may be ineffective. In order to simplify development, testing, validation, increase productivity and reduce time to final product a various software tools solution are presented such as LabView, Matlab, Caspoc, Proteus, solidThinking Embed and Saber.

The application of power electronic devices such as IGBT, especially in the AC motor control, has resulted the Voltage Source Converter (VSC) technology to become the new gold standard in industry [2]. Apart from motor control, new topologies take place in modern distribution systems as a part of distributed power sources [3]. The passive diode rectifier topology became overcome with the new convertor-inverter topology also called Active Front End due to its ability to recovery energy back in the power grid. The number of distributed power sources like photovoltaic solar farms, wind farms, electric cars, and energy storage is significant and combined with household/building appliances resulting a dramatic increase in the number of prosumers ('producer + consumer'). VSI converters cover a wide range of power ratings with increased efficiency at very affordable prices [4]. Simultaneously with hardware development, new control algorithms are emerging and new topologies are conceived. The twenty-first century - the age of energy-efficient devices and new regulations regarding the quality of electricity and CO_2 emissions [5], conditioned the development of new control algorithms. Their task is wide, from the possibility of the bidirectional energy flow control, adequate synchronization to the power grid and to provide grid voltage or power injection support. In addition to the aforementioned, the control algorithm should provide IGBTs switching control that will reduce the losses, perform the data acquisition/processing and ensure the system's immunity to disruptions.

A graphical system design approach is used for modelling, control, signal processing, and algorithm integration. LabVIEW software provides a superior method for design demanding embedded control systems and gives tools for appropriate tasks monitoring. The entire process of the Active Front End (AFE) modelling has been divided into several separate subprojects. Each of them is modelled separately and tested before the final co-simulation and build-in of the FPGA.

II. LABVIEW BASED MODEL ACTIVE FRONT END CONTROL

A. Active Front End control algorithm in the LabVIEW

Converters represent a bridge between renewable energy sources or prosumers at one side and the power grid on the other. They are irreplaceable devices in smart grids and micro networks generating variable/fixed frequency from DC sources or rectifying AC input voltage to the DC voltage for further conversion (e.g. storage or battery charging). The energy conversion is accomplished with two functional modules called the control stage and the power stage. The power processor of the converter is responsible for the power transfer from the input to the output, or vice versa through power stage. The controller is responsible for operating of the switches according to the specific control algorithm, while voltages and currents are measured at the system input or output. The widespread use of the AFE technology and its application is illustrated in the Fig. 1. Its possibility of bidirectional energy flow constitute the most important feature of this topology. DC-link converters connect motor/generator/load side inverters and/or battery storage or photovoltaic chopper with power grid enabling bidirectional energy flow. Such structure can also operate

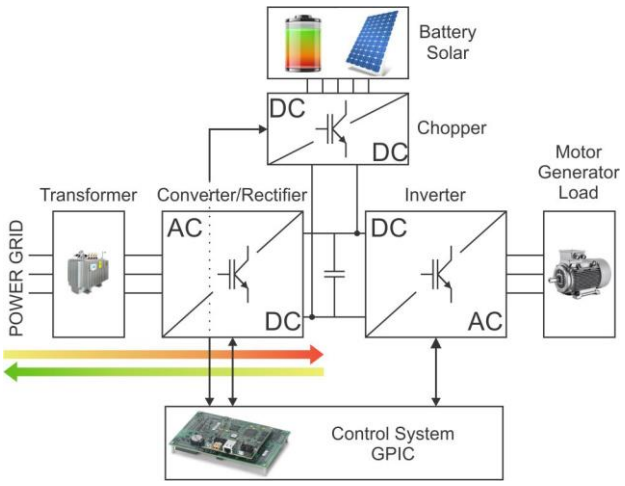


Fig. 1. Active Front End topology in micro grid

as a micro grid with back-to-back converter control strategy for an operation in both stand-alone and grid-connected cases. If we are considering one household/building as a small micro grid, two operating modes are possible. First mode is the stand-alone work regime of the micro grid, where the primary objective of the converter control is to provide desired voltage amplitude and frequency. The second mode is more complex because it requires synchronization of the converter outputs to the main power grid, but also offers more possibilities for various control algorithms.

In the initial stage of control algorithm development, the first mode of operation was taken in consideration. The basic requirements, that the control algorithm should satisfy, are to ensure the stability of the given voltage and frequency in the stand-alone operation. The control algorithm should minimize IGBTs switching losses and provide safe operation (prevents a short-circuit between top and bottom IGBTs in same branch). So extensive testing of the control algorithm should be carried out before the real hardware implementation is performed, in order to ensure functionality of the IGBT switching devices. The principal block-scheme for convertor control algorithm is presented in the Fig. 2.

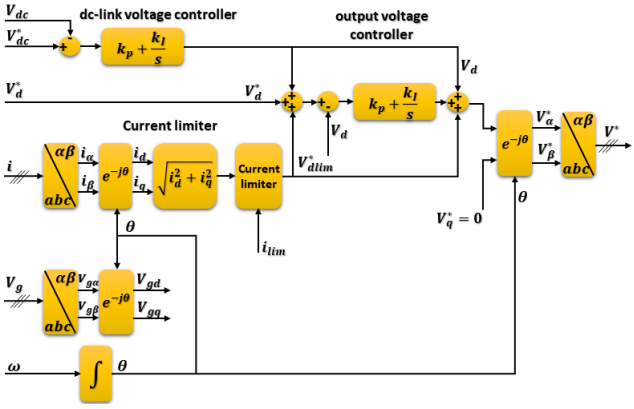


Fig. 2. The control algorithm scheme for the Active Front End converter

LabVIEW provides the convenience of modularity in modelling and programming creating a hierarchical structure of the virtual instruments (VIs) with no limit in the number of layers. Each VI has three components: a block diagram, a front panel and a connector panel. Within LabVIEW, program modularity means creating smaller sections of the control algorithms, known as subVIs. SubVIs are the same as VIs. They contain front panels and block diagrams, but they could be called from a main VI. A subVI is similar to a subroutine in text-based programming languages.

Control algorithm is divided in to several subVIs for easier testing and implementation at different hierarchical levels. An implementation of the control algorithm included three steps. The first one include FPGA development environment for control algorithm execution. The second includes complex mathematical operations and can be performed on the Rael Time platform with DSPs in order to optimize FPGA performance. The third level is the desktop computer for monitoring and control hardware over appropriate application generated in LabVIEW.

SubVIs of the control algorithm are:

- Triangle carrier signal (Triangle.vi) for PWM
- Saw carrier signal (Saw.vi) for PWM
- Dead time implementation for IGBT (DeadTime.vi)
- Filter for electrical signal (SigFilter.vi)
- Calculation angle θ based on ω (ThetaCalc.vi)
- ABC to DQ transformation (ABCtoDQFXP.vi)
- DQ to ABC transformation (DQtoABCFXP.vi)
- PID controller (PIDFXP.vi)

Main VI consists of four structures, which are used in the co-simulation through the FPGA Desktop Execution Node, as it is presented in Fig. 4:

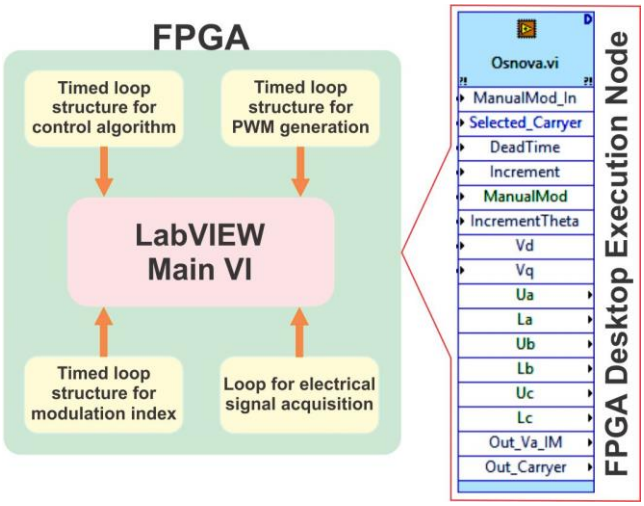


Fig. 4. The control algorithm for the Active Front End converter in the island mode

Before co-simulation is performed, every part of the control algorithm is tested in LabVIEW independently and optimised in respect with execution speed and memory usage. The optimisation is done for an NI Single-Board RIO 9863 Mezzanine with Embedded Device LX45 FPGA

sbRIO-9606. This combination give a possibility of DSP and FPGA usage for control algorithm implementation. Co-simulation is performed using add on LabVIEW Control Design and Simulation Module and NI LabVIEW Co-Simulation Plugin with Multisim in order to evaluate the performance of the control algorithm on the desktop. In order to simulate precise time behaviour of the control algorithm implemented on the FPGA, a discrete time execution rate of the Desktop Execution Node must be configured. This comes at the cost of performance, since a simulation step-size must be chosen to match the 40 MHz FPGA clock period. The simulation tasks was solved using the Runge-Kutta 23 ordinary differential equation (ODE) solver. The control and simulation loops for co-simulation in the form of block diagram is presented on the Fig. 5.

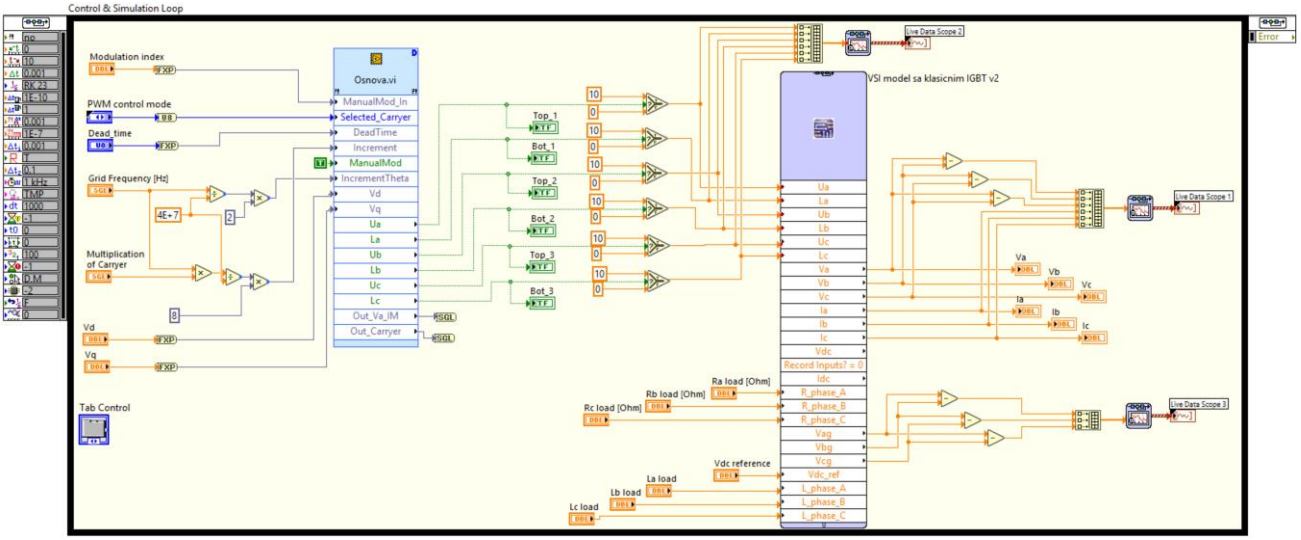


Fig. 5. The control and simulation loops for co-simulation in LabVIEW

B. Voltage Source Inverter, LC filter and load side model in the LabVIEW Multisim

In order to test control algorithm using co-simulation a Voltage Source Invertor (VSI), LC filter and load side model are designed in Multisim. Fig. 6 shows the model used for co-simulation. The load type can be changed from pure active to mixed inductive, in order to observe the current and voltage waveform quality. A reference voltage of the DC link could be set through V_{dcref} and load parameters through $R_{\_phase}$ and $L_{\_phase}$. Measured values for voltages on the output side of the converter and the grid side could be shown with current on the grid side.

The switches S1-S6 where represented with Infineon IGBT model SKM 50GB123D. The gate drivers were not considered in this co-simulation, because their effect on the

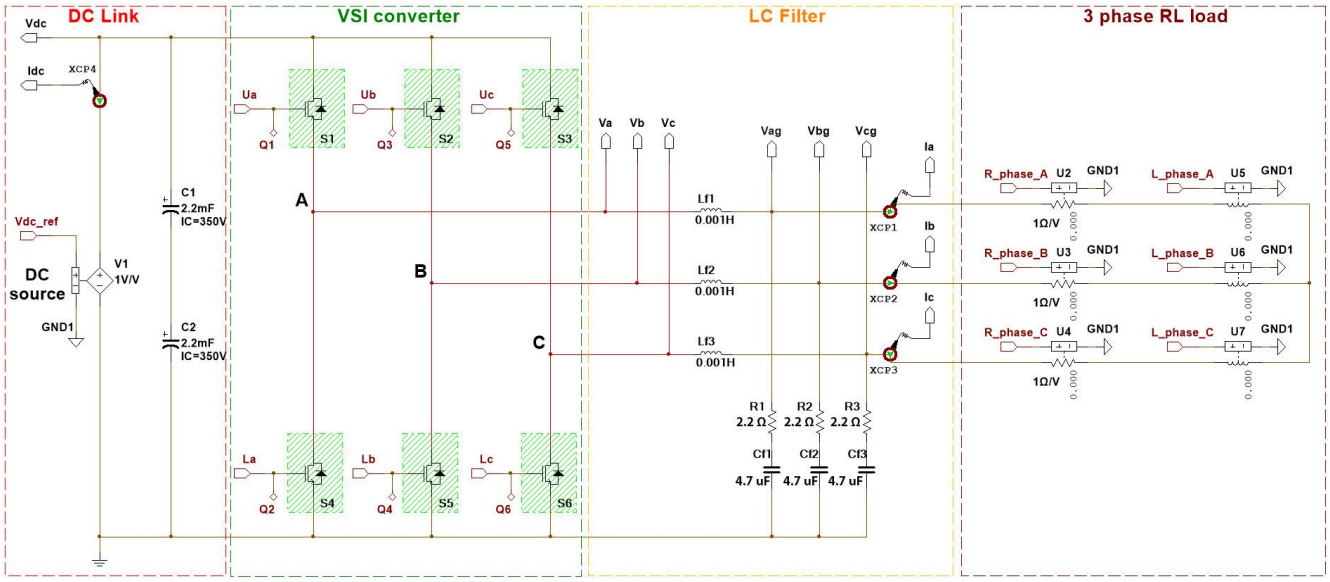


Fig. 6. The power part of the Active Front End converter in the island mode

system dynamic is minimal. The control signals for the IGBTs can be generated through Sine Pulse Width Modulation (SPWM) with saw or triangle carrier. These signals are transmitted to the transistors' gates through the ports U_a, U_b, U_c, L_a, L_b, and L_c. The frequency of the carrier should be select as an integer multiple of the nominal grid frequency f=50 Hz. The LC filter is designed corresponding to the nominal current of the VSI. This current according to the data sheet specification of the IGBTs is 40A for T_{case}=80 °C. In order to calculate inductor L for the filter the current ripple will be defined within ΔI_r=20% of nominal current. According to [6] the inductance L should be selected as:

$$L > 2\frac{U_{dc}}{2\cdot\pi\cdot\sqrt{2}\cdot f_{sw}\cdot 3\cdot\Delta I_r}, \quad (1)$$

where the DC link voltage U_{dc}=650 V and switching frequency f_{sw}=5000 Hz. Adopted value for inductor L is 1.5 mH. This value with nominal grid frequency and nominal load current should produce voltage drop below 10 % of the phase voltage. Voltage drop under rated load is given with:

$$\Delta U = I_n\cdot 2\cdot\pi\cdot f\cdot L \quad (2)$$

The value of the filter capacitors C must provide appropriate resonant frequency f_{res} of the output filter that should satisfy the condition given with Eq. (3) [6]:

$$10\cdot f_{out1h} < f_{res} < \frac{1}{2} f_{sw} \quad (3)$$

For the assumed value of the resonant frequency f_{res}=2000 Hz, the capacitor value C should be calculate with Eq. (4) according to [6]:

$$C = \frac{1}{4\cdot\pi^2\cdot f_{res}^2\cdot L} \quad (4)$$

The nearest standard value of the capacitance C=4.7 μF is selected. For the given parameters L and C, the resonant frequency is given with Eq. (5):

$$f_{res} = \frac{1}{2\pi\sqrt{LC}} \quad (5)$$

and satisfy the condition given with Eq. (3).

Taking into the account the filter quality factor Q=8 for characteristic impedance and damping resistor R_d should be calculated with Eq. (6) and (7) according to [6]:

$$Z_0 = \sqrt{\frac{L}{C}} \quad (6)$$

$$R_d = \frac{Z_0}{Q} \quad (7)$$

The nearest standard value of the resistor is R_d=2.2 Ω.

III. CO-SIMULATION RESULTS

Front end user interface in the LabVIEW co-simulation is presented in the Fig. 7. Through the desktop graphical user interface, it is possible to set the reference values for the grid voltage components and frequency, DC link voltage, dead time for IGBTs and load values in each phase. The measured voltages and currents are shown over the graphic charts while visual presentation of the IGBTs turn on/off state is given. There are also signal charts for the sine PWM signals for each of the six IGBT switches.

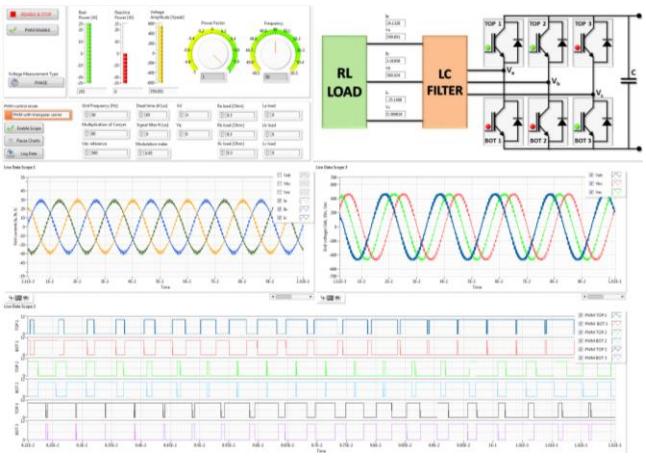


Fig. 7. Front end user interface in the LabVIEW co-simulation

The co-simulation is performed for nominal current of the convertor with the three phase active load of R_{phase}=5.74 Ω and unity power factor. The active power of the consumer is 27.5 kW. An output voltage RMS value from the converter is set to 400 V, grid frequency to 50 Hz, modulation index to 1, dead time of the IGBT to 1.125 μs and the reference DC link voltage to 650 V. In the Fig. 8. instantaneous values of phase currents I_a, I_b, I_c are presented in front of the LC filter. It can be noticed there is no leg shoot through caused with control algorithm. Currents are sinusoidal with frequency of 50 Hz. There is a 20% current ripple of nominal value in the waveform, which makes the current waveform non-ideal sinusoid. In the Fig. 9. instantaneous values of line voltages U_{ab}, U_{bc}, U_{ac} are presented. Voltages at the convertor output in comparison with the voltages in front of the LC filter are smoothed due to the presence of capacitors. Fig. 10. shows generated sine PMW signals with triangular carrier for all six IGBT switches.

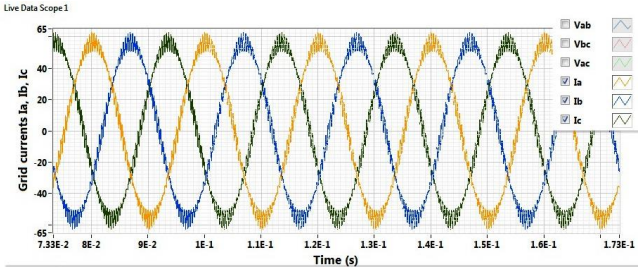


Fig. 8. Waveforms of phase currents I_a, I_b, I_c

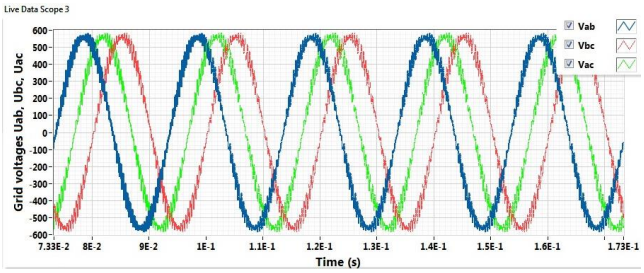


Fig. 9. Waveforms of line voltages U_{ab}, U_{bc}, U_{ac}

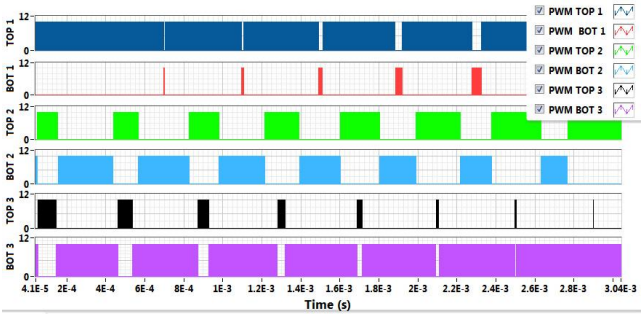


Fig. 10. Generated sine PWM signals

In the second case, a half of the nominal current load is used. It was a three phase active load of R_{phase}=11.4 Ω resistance. In this case, active power of the consumer is around 13.9 kW. The current and voltage waveforms and their instantaneous values are presented in Fig. 11. and 12. It is clear that this control strategy work in conditions of load deviation from nominal ones.

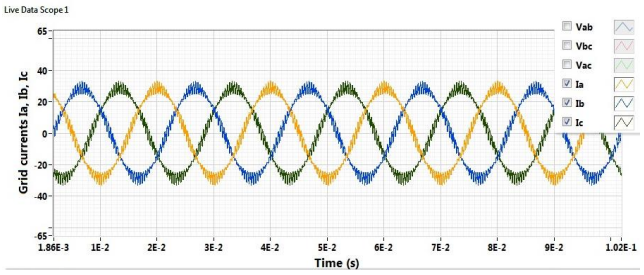


Fig. 11. Waveforms of phase currents I_a, I_b, I_c

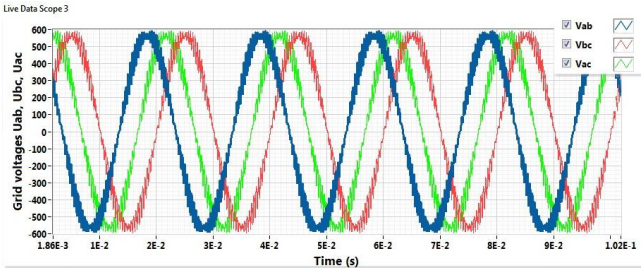


Fig. 12. Waveforms of line voltages U_{ab}, U_{bc}, U_{ca}

The control algorithm and converter model are tested also in the case of the induction load. The load parameters are: phase resistance R_{phase}=4.59 Ω, inductance L_{phase}=11 mH, power factor 0.8 and phase current I_{pRMS}=40 A. The active power of the consumer in this case is 22 kW. The instantaneous values of phase currents I_a, I_b, I_c are shown on the Fig. 13. It could be seen that the current waveforms are almost ideal sinusoid due to value of the inductance L_{phase}. A higher inductance values act as a filter on the current waveforms. Due to the fixed parameters values of the LC filter, which is a real case, the type of the consumer plays a significant role in the current and voltage waveforms. Due to fixed value of capacitance in the LC filter, the consumer load in this case caused a much lower quality voltage waveform. The waveforms of the U_{ab}, U_{bc}, and U_{ac} are presented in Fig. 14.

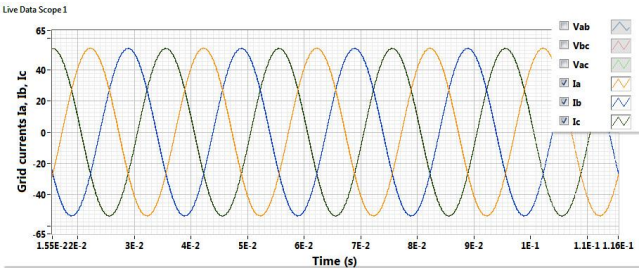


Fig. 13. Waveforms of phase currents I_a, I_b, I_c

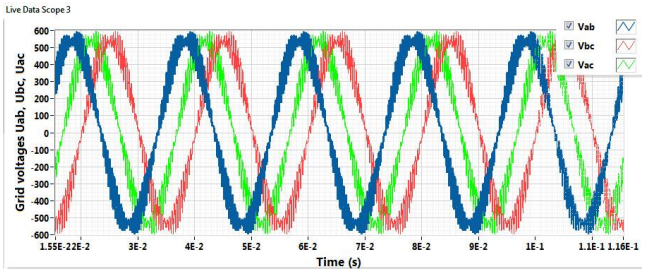


Fig. 14. Waveforms of line voltages U_{ab}, U_{bc}, U_{ac}

FPGA Desktop Execution Node has severe impact on the speed of simulation execution, and twofold impact on the simulation experience. On the one hand, 100 ms of simulation time can take up to an hour on decent desktop PC. On the other, this provides a possibility of online parameter change. Parameters of the load, for example, can be changed during simulation and results can be seen instantaneously. Simulation can be speeded up further more with step size parameters or maximum acceptable error trade off.

IV. Conclusion

This paper presents an Active Front End converter initial design procedure overview. Although current stage of development enables only islanding operation, in future work development of control algorithm part for grid-tie operation will be included. Its control algorithm is developed using LabVIEW and tested using LabVIEW add-ons.

Co-simulation showed up to be a valuable tool for the hardware and control system design, calculating parameters of the control algorithm and fine-tuning without a risk of damaging equipment. Time of development is reduced significantly and LabVIEW is suitable for this application, since algorithms that are executed on FPGA and desktop are written using the same graphical programming language.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Science and Technological Development, Republic of Serbia (Project number: III 44006).

REFERENCES

[1] E. Ohno, *"Power electronics and its applications to industry and society"*, Industrial Technology, 1996. (ICIT '96), Proceedings of The IEEE International Conference on, Shanghai, China, 2-6 Dec. 1996, ISBN: 0-7803-3104-4.

[2] Gerard-Andre Capolino, "*Recent advances and applications of power electronics and motor drives - Advanced and intelligent control techniques*", Industrial Electronics, 2008. IECON 2008. 34th Annual Conference of IEEE, Orlando, FL, USA, 10-13 Nov. 2008, ISSN: 1553-572X.

[3] Deepak Divan and Prasad Kandula, "*Distributed Power Electronics: An Enabler for the Future Grid*", CPSS Transactions on Power Electronics and Applications, December 2016. Vol. 1, No. 1, pp: 57-64.

[4] E. I. Carroll, "*Power Electronics for Very High Power Applications*", IEE/PEVD, London, September 21-23, 1998, pp: 1-6.

[5] Masashi Toyota, Zhi Long Liang, Yoshitoshi Akita, Hiroaki Miyata, Shuji Kato, Toshiki Kurosu, "*Application of Power Electronics Technology to Energy Efficiency and CO2 Reduction*", Hitachi Review Vol. 59 (2010), No. 4, pp: 141-148.

[6] Jaroslaw Guzinski, Haitham Abu-Rub, Patryk Strankowski, "*Variable Speed AC Drives with Inverter Output Filters*", 2015 John Wiley & Sons, Ltd, ISBN: 9781118782897

Rapid prototyping approach for dynamic load emulation

Lazar Sladojević, Milutin Petronijević, Miodrag Stojanović

Abstract – In this paper the new solution to the problem concerning dynamic load emulation is presented. This paper represents the improvement of a compensator method application with linear PID controller. Detailed simulation results provide good starting point for controller tuning and verification of the main idea. New technique, based on rapid prototyping approach is described and the experimental results are presented. Simulation and experimental results are compared for emulation of linear and nonlinear loads. Some limitations in use of proposed controller are gained and future improvements are noticed.

Keywords – Dynamic load emulation, Mechanical load, Rapid prototyping, Matlab simulation.

I. INTRODUCTION

In order to anticipate the performance of a designed system, engineers regularly use some of the well-known simulation tools. Today, the most of the simulations are carried out on digital computers. Mathematical models of systems are created and their behavior is inspected.

However, in order to overcome a number of uncertainties, sometimes the laboratory prototyping is recommended. Unfortunately, this is not always possible to achieve. A good example of this would be an electrical drive that is designed to work with some kind of nonlinear load. This is why the dynamic load emulation was developed. It is a mean of "hardware simulation" of any drives load condition that can be mathematically described. The dynamic load emulation procedure consists of two drives rotating on the same shaft with rigid coupling. One of them is the drive under test (DUT), the one whose behavior under certain conditions is inspected. The other one is the drive "acting" as the load, often called dynamometer (DYN). It doesn't have to be the same kind of drive as the drive under test, but it has to be torque-controllable.

There are four most commonly used methods for achieving the dynamic emulation:

1. The compensation method [1]
2. The feedforward regulation method [2]
3. The feedback regulation method [3]
4. The PI estimation method [4]

Lazar Sladojević, Milutin Petronijević and Miodrag Stojanović are with the Department of Power Engineering, Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, Niš
E-mail:
lazar.sladojevic@elfak.rs, milutin.petronijevic@elfak.ni.ac.rs, miodrag.stojanovic@elfak.ni.ac.rs

This paper will address the compensation method given in [1] and a way of its improvement. The theoretical hypothesis will be simulated using the Matlab/Simulink simulation software and experimentally proven using the experimental set-up described in section IV. The algorithm for dynamic emulation is developed in Simulink and transferred to a dSpace development platform which controls both machines. The DUT is a vector controlled, squirrel cage induction motor. The DYN is also a squirrel cage induction motor, but its torque is controlled directly, via a DTC drive. The proposed emulator will later be used for testing of speed controller robustness and performances.

II. Traditional compensation method

A. The inverse dynamics model

The most basic idea of a dynamic emulation algorithm is shown in Fig 1. This is so called inverse dynamics approach. The idea is very simple. The shaft speed is estimated from position encoder signal and used for calculation of a necessary load torque which is to be applied by a dynamometer. The flaw of this approach is the fact that it needs the inverse mechanical characteristic, i.e. the load torque as a function of speed:

$$T_l = f(\omega_r) \tag{1}$$

where T_l is the load torque and ω_r is the angular speed of the rotor.

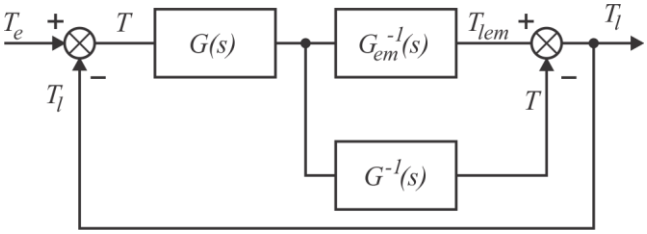


Fig. 1. The inverse dynamics approach to dynamic load emulation

On Fig. 1, $G_{em}(s)$ is the transfer function of an emulated load, while $G(s)$ is the transfer function of a mechanism - two rotors and a shaft.

$$G(s) = \frac{1}{J_m s + B_m} \tag{2}$$

The characteristic from Eq. (1) is practically unachievable since it requires the derivative of speed over time. For example, a simple linear load inverse mechanical characteristic is given in Eq. (3):

$$T_l = J\frac{d\omega_r}{dt} + B\omega_r \tag{3}$$

where J is the moment of inertia and B is the viscous friction coefficient of a linear load.

Such an approach would be impractical because measurement of acceleration is very submissive to noise. Further, it can also exhibit stability problems when the algorithm is implemented digitally. This is why the different approach was developed.

B. Regulating the speed instead of torque

The different approach is depicted in Fig 2. Consider for the beginning that $G_{comp}(s) = 1$. The transfer function of speed with respect to the driving torque is as follows:

$$\frac{\omega_r(s)}{T_e(s)} = G_{em}(s)\frac{G(s)G_t(s)}{1+G(s)G_t(s)} \tag{4}$$

where $T_e(s)$ is the driving motor torque, $G_{em}(s)$ is the emulated load transfer function, $G(s)$ is a transfer function of a mechanism and $G_t(s)$ is an emulated speed PI controller as shown in Fig. 2. The desired speed-torque characteristic is, however, given with Eq. (5).

$$\frac{\omega_r(s)}{T_e(s)} = G_{em}(s) \tag{5}$$

Eqs. (4) and (5) have different expressions on their right-hand sides, so the compensator is introduced to neglect these differences. Its transfer function is obviously:

$$G_{comp}(s) = \frac{1+G(s)G_t(s)}{G(s)G_t(s)} \tag{6}$$

By evaluating this transfer function, however, it can be seen that that it is improper. The degree of numerator is higher than degree of denominator. One possible solution of this problem is shown in this paper.

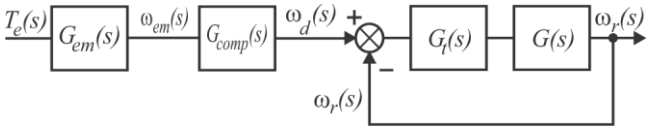


Fig. 2. The traditional compensation principle

C. Existing solution of a compensator problem

Before the suggested solution is explained, the existing solution has to be reviewed. It is adopted form [1] and consists of a modification of a compensator transfer function in discrete domain. Instead of continuous, the discrete compensator is used, with its discrete transfer function being:

$$G_{comp}(z) = \frac{1}{z}\frac{1+G(z)G_t(z)}{G(z)G_t(z)} \tag{7}$$

In order to compensate for the introduced delay, the transfer function of an emulated load has to be modified too. It now becomes:

$$G_{em}(z) = z \cdot Z\{G_{h0}(s)G_{em}(s)\} \tag{8}$$

where $G_{h0}(s)$ is the transfer function of a zero-order hold function used to discretize $G_{em}(s)$. The newly formed compensator can be realized in the digital processor which is normally used to execute the algorithm.

III. THE SUGGESTED DYNAMIC LOAD EMULATION APPROACH

This chapter covers the suggested approach to solving the compensator problem. The simplest way of designing the emulated speed controller, $G_t(s)$ is also shown.

By introducing a differential effect into a PI controller, $G_t(s)$ i.e. creating a PID controller, the degree of denominator in the compensator transfer function raises by one, and is now equal to the degree of the numerator. Such a compensator can be realized and simulated, even in continuous time domain. The differential effect also enhances dynamics of the system, especially when sudden changes in reference occur. The new scheme is very similar to the one from Fig. 2. The main difference is in the new transfer function of a controller, which is given in Eq. (9). Also, one more block is added to the regulation loop which is shown on the right side of Fig. 2. This block is mainly added for a purpose of calculating the parameters of the PID controller, but it also brings the whole model closer to reality. It accounts for any delay caused by the time necessary to execute the control algorithm, time needed for acquisition of data, calculating the necessary variables, delay caused by the inner current loops etc. This delay is described by the term T_{ek}. The new loop is shown in Fig. 3. The transfer function of the added block is simply:

$$G_{alg}(s) = \frac{1}{1+sT_{ek}} \tag{9}$$

The transfer function of an ideal PID controller is given with:

$$G_t(s) = G_{PID}(s) = K_p + K_i \frac{1}{s} + K_d s \quad (10)$$

where K_p is the proportional gain, K_i is the integral gain, and K_d is the differential gain of a PID controller.

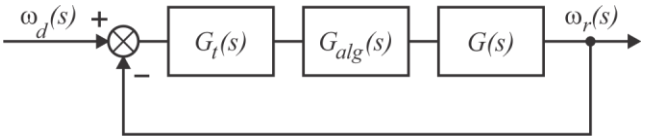


Fig. 3. The control loop of a suggested method

Although an ideal controller can't be made, it is possible to use Eq. (10) to calculate parameters of a real controller. The real PID controller's transfer function is given in Eq. (11):

$$G_{PID-real} = K_p + K_i \frac{1}{s} + \frac{K_d}{K_p} \cdot \frac{s}{1 + sT_D} \quad (11)$$

where T_D is filtering time constant of a differential term.

In order to calculate the parameters, the scheme on the Fig. 3 is used. Since the controller regulates the emulated speed, ω_{em}, that is, tries to make the real rotor speed equal to the emulated one, a simple feedback loop on the Fig. 3 provides for a characteristic equation of a controlled system. It is derived from Eq. (12):

$$1 + W_0 = 0 \quad (12)$$

where $W_0(s)$ is the open loop transfer function of a loop from Fig. 3. The characteristic equation in its extended form is given in Eq. (13):

$$s^3 + \frac{J_m + T_{ek}B_m + K_d}{J_m T_{ek}} s^2 + \frac{B_m + K_p}{J_m T_{ek}} s + \frac{K_i}{J_m T_{ek}} = 0 \quad (13)$$

In the previous equation, J_m is the moment of inertia and B_m is a viscous friction of a motor drive mechanism, respectively.

The previous equation is a characteristic equation of a third order system. It can be mathematically represented as a juncture of two systems: one of the order of two, and one of the first order. The goal is to have the poles of an order two system complex and dominant and the third pole should be negative, non-dominant and as far from imaginary axis as possible. Having in mind previous, the Eq. (13) can also be written in the following form:

$$(s - s_3) \cdot (s - s_2) \cdot (s - s_1) = 0 \quad (14)$$

where s_1, s_2 and s_3, are the roots of the Eq. (13) and are given in their well-known forms:

$$s_{1/2} = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2} \quad (15)$$

for the second order system and:

$$s_3 = -k_3\omega_n \quad (16)$$

The ζ and ω_n are the dumping coefficient and a natural frequency of a second order system, respectively. The value k_3=40 in the Eq. (16) is selected based on detailed simulation of whole system.

The recommended values of ζ and ω_n are given in [5]. They are:

$$\xi > 0.6 \cdot (1 - M_p) \quad (17)$$

and:

$$\xi\omega_n \geq \frac{4.6}{t_s} \quad (18)$$

where M_p is the overshoot in pu and t_s is the settling time in seconds. Those two are the common parameters that are optimized in the process of adjusting the parameters of a controller.

With ζ and ω_n determined, the only thing left to do is to replace those values in Eqs. (15) and (16), calculate the poles of the system, and replace those values into Eq. (14). By comparing the Eqs. (13) and (14), the parameters of PID controller are obtained.

IV. Simulation Results

The simulations were carried out for the torque regulation of a linear load and a two-mass system with backlash. The machines that were used for simulations are the ones that were used in real experiments. Linear load is simulated simply, by using its Torque - speed characteristic given with Eq. (19):

$$G_{em}(s) = \frac{1}{J_{em}s + B_{em}} \quad (19)$$

The reference q-axis current of vector-controlled DUT is a ±1 A square signal with the period of 0.8 seconds and duty cycle of 50%. It is assumed that there is nominal magnetic flux in the machine.

Simulation results for a linear load of different inertias and friction coefficients are given in figure 4. Fig. 4 a) represents linear load with the parameters $J_{em} = J_m$ and $B_{em} = B_m$, Fig. 4 b) load with parameters $J_{em} = 2J_m$ and $B_{em} = 20B_m$, while Fig. 4 c) shows the response of a linear load with parameters $J_{em} = 5J_m$ and $B_{em} = 50B_m$.

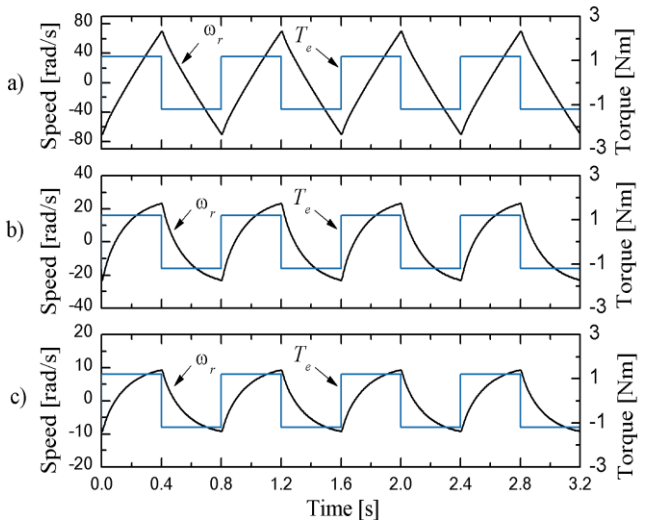


Fig. 4. Speed response of an emulated linear load

The next system for emulation is the simplest two-mass system with backlash as described in [6]. Mathematical model of a two mass-system is given by the following set of equations:

$$\frac{d\omega_{em}}{dt} = \frac{1}{J_m}\left(T_e - T_s - B_m\omega_{em}\right) \tag{20}$$

$$\frac{d\omega_{em2}}{dt} = \frac{1}{J_{em}}\left(T_s - T_{ext} - B_{em}\omega_{em2}\right) \tag{21}$$

$$\omega_{diff} = \omega_{em} - \omega_{em2} \tag{22}$$

where ω_{em} is the goal, emulated speed, T_e is the driving motor (DUT) torque, ω_{em2} is the dynamometer (DYN) rotor speed, T_{ext} is the external torque applied to the system (if it exists) and J_{em} and B_{em} are the emulated moment of inertia and viscous friction, respectively. The following equations describe the backlash in the system:

$$T_s = k_s D_a\left(\varphi_{diff}\right) \tag{23}$$

and:

$$D_a\left(\varphi_{diff}\right) = \begin{cases} \varphi_{diff} - \alpha, & \varphi_{diff} > \alpha \\ 0, & \left|\varphi_{diff}\right| < \alpha \\ \varphi_{diff} + \alpha, & \varphi_{diff} < -\alpha \end{cases} \tag{24}$$

where k_s is the elasticity of shaft, $D_a(\varphi_{diff})$ is a dead zone function, with α being the width of a dead zone and φ_{diff} is a difference in shaft angles of the two machines. The term φ_{diff} is obtained by integrating ω_{diff}, Eq. (22). Simulink model of a two-mass system is given in Fig. 5.

Simulation of a two-mass system was carried out with the following parameters:

- Fig. 6 a): $J_{em} = J_m$, $B_{em} = B_m$, $k_s = 2$, backlash of 15 degrees
- Fig. 6 b): $J_{em} = J_m$, $B_{em} = 20B_m$, $k_s = 2$, backlash of 15 degrees

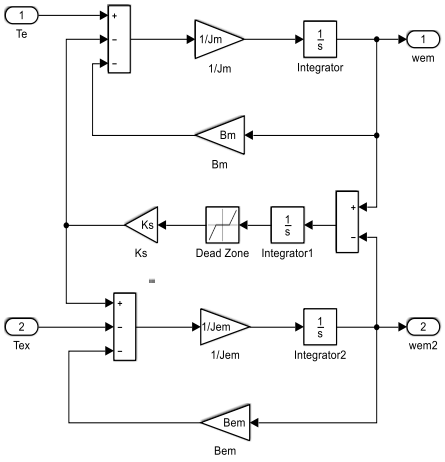


Fig. 5. Simulink model of a two-mass system with backlash

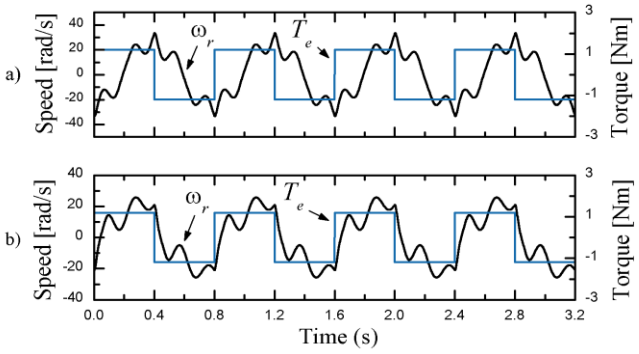


Fig. 6. Speed response of an emulated two-mass load with backlash.

It is to be observed from Figs. 5 and 6 that the system is acting as expected, which means it is well modeled. But this is yet to be confirmed with the experiments. Experimental setup and obtained results are shown in the following chapter.

V. Experimental results

The experiments were carried out according to the setup from Fig. 7. The emulation algorithm was developed in Simulink and transferred to a dSpace platform [7] which is used to execute it. This is so called rapid prototyping approach. Rapid prototyping is a great way of developing all kinds of control algorithms and bridging the gap to the practical application. After building a block scheme in Simulink, it can easily be translated into execution code and transferred to controller. Once loaded into the DSP, the

algorithm is ready for execution. The experimental software – Control Desk, also allows creating a custom control panel for the experiment monitoring from which the important variables can be observed and acquired. DSP system controls both drives.

The drive under test is a vector controlled induction motor, while the drive emulating the load is a Direct Torque Controlled (DTC) drive. Same emulated load conditions as in simulations were used.

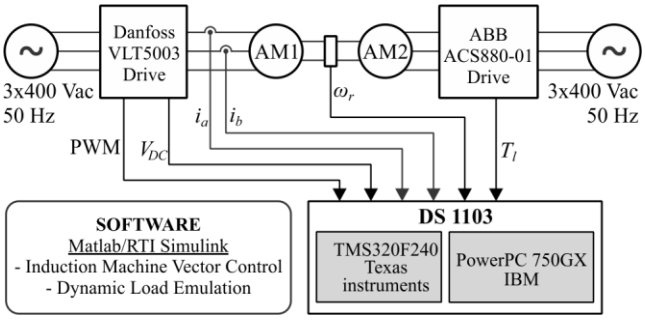


Fig. 7. Laboratory setup for the experiments

The machines that were used are standard squirrel-cage induction machines. Their most important parameters are given in Table I.

TABLE I
PARAMETERS OF THE MACHINES USED IN EXPERIMENTS

Rated power - P_n [W]	1500 W
Rated speed - n_n [o/min]	2860 o/min
Pole pairs – p	1
Rated phase voltage - U_{sn} [V]	230 V
Rated phase current - I_{sn} [A]	3.4 A
Stator resistance - R_s [Ω]	5.45 Ω
Stator leakage inductance - L_{ls} [mH]	11.8 mH
Rotor resistance - R_r [Ω]	3.18 Ω
Rotor leakage inductance - L_{lr} [mH]	11.8 mH
Magnetizing inductance - L_m [mH]	294.2 mH
Drive total inertia - J_m [kgm^2]	0.0035 kgm^2
Friction coefficient – B_m [Nms]	0.0022 Nms

Fig. 8 shows experimental results for linear load with the following parameters:

- Fig. 8 a) - $J_{em} = J_m$ and $B_{em} = B_m$
- Fig. 8 b) - $J_{em} = 2J_m$ and $B_{em} = 20B_m$
- Fig. 8 c) - $J_{em} = 5J_m$ and $B_{em} = 50B_m$

Experimental results for a two-mass system with backlash are shown in Fig. 9 in the following order:

- Fig. 9 a): $J_{em} = J_m$, $B_{em} = B_m$, k_s = 2, 15 degrees backlash
- Fig. 9 b): $J_{em} = J_m$, $B_{em} = 20B_m$, k_s = 2, 15 degrees backlash.

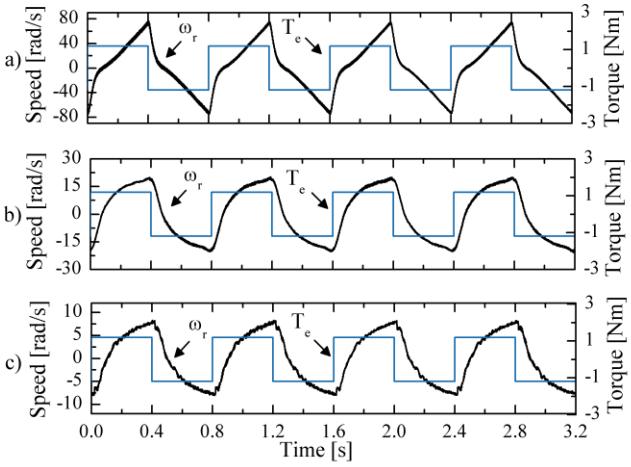


Fig. 8. Experimental speed response of an emulated linear load

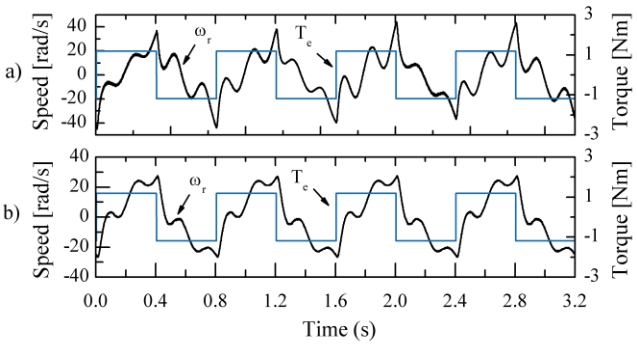


Fig. 9. Experimental Speed response of an emulated two-mass load

From Figs. 4 a) and 9 a) it can be seen that the experimental results do not fully correspond to the simulation results for the low values of inertia and friction. That reason for that lies in the fact that the parameters of PID controller do not provide exact tracking capabilities. The approach used to calculate those parameters utilizes the transfer function of an ideal PID controller, which doesn't fully relate to the real one. In practice, the controller is realized digitally, with low pass filtration applied to differential effect of the controller. This part has high influence on the dynamics of the system and must be designed with great care. If this is not the case, the impact of unmodeled dynamics is higher, and the results obtained do not match the simulation results.

For higher values of inertia and friction, conducted experiments give much better results. Both shape and maximal values of the speed response from Fig. 4 b) and c) are very similar to ones from Fig. 9 b) and c). Unmodeled dynamics do not have as high impact as before because the

whole system is slower.

Similar conclusion can be made for two-mass system with backlash. Speed responses from Figs. 6 and 10 are of similar shape and values, but there is some unmodeled dynamics that is not compensated. Therefore, the parameters of PID controller should be optimized or the application of nonlinear controller can be a promising solution for better tracking.

VI. Conclusion

In this paper, the rapid prototyping approach was used for dynamic load emulation. A short review of a design process is given in order to give initial parameters of the controller. The used algorithm represents a modified compensation method. Instead of PI, a PID controller was designed.

First stage in design are simulations in the Matlab/Simulink software. These simulations contain all elements that are described in this paper and later used in practical implementation. Linear load with different parameters was simulated first. Three different sets of parameters were adopted and the simulation results were shown. After that, the two-mass system with backlash was simulated. Two different sets of parameters were chosen and the corresponding results are presented. The results show that the system is acting as expected.

Experimental verification of results is depicted in Chapter V. The same load models with the identical parameters as in simulations were used. The experimental results show good agreement with the simulation results for higher values od emulated inertia and friction. For lower values of inertia and friction, however, the differences between simulations and experiment are bigger. This implies that there is a certain amount of unmodeled dynamics in the simulation model that the provided PID controller is not able to compensate. Further research in this field is required in order to improve the performance of a tracking controller. The experimental results for two-mass system with backlash show the similar performance.

Nevertheless, the whole research shows a great potential of the dynamic load emulation with rapid prototyping approach. Many different types of load can be emulated and tested. Switching from one type to another is very easy and intuitive, and only takes several minutes. This technique will later be used for testing of more advanced speed and torque controllers for motors.

REFERENCES

[1] Z. Hakan Akpolat, Greg M. Asher, Jon C. Clare *"Dynamic Emulation of Mechanical Loads Using a Vector-Controlled Induction Motor–Generator Set"*, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, VOL. 46, NO. 2, APRIL 1999.

[2] M. Rodič, K. Jezemik, M. Trlep, „A Feedforward Approach to the Dynamic Emulation of Mechanical Loads“, 35th Annual IEEE Power Elecrronics Specialists Conference, 2004

[3] M. Rodič, K. Jezemik, M. Trlep, „Dinamična emulacija mehanskih bremen”

[4] M. Rodič, K. Jezemik, M. Trlep, “Dynamic emulation of mechanical loads - approach based on nonlinear control“

[5] B. Veselić, „Metode digitalnog upravljanja i estimacije“, Elektronski fakultet Niš, 2017 godina

[6] K. Kyslan, E. Kušnír, V. Fedák, M. Lacko, F. Ďurovský, „Dynamic Emulation of Mechanical Loads with Backlash Based on Rapid Control Prototyping“, 16th International Power Electronics and Motion Control Conference and Exposition, Antalya, Turkey 21-24 Sept 2014

[7] DS1103 Hardware Installation and Configuration for release 5.2, dSpace Gmbh, 2006

Real-time System for Nonlinear Load Analysis in 50A Current Range

Marko Dimitrijević, Dejan Stevanović and Miona Andrejević-Stošović

Abstract – In this paper we will present 50A current range extension of the system for nonlinear load analysis based on virtual instrumentation paradigm. The presented system is capable of real-time operation, allowing measurements of various parameters, including *THD*, total power factor and various definitions of reactive and distortion power. The system is aimed for three-phase measurements, up to 50A. It is suitable for measurements of multiple small loads connected to one power source (one phase), the common case in households and offices.

Keywords – virtual instrumentation, real-time operation, extended current range

I. Introduction

Power quality analysis of electrical loads is related to measurement of power factor, distortion power and total harmonic distortion. In circuits consisting of linear loads, the currents and voltages are sinusoidal and the power factor effect depends only from the difference in phase between the current and voltage waveforms. In single phase system it is referred to as the displacement power factor or cos(φ) [1]. The power concepts developed for single-phase circuits with sinusoidal voltages and currents can be extended to polyphase circuits. Such circuits can be considered to be divided into a group of two-wire sets, with the neutral conductor (or a resistively derived neutral for the case of a delta-connected, three wire circuit) paired with each other conductor [2].

In the case nonlinear loads are present we should introduce new quantities in the calculations emanated by the harmonics and related power components [3]. Now the power factor can be generalized to a total or true power factor where the apparent power, involved in its calculations, includes all harmonic components. This is of importance in characterization and design of practical power systems which contain non-linear loads such as switched-mode power supplies [4].

The presented system represents extension of previously described system, presented in PhD thesis [5] and our previous papers [6].

Marko Dimitrijević and Miona Andrejević-Stošović are with the University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia. Email: {marko.dimitrijevic, miona.andrejevic}@elfak.ni.ac.rs.

Dejan Stevanović is with ICNT, Bulevar Nikole Tesle 61, 18000 Niš, Serbia. E-mail: dejan.stevanovic@icnt.rs.

The extension of system's measuring range is necessary for the analysis of loads with higher nominal power, as well as for the use of non-invasive load monitoring algorithms (NILM), which are applied to a group of loads connected to the same power source (phase) [7-10]. The system consists of an acquisition subsystem and a virtual instrument for calculating parameters and presenting the measurement results.

The system is based on virtual instrumentation paradigm, augmented with real-time component, ensuring determinism in measurement. The system hardware is implemented on field programming gate array (FPGA), it is in control of data acquisition. The system software has two components: application executing on real-time operating system (RTOS) aimed for basic parameter calculation, and user interface application for data analysis, manipulation and visualization that runs on simple operating system (GPOS). The system implemented in such a manner allows measuring and calculating a number of quantities that characterize loads, unobtainable with classical instruments. It offers many advantages – scalability, open architecture and flexibility: it can be simply extended in functionality – by number of calculated parameters and in the number of measurement channels (i.e. phases). It can be employed as a subsystem in harmonic compensation system [11] or used in power-in-the-loop and hardware-in-the-loop applications.

Flexibility is achieved through implementations on different platforms and through usage for different purposes: laboratory equipment for real-time operation (controller equipped with PXI-7813R FPGA card and expansion chassis), compact industrial PC for real-time operation (installed on programmable controller – PAC) or portable instrument connected by USB interface.

The paper is organized as follows: in the second section we will present signal acquisition system, third sections describes algorithms for harmonic analysis and parameter calculation, section four presents virtual instrument for data presentation and analysis and fifth section concludes the paper.

II. Acquisition

Acquisition subsystem consists of connection board with current sensors, acquisition modules and data interface. The layout of connection board is shown in Fig. 1. The function of the circuit is to connect the power source, the system and the load under test. The circuit consists of four Hall current sensors LEM LA-55P [12], which converts the current signal into a voltage. The conversion ratio is 2000A/V.

Hall sensors are chosen due to good linearity and accuracy at low frequencies, especially when measuring the DC current component. The disadvantage of Hall sensors, compared to the current transformer, is the necessity for external DC power supply.

Current sensor outputs are connected to *National Instruments* NI9215 [13] acquisition module. The acquisition module has four input channels for simultaneous voltage sampling with 16-bit resolution, 100kS acquisition rate and 250V_{RMS} isolation between the acquisition channel and the mass. The NI9215 acquisition module is isolated.

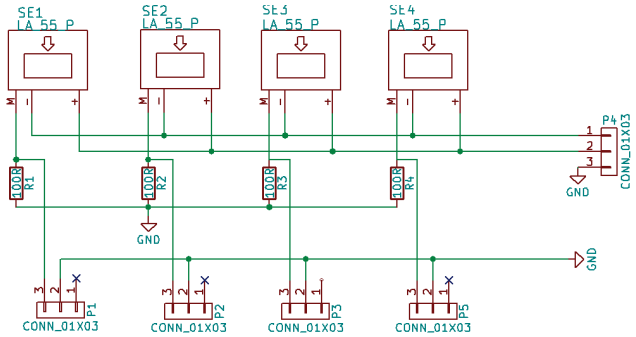


Fig. 1 Connection circuit diagram. The outputs of the LEM LA-55P sensors are connected to the NI9215 acquisition module

The voltage acquisition is implemented directly by connecting the *National Instruments* NI9225 [14] acquisition module. The resolution of the A/D conversion is 24-bit and the sampling rate is 50kS/s. The maximum isolation voltage between the acquisition channel and the ground is 600V_{RMS}. The measurement range is 300V_{RMS}. The voltage source is connected differentially. The input impedance of each channel is 1 MΩ. Detailed information regarding acquisition modules are given in Table 1.

TABLE I

INPUT CHARACTERISTICS NI 9225 AND NI 9215 MODULES

Parameter		NI9225	NI9215
Channels		3	4
Resolution		24	16
Sampling rate (kS/s)	min	1.613	1.613
	max	50	100
	range	50/n n=1,2,…,31	100/n n=1,2,…,31
Master time base	f_M	12.8	12.8
	Accuracy	±100ppm	±100ppm
Impedance		1 MΩ	1 GΩ
Noise (50kS/s)		2.2 mV	0.18 mV
Scaling coefficient		50.66 µV/LSB	305.18 µV/LSB

The input signals are conditioned, filtered and connected to the input of a 24-bit (16-bit for NI9215) sigma-delta A/D converters. The channels are completely independent and isolated, allowing simultaneous acquisition of three voltage and four current signals.

The sigma-delta A/D converter consists of an integrator, a comparator, a one-bit digital-to-analog converter (DAC) connected in a negative feedback loop and a decimation filter. The output signal of the integrator is compared in a comparator with a reference signal, which produces one bit. The sampling rate f_{OS} is 256 times the minimum frequency defined by the Shannon-Nyquist criterion. A series of bits from the comparator output is fed to the decimation filter input, which average and decimate the signal, giving a 24-bit digital signal with rate f_S. The decimation filter has a low-pass filter function, which reduces the quantization noise and aliasing. The decimator is realized as a FIR filter with a cutoff frequency equal to 0.47f_S. The bandwidth of the entire acquisition channel is shown in Fig. 2.

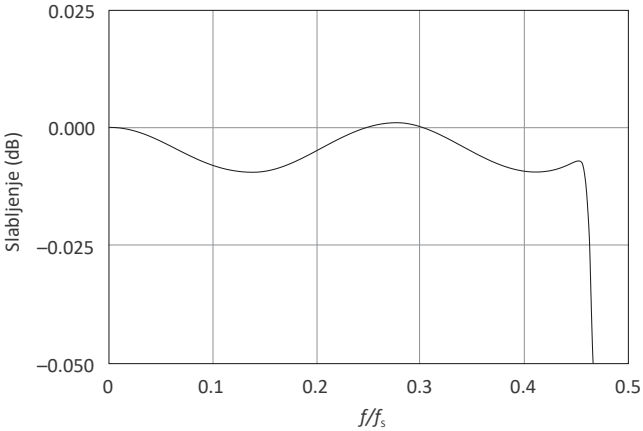


Fig. 2. Channel bandwidth

Sigma-Delta ADC uses an internal time base (Master Time Base) whose frequency is f_M=12.8 MHz. The time base can be synchronized with other acquisition modules, thus achieving synchronized sampling of different signals and minimal jitter. Oversampling frequency f_{OS} can be selected depending on the required sampling rate in such way the digital signal at the output of the decimation filter X_n has rate

$$f_S = \frac{f_M}{256 \cdot n}, \quad n = 1,2,...,31. \tag{1}$$

In this implementation the internal time base without dividers is used as oversampling frequency for all acquisition modules. The output digital signal of the ADC is 50kS/s, which corresponds to the bandwidth of the 25kHz input signal. Sigma-Delta ADC, using the oversampling method of acquisition, has several advantages, such as effectively eliminating alliances.

III. HARMONIC ANALYSIS AND PARAMETER CALCULATION

A. Definitions

We will first introduce the basic definitions that are expressing how the measured quantities are calculated from the current and voltage waveforms according to IEEE Std 1459-2000 [15] and IEEE Std. 1459-2010 [16] standards.

In the presence of nonlinear loads, the system no longer operates in sinusoidal condition and use of fundamental frequency analysis does not apply any more. Traditional power system quantities such as effective value, power (active, reactive, apparent), and power factor need to be numerically calculated from sampled voltage and current sequences by performing DFT or FFT algorithm.

The RMS value of some periodic physical entity X (voltage or current) is calculated according to the well-known formula:

$$X_{\mathrm{RMS}} = \sqrt{\frac{1}{\mathrm{T}} \int_{t_0}^{t_0+\mathrm{T}} \left(x(t)\right)^2 dt} \tag{2}$$

where $x(t)$ represents time evolution, T is the period and t_0 is arbitrary time. For any periodic physical entity $x(t)$, we can give Fourier representation:

$$x(t) = a_0 + \sum_{k=1}^{+\infty} \left(a_k \cdot \cos\left(k\omega t\right) + b_k \cdot \sin\left(k\omega t\right) \right) \tag{3}$$

or

$$x(t) = c_0 + \sum_{k=1}^{+\infty} c_k \cdot \cos\left(k\omega t + \psi_k\right) \tag{4}$$

where $c_0 = a_0$ represents DC component, $c_k = \sqrt{a_k^2 + b_k^2}$ magnitude of k^{th} harmonic, $\psi_k = \arctan\dfrac{b_k}{a_k}$ phase of the k^{th} harmonic and $\omega = \dfrac{2\pi}{\mathrm{T}}$, angular frequency.

Fourier coefficients a_k, b_k are:

$$a_0 = \frac{1}{\mathrm{T}} \int_{-\mathrm{T}/2}^{+\mathrm{T}/2} x(t)dt, \quad a_k = \frac{2}{\mathrm{T}} \int_{-\mathrm{T}/2}^{+\mathrm{T}/2} x(t) \cdot \cos\left(\frac{2k\pi t}{\mathrm{T}}\right) dt \tag{5}$$

and

$$b_k = \frac{2}{T} \int_{-\mathrm{T}/2}^{+\mathrm{T}/2} x(t) \cdot \sin\left(\frac{2k\pi t}{\mathrm{T}}\right) dt. \tag{6}$$

The RMS value of k^{th} harmonic is

$$X_{\mathrm{k,\,RMS}} = c_k \Big/ \sqrt{2}. \tag{7}$$

We can calculate total RMS value

$$X_{\mathrm{RMS}} = \sqrt{\sum_{k=1}^{M} X_{k,\,\mathrm{RMS}}^2} = \sqrt{X_{\mathrm{1,\,RMS}}^2 + X_{\mathrm{H,\,RMS}}^2} \tag{8}$$

where M is highest order harmonic taken into calculation. Index "1" denotes first or fundamental harmonic, and index "H" denotes contributions of higher harmonics.

Equations (2) – (8) need to be rewritten for voltage and current. Practically, we operate with sampled values and integrals (5) and (6) are transformed into finite sums.

For a single-phase system where k is the harmonic number, φ_k phase difference between voltage and current of k^{th} harmonic and M is the highest harmonic, the total active power is given by:

$$P = \sum_{k=1}^{M} I_{k,\mathrm{RMS}} \cdot V_{k,\mathrm{RMS}} \cdot \cos\varphi_k = P_1 + P_{\mathrm{H}}. \tag{9}$$

The first addend in the sum (9), denoted with P_1, is fundamental active power. The rest of the sum, denoted with P_{H}, is harmonic active power.

Total reactive power is given by:

$$Q = \sum_{k=1}^{M} I_{k,\mathrm{RMS}} \cdot V_{k,\mathrm{RMS}} \cdot \sin\varphi_k = Q_1 + Q_H. \tag{10}$$

It should be noted that the actual contribution of harmonic frequencies to active and reactive power is small (usually less than 3% of the total active or reactive power). The major contribution of higher harmonic to the power comes as distortion power D:

$$D^2 = \sum_{\substack{j \neq k \\ j=1,k=1}}^{M} I^2_{\,j,\mathrm{RMS}} \cdot V^2_{\,k,\mathrm{RMS}} \tag{11}$$

The apparent power can be written:

$$\begin{aligned} S^2 = &\underbrace{I^2_{\,1,\mathrm{RMS}} \cdot V^2_{\,1,\mathrm{RMS}}}_{S_1^2} + \underbrace{I^2_{\,1,\mathrm{RMS}} \cdot V^2_{\,\mathrm{H,RMS}}}_{D_V^2} + \\ &+ \underbrace{V^2_{\,1,\mathrm{RMS}} \cdot I^2_{\,\mathrm{H,RMS}}}_{D_I^2} + \underbrace{V^2_{\,H,\mathrm{RMS}} \cdot I^2_{\,\mathrm{H,RMS}}}_{S_{\mathrm{H}}^2} \end{aligned} \tag{12}$$

where S_1 represents fundamental apparent power, D_V voltage distortion power, D_I current distortion power and S_{H} harmonic apparent power. S_1 and S_{H} are

$$S_1 = \sqrt{P_1^2 + Q_1^2},\ S_{\mathrm{H}} = \sqrt{P_{\mathrm{H}}^2 + Q_{\mathrm{H}}^2 + D_{\mathrm{H}}^2} \tag{13}$$

where D_{H} represents harmonic distortion power. The total apparent power is

$$S = \sqrt{P^2 + Q^2 + D^2}. \tag{14}$$

The total harmonic distortions, THD, are calculated from the following formula:

$$THD_I = \frac{I_{\mathrm{H,\,RMS}}}{I_{\mathrm{1,\,RMS}}} = \frac{1}{I_{\mathrm{1,\,RMS}}} \sqrt{\sum_{j=2}^{M} I_{j,\,\mathrm{RMS}}^2} = \sqrt{\frac{I_{\mathrm{RMS}}^2 - I^2_{\,\mathrm{1,\,RMS}}}{I^2_{\,\mathrm{1,\,RMS}}}} \tag{15}$$

and

$$THD_V = \frac{V_{\mathrm{H},RMS}}{V_{\mathrm{1},RMS}} = \frac{1}{V_{\mathrm{1},RMS}} \sqrt{\sum_{k=2}^{M} V_{k,\,\mathrm{RMS}}^2} = \sqrt{\frac{V_{\mathrm{RMS}}^2 - V^2_{\,\mathrm{1,\,RMS}}}{V^2_{\,\mathrm{1,\,RMS}}}} \tag{16}$$

where I_j,V_k j,k=1, 2, …, M stands for the harmonic of the current or voltage. It can be shown that:

$$\begin{aligned} D_I &= V_{1,\,\mathrm{RMS}} \cdot I_{\mathrm{H,\,RMS}} = S_1 \cdot THD_I \\ D_V &= V_{\mathrm{H,\,RMS}} \cdot I_{1,\,\mathrm{RMS}} = S_1 \cdot THD_V \\ S_H &= S_1 \cdot THD_I \cdot THD_V . \end{aligned} \tag{17}$$

Fundamental power factor or displacement power factor is given by the following formula:

$$PF_1 = \frac{P_1}{S_1} = \cos\varphi_1 . \tag{18}$$

Total power factor *DPF*, taking into calculation (9) and (12), is

$$TPF = \frac{P}{S} = \frac{P_1 + P_\mathrm{H}}{\sqrt{S_1^2 + D_I^2 + D_V^2 + S_\mathrm{H}^2}} \tag{19}$$

and substituting (17) and (18):

$$TPF = \frac{\left(1 + \dfrac{P_\mathrm{H}}{P_1}\right)\cos\varphi_1}{\sqrt{1 + THD_I^2 + THD_V^2 + \left(THD_I THD_V\right)^2}} \tag{20}$$

In real circuits, $P_\mathrm{H} << P_1$ and voltage is almost sinusoidal (THD_V<5%), leading to simpler equation for *TPF*:

$$TPF = \frac{\cos\varphi_1}{\sqrt{1 + THD_I^2}} . \tag{21}$$

B. Harmonic analysis implementation

Frequency analysis of the signal is realized in two ways: in hardware, using the FPGA (*Field Programming Gate Array*) when the system is operating in real time, and in software when time determinism is not crucial point.

The function of the FPGA circuit is the determination of the fundamental frequency and the harmonic analysis of the current and voltage waveforms, i.e. calculation of harmonic amplitudes and phases. Algorithms are implemented on the Xilinx Vitex-II FPGA integrated circuit. FPGA is placed on the National Instruments PXI-7813R card connected to the NI-8014 PXI controller, via PCI interface [17]. The FPGA circuit clock is 40MHz. The card has 160 bidirectional digital channels, which support 3.3 V, 5 V, LVTTL and TTL logic levels. It supports 160 64-bit resolution and 40 MHz frequencies, with a maximum error of 100ppm. The card has 196 KB of memory, and the total number of DMA channels is 3.

Bidirectional channels are aggregated into four groups, each representing a separate physical connector. Through one 40-bit connector, the NI9151 extension housing, which includes the NI9225 and NI9215 acquisition modules is connected to the card. FPGA is connected to the PXI controller via PCI bus.

The FPGA circuit performs three parallel processes. The function of the first process is to control the acquisition of the signal. The process is performed in a programming loop that is repeated at time intervals equal to the integer duration of the cycle of the clock. The sampling rate of the signal is equal to the loop repetition frequency for acquisition.

The second process performs harmonic analysis of the selected signals using Goertzel algorithm [18]. The iterative part of the algorithm is executed on a block of 2000 samples, which at the maximum sampling speed corresponds to 40ms in time domain. The real and imaginary part of the spectrum are calculated in the non-iterative part of the algorithm:

$$\begin{aligned} \mathrm{Re}\left(X_k\right) &= S_k^{(\mathrm{N}-1)} - S_k^{(\mathrm{N}-2)} \cdot \cos\left(\frac{2\pi k}{\mathrm{N}}\right) \\ \mathrm{Im}\left(X_k\right) &= S_k^{(\mathrm{N}-2)} \cdot \sin\left(\frac{2\pi k}{\mathrm{N}}\right). \end{aligned} \tag{22}$$

Values for $\sin\left(\frac{2\pi k}{\mathrm{N}}\right)$ and $\cos\left(\frac{2\pi k}{\mathrm{N}}\right)$ are stored in the block memory (lookup table).

The function of the third process is transferring data from the FPGA to the data processing subsystem. The transfer takes place over three 64-bit DMA channels. In one process cycle, 192 bits can be transferred.

In the case of software-implemented frequency analysis, acquisition modules are connected via National Instruments cDAQ-9174 chassis and USB interface to a computer. sampled values of current and voltage are transmitted via the DAQmx driver to a virtual instrument for further analysis. Frequency analysis is performed using standard algorithms, available as LabVIEW modules [19]. These functions can be implemented on a real-time operating system [20] or a general-purpose operating system.

C. Parameter calculation

The software component of the system is implemented using National Instruments LabVIEW development package, which provides the possibility of simple realization of virtual instruments [21]. The virtual instrument consists of an interface to the acquisition module, a procedure for calculating numerical values and a user interface.

The interface to the acquisition modules is implemented through the DAQmx driver. Physical channels are represented as virtual channels. A virtual channel is a set of attributes assigned to a specific measurement and includes the name of the channel, the physical channel through which the measurement is performed, the inbound connection terminals, the type of measurement and scaling. The physical channel determines the terminal on the acquisition module to which the receiving signal is fed. A virtual channel can be configured globally at the system level, or locally in the application itself through an application interface. Each physical channel has a unique name.

It is possible to aggregate multiple virtual channels that define the same types of measurements in one measurement process. Processes, similar to virtual channels, can be defined locally at the application level and globally at the system level.

The application implemented using LabVIEW

developing package consists of three threads, exploiting contemporary multicore capabilities. First thread receives data from FPGA level. Second thread calculates power quantities: apparent power, active power, reactive power and energy; power quality parameters including total harmonic distortion (THD_I, THD_V), crest factors, total power factor and displacement power factor. Third thread handles TCP/IP communication with user interface.

The application is capable of logging measured values and calculated parameters and events locally on a hard disk drive (Fig. 3). All measured and calculated parameters, transient analysis, as well as triggered events are stored on hard disk or other type of data storage. The data quantity and therefore the measurement time, depends on the storage capacity. A separate thread is responsible for network communication with the user PC.

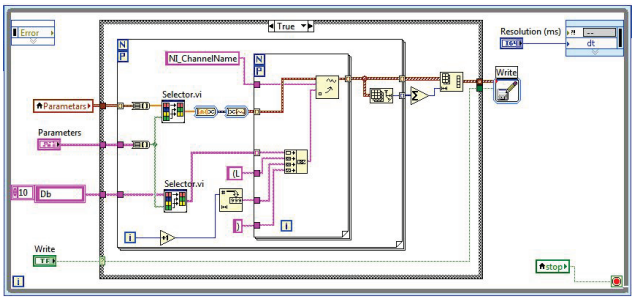


Figure 3. A part of the real-time application

IV. Virtual Instrument

The user interface virtual instrument is implemented in National Instruments LabVIEW developing package, which provides for simple creation of virtual instruments [21]. The virtual instrument is an application running on general purpose operating system (GPOS) such as Microsoft Windows, Linux or MacOS, connected with RTOS over the network. It consists of an interface to real-time application and a graphic user interface. The communication between the user interface and the real-time application is implemented using TCP/IP protocol.

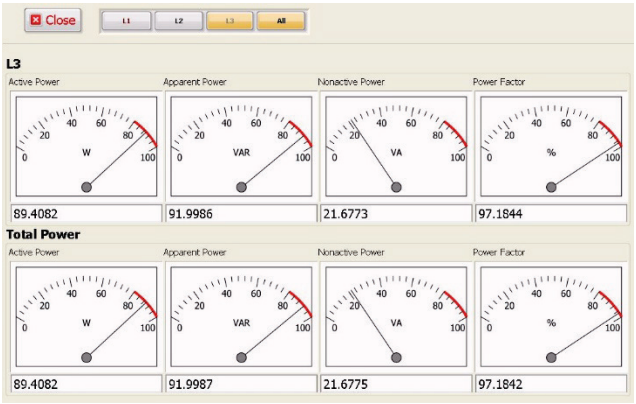


Fig. 4. Virtual instrument graphic interface displaying active, apparent and non-active power, as well as power factor for L3 phase and all phases

The user interface of the virtual instrument consists of visual indicators. It provides basic functions for measurement. The indicators – gauges and graphs – show measured values. User interface also provides controls for data manipulation and saving the measured values (Fig. 4).

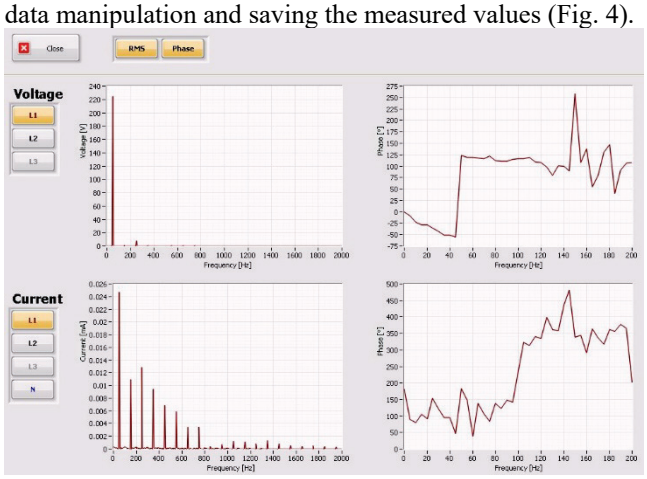


Fig. 5. Virtual instrument shows voltage and current spectra for L1 phase

The virtual instrument shows waveforms and spectra of measured voltages and currents (Fig. 5 and 6). Measured electrical quantities such as RMS and DC values, as well as calculated power quantities are shown in the front panel using virtual gauges and numeric indicators. Power factor and other power quality parameters are shown numerically. The virtual instrument can operate in one phase or three phase mode.

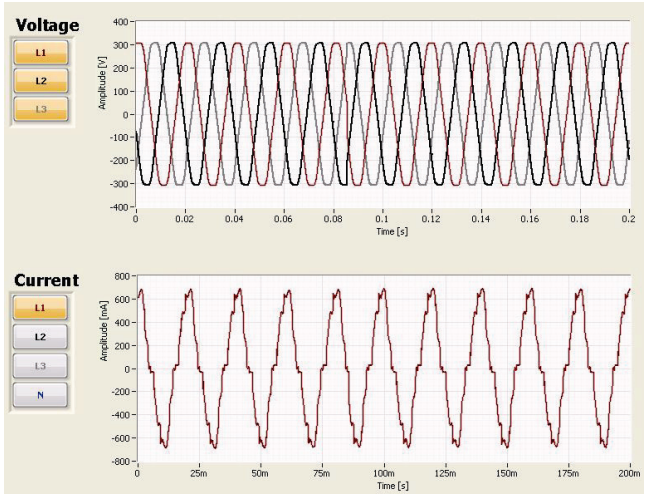


Fig. 6. Voltage waveforms for all phases, current waveform of L1

Harmonic magnitudes are shown in a table and each magnitude can be represented separately.

IV. Conclusion

A new approach to power factor and distortion analysis in polyphase systems was presented. It aggregates the

advantages of virtual instrumentation and the real-time response of classical instruments, capable for real-time sampling and measuring voltage and current of the device under test, providing possibility for transient analysis in time and frequency domain. FPGA and RTOS are taking the main role in data processing of all kind. In this way we introduced a flexible and versatile system with practically unlimited possibilities. The network PC has the control and user interface function.

The system elaborated in this paper is implemented as laboratory instrument using PXI-8105 controller running PharLap OS, equipped with PXI-7813R Virtex-II 3M gate FPGA programmable card and cRIO-9151 expansion chassis. Alternatively, it can be realized using programmable automation controllers (PAC) such National Instruments CompactRIO series. The small sized CompactRIO system includes acquisition modules, FPGA chassis and a real-time controller running VXWorks RTOS. It is suitable for long time executions and hard-ware-in-the-loop tests.

The presented application is capable for three-phase operation. However, the system is exceedingly scalable – it can be easily extended by adding additional acquisition modules into the expansion chassis.

ACKNOWLEDGEMENT

This research was partly funded by The Ministry of Education and Science of Republic of Serbia under contract No. TR32004.

REFERENCES

[1] H. W. Beaty, D. G. Fink, *Standard handbook for electrical engineers* (McGraw-Hill, New York, 2007).

[2] John G. Webster, *The Measurement Instrumentation and Sensors Handbook* (CRC Press, 1999).

[3] T. H. Tumiran, M. Dultudes, *"The Effect of Harmonic Distortion to Power Factor"*, Proceedings of the International Conference on Electrical Engineering and Informatics, 2007, pp. 834–837 Institute Teknologi Bandung, Indonesia

[4] G. Moschopoulos, *Single-Phase Single-Stage Power-Factor-Corrected Converter Topologies*, IEEE Trans. on Industrial Electronics 52, 2005, pp. 23–35.

[5] M. Dimitrijević, *Elektronski sistem za analizu polifaznih opterećenja baziran na FPGA*, PhD Thesis, Faculty of Electronic Engineering, University of Niš, Niš, Serbia, 2012.

[6] M. A. Dimitrijević, V. B. Litovski, *Power Factor and Distortion Measuring for Small Loads Using USB Acquisition Module*, Journal of Circuits, Systems, and Computers (JCSC), Vol. 20, No. 5, August 2011

[7] M. Andrejević-Stošović, M. Dimitrijević, D. Stevanović, *"Application of ELM Algorithm in Characterization of Nonlinear Loads"*, Proc. of the 2nd IcETRAN conference, pp. EL1-EL4 Kladovo, 2017

[8] M. Andrejević-Stošović, D. Stevanović, M. Dimitrijević, *"Monitoring and Classification of Nonlinear Loads Based on Artificial Neural Networks"*, 13th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS), Niš, 978-1-5386-1799-1, doi: 10.1109/TELSKS.2017.8246320

[9] M. Andrejević-Stošović, D. Stevanović, M. Dimitrijević, *"Classification of Nonlinear Loads Based on Artificial Neural Networks"*, IEEE 30th International Conference on Microelectronics (MIEL), Niš, 978-1-5386-2561-3, doi:10.1109/MIEL.2017.8190107

[10] Djordjević, S., Dimitrijević, M., Litovski, V., A Non-Intrusive Identification of Home Appliances Using Active Power and Harmonic Current, Facta Universitatis, Series: Electronics and Energetics, Vol 30, No 2, Niš, ISSN 0353-3670, doi: 10.2298 /FUEE1702199D

[11] M. Dimitrijević, D. Stevanović, M. Andrejević Stošović, M. Petronijević, P. Petković, *"Compensation of Harmonics in Smart Grid Using Photovoltaic Energy Source"*, XI International Symposium on Industrial Electronics - INDEL 2016, Banja Luka, 03.11.-05.11. 2016, pp. 1-6.

[12] LEM: Current Transducer LA 55-P/SP1 Datasheet

[13] National Instruments: NI 9215 Operating Instructions and Specifications.

[14] National Instruments: NI 9225 Operating Instructions and Specifications.

[15] IEEE Std 1459-2000 – Standard Definitons for the Measurement of Electric Power Quantities Under Sinusoidal, Nonsinusoidal, Balanced or Unbalanced Conditions (The Institute of Electrical and Electronics Engineers, 2000).

[16] IEEE Std 1459-2010 – Revision Standard Definition for the Measurement of Electric Power Quantities Under Sinusoidal, Nonsinusoidal, Balanced or Unbalanced Conditions (The Institute of Electrical and Electronics Engineers, 2010).

[17] National Instruments: NI PXI-7813R R Series Digital RIO with Virtex-II 3M Gate FPGA.

[18] Goertzel, G.: *"An Algorithm for the Evaluation of Finite Trigonometric Series."* The American Mathematical Monthly, January 1958, No. 1, Vol. 65, pp. 34-35.

[19] National Instruments: NI PXI-7813R R Series Digital RIO with Virtex-II 3M Gate FPGA.

[20] Jarvis, C., Kinsella, K., Timpanaro, P.: *"Phar Lap ETS™ – An Industrial-Strength RTOS White Paper."*

[21] National Instruments: LabVIEW System Design Software.

A Surface-Potential Based Sub-Circuit Model of I-V Characteristics In AlGaN/GaN HEMTs

Nebojša Janković, Soroush Faramehr and Petar Igić

Abstract - In this paper, the surface-potential based sub-circuit model of AlGaN/GaN HEMT yielding accurate and continuous I-V characteristics in the whole operation range is described. The model includes some important physical phenomena affecting device DC operation like 2DEG charge quantization, electron velocity saturation, high-field mobility degradation, channel length modulation, polarization charge and self-heating effects. A good fit between the sub-circuit model and the experimental I-V characteristics of Al0.25Ga 0.75N/GaN HEMT is demonstrated with SPICE simulations.

Keywords-Surface-Potential,Sub-Circuit Model, AlGaN/GaN, HEMTs

I. INTRODUCTION

AlGaN/GaN high electron mobility transistors (HEMTs) are considered as the best candidate for high-speed, microwave and high-power applications due to III-Nitride material properties such as high electron mobility, wide band-gap and large polarization charge enabling high current densities [1]. With significant progress in the improvements of AlGaN/GaN structure quality and performance of GaN HEMTs, the exploitation of full potential of these devices requires the advanced electrical models for circuit simulations. Among various compact models [2,3], the surface-potential-based (SP) models of GaN HEMTs appears advantageous because of the mostly physics-based model parameters and the ability to precisely describe the I-V characteristics in the whole operation range. In addition, the SP models have been proven very successful in the past for modeling advanced MOSFETs [4,5,6].

The majority of compact SP models of GaN HEMTs proposed so far [7-15] are based on deriving the approximate analytical expressions either for the electrostatic potential ψ_s of quantum-well-formed channel [7-11,13,14] and/or the mobile charge density n_s of two-dimensional electron gas (2-DEG) [8,9,12]. The analytical relations for ψ_s and/or n_s are subsequently used in drift-diffusion continuity equation with certain carrier transport models for obtaining the I-V characteristics of GaN HEMTs [8-15]. The main difficulty in deriving the analytical expressions of ψ_s and/or n_s stems from their complicated variations with the applied gate-source V_{GS} and drain-source V_{DS} voltages. simplifying the Schrodinger's and/or Poisson's equations in different Consequently, all SP-based compact models [7-15] are essentially developed by regions of device operation and using the suitable chosen smoothing functions to cover the whole V_{GS} and V_{DS} ranges.

In this paper, the SP-based sub-circuit (SPSC) model of AlGaN/GaN HEMT capable to accurately produce continuous I-V characteristics in the whole device operational range is presented and implemented in SPICE [23]. It avoids the need for the derivation of complex analytical expressions in compact modeling and allows for easy inclusion of some important physical phenomena such as 2DEG charge quantization, electron velocity saturation, high-field mobility degradation, channel length modulation, polarization charge and self-heating effects. A good fit of the SPSC model with experimental I-V characteristics of Al0.25Ga 0.75N/GaN HEMT is demonstrated in this work using SPICE simulations.

II. MODEL DESCRIPTION

Fig.1.a and Fig.1.b show the cross section of AlGaN/GaN HEMT and the schematics of the corresponding SPSC model, respectively.

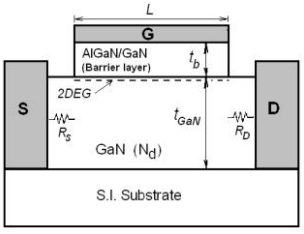


Fig. 1.a

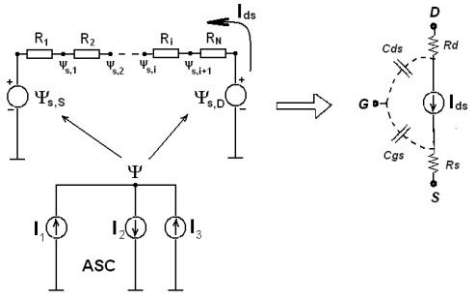


Fig. 1.b

Fig. 1. a) The AlGaN/GaN HEMT cross section with relevant parameters. b) The schematics of SPSC model with ASC

Nebojsa Jankovic: Faculty of Electronic Engineering Nis, University of Nis, Aleksandra Medvedeva 14, Nis, Serbia; e-mail: nebojsa.jankovic@elfak.ni.ac.rs
Soroush Faramehr, and Petar Igić: Electronic Systems Design Centre (ESDC), College of Engineering, Swansea University, Swansea SA1 8EN,UK.

Some important device structural parameters depicted in Fig.1.a include a gate length of L, a gate barrier of thickness t_b, a hetero-interface 2DEG sheet charge, an unintentionally doped GaN layer of thickness tGaN of doping N_d sitting on a semiinsulating (S.I.) substrate, and source and drain contact regions with parasitic resistances RS and RD. The channel-equivalent circuit is the core of the SPSC model as shown in Fig.1.b. It consists of N identical segmental channel resistors Ri (i=1,N) whose non-linear resistivity are the function of local node surface potentials $\psi_{s,i}$. The two voltage generators $\psi_{s,S}$ and $\psi_{s,D}$ shown in Fig.1.b denote the boundary surface potentials at the source and the drain ends of the channel, respectively, determined by the gate, source and drain biasing voltages V_S, V_{GS} and V_{DS}, respectively. The drain current I_{DS} is obtained from solving the channel equivalent sub-circuit with SPICE simulations. It can be copied via an ideal current source into the main sub-circuit model with necessary added AC or other parasitic external elements (resistance, capacitances and/or inductances) [14] for forming the final large-signal GaN HEMT as illustrated also in Fig.1.b.

2.1 Extracting the channel boundary potentials

Recently, Jana et al. [13] have derived the implicit surface potential equation used for calculating the channel surface potential ψ_s of AlGaN/GaN HEMTs with explicit appearance of the polarization charge σ_π. It was obtained from solving the Poisson's equation with charge sheet approximation and Boltzmann electron distribution function [16].

$$\frac{1}{2}\left(\frac{V_{gp}-\Psi_s}{t_b}\right)^2+\frac{q\sigma_\pi}{\varepsilon_s}\left(\frac{V_{gp}-\Psi_s}{t_b}\right)+\frac{q}{\varepsilon_s}\left[\begin{array}{l}\Psi_s N_d+\vartheta_t p_0\left(1-e^{-\frac{\Psi_s}{\vartheta_t}}\right)\\ -\vartheta_t N_d e^{-\frac{V_x}{\vartheta_t}}\left(e^{-\frac{\Psi_s}{\vartheta_t}}-1\right)\end{array}\right]=0 \quad (1)$$

where $V_{gp}=V_{GS}-V_p$ is the effective gate voltage, V_p is the channel cut-off voltage and V_x is the local channel potential under the influence of drain voltage V_{DS}. The meanings of other related symbols appearing in (1) are shown in Table I. The boundary surface potentials $\psi_{s,S}$ and $\psi_{s,D}$ are obtained in the SPSC model from SPICE simulations of the auxiliary sub-circuit (ASC) that is also shown in Fig.1.b. It consists of three non-linear voltage controlled current sources I_1, I_2 and I_3 expressed with respect to a voltage variable ψ_s as :

$$I_1=\frac{q}{\varepsilon_s}\vartheta_t N_d e^{-\frac{V_x}{\vartheta_t}}\left(e^{\frac{\Psi_s}{\vartheta_t}}-1\right),$$

$$I_2=\frac{q}{\varepsilon_s}\vartheta_t p_0\left(e^{-\frac{\Psi_s}{\vartheta_t}}-1\right), \quad (2)$$

$$I_3=\frac{1}{2}\left(\frac{V_{gp}-\Psi_s}{t_b}\right)^2+\frac{q\sigma_\pi}{\varepsilon_s}\left(\frac{V_{gp}-\Psi_s}{t_b}\right)+\frac{q}{\varepsilon_s}\Psi_s N_d$$

The functional expressions of I_1, I_2 and I_3 in (1) are obtained after re-grouping the additive terms of the implicit closed form equation (1). The two identical ASC are necessary to use in the SPSC model of Fig.1 for obtaining $\psi_{s,S}$ and $\psi_{s,D}$, with $V_x=V_S$ and $V_x=V_{DS}$ replaced in (1) for the source and the drain potential at the channel ends, respectively. The node potential ψ depicted in ASC of Fig.1 represents the solution of the surface potential equation, since the Kirkhoff's zero net current condition equality $I_1-I_2+I_3=0$ actually recovers the original equation (1).

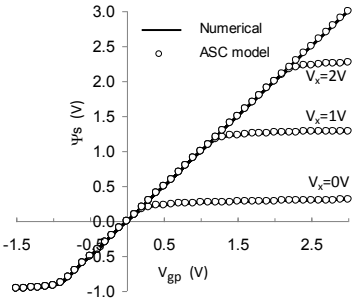


Fig. 2. Surface potential ψ_s as a function of the effective gate voltage V_{gp} for different local channel potentials V_x obtained with SPICE simulation of ASC and from numerically solving the original implicate equation [13].

Fig.2 shows an excellent agreement between ψ_s versus V_{gp} dependences obtained for different local channel potentials V_x with SPICE simulations of ASC and from numerically solving the implicate equation (1). The full overlap between the numerical and SPICE simulated curves as shown in Fig.2 validates the efficiency of ASC for extracting $\psi_{s,S}$ and $\psi_{s,D}$ in the SPSC model.

2.2 Modeling the 2DEG charge density n_s

The quantum-mechanical effects are not considered in deriving the original channel potential equation (1) for the sake of derivation simplicity [13]. In addition, the quantum corrections of 2DEG well potential has a minor effect on the output characteristics of AlGaN/GaN HEMTs due to

pre-dominance of polarization charge σ_π [17]. However, the quantization of 2DEG charge density n_s is useful to include in the SP -based HEMT models since it improves the overall device modeling accuracy [9,10,11,12]. A few complex analytical expressions of n_s with charge quantization have been reported recently [9,10,12,19] employing various interpolation functions between different regions of device operation. In contrast, a simple analytical relation between n_s and ψ_s is derived in this work and employed in the SPSC model that avoid possible convergence problems in SPICE simulations. Based on the compact analytical relation for 2DEG charge density at hetero-interface [18], an approximate implicit equation for calculating n_s is developed as shown in the Appendix (eq. A.5). It is repeated here as:

$$V_{gp} - \psi_s - \gamma_0 n_s^{\frac{2}{3}} - \frac{t_b q}{\varepsilon_{GaN}} n_s = \vartheta_t \ln\left(e^{-\frac{n_s}{D\vartheta_t}} - 1 \right) \quad (3)$$

where the meaning of physical parameters q, D, γ_0, ε and υ_t are listed in Table I.

The new implicit equation (3) shows that n_s is a complex function of V_{gp} and V_x due to a voltage dependence of ψ_s as can be observed in Fig.2. The values of ψ_s obtained for different V_{gp} and V_x from SPICE simulations of ASC as for Fig.2 are also used for numerically solving (3) over the variable n_s.

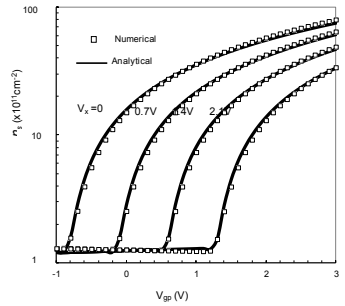


Fig. 3.a

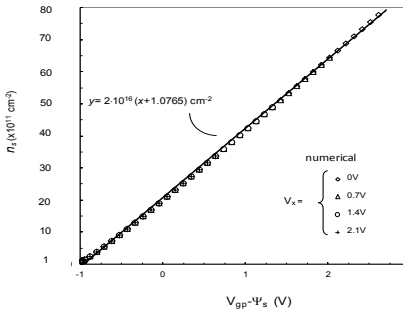


Fig. 3.b

Fig. 3. a) Comparison of the numerical calculations and the linear analytical model (Fig.3.b) of n_s versus the gate effective voltage V_{gp} for different local channel potentials V_x. b) A unique linear dependence of n_s versus the effective channel surface potential (V_{gp}-ψ_s) regardless of different local channel potential V_x.

Fig.3.a shows a non-linear behavior of n_s with V_{gp} obtained from numerically solving (3) for different constant values of V_x. However, when re-plotting the same numerical $n_s(V_{gp})$ curves from Fig.3.a, but now against the voltage difference (V_{gp}-ψ_s) on the x-axes, a unique linear dependence of n_s is obtained regardless of V_x as shown in Fig.3.b. It can be well approximated with the function $n_s = 2 \times 10^{12} (V_{gp} - \psi_s + 1.0765)$ cm $^{-2}$ depicted by solid line in Fig.3.b. The accuracy of the later linear approximation can be additionally validate with the results in Fig.3.a showing an excellent agreement between the numerical and the analytical values of n_s obtained with linear approximation from Fig.3.b. Following these results, the charge density $n_{s,I}$ of the i-th segment of the SPSC model can now be expressed as:

$$n_{s,i} = a\left(V_{gp} - \psi_{sm}\right) + b, \quad (4)$$

where a and b become new model fitting parameters. The mid-potential $\psi_{sm}=(\psi_{s,i}+\psi_{s,i-1})/2$ of the i-th segment is employed in (4) in order to improve the simulation accuracy of SPSC model.

2.3 Expressing the segmental resistance R_i

Using the charge-sheet approximation [16], the channel segmental resistance R_i appearing in Fig.1 can be analytically expressed as:

$$R_i = \frac{w_i}{W q \mu_n n_{s,i}} \quad (5)$$

where μ_n is the channel carrier's mobility, $w_i = L/N$ is the resistor length, L is the gate length, W is the channel width and q is the electron charge. The longitudinal and vertical electric filed dependences of the electron mobility in 2DEG are modeled in (4) by the following approximate expressions [10]:

$$\mu_n\left(E_x\right) = \frac{\mu_{LF} E_x}{1 + \dfrac{u_a \left| E_x \right|}{E_T}}, \qquad \mu_{LF}\left(E_{y.eff}\right) = \frac{\mu_0}{1 + p_1 \left| E_{y.eff} \right| + p_2 E_{y.eff}^{\,2}}, \quad (6)$$

where μ_{LF} is the low longitudinal field mobility, μ_0 is the low vertical field mobility, E_x is the longitudinal electric field along the channel, E_T is the critical electric field, $E_{y,\mathrm{eff}}$

denotes the effective vertical field in the GaN layer, and u_a p_1, p_2 are the fitting parameters. The mobility formulas (6) holds for relatively low values of V_{DS} e.g as long as $E_x < E_T$. Following the average-potential approximation [9, 10], the electric fields E_x and $E_{y,\text{eff}}$ are easily expressed in the SPSC model as:

$$E_x = \frac{\psi_i + \psi_{i-1}}{w_i}, \tag{7}$$

$$E_{y,eff} = \varepsilon_s \frac{V_{gp} - \frac{\psi_{s,S} + \psi_{s,D}}{2}}{d\varepsilon_{GaN}}, \tag{8}$$

where ε, t_b and ε_{GaN} are physical parameters with meanings listed in Table I.

Finally, for obtaining realistic I-V characteristics that reasonably agree with measured output curves of fabricated GaN HEMTs, the channel length modulation (CLM) phenomena and the self-heating effects (SHE) must be taken into account since they largely influence the device output current I_{DS}. These important physical effects are included in the SPSC model with semi-empirical relations modifying (4) as:

$$R_i = \frac{w_i\left(1 - \lambda V_{DS}^m\right)}{Wq\mu_n n_{s,i}}\left(1 + \beta \cdot \left|V_{gp} V_{DS}\right|\right). \tag{9}$$

The CLM phenomena is expressed in (9) with simple term $(1 - \lambda \cdot V_{DS}^m)$ modulating w_i,where λ and m are fitting parameters [6]. The SHE is also included in (9) by adding a new multiplicative term $(1+\beta \cdot |V_{gp} \cdot V_{DS}|)$, with β as fitting parameter. The latter effectively increases R_i at higher V_{gp} and/or V_{DS} causing the decrease of I_{DS} when the device operates at higher power level. A highly simplified SHE model implemented in (9) is based on the assumption that SHE are dominantly expressed as degradation of the mobility of 2DEG electrons owning to a local increase of channel temperature [20]. Note that a drain-induced barrier lowering (DIBL) effect important for highly-scaled GaN HEMTs, is not considered in this work since the referent experimental device was with relatively long gate (L=1μm). However, the DIBL effect can be easily included in the SPSC model with any suitable analytical expression describing the shift of channel cut-off voltage V_p with V_{DS} and/or L [25].

III. Results and Discussion

Fig.4.a and Fig.4.b compares the simulated I_{DS}-V_{DS} and I_{DS}-V_{GS} characteristics of conventional $Al_{0.25}Ga_{0.75}N$/GaN HEMT against the experimental DC measurements [26]. The numerical values of all model parameters used in SPICE simulations are shown in Table I. A fairly good matching between simulated and measured [26] characteristics is observed in Fig.4 in spite of using a highly simplified expressions for CLM and SHE effects in the SPSC model. To further improve modelling accuracy, the influence of access regions parasitic resistances (Fig.1.a) is included in SPICE simulations by adding two fixed resistors R_s and R_d in the external circuit as shown in Fig.1.b.

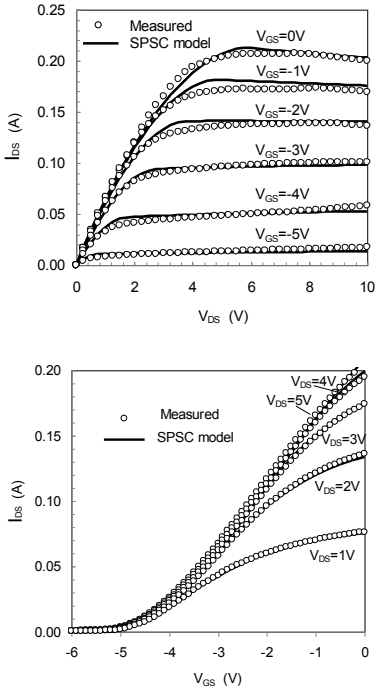


Fig. 4. Compares of the modelled: (a) I_{DS}-V_{DS} and (b) I_{DS}-V_{GS} characteristics with experimental data of $Al_{0.25}Ga_{0.75}N$/GaN HEMT with $L=1\mu m$ [26]

The Gummel symmetry test [24] is also performed to validate the correctness of the SPSC model with respect to drain/source symmetry. A voltage source V_m is applied at the drain terminal, whereas $-V_m$ is applied at the source terminal. V_m is varied from −1 to 1 V . Fig.5 shows the first derivative of I_{DS} versus V_m extracted for various V_{GS} from SPICE simulations. The results in Fig. 4 shows that the SPSC model passes the basic test condition of the symmetrical and continuous first derivative of I_{DS} with respect to V_m and at V_m=0, respectively [24].

Fig. 6 shows the dependence of relative numerical error of I_{DS} and g_m versus the number of channel segments N used in simulations with the SPSC model. The relative error is defined here as the difference between the values of I_{DS} and g_m obtained for N <10 against those simulated with N=10, e.g. rel.err.=$(I_{DS|\,N<10} - I_{DS|\,N=10})/I_{DS|\,N=10}$x100%.

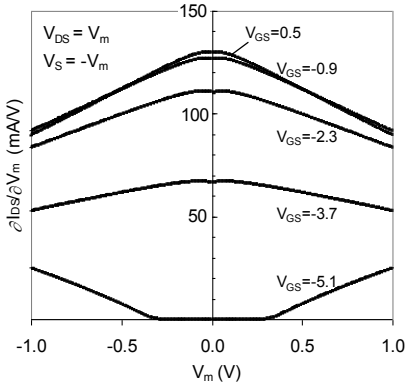


Fig. 5. Gummel symmetry characteristics of the SPSC model simulated for different V_{GS} and V_m= -1 to 1V, 0.01V steps, $V_{DS} = V_m$ and $V_{DS} = V_m$

In the example of GaN HEMT with L=1um, Fig.6 shows that only N=6 is sufficient to use in the SPSC model for achieving the relative numerical error smaller than 1%.

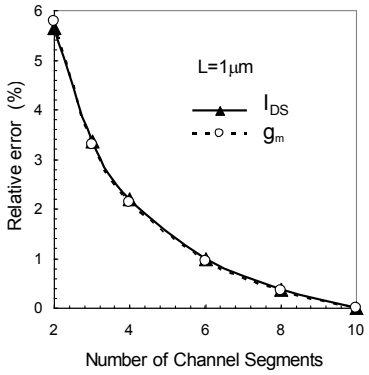


Fig. 6

Fig. 6. Relative numerical error of I_{DS} and g_m versus the number of channel segments N used in SPICE simulations with the SPSC model.

A steep decrease of relative error of I_{DS} and/or g_m with increasing N, which is observed in Fig.6, also indicates the high efficiency of SPSC model with respect to fast numerical converging and short CPU time required in SPICE simulation.

IV. Conclusions

In this paper, the surface-potential-based sub-circuit (SPSC) model of AlGaN/GaN HEMT yielding continuous I-V characteristics in the whole operational range is described and implemented in SPICE. It allows for easy inclusion of important physical phenomena such as electron velocity saturation, high-field mobility degradation, channel length modulation and self-heating effects and avoids the need for the derivation of complex analytical expressions in compact modeling. A good fit of the SPSC model with experimental I-V characteristics of $Al_{0.25}Ga_{0.75}N$/GaN HEMT is demonstrated with SPICE simulations.

TABLE I
PARAMETERS USED IN SPICE SIMULATION

Simbol	Description	Value
L	Gate length	1μm
W	Gate width	400 μm
N	Number of segments	10
V_{off}	Cut-off voltage	-5.9V
R_S	Source contact resistance	0.45 Ω
R_D	Drain contact resistance	1 Ω
t_b	Thickness of the gate barrier layer	33nm
σ_π	Polarization sheet charge at the heterointerface	$1e^{17}$ m^{-2}
n_i	Intrinsic carrier concentration	$2.9e^{-4}$ m^{-3}
N_d	Doping of unintentional doped GaN	$1e^{22}$ m^{-3}
D	Density of states at the conduction band edge	$1.001e^{18}cm^{-3}$
q	Electron charge	$1.609e^{-19}$ C
υ_t	Thermal voltage	0.0259 V
ε_s	Permitivity of AlGaN barrier layer	$7.965e^{-11}$
ε_{GaN}	Permitivity of GaN layer	7.88 e^{-11}
E_T	Critical electric field	1.68 e^7 V/m
γ_0	Experimentally determined parameter	$2.12e^{12}$ $m^{4/3}$
a, b	Fitting parameters of the n_s model	$2e^{16}$ $V^{-1}cm^{-2}$, $2.135e^{16}$ $V^{-1}cm^{-2}$
λ , m	SCE model parameters	0.105 V^{-1}, 0.5
β	SHE model parameter	9.4e-3
μ0	Low field mobility	$3.2e^{-2}m^2V^{-1}s$
u_a , p_1, p_2	Mobility degradation parameters	2.1, $7e^{-9}$ mV^{-1} $3.9e^{-17}m^2V^{-2}$
p_0	n_i^2/N_d, The equilibrium hole charge of GaN	
n_s	The density of the 2DEG	
E_f	Position of Fermi level	
E_0	Position of first energy level	
μ_0	Low vertical field mobility	
$E_{y,eff}$	Effective vertical electric field	
E_x	Longitudinal electric field	

V. APENDIX I

Here the derivation of (2) is explained in more detail. Based on the self consistent solution of Schrodinger's and Poisson's equations in the wide band-gap semiconductor, the analytical relation of 2DEG charge density for HEMTs has been derived using the triangular quantum well approximation and considering only the two lowest subband energy levels E_1 and E_0 [18,22]. Since at the AlGaN/GaN hetero-interface, the upper energy level E_1 is larger than the Fermi energy E_f for the complete range of V_{GS} and because E_1 is significantly larger than E_0 over the same range ($E_1 \approx 3E_0$), the electron contributed by E_1 and higher energy band to n_s can be safely ignored [10,12]. Then, the analytical expression for 2DEG charge density [18,22] can be simplified as:

$$n_s = Dv_{th}\left[\ln\left(e^{\frac{E_f - E_0}{v_{th}}} + 1\right)\right] \tag{A.1}$$

where D is the 2DEG density of states. The dependence of lowest energy sub-band E_0 versus n_s is approximated by [21]:

$$E_0 = \gamma_0 n_s^{\frac{2}{3}} \tag{A.2}$$

where γ_0 is the experimentally determined parameter with value shown in Tab I. Assuming that the AlGaN layer is completely ionized, the following relation exists:

$$n_s = \frac{\varepsilon_{GaN}}{qt_b}\left(V_{gp} - \psi_s - E_f\right) \tag{A.3}$$

From (A.3) , E_f is expressed as.

$$E_f = \frac{qt_b}{\varepsilon_{GaN}} n_s - V_{gp} + \psi_s \tag{A.4}$$

Replacing E_0 and E_f from (A.2) and (A.4) into (A.1), the implicit equation for calculating n_s with ψ_s as variable is obtained as:

$$V_{gp} - \psi_s - \gamma_0 n_s^{\frac{2}{3}} - \frac{qt_b}{\varepsilon_{GaN}} n_s = \vartheta t \ln\left(e - \frac{n_s}{D\vartheta t} - 1\right) \tag{A.5}$$

REFERENCES

[1] Johnson, W., Piner, E.L., *"GaN HEMT Technology in GaN and ZnO-Based Materials and Devices"*, edited by Pearton, S.J., Springer, Heidelberg, 2012, pp. 209-238.

[2] Chen, Q., "*Latest advances in gallium nitride HEMT modeling*", 12th IEEE International Conference on Solid-State and Integrated Circuit Technology (ISCIST), 2014, pp: 1-4.

[3] Shanthi, J., Vimala, P., Shanthi, J. Vimala, P, "*High Electron Mobility Transistor: A Review on Analytical Models*", International Journal for Innovative Research in Science & Technology (IJIRST), Vol. 3, 2016, pp: 103-114.

[4] Paydavosi N., Venugopalan S., Chauhan Y.S., Duarte J.P., Jandhyala S., Niknejad A.M., Hu C.C., "*BSIM-SPICE Models Enable FinFET and UTB IC designs*", *IEEE Access*, vol. 1, pp. 201–215, 2013.

[5] Jankovic N., Pesic T., "*Non-quasi-static physics-based circuit model of fully-depleted double-gate SOI MOSFET*", Solid-State Electronics, Vol. 49, 2005, pp: 1086-1089.

[6] Pesic T., Jankovic N., "*A Compact Non-Quasi-Static MOSFET Model based on the Equivalent Non-Linear Transmission Line*", IEEE Trans. Computer-Aided-Design Integrated Circ. Systems, Vol. 27, 2005, pp: 1550-1561.

[7] Khandelwal S., Chauhan Y.S., Fjeldly T.A., "*Analytical Modeling of Surface-Potential and Intrinsic Charges in AlGaN/GaN HEMT Devices*", IEEE Trans. Electron Devices, Vol. 59, 2012, pp: 2856–2860.

[8] Khandelwal S., Fjeldly T.A., "*A surface-potential-based drain current model for study of non-linearities in AlGaAs/GaAs HEMTs*", in Proc.Compound Semicond. IC Symp., 2012, pp: 1-4.

[9] Khandelwal S., Fjeldly T.A., "*A physics based compact model of I–V and C–V characteristics in AlGaN/GaN HEMT devices*", Solid-State Electronics, Vol. 76, 2012, pp: 60-66.

[10] Cheng, X.,Wang, Y., "*A surface-potential-based compact model for AlGaN/GaN MODFETs*", IEEE Transactions on Electron Devices, Vol. 58, 2011, pp: 448-454.

[11] Patrick M., Hahe R. HSP, "*A surface-potential-based compact model of AlGaN/GaN HEMTs power transistors*", 10th MOS-AK/GSA ESSDERC/ESSCIRC Workshop Bordeaux, 2011.

[12] Khandelwal S., Goyal, N., Fjeldly T.A., "*A physics-based analytical model for 2DEG charge density in AlGaN/GaN HEMT devices*", IEEE Transactions on Electron Devices, Vol. 58, 2011, pp: 3622-3625.

[13] Jana R., Jena D., "*A surface-potential based compact model for GaN HEMTs incorporating polarization charges*", IEEE 70th Annual Device Research Conference, 2012, pp: 147-148.

[14] Jun, L., Zhiping, Y., Lingling, S., "*A complete and accurate surface-potential based large-signal model for compound semiconductor HEMTs*", Journal of Semiconductors, Vol. 35, 2014.

[15] Khandelwal, S., Yadav, C., Agnihotri, S., Chauhan, Y.S., Curutchet, A., Zimmer, T., Jean- Jaeger D.C.,

Defrance N., Fjeldly, T.A., "*A. Robust surface-potential-based compact model for GaN HEMT IC design*", IEEE Transactions on Electron Devices, Vol. 60, 2013, pp: 3216-3222.

[16] Brews J. R., "*A charge-sheet model of the MOSFET*", Solid-State Electronics, Vol. 21, 1978, pp: 345–355.

[17] Yamakawa S., Goodnick S., Aboud S., Saraniti M., "*Quantum corrected full-band cellular Monte Carlo simulation of AlGaN/GaN HEMTs*", J. Comput. Electron, Vol. 3, 2004, pp: 299–303.

[18] Lee K., Shur M.S., Drummond T.J., Morkoc H., "*Electron density of the two dimensional electron gas in modulation doped layers*", J.Appl. Phys, Vol. 54, 1983, pp: 2093–2096.

[19] Pu J., Sun J., Zhang D., "*An accurate polynomial-based analytical charge control model for AlGaN/GaN HEMT*", Semiconductors, Vol. 45, 2011, pp: 1205–1210.

[20] Darwish A., Bayba A.J., Hung, H.A., "*Channel temperature analysis of GaN HEMTs with nonlinear thermal conductivity*", IEEE Transactions on Electron Devices, Vol. 62, 2015, pp: 840-846.

[21] Delagebeaudeuf D., Linh N.T., "*Metal-(n) AlGaAs-GaAs two-dimensional electron gas FET*", IEEE Transactions on Electron Devices, Vol. 29, 1982, pp: 955-960.

[22] Kola, S., Golio, J. M., Maracas, G. N., "*An analytical expression for Fermi level versus sheet carrier concentration for HEMT modeling*", IEEE Electron Device Letters, Vol. 9, 1988, pp: 136-138.

[23] Cadence Design Systems, "*ORCAD PSPICE A/D Reference Manual*", Version 9.2, 2000.

[24] McAndrew C.C., "*Validation of MOSFET model for source/drain symmetry*", IEEE Trans. Electron Devices, Vol. 53, 2006, pp: 2202–2206.

[25] Kumar, S.P., Agrawal, A., Chaujar, R., Kabra, S., Gupta, M., Gupta, R.S., "*Threshold voltage model for small geometry AlGaN/GaN HEMTs based on analytical solution of 3-D Poisson's equation*", Microelectronics Journal, Vol. 38, 2007, pp: 1013-1020.

[26] S. Faramehr, P. Igić and K. Kalna., "*Modelling and optimization of GaN capped HEMTs*", The Tenth International Conference on Advanced Semiconductor Devices and Microsystems, Smolenice, 2014, pp: 1-4.

A New Simplified Spice Modelling of Memristor

Miljana Milić, Miljan Petrović

Abstract - In this paper we will introduce a new memristor model for Spice simulations that is computationally very inexpensive. It has four adjustable parameters, including low and high resistances and switching voltage threshold, and perfectly expresses memristive behaviour during simulations in transient domain. It is verified in LTspice simulations and shows good characteristics.

Keywords – Hysteresis, Memristor, Modelling, Simulation, Spice.

I. Introduction

Memristor as a component was first introduced by Leon Chua in 1971. as the fourth fundamental passive circuit element besides, resistors, capacitors and inductors [1]. It is characterized by the new physical parameter that is referred to as memristance, which creates dependence between the charge and the flux of the device. It took thirty-seven years after this prediction, for Hewlett Packard team lead by Stanley Williams to develop a first functioning solid-state memristor [2,3].

The relations between basic fundamental circuit elements and all electromagnetic quantities are shown in Fig.1.

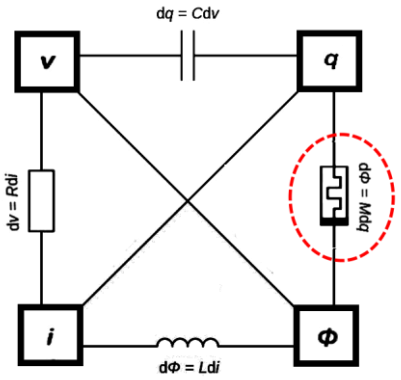


Fig. 1. Illustration of memristor as the missing circuit element

The research in the field of memristive devices has a significant importance due to a large potential of its applications in different areas of science: electronic, electrical, computational, bioengineering, neural etc.

Currently there are many unsolved questions related to the production of memristor, its modelling, simulating in a circuit and its applciations. All these problems depend on each other.

In this paper we will try to introduce a novel and a very simple Spice model of the memristor device, that can be used in a simulation of some simple circuit at first ie. for educational purposes. There are many models of memristor available in literature. There are few aspects of memristor modelling in which all these models compete: accuracy, complexity, possibilty to control threshold, possibility to be applied in different domains of circuit analysis, etc.

In order to develop a new model of the memristor and apply it into some novel application, it is necessary to have a good theoretical and applicative knowledge of the device. This will be done in the following section where we will give a brief overview of the memristor structure, physical, electrical and matematical behaviour, implementations, currently avaliable models and many different areas of applications. The third section describes a new Spice model of memristor. After that, we will verify the model in time domain simulations. The paper will be concluded with the ideas of its further development and applications.

II. Memristor Background

The first HP memristor was made of Titanium Dioxide creating two chemically different layers: one with high impedance and other that is conductive [4]. They were placed between two platinum electrodes. This is shown in Fig. 2. Titanium Dioxide normally behaves as an insulator, unless oxygen vacancies are introduced. Then it behaves as an n-type semiconductor. These oxygen vacancies affect the conductivity of the TiO_2, since they can drift when applied to varying electric field [5].

The mathematical model of HP memristor has two resistors connected in series, *Roff* and *Ron*, representing the undoped and doped fields of the component, respectively. It can be seen from Fig. 2 that the width of the memristor is denoted with *D*, while the width of the doped region is denoted with *w*. The width *w* represents a variable that changes with the charge [5].

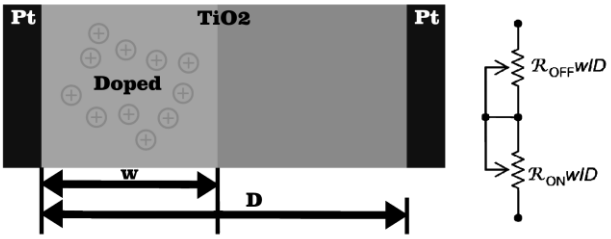


Fig. 2. HP memristor structure and its equivalent circuit

Miljana Milić and Miljan Petrović are with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia, E-mail: miljana.milic@elfak.ni.ac.rs, miljan.petrovic@elfak.ni.ac.rs.

The equivalent circuit for described structure is also shown in Fig. 2, while the corresponding *i-v* function for such a circuit is expressed with the two following equations:

$$v(t) = [R_{ON}\frac{w(t)}{D} + R_{OFF}(1-\frac{w(t)}{D})]\cdot i(t) \quad (1a)$$

$$\frac{dw}{dt} = \frac{\mu_v R_{ON}}{D} i(t) \quad (1b)$$

where μ_v denotes the mobility of ions [4, 5].

After the first Titanium Dioxide memristor, alternative materials were introduced for its implementation over last few decades. In 1963., a memristor made of ferroelectric material appeared [6]. In 2004. a polymeric (ionic) memristor was proposed [7]. A memristive behaviour occurred during the application of a resonant-tunnelling diode [8, 9]. Alternative memristor materials included also Graphene Oxide [10] and Silicon Oxide [11], while one of the alternative implementation is the spintronic memristive system described in [12].

Since its first appearance, many researchers were trying to demonstrate possible new applications of the new device [4]. All these applications can be categorized in a few groups. The first of them represents analogue circuits such as oscillators [13], programmable amplifiers [14], filters [15] and PLLs [16]. The second group is related to the neuromorphic circuits [17]. Chaotic systems are the fourth group of applications [18], whereas the last group incorporates the applications in digital circuits such as memories [19] and implementation of logic gates [20].

There are numerous models of memristor designed to emphasize different characteristics of the component. Some take into account the physical structure of the memristor [21, 22], and some exhibit only the memristor's functionality using generic Spice components [23, 24]. Although most of solutions are based on nonlinear modelling of memristor's characteristics [25, 26, 27], in this paper a linear model is proposed. Namely, memristor is modelled as a two-state resistor with ability of instant switching. This is useful for initial circuit simulation, especially for digital systems, such as memories.

III. SPICE MODEL

The memristor model described in this paper is a simplified model which provides a memristive behavior to circuit simulations. It does not take into account the physics and nanoelectronics of the real memristor component. Instead, it behaves as a two-state resistor and it is designed accordingly.

The schematic of the model is given in Fig. 3. As it can be seen, it consists primarily of behavioural voltage sources and some digital circuitry. Four parameters are defined as externals, meaning that their values are adjustable in the circuit which incorporates memristor component (they can vary from instance to instance). Those are *Ron*, low resistance of the memristor, *Roff*, high resistance of the memristor, *Vlim*, voltage threshold for switching between states, and *ton*, clock interval for the flip-flop which checks for the state of the memristor in the previous time instant.

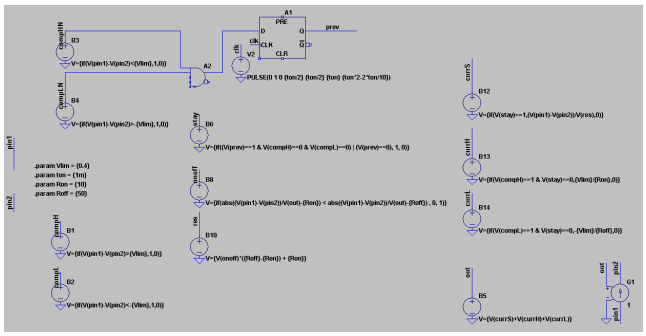


Fig. 3. Schematic of the described memristor model

Two nodes (*pin1* and *pin2*) represent external pins of the component. The memristor is modelled as voltage--controlled current source. However, all the behavioural sources in Fig. 3 are voltage sources, since Spice simulator does not allow nodes connected to current sources to "float". It is only after all the calculations, that the voltage-controlled current source (G1) with gain equal to 1 converts the output voltage into a current between nodes *pin1* and *pin2*.

Behavioural sources B3 and B4 function as comparators. They inspect whether the input voltage across the memristor is lower than the higher limit (*Vlim*), and higher than the lower limit (*–Vlim*, since the memristor *i-v* characteristic is modelled to be centrally symmetrical), and produce 1V output, accordingly. These two signals are passed through an AND circuit and fed into a D flip-flop controlled by an internal clock with period *ton*. On the other hand, behavioural sources B1 and B2 are comparators with slightly different functions, and they produce 1V if voltage across memristor is higher than *Vlim*, and lower than *–Vlim*, respectively, thus, denoting the current voltage across the memristor. Combining information from these comparators and "historical" information from the flip-flop, behavioural source B6 generates the voltage at a node *stay*, which is equal to 1V, only if it is not necessary (depending on the voltage conditions) to change the memristor's resistive state.

Further, the current resistance of the memristor is estimated with a behavioural source B8. It checks if the current resistance is closer to the value defined by *Ron*, or *Roff*, and sets the voltage at node *onoff* to 1V if the memristor is in the on state, and to 0V if the memristor is in the off state. The source B10 "translates" the information about the state (0 or 1 volts) to the voltage resistance. Namely, voltage at node *res*, is equal to *x* volts if the current resistance (*Ron* or *Roff*) equals *x* ohms.

Similarly, three voltage sources B12, B13, and B14 generate voltages at nodes *currS*, *currH*, and *currL*, respectively, and those are numerically equal to the appropriately generated currents. Source B12 gives the "current" equal to the ratio of voltage across the memristor and current resistance *res*, only if the component keeps its state. Sources B13, and B14 produce "current" if the state should be switched from high to low resistance (B13) or vice versa (B14). The "currents" are then limited as if they are produced by voltages *Vlim*, and *–Vlim*, in order to avoid extremely low or high values of current and voltage at switching time instances in simulation.

Finally, these "currents", modelled as voltages, are summed in the behavioural source B5, and, as it was aforementioned, converted into a current using component G1. Behaviour of the model in time domain is analysed in the following section of this paper.

III. SIMULATION

The memristor component is modeled in Spice as described in the previous section of this paper. Now, the model should be verified with a simple testbench. The circuit is fed with several different voltage waveforms, and the obtained behavior of the memristor is then analyzed.

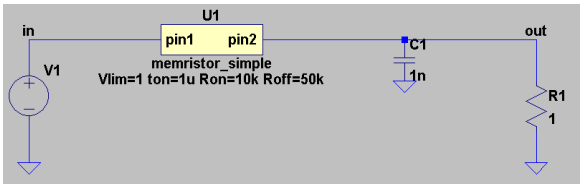


Fig. 4. Schematic of the testbench for memristor model simulations

The schematic of the testbench is shown in Fig. 4. It incorporates a voltage source V1 at the circuit input, a resistive load (1 Ω) at the output, while the main component i.e. the memristor is instantiated as the subcircuit U1. It is necessary to include a capacitor C1 (its initial value is 1nF) in order to slow down the ideal, instantaneous switching, controlled by behavioural sources inside the memristor, since that switching can lead to infinitesimal time step of the transient simulation, and further to a simulation crash. Values of memristor parameters are chosen as follows: voltage threshold is 1V, on and off resistances are 10kΩ and 50kΩ, respectively, and clock period is 1μs.

The first examined response of memristor is the response to the triangular input voltage. Since memristor's switch voltage was set to 1V, the amplitude of the input voltage is set to 1.005V. Four periods of the signals are plotted in Fig. 5. Voltage across the memristor is plotted with black line, and current through memristor with grey line. It can be seen that the slope of the current waveform alternates between lower and higher one. This is due to the switching of resistance that happens at each moment when the voltage signal reaches above 1V or below –1V, except at the first peak of 1.005V. The first voltage peak does not switch the memristor's state because the model still does not "know" what its previous state was. With the next ramp the memristor starts functioning. Further, resistance of memristor during this transient analysis is shown in Fig 5. It is obvious that the periods where resistance is either 10kΩ or 50kΩ correspond to the piece-wise linear parts of the triangular voltage signal.

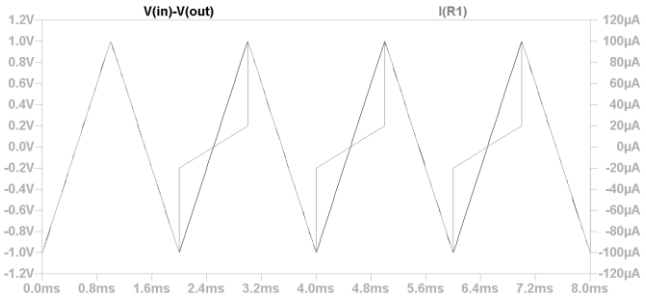


Fig. 5. Voltage across memristor and current through it for triangular voltage excitation

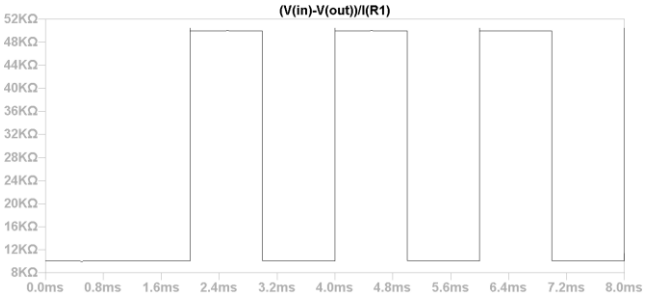


Fig. 6. Resistance of memristor for triangular voltage excitation

In order to get the current-voltage (*i-v*) characteristic of the memristor, one cannot use DC sweep, because of the presence of digital circuitry that memorize the previous state of the device model. Hence, this can be done only in transient analysis. Triangular input voltage signal is the appropriate one for it. Thus, we get the *i-v* characteristic as in Fig. 7. The well-known hysteresis of memristor can be recognized. Switching pieces of characteristic are close to ideal in the sense of infinite derivative (vertical lines in Fig. 7).

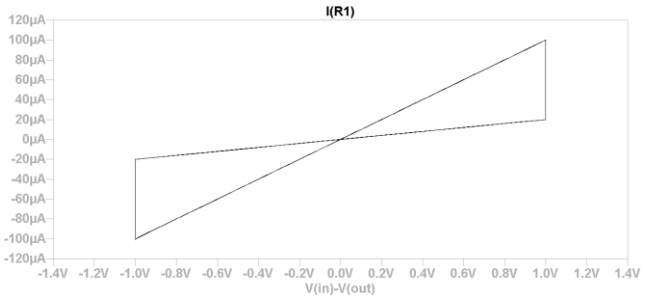


Fig. 7. I-V characteristic of memristor model

Sine and pulse voltage excitation are also applied in the testbench. Input voltage and output current of memristor for the sine wave input can be seen in Fig. 8. Input voltage and output current of memristor for the pulse wave input can be seen in Fig. 9. Amplitude of sine voltage is 1.005V, and the pulse wave switches between 1.005V and −1.005V.

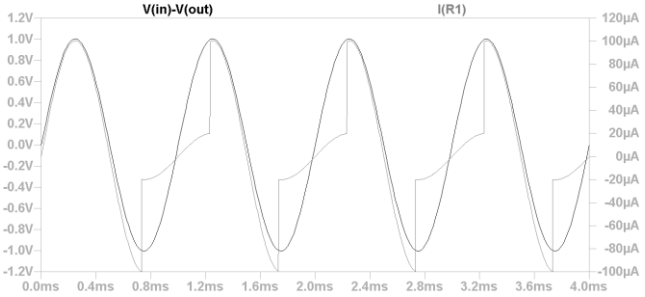


Fig. 7. Voltage across memristor and current through it for sine voltage excitation

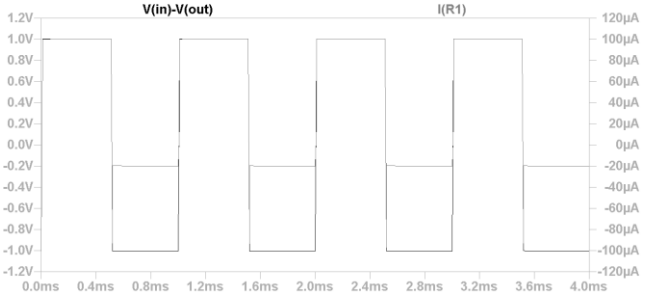


Fig. 8. Voltage across memristor and current through it for pulse wave voltage excitation

Observing the obtained output current waveforms leads to several conclusions. First, glitches are seen at the points of switching from low resistance to high resistance and not from low to high resistance, for both excitations, but more frequently and with higher amplitude in the case of pulse wave. In the case of square pulse input, one can notice that the memristor starts functioning earlier, i. e. switches its resistive state at the first change of voltage level (from higher to lower), but this is only because at the beginning of the transient analysis, the voltage rises from 0V (initial condition) to the set high level of 1.005V, so there is actually another change of voltage that should but does not induce a change of resistance.

IV. Conclusion

In this paper a new and simplified Spice model of memristor is introduced. It relies on functionality of an ideal memristor as a two-state (resistance) switch. It is parameterised by both on and off resistances and the switching voltage threshold, which is immensely useful for behavioural analysis in the initial phases of the circuit design. Also, the model accurately expresses the basic characteristics of a memristor component, and can be used for educational purposes. A disadvantage of the model is identified as a possibility to induce a significantly slow transient analysis in a more complex design, because of the flip-flop based circuitry.

Further research regarding this topic would be the simulation and the verification of some circuits incorporating this memristor model. A good start would be a simple analogue circuit, such as MC (memristor-capacitor) filters, or more complex circuits such as programmable analogue amplifier. Further, the model could be made more sophisticated if flip-flops which slow down the transient analysis are to be replaced with a faster alternative.

Appendix

The netlist of the Spice model of memristor described in this paper is as follows:

```
.SUBCKT memristor_simple pin1 pin2
+PARAMS: Vlim=0.4 ton=1m Ron=10
+Roff=50
Gout pin1 pin2 out 0 1
B1 compH 0 V={if(V(pin1)-
+V(pin2)>{Vlim},1,0)}
B2 compL 0 V={if(V(pin1)-V(pin2)<-
+{Vlim},1,0)}
A1 N001 0 clk 0 0 0 prev 0 DFLOP
V2 clk 0 PULSE(0 1 0 {ton/2} {ton/2}
+{ton} {ton*2-2*ton/10})
B3 compHN 0 V={if(V(pin1)-
+V(pin2)<{Vlim},1,0)}
B4 compLN 0 V={if(V(pin1)-V(pin2)>-
+{Vlim},1,0)}
A2 compHN compLN 0 0 0 0 N001 0 AND
B6 stay 0 V={if((V(prev)==1 &
+V(compH)==0 & V(compL)==0) |
+(V(prev)==0), 1, 0)}
B8 onoff 0 V={if(abs((V(pin1)-
+V(pin2))/V(out)-{Ron}) <
+abs((V(pin1)-V(pin2))/V(out)-{Roff})
+, 0, 1)}
B10 res 0 V={V(onoff)*({Roff}-{Ron})
++ {Ron}}
B12 currS 0 V={if(V(stay)==1,
+(V(pin1)-V(pin2))/V(res),0)}
B13 currH 0 V={if(V(compH)==1 &
+V(stay)==0,{Vlim}/{Ron},0)}
B14 currL 0 V={if(V(compL)==1 &
+V(stay)==0,-{Vlim}/{Roff},0)}
B5 out 0 V={V(currS)+V(currH)+
+V(currL)}
.ENDS memristor_simple
```

Acknowledgement

This research is funded by The Ministry of Education and Science of Republic of Serbia under contract no. TR32004.

REFERENCES

[1] Zhu, G., Singh, C., *"Improving students' understanding of quantum measurement. II. Development of research-based learning tools"*, Phys. Rev. ST Phys. Educ. Res, Vol. 8, No. 1, Apr., 2012, pp. 010118-1-8.

[2] Addee, S., *"Without Chua's circuit equations, you can't make use of this device,' says Williams"*, IEEE Spectrum, May, 2008, Archived from the original on 2008-05-11.

[3] Johnson, R. C., *"Missing link' memristor created: Rewrite the textbooks?"*, EE Times, 30.04.2008.

[4] Radwan, A. G., Fouda, M. E., "*Memristor: Models, Types, and Applications*", in On the Mathematical Modeling of Memristor, Memcapacitor, and Meminductor. Studies in Systems, Decision and Control, vol 26. Springer, 2015.

[5] Chawan, P. P., Gupta, P., *"A Review on Memristor Applications"*, International Journal of Electronics, Circuits, systems and signal processing, Vol. 32, No. 3, 2013, pp. 1013-1023.

[6] Moll, J., Tarui, Y., *"A new solid-state memory resistor"*, IEEE Trans Elect Dev, Vol. 10, No. 5, 1963, pp. 338.

[7] Kriegerand, J., Spitzer, S. M., *"Non-traditional, non-volatile memory based on switching and retention phenomena in polymeric thin films"*, Non-Vol. Mem. Tech. Symposium, IEEE, 2004, pp. 121-124.

[8] Buot, F., Rajagopal, A., *"Binary information storage at zero bias in quantum-well diodes"*, J. Appl. Phys., Vol. 76, No. 9, 1994, pp. 5552.

[9] Yilmaz, Y., Mazumder, P., *"Programmable quantum-dots memristor based architecture for image processing"*, 12th IEEE Conf Nanotech, 2012, pp. 1-4.

[10] Jeong, H. Y., Kim, J. Y., Kim, J. W., Hwang, J. O., Kim, J. E., Lee, J. Y., Yoon, T. H., Cho, B. J., Kim, S. O., Ruoff, R. S., Choi, S. Y., *" Graphene oxide thin films for flexible nonvolatile memory applications "*, Nano Lett, Vol. 10, No. 11, Nov., 2010, pp. 4381-4386.

[11] Mehonic, A., Cueff, S., Wojdak, M., Hudziak, S., Jambois, O., Labbé, C., Garrido, B., Rizk, R., Kenyon, A. J., *"Resistive switching in silicon suboxide films"*, J App. Phys., Vol. 111, No. 7, Feb., 2012, pp. 074507.

[12] A. Chanthbouala, A., Matsumoto, R., Grollier, J., Cros, V., Anane, A., Fert, A., Khvalkovskiy, A. V., Zvezdin, K. A., Nishimura, K., Nagamine, Y., Maehara, H., Tsunekawa, K., Fukushima, A., Yuasa, S., *"Vertical-current-induced domain-wall motion in MgO-based magnetic tunnel junctions with low current densities"*, Nature Physics, Vol. 7, No. 8, Mar., 2011, pp. 626-630.

[13] Zidan, M. A., Omran, H., Radwan, A. G., Salama, K. N., *"Memristor-based reactance-less oscillator"*, Elec Lett, Vol. 47, No. 22, Oct., 2011, pp. 1220-1221.

[14] Shin, S., Kim, K., Kang, S. M., *"Memristor-based fine resolution programmable resistance and its applications"*, Int Conf. on Commun. Circ. Syst, July, 2009.

[15] Driscoll, T., Quinn, J., Klein, S., Kim, H. T., Kim, B. J., Pershin, Y. V., Di Ventra, M., Basov, D., N., *"Memristive adaptive filters"*, Appl. Phys. Lett., Vol. 97, No. 9, Aug., 2010, pp. 093502-1-3.

[16] Zhao, Y. B., Tse, C.K., Feng, J. C., Guo, Y. C., *"Application of memristor-based controller for loop filter design in charge-pump phase-locked loops"*, Appl. Phys. Lett., Vol. 97, No. 9, Aug., 2010, pp. 093502-1-3.

[17] B. J., Pershin, La Fontaine, S, Y. V., Di Ventra, M., *"Memristive model of amoeba learning "*, Physical Review E, Vol. 80, No. 2, Jan., 2009, pp. 019904-1-6

[18] Itoh, M., Chua, L., *"Memristor oscillators"*, International Journal of Bifurcation and Chaos, Vol. 18, No. 11, 2008, pp. 3183-3206.

[19] Ho, Y., Huang, G. M., Li, P., *"Nonvolatile memristor memory: Device characteristics and design implications"*, Int Conf on Comp Des of Tech Papers, 2009, pp. 485-490.

[20] Borghetti, J., Snider, G. S., Kuekes, P. J., Yang, J. J., Stewart, D. R., Williams, R. S., *"'Memristive' switches enable 'stateful' logic operations via material implication"*, Nature, Vol. 464, No. 7290, Apr., 2010, pp. 873-876.

[21] Shin, S., Kim, K., Kang, S., *"Compact models for memristors based on charge-flux constitutive relationships"*, IEEE Trans Comp Des Integ Circ Syst, Vol. 29, No. 4, Apr., 2010, pp. 590-598.

[22] Abdalla, H., Pickett, M. D., *"SPICE Modeling of Memristors"*, IEEE International Symposium on Circuits and Systems, Rio de Janeiro, Brazil, 15-18 May, 2011.

[23] Vourkas, I., Batsos, A., Sirakoulis, G. C., *"SPICE modeling of nonlinear memristive behavior"*, International Journal of Circuit Theory and Applications, Vol. 43, 2015, pp. 553-565.

[24] Rak, A., Cserey, G., *"Macromodeling of the Memristor in SPICE"*, IEEE Transactions on Computer-aided Design of Integrated Circuits and Systems, Vol. 29, No. 4, Apr., 2010, pp. 632-636.

[25] Biolek, Z., Biolek, D., Biolkova, V., *"SPICE model of memristor with nonlinear dopant drift"*, Radioengineering, Vol. 18, No. 2, Apr., 2009, pp. 210-214.

[26] Mladenov, V., Kirilov, S., *"A Nonlinear Drift Memristor Model with a Modified Biolek Window Function and Activation Threshold"*, Electronics, Vol. 6, No. 77, Oct., 2017, pp. 1-15.

[27] Prodromakis, T., Peh, B. P., Papavassiliiou, C., Toumazou, C., *"A Versatile Memristor Model With Nonlinear Dopant Kinetics "*, IEEE Transactions on Electron Devices, Vol. 58, No. 9, Sept., 2011, pp. 3099-3105.

Development of Arduino Based Thermal Inspection System for Hot Spots Detection in Power Lines

Miljana Milić, Sandra Đošić, Miloš Ljubenović, and Duško Lukač

Abstract – Hot spots in power lines are very common cause of damage in Power distribution and industrial environment. By detecting, locating and eliminating those critical places in time, complex and expensive equipment can be preserved, and the loss of money and power prevented. In this paper we will suggest a simple, efficient and cheap solution that solves this task. We will describe the development of one thermal inspection system for detection of hot spots in power lines, which uses a very popular open source platform for electronic systems prototyping – Arduino. The idea presented in this paper is still under implementation and though we will present the most important simulations of hardware and software parts of the system. Key benefits of such a system are: low cost, mobility, independence, low power consumption, light weight in order to be carried by a drone.

Keywords – Power lines inspection, Hot spots, Arduino, drone.

I. Introduction

Beside the application of thermovision cameras, there were not many new methods and equipment in the field of identification and diagnostics of failures in power lines [1]. Thermovision scans are regularly applied as the basic as well as the additional method for analysis of the entire power distribution systems.

Early implemetation of this method required long trainings of the inspection stuff. Beside this, successfully trained engineers were directly exposed to high voltages in the power lines and transformers. At the beginning of the year 2017, *Hydro - Québec* [2], a company that manufactures drones, developed a special kind of drone that can eliminate the problem of direct exposure of inspection stuff to high voltages. The problems that arise with this solution posed a question of the duration and the manner of the drone flight. Beside the high accuracy of fault diagnostics at the cable, some sudden and unexpected interrupts of the drone communication with the operator on the ground appeared, during the sliding of thedrone along the power cable in order to detect thermal dissipation, corrosion and other cable deviations.

Poblems with this methodology appeared in crossing points, i.e. points that represent bounds between the healthy part of the conductor, and the one that dissipate voltage (current). The current arc develops in these points, and it can lead to communication disruption between the drone and the base at the ground, which can further lead to a drone drop and damage. A galvanic separation of drone from the cable of the transmission line can overcome these side effects. Alternative approach to this kind of thermal inspections of the power line cables could be the drone that is equipped with a thermovision camera and flies in a slow flight over the cables in order to scan the temperature of them. Nevertheless, because of the inertness of the camera, a speed of the drone should be reduced in order to achieve a usable resolution of the recording, and to enable a real-time monitoring of the situation at the cable in the screen or display of the control unit on the ground. The weight of the small thermovission cameras could be another problem, since it can decrease the duration of flight by increasing the energy consumption of the drone's battery. Drones that can satisfy request for longer autonomy and possibility to carry heavy cargo, are far more expensive, unaffordable and often unavailable. With advances in electronics and thermal imaging technology, flight time duration has risen drastically, but it still involves training of the stuff to identify dissipation in the thermal image. In addition, it still requires a lot of money, because such a diagnostics set, besides the thermal camera requires the inclusion of a real-time high-resolution camera, an additional set of batteries, etc.

The aim of this paper is to present and idea of far cheaper, but equally efficient solution to thermal inspection system for drones, which is based on the Arduino environment. It could be easily affordable for small companies. The key benefits of the suggested solution are small power consumption, low weight, faster and more stable communication and real-time monitoring of the cables at the control unit on the ground.

The paper is organized in the following manner. The next section compares a traditional method for thermal inspection of the power line cables with modern, drone base ones. The third section describes the suggested concept and explains in short, the requirements of the system. Fourth section explains in detail the development of the system that includes its software and hardware parts. Results of simulations and emulations are given in the fifth section. Final remarks and conclusions, as well as the ideas for further improvement of the system are given in the conclusion.

M. Milić, S. Đošić, and M. Ljubenović are with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, Niš, miljana.milic@elfak.ni.ac.rs, sandra.djosic@elfak.ni.ac.rs, traputopia@elfak.rs.

D. Lukač, IEEE Germany Section, Kaldauer Strasse 82 a, 53721 Siegburg, Germany, dlukac@aol.com

II. Traditional Versus Drone Based Thermal Inspections

The traditional method of measuring one such dissipation involves at least two trained persons that work on the ground; cameraman and an engineer that needs to analyze the results. This is shown in Fig. 1, [3]. The camera used in this analysis exceeds the price of 10000 euros. The measurement consists of a simple thermovision recording from the ground, and displaying of the obtained results with the proper color scale and resolution in real time on the screen observed by the person performing the analysis.

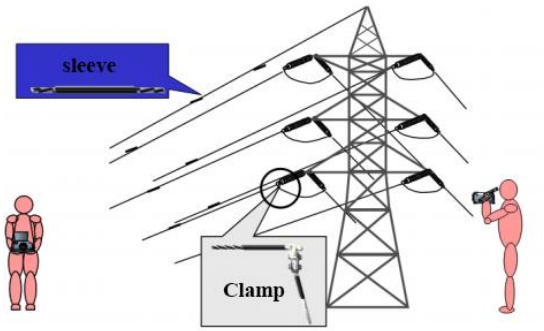


Fig. 1. Traditional method of power line dissipation analysis

Since the camera scans the area from the ground, it is possible to see only the temperature distribution, but not the cause of it, and consequently it is not possible to guarantie that the cable replacement would eliminate the detected cable fault.

After this kind of recording, a thermogram is created on a computer. The software for one such analysis costs a couple of hundred euros. The thermo-analysis results from such software are shown in Fig. 2.

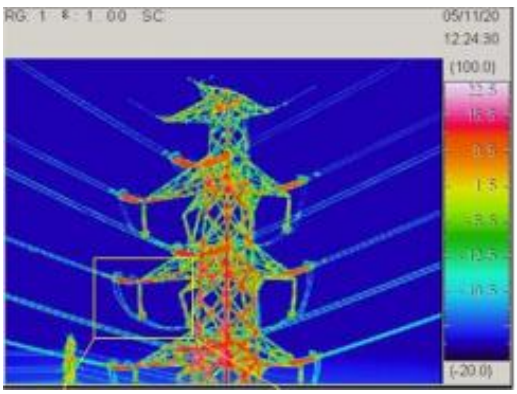


Fig. 2. Software thermo-analysis real-time view

The cost of a thermal camera plays a determining role in traditional measurement, and ranges from a minimum of $13,950 for the FLIR T540 (Fig. 3), up to several hundred thousand dollars; in electronics diagnostics, the cheapest thermal scanner costs about $500.

Fig. 3. FLIR T540

The comparison of some popular thermal cameras is shown in Table 1. Shown data are from the manufacturers' website, on December 25th. 2017.

TABLE I
COMPARISON OF DIFFERENT THERMAL CAMERA SPECIFICATIONS

Type of the camera	Price	IR sensor	Resolution	Precision
FLIR c2	$500,00	80x60	320 x 240	±2°C or 2%
Flir Ex-series	$959,00	320x240	320 x 240	±2°C or 2%
Flir Exx – series	$2,799	320x240	320x240	±2°C or 2%
FLIR T540	$13,995	320x240	320x240	±2°C or 2%

A. Drone based measurement

Due to many drawbacks, a traditional method is nowadays avoided. Some of the most important ones are: the inability to determine the cause of the breakdown in the transmission line. Therefore, drones have to carry two cameras, one thermal and one camera for real-time image display. Another drone based method is one that creates a direct contact of the drone with a conductor. The thermovision cameras that must be carried by the drone can be very heavy, and thus to reduce the flight time. One of the cameras that can be carried by a drone is FLIR T640BX, and its price is $27995. The resolution of this camera is 640 x 480 pixels. The error in thermal imaging is still very similar to the above mentioned cameras.

Fig. 4. FLIR Thermocopter

The weight of this camera is 1.3kg, although it is ranked as a lighter camera, it significantly affects the selection of drones that are able to carry it, and consequently the price of the entire system. The drone carrying a thermal camera is illustrated in Fig. 4.

B. Results of thermovision measurements

During the recording, the camera forms a thermal image by measuring the infrared radiation of the object [4]. Data obtained in such a recording must be sent and processed in the computer, or directly on the camera itself. Some cameras of higher class support these features. Often, the data processing software presents a bottleneck in the analysis of the damage of the transmission line. It is very expensive, and the time to obtain the appropriate thermogram is much longer than the time for collecting the data at the transmission line.

III. The System Model

The system presented in this paper consists of two Arduino platforms that communicate with each other using 433MHz RF module. One Arduino platform is located on the drone and it is responsible of power line thermal dissipation recording. The second Arduino platform is used for data collection, acquisition and processing. We use the Processing software package for data visualization, i.e. for forming the power line thermal dissipation image.

A. The drone model

The drone model consists of the Arduino Nano platform, three contactless IR temperature sensors MLX9061, RF transmitters and RF receivers, (both operating at 433MHz) and Li-Ion batteries. We use two servo motors to change the sensors position and to provide measurement at different points.

The measured data are sent to the receiving unit on the ground, thus creating the real-time image.

B. The ground model

The ground model consists of the Arduino UNO platform, RF receiver and RF transmitter (both operating at 433MHz), buttons and switches. It is also a system that needs the battery supply. The system's task is to receive data from the drone, and to forward them to the local computer or an Android device. The communication between the local computer and an Android device with Arduino UNO platform is serial. The communication is fast enough to realize the real-time image, with the minimum acceptable delay.

C. The processing software

The data is processed by Processing software package, based on Java. The software is used to collect data from serial communication, store them in temporary fast buffers, and plot the sensor's pixel temperature image. The sensors can be moved by the servo-motors, so the software knows the pixel position and based on the measured temperature, knows the color as well as the pixels intensity. The additional time saving is done with special pixel drawing: from left to right, and from right to left. The so obtained thermal image contain data memorized in a discrete time interval.

IV. The System Functionality

The functionality of the system will be described using UML (The Unified Modeling Language).

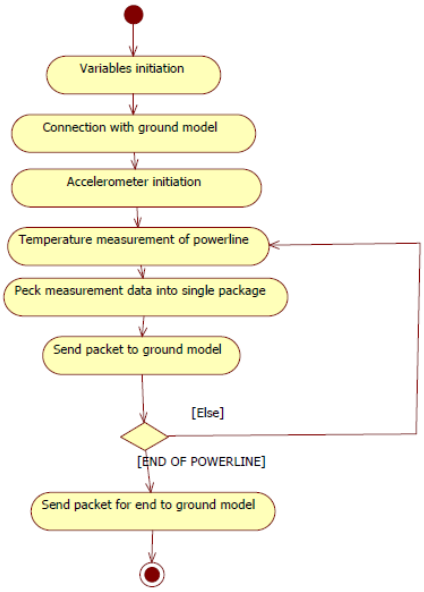


Fig. 5. Controller activity state diagram

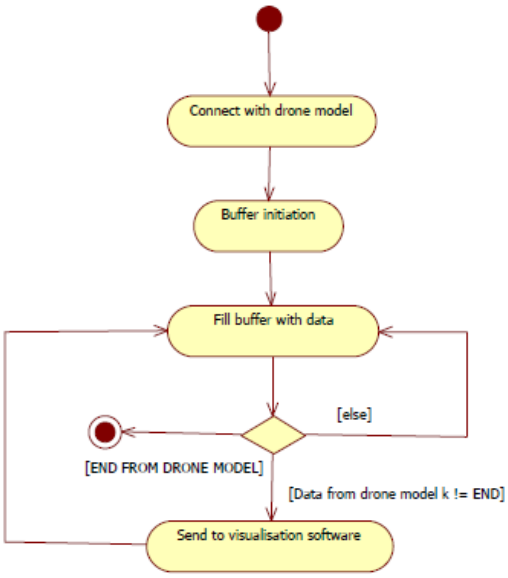


Fig. 6. The ground controller activity

UML is a family of graphical notations that help in describing and designing software systems, particularly software systems built using the object-oriented style [5]. The UML was created by the unification of many object-

oriented graphical modeling languages that were succesfull in the late 1980s and early 1990s. Since the OMG (Object Management, Group), [6] adopted it as a standard, UML began to conquer the market. Today, UML is one of the most preferred languages for specification, visualization, construction and documentation of the software systems development.

Fig. 5 describes the controller activity above the power line which measures the conductor's temperature in real time, and sends the data to the processor on the ground. Fig. 6 describes the ground controller activity, which has to accept the data and forward it to the computer, or the Android device for visualization.

V. Simulation Results

The simulation was done using Proteus [7] and Fritzing [8] software packages. Results of the simulation were processed in the Fritzing software package (for displaying the entire system on the prototype), Fig. 8. Due to the complexity of the MLX90614, [9], sensor (Fig. 7), and lack of the model in Proteus software package, the LM35 sensor was used to achieve the desired dynamics and the temperature change rate in order to perform performance testing, as well as the thermal design visualization. Fig. 9.

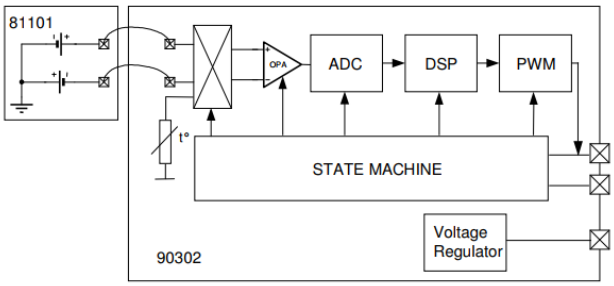


Fig. 7. Module MLX90614

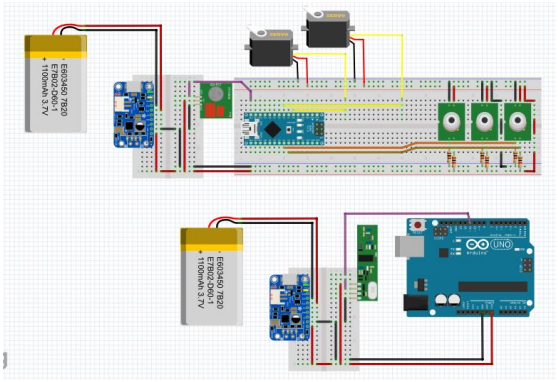


Fig. 8. Fritzing emulator

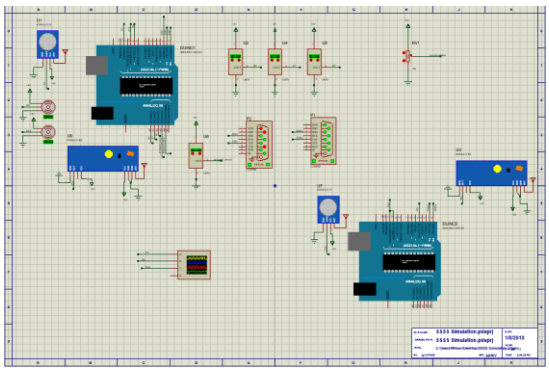


Fig. 9. Proteus emulator

Simulation models for Arduino as well as RF transmitters and receivers had to be developed additionally. The simulation inculdes collecting data from the sensors, storing them in the temporary buffer, and sending them via RF links. The first RF communication link carries information about the measurement, that is, the current temperature read by each of the three sensors individually. The second RF communication link is between pair of receiver and transmitter and it carries the control bits that allow the certain part of the image to be scanned. That was done by sending the servo engine coordinates from the ground to the drone. The first RF communication link operates at 433MHz, while the second link operates at 315MHz. ASK (Amplitude-shift keying) modulation is used for both communications links, since it gave the best results during simulations [10].

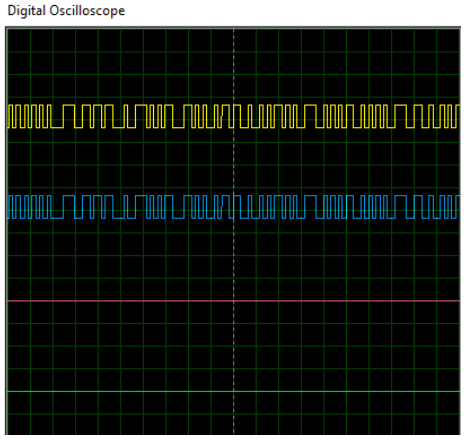


Fig. 10. Measured exchanged data: Yellow - send data, Blue - received data

Fig. 10. shows communication's speed between the drone model and the ground model. It is shown that data is accepted almost immediately after sending. This is a very important feature of the system, especially during linear scanning of the power line. In the scanning process is not desirable to skip any part of power line.

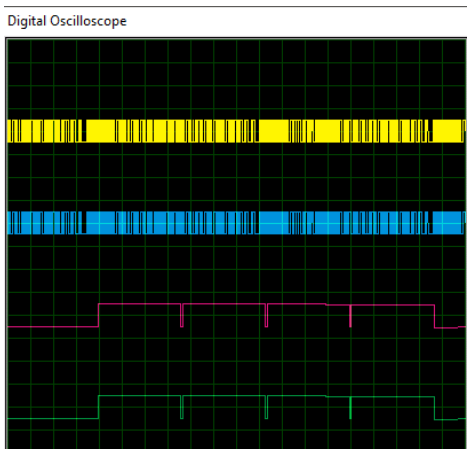


Fig. 11. Two parallel communications without crosstalk: Pink - Send control data, Green - Received control data

Fig. 11. shows communication without crosstalk, because we use buffers on both sides. The buffers are controlled by Arduino interrupts, and they depend on the communication.

Fig. 12. shows errors, when, for example, the temperature suddenly changes from 240 to 300 degrees; the error is $\pm 4^0$C [7].

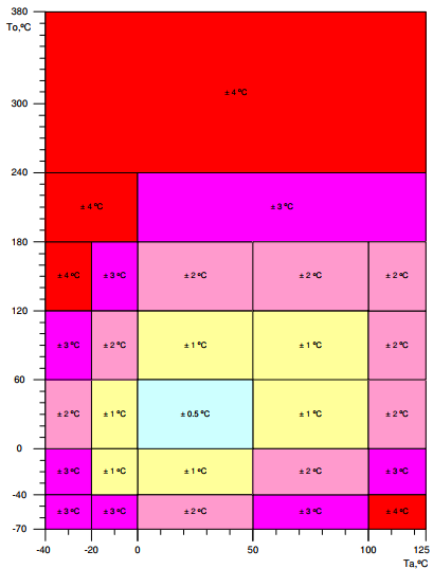


Fig. 12. Error display MLX90614 [7]

Measurement visualization software was written in the software environment Processing [11]. Processing is a Java customized software for processing data from a serial port, or from a microcontroller. The visualization software has two modes of operation. The first mode (Fig. 13) deals with static parts of the transmission line. The second mode represents a linear analysis of conductors. The software has the ability to display image from the drone's camera in a real-time, to save the image, to capture the image, and to set the thermal image over the real-time camera image.

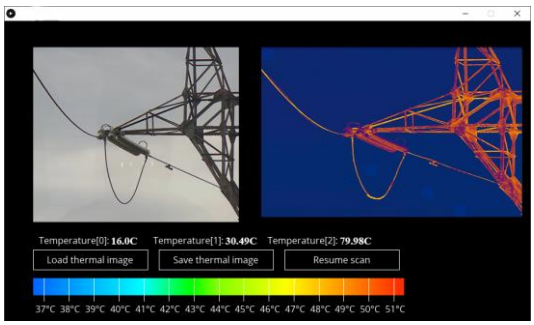


Fig. 13. The static part of the transmission line analysis

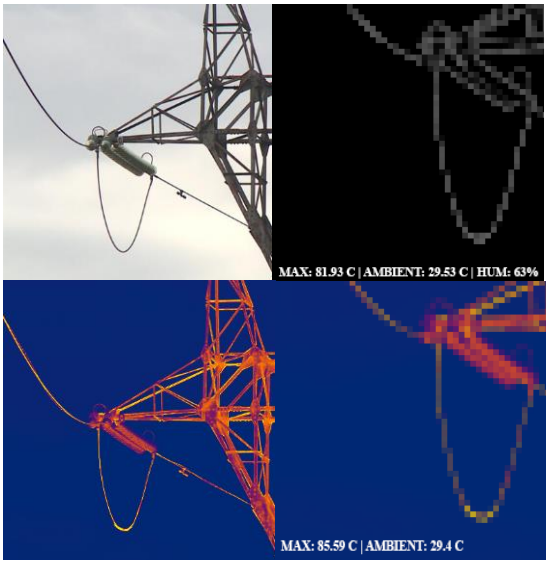


Fig. 14. Prototype images

Figure 14. shows the real-time camera pictures and the image from our model.

VI. Conclusion

In this paper, we have described the Arduino based thermal inspection system for hot spots detection in the power lines. The most important advantage of the presented system is its low price. Nowadays equipment for test ing the transmission line is very expensive which does not allow to be widely used. Considering the price of our presented system, it can be used to examine the thermal dissipation inside the house or for investigating dissipation from the central heating pipes, etc. Also, the presented solution is mobile and light weights, which allows work on different terrains especially in inaccessible places.

Acknowledgement

This paper is partially funded by The Ministry of Education and Science of Republic of Serbia under contract no. TR32004, and by Project Grant III44004.

REFERENCES

[1] Company FLIR, "Thermal imaging guidebook for industrial applications", FLIR Systems AB, 2011.
[2] http://www.hydroquebec.com/residential/
[3] Šaraba P, Popović B, Ljuboje Z, „Izrada izvještaja termovizijskih ispitivanja u elektroenergetskim postrojenjima" , INFOTEH-JAHORINA Vol. 15, March 2016.
[4] Kompanija NEC, "Primeri primena termovizijskih kamera", Mel co Buda d.o.o
[5] Fowler, M., UML Distilled: A Brief Guide to the Standard Object Modeling Language, Addison-Wesley,September, 2003.
[6] http://www.omg.org/
[7] https://www.labcenter.com/
[8] http://fritzing.org/home/
[9] MLX90614 DATA SHEET
https://www.sparkfun.com/datasheets/Sensors/Temperature/MLX90614_rev001.pdf , page 25
[10] https://en.wikipedia.org/wiki/Amplitude-shift_keying
[11] https://processing.org

List of Authors

35. Melikyan, Vazgen	37*
36. Milenković, Srdjan	29*, 33
37. Milić, Miljana	55, 96*, 101*
38. Milojković, Jelena	24*
39. Mitrović, Nebojsa	71
40. Mon Kueh, Si	13*
41. Nieto Taladriz Garcia, Octavio	29
42. Panić, Goran	9*
43. Petronijević, Milutin	71, 77
44. Petrović, Miljan	96
45. Piliposyan, Gor	37
46. Savić, Milan	18, 21*
47. Sladojević, Lazar	77*
48. Stevanović, Dejan	83
49. Stojanov, Riste	1
50. Stojanović, Miodrag	77
51. Trajanov, Dimitar	1*
52. Trajković, Slaviša	55
53. Vučković, Vladan	42*
54. Zdraveski, Vladimir	1
55. Zwolinski, Mark	63*