Studies in Systems, Decision and Control 74

Vyacheslav Kharchenko
Yuriy Kondratenko
Janusz Kacprzyk *Editors*

Green IT Engineering: Concepts, Models, Complex Systems Architectures

Springer

Studies in Systems, Decision and Control

Volume 74

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series "Studies in Systems, Decision and Control" (SSDC) covers both new developments and advances, as well as the state of the art, in the various areas of broadly perceived systems, decision making and control- quickly, up to date and with a high quality. The intent is to cover the theory, applications, and perspectives on the state of the art and future developments relevant to systems, decision making, control, complex processes and related areas, as embedded in the fields of engineering, computer science, physics, economics, social and life sciences, as well as the paradigms and methodologies behind them. The series contains monographs, textbooks, lecture notes and edited volumes in systems, decision making and control spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

More information about this series at http://www.springer.com/series/13304

Vyacheslav Kharchenko
Yuriy Kondratenko · Janusz Kacprzyk
Editors

Green IT Engineering: Concepts, Models, Complex Systems Architectures

Springer

Editors
Vyacheslav Kharchenko
Department of Computer Systems and Networks
National Aerospace University “KhAI”
Kharkiv
Ukraine

and

Centre of Safety Infrastructure Oriented Research and Analysis
RPC Radiy
Kirovohrad
Ukraine

Yuriy Kondratenko
Department of Electrical Engineering and Computer Science
Cleveland State University
Cleveland, OH
USA

and

Department of Intelligent Information Systems
Petro Mohyla Black Sea National University
Mykolaiv
Ukraine

Janusz Kacprzyk
Intelligent Systems Laboratory
Systems Research Institute of Polish Academy of Sciences
Warsaw
Poland

ISSN 2198-4182 ISSN 2198-4190 (electronic)
Studies in Systems, Decision and Control
ISBN 978-3-319-44161-0 ISBN 978-3-319-44162-7 (eBook)
DOI 10.1007/978-3-319-44162-7

Library of Congress Control Number: 2016947507

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

This volume provides a comprehensive state-of-the-art overview of a series of advanced trends and concepts that have recently been proposed in the area of green information technologies engineering (green IT engineering) as well as of design and development methodologies for models and complex system architectures and their intelligent components. These new research directions have attracted much attention in recent years and are considered to be very promising.

All the contributions included in the volume have their roots in the authors' presentations, and vivid discussions that have followed the presentations, at a series of workshop and seminars, held within the international TEMPUS-project GreenCo project in the UK, Italy, Portugal, Sweden, and the Ukraine, during 2013–2015 and at the 1st–5th Workshops on Green and Safe Computing (GreenSCom) held in Russia, Slovakia, and the Ukraine.

The main purpose of this volume is to present a systematic exposition of research on principles, models, components, and complex systems (software and the PLC-based systems, and networks and IT-infrastructures) and a description of industry- and society-oriented aspects of the green IT engineering. The green IT engineering is a special kind of engineering based on energy-saving and efficient information technologies. It is represented in a form of services focused on the improvement of energy efficiency, safety, and environmental performance of industrial processes and products.

The book aims at motivating researchers and engineers representing different IT domains, and also graduate and Ph.D. students, and "postdocs," to explore and pay attention to an acute need for the promotion and propagation of green values in the analysis and solutions of complex problems that are relevant to the society and national economy.

A chapter-oriented structure has been adopted for this book following a "vertical view" of the green IT, from hardware (CPU and FPGA) and software components to complex industrial systems. The chapters are prepared according to a general paradigm and unified scheme of content, and step by step describe elements of a green IT engineering taxonomy by logically adding new values and proposals.

In terms of the structure, the 15 chapters of the book are grouped into five parts: (1) Methodology and Principles of Green IT Engineering for Complex Systems, (2) Green Components and Programmable Systems, (3) Green Internet Computing, Cloud and Communication Systems, (4) Modeling and Assessment of Green Computer Systems and Infrastructures, and (5) Green PLC-Based Systems for Industry Applications.

The chapters have been intended to provide an easy to follow, comprehensive introduction to the topics that are addressed, including the most relevant references, so that anyone interested in them can start the study by being able to easily find an introduction to the topic through these references. At the same time, all of them correspond to different aspects of the work in progress being carried out by various research groups throughout the world and, therefore, provide information on the state of the art of some of these topics, challenges, and perspectives.

Part I, "Methodology and Principles of Green IT Engineering for Complex Systems," includes three contributions. The first, "Concepts of Green IT Engineering: Taxonomy, Principles and Implementation," by V. Kharchenko and O. Illiashenko, summarizes concepts and provides a taxonomy of the green IT engineering. Main challenges and solutions regarding applications of energy-saving IT-based components and systems as well as principles of development and implementation for green computing are analyzed and discussed. The deliverables of the TEMPUS-project GreenCo are described.

I. Shostak, M. Danova, and Y. Kuznetsova, in "Foresight-Research for Green IT Engineering Development," study the basic principles and development directions of the green IT engineering, as well as ecological research in the software industry. A generalized procedure for foresight project implementation is proposed to determine scientific–technical development directions and perspectives of the green IT engineering using the proposed information technology. The authors propose a new approach to the construction of a foresight-type technology system model in the form of a two-level hierarchical system consisting of the functional and methodological levels. As a formal basis, methods of bibliometrics and scientometrics (calculation of the number of publications, an analysis of citations), multi-criterion decision-making problems (t-ordering, Pareto optimality), and patent analysis (trace analysis of the dynamics of inventive activity) are chosen. The result of the foresight-research implementation will be a number of priorities for the development of the green IT engineering.

In "Green IT Engineering in the View of Resource-Based Approach," J. Drozd, A. Drozd and S. Antoshchuk consider the development of the green IT engineering from the point of view of a resource-based approach which analyzes the integration of an artificial world created by a human being into the natural world (NW). The solution is based on the attainment of a certain throughput, trustworthiness, and the investment in resources: models, methods, and means. The throughput is aimed at receiving the maximum return from resources and the trustworthiness—at a harmonization with the NW. Resources structured under parallelism and fuzziness of the NW show in the development process three levels: replication, diversification, and autonomy. They are serviced by methods improving the throughput, trustworthiness, and access to resources, respectively. Green technologies as a model occupy the autonomy level. The models and methods used belong to the level of diversification. Means are built with the replication of operational elements. The authors discuss the methods of multiple effect and its focusing on the parameters which are important for the green technology that is based on increasing a level of resource development.

Part II, "Green Components and Programmable Systems," includes three contributions. In the first, "Green Logic: Models, Methods, Algorithms," S. Tyurin and A. Kamenskih investigate the green computing systems as energy-aware or energy-efficient and naturally reliable computing systems. A combination of these approaches gives many advantages for applied computing systems but estimations for them should be done very carefully. Delay-insensitive circuits occupy an important role in the design of the green hardware, but reliability improvement techniques of delay-insensitive circuits should be further investigated. The authors propose new indexes for estimating the efficiency of computing systems which can help developers to perform complex estimations of computing systems operating in a wide ranges of supply voltage and temperatures. The semi-modularity and the synthesis technique of fault-tolerant delay-insensitive circuits is discussed in this chapter.

In "Energy-Efficient Scheduling for Portable Computers as Bi-criteria Optimization," I. Turkin and A. Vdovitchenko consider the main problem of portable computers, that is a short duration of operation of the device in the standalone mode, while ensuring the quality of service, i.e., a subjective user satisfaction. There is no single way of measuring the subjective user satisfaction as the current quality of service requirements are in conflict with the requirements to ensure uptime portable computers. A power consumption model of a portable computer is presented as a 3-level power graph. The developed model of the multitasking system's scheduling in a portable computer battery life is based on the assumption that the problem under consideration belongs to the soft real-time class. The synthesis of schedule's multitasking system in a portable computer battery life is carried out by solving a bicriteria optimization problem by deriving the Pareto-optimal solutions. These criteria include the minimum penalty for the decline in the quality of service and the maximum battery life with the current profile of power consumption. For experimental investigation of the power consumption, PCMark-7 and Microsoft Joulemeter were used. As a result of the experiments, it has been found that during the operation on portable computers the minimum and maximum energy levels differ by more than 2 times. A time transition to a low- or high-voltage processor, respectively, can lead to a twofold increase in the battery life.

D. Maevsky, E. Maevskaya and E. Stetsuyk, in "Evaluating the RAM Energy Consumption at the Stage of Software Development," propose a new method of absolute value estimation of the computer energy consumption while running performing. The evaluation is done on the basis of the program source code and can help to choose the optimal solution from the viewpoint of energy saving at the software development stage. The method is based on the indication of energy consumption by the computer Random Access Memory (RAM) depending upon how intensive the RAM is used by the software. The method for determining the estimated computer power consumption based on assembler source code is proposed based on the two proposed mathematical models which allow to create a green software with the control of a degree of being “green” on all stages of its development.

Part III “Green Internet Computing, Cloud and Communication Systems” includes three contributions. In the first one, “Impact of the Internet Resources Structure on Energy Consumption while Searching for Information,” by V. Dubovoi and O. Moskvin, a new model of the effect of the Internet resources structure impact on the energy efficiency of information search and the state of the art of search engines energy consumption, characteristics of hypertext systems, and search engines are presented. Moreover, a model of the relationship between the hypertext characteristics and the number of information search steps is developed. For this purpose, the impact of the hypertext structure on the number of steps while searching for relevant and pertinent information and the impact of the number of steps on the energy consumption are studied. As a result, approaches for the optimization of hypertext structure in the conditions of uncertainty are formulated. A simulation model makes it possible to test the adequacy of the developed model of the effect of the structure of distributed hypertext systems on the energy efficiency of information search.

A. Iqbal, C. Pattinson, and A.-L. Kor, in “Introducing Controlling Features in Cloud Environment by Using SNMP,” discuss the deployment Simple Network Management Protocol (SNMP) for the monitoring and control of a type 1 hypervisor (in a cloud environment). This is followed by the customization of the MIB and net-snmp with Agent X to provide more SNMP management features. The work provides results of a rigorous physical experimentation involving the SNMP monitoring and management for the type I hypervisor Xen. The research results allow load balancing in the cloud environment during peak hours leading to a reduced power consumption helping the green IT cause.

In “Efficient Error Detection and Correction in Block Data Transmission,” N.G. Bardis proposes a collection of techniques for correcting transmission burst errors in data transmitted over signal channels suffering from strong electromagnetic interference, such as those encountered in distributed and embedded systems. Efficiency is achieved by separating the error detection from the correction process and using different codes for each case. The proposed error control techniques are based on simple mathematical operations and are suitable for implementation in the FPGA devices. Therefore, it makes it possible to replace energy-demanding retransmission operations, including the overheads they entail with energy-efficient local error correction calculations. The techniques employed are shown to be more efficient than the existing ones, according to criteria that are relevant to current applications. These techniques reduce the need for error recovery by retransmission and hence the environmental effect of data transmission in terms of energy consumption and electromagnetic emissions.

Part IV “Modeling and Assessment of Green Computer Systems and Infrastructures” includes three contributions. In the first one, D. Basile, F.D. Giandomenico, and S. Gnesi in “Model-Based Evaluation of Energy Saving Systems” focus on a stochastic model-based approach as a support of the analysis of energy-saving systems, in combination with other non-functional properties, such as reliability, safety, and availability. The authors discuss general guidelines for building a model-based framework to analyze critical cyber-physical systems in which an efficient energy consumption is required, while assuring imposed levels of resilience. Also, an overview of the most commonly employed methodologies and tools for the model-based analysis is provided, and an extensive literature is provided as pointers to relevant research activities performed on this attractive topic over the last decades. Finally, in order to illustrate the proposed framework, a case study in the railway domain is shown. By adopting the Stochastic Activity Network formalism, the framework is instantiated to analyze effective trade-offs between the energy consumption and satisfaction of other dependability-related requirements.

In “MSS Models of Smart Grids with Multi-level Degradation and Recovery,” E. Brezhnev, H. Fesenko, V. Kharchenko, V. Levashenko, and E. Zaitseva discuss the digital substations (DS) of a smart grid that are complex multi-component maintained systems consisting of a lot of hardware and software components. Failures of the components cause functional and parametric degradation of the substations. According to the DS structure, the reliability-block diagram, the structure function, and the Direct Partial Logical Derivatives (DPLDs) for the RMSS “electronic transformers—merging unit” are considered. The principal condition of the DPLD application in reliability analysis is the representation of a system under consideration by a structure function. The authors consider the calculation of some of these measures and their criticality for the analysis of the electronic transformers—merging unit. The structure function of this unit is based on the operation conditions of this system (unit).

Ye. Bodyanskiy, O. Vynokurova, I. Pliss and D. Peleshko, in “Hybrid Adaptive Systems of Computational Intelligence and Their On-line Learning for Green IT in Energy Management Tasks,” consider a crucial problem of intelligent energy management which arises in the context of an intensively developed science direction, the Green IT. A hybrid neuro-neo-fuzzy system and its high-speed learning algorithm are proposed. This system can be used for online prediction of essentially nonstationary nonlinear chaotic and stochastic time series which describe electrical load producing and consuming processes. The considered hybrid adaptive system of computational intelligence has some advantages over the conventional artificial neural networks and neuro-fuzzy systems. The proposed hybrid neuro-neo-fuzzy prediction system provides a high-quality load prediction that is very important for the power systems.

Part V “Green PLC-Based Systems for Industry Applications” includes three contributions. The first one, Y. Kondratenko, O.V. Korobko and O.V. Kozlov, “PLC-Based Systems for Data Acquisition and Supervisory Control of Environment-Friendly Energy-Saving Technologies,” presents the development of PLC-based systems for data acquisition and supervisory control of environment-friendly energy-saving complex high-tech technologies. The functional structure and main components of the PLC-based SCADA systems for environment-friendly energy-saving technological processes are given. The examples of SCADA applications in the design of the PLC-based systems for the monitoring and automatic control of (a) ecopyrogenesis (EPG) and (b) thermoacoustic technological processes are presented. The authors consider the criteria of energy and economic efficiency of the EPG technological process. The functional structures, software, and hardware implementations as well as multi-level human–machine interfaces of the developed PLC-based systems for data acquisition and supervisory control are shown. A considerable attention is given to particular qualities of the computation of technological parameters of the ecopyrogenesis and thermoacoustic processes by the proposed SCADA systems. The developed PLC-based SCADA systems provide: (1) a significant increase of values of the energy and economic efficiency criteria of the EPG and TAD complexes, (2) a high precision control of both technological processes, (3) monitoring of current technological parameters using the indirect methods for the measurement of parameters and identification, and (4) automatic control with high values of quality indicators and optimal parameters.

In "Assessment of Energy Consumption for Safety-Related PLC-Based Systems," V. Sklyar, O. Odarushchenko, E. Bulba, R. Horbenko, A. Ivasyuk, and D. Kotov describe an approach to measure the energy consumption of programmable components (MCU, FPGA, etc.) which are a core of the Programmable Logic Controller (PLC). An analysis of dependence of energy consumption on the type of programmable components as well as on features of software gives a road map for the implementation of energy efficiency measures to make the PLCs "greener." A special toolset named GreenCo Controller has been designed, implemented, and tested for the above-proposed solution. The architecture designed includes the following components: a commercially available motherboard (Arduino UNO) programmed with software to perform data processing with respect to the measurement of energy consumption of the target board-shield (GreenCo board); a GreenCo board on the basis of Microchip dsPIC30F3011 MCU programmed with different types of application software as a subject of energy consumption investigation; a software monitor for a PC desktop using the GUI of the GreenCo Controller mentioned. "Green" features of the PLC are analyzed after that formal optimization problems against the "energy consumption level/safety level" criterion. Such problems can be solved using dynamic programming. The concept and design of the GreenCo Controller are described. Some suggestions for a further research on the GreenCo Controller are proposed as a conclusion.

Finally, in A. Shamraev, E. Shamraeva, A. Dovbnya, A. Kovalenko, O. Ilyunin, "Green Microcontrollers in Control Systems for Magnetic Elements of Linear Electron Accelerators," a new approach to the design of industrial control systems is presented which is based on the application of "green" microcontrollers (with a low power consumption) used in complex real-time control systems. A wide range of capabilities of modern microcontroller peripheral units makes it possible to implement a system with a wide functionality based on a single microcontroller chip. The authors describe the order of connection of microcontroller pins which reduces noise influence on the results of signal processing. As an example, the proposed approach is considered in the context of development of a power supply control system for magnetic elements of linear electron accelerators.

The papers selected for this book provide a broad and comprehensive overview of some of the most up-to-date and relevant problems in the area of green IT engineering and the approaches and techniques that relevant research groups within this area propose, develop, advocate, and—finally—use for the solution of real world problems. We would like to express our appreciation to all authors for their contributions as well as to reviewers for their timely and interesting comments and suggestions. We certainly look forward to working with all contributors again.

We also wish to thank Dr. Tom Ditzinger, Dr. Leontina di Cecco, and Mr. Holger Schaepe from Springer for their dedication and help to implement and finish this publication project on time maintaining the highest publication standards.

Kharkiv, Ukraine — Vyacheslav Kharchenko
Mykolaiv, Ukraine — Yuriy Kondratenko
Warsaw, Poland — Janusz Kacprzyk

Contents

Part I
Methodology and Principles of Green IT Engineering for Complex Systems

Concepts of Green IT Engineering: Taxonomy, Principles and Implementation

Vyacheslav Kharchenko and Oleg Illiashenko

Abstract Among the challenges faced by modern civilization, are the key challenges associated with the shortage of energy resources, environmental issues and safety. Information technologies are an important factor of energy consumption, safety and green culture. The paper describes concepts and taxonomy of green IT engineering. Main principles of development and implementation, indicators and values of green computing are analyzed. Description of EU project GreenCo is provided.

Keywords Green IT engineering · Green computing · Sustainable development · Taxonomy · Green values · Green metrics · Education project

1 Introduction

1.1 Duality of IT

Lack of energy and resources in general is caused by the growth of human consumption and the development of intensive production. Technogenic activities of mankind increase the load on the environment and increase the environmental risks. These global processes occur in the background extra-dynamical developments of IT technologies, which occupy a special place in the technological race, are in the leading group. Therefore question naturally arises whether the IT linked with these civilizational challenges or not, but if linked, how? ITs by providing a generally

V. Kharchenko · O. Illiashenko (✉)
National Aerospace University "KhAI", 17 Chkalov Str., Kharkiv, Ukraine
e-mail: o.illiashenko@csn.khai.edu

V. Kharchenko
Centre of Safety Infrastructure Oriented Research and Analysis,
RPC Radiy, 29 Tamm Str., Kirovograd, Ukraine
e-mail: V_S_Kharchenko@ukr.net

V. Kharchenko et al. (eds.), *Green IT Engineering: Concepts, Models, Complex Systems Architectures*, Studies in Systems, Decision and Control 74,
DOI 10.1007/978-3-319-44162-7_1

positive impact on the search for answers to those challenges pose specific problems which may significantly neutralize the positives in the end of the day, if not overlap them at all. This can be defined as duality factor. It is due to the fact that IT-systems are both the means and the object of energy saving.

Information technologies do not exist by themselves, and are implemented by IT-systems which store, display, transform and transfer the information in the interests of users or other systems. These IT-systems are a mandatory part of the energy systems that provide the generation, transformation, and delivery of various types of energy, primarily electricity.

For the following classes of systems the duality factor is particularly acute [1–3]:

- Energy-critical systems, where information technologies and IT-systems control and consume energy;
- Instrumentation and control systems used in nuclear power plants are protection means and can decrease and increase emergency risks;
- Aerospace (aviation, rocket and space), medical and other critical computer based systems influence on safety and power consumption.

The duality of IT-systems causes necessity of comprehensive review of all aspects (energy, environmental and safety—in the context of functional safety and information security) in the relationship and interdependence.

1.2 Green Technologies and Principles of Sustainable Development

Green technologies are part of a broader concept of *sustainable development* or *sustainability* of our civilization [4–6]. Its formulation is given in the report "Our Common Future" prepared by the World Commission on Environment and Development Organization of the United Nations in 1987. At the UN conference in Rio de Janeiro in 1992, representatives of 179 countries adopted the concept of sustainable development in the 21st century as the principle of the survival of the status and existence of an Earth civilization.

A development is called *sustainable* when it meets the needs of the present generation without infringements of the future generations to meet their own needs [7].

The concept of *sustainable development* is seen as a prerequisite for long-time progress of mankind, followed by augmentation of the capital understood in the broad sense and improving environmental conditions. This concept also implies a purposeful movement of financial resources from rich to poor regions, a broad exchange of environmental knowledge and information. Some experts believe that the term "sustainable development" contains a contradiction, which consists in the fact that, on the one hand, it highlights the need for constant development (including the raising of the material level), and on the other—intended to limit this development. This limitation is based on the provisions of the theory of evolution, which has a controversial position to a combination of the terms “sustainability” and “development”.

However, this approach cannot be considered generally accepted. Rather, it points to the complexity of the implementation of economic development, which is not accompanied by an improvement of environmental conditions.

In works [8, 9] the principles of sustainable development are given:

– *the principle of eco-efficiency*, which implies the addition of traditional efficiency formula as the “get more for less” by the mandatory additional element of “decreasing impact on the environment and a competitive and quality product”;
– *the principle of environmental justice*, according to which the distribution of resources between the living and future generations should be equal. Its actual implementation is possible only by reducing consumption of resources by present generations;
– the principle of environmental efficiency, which is the most radical and assumes full termination of the negative impact on the environment.

The following areas [10] are important for sustainable development in the context of green technology:

– the forward looking development and implementation of effective measures to prevent environmental degradation, prevention of environmental and man-made disasters are required;
– economic activity must become socially and environmentally friendly, and be accompanied by a decrease in the differences in the level of people’s lives, poverty and misery, strengthening the relationship of economy and ecology, the formation of a single (balanced) ecologized system of economic development. The projects that can either cause irreparable damage to the environment or lead to poorly understood consequences should be abandoned;
– sustainable development must take into account all aspects of the safety of individuals, region, country and others.

1.3 Green IT for Sustainable Development: The Positives and Challenges

Green Information Technologies (green IT)—is one of the kinds of green technologies. Green IT can be considered in several aspects:

- as a tool for the implementation of the concept of sustainable development (ecology, safety, energy efficiency);
- as an object for providing such sustainability.

IT and its components (software and hardware) are called sustainable when their direct or indirect negative impact on the economy, society, and human environment due to their development and use is minimal and/or has a positive impact on sustainable development [11]. This definition incorporates several views. The key here is the concept of sustainability, which covers all aspects of green IT.

In [12] the stability of the IT and software is seen as a generalization of the quality model in which the traditional components are grouped based on the concept of sustainable development.

In [13] a model of sustainable information technologies (software) seems a star chart of the five elements. Thus green software is positioned only with environmental sustainability, which reduces the scope of this term. In our view, a more constructive model would be a metric radial diagram (kiviat radar or chart) in which sustainability attributes are displayed on its axis in normalized form. In this case the calculation of private and common metrics of IT' sustainability or its components is possible, which would allow comparing different technologies more objectively.

In addition to the above, sustainable development involves a number of economical, technical, social and specific components associated with IT and software. In one of the latest and most comprehensive monographs on green IT [14], they are defined as technologies that combine three areas:

- «greening» of IT-systems itself (increasing their energy efficiency and reduce the negative effects of their use for energy and resource component);
- support of environmental sustainability (environmental and safety components);
- an increase of awareness of the value of green culture, green business through IT (social component).

The impact of information technologies on the implementation of control and supervisory functions in the technical and organizational and technical systems on resource-, energy-saving and the environmental component cannot be overestimated because there are already exist and are widely used:

- green cars, running on electricity and controlled by onboard computers that can interact with the transport of IT infrastructure;
- power plants running on green energy and controlled by embedded computer means which ensure their effectiveness and safety;
- smart grids, which has a computer monitoring of settings, the correction of functions and operational reconfiguration of systems.
- smart green buildings, smart offices, smart universities, smart cities, in which all the business processes of life support and control systems are implemented by all the deeper penetration of information technologies and materials, including such modern IT as cloud computing, Internet things (IoT) and others. The next step will be the green regions and countries.

The dynamic development of information technologies and components leads to a number of problems related to the fact that:

- they are more powerful consumers of electricity. For their manufacture, first of all, for the hardware components (chips, circuit boards, computers, peripheral devices) the expensive and non-harmful materials are required;

– IT-systems directly or indirectly have a negative impact on the environment by increasing the CO_2 emissions associated with their development and application. Disposal of old equipment is expensive. The problem of e-waste becomes global in nature and requires coherent and comprehensive legal, economic and technological solutions, without which it will increase risks for sustainable development;
– stiff competition and a huge market cause unreasonably fast refresh of mass application and entertainment software that requires greater resources for their creation, use and recycling;
– many hours of daily work of IT professionals and specialists leads to serious diseases and changes the psyche of people.

Thus, we can conclude that green ITs are becoming more voluminous and important component of sustainable development, so they can be considered in two sections: as an object of analysis (how to ensure the sustainability of IT as a component of sustainable development of technology in general, and civilization as a whole) and as an object of the development and application (for sustainable development) within the framework of this concept.

The purpose of this work is to systematize concepts related to green IT engineering and to discuss some aspects of its development and implementation.

The work is structured in a following manner: Sect. 2 describes main concepts, definitions of green engineering and green IT. The taxonomy of green computing is provided. Section 3 contains description of values and principles of green IT implementation as well as risks associated with introduction of green ITs. Section 4 briefly described joint initiative on educational aspect of green IT engineering GreenCo project funded by EU.

2 Concepts and Taxonomy of Green IT Engineering

2.1 Taxonomy of Green Computing

ITs can be described and viewed in the context of sustainable development both in the narrow and broad sense. This section contains refinement of research done in [15]. In a narrow sense they are energy-saving and energy-efficient ITs and systems, and in broad one they take into account environment, safety, security and sustainable development in general. Thus, main entities in the field of sustainability and green IT, as an "IT Sustainability Set" or ITSS, could be described by the Cartesian product of two subsets:

- Sustainability Development Set (*SDS*) of: energy and resource (EnF), ecology (EcF), safety (SF) и social and economic (CF);
- Two-element set (Means-Object Set—*MOS*):

Table 1 Green IT tasks in sustainability context

Set, *MOS*		Sets of sustainability factors, *SDS*			
		EnF	EcF	SF	CF
Means		Means * EnF	Means * EcF	Means * SF	Means * CF
Object, IT-components	HW	EnF * HW	EcF * HW	SF * HW	CF * HW
	SW	EnF * SW	EcF * SW	SF * SW	CF * SW
	NW	EnF * NW	EcF * NW	SF * NW	CF * NW
	IS	EnF * IS	EcF * IS	SF * IS	CF * IS
	IT	EnF * IT	EcF * IT	SF * IT	CF * IT

$$\mathrm{ITSS} = SDS * MOS = (\mathrm{EnF}, \mathrm{EcF}, \mathrm{SF}, \mathrm{CF}) * (\mathrm{Means}, \mathrm{Object}) \quad (1)$$

Table 1 discloses this multiplication in a form of matrix. Each cell contains a particular task of green IT. String "Means" is not parted to the components because when IT is used as a tool, the combinations of components are used as a rule. Strictly speaking both net and infrastructure means (or tools) include hardware and software.

Here HW relates to hardware, SW to software, NW to netware, IS to information system, and IT to information technology. One more string was added to section "Object, IT-components", which brings together all the components. This matrix allows conducting preliminary analysis of green IT directions and corresponding facts. It generalizes the similar matrices proposed by the project participants TEMPUS GreenCo in [2, 3]. The analysis performed allows developing of taxonomic scheme of green information technologies. Here we confine our analysis to the taxonomy of green IT in the narrow sense: define the hierarchy, clarify the basic concepts and logical connections between them.

2.2 *Main Definitions of Green Computing*

Green engineering is a special kind of engineering which is based on green technologies as it is often understood. It could be presented in a form of services which tend to improve energy efficiency, safety, security and environmental performance of processes and products in industry.

Green IT is an adaptation of ITs development practice and their application in order use ITs in a more effective manner. It forms a vector of development.

Green computing is special kind of computing. It is necessary to understand the nature and evolution of computing paradigm in general in order to define the concept of green computing precisely. It could be truly said that today there is a concept of computing in a narrow, broad and global sense:

- *in a narrow sense* it is calculations (the set of transformations that are performed with application of a finite number of pre-defined rules) that runs on a computer or in a computer system;
- *in a broad sense* it is a collection of scientific engineering methods and activities aimed at the development and application of computer technology, including hardware and software in different areas for the purpose of information and automation;
- *in a global sense,* computing or *noocomputing* is part of the noosphere, a theory which has developed a great scientist Vernadsky [9]. Nowadays due to globalization process the noocomputing should play a key role. Its formation can be completed within this decade [10].

Definition of green communications in a similar manner is projected onto a computer networks and telecommunications and can be considered as a component of green computing.

Green information technology (green IT) is a set of processes, methods and tools for data collection, storage, processing, supply, distribution of information and methods of their implementation, aimed at improving energy efficiency, safety and environmental technologies themselves and the systems in which they are applied, as well as dissemination of the relevant values in the society.

Green IT engineering is a kind of engineering based on the development and application of green IT in different types of human activity. Therefore, it is possible to distinguish and properly interpret the green computer engineering and green software.

Based on discussed concepts the following components of green IT engineering can be determined:

- *green hardware* which minimizes power consumption and the risk of dangerous failures in safety-critical systems. Thus we can speak about green chips, green microprocessors, green hardware modules, etc.;
- *green software* which minimizes the information and energy systems. They are based on the code which is optimized on the energy metrics.

IT-systems (infrastructures), based on green hardware and/or green software may be called green IT-systems (infrastructures).

Green cloud computing can be defined as technology that can provide potential benefits for the environment and energy savings in the provision of services via the Internet or other distributed computing technologies.

Greenware as the whole is a combination of green hardware and green software, and services that allow the user to minimize the effect of using a computer or computer system on the environment and the cost of expenses for their use and maintenance.

IT greening (or *greenwashing* by Murugesan and Gangadharan [14]) and *green IT reengineering* are terms in the sense of the process and they are particularly important when it comes to developing of energy efficient software, as well as modernization of existing IT-systems of different nature in the interests of reducing power consumption, etc.

Green IT culture forms the values and norms of behavior related and directed on preservation and enhancement of all components of the environment, resources, energy, safety and security by improving the development and implementation of green technologies and information systems, as well as methods of professional and social activities, IT specialists in forming, development, dissemination and adoption of these values and norms.

Green IT business aims at developing and implementation of IT and green technologies, which affirmed the value of green culture, implementing a business organization that minimizes the use of energy and other resources, and direct or indirect CO_2 emissions in creating products and services.

Green IT policy—is a set of goals and activities which regulate the achievement of rational results in the field of green IT, specific indicators on energy efficiency and resource conservation.

2.3 Green Computing Metrics

Energy saving and *energy efficiency* are the main characteristics of green computing. The first one is determined by the consumption of energy, and the second—how efficiently the energy is used. It is about power being measured by capacity P_e.

Thus energy E_e is the complex indicator that can be calculated as the ratio of capacity to be achieved, precision, and other characteristics, or their growth ΔP using technology or IT-systems per watt P_e or its changes like:

$$E_e = \Pi/P_e \quad \text{or} \quad E_e = \Delta\Pi/\Delta P_e \tag{2}$$

Energy saving has a broad meaning and indicates not only the quantitative value of energy savings by using green IT, but also on a set of measures aimed at reducing consumption.

In order to estimate the share of energy P_e, which is consumed by equipment of such system with respect to the total energy P_s of technical complex, enterprise, or any object that is embedded in the software and hardware the simple index MEI (power IT-system metric) is used:

$$\mathrm{PIM} = P_e/P_s \tag{3}$$

When talking about data centers (computing clusters, cloud infrastructures) one should use used metric *PUE* (*power usage effectiveness*):

$$\mathrm{PUE} = P_s/P_e \tag{4}$$

The following metrics could also be used [16–19]:

- *Green energy co-efficient* (*GEC*) which defines the part of the energy P_{er}, which is derived from renewable sources:

$$GEC = P_{eg}/P_e \tag{5}$$

- *Energy reuse factor* (*ERF*) determines the fraction of energy (heat primarily), released during the calculations and work of data centers as a whole, beneficial use in the future P_{er} (e.g. for heating, greenhouses, etc.):

$$EPF = P_{er}/P_e \tag{6}$$

Carbon usage effectiveness takes into account the impact of data centers and similar systems on the environment, the measured CO_2 emissions, and is determined by the ratio of emissions CE caused by a common data center power P_s to the volume of energy consumed for processing information P_e [16]:

$$CUE = CE/P_e \tag{7}$$

For the IT system which components and processes associated with the development and application, the particular indicators, which are sometimes called *green metrics* [17] could be calculated. They are described more detailed in [18, 19] and in broad sense are based on so called GAMES-approach (Green Active Management of Energy in IT Service Centres). Such metrics indicate:

– the proportion of processes aimed at reducing the use of resources, including reducing energy consumption and improving energy efficiency;
– assessment of the relative influence of each of these processes and project activities on resources, energy consumption and energy efficiency;
– degree of improvement in resource characteristics of the products obtained at different stages, etc.

2.4 A Taxonomy of Green IT Engineering

Figure 1 shows taxonomical scheme which describes the links between main concepts of green computing and green IT engineering. In the left part of the scheme four groups of concepts related to the following issues are identified:

- Sustainability and its components (environmental, safety, and resource);
- Engineering, technology and systems;

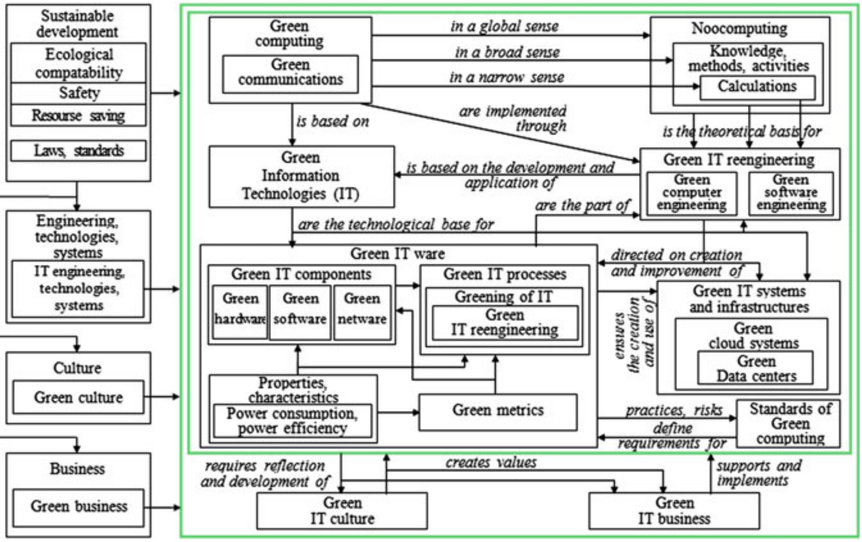


Fig. 1 Green computing taxonomy scheme

- Culture, including green culture;
- Business, including green business.

On the right side, in addition to the above terms the logical connections between them are specified:

- Green computing, noocomputing (noosphere computing);
- Green IT and green IT engineering;
- Green IT components (processes, properties, metrics) and green IT-systems;
- Green IT culture and green IT business.

The work done by Kern et al. [20] summarizes up the aspects that are relevant while creating a label for sustainable software products. They proposed especially criteria and suitable representations for a sustainability label or rather an eco-label and the model proposed in [21] (named as the GREENSOFT Model which is a conceptual reference model for “green and sustainable software”, which has the objective to support software developers, administrators, and software users in creating, maintaining, and using software in a more sustainable way) could be possible used jointly within the framework of proposed taxonomy. Although this model tries to look deep in the process of software development and maintenance, it is limited to software only and should be expanded to cover the hardware of IT, because software cannot be green without green hardware.

3 Values and Principles of Green IT Implementation

3.1 *The Model of Green IT Values*

Information technology and IT-systems create new values by saving human time, by in providing of assistance in making the best decisions in improvement of the quality of products and services, which are understood in the broadest sense of the word.

They are measured by financial indicators such as return on investment (ROI), total cost of ownership (TCO), net present value (NPV), internal rate of return (IRR), etc. [22]. Green IT may allow creation of new social and economic values; it is becoming the norm, which has an impact on human behavior and new hopes. The development of Green IT and IT-systems is the core of a stable environment in the formation of the sustainable environment. Model of green IT values was proposed by Chou [23] and it's based on the principle of so called "value chain", implying the need to analyze each step in any business process in terms of its values what was described by Porter [24].

The value chain includes four components: awareness, translation, comprehension, Green IT values. Analysis of the elements of the model and organizational measures for their implementation in the work of D. Chou allows us to conclude that the model of values focused primarily on project managers and IT companies in general.

3.2 *Risks of Introduction of Green IT in the Context of the Model Values*

The implementation of the described model of values involves a number of risks. Let us specify the background to the analysis of risks in the practice of green IT:

- there are many risks and uncertainties, which can affect and interfere with the practical implementation of green IT and lead to a deviation in the way of increasing the sustainability of development;
- the possible risks that may occur, should be identified and monitored;
- considering that the practice of introducing of green IT becomes more complicated, a variety of measures is needed to include;
- IT companies must follow the law and monitor developed products and complex for their technical support regularly;
- the government must promote the development of social interest in the achievement of sustainable development.

The risk is known to be defined as a combination or multiplication of the probability P (Xi) of an event Xi with its severity events (caused damage) L (Xi):

$$\mathrm{Risk}\ (X_i) = P(X_i) * L(X_i) \tag{8}$$

If necessary calculation the risk for n events (the risk components), the overall risk can be calculated:

$$\mathrm{Risk}(X) = \mathrm{Risk}(X_1) + \mathrm{Risk}(X_2) + \cdots + \mathrm{Risk}(X_n) \tag{9}$$

Analysis of risk factors based on model of the value of green IT in [23] allows us to conclude that these risks are addressed to a greater degree to managers of the companies that develop green IT-systems. For engineering practice, they can be summarized to the following events:

– the expected benefits will not be achieved in full or require additional expenses for its achievement (risk 1);
– the expected effect on energy savings (resources), and decreasing the impact on the environment or safety/security will not be achieved at all (risk 2);
– a positive effect on green characteristics will be followed by a significant or unacceptable deterioration of other technical characteristics (risk 3);
– a positive effect on green characteristics at the particular stage of the life cycle can be absorbed by the negative effect of the introduction of technology or system on another stage of life cycle (risk 4).

These risks need to be analyzed and designed in the early stages, and then refined as the development and implementation of green IT and IT-systems.

3.3 Principles of Green IT-Systems Implementation

In order to mitigate mentioned risks it is important to implement three main principles of Green IT-systems development—"principles of three balances":

3.3.1 The Principle of Balance Based on the Life Cycle Stages of Green IT-Systems

The effect, which is planned to be achieved by implementing green IT, should be calculated taking into account all stages of the life cycle: from the development of technology (systems), its implementation, use and disposal. Thus all the positives and negatives at these stages should be taken into account and evaluated by the integral effect in terms of reducing energy consumption, use of resources, the impact on safety and the environment. Negative effects may be related primarily to the disposal stage. When analyzing it was found that for such stage the materials whose disposal poses additional risks and costs could be used.

It is therefore necessary, on the one hand, to conduct a thorough economic and environmental analysis to provide an integral green effect and make the risks acceptable. IT-systems which provide recycling should be energy efficient themselves and the costs of their creation and use should be considered in the overall ecological and economic calculations.

3.3.2 The Principle of Balance on the Green Characteristics and Other Technical Characteristics of IT-Systems

The effect that is achieved through the green characteristics should be accompanied with suitable changes in the values of other characteristics—by their improvement, which is usually difficult to achieve, maintain or their acceptable decrease.

Acceptability reduction must be justified in terms of cost and /or safety characteristics depending on the type of system.

Reduction of power consumption for energy-critical systems can be achieved by various methods, including through the work of embedded micro-processor or FPGA-based solutions at lower voltages. This leads to increased risk of failures due to unstable operation in such modes. Thus, the power reduction can lead to a reduction of reliability metrics.

On the other hand, there are large and complex dependencies. For example, in the case of redundant (in the simplest case of duplicate systems) the risks of an unstable operation at low voltage channels and the related increase in the intensity of failures can be compensated by the redundancy in a situation when error conditions are not superimposed.

In addition, the work on the low voltage leads to a decrease in heating circuits, and thus reduces the likelihood of sustained failures. Consequently, the design decisions should provide the desired balance of all properties. Even more interesting and complex problems arise when power and reliability is optimized in conjunction with the performance. It is necessary to take into consideration the effect of various system operating frequencies on power consumption and failure operation.

3.3.3 The Principle of the Balance of the Methods and Measures Aimed at the Development and Implementation of Green IT and IT-Systems

The effect of introducing such technologies is possible only with a systematic approach to their development. In fact, green IT and IT-systems should be created using the principles of green computing.

Software and hardware IT-systems, which are introduced for their greening, should consume less energy than it saves during their implementation. For example, to reduce energy consumption and the level of internal interference by reducing the total number of simultaneous triggering elements the management tools which ensure the possibility of diversity of project parts of FPGA (e.g. independent channels of measurement parameters) could be used.

It is obvious that the final positive effect is only possible under condition that the energy, which will be consumed by additional timing control input, is smaller than the difference between the energy consumed by the source and upgradable devices.

A similar problem arises when in the solvers for management of sending (uploading) of tasks in the cloud on energy efficiency criteria is embedded in the mobile solutions. This requires accurate measurement of the energy consumed by the solver to understand the final effect.

Thus, this principle can be analyzed in the context of well-known problems of technical diagnostic ("who will guard the guards?"—which tools will control the means of control …), redundancy theory ("who will provide surgery treatment to the surgeons?"—which tools will be provide redundancy and reconfigure of the tools that support redundancy and reconfiguration), etc. In this case, this problem could be stated as the problem of who will green the "greeners", i.e. which tools will provide reducing of power consumption (energy efficiency) of power consumption tools of the initial system?

It should be noted that the problem of "guards" and "surgeons" also have precisely expressed energy component.

4 Educational Aspect of Green IT Engineering: TEMPUS GreenCo Project

Nowadays there are no a lot of teaching programmes of BSc, MSc and PhD level in green IT in Ukraine and other countries. There is a big need to fill in this gap. Concept, main activities, tasks and structure of project on Green Computing and Communication (shortly GreenCo), were mainly developed at department of computer systems and networks of National Aerospace University KhAI together with colleagues from Newcastle University. Detailed description of consortia, news, meetings could be found on the official website of the project [25]. Detailed evolution of another joint TEMPUS projects is described in [26]. GreenCo project is financed under Erasmus (formerly Tempus) programme beginning 2012 year and its specific outcomes are:

1. To introduce a green computing and communications programme for master students in universities;
2. To introduce a green computing and communications programme for doctoral students in universities;
3. To facilitate intensive capacity building measures for IT tutors;
4. To establish two PhD incubators in Ukraine and Russian Federation on green computing and communications.

Main activity of such TEMPUS-funded project is development of teaching courses, their dissemination for MSc, PhD and LLL (long-life learning) level.

In frame of TEMPUS GreenCo project the teaching courses on the following directions was developed:

- *MSc courses* which form a conceptual base of green ITs and provide review of existed technologies in the field of energy-saving computing, network decisions and decisions for automated process control system and robotic systems. They are intended for IT bachelor graduates to gain an understanding of the green computing methodologies and paradigms, energy efficient system level software such as compilers, hypervisors, monitoring and profiling tools, workload managers, and programming environments, energy aware large scale distributed systems, such as grids and clouds. They are suitable for those aspiring to be software developers, software architecture designers, FPGA developers, experts on distributed infrastructures.
- *PhD courses* which include materials of research and development of "hot-topic" directions in green IT on the level of microcircuits, systems, networks and infrastructures. Moreover the normative regulation issues in the field of green IT and related fields are reviewed.
- *Long-life learning* (*LLL*) *courses* are aimed at practical aspects – engineering techniques and tools for development and management of green IT-systems.

5 Conclusion

Information technologies perform increasingly significant social functions, because of the following:

- the range of services provided to users of all ages and professions is constantly increasing;
- they protect people's health by implementing monitoring functions in embedded medical devices, as well as storage, processing and transmission functions of patient data;
- they manage traffic and control traffic safety;
- support banking operations and payment of goods and services, etc.

Recently the term "pervasive" network, which is based on sensor technology, is appeared. Now we can talk about the pervasive information technologies in general, because they not only affect all aspects of society, but also make it with a more profound impact. At the same time such an impact and pervasiveness become not only natural, but also invisible—virtual.

Given the duality of IT-systems in terms of their impact on safety, security, environment and energy resources, professionals and active users of information technologies should support and promote green culture—a culture of careful attitude to the environment, nature and created resources.

Green IT culture, being part of a common green culture and culture in general, plays a particular, an increasingly important role. This is due to the following reasons:

- the number of IT professionals a variety of professional associations as the elements of civil society is growing rapidly increasing their impact on society (this factor is objective and long-term);
- IT-professionals are one of the most progressive parts of the not only engineers, but of the society as a whole due to the rapid progress of information technologies, their integration with other subject areas, including security, the environment and nature management;
- IT engineering and IT business are important components of globalization, which creates conditions for increasing the influence of professionals working in this area, to all the problems of civilization, including in the area of green technology and green culture;
- since people get acquainted with information technology at an early age, their impact on children and young people especially, and their role and capacity to dissemination of the culture of green values are crucial.

Thus, we can say that information technologies are, on the one hand, an important and powerful tool for education and the promotion of green culture, and IT professionals can become its vanguard. On the other hand, this part of society and technology which it uses are subject to form and develop of their own specific cultural values related to saving energy and resources related to creation of products and implemented processes.

References

1. Feng, W. (ed.): The Green Computing Book, 337 p. Taylor & Francis Group, London (2014)
2. Kharchenko, V., Gorbenko, A., Sklyar, V., Phillips, C.: Green computing and communications in critical application domains: challenges and solutions. In: Proceeding of Digital Technologies (DT'2013), pp. 191–197, Zilina, Slovak Republic, 29–31 May 2013
3. Kharchenko, V., Boyarchuk, A., Brezhnev, E., Gorbenko, A., Phillips, C., Sklyar, V.: Green information technologies: the trends in research, development and education domains. In: Proceeding of ACSN Conference, 2013, 4 p., 16–19 Sept 2013, Lviv, Ukraine
4. The Information Systems Examination Board (ISEB): Project Manager Today, http://www.pmtoday.co.uk/uploads/pmqualificationpageiseb.pdf. Access date: April 2016
5. Atkinson, G., Dietz S., Neumayer E.: Handbook of Sustainable Development. Edward Elgar Publishing (2009). ISBN 1848444729
6. Chen A.J.W., Boudrean M.-C., Watson R.T.: Information systems and ecological sustainability. J. Syst. Inf. Technol. **10**(3), 186–201 (2008)
7. Report of the World Commission on Environment and Development: Our Common Future. Oxford Press (1987). http://www.un-documents.net/our-common-future.pdf. Access date: April 2016
8. Sidorov, N.: Green information systems and technologies. Softw. Eng. **3**(7), 5–12 (2011) (in Ukrainian)

9. Anoprienko, A.: Civilization, noosphere and noorythms. In: Noosphere and Civilization, Scientific Journal, vol. 7(10), pp. 62–69, Donetsk (2009) (in Ukrainian)
10. Wolfengagen, V.: Computing: range of issues and characteristics. http://jurinfor.ru/elibcs/articles/vew09s02/vew09s02.pdf. Access date: April 2016
11. Dick, M., Naumann, S., Kuhn, N.: A model and selected instances of green and sustainable software. In: IFIP Advances in Information and Communication Technology, vol. 328, pp. 248–259 (2010)
12. Kern, E., Dick M., Naumann, S., Guldner, A., Johann, T.: Green software and green software engineering—Definitions, measurements, and quality aspects. In: Proceedings of the 1st International Conference on Information and Communication Technologies for Sustainability (ICT4S2013), pp. 87–94, ETH Zurich, 14–16 Feb 2013
13. Rahma, A., Saoud, B.: Towards a generic sustainable software model. In: Proceedings of 4th International Conference on Advances in Computing and Communications (ICACC), pp. 231–234, 27–29 Aug 2014
14. Murugesan, S., Gangadharan, G.R. (eds.): Harnessing Green IT: Principles and Practices, 389 p. Wiley, New York (2012)
15. Kharchenko, V., Illiashenko, O., Phillips, C., Vain, J.: Green computing within the context of educational and research projects. In: Recent Advances in Computer Science, 19th International Conference on Circuits, Systems, Communications and Computers, (CSCC 2015), pp. 513–518, Zakynthos Island, Greece, 16–20 July 2015. http://www.inase.org/library/2015/zakynthos/bypaper/COMPUTERS/COMPUTERS-84.pdf. Access date: April 2016
16. Maughan, A., PUE, CUE, DCeP.: Can Metrics Rescue Green IT?, pp. 1–15. Morrison & Foerster LLP, 15 April 2014
17. The Guardian: Social media explosion powered by dirty energy, report warns. http://www.theguardian.com/environment/2014/apr/02/social-media-explosion-powered-dirty-coal-greenpeace-report. Access date: April 2016
18. Belady C. (ed.): Carbon usage effectiveness (CUE): a green grid data center sustainability metric. In: The Green Grid, 8 p (2010)
19. Kipp, A., Jiang, T., Fugini, M.: Green metrics for energy-aware IT-systems. In: Complex, Intelligent and Software Intensive Systems (CISIS), International Conference on Date of Conference, pp. 241–248, 30 June–2 July 2011
20. Kern E., Dick, M., Naumann, S., Filler, A.: Labelling sustainable software products and websites: ideas, approaches, and challenges. In: Proceedings of 29th International Conference on Informatics for Environmental Protection (EnviroInfo 2015), Third International Conference on ICT for Sustainability (ICT4S 2015), pp. 82–91. Atlantis Press, Amsterdam (2015). http://www.atlantis-press.com/php/download_paper.php?id=25836154. Access date: April 2016
21. Naumann, S., Dick, M., Kern, E., Johann T.: The GREENSOFT model: a reference model for green and sustainable software and its engineering. Sustain. Comput. Inform. Syst. **1**(4), 294–301 (2011)
22. Lawrence, A.: How to Develop a Green IT Culture, 29 April 2008. http://www.greenbiz.com/blog/2008/04/29/how-develop-green-it-culture. Access date: April 2016
23. Chou D.C.: Risk identification in Green IT practice. Comput. Standards Interf. **35**(2), 231–237 (2013)
24. Porter, M.E.: Competitive Advantage: Creating and Sustaining Superior Performance. Simon and Schuster, New York (2013)
25. Tempus GREENCO project official website. http://my-greenco.eu/. Access date: April 2016
26. Kharchenko, V., Illiashenko, O., Boyarchuk, A., Phillips, C., Vain, J., Krispin, M.: FPGA-based critical computing: TEMPUS and FP7 projects issues. In: Proceedings of the 10th European Workshop on Microelectronics Education (EWME), pp. 74–79. IEEE, Tallinn (2014)

Foresight-Research for Green IT Engineering Development

Igor Shostak, Mariia Danova and Yuliia Kuznetsova

Abstract It is given the basic principles and directions of Green IT Engineering development, as well as the principles of ecological research in the software industry. It is performed the generalized procedure for foresight-project implementation to determine scientific-technical development Green IT Engineering perspective directions using the proposed information technology. It is proposed the approach to construction of foresight technology system model in the form of the two-level hierarchical system consisting of the functional and methodical levels. As a formal basis, methods of bibliometrics and scientometrics (calculating the number of publications, analysis of citing), multi-criterion decision-making problems (t-ordering, Pareto optimality) and patent analysis (trace analysis of the dynamics of inventive activity) are chosen. The result of the foresight-research implementation will be a number of priorities for Green IT Engineering development. In future members of foresight-projects will implement the theoretical development in the form of interactive decision-making support system.

Keywords Green IT engineering · Foresight-research · t-ordering · Pareto optimality · Priority directions development

1 Introduction

In the era of globalization, one of the fundamental priorities of scientific-technical and innovation policies in most countries include ensuring global leadership in high-tech industries and innovative technologies. At that, the efforts of the states are aimed both to promoting research and development activities in the advanced fields of science and technologies and setting national priorities for scientific and technical development. Practically all developed countries periodically create special programs that set priority development areas of science and technology. The

I. Shostak (✉) · M. Danova · Y. Kuznetsova
National Aerospace University "KhAI", 17 Chkalova Street, Kharkiv, Ukraine
e-mail: iv.shostak@gmail.com

V. Kharchenko et al. (eds.), *Green IT Engineering: Concepts, Models, Complex Systems Architectures*, Studies in Systems, Decision and Control 74,
DOI 10.1007/978-3-319-44162-7_2

methods used in development process of these programs proved to be the most effective tool of choosing the priorities in the field of science and technologies.

The concept of "foresight" arose in the 1950s in the "Rand" American Corporation that solved the problems of determining the perspective military technologies. In view of insufficiency of traditional predictive methods (quantitative models, extrapolation of current trends, etc.), "Rand" experts developed Delphi method [1], which became a basis of most foresight-researches. During the 1960s large-scale operations on predicting were carried out by Naval and Air-Force establishments of the USA. Since the 1980 the foresight has already been used in the European countries [1]. The foresight reached its peak as an analysis method in the mid-nineties within national programs of technological prediction, which still remain the main scope of application of foresights. For today, the foresight is used as a system instrument for future development that allows one to consider changes in all areas of society, namely science and technologies, economy, social and public relations, culture.

1.1 Background of the Work

Today, in most countries of the world (USA, Japan, Great Britain, France, Sweden, Russia and others) and in Ukraine, in particular, the foresight technology proved to be the most effective tool for selection of priority directions in the field of science and technology.

In Japan since 1971 the foresight results determine the development forecast of science and technologies in the country for the next 30 years [2]. Thus, the results are "adjusted" each 5 years. The main stages of foresight in Japan included the analysis of trends in the world of science and technology; making the list of perspective "subjects" of economic, scientific, technical and social development; two-round survey of experts according to the Delphi method; ranging the chosen subjects according to the extent of their innovative importance; making the list of national scientific and technical priorities as well as critical technologies.

In the USA in recent years the main efforts at the federal level were aimed to drawing up a list of technologies that are critical to the national economy. Due to the fact that the researches are not general one of the most difficult challenges facing the US government in the analysis of foresight research was bringing together the different lists of critical technologies and identifying similarities between them, as different researches used techniques different in the level and quality (survey of experts, expert panels, technology road maps), which were often formal and were not the result of original research [3].

In Great Britain there were carried out three foresight-programs organized by the Government [4]. "Foresight-1" (1994–1999) program took into account only the technological and market prospects. "Foresight-2" (1999–2002) and "Foresight-3" (2002–2004) are characterized by the integration of technological, market and social problems, as well as the involvement of a wide range of participants. The

Government committee heads the program that includes representatives of 17 ministries and departments. Each program consists of three interconnected stages, namely the analysis, distribution of information and application of results, preparation for the following program. As a result the state priorities are set in scientific and technical programs, in staff training and methods of state regulation. The budgetary priorities are formed based on five-year plans and since the 1990—with regard to long-term (15–30 years) forecast priorities of foresight and scenarios. The main mechanism of implementation is a LINK program of promotion of the cooperation and the methods are Delphi, expert panels, scanning of technologies and scenarios.

In Germany, two rounds of Delphi were carried out in 1993 and 1998 [5], the results of which were used in the formation of national research and development policy. Their initiator was the Federal Ministry of Education, Science, Research and Technology, and Institute of systems and innovative researches Fraunkhofer carried them out in Karlsruhe. The general objective of these researches was not predicting the future, but preparing possible scenarios and gathering information for decision-making, as well as preparing the scientific basis for national debates on the development of the country future. The result of foresight studies in close cooperation with Japanese colleagues ("Delphi-93" and "Delphi-98") has become its own unique program FUTUR (another name—"The German Research Dialogue") that allows one to create a strategic vision for the Ministry of Education and science for the next 20 years.

Implementation of technological foresight in France in 1999–2000 was carried out in four stages [6]. A distinctive feature of the French foresight involves its organization in two mutually parallel directions. The first of them was implemented by the Ministry of Higher Education and Research, and was carried out by Delphi method through most experts (about 3.5 thousand). Its strategic objective laid in studying important technological development in various areas that would interest economy and society. The second direction was in charge of the Ministry of Industry and its objective laid in studying the technologies that were critical for the French industry in the next 30 years. As a methodological basis of the research, there were used bibliometric and patent analysis, as well as expert evaluation.

The history of foresight-researches in Russia is nearly 15 years. Over the years, various departments and research groups implemented projects aimed at determining the prospects for the development of various scientific and technological areas, sectors of economy and regions. However, the long-term forecast of scientific and technological development in the Russian Federation for the period to 2025 prepared by the Ministry of Education and Science of the Russian Federation in 2007–2008 [7] became the most large-scale and rather complex research. The forecast was developed within the Federal target program called "Research and Development in the Priority Directions of Development of Scientific and Technological Complex of Russia for 2007 – 2012" with the application of the methods, namely "critical technologies", "expert panels", "scenarios", as well as with the direct support of Federal agency on science and innovations.

The forecasting and analytical research carried out in Ukraine, within the State Program on Forecasting of scientific and technological innovation and development approved by Cabinet of Ministers involved more than 700 experts. The Ukrainian foresight methodology [8], which objective was forming state scientific-technological and innovative priorities provides implementation for a set of stages, by questioning a group of experts based on the Delphi method [9]:

- the first stage involves solving the problem of selecting of experts to participate in the foresight-research; when forming the group of experts the method of "snowball" is used;
- the second stage, using of expert surveys, as well as "benchmarking" (comparison with other countries or regions) involves forming a preliminary list of subject directions and the criteria for their evaluation of the main objectives of the country (scientific-technical development) STD;
- the third stage involves estimating by experts the obtained subject directions by the set of criteria. This procedure is carried out by means of Delphi method that assumes survey (questioning) of experts and organization of feedback (through carrying out three rounds of survey). Results of research include the evaluation summary for each topic, as well as analytical reviews of the subject directions;
- the final stage involves coordinating and approving the lists of subject directions of the country STD obtained in the previous stage.

1.2 Work Related Analysis

Recently a number of publications devoted to research of theoretical and practical aspects of the foresight (the technology of future development) is increasing [9–12]. The foresight was started to be used since the beginning of the 1990s by the governments of the U.S., United Kingdom, Germany, Japan and Australia, and by 2000 the number of such countries exceeded 30. At present this technology has been adopted not only in Western Europe, U.S. and Japan, but also in a number of developing countries and countries with economies in transition (Poland, Czech Republic, Russia, Belarus, Ukraine, etc.) [1–12].

The foresight is the systematic process that is aimed at identifying new strategic directions of scientific and technological achievements, which in the long run will have a serious impact on the economic and social development of the country [1]. The foresight technology has the following main characteristics: the foresight is a systematic process; the central place in this process is taken by scientific and technological directions; priorities are considered with regard to their influence on social and economic development of the country; time horizon is considered in the average and long terms.

In the framework of foresight there is a question about an assessment of possible prospects of innovative development connected with the progress in science and technologies, as well as about possible technological horizons that can be reached

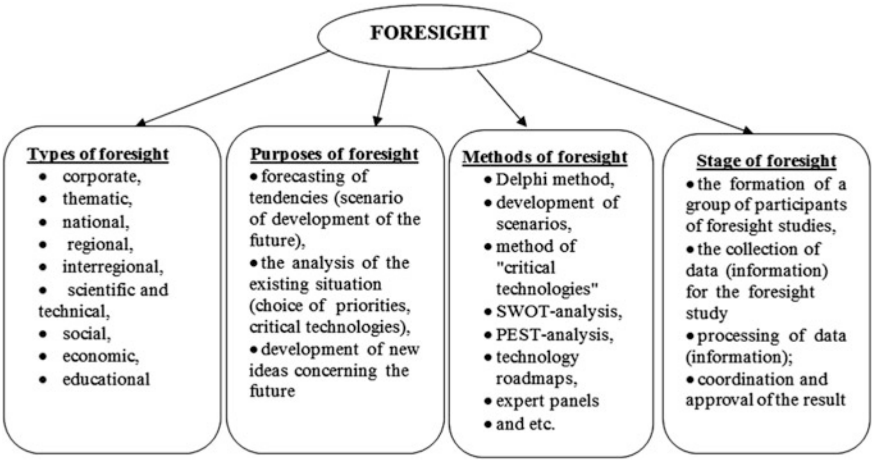


Fig. 1 Structural elements of technology foresight

in case of investment of certain funds and organization of systematic work, and probable effects for economy and society.

In practice, there are various types of the foresight [1, 13] as each foresight-project differs from others in various parameters: the contents, depth of analysis, scales, time frames, territory of coverage, number of participants, available resources, etc. Because of the analysis of the foresight in different world countries, it is possible to allocate its basic structural elements (Fig. 1).

Analysis of the publications from domestic and foreign sources [1–15] devoted to the study of theoretical and practical aspects in foresight methodology application showed that there was a need for further study of the problem related to the automation of foresight technology. The authors analyzed the way of automation for each stage in the national foresight-project [8, 14]. As a result of the analysis it was offered information technology support of foresight projects in Ukraine [15], based on the specialized techniques for selecting the priorities [8]. The adoption of specified technology will allow increasing the efficiency of national foresight-projects due to their complete computerization based on formal methods.

1.3 Goals and Approach

A **goal of the paper** is to consider the generalized procedure for foresight-project implementation to determine STD Green IT Engineering perspective directions using the proposed information technology. The result of the foresight-research implementation will be a number of priorities for Green IT Engineering development. As a formal basis, methods of bibliometrics and scientometrics (calculating the number of publications, analysis of citing), multi-criterion decision-making problems (t-ordering, Pareto optimality) and patent analysis (trace analysis of the dynamics of inventive activity) are chosen. The complex automation approach of foresight-projects is expounded. In future members of foresight-projects will implement the theoretical development in the form of interactive decision-making support system.

The technology of carrying out the foresight-research consists of the following steps:

1. Formation of expert panels, i.e., number and structure of expert members participating in forecasting based on the evaluation of their competence level is formed.
2. Forming the initial list of the Green IT Engineering directions. It is necessary to analyze the status and prospects of Green IT Engineering development using methods of bibliometrics (calculating the number of publications), scientometrics (analysis of citing) and patent analysis (trace analysis of the dynamics of inventive activity). Then, for the obtained lists of trends of the Green IT Engineering development it is calculated value of criteria for their assessment provided by technique of foresight-researches. Thus, all “leading” directions have quantitative assessments for each of the criteria, which will help to determine the number of priority.
3. Selection of priority directions for Green IT Engineering development. Initial data for the selecting of priorities is a list of directions of Green IT Engineering, as well as a set of values of criteria assessment for each direction. The procedure for the selection of priority directions of Green IT Engineering development is in ranking these directions criteria using the Pareto optimality principle and t-ordering method.
4. Coordination and approval of Green IT Engineering priority directions. In accordance with the current technique the strictly regulated procedure for coordination and approval of priorities is carried out.

2 Research of Foresight Technology Structure as a Complex System

For computer implementation of foresight technology it is necessary to formalize it, by development of the model, with a known level of adequacy reflecting all stages of the foresight-research. The foresight technology consists of a set of interacting components—subprocesses (foresight-research stages) and possesses the following row of the main properties:

- hierarchies—actions of subsystems of the top level depend on the actual execution by the bottom levels of their functions;
- discretizations—correlation between elements of system (subprocesses) is carried out sequentially;

– determinancies—all transitions between elements of system are strictly defined, i.e. transition to the next happens on condition of successful execution of the previous subprocess.

To select a special mathematical apparatus that would consider the selected main characteristics of foresight-researches, it is appropriate to conduct modelling which will allow to receive an holistic, systemic view model, allowing to reveal correlations of elements foresight technology, possible states of each element and the relation between them.

Foresight technology representable in the form of multilevel structure described from positions of systems theory [16]. This structure should reflect the most important characteristics of the simulated system, namely:

– that the foresight technology consists of interrelated subsystems, which have a right to make decisions;
– that these subsystems form a hierarchy.

On the assumption of that, theoretical-system model of foresight technology represented in terms of functional-method relations, in which a function is understood as subject needs for receiving some result, and a method—some process by which this result is obtained. In this case, the upper level of foresight technology is set by the functional part and the lower—the methodical.

On the functional level displayed the main purpose of the foresight technology (analysis of the existing situation (the selecting of priorities, the search for critical technologies), prediction of tendencies (described of development scenarios), etc.), which in turn can be decomposed into sub-goals (stages), depending on the type of foresight (national, corporate, scientific and technical, social, etc.). The methodical level reflects the transition way from the input to the output elements and contains as methods of achieving the purpose (set of methods for foresight technology), and conversion technology from input to output elements (certain methods and order for their implementation).

The provided description of foresight technology is initial for its formalized description on the set-theoretic level, at that the general problem statement of synthesis the models of computer implementation of foresight technology, given on the set-theoretic level of description, is similar in its structure and content to the general problems statement of decision-making problems in the conditions of a multicriteriality. A formal set-theoretic description is carried out using the following model:

$$\Phi = \langle \Theta, G, \Lambda, \Omega \rangle, \tag{1}$$

where: $\Phi = \{\varphi_g\}, g = \overline{1, m}$—the set of types in foresight technology; $\Theta = \{\theta_i\}, i = \overline{1, k}$—set of goals in foresight technology that indicate a given state of foresight technology; $G \subset G_0 \cup G_1$—set of initial G_0 and final G_1 states of foresight technology, at that $G_0 \cap G_1 \neq \varnothing$; $\Lambda = \{\lambda_q\}, q = \overline{1, z}$—set of stage in foresight technology; $\Omega = \{\omega_j\}, j = \overline{1, n}$—plural sets of methods foresight technology.

For the description of dynamic properties by means of system model, we will enter the following concepts:

1. Time T is a linear ordered by the relation "$\leq$" a set of time points t:

$$T = \langle T^o = \{t\}, \leq \rangle : t \in [t_0, t], \tag{2}$$

where: t_0—initial time point in the interval [t_0, t];

2. The instantaneous state of foresight technology Φ for a period of time T:

$$\Phi^T \subset \Phi \times T : \Phi^T = \{\Phi^t\} \wedge \Pi_T(\Phi^T) = \{t\} \subseteq T, \tag{3}$$

where: $\Pi_T(\Phi^T)$—projection of a set of the instantaneous statuses Φ^T to a set of time points T.

$\Pi_T(\Phi^T)$ is understood as a set of those elements from T which are projections of elements from (Φ^T) on *T*. For a couples $\Phi^t = (\Phi, t)$ projection element Φ^T on the set T serves the element t. Instant state of foresight technology stage $\Phi \in \Phi^T$—making in a time point t some event (e.g. appearance of data at the input or output of stage foresight technology), relating to this stage.

The initial G_0^T and final G_1^T states of foresight technology represent the sets, including prehistory of its instantaneous states for the period of time T to some initial and final time points.

Using the introduced concepts, build the functional part $\Phi^{\Phi T}$ of a system model of the foresight technology as follows:

$$\Phi^\Phi \subset P(G_0^T \times \theta_i^T \times G_1^T). \tag{4}$$

In the particular case Φ^Φ can be given mapping

$$\Phi^\Phi : (G_0^T \times \Theta^T) \to G_1^T, \tag{5}$$

where: G_0^T, G_1^T—initial and final states of foresight technology for the time period T; Θ^T—set of foresight technology goals for the time period T.

Mapping (5) shows, that the final status of any of the stages of foresight—a function of its initial state and specified goal of this stage, specifying these states.

Methodical part of the model Φ^M is intended to describe the possibilities on any stage in foresight technology to reach certain values of the output value for the fixed initial state and a specified goal, i.e. reflects a transition way from input elements to output:

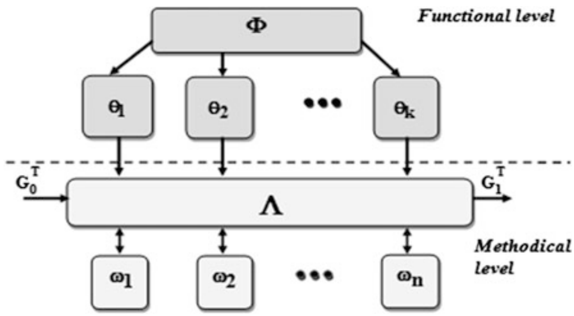


Fig. 2 Structural diagram of foresight technology system model

$$\Phi^{M} : (\Theta \times \Lambda) \to \Omega, \quad (6)$$

where: $\Theta = \{\theta_i\}, i = \overline{1,k}$—set of foresight technology goals that indicate a given foresight technology state; $\Lambda = \{\lambda_q\}, q = \overline{1,z}$—set of foresight technology stage; $\Omega = \{\omega_j\}, j = \overline{1,n}$—plural sets of foresight technology methods.

As a result, the system model (SM) of foresight technology will be the following:

$$\begin{aligned} &\Phi = (\Phi^{\Phi} \times \Phi^{M}) \\ &\Phi^{\Phi} : (G_0^T \times \Theta^T) \to G_1^T, \\ &\Phi^{M} : (\Theta \times \Lambda) \to \Omega \end{aligned} \quad (7)$$

where: Φ^{Φ}, Φ^{M}—functional and methodical parts of foresight technology SM; G_0^T, G_1^T—initial and final states of foresight technology for the time period T; Θ^T—set of foresight technology goals for the time period T.

Foresight technology SM is shown in Fig. 2. On the input of stage foresight technology when he is in some initial state, incoming input action in the form of given purpose, which upon reaching of it this stage should go to the desired finite state.

3 Approach to Complex Automation of Foresight-Projects Defining Scientific and Technological Priorities in Green IT Engineering

Results of conducted foresight—researches in majority developed countries of the world showed that one of the priority directions development are energy efficiency and energy saving, as well as decrease in negative influence on environment and improvement of an ecological condition of the planet. The volume of energy consumption is not only important indicator of economic development, but also scales of environment pollution, including of atmospheric air pollution. A main objective of Green Engineering [17] is decrease in negative influence on environment at the help of science and equipment [18–20]. In [21] 12 basis principles of Green Engineering are given:

1. Designers need to strive to ensure that all material and energy inputs and outputs are as inherently nonhazardous as possible.
2. It is better to prevent waste than to treat or clean up waste after it is formed.
3. Separation and purification operations should be designed to minimize energy consumption and materials use.
4. Products, processes, and systems should be designed to maximize mass, energy, space, and time efficiency.
5. Products, processes, and systems should be "output pulled" rather than "input pushed" through the use of energy and materials.
6. Embedded entropy and complexity must be viewed as an investment when making design choices on recycle, reuse, or beneficial disposition.
7. Targeted durability, not immortality, should be a design goal.
8. Design for unnecessary capacity or capability (e.g., "one size fits all") solutions should be considered a design flaw.
9. Material diversity in multicomponent products should be minimized to promote disassembly and value retention.
10. Design of products, processes, and systems must include integration and interconnectivity with available energy and materials flows.
11. Products, processes, and systems should be designed for performance in a commercial "afterlife".
12. Material and energy inputs should be renewable rather than depleting.

Thus, in order to harmonize economic, social and environmental factors affecting the sustainable current state of society and the satisfaction of its potential needs it is appropriate to study the processes of ecosystems that are software products. Application of ecological researches in the software industry shows that their distribution goes on the basis of three principles [20–22]:

– eco-efficiency—assumes a combination of the efficiency purposes in traditional sense with the ecological purposes and the direction of their achievement on improvement of life quality and reduction of environment influence, i.e. where and how the software has to be applied to reduce use of natural resources and to lower an adverse effect on environment;
– eco-equity—assumes distribution of resources between the living and future generations taking into account the concept of a sustainable development [23];
– eco-effectives—assumes creation of "pure" (waste-free) systems and technologies in which waste of one processes is sources for other processes and finally their application with advantage is carried out. An applying of the principle is directing on solution of the pollution problems by reducing waste.

As the directions for foresight-research of Green IT Engineering STD definition is expediently to use the following directions [23–25]:

- Efficiency improvements of the software and algorithms (“green” software development):

  energy-saving green software—software, that focuses on the principles of software using, ensuring reduction of harmful effects (direct or indirect) on the environment;
  green software engineering—software development with the least harmful effects on the environment;
  software ecosystem—an artificial complex that includes the software, environment of its development, operation, maintenance and disposal, which are connected by sharing software and intelligence.

- Effective power supplies development (with high efficiency);
- Green telecommunications development;
- Use of the lowered supply voltage, voltage control and synchronization frequency of the equipment;
- Virtual calculations development and use of Plug computers.

3.1 Formation of Expert Panels

Expert members participating in forecasting—the key parameter defining quality of future forecast. Thus, for high-quality examination it is necessary to solve the following problems:

1. to assess the competence level of expert’s;
2. to determine the number of expert’s;
3. to form the final list of expert members participating in forecasting.

Denote by Q the experts set, then to evaluate the competence level of each i-th expert (i = 1, …, m) will use the generalized indicator of the competence level (K_i) given in [26], which takes into account both the professional activity, and personal qualities of experts:

$$K_i = \frac{1}{4}\sum_{j=1}^{4} K_{ij}, \tag{8}$$

where K_{i1}—coefficient reflecting the level of professional training and knowledge of the ith expert (it takes into account the levels of qualification such as “Doctor of Science”, “PhD” etc., and is measured in points $0.5 \le K_{i1} \le 1$);

K_{i2}—coefficient reflecting the level of the basic coherence of ith expert when decision making (it takes into account factors such as intuition, production experience, theoretical analysis etc., and is measured in points $0.05 \le K_{i2} \le 1$);

K_{i3}—coefficient reflecting the personal qualities of the ith expert and on the basis of self-assessment is calculated ($0 \le K_{i3} \le 1$):

$$K_{i3} = \frac{1}{n}\sum_{j=1}^{n} K_{i3_j}, \tag{9}$$

where K_{i3_j}—coefficient reflecting a self-assessment of the ith expert in existence of his jth personal qualities; n—number of expert personal qualities;

K_{i4}—coefficient reflecting the personal qualities of the ith expert and by colleagues experts is calculated ($0 \le K_{i4} \le 1$):

$$K_{i4} = \frac{1}{n \cdot m}\sum_{l=1}^{m}\sum_{j=1}^{n} K_{i4_{jl}}, \tag{10}$$

where $K_{i4_{jl}}$—coefficient given lth expert about existence of jth-personal qualities in the ith expert; n—number of expert personal qualities; m—number of experts participating in the assessment of the ith expert.

As a criterion for assessment of the required number of experts, the following formula is used:

$$N_{min} = 0.5(3/\varepsilon + 5), \tag{11}$$

where N_{min}—the minimum required number of experts; ε—parameter setting a minimum level of an error expertize ($0 < \varepsilon \le 1$).

In case of an admissible error expertise in 5 % (ε = 0.05) the number of experts should be at least 32. According to [26] the necessary number of experts for the assessment must be at least 7–9 people, therefore, the number of experts involved in forecasting, is in limits $7 \le N \le 32$.

To receive the final list of all experts who passed certification is range on the competence level (value of the generalized indicator K_i) and according to the relation (11) the list of the expert members is formed.

3.2 The Initial List Formation of the Subject Directions in Terms of Green IT Engineering

Within this stage, there is a formation of the initial list of the directions, for its subsequent assessment, and values of criteria of their assessment are calculated.

We will break the task set above into a row of subtasks:

1. to form the initial list of subject directions (SD), for their subsequent assessment;
2. to calculate the values of the quantitative criteria assessment on each of the directions;
3. to get the values of qualitative criteria of an assessment on each of the directions.

It is necessary for the solution of first formulated subtasks to use methods of bibliometrics, scientometrics and patent analysis to form the initial list of SD. For solution of two remaining tasks, it is necessary to calculate values criteria from system criteria assessment for each SD. As a result, more perfect lists of STD SD, each of which has the quantitative assessment based on the received values of criteria of their assessment that further will allow defining a row of the priority is formed.

Formation of the list of SD will made by methods of bibliometrics, scientometrics and patent analysis.

Bibliometric Analysis. "*Calculation the number of publications*" *method* consists in the following: need to made the quantitative assessment of the document flow that organized within one of the accepted classifications i.e. the analysis of amount of abstracting scientific documents number from different knowledge domains is carried out. Thus, it is selected the areas of science and separate sections, which take the leading place in structure of scientific knowledge on the number of scientific publications. After that, it is compared the amount of publications on separate branches to select the "leading" branches of knowledge (as a percentage of total quantity of the abstracting publications for this period).

Scientometrics Analysis. The *analysis of citing* is carried out by research of bibliographic links in publications of scientific periodical press databases (Web of Science, SCOPUS, the Russian index of scientific citing, National library of Ukraine of V. I. Vernadsky and etc.) for the quoting of the publications detection that form a certain direction of science (the direction of science, number of links [self-citing is excluded], total quantity of publications in the direction). The leading scientific directions is selected upon results analysis of citing.

Patent Analysis. In the patent analysis, for obtaining the quantitative characteristics of the certain directions of science and technique development, statistical techniques of patent information arrays processing are used. These include the trace analysis of the dynamics of inventive activity in each scientific and technical direction, which consists of the cumulative rows patenting construction characterized by increase of summary number of the patents relating to this direction. Intensity of the researched direction development is define by a slope ratio of curve dynamics of patenting to a timebase. On its basis, calculate the coefficient of ponderability for the researched SD, which is accepted as the quantitative index of the researched direction prospects:

$$D_i = \frac{tg\alpha_i}{\sum_{i=1}^{n} tg\alpha_i}, \qquad (12)$$

where D_i—ponderability coefficient of ith SD, and $\sum D_i = 1$; $tg\alpha_i$—a slope ratio of tangent line to a curves of the dynamics of inventive activity of ith SD; $\sum tg\alpha_i$—sum of slope ratios tangent lines to curves of the dynamics of inventive activity of all SD.

The ranged list of SD for further calculation of criteria values of their assessment we will receive because of the first delivered subtasks solution.

In the procedure of initial lists SD formation the following step is the solution of the second and third subtasks, namely—calculation the values of the criteria assessment on each SD. The specified criteria, according to their values, divide into the quantitative and qualitative, received by expert estimation.

The quantitative criteria (expressed by the quantitative values) are—financing (K_f); presence of scientific experts (K_{se}); productivity of scientific and technical activities (K_{sta}); existence of production capacities (K_{pc}) and existence of experimental and production basis (K_{epb}).

The qualitative criteria is received by expert estimation include applicability (K_a), presence of the potential investor (K_{pi}), output perspective to the world market (K_{wm}), and presence of the scientific leader belong (K_{sl}).

For determination of values on each of qualitative criteria experts need to fill in the polling questionnaire. For processing of group expert estimates it is necessary to check a level of coherence of experts judgments, and then the generalized (aggregated) group assessment for each SD by each of criteria is determined.

As exponents of coherence of experts judgments apply the variation coefficient characterizing the relative dispersion of result:

$$V_j = \frac{\sqrt{\frac{\sum_{i=1}^{m_j} (x_{ij} - \overline{x_j})^2}{m_j - 1}}}{\overline{x_j}} \cdot 100\ \%, \tag{13}$$

where V_j—variation coefficient of estimates on jth SD; m_j—amount of the experts estimating jth SD; x_{ij}—an assessment the ith expert of jth SD in points; $\overline{x_j}$—average value an assessment of SD in points:

$$\overline{x_j} = \frac{\sum_{i=1}^{m_j} x_{ij}}{m_j}. \tag{14}$$

The coherent a consensus of experts when value of coefficient variation is less. If coherence of experts judgments is absent, repeated questioning is carried out. To experts, except polling questionnaires, additional information on a subject of expertize is sending and they, as a rule, adjust the estimates. The corrected information comes to analytical group for coherence check again.

For calculation of the aggregated group assessment, we will use method of average ball estimates. Taking into account the weight coefficients of the experts, group evaluation of jth SD is calculated as a weighted average:

$$x_j^{CB} = \sum_{i=1}^{m_j} (K_i \cdot x_{ij}), \tag{15}$$

where K_i—weight factors of experts competence; m_j—amount of the experts jth TD estimating; x_{ij}—an assessment in points by ith expert of jth SD.

3.3 Selection of Priority Directions in Terms of Green IT Engineering

At this stage, it is necessary to select the STD of priority directions by an assessment among themselves of SD received at the previous stage i.e. to evaluate SD on values of their criteria. The assessment of SD occurs by principle of Pareto optimality and a method of t-ordering (for narrowing of Pareto area). Priority directions are SD, whose vector estimates make the Pareto set. We will consider in detail each of methods.

The mathematical model of the task about a choice most of the priority directions of STD can be presented in the form

$$B_f = \langle X, f_1, f_2, \ldots, f_m \rangle,$$

where X—a set of SD; f_j—the numerical function defined on a set X, at that $f_j(x)$ is assessment a SD $x \in X$ on jth criterion assessment (j = 1, m).

The purpose of the given problem solution is consisted of getting the directions having as much as possible highest marks by each criterion i.e. in allocation of Pareto set [8, 15]. All criteria functions f_j reflect usefulness of SD $x \in X$ from the viewpoint of different criteria assessment and shall be commensurable, i.e. values of each criteria function change in the same limits [a, b]:

$$\forall x \in X : 0 \leq a \leq f_j(x) \leq b, j = \overline{1, m}. \tag{16}$$

At that the least preferable SD on any of private criteria $f_j(x)$ will receive a mark **a** and the most preferable—a mark **b** (a = 0, b = 1).

The numerical functions $f_j(x)$ (j = 1, m), stated above, is formed the vector criterion $f = (f_1(x), f_2(x), \ldots, f_m(x)) \in R^m$, where R_m—field of m-dimensional vectors. For any of SD $x \in X$ a set of its marks by all criteria, i.e. the set $(f_1(x), f_2(x), \ldots, f_m(x))$ is a vector mark of SD x. All possible vector marks will formed a set of possible marks $Y = f(X) = \{y \in R^m | y = f(x)$ in case of some $x \in X\}$.

The domination relation by Pareto is determined as follows: SD x_i dominates according to Pareto of TD x_j, if the vector mark $f(x_i) = (f_1(x_i), \ldots, f_m(x_i))$ of SD x_i dominates according to Pareto a vector mark $f(x_j) = (f_1(x_j), \ldots, f_m(x_j))$ of SD x_j, i.e. if the inequality $f(x_i) \geq f(x_j)$ is executed, consequently $x_i \geq x_j$. The domination condition on Pareto in informative means, that SD x_i isn't worse, than SD x_j on any of the considered criteria, and at least, on one of these criteria x_i is better than x_j.

So, according to values criteria functions of SD we will receive their vector marks. For searching of a set of Pareto-optimal vectors P(Y) we compare them among themselves by the rule described above. If the received couples are incomparable on Pareto relation, then setting up the task of narrowing of the initial set by means of a t-ordering method, for choice of several SD as final output.

Setting up the task for choice of several SD as final output by means of narrowing of Pareto set. One of such methods is the method of t-ordering [8, 15] using ordinal information of the person making decision (PDM) on the relative significance of criteria.

As the initial information for a method of t-ordering the set S of expressions PDM about the relative importance of criteria assessment of a look is accepted:

$$S = \{f_k = f_j; \ldots; f_q > f_p\}, \tag{17}$$

which needs to be expanded due to adding of the new transitive expressions which are consequences of the already available.

Taking into account the received set (17) when comparing two vector marks of SD the preference relation on a t-ordering method is built:

$$\begin{gathered} f(x_z)^t \succ f(x_w) \leftrightarrow \left[\exists f''(x_w) \in f(x_w)I : f(x_z) \overset{p}{\underset{\sim}{\succ}} f''(x_w)\right], \\ f(x_z)^t \overset{p}{\underset{\sim}{\succ}} f(x_w) \leftrightarrow \forall j \in f[1:m] : f_j(x_z) \geq f_j(x_w)] \end{gathered} \tag{18}$$

where $f(x_Z)$, $f(x_W)$—SD vector estimates ($f(x_Z) = (f_1(x_Z), \ldots, f_m(x_z))$;

$f(x_W) = (f_1(x_W), \ldots, f_m(x_W)))$;

$f(x_W)I$—set $f(x_W)$—the improved vectors ($f_k = f_j; f_q > f_p$).

Thus, basic data of the task are a set (16) of ordinal informations on the relative importance of criteria and a set of vectors, incomparable on Pareto. Selecting a couple of vectors $f(x_i)$ and $f(x_j)$ for comparing of their vector estimates. The vector f (x_i) is fixed, and on a vector $f(x_j)$ we receive sets of the improved vectors $f(x_j)I$ according to (18). After conversions we receive the vectors, comparable according to Pareto. As a result we receive a set of Pareto-optimal vectors, and so—the priority directions of STD.

4 Results and Analysis

As initial data in the foresight-research for the selection of priority directions of STD Green IT Engineering in Ukraine are used statistical data for 2010–2015 years.

For an objective analysis of each direction of Green IT Engineering in the beginning of foresight-research the following assumptions of sufficiency were entered: information from the selected sources; specified time intervals; values of the basic criteria calculation for their assessment.

The result of the foresight-research implementation has become a number of Green IT Engineering priorities in Ukraine.

Course of foresight-research according to the developed applied IT [15], based on the foresight-research system model implementation and decision-making

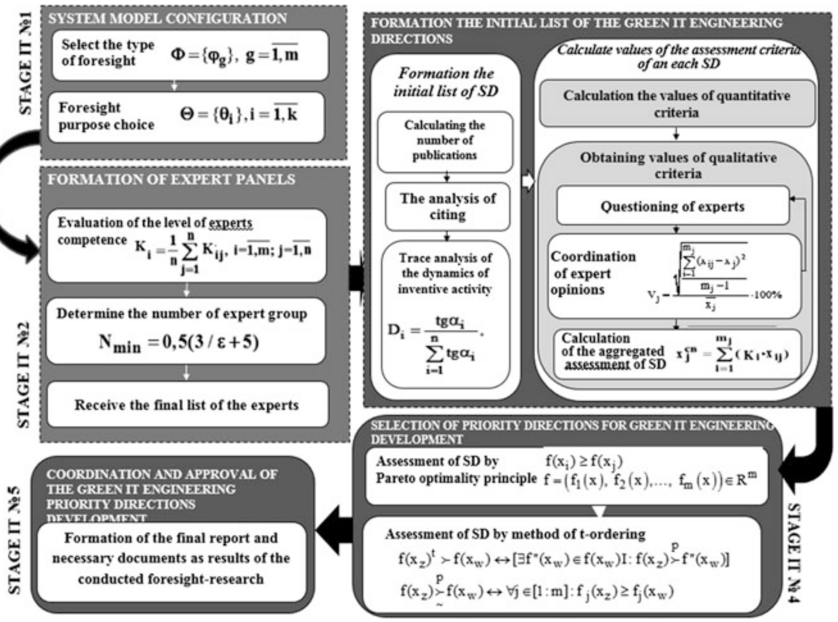


Fig. 3 The structure of applied information technology for foresight-researches of priority directions selection of STD Green IT Engineering

support methods for STD priorities selection that are proposed in the previous sections, consists of five main parts (stages) (Fig. 3): SM configuration; formation of expert panels; formation of the initial list of Green IT Engineering directions; selection the priority directions of Green IT Engineering; coordination and approval of the Green IT Engineering priority directions.

The first stage provides the configuration of foresight-researches and consists of the procedures of selection the type and objectives of foresight-research. At the second stage it is carried out operations such as calculations of required number of experts for research and generalized indicator in level of competence for each expert based on the coefficients reflecting both training and personal qualities of the expert. Further experts are ranged on value of the generalized indicator level of their competence and it is formed the final list of experts taking into account the number of expert group.

Thus, in the context of foresight-research there were attested 12 candidates in experts, 7 of them were selected for further involvement in expertize that enters the calculated earlier allowable range.

Within the third stage there is a formation of the initial list of the SD and calculating the values of criteria for their assessment. The initial data of this stage, in view of the assumptions about sufficiency made at the beginning of the research, are the following directions of Green IT Engineering development [23–25]:

1. Energy-saving green software.
2. Green software engineering.
3. Software ecosystem.
4. Development of green telecommunications.

To form the initial list of SD it is used the methods of counting the number of publications to analyze of the Ukrainian documentary flow in the Green IT Engineering directions. The results of this analysis for the 5-year time period (2010–2015 years) are shown in Table 1.

Thus, according to the results of the bibliometric analyze in the initial list for identifying the priorities of the Green IT Engineering included those that received the highest marks—namely, energy-saving green software, green software engineering and development of green telecommunications.

The next step is calculation of criteria values provided by the developed IT [15], for each initial direction, taking into account the assumptions made by us (Table 2).

At the fourth stage there is processing of the results obtained at the previous stage taking into account ordinal information from decision-making person (DMP) on the relative importance of assessment criteria of the SD. Comparison of the approved SD by means of the principle of Pareto-optimality is carried out. If comparison of the SDs across Pareto is impossible, we apply the t-ordering method considering ordinal information of the DMP on the relative importance of assessment criteria for the SDs. As a result we obtain the ranged list of priority directions of STD Green IT Engineering.

Table 1 Distribution of scientific publications in the Green IT Engineering directions for 2010–2015 years

№	Direction	Number of scientific papers (articles, abstracts and books) (pcs)	Specific weight in the total number of publications (%)
1.	Energy-saving green software	87	28
2.	Green software engineering	112	36
3.	Software ecosystem	51	16
4.	Development of green telecommunications	64	20
Total		314	100

Table 2 Values of main coefficients for Green IT Engineering directions

№	Direction	K_{se}	K_a	K_{wm}	K_{sl}
1.	Energy-saving green software	0.31	25.34	9.78	10.89
2.	Green software engineering	0.44	29.36	8.56	11.3
3.	Development of green telecommunications	0.17	20.11	5.09	7.23

To find the set of Pareto-optimal vectors P(Y) obtaining the vector estimates SDs, the method described in Sect. 3.3:

$$\begin{aligned} y_1 &= (0.01; 0.86; 0.33; 0.38) \\ y_2 &= (0.01; 1.00; 0.29; 0.38), \\ y_3 &= (0.00; 0.68; 0.17; 0.24) \end{aligned}$$

where y_1—vector estimate of “Energy-saving green software” direction; y_2—vector estimate of “Green software engineering” direction; y_3—vector estimate of “Development of green telecommunications” direction.

As a result, we obtain the following set of Pareto-optimal vectors P(Y) = $\{y_1, y_2\}$.

Based on the received result it is possible to draw a conclusion that priority directions of development Green IT Engineering, from selected by us, are—energy-saving green software and green software engineering.

5 Conclusion

According to the analysis of domestic and foreign publications on results of the carried-out foresight-research in most countries of the world (Japan, USA, France, Germany, etc.)—one of the development priority directions of the country is the sustainable growth through green innovation, which treats also Green IT Engineering, which has a set of the main purpose of reducing the negative impact on the environment through science and technology.

In order to harmonize economic, social and environmental factors affecting the sustainable current state of society and the satisfaction of its potential needs it is appropriate to carrying out the foresight-research on selection the priorities of the STD Green IT Engineering in Ukraine.

The problem description and formal statement of synthesis model problem of foresight-technology computer implementing are produced in the paper. The foresight-research implementation system model based on event-driven approach to decision-making, which, unlike the existing ones, allows to display the complex nature of foresight and provides a transition from heuristic procedures for their formal description for further information of foresight technology is created.

The complex automation approach of foresight-projects in Ukraine is expounded. Within the solution of the problem of the “Formation of Expert Panels” stage automation, it is proposed to perform selection of experts involved in the assessment on the basis of evaluation of their competence level by means of the generalized indicator of the competence level of each expert. For the conducted research on determination of the priority directions of Green IT Engineering development 7 experts are selected that enters the calculated allowable range.

In order to form the initial list of Green IT Engineering areas, technologies of bibliometrics, scientometrics and patent analysis are proposed to be used, which ensures objectivity of initial data of the foresight-research. In the current foresight-research the formation of the initial list of SD produced by using bibliometric analysis for the 5-year time period (2010–2015 years). As a result of this analysis, the initial list, for the subsequent determination of priority directions of Green IT Engineering development, includes SD's with the highest number of scientific publications in the analyzed period—namely, energy-saving green software (87 pcs.), green software engineering (112 pcs.) and development of green telecommunications (64 pcs.).

At the stage of priority directions selection for Green IT Engineering development it is offered to use the t-ordering method and the Pareto-optimality principle that give the chance to increase adequacy of an expert assessment in directions selecting of Green IT Engineering development. According to results of the foresight-research priority directions of Green IT Engineering development in Ukraine are the energy-saving green software and the green software engineering since their vector estimates make Pareto set and have values $y_1 = (0.01; 0.86; 0.33; 0.38)$ and $y_2 = (0.01; 1.00; 0.29; 0.38)$ respectively.

Proceeding from the received results of the foresight-research which is carried out by authors it is possible to claim that the specified sector in Ukraine is growing rapidly, however has no sufficient state support. In this connection, it is appropriate to carry out a national Green Foresight-Research which will allow to reveal the leading Green IT Engineering directions for further public financing of scientific researches in the selected priority directions, thereby having increased competitiveness of the state.

References

1. Technology Foresight Manual: United Nations Industrial Development Organization, Vienna (2005)
2. Kuwahara, T.: Technology foresight in Japan—the potential and implications of DELPHI approach. In: The Approach to and the Potential for New Technology Foresight. International Conference on Technology Foresight, Tokyo, March 2001. http://www.nistep.go.jp/achiev/ftx/eng/mat077e/html/mat077ee
3. Keenan, M., Popper, R.: Comparing foresight 'style' in six world regions. J. Foresight **10**(6), 16–38 (2008)
4. Loveridge, D., Georghiou, L., Nedeva, M.: United Kingdom Foresight Programme. PREST, Manchester (1995)
5. Cuhls, K.: Foresight in Germany. In: Georghiou, L., Cassingena Harper, J., Keenan, M., Miles, I., Popper, R. (eds.) The Handbook of Technology Foresight: Concepts and Practice, pp. 131–153. Edward Elgar, Cheltenham (2008)
6. Shelyubskaya, N.V.: Foresight—a mechanism for determining priorities for the formation of a knowledge society in Western Europe. Phoenix, Kiev, Ukraine (2007)
7. On approval of the action plan to implement the policy framework of the Russian Federation in the field of science and technology for the period up to 2010 and further prospect. Act 2002 (Pr-576). http://www.scrf.gov.ru/documents/22.html

8. Kyzym, M.O., Matyushenko, I.Y., Shostak, I.V., Danova, M.O.: Prospects for foresight forecasting of priority directions of development of nanotechnology and nanomaterials in countries of the world and Ukraine: monograph. INZHEK, Kharkiv, Ukraine (2015) (17)
9. Sokolov, A.: Long-term forecasting trends of development the education by methods of foresight. J. Educ. **3**, 66–76 (2004)
10. International Practice in Technology Foresight. UNIDO, Vienna (2002)
11. Shostak, I.V., Danova, M.A.: The approach to automating the process of forecasting scientific and technological development of the region based on technology foresight. Collection scientific papers military institute KNU named after T.G. Shevchenko, vol. 38, pp. 151–154 (2012)
12. Saritas, O.: A systems analysis of British, Irish and Turkish foresight programmes. In: Seminar: New Technology Foresight, Forecasting & Assessment Methods. Seville, pp. 44–64 (2004)
13. What is Foresight? International Research and Educational Foresight Centre. https://foresight.hse.ru/whatforesight
14. Kuznetsova, Y.: The evaluation model of software usability. In: Information Technologies and Systems. International scientific-practical conference, Minsk, Byelorussia, pp. 62–63 (2015)
15. Danova, M.A., Shostak, I.V.: Information technology support foresight projects in Ukraine. Collection scientific papers military institute KNU named after T.G. Shevchenko, vol. 43, pp. 211–217 (2013)
16. Volkov, V.I., Denisov, A.A.: Fundamentals of the theory of systems and system analysis. SPbGTU, St. Petersburg (1997)
17. Anastas, P.T., Heine, L., Williamson, T.C.: Green Engineering. American Chemical Society, Washington (2000)
18. Ehrenfeld, J.R.: Industrial ecology: a framework for product and process design. J. Clean. Prod. **5**(1–2), 87–95 (1997)
19. Fiksel, J.: Design for Environment: Creating Eco-efficient Products and Processes. McGraw-Hill, New York (1998)
20. Kuznetsova, Y.: Green software engineering for visualization of controlling algorithms. In: Monograph. National Aerospace University named after N. Ye. Zhukovskiy “KhAI”, Kharkiv, Ukraine (2016)
21. Anastas, P.T., Zimmerman, J.B.: Design through the 12 principles of green engineering. J. Environ. Sci. Technol. 95–101 (2003)
22. Sidorov, N.A.: Software ecology. J. Softw. Eng. **1**, 53–61 (2010)
23. Handbook of Energy-Aware and Green Computing. (2012). http://www.crcpress.com/product/isbn/9781466501164
24. Murugesan, S.: Harnessing green IT: principles and practices. IEEE IT Prof. **10**(1), 24–33 (2008)
25. The Green Computing Book: Tackling Energy Efficiency at Large Scale. (2014). http://www.crcpress.com/product/isbn/9781439819876?source=crcpress.com&utm_source=productpage&utm_medium=website&utm_campaign=RelatedTitles. Accessed 25 Sept 2015
26. Postnikov, V.N.: Analysis of approaches to formation of an expert group focused on preparing and making decisions. J. Sci. Educ. **5**, 333–346 (2012)

Green IT Engineering in the View of Resource-Based Approach

Julia Drozd, Alex Drozd and Svetlana Antoshchuk

Abstract State-of-Art and development of Green IT engineering in the view of resource-based approach which analyzes integration of the artificial world created by human into the natural world (NW) are considered. The challenges solved in integration process include problems of the green technologies. The solution is based on achievement of a certain throughput, trustworthiness and investment of resources: models, methods and means. Throughput is aimed at receiving the maximum return from resources, and trustworthiness—at harmonization with the NW. Resources structured under parallelism and fuzziness of the NW show in development three levels: replication, diversification and autonomy. They are serviced by methods improving throughput, trustworthiness and access to resources, respectively. Green technologies as model occupy autonomy level. The used models and methods belong to the level of diversification. Means are built with replication of operational elements. The methods of multiply effect and its focusing on parameters important for green technology are offered. These methods are based on increasing a level of resource development.

Keywords Green IT engineering · Resource-based approach · Parallelism and fuzziness · Levels of resource development · Methods of multiply effect and its focusing

J. Drozd · A. Drozd (✉) · S. Antoshchuk
Odessa National Polytechnic University, 1, Shevchenko Avenue, Odessa, Ukraine
e-mail: drozd@ukr.net

J. Drozd
e-mail: dea_lucis@ukr.net

S. Antoshchuk
e-mail: svetlana_onpu@mail.ru

V. Kharchenko et al. (eds.), *Green IT Engineering: Concepts, Models, Complex Systems Architectures*, Studies in Systems, Decision and Control 74,
DOI 10.1007/978-3-319-44162-7_3

1 Introduction

1.1 Motivation

The Computer World (CW) for a short time of the existence became one of the most powerful factors defining development of human society. It explains why the modern level of CW development and its prospects are an important object of research. The fact of such researches testifies that the human creates the CW unconsciously, without controlling this process. As a rule, the human researches the CW development history which already came true in his execution. Many observations and generalizations concerning development of IT-technologies are known: Moore's Law (the number of transistors in a dense integrated circuit doubles approximately every two years), Rock's law or Moore's second law, which says that the cost of a semiconductor chip fabrication plant doubles every four years, Machrone's Law (The PC you want always costs $5000), Metcalfe's Law (Communications networks increase in value as they add members), Wirth's Law (Software is slowing faster than hardware is accelerating) [1]. They allow predicting the next achievements based on data accumulated on prior stages of development. The following stages adjust the quantitative estimates of predictions, but save the general regularities, proving their existence.

We offer the concept of the resource-based approach allowing to coordinate many observations on a unified basis by researching process of integration of the artificial world created by the human into the natural world (NW) [2]. To be integrated into the NW, it is necessary to understand and accept rules of this process and also to learn to play by these rules. On this way we see two main hindrances: the NW is too various, and human life is short. However, the CW repeats development of the NW, but in a short space of time which allows the human to understand better these rules and to be trained in them, comparing these patterns. The challenges solved in integration process include problems of the CW and the green technologies developed in it [3].

As a first approximation, we can consider that a problem will be solved in case of execution of three conditions: achievement of a certain throughput for execution of all amount of works for limited time; receiving authentic results; investment of resources, which provide the above throughput and trustworthiness.

Resources comprise all the necessary for problem solving: models, methods and means. The models are human understandings of the NW and its components. Methods serve for conversion and an assessment of resources. Models and methods form an informational part of the resources but means belong to the technological one.

Means are the material carriers of informational resources. Models and methods are recorded and kept in structure and described by the process of functioning the means. For example, the model of the coming winter is coded in onion structure, and the law of universal gravitation (i.e. the method of interaction of resources) is described in the process of falling an apple from the tree. In order to read these records, we are to be prepared i.e. to get access to the resources of certain levels of development.

1.2 Related Works

The known works related to the development of resources in Green IT engineering can be partitioned into two groups which show:

- features of the modern green hardware;
- models, methods and means in development.

The analysis of the green hardware decisions used in the organization and functioning of controllers, processors and FPGA projects shows a number of features.

1. As a rule, these decisions exploit only one idea of undervoltage of a supply up to its switching off for temporarily inactive elements of structure [4].
2. This idea is often presented in the distorted look as lowering of clock frequency based on a formula of a dynamic component of power consuming: $I_{DYN} = CV^2 f$, where C is loading capacitance; V—supply voltage; f—clock frequency [4].

However frequency only specifies the instantaneous character of this assessment brought to a unit of time. Lowering of frequency not only reduces amount of the consumed energy but also in so many time reduces amount of the executed clock cycles. Thus, the solution of the computing task for a certain amount of clock cycles on the lowered clock frequency doesn't change amount of the consumed energy but only decelerates receiving of results. Power consumption will be decreased in case of reduction of supply voltage (in square dependence). Lowering of clock frequency is a compelled action caused by supply undervoltage. Really, reduction of supply voltage decelerates execution of computation in clock cycle. It requires increase in its duration, i.e. lowering of frequency.

3. Different types of a diversity of both structures and operation modes integrated into green technologies are the cornerstone of energy saving decisions.

These features can be shown on the example of STM8L microcontrollers [5].

Microcontrollers of this family use two internal voltage controllers of a core supply: MVR and LPVR: at the level of 1.8 V (the main mode, current to 25 mA) and 1.55 B (for the modes of the under energy consumption with current to 0.2 mA). Some types of clock pulse sources are used for clocking of the microcontroller and its periphery: external high speed oscillator 1 … 16 MHz and the generator to 16 MHz, internal high speed oscillator of HSI of 16 MHz, external and internal low-speed oscillators of 32.768 and 38 kHz. In the same family of microcontrollers, two types of memory are used: 8-bit RAM and 32-bit Flash-memory which supports the IDDQ mode of low energy consumption (in an inactive state). In this case the program is read from 8-bit RAM of memory with decline in the throughput.

A line of STM8L15x/162 microcontrollers provides 5 modes of energy saving:

- Wait—CPU is stopped, peripheral devices work;
- Low power run—the program is executed from RAM with clocking by slow oscillators (LSI, LSE), the selected peripheral devices are active;
- Low power wait—CPU is stopped, the selected peripheral devices are active;
- Active-halt—CPU is stopped, AWU, WDT and RTC continue to work if are switched on;
- Halt—clocking of CPU and peripheral devices is stopped.

The energy saving technologies: Cool‘n’Quiet AMD and Enhanced SpeedStep Intel (checked on mobile processor systems) offer systematic lowering of frequency and supply voltage in process of falling of computing loading. It is shown in a series of Athlon 64, Opteron, Phenom, Phenom II and Pentium 4-M, Pentium M processors. Each model of the Athlon 64 processor contains the table of the possible states characterized by a combination of frequency and supply voltage. Phenom and Phenom II processors can change the frequency of each core irrespective of others [6, 7]. The 8-core Exynos S Octa processor is completed on the big.LITTLE ARM technology with two types of cores: 4 cores of Cortex-A7 with low energy consumption and 4 cores of Cortex-A15 with high performance. This provides energy saving at the level of 70 % [8].

The carried-out analysis allows to conclude the following.

- Green technologies show aspiration to lowering of dependence of hardware in its functionality from energy consumption in the direction of raising mobility of computer systems.
- The main feature of green hardware is variety of units and modes of their functioning at the system level.

The works related to the development of models, methods and means are considered in the course of resource-based approach presentation.

1.3 Goal and Structure

The purpose of this research consists in estimating a current state of green IT engineering and in showing the prospects of its development. The technologies created by the human permanently take exam concerning their compliance to the NW. Green technologies can be considered as one of approaches to integration of the artificial world into the NW. Therefore the assessment of a current state of green IT engineering and prospects of its development is executed from a line item of the integration process analyzed with use of resource-based approach.

The main concepts of resource-based approach and development of resources are considered in Sect. 2. Conceptual framework includes levels and multistage model of resource development shown for models, methods and means of the CW. We consider models of numerical data, of a result of computation and an arithmetical operation. Methods are analyzed for on-line testing area. Development of means is tracked on the example of personal computers (PC). Section 3 considers prospects of green technologies development with the accounting of features of resource integration into the NW. Methods of the multiple effect and its focusing on parameters most important for green technologies are offered and evaluated.

2 Development of Resources

2.1 The Main Concepts of Resource-Based Approach

The solution of a problem is characterized by two components: process and result. As a first approximation, throughput and investment of resources were considered in relation to process, and trustworthiness—to result of the solution. However, throughput and trustworthiness show more general character in a solution. Throughput is the efficiency in use of resources in the course of a solution and the efficiency of resources in the received result. Trustworthiness is a level of adequacy of process or the result in relation to the NW. Naturalness of process and result determines their applicability in practice. Thus, throughput is aimed at receiving the maximum return from resources, and trustworthiness—on harmonization with the NW. We know many cases when high throughput did the considerable harm to mankind and the NW, for example, during military operations or technogenic accidents.

Conditions of a solution determine its key parameters in process and result: throughput, trustworthiness and turnover of resources: consuming and receiving of resources.

Resources are represented organizationally and functionally as system of elements and as an element of system, respectively. All resources are elements of the NW. Therefore they and their elements develop from simple to real in the path of structuring under features of the NW. The most evident features in the CW are parallelism and fuzziness. Simple resources are exact and sequential. They appear in accordance with the initial understandings and opportunities of a human. Real forms are parallel and approximate according to features of the NW.

In the resource development occurring under the influence of parallelism and fuzziness, the resource-based approach identifies a number of levels: replication (stamping of clones, identical operational elements), diversification (transformation of clones and identical elements in an individual and the version) and autonomy. In the hierarchy, they occupy the low, L, the middle, M and the high, H levels respectively. Each preceding level serves for the next one. In all the levels, from

L to *H*, the goal is surviving (integration into the NW) which is provided by different methods: increasing throughput, trustworthiness and access to resources (self-sufficiency), respectively.

Levels of resource development are shown in Fig. 1.

Replication is the simplest method of parallelism distribution and the most resource-demanding. In the wildlife the replication exists under the slogan: “It is more to give birth, than will die” and is an extensive method of resource capture at a stage of filling the ecological, technological, market and other resource niches where the throughput becomes a decisive parameter. This way is necessary but limited. It leads to closing of a resource niche. The cholera bacterium replication leads to dehydration of organism which will pass away together with the cholera bacterium. Replication of herbivorous animals and predators results in pauperization of their nutrient medium. They only survive under the conditions of the resource balance, i.e. the restriction of replication. In case of the unfilled resource niche the best conditions for a surviving are created by replication.

The stimulus for diversification of resources (known as transition of quantity to quality) is filled resource niche, when the further escalating of throughput becomes ineffective and leads (for example in economy) to the crises of overproduction. Under these conditions the survival demands the development of individuality in clones. The operational elements should be transformed into versions. Such transformation is directed to increase of trustworthiness in safety: during epidemics clones are doomed to extinction, but the versions have chances to survive due to the particularities. The majority system choosing the result by voting from the outputs of its several replicated channels is as fault tolerant and the results are authentic as these channels are different.

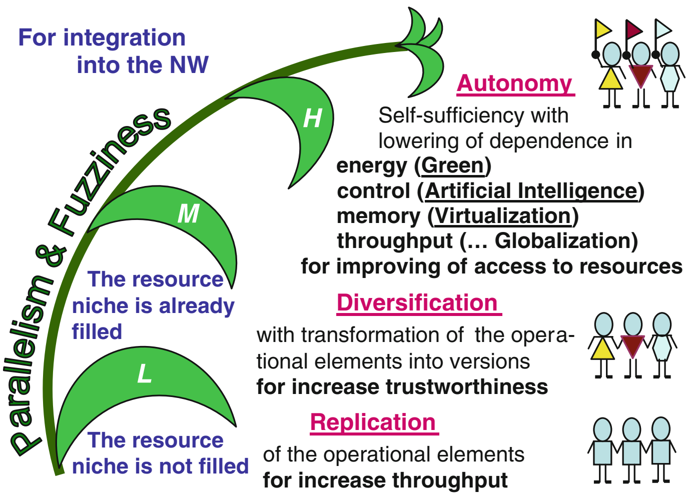


Fig. 1 Levels of resource development

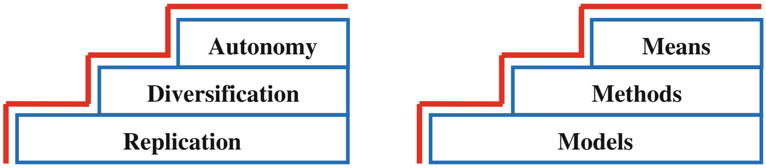


Fig. 2 Multistage models

Self-sufficiency can be reached by enhancement of access to resources, including enrichment by resources of an inner pattern. In the NW the aspiration to autonomy is watched in creation of the geological, ecological, biological and other isolated systems showing self-sufficiency in many respects.

In the CW the aspiration to autonomy is shown in the following developments: the object-oriented approaches selecting objects and allocating with their certain independence; the virtual computing resources for lowering of dependence on throughput and memory size; artificial intelligence for increase of system independence level in control. Development of the green technologies aimed at reduction of energy dependence is also an example of aspiration to autonomy.

Multi-level development of resources reflects model of multistage development of the NW when the sequence of stages is created, and everyone previous is the base for all following. The multistage model describes also the relations operating between types of resources. The first stage is made by models. The methods forming the second stage of resources development are built on base of models. The means will be created at the third stage in case of the presence of the first two stages. Existence, development and knowledge also form the multistage model.

Examples of multistage models are shown in Fig. 2.

2.2 *Development of Models*

We can consider development of models on the examples of numerical data, result of computation and arithmetical operation.

In the NW everything exists and is evaluated in tolerances. For example, the linear sizes of the desktop have no exact values. Really, its boundary points at the level of elementary particles show the wave nature, i.e. don't possess spatio-temporal coordinates. Besides, any estimates are restricted in accuracy to an instrumental error. Therefore the human obtains approximate numeric data which describe results of measurements in the NW. However, the human thinks by using the exact data, i.e. integers by the nature which include the numbers of set items. For this reason, the human often appears under the influence of a model of exact data when all numbers irrespective of their real nature are considered as exact data. Coding of numeric data is their numbering. That is why only exact data can be written in the computer. It increases influence of the exact data model on the researches conducted in the CW. Therefore, the human, adapting for features of the NW, has no other alternative except enhancement of numeric data model by developing it in the direction of increase in level of parallelism and fuzziness. For exact numbers, their size of n uniquely determines the precision of n and volume of the range 2^n in data representation, rigidly connecting these characteristics among themselves.

Different requirements imposed by the human to the range and the precision of numbers are caused by features of the NW. They are shown in aspiration to solve problems in case of short precision and the increased range, for example, computation of the sum of numbers 800 and 100 which are in the range 0 … 1000, using a 8-b codeword with the range 0 … 255. The ratio of volumes of these ranges determines the scaling factor $k = 4$. The initial numbers are normalized by division into k: 800/4 = 200 and 100/4 = 25. Addition of the normalized numbers 200 and 25 in a 8-b codeword determines result 225 which can be transformed by multiplication by k to the problem range: $225 \cdot 4 = 900$, where 225 is mantissa m in a normal form of number representation: $m \cdot k$.

The described example shows development of the CW which led to representation of number in a floating-point format in the form of the product $m \cdot B^E$ containing two independent (i.e. existing in parallel) components: a mantissa m and the scaling factor $k = B^E$ with use of an exponent E and numerical notation base B. Sizes of a mantissa m and an exponent E independent from each other determine respectively a precision and volume of the range in representation of data. Thus, the model of number passed a path of development from the single-component exact form to the approximate diverse representation parallelized to two components: mantissa and exponent. The floating-point formats supporting such approximate parallel representation of data gained the greatest distribution in the CW. Now they are extended for use in decimal arithmetic [9].

The result of computing operation also gained development from correct exact number to the authentic approximate one with parallelizing of its structure on two diverse types of bits: most significant bits (MSB) and least significant bits (LSB) [10]. Faults of digital circuits cause in these bits also two diverse types of errors which are essential and inessential for trustworthiness of result, respectively. Therefore in case of an inessential error, the erroneous result is authentic.

In development of arithmetical operation model it is necessary to mark a key role of multiplication which is present at the record of approximate number in the floating-point formats. That's why multiplication in this or that look is present at all operations with mantissas, and their results will inherit properties of product. In particular, the result of two-operand operations as product doubles size in comparison with size of an operand.

Operations with single precision (in relation to the operand size) belong to homogeneous in which the floating-point operands and floating-point result are all of the same format [9]. In the conditions of single precision the complete arithmetical operation becomes redundant and gains development in the form of the truncated one saving single precision at the smaller volume of computation [11, 12].

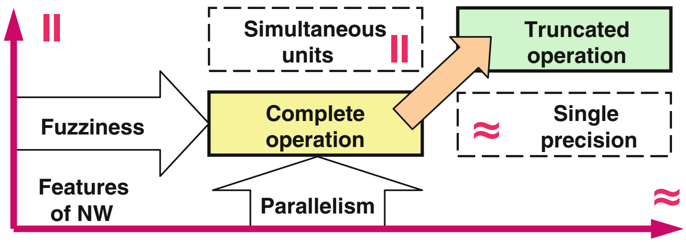


Fig. 3 The natural path of the arithmetic operation development

Paralleling of circuits to the level of simultaneous units (the parallel adders and shifters, iterative array multipliers (IAM) and dividers) executing data processing in parallel codes brings the volume of computation and hardware of its implementation into accord. As a result, the truncated operation implemented with smaller expenses becomes the main method for processing of mantissas.

The natural path of the arithmetical operation development from the completed form to truncated one is shown in Fig. 3.

A model of arithmetic operation increased the level of both parallelism in array circuits and fuzziness in single precision.

2.3 Development of Methods

We consider methods on the example of on-line testing which is aimed at increase of trustworthiness, playing a key role in safety of systems in area of critical application. Unfortunately only negative events testify to an energy saving directivity of on-line testing. The Chernobyl catastrophe, accident on Sayano-Shushenskaya hydroelectric power station led to a real loss of energy as a result of a stop of its generation.

Development of on-line testing can be divided into three main stages: the initial stage of the borrowing of methods and means from the theory and practice of error-correcting data transmission on the distance; formation stage characterized by the distributing of on-line testing on own means within the exact data processing; modern stage distributing on-line testing for approximate calculations.

The formation stage is the longest. It passed in development of the theory and practice of designing the self-checking digital circuits. If the circuit is both fault-secure and self-testing it is said to be totally self-checking for a given set of faults. Fault-secure circuit defines information redundancy which is necessary for detection of single faults from a given set with the first error. Self-testing of circuit defines the condition of detection of the first fault before the occurrence of the second one. It is based on the existence of at least one input word, where a fault is detected, and the high probability of passing all input words in time interval between neighboring faults. Self-checking circuits have been developed for almost all types of digital units: combinational circuits and machine with memory, control and synchronization circuits, as well as arithmetic units that have made a significant contribution to the development of on-line testing with real practical application [13].

The reasons of the on-line testing transition to the modern stage of development and also features of this stage can be revealed with the help of the resource-based approach. Development of on-line testing is based on improvement of models for numerical data, an arithmetical operation and their respective objects and online testing purpose.

According to the definitions of totally self-checking circuits, on-line testing and testing are aimed at assessing means of solving the problem by detection of faults in a digital circuit. In conditions of approximate calculations, on-line testing demonstrates its own purpose: to evaluate the trustworthiness of the calculated results. In the case of exact data processing, these purposes are the same, because detected error indicates at once a fault in the circuit and non-trustworthiness of erroneous result. For approximate calculations the purposes are diversified, because the erroneous result is authentic in case of inessential error.

The on-line testing method shows trustworthiness in checking the result detecting essential errors and skipping inessential ones by the following formula [14]:

$$D = P_E P_D + (1 - P_E)(1 - P_D), \tag{1}$$

where

P_E is a probability of an essential error;

P_D is a probability of error detection.

The probabilities P_E and P_D characterize the object and method of on-line testing, respectively. In the case of exact data all errors are essential, i.e. $P_E = 1$, and trustworthiness $D = P_D$ does not depend on the object of on-line testing. In conditions of approximate calculations, the trustworthiness D considers two parameters: P_E and P_D. This makes the assessment of the trustworthiness D more adequate to the fuzzy NW.

The features of approximate calculations significantly reduce the probability P_E. A low half of the result calculated in homogeneous operations will be discarded. This twice reduces P_E. Only the truncated operations reduce the probability P_E to a lesser degree. However denormalisation and normalization of mantissas additionally reduce the probability P_E in results of all previous and following operations, accordingly [14].

On-line testing methods developed in the framework of the theory of the totally self-checking circuits have become traditional. They are based on model of the exact data, the adequacy of which can be shown with the use of a square with unit size [14]. The horizontal side of the square contains a sum of probabilities P_E and $P_N = 1 - P_E$. The vertical side shows a sum of error detection probability P_D and probability $P_S = 1 - P_D$ of error skipping. The case typed for traditional methods with the high probability P_D and a low probability P_E following from execution of a complete arithmetical operation is shown in Fig. 4.

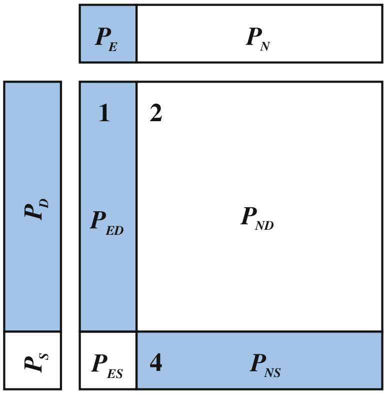


Fig. 4 Trustworthiness of the traditional on-line testing methods

Areas $P_{ED} = P_E\ P_D$ and $P_{NS} = (1 - P_E)\ (1 - P_D)$ of the square parts 1 and 4 which are the components of trustworthiness D point to its low level. Thus, the trustworthiness of the on-line testing methods can't be correctly estimated within model of exact data.

The fuzziness of on-line testing methods is determined by the presence of an area of inessential errors, which occupies the largest part 2 of the square and shows the new property of the methods to make a mistake, rejecting erroneous, but authentic results. In addition, approximate character of on-line testing methods follows from the fuzzy threshold between essential and inessential errors. The methods are fuzzy also in the substitution of concepts referring the result to non-authentic or authentic on the basis of error detection or skipping, but the actual symptom is that the error is essential or inessential. The on-line testing methods always have been approximate for the approximate data processing. However, our understanding of these methods as approximate opens up new opportunities for their development in path of improving the adequacy to the NW.

The trustworthiness D of the methods can be increased by three ways: an increase of the probability P_E; a decrease of the probability P_D; the detection of the essential and inessential errors with different probabilities P_{D-E} and P_{D-N}, where $P_{D-E} > P_{D-N}$. Analysis of formula (1) shows that high trustworthiness $D > 0.5$ can be achieved, if both parameters are on one side of 0.5. This defines two ways of increasing the trustworthiness of the methods developed for both cases: $P_E > 0.5$ and $P_E < 0.5$. These ways are shown in Fig. 5.

The first way allows to develop traditional on-line testing methods with high detection probability $P_D > 0.5$, including, first of all, wide spread residue checking (Fig. 5a). The trustworthiness D is improved by the increasing the part 1 of the square (the probability P_{ED}). Implementation of this way is based on use of truncated arithmetical operations as only they allow to provide probability $P_E > 0.5$. The on-line testing methods are developed for all truncated arithmetic operations [15, 16]. The second way improves trustworthiness D by reducing of probability P_D for increase of the part 4 in probability P_{NS} (Fig. 5b). The main requirement to

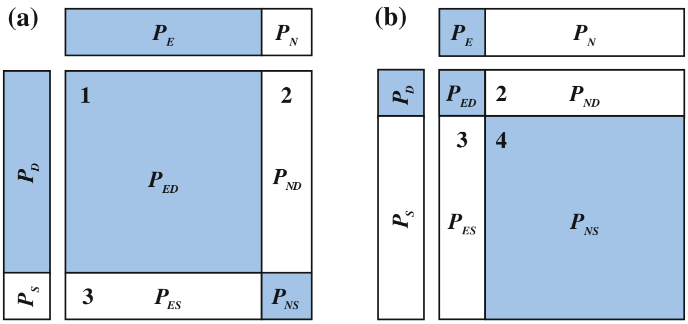


Fig. 5 Two ways to increase trustworthiness: **a** $P_E > 0.5$, **b** $P_E < 0.5$

implementation of the second way consists in saving a set of faults which will be detected when lowering probability P_E. It can be executed using natural information redundancy of results in arithmetic operations. The on-line testing methods using a part of this redundancy, which provides the simple error detection and procedure of its design are offered for IAM and squarer. The on-line testing methods checking the approximated result by simplified operation create information redundancy by limiting a set of considered input words of the circuit [16].

In case of third way, the formula (1) assumes the following form:

$$D = P_E P_{D-E} + (1 - P_E)(1 - P_{D-N}),$$

where P_{D-E} and P_{D-N} are the detection probabilities of essential and inessential errors, respectively; $P_{D-E} > P_{D-N}$.

The third way is shown in Fig. 6, where $P_{S-E} = 1 - P_{D-E}$ and $P_{S-N} = 1 - P_{D-N}$.

This way is executed in on-line testing methods, which estimate size of result and its error, as the following: logarithm checking; checking by inequalities; checking by segments [16].

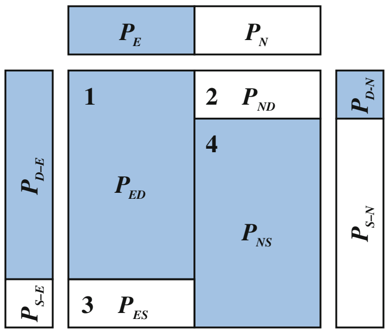


Fig. 6 The third way for increase in trustworthiness D

Table 1 Development of data model

Model of exact data	Model of approximate data
Existing on-line testing is applicable to any type of data	Existing on-line testing is applicable to the exact data only
A purpose of on-line testing is to estimate reliability of computing circuits	A purpose of on-line testing is to estimate trustworthiness of calculated results
All processed numbers are considered as the exact data	Processed numbers are in most cases approximate data
All errors are essential for trustworthiness of calculated results	Basically, the errors are inessential for trustworthiness of calculated results
Traditional on-line testing methods have high trustworthiness: detect almost all errors and faults	Traditional on-line testing methods have low trustworthiness: mainly detect inessential errors
Only one way for increasing the trustworthiness of on-line testing methods: to increase probability of error detection	Three ways to increase the trustworthiness of on-line testing methods by distinguishing essential and inessential errors

The assessment of on-line testing method trustworthiness and ways of its increase allow to estimate development of data model from exact numbers to approximate form, addressing to the Table 1.

2.4 Development of Means

Development of means can be considered on the example of the PC showing successful integration into the NW. External manifestations of successful development are wide demand, mass production and distribution. The internal content of successful integration is increase of level of both parallelism and fuzziness in the hardware support of approximate calculations. The PC have passed a way from co-processors of non-obligatory delivery (Intel 8087/287/387) to several floating point pipelines in the structure of the Pentium CPU and to many thousand floating point pipelines in the graphic processor with its use for execution of parallel calculations according to the technology CUDA [17].

This natural path of development of the PC is shown in Fig. 7.

3 Prospects of Green Technologies Development

3.1 The Place of Green IT Technologies in Integration Process

Resource-based approach shows that green technologies take different positions in hierarchy of levels of resource development:

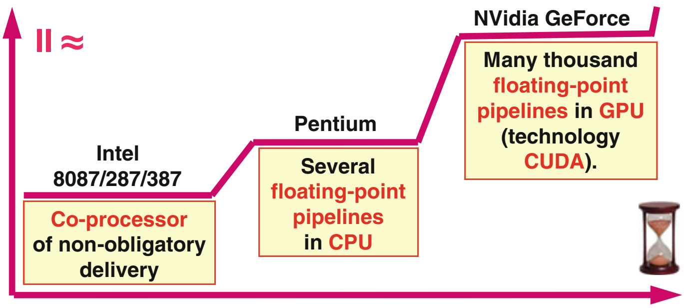


Fig. 7 Development of the PC

- Level *H* of autonomy as the model developing self-sufficiency in energy consumption by an evolutionary method;
- Level *M* of diversification in the used models and methods profiling resources under features of the solved problem;
- Level *L* of replication in design of the means with array circuits of digital units.

The evolutionary method is based on such gradual changes which allow the NW to adapt for them. Therefore this method initially contains vulnerability to which it is subject from the sharp changes caused by a cataclysm. The cataclysm of the Permian Period which occurred 250 million years ago destroyed 90 % of wildlife. Disappearance of dinosaurs and mammoths is also explained with sharp changes in the NW.

Development of the method in the direction of autonomy creates immunity for overcoming of such vulnerability in a type of the resources counteracting it. The multistage model of the NW is built. The first stage is flora which is completely busy with energy storage of the sun in vegetable adipose, proteins and carbohydrates. The second stage is herbivorous animals which are busy by conversion of the resources received from plants to animal adipose, proteins and carbohydrates. The third stage is creophagous which build themselves of ready materials of an animal origin. It gives them a certain energy independence. They for the first time release themselves for games, training and research of different scenarios of the NW development.

The first step to mitigation of sharp changes is their prediction at early stages. However how many attention the mankind pays to prediction? Human life is generally spent for physiological processes, training, a labor duty, a relaxation, the family and public relations. Therefore it is important that the human creates the following step: the artificial world, the CW, human-machine complexes of the continuous action which execute basic functions without distraction on a dream and food. They include the safety-related instrumentation and control systems directly aimed at prediction, anticipation and mitigation of the sharp changes happening owing to accidents, catastrophes and cataclysms.

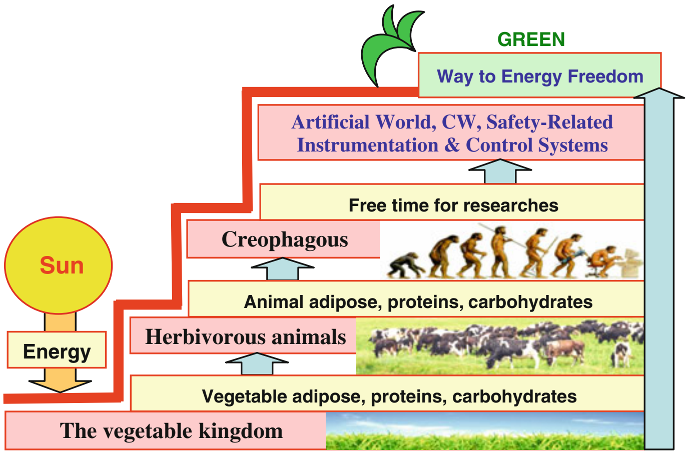


Fig. 8 Stages of evolutionary method development

The stages of evolutionary method development are shown in Fig. 8.

The evolutionary method creates protection against cataclysms on the basis of energy release which at the current stage is represented by development of green technologies.

Models and methods used in green technology are diversified being implemented in structure of hardware and the modes of its functioning.

The diversity implemented in hardware generally is formed at the system level by three version types of elements: the elements which ensure functioning of other elements at several levels of supply voltage and clock frequency corresponding to these levels; the elements, every one of which works at the its voltage level of a supply and the corresponding frequency; the elements which support operation on several voltage levels of a supply and the appropriate clock frequency.

The diversity implemented in operation modes of hardware is shown in setting: a certain configuration of the elements switched on and off; levels of switch-off of elements; voltage levels of a supply and appropriate clock frequency.

Thus, green technologies generally develop in the path of resource profiling under features of the solved task by diversification of both models and methods implemented in hardware and modes of its functioning at the system level. As a rule, process of profiling is effective on the first step. The following steps lead to increase of expenses and lowering of return up to their alignment. Design of means is executed by replication of identical operational elements in array circuits of single-cycle digital units. Features of replication in green hardware are considered further. Replication of means determines the low level L in implementation of models and methods into green technologies.

3.2 Assessment of Methods in Development

For enhancement of a resource basis of green technologies, it is expedient to estimate development of methods in the solution of certain problems. Conditions of a solution determine its key parameters in process and result: throughput, trustworthiness and turn of resources (consuming and receiving of resources).

Development of methods can be analysed with use of relative indexes of the basic parameters of a solution in comparison with the underlying methods possessing the conventional high levels of parameters. The relative index of development of a method in throughput is estimated as $K_{TP} = P/P_B$, where P and P_B—throughput of the offered and underlying method, respectively. As a rule, trustworthiness of result is estimated with use of complementary values. This defines its relative index as $K_{TW} = (1 - W_B)/(1 - W)$, where W_B and W—trustworthiness for the underlying and offered method, respectively. The relative index of resource consumption is determined by the formula: $K_{RC} = C_B/C$, where C_B and C—the cost of the invested resources for the underlying and offered method, respectively. Development of methods can be estimated with use of the integrated assessment considering all basic parameters of a solution: $K_{IA} = K_{TP} \cdot K_{TW} \cdot K_{RC}$.

3.3 Traditional Methods and Methods of the New Nature

Development of methods is aimed at improving of solution parameters. The most simply separate parameters can be improved at the level of replication. Stamping additional elements, we can increase throughput, reduce power consumption on the reduced values of supply voltage and clock rate or increase trustworthiness by duplicating of computation. The part of elements can be reoriented for increase of trustworthiness instead of throughput or on the contrary. Such practice led to rooting in human consciousness of a method "Robbing Peter to pay Paul" rearranging a patch from less significant hole on more significant one. Thereby, improving of parameters traditionally associates with need of deterioration of others. The traditional question asked at project presentations is: "What has been lost in exchange for improvements?".

Replication, the unfilled resource niche, increase of throughput for integration into the NW, the "Robbing Peter to pay Paul" method—all this is manifestations of the same phenomenon when still there is no conflict to the NW, i.e. boundaries, constraint-driven of the NW, aren't reached yet. Therefore each manifestation testifies to existence of all other manifestations.

The first response to completion of filling of a niche is attempts of its extension. Example of successful extension of a resource niche is long-term increase of throughput and memory size of computer systems. It saves development of the software at the level of replication by compilation of the replicated excess modules [18].

Level M of diversification is also characterized by a number of manifestations, including the filled resource niche, increase of trustworthiness for integration into the NW and methods of the new nature. They show the multiple effect of simultaneous improving of several parameters without loss in others. Improving is reached thanks to development of resources. For example, development of the PC which succeeded in increase of parallelism level in approximate calculations represented in Fig. 7 shows the multiple effect of the considerable improving of key parameters. In the last 20 years, the PC have increased throughput by raising clock rate from KHz up to GHz. Concurrently they increased the memory size from Mb up to Tb, i.e. key parameters at once increased a million times. Besides, a number of other factors including the price are improved without any loss. What nature of such considerable improving is?

We shall distinguish the price of the invested resources and value of the received resources. Price of the resource is made up of its prime cost and profit, the size of which is determined by the difference between the value and the prime cost at the time of pricing. The price remains fixed for the majority of the paid resources. The value is formed by features of the resource which are shown in its internal organization and functioning and are evaluated in relation to features of other resources. At the time of pricing, the features for which the resource is created are generally considered. Thus a set of all features of the resource is not less infinitely, than a lot of other resources surrounding it. The value of the resource is increasing with new manifestations of features with respect to the appearing resources. This explains the ever-growing value of art works and other rarities, as well as models and methods in relation to the extending set of other resources. Transition to the level of diversification stimulates development of features of resources in the direction of integration into the NW and creates conditions for development of the multiple effect methods. These methods characterized by an integral index $K_{IA} > 1$ reduce the mutual contradictions (typical for key parameters in traditional methods) according to growth of K_{IA}.

3.4 Development of Methods of the Multiple Effect and its Focusing

Basis for development of the multiple effect methods is the natural path of increase in level of both parallelism and fuzziness from replication of resources to their diversification in setting and the solution of problems that is objective process. How methods of the multiple effect can be developed? As a rule, the multiple effect methods are developed blindly and unconsciously but they can be generated according to the plan.

The case of development as "blindly" can be considered on the example of arithmetical operation improvement shown in Fig. 3.

For n-bit mantissas of operands, truncated multiplication is executed by an exception from processing of $k = n - \log_2 n$ low columns in a matrix of product conjunctions. It reduces quantity of operational elements (full adders) of the IAM from $C_B = n^2 - 2n$ to $C = n^2 - 2n - k\,(k - 1)/2$. Operation time can be estimated as proportional to size of the calculated result which decreases in case of truncated product from $T_B = 2n$ to $T = 2n - k$ in comparison with complete product. It determines the relative indexes of resource investment at a design stage and throughput as $K_{RC} = C_B/C$ and $K_{TP} = T_B/T$, respectively. Growth of trustworthiness can be estimated by increase of reliability in the truncated circuit at a design stage from $W_B = p^{C_B}$ to $W = p^{\ C}$ where p is reliability of an operational element. Trustworthiness can be presented in the form $W_B = 1 - C_B\ (1 - p)$ and $W = 1 - C\,(1 - p)$ taking into account $p \rightarrow 1$. It determines the relative index of trustworthiness as $K_{TW} = C_B/C$ or $K_{TW} = K_{RC}$. Development of arithmetical operation shows the multiple effect evaluated integrally as $K_{IA} = 2n\,(C_B/C)^2/(2n - k)$. For $n = 16$ and $n = 32$, parameters will be improved by $K_{IA} = 3.2$ and $K_{IA} = 4.3$ times, respectively. Improving of parameters was for the first time marked for each of them separately, without generalization in the multiple effect [19, 20]. The experimental confirmation of the multiple effect in a significant area, power and delay improvement is described in [12].

The multiple effect can be focused on improving of the most important parameters at the expense of others by method “Robbing Peter to pay Paul”. It creates conditions for overcoming of earlier reached boundaries of possible decisions and essential extension of a set of the solved problems.

Focusing of the multiple effect in the truncated operation allows to use simplification of the circuit and reduction of operation time for increase of throughput or lowering of power consumption. Reduction of circuit complexity allows processing input data in parallel on two IAM with truncated calculations instead of one executing full operation. Lowering of operation time allows increasing clock frequency and respectively throughput of circuit. Duplicating of the IAM increases throughput specified to complexity of the initial circuit by K_{RC} times, and reduction of operation time—by K_{TP} times that as a result gives a throughput benefit $K_{RP} = K_{RC}\ K_{TP}$ times. For example, for $n = 16$ and $n = 32$ throughput increases respectively over 2.3 and 2.7 times. Power consumption is in square dependence on supply voltage, reduction of which increases operation time and require the appropriate lowering of clock frequency. Duplicating of the circuit allows increasing operation time by K_{RP} times in case of saving the initial throughput.

Undervoltage increases operation time at first linearly, and then with square dependence. Let the operation time increases by $H < K_{RP}$ times linearly, and then in square. Then power consumption can be reduced by H^2 times on the linear section and K_{RP}/H times on square one that as a result gives a scoring in power consumption $K_{RP}\ H$ times. For example, in case of $H = 2$ and $H = K_{RP}$ (for $n = 32$), we can expect additional lowering of power consumption over 5.4 and 7.3 times, respectively.

The multiple effect methods developed unconsciously can be shown on the example of multiplication in the context of its key role in processing of the approximate data peculiar to the NW. Features of the multiplication which is present at the record of approximate data in floating-point formats will be transferred to other operations with mantissas, These operations (as well as multiplication) are executed in digital circuits with dominance of the “0” values of signals. The domination of zero increases with improving level of the circuit parallelism and consequently throughput. It leads to the masking of stuck-at “0” faults and shorts, and thus naturally improves the trustworthiness of results. At the same time, it reduces the amount of switching events (transition), which is the maximum for the same number of the “0” and the “1”. It leads to lower power consumption in the dynamic component, i.e. it increases the rate of resource saving.

We researched domination of zero on the example of a single-cycle arithmetical shifter of mantissas (AS) which is widely applied to a denormalization of operands in floating-point adder. The AS executes right shift with the normalized N-bit mantissa A and calculates N-bit mantissa of result with single precision. Shift is performed under control of R-bit code of shift size $B = 0, \ldots, 2^R - 1$. This operation is a special case of multiplication by 2^{-B}. The program model creates R versions 1, 2, …, R of the AS circuit with different level of the circuit parallelism, using 2, 4, …, $2R$ of gate levels. Increase of parallelism level from version R to version 1 increases AS throughput by R times. For $R = 5$, the circuit becomes complicated by Quine from $C_B = 1581$ to $C = 7595$ that characterizes investment of resources in the course of circuit design. Domination of zero grows in circuit points from 58.4 % to 97.6 %. It increases trustworthiness of the result, calculated in case of stuck-at fault insertion, from $W_B = 80$ % to $W = 86$ %. The number of transitions decreases from 42.8 % to 3.5 % that characterizes power consuming, i.e. investment of resources in the course of circuit use. The relative indexes of parameters improving accept the following values: $K_{TP} = 5$, $K_{TW} = 1.4$, $K_{RC} = 2.5$, where $K_{RC} = K_{RC\ D} \cdot K_{RC\ \ U}$, $K_{RC\ D} = 0.2$ and $K_{RC\ U} = 12.2$ are relative indexes of resource consuming in case of design and operation of the circuit. An integral assessment $K_{IA} = 18$ testify to the multiple effect. This example shows the development of the multiple effect methods and green technologies with unconscious assistance of human, and it proves objective nature of such development.

Ordering of parallelism types according to their level opens possibilities of plan development of the multiple effect methods. It is enough to take a step to the following level. We suggest taking a step from the matrix parallelism reflecting replication level to the pipeline which corresponds to diversification. Now digital components of computer systems are built in the form of pipelines. But their sections are single-cycle units with matrix parallelism for data processing in parallel codes. But these codes will be processed sequentially from LSB to MSB. This shows dependence of matrix parallelism by data. The single-cycle n-bit IAM in version of the fastest circuit executes operation for $2n - 2$ time delays of full adders [21], i.e. their such quantity is connected sequentially, and each of almost n^2 of full adders is used with coefficient $K_U = (2n - 2)^{-1}$. For $n = 32$, more than 10^3 full adders are used only with $K_U = 1.6$ %. With transition to 64-b platform, K_U deteriorates twice and becomes less than 1 % of operation time.

Besides, parasitic transitions of signals owing to different delay of their propagation (critical races) additionally increase power consumption. It is known, that 8-b ripple-carry adder with uniformly distributed set of random input pattern will typically consume an extra 30 % in energy [4]. Research of the IAM on program model shows that execution of one multiplication changes values of signals in 30 % of the digital circuit points. This necessary number of transitions increases 3, 3.5, 4, 4.5 times for the 8-b, 10-b, 12-b, 14-b multiplier, respectively, i.e. the number of parasitic transitions grows for 50 % with increase in the size of an operand by two bits.

To increase the level of parallelism it is advisable to use the bitwise pipelined organization of computation and minimize the matrix parallelism of pipelined sections, bringing it to single operational element, which processes the numbers bit-by-bit. Such a parallelization in serial codes can be implemented in the form of a multi-threaded processing. In this situation, the matrix parallelism is used at macro level of the system but in the best form, as a set of independent pipelines which operate simultaneously, i.e. there is no dependency on data. In this case, the pipeline becomes the bitwise unit with data processing in sequential codes. Throughput increases together with parallelism level: all bits will be processed at the same time on different sections of the pipeline. Complexity of the unit decreases from square dependence on size of operands to the linear. Propagation delays of signals will be aligned. It eliminates losses of energy in critical races. Besides, the quantity of inputs and outputs of the multiplier repeatedly decreases, raising a level of its self-sufficiency (independence from external units) in the path to the level of autonomy in resources.

The bitwise pipeline executes multiplication for $2n$ clock cycles. Duration of one clock cycle is defined by a summary time delay of an operational element and the register of the pipeline. The IAM executes operation for one clock cycle, duration of which is determined by the amount of delays of $2n - 2$ operational elements and the register. The relative index of throughput in comparison with the IAM can be evaluated as $K_{TP} = (z + 1)^{-1}$, where z is a ratio of time delays of the register and an operational element. Complexity $C_B = n^2$ and $C = 5n$ of iterative array and bitwise pipeline multipliers defines the relative index of the resources used at a design stage as $K_{RC} = n/5$. Relative index of the trustworthiness determined by reliability of the circuit is evaluated in view of complexity as $K_{TW} = K_{RC}$. The integral assessment $K_{IA} = (n/5)^2/(z + 1)$ characterizes bitwise pipelining as a method of the multiple effect.

The multithreaded system opens a resource niche for replication of bitwise pipelines and focusing of the multiple effect on improving of the parameters important for enhancement of green technologies using a method "Robbing Peter to pay Paul".

Focusing of the multiple effect on power saving can be executed by use of a scoring in the ratio of throughput to complexity for lowering of clock frequency. It allows reducing power consumption by lowering of the supply voltage. This is followed by deceleration of computation and convergence of levels of the “0” and “1”. Such focusing of the multiple effect can be executed within lowering of clock frequency and requirements to noise immunity. In case of saving the ratio of throughput to complexity reached in the IAM, the multithreaded system can work at the frequency reduced by $K_{RP} = K_{RC}\ K_{TP}$ times. This allows the appropriate undervoltage of a supply. For example, in case $z = 2$, clock frequency can be reduced by $n/15$ times that determines $K_{RP} = 2.1$ for $n = 32$ and allows to lower energy consumption by 4.4 times for its square dependence on supply voltage.

The multiple effect can be focused on the trustworthiness at a design stage or during functioning by methods and means of on-line testing. In the first case, trustworthiness of results increases with growth of reliability due to simplification of the multiplier circuit. In case of saving throughput reached in the IAM, complexity C/K_{TP} of multithreaded system defines the relative index of trustworthiness as K_{RP}. In the second case, trustworthiness increases by elimination of non-authentic results with probability equal to trustworthiness of a method of on-line testing.

The IAM and the bitwise pipeline multiplier were designed in the CAD Altera Quartus II v.13 and implemented for the operand size $n = 16$, 24, 32 and 40 in EP2C35F672C6 Altera Cyclone II FPGA chip [22].

The experimental results show multiply effect $K_{IA} > 1$ for $n > 32$, where the pipeline method becomes more efficient than the matrix one. However, the experimental estimation many times loses to the expected one. This effect can be explained by that the FPGA design of the IAM is based on the method of result preparing, which pertains to the level H of the resource development [16, 23]. The method provides a fast propagation of the carry and sum in the IAM owing to use of the conditional carry technique and prepared circuits of carry propagation between the adjacent logical elements. It repeatedly reduces time of computation in the IAM. This fact shows the significant role of the method of result preparing in efficient design of the digital circuits.

4 Conclusions

All history of development of the CW and green technologies is an illustration of successful integration of the artificial world created by the human to the NW. The challenges solved in integration process include problems of the CW and the green technologies developed in it. The solution (process and result) is based on achievement of a certain throughput and trustworthiness (both in process and in the obtained results) and also on investment and receiving of resources which are models, methods and means. Throughput is aimed at receiving the maximum return from resources, and trustworthiness characterizes a harmonization level with the NW. Resources develop from simple forms to real by structuring under the NW features among which parallelism and fuzziness are the most shown in the CW. Development of resources shows three levels: replication, diversification and autonomy. They are serviced by different methods which are aimed at improving of throughput, trustworthiness and access to resources, respectively.

Models, methods and means reflect a multistage model of development where each preceding stage serves for the next one. Replication, diversification and autonomy are also stages of a multistage model. Green technologies as model occupy autonomy level on the way to energetic self-sufficiency. The used models and methods belong to the level of diversification of both hardware and operation modes by their profiling to features of the problem. Means are built with replication of operational elements.

Conditions of a solution determine its parameters: throughput, trustworthiness and turnover of resources. Traditional methods improve parameters due to deterioration of others by the principle of "Robbing Peter to pay Paul". They reflect replication level of resource development in case of the unfilled resource niche.

Transition to diversification opens methods of the new nature. They provide the multiple effect in improving of a number of parameters without loss in others only due to development of resources. Basis for development of the multiple effect methods is the natural way in increasing a level of parallelism and fuzziness. But now, as a rule, these methods develop blindly or unconsciously. Execution of the truncated operation almost twice simplifies the circuit of the IAM of mantissas and reduces time of computation. However these types of improving were for the first time noted by different researchers. Domination of zero values in the parallel and approximate computer world provides development to many methods of the multiple effect. However their multiple effect remains in the aspiration shadow to increase only of throughput.

Evaluating a resource level of development, we can generate methods of the multiple effect in case of approximation to the following level, for example, strengthening the pipeline principle of approximate data processing. The reached multiple effect opens resource niches for replication with focusing on the most important parameters with use of a method "Robbing Peter to pay Paul". For green technologies, such parameters are trustworthiness as a safety basis, and an energy component of consuming and creation of resources.

References

1. Ross, P.E.: 5 Commandments, IEEE Spectrum, Dec. 30–35 (2003)
2. Kharchenko, V. (ed.): Green IT-Engineering. Vol. 1. Principles, Components and Models—593 p. National Aerospace University KhAI, Ukraine (2014) (in Russian)
3. Murugesan, S., Gangadharan, G.R.: Harnessing Green IT. Principles and Practices. Wiley and Sons Ltd., UK (2012)
4. Chandracasan, A.P., Sheng, R., Brodersen, S.: Low-power cmos digital design. IEEE J. Solid-state Circ. **27**(4), 473–484 (1992)
5. STMicroelectronics STM8L Family Power Management (03.2010) [AN3147]. url: http://www.bdtic.com/DownLoad/ST/AN3147.pdf

6. Cool'n'Quiet™ Technology Installation Guide for AMD Athlon™ 64 Processor Based Systems [2004 Advanced Micro Devices, Inc]. url: http://web.archive.org/web/20070409045621
7. Larabel, M.: Intel EIST SpeedStep. [Phoronix] (2006) url: http://www.phoronix.com/scan.php?page=article&item=397&num=1
8. Greenhalgh, P.: big.LITTLE Processing with ARM Cortex-A15 & Cortex-A7. White paper ARM, 1–8 (2011) url: http://www.arm.com/files/downloads/big_LITTLE_Final_Final.pdf
9. IEEE Std 754™-2008 (Revision of IEEE Std 754-1985) IEEE Standard for Floating-Point Arithmetic. IEEE 3 Park Avenue New York, NY 10016-5997, USA (2008)
10. Kekre, H.B., Mishra, D., Khanna, R., Khanna, S., Hussaini, A.: Comparison between the basic LSB replacement technique and increased capacity of information hiding in LSB's method for images. Int. J. Comput. Appl. **45**(1), 33–38 (2012)
11. Lim, Y.: Single-precision multiplier with reduced circuit complexity for signal processing applications. IEEE Trans. Comput. **41**(10), 1333–1336 (1992)
12. Garofalo, V.: Truncated binary multipliers with minimum mean square error: analytical characterization, circuit implementation and applications. Ph.D. Dissertation. University of Studies of Naples "Federico II", Naples, Italy (2008)
13. Nicolaidis, M., Zorian, Y.: On-line testing for VLSI—a compendium of approaches. Electron. Test. Theory Appl. (JETTA) **12**, 7–20 (1998)
14. Drozd, A., Lobachev, M., Drozd, J.: The problem of on-line testing methods in approximate data processing. In: Proceedings of 12th IEEE International On-Line Testing Symposium, Como, Italy, pp. 251–256 (2006). doi:10.1109/IOLTS.2006.61
15. Drozd, A., Lobachev, M.: Efficient on-line testing method for floating-point adder. In: Proceedings of IEEE Design, Automation and Test in Europe (DATE), Munich, Germany, pp. 307–311 (2001). doi:10.1109/DATE.2001.915042
16. Drozd, A., Kharchenko, V. (eds.): On line testing of the safe instrumentation and control systems—641 p. National Aerospace University named after N.E. Zhukovsky "KhAI", Ukraine (2012) (in Russian)
17. NVIDIA CUDA Compute Unified Device Architecture. Programming Guide/Version 1.0, NVIDIA Corporation (2007)
18. Maevsky, D.A.: New Approach to Software Reliability. Lecture Notes in Computer Science: Software Engineering for Resilient Systems, vol. 8166, pp. 156–168 (2013)
19. Rabinovich, Z.L., Ramanauskas, V.A.: Typical Operations in Computers. Technika, Kiev (1980). (in Russian)
20. Savelyev, A.Y.: Applied Theory of Digital Machines. High School, Moskow (1987). (in Russian)
21. Melnik, A.O.: Architecture of Computer: Volinska oblasna drukarnja, Lutsk, Ukraine (2008) (Ukrainian)
22. Cyclone FPGA Family Data Sheet. Altera Corporation (2003), http://www.altera.com
23. Drozd, J., Drozd, A.: Models, methods and means as resources for solving challenges in co-design and testing of computer systems and their components. In: Proceedings of the 9th International Conference on Digital Technologies, May, 29–31, Zhilina, Slovakia, pp. 176-180 (2013)

Part II
Components and Programmable Systems

Green Logic: Models, Methods, Algorithms

Sergey Tyurin and Anton Kamenskih

Abstract The green computing systems should be not only energy-aware or energy-efficient but also naturally reliable computing systems. Combination of these approaches gives many advantages for applied computing systems. However, estimations for them should be done very carefully. Delay insensitive circuits occupy important role in designing of green hardware, but reliability improvement techniques of delay-insensitive circuits should be further researched. In this chapter, the new indexes of estimating efficiency of computing systems are proposed. These indexes help developers in complex estimation of computing systems operating in wide supply voltage range and temperatures. The semi-modularity for fault-tolerant circuits is analyzed. The synthesis technique of fault-tolerant delay-insensitive circuits is proposed. The technique helps to researchers and developers of Green Hardware in optimization energy consumption, complexity, reliability and performance.

Keywords Green computing · Energy-efficiency · Energy-modulated computing · Reliability · Fault-tolerance · Delay insensitive · Self-timed · Energy-reliability

1 Introduction

An interesting way to enhance energy-efficiency of digital devices without using of "hibernation" or "disconnection" of blocks that are not operating is to reduce voltage supply. At the same time, it can be used not only to reduce energy-consumption when reducing supply voltage but also to provide possibilities of dynamic changing for energy-consumption and performance (energy-modulation) [1]. Modern digital devices work under supply voltages close to 1 V but some

S. Tyurin (✉) · A. Kamenskih
Perm National Research Polytechnic University, 29 Komsomolsky Avenue, Perm, Russia
e-mail: Tyurinsergfeo@yandex.ru

A. Kamenskih
e-mail: Kamenskikh.anton@gmail.com

V. Kharchenko et al. (eds.), *Green IT Engineering: Concepts, Models, Complex Systems Architectures*, Studies in Systems, Decision and Control 74,
DOI 10.1007/978-3-319-44162-7_4

already use less than 0.5 V, what is very close to transistor thresholds. The development of energy-modulated computing systems is important not only in fields such as mobile devices or PC but also in energy-critical systems such as aerospace microelectronics [2] or remote computing systems with limited power generation (for example, distributed control systems) [3]. Energy-modulated computing systems, unlike systems with "hibernation", can perform planned tasks slower and with less energy-consumed when there is a reduction in the voltage supply. Apart from energy-consumption optimization, this can also be used for an emergency shutdown or another necessary task in case of a supply fault.

The systematic development of technologies results in the need for qualitative change in design approaches. This became a reason for the intensive development of an asynchronous approach. It will be actively forcing out usual synchronous decisions according to the forecast of ITRS [4]. The synchronous computing system is always limited by frequency of clock signal. Therefore, energy-modulation cannot be done in real-time. This restriction is not present in asynchronous computing systems. Nowadays, asynchronous approach actively is being developed in the direction of Green computing systems, development of computer-aided design systems and high-reliable computing systems.

The reduction of supply voltage results in the intensive decrease of performance and increase of failure rate. Therefore, designers and developers of energy-modulated computing systems should carefully evaluate the efficiency of a system considering all-important parameters: energy-consumption, performance, reliability, complexity and their auxiliary parameters such as noise stability. It is especially necessary for energy-critical and highly reliable computing systems. The complex approach of "energy-reliability" is proposed to help to evaluate and compare computing systems for energy-critical applications. This approach integrates requirements for energy-efficiency and reliability. It has influence not only on evaluations of devices' parameters but also on designing techniques.

Delay-insensitive circuits (self-timed circuits according to Russian terminology) are a basic solution in designs of energy-modulated computing systems. The self-timed circuits engineering (STCE) became a perspective approach in view of requirements of compensation of negative effects that are related with energy-saving techniques. Self-timed circuit (STC) as asynchronous circuit is operating with real delays and can be easily implemented for energy-modulation. Moreover, strictly self-timed circuits (SSTC) have stable work under ultra-low-supply-voltage (ULSV) and ave self-checking possibilities [5, 6]. All these properties allow controlling energy-consumption and performance when changing supply voltage dynamically.

It should be noted that self-timed (delay insensitive) logic could not be implemented in synchronous hardware. An asynchronous device has limited opportunities in hardware realization. Some designs were oriented on Uncommitted logic Array (ULA). At the same time, there are the asynchronous FPGA architectures [7, 8] oriented on specific self-timed designs. In general, asynchronous FPGA is much less available than synchronous. In addition, not all self-timed implementations and sub-classes will be suitable for « Green » or highly reliable computing systems [9].

The scientific groups of A. Yakovlev from Newcastle University, Y. Stepchenkov from IPI RAN (Institute of information problem at Russian Scientific Academy), S. C. Smith from University of Arkansas and others are researching and developing this field continuing the fundamental works started by Muller [10] and Varshavksy [11]. These groups focus on the development of self-timed designing theory and designing of self-timed computing systems. However, reliability problems of self-timed devices in their works were presented only in general. To successfully compete with the synchronous approach, the asynchronous approach should have the same opportunities in circuits' synthesis. Insufficient research of self-timed circuits' reliability problems resulted in self-timed circuits not taking a worthy place in microelectronics for critical applications. This is why the authors decided to research self-timed circuits in terms of reliability and then connect results with existing practices in the designing of « Green » computing systems. For this purpose, authors propose an approach based on STCE that provides optimal combination of energy-consumption, performance and reliability.

It should be noted that features of self-timed approach make actively fault-tolerant techniques (self-repair, reconfiguration) more perspective than passively fault-tolerant techniques, but some computing systems require exactly passively fault-tolerant circuits.

A **goal of the paper** is to improve existing analysis and synthesis techniques of green hardware via research and development of delay-insensitive approach in combination with reliability theory. The second part of the paper is devoted to development of optimization indexes that will help in synthesis of green computing systems. The third part of the paper presents the complex analysis and the synthesis technique of fault-tolerant delay-insensitive circuits.

2 Principles of Energy-Reliable Computing

When considering the calculation process, the three main characteristics can be allocated [12]. They are performance, validity and the used resources including energy, time, equipment etc. [13]. Performance shows the volume of calculations that can be done within a unit of time. Validity is the property showing that the calculation process result is in tolerance limit. We should remember that if the result is not valid, then it cannot be used.

The traditional definition of energy-efficiency is complex index that shows how rationally the task is performed in terms of energy-consumption. In general, this index considers such parameters as energy consumption and performance, and other parameters are excluded as insignificant. When using the energy-modulation technique, failure rate cannot be excluded as insignificant because it has important influence on energy-efficiency. The failure rate increases very intensively with decreasing supply voltage close to 0.5 V. Thus, results will be valid less and less, and resource expenses can exceed the saved amount of energy.

It is known from the reliability theory that failure rate depends on temperature, supply voltage and other environment parameters. Reduction of the supply voltage increases failure rate but decreases heat dissipation that results in temperature decreasing, which in turn decreases failure rate. Therefore, all parameters of a «Green» device take important role in estimations of energy efficiency. It is the main principle of the proposed energy-reliability approach.

Summarizing, the result of calculation is not valid in case of failure. Therefore, it should be recalculated, and the *efficiency of resources use is equal to zero* in this work cycle. Thus, reliability (reliability probability, validity, fault tolerance etc.) have a great influence on energy-efficiency and should be taken into account in « Green » and energy-critical computing systems.

2.1 Energy-Reliability Optimization

It is well known that decrease of the supply voltage causes energy-consumption and performance decreasing, but the influence of decrease of the supply voltage at reliability is investigated not so well. Let us do a short review of dependences in digital circuit parameters (energy-consumption, performance, reliability, complexity). It should be noted that features of self-timed circuits often do not allow using well-known formulas.

The circuit complexity will be estimated in number of transistors. Self-timed circuits have two data channels, indicators and other gates that can have different number of transistors depending on design technique. Therefore, comparison should be done at the lowest level of transistors number then it will be correct. For example, Muller's C-gate in static realization has 12 transistors and 8 in semi-static.

The reliability probability calculation is given in exponential model of distribution of time before failure, formula (1). The failure rate is presented as a function of environment conditions (ε) that includes supply voltage and temperature:

$$P(t) = e^{-n\lambda(\varepsilon)t}, \tag{1}$$

where t—time in hours, n—number of transistors, λ(ε)—failure rate.

The frequency is still main performance parameter in spite of the fact that it is not constant in self-timed circuits. Self-timed circuits use architecture with "request and answer" that means one period is equal to time of a single operation. Thus, it excludes empty switching (periods without operation) that reduces energy-consumption.

However, to estimate energy-consumption, it is necessary to present the energy consumption (power) divided by frequency (PDF), formula (2) [13]. The index (2) shows energy-consumption of operation that gives objective characteristic to the energy-efficiency in both synchronous and self-timed circuits. The frequency of request signal is accepted for operation frequency in self-timed circuits:

$$W_F = \frac{I \cdot U}{F} \tag{2}$$

The following picture is received summarizing simulation results in CAD MultiSim. The function of PDF has linear relation with frequency. The more performance is required the more energy is spent for each calculation step, Fig. 1.

The decrease in the voltage supply from 5 to 3.5 V reduces energy-consumption on operation almost two times, but it results in the increase of failure rate that reduces reliability probability. In [14, 15] the impact of supply voltage and temperature on failure rate is researched. The research results and simulation data allow the extrapolation of simulation results received for self-timed circuits. It should be noted that extrapolation for higher supply voltage than that used in [14, 15] is a linear part of function while the exponential growth of failure rate is observed at values close to 0 V. Thus, ε increases the failure rate by ten per each 0.5 V drop in the interval from 5 to 3.5 V, Fig. 2.

The combination of previous indexes will give the indexes of energy-reliability. The gradient of energy-reliability is index that shows the reduction rate of PDF per unit of reliability probability, formula (3). The slower reduction of reliability probability leads to the better gradient of energy-reliability. In addition, ΔW_F increases energy-reliability because decreasing of energy-consumption gives a positive effect while decreasing of reliability probability gives negative. This index

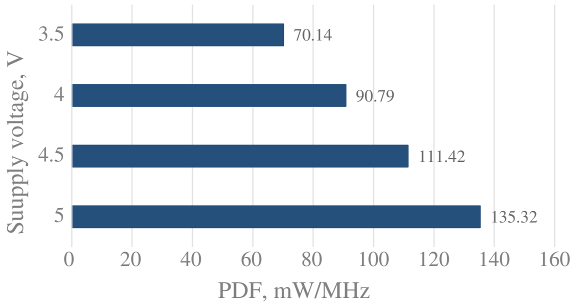


Fig. 1 PDF function of supply voltage

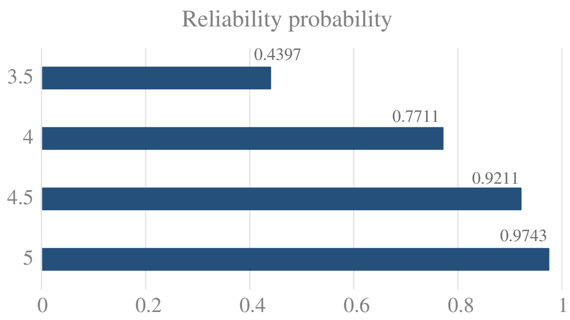


Fig. 2 Reliability probability function of supply voltage

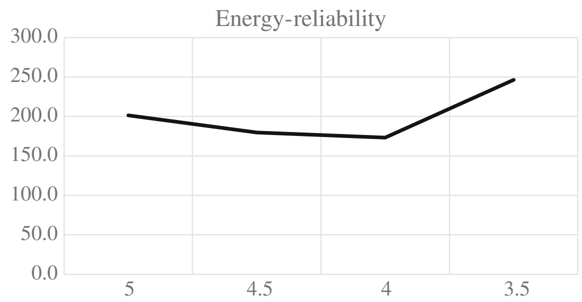


Fig. 3 Energy-reliability of not redundant self-timed circuit

shows the potential of circuit in energy-modulation in view of reliability loss that helps to estimate efficiency of designing techniques:

$$\mathrm{E}'_P = \frac{\delta W_F(u)}{\delta P(u,t)} \tag{3}$$

To exclude obviously losing decisions, the index (4) is proposed. It shows the reliability of energy using in moment t. This index helps to find optimal operation mode in which fault-tolerant circuit can use less energy than not redundant one. Both indexes can be used to compare self-timed circuits with synchronous. The energy-reliability has inverse relation of reliability. It means that the lower values correspond to the better conditions, Fig. 3.

$$\mathrm{E(U)} = \mathrm{W_F}\frac{1}{P(t)} \tag{4}$$

The simulation shows that in terms of energy-reliability the energy-modulation is efficient only in small range of voltage changing. Thus, the rapid increase of failure rate has to be compensated to allow using of energy-modulation in computing systems for highly reliable applications. The fault-tolerant designing techniques increase reliability probability, but energy consumption is also increased. The self-repair and reconfiguration techniques are less redundant but mostly are oriented to irreversible faults. In next part, it will be shown how redundancy influences on energy-reliability.

2.2 Influence of Redundancy on Energy-Reliability

The reciprocal influence of reliability and energy-efficiency was described in previous part of this chapter. Now we should investigate how redundancy influences on energy-reliability. Can redundancy improve it or all advantages will be eliminated by increase of energy consumption?

Table 1 Energy-reliability calculation results

	ΔRP			ΔPDF			E'		
ΔV	NR	TR	RG	NR	TR	RG	NR	TR	RG
0.5	−0.053	−0.036	−0.009	−30.7	−65.7	−33.1	577	1809	3801
1	−0.150	−0.168	−0.027	−32.1	−77.0	−36.8	214	460	1384
1.5	−0.331	−0.485	−0.080	−25.2	−84.4	−24.3	76	174	306

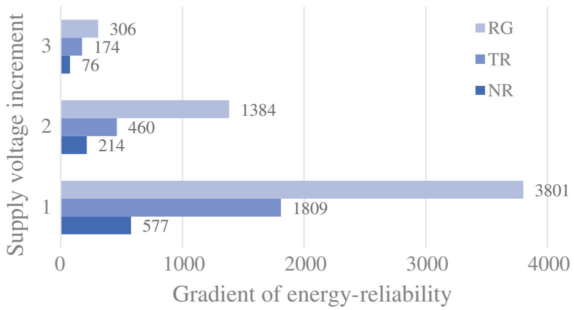


Fig. 4 The comparison of energy-reliability gradient

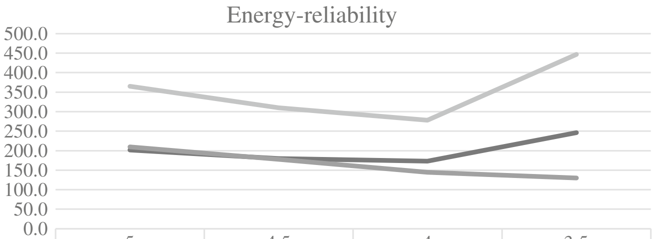
The different redundancy techniques have different influence on energy-consumption. The comparison of circuits with no redundancy, with redundancy at transistor-level and n-of-m redundancy is presented here. The redundancy techniques in more complete description are considered in part 3.

All simulation data (Table 1) are received for equal circuits that use different fault-tolerant designing techniques. NR—not redundant (initial) circuit, TR—triplication redundancy (2-of-3 voting), RG—redundant gates (transistor-level redundancy), Fig. 4.

The TLR on average improves gradient of energy-reliability 2.3 times in comparison with 2-of-3 redundancy and 5.7 times in comparison with the not redundant circuit. In this example, the advantage is caused by the very good dynamics of TLR reliability probability.

When considering static parameters of reliability probability and energy-consumption, the TLR is better than 2-of-3 redundancy in terms of energy-reliability in all range of supply voltage changing, but it worse than not redundant circuit in initial range from 5 to 4.5 V, Fig. 5.

From the Fig. 5, it is visible that 2-of-3 redundancy is worse than not redundant circuit. It happens because 2-of-3 redundancy gives improvement of reliability probability only in small range, and this type of redundancy significantly increase energy consumption. It means that only specific redundancy techniques can make energy-modulation perspective for highly reliable or energy-critical computing systems. Redundancy at transistor-level corresponds with this characteristic because it is better in terms of reliability than not redundant circuit in all time range. Moreover, TLR slightly increases PDF. However, it is correct only for “deep” reservation if redundancy was added at high-level (for example, three microkernels



	5	4.5	4	3.5
NR	201.4	179.7	173.1	246.3
TR	364.8	309.8	277.8	446.5
RG	209.4	177.7	144.3	129.8

Fig. 5 The comparison of energy-reliability

with 2-of-3 voting) then redundant circuit will be better at some interval. The range where function has the lowest values shows efficient operation mode in terms of energy-reliability, but it is necessary to remember about performance decreasing.

3 The Synthesis Technique of a Fault-Tolerant Delay-Insensitive Circuit

The synthesis task is formalized as following: The number of possible faults *k* and the maximum of sequentially connected transistors *q* is given. A logic circuit is set as in DNF for variables *x* that defines pull-up network (with dual-rail inputs):

$$f_{d+} = \bigvee_{i=1}^{v} \mathop{\&}_{j=1}^{n_{d+}} (\bar{x}_{ij}) \tag{5}$$

Where υ_{d+}—number of disjunction arguments, n_{d+}—number of conjunction arguments. The dual function for pull-down network:

$$\bar{f}_{d+} = \bar{f}_{d-} \bigvee_{i=1}^{v_{d-}} \mathop{\&}_{j=1}^{n_{d-}} (x_{ij}) \tag{6}$$

To receive: A passively fault-tolerant self-timed circuit (FTSTC) using combined redundancy with optimization task:

Task 1: To find $\omega = \{i_{d.h}\}$, $L_\omega \rightarrow \min$ with $P(t)_\omega \geq P(t)_{set}$

Task 2: To find $\omega = \{i_{d.h}\}$, $P(t)_\omega \rightarrow \max$ with $L_{set} \geq L_\omega$

Where ω is a vector of circuit's sections with redundancy, d—is the number of circuit section, h—the number of redundancy type.

The design of self-timed circuits is integrated with problems of analysis. The lack of signals competitions should be proved. With increasing of gates number and circuit's parallelism, the analysis becomes a difficult computing task. The both parameters are increased by multiple in a redundant circuit in comparison with not redundant one. Moreover, requirements to reliability (redundancy, fault-tolerance etc.) conflict with requirements taking into account self-timed theory (semi-modularity).

3.1 About Non Semi-Modularity of Passively Fault-Tolerant Circuits

Let us see why an n-of-m redundant circuit cannot be semi-modular. Muller's C-gate is a basic element of all delay insensitive circuits. The logic function (output) of this gate holds the value of arguments. For example, when all inputs are asserted then a gate will be asserted otherwise a gate hold current value. The gate will be deasserted if all inputs will be deasserted. Asynchronous process can be described as inverters within restrictions of Muller's model, Fig. 6.

It is easy to see that such process is hazard-free because each gate waits for other gates. Moreover, this circuit is semi-modular [16].

Definition 1 A logic circuit $F(x_1, x_2 \ldots x_n)$ is semi-modular if its unstable variables cannot become stable without changing their value.

In [17] semi-modularity is described as a property inherited from mathematics. It can be simply defined using grid with cells (gates). Each cell has switching counter. When gate activated, the value of counter is increased by one. When gate is deactivated due to change of input set, the value of counter is decreased. When circuit is operating, and if no cells in the grid are decreasing their value then circuit is semimodular.

In practice, semimodularity of circuits depends on their initial state. That is why all programs for analysis are working with the pair—logic equations of circuit

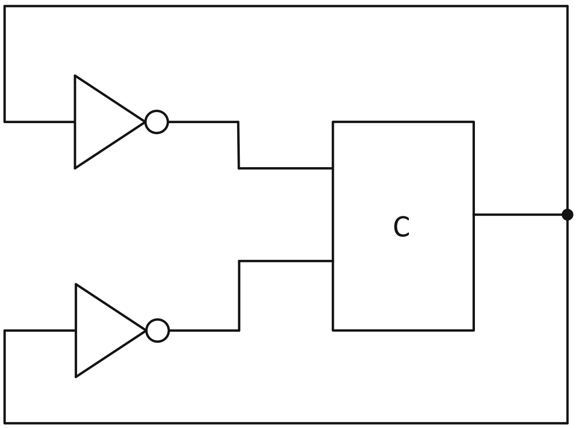


Fig. 6 Synchronization of processes using Muller's C-gate

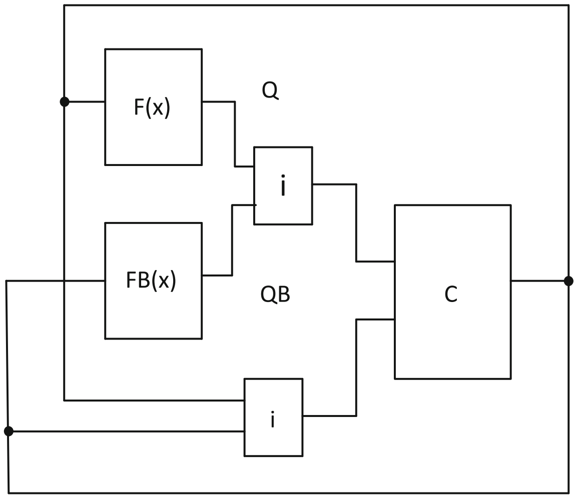


Fig. 7 Synchronization in strictly self-timed circuits, where i—an indicator's logic function

F(x_1, x_2…x_n) and initial state *A* = {x_1, x_2…x_n}, *A* belongs to set of work states or null (spacers).

Strictly self-timed circuits are designed with architecture that is more complex. It uses the indicators of transition process completion and the special state between data states, Fig. 7.

An attention should be paid to the fact that a circuit will not be semi-modular without indication of both inputs and outputs.

Definition 2 A logic circuit F (x_1, x_2…x_n) is passively fault-tolerant if in the case of internal fault Q ($x_{1\text{-}n}$) it saves initial function without using reconfiguration or repair.

For example, inverter F(x) has logic function NOT x. Fault-tolerant inverter FT (x) has logic function $FT(x_{1-4}) = \overline{x_1 x_2} \vee \bar{x}_3\bar{x}_4$. In the case of stuck at one fault $Q^1(x_n)$ for one input the logic function F(x) will be stuck-at null $F(Q^1(x)) = 0$ and the logic function FT(x) will be $FT(Q^1(x_2), x_{1,3,4}) = \bar{x}_1\bar{1} \vee \overline{x_3 x_4} = \bar{x}$.

Proposition 1 *Let FT (x_1, x_2…x_n) be a passively fault-tolerant delay insensitive circuit designed using redundancy at gate-level (or higher). Then it is not a semi modular.*

Proof High-level synthesis uses logic equations of gates from special libraries to design circuit. Muller's model of asynchronous circuit has restriction "one equation – one gate". If gate have stuck-at fault then its value cannot be changed. Therefore, semi-modularity property cannot be executed (Definition 1). It is normal situation for fault-tolerant circuit because of redundancy FT(Q($x_{1\text{-}n}$), x_1, x_2… x_n) ≠ const, but it is mistake in terms of Muller's model.

Proposition 1 does not allow using existing CAD techniques to analyze a fault-tolerant circuit on semi-modularity. Therefore, it cannot be proved that a fault-tolerant circuit belongs to strictly self-timed circuits.

Definition 3 *A logic circuit $F(x_1, x_2 \ldots x_n)$ is delay-insensitive (self-timed) if its diagram of transitions (changes) does not contain: 1) deadlock states 2) logic and functional hazards.*

Proposition 2 *If all local functions of circuit $F(x_1, x_2 \ldots x_n)$ are semi-modular and the behavior in bundle is delay-insensitive then it is hazard-free and delay-insensitive.*

Proof To prove proposition 2, the concept of local semi-modularity will be used. It is the same semi-modularity but for a limited group of elements (gates). It is easy to separate circuits on N sections in the context of fault-tolerant design. The first section contains only initial not redundant circuit. The second section contains subsections with number of redundant channels. The third section contains ancillary equipment. It is easy to see that Sects. 1–2 are simple self-timed circuits that can be analyzed with the *identical signals discipline* by existing CAD. The problem of analysis appears when all sections are connected, and completion of transition process in one of them is ignored. Thus, strict compliance of signals discipline is required in Sect. 3. It should be delay-insensitive and analyzed depending on signals discipline of Sects. 1–2. When performing these conditions, the «safety» of calculations is guaranteed in CMOS technologies despite of semi-modularity losses. Thus, compromise between restrictions of reliability and delay-insensitiveness is provided.

3.2 Redundancy at the Transistor-Level

The redundant gates can be used to provide passive fault-tolerance except n-of-m redundancy. It is special fault-tolerant gate designed with redundancy at transistor-level. According to the analysis of problems and complexity (the problem of "control of control"), transistor-level redundancy becomes interesting solution because it hides all redundancy from top-level allowing analysis of a circuit on semi-modularity. Transistor-level redundancy (TLR) makes a circuit fault-tolerant, and if it was designed as a delay-insensitive, it will save this property.

TLR uses functionally full tolerant (FFT) functions to provide fault-tolerance of every single transistor. The FFT-functions can be different depending on the quantity of admissible faults k. For saving operability in case of stuck-at fault, TLR adds some redundant transistors in pull-up and pull-down networks. Therefore, these networks are called redundant transistor networks (RTN), and logic gates designed with TLR are called redundant gates (RG). Figure 8 shows different RGN for $k = 1$.

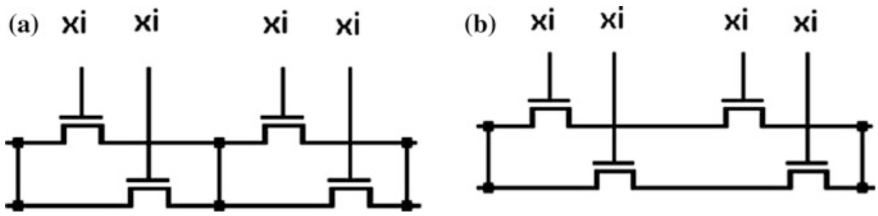


Fig. 8 Redundant transistor networks for saving functions in case of single stuck-at fault

$$(a)\quad f_{1.1} = (x_i \vee x_i)(x_i \vee x_i) \tag{7}$$

$$(b)\quad f_{1.2} = x_i x_i \vee x_i x_i \tag{8}$$

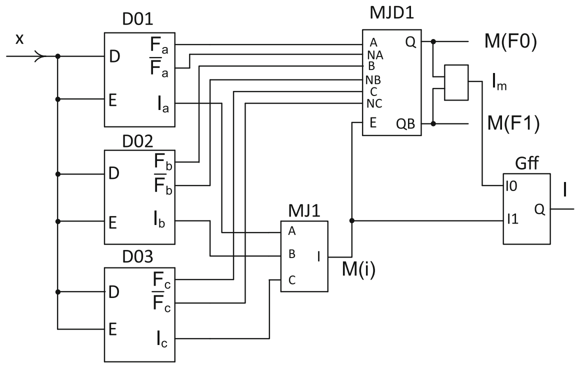
Let us tell a little more details about FFT-functions. Each FFT-function has full form and set of residual forms. For example, in case of stuck-at fault FFT-function (8) has two residual basis. The first, $f_{1.2} = x_1x_2$ v $x_30 = x_1x_2$ and the second, $f_{1.2} = x_1x_2$ v $x_31 = x_1x_2$ v x_3. Thus, the fault-tolerance is provided by the existence of residual base, and FFT-functions are different in reliability probability. However, the implementation of these functions also occupies different space on a crystal.

Let us compare the reliability probability (P) and complexity (L) of RTN shown in Fig. 6 with 2-of-3 redundancy (TR). The example of TR is shown in Fig. 9.

It can be seen that RTN multiply an initial circuit complexity (L_i) by four ($L_{RTN} = 4L_i$), but it does not need any additional equipment. In general, an initial circuit complexity will be multiplied by $k^2 + 2k + 1$. One main channel and two redundant are used in 2-of-3 redundancy, but also the complexity of "shadow" equipment (L_{SE}) such as major-gate and latches (or flip-flops) should be taken into account ($L_{TR} = 3L_i + L_{SE}$). In general, an initial circuit complexity will be multiplied by $2k + 1$ and plus L_{SE}. If $L_{SE} > L_i$ then complexity of RTN less than complexity of TR otherwise RTN more redundant.

The reliability probability is defined in exponential model of distribution of time before failure. A RG is operational if it is not less than $k^2 + k+1$ transistors are

Fig. 9 The fault-tolerant self-timed register cell using 2-of-3 redundancy



operational in each RTN. For example, if k = 1 then not less than three transistors should operate in each RTN. Thus reliability probability of RG for k = 1 will be equal to (9):

$$P(t)_{RTN1} = [e^{-4\cdot\lambda\cdot t} + 4\cdot e^{-3\cdot\lambda\cdot t}(1 - e^{-\lambda\cdot t})]^{L_i} \tag{9}$$

A circuit with 2-of-3 redundancy is operational if not less than two channels is operational. Thus reliability probability of TR for k = 1 will be equal to (10):

$$P_3 = (3\cdot e^{-2\cdot L_i\cdot\lambda\cdot t} - 2\cdot e^{-3\cdot L_i\cdot\lambda\cdot t})\cdot e^{-L_{se}\cdot\lambda\cdot t} \tag{10}$$

For example, comparison of reliability probability for fault-tolerant self-timed full-adders is shown in Fig. 10.

The TLR has possibilities in reservation of power supply. However, taking into account reliability of power supply would not have essential effect on the ratio of reliability probability. The redundant power supplies can be used for channels but not for shadow equipment in n-of-m redundancy. Therefore, often the reliability of n-of-m redundancy is improved using redundant shadow equipment, but it increases complexity and energy consumption. In this case, the reliability probability will be equal to (11):

$$P_3 = (3\cdot e^{-2\cdot L_i\cdot\lambda\cdot t} - 2\cdot e^{-3\cdot L_i\cdot\lambda\cdot t})\cdot(3e^{-2L_{se}\cdot\lambda\cdot t} - 2e^{-3L_{se}\cdot\lambda\cdot t}) \tag{11}$$

The simulation results shows that energy-consumption of circuits with redundancy at transistor-level is less than in n-of-m redundant circuits. The energy-consumption was measured with circuit's autonomous operation frequency

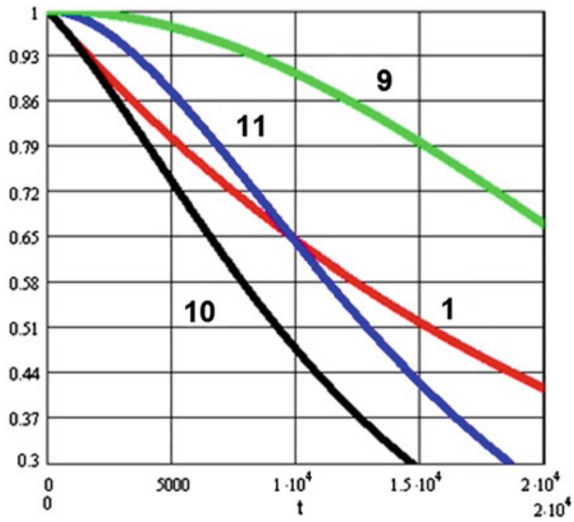


Fig. 10 Comparison of self-timed full-adders reliability probability

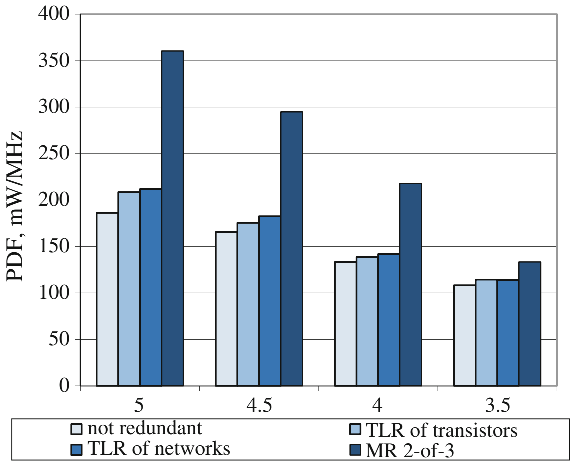


Fig. 11 The comparison of energy consumption of self-timed circuits designed in different fault-tolerant styles

and only for the target circuit. Therefore, it does not consider energy-consumption of ancillary and measuring equipment, Fig. 11.

In conclusion, transistor-level redundancy is actually used only as alternative to deep n-of-m redundancy in view of complexity. Rapid growth relatively numbers of possible faults is compensated by the lack of “shadow” equipment such as major-gate. The transition to lowest possible level of single transistor gives best increase of reliability probability. In addition, it is not unimportant that energy-efficiency of TLR is much better than n-of-m redundancy.

3.3 The Synthesis Technique of a Fault-Tolerant Self-timed Circuit

When synthesizing FTSTC, the technique uses three possible redundancy type. The first type is transistor-level redundancy (TLR) if it is allowed by combination of q and k parameters. In case of transistor-level redundancy impossibility, the next step is decomposition (dTLR) of initial circuit to satisfy restrictions q and k. This step should be done very carefully because decomposition can make circuit worse than initial circuit in all parameters. The distributive property and other properties of logic function (like De Morgan’s law) are used to decompose circuit, but semi-modularity property can demand to decompose circuit several times before correct semi-modular (SM) circuit will be found. We will designate the set of decomposition circuit versions (V) and the set of remained decomposition circuit versions (VR).

The last used approach is n-of-m redundancy (MR) using FCT-functions. This type of redundancy is used to limit the second type. If circuit’s parameters became

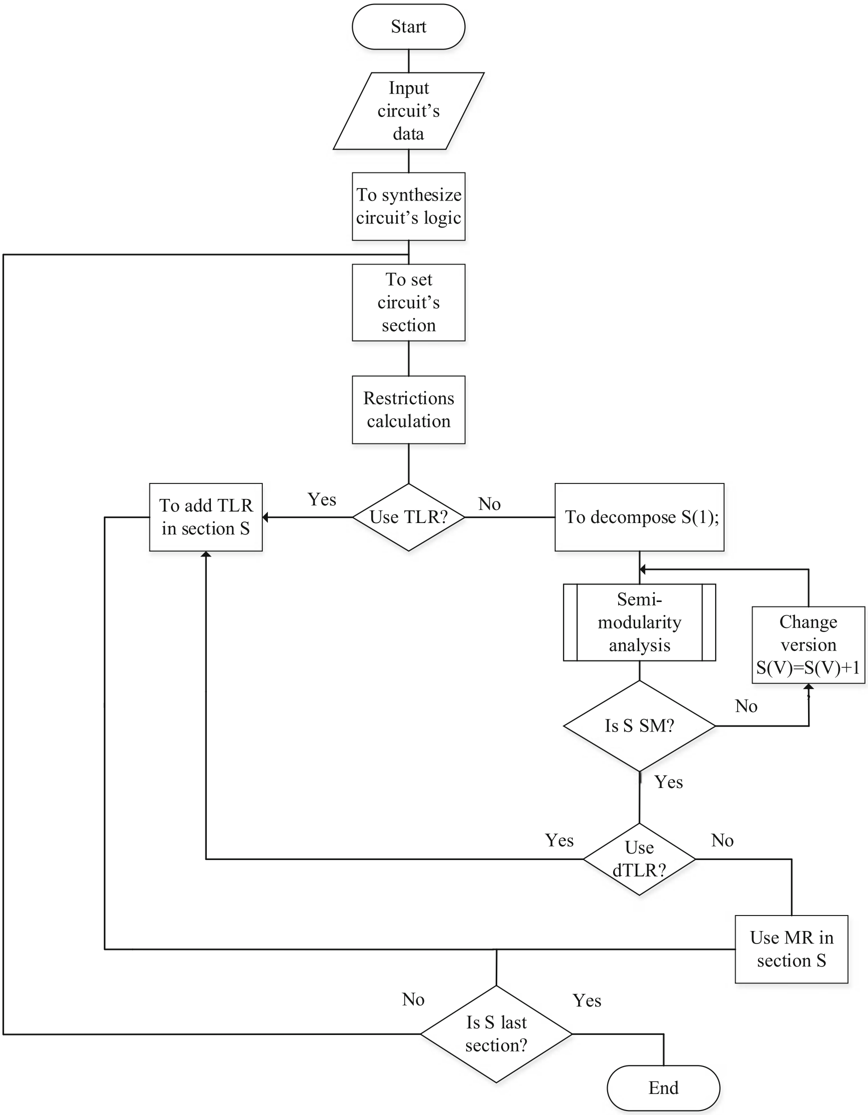


Fig. 12 The synthesis algorithm of a fault-tolerant self-timed circuit

worse than in n-of-m redundancy in case of decomposition, the transistor-level redundancy is prohibited, Fig. 12.

The synthesis algorithm presented on Fig. 12 is used when the complexity of reserved elements is less than 50 transistors in other case the gradient method is used. The algorithm's execution procedure is based on the analysis of fault-tolerant designing techniques in context of self-timed circuits. It does not contain optimization and just synthesize fault-tolerant self-timed circuits because transistor-level redundancy guarantees the maximum increment of reliability probability for strictly delay-insensitive circuits [18]. Thus, TLR should be used if it is possible for the optimization task 2. Gradient method is used if TLR cannot be implemented for some reasons or optimization task sets the minimum complexity.

The gradient method uses the same types of redundancy, but redundant elements are added gradually in sections with best reliability increment. The first step is calculation of initial table of circuit parameters. The next step is to compare parameters of circuit sections with preset-level. The last step is to gradually increase redundancy of circuit sections until circuit parameters will correspond to the required level. For this purpose, the gradient (12) is calculated for all types of redundancy and all sections. Then, a section and type of redundancy with max gradient is selected. The gradient (12) can be modified in order to take into account energy-reliability and performance.

$$\delta_i^j = \frac{P_{\omega.i}^{j+1}(t) - P_{\omega.i}^{j}(t)}{L_{\omega.i} \cdot P_{\omega.i}^{j+1}(t)} \tag{12}$$

where P—reliability probability, L—complexity, j—iteration number.

To compare efficiency of a circuit in difficult situations, the set of Pareto is used. For example, we have two solutions $F_1(x_1, x_2 \ldots x_n)$ and $F_2(x_1, x_2 \ldots x_n)$ both has the same reliability, but $F_1(x_1, x_2 \ldots x_n)$ is better in performance, and $F_2(x_1, x_2 \ldots x_n)$ is better in complexity. In this case, designer should choose better solution.

In [19] the program of reliability calculation was developed to help designers of fault-tolerant logic circuits. In general, the gradient of energy-reliability is used but developed CAD allows changing it. Moreover, it allows adding new redundancy techniques. Pareto set is used, and designer should choose type of redundancy independently if influence of redundancy on other parameters is not equal for different redundancy types. The synthesis technique uses many assessments discussed in [19, 20].

4 Conclusions

Self-timed approach is perspective solution for the problems of microelectronics for aerospace field of application and other similar fields because self-timed circuits are stable to environmental variations (Voltage droop or increasing temperature). In addition, it is perspective in the microelectronics designing for remote, energy-critical, energy-aware, energy-modulated and green computing systems.

The comparison of green computing systems should take into account all parameters including their dynamic components. Thus, complex indexes of energy-reliability were proposed in addition to existing. Energy-reliability gives the possibility of accurate estimation of the efficiency of computing systems. Moreover, it helps to compare different synthesis techniques of fault-tolerant circuits (or highly reliable).

The development of green approach in the field of reliability theory allows using of good practices in computing systems with reliability restrictions. In addition, taking into account the reliability improves estimations of quality of green computing systems. Thus, while doing research, the synergy effect appears.

It should be noted that not only the research and development of theory is needed for dissemination the good practices of Green IT in industries but also the development of educational courses [13].

In this chapter, the solution of the actual practical task (research and development of passively fault-tolerant self-timed circuits) is presented. The features of self-timed circuits in terms of fault-tolerant design were investigated. The theoretical model was expanded to allow analysis of fault-tolerant self-timed circuits. The synthesis technique of fault-tolerant self-timed circuit was proposed. The proposed technique uses new redundancy types including the combined redundancy that significantly increases the energy-efficiency of highly reliable computing systems.

The features of strictly self-timed circuits make reconfiguration and self-repair techniques more preferable. Therefore, in future we plan to develop a technique of dynamic diagnostics of faults using the features of strictly self-timed circuits. In case of self-repair, the indicators of completion of transition process receive additional function of faults diagnostics. It has both advantages and disadvantages. As advantage, it does not need any additional equipment and time to reconfiguration if the “hot” redundancy is used. However, the validity of calculation process will not be equal to indices of passively fault-tolerant circuits.

The developed TLR is very perspective in context of energy-reliability, but it has problems with performance and noise-stability. The TLR already has better characteristics than “deep” 2-of-3 redundancy. However, the modern norms of design can make TLR even better.

References

1. Yakovlev, A.: Energy-modulated computing. In: Design, Automation & Test in Europe Conference & Exhibition (DATE), pp. 1–6. IEEE (2011)
2. Smith, S. C. et al.: Delay-insensitive asynchronous ALU for cryogenic temperature environments. In: Circuits and Systems (MWSCAS 2008) 51st Midwest Symposium, pp. 322–325. IEEE (2008)
3. Lin, T., et al.: An ultra-low power asynchronous-logic in-situ self-adaptive system for wireless sensor networks. Solid-State Circuits, IEEE J. **48**(2), 573–586 (2013)
4. International Technology Roadmap for Semiconductors 2007. Design, http://www.itrs.net
5. Stepchenkov, Y., et al.: Self-timed sequential circuits: development experience and design guideline. Syst. Means of Inform. **17**(1), 503–529 (2007). (In Russian)
6. Kuang, W. et al.: Supply voltage scalable system design using self-timed circuits. In: Proceedings. IEEE Computer Society Annual Symposium, pp. 161–166. IEEE (2003)

7. Hauck, S. et al.: MONTAGE: an FPGA for synchronous and asynchronous circuits. In: Field-Programmable Gate Arrays: Architecture and Tools for Rapid Prototyping, pp. 44–51. Springer, Berlin, Heidelberg (1993)
8. Payne, R.: Self-timed FPGA systems. In: Field-Programmable Logic and Applications, pp. 21–35. Springer, Berlin, Heidelberg (1995)
9. Kamenskih, A.N., Tyurin, S.F.: The development of basic logic element for strictly self-timed FPGA. In: Information and Digital Technologies (IDT-2015), pp. 128–131. IEEE (2015)
10. Muller, D.E., Bartky, W.S.: A theory of asynchronous circuits I. University of Illinois, Graduate College, Digital Computer Laboratory (1955)
11. Varshavsky, V. et al.: Concurrent hardware: the theory and practice of self-timed design. John Wiley & Sons, Inc. (1994)
12. Drozd, A. et al.: The levels of target resources development in computer systems. In: Design & Test Symposium (EWDTS), 2014 East-West, pp. 1–5. IEEE (2014)
13. Kharchenko, V. (ed.): Green IT-Engineering. In 2 volumes. Vol. 1. Principles, Components and Models—593 p. Vol. 2. Systems, Industry, Society.—628 p. Ukraine: National Aerospace University KhAI (2014) (In Russian)
14. Mukhopadhyay, S. et al.: Modeling and estimation of failure probability due to parameter variations in nano-scale SRAMs for yield enhancement. In: VLSI Circuits, 2004. Digest of Technical Papers. 2004 Symposium, pp. 64–67. IEEE (2004)
15. Mehta, N. An ultra-low-energy, variation-tolerant FPGA architecture using component-specific mapping, Dissertation, California Institute of Technology (2012)
16. Brzozowski, J.A., Zhang, H.: Delay-insensitivity and semi-modularity. Formal Methods in Syst Des. **16**(2), 191–218 (2000)
17. Plekhanov, L.P.: Functional method to analyze any self-timed circuits. In: Problems of development of perspective micro-and nano-computing systems, pp. 107–112 (2012). (In Russian)
18. Kamenskikh, A.N., Tyurin, S.F.: Features that provide fault tolerance of self-synchronizing circuits. Russ. Electr. Eng. **85**(11), 677–682 (2014)
19. Kamenskih, A.N., Tyurin, S.F.: Application of redundant basis elements to increase self-timed circuits reliability. In: Proceedings of the Electrical and Electronic Engineering Conference (ElConRusNW) 2014, IEEE NW Russia Young Researchers, pp. 47–50. IEEE (2014)
20. Kamenskih, A.N., Tyurin, S.F.: Advanced approach to development of energy-aware and naturally reliable computing systems. In: Young Researchers in Electrical and Electronic Engineering Conference (EIConRusNW), 2015 IEEE NW Russia, pp. 75–77. IEEE (2015)

Energy-Efficient Scheduling for Portable Computers as Bi-Criteria Optimization Problem

Igor Turkin and Aleksandr Vdovitchenko

Abstract Many of today's portable computers are not inferior to the computing power of a desktop computer. The main problem of portable computers is short duration of activity of the device in standalone mode, while ensuring the quality of service, i.e., subjective user satisfaction. There is no single way of measuring subjective user satisfaction as the current quality of the service requirements conflicts with the requirements to ensure uptime portable computers. Power consumption model of the portable computer is presented as 3-level power graph. The developed model of the multitasking system's scheduling in a portable computer battery life is based on the assumption that the problem under consideration belongs to the soft real-time subject. Synthesis of schedules multitasking system in a portable computer battery life is carried out by solving the bi-criteria optimization problem with the release of the Pareto-optimal solutions. These criteria includes minimum penalty for the decline in the quality of service and maximum battery life with the current profile of power consumption. For experimental investigation of the power consumption PCMark-7 and Microsoft Joulemeter were used. As a result of experiments it is found that during operation on portable computers minimum and maximum energy levels differ by more than 2 times. A time transition to a low or high voltage processor, respectively, can lead to a twofold increase in battery life. The latter is, of course, the most optimistic estimate, since an increase in battery life portable computers only 10–20 % will give the user quite noticeable and necessary advantages.

Keywords Constraint programming · Dynamic voltage scaling · Power model · Pareto-optimal solutions · Portable computers · Scheduling algorithms

I. Turkin (✉) · A. Vdovitchenko
National Aerospace University "KhAI", 17 Chkalova Street, Kharkiv, Ukraine
e-mail: igor.turkin.khai@gmail.com

V. Kharchenko et al. (eds.), *Green IT Engineering: Concepts, Models, Complex Systems Architectures*, Studies in Systems, Decision and Control 74,
DOI 10.1007/978-3-319-44162-7_5

1 Introduction

1.1 Motivation

The rates of desktop computers sales stopped to grow a few years ago, the demand has shifted to mobile devices. Portable computers include both laptops (also known as notebooks) and netbook computers (smaller, less powerful laptops), but excludes mobile computing devices such as smart phones and tablets. According to the report on the energy consumption by consumer electronics in the United States in 2013, presented by Fraunhofer USA Center for Sustainable Energy Systems, on average a device consumes from the power network 53 kWh a year [1], but the assessment of average battery of these computers uptime were not fulfilled.

For mobile devices, the task of creating software (SW), which uses resources of computing platform sparingly is particularly relevant because such an important parameter as the runtime of the device from an independent power source and as a result the consumption of battery life depends on the work of the software. Fast discharging of the battery during the usage time of a mobile computing device without connecting to the power supply in addition to the inconvenience of use has one problem—recycling of batteries after the battery life. Reducing the total power consumption of mobile devices leads not only to increase battery of mobile devices uptime, but also to extend the life of the battery, which in turn reduces the overall costs of the recycling of the battery as a whole. The last thing contributes directly to the implementation of the concept of sustainable human development.

1.2 Goal and Approach

A goal of the paper is carrying out the analysis of the known approaches to energy saving of the PC, in particular based on DVSF. We investigate the peculiarities of the task solution influence on the power consumption of the PC. The results of this analysis will allow to use the modes of the energy consumption of OS efficiently, according to the solving problem.

Stages of the research are the following:

1. The analysis of the existing approaches to a management of mobile devices' energy consumption;
2. The analysis of planning the tasks in the real-time systems;
3. The determination of features of the energy consumption's process for PC
4. Providing the experiment to research the influence of the current supply schemes in OS Windows, on the task solution and energy consumption;
5. The analysis of the experiment results.

2 Related Work

Methods and means of preventing the fast discharging of the battery during operation without connection to the power supply fulfill a control of processor and system bus, peripherals, charge and discharge of the battery to compromise the current requirements for productivity of the computing system and the amount of the energy consumption.

The SmartReflex™ of Texas Instruments (TI) Company—the technology management of energy consumption, which combines intelligent and adaptive silicon, circuit design and software to solve power and performance management challenges at smaller process nodes [2]:

SmartReflex 2 power and performance technologies:

- dynamically adjusts transistor performance versus leakage;
- dynamically lowers voltages for idle memory banks;
- automates the application of SmartReflex™ technologies in the design process [2].

SpeedStep, PowerNow!, Cool'n'Quiet—are energy-saving technologies from Intel and AMD based on dynamic frequency and voltage of the processor [3–5]. At the basis of these energy-saving technologies there is a method of dynamic voltage scaling (DVS) [6]. The applied voltage determines the frequency of the processor (the delay between the execution of instructions), and thus affects the energy consumption. At the same time this switching voltage is energy intensive and quite lengthy (expensive) transition processes. Accordingly, the identification of the optimum frequency of the processor for performing the incoming tasks is an urgent problem. DVS' algorithms are schedulers of voltage in real time, which uses a known dependence of the power consumption of the processor on the operating parameters [7, 8]:

$$P = ACV^2 f + \tau AVI_{short} f + VI_{leak}, \tag{1}$$

where the first item measures the dynamic power consumption caused by the charge and discharge of the capacitive load at the output of each key, which is equal to the product of the capacity—C, the square of supply voltage—V, the processor's frequency—f, and the coefficient A, which characterizes the activity of the keys in the system;

the second item is the power expended as a result of short-circuit current, which takes place at the time of switching the logical element;

the third item are losses due to current leakage.

In general case methods based on the minimization of a function value are used [9]. At these methods each task determines a value that is considered in the case of the mission failure before the time target. Accordingly, the optimization is the minimization of the loss function. We divided the known algorithms of energy-saving planning into 3 groups according to criterion of optimization.

Minimizing the mean time of response. Pruhs et al. [10] proposed an effective Off-line scheduling algorithm that minimizes the mean time for a restriction on the amount of energy. They noted that for any locally optimal plan each work must be carried out in proportion to the amount of unfinished work which will be detained, if the work remains unfulfilled. Albers and Fujiwara proposed the online-version of the scheduling algorithm and analyzed its work in batch jobs' condition [11]. They also formulated its effective autonomous implementation by dynamic programming. Bansalet al. [12] examined the algorithm when processing power is proportional to the volume of uncompleted tasks.

Minimizing the amount of energy or energy efficiency under constraints on time of response or throughput [13].

Minimizing the weighted sum of the execution time and expended energy is more general formulation. This category includes algorithms of interactive dynamic control on basis of Shortest Remaining Processing Time (SRPT), or Highest Density First (HDF) [14].

In one of the first articles on the efficient power management [15] an Online-algorithm of energy-saving control of device with several states: Sleep, Stand-by, Idle, Active is proposed. It has been proved that the stochastic teachings in Online mode have advantages over deterministic strategies. In the article [16] this approach is developed with the help of simultaneous considering of two different mechanisms of energy saving mobile devices. The first mechanism consists in that the unloaded system can be placed in a sleep state, which will require a fixed amount of power to return into the active state. The second mechanism consists in the management of the speed of the processor. The problem is to plan the jobs which are entering the system in the way to minimize total power consumption and guarantee the execution of all operations in given period. The dynamic scheduling algorithm which is proposed by authors with taking into account the possibility of the sleep states solves the problem just partly because of its algorithmic complexity.

At the level of the operating system (OS), the following solutions assigned for use in mobile devices are known [17].

Zeng et al. [18, 19] proposed the framework ECOsystem for control of energy as a resource of the first level in the OS battery powered devices. Authors think that the software applications play a key role in the possibility of a fair distribution of energy, so ECOsystem provides an interface to determine the uptime battery's and application priority. The authors divided the management of energy at the level of the OS for 2 dimensions: the set of system devices that can consume energy at the same time, and the applications that can use the system devices. The authors introduced a new unit called «currency», which characterizes the right of an application to consume a certain amount of energy for a fixed period of time. When the user sets the target of battery life and priorities of ECOsystem application converts the data to the appropriate number of «currency» and determines how many «currency» should be allocated to each application in each time frame. The system was implemented as a modified Linux kernel and got a positive experimental estimate.

Neugebauer and McAuley [20] developed the resource-based OS Nemesis to ensure the quality of service of real-time applications. Nemesis provides control and accounting of the electrical energy usage of all system resources: processor, memory, disk and network devices. To be able to calculate the use of resources for each process, the OS has been vertically structured: most of the system's functions, protocol stacks and drivers of devices are implemented in user-level of shared libraries which are called by application. This design allows you to record the energy consumption of individual applications. If the current energy consumption can lead to disparity of the user's expectations, the system informs the application, after that the application should adapt its behavior. Otherwise, the OS will cancel of the background tasks.

Sachs et al. [21] have created an Illinois project GRACE (Global Resource Adaptation through CoopEration), then they developed it in GRACE-2 [22], in which they proposed to save energy by means of coordinated adaptation at three levels of the system according to changes in demand. These three levels are:

- the global adaptation which takes into account all the applications run inside the system, and all the system levels; this adaptation is realized in case of significant changes in the system, such as launch or shutdown of the application;
- the adaptation of the application, which is held after the end of the time quantum which is allotted to the application;
- the internal adaptation is responsible for the adaptation to the real state of resources.

The ideas of GRACE-2 had been implemented as a part of the Linux kernel, and proved the possibility of saving up to 33 % energy.

Rajkumar et al. [23] proposed four alternative algorithms for implementation of DVFS in real-time systems, and have implemented a prototype as a modified kernel on OS Linux. System performs the choice of algorithm automatically as needed. The rules of selections are defined by the current state of the system taking into account an assessment of overhead charges to change the frequency and voltage. The results of experiment have shown the possibility to save the 50 % of energy.

Flinn and Satyanarayanan [24] investigated the problem of managing the limited computational resources and battery life in mobile systems taking into account the unstable network communications. As a result, they have developed two systems: Coda and Odyssey, which implement the adaptation at several levels of the system.

Coda realizes a distributed file system that does not require any modifications to existing applications, and Odyssey is responsible for the initiation and management of applications, taking into account the availability of resources and the required quality of the multimedia data.

Now the method of power management is implemented at the level of all modern operating systems, it is based on choosing a power scheme [25] of the available schemes:

– balanced scheme provides the most efficient use of the battery when you need it and saves power when the device is idle;
– power-saving scheme is the best way to maximize the long-term use without recharging the battery due to reducing the performance;
– scheme with high performance, which increases the performance of the system due to reducing the operating time of the battery.

3 Background of the Work

According to the recommendations by Workshop "Energy-Efficient Computing Systems, dynamic adaptation of Quality of Service and approximate computing" to the European Commission on the strategic directions needed for research, development, innovation and ecosystem creation for energy efficient computing systems: "One design point that fits all use cases won't be sufficient to scale down the energy consumption to the required levels. Therefore, design and run-time tools supporting dynamic adaptation and adoption of energy under quality of experience requirements will be needed. The full complexity of dynamic adaptation should not percolate up to the application programming level but be part of the system level. This will require comprehensive energy models for computing" [26].

The concept of adaptability is dominant at the modern theory of energy-saving control of mobile computing system. Actually first of all adaptability required to find a compromise between economical expenditure of residual energy of the battery of the mobile device and information processing speed, the response time or capacity.

It is necessary to solve two interrelated problems: the selection and installation of the operating mode of the processor and essentially planning of the works (processes), taking into account the restrictions for the time of their execution.

The quality of the system's functioning taking into account the restrictions for the execution time is determined by the timeliness of certain tasks, among which we can mark out 3 classes [27]:

– hard real-time, when the exceeding of maximum allowed time is equivalent to failure of the system;
– soft real-time, when the violation of certain time limits leads to a decrease in the quality of the integrated system;
– firm real-time is viewed as a hybrid category between hard and soft real-time. Firm real-time processes are always required to meet a certain amount of deadlines in a given time window, described by a certain number of consecutive invocations [28].

4 Proposed Solution

Power consumption is a process of electricity expenditure of mobile computing device. Since the change of energy consumption caused by changes in loading of processor, RAM, peripherals, thus an uneven load graph is formed. Characteristics of this graph are discussed further in the deterministic formulation.

To summarize the properties of mobile computing power lets examine the equivalent three-level graph of energy consumption, which uses the known of measured value of the minimum, average and maximum energy consumption values of the original graph: Nmin, Nav, Nmax.

Calculation formulas for determining the length of sections with minimum, average and maximum power consumption $\tau_{\min}, \tau_{av}, \tau_{\max}$ follow from the condition of the equivalence of baseline energy consumption and a three-level performance graph (Fig. 1):

$$\tau_{\max} = \frac{\int_{t=0}^{T} \max(0, N(\tau) - N_{av}) \cdot d\tau}{N_{\max} - N_{av}}, \quad \tau_{\min} = \frac{\tau_{\max} \cdot (N_{\max} - N_{av})}{(N_{av} - N_{\min})}, \quad \tau_{av} = T - \tau_{\max} - \tau_{\min}, \tag{2}$$

where T is the duration of the original graph.

Scheduling the tasks taking into account soft real-time constraints. In contrast to the hard real-time systems, where the violation of the time limits is inadmissible in this class of systems the violation of the restrictions is possible, but not desirable, because it leads to a decrease of the service's quality. In a multitasking system, the

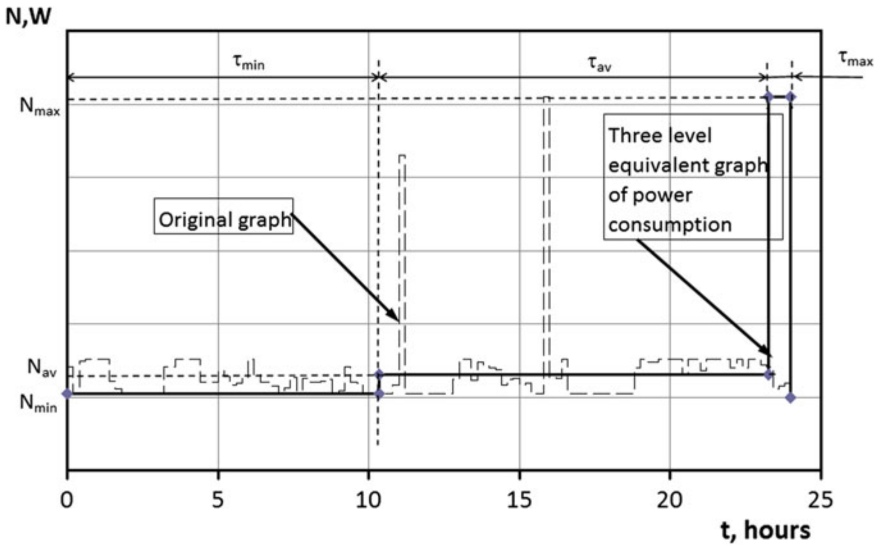


Fig. 1 The replacement of the original graph with random load by the equivalent three-level power consumption graph

mobile computing device can perform many tasks in parallel without time constraints. The tasks with the limitations include the voice telephony, video telephony, video conferencing, real-time games, on demand video etc. To ensure quality of service [29, 30], this system must find a compromise while distributing system resources in that way that all tasks were performed with a certain level of service.

These systems are often overloaded and there is a problem of non-compliance of definite terms for soft real-time tasks. In such cases, different techniques [31, 32] which allow to balance the load system are used.

Suppose, that each task is characterized by five parameters $\tau_i^* = \langle C_i, D_i, To_i, Tm_i, \Delta QoS_i \rangle$:

C_i—run-time;
D_i—the definite date of deadline;
To_i—optimal period of request's formation;
Tm_i—the maximum permissible period of request's formation;
$\boldsymbol{\Delta} QoS_i$—the penalty for loosing the quality of service in violation of the due date of execution per unit of time planning.

Then, taking into account the features of the current task's formulation of compilation scheduling we define the following criteria which are typical for soft real-time systems.

Criterion 1. Minimizing the average of summary penalty for violation of definite terms at the planning interval:

$$S^*(V, f) = \arg\min \sum_{i=1}^{N} \left(\Delta QoS_i \cdot \sum_{j=1}^{T_{sched}/T_i} \upsilon_{i,j}(S(V, f)) \right) \Big/ \lfloor T_{sched}/T_i \rfloor, \quad (3)$$

where—$S^*(V, f)$—optimal schedule, which minimizes the weighted sum of the definite terms violations—$\upsilon_{i,j}(S(V, f))$;

$\lfloor \ldots \rfloor$—rounding down operation;

V, f—voltage and frequency of the processor, arguments of function's power consumption.

Criterion 2. Minimizing the penalty for reducing the length of the processor's work in terms of battery life of mobile computer device multitask system:

$$\langle V^*, f^* \rangle = \arg\min(T_{\max}(V, f) - T_{3\max}(V, f)), \quad (4)$$

where $T_{\max} = \frac{E_{ak}}{N_{\min}}$, $T_{3\max} = \left\lfloor \frac{E_{ak}}{E_{3ak}} \right\rfloor \cdot T_{sched}$—the duration of battery life in both minimal and current loads.

Lowering the voltage of the processor leads to a reduction of penalties for the loss of service time's duration, but it may increase penalties for the loss of service's.

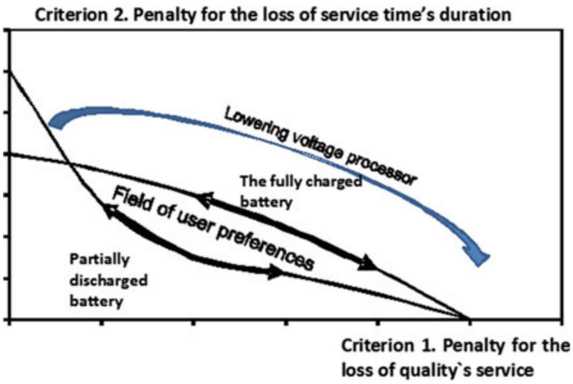


Fig. 2 Pareto-optimal solutions in the criteria space (plane) "penalties for violation of definite terms at the planning interval"—"the penalty for reducing the length of the processor's work in terms of battery life"

At different depths of battery discharge the transformation of the curve preference in the coordinate plane "Criterion 1"—"Criterion 2" happened (Fig. 2).

The area of user's preferences is a subset of Pareto optimal solutions. Its points are acceptable to the user for a combination of parameters "quality of service"—"the duration of battery life with the current power consumption" T_{3max}.

5 Simulation

The experiment was done at the sub-notebook Lenovo IdeaPad S205 (Fusion) 2012. Sub-notebook configuration is as follows:

– CPU: AMD E-450 APU with Radeon(tm) HD Graphics;
– GPU: AMD Radeon HD 6320 M;
– OS: Microsoft Windows 7 Service Pack 1 32-bit 7601;
– memory: 1640 МБ.

During the experiment the effect of the choice of the operating system's power supply plan on indices of performance and power consumption was studied. In Windows 7, the following three power plans are available:

– high performance, when the dynamic scaling of performance are disconnected due to the working load and levels of continuously increase performance are provided instead with a corresponding increase in energy consumption;
– balanced plan, when the dynamic scaling of performance is realize depending upon the requirements of the current working load;
– power saver plan designed to maximize the energy savings and suits well to increase battery life in mobile PCs.

As the loading application PCMark-7 the benchmark developed by Finnish company Futuremark is used, it is testing the stability and productivity of the processor's work, speed characteristics and capacity of operational and permanent memory, and also many other characteristics of computer components [33]. For experimental evaluation of the effectiveness of the OS of Windows LightWeight Test is selected, it contains a collection of loads for measuring the productivity of systems with low productivity, such as personal computers of elementary level, tablets and laptops.

For identification of the statistical model of power plan the Joulemeter—Energy profiler which uses metrics of productivity to assess the energy expended [34], which takes into account the cost CPU, HDD, GPU, network devices and the screen is used. At the basis of calculation Joulemeter the consumption data system hardware resources—CPU, hard drives, memory, display, and others are laid. These data are converted into figures of actual energy consumption, which is achieved by reading the values of some registers that are model-dependent (Model-Specific Registers, MSR)—special registers processor of x86 architecture, the existence and purpose of which varies from model to model of processor.

6 Results and Analysis

At the result of experimental data' processing is revealed, that the transition of PC's power consumption from power saving plan to the plan with high performance leads to:

- an increase by 40 % the energy expended in the calculation of the CPU in a test mode (Fig. 3);
- reduction in the test time by 20 %;
- increase the rate of performance by 35 %. according to the results of the PCMark-7 (Fig. 4).

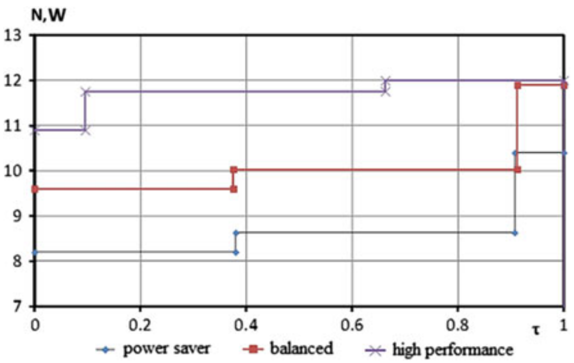


Fig. 3 Equivalent three-level graphics of power consumption for three plans of power

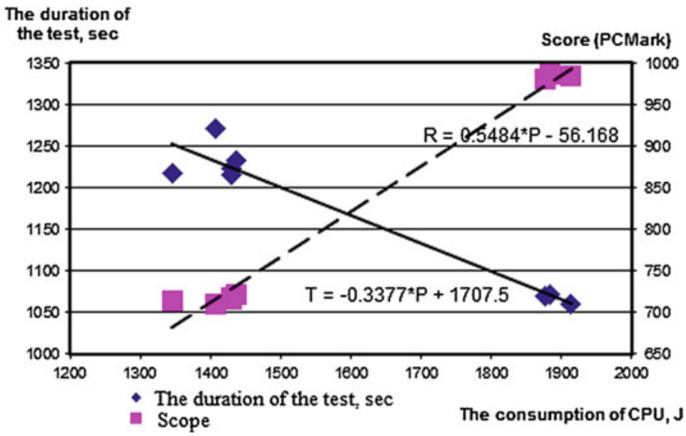


Fig. 4 Experimental data and modeling dependings of time-consuming on the duration of the test (test duration) and performance rating (Scope) of the total consumption of the CPU according to the results of the PCMark-7 and Joulemeter

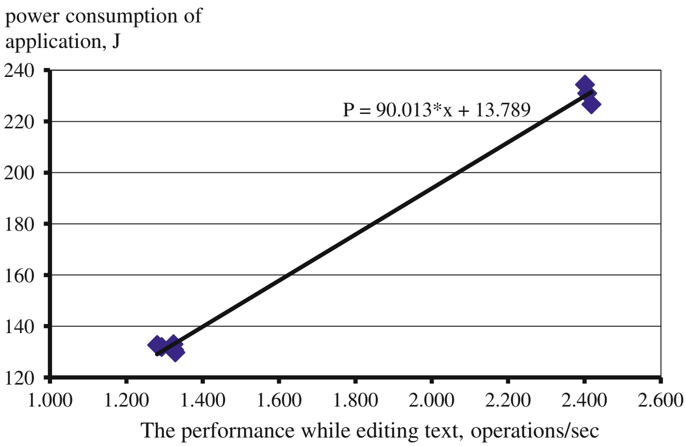


Fig. 5 Dependence of energy costs on the performance while editing text

While estimating the characteristics of individual energy-consuming applications, such as a program for editing the text or Web-surfing the above figures may be much higher: there is a doubling of performance during a doubling of electricity consumption (Fig. 5).

7 Conclusion

Among the known approaches to the power management the main group is a number of approaches based on the DVSF.

The process of power consumption of a PC is presented as an equivalent three level power consumption graph which simplifies the further modelling and analysis. The task scheduling in the real time systems are investigated. The bi-criterial optimization problem with a number of Pareto-optimal solutions is proposed. To the number of criteria we concern minimum penalty for the decline in the quality of service and maximum battery life with the current profile of power consumption. The further narrowing of the Pareto-optimal solutions number is possible on the basis of taking into account the users' preferences.

The results confirm the possibility of a significant increase of energy efficiency, and, consequently, the duration of battery life without access to the modern mobile computing devices due to the usage of adaptive approaches in the management of the processor's modes. Further research would be appropriate in the direction of the development of methods and algorithms of operating modes adaptive control for mobile computing devices.

References

1. Urban, B., Shmakova, V., Lim, B., Roth, K.: Energy consumption of consumer electronics in U.S. homes in 2013. In: Consumer Technology Association. http://www.cta.tech/CorporateSite/media/environment/Energy-Consumption-of-Consumer-Electronics.pdf. (2014). Accessed 26 Dec 2015
2. Carlson, B., Giolma, B.: Smartreflex. power and performance management technologies: reduced power consumption, optimized performance. In: Texas Instruments. http://www.ti.com/lit/wp/swpy015a/swpy015a.pdf. (2008). Accessed 29 Dec 2015
3. Intel.: Enhanced IntelB® SpeedStepB® Technology for the IntelB® PentiumB® M Processor. http://download.intel.com/design/network/papers/30117401.pdf. (2004). Accessed 9 Jan 2016
4. Larabel, M. (2006) Intel EIST SpeedStep - Phoronix. In: Phoronix.com. http://www.phoronix.com/scan.php?page=article&item=397&num=1. Accessed 12 Jan 2016
5. Wikipedia.:Cool'n'Quiet. https://en.wikipedia.org/wiki/Cool%27n%27Quiet. (2004). Accessed 5 Jan 2016
6. Snowdon, D., Ruocco, S., Heiser, G.: Power management and dynamic voltage scaling: myths and facts. In: Proceedings of the National ICT Australia and School of Computer Science and Engineering University of NSW. pp 15–19 (2005)
7. Badam, A., Skiani, E., Chandra, R., et al.: Software defined batteries. In: Proceedings of the 25th Symposium on Operating Systems Principles—SOSP '15 (2015). doi:10.1145/2815400.2815429
8. Mudge, T.: Power: a first-class architectural design constraint. Computer **34**, 52–58 (2001). doi:10.1109/2.917539
9. Pering, T., Burd, T., Brodersen, R. The simulation and evaluation of dynamic voltage scaling algorithms. In: Proceedings of the 1998 international symposium on Low power electronics and design—ISLPED '98. (1998). doi:10.1145/280756.280790

10. Pruhs, K., Uthaisombut, P., Woeginger, G.: Getting the best response for your erg. ACM transactions on algorithms **4**, 1–17 (2008). doi:10.1145/1367064.1367078
11. Albers, S., Fujiwara, H.: Energy-efficient algorithms for flow time minimization. ACM Trans. Algorithms 3:49-es. (2007). doi:10.1145/1290672.1290686
12. Bansal, N., Pruhs, K., Stein, C.: Speed scaling for weighted flow time. SIAM J. Comput. **39**, 1294–1308 (2010). doi:10.1137/08072125x
13. Chan, H., Lam, T., Li, R.: Tradeoff between energy and throughput for online deadline scheduling. Sustain. Comput.: Inform. Syst. **1**, 189–195 (2011). doi:10.1016/j.suscom.2011.05.002
14. Andrew, L., Wierman, A., Tang, A.: Optimal speed scaling under arbitrary power functions. ACM SIGMETRICS Perform. Eval. Rev. **37**, 39 (2009). doi:10.1145/1639562.1639576
15. Irani, S., Shukla, S., Gupta, R.: Online strategies for dynamic power management in systems with multiple power-saving states. TECS **2**, 325–346 (2003). doi:10.1145/860176.860180
16. Irani, S., Shukla, S., Gupta, R.: Algorithms for power savings. ACM Trans. Algorithms 3:41-es (2007). doi:10.1145/1290672.1290678
17. Kaur, T., Chana, I.: Energy Efficiency Techniques in Cloud Computing. CSUR **48**, 1–46 (2015). doi:10.1145/2742488
18. Zeng, H., Ellis, C., Lebeck, A.: Experiences in managing energy with ECOSystem. IEEE Pervasive Comput. **4**, 62–68 (2005). doi:10.1109/mprv.2005.10
19. Zeng, H., Ellis, C., Lebeck, A., Vahdat, A. Currentcy: a unifying abstraction for expressing energy management policies. In: Computer Science and Engineering. http://cseweb.ucsd.edu/~vahdat/papers/usenix03.pdf. Accessed 21 Dec 2015
20. Neugebauer R, Mcauley, D.: Energy is just another resource: energy accounting and energy pricing in the Nemesis OS. In: Proceedings Eighth Workshop on Hot Topics in Operating Systems. (2001). doi:10.1109/hotos.2001.990063
21. Sachs, D., Yuan, W., Hughes, C., et al.: GRACE: a hierarchical adaptation framework for saving energy. 2004–2409
22. Vardhan, V., Yuan, W., Harris, A., et al.: GRACE-2: integrating fine-grained application adaptation with global adaptation for saving energy. IJES **4**, 152 (2009). doi:10.1504/ijes.2009.027939
23. Jeffay, K., Zhang, H.: Readings in multimedia computing and networking. Morgan Kaufmann Publishers, San Francisco (2002)
24. Flinn, J., Satyanarayanan, M.: Managing battery lifetime with energy-aware adaptation. ACM Trans. Comput. Syst. **22**, 137–179 (2004). doi:10.1145/986533.986534
25. windows.microsoft.com.: Conserving battery power—Windows Help. http://windows.microsoft.com/en-us/windows/conserving-battery-power#1TC=windows-7. (2013) Accessed 6 Jan 2016
26. European Commission.: Energy-efficient computing systems, dynamic adaptation of quality of service and approximate computing.pp. 1–44 (2014)
27. Pinedo, M.: Scheduling. (2012). doi:10.1007/978-1-4614-2361-4
28. Kaldewey, T., Lin, C., Brandt, S. Firm real-time processing in an integrated real-time system. In: Work in Progress Session of the 12th IEEE Real-Time and Embedded Technology and Applications Symposium vol. 4 (2006)
29. Cucu-Grosjean, L., Buffet, O.: Global multiprocessor real-time scheduling as a constraint satisfaction problem. In: International Conference on Parallel Processing Workshops. (2009). doi:10.1109/icppw.2009.31
30. Huang, D., Shi, J.: Quality of service scheduling in wireless multimedia communications. In: Proceedings on IEEE VTS 53rd Vehicular Technology Conference, Spring 2001 (Cat No01CH37202). doi:10.1109/vetecs.2001.945092
31. Buttazzo, G., Lipari, G., Abeni, L.: Elastic task model for adaptive rate control. In: Proceedings on 19th IEEE Real-Time Systems Symposium (Cat No98CB36279). doi:10.1109/real.1998.739754

32. Kuo, T., Mok, A.: Load adjustment in adaptive real-time systems. In Proceedings Twelfth Real-Time Systems Symposium. (1991) doi:10.1109/real.1991.160369
33. Futuremark.com.: Test your PC and Android device. http://www.futuremark.com/benchmarks/pcmark. (2015). Accessed 8 Nov 2015
34. Research at Microsoft.: Joulemeter: computational energy measurement and optimization. http://research.microsoft.com/en-us/projects/joulemeter/. (2014). Accessed 3 Dec 2015

Evaluating the RAM Energy Consumption at the Stage of Software Development

D.A. Maevsky, E.J. Maevskaya and E.D. Stetsuyk

Abstract A method of absolute value estimation of the computer energy consumption in performing the programs is proposed in the chapter. The evaluation is made on the basis of the program source code and can help to choose the most optimal solution from the viewpoint of energy saving at the Software development stage. The method is based on the indication of energy consumption by the computer Random Access Memory (RAM) depending upon how intensive the RAM is used by Software. Selection of RAM due to the fact that it is a necessary equipment, and any computing device cannot function without it. Evaluation of energy consumed by RAM is made in using the two proposed mathematical models. The method for computer's estimating power consumption based on assembler source code was made on the basis of the proposed models. This allows you to create a green software with the control of "green" degree on all stages of its development.

Keywords Green software · Energy consumption · Energy-efficient software development · Green computing · Software engineering

1 Introduction

During the past years the efforts of scientists all over the world have been focused on the development of the so-called "green" technology. Its main goal is to reduce the amount of energy the mankind is consuming and, consequently, the GHG emission into the atmosphere. According to data of the International Energy Agency (IEA) if no measures are taken the energy infrastructure of our planet will

D.A. Maevsky (✉) · E.J. Maevskaya · E.D. Stetsuyk
Odessa National Polytechnic University, 1 Boulevard of Shevchenko, Odessa, Ukraine
e-mail: Dmitry.A.Maevsky@gmail.com

E.J. Maevskaya
e-mail: E.J.Maevskaya@gmail.com

E.D. Stetsuyk
e-mail: Lmd29@ukr.net

V. Kharchenko et al. (eds.), *Green IT Engineering: Concepts, Models, Complex Systems Architectures*, Studies in Systems, Decision and Control 74,
DOI 10.1007/978-3-319-44162-7_6

exceed all the permissible limits of GHG emission till the end of 2017. It is expected that after this the global heating process will become irreversible and lead to all ecological systems destruction and worldwide disaster.

1.1 Why Software Should Be Green

Green software engineering so is critical in our drive towards a sustainable development of civilization. This is evidenced by simple calculations. By the data of Ericsson Mobility Report [1], in connection with the development of Internet of Things, around 26 billion connected computing devices are expected by 2020. Each of these devices uses electric energy. It is not easy to estimate the general volume of energy consumed as the energy consumption levels are different for various devices [2]. However, we can estimate the electrical energy saving. Assume that as the result of Software optimization of these devices each of them starts consuming 0.001 W less power. Assume as well that each of the devices operated 20 h a day 360 days a year. Then each of the devices consumes 7.2 Watts per hour (Wh) less.

Taking into consideration the number of devices the general electrical energy saving is 201.6 billion Wh or 201,600 MWh! In order to produce such amount of energy at the thermal power station, 21.7 million cubic meter of gas should be burnt [3]. If to take into consideration the gas price in Europe in 2015 [4], $5,250,000 should be used [5]. Wherein the harmful substances emission in the atmosphere is about 40 thousand tons.

Let us consider these numbers from the other side. Earlier we assumed that the Software optimization had been produced. Assume as well that the optimization has been performed by the team of five programmers during two months. According to the US Bureau of Labor Statistic [6, Occupation code 15-1130] information the average wage of Software Developers and Programmers is $47 an hour or $7500 a month. So the general cost of work of these programmers is $75,000.

Thus in optimizing the Software (pay attention, only once) $75,000 are expended, and $5,250,000 is only saved on gas. We do not consider the damage annually caused to the environment, which cannot be measured with money.

Therefore, the Green Software development, in usage of which the energy is saved, is profitable from the viewpoint of economy and necessary from the one of environment protection.

1.2 Green Software Engineering Today

Green software engineering is a very young field of science. Nowadays it is about ten years old. And comparatively a few papers (about 200) permit to trace the development of this new branch.

The earliest papers, e.g. [7–9] are still strongly connected to hardware and specific devices of the computer, e.g. to Wi-Fi routers. The further works are more and more abstracted from hardware. The authors are trying to trace the main trends in Software development and negative sides of this development [10, 11]. As a negative trend one can distinguish the so-called "Software Bloat", which is characterized with the existence of positive feedback between the hardware capabilities growth and the one of Software volume. E.g. the RAM volume increase leads inevitably to the less economical programs, which use the whole volume of it. Respectively, to further develop a program the larger memory volume is necessary, which stimulates the manufacturers to increase it. And this, in its turn, leads to more wasteful resources usage, and electrical energy, in particular.

At the same time a stage of experimental data accumulation is beginning. The level of energy consumed by various computing devices and their parts are investigated. E.g. in paper [5] a smartphone energy consumption in various modes is described. In paper [12] the unique experimental data on energy level necessary to transfer a cell of different types of RAM from "0" into "1" are represented. These experimental investigations are similar to those ones mentioned further in this section.

In the works published over the last years the attempts have been made to generalize the obtained results and move to forming the mathematical models. Article [13] is particularly devoted to this process, where the following is underlined: "Writing energy-efficient software is one of the most challenging issues in this area, because it requires not only a change of mindset for software developers and designers but also models and tools to measure and reduce the effect of software on the energy consumption of the underlying hardware".

However nowadays the mathematical models of electric energy consumption processes by various computer devices and, in particular, by RAM are absent. In the given section the first attempt of such kind of model construction is represented.

1.3 Purpose of the Study

The amount of electrical energy consumed by the computer is largely dependent on the program compilation quality and Software optimization degree. If a program source code is not optimal it inevitably leads to time increase, RAM overload and eventually to energy overconsumption. The Software having the functionalities similar to its analogue can be supposed more optimal (more "green") if the computer consumes less electrical energy in using such kind of Software. The program code optimality is indicated by relative evaluation in comparing to the code of programs similar as to their functionalities. But the performance of relative evaluation is impossible at the earlier stages of the Software development. We can observe here a kind of a contradiction. On the one hand the existing models and methods are only able to evaluate the Software optimality degree comparing it to the prototypes available. On the other hand, the comparison cannot be made at the earlier stages of development because of the absence of program to be compared to the prototypes. We are able to resolve the contradiction provided the absolute (i.e. in power measurement units) evaluation of energy consumption level is made, but not relative one. So at the earlier stages of Software development as actually the program code optimality is formed the immediate energy consumption measurement is not able to be made because the program is under the process of developing and is not completely operational. It is the program source code (probably still insufficiently debugged) that is available at that stage.

The goal of the present research is to develop the method of estimation of absolute level of energy consumed by RAM in performing the researched program. The estimation can be carried out on the basis of the source code of the program.

In order to achieve the goal, the following has been carried out in the section:

- a mathematical model for determination of the absolute level of power, which is necessary to transfer 1 Bit of RAM from "0" into "1", has been justified and developed. The electric parameters of RAM published in the corresponding guides are the initial data for the model;
- a mathematical model determining the expected value of the amount of bits, which transfer from "0" into "1" in recording the random data of the known length, has been developed;
- a method of indication of absolute level of RAM energy consumption in performing the program on the basis of its source Assembler code has been created on the foundation of these models.

2 Why Was RAM Chosen?

Let us start with the answer to the question—why it was RAM, which was chosen among all the variety of computing device components. At first sight it is not RAM one should begin with altogether if wants to decrease the energy consumption. Even a quick overview of information represented in Internet shows that RAM energy consumption is very small as compared to the other components of PC. Thus, e.g. in [14] the data about maximum and minimum levels of PC components energy consumption are given. They are shown in Table 1.

We can see from the table that the personal computer RAM consumes 20 times less energy than CPU and 40 times less energy than video card on average. The comparison leads to the obvious conclusions that in order to decrease the energy consumption level one should focus on these exact devices. But is it so in reality? Let us try to examine.

Firstly, as we can see from [1] of 26 billion devices, which are going to be connected to Internet in 2020, PC will be only 3 billion, i.e. only 11 %. And what about other 89 %?

The figure includes graphics pads and smart phones (about 9 billion) and the devices Internet of Things (other 17 billion). And the segment of powerful PC with

Table 1 Power consumption of PC components in watts

Component	Min power	Max power	Average
CPU	50	150	100
Motherboard	20	80	50
RAM	3	6	4.5
Video card	25	350	187.5
hard drive	1	9	5
Optical drive	15	30	22.5
Case fan	0.6	6	3.5

powerful video cards is going to decrease in full compliance with [15]. These figures tell that it is the most numerous mobile devices and the devices Internet of Things that will provide the main saving in the field of energy consumption.

Secondly, there are only two devices mentioned in the table without which any computing device cannot operate. They are CPU and RAM. All other devices are not obligatory. E.g. video card is necessary only in the case, when information output to display is intended. But the powerful servers of database and data-center frequently do not possess displays at all [16]! Their software is set up and tuned only once and after this a server operates without switch off and reboot of operation system. A hard disk is not the most necessary component of hardware either. Who saw a smart phone equipped with a hard disk? But as we have already mentioned the obligatory elements of any computing device are CPU and RAM. A processor performs a program but the program is exactly recorded in RAM. Mentioning about CPUs we should distinguish that their modern models possess in the hardware the means for energy consumption control [17, 18]. In order to suitably run with these hardware means a great number of various utilities have been developed [19–23] allowing to reflect power inputs of not only CPU as a whole but also its individual components in the form of figures and curves. But none of the types of RAM possesses such means that is why a direct control of energy consumed by RAM is impossible. For this reason, the development of models and methods of indirect estimation of energy consumption becomes topical.

And at last thirdly. As the experimental results have demonstrated, the energy of level consumed by CPU and RAM do not always differ greatly. In some of the moods of computing devices operation these levels are able to practically coincide. In the paper mentioned above [5] the energy consumption level measurements of smart phone components are described. The measurements have shown that in the standby mood the CPU energy consumption is 12.5 mW, and the RAM one is just 2.5 mW, i.e. five times less. However, in performing such kind of benchmark tests as chess program Crafty or the game Mystery Case Files (also known as MCF), RAM energy consumption can exceed the one of CPU—65 and 45 mW, respectively.

Besides the titles in Table 1 demonstrate that all the values of power are measured in Watts. It is known from the theory of electrical engineering that the active power is measured in watts. But along with the active power there exists the reactive one in electric circuits. This type of power in circuits with currents and voltages of sinusoidal form is conditioned with the shift of phases between them. In the circuits where currents and voltages do not change according to the sinusoidal law the reactive power is conditioned with the fact that the current law change is different from the voltage law change. It is that exact event which takes place in RAM. Reactive power does not do a useful work or emits the heat, but it is consumed from the source of power as well as active power. Acting together active and reactive powers create a total power, which can considerably exceed the active power value. So the active power value alone cannot characterize the general level of device energy consumption. As we will further see in paragraph 4.2 it is that event which takes place in RAM. So by virtue of these reasons we are going to pay the main attention to the theoretical models of energy consumed by RAM.

3 The Mathematical Model of Consumed Power Dependence upon RAM Electrical Parameters

RAM can be easily imagined as a set of a great number of elementary cells each of which stores a state of 1 RAM bit. In Fig. 1 a principle chart of one RAM cell, which consists of one field-effect transistor and one capacitor (1T1C-element) is shown [24].

Cells of such kind of a structure are used in the modern memorable devices. Field-effect transistor in a cell plays the role of a switch, which controls the capacitor charge and discharge. As the transistor turns on the capacitor appears to be connected to the information bus. This mood is used for the three goals:

- for the capacitor charging if a cell is to be moved to the state “1”;
- for reading the capacitor state (in this case the capacitor discharges);
- for refreshing after reading or during the process of storing.

In the situation when the transistor is closed (voltage is absent at the gate) the capacitor is off from the signal bus. However, in this case because of the small value of charge and the presence of leakage current the charge gradually decreases, and in some period of time the element state “0” transits to “1”. Besides the element state

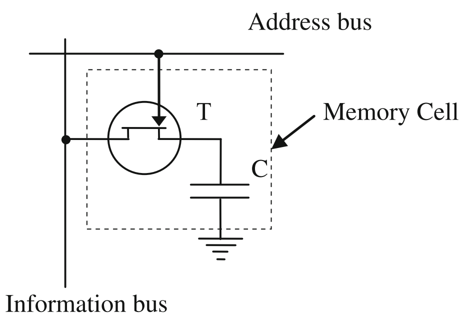


Fig. 1 Basic 1T1C DRAM cell structure

sensing is a destructive operation. Because the capacitor discharges in sensing. That is why the element charge is always to be regenerated. Such kind of process is known as a memory regeneration process. For DDR memory the period of regeneration is 32 or 64 ms depending on the type of memory.

The processes of capacitor charging and discharging are known to take some period of time because its voltage cannot change instantly. That is, the law of the capacitor voltage and current changes and, consequently, the law of change of instantaneous power consumed for the capacitor recharging is to be calculated as a transient process in memory component equivalent. That is, in order to create a model of electrical energy consumption level dependence on RAM electrical parameters one should construct the memory component equivalent and calculate the transient process in it.

To create the model of consumed power dependence upon the electrical RAM parameters we should consider just the process of charging the capacitor. In discharging the capacitor, the energy of supply is not used but refresh is automatically performed and the power consumed in this case does not depend on the Software. As we know the process of charging the capacitor always occurs within some period of time as the voltage in the capacitor cannot change instantaneously. Thus in order to create the model we should calculate the current change law in cell supply circuit as a transient process in its (cell) equivalent circuit. Knowing the supply current and voltage we can calculate the power consumed by a cell in moving from the state "0" to the state "1".

3.1 Calculation of Transient in the Equivalent Circuit of RAM Cell

As we know [24] a potential intermediate between logical "0" level and logical "1" level denoted in standard [25] as V_{REF} is constantly supported in RAM information bus. To make the cell move from the state "0" to the state "1" the control circuit is to perform the following actions:

- Logic 1 Voltage (V_{IH}) is supplied to the information bus;
- Positive potential is supplied to the field-effect transistor Gate and transistor turns on. Thus capacitor C via Source-Drain resistance of the transistor appears to be connected to the information bus. The process of charging the capacitor till the voltage V_{IH} starts.
- On completion charging the positive potential of the gate disappears and field-effect transistor turns off. Thus the capacitor C remains the logical "1" potential—V_{IH}.

The equivalent circuit shown in Fig. 2 corresponds to this process.

The circuit consists of: V_{IH}—logical "1" voltage; V_{REF}—voltage in the information bus; V_{IL}—logical "0" voltage; R_1—information bus resistance; R_2—Source-Drain

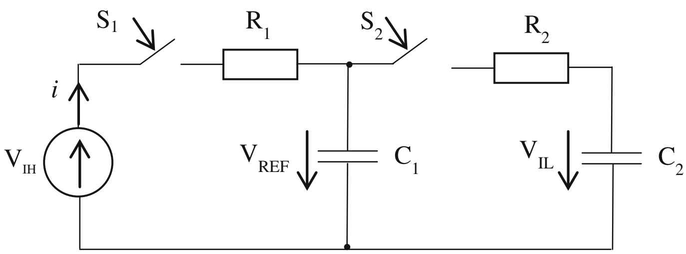


Fig. 2 Equivalent circuit of the basic DRAM cell

resistance of field-effect transistor when it turns on; C_1—information bus capacitance; C_2—capacitance of the capacitor in Basic DRAM Cell.

Switches S_1 and S_2 turn on simultaneously. We should determine the current i change law in the supply circuit. The transient calculation is made with the help of Classical approach. The forced response of current i is absent because voltage V_{IH} is unchangeable in time. Natural response i is to be obtained in the form [26]:

$$i = A_1 e^{\alpha_1 t} + A_2 e^{\alpha_2 t}, \tag{1}$$

because only non-oscillatory transient can arise in the circuit containing two capacitors [24].

Characteristic equation of this circuit obtained with the help of input impedance can be represented in the form:

$$\alpha^2 R_1 R_2 C_1 C_2 + \alpha(R_1 C_2 + R_2 C_2 + R_1 C_1) + 1 = 0,$$

and its roots:

$$\alpha_1 = \frac{-R_1C_2 - R_2C_2 - R_1C_1 + \sqrt{(R_1C_2 + R_2C_2 + R_1C_1)^2 - 4R_1R_2C_1C_2}}{2R_1R_2C_1C_2}, \tag{2}$$

$$\alpha_2 = \frac{-R_1C_2 - R_2C_2 - R_1C_1 - \sqrt{(R_1C_2 + R_2C_2 + R_1C_1)^2 - 4R_1R_2C_1C_2}}{2R_1R_2C_1C_2}. \tag{3}$$

Having determined the initial conditions, we obtain the formula for calculation of integration constants:

$$A_1 = \frac{R_1C_2(V_{REF} - V_{IL}) - (\alpha_2 R_1R_2C_1C_2 + R_2C_2)(V_{IH} - V_{REF})}{R_1\sqrt{(R_1C_2 + R_2C_2 + R_1C_1)^2 - 4R_1R_2C_1C_2}}, \tag{4}$$

$$A_2 = \frac{(\alpha\alpha_1 R_1 R_2 C_1 C_2 + R_2 C_2)(V_{IH} - V_{REF}) - R_1 C_2 (V_{REF} - V_{IL})}{R_1 \sqrt{(R_1 C_2 + R_2 C_2 + R_1 C_1)^2 - 4 R_1 R_2 C_1 C_2}}. \tag{5}$$

So the current i change law in transient is determined with the help of formula (1) where integration constants A_1 and A_2 are calculated according to formulas (4) and (5) and characteristic equation roots α_1 and α_2—according to formulas (2) and (3).

3.2 Calculation of Transient for DRAM4

Let us calculate this current for DDR4 type DRAM. In accordance with [25] it possesses the following electrical parameters: $R_1 = 1\ \Omega$, $R_2 = 2\ \Omega$, $C_1 = 300$ fF, $C_2 = 30$ fF, $V_{REF} = 1.225$ V, $V_{IL} = 1$ V, $V_{IH} = 1.45$ V, $V_{DD} = 1.5$ V.

The curve of transient for current *i(t)* calculated for these parameters is demonstrated in Fig. 3.

As we see in Fig. 3 the current in the source achieve the sufficient values—more than 1.6 amp—at the beginning of the transient process. The transient process itself continues about 2×10^{-12} seconds because of a very small value of capacitor. On the basis of the data obtained we are able to construct the mathematical models in order to determine the active and total powers consumed by the RAM component from the source.

4 Mathematical Models Determining the RAM Cell Power

4.1 Mathematical Model Determining the RAM Cell Complex Power

Taking into account the fact that supply voltage V_{DD} of RAM cell is direct and current i is non-oscillatory we use the complex power $S = V_{DD} \cdot I$ for calculating the power consumed by the supply to move a cell from the state "0" to the state "1". To calculate the complex power, we are to calculate the root mean square of the current i changed according to law:

$$I = \sqrt{\frac{1}{T}\int_0^T i(t)^2 dt} = \sqrt{\frac{1}{T}\int_0^T (A_1 e^{\alpha_1 t} + A_2 e^{\alpha_2 t})^2 dt}.$$

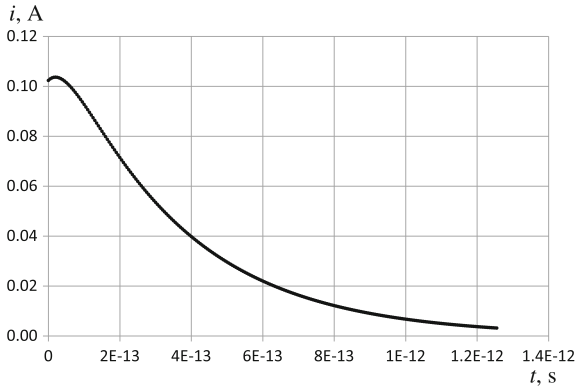


Fig. 3 The curve of current i upon time dependence

In this formula we should use RAM Refresh period as a period T. Let us write down the equation for the sum square, which is under the integral sign:

$$I = \sqrt{\frac{1}{T}\left[\int_0^T A_1^2 e^{2\alpha_1 t} dt + 2\int_0^T A_1 A_2 e^{(\alpha_1+\alpha_2)t} dt + \int_0^T A_2^2 e^{2\alpha_2 t} dt\right]}.$$

On integration and limit substitution we can obtain a mathematical model for calculation of the power consumed by a cell in moving from the state “0” to the state “1”:

$$S = V_{DD} \cdot \sqrt{\frac{1}{T}\left[\frac{A_1^2}{2\alpha_1}\left(e^{2\alpha_1 T} - 1\right) + \frac{2A_1A_2}{\alpha_1+\alpha_2}\left(e^{\alpha_{12} T} - 1\right) + \frac{A_2^2}{2\alpha_2}\left(e^{2\alpha_2 T} - 1\right)\right]}, \tag{6}$$

where $\alpha_{12} = \alpha_1 + \alpha_2$.

If we know electrical RAM parameters formula (6) allows to indicate RAM consumed power for the move from “0” to “1” of just 1 RAM bit. For DD4 type RAM Chips possessing electrical parameters accordance with [25] the calculated power value is 1.89×10^{-4} VA. At first glance the power value seems to be very small. But we should remember that 1.89×10^{-4} VA is consumed as we change the 1 RAM bit state only once. Certainly the move from “0” to “1” for two, three and more bits requires appropriately more power consumption than 1 bit. However, in real cases e.g. in rerecording of 1 RAM bite (8 bits) the number of bites moving from “0” to “1” is unknown. Depending upon the previous and subsequent bite states it can be any number of bits—from zero to eight. This amount is an incidental value.

4.2 Mathematical Model Determining the RAM Cell Active Power

The active power consumed by RAM component can be calculated as a mean value of the instantaneous power within the set period of time *T*:

$$P = \frac{1}{T}\int_{0}^{T} V(t) \cdot i(t) \cdot dt, \tag{7}$$

where *P*—active power, *V(t)* and *i(t)*—instantaneous values of current and voltage, respectively.

If we take into account that the power supply of RAM component is realized by the direct voltage source with the electromotive force *E*, the active value of voltage $V(t)$ remains constant and equal V_{DD} in it. On the basis of this fact the active power *P*, according to (7) which is consumed by the RAM component from the source may be represented as

$$P = \frac{V_{DD}}{T}\int_{0}^{T} i(t) \cdot dt. \tag{8}$$

Current *i(t)*—a current, which is generated in the source in the course of time while the capacitor charging (formula 1). Taking into account the process is repetitive periodically in RAM regenerating we can use the period of regeneration as time *T*.

To construct a mathematical model of active power determination consumed by one of the RAM component in recording the logical “1” or in regenerating the RAM we should use the expression (8), placing the current indicated in (1) in it. We have:

$$P = \frac{V_{DD}}{T}\int_{0}^{T} i(t) \cdot dt = \frac{V_{DD}}{T}\int_{0}^{T} A_1 e^{\alpha_1 t} + \frac{V_{DD}}{T}\int_{0}^{T} A_2 e^{\alpha_2 t}.$$

After integrating and substituting the limits we will obtain the mathematical model for the active power determination:

$$P = \frac{V_{DD}}{T}\left[\frac{A_1}{p_1} \cdot \left(e^{\alpha_1 T} - 1\right) + \frac{A_2}{p_2} \cdot \left(e^{\alpha_2 T} - 1\right)\right] \tag{9}$$

For the same values of electrical parameters of a RAM component, which were used in paragraph 4.1, the active power of the component was calculated. It value is 6.55×10^{-8} W. We can come to the conclusions that a RAM component actually uses exactly the reactive power. That is, one should use the mathematical model of complex power (according to 6) as a working model for determination of energy consumption level dependence on the RAM electrical parameters. Before this model usage it is necessary to calculate the values of characteristic equation roots (2, 3) and ones of integration constants (4, 5).

In order to calculate the RAM consumed power in executing the software we have to develop a mathematical model for determination of expected value of the changed bits amount depending upon RAM area-rerecorded length.

5 The Model Determining the Expected Value of the of the Number of Changed Bits

In the previous paragraphs the dependences, which allow to calculate the power consumed for one RAM component transition from “0” to “1” have been determined. Let us call such kind of power consumption as “a power consumption unit”. It is understandable that the total power consumed by RAM within definite period of time is determined with multiplication of the power consumption units by the total amount of bits, which transit from “0” to “1”.

The number of moves from the “0” state to the “1” state is only determined by the data recorded in RAM. At the stage of forming the model these data are unknown because the number of separate bit moves from the “0” state to the “1” state cannot be accurately calculated. That is why in order to calculate the bit amount we should apply the probability theory. On the basis of the theory we can calculate the expected value of the number of RAM bits, which moves from the “0” state to the “1” state in recording an incidental N bit combination. As an elementary event for probability calculation we take an event when k ($k = 0, 1, 2, \ldots, N$) RAM bits change the state. We use p_k to denote the event probability. Then the expected value of the number of bits, which move from the “0” state to the “1” state in recording N bits, can be calculated according to the formula

$$M = \sum_{k=0}^{N} k \cdot p_k. \tag{10}$$

In the modern programming languages, the smallest RAM usage unit is 1 byte, the next unit as to the value is a word (2 bytes) or a double word (4 bytes). So the expected value calculation is to be made exactly for these length values of RAM area—8, 16 and 32 bits.

The problem of expected value calculation of the amount of RAM bits transiting to “1” in random change of RAM area state is a rather complicated combinatorial task. Its difficulty is conditioned with a large number of transition variants even in the small parts of RAM. For example, in changing the part state of only one bite (8 bits) there are 2^8 possible states of that part. So the transition from one fixed state to other 255 states can occur (Fig. 4).

Thus, the total number of different possible variants of state changes is $256 \times 255 = 65280$. For example, as we can see in Fig. 4, from the byte state

Fig. 4 The example of transition from "0" to "1" of one byte of RAM

0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	1
2	0	0	0	0	0	0	1	0
3	0	0	0	0	0	0	1	1
4	0	0	0	0	0	1	0	0
5	0	0	0	0	0	1	0	1
6	0	0	0	0	0	1	1	0
7	0	0	0	0	0	1	1	1
8	0	0	0	0	1	0	0	0
9	0	0	0	0	1	0	0	1
10	0	0	0	0	1	0	1	0
…	.	.	.	.	.	.	.	.
25	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**1**
…	.	.	.	.	.	.	.	.
250	1	1	1	1	1	0	0	1
251	1	1	1	1	1	0	1	0
252	1	1	1	1	1	0	1	1
253	1	1	1	1	1	1	0	0
254	1	1	1	1	1	1	0	1
255	1	1	1	1	1	1	1	0
256	1	1	1	1	1	1	1	1

corresponding to number 25 (00011001 in binary digit system) the transition to the previous 24 and further 231 states is possible. Wherein, e.g. in transiting from state 25 to state 10 the only one bit is able to transit from "0" to "1" (with number 2), and to state 7—two bits (with numbers 2 and 3). For each of 65 280 possible changes one should calculate the probability of changes from "0" to "1" for one bit, simultaneously for two ones, simultaneously for three ones, four and so on up to eight bits.

For 4 bytes' part (32 bits) the amount of possible transition can be described with a number, which includes twenty digits—18446744069414584320, and besides in each of these transitions we are to trace the probability of one, simultaneously two, three and so on until the simultaneous change of all of 32 bits.

Taking into account a large amount of transition variants the computer program has been used in order to calculate the expected value of the amount of RAM bits, which transit to "1" depending on the part length.

Despite the fact that with the increase of amount of bits in RAM area the amount of possible variants of transitions increases in degree ($2^N \times 2^{N-1}$) dependence, the time, which is necessary to calculate the expected value, is increased in the same dependence. So, on the computer with processor Intel Core I5 the time for calculation when the length of 8 bits RAM area is about one-tenth of a second, and when the length doubles—up to 16 bits—it is increased to two hours. The calculation for the larger RAM area is unreal. E.g. according to the preliminary calculations for RAM of 32 bits length the time of calculation made by the same type of computer is about 900,000 years. That is why when the length achieved 16 bit, the calculations were stopped.

Table 2 Expected value

Length of area (bits)	Expected value
0	0
2	0.666667
3	0.857143
4	1.066667
5	1.290323
6	1.523810
7	1.763780
8	2.007843
9	2.254403
10	2.502444
11	2.751343
12	3.000733
13	3.250397
14	3.500214
15	3.750114
16	4.000061

The results of calculation of expected value dependence of the bits amount, which transit from "0" to "1, on the RAM area length are represented in Table 2.

We should note a very high accuracy of the expected value calculation—8 digits after the comma. The matter is that the table demonstrates the expected values of only one RAM state change that is why the high accuracy is of great importance. But for unit time, e.g. for a second, the amount of changes may be large enough.

At 100 MHz system bus operation frequency (the period is equal to 10 ns) the time of accessing to RAM to change its state is about 200 ns. That is, the amount of state changes is $^1\!/_{200} \times 10^{-9} = 5 \times 10^6$ (5 million) per second. And the amount of changes is a multiplier, by which we are to multiply the obtained expected value in order to calculate the amount of RAM bits, which change their state for a second.

On the basis of data represented in Table 2 as well as the calculation of complex power values (p. 4) we can build a mathematical model of expected value dependence of the RAM state amount changes on the part length. In order to build a model one should firstly identify the dependence of expected value on the bits amount, which transit to "1", on the RAM part length, i.e. to carry out the table data approximation with mathematical equation. As a result, the mathematical model of dependence expected value M(k) on the RAM state amount changes on the RAM part length (k), which varies, has been obtained. The model is represented with formula (11).

$$M(k) = -0.00011791 \cdot k^3 + 0.00444567 \cdot k^2 + 0.19515112 \cdot k + 0.22247721 \tag{11}$$

The model coefficients were calculated up to nine digits after the comma. The accuracy used is conditioned with the necessity to obtain the six correct digits as a final result.

The created mathematical models dependence of power consumption level on the RAM electric parameters, and the dependence of the RAM state amount changes on a part length can be used to develop the method of absolute energy consumption estimation on the basis of a program source code.

6 Method of Evaluation of Energy Consumption in Executing the Programs on the Basis of Their Source Code

The two above mentioned models allow to develop the method of evaluation of energy consumption in executing the programs on the basis of their source code.

The main goal of the method usage is to get an absolute level of computer energy consumption in operating the stated Software. We take the absolute level as the power value, which the computer consumes, expressed in the power units—watts.

As the input data we have got the program source code written in a certain programming language. In analyzing the source code, we must accurately determine both the time, which is spent on program execution, and the number of RAM usages. We should note that in analyzing the source code just a static analysis is possible, i.e. the one, which does not take into account the cycle repetition amount. The amount of repetitions is determined with the help of the input program data, which as a rule are unknown in analyzing the code. That is why in static analysis we cannot even theoretically assess the program execution time.

In order to overcome this disadvantage, the authors have proposed the following approach. On the basis of statistical analysis of the source code we assess the time spent on a single execution of all the program operators. At the same time the number of RAM state changes and RAM area-changed length are determined. Knowing the period of time of a single program execution we are able to calculate the number of such executions within a chosen time unit, e.g. a second.

And if we know the number of program repetitions per second as well as the number of RAM state changes, RAM area-changed length, and using the previously developed models we can calculate the absolute power values consumed by the computer for a second of program operation. Further we will call such kind of power a specific power and denote it with S_0.

Carrying out the analysis for several programs it is possible to choose the most optimal one on the basis of specific powers calculated for the programs. So the proposed approach allows without knowing the number of repetitious cycles but by comparing the specific powers to choose a program (among some programs possessing similar functionalities), which provides the least consumed power in executing it on the given computer.

Since in the process of analysis both the program execution time and RAM usage amount are to be evaluated as accurately as possible we have chosen Assembler as a programming language. The reasons are as follows:

Assembler is the most relevant to the computer instructions, which are executed by CPU. It gives the possibilities to accurately determine the amount of CPU clock signals necessary for each of the program instructions. Knowing a CPU clock signal frequency, we can assess the period of time of single program execution.

In analyzing Assembler instructions, we can accurately determine the length of operands in RAM, which are changed by instruction. Thus we can accurately determine the energy consumption for RAM state change with the help of the developed models.

All the modern Software systems, in particular, Visual Studio for the language C or Delphi for the language Pascal give the possibility to get the developed program source code in the language Assembler.

The described method consists of the following steps.

- To get Assembler code of the researched program.
- To carry out the analysis of each of the instructions of Assembler code and determine the number of clock signals, which are necessary for the instruction execution and RAM area length, which is changed by this instruction. The number of clock signals can be learnt in CPU specifications but the RAM area length corresponds to instruction operands length.
- To calculate the total number of CPU clock signals necessary for a single program execution on the basis of analysis described in previous paragraph. We use *Nt* to denote the calculated number of clock signals.
- To calculate the expected value of the bit amount changed by this instruction on the basis of RAM area length changed by each of the instructions and with the help of mathematical model developed in paragraph 5.
- To calculate the complex power spent in order to change the RAM state by the instruction according to the formula:

$$S_k = M(k) \cdot S_1,$$

where: k—RAM area length changed by the instruction; $M(k)$—expected value of the number of changed from “0” to “1” bits depending on the RAM area length calculated according to the formula (11); S_1—the complex power used in order to change 1 RAM bit state according to model (6).

- To calculate the complex power used in order to change all RAM area states during a single program execution as a sum of complex powers S_k for every separate instruction. We use S_Σ to denote this complex power.
- To calculate a specific power W by the computer for a one second of program execution according to the formula:

$$W = \frac{N_t}{f_t} \cdot S_{\Sigma}, \quad (12)$$

where: W—consumed energy, N_t—the number of CPU clock signals necessary for a single program execution, f_t—frequency of CPU clock signal;

Value W is a final result of the energy consumption evaluation method in executing the programs on the basis of their source code.

7 Case Study. Tools of Energy Estimation

The method of RAM energy consumption estimation described in paragraph 6 has been realized in the form of Software tool—ESTimation Energy Tool (ESTET). The purpose of the Software tool ESTET is to provide a user with a simple, comfortable and intuitively understandable instrument for power estimation consumed by RAM at the stage of Software development. The benchmark data for ESTET is an initial Software text as well as a type of RAM used and a central processor. The Software tool ESTET is realized in Java which allows to transmit it to any computing Platform.

In applying ESTET a user should perform only three steps. In the first step (Fig. 5) a choice of the type of the RAM used and central processor takes place.

The choice among the types kept in the library takes place. If there is no necessary type in the library a user has the opportunity to designate the parameters of RAM or central processor manually.

The second step provides the choice of file with the source code of the investigated program (Fig. 6).

In the current version ESTET the program analysis written in Assembler language of Intel-compatible processors is possible. However, that fact is not a serious limitation if we take into account that all the up-to-date systems of Software development allow to record the Assembler text of the investigated program.

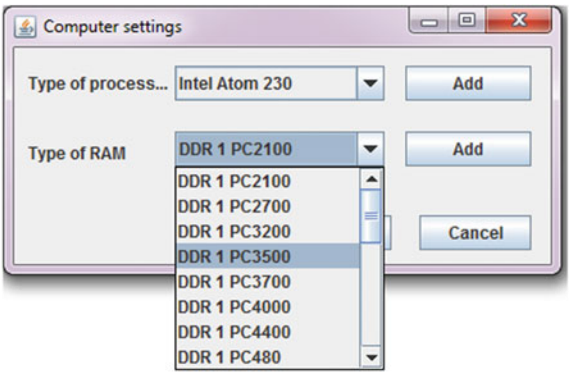


Fig. 5 The choice of the RAM and central processor type

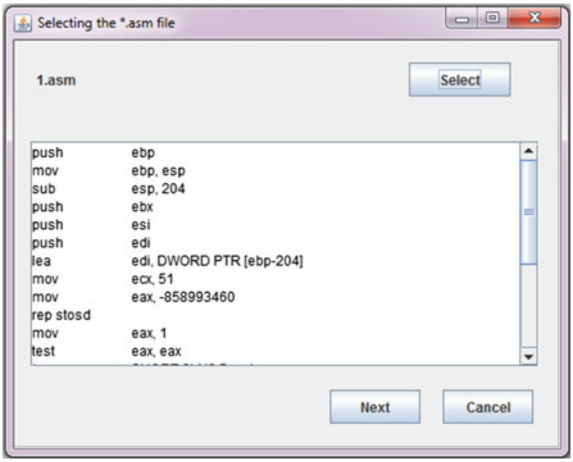
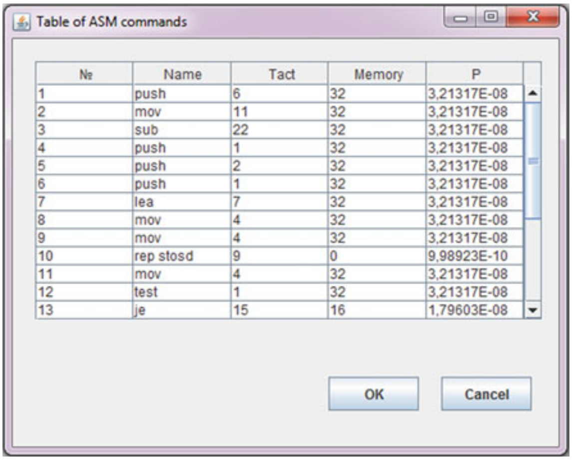


Fig. 6 The choice of the investigated program



№	Name	Tact	Memory	P
1	push	6	32	3,21317E-08
2	mov	11	32	3,21317E-08
3	sub	22	32	3,21317E-08
4	push	1	32	3,21317E-08
5	push	2	32	3,21317E-08
6	push	1	32	3,21317E-08
7	lea	7	32	3,21317E-08
8	mov	4	32	3,21317E-08
9	mov	4	32	3,21317E-08
10	rep stosd	9	0	9,98923E-10
11	mov	4	32	3,21317E-08
12	test	1	32	3,21317E-08
13	je	15	16	1,79603E-08

Fig. 7 Presentation of ESTET operation results

In the third step of ESTET operation a static analysis of the program source code and energy estimation consumed by RAM in performing it is realized. The results for each of the instructions separately and for the investigated program as a whole are represented in Fig. 7.

In this window a source Assembler code broken into instructions and represented in the form of a table is shown. For each of the instructions the amount of clock cycles, which is spent for its fulfillment as well as the amount of RAM bits changed by the instruction is shown in the table. In the last column of the table the total power estimation consumed by RAM in performing the instruction is represented.

The results of program operation are the amount of energy consumed by RAM (Watt on sec) are represented under the table.

Thus with the help of ESTET the Software developers possessing only the program source code can estimate the energy costs on its implementation. This gives the possibility to control and minimize the energy consumption level at the stage of Software development. Having several variants of one and the same algorithm the developers are able to choose the variant, which is the most optimal from the viewpoint of energy consumption.

8 Conclusions

8.1 Discussion of Outcomes

The models and method proposed in the paper are the first attempt to get an absolute value of power consumed by the computer on the basis of program source code. The power consumed only by the computer RAM is calculated. There are some reasons for this. Firstly, RAM along with CPU mounted on Motherboard is a necessary installation, and computer functioning is impossible without it. So RAM is included in and used by Software in any computer system.

Secondly, the usage intensity of RAM by Software indicates its energy consumption. The more frequently the program uses RAM the more electric energy the RAM consumes. The paper demonstrates that energy is consumed in moving the RAM bit from the “0” state to the “1” state.

In the third, RAM chips do not possess embedded means of energy consumption measurement unlike CPU chips. It makes using the indirect methods of evaluation of the energy consumption. The method proposed in the given paper can be referred to such kind of methods.

8.2 Future Research

Further experiment is to be continued in the two directions. First of all, the power values obtained on the basis of mathematical models are to be verified and proved in experimental way.

In performing the experimental verification of the power consumed by RAM one should take into consideration three substantial factors.

Firstly, we are to maintain the clarity of an experiment. In measuring one should make sure that RAM is only used by a test program at the moment and no other processes are carried out. The full guarantee of absence of irrelevant processes can be received by running the test program without operation system. It is quite possible but requires the complicated low level programming in order to create its own program-loader, which puts the program into RAM and gives the control to it. More simple way to solve the problem is to perform the test program under the control of the earlier single-task versions of operation system MS-DOS but not Windows. The system is simple, and program processes, which are run simultaneously with the test program fulfillment, are absent in it.

Secondly, one should be sure the measured power is consumed by RAM exactly. The physical changes on the motherboard of the computer, which are connected to the RAM power supply bus break and inclusion of the measuring device in the break, are necessary.

Thirdly, one should understand what power exactly is to be measured - active or total one. The standard power meter (wattmeter) cannot be used for measurement. All wattmeters only guarantees the correct readings in the case, when both voltage and current either remain unchanged or changed according to the sinusoidal law. However, as we see (Fig. 3) the current used by the RAM cell is neither constant nor sinusoidal. In this case in order to provide the accurate measurement we are to use two measuring devices—voltmeter and ammeter. Both of these devices are to send the measured values to the computer in real time. The voltmeter is to be connected in parallel to the RAM power supply clips and the ammeter—to the power supply bus break. According to the measured values of voltage and current (formulas 6 and 7) using the numerical integration we are to calculate the values of total and active power.

References

1. On the pulse of the networking society Ericsson mobility report. http://www.ericsson.com/res/docs/2015/mobility-report/ericsson-mobility-report-nov-2015.pdf (2015)
2. Mann, M., Spath, P.: The net CO_2 emissions and energy balance of biomass and coal-fired power systems. In: Fourth Biomass Conference of Americas, Oakland, California, 29 Aug–2 Sept 1999
3. Honorio, L., Bartaire, J-G., Bauerschmidt, R.: Efficiency in electricity generation. EURELECTRIC Report. http://www.eurelectric.org/Download/Download.aspx?DocumentID=13549 (2003)
4. Carroll, A., Gernot, H.: An analysis of power consumption in a smartphone. In: USENIX Annual Technical Conference (USENIXATC'10), pp. 21–21. Boston, USA, 23–25 June 2010
5. Bolla, R., Bruschi, R.: Performance and power consumption modeling for green cots software router. In: Communication Systems and Networks Workshops, pp. 1–8. Bangalore, India, 5–10 Jan 2009. doi:10.1109/COMSNETS.2009.4808883
6. Ikebuchi, D., Seki, N., Kojima, Y.: Geyser-1: A MIPS R3000 CPU core with fine grain runtime power gating. In: Solid-State Circuits Conference (A-SSCC 2009), pp. 369–370. Taipei, Taiwan, 16–18 Nov 2009. doi:10.1109/ASSCC.2009.5357257
6. Ciocoiu, N., Ciolac, C.: Automated framework for green IT classification using software agents. In: 2nd IEEE International Conference Computer on Science and Information Technology (ICCSIT 2009), pp. 279–283. Beijing, China, 8–11 August 2009. doi:10.1109/ICCSIT.2009.5234398
7. Xue, X., Qingshan, Z., Jun, M.: Development of an evaluation system software for manufacturing green product based on B/S&C/S mode combined. In: 4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '08), pp. 1–5. Dalian, China, 12–17 Oct 2008. doi:10.1109/WiCom.2008.2803
8. Bhattacharya, S., Gopinath, K., Rajamani, K.: Software bloat and wasted joules: is modularity a hurdle to green software? Computer **44**(9), 97–101 (2011). doi:10.1109/MC.2011.293

9. Shenoy, S., Eeratta, R.: Green software development model: an approach towards sustainable software development. India Conference (INDICON). pp. 1–6. Hyderabad, India, 16–18 Dec 2011. doi:10.1109/INDCON.2011.6139638
10. Harsha, M.: Energy comparison of cypress F-RAM and EEPROM. Cypress Semiconductor Corp. http://www.cypress.com/file/46746/download (2014)
11. Ardito, L., Procaccianti, G., Torchiano, M.: Understanding green software development: a conceptual framework. IT Prof. **17**(1), 44–50 (2015). doi:10.1109/MITP.2015.16
12. Power consumption of PC components in watts http://www.buildcomputers.net/power-consumption-of-pc-components.html (2012)
13. Drozd, J., Drozd, A.: Models, methods and means as resources for solving challenges in co-design and testing of computer systems and their components. In: Ninth International Conference on Digital Technologies (DT'2013), pp. 176–180. Zhilina, Slovakia (2013). doi:10.1109/DT.2013.6566307
14. Yoo, S.J.: Intra- and inter-data center networking architectures for future cloud computing. In: Proceedings of Communications and Photonics Conference (ACP-2012). Guangzhou, China, 7–10 November 2012. doi:10.1364/ACP.2012.AS4A.1
15. Binu, P.: Effectiveness of SPEC CPU2006 and multimedia applications on intel's single, dual and quad core processors. Dissertation. University of Texas, San Antonio (2009)
16. Kataoka, H., Duolikun, D., Enokido, T.: Power consumption and computation models of a server with a multi-core CPU and experiments. In: IEEE 29th International Conference on Advanced Information Networking and Applications Workshops (WAINA). Gwangju, Korea, 25–27 Mar 2015. (2015). doi:10.1109/WAINA.2015.127
17. Gurumurthi, S., Sivasubramaniam, A., Irwin, M.: Using complete machine simulation for software power estimation: the SoftWatt approach. In: Eighth International Symposium on High-Performance Computer Architecture, p. 141. Cambridge, 2–6 Feb 2002. doi:10.1109/HPCA.2002.995705
19. Jason, F., Satyanarayanan, M.: PowerScope: a tool for profiling the energy usage of mobile applications. In: Proceedings of the Second IEEE Workshop on Mobile Computer Systems and Applications (WMCSA '99). Washington, USA, 25–26 Feb 1999
20. Lange, K., Tricker, M.: The design and development of the server efficiency rating tool (SERT). In: Proceedings of the 2nd ACM/SPEC International Conference on Performance engineering (ICPE'11). Karlsruhe, Germany, 14–16 Mar 2011. doi:10.1145/1958746.1958769
21. Zhang, D., Thao, N.: RPM: a remote computer power management tool. Int. J. Comput. Appl. Technol. **22**(4), 161–171 (2005). doi:10.1504/IJCAT.2005.006954
22. Wang, D.T.: Modern DRAM Memory Systems: Performance Analysis and Scheduling Algorithm. Dissertation, 247 p. University of Maryland at College Park (2005)
23. Double Data Rate (DDR) SDRAM JESD79F. In: JEDEC Standard. http://cs.ecs.baylor.edu/~maurer/CSI5338/JEDEC79R2.pdf (2008)
24. Shenkman A.: Transient Analysis of Electric Power Circuits Handbook, 543 p. Springer, Netherlands (2005)
25. European Union Natural Gas Import Price YCHARTS. https://ycharts.com/indicators/europe_natural_gas_price. (2015)
26. National Industry-Specific Occupational Employment and Wage Estimates. NAICS 541500—Computer Systems Design and Related Services. http://www.bls.gov/oes/current/naics4_541500.htm#15-0000 (2014)

Part III
Green Internet Computing, Cloud and Communication Systems

Impact of the Internet Resources Structure on Energy Consumption While Searching for Information

Volodymyr Dubovoi and Oleksii Moskvin

Abstract The article presents a new model of the effect of the Internet resources structure impact on the energy efficiency of information search. Existing studies have revealed that the number of Internet queries is growing exponentially. The execution of each one consumes energy. The article also present the state of the art of search engines energy consumption, characteristics of hypertext systems and search engines. Besides, a model of the relationship between the hypertext characteristics and the number of information search steps is developed. For this purpose, the impact of hypertext structure on the steps number on searching for relevant and pertinent information; and the impact of the steps number on energy consumption were studied. As a result, approaches for optimization of hypertext structure in the conditions of uncertainty were formulated. A simulation model allowed testing the adequacy of the developed model of the effect of the structure of distributed hypertext systems on the energy efficiency of information search. To this end, a hypertext model was generated as a random hypergraph. The results can be used to create automated systems for hypertext systems optimization.

Keywords Internet resources structure · Information search · Energy consumption · Hypertext characteristics

1 Introduction

1.1 Motivation

The objectives of energy-saving technologies for information processing in the world are becoming increasingly relevant within the complex scientific and practical direction generally called "Green IT". In recent years, search engines have become one of the most common information procedures to find information on the

V. Dubovoi (✉) · O. Moskvin
Vinnytsia National Technical University, 95 Khmelnitske Shoes, Vinnytsia, Ukraine
e-mail: v.m.dubovoy@gmail.com

V. Kharchenko et al. (eds.), *Green IT Engineering: Concepts, Models, Complex Systems Architectures*, Studies in Systems, Decision and Control 74,
DOI 10.1007/978-3-319-44162-7_7

Internet. According to statistical data, the number of queries to search engines has increased exponentially [1]. The growth rate of the queries number has increased even more with the widespread use of mobile devices. Mobile devices with ever more powerful processors, GPS-sensors, and other tools "independently" send information, on the basis of which search engines offer contextually and spatially dependent information and advertising. This fact explains outpacing projected growth in Internet traffic from mobile devices about 1.6 times per year [2]. According to recent reports, currently, about 10 billion devices are connected to the network, which exceeds the Earth population.

According to available data [3], due to the high speed of query processing, a typical search engine produces 7 g of CO_2 to execute a single query. For example, Google search engine for an average query returns results in less than 0.2 s consuming about 1 kJ. It should be noted that the research results presented below confirm the somewhat higher time expenditure and substantially lower energy consumption [4].

The Internet resources structure as a hypertext system significantly influences the number of queries and the complexity of the search process. The discrepancy between the data link structure and semantic structure of information leads to search steps number increase and accordingly to queries number growth.

1.2 Analysis of Related Works

The task of energy consumption assessment and reduction when searching for information on the Internet is considered from different perspectives depending on stages and participants of the browsing process. Figure 1 shows the principal participants in the search process.

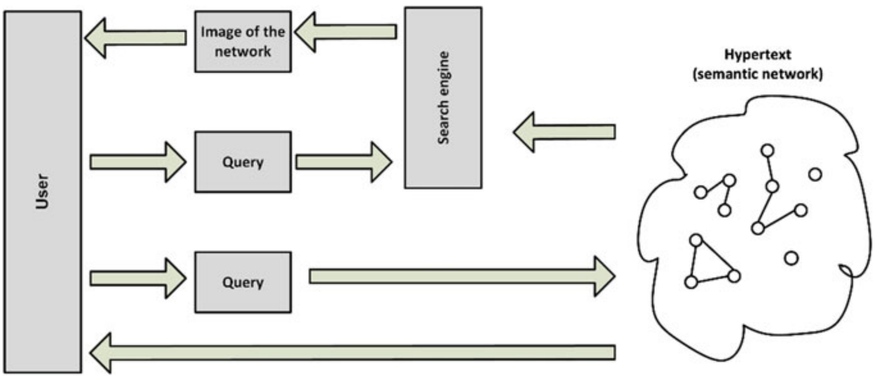


Fig. 1 Participants in the search process

The exemplary sequence of steps to find information is as follows:

- *A user* generates a query and sends it to a *search engine*;
- *The search engine* analyzes the indexed information and on its basis provides to *a user* a simplified *image* corresponding to a fragment of the *semantic network* of the hypertext;
- *The user* analyzes the image and decides whether to refine the search query or to request one of the *hypertext* pages;
- Further, depending on the received results the user decides whether to stop browsing or to go back to one of the previous stages.

Energy is spent on each stage. The existing studies focus on the analysis and minimization of energy consumption by the search engine and the transfer system transferring queries and hypertext pages. Catena [4] studied the creation of energy-saving software for search engines. The article [5] discusses the development of energy-efficient caching mechanisms in the search engine. The article [4] experimentally proved that the dynamic energy consumption of the search engine linearly depends on the search time with 5 J growth on 0.2 s of processing time. The amount of energy consumption is significantly less than that specified in the work [3], which may be a result of improvement of Google energy efficiency in recent years.

Nowadays the massive proliferation of mobile devices has led to an increase in the number of studies concerning energy consumption in wireless data transmission. Balasubramanian et al. [6] show that the energy consumption increases proportionally to the exponent of the logarithm of the transmitted data amount, that is in linear fashion. Hence, it follows, that it is necessary to reduce the number of transmitted pages, while maintaining the search quality.

The hypertext as the basis of the web network, used for browsing information on the Internet, reflects the web-page authors' views about the semantic structure of the information. The set of concepts, descriptions of their properties and the relationships between them is known as ontology. The semantic structure (semantic network) is the description of the relationships between concepts, usually presented in the form of a graph.

Analysis of the ontology is a difficult task of significant dimension, as it involves processing of large amount of data and logic statements contained in ontologies [7]. Systems based on the analysis of the activities of web resources users are used in analytical and reference systems, the primary task of which is mapping of statistical information of their visits. Google Analytics is an example of such system [8].

The search method, duplicates presence, keywords ambiguity, non-informative blocks presence on pages are factors that among others influence the relevance of information provided to a user and the number of user search steps.

The method of shingles [9] allows determining duplication of web pages. The main idea of this method consists in the breakdown of texts and comparison of selected sequences of words from the text (shingles), for each of which a checksum is calculated.

Search, based on information units, is based on the partition of web pages into information blocks, which are considered information independent units, in which the search is carried out. Lin and Ho [10] developed the InfoDiscover system, which divides a web-page into blocks. The work of [11, 12] proposed the VIPS method for splitting pages into blocks and a model for their evaluation [13].

The number of the user's steps in the search process depends on the search quality. To assess the quality of search results [9] the following characteristics are suggested: precision *SP;* recall *SR*; fall-out (probability of finding irrelevant resource) *SF*, calculated from experimental data; R_{rel} is the set of relevant resources in the network; R_{retr} is the set of resources found by the system; R_{nrel} is the set of irrelevant resources. However, in practice, such estimates are difficult to calculate, since the value R_{rel} cannot be accurately determined.

The work of [14, 15] proposed another method to assess the search results. They used Google PageRank (PR), an algorithm for calculating the web-pages credibility. Methods to improve PageRank, mainly using methods of hyperlinks classification and filtration, are studied in [16, 17].

There are several studies concerning the relation between search quality and hypertext structure. The work of [18, 19] suggests the use of a particular hypertext metric, which allows evaluating the hypertext system with formal indicators, based on the representation of hypertext with a graph $G(V, E)$:

Index of information compactness

$$Cp = \frac{ST_{\max} - PG}{ST_{\max} - ST_{\min}}, \tag{1}$$

where $ST_{\max}$—is the greatest possible number of steps that must be done on the links connecting all hypertext nodes; $ST_{\min}$—is the minimum possible number of steps, which connect all hypertext nodes (when all nodes are interconnected); *PG*—is the index of graph paths.

A complete digraph has the shortest distance $ST_{\min} = n_v^2 - n_v$, a totally disconnect graph should have the greatest distance $ST_{\max} = C \cdot (n_v^2 - n_v)$, where C is the maximum value of the distance matrix; n_v is the number of vertices.

Missing paths ratio

$$K_m = \frac{PG_r}{n_v^2 - n_v}, \tag{2}$$

where PG_r is the number of missing paths in the graph. The maximum number of missing paths is $(n_v^2 - n_v)$, and the minimum value is 0. Thus, the value of the missing paths ratio varies in the range [0; 1].

The cyclomatic number describes the difference of the graph structure from the treelike form

$$Cn = n_E - n_V + C_C, \tag{3}$$

where n_E is the number of edges, n_V is the number of vertices, C_C is the number of connected components of the graph. The cyclomatic number indicates the smallest number of edges that must be removed, so that it turns into a tree. The cyclomatic number determination allows finding the number of independent paths in the hypertext structure.

The index of stratification characterizes the linearity of hypertext. In fact, the stratification index allows estimating the connectivity level of elements, which belong to different levels of the hierarchy.

The absolute stratification is

$$AS = \sum_i |s_i| = \sum_i |a_i - b_i|, \tag{4}$$

where $d(u, v)$ is the distance between the vertices u and v in the digraph G; a_i is the sum of distances $d(v_i, u)$ from vertices v_i to all u in G. Thus, a_i is the sum of final values of the ith row of the distance matrix $M(G)$; b_j is the sum of distances $d(u, v_j)$ from all vertices u to v_j. Thus, b_j is the sum of final values of the jth column of the distance matrix $M(G)$.

Node centrality is intended to identify the index page of hypertext. Determination of centrality depends on the choice of links direction of the page (outgoing or incoming) of interest:

- The indicator of the relative external centrality is

$$ROC_i = CD/COD_i \tag{5}$$

- The indicator of the relative internal centrality is

$$RIC_i = CD/CID_i,$$

where *COD*, the outgoing transformed distance from the vertex *i*, is the sum of all elements of the row of the distance matrix $COD_i = \sum_j d_{ij}$; *CID,* the converted incoming distance to the vertex *i*, is the sum of all elements of a column of the distance matrix $CID_i = \sum_j d_{ji}$.

The central vertex is the element, for which the distance to all the other vertices is minimal. Consequently, the smaller the value of *COD is*, the more central the vertex is; the greater the value of *ROC is*, the greater the level of centrality of the vertex is.

The variability and uncertainty of structure (pages characteristics, their quantity and relations, quality of search) interfere with the successful model creation of the hypertext optimal structure. Currently, the problem of optimal structure of Internet resources in terms of information search energy efficiency under uncertainty of processes in the Internet has not yet been solved.

1.3 Goals and Structure

The aim of the article is to develop a model of the effect of the Internet resources structure on the information search energy efficiency.

This objective is achieved by solving three tasks:

1. Analyzing the effect of the steps number on energy consumption, when the user searches for relevant and sufficient information;
2. Development of a model of the relationship between the hypertext characteristics and the number of search step;
3. Development a method to optimize the hypertext structure to reduce energy consumption in the condition of uncertainty.

2 Methods

Energy consumption when performing a query consists of three components: consumption when a user analyzes the relevance and pertinence of the results, search engine consumption and consumption in data transfer. Search engine energy consumption depends on the complexity of the query in the given context and on the user's search steps number. Consumption in data transferring depends on the search steps number, and on the volume of information transmitted at each step. Energy consumption due to user's activities depends on the number of search steps, the volume of transmitted information and the speed of its analysis by the user.

The fewer "additional" search steps the user makes, in the process of obtaining and analyzing the hypertext data, and the less redundant information sent to him is, the less the energy consumption on searching for relevant information is. Thus, a connection is established between the characteristics of hypertext structure and the consumption of energy on searching for information.

In the construction of the model of the relationship between the characteristics of hypertext and the number of search steps the following characteristics of hypertext structure were considered:

1. Index of information compactness Cp.
2. The ratio of missing paths Km.
3. Cyclomatic number Cn.
4. Index of stratification AS.
5. Node centrality ROC.

Besides structural characteristics, the particularities of the browsed web-pages influence the user's search steps number. Let's use such indicators for assessment of web-pages:

1. The amount of relevant information on the page I_{rel}.
2. The amount of information expected by the user I_T.
3. Index of complexity of the search query CQ.
4. The average amount of transmitted information in the performance of a single query I_Q.

We consider the following fundamental systematic prerequisites of the model construction:

1. User's temperament, which is expressed by the average number of browsed links p_e on the page, also influences on the statistics of visits.
2. The amount of relevant information on the page I_{rel} is proportional to the number of keywords in the query k_{si}, located on the page (considering synonyms), and to the number of links between them.
3. The probable amount of relevant information I_{rel} on the page depends on the number of keywords homonyms k_{om}.
4. The amount of relevant information I_{rel} on the page at a given number of keywords and keywords homonyms k_{om} is normally distributed.

The complexity of search queries is determined with relation to the information semantic structure. For evaluating the query, in the context of the analysis of its impact on the search energy characteristics, we suggest measuring the compactness of the graph, which is obtained from the application of the query operator on the semantic network.

The query complexity directly affects the user's steps number, when searching for information. It is assumed, that the user must take a few steps in almost every element of the compactness of the search results set for the analysis of the relevance and adequacy of the search. The number of browsed elements of compactness and the steps number in each element depend on the user's temperament and motivation. The dependence of the number of steps and the transferred information volume can be approximated by exponential dependence with saturation for each user.

The optimal browse path according to the criterion of minimizing the steps number required to produce a given amount of information I_T corresponds to the "greedy" algorithm. However, the user cannot use such algorithm due to two factors:

1. There are no hyperlinks on the path of the greedy algorithm;
2. Having some references on the web-page, the user does not know which of them conducts on the path with the greatest weight.

Obviously, for reducing the impact of the 1st factor is necessary to increase the of links number, seeking to get a complete connect hypergraph; and for reducing the effect of the 2nd factor is needed to decrease the links number to increase the probability the user to select the optimal path.

3 Solution

3.1 Relations of the Concepts

The hypergraph $G(W, H)$ represents the hypertext and the subgraph W represents the web-pages $W\{c, h\}$. The hypergraph vertices represent the keywords and concepts of the semantic subnetwork. In the following presentation, the interrelated concepts will be used, the notations of which are given in Table 1.

Like any graph, connections in this hypergraph are described by the matrix of adjacency. If the connections E_P within a page adequately reflect the semantic structure (as the user builds it himself in his mind during the reading of the page), the connections E_G between pages are set "manually" via hyperlinks and may differ from the semantic structure.

We will characterize subgraphs of the hypergraph with the amount of information I_p on the page. The amount of information is defined by the amount of concepts of the semantic network and communications between them $I_p = i \cdot ln(|\{V, E\}|)$. For our model, we will consider conditional information in relation to a search query $I_{p/Q}$.

3.2 Search for Information Model

The number of information search paths browsed by the user depends on the complexity of the query CQ. The complexity of the search query depends on the of keywords number and other conditions, which leads to a p-complete problem. Assuming that the query has no mistakes, which the search engine will try to correct, the query structural complexity will be

$$CQ_S \sim (N_k)^L,$$

where L is the number of operations in the query.

The keywords number determines the size of the semantic core, the number of basic vertices of the semantic network. The set of search results depends on the popularity of the keyword (the frequency of use v_0 or the cardinality of the set of web-pages with these keywords.

Table 1 The interrelated concepts

Notation	Related concepts			
G	Hypergraph	World wide web	Semantic web	Hypertext
P	Compact subgraph	web-page	Compact sub-network	Page
E	Edge	Hyperlink	Relationship	Keyword
V	Vertex	Object	Concept	Fragment

Table 2 Example of search results for the query with logical operators

(w) 4, (+) 0, (−) 0	(w) 4, (+) 1, (−) 0	(w) 4, (+) 0, (−) 1	(w) 4, (+) 1, (−) 1	(w) 4, (+) 1, (−) 2
Titanic crash iceberg video 173,000 results (0.36 s)	+Titanic crash iceberg video 119 results (0.50 s)	−Titanic crash iceberg video 543,000 results (0.36 s)	+Titanic crash iceberg −video 50 results (0.37 s)	−Titanic crash iceberg −video 543,000 results (0.28 s)
	Titanic +crash iceberg video 39 results (0.39 s)	Titanic −crash iceberg video 463,000 results (0.30 s)	Titanic +crash iceberg −video 6 results (0.36 s)	−Titanic −crash iceberg +video 9180 results (0.29 s)
	Titanic crash +iceberg video 27 results (0.43 s)	Titanic crash −iceberg video 543,000 results (0.33 s)	Titanic crash +iceberg−video 6 results (0.28 s)	−Titanic crash −iceberg +video 268,000 results (0.31 s)

Table 3 Example of search results for the query with different of keywords number

(w) 0	(w) 1	(w) 2	(w) 3	(w) 4
qqsdvhu 2 (0.43 сек.)	Titanic 87,800,000 (0.62 s)	Titanic crash 2,730,000 (0.48 s)	Titanic crash iceberg 241,000 (0.58 s)	Titanic crash iceberg video 173,000 results (0.36 s)
	Crash 388,000,000 (0.57 s)	Titanic iceberg 670,000 (0.58 s)	Titanic crash video 7,310,000 (0.69 s)	
	Iceberg 220,000,000 (0.86 s)	Crash iceberg 645,000 (0.48 s)	Titanic iceberg crash 194,000 (0,52 s)	
	Video 8,170,000,000 (0.45 s)	Iceberg video 19,200,000 (0.45 s)	Iceberg crash video 517,000 (0.50 s)	

The search engine algorithm has a fuzzy nature, in particular, it takes into account the search history and other factors. Tables 2 and 3 show an example of the numerical characteristics of the search results for the query “*Titanic crash iceberg video*” with logical operators “+” (the word must be present in the result), and “-” (the word should not appear in the results).

For simplicity, we can assume that the set of search results is determined by the cover of the intersection of the set of compact subgraphs (hypertext pages) containing the query keywords as subgraphs vertices or their complements (if there is a word before the operation "−") and arcs satisfying the logical operations of the query.

$$R_{retr} : G(v \in R_{retr}, E) = \bigcup_{\substack{i=1 \\ j=1 \\ i \neq j}}^{N_k} G_1\left[L_i(w_i), L_{ij}\left(E_{ij}\right)\right],$$

where w_i is the query keyword; L_i are the logic operations specifying the presence or absence ith keyword in the results; L_{ij} are the logical operations clarifying the relationship between the keywords.

The search results number can be estimated assuming that the intersection of subgraphs are equiprobable.

$$|R_{retr}| \approx \min_{i \in [1, N_k]} \frac{v_i}{2e^{k_s\left(L_i + L_{ij}\right)}}, \tag{6}$$

where k_s are coefficients.

If the cyclomatic number satisfies the condition $Cn = 0$, then for every vertex of the core, the of outgoings number from the vertices edges m_{si} are determined by the position of the related keywords in the frequency dictionary of the subject domain.

$$m_{si} = k_{sa} v_i, \tag{7}$$

where v_i is the frequency of use of the word; k_{sa} is a coefficient depending on the cardinality of the set of domain concepts, $k_{sa} = n$.

Thus, the average search paths number browsed by the user is

$$S = \left[\prod_{i=1}^{N_k} (k_{sa} v_i \cdot ROC_i)\right]^{p_e} \tag{8}$$

The scheme of information search paths at $ROC_i = 0$ is shown in Fig. 2.

Let's define the concept of information search path weight

$$W = \sum_{s=1}^{p_e} I_{relsi},$$

where I_{relsi} is the amount of relevant information on the sth page of the search path. Modern search engines sort search results in decreasing order of relevance (if the of keywords number $N_k > 1$). However, on the pages, the links are not sorted. Therefore, we can assume the following:

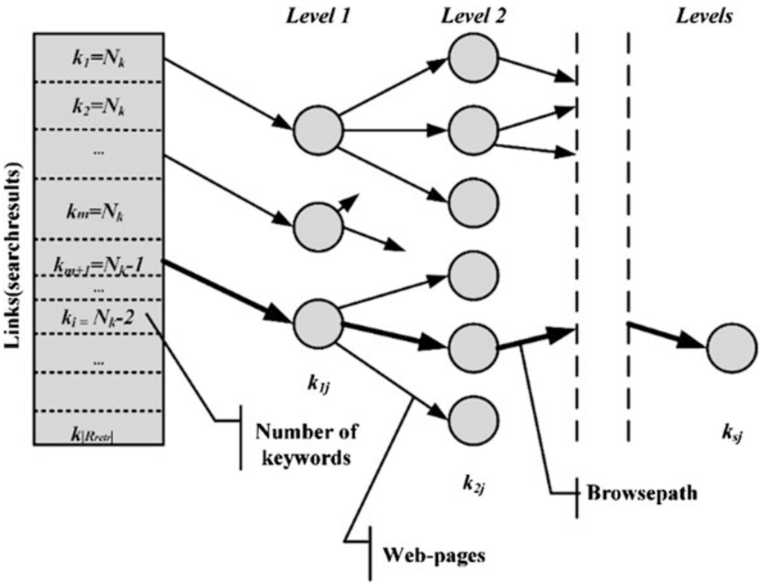


Fig. 2 The scheme of information search paths

- If the average of homonyms number of each keyword is k_{om}, one can expect a of keywords combinations number equal to $(k_{om})^{N_k}$ on the index page. From them only one combination corresponds to the search target.
- The probability of obtaining relevant information in the first page (after the index page) of the selected search path decreases as the user moves through the list of search results and the coincident keywords number decreases.
- Let us estimate the keywords number of query with conditions

$$\left| \begin{array}{ll} k_i = N_k & i = 1 \\ k_i = 1 & i = |R_{retr}| \end{array} \right.,$$

which implies $k_i = round\left[N_k \cdot e^{-(i-1)}\right]$.

Thus, the average probability of usefulness of a given link (that is, the fact that this link may serve as the beginning of the browse path)

$$P_u(k_i) = \frac{k_i}{n} \Big/ (k_{om})^{N_k}, \tag{9}$$

where i is the of links number in the search list.

- If the average relevant information quantity on the page is proportional to the relevant keywords number, then after p_e steps on the links of the index page the user receives $I_{\mathrm{rel}} = \sum_{i=1}^{p_e}[P_u(k_i) \cdot k_i \cdot m_I]$ information, where m_I is the average amount of information, corresponding to one keyword. Then on average from one search page from the links of the index page the user gets a volume of information $I_0 = {}^{I_{\mathrm{rel}}}\!/_{p_e}$.

- If $I_{\mathrm{rel}} < I_{\mathrm{T}}$, then the user continues browsing links from relevant pages, moving away from the index page. It is possible to expect that the subsequent pages still belong to the relevant subject domain. However, as the distance from the first page increases, these pages will have less search keywords number. Accordingly, the amount of relevant information will be reduced. If there are $k_{si} = k_i e^{-\alpha(s-1)}$ relevant keywords on the page, where $1 \le s \le p_e$ is the steps number along the browse path, beginning at the index page, the weight of the path will be

$$W_i = \sum_{j=2}^{s} m_I k_{si} \tag{10}$$

Then the total amount of browsing information in the search page is

$$I = I_\Sigma + \sum_{i=1}^{u} W_i, \text{ and } u : I \le I_T \tag{11}$$

Thus, the total user's steps number is

$$N = p_e + \sum_{i=1}^{u} s_i \tag{12}$$

3.3 Browsing Redundancy

While browsing the current page of the search path, the user selects a link for the next step more or less randomly.

In addition to the useful information, redundant information is also provided to the user, the transmission of which requires extra energy and which contains redundant links in the context of the query. Let the average page size be I_p. If for on average for one link (keyword) there is an amount of information m_I, it is possible to expect, that there will be an average $m_p = int\left[{}^{I_p}\!/_{m_I}\right]$ links number on the page, and only k_{si} of them will be relevant, where k_{si} is keywords number initial request on step s of path from the link i of index page of search results. Browsing the $m_{red\,i}$ remaining links leads to an excessive steps number

$$m_{red\,si} = \left| \begin{array}{ll} m_p - min[p_e, k_{si}] & \text{if } m_p > min[p_e, k_{si}] \\ 0 & \text{if } m_p \leq min[p_e, k_{si}] \end{array} \right. \tag{13}$$

and receiving redundant data (with redundant information based on the start page of the browse path)

$$I_{red\,si} = \left| \begin{array}{ll} (I_p - m_I) + I_p \cdot m_{red\,si} & \text{if } I_p > m_I \text{ or } m_p > 1 \\ I_p \cdot m_{red\,si} & \text{if } I_p \leq m_I \text{ or } m_p \leq 1 \end{array} \right. \tag{14}$$

and relevant data

$$I_{rel\,si} = \left| \begin{array}{ll} m_I \cdot \frac{p_e}{m_p} \cdot k_{si} & \text{if } m_p \geq p_e (\text{T.e.} \frac{I_p}{m_p} \geq p_e) \\ m_I \cdot k_{si} & \text{if } m_p < p_e \end{array} \right. \tag{15}$$

Let us estimate the query keywords number in the search engine results, depending on the number of the result.

Let the number of results, which contain all the N_k keywords is m, which characterizes the popularity of the subject. Then $k_{i \in [1,m]} = N_k$.

In the following search results, the keywords number of the query decreases, but increases the number of keywords combinations, i.e.

$$k_{i \in [a_j, b_j]} = N_k - j,$$

where $j \in [1, N_k - 1]$; $a_j = 1 + m \sum_{l=0}^{j-1} \binom{N_k - l}{N_k}$; $b_j = a_j + m \binom{j}{N_k} + 1$.

3.4 Energy Consumption for Information Search

The search engine energy consumption in performing a query is proportional to the complexity of the semantic web and the complexity of the query.

The energy consumption when browsing information links of the search page is proportional to the browsing steps number and the amount of required information.

Let us estimate ***search engine energy consumption on searching for information***.

As indicated above, energy consumption is proportional to the processing time of the query. In its turn, the processing time is proportional to the complexity of the query and the resources number that the search engine finds relevant.

To assess the time complexity of the query execution is more difficult because modern search engines can perform a significant amount of work on correcting grammatical errors, making recommendations based on artificial intelligence to optimize the query. Let us assume that query intellectual processing time is proportional to the structural complexity. Another component of the time complexity depends on the found resources number. Since search engines search uses a tree index system, the time complexity of this component is proportional to $ln|R_{retr}|$. Thus, the time complexity is

$$CQ_T = a_{QT} \cdot ln|R_{retr}| + b_{QT}, \tag{16}$$

where a_{QT}, b_{QT} are coefficients.

Accordingly, the energy consumption for the search is

$$\begin{aligned} E_S = k_{ES} CQ_T &= k_{ES}[a_{QT} \cdot ln|R_{retr}| + b_{QT}] \\ &= k_{ES}\left[a_{QT} \cdot \min_{i \in [1,N_k]} \frac{v_i}{2e^{k_s(L_i + L_{ij})}} + b_{QT}\right], \end{aligned} \tag{17}$$

where k_{ES} is the specific energy consumption.

The energy consumption to transmit information is proportional to information volume. If a user searches for as long as it gets the relevant information I_T then from (16) to (17) we can write:

$$\left| \begin{array}{l} I_{rel} = \sum\limits_{\substack{i=1}}^{\substack{min[p_e, R_{retr}] \\ I_{rel} \leq I_T}} \left\{ \sum\limits_{s=1}^{min[p_e, R_{retr}]} I_{rel\,si} \right\} \\ I_{red} = \sum\limits_{\substack{i=1}}^{\substack{min[p_e, R_{retr}] \\ I_{rel} \leq I_T}} \left\{ \sum\limits_{s=1}^{min[p_e, R_{retr}]} I_{red\,si} \right\} \end{array} \right. \tag{18}$$

The total amount of transmitted information is

$$I_0 = I_{rel} + I_{red} \tag{19}$$

Accordingly, the cost of energy to transmit this amount of information is

$$E_T = k_{ET} I_0, \tag{20}$$

where k_{ET} is the specific energy consumption.

Energy consumption due to user's analysis of information is also proportional to information volume. This consumption is related to the necessary time for the user to work with the computer.

$$E_A = k_{EA} I_0, \tag{21}$$

where k_{EA} is the specific energy consumption.

The total amount of energy consumption

$$E_0(I_p, m_p) = E_S + E_A + E_T \tag{22}$$

Energy consumption estimation algorithm is shown in Fig. 3.

3.5 Experimental Investigations and Analysis of the Model

For modeling the energy consumption, it is necessary to estimate the algorithm parameters defined in Fig. 3. For this purpose experimental studies, involving 74 students of the "Systems theory and systems analysis" course, were carried out. In the experiment, students were asked to find information about individual topics to write essays of given dimensions. A set of keywords was formed from the topics. Google search engine was used for the experiment because this system has the largest quantity of statistical data published. In the process of searching for information, students saved all browsed pages that allowed them to make a semantic network for each topic.

The coefficients of time complexity a_{QT} and b_{RT} calculated from the data displayed by Google in the search results. An example of typical results is shown in Fig. 4, and the results in Tables 2 and 3. The experimental results are shown in Table 4.

We estimated the coefficients a_{QT}, b_{QT}, k_{ET}, k_{ES} k_{EA} based on the data on [4, 6], and the data on Notebook power consumption, used in the analysis of search results: $E_S \in [0, 50J]$, $E_T \in [20, 100J]$, $E_A \in [0, 250J]$.

Thus,

$$k_{ES} b_{QT} = 7.5,\ k_{ES} a_{QT} = \frac{50J}{10^9} = 0.05 \times 10^{-6},$$
$$k_{ET} = \frac{100J}{10^8} = 1 \times 10^{-6},\ k_{EA} = \frac{250J}{10^8} = 2.5 \times 10^{-6}$$

3.6 Analysis of Simulation Results

The resulting model and simulation algorithm allow investigating the dependence of energy consumption on the page size and the links number on them. These characteristics are particularly useful for changes in the process of optimization of the structure of hypertext system. These characteristics are shown in Fig. 5.

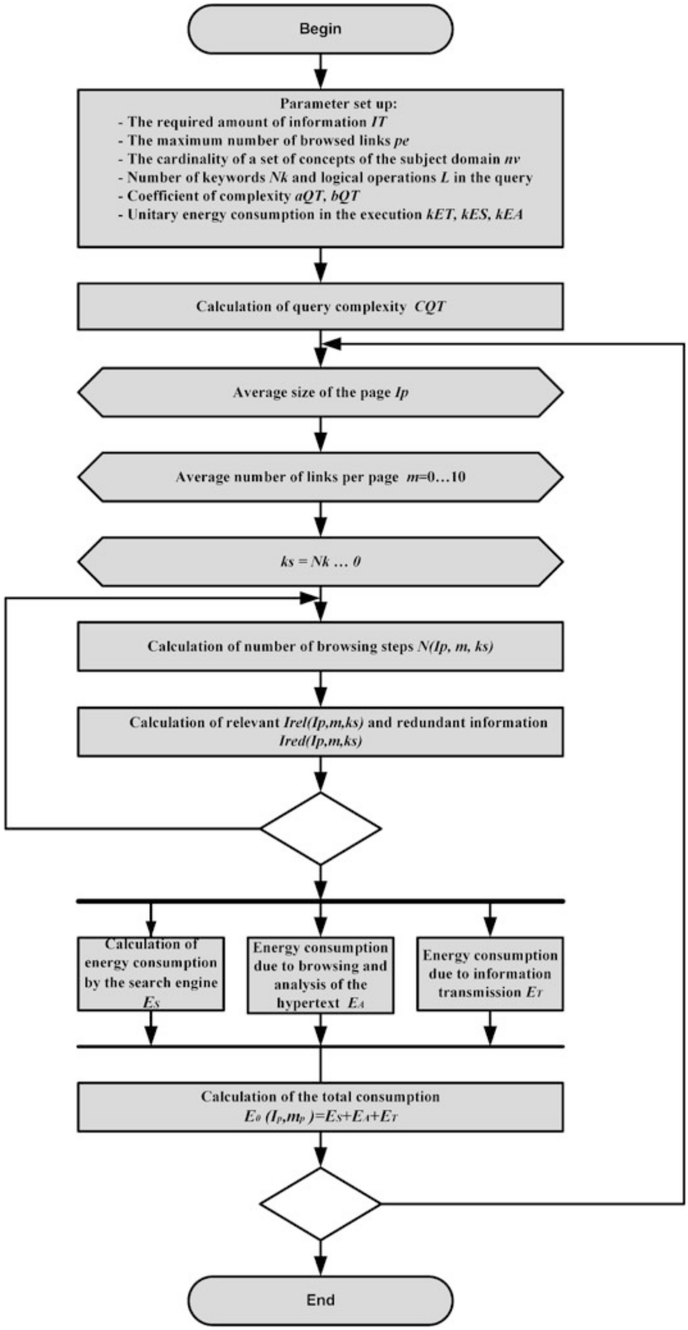


Fig. 3 Energy consumption estimation algorithm

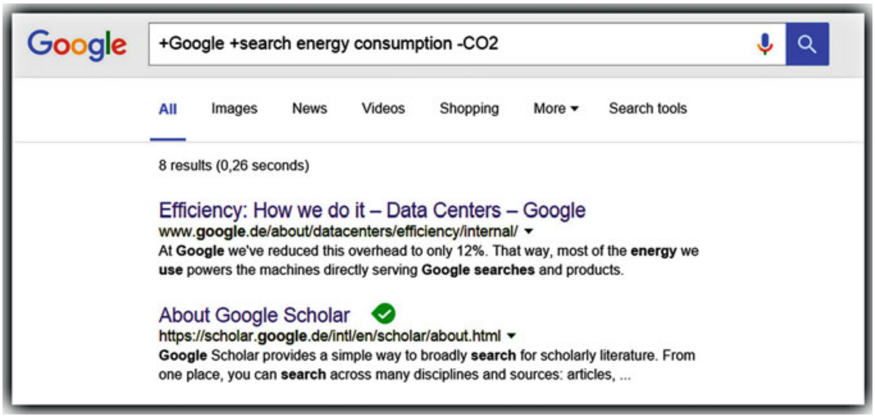


Fig. 4 Typical Google search results

Table 4 Statistical results of experimental search

Parameter	Minimum value	Maximum value	Average value	Dispersion
p_e	0	7	3	1.36
n_V	3	27	22	16
a_{QT}	0.009	0.013	0.011	0.4×10^{-6}
b_{QT}	0.28	0.42	0.33	0.22
$m\ (N_k = 3, L = 0)$	80×10^4	9×10^6	4×10^6	2×10^{12}
$m\ (N_k = 3, L = 1)$	79	149	120	81
$m\ (N_k = 4, L = 2)$	6	112	80	19
$m\ (N_k = 5, L = 4)$	3	8	6	4
I_p	10^6	8×10^6		
m_p	0	26		

Follows from the analysis of the figures that energy consumption are minimum when the links number on the page is $m_p \approx 4$.

3.7 *Optimization of the Hypertext Semantic Structure*

The simulation showed that optimization of both the structure of web-pages (in the first place, their size) and the structure of the connections between them (hyperlinks) is necessary for the increase of efficiency of information search and the decrease in energy consumption. This problem differs from the usual SEO (Search Engine Optimization), because the focus is not on the search engine, but on the Internet resources.

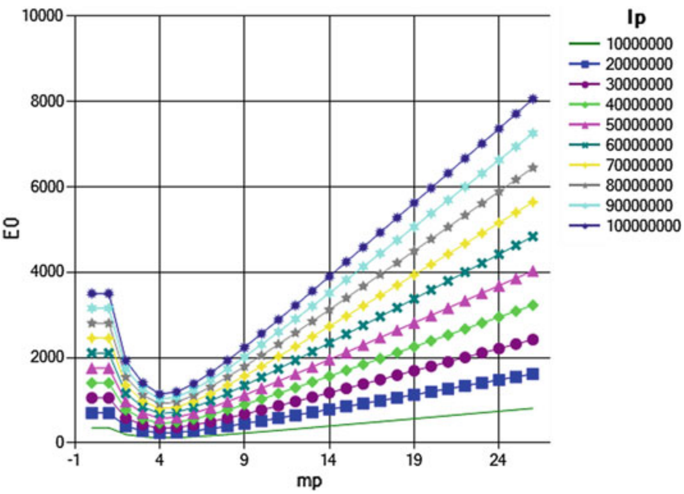


Fig. 5 Dependence of energy consumption on the page size and the links number

As separate segments of the hypertext are located on different uncoordinated servers change, the multi-agent approach is suggested for the solution of the problem of optimization [20]. The functions of the interacting agents on the servers should include the analysis of users' actions and the refinement of the model parameters. The latter cannot be set on preliminary statistical analysis and are not part of the optimization problem. Based on the statistical analysis of users' activities in every segment of the network the following is carried out: gathering of information about the structure of the controlled and semantically adjacent network segments, calculation of hypertext structure indicators, optimization of the link structure and formation of recommendations to the administrator segment.

The evaluation of the model parameters takes into account the main characteristics of user behavior in the hypertext. The main assumptions in the method of estimating the model parameters are the following:

- If the user opens a page, but actually does not browse it, testifies about the irrelevance of the page.
- The relevance depends on the level of centrality of the page (if the page is an indicator, i.e. with a high value of coefficient ROC, the browse path begins from it, and the page is considered relevant).
- The basis of assessment is the nature of the user's actions in the selection of pages, corresponding to the objectives of his information search, in which he spends more or less time browsing some pages, returning to certain pages several times and stopping the search if it becomes too long.

The process of optimization consists of two parts: optimization on the basis of semantic metadata and optimization on the basis of the analysis of the activity of users. Then the optimization results are generalized and provided to the system administrator in the form of recommendations for structure modification of the Internet resource.

The optimization problem has a limitation: changes in the hyperlinks number should provide good reachability of pages. Reachability of pages is based on the reachability matrix of the hypergraph [21]. Considering the mentioned restrictions caused by temperament of the user,

$$\forall d_{1j} \leq p_e,$$

where 1 corresponds to an index page, and j to the last page browse path, $d_{ij} \in M(G)$, $M(G)$ is the distance matrix, and then the reachability matrix is

$$M_{ach}(G) = \left[M_{adj}(G)\right]^{p_e},$$

where M_{adj} is the adjacency matrix of the hypergraph.

We assume the reachability is reliable if at least two paths from the index page lead to every website, and after the removal of any vertex of the hypergraph (unavailability of any page) the paths number to other pages becomes less than 1, i.e.

$$\forall\left[a_{1j} \in M_{adj}(G)\right] \geq 2 \text{ and } \forall\left[a_{1j} \in M_{adj}(G \backslash \exists v\right] \geq 1$$

The algorithm for optimization of the structure of Internet resources includes the following steps:

1. *Determination the hypergraph connectivity, information compactness index, stratification, and ratio of missing paths*.
2. *IF hypergraph is disconnected THEN execution of semantic analysis and formulation of semantic networks of all disconnected elements*.
3. *Combining semantic networks of hypertext resources belonging to the optimized segment of the network*.
4. *Construction of a given semantic network hypergraph*.
5. *SEARCH for an optimal hypertext structure WHILE energy consumption is reduced*:

   5.1. *Calculation of reachability matrix*.
   5.2. *Evaluation of energy consumption on searching the information*.
   5.3. *Remove edges (links) between the vertices with ROC* < 3 *and* $m_S > 3$.
   5.4. *Adding edges (links) between the vertices with c ROC* < 3 *and* $a_{ij} < 2$.

6. *End*.

The adequacy of the model of information search and optimization algorithm was tested on a simulation model. For this purpose, the hypertext model was generated in the form of a random hypergraph. The connected subgraphs number is n = 100, and in each subgraph, there are from 1 to 10 vertices. Each vertex is set in correspondence with a random amount of information I_W and relevant $I_{rel} \in [0, I_w]$ with probability $f(I_{rel}) = \frac{1}{I_{W1}(i)} e^{-\frac{I_{rel}}{I_{W1}(i)}}$, where $I_{W1}(i) = I_W e^{-\frac{i}{|R_{retr}|}}$.

Agents are implemented as objects of modeling software. From the software of agents only the algorithm of optimization was programmed. Each agent is associated with one of the central nodes and receives information about the structure of the hypergraph (the nodes number and adjacency matrix) from the agents related to the nodes, the distance to which is $d_{ij} = 1$. In the presence of disconnected subgraphs, a pair of vertices having a semantic link are randomly selected (the use of the known algorithm of the semantic analysis was supposed).

The analysis of the data confirmed the correctness of the optimization algorithm under given simulation conditions. Experimental system [22] had analyzed user activity and steps number while searching for information. The estimation of energy consumption was done on such base show that the optimization allowed reduce energy consumption up to 20 % in some cases. The largest decrease in energy consumption was obtained on the consumption due to the analysis of information. Optimization of the hypertext structure has no effect on search engine energy consumption.

4 Conclusions

In real conditions, the reduction in energy consumption appears to be somewhat lower, because search engines provide increased relevance of information by intelligent processing. Besides, during the work of a multi agent system there are additional flows of information between agents increasing energy consumption on communication. However, this component of energy consumption is not significant. Since communication between agents is required only when changing the structure of the hypertext.

The task of developing a model of the impact of the structure of Internet resources in the energy efficiency of information search is relevant in the context of the general problem of reducing energy consumption.

The developed model of the effect of the structure of Internet resources on energy efficiency allows formulating a multi-agent optimization algorithm. The use of the algorithm can provide a significant reduction in energy consumption on searching for information, but its solutions are of advisory nature for administrators and owners of Internet resources. In this regard, the application of modeling and optimization of structure of Internet resources encounters a number of organizational obstacles, the main thing being the unwillingness of many administrators to spend time on the solution of the general problem of energy efficiency.

5 Direction of the Future Research

Future work will continue in this direction, for instance in evaluating the influence of web-structure uncertainty on optimizing process. We also believe there is still space to further improve the multi-agent technology to investigate and optimize of web-structure.

References

1. Tom Daly: Dyn,http://dyn.com/blog/evaluating-the-growth-of-internet-traffic/ (2011)
2. Julien, D.: Retour sur les décisions, les projets et les polémiques de Mozilla des dernières années, http://linuxfr.org/users/julien-d/journaux/retour-sur-les-decisions-les-projets-et-les-polemiques-de-mozilla-des-dernieres-annees (2015)
3. Hölzle, U.: Powering a Google search. Official Blog, https://googleblog.blogspot.com/2009/01/powering-google-search.html
4. Catena, M.: Energy efficiency in web search engines. In: Proceedings of the 6th Symposium on Future Directions in Information, Published by BCS Learning and Development Ltd. doi: http://dx.doi.org/10.14236/ewic/FDIA2015.1 (2015)
5. Catena, M., Tonellotto, N.: A study on query energy consumption in web search engines, http://ceur-ws.org/Vol-1404/paper_20.pdf (2013)
6. Balasubramanian, N., Balasubramanian, A., Venkataramani, A.: Energy consumption in mobile phones: a measurement study and implications for network applications. University of Massachusetts Amherst, http://people.cs.umass.edu/~arun/papers/TailEnder.pdf (2009)
7. Berger, P.: How to evaluate websites for better or worse. Inf. Searcher **16**(2), 1–10 (2006)
8. Clifton, B.: Advanced Web Metrics with Google Analytics, 384 p. Wiley (2008) ISBN: 978-0-470-25312-0
9. Broder, A.Z., Glassman, S., Manasse, M.S., Zweig, G.: Syntactic clustering of the web. SRC Technical Note #1997-015, http://www.std.org/~msm/common/clustering.html
10. Lin, S.-H., Ho, J.-M.: Discovering informative content blocks from web documents. In: The Proceedings of eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'02), pp. 588–593(2002)
11. Cai, D., Yu, S., Wen, J.-R., et al.: Block-based Web Search. SIGIR'04, pp. 456–463. Sheffield, South Yorkshire, UK (2004)
12. Song, R., Liu, H., Wen, J.-R. et al.: Learning block importance models for web pages. In: The Proceedings of the 13th international conference on World Wide Web, pp. 203–211. New York, NY USA (2004)
13. Chakrabarti, D., Kumar, R., Punera, K.: A graph-theoretic approach to webpage segmentation. In: The Proceeding of the 17th International Conference on World Wide Web, pp. 377–386 (2008)
14. Page, L., Brin, S., Motwani, R., et al.: The PageRank citation ranking: bringing order to the web, 17 p. Stanford, Stanford InfoLab (1999)
15. Berkhin, P.: A survey on PageRank computing. Internet Math. **2**, 73–120 (2004)
16. Saito, K., Nakano, R.: Improving convergence performance of PageRank computation based on step-length calculation approach. Lecture Notes in Computer Science, vol. 4252, pp. 945–952 (2006)
17. Dubovoy, V., Moskvin, O.: Development of the hypertext structures fuzzy classification system. Scientific Works of Vinnytsia National Technical University, no. 1, http://works.vntu.edu.ua/index.php/works/article/view/146 (2008)

18. Botafogo, R.A., Rivlin, E., Shneiderman, B.: Identifying hierarchies and useful metrics. ACM Trans. Inf. Syst. (TOIS) (2), 142–180 1629 (1992). ISSN 1042-1629
19. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American Magazine, http://www.sciam.com/article.cfm?id=00048144-10D2-1C70-84A9809EC588EF21
20. Glon, O., Dubovoy, V., Moskvin, O.: Site structure optimization in conditions of incomplete information. Scientific Works of Vinnytsia National Technical University, no. 1, http://works.vntu.edu.ua/index.php/works/article/view/48 (2008)
21. Harary, F., Norman, R., Cartwright, D.: Structural models. An Introduction to the Theory of Directed Graphs, 415 p. Wiley, New York (1965)
22. Moskvin, O.M., Sailarbek, S., Gromaszek, K.: User behavioral model in hypertext environment. In: Proceedings SPIE 9816, Optical Fibers and Their Applications, 98161. doi:10.1117/12.2229137 (2015)
23. Sazoglu, F.B., Cambazoglu, B.B., Ozcan, R., et al.: A financial cost metric for result caching. In: Proceedings of SIGIR, ACM, pp. 873–876 (2013)

Introducing Controlling Features in Cloud Environment by Using SNMP

Asif Iqbal, Colin Pattinson and Ah-Lian Kor

Abstract Simple Network Management Protocol (SNMP) is an application-layer protocol that is used to monitor IP based devices. Also it can monitor and manage IP based devices without impacting their performance. SNMP is also used in a cloud computing environment to monitor and control virtual machines. However, currently, there is limited published research work on the deployment of SNMP for monitoring and controlling virtual machines in such type of environment. This would allow load balancing in cloud environment during peak hours leading to reduced power consumption helping in Green IT cause. This paper discusses the deployment SNMP for monitoring and controlling Type 1 hypervisor (in a cloud environment). This is followed by the customization of MIB and net-snmp with Agent X to provide more SNMP management features. The completed research work will provide a rigorous physical experimentation involving SNMP monitoring and management for Type I hypervisor Xen.

Keywords Cloud environment · SNMP · Type-1 hypervisors

1 Introduction

Simple Network Management Protocol (SNMP) is an application-layer protocol that is used to monitor IP based devices. SNMP is defined by Internet Architecture Board (IAB) in RFC1157 [1] that helps users to exchange management information between different network devices. SNMP is part of TCP/IP (Transmission Control Protocol/Internet Protocol) protocol suite. Simple Network Management Protocol is

A. Iqbal (✉) · C. Pattinson · A.-L. Kor
Leeds Beckett University, Leeds, UK
e-mail: a.iqbal8836@student.leedsbeckett.ac.uk

C. Pattinson
e-mail: c.pattinson@leedsbeckett.ac.uk

A.-L. Kor
e-mail: a.kor@leedsbeckett.ac.uk

V. Kharchenko et al. (eds.), *Green IT Engineering: Concepts, Models, Complex Systems Architectures*, Studies in Systems, Decision and Control 74,
DOI 10.1007/978-3-319-44162-7_8

usually used to monitor and manage network devices. SNMP has the capability to unobtrusively control or manage the IP based devices with minimal impact on their performances. SNMP Agents are used to communicate with Network Management System (NMS) but these agents must be enabled and configured first [2].

While working in cloud computing environment where virtual machines are running, SNMP can be a good solution for management and control of Virtual Machines. However, currently, there is limited published research work on the deployment of SNMP for monitoring and controlling virtual machines in such type of environment. In this research, it demonstrates how a Type I hypervisor (in a cloud environment) could be monitored and managed with the use of SNMP.

1.1 Motivation for Green IT

Green Information Technology (Green IT) is concept based on practice to use information technology resources in environment friendly way. Main goal of Green IT study is to reduce the negative influence on the environment caused by IT resources. This can be achieved by reducing the amount of energy consumed by resources [3, 4]. The research carried out in this paper provides us monitoring information of virtual machines in the cloud environment. With the use of this monitoring information, we can balance load in peak hours when load at one virtual machine is too much consuming more power than others. We will able to shut down or power up different virtual machines as per our requirements helping us to reduce energy consumption causing greener environment. This will be part of future work.

This is followed by the customization of MIB and net-snmp with Agent X to provide more SNMP management features. The completed research work will provide a rigorous physical experimentation involving SNMP monitoring and management for: (1) Type I hypervisor Xen; (2) SNMP Agent with sub Agent; (3) Virtual Machines; (4) Virtual Machine Manager (VMM).

The rest of this paper is organized as follows. Section 2 provides literature review of Cloud Environment, Hypervisor, SNMP and Libvirt. Section 3 outlines Libvirt experiments with results followed by future work. Section 4 concludes this paper.

2 Literature Review

This section is divided into four sub-sections. The first sub-section provides a brief introduction of cloud computing and virtualization. Types of hypervisors are discussed in the second sub-section. The third sub-section provides a discussion on SNMP and basic components of SNMP. A brief introduction of net-snmp is also part of this sub-section. The last sub-section describes Libvirt and Libvirt-snmp.

2.1 Cloud Environment

Cloud Environment or Cloud computing is an internet-based on-demand computing concept where computers and other devices can get access to shared resources on-demand. Cloud computing model enables ubiquitous, on-demand access to a shared pool of configurable computing resources [5, 6]. Storage solutions provided by cloud computing help users save and conduct operations on third party data centers [7]. The main concept of cloud computing is to provide consistent and economies of scale which is similar to utilities over a network. However, the base idea for cloud computing is providing shared services shared via a converged infrastructure. With nominal controlling effort, one can rapidly release and provision configurable resources such as servers, networks, storage and services etc. [6].

Virtualization is the main supporting technology for cloud computing. Separation of physical computing device into various "virtual" devices is done using virtualization software. Each virtual device can perform computing tasks with minimal management efforts. Operating system-level virtualization allows idle computing resources to be used more efficiently via effective allocation. Scalable system of multiple independent computing devices is created with operating system-level virtualization [8].

2.2 Hypervisors

A hypervisor is a software or firmware that is used to create and run virtual machines. It is also known as Virtual Machine Monitor (VMM). Host machine is one on which hypervisor run, virtual machines known as guest machines. A virtual operating system is presented to guest operating systems to manage their executions. Virtualized hardware resources may be shared by various operating systems.

According to an article "Formal Requirements of Virtualizable Third Generation Architectures", Hypervisors can be classified into two types: Native and Hosted [9].

Type-1: Native or Bare-Metal Hypervisors.

Native or bare-metal hypervisors run directly on the hardware of a host. They control the hardware and manage the guest operating systems. Guest operating system basically runs as a process on the host. IBM first developed native hypervisors such as SIMMON and CP/CMS [10]. Latest type-1 hypervisors include VMware ESX/ESXi, Citrix Xen Server, MS Hyper-V 2008/2012 and Oracle VM Server for SPARC.

Type-2: Hosted Hypervisors.

Type-2 hypervisors just like other computer programs run on a conventional operating system. They are also known as hosted hypervisors. Type-2 hypervisors abstract guest operating systems from the host operating system. Some examples of the hosted hypervisors include QEMU, VMware Workstation and Virtual Box.

Type-1 and Type-2 Hypervisors are illustrated in Fig. 1.

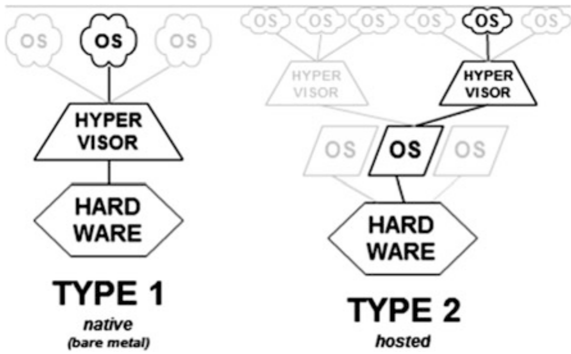


Fig. 1 Type-1 versus Type-2 hypervisors [11]

XEN.

First developed at University of Cambridge Computer Laboratory, XEN project or simply Xen is an example of Type-1 hypervisor that uses a microkernel design. Xen provide services that allow multiple operating systems to run simultaneously on the same computer hardware. Xen project is maintained and developed by Xen Project community as free and open-source software [12].

2.3 SNMP

Simple Network Management Protocol (SNMP) is an application-layer protocol that is used to monitor IP based devices. SNMP is defined by Internet Architecture Board (IAB) in RFC1157 [1] that helps users to exchange management information between different network devices. SNMP is part of TCP/IP (Transmission Control Protocol/Internet Protocol) protocol suite. Simple Network Management Protocol is usually used to monitor and manage network devices. SNMP has the capability to unobtrusively control or manage the IP based devices minimal impact on their performances. SNMP Agents are used to communicate with Network Management System (NMS) but these agents must be enabled and configured first [2].

Basic components of SNMP and their functionality are given below:

- **SNMP Manager**: SNMP manager is used to run one or more NMS making it a separate entity responsible for communication with SNMP agent implemented network devices. Some of the key responsibilities carried out by SNMP: manage query agents, set variables in agents and get response from agents.
- **Managed Devices**: Network elements that require some kind of management and monitoring are known as managed devices. Routers, servers, printers and switches are some examples of managed devices.
- **SNMP Agent**: SNMP agent is an important component of SNMP. It is just a program package inside a network element. Agent collects the management information from the local device and sends it to SNMP manager, when it is

requested for. Net-snmp is an example of standard agent while HP insight agent is an example of vendor specific agent. SNMP Agent also acts as a proxy for some non-SNMP manageable network node.

An information database is maintained by every SNMP agent that describes the parameters of devices being managed. This database is used by SNMP manager to request the agent for particular information. SNMP manager further translates this information as per requirements of NMS. Communication between SNMP manager, managed devices and SNMP agent is illustrated in Fig. 2.

NET-SNMP.

Net-SNMP is an application suite that implements various versions of SNMP such as SNMP v1, v2c and v3 using both IPv6 and IPv4. The suite comes with a command line application to obtain information from an SNMP capable device. MIB browser, daemon application, an extensible agent and a library for developing

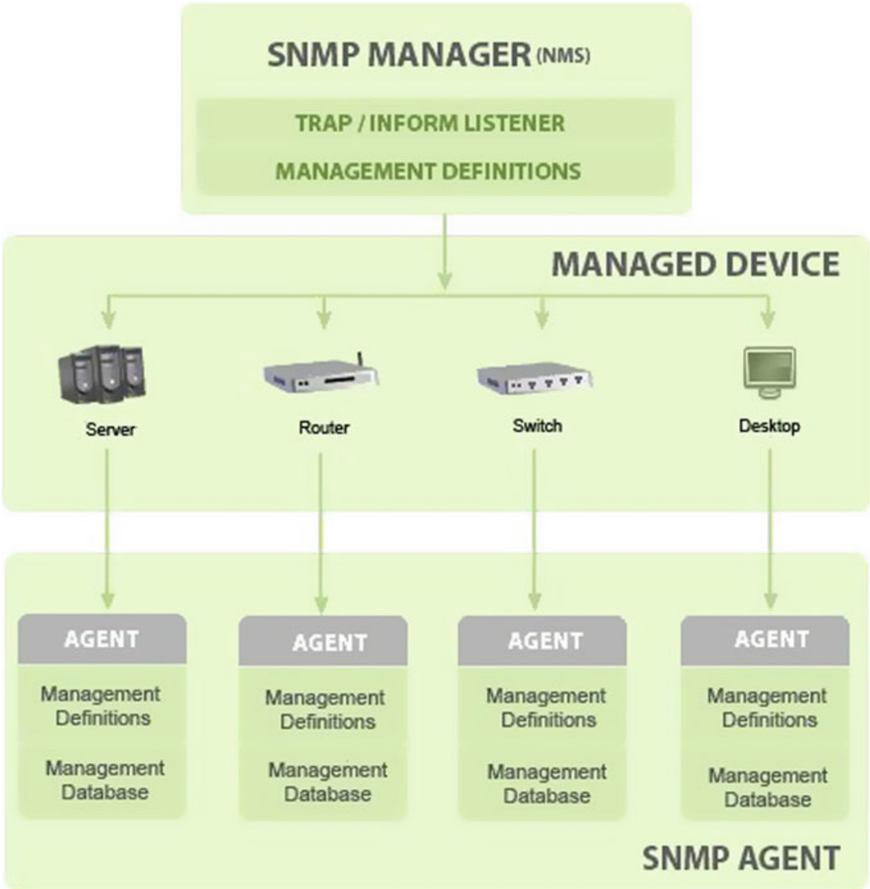


Fig. 2 Basic SNMP communication diagram [13]

new SNMP applications are part of the net-snmp suite. Net-snmp is available for both Linux and Windows operating systems [14].

2.4 Libvirt

Libvirt is an open source API developed by Red Hat to manage platform virtualization. It works like a daemon and management tool and can be used to manage various virtualization technologies such as KVM, Xen, QEMU and VMware ESX. In the development of cloud based solutions, Libvirt API is extensively used in the composition layer of hypervisors. Libvirt is supported by various management solutions like virsh, virt-manager, OpenStack and oVirt. Libvirt supports several Hypervisors like KVM, LXC, XEN and ESX etc. Hypervisors supported by Libvirt and management solutions that support Libvirt are displayed in Fig. 3.

Various virtualization platforms and programs use Libvirt. Both graphical and command line interface are provided. Virtual Machine Manager (VMM) is the most common graphical tool where Virsh is command line tool which is displayed in Fig. 4 [16].

Libvirt provides

- Local access control
- Remote Management and Remote management authentication
- Zero-conf discovery
- Portable client API for Linux and Windows
- Management of virtual machines

Libvirt-snmp

Libvirt-snmp is a sub-project of Libvirt providing SNMP functionality for Libvirt. With the help of Libvirt-snmp, users can monitor domains and their

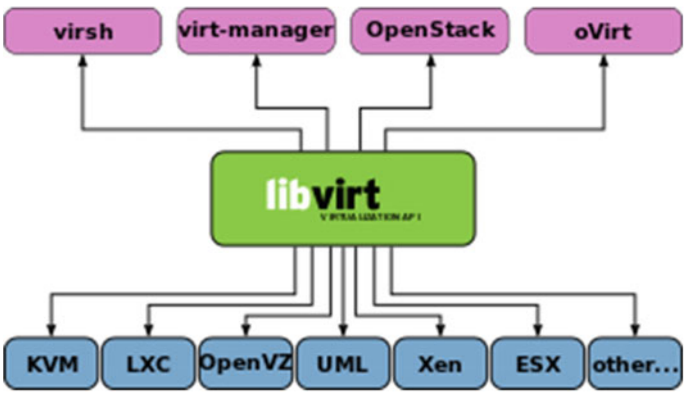


Fig. 3 Libvirt and supported hypervisors [15]

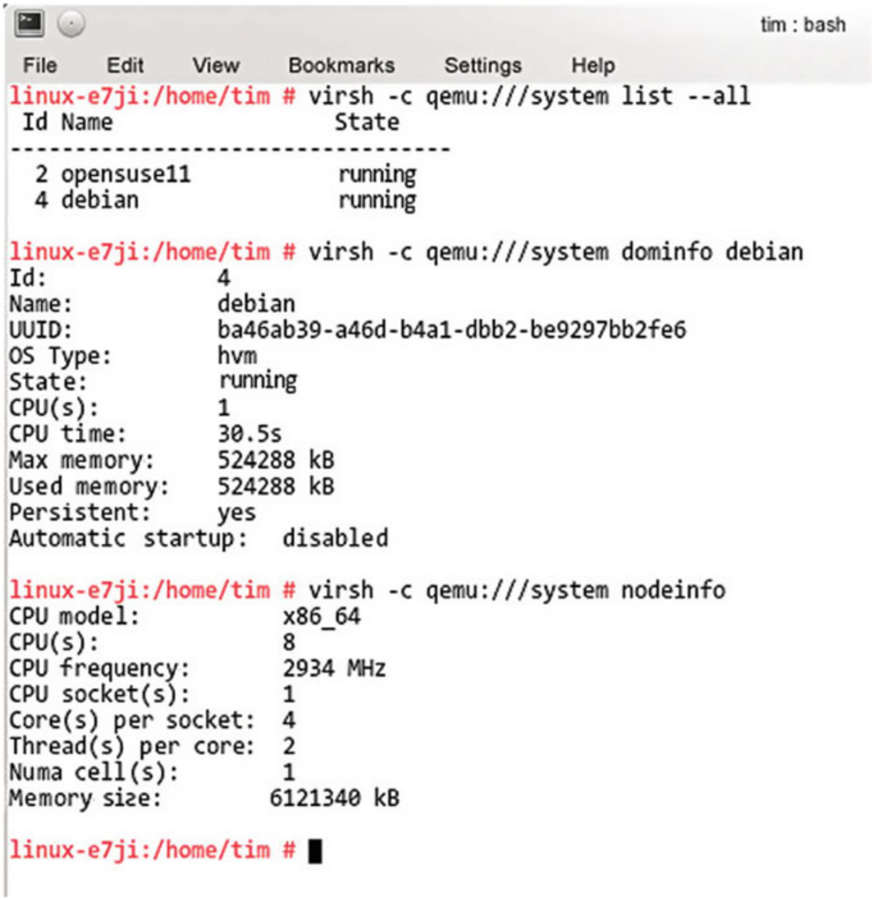


Fig. 4 Virsh command line tool example [16]

attributes over SNMP. Features provided by Libvirt-snmp include: allowing users to request information about domain status, control domain lifecycle and be informed events of domain lifecycle [17].

The basic working concept of Libvirt-snmp is that a snmp agent connects with an already running Libvirt daemon and then listens for all the incoming requests. Whenever a new request arrives, listening agent responds to that request by calling appropriate Libvirt function. Once results are gathered, they are sent back in a typical request-response mechanism. Libvirt-snmp also monitors domain activity and it sends asynchronous traps to the network management system. NMS is immediately notified of unexpected domain crashes. First part of Fig. 5 illustrates a typical request-response mechanism between Agent and NMS while second part of Fig. 5 depicts asynchronous traps sent by Agent [17].

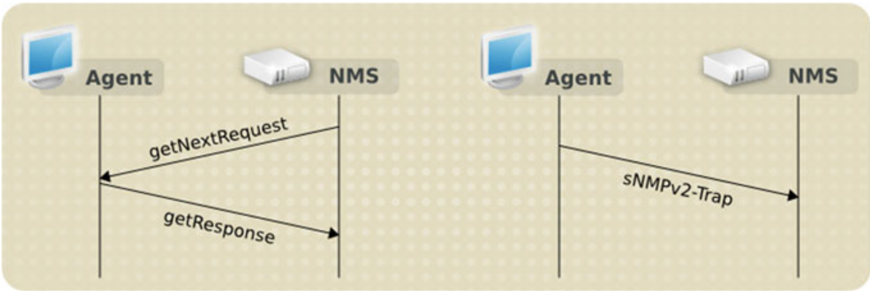


Fig. 5 Libvirt-snmp basic concept

3 Libvirt Experiments

This section outlines the configuration of test environment and experiments done using Libvirt SNMP on virtual machines in a test environment. Test results in the form of screenshots are provided.

3.1 Configuration of Test Environment

This sub-section outlines the procedures required to configure a testing environment so that it could work with Libvirt-snmp. Note that Fedora 22 is used as a base operating system for the configuration.

1. Install XEN Hypervisor using YUM command
2. Once successfully installed, restart computer
3. Start computer with Fedora 22 on XEN Hypervisor option
4. Download Libvirt-snmp source via git command
   `$ git clone git://libvirt.org/libvirt-snmp.git`
5. Install couple of packages before installing Libvirt-snmp
   `$ sudo yum install net-snmp-perl net-snmp net-snmp-utils net-snmp-devel libvirt-devel`
6. Use autobuild.sh command to build Libvirt-snmp
   `$ ./autobuild.sh`
7. On successful compilation, executable binaries will be generated in RPMs directory. Move to that directory
   `$ cd $HOME/rpmbuild/RPMS/<arch>/`
8. Install RPMs using yum command
   `$ sudo yum –nogpgcheck localinstall libvirt-snmp-*.rpm`
9. After successful installation, run the following command to test your environment
   `$ sudo libvirtMib_subagent -H -L`

10. Start/Restart service using
    `$ systemctl restart snmpd.service`
11. Connect to Agent X using
    `$ LIBVIRT_DEFAULT_URI="xen:///system" MibTable_subagent -f –L`
12. Open a new tab in terminal and test different commands as given in the following sub-sections.

3.2 Experiments

1. **SNMP Walk Test**

SNMP Walk test run using Libvirt SNMP provides you details of all virtual machines. Information given by snmpwalk command is as follows:

- **Name**: Name of active virtual guest
- **State**: Current state of virtual guest (running, paused, blocked, shutdown, shutoff, crashed)
- **CPU Count**: Number of virtual CPUs the virtual guest uses
- **Memory Current**: Current amount of memory (in MiB) used by virtual guest
- **Memory Limit**: The maximum amount of memory (in MiB) that can be used by the virtual guest
- **CPU Time**: The CPU time used by the virtual guest, in nanoseconds
- **Row Status**: Status of the virtual guest.

– **Test Command**:
  `snmpwalk -m ALL -v 2c -c public -OX localhost libvirtMIB`

This command will print information of all the guests running on the localhost. For each guest, properties such as name, state, CPU count and current memory are displayed.

– **Result**:

Here it can be seen in Fig. 6, a single guest Virtual Machine with the name fedora21a, is running. Memory usage and memory limit is 2 GB where CPU count

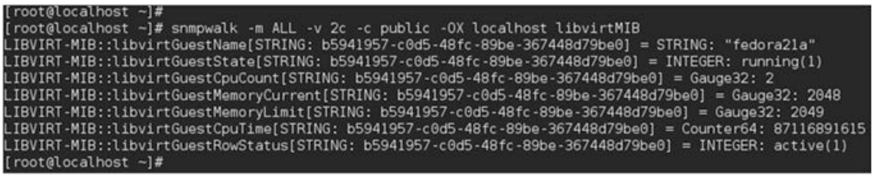


Fig. 6 Snmpwalk command result

is 2. Finally, we have a UUID `b941957-c0d5-48fc-89be-3e-367448d79be0` which is used to refer to this domain without any ambiguity. Information provided by this command such as memory usage and memory limit can be helpful for further experimentation for virtual machines.

2. **SNMP Table Test**

SNMP Table command gives a table of objects that represent active virtual guests on a host. Information provided by this command is the same as of SNMP Walk command.

– **Test Command**:
`snmpwalk -m ALL -v 2c -c public -Cb localhost libvirtGuestTable`

This command prints the same information as the snmpwalk command but in a better tabular format.

– **Result**:

Results of this command (Fig. 7) show that there is one guest virtual machine running with the name, fedora21a. CPU count is 2 where memory usage and limit is 2 GB. It is the same as the snmpwalk command but the difference is that in this command, the output is displayed in a table. UUID in this command is skipped which can be obtained by running snmpwalk command again. If there are many virtual machines running than this command can be very helpful getting statistics of each virtual machine in a Table.

3. **SNMP Set—Paused Test**

SNMP Set command allows the state of the active virtual guest to be changed. In this test, state of active guest was changed from Running to Paused.

– **Test Command**:
`snmpwalk -m ALL -v 2c -c public -Cb localhost libvirtGuestTable. \'b941957-c0d5-48fc-89be-3e-367448d79be0\'` =paused

This command can change state of active guest from Running to Paused given by UUID.

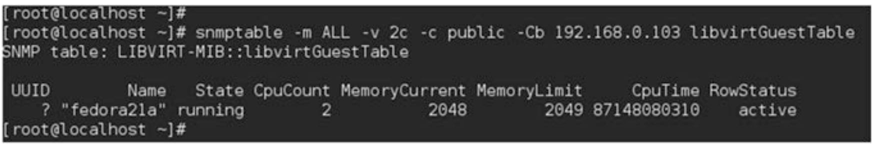


Fig. 7 Snmptable command result

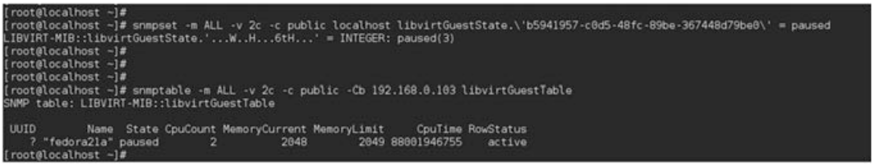


Fig. 8 Snmpset—paused command result

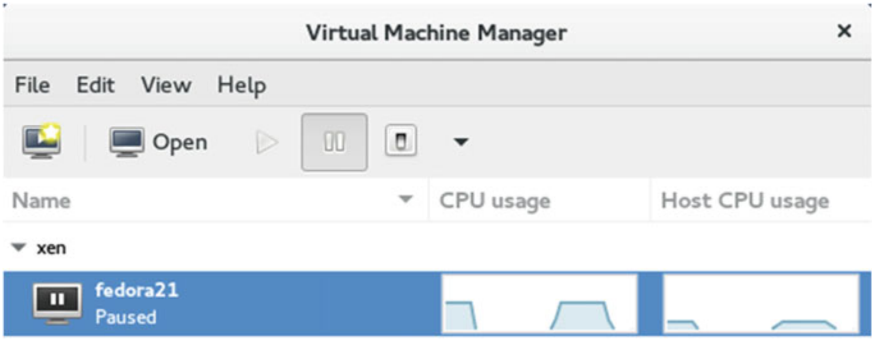


Fig. 9 VMM paused state result

– **Result**:

The results (Fig. 8) show that the state of the active guest with UUID b941957-c0d5-48fc-89be-3e-367448d79be0 has been changed from Running to Paused. Result of the command (snmptable) is displayed in (Fig. 8), and updated state is shown in VMM (Fig. 9). Virtual Machine Manager (VMM) is software that gives graphical interface to view statistics of all the guests created for single host. The guest was active earlier but now it is paused which means all the activities that were being performed by this guest are now halted. UUID is important to successfully execute this command which can be obtained by running snmpwalk command.

4. **SNMP Set—Running Test**

In this test, the state of the active guest was changed from Paused to Running.

– **Test Command**:

```
snmpwalk -m ALL -v 2c -c public -Cb localhost libvirtGuestTable.
\'b941957-c0d5-48fc-89be-3e-367448d79be0\'=running
```

This command can change state of active guest from Paused to Running given by UUID.

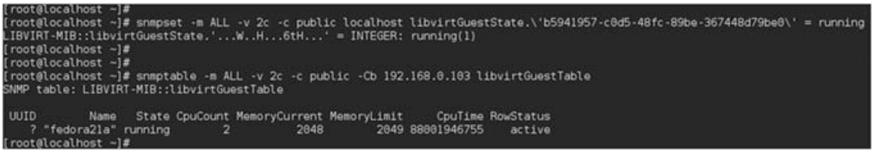


Fig. 10 Snmpset—running command result

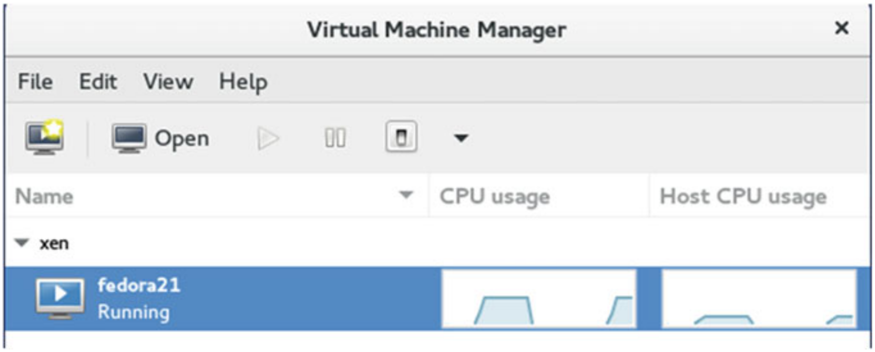


Fig. 11 VMM running state result

– **Result**:

The results (Fig. 10) show that the state of the active guest with UUID b941957-c0d5-48fc-89be-3e-367448d79be0 was changed from Paused to Running. And updated state is shown in the terminal by using snmptable command in Fig. 10. Moreover it can be seen in Fig. 11 too where VMM showing the graphical version of VMs states. When state of guest was changed to pause state all the activities halted. But setting it to running state resumes the activities again.

5. **SNMP Set—Shutdown Test**

In this test, the state of the active guest was changed from Running to Shutdown.

– **Test Command**:

```
snmpwalk -m ALL -v 2c -c public -Cb localhost libvirtGuestTable.
\'b941957-c0d5-48fc-89be-3e-367448d79be0\' =shutdown
```

This command can change state of active guest from Paused/Running to Shutdown given by UUID.

– **Result**:

The results show that the state of the active guest with UUID b941957-c0d5-48fc-89be-3e-367448d79be0 was changed from Running to Shutdown. Output of the command is displayed in Fig. 12 and VMM (Fig. 13) is

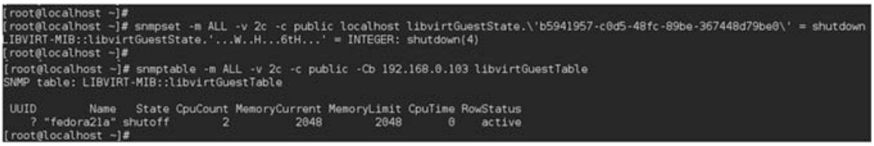


Fig. 12 Snmpset—shutdown command result

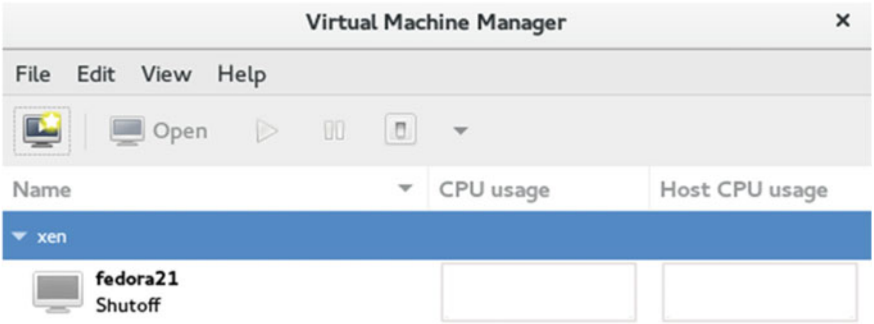


Fig. 13 VMM shutoff state result

also showing that the state has been changed. To shutdown a guest machine in certain situations, this command can help. It clears all the running processes and goes into shutoff state.

4 Conclusion

This paper demonstrated the use of a Simple Network Management Protocol (SNMP) to monitor and manage a Type 1 Hypervisor in a cloud environment. Additionally, we conducted an investigation on load balancing within a cloud environment during peak hours.

SNMP Agent was installed and configured on a Linux Operating System with XEN as (Type 1) Hypervisor. Virtual Machines were created and controlled using a SNMP Agent. State of VMs running on XEN was changed using Libvirt SNMP. Experimentations like changing state from Running to Paused, Paused to Running and Running to Shutdown were done and results are presented in this paper.

Current work also includes monitoring of Virtual Machines to obtain useful information such as: CPU count, current memory and memory limit. This will enable us to monitor virtual machines during peak hours and use this information to balance load on each machine. Future work will include load balancing during peak hours.

5 Future Work

Current research includes monitoring and managing of Virtual Machines running under Type 1 Hypervisor. Rigorous experiments involving state transition were successfully carried out in this research, which would lead us to further exploration in this field.

Future work will include research on load balancing within a cloud environment during peak hours. This can further lead to customization of MIB and net-snmp with Agent X to provide more SNMP management features. Load balancing in cloud environment will help to reduce power consumption leading to greener environment and helping in Green IT cause.

References

1. Case, J., Fedor, M., Schoffstall, M., Davin, J.: A Simple Network Management Protocol (SNMP). https://www.ietf.org/rfc/rfc1157.txt (1990)
2. Ren, J., Li, T. Network management. In Bidgoli, H. (ed.) Handbook of Technology Management. John Wiley & Sons, Inc. (2010)
3. Murugesan, S., Gangadharan, G.R.: Harnessing Green IT. Principles and Practices. Wiley and Sons Ltd., UK (2012)
4. Green Technology: www.greenevolution.ru/enc/wiki/zelenye-texnologii
5. Hassan, Q.: Demystifying cloud computing. J. Defense Softw. Eng. 16–21. (2011)
6. Mell, P. M., Grance, T. SP 800–145. The NIST Definition of Cloud Computing. National Institute of Standards & Technology, Gaithersburg (2011)
7. Haghighat, M., Zonouz, S., Abdel-Mottaleb, M. CloudID: trustworthy cloud-based and cross-enterprise biometric identification. Expert Syst. Appl. 7905–7916 (2015)
8. Hamdaqa, M.: Cloud Computing Uncovered: A Research Landscape, pp. 41–85. Elsevier Press (2012)
9. Popek, G.J., Goldberg, R.P.: Formal requirements for virtualizable third generation architectures. Commun. ACM 412–421 (1974)
10. Meier, S. IBM systems virtualization: servers, storage, and software. 15–20 (2008)
11. Scsami: The two types of hypervisors for virtualization. https://commons.wikimedia.org/wiki/File:Hyperviseur.png
12. Best Quality and Quantity of Contributions in the New Xen Project 4.6 Release. https://blog.xenproject.org/2015/10/13/xen-4-6
13. ManageEngine. SNMP tutorial. https://www.manageengine.com/network-monitoring/what-is-snmp.html
14. NET-SNMP. http://www.net-snmp.org/
15. ScotXW. Libvirt support: https://commons.wikimedia.org/wiki/File:Libvirt_support.svg
16. Schürmann, T.: Virsh Libvert tool. http://www.admin-magazine.com/layout/set/print/Articles/Virsh-Libvert-Tool (2015)
17. Libvirt-snmp concepts: http://wiki.libvirt.org/page/Libvirt-snmp

Efficient Error Detection and Correction in Block Data Transmission

Nikolaos G. Bardis

Abstract The rapid advances of communication technologies that aim to increase the information transmission speeds, aggravate problems of reliable data exchanges. Especially the expansion of the use of wireless telecommunications technologies is accompanied by a noticeable increase of the intensity of the electromagnetic field and consequently by an increase in the number of errors caused by external interference. The importance of the classical criteria, such as the number of control bits, is reduced and more attention is paid to other parameters, such as the computational and temporal complexity of the procedures for correcting errors, as well as transmission energy requirements. The above factors dictate the necessity for sufficient developments of the means for ensuring the reliability of communication systems, including methods for data transmission error correction. This chapter proposes a collection of techniques for correcting transmission burst errors in data transmitted over signal channels suffering from strong electromagnetic interference, such as encountered in distributed and embedded systems. Efficiency is achieved by separating the error detection from the correction process and using different codes for each case. The proposed error control techniques are based on simple mathematical operations and are suitable for implementation in FPGA devices. It hence becomes possible to replace energy demanding retransmission operations, including the overheads they entail with energy efficient local error correction calculations. The techniques employed are shown to be more efficient than existing ones, according to criteria that are relevant to current applications. These techniques reduce the need for error recovery by retransmission and hence the environmental effect of data transmission in terms of energy consumption and electromagnetic emissions.

Keywords Burst error · Error detection · Error correction · Erasure codes

N.G. Bardis (✉)
Department of Mathematics and Engineering Sciences, Hellenic Military Academy,
Varis—Koropiou Avenue, 116 72 Vari, Greece
e-mail: bardis@ieee.org

V. Kharchenko et al. (eds.), *Green IT Engineering: Concepts, Models, Complex Systems Architectures*, Studies in Systems, Decision and Control 74,
DOI 10.1007/978-3-319-44162-7_9

1 Introduction

Over the last decade the speed of data transmission has increased by several orders of magnitude, advances in signal processing have enabled the implementation of complex technologies such as wavelength division multiplexing (WDM) so as to maximize the usability of communication channels and wireless channels have widely expanded their use. The rapid development and integration of technology has also led to an increase of the complexity required for the implementation of the procedures mentioned above as well as to the synthesis of advanced protocols of data exchange. Under those circumstances, the occurrence of errors during data transmission has become a very critical problem. Even more significant, the dynamic development of the means for transmission also had an impact on the nature of the errors that occur during data transmission. For example, the wide use of wireless channels leads to greater exposure to external noise and consequently to an increase in transmission errors that cluster into bursts [1]. Consequently, the problem of data integrity changes faces constantly and becomes more and more challenging. However, in order to maximize the efficiency of communication channels one needs to ensure the control and the processing of the errors during data transmission.

The particularities of the specific channels are the lack of digital data modulation in high frequencies of sinusoidal signals and the lack of a natural separation of the data in symbols—packets of bits that modulate one signal. Another case examined is data integrity for storage units where various problems come up especially with the use of run length limited (RLL) modulation [2]. Indeed, errors can often occur in RAM affecting a lot of memory addresses at the same time. In both cases the prevalent error pattern is the multiple distorted bits that formulate bursts.

Burst error control in data-transmission systems is particularly important in military applications. As an example one can mention the application in the weapon control systems where strong electromagnetic interferences may become sources of errors in some cases during data transmission among the components of the system due to their nature of operation, and in other cases during channel transmission between the computer control weapon system components especially in ships and aircraft [3].

Increases in transmission rates inevitably lead to increases in error rate values, since data transmission is one of the least reliable subsystems of modern computer hardware. The reason for this is that a large number of complex physical processes are simultaneously happening in a means of data transfer and lead to distortions and damage of the transmitted information. In general, the larger the transmission rate is, the more are the physical processes and phenomena that appear and may seriously affect the transmission process.

Apart from the above considerations, during the operation of any physical means of data transfer, noise interference that will be added to the transmitted data and distort it are inevitable. These noises are usually of different intensity, random and their structure resembles that of the information signals, thus making them difficult to detect [4]. Therefore, the physical processes imply that the satisfaction of the need for larger transmission rates needs to be balanced by the design of innovative methods for the detection and correction of errors that enable the safe and reliable information transmission at high transfer rates [3, 5].

The problem of guaranteeing information reliability in control computer systems is even more critical. In such systems, data transfer errors may have significant negative consequences. Hence the exercise of achieving increased efficiency in timing error detection and correction in the context of asynchronous digital communication lines, is currently significant for the field of development of information technology.

The rapid development and integration of digital data transmission systems plays an important role in the contemporary phase of the expansion of computer and information technologies.

Digital data transmission has applications in areas such as mobile telephony (EDGE, CDMA2000, 1xEV-DO), in digital video broadcasting (DVB-T/H/C/S/S2, ATSC), in wireless digital data transmission systems (Wi-Fi, Wi-MAX, Bluetooth), in computer networks etc. [4–6].

The use of different forms of spectral modulation (e.g. Phase Shift Keying—PSK or Quadrature Amplitude Modulation—QAM) for the increase of channel capacity is a characteristic property of modern data transmission systems. In this case, a change of phase may convey more than one bit of information, giving rise to a symbol-oriented channel rather than a bit-oriented channel. For example: QPSK—the channel symbol may convey 2 bits, 8-PSK 3 bits, 16, 32, 64, 128 and 256 QAM—4,5,6,7 and 8 bits respectively.

The increase in channel capacity for the transmission of digital data is impeded by interference during transmission and requires the sufficient development of means for the detection of errors. In a large number of digital data transmission systems, the correction of errors is performed implicitly via the repeated transmission of the packets. For such systems, in contrast to correction codes, the error detection and error correction phases are distinct and spread in different parts of the overall system.

This chapter presents an overview of methods concerned with error detection and error correction that have been developed by the author. More specifically it presents a series of research results that propose solutions to the problems of error detection and correction under various constraints that are significantly better than those of benchmark techniques. Fundamentals of the error detection and correction problems are presented.

2 Foundations of Error Detection and Correction

Information transmission is one of the fundamental processes during the operation of a computer system. This process traverses individual units of the computer system and reaches remote peripheral devices or even remote systems. For the transmission of data at the physical level, a series of encoding schemes have been developed such as Non Return to Zero (NRZ), Alternate Mark Inversion (AMI), Manchester encoding etc. A drawback of most methods is the lack of self-synchronization, i.e. a tendency for suffering from timing errors [4].

The appearance of these errors is due to the fact that the timing period τ_S, at the transmitter is different by a random value δ from the corresponding timing period τ_R, appearing at the receiver, so that:

$$\tau_R = \tau_S + \delta. \tag{1}$$

During the transmission of a series of q unit ("1") bits that is timed at the transmitter and the receiver, the timing difference is $q \cdot \delta$. If $q \cdot \delta$ is proportional to τ_S, there exists a probability that the transmitted sequence of q unit bits is perceived by the receiver as a series of $q + 1$ or $q - 1$ unit bits. It may be shown that the probability for this type of error that is caused from a timing violation increases with the length of consecutive unit values in the transmitted packet [4].

Practical error correcting codes are usually designed to correct several scattered errors and much longer burst errors. However, burst-error mechanisms are often badly understood and there may be no generally accepted models that fit the real behaviour [9]. In general, codes for correcting random errors are not efficient for correcting burst errors. Therefore, it is desirable to construct codes or modify existing techniques specifically for correcting burst errors.

To be more specific, for the most part, research community has attempted a random bit-error correcting code approach on a burst-error channel, and the technique that is mostly used is interleaving, namely the shuffling of the symbols at the receiver, so as to break up and spread around the burst of errors thereby creating an effectively random channel [10].

Generally, many sources can cause different types of errors in data transmission channels. Indicatively we refer to thermal noise that is present in every electronic device, cross talk between signals, noise or interference signals caused by events external to the transmission path (solar flares affecting satellite electronics or radio waves, heavy electrical machinery etc.), frequency, amplitude, or phase distortion of the transmitted signal, atmospheric fading etc. [11]. Unlike the errors that are caused by thermal noise, all other sources tend to give errors that formulate bursts. About the frequency of the error events only empirical conclusions can be drawn by testing under real conditions.

The term burst errors suggests that those errors are correlated, i.e. if one bit is erroneous, it is quite likely that the adjacent bits have also been corrupted. When one refers to the term burst error of size m, what is meant is that the distance in bits from the first to the last error in the frame is at most m − 1 while the intermediate bits may or may not be corrupted. One thing worth mentioning is that for the same bit error rate, burst errors are in principle easier to deal with than random errors. For example, consider the case of frames consisted of 1000 bits and an overall error rate of 1 in 1000. If all errors were random, almost all frames would be corrupted. However, if dealing with burst error events of size 1000, only 1 frame in 1000 would be corrupted [12]. The disadvantage that comes with the appearance of burst error events is that error detection and correction as well as analytical modeling of the specific patterns is much more difficult to achieve compared with the case of random errors.

When random error is the case, the theoretical model of the memoryless Binary Symmetric Channel (BSC) is being used [13]. The BSC is not appropriate for the description of burst error events as a channel with memory is necessary [6].

The probability that two burst errors occur is lower than the probability of the occurrence of a singular burst error. Thus, burst error multiplicity is usually considered to be equal to one in the state of the art.

With the use of special coding it is possible to distinguish two approaches for the control of the appearing errors:

- the error detection by special codes and their correction by retransmitting the data block (ARQ—Automatic Repeat Request),
- the correction of the appearing errors by applying correcting codes without the repeated transmission (FEC—Forward Error Correction).

The choice of the appropriate technology is determined by the features of the data transmission channel as well as the particularities of the specific use. Thus, in wired systems for digital data transmission, in which the intensity of errors is several orders lower in comparison with the wireless channels, the use of ARQ is considered to be more effective. On the contrary, in the wireless channels the prevailing source of the transmitted errors is the externally produced noise and in this case the intensity of the appearing errors is high enough to consider the application of FEC technologies to be more preferable [14].

In general, ARQ technology is being preferred over FEC technology as the latter demands more complicated decoding hardware and greater redundancy. This makes sense if one takes into consideration that FEC technology demands not only the knowledge of the occurrence of at least one error but also of the exact number of erroneous bits and of course of the exact positions of the corrupted bits in the affected block. Therefore, the drawbacks of error correction codes include the relatively large volume of control information, the complexity of the procedures and the decoding process. As a matter of fact, the number of control bits increases exponentially as the multiplicity of errors that are required to be corrected increases. Additionally, in the case of correction codes the computations for the syndrome calculation take place even if no corruption has occurred and that causes a considerable cost in complexity [15].

3 Sample Implementations and Applications of the Error Correction Framework

3.1 Burst Error Correction Based on the Binary Weighted Checksum

A method for error correction for application on magnetic disks.

The primary idea on which the method is based makes use of a more sophisticated problem solving approach that suggests the analysis of the initial problem in order to produce a set of sub-problems such that solutions to the sub-problems would imply a solution to the original problem [16–18]. Thus, in the examined case, the original problem of burst error correction is broken down into the two sub-problems below [19–24]:

- Localize the distorted bits in the burst,
- Localize the bit where the burst originates.

This technique will enable us to reduce the size and the complexity of the initial problem while in the Reed–Solomon case, the two sub-problems are so closely interrelated that the computational cost becomes extremely high. Each sub-problem will be dealt with separately using the most appropriate method for each case. Thus, in order to locate the corrupted bits in the burst one could simply perform the conventional checksum where the size of the checksum must be approximately twice the maximum size of the burst. For the actual localization of the start of the burst the weighted checksum is used, namely the checksum version that takes into consideration both bit values and their position.

Consequently, one firstly needs to localize the bits of the burst that are corrupted or in other words to compute the m-dimensional error vector of the burst $E = \{e_1, e_2,\ldots, e_m\}$, $\forall l = 1, \ldots, m$: $e_l \in \{0,1\}$ and then to localize the first bit of the burst. The steps to follow are described below:

Let us consider a control packet B that consists of N bits: $B = \{b_1, b_2, \ldots, b_N\}$, $b_l \in \{0,1\}$ $\forall l \in \{1, \ldots, N\}$. One could consider the sequence of bits as n symbols of size $2 \cdot (m - 1)$ bits, where m is the maximum size of the burst. Each one of the N bits of the control packet is assigned to a weight coefficient W_i, $i \in \{1, \ldots, N\}$ of size $q = \lfloor \log_2 N \rfloor$. The set of weight coefficients $W_1, W_2, \ldots, W_N$ is actually a sequence of elements $\alpha, \alpha^2, \ldots \alpha^N$ of the Galois field with generator polynomial an irreducible polynomial $P(x)$ of degree q. For example, for $N = 15$, $q = 4$ και $P(x) = x^4 + x + 1$ the values of weight coefficients for the 15 bits are presented in Table 1.

Table 1 Units for magnetic properties

Number j (bit)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Weight factor W_j	1	3	7	15	14	13	10	5	11	6	12	9	2	4	8

The control code comes up via the computation of two binary components C_1 and C_2, first of which is of size $2 \cdot (m - 1)$ and is defined as the XOR—checksum of the n symbols of the packet.

$$C_1 = X_1 \oplus X_2 \oplus \cdots \oplus X_n \tag{2}$$

The latter component, namely C_2, is formulated as the weighted checksum of the bits and the aforementioned weight coefficients. Namely,

$$C_2 = b_1 \cdot W_1 \oplus b_2 \cdot W_2 \oplus \cdots \oplus b_N \cdot W_N \tag{3}$$

Returning to the example above, in order to control a 15-bit packet B = 1001 1010 1100 101 with $m = 4$, the packet divides into three 6-bit symbols (the missing bits are assigned to 0).

$$X_1 = 100110, \quad X_2 = 101100 \quad \text{and} \quad X_3 = 101000$$

Consequently, for the first component of the control code we have:

$$C_1 = X_1 \oplus X_2 \oplus X_3 = 100110 \oplus 101100 \oplus 101000 = 100010$$

The second component of the control code is computed as below:

$$\begin{aligned} C_2 &= b_1 \times W_1 \oplus b_2 \times W_2 \oplus \cdots \oplus b_{15} \cdot W_{15} \\ &= 1 \oplus 15 \oplus 14 \oplus 10 \oplus 11 \oplus 6 \oplus 2 \oplus 8 = 13_{10} = 1101_2 \end{aligned}$$

Consequently, the control code is $\langle 100010{,}1101 \rangle$.

To the receiver's end the control code of the weighted checksum is computed in a similar way. For example, let us consider that during the transmission external noise caused a burst of errors that originates in 5th bit and consists of a corrupted 5th, 6th and 8th bit. Consequently, the receiver received packet B' as follows: B' = 1001 **0111** 1100 101. Thus, for the receiver $X_1' = 100101$, $X_2' = 111100$ and $X_3' = 101000$. Consequently,

$$\begin{aligned} C_1' &= X_1' \oplus X_2' \oplus X_3' = 100101 \oplus 111100 \oplus 101000 = 110001. \\ C_2' &= b_1' \cdot W_1 \oplus b_2' \cdot W_2 \oplus \cdots \oplus b_{15}' \cdot W_{15} \\ &= 1 \oplus 15 \oplus 13 \oplus 10 \oplus 5 \oplus 11 \oplus 6 \oplus 2 \oplus 8 = 11_{10} = 1011_2 \end{aligned} \tag{4}$$

The receiver calculates the difference Δ using the two components computed in their end and those that were received. Indeed, $\Delta = \langle \Delta_1, \Delta_2 \rangle$ where $\Delta_1 = C_1 \oplus C_1'$, $\Delta_2 = C_2 \oplus C_2'$.

- If both components of code Δ are equal to zero, namely $\Delta_1 = 0$ και $\Delta_2 = 0$, then the packet has been transmitted without errors.
- If the opposite is the case, code Δ_1 is analyzed:

– If Δ_1 does not include $m-2$ $(2(m-1)-m)$ successive zeros (taking into account that last bit in Δ_1 is accessed before the first) then one concludes that more than a burst error occurred during the transmission, namely the errors cannot be corrected. In that case, a retransmission of the packet is required by the receiver.
– If $\Delta_1 \neq 0$ and it includes $m-2$ successive zeros then the case is categorized as a common single burst error which can be corrected following the steps explained below:

In the context of the example developed above we have $\Delta_1 = 100010 \oplus 110001 = 010011$ and $\Delta_2 = 1101 \oplus 1011 = 0110$.

Indeed, this is the usual case of a single burst error. The correction takes place in the two aforementioned levels:

- The determination of the error vector E of the burst,
- The localization of the bit which is the start point of the burst.

For the first level of the localization of the corrupted bits we define the position of the first distorted bit of the burst as the 1 that is located exactly before $m-2$ consecutive zeros (taking into account that the last bit of Δ_1 is accessed before the first bit). The software implementation makes use of shift operations as well as masking while the hardware implementation makes use of shift registers and units of logic elements OR with m – 2-bits data input that detect the existence of the $m-2$ last zeros. Thus, for the example where $\Delta_1 = 010011$ and after the required arithmetic shifts, the last $m-2$ digits are equal to zero. This implies that the components of vector E are the first $m = 4$ bits of code Δ_1, namely $E = \{1,1,0,1\}$.

Next, determine the weight coefficient $W_j = \{w_{j,1}, w_{j,2}, \ldots, w_{j,q}\}, j \in \{1, \ldots, N\}$ that corresponds to the first distorted bit in the burst based on the already known vector $E = \{e_1, e_2, \ldots, e_m\}$ and code $\Delta_2 = \{\delta_1, \delta_2, \ldots, \delta_q\}$. Indeed, a linear system the solution of which gives the weight coefficient W_j will be presented with the help of the function $\lambda_1(w_{j,1}, w_{j,2}, \ldots, w_{j,q})$ that is defined as follows:

The linear function $\lambda_1(w_{j,1}, w_{j,2}, \ldots, w_{j,q})$ is determined by the irreducible polynomial generator $P(x)$ and given the components of weight coefficient W_j it computes the qth component of weight coefficient $W_{j+1.}$ Namely,

$$w_{j+1,q} = \lambda_1\left(w_{j1}, w_{j2}, \ldots, w_{jq}\right)$$

Because of the particular nature of Galois fields, for all other $q-1$ components of weight coefficient W_{j+1} we have:

$$w_{j+1,1} = w_{j,2}, w_{j+1,2} = w_{j,3}, \ldots, w_{j+1,q-1} = w_{j,q}$$

Thus,

$$W_{j+1} = \left\{w_{j2}, w_{j3}, \ldots, w_{jq}, \lambda_1\left(w_{j1}, w_{j2}, \ldots, w_{jq}\right)\right\}$$

For example, in the case examined above, where

$$P(x) = x^4 + x + 1$$

the linear function is

$$\lambda_1\left(w_{j,1}, w_{j,2}, w_{j,3}, w_{j,4}\right) = w_{j,1} \oplus w_{j,4} = w_{j+1,4}.$$

This means that if W_{11} = {1,1,0,0} then W_{12} = {1,0,0, 1 $\oplus$ 0}.

In a much similar way,

$$\begin{aligned} W_{j+2} = \big\{ & w_{j,3}, w_{j,4}, \ldots w_{j,q}, \lambda_1(w_{j,1}, w_{j,2}, \ldots, w_{j,q}), \\ & \lambda_1(w_{j,2}, \ldots, w_{j,q}, \lambda_1(w_{j,1}, w_{j,2}, \ldots, w_{j,q}))\big\} \end{aligned}$$

If $\lambda_2(w_{j,1},w_{j,2},\ldots,w_{j,q}) = \lambda_1(w_{j,2},\ldots,w_{j,q},\lambda_1(w_{j,1},w_{j,2},\ldots,w_{j,q}))$,

$$W_{j+2} = \big\{w_{j,3}, w_{j,4}, \ldots w_{j,q}, \lambda_1(w_{j,1}, w_{j,2}, \ldots, w_{j,q}), \lambda_2(w_{j,1}, w_{j,2}, \ldots, w_{j,q})\big\}$$

Returning to the example,

$$\lambda_2 = w_{j,2} \oplus \lambda_1(w_{j,1}, \ldots, w_{j,4}) = w_{j,2} \oplus w_{j,1} \oplus w_{j,4}$$

and the weight coefficient can be presented as below:

$$\begin{aligned} W_{13} &= \{w_{11,3}, w_{11,4}, w_{11,1} \oplus w_{11,4}, w_{11,2} \oplus w_{11,1} \oplus w_{11,4}\} \\ &= \{0, 0, 1 \oplus 0, 1 \oplus 1 \oplus 0\} = \{0, 0, 1, 0\} \end{aligned}$$

Generalizing, one can deduce that components of the weight coefficient W_{j+m} that corresponds to the last bit affected by the burst error can be expressed with the help of the components of weight coefficient W_j that corresponds to the first bit in the burst. Indeed,

$$\begin{aligned} W_{j+m} &= \big\{w_{j,m}, \ldots, w_{j,q}, \lambda_1(w_{j,1}, \ldots, w_{j,q}), \ldots, \lambda_{m-1}(w_{j,1}, \ldots, w_{j,q})\big\} \text{ for } m < q \\ W_{j+m} &= \big\{w_{j,q}, \lambda_1(w_{j,1}, \ldots, w_{j,q}), \ldots, \lambda_{q-1}(w_{j,1}, \ldots, w_{j,q})\big\} \text{ for } m = q \\ W_{j+m} &= \big\{\lambda_{m-q+1}(w_{j,1}, \ldots, w_{j,q}), \ldots, \lambda_m(w_{j,1}, \ldots, w_{j,q})\big\} \gamma\iota\alpha m > q \end{aligned} \quad (5)$$

Generally speaking, each function λ_l with l = 2, …, m − 1 is defined in a recursive way based on previous functions λ_1, λ_2, …, λ_{l-1} in the following way:

$$\lambda_l(w_{j,1}, \ldots, w_{j,q}) = \lambda_1(w_{j,i}, \ldots, w_{j,q}, \lambda_1(w_{j,1}, \ldots, w_{j,q}), \ldots, \lambda_{l-1}(w_{j,1}, \ldots, w_{j,q})) \quad (6)$$

The computations so far have yielded the:

- Component Δ_1 of code of difference Δ namely error vector E,
- Component Δ_2 of code of difference Δ where

$$\Delta_2 = C_2 \oplus C_2' = (b_1 \oplus b_1') \cdot W_1 \oplus (b_2 \oplus b_2') \cdot W_2 \oplus \cdots \oplus (b_N \oplus b_N') \cdot W_N$$

- Formulae (3) that suggest that components of weight coefficient W_{j+m} (that corresponds to the last bit affected by the burst error) can be expressed with the help of the components of weight coefficient W_j (that corresponds to the first bit in the burst).

If we assume that the burst of errors originated in the bit of order j, component δ_1 of code Δ_2 is the checksum of all first components of weight coefficients that correspond to the positions of 1's in error vector E.

$$e_1 \cdot w_{j,1} \oplus e_2 \cdot w_{j+1,1} \oplus e_3 \cdot w_{j+2,1} \oplus, \ldots, \oplus e_m \cdot w_{j+m,1} = \delta_1 \tag{7}$$

Formula (5) can also express all other components $\delta_2,\ldots,\delta_q$ of difference Δ_2 and that is how the following system comes up:

$$\begin{array}{l} e_1 \cdot w_{j,1} \oplus e_2 \cdot w_{j+1,1} \oplus e_3 \cdot w_{j+2,1} \oplus \cdots \oplus e_m \cdot w_{j+m,1} = \delta_1 \\ e_1 \cdot w_{j,2} \oplus e_2 \cdot w_{j+1,2} \oplus e_3 \cdot w_{j+2,2} \oplus \cdots \oplus e_m \cdot w_{j+m,2} = \delta_2 \\ \ldots \\ e_1 \cdot w_{j,q} \oplus e_2 \cdot w_{j+1,q} \oplus e_3 \cdot w_{j+2,q} \oplus \cdots \oplus e_m \cdot w_{j+m,q} = \delta_q \end{array} \tag{8}$$

Taking into account that each component of the weight coefficients W_{j+1}, W_{j+2}, …, W_{j+m} can be expressed as a linear function of weight coefficient components W_j as the set of equations (3) suggests we formulate the system:

- **For $m \geq q$,**

$$\begin{array}{l} e_1 \cdot w_{j,1} \oplus e_2 \cdot w_{j,2} \oplus \cdots \oplus e_m \cdot \lambda_{m-q+1}(w_{j,1}, \ldots, w_{j,q}) = \delta_1 \\ e_1 \cdot w_{j,2} \oplus e_2 \cdot w_{j,3} \oplus \cdots \oplus e_m \cdot \lambda_{m-q+2}(w_{j,1}, \ldots, w_{j,q}) = \delta_2 \\ \ldots \\ e_1 \cdot w_{j,q} \oplus e_2 \cdot \lambda_1(w_{j,1}, \ldots, w_{j,q}) \oplus \cdots \oplus e_m \cdot \lambda_m(w_{j,1}, \ldots, w_{j,q}) = \delta_q \end{array} \tag{9}$$

- **For $m < q$,**

$$\begin{array}{l} e_1 \cdot w_{j,1} \oplus e_2 \cdot w_{j,2} \oplus \cdots \oplus e_m \cdot w_{j,m} = \delta_1 \\ e_1 \cdot w_{j,2} \oplus e_2 \cdot w_{j,3} \oplus \cdots \oplus e_m \cdot w_{j,m+1} = \delta_2 \\ \ldots \\ e_1 \cdot w_{j,q-m+1} \oplus e_2 \cdot w_{j,q-m+2} \oplus \cdots \oplus e_m \cdot w_{j,q} = \delta_{q-m+1} \\ e_1 \cdot w_{j,q-m+2} \oplus e_2 \cdot w_{j,q-m+3} \oplus \cdots \oplus e_m \cdot \lambda_1(w_{j,1}, \ldots, w_{j,q}) = \delta_{q-m+2} \\ \ldots \\ e_1 \cdot w_{j,q} \oplus e_2 \cdot \lambda_1(w_{j,1}, \ldots, w_{j,q}) \oplus \cdots \oplus e_m \cdot \lambda_{m-1}(w_{j,1}, \ldots, w_{j,q}) = \delta_q \end{array} \tag{10}$$

Systems (7) and (8) are sets of q linear equations with $w_{j,1}$, $w_{j,2}, \ldots, w_{j,q}$ as the q binary unknowns and are simple enough to be solved with common methods. The solution to each system clearly points out to the weight coefficient W_j, namely to the bit where the burst originated.

Applying the results mentioned above to our example, where $E = \{1,1,0,1\}$, which means $e_1 = 1$, $e_2 = 1$, $e_3 = 0$ and $e_4 = 1$, $\Delta_2 = \{0,1,1,0\}$ namely $\delta_1 = 0$, $\delta_2 = 1$, $\delta_3 = 1$ και $\delta_4 = 0$ and $q = m = 4$, the following set of equations are formulated:

$$\begin{aligned} w_{j,1} \oplus w_{j+1,1} \oplus w_{j+3,1} &= 0 \\ w_{j,2} \oplus w_{j+1,2} \oplus w_{j+3,2} &= 1 \\ w_{j,3} \oplus w_{j+1,3} \oplus w_{j+3,3} &= 1 \\ w_{j,4} \oplus w_{j+1,4} \oplus w_{j+3,4} &= 0 \end{aligned} \tag{11}$$

But according to formula (3),

$$\begin{aligned} & w_{j+1,1} = w_{j+1,2} = w_{j,3}, \quad w_{j+1,3} = w_{j,4} w_{j+1,4} = w_{j,1} \oplus w_{j,4} \\ & w_{j+2,1} = w_{j+1,2} = w_{j,3}; \quad w_{j+2,2} = w_{j,4}; \\ & w_{j+2,3} = w_{j,1} \oplus w_{j,4}; \\ & w_{j+2,4} = w_{j+1,1} \oplus w_{j+1,4} = w_{j,2} \oplus w_{j,1} \oplus w_{j,4} w_{j+3,1} = w_{j,4}; \\ & w_{j+3,2} = w_{j,1} \oplus w_{j,4}; \\ & w_{j+3,3} = w_{j,2} \oplus w_{j,1} \oplus w_{j,4}; \\ & w_{j+3,4} = w_{j+2,1} \oplus w_{j+2,4} = w_{j,3} \oplus w_{j,2} \oplus w_{j,1} \oplus w_{j,4} \end{aligned} \tag{12}$$

Consequently, after the required substitutions and the calculation of several basic transformations the final system to solve comes up:

$$\begin{aligned} w_{j,1} \oplus w_{j,2} \oplus w_{j,4} &= 0 \\ w_{j,2} \oplus w_{j,3} \oplus w_{j,1} \oplus w_{j,4} &= 1 \\ w_{j,3} \oplus w_{j,1} \oplus w_{j,2} &= 1 \\ w_{j,4} \oplus w_{j,2} \oplus w_{j,3} &= 0 \end{aligned} \tag{13}$$

The solution of the system consists of the values $w_{j,1} = 1$, $w_{j,2} = 1$, $w_{j,3} = 1$ και $w_{j,4} = 0$. Thus, $W_j = 1110_2 = 14$ which means that the start of the burst error is located in the 5th bit. According to the values of the components of vector $E = \{1,1,0,1\}$, the correction of the burst error is obtained with the inversion of the 5th, the 6th and the 8th bit of the packet.

In terms of computational complexity the most efficient way to solve linear systems with binary coefficients ((7) or (8)) is the use of a memory table. For a fixed polynomial $P(x)$ it is quite easy to express the components of weight coefficient W_j as linear functions of the components of code Δ_2 for all $2^m - 1$ possible values of E.

Thus, for our example, if $E = \{1,1,0,1\}$ then the value of the component W_j will be:

$$
\begin{aligned}
w_{j,1} &= \delta_2 \oplus \delta_4 \\
w_{j,2} &= \delta_1 \oplus \delta_3 \\
w_{j,3} &= \delta_1 \oplus \delta_2 \\
w_{j,4} &= \delta_2 \oplus \delta_3
\end{aligned}
$$

The total size of the memory table does not exceed $2^m \cdot q$ bits. Taking into account that practically almost always $m \leq 8$, $q \leq 16$, the size of the memory table is of order of 10 kB. This means that the memory table is a viable solution when it comes to both software and hardware implementation.

3.2 General Error Control Based on the Weighted Checksum

It is widely known that checksums are the fastest error control method. Among the various modifications of checksums, the most efficient ones are the Weighted Checksums (WCS) [6–8]. Weighted checksum works in a way similar with the normal one. The specific distinction between those two methods is that while the normal version is based on a simple summation of the bit values, the weighted version takes both data value and position into consideration. Their basic advantage is that they are able to adjust to special types of error patterns. Based on the specific advantage, the goal can be achieved with the development of a modification of the normal WCS method which will be able to adjust for the control of bursts of errors by one or two interference incidents [25].

Let us consider a control packet B that consists of n bits: $B = \{b_1, b_2, \ldots, b_n\}$, $b_l \in \{0,1\}\ \forall l \in \{1, \ldots, n\}$. The length q of the control packet may then be calculated as $q = n/m$ m-bit channel symbols: $B = \{X_1, X_2, \ldots, X_q\}$. The symbol at the jth position X_j, $j \in \{1, 2, \ldots, q\}$ contains m mixed bits of the packet being checked for errors. Namely,

$$
\forall \mathrm{j} \in \{1, \ldots, \mathrm{q}\} : \ \mathrm{X_j} = \{\mathrm{b}_{(\mathrm{j}-1)\times \mathrm{m}+1}, \mathrm{b}_{(\mathrm{j}-1)\times \mathrm{m}+2}, \ldots, \mathrm{b}_{\mathrm{j}\times \mathrm{m}}\}
$$

Then, packet B can be studied as a matrix X of dimensions $q \times m$ with each row representing one of the q symbols and each column representing one of m bit positions.

The control code that is proposed is a vector $C = \{S_0, S_1, S_2, \ldots, S_m, P\}$. Below each of the components above are analysed:

$$
\forall i \in \{1, 2, \ldots, m\}, \quad p_i = x_{1i} \oplus x_{2i} \oplus \cdots \oplus x_{qi} = b_i \oplus b_{m+i} \oplus \cdots \oplus b_{n-q+i} \quad (14)
$$

This is the component that detects all errors of odd multiplicity in the ith position of the symbols.

The k-bit ($k = \log_2 q$) code S_0 makes use of weight coefficients Wj. A typical way to assign weight coefficients is to give the jth symbol of the packet the weight W_j, namely $W_j = j$. Then S_0 is calculated in two steps. First, we formulate the q terms by summing up the bits of each symbol and then by multiplying by the respective weight coefficient. Secondly, the q calculated terms are XORed.

$$S_0 = \bigoplus_{j=1}^{q} W_j \cdot \left(x_{j1} \oplus x_{j2} \oplus \cdots \oplus x_{jm}\right) \tag{15}$$

The component above points out to the symbols that have been affected by errors of odd multiplicity.

For the formulation of m components $S_1, S_2, \ldots, S_m$ of the control code, the set of weighting factors $U_1, U_2, \ldots, U_m$ is used. The ith weighting factor U_i contains q components $U_i = \{u_{i1}, u_{i2}, \ldots, u_{iq}\}$ satisfying specific properties. S_i is formulated as it is shown below:

$$\forall i \in \{1, 2, \ldots, m\}, \quad S_i = \bigoplus_{j=1}^{q} u_{ij} \cdot x_{ji} \tag{16}$$

The set of weighting factors $U_1, U_2, \ldots, U_m$ is such that the following conditions are satisfied:

1. $\forall i \in \{1, \ldots, m\}, \forall j, e \in \{1, \ldots, q\}, j \neq e$: $u_{ij} \neq u_{ie}$.
2. For each pair (U_i, U_v) $\forall i, v \in \{1, \ldots, m\}$ holds that $\forall j,e,z,y \in \{1, \ldots, q\}, j \neq e, j \neq z, j \neq y, e \neq z, e \neq y, z \neq y$:

$$(u_{ij} \oplus u_{ie}) + 2^{k+1} \cdot (u_{vj} \oplus u_{ve}) \neq (u_{iz} \oplus u_{iy}) + 2^{k+1} \cdot (u_{vz} \oplus u_{vy}). \tag{17}$$

The components above, point out to the symbols that have been affected by one or two interference incidents.

A set of values U_1, U_2, U_3, U_4 that satisfy the proposed condition for $m = 4$ and $q = 10$ are given in Table 2.

From the second condition one can deduce that an appropriate set $U_1, U_2 \ldots U_q$ offers unique localization of two symbols when two errors have occurred in the same bit positions for both symbols. For the example presented in Table 2, let us consider that corruptions have occurred in 2nd and 3rd position for two symbols $r, s \in \{1,\ldots,10\}$. As it is analyzed in Table 3, the value of $\langle u_{2,r} \oplus u_{2,s}, u_{3,r} \oplus u_{3,s}\rangle$, that comes up through codes of difference ΔS_2 and ΔS_3 allows the unique identification of one of the 45 possible pairs $\langle r, s\rangle$.

Table 2 Set of weighting factors

Serial number of symbol	Set of weighting factors			
	U_1	U_2	U_3	U_4
1	2	2	2	2
2	3	1	1	1
3	4	4	4	4
4	5	3	5	6
5	6	7	3	8
6	7	8	8	3
7	8	5	6	9
8	9	9	11	5
9	10	10	15	10
10	11	14	13	15

Let us consider a data packet which is consisted of 40 bits and that external noise does not last more than the transmission duration of one 4-bit symbol. Hence, we have $m = 4$ and $q = 10$: $X_1 = \{1001\}$, $X_2 = \{0011\}$, $X_3 = \{1000\}$, $X_4 = \{1111\}$, $X_5 = \{1011\}$, $X_6 = \{0001\}$, $X_7 = \{1100\}$, $X_8 = \{0110\}$, $X_9 = \{0101\}$, $X_{10} = \{0010\}$. If we use the set of weighting factors U_1, U_2, U_3 and U_4, of Table 2 then the components of the control code will be computed as such:

$$S_0 = \{1010\}, \quad S_1 = \{1101\}, \quad S_2 = \{0101\},$$
$$S_3 = \{0001\}, \quad S_4 = \{0100\}, \quad P = \{1010\}$$

3.3 Error Correction Based on Binary Multiplication Without Carry

The correction procedure consists of two levels [26]:

1. Determination of the number r of the low order bits, from which the distortion of the symbol or pair of symbols begins. In the case of the distortion of two adjacent symbols, the sequence number d of the lowest order bit between those two is also determined. If only a single symbol has been distorted. then r corresponds to the sequence number of the first bit with value 1 within Δ_1 or Δ_2. If two adjacent bits have been distorted, then r corresponds to the sequence number of the first bit with a value of 1 in Δ_2 and d-the sequence number of the first bit with a value of 1 in Δ_1, if $\Delta_1 > \Delta_2$ (an odd numbered symbol is the first one to be distorted). Otherwise, r corresponds to the first bit with a value of 1 in Δ_2, and d-the sequence number of the bit with a value of 1 in Δ_1, if $\Delta_2 > \Delta_1$ (the first distorted symbol is an even numbered one). For the particular example analyzed earlier on, since $\Delta_1 > \Delta_2$, r is equal to the sequence number of the

Table 3 Use of U_2 and U_3 for localizing the pair of symbols that where corrupted in 2nd and 3rd bit

$u_{2,r} \oplus u_{2,s}$, $u_{3,r} \oplus u_{3,s}$	r, s	$u_{2,r} \oplus u_{2,s}$, $u_{3,r} \oplus u_{3,s}$	r, s
1–2	3, 7	9–9	2, 6
1–3	6, 8	9–10	4, 9
1–7	1, 4	9–14	5, 10
2–4	2, 4	10–9	3, 10
2–5	5, 7	10–10	1, 6
2–7	6, 9	10–14	4, 8
3–3	1, 2	11–9	1, 8
3–4	8, 9	11–11	7, 10
3–7	3, 5	11–13	4, 6
4–2	9, 10	11–14	2, 9
4–6	4, 5	12–12	3, 6
4–7	2, 7	12–13	7, 8
5–1	1, 5	12–15	1, 10
5–5	2, 3	13–8	4, 10
6–2	2, 5	13–12	5, 9
6–3	4, 7	13–14	6, 7
6–5	6, 10	13–15	3, 8
6–6	1, 3	14–8	5, 8
7–1	3, 4	14–11	3, 9
7–4	1, 7	15–9	7, 9
7–6	8, 10	15–11	5, 6
8–10	2, 8	15–12	2, 10
8–13	1, 9		

lowest order bit with a value of 1 in $\Delta_2 = 0100$, i.e. $r = 2$ and d is equal to the sequence number of the lowest order bit with a value of 1 in $\Delta_1 = 0010$, i.e. $d = 3$.

2. Determination of the sequence number j of the first of the pair of adjacent distorted symbols.

The determination of the number r is determined in the following manner: if $\Delta_1 < \Delta_2$ and $\Delta_1 \neq 0$, then r is equal to the sequence number of the first non-zero bit of the code Δ_1. If $\Delta_1 > \Delta_2$ and $\Delta_2 \neq 0$, then r is equal to the sequence number of the firs non-zero bit of the code Δ_2.

The determination of j, the k-bit sequence number $(k = \lceil \log_2 n \rceil)$ of the first of the pair of adjacent distorted symbols, this is determined as follows: Denote as w_1, w_2, …, w_k the bits of the number j, so that $j = w_1 + 2 \cdot w_2 + 4 \cdot w_3 + \cdots + 2^{k-1} \cdot w_k$. Denote as $c_2, c_3, \ldots, c_{k+1}$ the values of the carry bits that arise during the augmentation of j, so that, for example c_2—corresponds to the carry bit carried over for the second bit. The bits of the parts Δ_1, Δ_2 and Δ_3 are determined in the form: $\Delta_1 = \{\lambda_1, \lambda_2, \ldots, \lambda_m\}$ $\Delta_2 = \{\mu_1, \mu_2, \ldots, \mu_m\}$, $\Delta_3 = \{\delta_1, \delta_2, \ldots \delta_{k+m-1}\}$. For $i = 1, \ldots, m$

the bit $\xi_i = 1$, if $i < d$ and $\xi_i = 0$, if $i \geq d$ (if a single symbol were distorted and for $l > m$ $\xi_l = 0$) or alternatively, the value of the bit $\alpha_l = \lambda_l \oplus \mu_l$, if $l \leq m$ and $\alpha_l = 0$, if $l > m$).

The procedure for determining $w_1, w_2, \ldots, w_k$—the bits of the number j, consists of the following steps:

1. If $\Delta_1 > \Delta_2$ (an odd numbered symbol was distorted first), then $w_1 = 1$ and $c_2 = 1$. Otherwise $w_1 = 0$ and $c_2 = 0$. Number $i = 2$, $c_1 = 1$.
2. The values of the bits of the coefficients w_{i-r+1} belonging to the number j, of the first distorted symbol, are determined using the formula:

$$w_i = \delta_{i+r-1} \oplus \xi_r \cdot c_i \oplus \delta_{i+r-2} \oplus w_{i-1} \oplus \xi_r \cdot c_{i-1} \oplus \alpha_{i+r-1} \cdot (w_1 \oplus \xi_{i+r-1})$$
$$c_{i+1} = w_i \cdot c_i$$

   If $w_i = 1$ and $c_i = 1$, then $c_{i+1} = 1$, otherwise $c_{i+1} = 0$.

3. If $i < k$, then $i = i + 1$ and return to step 2, otherwise the algorithm finishes.

For the above example $\Delta_1 = 0010$, $\Delta_2 = 0100$ and $\Delta_3 = 00100$. The value r = 2 and d = 3. As $\Delta_1 > \Delta_2$, $w_1 = 1$, $c_2 = 1$, $c_1 = 1$.

For $i = 2$: $\delta_3 = 1$, $\xi_3 = 0$, $\delta_2 = 0$, $\xi_2 = 1$, $\alpha_3 = 1$. Hence it follows that: $w_2 = 1 \oplus 1 \cdot 1 \oplus 0 \oplus 1 \oplus 1 \cdot 1 \oplus 1 \cdot (1 \oplus 0) = 1$. While $w_2 = 1$ and $c_2 = 1$, then $c_3 = 0$. $c_3 = c_2 \cdot w_2 = 1$.

For $i = 3$: $\delta_4 = 0$, $\xi_4 = 0$, $\delta_3 = 1$, $\xi_2 = 0$, $\alpha_4 = 0$. It hence follows that: $w_3 = 0 \oplus 1 \cdot 1 \oplus 1 \oplus 1 \oplus 1 \cdot 1 \oplus 0 \cdot (1 \oplus 0) = 0$. Therefore, $j = w_1 + 2 \cdot w_2 + 4 \cdot w_3 = 3$.

The correction of the distorted symbols is achieved with the corresponding expression:

$$\text{If } \Delta_1 > \Delta_2, \quad X_j = X_{jR} \oplus \Delta_1 \quad \text{and} \quad X_{j+1} = X_{j+1,R} \oplus \Delta_2$$
$$\text{If } \Delta_1 < \Delta_2, \quad X_j = X_{jR} \oplus \Delta_2 \quad \text{and} \quad X_{j+1} = X_{j+1,R} \oplus \Delta_1$$

The proposed procedure that was explained above may be used for the correction of the correction of a single erroneous symbol or a pair of adjacent distorted symbols. The total number K of the required control bits, in contrast to the Reed–Solomon codes, does not depend on the bits m of the symbols and is determined by the expression:

$$K = 3 \cdot m + k - 1 \tag{18}$$

Since for the Reed–Solomon codes it may be taken that $k = m$, then Eq. (18) becomes:

$$K = 4 \cdot m - 1 \tag{19}$$

In order to guarantee the detection of errors caused by the distortion of the timing between the transmitter and the receiver [4, 5], a method id proposed whose essence is the control of the unity bit sequence length whose transmission is not synchronized. Assume that the information consists of a packet of m bits, $B = \{b_1, b_2,\ldots, b_m\}$, $b_l \in \{0,1\}$, $\forall l = 1,\ldots, m$. In the packet it is proposed that a sequence of values is contained, whose transmission is also asynchronous. Without loss of generality, it may be assumed that a series of ones is transmitted. This case corresponds to the USB mode of operation. The symbol E_e^k is defined as the sequence of e units that may correspond with an order k in the control packet.

This implies that until the E_e^k sequence the transmitted packet consists of an order k-1 sequence of units of the same length e. Similarly, $E^k_{>e}$ signifies the kth sequence in the controlled packet of the following order, the length of which exceeds e. Assuming that the kth sequence contains $L_{ke} > e$ units and that the total number of sequences of the subsequent order that contain more than e units is K_e, then taking inti consideration the definition as it has already been stated, the total number of units in the transmitted packet may be presented in the form:

$$\sum_{i=1}^{m} b_i = \sum_{k=1}^{K_0} L_{k0}.$$

In order to guarantee the detection of errors that are caused by the violation of the timing during the transmission of a series of ones, a method is proposed that focuses in changes in the lengths of the sequences, since a timing violation may cause the length of the sequences of one to be increased or reduced at the receiver, compared to the one that was transmitted. The method involves the control of the changes in the length of the sequences for which there exists an increased risk of synchronization violation. In reality, the sequences contained in this class are the sequences $E_{>u}$, whose lengths exceed a predefined threshold u. For example, in the case of USB it is advisable that sequence length changes of sequences that contain no more than four (4) ones.

The control code V is proposed in order to calculate a weighted checksum. For this purpose, each one of the k units of the $\mathrm{E}^k_{>u}$ sequence that is controlled, is assigned a weight W_k, that uniquely corresponds to the symbol k in the sequence $\mathrm{E}^k_{>u}$ in the packet that is transmitted.

During the transmission to the receiver of the packet, the parity bit is calculated for every k sequence of length L_k. The code V that represents the weighted checksum is calculated at the receiver and the transmitter as the XOR sum of the parity bits of the length of the sequence that is being controlled and as the subsequent XOR of the produced parity bit of the length of the sequence with the corresponding weighted sum:

$$V = \bigoplus_{k=1}^{K_u} (L_{ku} \bmod 2) || \bigoplus_{k=1}^{K_u} W_k \cdot (L_{ku} \bmod 2). \tag{20}$$

If the values calculated at the sender and the receiver using expression (2) are symbolized as V_s and V_R, then the error control code is calculated as $\Delta = V_s \oplus V_R$.

If $\Delta = 0$, then it is considered that errors that alter the length of sequences of ones do not exist.

The selection of weighting coefficients W_k is performed based on the given multiplicity for guaranteed detection of errors due to timing violations. Due to the first factor in expression (2) the detection of all timing violation errors of odd multiplicity is guaranteed. For $W_k = k$, apart from errors of odd multiplicity, all double errors are also guaranteed to be detected. Indeed if the length of odd multiplicity changes, for example for the qth and the dth sequence, i.e. $E^{q}_{>u}$ and $E^{d}_{>u}$, from the sets of units of the control packet $q, d \in \{1, \ldots, K_u\}$, then the difference between the codes of the sender and the receiver will be equal to $\Delta = W_q \oplus W_d$. As $W_q \neq W_d$ which is in turn due to the fact that $d \neq q$, then $\Delta \neq 0$ and this implies that all double errors are guaranteed to be detected.

In order to detect errors that are caused by timing violations and are of even multiplicity, weighting factors that represent partially orthogonal codes may be used [3]. With reference to the K factors of the set Ω of the partially orthogonal codes of class h, is the set $\Omega = \{W_1, W_2, \ldots, W_K\}$, consisting of K codes such that for the XOR sum of any subset ϑ, containing no more than h such codes that are non-zero, it is true:

$$\begin{gathered} \forall\vartheta = \{X_1, X_2, \ldots, X_{q_\vartheta}\} \subset \Omega, q_\vartheta \leq h : \\ X_1 \oplus X_2 \oplus \cdots \oplus X_{q_\vartheta} \neq 0. \end{gathered} \tag{21}$$

In reality this means that any subset $\vartheta \subset \Omega$, $|\vartheta| \leq h$, is an orthogonal base in an h-dimensional space. In Table 4, the values of Ω are presented for $K = 64$, $h = 4$, whose usage enables the guaranteed detection of errors with even multiplicity that does not exceed 4.

The analysis of the bits referred to in Table 4 shows that the numbers with length q bits are partially orthogonal codes that are used for the guaranteed detection of errors of multiplicity h (h even) depends approximately logarithmically from the length m of the control packet [3], i.e. the dependence of the bits q_w of the weighting coefficient on the number K_u of the sequence of ones may be approximately expressed as:

$$q_W \approx a_h \cdot \log_2 K_u \tag{22}$$

The weighting factors in Table 4 are presented in hexadecimal form. The theoretical and experimental studies that have been carried out show that the actual arithmetic values for coefficients α_4 and α_6 do not remain constant and tend to decrease as the number K_u of the sequence of units of the data control packet increases. However the length of the packet remains restricted in reality, therefore the approximations that for $h = 4$ the coefficient becomes $a_4 \approx 2.3$, for $h = 6$ the coefficient becomes $a_6 \approx 3.65$ and for $h = 8$ the coefficient is $a_8 \approx 7.1$, are sensible and safe.

Table 4 Weighting factors for guaranteed detection of errors with even multiplicity <4

k	W_k	k	W_k	k	W_k	k	W_k
1	0001	17	00F7	33	0458	49	0CD4
2	0002	18	0100	34	0491	50	0D9B
3	0004	19	0117	35	0535	51	0DE5
4	0009	20	0129	36	0541	52	0EB3
5	000F	21	01C7	37	0654	53	1000
6	0010	22	0200	38	074A	54	101D
7	0020	23	0219	39	0800	55	1027
8	0033	24	022D	40	081C	56	104E
9	0040	25	0252	41	0825	57	1153
10	0055	26	0283	42	0889	58	116B
11	006A	27	0323	43	0944	59	1175
12	0080	28	035F	44	0998	60	1191
13	0096	29	03E6	45	0A3A	61	1290
14	00AB	30	0400	46	0AF0	62	1487
15	00BD	31	041B	47	0B39	63	1664
16	00ED	32	042E	48	0C37	64	16AC

A fundamental aspect of the evaluation of the effectiveness of the proposed error control method, for errors that dominate serial interfaces of information systems with asynchronous data encoding, is concerned with the calculation of the number of control bits.

It is apparent that the number s of the control bits is determined by the length of the weighting factors W which in turn depend on the number K_u of the sequences, whose transmission is not synchronized and are controlled by the proposed method. It is hence important to determine the dependence of the number K_u of the controlled sequences on the minimum length $u + 1$ and on the length m of the packet being transmitted.

Assuming that the appearance of a one—symbol and a zero—symbol in the sequence being controlled is equi-probable and independent of the remaining bits, then the probability that a sequence consists of a single "one" is determined by the fact that the next symbol must be a zero and is 0.5. The probability P_2 that the sequence consists of two "ones" is determined by the fact that the second bit needs to be "one" and the third "zero" and is equal to 0.5^2. Similarly, it may be shown that the probability of a sequence having a length of e "ones" is calculated as 0.5^e. The average length of a sequence of "ones" is defined as L_a and this may be calculated as the sum of the following series:

$$L_a = \sum_{i=1}^{m} \frac{i}{2^i} \approx \sum_{i=1}^{\infty} \frac{i}{2^i} = \frac{1}{(1 - 0.5)} = 2 \tag{23}$$

As in the assumptions for the above calculation of the average length of sequences of "ones" in the transmitted packet are balanced for "zeroes" as well, the probability distribution of the length of sequences of "zeroes" is similar to the above one for sequences of ones. Hence the mean number of sequences of "zeroes" is approximately equal to the mean of the sequences of "ones". From expression (5) it follows that the mean length of the sequence is 2 and that respectively, the mean number of transmitted packets is $m/2$, half out of which (i.e. $m/4$) are sequences of "ones". From the above analysis it may be deduced that half of the sequences of "ones" contain just one unit, ¼ contain two units etc. Hence the mean number v_e of sequences that contain e "ones" in the m bit packet are determined by the formula:

$$v_e = \frac{m}{2^{e+2}} \tag{24}$$

It follows immediately that:

$$\sum_{e=1}^{m} v_e \cdot e \approx \sum_{e=1}^{\infty} v_e \cdot e = \frac{m}{4} \cdot \sum_{e=1}^{\infty} \frac{e}{2^e} = \frac{m}{2} \tag{25}$$

The mean number of the sequences K_e that contain more than e "ones" are calculated as the sum:

$$K_e = \sum_{i=e+1}^{m} v_i \approx \sum_{i=e+1}^{\infty} v_i = \frac{m}{2^{e+2}} \cdot \sum_{l=1}^{\infty} \frac{1}{2^l} = \frac{m}{2^{e+2}} \tag{26}$$

The number s of control bits that are used depends on two factors: the length m of the control packet and the multiplicity of the errors that should be detected. Hence for the detection of all errors with odd multiplicity and with double multiplicity, the length of the weighting factor is equal to $\log_2 K_e$. More specifically, for serial USB interfaces the length of the packet has been determined at 1024 bytes (m = 8192) [1]. With the use of bit stuffing, it is necessary that changes in sequence s with lengths 1–6 are controlled. The number K of such sequences is defined as the sum:

$$K = \sum_{i=1}^{6} v_i = K_0 - K_7 = \frac{m}{4} - \frac{m}{512} \approx \frac{m}{4} \tag{27}$$

The length of the weighting factor for the guaranteed detection of double errors is calculated as $\log_2 m - 2 = \log_2 8192 - 2 = 11$ bits. Taking into account the parity bit of the total length of the transmitted packet control code as well, the total number of control bits becomes 12.

4 Evaluation of the Performance Improvement

The above imply that both the proposed method and the Reed–Solomon codes use for the correction of the distortions that appeared form a burst of errors in an impulse modulation channel approximately the same number of bits. The fundamental advantage of the proposed method is that for both software and hardware implementations, it is significantly simpler, compared to the Reed–Solomon approach. Data for the comparison of the computational complexity and the time required for the completion of the algorithms, in the case of error correction of burst errors in low frequency transmission lines, are presented in Table 1.

Comparison of the computational and temporal complexity.

The analysis presented in Table 5, demonstrates that the proposed method enables the simplification and acceleration of the error detection and correction processes by up one or more orders of magnitude, compared to the Reed–Solomon codes. This is particularly important for applications such as military applications. Especially in military embedded computational systems, where data transmission is implemented by using impulse code modulation in conditions of extremely strong electromagnetic interference, the proposed method manages to satisfy the strictest requirements in the time required for detecting and correcting errors that have appeared.

From a mathematical point of view, the proposed method is based on convergence of the correction of the burst of errors to the solution of a system of equations. Such a system may be solved algebraically without tests. This mainly reduces the computational complexity of the correction of the burst by n times compared to every cyclic code, including the RS codes.

Unfortunately, the way this problem has been set, so that the correction of the burst converges to the solution of the system of equations, is feasible only in the case of a burst.

The characteristics of the number of control bits and the capability for guaranteed detection of errors via the proposed method, to be used in current USB technology for CRC, are presented in Table 6.

The data analysis of the Table 6, shows that the developed method allows the comparison with known methods. The proposed method significantly improves the efficiency of error detection in the serial interfaces which faulted synchronization of data transmission.

Table 5 Comparison of the computational and temporal compexity for BMWC

Action	Computational complexity		Temporal complexity	
	Proposed method	Reed Solomon method	Proposed method	Reed Solomon method
Encoding	$O(4 \cdot n)$	$O(12 \cdot n \cdot m)$	$O(\log_2 n)$	$O(12 \cdot n \cdot m)$
Error detection	$O(4 \cdot n)$	$O(20 \cdot n \cdot m)$	$O(\log_2 n)$	$O(20 \cdot n \cdot m)$
Correction of burst errors	$O(8 \cdot \log_2 n)$	$O(4 \cdot n \cdot m)$	$O(8 \cdot \log_2 \cdot n)$	$O(4 \cdot n \cdot \log_2 m)$

Table 6 Comparative characteristics of efficient errors detection caused by the insertion of the sync USB

Method of errors detection	Quantity of guaranteed error detection caused by the faulted of synchronize	Quantity control bits
Proposed error control method based on the length change of asynchronous serial bits	All odd errors multiplicity	1
	Odd errors multiplicity and double errors	12
	Odd errors multiplicity and even errors multiplicity less than 6	24
CRC	Guaranteed of not detected errors	16

All the above results demonstrate that the use of the proposed techniques result in the following advantages:

- In a significant number of cases, errors are corrected by the receiver, without need for retransmissions.
- Incidents where retransmission is necessary are efficiently and effectively detected by the receiver and hence there is a significant reduction in the need for retransmissions.
- The computational cost of the operations is significantly lower than current techniques

Apart from the improvements in the channel usage that these results imply, they all result in smaller burdens for the environment, since the power required for calculations and retransmissions and the amount of electromagnetic radiation emitted on average appear significantly reduced.

5 Conclusions

This chapter has reviewed the topic of error detection and correction techniques. Such techniques have been developed that in a variety of commonly encountered conditions exhibit performance that is significantly better compared to baseline techniques.

The first approach developed enhances the efficiency of single burst error control in low frequency channels and data storage, by localizing the positions of the corrupted bits in the burst and the start point of the burst that has occurred. A conventional checksum method and the use of the weighted version of the checksum for the solution of the first and the second problem respectively. The start point of the burst is located via the solution to low complexity linear system. The number of control bits required is $2 \cdot (m - 1) + \log_2 N$, a number that represents a significant reduction in the computational complexity necessary for the correction of the burst error comparing to Reed–Solomon codes. Additionally, using a number of control bits that is significantly smaller than that of competing techniques, correctable errors are distinguished from non-correctable ones, hence eradicating the number of retransmissions.

An additional error detection and correction procedure was presented that is based on the binary multiplication without carry and expands the ideas of weighted checksum based error correction and detection. Error detection and correction using the proposed techniques was presented. Finally the expected benefits of the use of the new technique with respect to complexity reductions were presented. Both the above results imply a significant reduction of the environmental effects of transmission in terms of energy consumption and electromagnetic radiation emission.

It is expected that the proposed methods may be additionally developed to detect and correct errors with larger multiplicities and research is currently under way in this direction.

References

1. Bardis, N.G., Markovskyi, O.P., Cepisca, C., Argiri, K.: Burst error control based on weighted checksum. In: Proceeding ICCOM'10 Proceedings of the 14th WSEAS International Conference on Communications World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA ©2010
2. Mee, C.D., Daniel, E.D.: Magnetic Storage Handbook, 2nd edn. McGraw Hill, New York (1996). ISBN 0070412758
3. Bardis, N.G., Drigas, A., Markovskyi, O.P.: Performance increase of error control operation on data transmission. In: 3rd International Conference on New Technologies, Mobility and Security, IEEE. IEEE Communications Society, Egypt, Cairo, 20–23 Dec 2009
4. Klove, T., Korzhik, V.: Error Detecting Codes: General Theory and Their Application in Feedback Communication Systems, 433 p. Kluwer, Norwell, MA (1995)
5. Bardis, N.G., Markovskyy, A.P.: Utilization of avalanche transformation for increasing of Echoplex and checksum data transmission control reliability. In: ISITA 2004—International Symposium on Information Theory and its Applications, IEEE/SITA, Parma, Italy, 10–13 Oct 2004. ISBN 4-902087-08-1
6. Sklar, B.: Digital Communication. Fundamental and Application, p. 1104. Prentice Hall PTR, Englewood Cliffs (2001)
7. Nakassis, A.: Fletcher's error detection algorithm: how to implement it efficiently and how to avoid the most common pitfalls. In: ACM SIGCOMM Computer Communication Review, vol. 18, Issue 5, pp. 63–88, Oct 1988
8. Maxino, T.C., Koopman, P.J.: The effectiveness of checksums for embedded control networks. IEEE Trans. Dependable Secure Comput. **6**(1), 59–72 (January–March 2009)
9. Sweeney, P.: Error Control Coding From Theory to Practice. Wiley, New York (2002)
10. Moon, T.K.: Error Correction Coding: Mathematical Methods and Algorithms. Wiley, New York (2005)
11. Mahan, R.E.: The Data Link Layer—Error Detection & Correction (2002)
12. Tanenbaum, A.S.: Computer Networks, 4th edn. Prentice-Hall PTR, Englewood Cliffs (2003)
13. Sklar, B.: Digital Communication, Fundamental and Application, p. 1104. Prentice Hall PTR, Englewood Cliffs (2001)
14. Bardis, N., Drigas, A., Doukas, N., Karadimas, N.: Optimised information system reliability techniques for knowledge society acceptance. Int. J. Knowl. Learn. **5**(3/4), 207–221 (2009)
15. Blahut, R.E.: Theory and Practice of Error Control Codes, 452 p. Addison-Wesley Publishing Company, Reading (1983)

16. Clark, G.C., Cain, J.B.: Error-Correction Coding for Digital Communications, p. 349. Plenum, New York (1981)
17. Stone, J., Greenwald, M., Patrige, C., Hughes, J.: Performance of checksum and CRC's over real data. IEEE/ACM Trans. Networking **6**(5), 529–543 (1998)
18. Feldmeier, D.C.: Fast software implementation of error detection codes. IEEE/ACM Trans. Networking **3**(6), 640–651 (1995)
19. Bardis, N.G., Markovskyy, A.P.: Utilization of avalanche transformation for increasing of echoplex and checksum data transmission control reliability. In: ISITA 2004—International Symposium on Information Theory and its Applications. IEEE/ISITA, Parma, Italy, 10–13 Oct 2004. ISBN 4-902087-08-1
20. Desaki, Y., Kazuhiko, I., Miura, Y., Yokota, D.: Double and triple error detecting capability of internet checksum and estimation of probability of undetectable error. In: PRFTS, 1997 Pacific Rim International Symposium on Fault-Tolerant Systems (PRFTS'97), pp. 47–51 (1997)
21. Mezentseva, O., Alekseev, A.: Regarding the (N, k)-threshold schemes realization based on modular arithmetic algorithms. J. Appl. Math. Stat. Inform. (JAMSI) **4**(2) (2008)
22. Bardis, N.: Coding of checksum components for increasing the control reliability of data transmission for military applications. WSEAS Trans. Commun. **11**(7), 1122–1131 (2008)
23. Doukas, N., Bardis, N.: Effectiveness data transmission error detection using check sum control for military application. In: Proceedings of the 10th WSEAS International Conference on Mathematical Methods, Computational Techniques and Intelligent Systems, pp. 498–502, Corfu, Greece (2008). Available online: http://portal.acm.org/citation.cfm?id=1562334.1562421
24. Bardis, N.G., Doukas, N, Markovskyi, O.P., Drigas, A.: Two level efficient user authentication scheme. In: 2010 4th IEEE International Conference on Digital Ecosystems and Technologies (DEST), pp. 470–474. IEEE (2010)
25. Bardis, N.G., Markovskyi, O., Doukas, N.: Efficient burst error correction method for application in low frequency channels and data storage units. In: 2011 17th International Conference on Digital Signal Processing (DSP), pp. 1–6. IEEE (2011)
26. Bardis, N.G., Doukas, N., Markovskyi, O.P.: Burst error correction using binary multiplication without carry. In: Military Communications Conference, 2011-MILCOM 2011, pp. 305–309. IEEE (2011)

Part IV
Modeling and Assessment of Green Computer Systems and Infrastructures

Model-Based Evaluation of Energy Saving Systems

Davide Basile, Felicita Di Giandomenico and Stefania Gnesi

Abstract Nowadays, there is a great attention towards cautious usage of energy sources to be employed in disparate application domains, including critical infrastructures, to save both in financial terms and in environmental impact. This chapter focuses on stochastic model-based as a support to the analysis of energy saving systems, in combination with other non functional properties, such as reliability, safety and availability. We discuss general guidelines to build a model-based framework to analyse critical cyber-physical systems, where effective energy consumption is required, while assuring imposed levels of resilience. Also, an overview of the most commonly employed methodologies and tools for model-based analysis is provided, and extensive literature is indicated as pointers to relevant research activities performed on this attractive topic over the last decades. Finally, in order to corroborate the proposed framework, a case study in the railway domain is proposed. By adopting the Stochastic Activity Networks formalism, the framework is instantiated to analyse effective trade-offs between energy consumption and satisfaction of other dependability related requirements.

Keywords Energy-saving · Reliability · Quality models · Stochastic analysis

D. Basile (✉) · F. Di Giandomenico · S. Gnesi
National Research Council (CNR), Institute of Information Science
and Technologies (ISTI), via G. Moruzzi 1, 56124 Pisa, Italy
e-mail: davide.basile@isti.cnr.it

F. Di Giandomenico
e-mail: felicita.digiandomenico@isti.cnr.it

S. Gnesi
e-mail: stefania.gnesi@isti.cnr.it

V. Kharchenko et al. (eds.), *Green IT Engineering: Concepts, Models, Complex Systems Architectures*, Studies in Systems, Decision and Control 74,
DOI 10.1007/978-3-319-44162-7_10

1 Introduction

Energy management and the related environmental issues are emerging as a relevant technological, political, and societal concern. Reducing energy consumption is nowadays one of the primary goals in the design of IT systems [1]. The benefits of adopting strategies of energy management are many, as for example: *reduction of costs*, by adopting strategies of energy management most organisations can save in economic terms; *mitigating adverse environmental effects*, which helps organizations in gaining a good reputation and avoid penalties of governments aiming at reducing the environmental pollution; *reduction of risks* due to energy shortages and prices fluctuations.

For being effective, policies of energy saving need to be integrated in the design of an energy-saving system. When critical systems are considered, that is employed in sectors where *dependability* requirements are stringent, the challenge of energy saving needs to be mediated with satisfaction of properties such as *reliability*, *safety*, *availability*. Unfortunately, reducing the energy consumption and increasing reliability are in general contrasting requirements. Hence, it is important to find a good trade-off between the two aspects. To this purpose, resorting to *quantitative assessment* of policies devised for energy management is extremely helpful to support the choice of the one that better suits the overall set of system requirements. Stochastic model based analysis, which is widely recognized as a powerful approach for quantitatively assessing non functional properties since the early phases of a system development, is addressed in this work.

In particular, this chapter offers the following contributions. First, guidelines to the set up of a stochastic model-based framework, suitable to evaluate trade-offs between energy consumption and dependability-related properties in *critical energy saving cyber-physical systems*, are discussed. The targeted systems are characterised by physical entities that are controlled by computational units, where an energy management policy is in place to supply the physical components as needed to reliably perform their work. Comparison among different energy management strategies is possible, in terms of indicators representative of the energy consumption and dependability-related indicators, to select the most effective one.

Second, a review of widely adopted formalisms and tools for stochastic model based evaluation is presented, to provide a basis of existing instruments to implement the proposed analysis framework.

Third, a realistic case study is proposed to exploit our modelling framework, which is taken from the railway transport domain, a critical domain where failures can lead to catastrophic consequences, as for example the derailment of trains. In particular, we evaluate policies of energy consumption for a system of railroad switch heaters. By adopting the proposed method, evaluation of the measures of interest is used for tuning the parameters of the considered energy-saving policies in order to reduce energy consumption while satisfying reliability constraints.

Related work
Model-based analysis of energy-saving reliable systems is recently gaining interest from the research community, hence the literature concerning this subject is relatively heterogeneous; in the following we briefly review recent works that analyse and optimise the energy consumption in several application domains. Generalized Stochastic Petri Nets [2] are used to solve the dynamic power management problem for systems with complex behaviour in [3]. Dynamic power management addresses reduction of power dissipation in embedded systems, with a selective shut-off or slow-down of system components that are idle or underutilized. A time-out policy is used for power saving, which turns on a component when it is used and turns it off when it is not used for a certain amount of time. Comparisons are also performed with other models based on Markov Decision Processes (MDP) [4]. The dynamic power management problem is interpreted as a hybrid automaton control problem and integrated stochastic control in [5]. Hybrid automata mixed both a discrete state, representing the power mode of the system, and a continuous one, representing the consumed power. Two strategies are compared: on demand wake-up of a component (that was previously turned off) and pre-emptive wake-up. The former provides better results for conservation of energy and prevention of latency. The applicability of self-organizing systems for different fields of power system control is discussed in [6]. Agent-based decentralized power flow control is compared with current practice based on central decision making. The authors study how to balance the voltage and frequencies stability of the network to meet the demand of energy. These parameters are linked to reliability and safety of the system. It is shown how a decentralized control can improve reliability, safety and efficiency by providing a real-time adaptivity to changes in the network (failure of a node, blackout). The survivability of a smart house is analysed in [7], that is the probability that a house with locally generated energy (photovoltaic) and a battery storage can continuously be powered in case of a grid failure. Hybrid Petri Nets [8] are used for modelling this scenario. Different strategies of battery management are considered. In the first one, all the battery is consumed when needed, in the second there is a minimum threshold of energy saved in case of grid failure. In the third case the battery is also charged to a maximum threshold when the grid is operating. It is shown how the third strategy is better both for the local usage of energy and for the survivability of the smart house. The authors consider a randomly chosen probability of failure and fixed thresholds. The trade-off between energy saving and reliability is studied in [9], by managing frequencies and voltage of the delivered energy. In particular, lower frequencies result in higher reliability while for voltage scaling the reliability decreases dramatically. In our approach the energy consumption is managed by changing the power consumed by the system.

Power management in smart grids is discussed in [10], where it is shown that, by adjusting the power supplied to the different clients, it is possible to obtain a good trade-off between power utilization and customers satisfaction. Services negotiation of energy and reliability requirements is the selected case study in [11], where an energy provider, an energy consumer and a mediator try to find an agreement on the amount of energy delivered, its reliability and price. In Sect. 2.3 we will review some of the previously discussed models for energy assessment.

Structure of the chapter.

Section 2 discusses general guidelines for supporting the design and evaluation of energy-saving systems and provides a brief survey on the modelling methodologies and tools for evaluating dependability measures and energy consumption of systems. A case study is proposed in Sect. 3 where the proposed method is applied for evaluating reliability and energy consumption indicators for the system under analysis. Conclusions are discussed in Sect. 4.

2 Supporting the Design of Energy Saving Systems

In this section, we introduce general guidelines toward the building of a modelling framework to support the design of critical energy saving systems. To provide the context useful to understand the developments, background on model-based analysis is firstly introduced. Then, a general overview of the proposed modelling framework is presented, followed by an overview of the pertinent formalisms and tools from the literature. To concretely demonstrate the potentialities of the proposed approach, instantiation of the general framework in a use case from the railway domain is then addressed, as extensively described in Sect. 3.

2.1 Model-Based Analysis

Quantitative assessment of non-functional properties, especially dependability and performance related ones, is an important activity when developing systems to be employed in critical domains, where strict requirements on these properties need to be met. Such evaluation typically consists in probabilistically estimate the occurrence of faults and their impact on the ability of the system to operate correctly, in presence of possible measures to enhance system resilience (such as fault tolerance mechanisms). Several approaches are available in literature to perform system assessment [12], mainly testing, fault injection and model-based evaluation. For quantitative evaluation of dependability indicators, stochastic model-based analysis [13] has been proven to be particularly useful, versatile and cost-effective for manufacturers [14, 15]. To keep the model manageable, the system needs to be represented at a properly identified abstraction level. Indeed, depending on the properties to be analysed the emphasis on the system representation is focused on those aspects that are relevant for the analysis purposes, while irrelevant aspects are neglected. Therefore, a wide variety of models are used in practice, to tailor the right abstraction level for the system under analysis, in accordance with the properties to be assessed, the desired degree of accuracy and available resources to manage models development and solution.

Stochastic model-based approaches [13] are useful to support the development of systems, in all the phases of their life cycle. In the early design phases, they can be helpful for avoiding waste of time and resources in the development phase. This can be done by pointing out the properties and the requirements that the system must satisfy, as for example the energy consumed, and build a model that captures the relevant structural and behavioural aspects of the system under development.

Such early modelling phase is therefore exploited to highlight problems in the design of the system, such as bottlenecks, to perform comparisons of different alternative architectural solutions and selection of the most suitable one. For an already existing system, an a posteriori analysis of properties such as dependability or performance is useful to understand and learn about specific aspects, to detect possible design weak points, to perform a late validation of the dependability requirements, and it is especially useful to improve the system in its future releases. Moreover, with a model-based analysis it is possible to predict future behaviours to plan the maintenance and the upgrading of the system.

2.2 Guidelines to Model Critical Cyber-Physical Energy Saving Systems

By resorting to a model-based approach, we propose general guidelines to develop a modelling framework for supporting the design of critical energy saving systems. We target energy consumption in cyber-physical systems (CPS) [16], where physical entities are controlled by computational units. Among the functionalities under the supervision of the cyber-control, there is the strategy for supplying energy to components of the physical system, necessary to keep them effective and reliable in the service they accomplish. Our interest is in assessing measures that are representative of the energy consumption, to be combined with other dependability-related properties dictated by the critical domain the CPS is employed in. The results from this kind of analysis are very helpful for: (i) gaining insights on benefits from different energy supply strategies and (ii) properly tuning the parameters of these strategies toward most rewarding configurations.

A modular and compositional framework is pursued, to promote wide applicability to a variety of scenarios relevant to cyber-physical systems, as well as ability to undergo future refinements and extensions. A diagram of the proposed framework is depicted in Fig. 1. The proposed analysis framework is built around three major modules, as outlined in the following.

- *Physical-aspects module*: this module focuses on the physical components of the system and on their characterisation in terms of relevant aspects from the energy viewpoint. It includes models representing phenomena related with energy supply, which depend on the fabric of the supplied components and

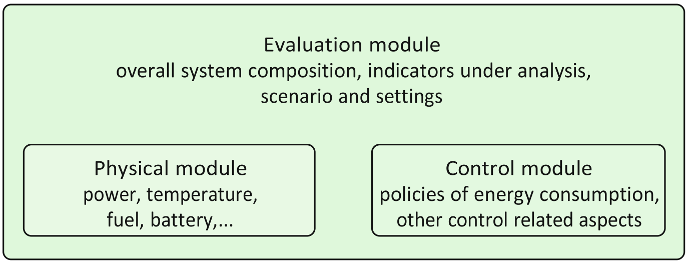


Fig. 1 The proposed analysis framework for energy saving CPS

environmental conditions impacting on the energy consumption. Examples are: (i) internal and external temperatures, properly modelled taking into account their evolution in time, given the different means involved (such as iron or copper for the physical components, winter or summer days for the external air); (ii) fuel consumption, represented by properly considering the engine parameters, aerodynamic drag, weight, and other relevant parameters; (iii) supplied power, represented by properly considering the laws regulating the involved real process; (iv) battery charging and others.

- *Control-aspects module*: this module deals with the policies that dynamically regulate the energy consumption of the physical components. To manage potential complexity while assuring adequate accuracy of the analysis, the representation of such policies is abstracted at the level of their impact on the energy parameters of the controlled physical components. As already mentioned, the primary objective of the proposed model-based approach is to assist the system designer in identifying the best policy to employ among several alternatives, in accordance with pre-established dependability and cost requirements. Trade-offs between energy consumption and dependability requirements are mandatory in critical domains, where, e.g., the energy should not be reduced in safety critical situations. A typical family of energy supply algorithms is of kind on-off, where energy is supplied or turned off on the basis of values assumed by parameters that depend on physical conditions of the system and of the environment.
- *Evaluation module*: Finally, the third module deals with the composition of the several models from the previous two modules, to end up with the overall evaluation framework. Exercising the overall composed model, energy supply policies trading energy saving and dependability properties can be quantitatively evaluated and compared in terms of properly defined indicators.

The generality of the above outlined approach allows assessing a variety of measures of interest to final customers, service providers and operators, in accordance with the specific application domain where the CPS system under analysis is utilised. Given the aim of trading energy consumption with dependability, typical indicators are energy supplied to individual system components or to the overall system in a certain time interval, as well as failure probability of an individual component or of the overall system due to lack of supplied energy.

In Sect. 3 we will provide an example of an energy saving system designed with this method [17, 18].

2.3 Methodologies and Tools for Model-Based System Analysis

Assessing the ability of a developed system to fulfil its requirements (both functional and non-functional) is a fundamental phase in the system engineering development process. As already mentioned, model-based evaluation is a versatile support since the early phases of system development and a large variety of modelling formalisms and tools have been developed over the past decades. The interest towards energy consumption analysis gained great attention in very recent years, so there is a relatively limited number of works addressing this aspect. The review below offers a reference to existing modelling formalisms and tools which can be fruitfully employed also to this purpose.

In particular, we briefly describe a subset of *state-space oriented* modelling formalisms, that are used for quantifying non-functional properties (typically, dependability and performance indicators) at different stages of a system development process.

Both *deterministic* and *stochastic* modelling formalisms have been defined, to properly address the different deterministic or probabilistic behaviours of systems to be analysed. Concerning the evaluation of energy saving systems and dependability attributes, stochastic models are more adequate. Indeed, they are capable of expressing the uncertainty about the occurrence of events as failures, and for modelling external factors affecting the energy consumption, as for example weather conditions.

According to the underlying stochastic process, stochastic models can be further classified into *Markovian* and *non-Markovian* [19]. Markovian models are those satisfying the *Markov property*, that is the conditional probability distribution of future states (conditional on both past and present values) depends only upon the present state. Indeed, in Markovian models the occurrence of events always has an exponential distribution of time. For modelling systems without this property non Markovian models are used. For the discussed modelling methodologies, several tools are available for building and solving them [20–22].

2.4 Markovian Models

Markovian models are stochastic models where it is assumed that future states depend only on the present state. Among the others we discuss *Markov chains, Stochastic Petri Nets, Generalised Stochastic Petri Nets*.

Markov Chains

Markov chains [19] are capable of modelling several types of systems and they are equipped with efficient solution algorithms implemented in many automated tools. With this formalism the states of a system are modelled with a random variable $X(t)$ that changes through time. Thanks to the Markov property, the distribution for this variable depends only on the distribution of the previous state.

When only the elapsed time is important and not the actual instance of time, then the Markov chain is said to be homogeneous; otherwise, it is said to be non homogeneous. If the parameter t that indexes the Markov chain is continuous, we have a continuous-time Markov chain; otherwise, we have a discrete-time Markov chain.

A graphical representation of a Markov chain is a directed graph called *state transition diagram*, where nodes represent states and arcs represent state transitions. Because of the memoryless property, each transition occurs in an exponentially distributed time, and the rate is exactly the inverse of the expected time to the transition, that is, the rate of the corresponding exponential distribution. This is the most severe constraint that limits the applicability of Markov chains. Markov chains can be solved in terms of transient or steady-state analyses. Once the state occupation probabilities of the Markov chain are obtained, the values of the most important dependability measures can be evaluated [14]. A reward structure can also be defined, which assigns reward values to the states (or to the transitions) of the Markov chain model, obtaining the Markov reward models [23], which among the others can be used to evaluate the energy consumption of systems.

Stochastic Petri Nets

Petri Nets are widely used as modelling paradigm thanks to the efficiency in describing the qualitative and quantitative aspects of complex systems and because of their intuitive and appealing graphical representation. Originally Petri Nets were developed for representing in a compact way concurrency among processes [24].

Stochastic Petri Nets [25] are a popular timed extension of place-transition Petri Nets [24], where places are represented by circles and timed exponential transitions by white or black rectangular boxes connected by directed arcs. Arcs have positive integer weights associated and are of two types: input arcs connect input places to transitions while output arcs connect transitions to output places. Places contain tokens and the whole state of the net, called *marking*, is represented by the number of tokens in each place. Each transition has associated a random firing delay with negative exponential distribution time (Markov property).

When the marking of the input places matches the weights on the input arcs then the transition is enabled, and a timer starts counting the sampled time. If the transition remains continuously enabled the whole time, it will fire by removing the tokens from the input places and adding them to the output places according to the weights of the arcs. Once the marking changes, some transitions can be enabled and others aborted. The set of all possible reachable markings is called the *reachability set*. The *reachability graph* describes all the possible evolutions of the network starting from the initial marking, where nodes are possible markings and arcs are transitions causing the change of marking.

The reachability graph of a network is isomorphic to the correspondent continuous time Markov chain, where states and transitions are in one-to-one correspondence with those of the reachability graph, and the rates of the transitions are equal to the firing rates of the transitions of the network which cause the change of markings. Hence, in order to solve a stochastic Petri Nets it suffices to analyse the associated Markov chain, and the corresponding reward structure [23].

Generalized Stochastic Petri Nets

Generalized Stochastic Petri Nets [2] enrich stochastic Petri Net with instantaneous transitions and inhibitor arcs. The first, graphically represented as thin bars, once enabled fire in zero time. Conflicts among instantaneous transitions are solved by associating priorities and weights to transitions with the same priority. Inhibitor arcs, graphically depicted as arcs terminating with a circle, inhibit the firing of the connected transition when the input place of the inhibitor arc has one or more tokens. The reachability graph of this type of networks is not in correspondence with Markov chains because of the instantaneous transitions. Indeed, now the network spends zero time in some markings, which are called *vanishing*. Nevertheless, it is possible to eliminate those vanishing markings by performing some approximations, in order to obtain a reduced reachability graph, which is isomorphic to a Markov chain. The reward structure can be associated to the underlying Markov chain (Markov reward models) [23].

2.5 Non-Markovian Models

In Generalised Stochastic Petri Nets, when transitions with non-exponential duration are represented (i.e. instantaneous activity) then an approximation is introduced in order to solve them. This constraint (Markov property) may be too strict for certain classes of systems, and for them several classes of non-Markovian models have been introduced in literature [26], for which it is possible to represent generally distributed activities and that can be solved by simulation, or analytically; by approximating non-exponential random variables [13, 26, 27] with exponential ones. Here we discuss *Semi-Markov Stochastic Petri Nets* and *Stochastic Activity Network*.

Semi-Markov Stochastic Petri Nets

Semi-Markov Stochastic Petri Nets are an extension of Generalised Stochastic Petri Nets for which the firing time of a transition is generally distributed (non-Markovian property), and with the *resampling memory policy* for transitions, where each time a transition fires the remaining firing time of the other transitions is resampled. These nets are in correspondence with Semi-Markov Stochastic

Processes [28, 29], where the holding time of states is generally distributed and the instants of time a timed transition fires are regeneration points. Indeed, after a regeneration point the future of the marking process becomes a replica of the process at time zero. The evolution of the Semi-Markov Stochastic Process is studied for transient and steady-state analyses [29].

Stochastic Activity Network

Stochastic Activity Networks [30] are a formalism widely used for performance, dependability and performability evaluation of complex systems. The SAN formalism is a variant of Stochastic Petri Nets [25], and has similarities with Generalised Stochastic Petri Nets [2]. A SAN is composed of the following primitives: *places, activities, input gates* and *output gates*. Places and activities have the same interpretation as places and transitions of Petri Nets. Input gates control the enabling conditions of an activity and define the change of marking when an activity completes. Output gates define the change of marking upon completion of the activity. Each enabled activity may complete. Activities are of two types: *instantaneous* and *timed*. Instantaneous activities complete once the enabling conditions are satisfied. Timed activities take an amount of time to complete following a general temporal stochastic distribution function. An enabled activity is aborted, i.e. it cannot complete, when the SAN moves into a new marking in which the enabling conditions of the activity no longer hold. Cases are associated to activities, and are used to represent probabilistic uncertainty about the action taken upon completion of the activity. When an activity completes, the following steps are executed:

- one of the cases of the activity is chosen according to its marking-depending probability;
- the function of each input gate of the activity is executed;
- the function of each output gate linked to the case selected at first step is executed.

The primitives of the SAN models are defined using C++ code.

2.6 *List of Tools*

We now overview a small subset of the vast body of tools for automatising the description and evaluation of stochastic models.

- SHARPE (Symbolic Hierarchical Automated Reliability and Performance Evaluator, http://sharpe.pratt.duke.edu/node/1) [27] is a toolkit that provides a specification language and solution methods for models described as, among the others, Markov chains and Generalized Stochastic Petri Nets. Steady state, transient and interval measures can be computed. For most of the model types, SHARPE provides more than one analysis algorithm from which the user can choose. SHARPE allows the user to combine models in a hierarchical fashion. The model types include state-space ones such as Markov and semi-Markov reward models and Stochastic Petri Nets;

- SPNP (Stochastic Petri Net Package) [31] supports models as Stochastic Rewards Petri Nets and Generalised Stochastic Petri Nets. It supports transient and steady-state analytical evaluations for Stochastic Rewards Petri Nets models and discrete-event simulation for non-Markovian models. More informations at: http://people.ee.duke.edu/%7Ekst/chirel/IRISA/spnp.html;
- SURF-2 (http://homepages.laas.fr/surf4tst/what-uk.html) is a dependability evaluation tool for hardware and software systems, based on strict construction, validation and numerical resolution of Markov models. The available modelling formalisms are Markov chains and Generalized Stochastic Petri Nets. Reward structures are used to obtain measures of dependability, performance or cost. It supports transient and steady-state analysis;
- GreatSPN (http://www.di.unito.it/%7Egreatspn/index.html) [32] supports formalisms as Generalized Stochastic Petri Nets and Stochastic Well-formed Nets, and it allows the composition of different models. An algorithm for the fast computation of performance bounds based on linear programming techniques is available; together with algorithms for the analysis of stochastic well-formed nets. Steady state and simulation solvers are available;
- Oris (http://www.oris-tool.org/) [33] is a tool which supports a variety of timed extensions of Petri Nets, and it implements a symbolic state space analysis which enables an integrated approach to qualitative verification and quantitative evaluation;
- Möbius (https://www.mobius.illinois.edu/) [34] is a tool that supports various formalisms such as SAN, PEPA (a process algebra), Fault Tree, and different analytical and simulative solvers. Möbius can be used for studying the reliability, availability, and performability of systems. It follows a modular modelling approach, with proper operators *Rep* and *Join* to compose atomic models into an overall composed model.

3 A Case Study of Energy Saving System

In this section, we tailor the general method to build evaluation frameworks discussed in Sect. 2.2 to a scenario of energy saving in the railway domain. Following the spirit of the proposed method, a modular and parametric approach is followed, to assure usability of the developed analysis framework in a variety of system configurations, as well as to promote extension and refinements of the model itself to account for further involved aspects/phenomena and so enhance its adherence to sophisticated and realistic implementations with respect to the current version.

Specifically, the considered case study is a railroad switch heating system. A railroad switch is a mechanism enabling trains to be guided from one track to another. It works with a pair of linked tapering rails, known as points. These points can be moved laterally into different positions, in order to direct a train into the straight path or the diverging path. Such switches are therefore critical components in the railway domain, since reliability of the railway transportation system highly depends on their correct operation, in absence of which potentially catastrophic consequences may be generated.

Unfortunately, during winter, snow and ice can prevent the switches to work properly, exposing the railway network to potential failures. In the past, the switches were kept clean manually by employees who were sweeping the snow away. Nowadays, heaters are used so that the temperature of the railroad switches can be kept above freezing. The heaters may be powered by gas or electricity; for simplicity, in the next description we refer to electrical supply.

Different policies may be adopted to power the heaters, as for example to heat a selection of switches for a given amount of time or to heat all the switches together. Of course, different policies imply different energy consumption and different probabilities to expose the switches to a malfunction because of lack of sufficient heating.

The proposed analysis contributes to gain insight on the interplay between energy consumption and reliability in order to select an appropriate policy for the heating of the switches, which guarantees a satisfactory trade-off. In particular, the (electric) energy consumption of the system is the amount of power consumed by the system in a unit of time, that is measured in Watt per hours, while reliability is defined as the continuity of correct service, i.e. the ability of the system to avoid service failures that are more frequent or more severe than is acceptable [12].

In the following, we describe the steps to set up and exercise the modelling framework for the evaluation of reliability and energy consumption indicators for a system of (remotely controlled) railroad switch heaters. An on-off policy is considered for heating the switches, with parametric thresholds representing the temperatures triggering the activation/deactivation of the heating. The management of the heaters is automatic, and is remotely controlled by a central computational unit. In a railway station, there are tracks that are less important than others, for example the side tracks. In case of extremely cold conditions, the total amount of energy available could not be sufficient to heat the overall system, hence it is important to duly choose the heaters that must be primarily heated and those that may be heated later on. Those railroad switch heaters whose temperature cannot be kept above the freezing thresholds will experience a failure.

Since the analysed system is characterized by stochastic phenomena that cannot be accurately represented by the exponential distribution, (e.g. the time regulating changes of the temperature of the switches), we have resorted to non-Markovian models, in particular the ones defined by Stochastic Activity Networks (SAN) [30], and the Möbius tool (see Sect. 2.3) for evaluating the measures of interest. We remark that the proposed framework allows to model and analyse the system with different formalisms; however due to lack of space we only resort to SAN models.

3.1 Physical-Aspects Module

We briefly describe the real-world devices that constitute the physical components of the system under analysis. The heating system for a single heater consists in a series of tubular flat heaters along the railroad track, which warm up the railroad by induction heating. To accomplish its task, the railroad switch heater system reads through sensors the temperatures of the air and of the railroad [35].

The physical behaviour of the railroad is modelled in terms of temperature decay and increase, when the heating is switched off and on, respectively. For the temperature of the air, statistic data about cold winter nights are used, while for modelling the internal temperature of the device and to estimate the energy consumption of the heater, the exchange of heat through convection is modelled. Indeed the railroad gets cooled by the external temperature and warmed by the heaters.

To make the needed calculation we consider the portion of the railroad track to be heated, which for simplicity is represented by an iron bar. We assume that the bar is exposed to the external temperature both from the top and from the bottom.

The heater is represented by a resistance that passes through the railroad in different points in order to warm up the iron. The set-up for the heating device is based on patents of heating switches [36], which also contain data about the power consumed by a single heater and the increment of the temperature of the track in cold winter nights. We assume that the power used by the heater is constant, in order to estimate the kilowatt per hours consumed during the time interval that we consider.

Every hour new data for the internal temperature of the railroad track are computed. Assuming that the values of the external temperature of the surrounding area T_e and the previous internal temperature T_i are known, we foresee the updated internal temperature T of the railroad by solving the differential equation (balance of energy) representing the exchange of heat by convection [37]:

$$mc\frac{\partial T}{\partial t} = -uA(T - T_e) + Q^{\cdot}$$

where u is the coefficient of convective exchange, c the heat capacity of iron; A the surface area exposed to the external temperature; t the interval of time, one hour in our case; m the mass of the iron bar; $Q^{\cdot}$ the power used when the heater is turned on, if the heater is turned off this value will be zero.

We remark that in the physical phenomenon we have studied (i.e. heating exchange), the unknown variable is the internal temperature at instant of time $(k + 1) * t$, while the external temperature and the internal temperature at time $k * t$ are known.

3.2 Control-Aspects Module

We identify two main logical components in the control subsystem: the *heater* and the *central coordinator*. The network of heaters is realised by replicating the heater component, and the activation/deactivation of each heater is controlled by the central coordinator. We now present these two main components.

- *Heater*: we based the policy employed to activate/deactivate the heating on two threshold temperatures:
  - *warning threshold* (T_{wa}): this temperature represents the lower temperature that the track should not trespass. If the temperature is lower than T_{wa}, then the risk of ice or snow can lead to a failure of the railroad switch and therefore the heating system needs to be activated;
  - *working threshold* (T_{wo}): this is the working temperature of the heating system. Once this temperature is reached, the heating system can be safely turned off in order to avoid an excessive waste of energy.

  The energy consumption of the overall system depends on the value of T_{wa} and T_{wo}. Reducing the gap between these thresholds will result in a more frequent activation of the heating system, but for a shorter period of time.
- *Coordinator*: the coordinator collects the requests of activation from the pending heaters, and it manages the energy supply according to a FIFO prioritized order. Indeed, the first heater that asks to be turned on will be the first to be activated. We assign priorities to switches based on their criticality on the track; the purpose of considering priorities is to guarantee higher reliability to those switches that are vital for the correct functioning of the overall station. The percentage of heaters that can be turned on at the same time is called NH_{max}. This value represents the maximum amount of energy deliverable by the system, and cannot be exceeded. If there is no energy available, each request will be enqueued in the queue of pending heaters.

We now give details about the exchange of messages between the network of heaters and the central coordinator:

- *Heater*: At starting time, each heater h_i is switched off and its internal temperature is above T_{wa}. Once the internal temperature goes below T_{wa}, h_i asks the coordinator to be turned on and waits. When notified, h_i is turned on. After that, two events can happen:
  - the heater h_i reaches an internal temperature above T_{wo}, communicates to the central coordinator the termination of the heating phase and is switched off;
  - a second component h_j with a higher priority asks to be turned on. The energy delivered to h_i is turned off, even though it has not yet reached an internal temperature above T_{wo}. If the temperature is below T_{wa}, h_i will issue a new request of activation to the coordinator.
- *Coordinator*: at starting time, the central coordinator is waiting for a message from one of the heaters h_i in the network. Two messages can be received:

- h_i asks to be activated. This request is inserted in the queue of pending requests in case there is no energy available and the priority of h_i is not higher than that of the already activated switches. Otherwise, the request is accepted and the coordinator switches to a busy state. In this state, two events are possible. In case there is energy available, h_i will be activated by issuing a notification. Otherwise, if no energy is available but h_i has a priority higher than one of the activated heaters, firstly the heater with lower priority will be turned off with a negative notification, and then the activation is notified to h_i;
- h_i asks to be deactivated. After the deactivation, if there are no heaters that are waiting for being activated then no action is performed. Otherwise, one of the pending heaters h_j (the first in the queue of pending heaters with higher priority) is activated by issuing a notification to it.

SAN models

The overall model is obtained by the composition of the atomic models, using the *Join* and *Rep* operators (see Sect. 2.6), as shown in Fig. 2.

The atomic model *Coordinator* represents the central coordinator. In addition to the models representing the physical and heating policy of the railway switch heating system, other models accomplish supporting operations. The atomic SAN models *ProfileSelector*, *LocalitySelector*, and *SwitchIDSelector* represent, respectively, the selector for the weather profile, the location of the switch and the unique identifier of each switch. The SAN model *RailRoadSwitchHeater* represents the railroad switch heater. The submodel *HeaterModuleM* represents an instance of a single heater module, obtained by the composition, using the join operator, of the four atomic SAN models. Those atomic models share the places relative to the locality of the device, its weather profile and the unique ID. The submodel *HeatersNetM*, obtained by replicating *numRep* times the model *HeatherModuleM*,

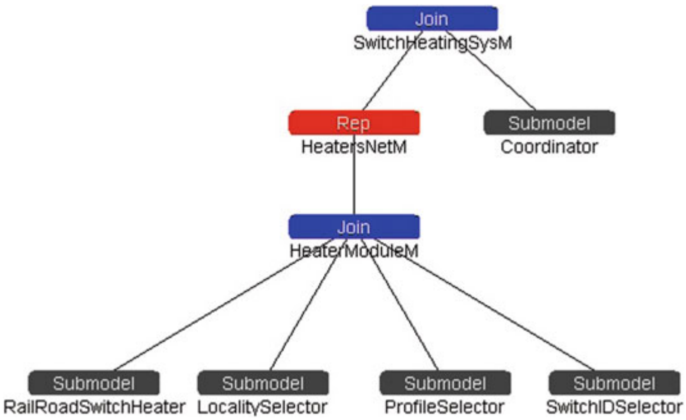


Fig. 2 The composed model

represents the network of heaters, where the parameter *numRep* identifies the number of devices composing the network. Finally, the model *SwitchHeatingSysM*, obtained using the join operator, represents the overall system. Indeed all the submodels share the same coordinator.

All these SAN models interact through *shared places*, a feature available in the Möbius tool [34] for joining different SAN models thus allowing composability.

Railroad Switch Heater SAN model

In Fig. 3 the SAN model representing the railroad switch heater is depicted. We identify three logical components inside this SAN model: the *init sub-net*, the *clock sub-net* and the *heater sub-net*.

Init sub-net. The *init* subnet initialises the C++ data structures based on the values of the places *locality*, *profileID*, *SwitchID* shared with the SAN selector models (there is a bijection between the marking of the network and the data structures used in C++), and assigns a priority to the railroad switch heater.

Clock sub-net. The clock sub-net models the evolution of time (during one day in our analyses), and it is used to load the external temperature and the temperature of the railroad track. We consider as unit of time one hour. The activity *clock* has a deterministic distribution of time (non-Markovian) and completes each hour and updates (among the others) the places *TimeOn* and *TimeReady* of the heater subnet that are respectively counting the time that a heater is activated and the time that a heater is waiting. Moreover, when *clock* completes, the place *Temperature* is updated: if the heater is turned on then the temperature increases, otherwise the temperature will be updated according to the temperature of the environment. The function representing the heating exchange (see 3.1) is defined in C++, and it is called by the output gate $O1_{clock}$ of this SAN model to update the temperature of the

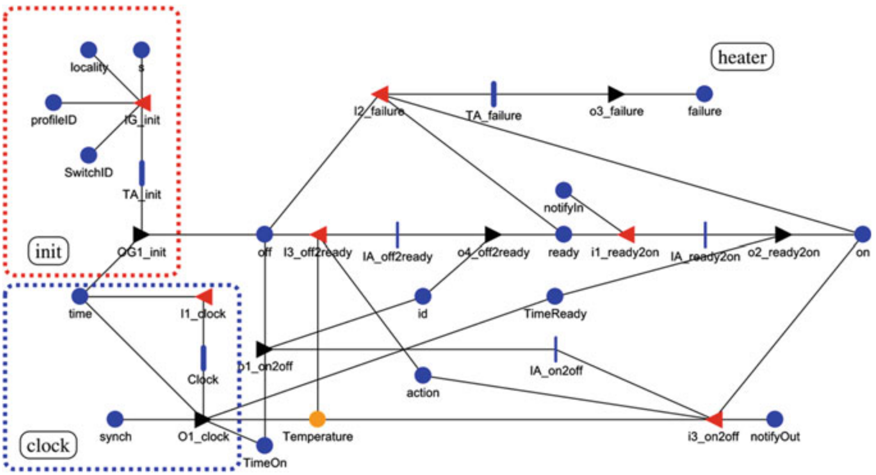


Fig. 3 The SAN model RailRoadSwitchHeater, logically divided into three subnets: the init sub-net, the clock sub-net and the heater sub-net

railroad track each interval of time t. We note that in this particular case study there is no neat distinction between the SAN models concerning the physical module and the one for the control module. Indeed the discussed SAN model captures aspects concerning both modules (i.e. energy policies, temperature).

Heater sub-net. The heater sub-net represents the status of the railroad switch heater. The heater can be activated (one token in the place *on*), waiting for being activated (one token in the place *ready*), turned off (one token in the place *off*), or failed (one token in the place *failure*). Indeed, according to the heating policy, once the system temperature goes below a pre-defined warning threshold (T_{wa}), the heating needs to be activated, otherwise the associated switch fails. Then, once the temperature raises and reaches the working threshold (T_{wo}), the heating system can be safely turned off.

The heater sub-net interacts with the Coordinator SAN model through places shared among all the replicas of the heater model and the Coordinator model implementing the logic described above. For example, if the heater is on state ready, in order to be turned on, the input gate $i1_{ready2on}$ checks if the marking of the shared place *notifyIn* is equal to the marking of the place *SwitchID*, which means that the coordinator has notified the heater to be turned on.

The activity $TA_{failure}$ models the failure of a component. It has an exponential distribution of time based on the temperature of the railroad track: the more the temperature is below the freezing threshold the more is probable that the activity will fire (the activity is not activated if the temperature is positive).

For a comprehensive description of all the SAN models we refer the interested reader to [17, 18].

3.3 Evaluation Module

In the evaluation module the overall composed model is instantiated with varying values of T_{wa}, T_{wo} and NH_{max}, in order to find the best compromise between reliability and energy consumption. We analyse two different measures of interest. The first concerns the energy consumption while the second addresses the reliability of the system under analysis.

1 $CE(t,l)$: the time (number of hours) an heater is activated in the time interval $[t,t + l]$. This measure is defined by accumulating the marking of the place *on* of the *RailRoadSwitchHeater* net in the interval $[t,t + l]$, that is the time that each replica of the SAN model *RailRoadSwitchHeater* spends in the marking represented by one token in the place *ON*. Hence $CE(t,l)$ is the hours that an heater is active. By multiplying $CE(t,l)$ for the power consumed (kilowatt per hour) it is possible to derive the energy consumed by the system;
2 $PFAIL(t,l)$: the probability that at least a switch fails (becomes frozen) at time $t + l$, given that at time t is not failed. This measure is defined as the probability that at time $t + l$ there is one token in the place *failure* of the SAN model *RailRoadSwitchHeater*.

We remark that reliability is computed as the probability that no failure occurs in the interval of time under analysis [38], that is $1 - PFAIL(t,l)$. Simulation-based evaluations have been performed using Möbius tool [34], considering from a minimum of 1000 batches to a maximum of 10,000 batches. The measures of interest were estimated within an interval of confidence of 0.95, and the model is instantiated to real-world data, available at [17, 18].

3.4 Results Presentation

Figures 4b, a depicts respectively the measures of interest $PFAIL(t,l)$ and $CE(t,l)$ at the varying of the gap between T_{wa} and T_{wo}, while Fig. 4d, c depicts respectively the measures of interest $PFAIL(t,l)$ and $CE(t,l)$ at the varying of the available energy NH_{max}.

From Fig. 4 it is possible to observe the results of our evaluation. For values of the thresholds T_{wa} = 7C, T_{wo} = 8C, and energy available NH_{max} = 50 % we obtain the optimal trade-off between energy consumption and reliability. Indeed, for values of T_{wa} lower than 7C, as a result of the environment temperature, the internal temperature of the heater will fall quickly under zero, hence increasing the probability of failure. Instead, for values greater than 7C, the active heaters jeopardize all the energy available, leaving the other pending heaters waiting for too long, hence increasing the probability of a failure. Since the kilowatts consumed by each heater are constant, an increment in the gap between the thresholds will result in a longer period of activation of the heater. Finally, in Fig. 4 we only have analysed the high-priority switches, which are predominant for this case study, and in particular NH_{max} = 50 % is enough for heating all the high-priority switches; while for lower values of NH_{max} we have an increment in the probability of failure. More details and measures of interest for this case study are available at [17, 18].

4 Conclusions

We have discussed techniques for modelling and evaluating energy saving reliable CPS, where the management of the energy resources is handled by specific computational units that control physical entities. Policies of energy saving are generally adopted for implementing energy saving techniques. While applying these policies, computational units must also guarantee the overall reliability of the system and other dependability related requirements. Energy saving and reliability aspects are often opposite requirements, and it is important to find a good trade-off between these measures.

Hence, we have proposed a model-based approach to the analysis of CPS, which is very useful for evaluating the measures of interest in order to find a good setup

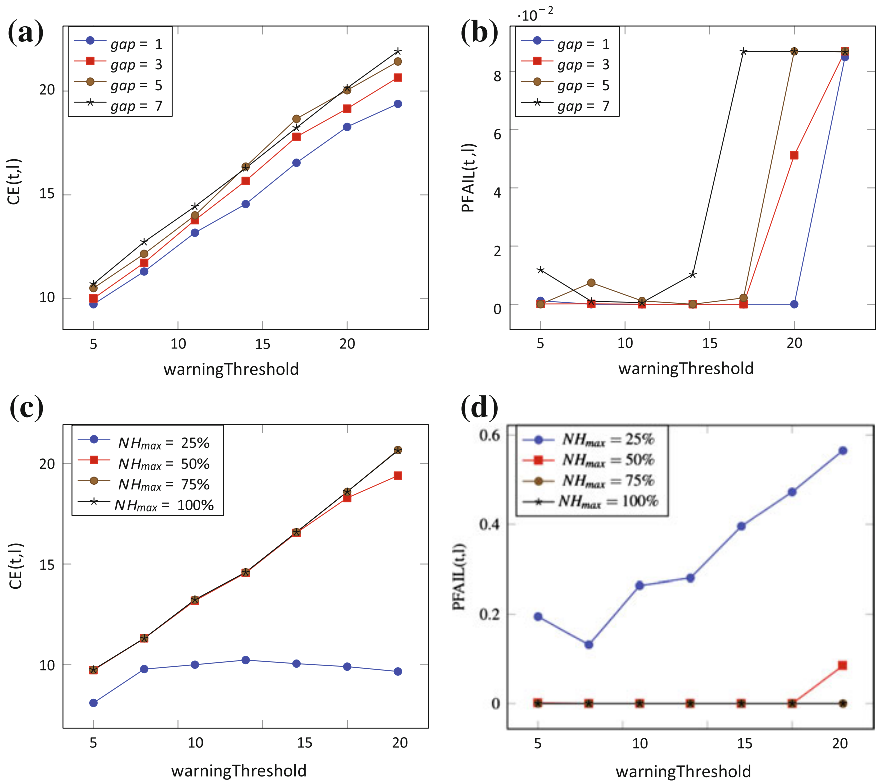


Fig. 4 The graphs for the measures of interest $CE(t,l)$ and $PFAIL(t,l)$. **a** The amount of energy consumed, where the different lines correspond to different gaps between the thresholds T_{wa} and T_{wo}. **b** The probability of failure, where the different lines correspond to different gaps between the thresholds T_{wa} and T_{wo}. **c** The amount of energy consumed, where the different lines correspond to different levels of energy available to the whole system (NH_{max}). **d** The probability of failure, where the different lines correspond to different levels of energy available to the whole system (NH_{max})

for the parameters of the policies of energy saving, so to satisfy both energy and reliability requirements. By adopting a model-based approach it is possible to evidence, far before they become obvious, weakness points in the system under analysis, saving both in time and economic terms.

We have presented widely adopted modelling formalisms for the evaluation of energy-saving aspects, focusing on Markovian models, non-Markovian models and extensions of Petri Nets capable of expressing probabilistic aspects of the systems under analysis. A list of available tools for supporting the design and evaluation of models of energy saving systems has been presented, with pointers to the related literature.

To illustrate and corroborate our proposal, we have illustrated a case study taken from the railway transport domain. In particular, we have tailored our modelling framework to the evaluation of reliability and energy consumption indicators for a system of railroad switch heaters, by using stochastic activity networks and the Möbius tool. The adopted policy of energy saving is based on an on-off mechanisms for the heating of the switches, with parametric thresholds representing the temperatures triggering the activation/deactivation of the heating.

We address some lines of future research concerning energy saving CPS. Firstly, due to their physic nature and their interactions with the environment, the behaviour of these systems is in general unpredictable. New control techniques are necessary, for restricting their possible behaviours in order to predict and avoid possible failures, improve and verify their dependability [39]. Moreover, defining the energy requirements that a CPS must satisfy, verifying that the proposed model satisfy them or proving that such requirements are not satisfiable is a hot research topic [40].

Likewise, energy saving solutions for CPS are expected to be impacted by these foreseen evolutions. In turn, analysis support to help defining and selecting appropriate energy saving policies which satisfy required dependability properties needs to keep the pace with this future prospect. Actually, the framework outlined in this paper is amenable to undergo extensions and adaptation. Of course, the implementation aspects will be highly impacted and may raise significant challenges, especially in terms of scalability and efficiency. However, research advancements in model/based solution techniques in relation with modern and future complex systems are expected to alleviate such potential weaknesses of model/based analysis.

Concerning the proposed case study, several extensions have been identified. It would be interesting to study how the energy consumption is modified by changing parameters of the underlying physical model. Indeed, the obtained results may suggest that, by changing the material the heaters are composed of, its length or the power consumed, a better trade-off between reliability and energy optimization can be obtained. It would also be interesting to let the power consumed by the system vary at different weather conditions. This may help to improve the reliability of the system. Indeed, in case of emergency a major throughput may prevent a failure. Adapting the thresholds to the different class of priorities may increase the reliability of the system. Finally, modelling and evaluating the case study with different formalisms and tools would help to validate the obtained results.

References

1. Friedler, F.: Process integration, modelling and optimisation for energy saving and pollution reduction. Appl. Therm. Eng. **30**(16), 2270–2280 (2010). http://www.sciencedirect.com/science/article/pii/S1359431110001936. In: Selected Papers from the 12th Conference on Process Integration, Modelling and Optimisation for Energy Saving and Pollution Reduction

2. Balbo, G.: Introduction to generalized stochastic petri nets. In: Bernardo, M., Hillston, J. (eds.) Formal Methods for Performance Evaluation, LNCS, vol. 4486, pp. 83–131. Springer, Berlin (2007)
3. Qiu, Q., Wu, Q., Pedram, M.: Dynamic power management of complex systems using generalized stochastic petri nets. In: DAC. pp. 352–356 (2000)
4. Qiu, Q., Wu, Q., Pedram, M.: Stochastic modeling of a power-managed system: construction and optimization. In: Proceedings of the 1999 International Symposium on Low Power Electronics and Design, 1999, San Diego, California, USA, 16–17, Aug 1999. pp. 194–199 (1999)
5. Erbes, T., Shukla, S.K., Kachroo, P.: Stochastic learning feedback hybrid automata for dynamic power management in embedded systems. In: SMCia/05, IEEE Mid-Summer Workshop on Soft Computing in Industrial Applications, June 2005 (2005)
6. Muller, S.C., Hager, U., Rehtanz, C., Wedde, H.F.: Application of self-organizing systems in power systems control. In: Dieste, O., Jedlitschka, A., Juzgado, N.J. (eds.) PROFES 2012 Proceedings LNCS, vol. 7343, pp. 320–334. Springer, Berlin (2012)
7. Ghasemieh, H., Boudewijn, R. Haverkort, M.R.J., Remke, A.: Energy resilience modeling for smart houses. In: DSN (2015) (to appear)
8. David, R., Alla, H.: On hybrid petri nets. Discrete Event Dyn. Syst. **11**(1–2), 9–40
9. Zhu, D., Melhem, R., Mosse, D.: The effects of energy management on reliability in real-time embedded systems. In: IEEE/ACM International Conference on Computer Aided Design, 2004, ICCAD-2004, pp. 35–40, Nov 2004 (2004)
10. Misra, S., Krishna, P.V., Saritha, V., Obaidat, M.S.: Learning automata as a utility for power management in smart grids. IEEE Commun. Mag. **51**(1), 98–104 (2013)
11. Cau sevic, A., Seceleanu, C., Pettersson, P.: Distributed energy management case study: a formal approach to analyzing utility functions. In: Margaria, T., Steffen, B. (eds.) Leveraging Applications of Formal Methods, Verification and Validation. Specialized Techniques and Applications, Lecture Notes in Computer Science, vol. 8803, pp. 74–87. Springer, Berlin (2014)
12. Avizienis, A., Laprie, J.C., Randell, B., Landwehr, C.: Basic concepts and taxonomy of dependable and secure computing. Dependable Secure Comput. IEEE Trans. **1**(1), 11–33 (2004)
13. Front matter: In: Karlin, H.M.T. (ed.) An Introduction to Stochastic Modeling (Revised Edition), pp. iii. Academic Press (1994). http://www.sciencedirect.com/science/article/pii/B978012684885450001X
14. Bernardi, S., Merseguer, J., Petriu, D.C.: Model-Driven Dependability Assessment of Software Systems. Springer, Berlin (2013)
15. Diab, H.B.; Zomaya, A.Y.: Dependable Computing Systems: Paradigms, Performance Issues and Applications. Wiley, USA (2005)
16. Lee, E.A.: Cyber physical systems: design challenges. In: Proceedings of the 2008 11th IEEE Symposium on Object Oriented Real-Time Distributed Computing. ISORC'08, pp. 363–369. IEEE Computer Society, Washington DC (2008). http://dx.doi.org/10.1109/ISORC.2008.25
17. Basile, D., Chiaradonna, S., Di Giandomenico, F., Gnesi, S.: Stochastic model-based evaluation of reliable energy-saving railroad switch heating systems. Tech. Rep., Istituto di Scienze e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa (2016), http://puma.isti.cnr.it/dfdownload.php?ident=/cnr.isti/2016-TR-009
18. Basile, D., Chiaradonna, S., Di Giandomenico, F., Gnesi, S., Mazzanti, F.: Stochastic model based analysis of energy consumption in a railroad switch heating system. In: Software Engineering for Resilient Systems—Proceedings of 7th International Workshop, SERENE 2015, Paris, France, 7–8 Sept 2015. pp. 82–98 (2015)
19. Haverkort, B.R.: Lectures on formal methods and performance analysis (Chap). In: Markovian Models for Performance and Dependability Evaluation, pp. 38–83. Springer, New York (2002). http://dl.acm.org/citation.cfm?id=567305.567307
20. Lyu, M.R. (ed.): Handbook of Software Reliability Engineering. McGraw-Hill Inc, Hightstown (1996)

21. O'Connor, P.P., Kleyner, A.: Practical Reliability Engineering, 5th edn. Wiley Publishing (2012)
22. Trivedi, K.S., Malhotra, M.: Messung, Modellierung und Bewertung von Rechen und Kommunikationssystemen (Chap). In: Reliability and Performability Techniques and Tools: A Survey, pp. 27–48. Springer, Berlin (1993)
23. Reibman, A., Smith, R., Trivedi, K.: Markov and Markov reward model transient analysis: an overview of numerical approaches. Eur. J. Oper. Res. **40**(2), 257–267 (1989). http://www.sciencedirect.com/science/article/pii/0377221789903354
24. Peterson, J.L.: Petri nets. ACM Comput. Surv. **9**(3), 223–252 (1977)
25. Bause, F., Kritzinger, P.S.: Stochastic petri nets: an introduction to the theory. SIGMETRICS Perform. Eval. Rev. **26**(2), 2–3 (1998)
26. Boguna, M., Lafuerza, L.F., Toral, R., Serrano, M.A.: Simulating non-markovian stochastic processes. Phys. Rev. E **90**, 042108 (2014)
27. Sahner, R.A., Trivedi, K., Puliafito, A.: Performance and Reliability Analysis of Computer Systems: An Example-Based Approach Using the SHARPE Software Package. Springer Publishing Company, Incorporated (2012)
28. German, R.: Performance Analysis of Communication Systems with Non-Markovian Stochastic Petri Nets. Wiley, New York (2000)
29. Front matter: In: Grabski, F. (ed.) Semi-Markov Processes: Applications in System Reliability and Maintenance, pp. i–ii. Elsevier (2015). http://www.sciencedirect.com/science/article/pii/B9780128005187099889
30. Sanders, W.H., Meyer, J.F.: Stochastic activity networks: formal definitions and concepts. In: Brinksma, E., Hermanns, H., Katoen, J. (eds.) Lectures on Formal Methods and Performance Analysis, First EEF/Euro Summer School on Trends in Computer Science 2000, Revised Lectures. LNCS, vol. 2090, pp. 315–343. Springer, Berlin (2000)
31. Ciardo, G., Muppala, J., Trivedi, K.: Spnp: Stochastic petri net package (1989)
32. Chiola, G., Franceschinis, G., Gaeta, R., Ribaudo, M.: Greatspn 1.7: Graphical editor and analyzer for timed and stochastic petri nets. Perform. Eval. **24**, 47–68 (1995)
33. Bucci, G., Carnevali, L., Ridi, L., Vicario, E.: Oris: a tool for modeling, verification and evaluation of real-time systems. STTT **12**(5), 391–403 (2010). doi:10.1007/s10009-010-0156-8
34. Clark, G., Courtney, T., Daly, D., Deavours, D., Derisavi, S., Doyle, J.M., Sanders, W.H., Webster, P.: The möbius modeling tool. In: Proceedings of the 9th International Workshop on Petri Nets and Performance Models. pp. 241–250 (2001)
35. http://www.railsco.com/ ~ Eelectric_switch_heater_controls.htm
36. Brodowski, D., Komosa, K.: A railroad switch and a method of melting snow and ice in railroad switches (2013). https://data.epo.org/publication-server/rest/v1.0/publication-dates/20131225/patents/EP2677079NWA1/document.html
37. Cannon, J.R.: The One-Dimensional Heat Equation (Cambridge Books Online). Cambridge University Press (1984)
38. Trivedi, K.S.: Probability and Statistics with Reliability, Queuing and Computer Science Applications. Wiley, NY (2008)
39. Antsaklis, P.: Goals and challenges in cyber-physical systems research editorial of the editor in chief. IEEE Trans. Autom. Control **59**(12), 3117–3119 (2014)
40. Banerjee, A., Venkatasubramanian, K.K., Mukherjee, T., Gupta, S.K.S.: Ensuring safety, security, and sustainability of mission-critical cyber-physical systems. Proc. IEEE **100**(1), 283–299 (2012)

MSS Models of Smart Grids with Multi-level Degradation and Recovery

Eugene Brezhnev, Herman Fesenko, Vyacheslav Kharchenko, Vitaly Levashenko and Elena Zaitseva

Abstract Digital substations (DS) of smart grid are complex multi-component maintained systems, consisting of a lot of hardware and software components. Failures of the components cause functional and parametrical degradation of the substations. According with systems (DS) structure the reliability-block diagram (RBD), the structure function and the structure function and the Direct Partial Logical Derivatives (DPLDs) for RMSS "electronic transformers—merging unit" are considered. The mathematical tool of logical differential calculus and DPLD in particular are used in many application problems. One of them is reliability engineering. The principal condition of the DPLD application in reliability analysis is the representation of system under investigation by the structure function. We consider the calculation some of these measures as structural, Birnbaum's, and criticality for the analysis of the electronic transformers—merging unit. The structure function of this unit based on the operation conditions of this system (unit). The construction of this function allows estimating the most important

E. Brezhnev (✉) · V. Kharchenko
National Aerospace University KhAI, Kharkiv, Ukraine
e-mail: milestone@list.ru

V. Kharchenko
e-mail: v_s_kharchenko@ukr.net

E. Brezhnev · V. Kharchenko
Centre for Safety Infrastructure-Oriented Research and Analysis, Kharkiv, Ukraine

H. Fesenko
O.M. Beketov National University of Urban Economy in Kharkiv, Kharkiv, Ukraine
e-mail: fesennko@mail.ru

V. Levashenko · E. Zaitseva
University of Žilina, Žilina, Slovakia
e-mail: vitaly.levashenko@fri.uniza.sk

E. Zaitseva
e-mail: lena.zaitseva@fri.uniza.sk

V. Kharchenko et al. (eds.), *Green IT Engineering: Concepts, Models, Complex Systems Architectures*, Studies in Systems, Decision and Control 74,
DOI 10.1007/978-3-319-44162-7_11

components of this system in stationary state through structural, Birnbaum's, and criticality importance measures. According with developed technique an example of DS availability model is presented and discussed.

Keywords Recovered multi-state system · Structure function · Reliability-block diagram · Direct partial logic derivative

1 Introduction

1.1 Motivation

Smart grid are widely use in order to effectively meet the dynamically changing needs of the consumers without dramatically increasing the coast, reliability and quality of services provided. Using digital information, automated management and stand-alone systems, Smart Grids provides reliable power supply with less short-term power outages.

One of the most important smart grid parts are DS. It is complex multi-component maintained systems consisting of a lot of components. Different failed components can cause significant deviations from normal value some parameters, such as power frequency, supply voltage etc. Therefore DS are systems with multi-level energy (power) degradation or multi-state system (MSS). Besides, a DS as recovered systems, and a set of on-line testing and maintenance means and procedures assure step by step recovery. Hence such systems can be described by model of a recovered multi-state system (RMSS).

1.2 Work Related Analysis

Hence such systems can be described by model of a RMSS. The joint redundancy and maintenance replacement schedule optimization problem generalized to RMSS are formulated in [1]. A maintenance policy for a RMSS is proposed in [2]. A two-stage preventive maintenance (PM) policy for the multi-state deterioration system under periodic inspection and with multiple candidate actions for PM is presented in [3].

A model for evaluating the availability, the production rate and the reliability function of multi-state degraded systems subjected to minimal repairs and imperfect preventive maintenance is developed in [4].

Also note the papers in [5–9]. These papers present latest studies and advances about RMSS reliability evaluation, RMSS optimization and RMSS maintenance.

1.3 Goals and Structure

Goal of the paper is development and research of availability models for complex systems similar DS as a model of RMSS. Structure of the paper is the following.

Second section structure describes functions of typical MSSs and presents MSS importance analysis.

Third section considers the RBD, the DD, the structure function and the Direct Partial Logical Derivatives (DPLD) for RMSS "electronic transformers—merging unit".

Last section concludes and discusses the next research steps.

2 General Approach to Analysis and Modelling RMSS

Reliability is considered as an important characteristic of any modern system. However, modern systems are very complex and, therefore, their reliability analysis is a challenging task. The complexity of these systems results from the fact that they consist of a huge amount of elements with various behaviour. Examples of such systems include electrical transmission systems, gas grids [10, 11] or healthcare systems [12]. Distribution grids are typical examples of network systems. They represent large systems consisting of many (not necessarily) hardware elements with different properties, e.g. generating units, different types of transmission lines, and demand centres in the case of electrical grids. Similarly, a healthcare system is a typical instance of socio-technical systems composing of a lot of highly variable elements (components) that can be classified as hardware, software, organizational, and human [10]. This variety of components of complex systems causes that such systems can operate at many different levels of performance and, therefore, their analysis is more difficult than the analysis of other systems.

2.1 Classification of MSS

The reliability analysis of any system is provided based on the mathematical representation (description) of system. The mathematical description is interpreted as the system mathematical model. There some types of the system mathematical models in the reliability engineering. The type of the system mathematical model is caused by two conditions. The first of them is number of performance level considered in the system functioning in point of view reliability evaluation. The second is type of the mathematical approach that is used for the reliability evaluation.

There are two mathematical models in point of view of number of the performance level in the system functioning [10]. The first ones are known as *Binary-State Systems* (BSSs) because they permit defining only two states in system/components performance—functioning and failed. BSSs have been widely used in classical approaches of reliability engineering [13]. However, they are not very suitable for the analysis of complex systems because their use requires drawing the line between situations in which the system is functioning and when it is considered to be failed. Very often, this is impossible and, therefore, other types of mathematical models are used. According to [10], *Multi-State Systems* (MSSs) are one of the most prospective approaches. MSSs have been introduced in reliability engineering since 1975 [14–16]. In a MSS, both the system and its components may experience more than two states (performance levels). This indicates that they are very suitable for modelling of complex systems and investigation of the reliability behaviour in more details. But the application this mathematical model has the complication: growing number of states of system components results in dramatic increase of model size, what causes increase of time complexity of the analysis. Therefore, use of MSSs requires development of new and effective methods that could be used to analyse them.

The system evaluation based on BSS and MSS is possible based on four principal mathematical backgrounds [17]:

- extensions of Boolean methods,
- stochastic processes,
- universal generating function,
- Monte Carlo simulation.

Each of these approaches is used for some specific tasks relating to evaluation of MSSs. For example, stochastic processes (such as Markov processes) are used to analyse system state transition process [18], universal generating function is useful in optimization problems [19], while Monte Carlo simulation represents a universal tool for reliability assessment of systems consisting of a huge amount of components [20]. If we want to investigate influence of individual system components on the system activity, then the approach based on extensions of Boolean methods can be viewed as one of the most suitable.

The idea of extensions of Boolean methods for the analysis of MSSs lies in the fact that the tools of Boolean algebra have been widely used in reliability analysis of BSSs. This has resulted from the fact that the system structure function, which defines dependency between system state and states of system components, can be viewed as a Boolean function in the case of BSSs [13, 17]. However, this is not true for MSSs. Therefore, some works, e.g. [21, 22], have tried to transform the structure function of a MSS into a Boolean function using Boolean algebra with restrictions [21]. Such transformation allows us to use methods of Boolean algebra in reliability analysis of MSSs. However, this transformation can result in increase of model complexity. To avoid these complications, other authors have proposed using Multiple-Valued Logic (MVL) in reliability analysis of MSSs since there exists some correlation between the structure function of a MSS and a MVL function [23, 24].

2.2 *Indicators*

2.2.1 Structure Function

A MSS is mathematical model that is used for the description of the system of n components. The ith system component state is denoted as x_i (i = 1, …, n). Consider simple variant of a MSS where the system components and system have equal number of states and performance levels. Each component in such mathematical representation has m_i states that are indicated as 0 for the complete failure and as m_i-1 for perfect functioning. Suppose, that the system has M performance level too: from the complete failure (it is 0) to the perfect functioning (it is M-1). The system performance levels depend on components states and this dependency is defined by the structure function $\phi(\boldsymbol{x})$ identically:

$$\phi(\boldsymbol{x}) : \{0, 1, \ldots, m_1 - 1\} \times \ldots \times \{0, 1, \ldots, m_n - 1\} \to \{0, 1, \ldots, M - 1\} \quad (1)$$

In the mathematical point of view the structure function (1) agrees with the definition of incompletely specified MVL function [33, 34]. Note that the structure function of a MSS (2) is transformed into the structure function of a BSS if $m_i = m_j = M = 2$ (for all $i,j \in \{1, \ldots, n\}$, $i \neq j$): $\varphi(\boldsymbol{x}) : \{0, 1\}^n \to \{0, 1\}$.

2.2.2 Structure Functions of Typical MSSs

There are some typical structures in the reliability engineering: series, parallel, k-out-of-n and bridge. These structures have been introduced for BSSs and every of these systems has single description in form of RBD (Reliability Block Diagram) and the structure function (Fig. 1). The application of system with typical structure allows simplifying the analysis and estimation of complex system [27].

The extension of the conception of typical structures for MSS has some complications. First of all it is ambiguity of the structure function definition for MSS with typical structure. For example, in paper [24, 28, 29] the structure functions of series, parallel and k-out-of-n MSS are defined based on the interpretation of mathematical equations of these system in terms of MVL. Structure functions of these MSS in MVL terms are declared by OR (∨) and AND (∧) MVL functions: OR(a, b) = MAX(a, b) and AND(a, b) = MIN(a, b). According to [24, 28, 29] structure functions parallel, series and k-out-of-n MSS are declared

- for parallel MSS as:

$$\phi(\boldsymbol{x}) = \bigvee_{i=1}^{n} x_i = MAX(x_1, x_2, \ldots, x_n), \quad (2)$$

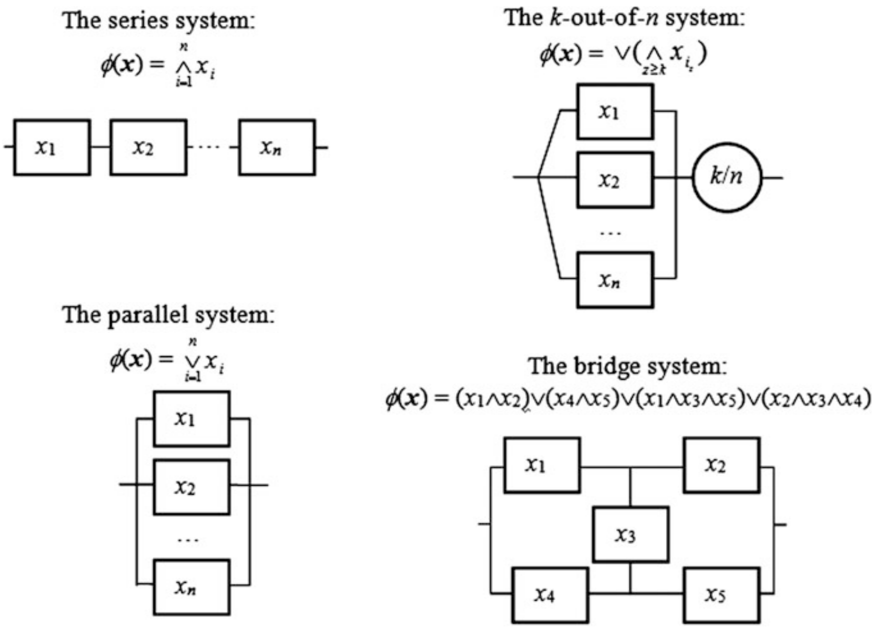


Fig. 1 Graphical and mathematical interpretation of typical structures of Binary-State System

- for series MSS as:

$$\phi(\boldsymbol{x}) = \overset{n}{\underset{i=1}{\wedge}} x_i = MIN(x_1, x_2, \ldots, x_n), \tag{3}$$

- for k-out-of-n MSS as:

$$\phi(\boldsymbol{x}) = \vee(\underset{z \geq k}{\wedge} x_{i_z}) = MAX(MIN_1(x_{i_1}, x_{i_2}, \ldots, x_{i_k}), \ldots, MIN_w(x_{i_1}, x_{i_2}, \ldots, x_{i_k})), \tag{4}$$

where $w = n!/(k! \cdot (n-k)!)$.

In papers [30, 31] there are other declarations of these MSS in which a system performance level depends on number of components with fixed states. For example, k-out-of-n MSS according this rule is defined as:

- if at least k components are in state m_i-1, then the system is in state M-1;
- else if at least k components are in state m_i-2 or better, then the system is in state M-2;
- …
- else if at least k components are in state 1 or better, then the system is in state 1;
- else the system is in state 0.

Table 1 The structure functions of parallel MSS ($n = 2$, $m_1 = m_2 = M = 3$)

Components states		Structure function of parallel MSS			
x_1	x_2	$\phi_1(\boldsymbol{x})$	$\phi_2(\boldsymbol{x})$	$\phi_3(\boldsymbol{x})$	$\phi_4(\boldsymbol{x})$
0	0	**0**	**0**	**0**	**0**
0	1	**1**	**1**	**1**	**1**
0	2	**2**	**1**	**1**	**1**
1	0	**1**	**1**	**1**	**1**
1	1	**1**	**2**	**1**	**1**
1	2	**2**	**2**	**1**	**2**
2	0	**2**	**1**	**1**	**1**
2	1	**2**	**2**	**1**	**2**
2	2	**2**	**2**	**2**	**2**

In case of BSS this definition of k-out-of-n MSS and definition according to the Eq. (4) are equal.

In Table 1 some structure functions of parallel MSS of two components ($n = 2$) with three states ($m_1 = m_2 = 3$) and three performance level ($M = 3$) are shown. The structure function $\phi_1(\boldsymbol{x})$ in Table 1 is defined based on the Eq. (2). The structure function of parallel MSS $\phi_2(\boldsymbol{x})$ is define according to rules proposed in papers [30, 31] as: the system performance level is 2 (working) if all components function; the system performance level is 2 (partly working) if some of components are functioning; and the system is failure if all component are skip down. But all structure functions in Table 3 agree with parallel BSS.

Therefore the typical structures of MSS have no single valued mathematical definition, because there are the set of structure functions of MSS that can be agreed with one BSS. The structure function allows defining MSS explicitly. Therefore the structure function is preferable form of MSS mathematical representation.

2.2.3 MSS Availability

The system evaluation in reliability engineering is implemented based on different indices and measures. The most used indices are the system availability and unavailability. Availability of system is defined as the probability that the system will perform its intended functions at a given time, under designated operating conditions, and with a designated support environment [32]. It is an important measure of the system performance—its ability to quickly become operational following a failure. Mathematically, availability is a measure of the fraction of the total time a system is able to perform its intended function. These indices are computed based on the system structure and probabilities of the system components states. Every system component is characterized by the probabilities of its state:

$$p_{i,s} = Pr\{x_i = s\}, s = 0, \ldots, m_i - 1. \quad (5)$$

The MSS availability and unavailability based on the conception of the structure function (1) and taking into account the components states probabilities (2) are defined as follows [8]:

- for parallel MSS as:

$$A(j) = Pr\{\phi(\boldsymbol{x}) \geq j\}, j = 1, \ldots, M - 1, \quad (6)$$

$$U = Pr\{\phi(\boldsymbol{x}) = 0\}. \quad (7)$$

But there is the special case of the definition of the system availability for MSS as [17, 24]:

$$A_j = \Pr\{\phi(\boldsymbol{x}) = j\}, j = 1, \ldots, M - 1 \quad (8)$$

In this paper definitions (7) and (8) for a MSS availability and unavailability will be used.

For example, consider the availability of the first performance levels for two parallel systems that are defined by structure functions $\phi_1(\boldsymbol{x})$ and $\phi_2(\boldsymbol{x})$ in Table 1. The availability of the partly working level A_1 (8) for the first function $\phi_1(\boldsymbol{x})$ is defined as:

$$A_1 = p_{1,0} \cdot p_{2,1} + p_{1,1} \cdot p_{2,0} + p_{1,1} \cdot p_{2,1}$$

and for the second function $\phi_2(\boldsymbol{x})$ is defined as:

$$A_1 = p_{1,0} \cdot p_{2,1} + p_{1,0} \cdot p_{2,2} + p_{1,1} \cdot p_{2,0} + p_{1,2} \cdot p_{2,0}$$

The two last equations are not equal; therefore they define the availability of different system. This example illustrates the importance of correct definition of the system mathematical representation based on MSS.

2.2.4 MSS Importance Analysis

The availability and unavailability are useful indices in reliability estimation of the system, but these indices represent and characterise static system state and they do not permit to identify the influence of individual system components on the proper work of the system. The MSS importance analysis is one of the possible directions for the estimation of MSS behaviour against the system structure and components states. This analysis allows the examining and quantifying the influence of change of the component state on the system performance level. Therefore, every system component has a measure of its importance for the system functioning and failure. This measure in reliability engineering is called *Importance Measure* (IM). IM is a probability of the MSS performance level change depending on the change of the system component state [17, 33]. There are different methods and algorithms for the calculation of IMs for the MSS. In paper [19] IMs, a system with two performance levels and multi-state components is considered. J.E. Ramirez-Marquez and D.W. Coit in [34] have generalized this result for MSS and have proposed a new type of IM called the composite importance measures. An actual review of IMs is presented in [35]. A new application of IMs for degraded MSS has been proposed in [36]. New methods for importance analysis of MSS are considered in [24, 27, 28]. These methods are based on the mathematical approach of *Multiple-Valued Logic* (MVL) such as Logical Differential Calculus. The Logical Differential Calculus is a mathematical approach that permits changes in MVL function to be analysed depending on changes of its variables [37]. Therefore, this tool can be used to evaluate the influence of every system component state change on the MSS performance level if the MSS structure function is interpreted as a MVL function.

There exist a lot of IMs, but the most often used are the *Structural Importance* (SI), *Birnbaum's Importance* (BI) and *Criticality Importance* (CI).

The MSS importance analysis is approach for estimation of MSS behaviour against the system structure and components states. There are some types of measures that show change of the system availability due the ith component state change.

The SI takes into account the topological specifics of the system [17, 22]. SI of MSS for the ith component state s is defined as probability of this system performance level j decrement if the component state changes from s_i to (s_i-1):

$$I_S^{(s_i,j)} = \frac{\rho_i^{s,j}}{m_1 \ldots m_{i-1} m_{i+1} \ldots m_n} \tag{9}$$

where $\rho_i^{s,j}$ is number of system states when the change component state from s to (s-1) results decrement of the system performance level.

In Zaitseva [24, 25] a new method for calculation of the number $\rho_i^{s,j}$ in (9) has been proposed. This method is based on the calculation of the *Direct Partial Logic Derivative* (DPLD) of MSS structure function. DPLD is method of Logical Different Calculation for estimation of the influence of change of variable value to the MVL function value change. DPLD background for MVL function has been considered in [33]. DPLD for MSS structure function is defined as [24]:

$$\partial\phi(j \to j-1)/\partial x_i(s \to s-1) = \begin{cases} 1, \ \mathit{if} \ \phi(s_i, \boldsymbol{x}) = j \ \mathit{and} \ \phi((s-1)_i, \boldsymbol{x}) = j-1 \\ 0, \quad \mathit{other} \end{cases} \tag{10}$$

where $\phi(s_i, \boldsymbol{x}) = \phi(x_1, \ldots, x_{i-1}, s_i, x_{i+1}, \ldots, x_n)$; $\phi((s\text{-}1)_i, \boldsymbol{x}) = \phi(x_1, \ldots, x_{i-1}, (s\text{-}1)_i, x_{i+1}, \ldots, x_n)$; $s_i \in \{1, \ldots, m_i\text{-}1\}$ and $j \in \{1, \ldots, M\text{-}1\}$.

Note the SI (10) investigates correlation and influence of system component from the topological point of view. But this measure does not take into account the probabilities of states of system components. This disadvantage is eliminated in the calculation of another IM that is known as the BI. The BI of a given component is defined as the probability that system is sensitive to inoperative of the ith system component [17, 38]. Let us consider the DPLD for calculation of the BI. In paper [38] the BI has been defined as

$$IB_i^{s,j} = \Pr\{\partial\phi(j \to h)/\partial x_i(s \to s-1) \neq 0\} \tag{11}$$

The Eq. (11) supposes the BI calculation as the probabilities of non-zero values of Direct Partial Logic Derivative $\partial\phi(j \to h)/\partial x_i(s \to s-1)$.

Consider the definition of CI which is the probability that the ith system component is relevant for the MSS performance decrement if it has failed or has a diminished state. For MSS, this measure can be defined as the probability that the MSS performance reduces if the state of the ith system component has changed from s to $s-1$ [19, 27]:

$$IC_i^{s,j} = IB_i^{s,j} \cdot \frac{p_{i,s-1}}{A_{j-1}} \tag{12}$$

where $p_{i,s-1}$ is the probability of the ith system component state $s-1$ (5) and A_{j-1} is the MSS probability state (8).

The CI measure (12) corrects BI for unreliability or a lower state of the ith component relative. This measure is useful, if the component has high BI and a low probability of the investigated state with respect to the MSS performance decrement. In this case, the ith component CI is low.

3 System «Electronic Transformers—Merging Unit» as MSS

Consider the analysis of the system “electronic transformers—merging unit” (Fig. 2) based on mathematical model of MSS structure function.

This system consists of 4 components (3 transformers and 1 merging unit) in point of view structure function based method. The Table 2 represents the functioning of this system. According to the system functioning description in Table 2 we can interpreted this system as system on 4 components ($n = 4$) that have two states: $x_i = 0$ is the component failure and $x_i = 1$ is the component working state ($i = 1, \ldots, 4$). The system has 3 performance levels ($M = 3$). Consider this system availability and importance analysis.

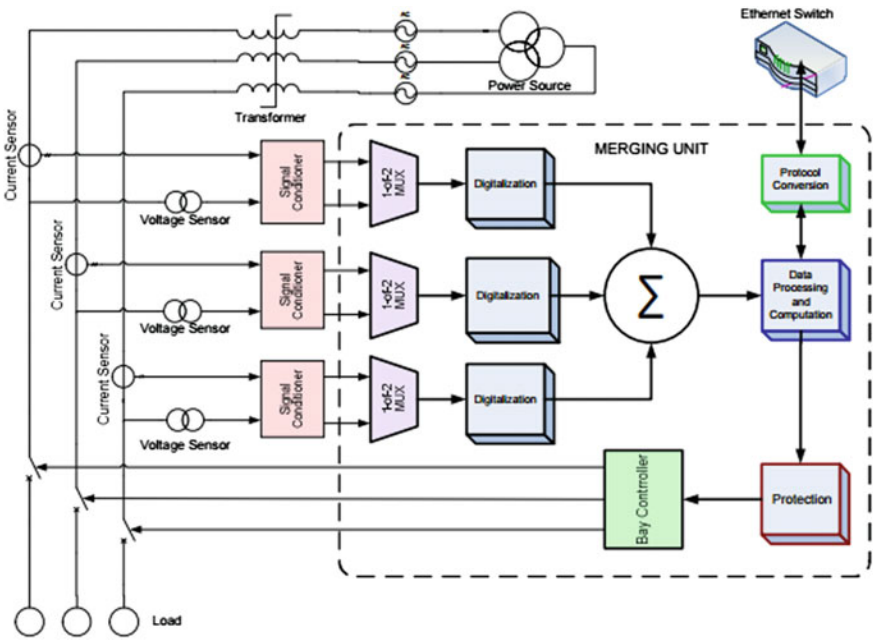


Fig. 2 The system "electronic transformers—merging unit"

Table 2 The system functioning

The system performance level	The level description	The system performance level definition
The performance level 2	The 3 electronic transformers and merging unit are functioning	**Functioning** (***M*** **= 2)**
The performance level 1	The 2 electronic transformers and merging unit are functioning	**Partly functioning** (***M*** **= 1)**
The performance level 0	The 1 electronic transformer functioning only or merging unit are not functioning	**Fault** (***M*** **= 0)**

3.1 System and Reliability Block Diagram

The structure function is defines univalent correlation of a system performance level and components states. Need to say that the structure function of BSS is Boolean function as a rule [13, 14]. The structure function method extension for the investigation of MSS is based on Boolean logic mostly [29, 33, 35]. In this case a system with some performance levels and some states of components is transformed into set of systems with two performance levels and two states of

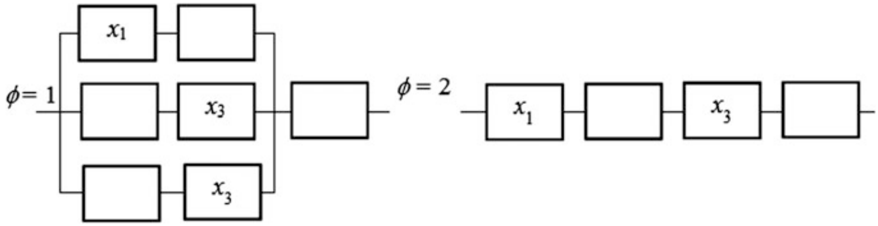


Fig. 3 RBD for the system "electronic transformers—merging unit"

components. It means representation of one MSS by set of BSSs: every performance level of MSS is considered separately and coding by 0 and 1 for all possible components states [29, 33, 39]. BSS of fixed performance level for components states has value 1 if this performance level confirms with equal components states of MSS.

For example, consider the representation of the system "electronic transformers—merging unit" (Fig. 2) by the Reliability Block Diagram (DBD) based on the methods proposed in [39]. The system can be presented by two RBDs for performance level 1 and 2 according to [39]. These diagrams are shown in Fig. 3.

This example illustrates that the extension of Boolean algebra to the representation of MSS structure function complicates the system analysis, decreases exactness, and increases a computational complexity. The main disadvantage of these methods is the great number of Boolean variables and functions in analysis. Therefore the application of Boolean algebra and the extension of this algebra to the multi-valued case are not effective.

3.2 *Structure Function of MSS*

Consider the representation of the system "electronic transformers—merging unit" (Fig. 1) by the structure function (1). According to the system description in Table 2 the structure function has 4 variables (n = 4), that have two values: $m_1 = m_1 = 2$. The structure function has 3 values (M = 3). The structure function of the system "electronic transformers—merging unit" according to definition (1) is shown in Table 3. The variable x_1, x_2 and x_3 interpret the states of 3 electronic transformers and the variable x_1 interprets the states of merging unit.

The structure function of the system in Table 3 can be used for the reliability analysis and allows to compute the system availability by (8) for the performance levels 1 and 2, the system unavailability based on (7) and importance measures for system components according to (9), (11) and (12).

Table 3 The structure function of the system "electronic transformers—merging unit" (Fig. 2)

x_1	x_2	x_3	x_3	$\phi(\boldsymbol{x})$
0	0	0	0	**0**
0	0	0	1	**0**
0	0	1	0	**0**
0	0	1	1	**0**
0	1	0	0	**0**
0	1	0	1	**0**
0	1	1	0	**0**
0	1	1	1	**1**
1	0	0	0	**0**
1	0	0	1	**0**
1	0	1	0	**0**
1	0	1	1	**1**
1	1	0	0	**0**
1	1	0	1	**1**
1	1	1	0	**0**
1	1	1	1	**2**

3.3 The System Estimation

3.3.1 The System Availability

Consider the system "electronic transformers—merging unit" availability and unavailability. This system has three performance levels. Therefore this system is characterised by availability of two levels and unavailability. The system structure function is defined in Table 4. This system unavailability is according to (7) is calculated as sum of situations that agree with zero-values of the structure function and is:

$$U = p_{40} + p_{10} \cdot p_{20} + p_{30} \cdot (p_{10} \cdot p_{21} + p_{11} \cdot p_{20}), \tag{13}$$

and system availability for performance levels "1" and "2" according to (8) are:

$$A_1 = p_{41} \cdot (p_{10} \cdot p_{21} \cdot p_{31} + p_{11} \cdot p_{20} \cdot p_{31} + p_{11} \cdot p_{21} \cdot p_{30}.) \tag{14}$$

$$A_2 = p_{11} \cdot p_{21} \cdot p_{31} \cdot p_{41} \tag{15}$$

According to (6) $A(2) = A_2$ and $A(1)$ is calculate as

$$A(1) = A_1 + A_2 = p_{41} \cdot (p_{11} \cdot p_{21} + p_{31} \cdot (p_{10} \cdot p_{21} + p_{11} \cdot p_{20})). \tag{16}$$

Therefore the indices of the unavailability (13) and availability (14)—(16) allow to estimate the system "electronic transformers—merging unit" stationary state

Table 4 Direct Partial Logical Derivatives $\partial\phi(j \rightarrow j–1)/\partial x_1(1 \rightarrow 0)$ for the system “electronic transformers—merging unit” (Fig. 1)

x_1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
x_2	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
x_3	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
x_4	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
$\partial\phi(2\rightarrow1)/\partial x_1(1\rightarrow0)$	–	–	–	–	–	–	–	–	0	0	0	0	0	0	0	1
$\partial\phi(2\rightarrow0)/\partial x_1(1\rightarrow0)$	–	–	–	–	–	–	–	–	0	0	0	0	0	0	0	0
$\partial\phi(1\rightarrow0)/\partial x_1(1\rightarrow0)$	–	–	–	–	–	–	–	–	0	0	0	1	0	1	0	0
$\partial\phi(2\rightarrow1)/\partial x_2(1\rightarrow0)$	–	–	–	–	0	0	0	0	–	–	–	–	0	0	0	1
$\partial\phi(2\rightarrow0)/\partial x_2(1\rightarrow0)$	–	–	–	–	0	0	0	0	–	–	–	–	0	0	0	0
$\partial\phi(1\rightarrow0)/\partial x_2(1\rightarrow0)$	–	–	–	–	0	0	0	1	–	–	–	–	0	1	0	0
$\partial\phi(2\rightarrow1)/\partial x_3(1\rightarrow0)$	–	–	0	0	–	–	0	0	–	–	0	0	–	–	0	1
$\partial\phi(2\rightarrow0)/\partial x_3(1\rightarrow0)$	–	–	0	0	–	–	0	0	–	–	0	0	–	–	0	0
$\partial\phi(1\rightarrow0)/\partial x_3(1\rightarrow0)$	–	–	0	0	–	–	0	1	–	–	0	1	–	–	0	0
$\partial\phi(2\rightarrow1)/\partial x_4(1\rightarrow0)$	–	0	–	0	–	0	–	0	–	0	–	0	–	0	–	0
$\partial\phi(2\rightarrow0)/\partial x_4(1\rightarrow0)$	–	0	–	0	–	0	–	0	–	0	–	0	–	0	–	1
$\partial\phi(1\rightarrow0)/\partial x_4(1\rightarrow0)$	–	0	–	0	–	0	–	1	–	0	–	1	–	1	–	0

depending of the probabilities of the system components states p_{ij} (i = 1,…, n and j = 0, …, $m_i - 1$).

3.3.2 The Importance Analysis of the System

Consider the importance analysis of the system "electronic transformers—merging unit" by the SI (9), BI (11) and CI (12). These indices are calculated based on the DPLD (10) of the structure function of this system. In particular, investigate the influence of the system components failure into the change of the system performance level. The indices for this analysis is calculated based on DPLD $\partial\phi(j \to k)/\partial x_i(1 \to 0)$ for j = 1, 2. All of these derivatives are shown in Table 4.

Values of SI (9) and intermediate values of $\rho_i^{s,j}$ are shown in Table 5. The SI values in the Table 5 permit to estimate the influence of the every components fault to the change of the system performance level. The maximal value has SI $IS_4^{1,1}$ that means maximal influence of the fourth component (merging unit) to the system failure. But this component fault has no influence into the system degradation from the level 2 to the level 1. Other components (electronic transformers) faults cause the system failure with the probability 0.250 and the system degradation with the probability 0.125.

Note the SI (9) investigates correlation and influence of system component in point of view of the reliability/availability investigation. But this measure doesn't take into account the probability of state of system components. These disadvantage is eliminated in the calculation other importance measure as BI. The BI of a given component is defined as the probability that system is sensitive to inoperative of the ith system component [17, 35].

According to data in Table 4 and Eq. (11) the BI measures for the first component are calculated as:

$$IB_1^{1,2\to 1} = Pr\{\partial\phi(2 \to 1)/\partial x_1(1 \to 0) = 1\} = p_{2,1}p_{3,1}p_{4,1}$$

$$IB_1^{1,2\to 0} = Pr\{\partial\phi(2 \to 0)/\partial x_1(1 \to 0) = 1\} = 0$$

$$IB_1^{1,1\to 0} = Pr\{\partial\phi(1 \to 0)/\partial x_1(1 \to 0) = 1\} = p_{2,0}p_{3,1}p_{4,1} + p_{2,1}p_{3,0}p_{4,1}$$

Table 5 Structural importance for the human module in Fig. 3

i	$\rho_i^{s,j\to k}$			IS_i		
	$\rho_i^{1,2\to 1}$	$\rho_i^{1,2\to 0}$	$\rho_i^{1,1\to 0}$	$IS_i^{1,2\to 1}$	$IS_i^{1,2\to 0}$	$IS_i^{1,1\to 0}$
1	1	0	2	0.125	0	0.250
2	1	0	2	0.125	0	0.250
3	1	0	2	0.125	0	0.250
4	0	1	3	0	0.125	0.375

Table 6 BI for the human module in Fig. 2

i	$IB_i^{1,2\to 1}$	$IB_i^{1,2\to 0}$	$IB_i^{1,1\to 0}$
1	$p_{2,1}\ p_{3,1}\ p_{41}$	0	$p_{2,0}\ p_{3,1}\ p_{4,1} + p_{2,1}\ p_{3,0}\ p_{4,1}$
2	$p_{1,1}\ p_{3,1}\ p_{4,1}$	0	$p_{1,0}\ p_{3,1}\ p_{4,1} + p_{1,1}\ p_{3,0}\ p_{4,1}$
3	$p_{1,1}\ p_{2,1}\ p_{4,1}$	0	$p_{1,0}\ p_{2,1}\ p_{4,1} + p_{1,1}\ p_{2,0}\ p_{4,1}$
4	0	$p_{1,1}\ p_{2,1}\ p_{3,1}$	$p_{1,0}\ p_{2,1}\ p_{3,1} + p_{1,1}\ p_{2,0}\ p_{3,1} + p_{1,1}\ p_{2,1}\ p_{3,0}$

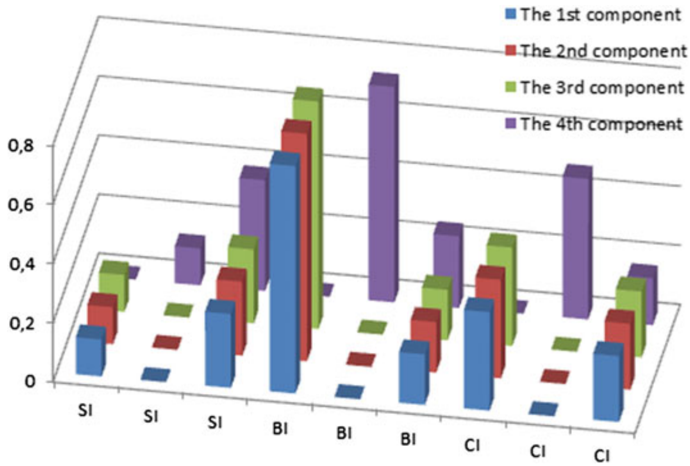


Fig. 4 IMs of the system "electronic transformers—merging unit" ($p_{1,1} = p_{2,1} = p_{3,1} = 0.9$ and $p_{4,1} = 0.95$)

The CI is calculated based on the BI (Table 6) according to Eq. (12). The values of the IM as SI, BI and CI for all components for $p_{1,1} = p_{2,1} = p_{3,1} = 0.9$ and $p_{4,1} = 0.95$ is shown in Fig. 4. Consider the BI of this system. The system degradation ($IB_i^{1,2\to 1}$) is more possible that the system fault ($IB_i^{1,1\to 0}$) depending of the 1st, 2nd or 3rd components failures according to BI. The probability of the component failure is not taken into account in this case. The CIs allow "improving" the estimation of the system. Based on the analysis of CI the system degradation ($IC_i^{1,2\to 1}$) and failure ($IC_i^{1,1\to 0}$) have similar values (Fig. 4). The influence of the component state probability to the value of the CI is illustrates the Fig. 5, where the p1, 1 = 0.6. The probability of the system degradation increases depending on the 1st component failure ($IC_1^{1,2\to 1}$) in comparison with other components (Fig. 5). Therefore the IMs allow investigate influence of all changes of the system performance level depending on the component fault for the system "electronic transformers—merging unit".

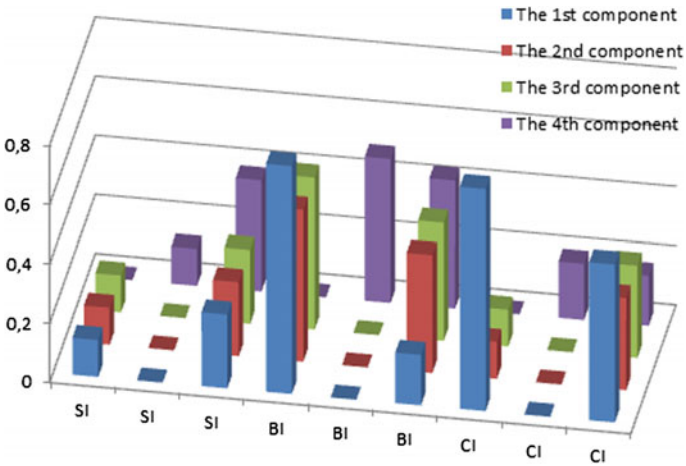


Fig. 5 IMs of the system "electronic transformers—merging unit" ($p_{1,1}$ = 0.6, $p_{2,1} = p_{3,1}$ = 0.9 and $p_{4,1}$ = 0.95)

3.4 Main Results of Using Proposed Models

Logical differential calculus has been developed for the analysis of dynamic behaviour of Boolean and MVL functions. The central concept of this tool is a logic derivative. There exist several types of logic derivatives from which the most interesting are DPLDs. In this paper, it has been shown that DPLDs are useful in finding correlation between degradation of system components and system degradation based on the analysis of the structure function. Therefore, they can be used to identify situations in which a degradation of a given component results in a decrease in system performance. Identification of such situations is very important for estimation of component influence on system activity, which is done via special reliability indices that are known as importance measures.

4 Conclusions

The RBD, the DD, the structure function and the DPLDs for RMSS "electronic transformers—merging unit" are considered.

The mathematical tool of logical differential calculus and DPLD in particular are used in many application problems. One of them is reliability engineering. Use of DPLDs in reliability analysis of MSSs has been considered in several works, e.g. [24, 25, 27]. The principal condition of the DPLD application in reliability analysis is the representation of system under investigation by the structure function (1).

We consider the calculation some of these measures as structural, Birnbaum's, and criticality for the analysis of the electronic transformers—merging unit (Fig. 2). The structure function of this unit is defined in Table 2 based on the operation conditions of this system (unit). The construction of this function allows estimating the most important components of this system in stationary state through structural, Birnbaum's, and criticality importance measures. But this analysis has been provided for the system in the stationary state only. The behaviour of the system in time cannot be considered under the structure function representation. Such analysis can be possible based on other system representation with application of other mathematical approaches for the system examination. As one of possible approaches can be used Markovian methods. We have been provided the analysis of the electronic transformers—merging unit based on Markovian methods to add the evaluation of this system depending on time characteristic. This evaluation will be presented in the second part of this paper.

References

1. Levitin, G., Lisnianski, A., Haim, H.: Redundancy optimization for series-parallel multi-state systems. IEEE Trans. Reliab. **47**(2), 165–172 (1998)
2. Gurler, U., Kaya, A.: A maintenance policy for a system with multi-state components: an approximate solution. Reliab. Eng. Syst. Saf. **76**, 117–127 (2002)
3. Huang, C., Yuan, J.: A two-stage preventive maintenance policy for a multi-state deterioration system. Reliab. Eng. Syst. Saf. **95**(11), 1255–1260 (2010)
4. Isaac, W., Nourelfath, M., Daoud, A.: Performance evaluation of multi-state degraded systems with minimal repairs and imperfect preventive maintenance. Reliab. Eng. Syst. Saf. **95**(2), 65–69 (2010)
5. Nahas, N., Khatab, A., Daoud, A.: Extended great deluge algorithm for the imperfect preventive maintenance optimization of multi-state systems. Reliab. Eng. Syst. Saf. **93**(11), 1658–1672 (2008)
6. Tan, C., Raghavan, N.: A framework to practical predictive maintenance modeling for multi-state systems. Reliab. Eng. Syst. Saf. **93**(8), 1138–1150 (2008)
7. Nourelfath, M., Dutuit, Y.: A combined approach to solve the redundancy optimization problem for multi-state systems under repair policies. Reliab. Eng. Syst. Saf. **86**(3), 205–213 (2004)
8. Yingkui, G., Jing, L.: Multi-state system reliability: a new and systematic Review. Procedia. Eng. **29**, 531–536 (2012)
9. Albasrawi, M., Jarus, N., Joshi, K., Sarvestani, S.: Analysis of Reliability and Resilience for Smart Grids. In: Proceedings of IEEE 38th Annual on Computer Software and Applications Conference (COMPSAC), pp. 529–534 (2014)
10. Zio, E.: Reliability Engineering: Old Problems and New Challenges. Reliab. Eng. Syst. Saf. **94**, 125–141 (2009)
11. Praks, P., Kopustinskas, V.: Monte-Carlo based Reliability Modelling of a Gas Network using Graph Theory Approach. In: 2014 Ninth International Conference on Availability, Reliability and Security. IEEE, pp. 380–386 (2014)
12. Zaitseva, E., Levashenko, V., Rusin, M.: Reliability Analysis of Healthcare System. In: 2011 Federated Conference on Computer Science and Information Systems FedCSIS 2011. IEEE, pp. 169–175 (2011)

13. Rausand, M., Høyland, A.: System Reliability Theory. John Wiley and Sons Inc, Haboken, NJ (2004)
14. Murchland, J.D.: Fundamental Concepts and Relations for Reliability Analysis of Multistate System. In: Reliability and Fault Tree Analysis, Theoretical and Applied Aspects of System Reliability. SIAM, pp. 581–618 (1975)
15. Barlow, R.E., Wu, A.S.: Coherent Systems with Multi-State Components. Math. Oper. Res. **3**, 275–281 (1978)
16. Hudson, J.C., Kapur, K.C.: Modules in Coherent Multistate Systems. IEEE Trans. Reliab. **32**, 183–185 (1983)
17. Lisnianski, A., Levitin, G.: Multi-state system reliability. assessment, optimization and applications. World Scientific, Singapore, SG (2003)
18. Xie, M., Dai, Y.-S., Poh, K.-L.: Multi-state system reliability. In: Computing system reliability. Models and Analysis. Kluwer Academic Publishers, New York, NY, 207–237 (2004)
19. Levitin, G., Lisnianski, A.: Optimization of imperfect preventive maintenance for multi-state system. Reliab. Eng. Syst. Saf. **67**, 193–203 (2000)
20. Zio, E., Marella, M., Podofillini, L.: A monte carlo simulation approach to the availability assessment of multi-state systems with operational dependencies. Reliab. Eng. Syst. Saf. **92**, 871–882 (2007)
21. Caldarola, L.: Coherent system with multi-state components. Nucl. Eng. Des. **58**, 127–139 (1980)
22. Veeraraghavan, M., Trivedi, K.S.: A combinatorial algorithm for performance and reliability analysis using multistate models. IEEE Trans. Comput. **43**, 229–234 (1994)
23. Reinske, K., Ushakov, I.: Application of graph theory for reliability analysis. Radio i Sviaz, Moscow, USSR (1988). (in Russian)
24. Zaitseva, E.: Reliability analysis of multi-state system. Dyn. Syst. Geom. Theor. **1**, 213–222 (2003)
25. Zaitseva, E., Levashenko, V., Kostlny, J., Kvassay M.: Analysis of Boundary States of Multi-State System by Direct Partial Logic Derivatives. In: Proceedings of 11th International Conference on ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer, CEUR Workshop, pp. 1356 (2015)
26. Miller, M.D., Thornton, M.A.: Multiple valued logic: concepts and representations. Morgan and Claypool Publishers, Synthesis Lectures on Digital Circuits and systems (2008)
27. Zaitseva, E., Levashenko, V.: Multiple-valued logic mathematical approaches for multi-state system reliability analysis. J. Appl. Logic **11**, 350–362 (2013)
28. Kvassay, M., Zaitseva, E., Levashenko, V.: Minimal cut sets and direct partial logic derivatives in reliability analysis, In: Proceedings of the Safety and Reliability: Methodology and Applications European Safety and Reliability Conference. CRC Press, pp. 241–248 (2014)
29. Kołowrocki, K.: Reliability of large and complex systems. Elsevier (2014)
30. Hung, J., Zuo, M.J.: Multi-State k-out-of-n System Model and its Applications. In: Ann. Reliability & Maintainability Symp. IEEE Pres, New York, 264–268 (2000)
31. Boedigheimer, R.A., Kapur, K.C.: Customer-Driven Reliability Models for Multistate Coherent Systems. IEEE Trans. Reliab. **43**, 46–50 (1994)
32. Nikolaidis, E. Ghiocel, D. M. Singhal, S. (eds.): Engineering Design Reliability, CRC Press (2005)
33. Natvig, B.: Multistate systems reliability theory with applications. Wiley, New York (2011)
34. Ramirez-Marquez, J.E., Coit, D.W., Tortorella, M.: A generalized multistate based path vector approach for multistate two-terminal reliability. IIE Trans. **38**, 477–488 (2006)
35. Kuo, W., Zhu, X.: Importance measures in reliability, risk, and optimization: principles and applications, Wiley (2012)
36. Nourelfath, M., Chatelet, E., Nahas, N.: Joint redundancy and imperfect preventive maintenance optimization for series–parallel multi-state degraded systems. Reliab. Eng. Syst. Saf. **103**, 51–60 (2012)

37. Miller, M.D., Thornton, M.A.: Multiple Valued Logic: Concepts and Representations. Morgan and Claypool Publishers, Synthesis Lectures on Digital Circuits and systems (2008)
38. Zaitseva, E.: Importance Analysis of a Multi-State System Based on Multiple-Valued Logic Methods. In: Lisnianski, A., Frenkel, I. (eds.) Recent Advances in System Reliability: Signatures, Multi-state Systems and Statistical Inference, pp. 113–134. Springer, London, UK (2012)
39. Wood, A.P.: Multistate Block Diagrams and Fault Trees, IEEE Trans on Reliability, R-34, pp. 236–240 (1985)

Hybrid Adaptive Systems of Computational Intelligence and Their On-line Learning for Green IT in Energy Management Tasks

Yevgeniy Bodyanskiy, Olena Vynokurova, Iryna Pliss and Dmytro Peleshko

Abstract In this book chapter, we have considered a topical problem of intelligent energy management, which arises in the context of an intensively developed science direction—Green IT. The hybrid neuro-neo-fuzzy system and its high-speed learning algorithm are proposed. This system can be used for on-line prediction of essentially non-stationary nonlinear chaotic and stochastic time series, which describe electrical load producing and consuming processes. The considered hybrid adaptive system of computational intelligence has some advantages over the conventional artificial neural networks and neuro-fuzzy systems. The proposed hybrid neuro-neo-fuzzy prediction system provides a high quality load prediction that is very important for power systems.

Keywords Computational intelligence · On-line learning · Green IT · Hybrid adaptive systems · Neuro-neo-fuzzy system

1 Introduction

Nowadays, Intelligent Energy Management (IEM) has become one of the major research fields both in Electrical Engineering and Information Technologies (IT) with the help of contemporary achievements of Computer Science (CS) and, first of all, Computational Intelligence approaches [1]. Energy management

Y. Bodyanskiy (✉) · O. Vynokurova · I. Pliss
Kharkiv National University of Radio Electronics, 14 Nauky av., Kharkiv 61166, Ukraine
e-mail: yevgeniy.bodyanskiy@nure.ua

O. Vynokurova
e-mail: olena.vynokurova@nure.ua

I. Pliss
e-mail: iryna.pliss@nure.ua

D. Peleshko
Lviv Politechnic National University, 12 S. Bandery st., Lviv 79013, Ukraine
e-mail: dpeleshko@gmail.com

V. Kharchenko et al. (eds.), *Green IT Engineering: Concepts, Models, Complex Systems Architectures*, Studies in Systems, Decision and Control 74,
DOI 10.1007/978-3-319-44162-7_12

problems are important both for industry and private homeowners and tenants under the development and operation of Smart Homes [2–15].

There is an evolution of a new research direction also known as Green Computing (notably, Green IT Energy Management). An important part of IEM is problem of on-line short time load forecasting which is usually solved by conventional forecasting methods from exponential smoothing, regression, correlation, spectral analysis up to adaptive forecasting models. But these methods proved their ineffectiveness in many cases. Time series which describe electricity production and consumption are essentially non-stationary, nonlinear, stochastic, chaotic and may contain abrupt changes and seasonal periodic components and can be influenced by many controllable (first of all, meteorological conditions) and uncontrollable disturbances.

Under these conditions, intelligent computational methods [4–20] and, first of all, artificial neural networks (ANNs) and hybrid systems such as neuro-fuzzy systems, wavelet-neuro-fuzzy-systems, neo-fuzzy systems and etc. have demonstrated their efficiency for solving prediction tasks. These systems combine flexibility, universal approximation properties and learning abilities (they take these properties from ANNs), results' transparency and interpretability (provided by fuzzy reasoning systems), the possibility of local features compact description for non-stationary signals (provided by wavelet systems).

These systems have demonstrated their advantage while solving Data Mining tasks [21] like electric load forecasting, but they don't work well for on-line information processing which is provided Data Streams Mining [22, 23].

Forecasting systems which are widely used nowadays in IEM are based on either multilayer ANNs (such as a multilayer perceptron, neuro-fuzzy-systems like ANFIS [24]) or some similar architectures. A learning process for these systems is implemented in a multi-epoch batch mode and it is assumed that a learning sample is given a priori, and its volume is not being changed during a process of parameters' adjustment process.

For data processing in the real-time mode so-called generalized additive models [25, 26] are suited in the best way, but however they aren't designated for the analysis of non-stationary nonlinear chaotic signals.

It should be noticed, most of these systems have only one output signal, while most of energy management real tasks are described by MIMO (multi inputs—multi outputs) models.

It is reasonable to synthesize multivariate adaptive systems of computational intelligence which are aimed at solving energy management tasks, and, first of all, prediction of nonlinear non-stationary stochastic and chaotic signals in many cases under conditions of small learning sample.

The main requirements for these systems are both high extrapolation properties and computational simplicity, which allows their implementation in embedded systems, high speed for the learning process, a small number of tuned parameters and the most important feature its possibility of on-line processing multivariate data stream, which are fed in a sequential mode.

2 An On-line Hybrid System of Computational Intelligence for Data Stream Adaptive Processing

In many cases, Takagi-Sugeno-Kang neuro-fuzzy system satisfies all these requirements [27, 28]. It takes place due to its relatively small number of synaptic weights to be tuned and possibility of adaptive linear identification learning algorithms, which are used for their tuning.

At the same time, using polynomial approximation that is implemented by the TSK-systems often cannot be effective for non-stationary and chaotic signals processing. It seems appropriate to develop hybrid systems which are not connected with polynomial or harmonical approximation and extrapolation.

This system has to combine universal approximation properties of artificial neural networks, interpretability and transparency of neuro-fuzzy systems and a high learning rate and flexibility of neo-fuzzy systems [29–32].

Figure 1 shows architecture of a hybrid neuro-neo-fuzzy system (HNNFS) that consists of three layers for information processing. This system is hybrid model of

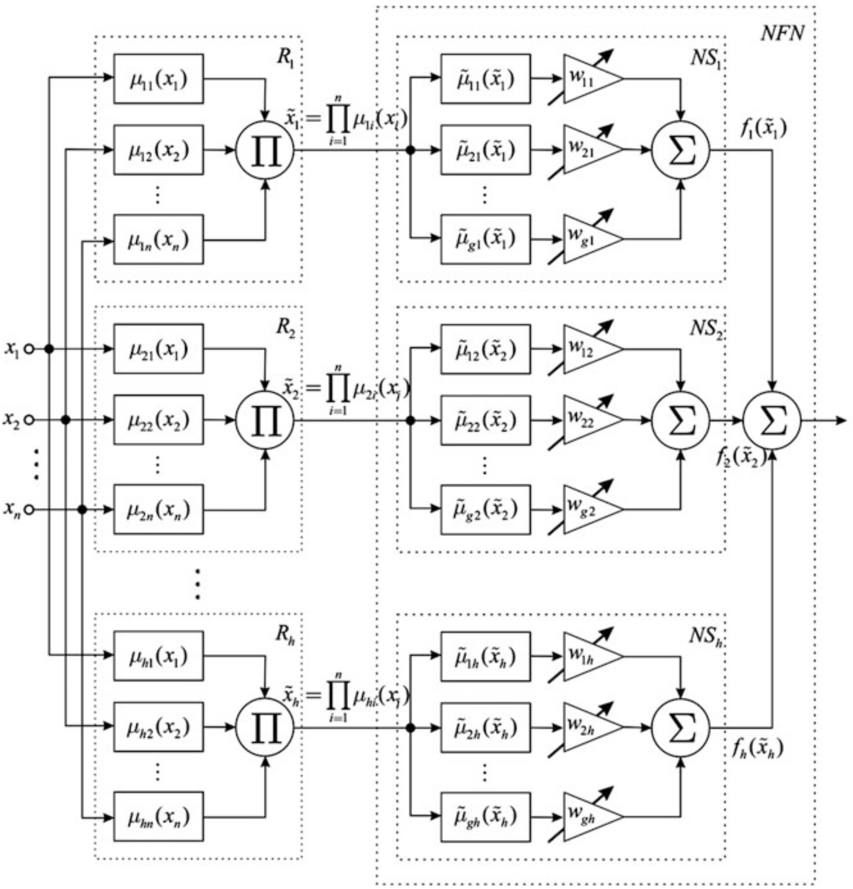


Fig. 1 The hybrid neuro-neo-fuzzy system

TSK-neuro-fuzzy system and the RBF-NFN architecture that was introduced in [33, 34].

A $(n \times 1)$-dimensional vector of input signals $x(k) = (x_1(k), \ldots, x_n(n))^T \in R^n$ are fed to the input layer, where $k = 1, 2, \ldots$ is current time. The first hidden layer consists of hn membership functions $\mu_{li}(x_i)$, $l = 1, 2, \ldots, h$ and provides fuzzification of the input variables.

The second layer implements aggregation of membership levels that are computed in the first layer and consists of h multiplier units. And, finally, the third (output) layer is formed by a neo-fuzzy neuron (NFN) [29–31] with h inputs and a scalar output signal $\hat{y}(k)$.

Therefore, if the vector signal $x(k)$ is fed to the system's input than the first layer's elements compute the membership levels $0 < \mu_{li}(x_i) \leq 1$ (the bell-shaped membership functions with nonstrict local receptive fields are used) for avoiding "gaps" appearance in the fuzzificated space under scatter partition of the input variables' space. Gaussian membership functions are commonly used in the first layer in the form

$$\mu_{li}(x_i(k)) = \exp\left(-\frac{(x_i(k) - c_{li})^2}{2\sigma_i^2}\right) \tag{1}$$

where c_{li}, σ_i are scalar parameters of centers and widths correspondingly.

Nodes in the second hidden layer compute aggregated values in the form

$$\tilde{x}_i(k) = \prod_{i=1}^{n} \mu_{li}(x_i(k)) \tag{2}$$

and we can write an expression for the Gaussian functions with equal values of width parameters σ in the form

$$\tilde{x}_i(k) = \prod_{i=1}^{n} \exp\left(-\frac{(x_i(k) - c_{li})^2}{2\sigma^2}\right) = \exp\left(-\frac{\|x_i(k) - c_{li}\|^2}{2\sigma^2}\right) \tag{3}$$

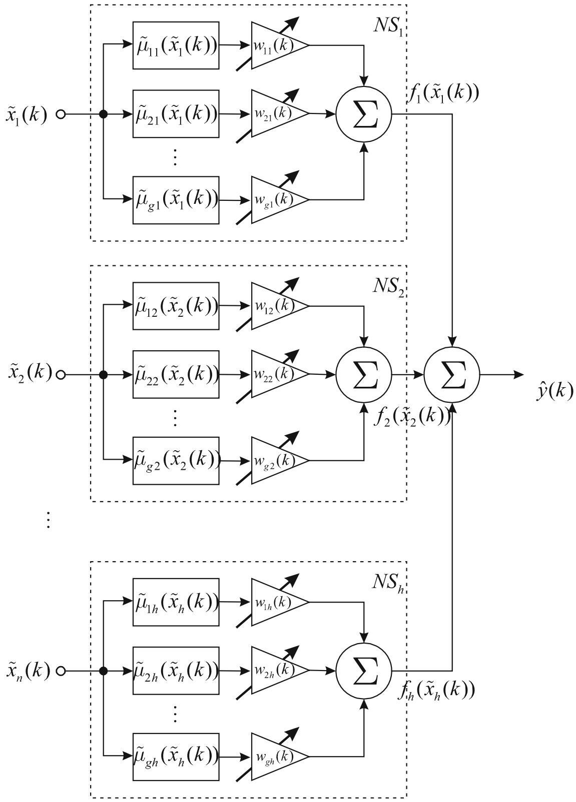
(here $c_l = (c_{l1}, \ldots, c_{li}, \ldots, c_{\mathrm{ln}})^T$), i.e. elements of the first and second hidden layers process the input signals similarly to R-neurons of Radial Basis Functions Neural networks (RBFNs).

These signals are fed to inputs of the nonlinear synapses $NS_1, NS_2, \ldots, NS_h$ which form the neo-fuzzy neuron's (NFN) architecture, which is shown in Fig. 2.

The neo-fuzzy neuron is a nonlinear learning system with multiple inputs and a single output, which provides the mapping [29]

$$\hat{y}(k) = \sum_{l=1}^{h} f_l(\tilde{x}_l(k)) \tag{4}$$

Fig. 2 The neo-fuzzy neuron



where each nonlinear synapse NS_l consists of g membership functions $\tilde{\mu}_{jl}(\tilde{x}_l)$, $l = 1, 2, \ldots, g$ and the same number of synaptic weights w_{jl} to be tuned.

Therefore, the transformation that is implemented by each nonlinear synapse can be written in form

$$f_l(\tilde{x}_l(k)) = \sum_{j=1}^{g} w_{jl}\tilde{\mu}_{jl}(\tilde{x}_l(k)) \tag{5}$$

and the neo-fuzzy neuron provides the nonlinear mapping:

$$\hat{y}(k) = \sum_{l=1}^{h}\sum_{j=1}^{g} w_{jl}\tilde{\mu}_{jl}(\tilde{x}_l(k)) = w^T\tilde{\mu}(\tilde{x}(k)) \tag{6}$$

where $w = (w_{11}, \ldots, w_{g1}, w_{12}, \ldots, w_{jl}, \ldots, w_{gh})^T$, $\tilde{\mu}(\tilde{x}(k)) = (\tilde{\mu}_{11}(\tilde{x}_1(k)), \ldots, \tilde{\mu}_{g1}(\tilde{x}_1(k)), \tilde{\mu}_{12}(\tilde{x}_2(k)), \ldots, \tilde{\mu}_{jl}(\tilde{x}_l(k)), \ldots, \tilde{\mu}_{gh}(\tilde{x}_h(k)))^T$ are $(gh \times 1)$-dimensional vectors.

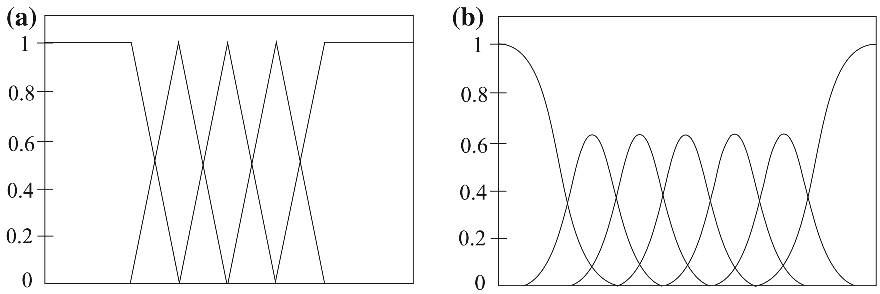


Fig. 3 B-spline membership functions of order 2 (**a**) and 4 (**b**)

Triangular functions are usually used as the membership functions in the neo-fuzzy neuron. Their values are determined by a distance between a current value of the input signal $\tilde{x}_l$ and center parameters c_{jl}:

$$\tilde{\mu}_{jl}(\tilde{x}_l) = \begin{cases} \frac{\tilde{x}_l - c_{j-1,l}}{c_{jl} - c_{l,l-1}}, \ \tilde{x}_l \in [c_{j-1,l}, c_{jl}], \\ \frac{c_{j+1,l} - \tilde{x}_l}{c_{l,l+1} - c_{jl}}, \ \tilde{x}_l \in [c_{jl}, c_{j+1,l}], \\ 0 \quad otherwise. \end{cases} \tag{7}$$

To solve more complex tasks, it is possible to use B-spline membership functions, which are shown in Fig. 3 for different values of a q parameter.

Using a sequence of ordered knots $\{c_1, c_2, \ldots, c_m\}$, the pth B-spline basis function of order q is defined as [26]

$$N_{p,q}(x) = \begin{cases} \left\{ \begin{array}{ll} 1, & for\ c_p \le x < c_{p+1}, \\ 0, & otherwise, \end{array} \right\} & if\ q = 1, \\ \left. \begin{array}{l} \frac{x - c_p}{c_{p+q-1} - c_p} N_{p,q-1}(x) \\ + \frac{c_{p+q} - x}{c_{p+q} - c_{p+1}} N_{p+1,q-1}(x), \end{array} \right\} & if\ q > 1, \\ p = 1, \ldots, m - q, \quad m \ge q. \end{cases} \tag{8}$$

It can be noticed that if $q = 2$ triangular membership functions are obtained, if $q = 3$—quadratic B-splines, if $q = 4$—cubic B-splines.

It is important to notice that this configuration of the membership functions provides the unity partitioning automatically

$$\sum_{j=1}^{g} \tilde{\mu}_{jl}(\tilde{x}_l) = 1, \ \forall l \tag{9}$$

that allows to reduce significantly the proposed system's realization by exception of the normalization-defuzzification layer in comparison to the conventional TSK-system.

The hybrid neuro-neo fuzzy system's learning actually deals with tuning the neo-fuzzy neuron's synaptic weights, which compose the system's output layer. Due to the fact that the system's output signal $\hat{y}(k)$ depends linearly on a vector of the synaptic weights w, a conventional exponentially weighted recurrent least squares method can be used for their learning. This method is actually the Gaussian-Newtonian optimization procedure of the second order:

$$\begin{cases} w(k) = w(k-1) + \frac{P(k-1)(y(k) - w^T(k-1)\tilde{\mu}(\tilde{x}(k)))\tilde{\mu}(\tilde{x}(k))}{\alpha + \tilde{\mu}^T(\tilde{x}(k))P(k-1)\tilde{\mu}(\tilde{x}(k))}, \\ P(k) = \frac{1}{\alpha}\left(P(k-1) - \frac{P(k-1)\tilde{\mu}(\tilde{x}(k))\tilde{\mu}^T(\tilde{x}(k))P(k-1)}{\alpha + \tilde{\mu}^T(\tilde{x}(k))P(k-1)\tilde{\mu}(\tilde{x}(k))}\right) \end{cases} \tag{10}$$

where $0 < \alpha \leq 1$ is a forgetting factor parameter, $y(k)$ is an external reference signal.

In a case of non-stationary chaotic signal processing, it's recommended to use a gradient learning procedure in the form [35]

$$w(k) = w(k-1) + \frac{(y(k) - w^T(k-1)\tilde{\mu}(\tilde{x}(k)))\tilde{\mu}(\tilde{x}(k))}{\tilde{\mu}^T(\tilde{x}(k))\tilde{\mu}(\tilde{x}(k))} \tag{11}$$

which is the adaptive optimal speed Kaczmarz-Widrow-Hoff gradient learning algorithm.

3 A Multivariate Hybrid Adaptive Neuro-Neo-Fuzzy System

The considered neuro-neo-fuzzy system implements the nonlinear mapping $R^n \to R^1$ like most of the well-known neuro-fuzzy systems, i.e. this system allows predicting only one-dimensional signals. Most of the energy management tasks assume the multivariate time series analysis. Of course, we can use some number of one-dimensional systems in a parallel way as it is proposed in [15], where a task of the multivariate signals analysis is solved with the help of an ensemble of ANFISs. However, both the computational complexity and a number of tuning parameters are increasing.

In this connection, we propose to use a multivariate neuro-neo-fuzzy system, which implements the nonlinear mapping $R^n \to R^m$ and its architecture is shown in Fig. 4.

The first two layers of the system are the same as the system's layers in Fig. 1. The output layer of the proposed system is formed by a generalized neo-fuzzy neuron (GNFN) [36], its components are multivariate nonlinear synapse MNS_l, $l = 1, 2, \ldots, h$. Each multivariate nonlinear synapses has one input and m outputs $\hat{y}_p$, $p = 1, 2, \ldots, m$ and consists of g membership functions $\tilde{\mu}_{jl}$, $j = 1, 2, \ldots, g$ and gm tuning synaptic weights w_{pjl}.

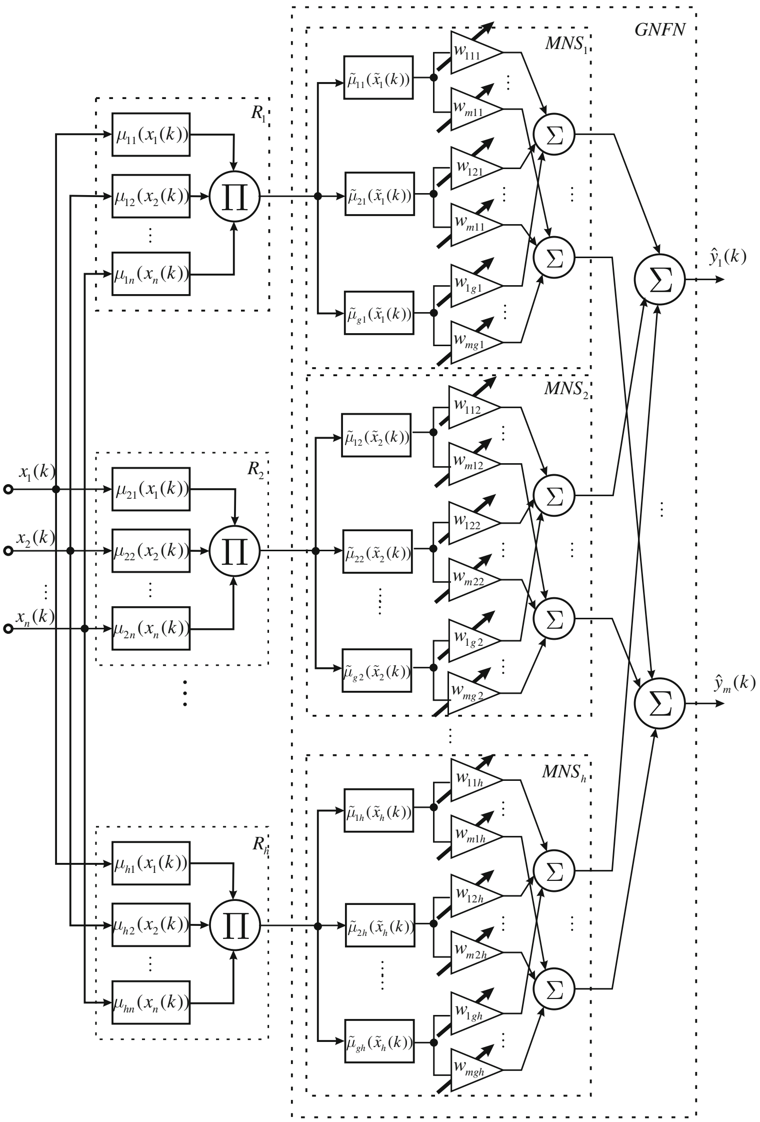


Fig. 4 The multivariate hybrid neuro-neo-fuzzy system

Therefore, when signals $\tilde{x}_1(k), \tilde{x}_2(k), \ldots, \tilde{x}_h(k)$ are fed to the input of the third layer, the membership functions of this layer compute values $\tilde{\mu}_{jl}(\tilde{x}_l(k))$, which are further multiplied by corresponding synaptic weights w_{pjl}.

Finally, the pth system's output can be written in the form

$$\hat{y}_p(k) = \sum_{l=1}^{h}\sum_{j=1}^{g} w_{pjl}\tilde{\mu}_{jl}(\tilde{x}_l(k)),\ p = 1, 2, \ldots, m. \tag{12}$$

Introducing a $(gh \times 1)$-vector of membership functions $\tilde{\mu}(\tilde{x}(k)) = (\tilde{\mu}_{11}(\tilde{x}_1(k)), \ldots, \tilde{\mu}_{g1}(\tilde{x}_1(k)), \tilde{\mu}_{12}(\tilde{x}_2(k)), \ldots, \tilde{\mu}_{jl}(\tilde{x}_l(k)), \ldots,\ \tilde{\mu}_{gh}(\tilde{x}_h(k)))^T$, a $(m \times 1)$-vector of the output's signals $\hat{y}(k) = \left(\hat{y}_1(k), \ldots, \hat{y}_p(k), \ldots, \hat{y}_m(k)\right)^T$, and a $(m \times gh)$-matrix of the synaptic weights

$$W(k) = \begin{pmatrix} w_{111}(k) & w_{121}(k) & \cdots & w_{1gh}(k) \\ w_{211}(k) & w_{221}(k) & \cdots & w_{2gh}(k) \\ \vdots & \vdots & \ddots & \vdots \\ w_{m11}(k) & w_{m21}(k) & \cdots & w_{mgh}(k) \end{pmatrix} \tag{13}$$

accordingly to previous notations, we can write a vector of the output's signal in a compact form

$$\hat{y}(k) = W\tilde{\mu}(\tilde{x}(k)). \tag{14}$$

Adjusting the weights' matrix W can be provided by a matrix modification procedure of the algorithms (10), (11), which can be rewritten in form

$$\begin{cases} W(k) = W(k-1) + \frac{(y(k) - W(k-1)\tilde{\mu}(\tilde{x}(k)))\tilde{\mu}^T(\tilde{x}(k))P(k-1)}{\alpha + \tilde{\mu}^T(\tilde{x}(k))P(k-1)\tilde{\mu}(\tilde{x}(k))}, \\ P(k) = \frac{1}{\alpha}\left(P(k-1) - \frac{P(k-1)\tilde{\mu}(\tilde{x}(k))\tilde{\mu}^T(\tilde{x}(k))P(k-1)}{\alpha + \tilde{\mu}^T(\tilde{x}(k))P(k-1)\tilde{\mu}(\tilde{x}(k))}\right), \quad 0 < \alpha \le 1 \end{cases} \tag{15}$$

and

$$\begin{aligned} W(k) &= W(k-1) + \frac{(y(k) - W(k-1)\tilde{\mu}(\tilde{x}(k)))\tilde{\mu}^T(\tilde{x}(k))}{\tilde{\mu}^T(\tilde{x}(k))\tilde{\mu}(\tilde{x}(k))} \\ &= W(k-1) + (y(k) - W(k-1)\tilde{\mu}(\tilde{x}(k)))\tilde{\mu}^+(\tilde{x}(k)) \end{aligned}. \tag{16}$$

where $\tilde{\mu}^+(\tilde{x}(k)) = \frac{\tilde{\mu}^T(\tilde{x}(k))}{\|\tilde{\mu}(\tilde{x}(k))\|^2}$ is a pseudo-inverse vector to $\tilde{\mu}(\tilde{x}(k))$.

It should be noticed that although the learning algorithms (10), (11), (15), (16), provide a maximum possible speed, they still have some disadvantages. So, the exponential weighted recurrent least square method can be numerically unstable for small values of the forgetting factor α. This fact doesn't allow using this method for processing the significantly non-stationary time series, which are the subject of analysis in IEM.

Although the Kaczmarz-Widrow-Hoff algorithm is the most high-speed gradient procedure, it is sensitive to the disturbance and noise influence that always occurs in real signals.

In this connection, it is appropriate to use a "hybrid" method of the exponential weighted recurrent least squares method and the Kaczmarz algorithm to tune the GNFN in the form [37]

$$\begin{cases} W(k) = W(k-1) + r^{-1}(k)(y(k) - W(k-1)\tilde{\mu}(\tilde{x}(k)))\tilde{\mu}^T(\tilde{x}(k)), \\ r(k) = \alpha r(k-1) + \tilde{\mu}^T(\tilde{x}(k))\tilde{\mu}(\tilde{x}(k)), \quad 0 \le \alpha \le 1. \end{cases} \tag{17}$$

This algorithm is stable for arbitrary values of a smoothing parameter and provides a compromise between smoothing and tracking properties of the learning process.

4 Experimental Results of the Hybrid Adaptive Neuro-Neo-Fuzzy Systems in the Energy Management Tasks

4.1 *Simulation of a Single Output Hybrid Adaptive Neuro-Neo-Fuzzy System*

To demonstrate the efficiency of the proposed hybrid neuro-neo-fuzzy system and its learning algorithm, some experiments have been performed in the area of hourly energy consumption for the prediction in one of Germany federal lands [38, 39].

A number of inputs proposed hybrid neuro-neo-fuzzy system is n = 7 so an input vector can be written in the form

$$\begin{aligned} x(k+h) = (&x(k), x(k-h), \Delta x(k), x(k+h-24), x(k+h-168), \\ &[k/42], [k/168], [k/(366*24)]) \end{aligned}$$

where h is a prediction horizon, k is a discrete time value, $x(k+h)$ is a value of energy consumption to be predicted, $x(k)$ is a current value of energy consumption, $x(k-h)$ is a value of energy consumption h-steps ago, $x(k+h-24)$ is a value of energy consumption twenty-four hours ago from the prediction value, $x(k+h-168)$ is a value of energy consumption one week ago from the prediction value, $[k/42]$ is a hour number in a day, $[k/168]$ is a hour number in a week, $[k/(366*24)]$ is a hour number in a year. The initial values of the synaptic weights were taken equal to zero.

The mean-square error (MSE) and mean absolute percentage error (MAPE) were used as criteria for the prediction quality.

Figure 5 shows results of energy consumption prediction. Two curves, representing an actual value (a dot line) and a forecasting value (a solid line), are almost indistinguishable.

Table 1 shows a comparative analysis of the energy consumption prediction which are based on different approaches.

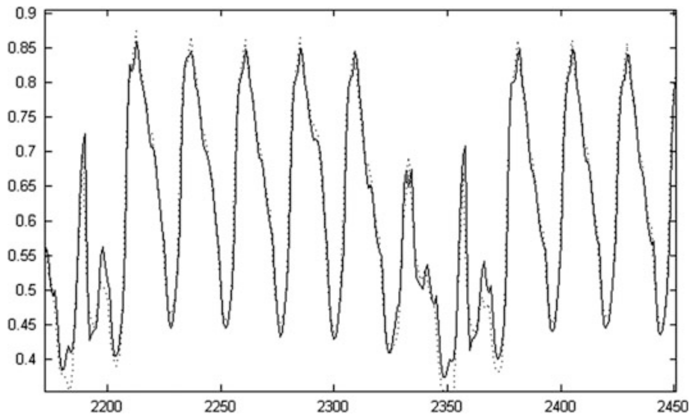


Fig. 5 The results of the energy consumption prediction

Table 1 The comparison analysis of the energy consumption prediction

Neural network/learning algorithms	MSE	MAPE (%)
Hybrid neuro-neo-fuzzy system/proposed learning algorithm	0.0008	3.9
Neo-fuzzy neuron/recurrent least squares learning algorithm	0.002	5.76
Adaptive neuro-fuzzy inferences system (TSK)/gradient learning algorithm	0.0101	10.3
Radial basis function neural network (RBFN)/recurrent least squares learning algorithm	0.0834	13.4

Thus, it can be seen from the experimental results that the proposed hybrid neuro-neo-fuzzy system and its learning algorithm provide the best prediction quality.

4.2 *Experimental Simulation of Multivariate Hybrid Adaptive Neuro-Neo-Fuzzy System*

Data used for this example are historical hourly temperatures, system loads and day-ahead electricity prices from the New England Pool region [40].

A number of inputs and outputs for the proposed adaptive multivariate hybrid neuro-fuzzy system is accordingly $n = 8$ and $m = 3$ so an input vector can be written in the form

$$[x_1(k+h), x_2(k+h), x_3(k+h)] = [x_1(k), x_1(k-h), x_1(k+h-24), x_1(k+h-168), x_2(k), x_2(k-h), x_2(k+h-24), x_2(k+h-168), x_3(k), x_3(k-h), x_3(k+h-24), x_3(k+h-168), hd, dw, hw]$$

where $x_1(k+h), x_2(k+h), x_3(k+h)$ are a prediction value for energy consumption, dry bulb temperature and dew point temperature; $x_1(k), x_2(k), x_3(k)$ are their current values; $x_1(k-h), x_2(k-h), x_3(k-h)$ are their values h steps ago; $x_1(k+h-24), x_2(k+h-24), x_3(k+h-24)$ are their values twenty-four hours ago from the prediction value; $x_1(k+h-168), x_2(k+h-168)$, $x_3(k+h-168)$ are their values one week ago from the prediction value; *hd* is a day hour; *dw* is a day of the week; *hw* is a holiday/weekend indicator (0 or 1).

The initial values of the synaptic weights were taken equal to zero. Figure 6 shows the results of energy consumption prediction, Fig. 7 shows the results of dry bulb temperature prediction and Fig. 8 shows the results of dew point temperature prediction. Two curves, representing an actual value (a dot green line) and a forecasting value (a solid blue line), are almost indistinguishable. A prediction error is presented by a solid red line.

Table 2 demonstrates a comparative analysis for time series prediction energy consumption, which are based on the different approaches.

Thus, it can be seen from the experimental results that the proposed adaptive multivariate hybrid neuro-fuzzy system and its learning algorithm provides the best prediction quality.

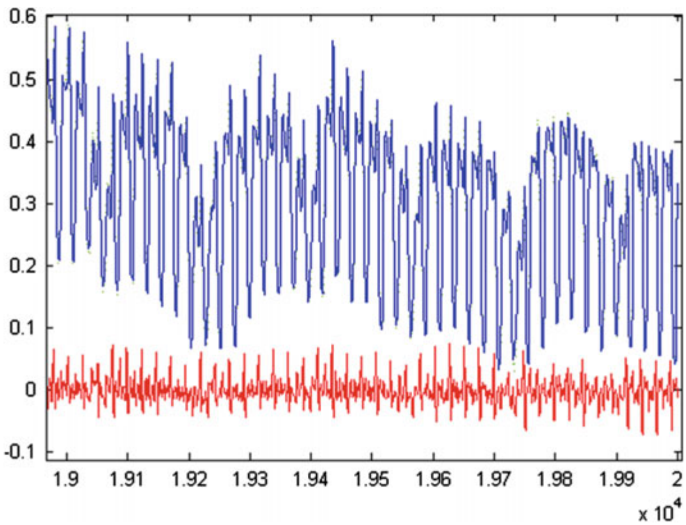


Fig. 6 The results of the energy consumption prediction

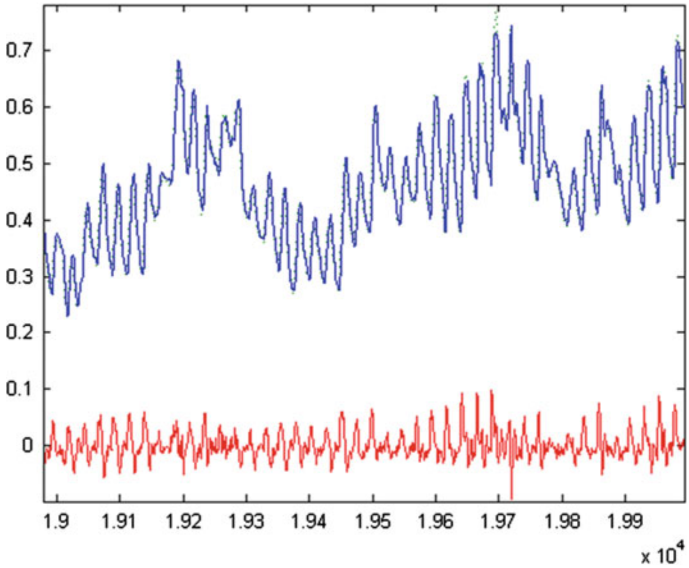


Fig. 7 The results of the dry bulb temperature prediction

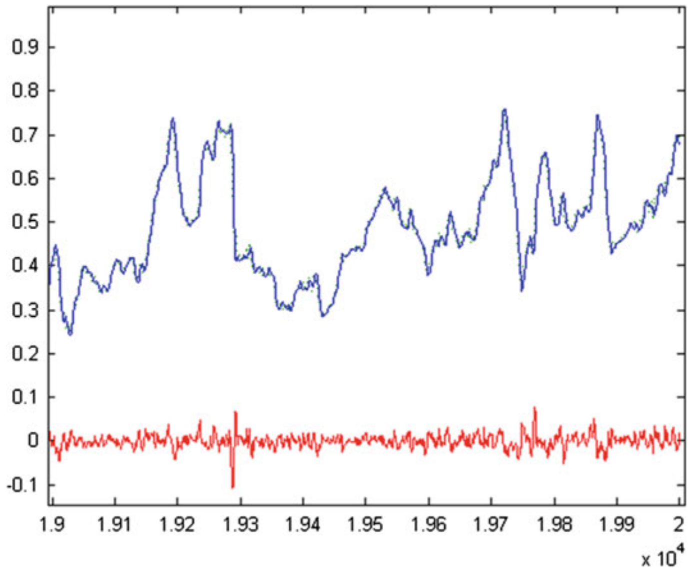


Fig. 8 The results of the dew point temperature prediction

Table 2 The comparative analysis of multivariate time series prediction

Neural network/learning algorithms	MSE			MAPE		
	x_1	x_2	x_3	x_1 (%)	x_2 (%)	x_3 (%)
Hybrid multivariate neuro-neo-fuzzy system/proposed learning algorithm	0.028	0.024	0.018	5.7	3.1	2.2
Multivariate neo-fuzzy neuron/recurrent least squares learning algorithm	0.078	0.072	0.043	7.6	6.9	5.7
Adaptive multivariate neuro-fuzzy inferences system (TSK)/gradient learning algorithm	0.058	0.042	0.023	4.6	5.1	4.7
Multivariate neuro-fuzzy system of Wang-Mendel/gradient learning algorithm	0.06	0.05	0.034	4.9	6.1	5.1

5 Conclusions

In this book chapter we have considered a problem, which connects with Green IT energy management and is the short-term adaptive prediction of load forecasting. The hybrid neuro-neo-fuzzy system and its high speed learning algorithm are proposed. This system can be used for on-line prediction of the essentially non-stationary nonlinear chaotic and stochastic time series, which describe the processes of electrical load producing and consuming. The considered hybrid adaptive system of computational intelligence has some advantages over the conventional artificial neural networks and neuro-fuzzy systems. First of all, its computational simplicity that allows its implementation in the embedded systems, and high learning speed that allows its tuning based on a short data sample with some noisy observations and outliers. In our future work, we assume to expand the proposed approach to processing high dimensional tasks under conditions of Big Data problems.

References

1. Abraham, A., Khan, M.R.: Neuro-fuzzy paradigms for intelligent energy management. In: Abraham, A., Jain, L., Zwaag, B.J. (eds.) Innovations in Intelligent Systems, Part 2, pp. 285–314. Springer, Berlin (2004)
2. Khotanzad, A., Hwang, R.C., Abaye, A., Maratukulam, D.: An adaptive modular artificial neural network hourly load forecaster and its implementation at electric utilities. IEEE Trans. Power Syst. **10**(3), 1716–1722 (1995)
3. Kiartzis, S.J., Zoumas, C.E., Theocharis, J.B., Bakirtzis, A.G., Petridis, V.: Short-term load forecasting in an autonomous power system using artificial Neural Networks. IEEE Trans. Power Syst. **12**(4), 1591–1596 (1997)
4. Khotanzad, A., Afkhami-Rohani, R., Maratukulam, D.: ANNSTLF-artificial neural network short-term load forecaster-generation tree. IEEE Trans. Power Syst. **13**(4), 1413–1422 (1998)

5. Abraham, A.: An evolving fuzzy neural network model based reactive power control. In: Proceedings of the Second International Conference on Computers in Industry, pp. 247–253, Bahrain (2000)
6. Abraham, A., Nath, B.: A neuro-fuzzy approach for forecasting electricity demand in victoria. Appl. Soft Comput. **1**(2), 127–138 (2001)
7. Hippert, H.S., Pedreira, C.E., Souza, R.C.: Neural networks for short-term load forecasting: a review and evaluation. IEEE Trans. Power Syst. **16**(1), 44–55 (2001)
8. Khan, M.R., Zak, L., Ondrusek, C.: Forecasting weekly load using a hybrid fuzzy-neural network approach. Int. J. Eng. Mech. **5**, 327–336 (2001)
9. Bodyanskiy, Y., Popov, S., Rybalchenko, T.: Multilayer neuro-fuzzy network for short term electric load forecasting. In: Hirsch, E.A., Razborov, A.A., Semenov, A., Slissenko, A. (eds.) Third International Computer Science Symposium in Russia, CSR 2008 Moscow, Russia, 7–12 June 2008. Lecture Notes in Computer Science, vol. 5010, pp. 339–348, Springer, Berlin (2008)
10. Bodyanskiy, Y., Popov, S., Rybalchenko, T.: Feedforward neural network with a specialized architecture for estimation of the temperature influence on the electric load. In: Proceedings Forth International IEEE Conference on Intelligent Systems, Golden Sands Resort, vol. 1, pp. 7.14–7.18, Varna, Bulgaria (2008)
11. Bodyanskiy, Y., Otto, P., Babenko, A., Popov, S.: Neural network approach to signals parameters estimation in electric power systems. In: Proceedings 55 International Wiss. Koll. “Crossing Borders within ABC Automation, Biomedical Engineering and Computer Science”, pp. 173–179, Ilmenau, TUI (2010)
12. Roessler, F., Teich, T., Franke, S.: Neural networks for smart homes and energy efficiency. In: Katalinic, B. (ed.) DAAAM International Scientific Book, pp. 305–314. DAAAM International Publishing, Vienna (2012)
13. Reaz, M.B.I.: Artificial intelligence techniques for advanced smart home implementation. Acta Technica Corvininesis Bull. Eng. **6**(2), 51–57 (2013)
14. Lee, S-H., Lee, S-J., Moon, K-I.: An ANFIS control system for smart home. Int. J. Adv. Comput. Technol. **5**(11), 464–470 (2013)
15. Lee, S.-H., Lee, S.-J., Moon, K.-I.: Smart home security system using multiple ANFIS. Int. J. Smart Home **7**(3), 121–132 (2013)
16. Kim, J.H., Lee, M.J. (eds.): Green IT: Technologies and Applications. Springer, Berlin (2011)
17. Rutkowski, L.: Computational Intelligence: Methods and Techniques. Springer, Berlin (2008)
18. Mumford, C.L.: Computational Intelligence Collaboration, Fusion and Emergence. Springer, Berlin (2009)
19. Kruse, R., Borgelt, C., Klawonn, F., Moewes, C., Steinbrecher, M., Held, P.: Computational Intelligence. A Methodological Introduction. Springer, Berlin (2013)
20. Du, K.-L., Swamy, M.N.S.: Neural Networks and Statistical Learning. Springer, London (2014)
21. Aggarwal, C.C.: Data Mining: The Textbook. Springer, Switzerland (2015)
22. Bifet, A.: Adaptive Stream Mining: Pattern Learning and Mining from Evolving Data Streams. IOS Press, Amsterdam (2010)
23. Aggarwal, C. (ed.): Data Stream: Models and Algorithms. Springer Science+Bussiness Media, LLC, N.Y. (2007)
24. Jang, R.J.-S.: ANFIS: Adaptive Network Based Fuzzy Inference Systems. IEEE Trans. Syst. Man Cybern. **23**(3), 116–132 (1993)
25. Hastie, T., Tibshirani, R.: Generalized Additive Models. Chapman and Hall, London (1990)
26. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer (2013)
27. Takagi, T., Sugeno, M.: Fuzzy Identification of Systems and its Applications to Modeling and Control. IEEE Trans. Syst. Man Cybern. **15**(1), 116–132 (1985)
28. Sugeno, M., Kang, G.: Structure identification of fuzzy model. Fuzzy Sets Syst. **28**(1), 15–33 (1988)

29. Yamakawa, T., Uchino, E., Miki, T., Kusanagi, H.: A neo-fuzzy neuron and its applications to system identification and prediction of the system behavior. In: Proceedings 2-nd International Conference on Fuzzy Logic and Neural Networks (IIZUKA-92), Iizuka, Japan, 17–22 July 1992, pp. 477–483 (1992)
30. Uchino, E., Yamakawa, T.: Soft computing based signal prediction, restoration and filtering. In: Ruan, D. (ed.) Intelligent Hybrid Systems: Fuzzy Logic, Neural Networks and Genetic Algorithms, pp. 331–349. Kluwer Academic Publishers, Boston (1997)
31. Miki, T., Yamakawa, T.: Analog implementation of neo-fuzzy neuron and its on-board learning. In: Mastorakis, N.E. (ed.) Computational Intelligence and Application, pp. 144–149. WSES Press, Piraeus (1999)
32. Bodyanskiy, Y., Setlak, G., Pliss, I., Vynokurova, O.: Hybrid neuro-neo-fuzzy system and its adaptive learning algorithm. In: Proceedings of Xth IEEE International Scientific and Technical Conference "Computer Science and Information Technologies", Lviv, Ukraine, Sept 2015, pp. 111–114 (2015)
33. Bodyanskiy, Y., Gorshkov, Y., Kolodyazhniy, V., Rvacheva, T.: Hybrid neuro-fuzzy network with variable shape basis function. In: Proceedings East West Fuzzy Colloquium, Zittau/Goerlitz, Germany, 13–15 Sept 2006, pp. 322–331 (2006)
34. Bodyanskiy, Y., Gorshkov, Y., Otto, P., Pliss, I.: Medical image analysis using neuro-fuzzy network. In: Proceedings of 54 International Scientific Colloquium. (IWK-2009) Information Technology and Electrical Engineering, Ilmenau, Germany, 9–12 Sept 2009, pp. 243–248 (2009)
35. Bodyanskiy, Y., Kokshenev, I., Kolodyazhniy, V.: An adaptive learning algorithm for a neo-fuzzy neuron. In: Proceedings of 3rd International Conference of European Union Society for Fuzzy Logic and Technology (EUSFLAT), Zittau, Germany, 10–12 Sept 2003, pp. 375–379 (2003)
36. Landim, R.P., Rodrigues, B., Silva, S.R., Caminhas, W.M.: A neo-fuzzy-neuron with real time training applied to flux observer for an induction motor. In: Proceedings of IEEE Vth Brazilian Symposium on Neural Networks, Belo Horizonte, 9–11 Dec 1998, pp. 67–72 (1998)
37. Ye, Bodyanskiy, Tyshchenko, O., Wojcik, W.: Multivariate non-stationary time series predictor based on an adaptive neuro-fuzzy approach. Elektronika **54**(8), 10–13 (2013)
38. Otto, P., Ye, Bodyanskiy, Kolodyazhniy, V.: A new learning algorithm for a forecasting neuro-fuzzy network. Integr. Comput. Aided Eng. **10**(4), 399–409 (2003)
39. Ye, Bodyanskiy, Vynokurova, O.: Hybrid type-2 wavelet-neuro-fuzzy network for businesses process prediction. Bus. Inform. **21**, 9–21 (2011)
40. Deoras, A.: Electricity load forecasting using neural networks. In: Electricity Load and Price Forecasting Webinar Case Study (2011). http://www.mathworks.com/matlabcentral/fileexchange/28684-electricity-load-and-price-forecasting-webinar-case-study/content/Electricity%20Load%20&%20Price%20Forecasting/Load/html/LoadScriptNN.html.DatasetfromtheNew EnglandISO http://www.iso-ne.com/

Part V
Green PLC-Based Systems for Industry Applications

PLC-Based Systems for Data Acquisition and Supervisory Control of Environment-Friendly Energy-Saving Technologies

Yuriy Kondratenko, Oleksiy V. Korobko and Oleksiy V. Kozlov

Abstract This paper presents the development of PLC-based systems for data acquisition and supervisory control of environment-friendly energy-saving complex high-tech technologies. The functional structure and main components of PLC-based SCADA-systems for environment-friendly energy-saving technological processes are given. The examples of SCADA applications in design of PLC-based systems for monitoring and automatic control of (a) ecopyrogenesis (EPG) and (b) thermoacoustic technological processes are presented. Paper considers the criteria of energy and economic efficiency of the EPG technological process. The functional structures, software and hardware implementation as well as multi-level human-machine interfaces of the developed PLC-based systems for data acquisition and supervisory control are given. Considerable attention is given to particular qualities of computing of ecopyrogenesis and thermoacoustic processes technological parameters by the proposed SCADA-systems. The developed PLC-based SCADA-systems provide: significant increasing of energy and economic efficiency criteria of the EPG and TAD complexes, high precision control of both technological processes, monitoring of current technological parameters, using the indirect methods for parameters measuring and identifying, and automatic control with high quality indicators and optimal parameters.

Y.Kondratenko
Department of Electrical Engineering and Computer Science,
Cleveland State University, Cleveland, OH, USA
e-mail: y.kondratenko@csuohio.edu; y_kondrat2002@yahoo.com

O.V. Korobko · O.V. Kozlov (✉)
Admiral Makarov National University of Shipbuilding,
9 Geroiv Stalingrada Av., Mykolaiv 54025, Ukraine
e-mail: oleksiy.kozlov@nuos.edu.ua

O.V. Korobko
e-mail: oleksii.korobko@nuos.edu.ua

Y.Kondratenko
Petro Mohyla Black Sea State University, 10, 68th Desantnykiv Str.,
Mykolaiv 54003, Ukraine

V. Kharchenko et al. (eds.), *Green IT Engineering: Concepts, Models, Complex Systems Architectures*, Studies in Systems, Decision and Control 74,
DOI 10.1007/978-3-319-44162-7_13

Keywords PLC · SCADA · Environment-friendly energy-saving technological processes · Ecopyrogenesis · Thermoacoustic · Data analysis · Control system · Intelligent controller

1 Introduction

Energetics is one of the basic sectors of the region economy, and energy efficiency is an important condition for its development [1, 2]. It is known that the current state of world energetics is characterized by the deficiency and high cost of natural organic fuel as well as environmental pollution with harmful waste [3]. The development strategy of the energy complexes of many countries and regions is related with the use of new (non-traditional) and renewable energy sources, which is the real way of successful solving of the energy problems of the world economy [4].

Especially effective area of energy development is the use of alternative fuels and utilization of excess energy [1]. Promising from an economic point of view is the involving of municipal solid wastes (MSW) and low temperature thermal excess energy in the region's energy balance as an alternative fuel and energy with environmental safety [3–5].

As demonstrated by international practice, there are more than 50 methods of MSW and thermal excess energy utilization known in the world. Each method has its advantages and disadvantages, but among the most advanced from environmental and economic points of view we should allocate two environment-friendly energy-saving technologies: ecopyrogenesis technology and thermoacoustic technology of low temperature thermal energy utilization.

EPG is a new technology of solid waste disposal in order to obtain an alternative fuel [6]. It combines the processes of multi-loop circulation pyrolysis (MCP) and the dual-zone thermostable gasification (DZTG) in a single process line, that allows complete utilization of the whole scope of the organic part of solid waste and low-grade coal in the environment-friendly and energy-saving modes according to the economically-profitable scheme [6]. Application of the EPG technology gives the opportunity to obtain substantial profit from the sale of output products, which are liquid and gaseous fractions of alternative fuels that can be used in combustion engines without any extra filtration [6].

Thermoacoustic (TA) technology [7] uses the mutual sound and heat energy conversions as the main working principle. Heat supplied to TA apparatus (TAA) is converted in acoustic energy [8] inside the resonator of TAA. Second part of TAA uses produced acoustic energy for heating/cooling [9] the temperature of coolant outlet. Main scope of TAA applications is the utilization [10] of waste heat resources (WHR). Usage of thermoacoustic plant in WHR utilization complex allow to increase the temperature of waste heat and reuse it in technological process. Ability to utilize low potential heat resources makes TAA profitable for industry and transport applications as well as allow decreasing the amount heat emissions into the atmosphere.

For realization of the given above environment-friendly energy-saving technologies specific technological complexes are used, which are, in turn, complicated multi-component technical objects. This technological complexes require the development of a new class of monitoring and control systems that are based not on static information of databases, but on current data, that is generated in real time during the occurrence of the relevant technological processes [11]. Typically, such systems have a hierarchical branched structure, the components of which are industrial computers, process controllers, operator automated posts, programmers and facilities that implement industrial networks.

At developing of monitoring and control systems of the EPG and thermoacoustic technological complexes the important task of compatibility of software complex with most, presented on the market, digital controllers appears. This guarantees to the user a wide scope for choosing the hardware part of the appropriate project and the ability of integration with different operating systems. The modern systems of such type are PLC-based systems for data acquisition and supervisory control [12, 13].

Automation of EPG and thermoacoustic technological complexes on the basis of PLC-based SCADA-systems allows to significantly increase the operation efficiency and economic parameters.

SCADA stands for “supervisory control and data acquisition”. It is a software application being used for process control, which involves collecting data from different locations in real-time. Conception of SCADA is caused by the whole course of development of control systems and the results of complex automation of complicated industrial and technological processes [12]. Application of SCADA-technology enables to achieve a high level of automation in solving the problems of the development of control systems, data acquisition, processing, transmission, storage and display of information. Friendliness of human-machine interface (HMI/MMI) of SCADA-systems is provided by completeness and clarity of the information that appears on the screen, the availability of control “levers”, convenience of using of prompts and help system, etc., which greatly increases the efficiency of interaction of dispatcher with system and nullifies his critical errors in control processes.

SCADA systems have a number of components [13] that ease the development of computerized data acquisition and analysis tools, such as communication protocols for external devices connection, built in signals filtering, support of SQL data exchange etc. This and many other features makes the SCADA software a perfect tool for implementation of control, data analysis and information processing algorithms in household, industry and science research.

SCADA systems [14, 15] can be divided into two big classes—modular and scattered. Scattered systems consists of the numerous applications and each application is aimed to solve one particular task (program controller, establish connection to DB etc.). Mainly such systems are the result of systematic evolution of early technological processes control programs. Modular systems are modern applications that combine all SCADA features in one program thus makes it easier for users to work in it and reduces the control system implementation time.

The main purpose of this work is development and implementation of PLC-based systems for data acquisition and supervisory control of environment-friendly energy-saving ecopyrogenesis and thermoacoustic technologies.

Structure and main components of the generalized PLC-based system for data acquisition and supervisory control of environment-friendly technologies are given in Sect. 2. The functional structures, software and hardware implementation as well as multi-level human-machine interfaces of the developed PLC-based SCADA-systems of the EPG and thermoacoustic technological complexes are presented in Sects. 3 and 4 respectively. Section 5, in turn, has conclusions, discussion of the results and further research on PLC-based system for data acquisition and supervisory control of environment-friendly technologies.

2 Structure and Main Components of Generalized PLC-Based System for Data Acquisition and Supervisory Control

A generalized PLC-based SCADA system has software as well as hardware components. The role of software is to collect data and feed into a PLC that has the specific software installed.

The block diagram of PLC-based SCADA system shown in Fig. 1 represents the basic SCADA architecture. Generally, a SCADA system consists of the following components.

First, analytical instruments that sense process variables and operating equipment connected to instruments. Second, one or more field data interface devices, usually remote terminal units (RTUs) or programmable logic controllers (PLCs). RTUs connect to sensors in the process, converting sensor signals to digital data and sending digital data to the supervisory system. PLCs are used as field devices because they are more economical, versatile, flexible and configurable than special-purpose RTUs. Usually, PLCs make the primary information processing and greatly accelerate the process of information exchange [16]. Third, communication infrastructure used to transfer data between field data interface devices and control units as well as supervisory system. The infrastructure can be radio, telephones, cables, satellites or any combination of these. Fourth, host computers are the center of human monitoring and control of the processes-storing databases and displaying statistical control charts and reports. Host computers are also known as master terminal units (MTUs) or the SCADA server. Fifth, a human-machine interface (HMI) is the apparatus, which presents process data to a human operator in order to control the process through the human operator monitors [13].

These systems can be relatively simple, such as one that monitors environmental conditions of a small office building, or very complex, such as a system that monitors all the activity in a nuclear power plant or the activity of a municipal water

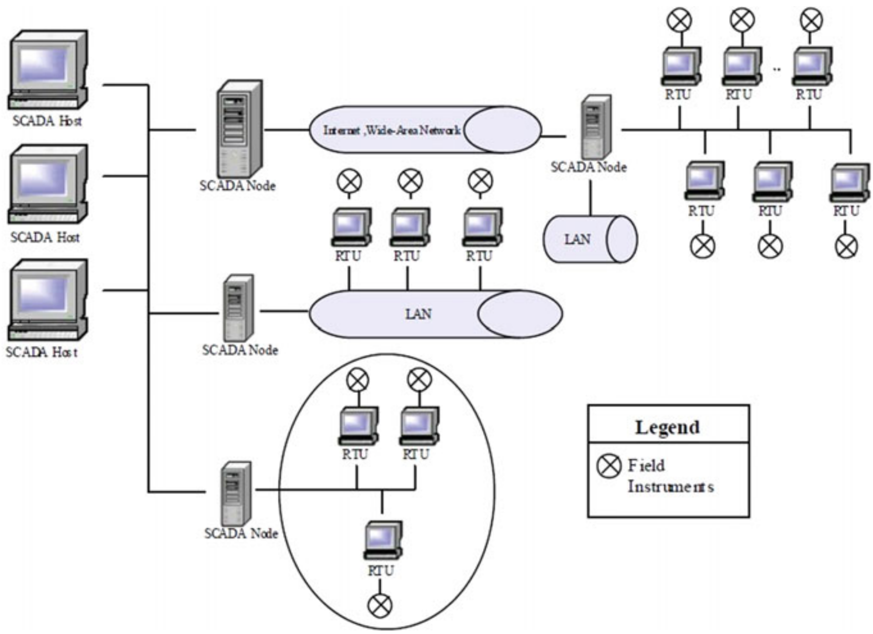


Fig. 1 Functional diagram of the generalized PLC-based SCADA system

system. Traditionally, SCADA systems have made use of the Public Switched Network (PSN) for monitoring purposes. Today many systems are monitored using the infrastructure of the corporate Local Area Network (LAN)/Wide Area Network (WAN). Wireless technologies [17] are now being widely deployed for purposes of monitoring. SCADA is not a full control system, but rather focuses on the supervisory level. Thus, it is a purely software package that is positioned on top of hardware to which it is interfaced, in general via Programmable Logic Controllers (PLCs), or other commercial hardware modules [18]. SCADA can be used in a variety of automated process control systems. In [19, 20] an internet based SCADA systems for power plants taking as case study the power system in Montenegro and Cameroon were presented. Researcher, Aydogmus [16], has presented a SCADA control via PLC for monitoring of water level in tank with fuzzy controller. Another good example is the introduction of effective use of SCADA-system [21] on the Kursk Cogeneration 1. Also SCADA can be used to control completely different actuators, such as AC and DC engines [22, 23]. There are many other examples of the use of SCADA, however, such systems evolve rapidly and are now penetrating the market of plants with a number of I/O channels of several 100K.

Let us consider the developed by the authors PLC-based systems for data acquisition and supervisory control of environment-friendly energy-saving EPG and thermoacoustic technologies.

3 PLC-Based SCADA-System for the Ecopyrogenesis Technological Complex

The basic diagram of the generalized EPG technological complex is presented in Fig. 2 [24, 25], where the following notations are accepted: *1*—pyrolysis reactor of the multicircuit pyrolysis plant (MPP); *2*—oil station (OS) of the wastes loading hydraulic drives (LHD) of the EPG technological complex; *3*—gas burner (GB) for the MPP pyrolysis reactor heating; *4*—solid polymeric wastes (SPW) LHD; *5*—MPP gas burner ventilation system; *6*, *7*, *8*—first, second and third cooling circuits of the MPP multicircuit circulation system (MCS), respectively; *9*—pipe for the flue gases outlet; *10*—dual-zone gas generator (DZGG); *11*—DZGG gas burners; *12*—upper (reverse) gasification zone; *13*—lower (direct) gasification zone; *14*—wet wastes (WW) LHD; *15*—slag (S) discharging hydraulic drive; *16*—pipe, through which generator gas (GG) enters the pyrolysis reactor.

A generalized EPG complex (Fig. 2) consists of two interrelated technological nodes: DZGG and MPP—for utilization of wet and solid polymeric wastes, respectively [24, 25].

To provide the continuous passage of EPG process and to get the liquid fuel with the required molecular mass as well as synthesis gas of high quality at the output it is necessary to perform the following functional tasks: control of processes of multicircuit circulation gasification and multicircuit circulation pyrolysis, monitoring of current technological parameters and automatic regulation of mechanisms that provide optimal process parameters [25].

The main indicators of energy and economic efficiency of the EPG complexes are: productivity Q_{EPG}, specific productivity B_{EPG}, total power consumption $P_{\Sigma EPG}$,

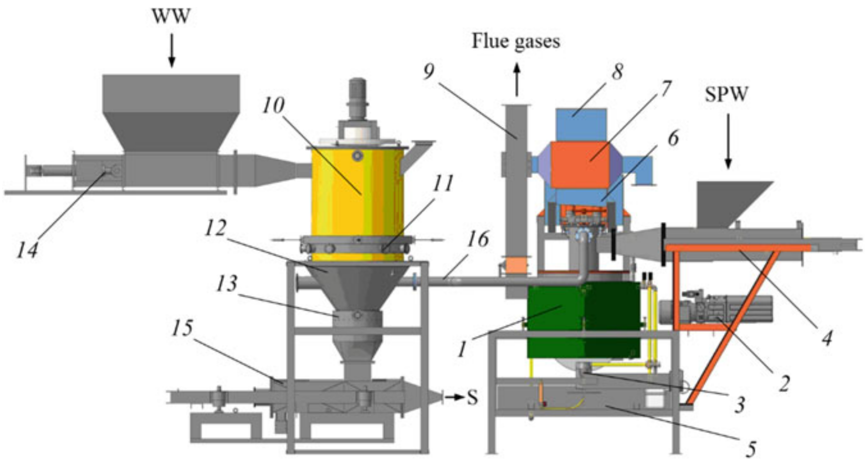


Fig. 2 Basic diagram of the generalized EPG technological complex

relative deviations of the real values of the controlled parameters ΔT_{R}, ΔL_{R} and ΔM_{AF}.

Let's consider the main indicators of energy and economic efficiency of the EPG complexes in details.

Productivity of the complex. The productivity Q_{EPG} of the EPG technological complex is calculated by the equation

$$Q_{\mathrm{EPG}} = \frac{m_{\mathrm{AF}}}{t_{\mathrm{OP}}}, \tag{1}$$

where m_{AF}—mass of the produced alternative fuel; t_{OP}—operation time of the complex. The increase of the complex productivity Q_{EPG} at the same energy costs gives the opportunity to significantly increase the energy and economic efficiency of the EPG technological complex. This can be achieved due to the optimal control of the technological processes of DZTG and MCP that can provide increasing of the speed of these processes at the same energy costs.

Specific productivity of the complex. The specific productivity B_{EPG} of the EPG technological complex is calculated by the equation

$$B_{\mathrm{EPG}} = \frac{Q_{\mathrm{EPG}}}{K} = \frac{m_{\mathrm{AF}}}{K t_{\mathrm{OP}}}, \tag{2}$$

where K—the number of operating personnel of the complex. The increase of the complex specific productivity B_{EPG} gives the opportunity to significantly increase the energy and economic efficiency of the EPG technological complex by reducing labor costs. This can be achieved by reducing the number of operating personnel of the complex at the expense of increasing the level of automation of the EPG technological complex and using fully automated equipment for the EPG technological process performing.

Total power consumption of the complex. The total power consumption $P_{\Sigma\mathrm{EPG}}$ of the EPG technological complexes is calculated by the equation

$$P_{\Sigma\mathrm{EPG}} = \frac{W_{\Sigma\mathrm{EPG}}}{t_{\mathrm{OP}}}, \tag{3}$$

where $W_{\Sigma\mathrm{EPG}}$—the total amount of consumed energy of the complex. The reduction of the complex total power consumption $P_{\Sigma\mathrm{EPG}}$ at the same value of productivity Q_{EPG} gives the opportunity to significantly increase the energy and economic efficiency of the EPG technological complex by reducing the production costs of alternative fuel. This can be achieved due to the optimal control of the technological processes of DZTG and MCP that can provide reducing energy costs for the implementation of these processes at the same value of productivity.

Relative deviations of the real values of the controlled parameters from their set values. The relative deviations Δ of the real values from the set values for such controlled parameters of the complex as heating temperature T_{R} of the reactor, loading level L_R of the reactor and molecular weight M_{AF} of the liquid fractions of the output alternative fuels are calculated by the following equations:

$$\Delta T_R = \frac{T_{SR} - T_{RR}}{T_{SR}} \cdot 100\ \%; \tag{4}$$

$$\Delta L_R = \frac{L_{SR} - L_{RR}}{L_{SR}} \cdot 100\ \%; \tag{5}$$

$$\Delta M_{AF} = \frac{M_{SAF} - M_{RAF}}{M_{SAF}} \cdot 100\ \%, \tag{6}$$

where T_{SR}, L_{SR}, M_{SAF}—the set values of the corresponding parameters of the complex; T_{RR}, L_{RR}, M_{RAF}—the real values of the corresponding parameters of the complex. The reduction of relative deviations of the real values of the controlled parameters of the complex from their set values gives the opportunity to significantly increase the energy and economic efficiency of the EPG technological complex by improving the quality of the obtained alternative fuel. The quality of the obtained at the output alternative fuel directly depends on the control accuracy of the EPG complex main parameters, such as temperature T_R, loading level L_R and molecular weight M_{AF}. To achieve the high quality indicators of control of the EPG complex main parameters it is advisable to apply high-precision automatic control system of the following parameters based on the intelligent principles and technologies [24–27].

To implement the given above functional tasks in order to achieve high performance values of the main energy and economic efficiency indicators of the EPG complexes it is appropriate to use the developed by the authors PLC-based SCADA-system, which has a multilevel branched structure.

Functional structure of the developed by the authors PLC-based SCADA-system for the EPG technological complex [25] is presented in Fig. 3, where the following notations are accepted: IC—industrial computer; CI—converter of interface (Rs485/Rs232); PLC—programmable logic controller; AOM—analog output module; DAM—data acquisition module; TS, LS—temperature and level sensors, respectively; S_n—sensor of certain parameter X_n (temperature, pressure, flow, linear position etc.); ACS—air cooling system; SGM—steam-gas mixture; u_{GB}, u_{LHD}, u_{ACS}—control signals of GB, LHD and ACS, respectively; u_{TS}, u_{LS}, u_n—sensor signals of corresponding parameters; P_{GB}—power values of pyrolysis reactor and DZGG gas burners, respectively; Y_{LHD}—values of linear motion of the pistons of LHD of WW and SPW, respectively; T_{GG}, T_R, T_{MSC}—temperature values of certain points of DZGG, pyrolysis reactor and MCS, respectively; L_{GG}, L_R—values of the DZGG and pyrolysis reactor load level, respectively; Q_A—the flow rate of cooling air of ACS.

The proposed specialized PLC-based SCADA-system has three hierarchical levels of monitoring and control:

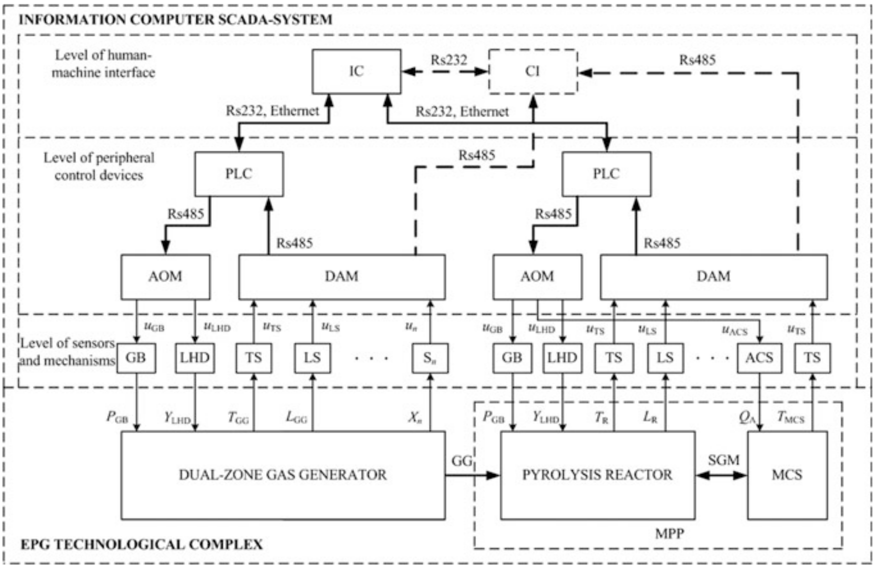


Fig. 3 Functional structure of the EPG technological complex's PLC-based SCADA-system

(1) lower level—the level of the sensors and mechanisms of the EPG complex;
(2) average level—the level of the peripheral control devices (includes modules for data acquisition and output as well as PLC);
(3) upper level—the level of human-machine interface.

Lower level of the developed PLC-based SCADA-system consists of the automatic control sub-systems (ACSS), which perform regulation of the given values of the EPG complex major controlled data according to a control program in terms of various disturbances operation [25].

Industrial computer serves as the operator station, it contains a control panel of the EPG technological process with the image of the main components and their technical parameters that are measured by various sensors. Also, the set values of the control parameters, that are inputs of the complex ACSS, are established on the IC. PLC processes the information received from the sensors and sends it to the IC. Also, it gets the set values of the complex control parameters and performs the functions of it automatic control. Data acquisition modules receive signals from sensors and transmit it to the PLC in a convenient form for further processing. Analog output modules receive signals from the PLC and produce corresponding control signals that go directly to the actuators.

One of the most important particularities of the developed specialized PLC-based SCADA-system is that it can implement the measurement of the pyrolysis reactor load level L_R with the help of the indirect method [26].

The reactor load level value L_R direct measurement with the level sensor causes many difficulties associated with the particularities of the pyrolysis technological process and with the operating conditions of the level sensor in the pyrolysis reactor. The main limitations of the use of the level sensor inside the reactor are as follows: temperature of at least 600 °C, high viscosity, changing of the aggregate state and density of wastes in the operating process, the possibility of wastes solidification and sticking to the sensing element, wastes boiling [26].

For solving this problem the developed specialized PLC-based SCADA-system uses the indirect method of level measuring based on the calculating of the loading level value by measuring of the other parameters of the EPG complex, such as: pyrolysis gas flow rate, liquid fuel flow rate, temperature in the reactor operating medium and the linear movement of the SPW LHD piston [26]. The current value of the pyrolysis reactor load level L_R is calculated by the proposed SCADA-system with the help of the following equations:

$$L_{\mathrm{R}} = \frac{M_{\mathrm{WR}}}{\rho_{\mathrm{WR}} S_{\mathrm{R}}}; \tag{7}$$

$$M_{\mathrm{WR}} = M_{\mathrm{LW}} - (M_{\mathrm{PG}} + M_{\mathrm{LF}}); \tag{8}$$

$$M_{\mathrm{LW}} = \rho_{\mathrm{LW}} S_{\mathrm{P}} Y_{\mathrm{LHD}}; \tag{9}$$

$$M_{\mathrm{PG}} = \rho_{\mathrm{PG}} V_{\mathrm{PG}} = \rho_{\mathrm{PG}} \int Q_{\mathrm{PG}} dt; \tag{10}$$

$$M_{\mathrm{LF}} = \rho_{\mathrm{LF}} V_{\mathrm{LF}} = \rho_{\mathrm{LF}} \int Q_{\mathrm{LF}} dt, \tag{11}$$

where M_{WR}—wastes mass in the reactor; M_{LW}—wastes mass that are loading to the reactor; M_{PG}—mass of the pyrolysis gas, that is allocated during the EPG complex operating process; M_{LF}—mass of the liquid fuel, that is allocated during the EPG complex operating process; ρ_{LW} and ρ_{WR}—density values of the wastes that are loaded with the help of the LHD and wastes that are in the reactor during operation, respectively; ρ_{PG} and ρ_{LF}—density values of the pyrolysis gas and liquid fuel, respectively; V_{PG}, Q_{PG}, V_{LF} and Q_{LF}—volume and flow rate values of the pyrolysis gas and liquid fuel, respectively; S_P and S_R—cross-sectional area values of the LHD piston and reactor, respectively.

Density values ρ_{LW}, ρ_{WR}, ρ_{PG} and ρ_{LF} are determined experimentally for each type of wastes (polyethylene, polypropylene, polystyrene); in turn density value of wastes that are in the reactor during operation ρ_{WR} also depends on its temperature T_R and aggregate state ($\rho_{WR} = f(T_R)$). Flow rate values of the pyrolysis gas and liquid fuel as well as linear movement value of the LHD piston Q_{PG}, Q_{LF} and Y_{LHD} are determined with the help of flow rate and linear movement (position) sensors [26].

To implement the EPG technological complex PLC-based SCADA-system it is expedient to use certain software and hardware, more detailed description of which is presented below [25].

The hardware facilities for the EPG technological complex PLC-based SCADA-system implementation are data acquisition modules, analog output modules and PLC. Data acquisition modules receive analog signals from different sensors and convert them into the digital data that is sent to SCADA-system by the RS485 interface. PLC WP-8131 [27] of the ICP DAS company performs the calculations of the level measuring indirect method, the automatic control of the EPG process variables and their transfer to the operator interface of IC. Analog output module, in turn, implements the conversion of digital signals from the PLC to analog ones that directly go to the actuators of the complex parameters ACSS.

To receive signals from sensors of different types in this SCADA-system 3 types of the ICP DAS company data acquisition modules are used: I-7018P 8 inputs modules—for data acquisition from thermocouples; I7017S 8 inputs modules—for data acquisition from sensors with current output (0–20 mA) and I-7051 16 inputs modules—for data acquisition from sensors with discrete output [27]. To implement the conversion of digital signals from the PLC to analog ones that directly go to the actuators of the complex parameters ACSS the I-7024 4 inputs module for analog output of the ICP DAS company is used.

The appearance of the control unit of the experimental EPG complex PLC-based SCADA-system, which is developed on the basis of the above modules is shown in Fig. 4.

To measure the main EPG complex technical parameters different sensors are used:

- temperature sensors—thermocouples of two types of the “TERA” company: L-type—for temperature measurement in range −40 to 600 °C and K-type—for temperature measurement in range −40 to 1050 °C;
- level sensors—float level sensors of the “OVEN” company—for measurement of water and liquid fuel level and radar level sensors of the “Rosemount”

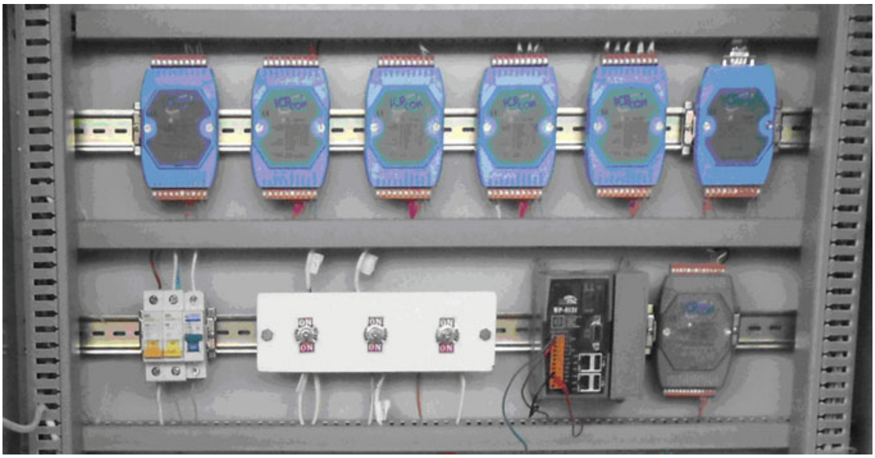


Fig. 4 Control unit of the experimental EPG complex PLC-based SCADA-system

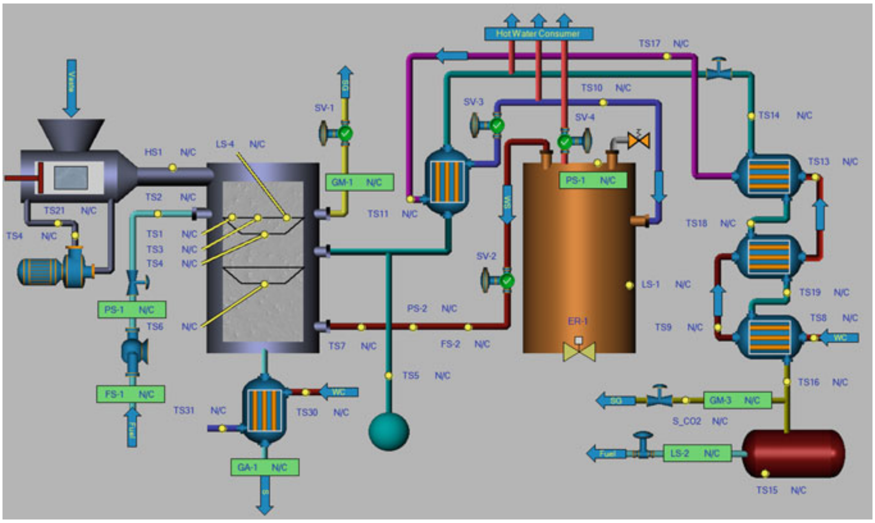


Fig. 5 Human-machine interface of the EPG complex PLC-based SCADA-system

company—for measurement of loading level of wastes in MPP pyrolysis reactor and DZGG;
- pressure sensors—of the "Dwyer Instruments" company of 673-2-type, etc.

As the software facilities of the developed EPG complex PLC-based SCADA-system the TRACE MODE 6 is used [25, 28], which belongs to the class of integrated systems that provide maximum comfort to designers and users.

The multi-level human-machine interface (Fig. 5) [25, 28] of the developed PLC-based SCADA-system provides a visual representation of the performance indicators of the EPG technological complex on the IC.

The developed PLC-based SCADA-system provides: high precision control of processes of DZTG and MCP, monitoring of current technological parameters, using the indirect method of level measuring, and automatic regulation with high quality indicators of mechanisms that provide optimal process parameters.

It also gives the opportunity to achieve high performance values of the main energy and economic efficiency indicators of the EPG technological complexes.

As an example, the experimental EPG technological complex "ECOPYR-18 M", developed at the Admiral Makarov National University of Shipbuilding can produce 96 l and 22 m^3 of liquid and gaseous fractions of high quality alternative fuels, respectively, while utilizing 120 kg of organic waste per day. The implementation of the proposed PLC-based SCADA-system in this technological complex makes it possible to increase its productivity Q_{EPG} by 15 %, reduce the maximum value of its total power consumption $P_{\Sigma\mathrm{EPG}}$ by 10.2 % and reach a zero value of the relative deviations of the real values from the set values for main controlled parameters of the complex ($\Delta T_{\mathrm{R}} = 0$; $\Delta L_{\mathrm{R}} = 0$; $\Delta\, M_{\mathrm{AF}} = 0$).

As for the specific productivity B_{EPG} of the given EPG complex it value is increased by 6 times by means of PLC-based SCADA-system implementation.

4 PC-Based System for Thermoacoustic Processes Supervisory Control

Thermoacoustic devices (TAD) are the newest type of unconventional heat machines [7, 8]. Their work is based on the mutual transformation of heat and acoustic energies. The main feature of thermoacoustic systems is that unlike the other heat machines, the powerful acoustic pulses [8], generated by TAD are the carrier of mechanical energy and the "executive mechanism" of thermodynamic cycle.

Considering the TAP reversibility, thermoacoustic devices can be subdivided into two main types [7]:

- thermoacoustic engines (TAE), which consume supplied heat energy and produce acoustic wave inside the resonator;
- thermoacoustic refrigerators (TAR) and heat pumps (TAHP) which consume energy from supplied acoustic wave and produce cooling or heating power, respectively.

The schematic of simple thermoacoustic refrigerator is shown in Fig. 6. It consists of sound wave electromechanical generator, hollow resonator filled with gas medium, heat exchange surfaces (TC, TH) for heat adding and subtraction and a special heat exchange surface (the stack), which is the main catalyst of acoustic and thermal energies mutual conversion in TAD.

Absence of moving parts [8] in such heat engines increases their reliability and reduces energy losses in the mechanical connections. It should be noted that such heat machines used as heat pumps or refrigerators are significantly eco-friendlier comparing to conventional plants due to absence of CFCs in working process. However, for high power thermoacoustic systems operations the acoustic pressure

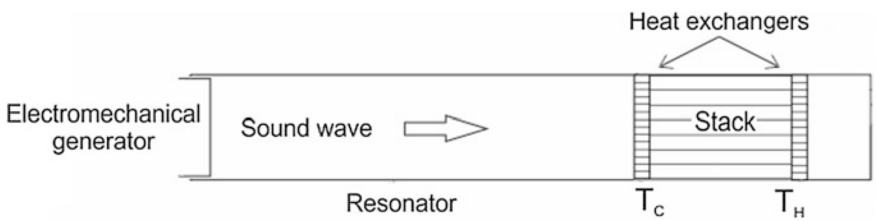


Fig. 6 Structure of the thermoacoustic refrigerator with electromechanical generator of sound waves

level of 150–180 dB must be maintained in resonator, which leads to the appearance of parasitic nonlinear effects [8, 29].

Zink et al. in work [30] provide comprehensive analysis of ecological impact of TAR implementation, instead of traditional refrigerating technologies. They consider the installation of waste heat driven thermoacoustic refrigerators in cars instead of vapor compression refrigerators. Such replacement could lower overall car emissions by eliminating the fuel loses on air conditioning. Using the statistics of USA transport departments (approx. 170 million vehicles), they calculated that the CO_2 emissions of car refrigeration systems reach 1.675×10^{10} kg CO_2 per year. This number corresponds to roughly 1 % of the total (energy-related) emission of CO_2 in the U.S. Replacing traditional air-conditioning systems with a thermoacoustic waste heat-driven system will directly eliminate these emissions. It should be noted that as the source of waste heat in such systems can be used the exhaust gases after the catalytic converters, or, taking into account latest results in low potential energy utilization with TAD, even the temperature of cooling water (about 90 °C).

This heat machines are characterized by high sensitivity to the working fluid properties (viscosity, density, thermal conductivity) as well as the environment effects (temperature, pressure, etc.). An effective work of TAD is possible only if the different nature values of the internal parameters (acoustic, hydrodynamic, thermal) are stabilized. For that reason, authors has developed the SCADA-based system for thermoacoustic processes (TAP) control. Designed system uses the specially implemented indirect method for measurement of acoustic parameters. This method is described next.

4.1 Acoustic Parameters Measurement by Means of Modified Two Sensor Method

The main characteristics that describe the sound wave [29] is the acoustic pressure $p(t)$ and particle velocity $v(t)$ of the environment in which sound is distributed. These parameters determine the intensity $I(t)$ (6) of sound wave, which is the [3, 18] time-averaged flow of acoustic energy in the TAD resonator and therefore numerically equals to the energy that the wave transfers during one second through the unit area.

$$I = \frac{1}{2}\mathrm{Re}(\tilde{p}v), \tag{12}$$

where $\tilde{p}$ is the complex conjugate value of $p(t)$ and $\mathrm{Re}(\cdot)$ denotes the real part of complex number.

The direct method [31, 32] that combines the simultaneous measurements of $p(t)$ using the pressure transducer and v(t) with the laser Doppler velocity meter is typically used for determination of the sound waves parameters in TAE driven thermoacoustic systems. This approach allows determining the values of $p(t)$ and $v(t)$ with high accuracy, but requires expensive equipment and the availability of transparent zone in resonator to undergo laser beams. These shortcomings limit the possibility of the direct method usage only by research systems, and make its implementation in TAD automated control systems unreasonable.

Designed computer system for control of thermoacoustic processes implements the modified version of Fusco's two sensors method [10, 32] for acoustic wave's parameters estimation in the resonator of thermoacoustic device. This approach allows the avoidance of the drawbacks and limitations of the above stated methodologies as well as makes possible the creation of the universal system for TAP control.

In two-sensor method [31, 32], pressure (13) and particle velocity (14) values in point "0" are calculated using the readings of two pressure sensors (A and B) located on a fixed distance Δx from each other [3].

$$p(0) = \frac{p_A + p_B}{2\cos(k\Delta x/2)}, \quad k = k_0\left\{1 + \frac{1-j}{2}\frac{\delta_v}{r_0}\left(1 + \frac{\gamma - 1}{\sqrt{\sigma}}\right)\right\}, \tag{13}$$

$$v(0) = \frac{kF}{\omega\rho}\frac{p_A - p_B}{2j\sin(k\Delta x/2)}, \quad F = 1 - \frac{1-j}{r_0/\delta_v}, \tag{14}$$

where k is the complex wave number, Δx is the distance between pressure sensors A and B, p_A, p_B are the A and B pressure sensors values, $k_0 = \omega/a$ is the wave number in free space, σ is the Prandtl number, δ_v is the viscous boundary layer thickness, r_0 is the duct radius, ρ is the average gas density, j is the imaginary unit, $\omega = 2\pi f$ is the sound wave angular frequency, γ is the adiabatic index of gas in resonator.

Based on Eqs. (13) and (14) and taking into account that acoustic intensity is defined as (12) the final formula for sound intensity as the function of two pressure sensors readings is defined as

$$I = \frac{1}{8\omega\rho}\left[\mathrm{Im}(H)\left(|p_A|^2 - |p_B|^2\right) + 2\mathrm{Re}(H)|p_A||p_B|\sin\theta\right], \tag{15}$$

$$H = \frac{kF}{\cos(\tilde{k}\Delta x/2)\sin(k\Delta x/2)}, \tag{16}$$

where $\tilde{k}$ is the complex conjugate of k, θ is the phase shift between p_A and p_B signals, $\mathrm{Im}(\cdot)$ denotes the imaginary part of complex number.

It should be noted that the main drawback of the Fusco's method is the difficulty of determination of the phase difference θ, the value of which is very small. This imposes certain restrictions on the principles of the two sensors method usage. Thus, for a precise definition it is necessary to use additional equipment such as phase lock amplifiers and spectrum analyzers [8, 30], install sensors near the nodal points of acoustic pressure or velocity, where phase fluctuations are largest.

To get rid of described downside of a Fusco's method the developed computer system uses the complex form of acoustic parameters. In combination with implemented digital signal processing algorithms, it allows the intensity of the sound wave to be calculated directly from the expression (6). Such approach avoid the difficulties associated with the phase difference θ determination, because the oscillating acoustic pressure and velocity are determined by the expressions (13) and (14), and their complex modulus and phase respectively calculated using expressions (17) and (18).

$$|x| = \sqrt{\frac{1}{N}\sum_{k=1}^{N} x_k^2}, \tag{17}$$

$$\varphi(x) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[arctg\left(\mathrm{Im}(x_k)/\mathrm{Re}(x_k)\right)\right]^2}. \tag{18}$$

4.2 Structure and Main Components of Specialized Computer System

Figure 7 shows the structure and main components of a specialized computer system (SCS) [10, 33] for the thermoacoustic processes control. Due to the thermoacoustic process features, the computer system combines rapid data (acoustic pressure actuations, current and voltage oscillations) and slow data signals (temperature variations).

The main elements of the system are: personal computer (PC); pressure ($PS1$,…, $PS4$), temperature ($TS1$,…,$TS4$), current (CS) and voltage (VS) sensors; programmable logic controller (PLC) ICPDAS μPAC 7186EX-SM that collects the thermal behavior data of TAD and microcontroller (MCU) STM32F407VGT6, that is used to collect rapidly changing data from the system pressure, current and voltage sensors and transfer it to PC.

STM32F4 microcontroller has three analog-to-digital converters, which are characterized by 12-bit resolution and conversion frequency of up to 2.4 Msps and hardware implemented floating-point unit and therefore is suitable for digital signal processing tasks [33]. This provides high performance and applicability of the proposed computer system for any thermoacoustic unit, since the operating frequency range of TAD does not exceed tens of kHz. PLC performs data acquisition of the status of TAD peripheral equipment and sends it to PC by the Ethernet network.

The designed system allows the communication between all components via RS232 network, also to enable the transfer of large amounts of data, PLC supports

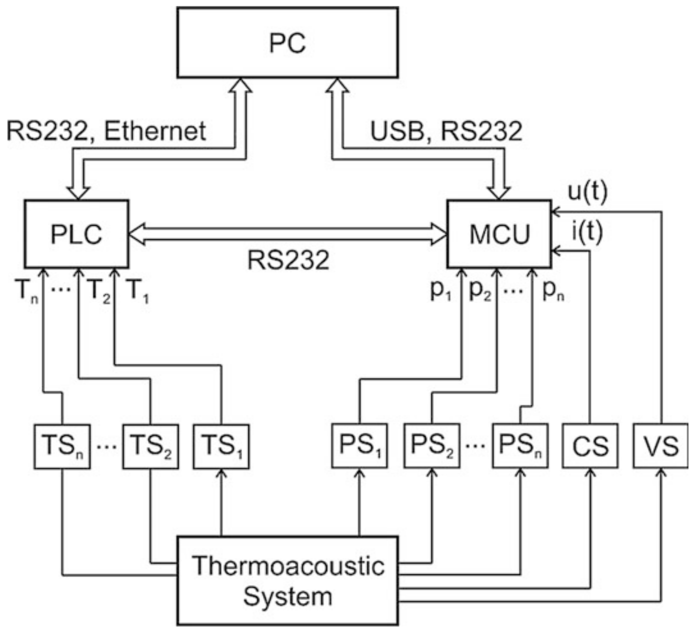


Fig. 7 Schematic of the specialized computer system for TAP control

the Ethernet connection with PC and MCU can communicate with PC via USB. This grants maximum flexibility to designed SCS.

Created on the basis of SCADA TraceMode 6 software for PC [6] allows storing the received from the controller information in the archive database, execution of the necessary algorithms for digital processing of measured signals, and displaying all available information on the user's screen.

As a result, the proposed structure of SCS allows registration and ongoing monitoring of the thermoacoustic processes and is suitable for the real time control algorithms implementation for both thermoacoustic engines and refrigerators.

4.3 Experimental Verification of Designed Specialized Computer System

The tests of proposed by the authors SCS were conducted using the experimental thermoacoustic engine (TAE) (Fig. 8) [10, 33], with a resonator length of 1010 mm and diameter of 46 mm, which was filled with air at atmospheric pressure. Engine was coupled with thermoacoustic heat pump (TAHP), which acted as a payload of the system.

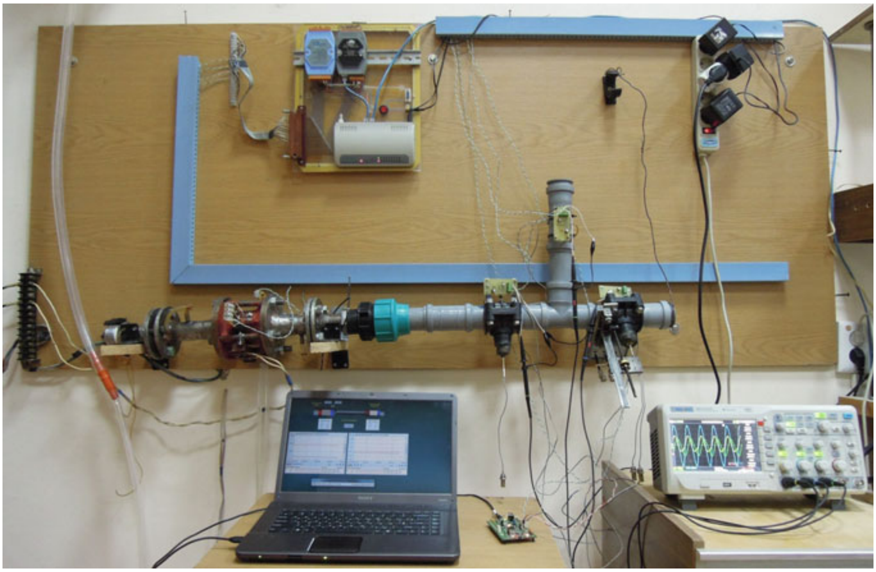


Fig. 8 Experimental setup for the study of PC-based thermoacoustic processes control system

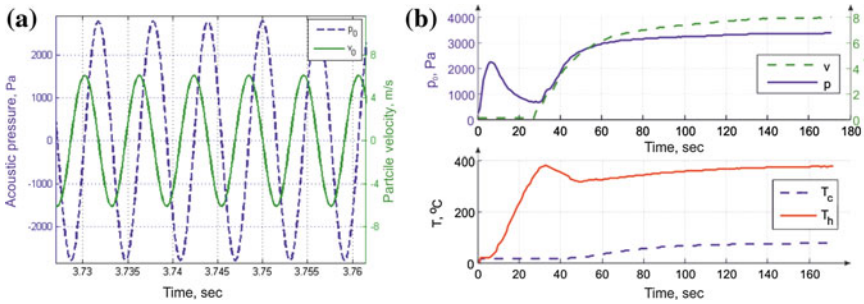


Fig. 9 Results of the experimental study of TAE working process. **a** Instant values of pressure $p0(t)$ and velocity $v0(t)$. **b** TAE starting process

Thermoacoustic engine starting process is shown in Fig. 9b where on the top are shown rooted mean square values of calculated pressure $p_0(t)$ and particle velocity $v_0(t)$ and on the bottom—temperature changes in points T_H and T_C (Fig. 6). Figure 9a demonstrates the results of calculated instant values of acoustic pressure $p_0(t)$ and particle velocity $v_0(t)$.

Conducted tests and their results show the adequacy of the developed system, its components and proposed indirect method of acoustic parameters measurements. Thus making it possible to design effective, environmentally friendly TAA structure.

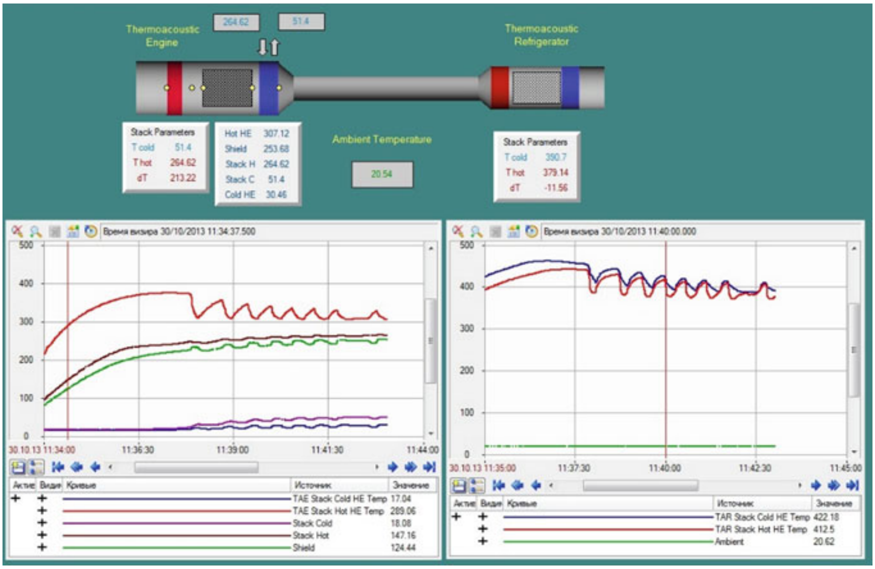


Fig. 10 HMI of PC-based system for thermoacoustic processes supervisory control

The human-machine interface (Fig. 10) of the system was created in the basic version of SCADA Trace Mode 6. It allows showing the graphs of temperatures, pressure and velocity in TAE resonator; displays the current values of measured parameters and stores the data to database for further analysis. Usage of SCADA software ensure that designed SCS can be integrated in power control systems of existing plants.

This approach facilitates the implementation of thermoacoustic systems into existing energy complexes and lowers the cost of TA waste heat utilization systems implementation.

5 Conclusions

The development of PLC-based systems for data acquisition and supervisory control of environment-friendly energy-saving complex high-tech technologies is presented in this paper.

Modern SCADA information technologies provide unique abilities to the developers of complex data processing and control systems. Range of functions that they provide eases the creation of distributed computerized systems. For example, implemented in SCADA communication protocols, such as ModBus, CAN etc. are standardized and so the data exchange processes can be established with any third-party software, that support any of this protocols.

The examples of SCADA applications in design of PLC-based systems for monitoring and automatic control of ecopyrogenesis and thermoacoustic technological processes are presented. The functional structures, software and hardware implementation as well as multi-level human-machine interfaces of the developed PLC-based systems for data acquisition and supervisory control are given.

The developed PLC-based SCADA-systems provide significant increasing of energy and economic efficiency criteria of the EPG and TAD complexes, high precision control of both technological processes, monitoring of current technological parameters, using the indirect methods for parameters measuring and identifying, and automatic control with high quality indicators and optimal parameters.

In particular, the implementation of the proposed PLC-based SCADA-system in experimental EPG technological complex makes it possible to increase its productivity by 15 %, reduce the maximum value of its total power consumption by 10.2 % and increase its specific productivity by 6 times.

References

1. Halkos, G.E., Tsilika, K.D.: Computing optimality conditions in economic problems. J. Comput. Optim. Econ. Finance. **3**(3), 143–155 (2011)
2. Kosmidou, K., Zopounidis, C.: Assessing the macroeconomic competitiveness of European countries using multicriteria analysis. J. Comput. Optim. Econ. Finance **2**(2), 99–108 (2010)
3. Anthony, R.V., Bilitewski, B., Beckmann, M., Christian, L.: Recycling, thermal treatment and recovery. In: Municipal Solid Waste Management, pp. 44–127. Springer, Berlin (2003)
4. Markina, L.M.: Development of new energy-saving and environmental safety technology at the organic waste disposal by ecopyrogenesis. J. Collect. Works NUS **4**, 8 (2011). (in Ukrainian)
5. Dvořák, R., Pařízek, T., Bébar, L., Stehlík, P.: Incineration and gasification technologies completed with up-to-date off-gas cleaning system for meeting environmental limits. Int. Clean Technol. Environ. Policy **11**(1), 95–105 (2009)
6. Ryzhkov, S.S., Markina, L.M.: Experimental researches of organic waste recycling method of multiloop circulating pyrolysis. J. Collect. Works NUS **5**, 100–106 (2007). (in Ukrainian)
7. Rott, N.: Thermoacoustics. Adv. Appl. Mech. **135** (1980)
8. Swift, G.W.: Thermoacoustics—A Unifying Perspective for Some Engines and Refrigerators. Acoustical Society of America (2002)
9. Ryan, T.S.: Design and control of a standing-wave thermoacoustic refrigerator. Dissertation, University of Pittsburgh (2006)
10. Kondratenko, Y., Korobko, V., Korobko, O.: Specialized computer system for control and parametric optimization of thermoacoustic processes. In: Computing in Science and Technology. Monographs in Applied Informatics, pp. 38–54. Warsaw University of Life Sciences and University of Rzeszow, Warsaw (2013)
11. Kondratenko, Y., Kozlov, O.V.: Mathematical model of ecopyrogenesis reactor with fuzzy parametrical identification. In: Proceedings of the 4th World Conference on Soft Computing, Berkeley, CA, USA, pp. 63–68 (2014)
12. Watson, D., Chakraborty, T., Rodgers, M.: The need for SCADA communication in a Wind R&D Park. Sustain. Energy Technol. Assessments **11**, 65–70 (2015)
13. Erez, N., Wool, A.: Control variable classification, modeling and anomaly detection in Modbus/TCP SCADA systems. Int. J. Crit. Infrastruct. Prot. **10**, 59–70 (2015)

14. Sulthana, S., Thatiparthi, G., Gunturi, R.S.: Cloud and intelligent based SCADA technology. Int. J. Adv. Res. Comput. Sci. Electron. Eng. (IJARCSEE) **2**(3) (2013)
15. Mehta, B.R., Reddy, Y.J.: Chapter 7—SCADA systems. In: Industrial Process Automation Systems, pp. 237–300 (2015)
16. Aydogmus, Z., Aydogmus, O.: A web-based remote access laboratory using SCADA. IEEE Trans. Educ. **52**(1), 126–132 (2009)
17. Kim, H.J.: Security and vulnerability of SCADA systems over IP-based wireless sensor networks. Int. J. Distrib. Sens. Netw. (2012) (Hindawi Publishing Corporation)
18. Md. Ansari, F.: Design and implementation of SCADA based induction motor control. Int. J. Eng. Res. Appl. **4**(3), 5–18 (2014)
19. Stojkovic, B., Vukasovic, M.: A new SCADA system design in the power system of Montenegro—ICCP/ TASE.2 and web-based real-time electricity demand metering extensions. In: Abstracts of the Power Systems Conference and Exposition, Atlanta, pp. 2194–2199 (2006)
20. Tanyi, E., Noulamo, T., Nkenlifack, M., Tsochounie, J.: A multi-agent design and implementation of an internet based platform for the remote monitoring and control of the cameroon power network. Eng. Lett. **13**(2), 195–203 (2006)
21. The use of an integrated platform for creating TRACE MODE SCADA and Production Management: Theory and Practice, 5th International Conference on Materials (2005)
22. Dulău, I.L., Abrudean, M., Bică, D.: SCADA simulation of a distributed generation system with storage technologies. Proc. Technol. **19**, 665–672 (2015)
23. Horng, J.H.: SCADA system of DC motor with implementation of fuzzy logic controller on neural network. Adv. Eng. Softw. **33**, 361–364 (2002)
24. Kondratenko, Y., Kozlov, O.V.: Mathematic modeling of reactor's temperature mode of multiloop pyrolysis plant. In: Modeling and Simulation in Engineering, Economics and Management: Lecture Notes in Business Information Processing, MS 2012, New Rochelle, pp. 178–187. Springer, Berlin (2012)
25. Kondratenko, Y., Kozlov, O.V.: Computerized monitoring and control system for ecopyrogenesis technological complex. In: Tribuna Plural: La revista cientifica, vol. 1, pp. 223–255. Reial Academia de Doctors, Barcelona (2014)
26. Kondratenko, Y., Kozlov, O.V., Klymenko, L.P., Kondratenko, G.V.: Synthesis and research of neuro-fuzzy model of ecopyrogenesis multi-circuit circulatory system. In: Jamshidi, M., Kreinovich, V., Kazprzyk, J. (eds.) Advance Trends in Soft Computing. Studies in Fuzziness and Soft Computing, vol. 312, pp. 1–14. Springer, Berlin (2014)
27. Grafkin, A.V.: Principles of the software control of modules ICP DAS I-7000 in the industrial automation, Samara, Russia, (2010) (in Russian)
28. Badlo, S.G.: TRACE MODE in metallurgy. Mir avtomatizatsii, pp. 68–70 (2007) (in Russian)
29. Strett, J.W.: Theory of Sound. The State Publishing of Technical Theoretical Literature, Moscow (1958)
30. Zink, F., Vipperman, J.S., Schaefer, L.A.: Environmental motivation to switch to thermoacoustic refrigeration. Appl. Therm. Eng. **30**, 119–126 (2010)
31. Biwa, T., Tashiro, Y., Nomura, H., Ueda, Y.: Experimental verification of a two-sensor acoustic intensity measurement in lossy ducts. J. Acoust. Soc. Am. **124**(3), 1584–1590 (2008)
32. Fusco, A.M., Ward, W.C., Swift, G.W.: Two-sensor power measurements in lossy ducts. J. Acoust. Soc. Am. **91**, 2229–2235 (1992)
33. Kondratenko, Y., Korobko, V., Korobko, O.: Distributed computer system for monitoring and control of thermoacoustic processes. In: Proceedings of the 7th IEEE International Conference IDAACS'2013, Berlin, Germany, vol. 1, pp. 249–253 (2013)

Assessment of Energy Consumption for Safety-Related PLC-Based Systems

Vladimir Sklyar, Oleg Odarushchenko, Eugene Bulba, Roman Horbenko, Alexander Ivasyuk and Dmitry Kotov

Abstract The paper describes an approach to measure energy consumption of programmable components (MCU, FPGA etc.) which are a core of Programmable Logic Controller (PLC). Investigation of energy consumption dependence from a type of programmable components as well as from features of software gives a road map for implementation of energy efficiency measures to make PLCs “greener”. Special toolset named GreenCo Controller has been designed, implemented and tested for the above research. A designed architecture includes the following components: a commercially available motherboard (Arduino UNO) programmed with software to perform data processing in respect to energy consumption measurement of the target board-shield (GreenCo board); GreenCo board on the base of Microchip dsPIC30F3011 MCU programmed with different type of application software as a subject of energy consumption investigation; a software monitor for PC desktop using named GreenCo Controller GUI. “Green” features of PLC are analyzed after that formal optimization problems in accordance with criterion “energy consumption level/safety level” have been stated. Such problems can be solved with dynamic programming method. Concept and design of GreenCo Controller are described. Areas of future researches with GreenCo Controller are proposed as a conclusion.

V. Sklyar (✉) · O. Odarushchenko · E. Bulba · R. Horbenko · A. Ivasyuk · D. Kotov
Research and Production Corporation Radiy,
29, Geroyiv Stalingrada Street, Kirovograd, Ukraine
e-mail: v.sklyar@radiy.com

O. Odarushchenko
e-mail: odarushchenko@gmail.com

E. Bulba
e-mail: evhenb@gmail.com

R. Horbenko
e-mail: romanich1987@gmail.com

A. Ivasyuk
e-mail: ivasiuk.radiks@gmail.com

D. Kotov
e-mail: kotovdmitry@ukr.net

V. Kharchenko et al. (eds.), *Green IT Engineering: Concepts, Models, Complex Systems Architectures*, Studies in Systems, Decision and Control 74,
DOI 10.1007/978-3-319-44162-7_14

Keywords PLC · Industrial control system · Arduino

1 Introduction

One from the main trends in automation of industrial plants is decreasing of functions performed by embedded systems including Programmable Logic Controllers (PLC). PLCs perform all the main functions of Industrial Control Systems (ICS) including safety-related functions. Energy efficiency of embedded systems is a hot topic taking into account greenhouse issues as well as lean production. At the same time there is a deficiency of sufficient tools for modeling and adjustment of energy efficient working of PLC-based ICS. Safety-related ICS is a separated knowledge domain. Such systems are not energy saving in their nature since for implementation we use different types of redundancy. Spending of extra resources for safety implementation requests a rational approach to efficiency including efficiency of energy consumption.

A goal of this paper is to develop an approach and tools for modeling of energy consumption for PLC-based systems including safety-related applications. After that investigation of impact of different PLC properties on energy consumption shall be performed. The follow up implementation of the investigated energy saving techniques would give us evidences of practical usability of the obtained results.

This paper describes a project directed to design, production and testing of a tool for measuring of energy consumption of embedded system. This tool has been named GreenCo Controller after the name of the TEMPUS GreenCo Project (Green Computing and Communication, 530270-TEMPUS-1-2012-1-UK-TEMPUS-JPCR), see http://my-greenco.eu/.

A full toolset includes Arduino UNO board, special shield for Arduino UNO named GreenCo board, interfaces, and desktop application for results monitoring of energy consumption measurement.

Parts of this chapter sequentially describe architecture of the tool as well of results of implementation.

GreenCo Controller implementation has been started from a concept with specification development. This specification describes requirements to GreenCo Controller as for a “black” box. After that product implementation these requirements have been used as an input for testing. Architecture of a tool set includes the following components:

- Arduino UNO (https://www.arduino.cc/en/Main/ArduinoBoardUno) serves as a motherboard with performing functions of energy consumption measurement as well as data transmission via USB-interface to a Personal Computer (PC);
- Arduino UNO software supports functions of the motherboard Arduino UNO;
- Specially designed GreenCo board with a target device which is a 16-bit Microchip dsPIC30F3011 microcontroller unit (MCU) (http://www.microchip.com/wwwproducts/en/dsPIC30F3011); this device is used as a connected shield

for Arduino UNO; MCU is programmed with a target code and MCU's current can be measured by Arduino UNO;
- Microchip PICkit3 In-Circuit Debugger is used for programming of Microchip MCU;
- A target GreenCo board software which is installed in Microchip MCU to investigate MCU's energy consumption with using of specific software and compiler options, MCU's frequency and other affecting parameters;
- A software monitor for PC desktop application named GreenCo Controller GUI (Graphical User Interface); this software provides service functions such as energy consumption data visualization and archiving.

Design description of the above components will be done below in the paper.

After that some ideas from code optimization theory will be discussed. A basic hypothesis is that more efficient code performance entails less power consumption of the target MCU. Microchip MCU voltage is setup with MCU initialization options. Microchip MCU current is measured by MCU ATmega on the board Arduino UNO. Electrical power of MCU can be calculated and compared based on measurement of voltage and current for competitive software codes. Also power consumption can be calculated for a target time of software execution or for single execution of non-cyclic software.

Results of experiments are given and discussed. Recommendations for achievement of PLC energy efficiency are given.

There is not an essential quantity of publications devoted to PLCs energy efficiency issues. Since so many data centers are operated around the world, a focus of research in Green Information Technologies (IT) is directed to energy efficiency of this subject. Books [1, 2] can be recommended for familiarization with Green IT aspects of data centers. A large review of existing problems in Green IT and approaches to solve its can be found in a monograph [3]. PLC reusable design is often considered as one from the possible approaches to efficiency in general [4]. There are also researches devoted to different aspects of electronic devices power consumption [5, 6]. It is essential to underline importance of university-industry cooperation in area of Green IT [7]. Thus the actual paper provides novelty in Research and Development (R&D) related to Green IT and energy efficiency of PLCs.

2 Green Features of PLC

As well as PLCs perform safety functions, there is an issue how does PLCs safety comply with PLCs ecological and green features. A general approach to combine safety, eco-friendliness and greenness for IT is presented on Fig. 1.

In this paper PLCs feature are discussed in the same paradigm as on Fig. 1. Safety systems are redundant what entails some extra resource spending for design, production, operation, maintenance, and extra energy consumption. At the first glance,

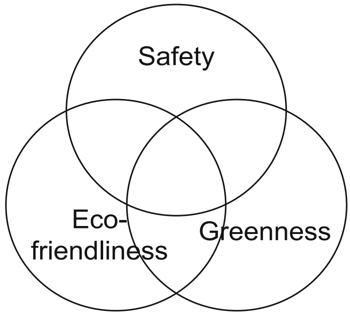


Fig. 1 Features composition in safe, eco-friendly and green IT

safety increasing entails decreasing of both eco-friendliness and greenness. However from the other hand, safety systems are used for environment risk reduction. Also safety increasing for power plants provides energy on the safe and stable basis what is a contribution for efficient energy using. So from this point of view safety increasing can improve eco-friendliness and greenness in the long term run.

At the same time we have a problem of energy consumption (EC) optimization for safety-related PLC-based systems. For safety measurement we can use one or some (for, example, N) weighted numerical safety indicators (SI). For example, in accordance with the standard IEC 61508, there are such SIs as average probability of a dangerous failure on demand of the safety function or average frequency of a dangerous failure of the safety function. Such problem has two ways of formulation. Both problems can be solved with dynamic programming method.

Problem 1 To find a PLC-based system for which $EC \rightarrow \min$ when $SI \geq SI_{required.}$ The target function is $f(EC) = \sum_{i=1}^{N} EC_i \cdot \frac{1}{SI_i} \rightarrow \min$ with limitations $\sum_{i=1}^{N} SI_i \geq SI_{required}$.

The problem to find a minimal EC with the required level of safety can be solved with a dynamic programming method. A solution is a set of programmable components chosen for a PLC implementation COMPONENTS ($EC \rightarrow \min$) = $\{Component_i\}$, with summary energy consumption $EC = \sum_{i=1} EC_i\left(\frac{EC_i}{SI_i} = \min\right)$ and a set of safety indicators $SI = \left\{SI_i\left(\frac{EC_i}{SI_i} = \min\right)\right\}$.

Problem 2 To find a PLC-based system for which $SI \rightarrow \max$ when $EC \leq EC_{required.}$ The objective function is $f(SI) = \sum_{i=1}^{N} SI_i \cdot \frac{1}{EC_i} \rightarrow \max$ with limitations $\sum_{i=1}^{N} EC_i \leq EC_{required}$.

The problem to find a maximal EC level of safety with the required EC can be solved with a dynamic programming method. A solution is a set of programmable components chosen for a PLC implementation COMPONENTS ($SI \rightarrow \max$) = $\{Component_i\}$, with summary energy consumption $EC = \sum_{i=1} EC_i\left(\frac{SI_i}{EC_i} = \max\right)$ and a set of safety indicators $SI = \left\{SI_i\left(\frac{SI_i}{EC_i} = \max\right)\right\}$.

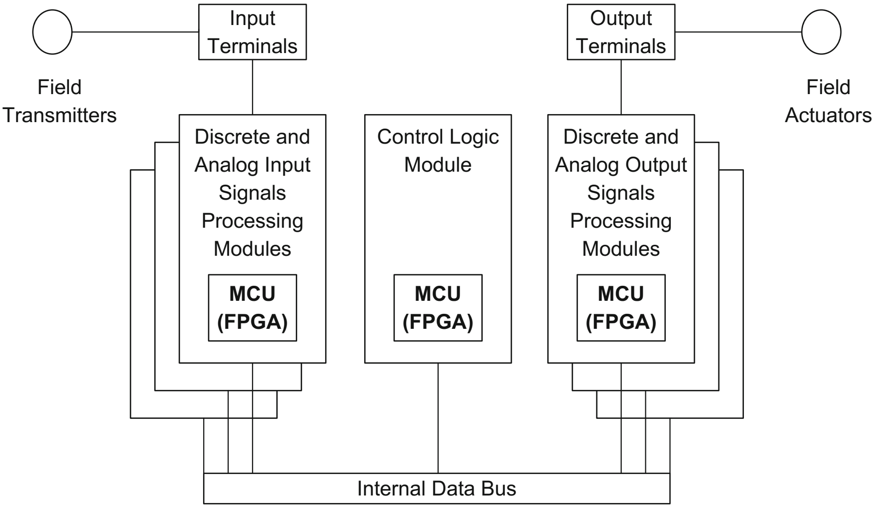


Fig. 2 Typical PLC structure

Let's consider a typical structure of PLC which is presented at Fig. 2.

PLC functions include processing of data received from field transmitters, control logic performance in accordance with implemented applications and control signals forming for actuators. So we have in the PLC Input/Output modules with Control Logic module. All modules have on the board MCU and/or Field Programmable Gates Array (FPGA) which are the most variable components from the point of energy consumption view. It is a reason why in our research we focus on energy consumption feature of MCUs to extend the results to PLCs.

From the prospective of EC optimization we may use two approaches. The first approach is to use energy efficient (green) MCU chips. The second approach is to use energy efficient (green) software for PLC functions implementation.

Microchip has been chosen for this research as one from the lider in area of energy efficient solutions. For example, Microchip Technology Inc. states the following low power EC metrics for MCU familiy named PIC (with 3 V of power supply range) supported with as named nanoWatt XLP (eXtra Low Power) Technology:

- Seven types of operating modes with different options of low EC depending on application scenario;
- Current range less than 1 μA in the sleeping mode including support of watchdog timer for the most part of MCUs;
- Current range down to 9 nA in the deep sleeping mode what means up to 30 years of operation with using typical on-boar battery with capacity of 225 mAh;
- Run currents down to 34 μA/MHz;

- Supporting with free tool XLP Battery Life Estimator which calculates life time for batteries with different capacities based on assumptions concerning Microchip PIC MCU type, operating modes as well as peripherals and input application currents.

Microchip dsPIC30F3011 has been chosen for implementation of GreenCo Controller and for the followed researches related with EC measurement.

3 Concept and Design of GreenCo Controller

The following ideas have been considered as inputs for GreenCo Controller concept:

- GreenCo Controller may be designed with using both commercially available kits and self developed Printed Circuit Boards (PCB);
- GreenCo Controller shall include power supply unit to provide different voltage ranges for the investigated PIC MCU; this feature also could be used to investigated MCU operation in different failure modes (for example, with high or low power supply voltage);
- GreenCo Controller shall include current measurement unit for monitoring of current and energy consumed by the investigated PIC MCU;
- GreenCo Controller shall support different clock frequencies for PIC MCU to investigate dependency between MCU EC and operability in different modes;
- GreenCo Controller shall include temperature measurement unit;
- GreenCo Controller shall support standard interface (USB) to be easy connected with PC.

The developed concept has been used as an input for design of GreenCo Controller. A functional diagram is presented on Fig. 3.

A design of GreenCo Controller includes Arduino UNO board connected with a shield on the base of Microchip dsPIC30F3011 (GreenCo board). This board supports interfaces with GreenCo board and interfaces with PC. Built-in USB interface supports data transmission to PC with Arduino board power supply and Arduino MCU programming. Two the last functions can be performed with alternative power supply and programming interfaces as it is presented on Fig. 3.

GreenCo board requests to use Microchip PICkit3 In-Circuit Debugger as programmer and power supply device. Alternative external power supply input with 9 VDC may be used if necessary. As an option, Microchip dsPIC30F3011 can drive different peripheral devices such as ADCs, LEDs, indicators etc.

An adjustable DC/DC convertor is used to provide stable voltage for the MCU in the range 2.5–5.5 VDC. At the DC/DC convertor output an additional precise resistor (shunt) is installed. This shunt provides both value of voltage and value current to be measured and converted to electrical EC. GreenCo board contains also temperature sensor and current measurement unit.

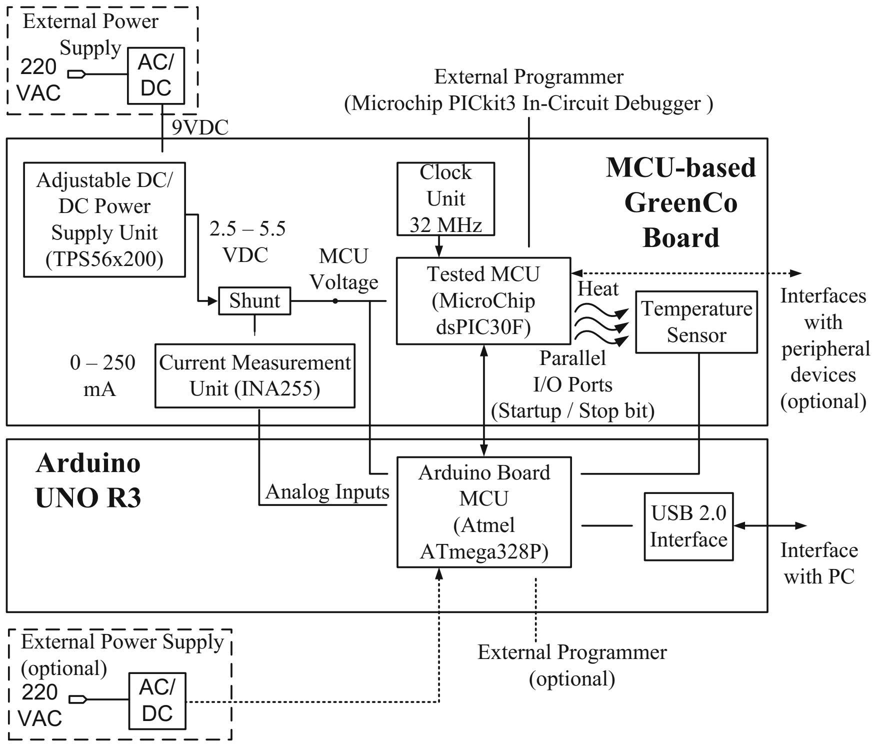


Fig. 3 Functional diagram of GreenCo Controller

Clock unit provides 32 MHz frequency for Microchip dsPIC30F3011. However the MCU is equipped with embedded feature of frequency control. It supports both low power internal resistor-capacitor (RC) oscillator with 512 kHz (LPRC) and fast internal RC oscillator 7.37 MHz (FRC). The FRC supports a feature of Phase Locked Loop (PLL) based frequency multiplication with factor values of 4, 8, and 16, e.g. 4×7.37 MHz = 29.48 MHz, 8×7.37 MHz = 58.96 MHz, 16×7.37 MHz = 117.92 MHz. The above features provide opportunities to investigate different frequency modes of the MCU.

The following interfaces connect GreenCo board with Arduino UNO board:

- MCU current value measured with current measurement unit;
- On board temperature value measured with temperature sensor;
- MCU voltage value measured at the shunt output;
- Startup (stop) bit for begin or end of measurement process from the Arduino UNO board site.

A 3-D model of GreenCo board is presented at Fig. 4.

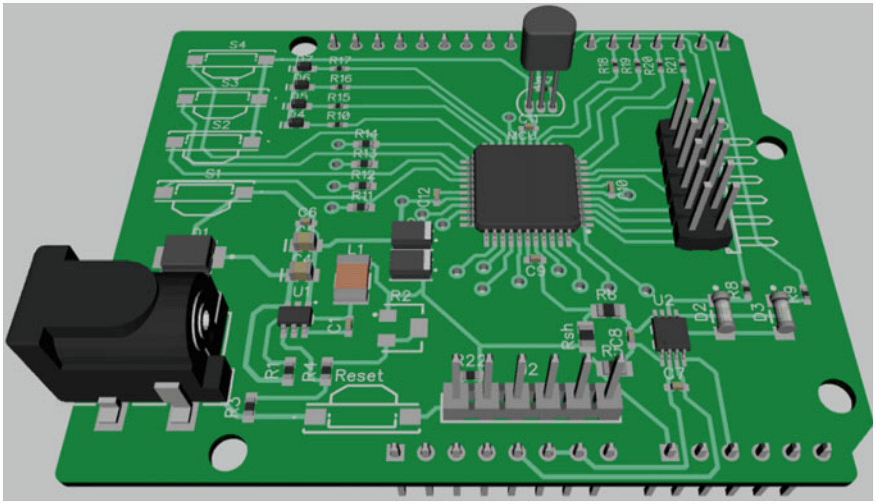


Fig. 4 A 3-D model of GreenCo board

4 Software Functions and Integration of GreenCo Controller

Software of GreenCo Controller includes the following parts:

- Arduino UNO software; this software performs functions of current, voltage and temperature measurement as well as measured data processing and transmission to PC via USB-interface; once developed, Arduino software is still not changeable (like firmware for the end user);
- GreenCo board software; this software may implement different kinds of algorithms which a subject of energy consumption measurement and research;
- GreenCo Controller GUI software; this software monitors in real time values of voltage and current which are measured by Arduino UNO board; this software also provides plots for voltage and current as well as calculates on this base instantaneous and average values of electrical power and energy consumption.

Arduino UNO software and GreenCo board software are written in C, and GreenCo Controller GUI software is written in Java.

General architecture of integrated parts of GreenCo Controller is presented on Fig. 5. GreenCo Controller as been successfully tested against input requirements and performance ability has been proven during some startup researches.

During experiments performance the end user may calculate automatically, for example, the following indicators:

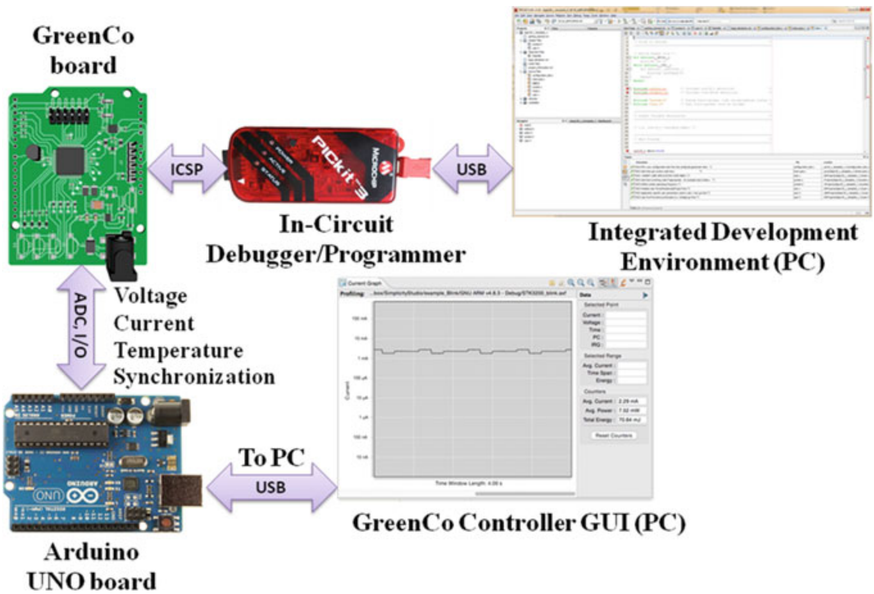


Fig. 5 Integrated GreenCo Controller

- Average value of instantaneous power is calculated as $P = \sum_{i=1}^{N} I_i \cdot U_i / N$, where I_i is a current value measured at the time t_i, U_i is a voltage value measured at the time t_i, N is a quantity of measurement (integration rate);
- Energy consumption value is calculated as $EC = \sum_{i=1}^{N} I_i \cdot U_i \cdot t_i$;
- Average power value for during calculation time t(N) is calculated as $P_{avg} = EC/t(N)$.

5 Case Study: Energy Consumption Improvement for Optimized Software Code

Let's consider as an example influence of software code optimization for a simple cycling algorithms. A code in C is presented below.

```
int data = 0;
int value, j, i;
for (j = 0; j < 200; j++) {
        value = rand()%20;
        for (i = 0; i < 10000; i++)
        { if (value> 10) data++;
            else data--; } }
```

This code can be optimized from the point of view of performed cycles quantity. For that let's put in the code a small update concerning consequence of cycle operator (see below).

```
int data = 0;
int value, j, i;
for (j = 0; j < 200; j++) {
    value = rand()%20;
    if (value> 10)   {
        for (i = 0; i < 10000; i++)
        data++;      }
    else { for (i = 0; i < 10000; i++) data--; }
```

After that source files were prepared for running in Microchip dsPIC30F3011 for both optimized and not optimized algorithms. GreenCo Controller provide features to measure current, voltage, power, temperature and EC (power consumption as at screenshots). Measurements have been performed for 5, 3.3, and 2.5 V of MCU power supply voltage. Obtained results for 5 V MCU power supply are presented at Figs. 6 and 7.

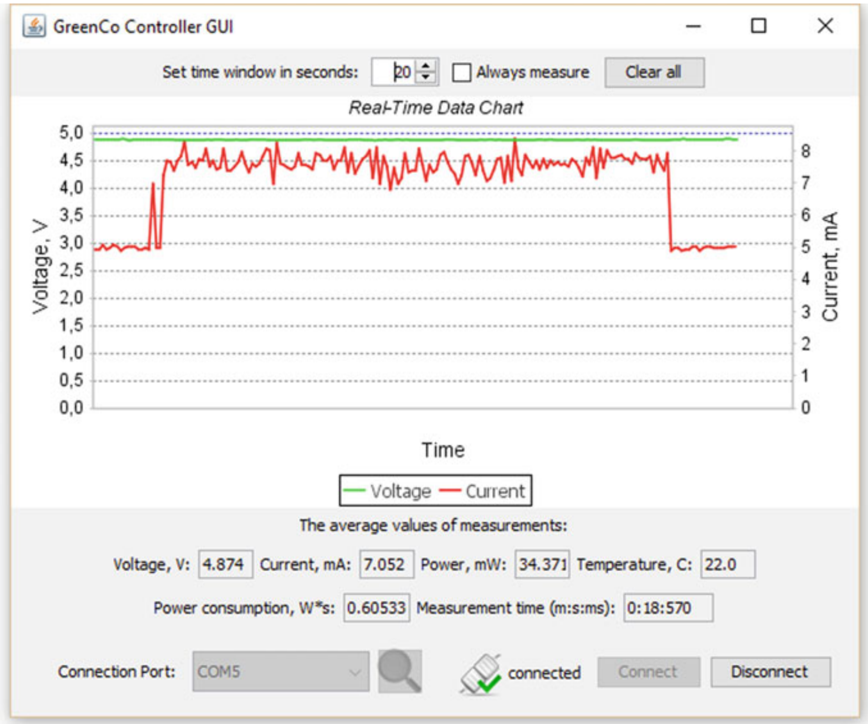


Fig. 6 Result of energy consumption measurement for not optimized code at 5 V power supply of the MCU

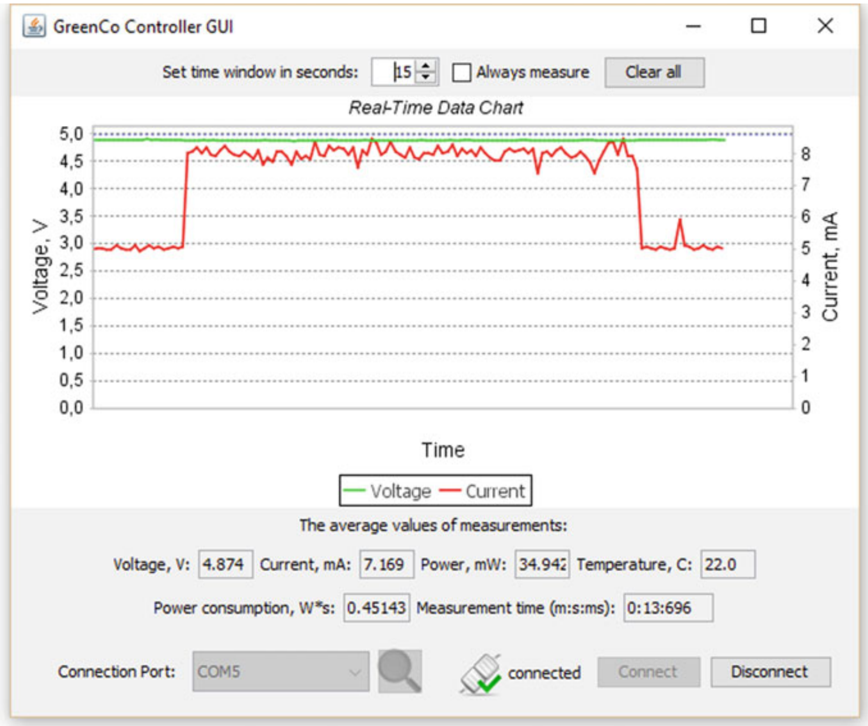


Fig. 7 Result of energy consumption measurement for optimized code at 5 V power supply of the MCU

Table 1 Results of energy consumption analysis at 2.5, 3.3, and 5 V power supply of the MCU

Experiment description	Voltage (V)	Current (mA)	Power (mW)	Temperature (°C)	Run time (s)	EC (W s)
2.5 V, not optimized code	2.36	3.24	7.63	22.0	18.46	0.134
2.5 V, optimized code	2.36	3.28	7.72	22.0	13.69	0.100
3.3 V, not optimized code	3.42	4.49	14.57	22.0	18.56	0.258
3.3 V, optimized code	3.42	4.56	14.77	22.0	13.69	0.191
5 V, not optimized code	4.87	7.05	34.37	22.0	18.57	0.605
5 V, optimized code	4.87	7.20	34.94	22.0	13.70	0.451

The same results have been obtined for power supply voltage of 3.3 and 2.5 V. All integrated results are collected in Table 1. For voltage, current, power and temperature we provide average value during MCU run time.

The following conclusions are obtained on the base of the measurements analysis.

There is not a variation in the average voltage values for optimized and not optimized code running, but there is a small increasing of the average current value for optimized code what entails a small decreasing for the average value of instantaneous power.

Average temperature is the same for all the considered cases.

Since all electrical parameters are the same for optimized and not optimized code, difference in EC is caused by shorter running time for optimized code. There is not a difference in a run time for all the values of MCU voltage. So value of relation between EC for optimized and not optimized code is the same for all the values of MCU voltage EC(optimized code)/EC(not optimized code) = 0.74. It means code optimization for the above case study provide 26 % of energy consumption saving.

6 Conclusions: The Future Researches

As well as valuable toolset for PLC EC modeling has been successfully designed and implemented a plan of future research is developing. We consider the following ideas for experiments.

The biggest issue is research of MCU EC depending on implementation of control algorithm with software code. This issue is like one from the "green" software research directions. It I known software EC s highly dependent on dynamic behavior so it can be simulated and investigated with control flow graph complexity of the software (see Fig. 8) [5]. Such approach is well known as code optimization so it seems valuable to learn MCU EC with software diversity.

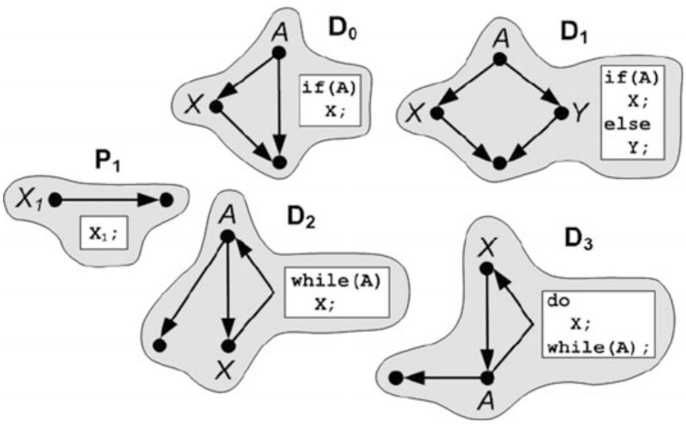


Fig. 8 Software control flow graphs depending on complexity of used operators

FPGAs and other programmable components type (CPLD, ASIC) using can provide the basis for selection of chips depending on requirements to EC.

GreenCo Controller also provides opportunities to investigate EC affection of used programming languages as well as compiler options setup.

One more research direction is investigation of EC affection of hardware feature such as operating voltage and frequency, temperature rate, I/O pins load and switching rate, oscillator frequency and others.

References

1. Velte, T., Velte, A., Elsenpeter, R.: Green IT: Reduce Your Information System's Environmental Impact While Adding to the Bottom Line, 281 p. McGrow Hill Companies (2008)
2. Poniatowski M.: Foundation of Green IT: Consolidation, Virtualization, Efficiency, and ROI in the Data Center, 352 p. Prentice Hall (2009)
3. Murugesan S., Gangadharan G.R., (eds.): Harnessing Green IT: Principles and Practices, 389 p. John Wiley and Sons Ltd. (2012)
4. Bonfatti, F., Gadda, G., Monari, P.D.: Re-usable software design for programmable logic controllers. In: LCTES'95 Proceedings of the ACM SIGPLAN 1995 Workshop on Languages, Compilers, & Tools for Real-Time Systems, vol. 30(11), pp. 31–40, Nov 1995
5. Chandrakasan, A., Sheng S., Brodersen, R.W.: Low Power Digital CMOS Design. Low-Power CMOS Design, IEEE Press, New York (1998)
6. Landman P.: High-level power estimation. In: Proceedings of the International Symposium on Low Power Electronics and Design. Monterey, CA, pp. 29–35 (1996)
7. Starov, O., Kharchenko, V., Sklyar, V.: Startup company and spin-off advanced partnership via web-based networking. In: Proceedings of the Universities-Industry Interaction Conference "UIIC2013". —Amsterdam, The Netherlands, pp. 294–310, May 27–29, 2013

Green Microcontrollers in Control Systems for Magnetic Elements of Linear Electron Accelerators

Anatoliy Shamraev, Elena Shamraeva, Anatoly Dovbnya, Andriy Kovalenko and Oleg Ilyunin

Abstract This chapter presents an approach to industrial control systems design, which is based on application of "green" microcontrollers (with low power consumption) used in complex real-time control systems. A wide range of capabilities of modern microcontroller peripheral units allows us to implement a system with wide functionality based on a single microcontroller chip. We describe the order of connection of microcontroller pins, which reduces noise influence on the results of signal processing. As an example, the proposed approach is considered in the context of development of the power supply control system for magnetic elements of linear electron accelerators.

Keywords Green microcontrollers · Energy consumption · Embedded systems · Linear accelerator · Energy spectrum measurement · Temperature regulator

A. Shamraev (✉) · E. Shamraeva
Belgorod State National Research University, 85, Pobedy str., Belgorod, Russia
e-mail: shamraev@bsu.edu.ru

E. Shamraeva
e-mail: shamraeva@bsu.edu.ru

A. Dovbnya
National Science Center "Kharkov Institute of Physics and Technology", 1, Academicheskaya str., Kharkov, Ukraine
e-mail: dovbnya@nik.kharkov.ua

A. Kovalenko
Centre for Safety Infrastructure-Oriented Research and Analysis, 37, Astronomicheskaya str., Kharkov, Ukraine
e-mail: a.kovalenko@csis.org.ua

O. Ilyunin
Kharkiv National University of Radio Electronics, 14, Nauki Ave., Kharkov, Ukraine
e-mail: khioo1967@ukr.net

V. Kharchenko et al. (eds.), *Green IT Engineering: Concepts, Models, Complex Systems Architectures*, Studies in Systems, Decision and Control 74,
DOI 10.1007/978-3-319-44162-7_15

1 Introduction

1.1 Motivation

For modern devices with autonomous power supply the microcontroller's power consumption is becoming more and more important. Power consumption is one of key parameters for modern electronic devices, which are becoming increasingly mobile and, in turn, often require autonomous power supply. This becomes even more important with various sensors, measuring and telemetry equipment, medical equipment; many of these devices are designed to be installed in hard-to-reach places and it makes their assembly and maintenance more complicated. Moreover, nowadays there is a trend to create devices that do not have built-in battery. They receive power supply directly from the environment—sunlight or artificial light, vibrations, etc. In order to choose the most efficient solution in terms of power consumption it is necessary to compare the microcontrollers' characteristics in all allowed working modes (including low-power modes), to consider synchronization techniques, characteristics of event processing system, built-in peripherals properties, the availability of power failure detection and protection in real-time, as well as the amount of leakage and efficiency of computer core.

Optimization of energy consumption is also an important task while designing control systems for industrial and scientific-technical objects, which include linear electron accelerators considered in this chapter. Accelerators are used both in scientific research and handling of applied problems (for example, in medicine for sterilization and radiological examinations of patients).

1.2 Work Related Analysis

Embedded digital data processing systems represent one of application domains for microcontrollers. The tasks to be solved by microcontrollers in such systems can be divided into the following three groups:

- discrete (relay) control;
- analog control;
- digital signal processing (DSP).

The first group includes digital machines, counting devices and programmators. These tasks are characterized by predominance of logical operations of data processing. The amount of calculation is small and can be solved within integer arithmetic. Input and output data are handled by devices of relay type. Control is driven by events, as well as in accordance with time diagram with a reaction time calculated in fractions of a second.

The tasks of analog control are the following: processing information received from analog sensors, related calculations and further issuing of control actions to actuating analog devices. Requirements to calculations accuracy are determined by sensors and actuators' accuracy, which is, as a rule, equal to 8–12 bits. Control is carried out with frequency that provides the required time of response to external stimuli in milliseconds range.

DSP tasks include a significant amount of calculations on data in real time with electronic signal converter's accuracy being, as a rule, equal to 12–24 bits. The intermediate calculations are performed with pinpoint accuracy in order to prevent accumulation of errors. Data processing is performed at signal sampling frequency from 10 kHz to 10 MHz and, possibly, higher.

The efficient architecture of a microcontroller core has to assure minimization of the amount of program code and cycles spent on its implementation. In addition, the processor architecture should:

- be based on the uniform use of a group of conceptual principles of architecting the system of commands that ensure an increase in data width and address space;
- provide a reduction in the semantic gap with the programming language;
- contribute to the achievement of the best characteristics of power consumption and cost.

Various manufacturers of microcontrollers solve these problems in different ways. Therefore, on the market there are many microcontrollers with different architectures at the same time. Until now 8-bit microcontrollers were the most popular ones in the field of embedded systems. In terms of price/performance ratio they are an excellent solution. However, nowadays the situation is different. In terms of 16-bit applications, manufacturers of microcontrollers are about to create a single multi-purpose universal device, for example, MSP430 microcontrollers' family by Texas Instruments is able to resolve the entire range of typical tasks for embedded systems. They differ from existing 32-bit microcontrollers by a number of advantages including:

- low price;
- ultra-low power consumption;
- availability of high-quality tools for developing and debugging.

The MSP430 microcontrollers incorporate a 16-bit CPU with reduced instruction set computer (RISC CPU), "smart" peripherals, and a flexible clock system that interconnect using a von-Neumann common memory address bus (MAB) and memory data bus (MDB) [1–4]. The combination of modern CPU and memory mapped analog and digital peripherals makes the MSP430 suitable for use in mixed signals processing applications.

Clock system. Microcontroller power consumption mostly depends on its clock system. The MSP430 clock system is specifically designed for use in battery powered devices. Low-frequency auxiliary clock (ACLK) is generated by watch crystals 32,768 Hz. The ACLK signal can be used to periodically "wake-up" background real-time clock. Built-in high-frequency digitally controlled oscillator can generate the main clock signal, used by the CPU and high-speed peripherals. Wake-up time of this oscillator is less than 1 μs at 1 MHz.

Solutions based on MSP430 MCU effectively use the high-performance 16-bit RISC CPU while very short time intervals, namely:

– low-frequency auxiliary clock is used to implement standby mode with ultra-low power consumption;
– high-frequency main clock is used in signal processing.

In addition, when analog-to-digital converter (ADC) is operating, the CPU core can be stopped, that will reduce the system consumption.

Power-Saving Modes. Availability of low-power modes is the most important feature, making it possible to satisfy the average current consumption requirements. The MSP430 family is designed for ultralow-power applications and uses five low-power modes [4]. In the Low-power mode zero (LPM0) CPU is switched off, but all other functions remain active. In LPM1 and LPM2 disabled list function adds different clock functions. LPM3 is the most commonly used low-power mode, at which only the low-frequency oscillator and peripherals, which are driven directly from this oscillator, are active. In LPM4 mode all oscillators are disabled, as result—all synchronous peripherals are turned off automatically. In this mode the analog peripherals can remain in an active mode, but they also may be turned off to decrease the microcontroller (even performing regeneration RAM) consumption to less than 100 nA.

Smart peripherals. MSP430 devices were originally designed as a low-power, and, therefore, their peripherals also have low consumption. ADC of the MSP430 microcontroller is an example of such smart low consumption peripherals. If the ADC does not perform any conversion, it does not consume current and blocks both internal oscillator and a digital processing circuit. Before starting the conversion ADC is automatically turned on or restarted. In addition, the peripheral devices can be controlled by interrupts from other units. For example, the ADC of the MSP430 microcontrollers can be run by timer «Timer A» or «Timer B», that allows to synchronize and execute ADC-sampling without using CPU-core. ADC can activate the processor core after the end of conversion. Some MSP430 devices have direct memory access (DMA) unit, which allows to process data automatically without the CPU intervention. Application of the DMA controller can not only increase the data processing speed, but, more importantly, reduce power consumption. Application of the DMA in automatic data recording tasks allows CPU-core to be disabled, and allows ADC to operate and to perform the conversion at this time: CPU-core is being activated after all necessary conversions are made.

Interrupt handling. Interrupts can wake-up microcontroller from low-power mode, therefore, the greater number of interrupts is used the more flexible microcontroller reduces overall consumption. The pins and peripherals states polling will inevitably lead to an increase in consumption, since it would require an additional microcontroller activity. MSP430 devices have interrupt handlers from the 16 I/O pins, as well as all peripherals. Some peripheral devices, such as, for example, timers «Timer A», «Timer B» and ADC, in order to provide greater flexibility, have multiple interrupts.

Data processing efficiency. CPU MSP430 contains 16 fully addressable 16- or 20-bit registers, and implements the most instructions within single cycle. Each of the registers, including the program counter, stack pointer, and status register can be used as a source (destination) pointer or as an operand. Architecture is orthogonal, as all instructions and addressing modes are the same for all of the internal MSP430 space. In addition, all commands support operation directly in memory without intermediate data loading into the CPU registers. A wide range of addressing modes, including byte, 16-bit and 20-bit addressing allows to provide high efficiency of branch algorithms, tables or data arrays processing. Such structure is optimized for high-level programming languages and allows to obtain a very compact code, which, in turn, leads to a decrease in consumption.

Microcontrollers market segmentation can be summarized as follows: 4-bit microcontrollers occupy a share of 10 % of the whole market. Mainly, these are mass products to be used in electronic watches, calculators, as well as controllers in household equipment and automobile electronics. 16-bit microcontrollers, in turn, occupy a share of about one-third of the whole market. However, recently due to the proliferation of digital signal processing devices they have started to give way to 32-bit microcontrollers, which provide more efficient production. The major share of microcontrollers market (about 50 %) is occupied by 8-bit microcontrollers due to the best price/performance ratio. Modern 8-bit microcontrollers can operate at frequencies up to 100 MHz, which allows them to solve not only the problems of logical and digital control but also the problems of digital signal processing.

In the world of 32-bit microcontrollers there have been several events that have changed the balance of power dramatically for the last years. The niche of exclusive application of 8-bit microcontrollers has happened to be under great pressure from the side of 16-bit microcontrollers and 32-bit ARM-devices (the name comes from "Advanced RISC Machines"). ARM microcontroller core was developed by a similarly-named British company founded in 1990. ARM Limited Company supplies its developments in electronic form, on basis of which customers design their own microcontrollers. The company's customers are more than 60 manufacturers of semiconductor components. Such popular companies as the Altera, Analog Devices, Atmel, Cirrus Logic, Fujitsu, MagnaChip (Hynix), Intel, Motorola, National Semiconductor, Philips, ST Microelectronics and Texas Instruments are among them. At the same time each manufacturer of ARM-microcontrollers has its own "flavors" in the field of embedded peripheral modules aimed at their use in various specific applications.

Nowadays ARM architecture takes a leading position and covers a share of 75 % of 32-bit embedded RISC-microprocessors market. The popularity of this core is due to its standardization, which allows a developer to use the software of both his own and third-party with more flexibility (during both transition to new ARM-core and migrations between different types of ARM-microcontrollers).

At the present time the following major families are developed: ARM7™, ARM9™, ARM11™, Cortex A™, Cortex R™, Cortex M™, SecurCore™ [5, 6].

The main requirement for real-time systems is fast and determinated response to external events. However, sometimes meeting this requirement leads to less efficient performance. Some cores and processors, designed for real-time systems, have no cache memory. The absence of cache memory provides determinated response to external events.

Data protection systems are mainly based on processors with the SecurCore™. The most important advantage of ARM core is its being standard, which allows us to use software from other compatible microcontrollers to get broader access to generic development tools and to arrange migration between microcontrollers more easily. Despite using the same ARM core in microcontrollers from different manufacturers each of them has its own characteristics that are achieved by the original set of peripherals. ARM-cores have a representative variety and dynamics of development. However, mainly the microcontrollers based on Cortex™-M0 and Cortex™-M3 cores are available for mass consumption.

Market of 32-bit microcontrollers has a high capacity that will grow rapidly in the coming years. The book [7] is devoted to a comprehensive description of the construction aspects and application of STM32 microcontrollers including their possible configurations, feasibility of communications and their applicability in display devices, storage devices and timers.

Particular attention is paid to such promising areas as use of this microcontroller family in processor modules, a variety of sensors and standardized communication modules.

It is important to note that STM32L1 family of STMicroelectronics Company has extremely low power consumption due to a number of contributing functions. Nowadays many configurations can be implemented; it contributes to support of applications with mutually conflicting requirements, especially in the aspects of cost and processing power. The main task for developers is to select the most suitable compromise as for power consumption and computing performance [8]. Many scientists in their works also try to summarize the whole range of factors influencing the choice of energy-efficient microcontrollers. Thus, researches in [9] are devoted to the study of the influence of microcontroller's operating frequency on its power consumption, the impact of the time ratio of microcontrollers' operations in different allowable working modes (and transitions between them) on the total energy consumption, as well as the impact of variations in supply voltage.

The authors of [1–4] considered the green-aspect of MSP430 microcontrollers' family when they are used for the implementation of high-speed USB-interfaces in portable equipment and data collection devices. They note the effectiveness of this approach, simplicity and ease of implementation of the data transfer functionality, dozens of supported device classes, as well as energy efficiency of microcontrollers' family.

Among the existing microcontrollers' families in the ultra-low power consumption segment it is possible to distinguish the following: MSP430F2/4x, MSP432P4x, STM32L1, EFM32GG [1–4, 8, 9]. Table 1 compares the main characteristics of these microcontrollers' families.

Table 1 Comparison of microcontrollers' families STM32L1, EFM32GG and MSP430F2/4x

	STM32L1	EFM32GG	MSP430F2/4x	MSP432P4x
Company	STMicro-electronics	Energy micro	Texas instruments	Texas instruments
Core	Cortex-M3	Cortex-M3	MSP430	Corte-M4F
Bit depth (bit)	32	32	16	32
Clock frequency	32 kHz–32 MHz	32 kHz–48 MHz	32 kHz–25 MHz	32 kHz–48 MHz
Timers	1 × 32 bitGP 6 × 16 bitGP 2 × Basic 2 × Watchdog RTC	4 × 16 bitGP 1 × LETIMER 1 × Watchdog 3 × PulseCounter RTC	4 × 16 bitGP 1 × Watchdog RTC	4 × 16 bitGP (5 capture, compare, PWM capability) 2 × 32 bitGP 1 × Watchdog, RTC
Communication interfaces	USB 5 × USART 3 × SPI/I2S 2 × I2C 1 × SDIO	USB OTG 3 × USART/SPI/I2S 2 × UART 2 × LEUART 2 × I2C	USB 4 × UART/LIN/ IrDA/SPI 4 × SPI/I2C	4 × eUSCI_A (UART/IrDA/SPI) 4 × eUSCI_B (SPI/I2C)
Analog peripherals	40Ch × 12 bit ADC 2Ch × 12 bit DAC	8Ch × 12 bit ADC 1Ch × 12 bit DAC	16Ch × 12 bit ADC 2Ch × 12 bit DAC	24Ch × 14 bit ADC 1-MSPS SAR ADC
DMA	1 × 7Ch + 1 × 5Ch DMA	1 × 12Ch DMA	1 × 6Ch DMA	1 × 8Ch DMA
Current consumption in the active mode (μA/MHz)	290	200	<100	90
Current consumption (μA) in stop mode	8	0.9	0.5	0.85
Current consumption (μA) in Standby mode	1.55	0.2	0.1	0.025
Wakeup time from standby mode (μs)	8	2	<5	<10
GPIO leakage current (nA)	±50	±25	±50	±20

1.3 Goals and Structure

One of the important tasks in the technique of linear accelerators is measurement of an energy spectrum of accelerated particles. The most accurate method is the measurement with use of analyzing magnet that is installed at the output of the linear accelerator. Measurement accuracy is determined by the quality of magnet control system, the effectiveness of which depends mostly on the quality of mathematical models used in its construction. Such models should reflect the properties of the objects more comprehensively and they should be convenient for the control algorithms implementation. The lack of complete information on the object's environment, as well as dynamic characteristics and the nature of the existing noise requires application both adaptive and robust approaches in controlling such objects in order to use simplified (in particular, linear) models in regulator's synthesis.

The main methods of construction of mathematical models for technical objects are:

- empirical method: based on the statistical treatment of real data acquired during operation of the object;
- analytic method: based on the use of physics and chemistry laws;
- combined method: based on integration of the rational planning of the experiment, statistical processing of the experimental data and basic physical and chemical laws;
- automatic construction of mathematical models: based on the application of digital calculator connected to the object through sensors and transducers.

Let us consider the solution of the dynamic identification problem related to analyzing magnet via using a specialized controller implemented on the basis of a personal computer. Such controller allows us to perform automated measurement of the energy spectrum of accelerated electron beam at the output of LU-40 linear accelerator [10].

In general, the problem of object's mathematical model developing is in choosing the model structure and evaluating its parameters in the way that, for a function reflecting difference between calculated and experimental data, application of the function minimum criterion should meet the requirement of the model to be quite rigorous to the process under survey.

In accordance with existing a priori data regarding analyzing magnet as a control object, it has been classified as a one-dimensional object with a self-leveling. The most effective parameters determining method of magnet model is the active experiment. To obtain the transfer function based on experiment data it is convenient to use a modified area method of Simoyu.

Based on results of analysis presented in previous section, MSP430 microcontrollers' family [1–4], a leader in the segment of low-power microcontrollers, was chosen. The combination of modern CPU and "smart" functional peripheries allows us to solve many processing tasks, in efficient way, for both analog and digital signals with minimal power consumption. Hence, for hardware implementation of control system for the linear electron accelerator magnetic elements a microcontroller of MSP430F1611, related to MSP430 family, was chosen.

Some time ago, the developers actively used MSP430F1611 microcontroller. However, after the release of a new MSP430F2/4x series, they turned to its active application in various projects and domains. The case described here was implemented on the basis of MSP430F2618. This microcontroller is pin-to-pin compatible with the original, but it has doubled performance, increased flash-memory size and additional peripherals functions.

The chapter is structured as follows. Section 2 describes the design process of control system for magnetic elements of linear electron accelerator, which is based on efficient green-microcontrollers. It describes the features of hardware and software solutions, which ensure improved energy efficiency of the control object. Section 3 is devoted to solving the problem of energy saving in heating. Here we considered the conceptual approach to the creation of the network temperature regulators, providing heating medium flow savings.

2 Green-Microcontrollers in Industrial Applications: Magnetic Spectrometer of Linear Electron Accelerator Case Study

2.1 Magnetic Spectrometer Control System

Figure 1 shows a part of generic digital control circuit related to magnetic spectrometer. The main part of the installation is analyzing magnet, and associated current produces the electrons beam scanning by pulses.

This control system is designed to maintain generator output current at a given level corresponding to a certain reference value, defined by the operator (using control computer). This impact is carried out by changing the output current in BP5-49 power supply source.

Actual beam current value, corresponding to particles distribution density on pulses, is taken from the susceptor, which is positioned behind analyzing magnet,

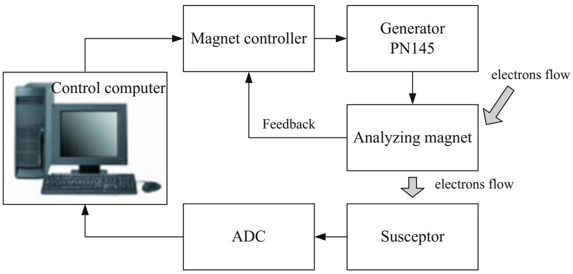


Fig. 1 The magnetic spectrometer control circuit

and is digitized by the help of eight-bit analog-to-digital converter (ADC). The data are received by control computer for further processing and archiving. Depending on the task, information about the beam current at the analyzing magnet output is provided either by a limited number of operator selectable parameters (amplitude, pulse width, time integral, etc.) or a complete waveform process. Processing of the complete impulse introduces additional features regarding analysis of the charged particles beam, in particular by using the "time of flight" method.

For the possibility of magnet remote control, operator sets the output current of the BP5-49 power supply, which is operating in a current source mode. Power supply in turn is connected to the excitation winding of PN145 generator, the anchor of which is connected to analyzing magnet. The generator is a DC generator with independent excitation. The total measurement error is determined mainly by the installation error of the excitation winding current and is equal +1.35 % at maximum magnet current (140 A) or +6.07 % at a minimum magnet current (10 A).

2.2 Proposed Upgrade for Control System

To overcome the shortcomings of the existing control system an automated control system for the LU-40 linear electron accelerator magnetic elements was developed; block diagram is shown in Fig. 2.

The basis of the magnet controller is a CPU module. Current sensor (CS) provides a measurement of the current flowing through the analyzing magnet coil. Precision resistor of 100 Ω is connected to the CS output. CS transfer ratio is equal to 1000:1. Thus, the CS output signals range (at a maximum analyzing magnet current variation from 0 to 150 A) is between +7.5 V. Depending on the switch state, CS output signal can be either positive or negative polarity. Precision

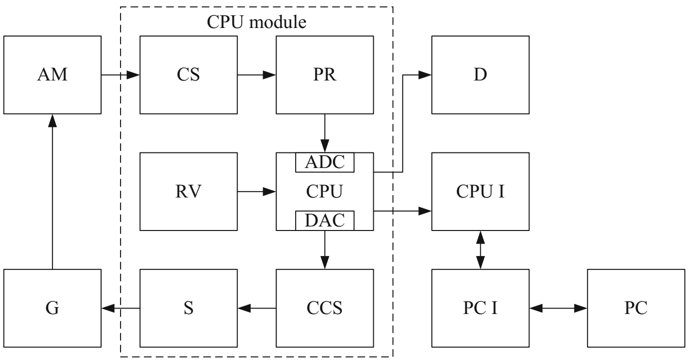


Fig. 2 Block diagram of automated control system for the LU-40 linear electron accelerator magnetic elements

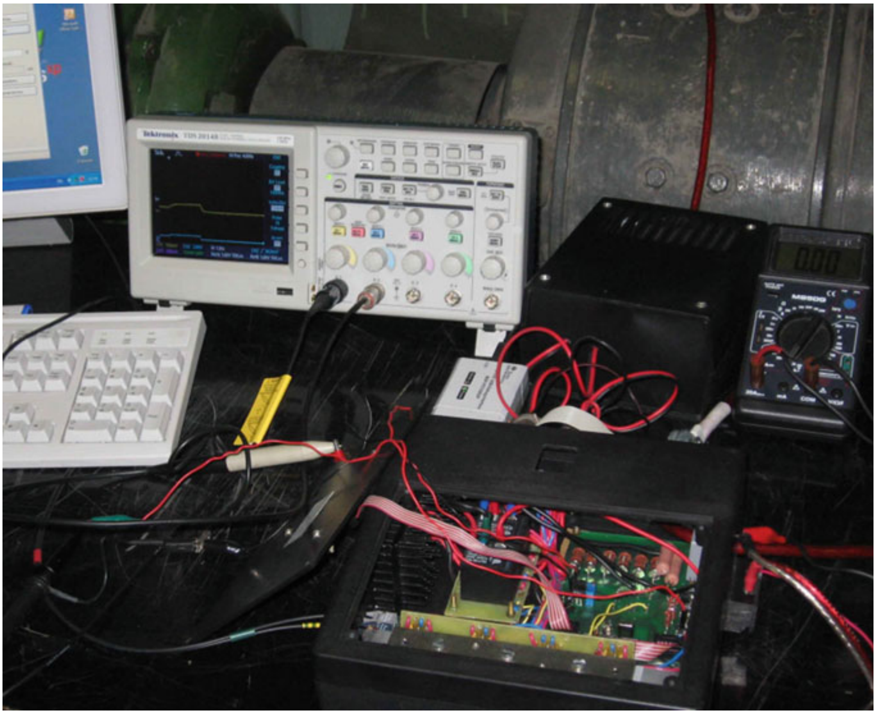


Fig. 3 Control system prototype

rectifier and scaling amplifier were applied in design to align with the valid ADC (0–2.5) input range. The rectifier output signal feeds to the ADC input and then it is processed by the controller to generate a control signal for controlled current source (CCS). The control signal is generated on the basis of proportional-integral-differential (PID) law. CCS compares the voltage at the DAC output to the voltage at sense resistor and provides a current stabilization in the excitation winding of the generator. The switch is intended to change the polarity of the current in the excitation winding of the generator and to reduce residual magnetization generator anchor.

CPU (Fig. 3) generates a reference voltage signal at the input of analyzing magnet current stabilization circuit. Voltage is generated by 12-bit DAC, so that step value of magnet current is equal to 0.1 A, that is less than 0.07 % of the maximum analyzing magnet current. The interface of the device is implemented using MSP430F2618 microcontroller, and main objectives are the formation of the control signals for the DAC, servicing local control and communication with the computer. Bidirectional communications with computer are carried over fiber-optic cable in a serial format using simplified RS232C protocol. The use of optical communication channel provides galvanic isolation between the controller and communication line that is required by the condition of ensuring the appropriate safety level and noise immunity when setting the reference voltage. Magnet electrical power supply is provided by a DC generator (PN145) with independent excitation.

In order to stabilize and adjust the magnet current, the excitation winding of the generator is connected to a controlled current source, the current in the excitation winding of the generator is given by the magnet controller. Controlled current source is based on a linear voltage source and a linear current source.

2.3 The Methodology of the Experiment

Tektronix TDS2014B digital storage oscilloscope was used to determine the transfer characteristics of the object; first oscilloscope channel was connected to the ADC input (analyzing magnet current measurement input) and the second channel—to the DAC output (control action, which sets the current in the generator excitation winding). The full current range (0 … 600 mA) in the excitation winding was divided into 12 intervals with 50 mA step, within which the controlled object can be considered as linear. Voltage oscillograms at these points are shown in Fig. 4 (the set point is changed from 200 to 250 mA).

The measurement results were used as input data for calculating the dynamic model parameters of the controlled object by the modified areas method of Simoyu [10]. Such method was used for active experiment data processing at different intervals within current variation range in the excitation winding of the generator. The result is a normalized transfer function of the following form:

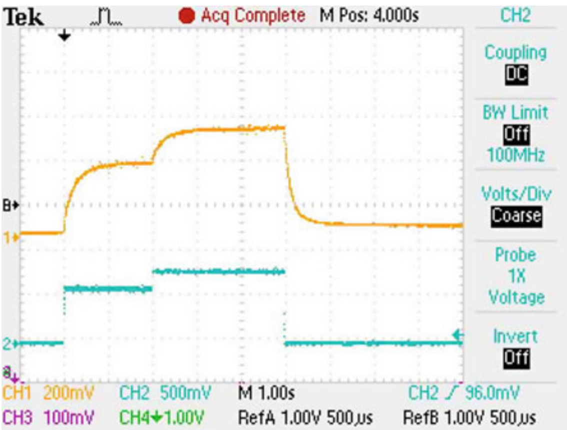


Fig. 4 Voltage oscillograms: *1*—response; *2*—set point

$$\overline{W}_M(s) = \frac{1 + b_1 s + b_2 s^2 + \ldots + b_m s^m}{1 + a_1 s + a_2 s^2 + \ldots + a_n s^n}. \quad (1)$$

The best approximation of dynamic properties of the identified object is achieved with the following values of coefficients: $a_0 = 1$, $a_1 = 2.764$, $a_2 = 2.136$; $b_0 = 1, b_1 = -0.152$.

The results of magnetic spectrometer modes research indicated quasistationarity of its dynamic characteristics. Possibility to change identified object characteristics in time can be explained due to the presence of uncontrollable external and internal disturbances, which influence the automatic measurement system. In this regard it seems appropriate to use the adaptation principles in the digital magnetic spectrometer control circuit. Quasistationary spectrometer properties make it possible to use an adaptive Bayesian identifier in digital regulator synthesis. To do this, an auxiliary data analysis procedure is used to assess the need of current analyzing magnet transfer function coefficients correction. As an initial magnet model authors used the obtained transfer function of the following form:

$$\overline{W}_M(s) = \frac{1 - 0.152s}{1 + 2.764s + 2.136s^2}.$$

2.4 Control System Schematic Diagram

MSP430F2618 green-microcontroller was selected as a basis for system implementation. Built-in modules of this microcontroller contain a lot of features allowing to implement many system functions without using CPU core that leads to reduction in energy consumption. So, initiation of ADC-conversion is performed by the «Timer B», the conversion results are stored using the DMA controller and interrupt handling is called only after the accumulation of a complete data set.

Current sensor S1 and the repeater connection diagram are shown in Fig. 5. MSP430F2618 microcontroller continuously monitors the spectrometer magnet current by reading the signal at the current sensor (S1 type CSNF661, Honeywell) output.

The current flowing through the controlled conductor generates a proportional magnetic field which is concentrated within the annular magnetic circuit and influences on a specialized linear Hall sensor. The sensor signal is amplified by DC amplifier, the load of which is a negative feedback coil. The coil creates in the magnetic circuit the opposite direction magnetic field, which completely compensates the original ones. The sensor output is the second terminal of the coil. Thus, the output signal is a current, which is proportional to the current in the monitored conductor and the number of feedback coil turns. DA2 chip in Fig. 5 is a non-inverting amplifier with a unity gain (repeater).

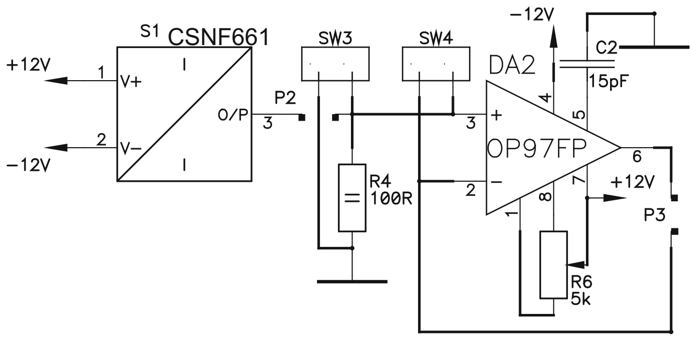


Fig. 5 Current sensor and repeater connection diagram

Precision rectifier is implemented using operational amplifiers (DA3, DA4), resistors (R8, R10, R11, R13, R14, R17, R19, R20) and diodes (VD3, VD4) and it is shown in Fig. 6. The operational amplifiers are used due to a necessity of low-voltage AC signals rectification. Diode VD3 performs a rectifier function and it is included into an amplifier feedback loop. This leads to the fact that the voltage acting on the circuit input is equal to an effective potential difference applied to the diode in the forward direction, divided by the amplifier gain without feedback. Thus, the rectifier is able to function properly with input voltages less than 1 mV. The second diode, VD4, and the associated resistors are required to provide

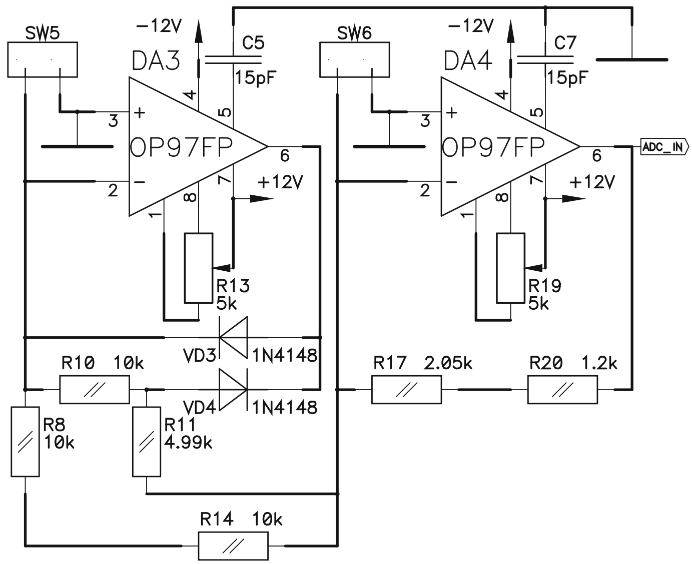


Fig. 6 Precision rectifier

feedback at the positive half cycle time of the input signal, thus avoiding overloading the amplifier.

The rectifier's output signal is digitized using 12-bit ADCs. As in a case of high-resolution ADC, appropriate printed circuit board layout and grounding techniques should be followed to eliminate ground loops, unwanted parasitic effects and noise. Ground loops are formed when return current from the A/D flows through paths that are common with other analog or digital circuitry. If care is not taken, this current can generate small unwanted offset voltages that can add to or subtract from the reference or input voltages of the A/D converter.

In addition to grounding, ripple and noise spikes on the power supply lines due to digital switching or switching power supplies can corrupt the conversion result. A noise-free design using separate analog and digital ground planes with a single-point connection is strongly recommended to achieve high accuracy [4].

16-bit MSP430F2618 microcontroller provides control for the current source, which is fed to the generator excitation winding. For this purpose, DAC_OUT signal is generated at the DAC output. The current value in the excitation winding is determined by the following relation:

$$I = \frac{U_{DAC}}{3.95}. \tag{2}$$

Controlled current source (CCS) schematic diagram is presented in Fig. 7. 2N3055 npn-transistor was chosen as a power transistor (VT4). To increase the overall gain it was chosen Shiklai scheme based on transistors, and power transistor

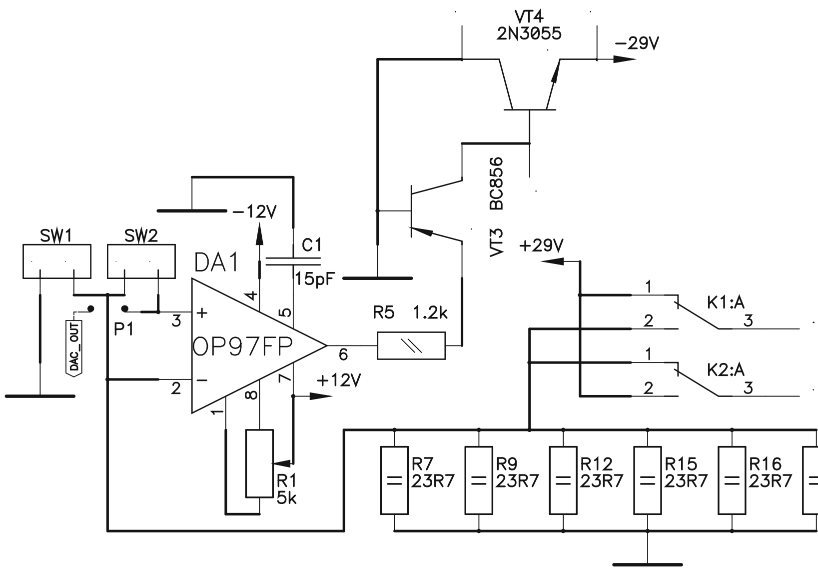


Fig. 7 Controlled current source

is controlled by the pnp-transistor (VT3) BC856B. Measuring resistor formed by parallel connection of the six precision resistors of 23.7 Ω + 1 % and maximal dissipated power of 2 W. Such connection allows to reduce temperature influence on resistors resistance, because the maximal dissipated power of each resistor is less than 0.24 W (which is 8 times less than the maximum allowable). The measuring resistor resistance is equal to 3.95 Ω.

OP97FP precision operational amplifier is used in the CCS circuit for providing high stability of the output current. Due to negative feedback at the inverting amplifier input, the voltage U_{DAC} is maintained and the load current is equal to $I = U_{DAC}/3.95$. Adjustable resistor R1 is used to eliminate the bias voltage. When the inverting and noninverting inputs are connected to the ground wire, the resistor R1 is used to set zero voltage at the amplifier output.

To verify circuit performance, the simulation of controlled current source, precision rectifier and scaling amplifier was carried out in OrCAD environment. As a test input signal a sine wave was used to verify circuit performance for precision rectifier and a trapezoidal signal was used to verify the controlled current source.

Communications between microcontroller and computer takes place asynchronously through a serial port at 115,200 baud speed. It uses USART1 serial communication interface operating in asynchronous UART mode. Plastic optical fiber (POF) is used as the transmission medium. This choice was driven by necessity to ensure transmission data reliability under noisy electromagnetic environment and to provide galvanic isolation between computer and magnet controller.

HFBR1522 optical transmitter and HFBR2522 receiver, which provide a maximum transfer rate of up to 1 MBd at link distance up to 45 m, were used as transceivers. Connection diagram of transceivers is shown in Fig. 8. To eliminate the distortion pulse width, which is caused by unequal propagation delays (tPHL) and (tPLH) of optical transmitter, a “pre-correction” circuit is applied using a capacitor C3.

2.5 Magnetic Spectrometer Control System Software

The system software includes “Accelerator” application running under Windows and application for MSP430. “Accelerator” was developed in Builder 6 integrated development environment using C++ language, and the MSP430 application was designed using C language in IAR Embedded Workbench integrated development environment for MSP430.

During the microcontroller software development, a modular approach was used to separate the functional parts of the program into blocks (header files and source files).

The main tasks assigned to the system software include the following:

– voltage measurement, data filtering (averaging);
– generator excitation winding current stabilization;

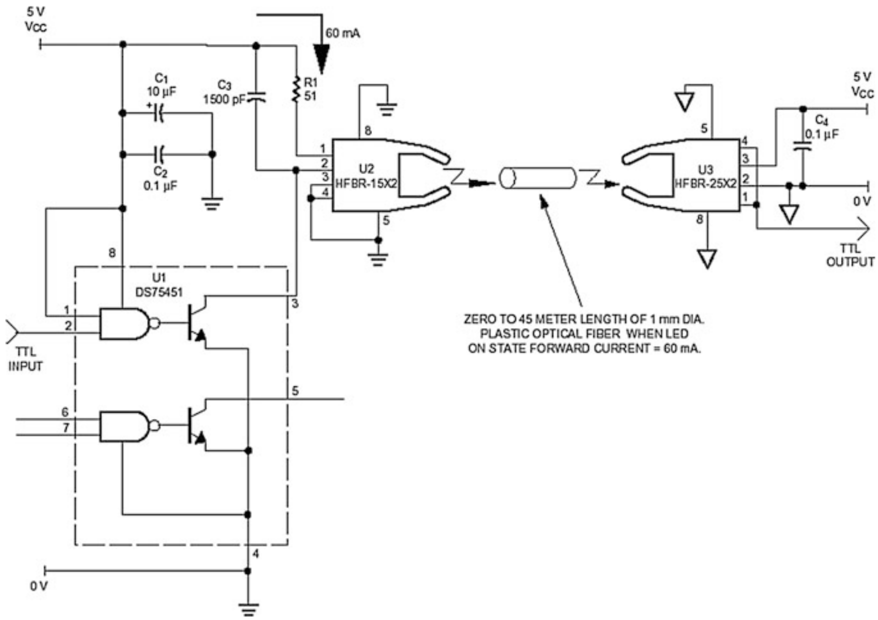


Fig. 8 Optical communication channel

– analyzing magnet current stabilization;
– communication protocol support.

The system operates using command handler principle: if the command from the computer is not received, the controller waits for it, and after receiving the command, controller processes it and then returns to its predefined state.

Microcontroller control application starts with:

– initialization of system resources—stop watchdog, initialize DAC and setting its output voltage to zero, initialize the I/O, system clock, ADC, timers A and B, UART and register variables;
– interrupts enabling and the subsequent transition to the main program loop.
– The body of the application is an infinite loop. The system can operate in the following modes:
– operating mode—magnet current stabilization mode;
– test mode—object parameters measurement mode for building its model;
– controlled microcontroller emulation mode—to debug the software.

The principle of the input signal ADC sampling is as follows: an analog-to-digital conversion is initiated with a rising edge of the sample hold input (SHI) signal. As a source for SHI the Timer_B Output Unit 1 is selected. Timer_B is configured to generate interrupt each 2 ms. The Set/Reset mode is selected for Timer_B Output Unit 1. The OUT1 signal is set when the timer counts to the

TBCL1 value. It is resetted when the timer counts to the TBCL0 value. At the end of ADC conversion cycle, the data from the internal ADC conversion memory registers using DMA controller are transferred into RAM. At DMA interrupt handler it is organized ring buffer for storing ADC measurement results (ten ADC samples are used to calculate the average current in the program). The magnet current is calculated after ten DMA transfers.

The feature of such system software structure is that all activities related to the magnet current measurement are performed without the use of a CPU core. Measurements are carried out on the basis of smart peripherals interaction, and only after accumulation of ten measurement results the processor core is being activated for calculations. This approach has provided an additional reduction of device power consumption.

System implementation and testing have led to the improvement of the accelerator parameter settings quality that resulted in significant reduction (8–12 %) of energy consumption. Testing of control system described here was carried out on the basis of the Science and Research Establishment “Accelerator” in National Science Center “Kharkov Institute of Physics and Technology”.

3 Green-Microcontrollers in the Energy Saving Tasks: Temperature Regulator and Pumps Control Device in Heating Systems

The trend of energy saving is increasingly important nowadays, and energy efficiency is defined as one of the innovative development priorities of the industrial countries’ economics. The objective of energy saving and energy efficiency enhancement programs is to ensure the sustainable process for improving energy consumption efficiency in various economic sectors.

Another important problem is in more efficient use of heat energy in heating and hot water supply for distributed buildings complexes [11].

Implementation of appropriate project, as a whole, provides the following benefits:

- increased reliability of heating;
- optimization for operation modes of heating systems through the introduction of automatic regulation and monitor tools;
- reducing the thermal energy cost.

One of the possible solutions for such problem can be in creation of automatic temperature regulators distributed network that controls the heat medium supply.

The authors proposed an implementation of temperature regulator on the basis of high-performance MSP432P401R green-microcontroller [12] (Fig. 9).

Disadvantages of the existing solutions were taken into account during designing, and the final design included not only the temperature control function,

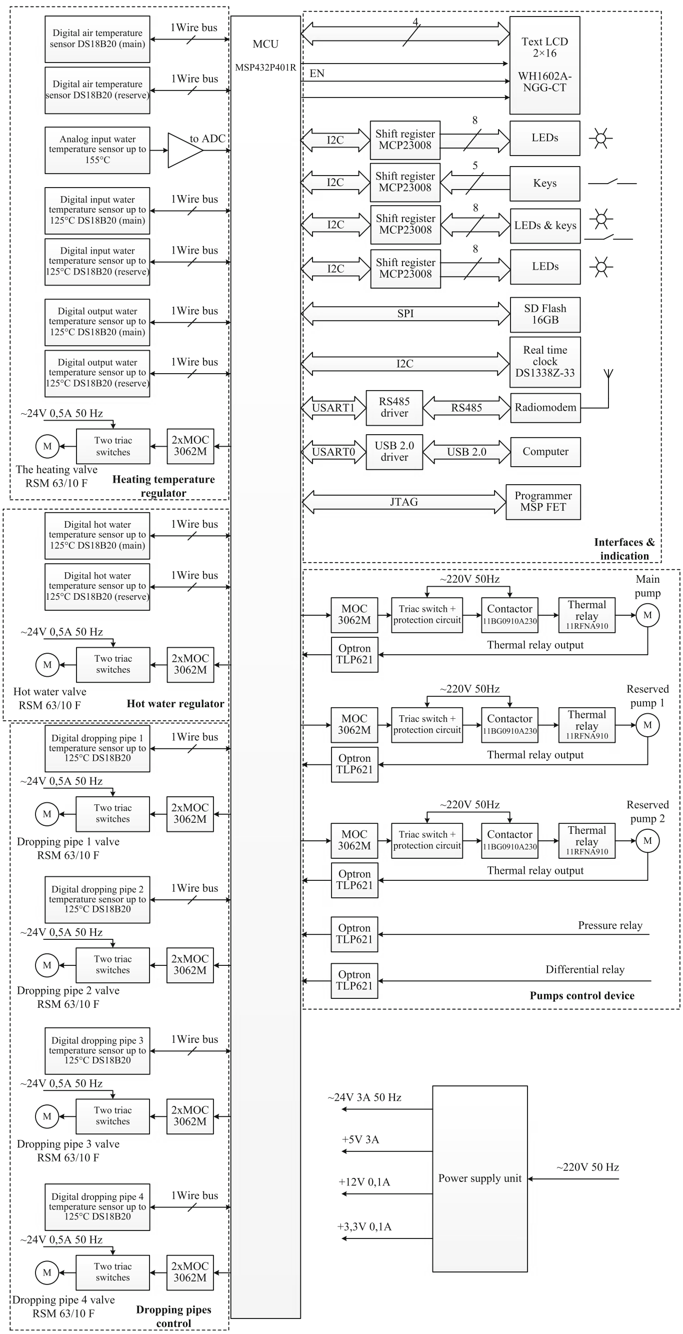


Fig. 9 The block diagram of temperature regulator and pumps control device

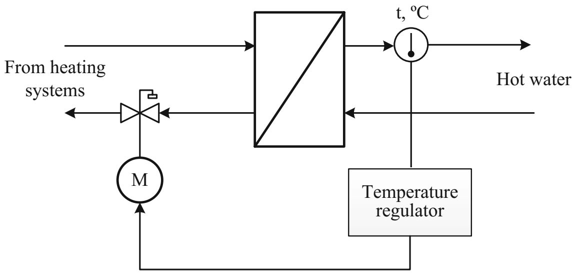


Fig. 10 Temperature regulator of hot water supply

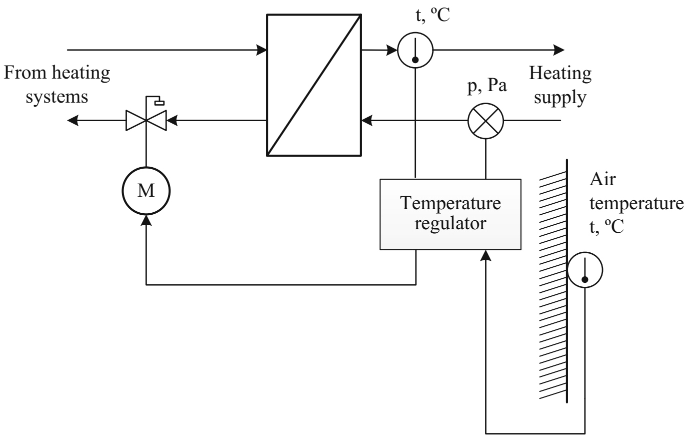


Fig. 11 Temperature heating regulator based on independent scheme

but also additional control functions of the pumping group, consisting of redundant pumps.

The proposed regulator refers to the lower level control element of the system and is intended for automatic temperature control by varying the heat medium flow rate. The electronic control unit (controller) includes scaling amplifiers, converters, microcontroller, electrical isolation and power supply elements.

According to operating principle, the control unit is an adaptive Fuzzy-regulator. Application of such regulator resulted in a qualitative regulation of heat exchange systems with different construction—starting from plate heat exchangers with quick response up to tubular sections with a diameter up to 800 mm, when the delay can be tens of minutes.

The basic temperature regulators connection diagrams are shown in Figs. 10 and 11.

The regulator consists of the following components: electronic control unit, temperature sensors and actuators with butterfly valve. Quantity of connected temperature sensors and actuators depends on the number of channels and used control loops types. Each channel includes one actuator and at least one temperature

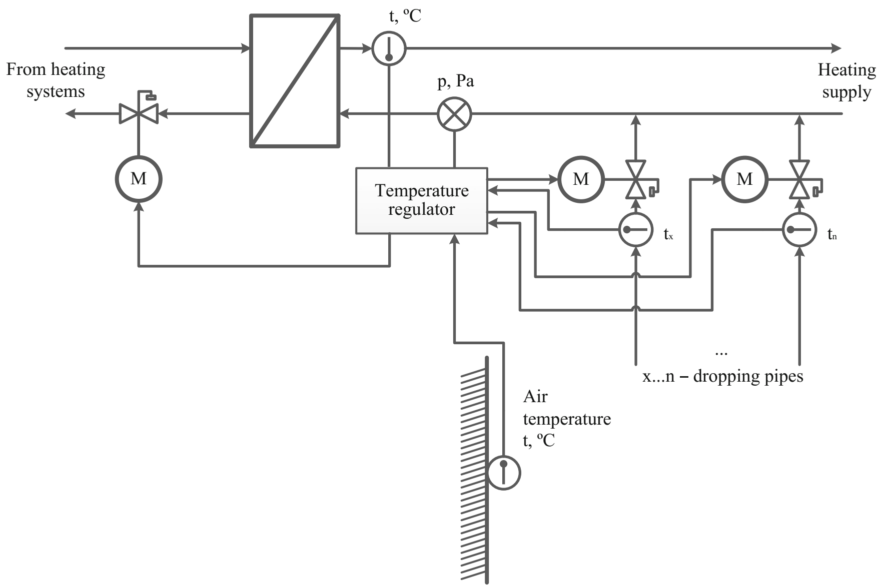


Fig. 12 Temperature heating regulator considering the temperature at dropping pipes

sensor in heat supply pipeline. Each of the channels, which envisages regulation takes into account the temperature at dropping pipe (Fig. 12), include one more sensor for heat carrier temperature measuring at dropping pipe. In addition, functionality can be extended by adding a temperature correction function depending on air temperature; in such case the electronic control unit should also be connected to the outside air temperature sensor.

Connection of pressure sensors to the electronic control unit allows, if secondary heat carrier circuit is not full filled (e.g., as a result of crash in the heat circuit or the termination of heat supply), automatically close the shut-off valve independently of temperature sensor data and thereby protect the heat exchanger from overheating.

Single-turn electric actuator with a maximum shaft rotation angle at 90° is used in the regulator. Angles of shaft drive end positions can be controlled over a wide range.

Electric drive is manually controlled. To switch to a manual mode, motor coupling clutch and gear is used. When coupling is disconnected, the shaft turns by hand-controlled wheel. Coupling also allows turnimg off the regulator without affecting the electrical switching elements. AC 24 V motors are used in electrical drives to safely use regulator in wet and unserved indoors, such as houses basements and individual heat points with no permanent staff, where the flooding probability is significant and there are strong requirements concerning electrical safety.

Currently, developed regulator is under bench testing at heat point of Belgorod State National Research University. Pilot operation has confirmed the effectiveness of the proposed solutions. Thus, for the hot water supply, it was achieved heat energy saving about 25 %, and for the heating system—up to 10 %. At the same time, developed regulator can be used in heating and hot water supply systems, in ventilation and air-conditioning systems, in technological processes with liquid or gaseous heating medium (heating water, steam, refrigerant gas, liquid fuel, etc.).

4 Conclusions

This chapter presents the results of the conducted analysis of the market of modern high-performance green-microcontrollers for embedded systems. Their main characteristics were considered and microcontrollers' market trends were analyzed.

It was shown the efficiency of green-microcontrollers application in industrial applications via detailed description of two cases. First of them represents control system for the linear electron accelerator magnetic elements based on MSP430F2618 mixed signal microcontroller, and second case represents temperature regulator and pumps control device in heating systems based on MSP432P401R ARM-microcontroller. In both cases, the reasonable use of green-microcontrollers has led to a significant reduction in system energy consumption up to 12 % in the first case and up to 25 % in the second case.

Further development of the control system for the linear electron accelerator magnetic elements is aimed at software functionality increasing (it is planned to implement an adaptive fuzzy-regulator in magnet current control circuit). Next development steps for temperature regulators may be related to implementation their network version and design of monitoring systems on their basis.

References

1. Harres, D.: MSP430-based Robot Applications: A Guide to Developing Embedded Systems. Elsevier, Waltham, USA (2013)
2. Nisarga, B., Quiring, K., Plant, M.: Designing with ultra-low power MSP430™ USB microcontrollers. White Paper. Texas Instruments. http://www.ti.com/lit/wp/slay026/slay026.pdf (2013)
3. Witt, T., Mena, R., Cornell, E., Pecze, H.: Integrating mixed signal capabilities on MSP430™ MCUs in industrial automation applications. White Paper. Texas Instruments. http://www.ti.com.cn/cn/lit/wp/slay030/slay030.pdf (2014)
4. Texas Instruments: MSP430x2xx Family User's Guide. http://www.ti.com/lit/ug/slau144j/slau144j.pdf (2013)
5. Predko, M.: Programming and customizing the ARM7 Microcontroller. McGraw-Hill/TAB Electronics (2011)
6. Yiu, J.: The Definitive Guide to ARM® Cortex®-M3 and Cortex®-M4 Processors. Elsevier, Kidlington, UK (2013)
7. Brown G.: Discovering the STM32 Microcontroller. Indiana University (2012)
8. Accurate power consumption estimation for STM32L1 series of ultra-low-power microcontrollers. Technical article TA0342. http://www.st.com/st-web-ui/static/active/jp/resource/technical/document/technical_article/DM00024152.pdf (2013)

9. Levy M.: Understanding the real energy consumption of embedded microcontrollers. http://www.digikey.com/en-US/articles/techzone/2012/jun/understanding-the-real-energy-consumption-of-embedded-microcontrollers (2012)
10. Dovbnya, A., Udovenko, S., Shamraev, A.: Defining magnetic spectrometer model parameters on experimental data base. J. Radio Electron. Comput. Sci. Control **1**(19), 45–49 (2008)
11. Boldyryev, S., Garev, A., Ilunin, O., Shamraev, A., Selyakov, O., Leshchenko, O., Kapustenko, P.: Modified criterion for economic efficiency estimation of heat pumps. J. Chem. Eng. Trans. **35**, 475–480 (2013)
12. Texas Instruments: MSP432P401x mixed-signal microcontrollers. http://www.ti.com/lit/ds/symlink/msp432p401r.pdf (2015)