



ISSN 1310-8271

JOURNAL

**OF THE TECHNICAL UNIVERSITY - SOFIA
PLOVDIV BRANCH, BULGARIA**

**FUNDAMENTAL SCIENCES
AND
APPLICATIONS**

Volume 22, 2016

**Journal of the Technical University - Sofia
Plovdiv branch, Bulgaria
“Fundamental Sciences and Applications”**

EDITORIAL BOARD

EDITOR-in-chief

Prof. Michail Petrov, PhD

Honorary EDITOR-in-chief

Prof. Marin Nenchev, DSc Eng., DSc Phys.

Journal Editorial Board

Prof. Sonia Tabakova, PhD

Prof. Valyo Nikolov, PhD

Prof. Teofil Iamboliev, PhD

Prof. Andon Topalov, PhD

Prof. Veselka Boeva, PhD

Prof. Galidia Petrova, PhD

Prof. Grisha Spasov, PhD

Prof. Angel Zumbilev, PhD

Prof. Dobrin Seizinski, PhD

Programme Editorial Board

Prof. Venelin Zhivkov, DSc

Prof. Georgi Andreev, DSc

Prof. Georgi Totkov, DSc

Prof. Emil Nikolov, DSc

Prof. Ivan Iachev, DSc

Prof. Marin Hristov, PhD

Prof. Ognian Nakov, PhD

Prof. Marc Himbert DSc

Prof. Tinko Eftimov DSc

Acad. Yuriy Kuznietsov DSc

Journal Scientific Secretary: Sevil Ahmed, PhD



CONTENTS

1. Witold Pawlowski, Jakub Sikorski and Lukasz Kaczmarek -	6
Space Applications of the Lightweight Gradient Materials Based on nc-Al Alloys	
2. Beladel Abdelkader, Kouzou Abdellah, Ahmed Hafaifa, Djilali Mahi	10
Dual Matrix Converter Feeding an Open-End Load Based on Modified Direct Transfer Function Approach	
3. Nadji Hadroug, Ahmed Hafaifa, Abdellah Kouzou, Mouloud Guemana, Ahmed Chaibet.....	17
Control Of Nonlinear Gas Turbine System using Strategies of Multidisciplinary Design Optimization	
4. Bachir Nail, Abdellah Kouzou, Ahmed Hafaifa, Belkacem Bekhiti	25
Stabilizing Linear Multivariable Gas Turbine Model via Sliding Mode Control	
5. Ahmed Hafaifa, Kouzou Abdellah, Guemana Mouloud, Nadji Hadroug	31
Reliability Analysis using Weibull Distribution Applied to a Booster Pump used in Oil Drilling Installations	
6. Nina Dobrinkova.....	38
ICT Technologies in Wildland Fire Modelling	
7. Georgi Ruzhekov, Teofana Puleva, Tsonio Slavov.....	44
Experimental System for Wind Modelling	
8. Kaddouri Ameer Miloud, Kouzou Abdellah, Hafaifa Ahmed	50
Performances Study of Multi-Layer Linear Gradient Index Antireflective Based On Oxynitride Silicon	
9. Dimitar Dimitrov, Atanas Iliev, Nevenka Kiteva - Rogleva	54
Validation of a Model For Generation of Hourly Ambient Temperature Time Series	
10. Nikola Georgiev.....	60
A Model of a Single-Phase Two-Rotor Axial Generator With Rare Earth Magnets	
11. Nikola Georgiev.....	65
Calculating Magnetic Induction in The Air Gap of a Three-Phase Two-Rotor Axial Generator with Rare Earth Magnets	
12. Vanya Plachkova.....	70
Fluorescent Spectroscopy Applications in Veterinary Medicine and Healthcare	

13. Svetoslav Ivanov, Yanka Ivanova	76
Electronic Converter for Fiber Optic Thermometer	
14. Margarita Deneva	80
Nd:YAG Laser in Two-Wavelength Coaxial Architecture	
15. Margarita Deneva, Marin Nenchev, Nadejda Kaymakanova, Ernst Wintner, Mark Himbert	86
Light Based Techniques as Possibility to Solve the Pollution Problems in Laser Ignition Combustion of Gaseous Mater. A Pilot Study	
16. Margarita Deneva, Marin Nenchev, Valko Kazakov, Nadejda Kaymakanova	92
Development of the Thermo-Sensitive Paper Application as Method for Laser Beam Spot Study	
17. Petar Mitsev, Angel Tanev, Tsvetelina Lazarova	96
Performance Sensing Solutions for a Green Environment Designed in Bulgaria	
18. Veneta Aleksieva, Ivaylo Zhelyazkov	101
Generator of Network DoS Attacks	
19. Simeon Tsvetanov	107
Developing of a Libcloud Driver for OwnCloud	
20. Ivan Kanev, Petya Pavlova, Dimitre Kromichev, Jordan Genoff	111
Generating 2D Gaussian Kernels for the Purpose of Studying FPGA-Based Linear Spatial Filters	
21. Lubomir Bogdanov	119
Look-Up Table-Based Microprocessor Energy Model	
22. Dobrinka Petrova	125
Big Data in Bioinformatics	
23. Hristo Valchanov	131
Communication System for LAN Searching	
24. Dimitre Kromichev	135
Technology of Computations in Canny Edge Detection for FPGA Implementation	
25. Nikolay Petkov, Totka Bakalova, Tetiana Cholakova, Lukáš Voleský, Petr Louda	141
Mechanical and Tribological Properties of AlN Hard Films Deposited by Reactive Magnetron Sputtering	
26. Nikolay Petkov, Totka Bakalova	145
Deposition of AlCN Thin Films by Reactive DC Magnetron Sputtering and Tribology	
27. Ilia Popov, Pepo Yordanov, Sabi Sabev	149
Experimental Determining of The Elasticity Module of Polymer Concrete Composites	

28. Krasimir Ambarev	155
An Experimental Study of The Influence of Advance Processor on Some Operating Performance Indicators of Spark Ignition Engine Working with LPG	
29. Atanaska Georgieva, Albena Pavlova	161
Lp (K) - Solution of Nonlinear Volterra-Hammerstein Integral Equations	
30. Ivan Badev, Georgi Zlatanov	167
Odd-Dimensional Spaces with Symmetric Affine Connection	
31. Ivan Badev, Marta Teofilova, Georgi Zlatanov	171
Cartesian Generalized Compositions in Spaces with a Symmetric Affine Connection	
32. Delyana Dimova	177
An Algorithm for Determining and Comparing the Levels of Economic Development in the Districts in Bulgaria	
33. Iliycho Iliev	183
Parametric Regression Models for CuBr Laser	
34. Toni Mihova, Valentina Nikolova - Alexieva	187
Principles and instruments of LEAN Methodology	
35. Thomas Moser	192
Causes, Prevention and Correction of Project Failures and Budget Overruns	
36. Thomas Moser	198
Will Agile Methods Resolve the Complexity Crisis in Software Engineering?	



SPACE APPLICATIONS OF THE LIGHTWEIGHT GRADIENT MATERIALS BASED ON NC-AL ALLOYS

WITOLD PAWLOWSKI, JAKUB SIKORSKI, LUKASZ KACZMAREK

Abstract: *In the paper the latest achievements of the lightweight gradient materials based on nc-Al alloys were utilized in order to create manipulators applied in cryogenic conditions of the outer space. The properties of the modern material based on nc-Al alloys are discussed. The planetary, compact, 3-stage gearboxes made of modern structural materials are presented. The 5 degree-of-freedom (DOF) space manipulator arm with rotary joints in which the gearboxes are applied is also proposed. The advantages of the new lightweight materials in the construction of the space mechanisms are outlined.*

Key words: *light-weight materials, cryogenic conditions, space manipulators*

1. Introduction

Extremely high cost of launching unit weight into space (around 10k\$/kg) is a driving force for the development of new lightweight structural materials (LwSM) for space vehicles which can sustain harsh cryogenic conditions. The aim of the research project being investigated is to develop LwSM based on nc-Al with deposited gradient C-coating with high wear and contact fatigue resistance under cryogenic conditions for space applications (e.g. gear boxes in extra-terrestrial vehicles).

This is highly motivated by the economic and technological requirements including cost effectiveness. According to the Russian Federal Space Agency, the cost of sending 1kg into Earth orbit is around 10k\$ and it dramatically increases when the load has to be delivered to another planet within the solar system. The reduction of spacecraft weight is among the most cost effective ways and it requires using of more Lightweight Materials (LwM) which exhibit appropriate properties even in the cryogenic conditions of deep space. Obviously, the use of steel as a structural material for the Unmanned Space Vehicles (USV) is highly uneconomical. Large number of extensive studies on the existing and new LwM is carried out to replace steel and to meet requirements for their mechanical and tribological properties under those conditions while having low specific weight.

2. Properties of the nc-Al alloys

For a number of years, a trend to lighten the weight of products has dominated global industry. This applies to electronic components used in the production of mobile phones, tablets but also to devices such as tanks and aircraft [1 – 5]. The drive to lighten the weight of vehicles is the sum of a number of different needs. Decreasing fuel consumption (improving fuel economy) and enhancing the maximum range of vehicles are of major importance. It would be also extremely profitable for the military industry where increased range and maneuverability could be critical. Alloys of lightweight metals such as aluminum, titanium, magnesium or composites based on them are used in order to make products lighter [6 – 7], and of those, aluminum and its alloys have become an especially popular choice in many industries [8 – 11]. Their strength-to-density ratio is their major advantage and therefore they are more and more frequently used in structural components where component weight is very significant. Nevertheless, their mechanical properties in most cases are insufficient to render them a reliable substitute for functional steel components e.g. gearwheels.

To improve mechanical properties of aluminum based alloys, heat treatment parameters are optimized and combinations of thermal and mechanical treatments are used [12 – 16]. The combination of shot peening and two-stage age hardening appears to be very promising as not only

does it lead to the hardening of the material and increased strength of the surface layer (due to the presence of core-shell type of phases in the continuous phase of some aluminum alloys) but it also introduces advantageous compressive stresses. Aluminum alloys subject to the combined treatments exhibit mechanical properties similar to those of steel, which significantly expands the range of their applications to include critical structural components as well as subcomponents working under friction conditions e.g. gears in unmanned vehicles. The synergistic effect of shot peening and two-stage age hardening has caused a surge in the number of studies on commercial applications of heat treated aluminum alloys. That is also the reason why the main objective of the foregoing study was to investigate the applicability of 7075 aluminum alloy treated with two-stage thermo-mechanical treatment to lightweight transmission gears. Such transmissions could be used in drones or military robots.

The novel core/shell precipitation hardened nc-Al alloys are among the most promising candidates with a number of advantages compared to the current Ti- and Mg-based alloys. The paper investigates the application possibility of these alloys in combination with carbon based coatings, which can potentially solve the main drawback of these materials insufficient wear and contact fatigue resistance under cryogenic conditions in space.

Unique properties of this material (low friction coefficient, higher yield strength and elongation with additional increase in fatigue resistance under cryogenic conditions) will allow its application for e.g., drive transmissions system with the reduce weight in space vehicles for planets exploration. Weight reduction subsequently contributes to the cost reduction of space transportation. This fully complies with the objectives of Russian Federal Space Agency and with NASA expedition program initiatives for the period 2014-2022 which are focused on the construction and transport of light mechanical assemblies for manipulators.

Besides space applications, the newly developed Al-based materials can be widely applied in general and even special aviation, e.g. for the construction of the advanced ultra-light drones, thus having wider societal impact.

3. Space construction requirements

The cost of launching unit weight of space vehicles into Earth orbit or even to the other planet is prohibitively high and there is a strong economic and societal pressure to reduce that cost. The

reduction of spacecraft weight is among the most effective approaches and it is limited by the availability of more lightweight structural materials which exhibit sufficient properties even in the harsh cryogenic conditions in the space. Rapidly growing aerospace industry permanently looks for lightweight functional and structural materials. The properties of that material should comprise good resistance to stress in cryogenic conditions as well as high fatigue strength in order to sustain hostile outer space conditions without being prone to defects. Thermal expansion coefficient of that material should also be as low as possible not to induce any tensions or strains between cooperating mechanical elements. What is also important, the material should be free of internal defects such as air bubbles or pores, which should be avoided or eliminated during technological process of molding or casting. Those defects can be of high negative influence in high vacuum and low temperature in outer space conditions.

4. Planetary gear transmission system

The first attempt has been undertaken to construct the planetary transmission system which could be successfully applied in the robotic space arm. The requirements for low mass and good cryogenic behavior properties induced the need of introducing the optimization procedures into the design process. The planetary gearbox design prerequisites comprised the following requirements:

- compact construction,
- minimum input power: 800 W,
- maximum output rotational speed: 16 rpm,
- minimum output torque: 400 Nm.

Having selected the appropriate electrical motor which could withstand the outer space conditions as well as could drive the construction of the manipulator, the following parameters could be additionally specified and assumed to be the prerequisites in the planetary gearbox design process:

- input power: 900W,
- nominal rotational speed: 4600 rpm,
- torque: 1.87 Nm

The resultant overall gear ratio which had to be achieved was minimum 287.5 (reduction). The assumed ratio was 280. Value of the ratio is high enough to plan the planetary transmission as the three-stage gearbox with the ratios of the subsequent stages respectively: 8x6x6.

The planetary gearbox which meets all these requirements has been constructed and it is presented in Fig. 1.

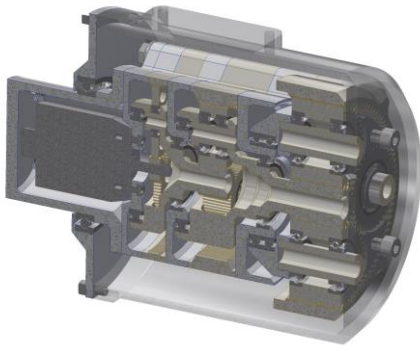


Fig. 1. The planetary transmission system

The exploded view of the planetary gearbox construction is presented in Fig. 2.

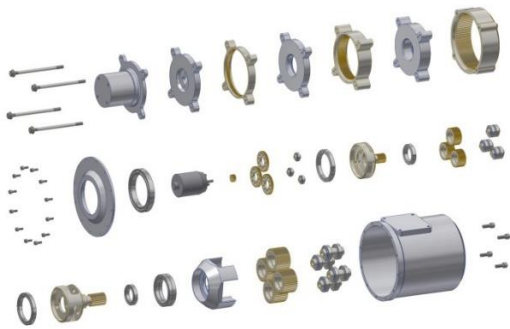


Fig. 2. The exploded view of the planetary gearbox

In the first version of the construction the structural material which was applied was steel. The total mass of the gearbox made of steel was 11.847 kg.

In order to make these construction more lightweight the optimization process was performed. It comprised the strain and stress analysis by means of finite element method (FEM) of the whole construction (assembly). Therefore the contact definition among every cooperating elements had to be created according to the character of their mutual interaction. Having performed the analysis the total mass was minimized down to 9.754 kg.

The next attempt was to apply structural lightweight materials which could withstand the conditions both the strain and stress as well as those of the severe outer space environment. Therefore the following materials were applied:

- the casing – nc-Al alloy (2.7 g/cm^3),
- the gears, shafts, pins – Ti alloy (4.43 g/cm^3),
- bolts, nuts – stainless steel (8 g/cm^3),
- fully ceramic bearings (3.18 g/cm^3).

Additionally the ceramic lightweight bearings were applied which makes the gearbox possible to operate without any lubrication. After introducing the appropriate changes in the whole

construction the overall mass of the planetary gearbox was minimized down to 4.250 kg.

The resultant properties of the newly designed planetary gearbox are listed below:

- input power: 900W,
- nominal rotational speed: 4600 rpm,
- total gear ratio: $8 \times 6 \times 6 = 288$,
- output torque: 484 Nm,
- output rotational speed: 16.0 rpm,
- overall mass: 4.250 kg.

The result of mass optimization resulted in over 2 fold mass reduction of the mechanism, the number of which in the whole robotic arm construction is 3.

According to the assumptions and optimized design procedures, the smaller size planetary gearbox has been constructed. The parameters of the box are as follows:

- input power: 300W,
- nominal rotational speed: 8500 rpm,
- total gear ratio: $8 \times 8 \times 8 = 512$,
- output torque: 155 Nm,
- output rotational speed: 16.6 rpm,
- overall mass: 2.095 kg.

The number of the smaller size gearboxes in the robotic arm construction is also three.

In the figure 2 the number of the elements is clearly visible. The casing and the plates constitute the main part of the mechanism that influences the total mass of the gearbox. That is why application the nc-Al alloy of low density resulted in that significant mass reduction.

5. Space manipulator arm

The planetary transmission systems have been applied in the space manipulator arm. The arm is designed to operate outside the satellite or spaceship in order to perform occasional amendments of the outer coating of the ship or any mechanisms which need repairing or servicing (antennas, solar panels, telescopes etc.).

The kinematics of the robotic arm comprises serial kinematics with rotary joints. The mechanical part of the arm is based on five rotary joints. Each joint is equipped with the planetary gearbox described above: three bigger size planetary gearboxes and the following two – smaller size gearboxes. The manipulation activities is performed by the gripper of parallel jaw motion. In the construction of the gripper the parallel lever mechanism is applied. In jaw clamping mechanism the planetary gearbox of small size was applied in order to save as much mass as it is possible. That results in lowering the mass placed on the long overhang in relation to the manipulator mounting plate. The arm construction is based on the typical

profiles made of lightweight nc-Al alloy. The 3D view of the manipulator is presented in Fig. 3. Zoomed view of the gripper construction, in which the small size planetary transmission system is applied is presented in Fig. 4.

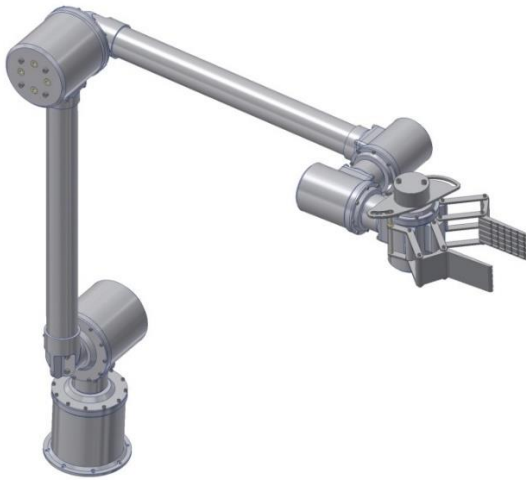


Fig. 3. The outer space robot arm manipulator

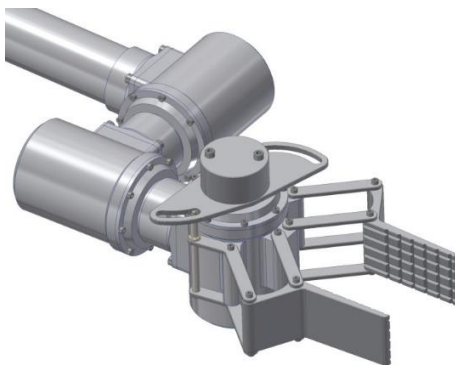


Fig. 4. The gripper construction

6. Conclusions

In the paper the space application of the construction of robot manipulator arm equipped with specially designed planetary transmission systems is presented. The high performance of the mechanism accompanies with lightweight construction made of modern structural materials e.g. nc-Al alloy. Total mass of the initial robotic arm made of steel is 55.2 kg whereas the optimized construction mass made of lightweight structural material is 23.25 kg. The optimization and appropriate material selection made the whole construction robust, efficient and lightweight. These properties are of great importance as far as the space application are being concerned. The analysis showed that reaching comparative properties the constructions made of modern structural materials can be more than twice as lightweight as the traditional design. It highly influences the level of funds that have to be found to prepare the successful space mission.

This work was supported by European Commission from ERA.Net RUS Plus programme under project: LIGHTMAT4SPACE – Lightweight nanocrystalline aluminium based material for space applications (modelling and technology verification).

REFERENCES

1. Bloch, K., Nabialek M., Pietrusiewicz P., Gondro J., Dospial M., Szota M., Gruszka K. (2012). *Acta Phys Pol A*. 126, 108.
2. Nabialek M., Szota M., Dospial M. (2012). *J Alloy Compd*. 526, 68.
3. Dospial M. J., Nabialek M. G., Szota M., Mydlarz T., Ożga K., Lesz S. (2012). *J Alloy Compd*. 536, 324,
4. Trško L., Guagliano M., Bokůvka O., Nový F. (2014). *Procedia Engineering*, 74, 246.
5. <http://www.european-aluminium.eu/>
6. Jiang F., Zhang H., Meng X., Li L. (2014). *Mater Design*, 55, 280.
7. Geng L., Huang L. (2014). *Acta Metall Sin. (Engl. Lett.)*, 27(5), 787.
8. Babalola P.O., Inegbenebor A.O., Bolu C.A., Inegbenebor A.I. (2015). *JOM*, 67(4).
9. Reilly C., Duan J., Yao L., Maijer D.M., Cockcroft S.L. (2013). *JOM*, 65(9).
10. Choi Y., Kim D. U., Kang B. Y., Park D. K., Lee D. J., Lee S., Shin H. T. (2013). *J Mech Sci Technol*. 27 (11), 3445.
11. Sakhrieh A., Al-Ghandoor A. (2013). *Energ Convers Manage*. 65, 715.
12. Kaczmarek Ł., Steglinski M., Sawicki J., Swiniarski J., Batory D., Kyzioł K., Kołodziejczyk Ł., Szymanski W. (2013). *Archives of Metallurgy and Materials*, 58 (2), 535.
13. Kreethi R., Verma P., Dutta K. (2015). *Trans Indian Inst Met*. 68(2), 229.
14. Emani S. V., Benedyk J., Nash P., Chen D. (2009). *J Mater Sci*. 44, 6384.
15. Byczkowska P., Sawicki J., Stegliński M. (2014). *Inż Mat*. 6(202), 459.
16. Farshidi M.H., Kazeminezhad M., Miyamoto H. (2013). *Mater Sci Eng A*. 580, 202.

Witold Pawlowski, PhD, DSc, TUL prof.
 Jakub Sikorski, MSc.
 Lukasz Kaczmarek, PhD, DSc, TUL prof.
 Lodz University of Technology (TUL)
 Address: 1/15 Stefanowskiego,
 90-924 Lodz, Poland
 E-mail: witold.pawlowski@p.lodz.pl
 E-mail: jakub--sikorski@wp.pl
 E-mail : lukasz.kaczmarek@p.lodz.pl

DUAL MATRIX CONVERTER FEEDING AN OPEN-END LOAD BASED ON MODIFIED DIRECT TRANSFER FUNCTION APPROACH

BELADEL ABDELKADER¹, KOUZOU ABDELLAH¹, AHMED HAFIFA¹, DJILALI MAHI²

Abstract: The present paper deals with the control of the three-phase-to-three-phase dual matrix converter feeding an open-end three-phase load. The main objective is the use of the Direct transfer function approach of Venturini to achieve the required control of the presented dual matrix converter. The both matrix converters are supplied from the same three phase power source. Whereas; the two output voltage systems of the dual matrix converter are shifted with 180 degrees. The performance and the advantages of the proposed control approach applied to the presented topology are finally discussed through simulation results. Furthermore, a short comparison between the single matrix converter close-end load topology and the dual structure applied to open-end load are also discussed.

Key words: Matrix Converter, Direct Transfer Function Approach (DTFA), Simulation Model, Open-end load, Dual matrix converter.

1. Introduction

Matrix converter (MC) and cyclo-converter are direct AC-AC power electronics converter, whereas a matrix converter is a forced commutated converter which can achieve variable amplitude and frequency at the output side. Indeed, the matrix converter presents many advantages in comparison with the classical indirect power electronics conversion that are based on two stages power conversion DC-AC and AC-AC. The main outstanding advantage of the matrix converter can be summarized as follows[1-2].

- ✓ The absence of the bulky DC-link and energy storage elements;
- ✓ sinusoidal input and output currents can be achieved;
- ✓ the easiness of the input power factor control for any load,
- ✓ Four-quadrant operation
- ✓ Simple and compact design

On the other side, due to the high integration capability of the semiconductor structures, the matrix converter topology is being recommended for several extreme and critical applications. However, the topology of the matrix converter itself has pushed the researchers to run after more favorable control strategies. In this context several modulation techniques have been developed to fulfil the requirement of the matrix converter control.

Among these techniques two main control approaches are being used, the scalar approach such as; the direct transfer function approach (DTF) proposed by Venturini [3-4], and the space vector approach such as; direct and indirect space vector modulation (DSVM and ISVM) [5-6].

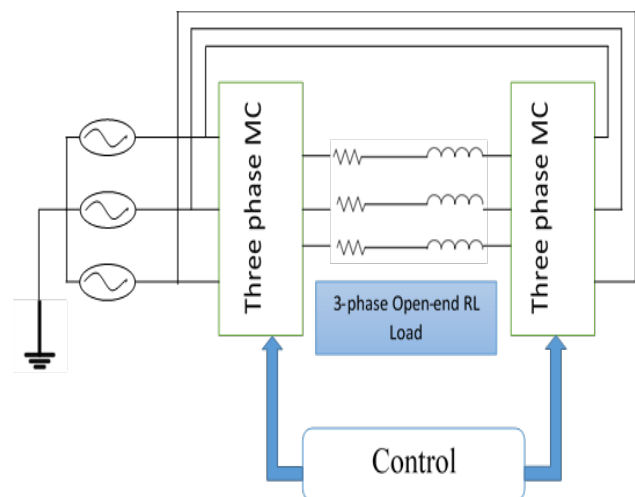


Fig. 1. Three phase Open-end Load drive system

In the present paper, the direct transfer function control approach is applied to the matrix converter for feeding an open-end three-phase load topology to overcome the main problem of the common mode voltage and to achieve the more control flexibility and the operation reliability [7-8]. The

principle of the mentioned topology is presented in Fig. 1. The both sides of the three-phase load are supplied via two three-phase to three-phase matrix converters. The two matrix converter have the same topology as it is shown in Fig. 2. The load is an open-end three-phase resistive-inductive load supplied by a dual matrix converter.

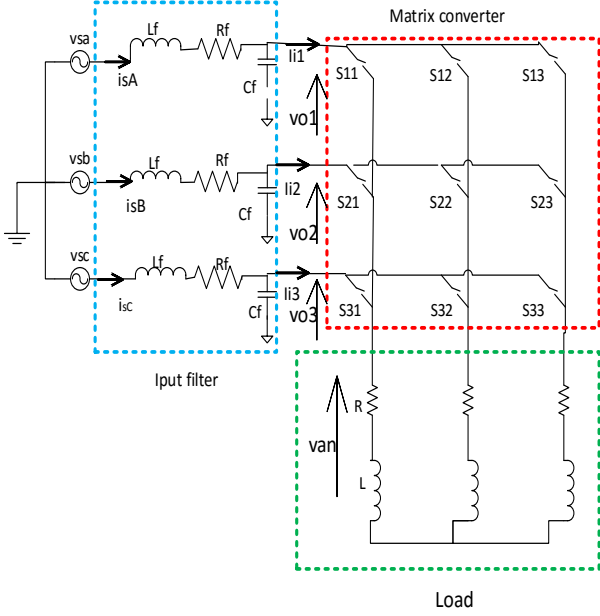


Fig. 2. Basics power circuit of Matrix converter with input filter

A comparison between the application of one matrix converter for star load topology and the application of a dual matrix converter for open-end of the same load is presented in this paper. Simulation tests were performed for the single and dual matrix converter feeding the same load, the main aim is to demonstrate the effectiveness of the dual matrix converter topology and the used control approach is such application.

2. The Matrix converter topology

The topology of a single three-phase input to three-phase output matrix converter is presented in Fig.2. Where V_{i1} , V_{i2} and V_{i3} are the input voltages and V_{o1} , V_{o2} and V_{o3} are the output voltages. Each leg has three bidirectional power switches to ensure the connection between one phase of the input and one phase of the output [9]. These switches are operating under the two main constraints, that the input should not be short-circuited and that the output side should not present an open-circuit at any instant. These are the two main important constraints imposed by the normal operation of the matrix converter to ensure its protection [10-16].

The vector of the input voltages are presenting the three phase power system network voltage that are expressed as follows:

$$\begin{bmatrix} V_{i1} \\ V_{i2} \\ V_{i3} \end{bmatrix} = V_{im} \begin{bmatrix} \cos(\omega_i t) \\ \cos(\omega_i t - 2\pi/3) \\ \cos(\omega_i t + 2\pi/3) \end{bmatrix} \quad (1)$$

The vector of the output voltages are given as follows:

$$\begin{bmatrix} V_{o1} \\ V_{o2} \\ V_{o3} \end{bmatrix} = V_{om} \begin{bmatrix} \cos(\omega_o t) \\ \cos(\omega_o t - 2\pi/3) \\ \cos(\omega_o t + 2\pi/3) \end{bmatrix} \quad (2)$$

S_{ij} represents the ideal bidirectional states of the switches, where i ($=1, 2, 3$) represents the index of the output side and j ($=1, 2, 3$) represents the index of the input side. The relation between the output voltages and the input voltages can be achieved via an intermediate matrix M , which is known as the modulation matrix, hence the relation between the input and output voltages can be expressed as follows:

$$[V_o] = [M] \cdot [V_i] \quad (3)$$

the relation between the input current $[I_i]$ and the output current $[I_o]$ can be deduced as follows :

$$[I_i] = [M]^T \cdot [I_o] \quad (4)$$

Where $[M]^T$ represents the transposed matrix of $[M]$. Equation (3) can be represented in a developed manner as follows:

$$\begin{bmatrix} V_{o1} \\ V_{o2} \\ V_{o3} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} V_{i1} \\ V_{i2} \\ V_{i3} \end{bmatrix} \quad (5)$$

where the elements of the modulation matrix presenting the modulation coefficients are expressed as follows:

$$m_{ij} = \frac{t_{ij}^k}{T_s} \quad (6)$$

During the switching process, the bidirectional switches can connect or disconnect the phase i of the input connected to the power source to/from the phase j of the output connected to the load. In general the modulation coefficient must respect the following rules [7]:

- At every instant t , only one switch S_{ij} ($i = 1, 2, 3$) is conducting in order to avoid the short-circuit between the phases.

$$\sum_{i=1,2,3} S_{ij}(t) = 1; j = \{1,2,3\} \quad \forall t \quad (7)$$

- At every instant t , at least two switches S_{ij} ($j = 1, 2, 3$) are conducting to ensure a closed loop load current.
- The switching frequency $f_s = \omega_s/2\pi$ must have a higher value than the maximum of f_i, f_o ($f_s \gg \max(f_i, f_o)$).
- During the period T_s , the sum of the conducting time which is being used to synthesize the same output phase must be equal to T_s this can be expressed by the following equation:

$$T_s = t_{1j} + t_{2j} + t_{3j} \quad \forall j \text{ with } j = 1,2,3 \quad (8)$$

3. Modified direct transfer function approach

The modified direct transfer function Approach was proposed in [1-3] and used in several previous research papers [11-14] to generate the appropriate firing pulses to the used bidirectional switches S_{ij} of the 3x3 matrix converter. Whereas, The main aim is to ensure the ability of producing a three-phase output voltages V_{o1}, V_{o2} and V_{o3} with variable frequency and variable amplitude following the reference voltages, in the same time the power factor of the matrix converter input side which is supplied by a fixed amplitude-frequency three-phase voltages should be kept nearly equal to the unity and the harmonics content of the input currents i_{i1}, i_{i2} and i_{i3} should be decreased. It is obvious that the output voltages are rich in harmonics, however their fundamental components have to follow the behaviour of the desired or the reference voltages. Taking into account t_{ij} as the conduction time of the S_{ij} (when S_{ij} is on) and T_s as the sampling time. The mean values of the output voltages over the K^{th} sequence can be expressed as follows:

$$\begin{aligned} v_{o1}^{(k)} &= v_{i1}^{(k)} \frac{t_{11}}{T_s} + v_{i2}^{(k)} \frac{t_{12}}{T_s} + v_{i3}^{(k)} \frac{t_{13}}{T_s} \\ v_{o2}^{(k)} &= v_{i1}^{(k)} \frac{t_{21}}{T_s} + v_{i2}^{(k)} \frac{t_{22}}{T_s} + v_{i3}^{(k)} \frac{t_{23}}{T_s} \end{aligned} \quad (9)$$

$$v_{o3}^{(k)} = v_{i1}^{(k)} \frac{t_{31}}{T_s} + v_{i2}^{(k)} \frac{t_{32}}{T_s} + v_{i3}^{(k)} \frac{t_{33}}{T_s}$$

The conduction time can be calculated function of each corresponding output, which means that for each phase there are three conduction times corresponding to the switches connected to the concerned phase. the details can be presented for each output phase as follows:

3.1. First phase

For the first phase, the conduction times of each switch is calculated as follows:

$$\begin{aligned} t_{11} &= \frac{T_s}{3} (1 + 2q \cos(\omega_m t + \theta_0)) \\ t_{12} &= \frac{T_s}{3} \left(1 + 2q \cos(\omega_m t + \theta_0 - \frac{2\pi}{3})\right) \\ t_{13} &= \frac{T_s}{3} \left(1 + 2q \cos(\omega_m t + \theta_0 - \frac{4\pi}{3})\right) \end{aligned} \quad (10)$$

3.2. Second phase

For the second phase, the conduction times of each switch are given by:

$$\begin{aligned} t_{21} &= \frac{T_s}{3} \left(1 + 2q \cos(\omega_m t + \theta_0 - \frac{4\pi}{3})\right) \\ t_{22} &= \frac{T_s}{3} (1 + 2q \cos(\omega_m t + \theta_0)) \\ t_{23} &= \frac{T_s}{3} \left(1 + 2q \cos(\omega_m t + \theta_0 - \frac{2\pi}{3})\right) \end{aligned} \quad (11)$$

3.3. Third phase

Finally, for the third phase, the conduction times of each switch are presented as:

$$\begin{aligned} t_{31} &= \frac{T_s}{3} \left(1 + 2q \cos(\omega_m t + \theta_0 - \frac{2\pi}{3})\right) \\ t_{32} &= \frac{T_s}{3} \left(1 + 2q \cos(\omega_m t + \theta_0 - \frac{4\pi}{3})\right) \\ t_{33} &= \frac{T_s}{3} (1 + 2q \cos(\omega_m t + \theta_0)) \end{aligned} \quad (12)$$

Where θ_0 is the initial phase angle.

The output voltage of the matrix converter at K^{th} sequence can be obtained using the modulation matrix at K^{th} sequence:

$$\left[V_o^{(k)} \right] = \left[M^{(k)} \right] \cdot \left[V_i^{(k)} \right] \quad (13)$$

where, the elements of the matrix $[M^{(k)}]$ are presenting the conduction times of each switch during one sampling time T_s , hence it is presented as follows:

$$[M^{(k)}] = \frac{1}{3}[\delta] + \frac{2q}{3} \begin{bmatrix} \cos(A) & \cos\left(A - \frac{2\pi}{3}\right) & \cos\left(A - \frac{4\pi}{3}\right) \\ \cos\left(A - \frac{4\pi}{3}\right) & \cos(A) & \cos\left(A - \frac{2\pi}{3}\right) \\ \cos\left(A - \frac{2\pi}{3}\right) & \cos\left(A - \frac{4\pi}{3}\right) & \cos(A) \end{bmatrix} \quad (14)$$

Where:

$$\begin{cases} \delta_{ij} = 1 \\ A = \omega_m t + \theta \end{cases} \quad (15)$$

4. Equations and symbols

In this paper, two simulation tests were performed to prove the advantage of the use of the open-end load fed by a dual matrix converter having the same topology and characteristics. The first simulation is presenting a star three-phase load fed by a single matrix converter, where the output voltage is characterized by a fundamental frequency $f_0 = 25$ Hz and an amplitude of $V_0 = 150$ V. The parameters of the input voltage and the balanced three-phase load are presented in Table. 1. On the other side, due to the distortion which may occur in the input current and to avoid the propagation of this kind of harmonics pollution toward the source an input LC filter is inserted as it is clearly shown in Fig. 2 and Table. 1.

Table 1. Simulation parameters

V_S	Input voltage	220 V
f_S	Input frequency	50 Hz
R_f	Input filter Resistance	0.5Ω
L_f	Input filter inductance	$0.003 H$
C_f	Input filter Capacitance	$2e-5 \mu F$
R_L	Load Resistance	10Ω
L_l	Load Inductance	$0.0112 H$

4.1. Single Matrix Converter feeding to star three-phase load

In this case, a single matrix converter is feeding a three-phase load under star connection. The output voltages, the output currents and the output phase to phase voltages of the matrix converter are presented in Fig. 3, 4, and 5.

respectively. The fundamental of the single phase output voltage is 153.9 V with a relatively high THD of 85.68% Fig. 6, in the same time a clear harmonics rays or a group of harmonics can be remarked clearly near the switching frequency $f_s = 2$ kHz, however and due to the inductive nature of the load the output current is characterized by a low THD of 4.17%, where the high order harmonics are eliminated Fig. 7.

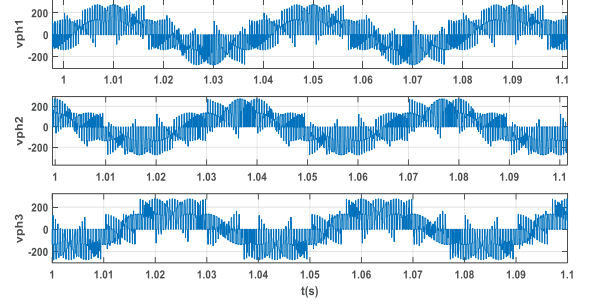


Fig.3 The output voltage of the single matrix converter.

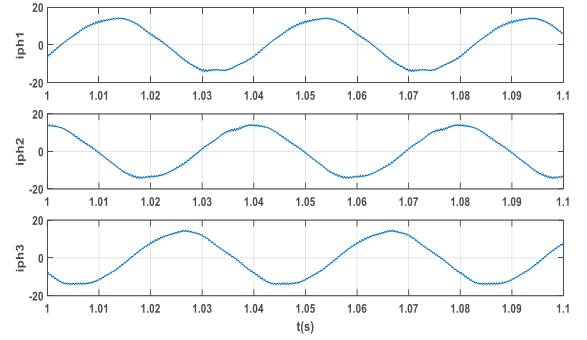


Fig. 4. The output currents of the matrix converter.

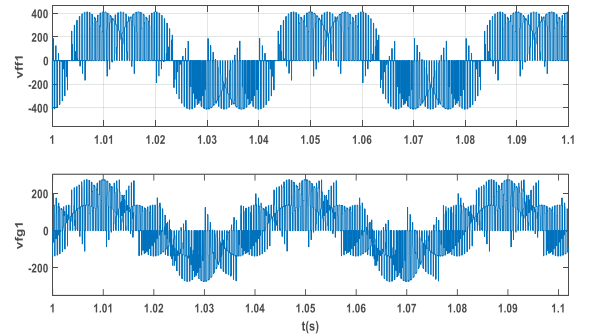


Fig. 5. The output phase to phase voltage and one phase voltage of the matrix converter.

On the other side, the effect of the insertion of the input filter can be shown clearly in Fig. 8, where the input current from the source side is nearly sinusoidal with neglected ripples with $THD = 11.37\%$, contrary the current from the matrix converter side has a $THD = 84.95\%$. Unfortunately, an important value of the common mode voltage at

the star point of the load is generated. It is due mainly to the components of voltage harmonics possessing a frequency multiply of three times of the fundamental frequency Fig. 9. It can present a serious risk for the equipment, the insulation, the operator its self, and the quality of power.

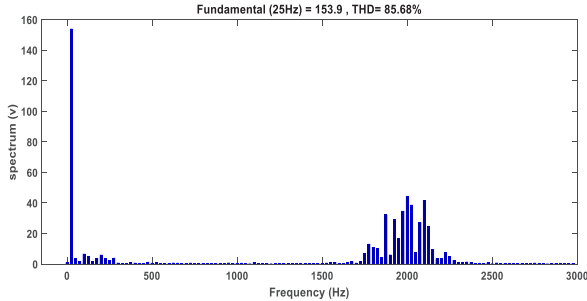


Fig. 6. The spectrum of the matrix converter output voltage.

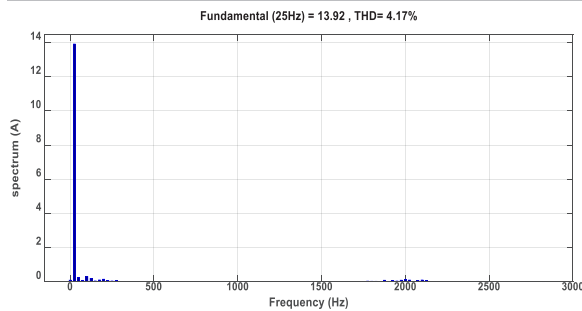


Fig. 7. The spectrum of the matrix converter output current.

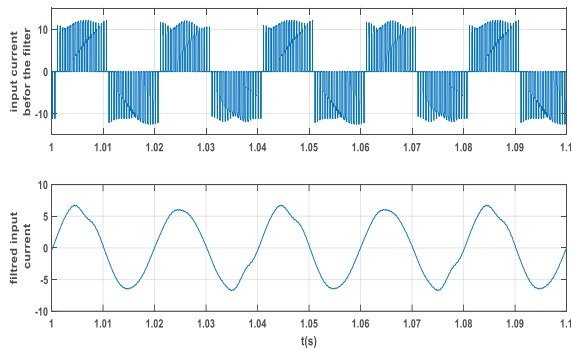


Fig. 8. The input current of the Matrix converter.

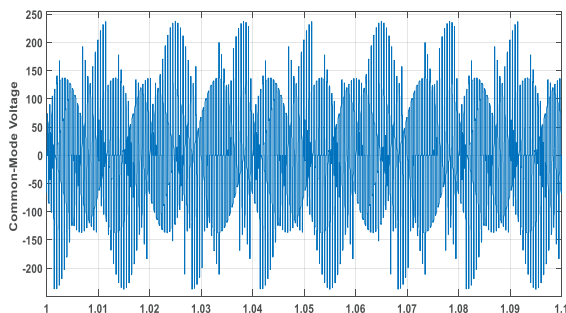


Fig. 9. The common-mode voltage waveforms

4.2. Dual Matrix Converter applied to open-end three-phase load

In this case, the both sides of the three-phase load are connected to two similar matrix converter (dual matrix converter) with the same characteristics as the single matrix converter aforementioned. The idea concerning this topology is that the voltage applied at the load will have a three-level voltage and the dual matrix converter will be seen from the load as a three-level converter. Under this topology the common mode voltage (CMV) will be totally removed and the three-phase load can be controlled independently, however due to the balanced nature of the load, the voltage applied at each load phase is the same, taking into account the eventual required shift phase. Figures 10. and 11. present the voltages between the two terminals of the three-phase load and the currents in the load respectively.

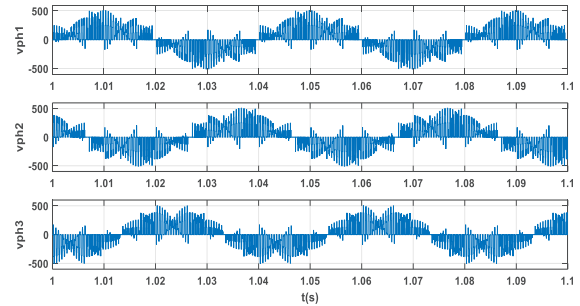


Fig.10. The voltage between the two terminals of the three-phase load.

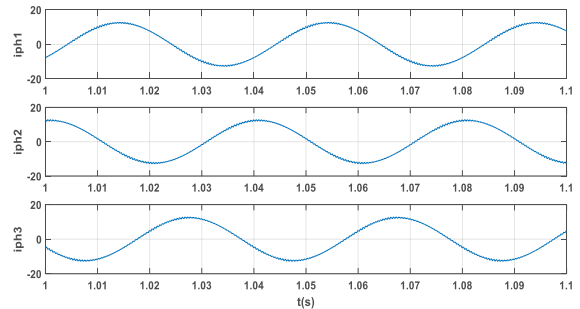


Fig.11 Current of the open-end three-phase load.

It is obvious that the fundamental magnitude of the voltage is 156.5 V with a THD of 91.59%, whereas, the fundamental current is 12.46 with a THD of 2.01 % which means that there is an improved current quality in comparison with a single matrix converter Fig. 12 and Fig. 13. On the other side, the first matrix converter and the second matrix converter have a fundamental magnitude voltages 78.38 V and 78.16 V respectively.

Whereas the output voltage to ground and the phase to phase voltage of the both matrix converter are presented in Fig. 14. and 15. respectively. The two level nature voltage is clearly observed and the three-level voltage nature at the load terminals is also obviously observed. the effect of the input filter can be seen clearly in Fig. 16, where the input current from the matrix converter side is rich in harmonics with THD = 89.86%, however the source current is nearly sinusoidal with THD = 13.57 % which is little bit great than the one of the single matrix converter, it may due to the level of the fundamental of the output currents and voltages.

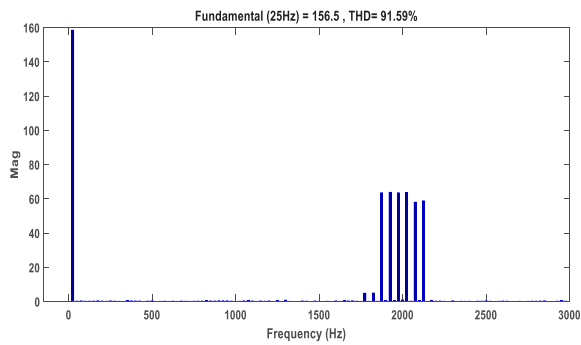


Fig.12. The spectrum of the matrix converter output voltage.

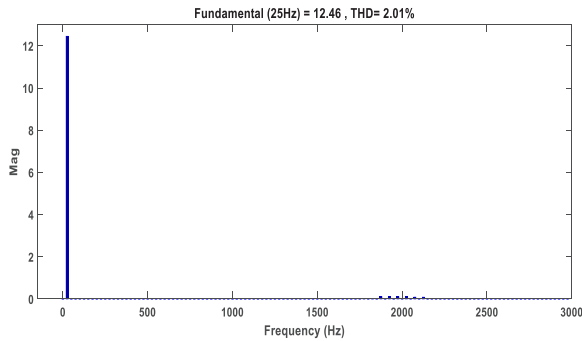


Fig.13. The spectrum of the matrix converter output current.

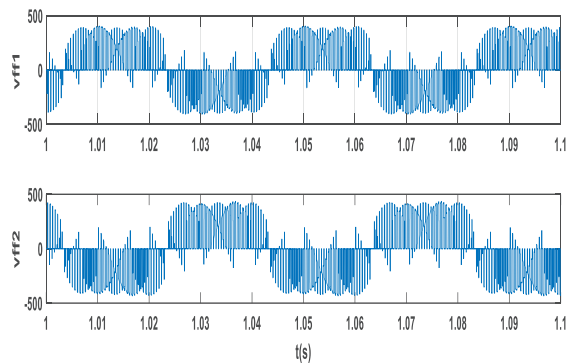


Fig.14. The output phase to phase voltage of MCI and MC2.

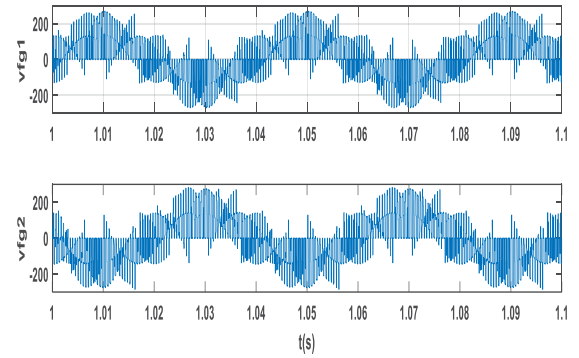


Fig. 15. The output voltage to ground of MCI and MC2.

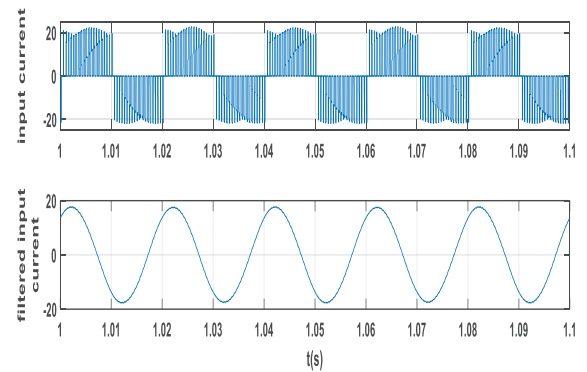


Fig. 16. The input currents before and after the input filter.

5. Conclusion

In the present paper the great advantage of the use of the dual matrix converter under the topology of open-end is proved clearly. The obtained results show the net improvement of the quality of the voltages applied to the load and the net elimination of the common mode voltage which is presenting a great problem in all three-phases load applications, especially in three-phases electrical machines. On the other side, and due to the multilevel nature of the voltage applied at the terminal of the open-end load, the current quality is also improved, where the THD is decreased at nearly 48% compared to the classical topology based on one matrix converter. An important issue, is that the use of a dual matrix converter is more reliable in case of fault on one or more switches, this problem can be solved by just adjusting the control of the switches. finally, it can be said to presented topology can have a large use in electric machines application in industrial plants based on the aforementioned advantages.

REFERENCES

1. P.W. Wheeler, P.W., Rodriguez, J., Clare, J.C., Empringham, L., and Weinstein, A: "Matrix Converters: A Technology Review", IEEE Trans. On Ind. Elect. vol. 49, no. 2, April, 2002, pp. 276-288.
2. René Vargas, José Rodríguez, Ulrich Ammann, and Patrick W. Wheeler, "Predictive Current Control of an Induction Machine Fed by a Matrix Converter With Reactive Power Control", IEEE Tran. On Ind. Electronics, vol. 55, no. 12, pp. 4362-4371, Dec. 2008.
3. M. Venturini, "A new sine wave in sine wave out, conversion technique which eliminates reactive elements," in POWERCON 7, 1980, pp. E3_1-E3_15.
4. A. Alesina, and M. Venturini, "Analysis and design of optimum amplitude nine-switch direct ac-ac converters", IEEE Trans. Power Elect. vol. PE-4, no. 1, pp. 101-112, 1989.
5. Sk. Moin Ahmed¹, Haitham Abu-Rub², Abdellah Kouzou³, "Predictive Simultaneous Power and Current Control in a Three-phase Direct Matrix Converter " .
6. MM. REZAOUL, L. NEZLI, MO.MAHMOUDI," Study of Output Voltages of a Matrix Converter Feeding a Five AC – Induction Motor Using the Strategy PWM Modulation to Five Intervals", Sixth International Conference on Electrical Engineering CEE 2010, Batna, Algeria (Oct 2010).
7. Sk. Moin Ahmed, Haitham Abu-Rub, Zainal Salam, "Common-Mode Voltage Elimination in a Three-to-Five-Phase Dual Matrix Converter Feeding a Five-Phase Open-End Drive Using Space-Vector Modulation Technique", IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, VOL. 62, NO. 10, OCTOBER 2015
8. Sk. Moin Ahmed, Haitham Abu-Rub, Zainal Salam, "Investigation of Space Vector Modulated Dual Matrix Converters Feeding a Seven Phase Open-end Winding Drive ", Industrial Electronics Society, IECON 2014 - 40th Annual Conference of the IEEE, Oct. 29 2014-Nov. 1 2014
9. Ghalem BACHIR, Azeddine BENDIABDELLAH," A Comparative Study between Two Control Strategies for Matrix Converter", Advances in Electrical and Computer Engineering, Volume 9, Number 2, 2009.
10. H. Altun, S. Sunter, "matrix converter induction motor drives modeling simulation and control", Electrical Engineering, December 2003, Volume 86, Issue 1, pp 25-33.
11. J.RODRIGUEZ, E.SILVA, and P.WHELLER "matrix converter controlled with the direct transfer function approach: analysis, modelling and simulation "in International journal of electronics Vol.92, No.2 2, February 2005, 63-85.
12. Abdelkader Djahar, Benyounes Mazari "High Performance Motor Drive Using Matrix Converter" acta electrotechnica et informatica No.2, Vol.7, 2007.
13. J.Rodriguez, E.Silva, F.Blaabjerg, P.Wheeler, J.Clare ,and J.Pontt, "Matrix converter controlled with the direct transfer function approach: Analysis, modelling and simulation," Int. J. Electron.,vol.92, no.2, pp.63–85, Feb.2005.
14. L.Zhang, C.Watthanasarn, and W.Shepherd, "Control of AC–AC matrix converters for unbalanced and/or distorted supply voltage," in Proc. 32nd Annu. IEEE Power Electron. Spec. Conf., 2001,vol.2, pp.1108–1113.
15. J.Rzasa, "Capacitor clamped multilevel matrix converter controlled with Venturini method," in Proc.13th EPEPEMC, Sep.2008, pp.357–364.
16. Y. Mei and L. Huang, "Improved switching function modulation strategy for three-phase to single-phase matrix converter," in Proc. IEEE IPEMC, May 2009, pp.1734–1737.

Authors' contacts:

Address:

¹Applied Automation and Industrial Diagnostics Laboratory (LAADI), Djelfa University departments of science and technology, Algeria

² University Amar Telidji of Laghouat

E-mail: Beladel_abdelkader@yahoo.fr

E-mail: kouzouabdellah@ieec.org



CONTROL OF NONLINEAR GAS TURBINE SYSTEM USING STRATEGIES OF MULTIDISCIPLINARY DESIGN OPTIMIZATION

NADJI HADROUG ⁽¹⁾, AHMED HAFIFA ⁽¹⁾, ABDELLAH KOUZOU ⁽¹⁾, MOULOUD GUEMANA ⁽²⁾
AND AHMED CHAIBET ⁽³⁾

Abstract: Gas turbines are known in recent years over a considerable development in many industrial applications. Especially in the field of transportation and reinjection of gaz. Despite their many advantages, their high sensitivity to accentual defects is very important because the influence of these defects can be critical to the proper functioning of these rotating machines. This paper is a contribution to the study of a gas turbine compressor station Type Titan 130, where this study proposes the analysis of the output parameter control strategy of the studied gas turbine, based on the multidisciplinary optimization of their operating variables, to meet the operating constraints and to maximize the life of this machine. The obtained results are satisfying and reliable for performing economic contributions to the studied system.

Keywords: Gas turbine, Multidisciplinary optimization, configuration control, power turbine.

1. Introduction

The study of the complex industrial systems have become a modern industrial challenge to analyze the sustainability of the operation of industrial systems. This certainly reduces the intervention costs and allows better meeting the objectives a priori for optimal system operation. For this purpose, this work proposes the study of the output parameters control strategy of a gas turbine based on optimizing their output variables to meet the operating constraints, to maximize the duration life of this machine, to prevent the aerodynamic limitations caused by the dropout problem and to ensure correct conduct of this system.

In general, the control problem consists of obtaining dynamic models of systems, then using these models to determine the control laws or the strategies to achieve the desired system response and performance. Indeed, the gas turbine system is made by sensitive organs to accidental faults that can be critical and costly for both sides, the humanly and the financially sides on the correct operation of the gas turbine used in pipeline. In order to build a predictive system capable of estimating the state of operation degradation of gas

turbines in operation even under severe degraded mode. However, the complexity of the design optimization depends on the complexity of the relevant disciplines, the sizing problem and the nature of objectives and constraints. In the case of the considered gas turbine, multidisciplinary optimization is proposed to the multivariable control of the studied gas turbine output variable.

2. Strategies of Multidisciplinary Optimization

Several multidisciplinary optimization strategies have been presented literature [1, 7, 9, 15, 16]. This optimization can facilitate the control of variables in interaction between system elements and can be used as decision support in regulation. In the case of the studied gas turbine, the choice of the parameters and the modeling and implementation of a control strategy and decision making are obviously important regarding the supervisory system configurations eligible for the gas turbine. The optimization problem can be presented by the minimization of the function $f(x, y, z)$ with respect to constraint functions, given by:

$$\begin{aligned}
& f(x, y, z) \\
& x, y, z \\
& g(x, y, z) \leq 0 \\
& h(x, y, z) = 0 \\
& \forall i, R_i(x_i, y_i, z_i) = 0 \\
& \forall j \neq i, y_{ij} = c_{ij}(x_j, y_j, z_j)
\end{aligned} \tag{1}$$

where x is the state variables, y are the coupling variables, z are the design variables, c is a function of coupling, f is an objective function, g is the inequality constraints, h are the equality constraints and R is residues function.

In this work, a more particular interest will be also transmitted to the defect detection phase which affect the system output variables of the studied gas turbine. Indeed, fault detection through the generation of a fault indicator called residue, this indicator is defined by a consistency relationship. It is designed to be zero on average in a case of operation. In contrast, in a case with default, the residue takes value away from its mean value and it is used to indicate the presence of a defect. The generation of residues characteristic of how the system, is the fundamental problem of diagnosis model based. This technique relies on the use of a correct assumed system model and generates residues that are the differences between the measured output signals and the estimated ones, as shown in Figure 1.

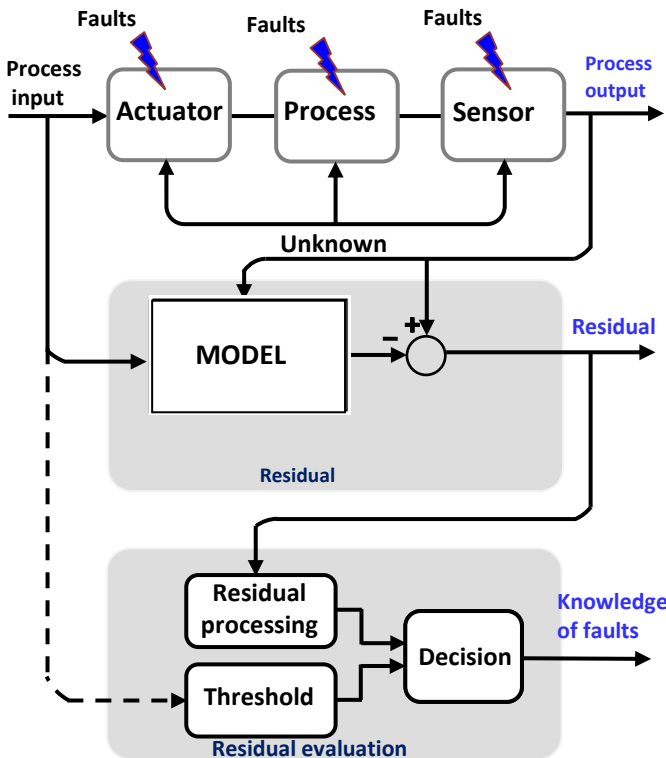


Fig. 1. Residual generation

For minimizing the equation error, the process model is written as:

$$y(k) = \Psi^T \theta \tag{2}$$

With θ and Ψ designating the parameters and regression vectors given by:

$$\theta^T = [a_1 \dots a_n, b_1 \dots b_n], \quad \Psi^T = [y(k-1) \dots y(k-n) \quad u(k-1) \dots u(k-n)]$$

The equation error $\varepsilon(\cdot)$ is introduced to the parameter estimation defining discrete transfer function of the system using the flowing formula:

$$\frac{y(z)}{u(z)} = \frac{B(z)}{A(z)} = \frac{\sum_{i=1}^{n_b} b_i z^{-i}}{1 - \sum_{i=1}^{n_a} a_i z^{-i}} \tag{3}$$

The error is defined by equation:

$$e(z) = \hat{B}(z)u(z) - \hat{A}(z)y(z) \tag{4}$$

Where $\hat{A}(z)$ and $\hat{B}(z)$ are the estimates of $A(z)$ and $B(z)$.

May provide a recursive form of estimated $\hat{\theta}$ using the recursive least squares algorithm [12]:

$$\hat{\theta}(k+1) = \hat{\theta}(k) + \gamma(k)e(k+1) \tag{5}$$

with :

$$\begin{cases}
\gamma(k) = \frac{1}{\Psi^T(k+1)P(k)\Psi(k+1) + 1} P(k)\Psi(k+1) \\
P(k+1) = [I - \gamma(k)\Psi^T(k+1)]P(k) \\
e(k) = y(k) - \Psi^T(k)\hat{\theta}(k-1)
\end{cases}$$

For the minimization of the output error, instead of as previously calculate the error equation, rather evaluates the output error, given by:

$$e(k) = y(k) - \hat{y}(\theta, k) \tag{6}$$

Where: $\hat{y}(\theta, z) = \frac{\hat{B}(z)}{\hat{A}(z)}u(z)$ and $\hat{y}(\theta, z)$ is

the model output.

The output of error $e(\cdot)$ in this case is being a nonlinear function of the parameters to be estimated, then numerical optimization methods are used to minimize the error $e(k)$ with respect to θ parameters. From the modeling of these residues, a strategy of the output control of the studied gas turbine parameters, based on the optimization of their operating variables which will be introduced in the implementation section, which are typically zero mean and a variance determined in the absence of malfunctions.

3. Industrial Application

A gas turbine is a combustion engine that can convert natural gas or other liquid fuels to mechanical energy, this energy then drives a generator. As part of this work, a gas turbine type two shafts TITAN 130 is studied, this gas turbine is installed at the gas compression station of Sonatrach SC3 in Djelfa, central Algeria. This gas compressor station is located on the pipeline of the GG1 line at a distance of 182 KM from the starting point to Hassi R'Mel southern Algeria, this station will enhance the transport capacity of the pipeline 07 to

13 billion m³ per year. It also has three turbochargers and a turbine generator, a control building equipped with DCS (Digital Control System). The *Titan 130* two-shaft is rated at 20,500 hp with 36.2% efficiency and its single shaft configuration is rated at 15.0 MWe with 35.2% at ISO operating conditions [17]. The turbomachinery comprises a turbine engine and a compressor mounted on a steel frame, with the support systems and ancillary anti-slip material, this turbomachinery is means; (Compressor Set / Mechanical Drive: All Compressor / Drive Mechanical), as shown in Figures 2.

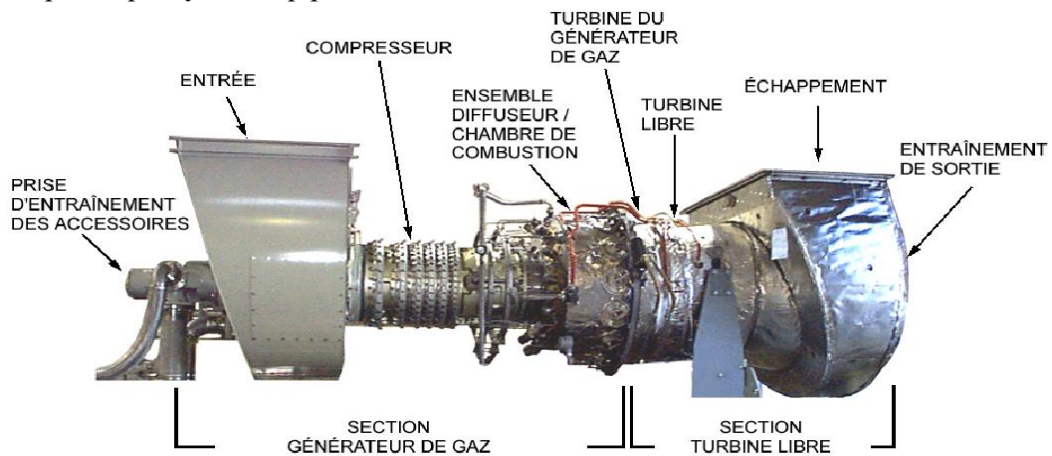


Fig. 2. Diagram of Titan 130 Two-Shaft Gas Turbine

The gas turbine provides the power to drive the compressor, has two main parts: the gas generator (GP) and the power turbine (PT). The gas generator (GP) converts the energy of the fuel gas at high temperature that takes the power turbine and

converts rotational energy. The fundamental principle of a gas turbine is converted fuel energy to mechanical energy by the application of a thermodynamic process known by the Brayton cycle, as shown in Figures 3.A and 3. B.

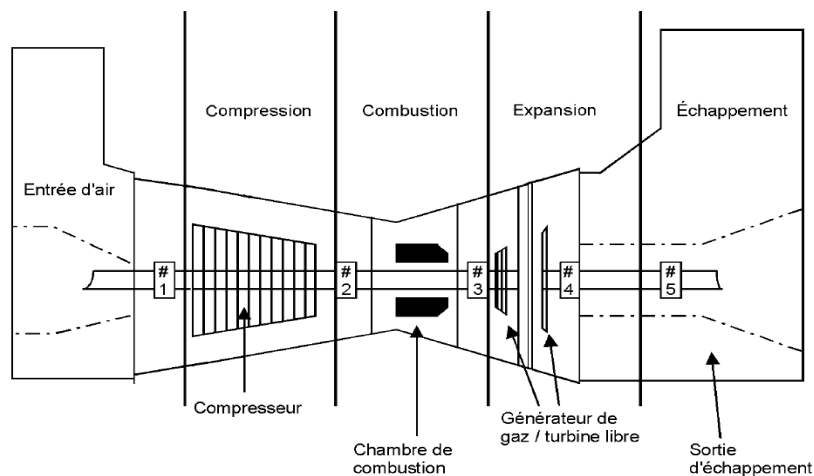


Fig. 3. A: States air flow of Brayton cycle

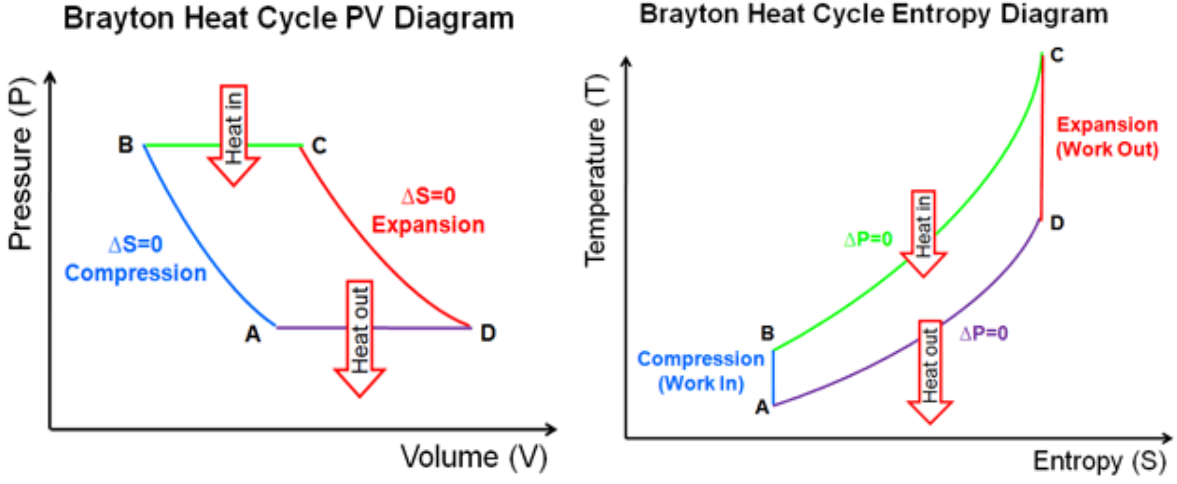


Fig. 3. B: Curves of Brayton cycle

3.1. A gas turbine control

The model of the turbine discussed in this work has two entries: room temperature (T_1) and mass flow rate of fuel flow (\dot{m}_c), sensors supplied from the turbine outputs include: the speed of HP high pressure shaft (NGP), the low pressure shaft (NPT) and the temperature of the blades (T_3). These measurements can be used to provide different pairs of input / output for the closed loop control. There are other important variables such as;

The air mass flow (\dot{m}_a), the pressure ratio at the outlet of the axial compressor (PCD) and the fuel / air ratio (f).

The proposed control configuration is shown in Figure 4, with the nonlinear model used for PI control configuration (proportional / integral). This closed loop control strategy is chosen for parameter optimization of a PI controller structure is designed for a nominal operating point of the discussed gas turbine.

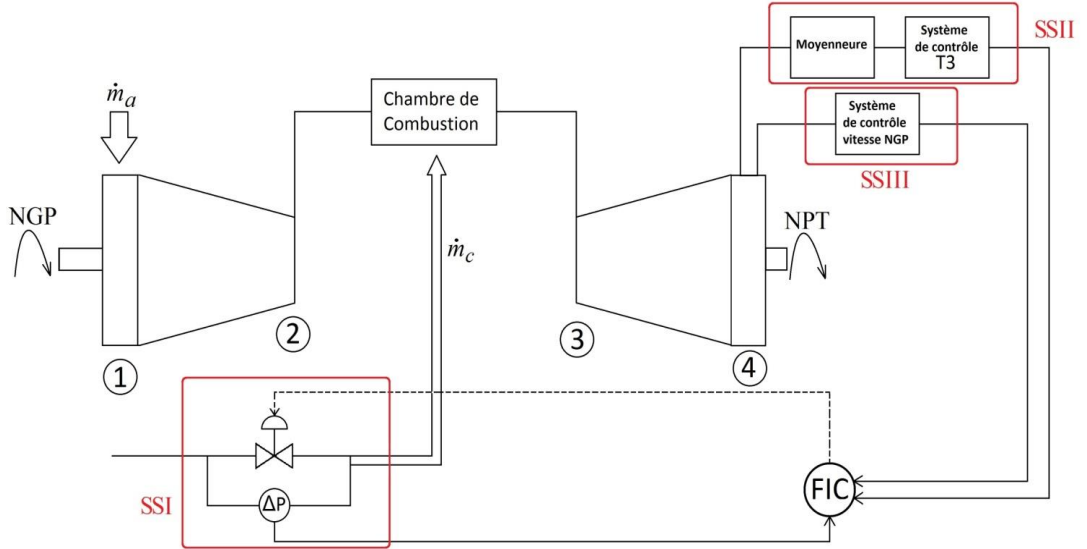


Fig. 4. Control setting of different subsystem of the TITAN 130 turbine

The mass flow rate of fuel flow (\dot{m}_c) is given by:

$$\dot{m}_c = \sqrt{\Delta P} \cdot \frac{1}{2} \rho \mu^2 = \frac{\sqrt{\Delta P}}{k} \quad (8)$$

The air mass flow (\dot{m}_a) is given by:

$$\begin{aligned} \dot{m}_a &= k'(P_2 - P_1) \\ \dot{m}_a &= k'(PVD - P_{atm}) \end{aligned} \quad (9)$$

The fuel / air ratio is calculated to ensure the control of the ratio f which is given by:

$$f = \frac{(T_3 - T_2)}{\eta_b \cdot CV + T_3} \quad (10)$$

3.2. Application results

The multidisciplinary optimization results developed in this work allows the control of gas turbine output variables. Figure 5 shows that the power absorbed by the compressor is proportional to the pressure ratio for an ambient temperature range ($0^\circ\text{C} < T_{\text{Amb}} < 50^\circ\text{C}$).

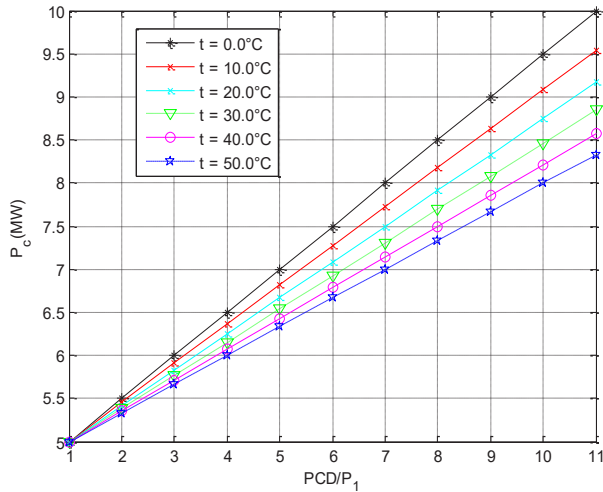


Fig. 5. Gas turbine power at different operating temperature

In Figure 6, when the ambient temperature is high ($> 30^\circ\text{C}$), it is noted that the performance of the profile tends to a maximum value of 7. Every time, it is not possible to exceed the value ($\varepsilon \approx 8.7091$) given by the manufacturer, because of the limitations imposed by the resistance of materials.

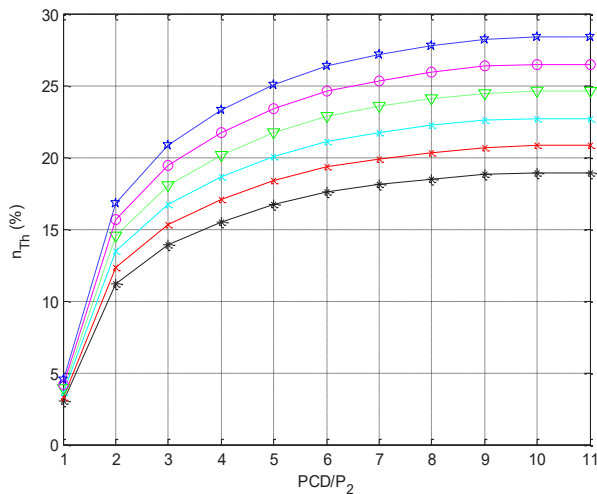


Fig. 6. Gas turbine efficiency at different operating temperature

On Figure 7 and 8, the power absorbed by the compressor and that supplied by the turbine will vary proportionally with the ambient temperature and the PT remains almost constant because the combustion temperature limit (given by the manufacturer).

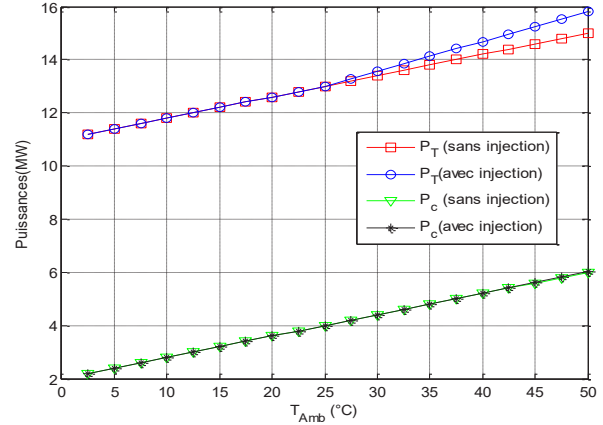


Fig. 7. Gas turbine power at different operating pressure

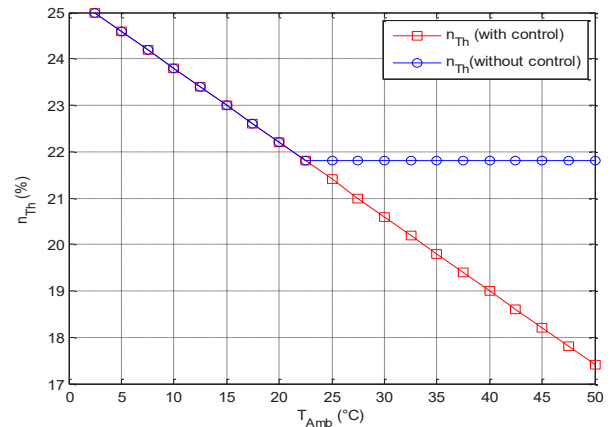


Fig. 8. Yield control of the examined gas turbine

Figure 9.A shows the discharge pressure variation of the fuel gas valve in the combustion chamber, and the variation of its residue is shown in Figure 9.B. Figure 10.A shows the discharge pressure change axial compressor and the variation of its residue is given in Figure 10.B, the control of the position of the fuel valve to the gas inlet of the combustion chamber is shown in Figure 11.

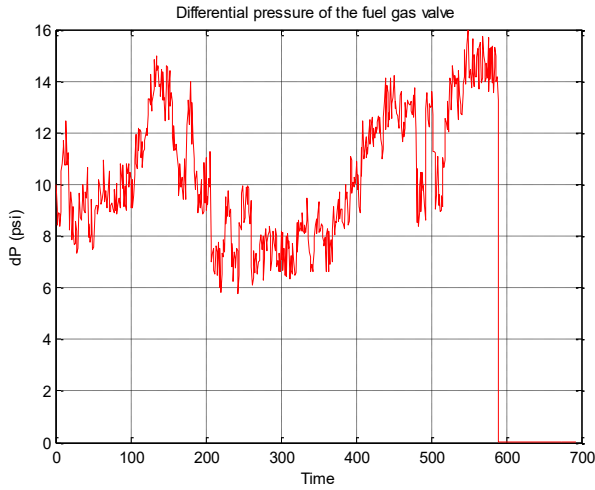


Fig. 9. A: Differential pressure of the fuel gas valve

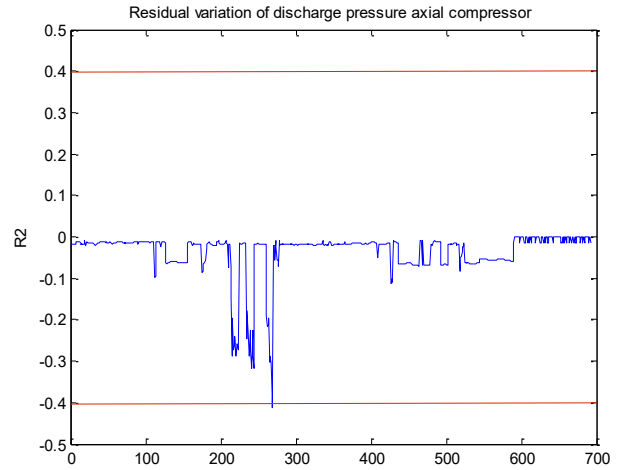


Fig.10. B: Residual variation of discharge pressure axial compressor

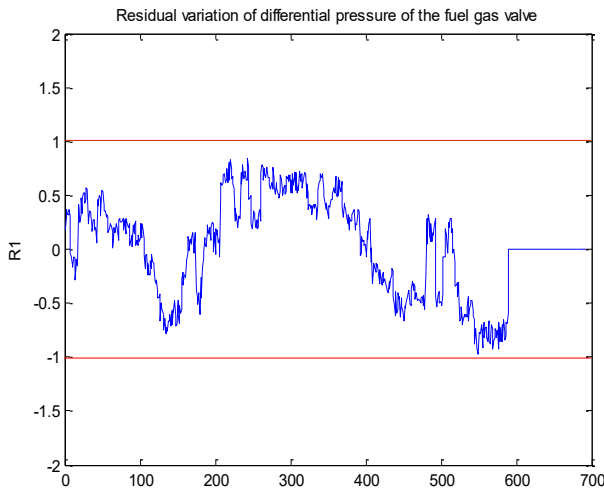


Fig. 9. B: Residual variation of differential pressure of the fuel gas valve

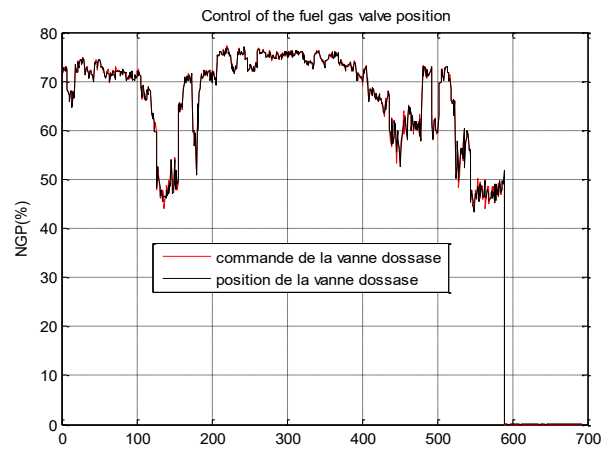


Fig. 11. Control of the fuel gas valve position

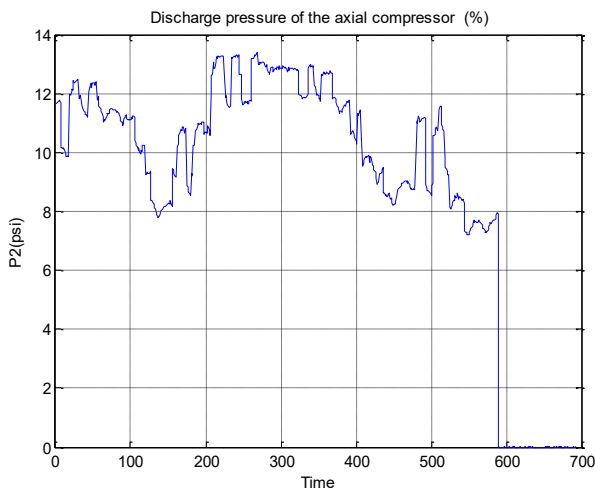


Fig. 10. A: Discharge pressure axial compressor

The turbine rotational speed variation NGP are controlled, after optimizing the manipulated variables pressure and mass flow rate suction and gas turbine system of delivery is given in Figure 8, the speed sensor is used in order to measure the speed of rotation of the HP shaft of the axial compressor, where the target value recommended by the manufacturer is 100%.

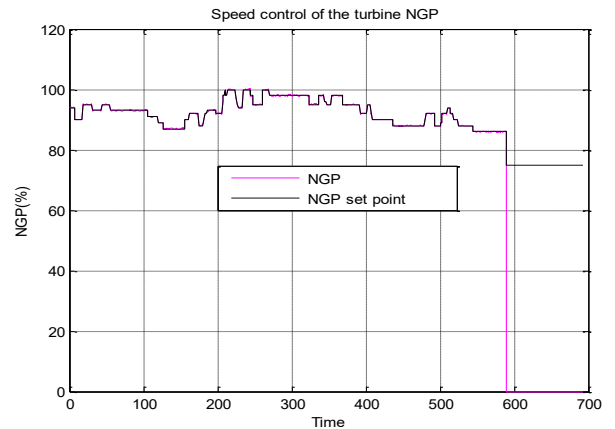


Fig. 12. Speed control of the turbine NGP

To confirm the control performance in more detail, validation tests were performed on the considered gas turbine, Figure 13 shows the variation of exhaust temperature of the examined turbine.

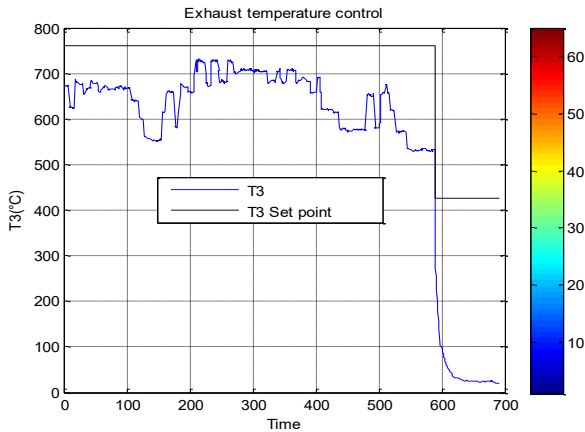


Fig. 13. Exhaust temperature control

This method leads to an improvement of the efficiency of a gas turbine, Figure 14 shows the variation of combustion efficiency. For normal operation of the studied gas turbine, when the ambient temperature rises above the reference machine performance decreases. The reduction becomes large when the room temperature is maximum, as shown in Figure 14.



Fig. 14. Combustion efficiency

Figure 15 shows converted fuel energy to mechanical energy by the application of a thermodynamic process known by the Brayton cycle.

The results of multidisciplinary optimization, applied to the control of the gas turbine output variable are satisfactory. They are used to solve some problems of supervision of this type of rotating machine. This allows the improvements in terms of studied gas turbine cost oversight, and the control configuration proposed

also led to the improvements in the operation of the studied gas turbine.

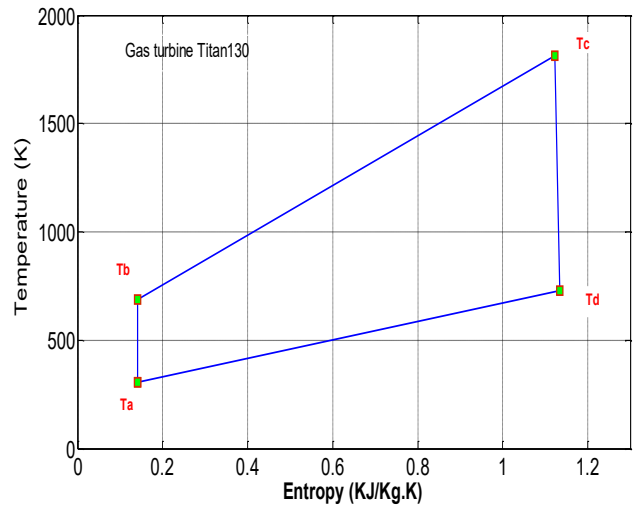


Fig. 15. T-S diagram of the real process in the studied gas turbine

4. Conclusion

The gas turbine system is composed of sensitive organs to accidental defects that can be critical and costly humanly and financially at the same time during the continuous operation of this system, especially in the gas transportation pipeline. Indeed, the complexity of the supervision in gas turbine plants is due to the strong interaction of the control loops in these industrial systems. For this purpose, the present paper presents the development of a multidisciplinary optimization which allows an efficient data representation of the studied gas turbine variables by integrating a control configuration for these variables. The improvements that have been obtained by using the optimized gas turbine output variables to simplify their control. These improvements are reflected in the performance of the power quality and output of the studied gas turbine.

REFERENCES

1. Abdelhafid Benyounes, Ahmed Hafaiifa and Mouloud Guemana, (2016). Gas turbine modelling based on fuzzy clustering algorithm using experimental data. *Journal of Applied Artificial Intelligence*, Taylor & Francis, vol. 30, no. 1, pp. 29-51.
2. Abdelhafid Benyounes, Ahmed Hafaiifa and Mouloud Guemana, (2016). Fuzzy logic addresses turbine vibration on Algerian gas line. *Oil & Gas Journal* | Jan. 4, pp. 22-28
3. Ahmed Hafaiifa, Guemana Mouloud and Saadat Boulanouar, (2016). Monitoring system based on real data acquisition for vibrations control

- in gas turbine system. *Revue de Nature & Technologie: A- Sciences fondamentales et Engineering*, vol. Jan. n° 14, pp. 13 – 18.
4. Boulanouar Saadat, Ahmed Hafaifa and Mouloud Guemana, (2016). Vibration analysis and measurement based on defect signal evaluation: Gas turbine investigation. *Journal of Advanced Research in Science and Technology*, vol. 3, no 1, pp.271-280
 5. Benrabeh Djaidir, Ahmed Hafaifa and Kouzou Abdallaha, Vibration detection in gas turbine rotor using artificial neural network. Proceedings of the International Conference on Acoustics and Vibration ATAVI'16, March 21-23, 2016, Hammamet -Tunisia
 6. Mohamed Ben Rahmoune, Ahmed Hafaifa, Mouloud Guemana, Fault diagnosis in gas turbine based on neural networks: Vibrations speed application. Proceedings of the International Conference on Acoustics and Vibration ATAVI'16, March 21-23, 2016, Hammamet -Tunisia
 7. Djamel Halimi, Ahmed Hafaifa and Elahmoune Bouali (2014). Maintenance actions planning in industrial centrifugal compressor based on failure analysis. *The quarterly Journal of Maintenance and Reliability*. Maintenance and Reliability, Vol. 16, no. 1, pp. 17–21
 8. Bindi Chen, Peter C. Matthews, Peter J. Tavner, (2013). Wind turbine pitch faults prognosis using a-priori knowledge-based ANFIS. *Expert Systems with Applications*, vol. 40, no. 17, pp. 6863-6876.
 9. Forbes G.L. and R.B. Randall, Simulated gas turbine casing response to rotor blade pressure excitation. Proceeding of the 5th Australasian Congress on Applied Mechanics, ACAM 2007, 2007, Brisbane, Australia.
 10. Grondahl C.M., M.E. Guiler, G.E. Jurczynski and R. Zell (1990), Performance and reliability improvements for MS3002 and MS5001 turbines in pipeline applications. Proceeding of the Petroleum and Chemical Industry Conference: Industry Applications Society 37th Annual, Houston, TX , USA, pp. 135 – 144.
 11. Gwo-Chung Tsai (2004), Rotating vibration behavior of the turbine blades with different groups of blades. *Journal of Sound and Vibration*, vol. 271, no. 3-5, 6, pp. 547-575.
 12. Hector Sanchez, Teresa Escobet, Vicenç Puig, Peter Fogh Odgaard, (2015). Health-aware Model Predictive Control of Wind Turbines using Fatigue Prognosis. *IFAC-PapersOnLine*, vol. 48, no. 21, pp. 1363-1368.
 13. Martha A. Zaidan, Andrew R. Mills, Robert F. Harrison, Peter J. Fleming, (2016). Gas turbine engine prognostics using Bayesian hierarchical models: A variational approach. *Mechanical Systems and Signal Processing*, vol. 70–71, pp. 120-140.
 14. Qinming Liu, Ming Dong, Wenyuan Lv, Xiuli Geng, Yupeng Li, (2015). A novel method using adaptive hidden semi-Markov model for multi-sensor monitoring equipment health prognosis. *Mechanical Systems and Signal Processing*, vol. 64–65, pp. 217-232.
 15. Ren-Jieh Kuo, Man-Hsin Huang, Wei-Che Cheng, Chih-Chieh Lin, Yung-Hung Wu, (2015). Application of a two-stage fuzzy neural network to a prostate cancer prognosis system. *Artificial Intelligence in Medicine*, vol. 63, no. 2, pp. 119-133.
 16. Yang Chen, Guobiao Cai, Zhenpeng Zhang, Yulong Huang, (2014). Multi-field coupling dynamic modeling and simulation of turbine test rig gas system. *Simulation Modelling Practice and Theory*, vol. 44, pp. 95-118.
 17. Saadatmand, M. R., Rocha, G., 1999, Design and Development of the Solar Turbines Single Shaft Titan 130 Industrial Gas Turbine” Power Gen Europe 1999.

Authors' contacts

Nadji Hadroug⁽¹⁾, **Ahmed Hafaifa**⁽¹⁾,
Kouzou Abdellah⁽¹⁾, **Mouloud Guemana**⁽²⁾
and Ahmed Chaibet⁽³⁾

⁽¹⁾ Applied Automation and Industrial Diagnostics Laboratory, Faculty of Science and Technology, University of Djelfa 17000 DZ, Algeria

⁽²⁾ Faculty of Science and Technology, University of Medea, Algeria

⁽³⁾ Aeronautical Aerospace Automotive Railway Engineering school, ESTACA Paris, France

Nadji Hadroug⁽¹⁾

E-mail: N_Hadroug@univ-djelfa.dz

Ahmed Hafaifa⁽¹⁾

E-mail: hafaifa.ahmed.dz@ieee.org

Kouzou Abdellah⁽¹⁾

E-mail: kouzouabdellah@ieee.org

Mouloud Guemana⁽²⁾

E-mail: guemana_m@yahoo.fr

Ahmed Chaibet⁽³⁾

E-mail: ahmed.chaibet@estaca.fr



STABILIZING LINEAR MULTIVARIABLE GAS TURBINE MODEL VIA SLIDING MODE CONTROL

BACHIR NAIL¹, ABDELLAH KOUZOU¹, AHMED HAFIFA¹, BELKACEM BEKHITI²

Abstract: In this paper, linear sliding mode control (LSMC) theory is used to stabilize the dynamic behavior of a linear multiple input multiple output (MIMO) system (Power Gas turbine). The main pros in this kind of control is the discontinuity behavior which can be applied to the uncertainties systems. Where, the first step must define the hyper-planes in the state space and calculate the equivalent control input the second step calculates the hyper-plane matrix G . The checking test of the robustness and the sensitivity under the variation of the dynamic and against the rejection of the external perturbation gives a clear proof of the presented control robustness in such application.

Key words: LSMC, linear MIMO System, Hyper-planes, MRLS, Power Gas Turbine, Robustness.

1. Introduction

The research on stability concept is attracting more attention and it occupies a large space in modern and classical control theory. Indeed, there are many theories in this area that have been introduced and used in several industrial control applications, such as; the Lyapunov stability theory based on energy [1], the bounded input bounded output (BIBO) stability theory [2], the input output stability, the hyper-stability [Dore Landau], the state feedback static gain stability based on pole placement, the linear quadratic regulator ..etc.[3],[4], All of these theories and others clarify that the stability study of systems is wide domain of research,. On the other side, the importance of this domain has its great impact in industrial systems applications, where the loss of stability in these systems can create dangerous and drastically problems, such as; the nuclear reactors, the jet engines, and the gas turbine which is the subject of the application presented in this paper.

Several papers has been written in recent years about the stability theories and their applications, particularly in linear MIMO systems such as; the optimal control [5], the multivariable feedback design [6], the Eigen structure assignment and controller optimization for mechanical systems [7], the optimal sliding mode control of a flexible spacecraft under stochastic disturbances [8], The left and right block pole placement comparison study applied to flight [9].

The main contribution of the present paper focuses mainly on the application of the linear sliding mode control, which is one of the control methods within the family of variable structure controls. This control strategy is applied on a gas turbine system, where the hyperbolic tangent function (\tanh) is used instead of the signum function (sign) [10].

2. Sliding Mode Control of a Linear System

Consider the following uncertain linear system:

$$\dot{x} = Ax + Bu \quad (1)$$

For m inputs, m hyper-planes in the state space are defined as:

$$s_i(t) = g_i^T x(t); \quad i = 1, 2, \dots, m \quad (2)$$

These hyperplanes pass through the origin of the state space $s_i = 0$ for From (1) and (2), the following expression is obtained:

$$\dot{s}_i(t) = g_i^T \dot{x}(t) = g_i^T Ax + g_i^T Bu; \quad i = 1, 2, \dots, m \quad (3)$$

furthermore equation (3) can be rewritten in the matrix formas follows:

$$\dot{s}(t) = GAx + GBu \quad (4)$$

Where

$$s = \begin{bmatrix} s_1 & s_2 & \dots & s_m \end{bmatrix}^T \quad (5)$$

And

$$G = \begin{bmatrix} g_1 & g_2 & \dots & g_m \end{bmatrix}^T \quad (6)$$

it is noted that:

$$s(t) = Gx(t) \quad (7)$$

and the vector $s(t) = 0$ represents the intersection of all the m sliding hyper-planes passing through the origin of the state space. The corresponding equivalent control input $u_{eq}(t)$ based on (4) it is expressed as follows (4)

$$u_{eq}(t) = -(GB)^{-1} GAx \quad (8)$$

Note that the matrix GB has been assumed to be nonsingular. The reaching condition for each hyper-plane is [8],[11],[12],[19]:

$$s_i \dot{s} < 0 \quad (9)$$

Conditions (9) are satisfied by the following control law:

$$u(t) = u_{eq}(t) - (GB)^{-1} \text{diag}(\eta) \text{sgn}(s(t)) \quad (10)$$

Where $\text{diag}(\eta)$ is a diagonal matrix with the i^{th} diagonal element equal to a positive number η_i

2.1. Computation of Sliding mode control (SMC) Hyper-plane Matrix G

Define a similarity transformation

$$q(t) = Hx(t) \quad (11)$$

Where

$$H = [N \quad B]^T \quad (12)$$

and the columns of the $nx(n-m)$ matrix N are composed of basis vectors of the null space [16] of B^T . From (1) and (11)

$$\dot{q} = \bar{A}q(t) + \bar{B}u(t) \quad (13)$$

Where

$$\bar{A} = HAH^{-1} \quad (14)$$

And

$$\bar{B} = HB \quad (15)$$

Because of the special structure of the matrix H , the first $(n-m)$ rows of \bar{B} turn out to be zeros. Hence, the vector q is decomposed as follows:

$$q = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} \quad (16)$$

Where q_1 and q_2 are $(n-m)$ and m dimensional vectors, respectively. Partitioning (13),

$$\begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \end{bmatrix} = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} \\ \bar{A}_{21} & \bar{A}_{22} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \bar{B}_r \end{bmatrix} u \quad (17)$$

and the sliding hyper-planes can be described as

$$s(t) = q_2(t) + Kq_1(t) \quad (18)$$

From (7) and (18),

$$G = \begin{bmatrix} K & I_m \end{bmatrix} H \quad (19)$$

Equation (19) indicates that the matrix G can be determined via the matrix K . The first objective behind the choice of the matrix G or the matrix K is to ensure the system stability on the intersection of all hyper-planes. When, $s(t) = 0$

$$q_2(t) = -Kq_1(t) \quad (20)$$

Equation (17) yields,

$$\dot{q}_1 = \bar{A}_{11}q_1 + \bar{A}_{12}q_2 \quad (21)$$

And

$$\dot{q}_2 = \bar{A}_{21}q_1 + \bar{A}_{22}q_2 + \bar{B}_r u \quad (22)$$

For the system (21), $q_2(t)$ can be viewed as the input and equation (20) as the state feedback law. Authors in [3] have shown that $(\bar{A}_{11}, \bar{A}_{12})$ is controllable if (A, B) is controllable. Therefore, the matrix K can be selected by using optimal control LQR [10]. From (21) and (8), the system dynamics on the intersection of sliding hyper-planes is given by

$$\dot{x} = \left[A - B(GB)^{-1} GA \right] x \quad (23)$$

When G is chosen according to Equation (19), $(n-m)$ eigenvalues of $\left[A - B(GB)^{-1} GA \right]$ are those of $(\bar{A}_{11} - K\bar{A}_{12})$ and remaining m eigenvalues are 0, which is a reflection of the fact that m degrees of freedom of the system is lost when $s(t) = 0$

2.2. Computation of K Matrix

Linear quadratic regulator (LQR) [4], is one of the optimal linear control methods that are used to stabilize the system by state feedback control $q_2(t) = -kq_1(t)$

Where k is obtained by minimizing the cost function:

$$J = \int (q_1^T Q q_1 + q_2^T R q_2) dt \quad (24)$$

Where, Q and R are positive semi-definite and positive definite symmetric constant matrices respectively. The LQR gain matrix K is given by :

$$K = R^{-1} \bar{A}_{12}^T P \quad (25)$$

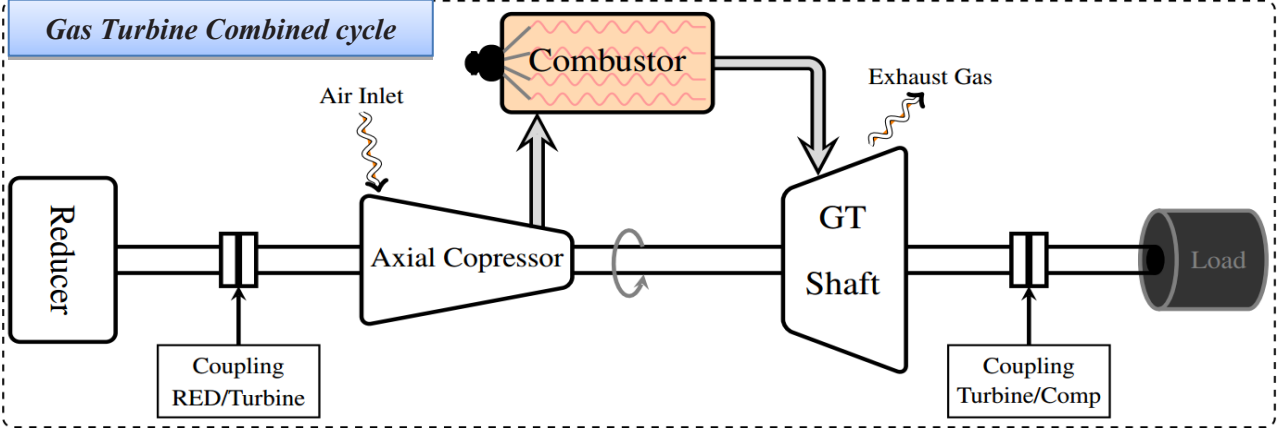
Where, P is a positive definite symmetric constant matrix obtained from the solution of algebraic matrix RICCATI equation (ARE):

$$\bar{A}_{11}^T P + P \bar{A}_{11} - P \bar{A}_{12} R^{-1} \bar{A}_{12}^T P + Q = 0 \quad (26)$$

Finally,

$$A_c = A - B(GB)^{-1} (GA + \eta G) \quad (27)$$

$$C_c = C - D(GB)^{-1} (GA + \eta G) \quad (28)$$



3. System model

The GE MS9001E Gas turbine model is obtained via the parametric identification of the real data, using MIMO Recursive least square (MRLS) estimator [24-26], for the validation of this model, the order of estimated model has to be changed every time from the order $n=6$ until the order $n=15$, this order gives the best models that present all frequencies of the gas turbine dynamic behavior, with almost negligible norm error between data outputs and estimated model outputs.

In this paper, it is mandatory to present the gas turbine model, by the important characteristics and key, «Pole-Zero map» Fig. 1, because the system is with a large scale, with 3 inputs and 3 outputs, which requires a big space.

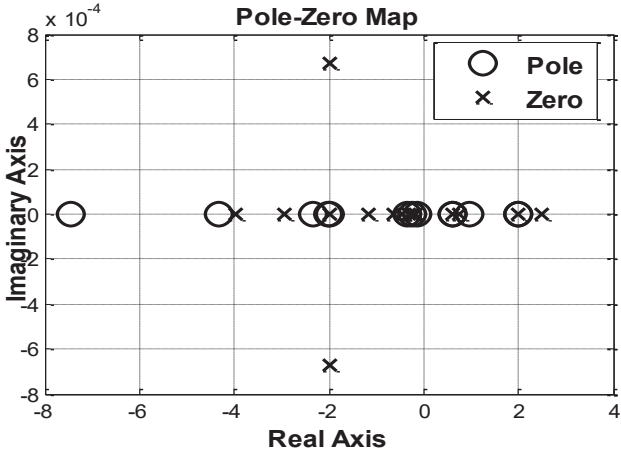


Fig. 1. System representation via pole-zero map

The values are chosen as follows:

$$\eta = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 7 \end{bmatrix}, H = [h_1 \quad h_2 \quad h_3]$$

Whereas, the elements of the matrices H, G, K are calculated based on equation (12), (19) and (25) respectively as follows:

$$h_1 = \begin{bmatrix} 0.0257 & -0.0965 & -0.1834 & 0.9549 & 0.0225 \\ -0.0129 & 0.0483 & 0.0917 & 0.0225 & 0.9887 \\ 0.8470 & 0.0577 & -0.0100 & 0.0021 & -0.0010 \\ -0.2253 & -0.1997 & 0.0122 & -0.0118 & 0.0059 \\ -0.1339 & -0.6529 & 0.0794 & -0.0303 & 0.0151 \\ 0.0590 & -0.2499 & 0.0818 & -0.0009 & 0.0004 \\ -0.0341 & 0.0614 & 0.2998 & 0.0668 & -0.0334 \\ 0.1075 & -0.4089 & -0.7441 & -0.1842 & 0.0921 \\ 0.0190 & -0.0714 & -0.1417 & -0.0346 & 0.0173 \\ 0.0048 & -0.0167 & -0.0300 & -0.0074 & 0.0037 \\ 0 & 0.0001 & 0 & 0 & 0 \\ 0 & 0 & -0.0001 & 0 & 0 \\ -2.3209 & 0 & 0 & 0 & 0 \\ 0 & 0.7071 & -0.7071 & 0 & 0 \\ 0 & 0 & 0 & 0.1118 & -0.0559 \end{bmatrix}$$

$$h_2 = \begin{bmatrix} -0.0105 & -0.0180 & -0.0630 & -0.0170 & 0.0618 \\ 0.0053 & 0.0090 & 0.0315 & 0.0085 & -0.0309 \\ 0.4718 & 0.1656 & 0.1701 & -0.0040 & 0.0050 \\ 0.1475 & 0.9319 & -0.1238 & -0.0291 & 0.0054 \\ 0.1074 & -0.1076 & 0.7049 & -0.1046 & 0.0092 \\ -0.0279 & -0.0178 & -0.0867 & 0.9584 & -0.0117 \\ 0.0160 & 0.0181 & 0.0626 & 0.0118 & 0.9041 \\ -0.0440 & -0.0747 & -0.2628 & -0.0718 & 0.2517 \\ -0.0077 & -0.0137 & -0.0475 & -0.0126 & 0.0476 \\ -0.0021 & -0.0030 & -0.0106 & -0.0029 & 0.0102 \\ -0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & -0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 5.2502 & -1.5817 & -1.4614 & 0.1335 & -0.0498 \\ -0.5707 & 0.3850 & 0.9300 & 0.3209 & 0.1145 \\ -0.0685 & 0.0546 & 0.1156 & 0.0105 & -0.1618 \end{bmatrix}$$

$$h_3 = \begin{bmatrix} -0.1851 & -0.0344 & -0.0075 & 0.0000 & 0 \\ 0.0926 & 0.0172 & 0.0038 & -0.0000 & 0 \\ 0.0088 & 0.0017 & 0.0001 & 0 & 0 \\ -0.0500 & -0.0087 & -0.0020 & 0 & 0 \\ -0.1300 & -0.0221 & -0.0053 & 0.0001 & 0 \\ -0.0060 & -0.0002 & -0.0003 & 0 & 0 \\ 0.2736 & 0.0512 & 0.0111 & 0 & 0 \\ 0.2434 & -0.1407 & -0.0307 & 0 & 0 \\ -0.1421 & 0.9736 & -0.0058 & 0 & 0 \\ -0.0306 & -0.0057 & 0.9988 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0.0016 & -0.0023 & 0.0027 & 0.0001 & -0.0001 \\ 0.0194 & -0.0034 & 0.0008 & -0.0002 & -0.0001 \\ 0.4574 & 0.0857 & 0.0185 & -0.0000 & 0 \end{bmatrix}$$

$$K = [k_1 \quad k_2]$$

$$k_1 = \begin{bmatrix} -0.1157 & 0.9110 & -1.7772 & 0.5903 & 0.2541 & -0.5139 \\ 0.0442 & 0.0409 & 0.0927 & -0.2920 & -1.3021 & -0.4220 \\ 2.1423 & 0.4090 & 0.1058 & 0.1186 & -0.2134 & -0.0962 \end{bmatrix}$$

$$k_2 = \begin{bmatrix} 0.4946 & -0.8322 & -0.1561 & -0.0352 & -0.0001 & 0.0000 \\ 0.0589 & -0.6737 & -0.1258 & -0.0328 & 0.0003 & 0.0000 \\ 0.1228 & -0.3037 & -0.0480 & 0.0099 & -0.0000 & 0.0001 \end{bmatrix}$$

$$G = [g_1 \quad g_2 \quad g_3]$$

$$g_1 = \begin{bmatrix} -4.1479 & 0.1796 & 0.8985 & 0.0842 & 0.8111 \\ 0.2174 & 2.0127 & -0.3161 & 0.2192 & -0.0466 \\ 0.0979 & 0.0937 & -0.1104 & 2.2378 & 0.3612 \end{bmatrix}$$

$$g_2 = \begin{bmatrix} 4.5922 & -1.2605 & -1.3196 & -0.3184 & 0.1475 \\ -0.6667 & 0.3287 & 0.2862 & 0.1112 & -0.0148 \\ -0.0261 & 0.1980 & -0.0557 & -0.0723 & -0.0094 \end{bmatrix}$$

$$g_3 = \begin{bmatrix} -0.0117 & -0.0058 & 0.0012 & 0.0000 & -0.0001 \\ 0.0733 & 0.0030 & -0.0026 & 0.0000 & -0.0000 \\ 0.0881 & 0.0251 & 0.0257 & -0.0001 & 0.0002 \end{bmatrix}$$

4. Simulation results

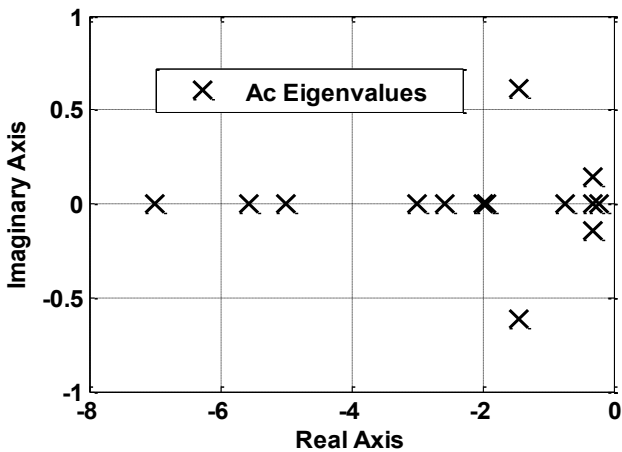


Fig. 2. Eigen-values of system Under control

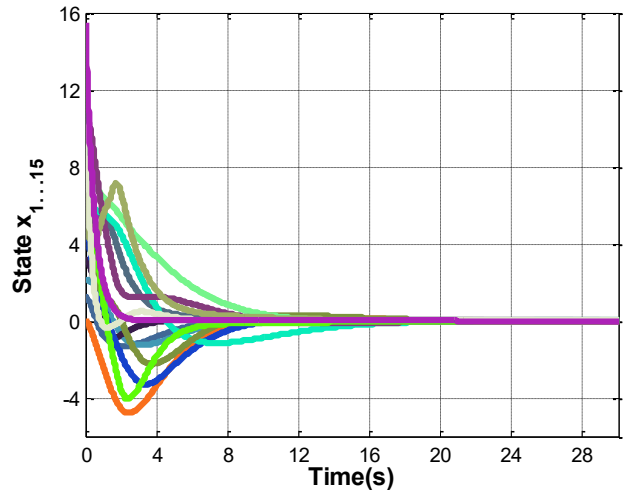


Fig. 3. The States After Stabilization

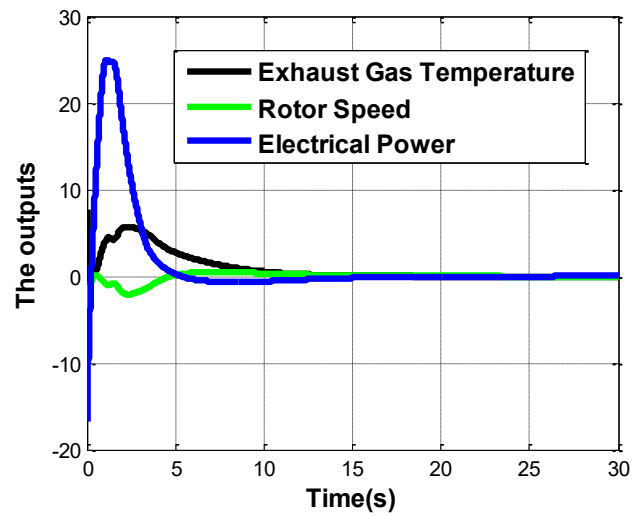


Fig. 4. The Outputs After Stabilization

Checking robustness, after Application incertitude $\Delta = 10^{-4} \times \text{magic}(15)$ in The System (Dynamic) matrix Ac and injection some noise and disturbance in states x_i , these results are obtained.

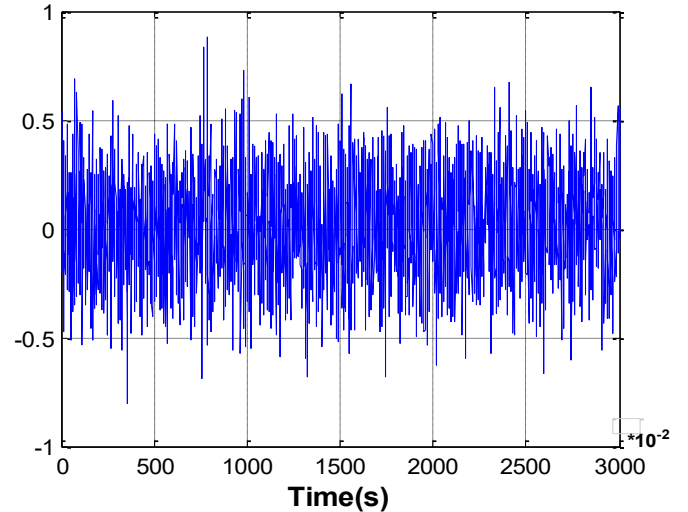


Fig. 5. Noise and Disturbance

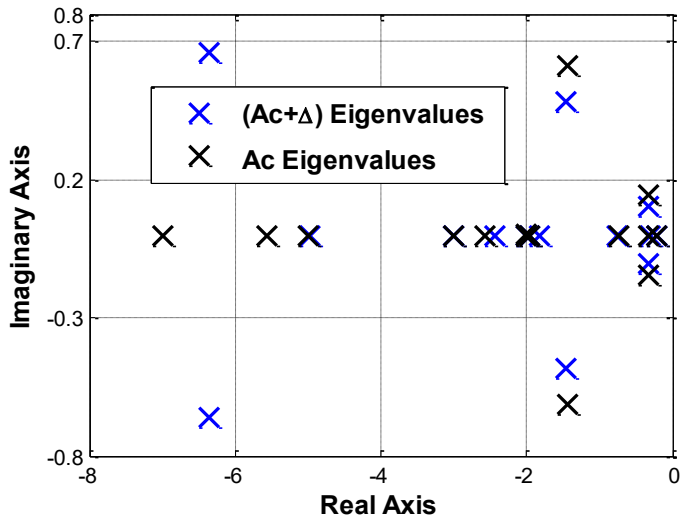


Fig. 6. Robustness test of Eigen-values

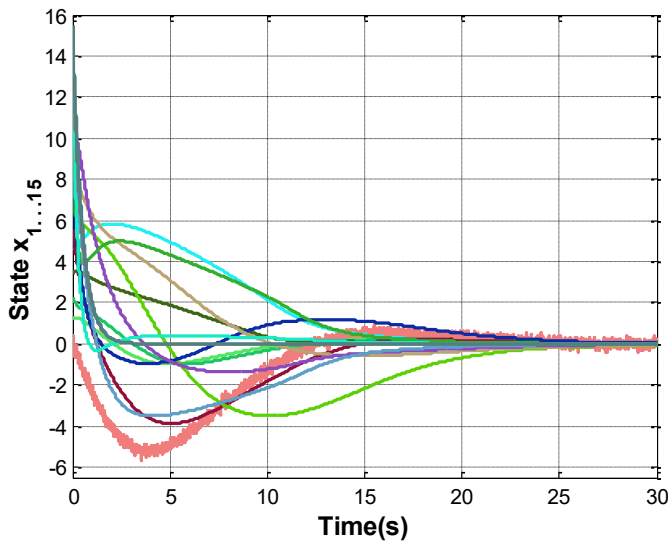


Fig. 7. The States under noise and disturbance

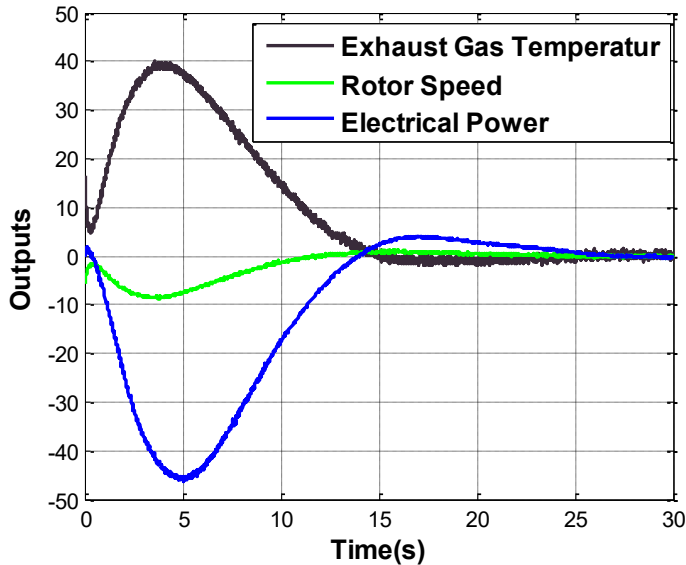


Fig. 8. The Outputs under noise and disturbance

From the obtained results, in (Fig 2) all eigenvalues are located in left half plane (negative real part). so the dynamic behavior of gas turbine

system is stabilized. Fig (3, 4), clarified the effect of the stability on all the states and the outputs that are ensuring the convergence to the origin in small time. Fig (6,7,8) showed the results under noise and disturbance, it can be seen clearly that there is no much change in the place of Eigenvalues after adding the incertitude to the origin system, the states and the outputs have been conserve their convergence towards the origin.

For the sensitivities of the eigenvalues, taking into account the distance between the origin and incertitude eigenvalues respectively, in real axis almost there is no shifting in real of eigenvalues, and from the imaginary axis there are some displacement of two eigenvalues but they do not have any effect on the stability and robustness of the system.

5. Conclusion

Based on the obtained results, it can be concluded that the application of the sliding mode control on a linear system gives best result. Whereas, a very interesting advantage can be ensured by this control which is the elimination of chattering problem due to the use of tanh function, on the same time the stability is guaranteed because the SMC is based on theory of lyapunov. On the other side, the robustness of this kind of control is observed, there is no big effect of the external noise and the uncertainties of system, this is due to the structure of this control which is independent on the system behavior.

However, it is important to clarify that there is a negative issue in this controller, which is common in all most the controllers that are linked to the feedback static matrix gain, this problem is due mainly to the large calculus and the dimension of the gain matrices K and G that are always depending on the size of the system order.

As a perspective to present work, future contributions will be oriented to the main aim of reducing the size of the matrices K and G to overcome the above mentioned drawback.

REFERENCES

1. H. K. Khalil, Nonlinear Systems, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
2. Gordon E. Carlson Signal and Linear Systems Analysis with Matlab second edition, Wiley, 1998, ISBN 0-471-12465-6.
3. M.H.Tu, C.M. Lin, Robust Stabilisation of Multivariable Feedback Systems with Desired Performance Requirement, IEE Proc. Vol. 139,N3, 1992.

4. K. Ogata, Modern Control Engineering, 5th ed, Pearson Education (New Jersey) Pvt. Ltd., USA, 2009, chapter 10-8.
5. Anderson, B.D.O. and Moore, J.B., Optimal Control: Linear Quadratic Methods, Prentice-Hall, Englewood Cliffs, NJ, 1990.
6. Doyle, J.C. and Stein, G., Multivariable feedback design: concepts for a classical/modern synthesis, IEEE Transactions on Automatic Control, Vol. AC-26, No. 1, February 1981.
7. Schultz, M.J. and Inman, D.J., Eigen-structure assignment and controller optimization for mechanical systems, IEEE Transactions on Control Systems Technology, Vol. 2, No.2, 88-100, 1994.
8. Sinha, A. and Miller, D.W., Optimal sliding mode control of a flexible spacecraft under Stochastic disturbances, AIAA Journal of Guidance, Dynamics and Control, Vol. 118, No. 3, 486-492, 1995.
9. Bekhiti Belkacem, dahimene Abdelhakim, Nail Bachir, and Hariche Kamel, The Left And Right Block Pole Placement Comparison Study : Application To Flight, Dynamics, Informatics Engineering, an International Journal (IEIJ), Vol.4, No.1, March 2016.
10. B.Nail, B.Bekhiti, K.Bdirina, A.Kouzou, and A.Hafaifa, Sliding Mode Control And Optimal GPC Algorithm for Coupled Tanks, 3rd International Conference on Control, Engineering & Information Technology CEIT'2015 Tlemcen, Algeria, 25-27 May, 2015.
11. Alok K. Sinha-Linear systems_ optimal and robust control-CRC Press (2007).
12. Utkin, V. I., Variable Structure Systems with Sliding Modes, IEEE-Trans.Automat, Contr. AC-22 No. 2 (Apr 1977), 212-222.
13. Stein, G., Formal Control System Synthesis with H_2/H_∞ Criteria, Lecture Notes for Multivariable Control Systems Course, MIT, Cambridge, MA, 1988.
14. BekhitiBelkacem, DahimeneAbdelhakim, Nail Bachir, and HARICHE Kamel, 2-DOF Block Pole Placement Control Application To : Have-Dash-Ii Missile , The International Journal of Information Technology, Modeling and Computing.
15. Panos J. Antsaklis, Anthony N. Michel, Linear system New York :McGraw-Hill, © 1997.
16. T. Kailath, Linear System, Englewood Cliffs, Prentice-Hall, NJ, 1980.
17. Chi-Tsong Chen, Linear Systems Theory and Design, Holt, Rinehart and Winston, NY, 1984.
18. P.Lancaster, Lambda-Matrices and Vibrating Systems, New York, Pergamon Press, 1966.
19. Slotine J. J. E., J. K. Hedrick and E. A. Misawa, On sliding observers for nonlinear systems, Trans. of the ASME Journal of Dynamic Systems, Measurement and Control, Vol. 109, pp. 245-252, 1987b.
20. Utkin, V. I., Variable Structure Systems with Sliding Modes, IEEE-Trans.Automat, Contr. AC-22 No. 2 (Apr 1977), 212-222.
21. J.A.Mohamed, Y.P. Kakad, Improvement of Robustness and Sensitivity Reduction for Multivariable Nonminimum Phase Systems, IEEE, 1995.
22. Magdi S. Mahmoud and Yuanqing Xia, Applied Control Systems Design, Springer-Verlag London Limited 2012.
23. Alok K. Sinha-Linear systems optimal and robust control-CRC Press (2007).
24. Ljung, L., System Identification : Theory for the user (Prentice Hall), 1999.
25. Katayama, T., Subspace methods for System Identification, Springer-Verlag, 2005.
26. Akroum, M., and Hariche K., A matrix fraction description-based identification algorithm for MIMO ARMAX models, submitted to Asian journal of control, 2008.

¹ The Applied Automation and Industrial Diagnostics Laboratory LAADI, University of DJELFA, Algeria.

² IGEE(Ex INELEC), BOUMERDES University.
 Address: Moudjbara Street BP 3117 Djelfa, Algeria.
 Phone :+213668602286
 E-mail: b.nail@univ-djelfa.dz
 E-mail: kouzouabdellah@yahoo.fr
 Email:hafaifa@hotmail.com
 Email:belkacem1988@hotmail.co.uk



RELIABILITY ANALYSIS USING WEIBULL DISTRIBUTION APPLIED TO A BOOSTER PUMP USED IN OIL DRILLING INSTALLATIONS

AHMED HAFIFA⁽¹⁾, KOUZOU ABDELLAH⁽¹⁾,
GUEMANA MOULOUD⁽²⁾, NADJI HADROUG⁽¹⁾

Abstract: *The safety exploitation in various industrial projects has become an important issue of ergonomics and economics fields. This work proposes the study of the dependability based on reliability analysis of a booster pump used in oil drilling installations. The achieved tests results, presented in this paper, improved the performance of the examined oil drilling sites and determine maintenance of this installation strategy.*

Keyword: *Reliability analyses, dependability system, reliability modelling, oil drilling installations, safety system, industrial process, booster pump.*

1. Introduction

The safe operation in various industrial installations has become an important issue in ergonomics and economics fields. This dependability is based on the lifetime of a material, which is characterized by three concepts; reliability, maintainability and availability, which can be added that security [5, 9, 14 and 17]. With the technology advancement, in these last years, industrial installations become too complex, it leads the stain dependability become more complex. Several studies have focused on reliability analysis in industrials fields based on the evolution of the failure rate, this tool is very important to characterize the behavior of equipment in different phases of their lifetime exploitation [2, 3, 19 and 24]. In this work, a booster pump used in oil drilling installation into southern Algeria project are examined, the dependability system are studied based on the proposed reliability analysis approach. This study is based on reliability modelling of the global system availability of water pumping in oil drilling, using the feeder pumps. In this investigation Weibull distribution are applied to the examined pump.

The achieved tests results, based on real lifetime data of the examined pump, allows to model many situations of wear and fatigue in the examined system. Also, this approach allows to characterize the pump system behavior, in order to improve the performance of oil drilling sites and to determine the maintenance strategy of this installation.

2. Dependability based on reliability function

The safe operation of industrial equipment is based on the lifetime of this equipment, which is characterized mainly on the reliability study, based on failure rate evaluation to characterize the behavior of equipment in different phases of their lifetime in exploitation.

The reliability of a device after t time in exploitation corresponds to the probability is that this device does not have a failure between 0 and t time in exploitation, expressed by the function $R(t)$, shown in figure 1.

$$R(t) = P(T \geq t) = 1 - F(t) \quad (1)$$

Where $F(t)$ is the distribution function of the variable t .

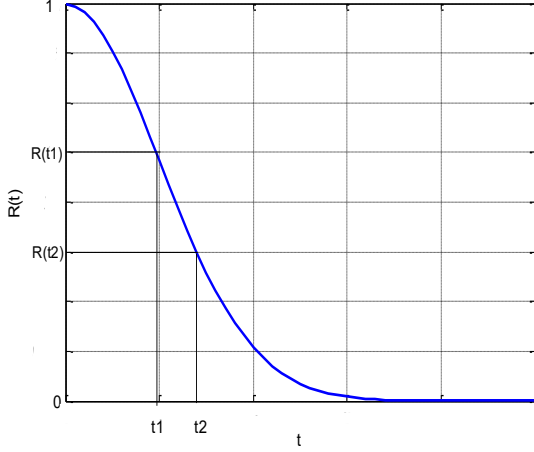


Fig. 1. Reliability function

In the reliability study the main indicator is the failure rate $\lambda(t)$ which is the number of failures per cell and per time unit [4, 6], this failure rate is high at the beginning of the lifetime of the device. Then it decreases quite rapidly with time (decreasing failure rate), this phase of life is called period of youth. After it stabilizes at a value we want as low as possible during a period called life period (constant failure rate). In the end, it goes back when the wear and aging effects are felt, it is the aging period (increasing failure rate). The failure rate $\lambda(t)$ is given as follows:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \left(\frac{1}{\Delta t} \cdot \frac{R(t) - R(t - \Delta t)}{R(t)} \right) \quad (2)$$

$$\lambda(t) = -\frac{dR}{dt} \cdot \frac{1}{R(t)} = \frac{dF(t)}{dt} = \frac{f(t)}{R(t)} \quad (3)$$

The failure rate $\lambda(t)$ knowledge is sufficient to determine the reliability using the following equation:

$$R(t) = \exp\left(-\int_0^t \lambda(t) dt\right) \quad (4)$$

In the most parts of industrial applications in reliability field, using the mean time between failure (MTBF), which is defined as the expected by T value, given as follows:

$$MTBF = E(t) = \int R(t) dt \quad (5)$$

This work examine an industrial pump used in oil drilling installations into southern Algeria site, the study is based on reliability and availability calculations of the global pump system using Weibull distribution with three parameters [1, 16, 20 and 24]. The probability density is given by the flowing expression [3, 15 and 22]:

$$f(t) = \frac{\beta}{\eta} \left(\frac{t-\gamma}{\eta} \right)^{\beta-1} e^{-\left(\frac{t-\gamma}{\eta}\right)^\beta}; \quad t \geq \gamma \quad (6)$$

where $\beta > 0$ is the shape parameter, $\eta > 0$ is the scale parameter and $\gamma > 0$ is the position parameter.

The reliability function is given by the equation (7) and the failure rate is given by equation (8).

$$R(t) = e^{-\left(\frac{t-\gamma}{\eta}\right)^\beta} \quad (7)$$

$$\lambda(t) = \frac{\beta}{\eta} \left(\frac{t-\gamma}{\eta} \right)^{\beta-1} e^{-\left(\frac{t-\gamma}{\eta}\right)^\beta} \quad (8)$$

$$f(t) = \frac{\beta}{\eta} \left(\frac{t-\gamma}{\eta} \right)^{\beta-1} e^{-\left(\frac{t-\gamma}{\eta}\right)^\beta}; \quad t \geq \gamma$$

In the equations (7, 8) η is the real scale parameter, when η is not equal to unity, a simple change of scale necessary. The reliability law is given by the relation (9), by derivative of this distribution the probability density function is given by the equation (10).

$$R(t) = e^{-\left(\frac{t}{\eta}\right)^\beta} \quad (9)$$

$$f(t) = F'(t) = -R'(t)$$

$$f(t) = \left[1 - e^{-\left(\frac{t}{\eta}\right)^\beta} \right]' = - \left[e^{-\left(\frac{t}{\eta}\right)^\beta} \right]' \quad (10)$$

The derivative of the formula (10) gives us the following:

$$f(t) = - \left[- \left[\left(\frac{t}{\eta} \right)^\beta \right]' e^{-\left(\frac{t}{\eta}\right)^\beta} \right]$$

$$= \left[\left(\frac{t}{\eta} \right)^\beta \right]' e^{-\left(\frac{t}{\eta}\right)^\beta} \quad (11)$$

$$f(t) = \beta \left(\frac{t}{\eta} \right)' \left(\frac{t}{\eta} \right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^\beta}$$

$$= \beta \frac{1}{\eta} \left(\frac{t}{\eta} \right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^\beta}$$

Finally, the probability density function is given by the relation:

$$f(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^\beta} \quad (12)$$

In take accounts the equations (10, 11, and 12), the failure rate is given by the equation (13), is the same to the instantaneous failure rate.

$$\lambda(t) = \frac{f(t)}{R(t)} = \frac{f(t)}{1-F(t)}$$

$$\lambda(t) = \frac{f(t)}{1-F(t)} = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} \cdot e^{-\left(\frac{t}{\eta}\right)^\beta} \cdot \frac{1}{e^{-\left(\frac{t}{\eta}\right)^\beta}} \quad (13)$$

$$\lambda(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1}$$

2.1. Reliability Data Processing

The data reliability studies usually come from historical failure records of exploited system or from test results of the investigated system. In all cases, the time between failures TBF is calculated and classified in ascending order of the number recorded N of the time between failures in sample size. The time between failures is grouped peer class, if $N > 50$ with the cumulative frequency given by [13, 21 and 22]:

$$F(i) = \frac{\sum n_i}{N} \quad (14)$$

where $F(i)$ is very similar to Weibull distribution $f(t)$.

The approximation formula of the average ranks given by equation (15) will be used in these work, where $20 < N < 50$.

$$F(i) = \frac{\sum n_i}{N+1} \quad (15)$$

In the case of $N < 20$, the approximation formula of median ranks given by the following equation in (16) will be applied for the examined pump [7, 8, 18 and 23].

$$F(i) = \frac{\sum n_i - 0,3}{N+0,4} \quad (16)$$

In order to improve the maintainability of the examined system by minimization of the time average techniques to repair (MTTR), given by $MTTR = \frac{\sum TTR}{N} = \frac{\sum \text{Intervention Time}}{\text{Number of failures}}$. There is an analogy between the statistical study of the

reliability and maintainability of the examined system, explained by the relationship between the variables reliability expressed by the probability density $g(t)$ and the duration of the reparation action $m(t)$ expressed by the probability of repairing in duration $TTR < t$ in the equation (17), with the repair rate $\mu(t)$ is given by the equation (18), [10, 11 and 12].

$$M(t) = \int_0^t g(t) dt \quad (17)$$

$$\mu(t) = \frac{g(t)}{1-M(t)} \quad (18)$$

In this examined case, to obtained a good availability, it must have the least possible downtime and quickly be repaired if it breaks down. The availability rate is steady as follows:

$$D = \frac{\mu(t)}{\mu(t) + \lambda(t)} \quad (19)$$

3. Application results

In this part of the present work, booster pumps used in oil drilling installations are examined. The real data of the uptime and duration breakdowns are grouped and sorted in ascending order, from historical records of the studied pump, during the operating period of 2000-2010, in the following table.

Table 1. Reliability data of the studied pump

Rang	TBF (h)	TTR (h)	F(t) (%)	R (t) (%)	$\lambda(t) 10^{-4}$
1	720	96	8,33	91,67	0,95
2	2160	108	20,23	79,77	1,12
3	2880	114	32,14	67,86	1,17
4	5760	48	44,04	55,96	1,30
5	6480	87	55,95	44,05	1,32
6	7200	56	67,85	32,15	1,35
7	16700	24	79,76	20,24	1,53
8	21600	67	91,66	8,34	1,59

The mean time between failures using Weibull distribution applied to the studied booster pump is 7925,7576 hours. The cumulative failure and the hazard function of the examined system, using Weibull distribution with tow parameters is shown respectively in the figures 2, 3 and the reliability function and Weibull probability paper in the figure 4 and 5.

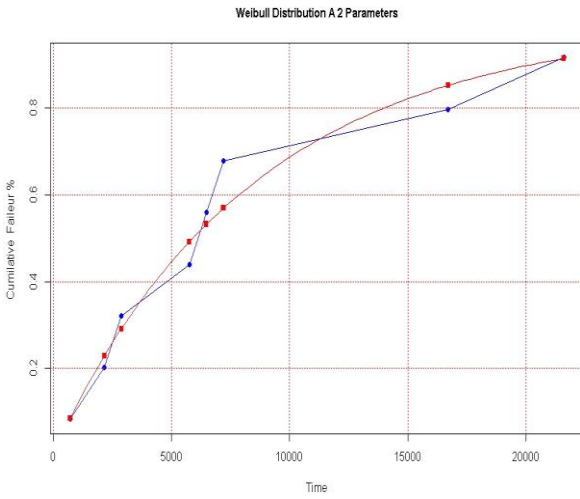


Fig. 2. Cumulative failure using Weibull distribution with tow parameters

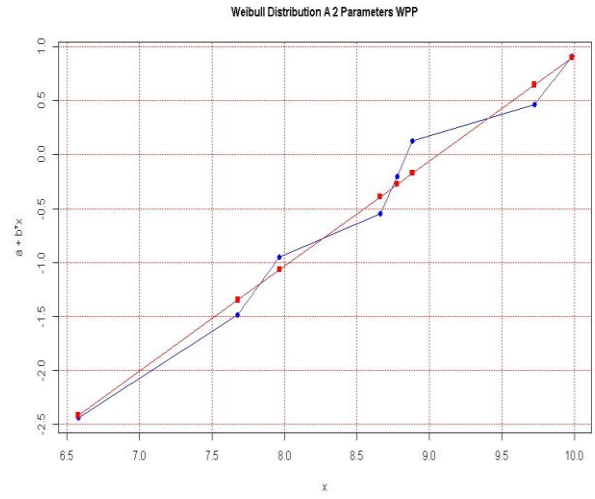


Fig. 5. Weibull probability paper using Weibull distribution with tow parameters

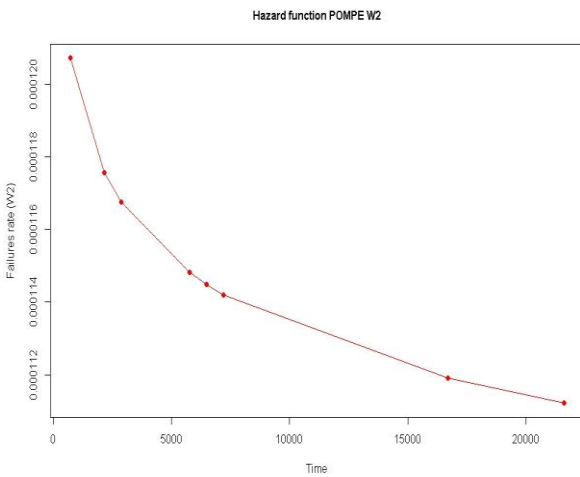


Fig. 3. Hazard function using Weibull distribution with tow parameters

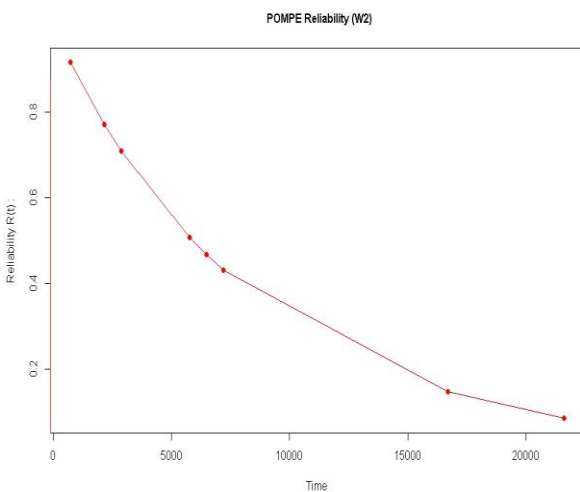


Fig. 4. Reliability function using Weibull distribution with tow parameters

The reliability curve decreases with time, which is normal and this is due to wear and failures are random. According to the graph the failure rate is strongly increasing at the beginning and then more slowly. Other validation tests on the examined pump were achieved using Weibull distribution with three parameters; this confirms the effectiveness of the proposed approach in this work.

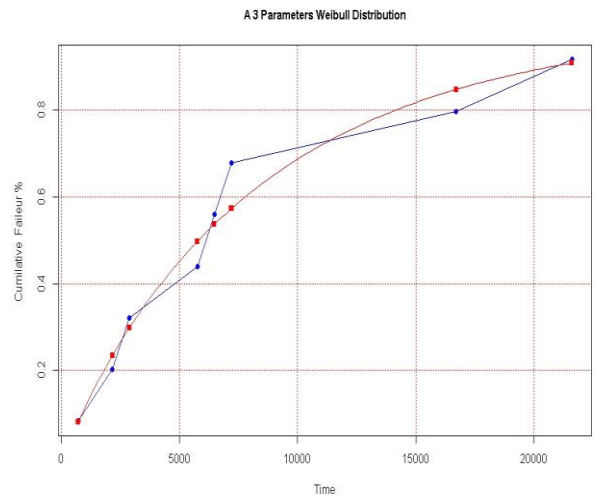


Fig. 6. Cumulative failure using Weibull distribution with tow parameters

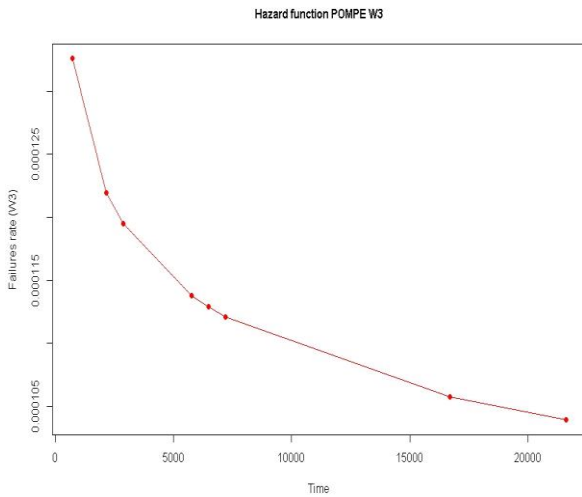


Fig. 7. Hazard function using Weibull distribution with two parameters

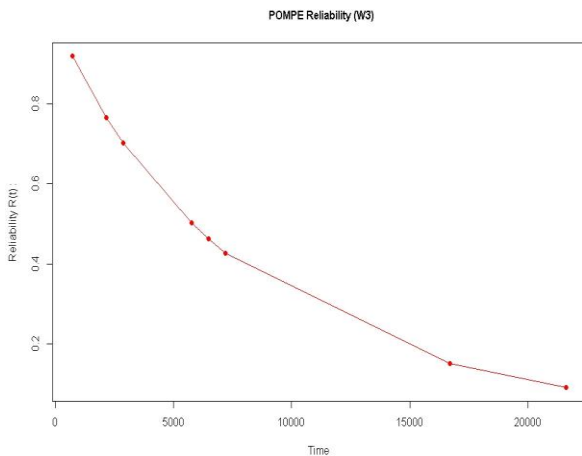


Fig. 8. Reliability function using Weibull distribution with two parameters

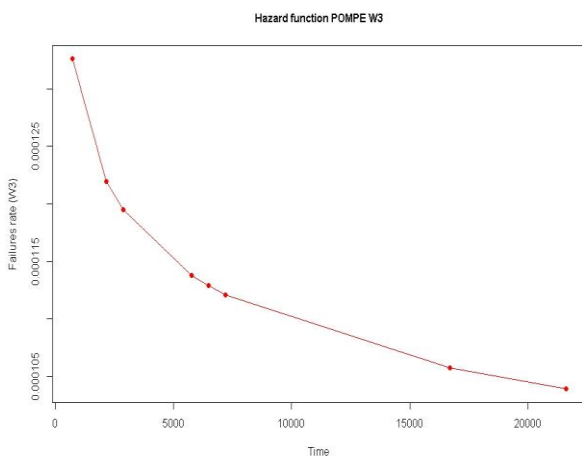


Fig. 9. Weibull probability paper using Weibull distribution with two parameters

On this basis, a reliability study was conducted; the average uptime was calculated on the basis of historical records. We note that the availability is high, as the booster pump. The calculating the reliability parameter using Weibull distribution with two and three parameters, figure 10 shows that the booster pump out of the period of the useful life and enters the period of old age.

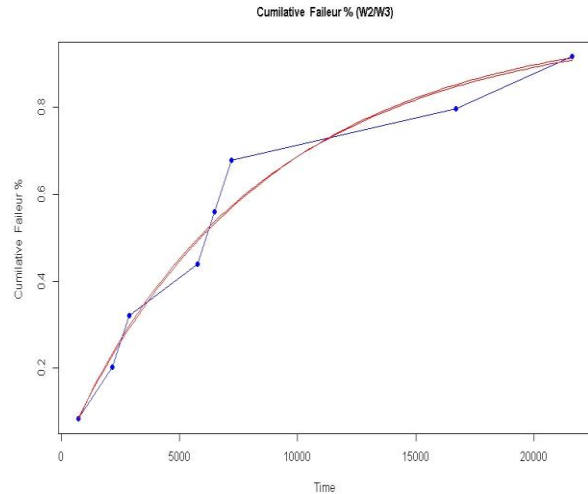


Fig. 10. Cumulative failure with Weibull distribution with two and three parameters

4. Conclusion

The reliability analysis is a very important tool to characterize the behavior of industrial systems in different phases of time life. In this paper, a booster pump used in oil drilling installations are examined, using Weibull distribution with two and three parameters to their reliability analysis. The achieved test results, improve the performance of the examined oil drilling sites and determine maintenance of this installation strategy.

The obtained results in this work, has allowed us to analyze the reliability of the examined pump. That allows us to take technical decisions to determine their intervention program and assess maintenance and consequently the degree of confidence in exploitation of this system. Indeed, the studied operational reliability in this work is based on the experience feedback, given the time guarantee measure and decrypts the lifetime of the examined booster pump. This allowed us a better exploitation of the pump in order to optimize their operations and meet production.

References

1. Asghar Moeini, Kouroush Jenab, Mohsen Mohammadi, Mehdi Foumani, (2013). Fitting the three-parameter Weibull distribution with

- Cross Entropy. *Applied Mathematical Modelling*, vol. 37, no. 9, pp. 6354-6363.
2. Chiacchio F., Cacioppo M., D'Urso D., Manno G., Trapani N., Compagno L., (2013). A Weibull-based compositional approach for hierarchical dynamic fault trees. *Reliability Engineering & System Safety*, vol. 109, pp. 45-52.
 3. Djamel Halimi, Ahmed Hafaifa and Elahmoune Bouali, (2014). Maintenance actions planning in industrial centrifugal compressor based on failure analysis. *The quarterly Journal of Maintenance and Reliability*, vol. 16, no. 1, pp. 17-21.
 4. Emad E. Elmahdy, Abdallah W. Aboutahoun, (2013). A new approach for parameter estimation of finite Weibull mixture distributions for reliability modeling. *Applied Mathematical Modelling*, vol. 37, no. 4, pp. 1800-1810.
 5. Emilio Sarno, Vipin Kumar, Wei Li, (2005). A hybrid methodology for enhancing reliability of large systems in conceptual design and its application to the design of a multiphase flow station. *Research in Engineering Design*, vol. 16, no. 1-2, pp. 27-41.
 6. Govind S. Mudholkar, Kobby O. Asubonteng, Alan D. Hutson, (2009). Transformation of the bathtub failure rate data in reliability for using Weibull-model analysis. *Statistical Methodology*, vol. 6, no. 6, pp. 622-633.
 7. Hafaifa Ahmed, Belhadeef Rachid, Guemana Mouloud, Reliability model exploitation in industrial system maintainability using expert system evaluation. Proceeding of the 4th International Conference on Integrity, Reliability and Failure IRF2013, 23-27 June 2013 in Funchal, Madeira, Portugal.
 8. Hafaifa Ahmed, Mouloud Guemana and Attia Daoudi, Spectral analysis approach applied to the vibrations detection and isolation in centrifugal pump used in petroleum industry. Proceeding of the 5th International Congress on Technical Diagnostics, September 2012, Univeristy of Science and Technology, Kraków, Poland.
 9. Hafaifa Ahmed, Ferhat Laaouad and Moulod Guemana, (2009). A New engineering method for fuzzy reliability analysis of surge control in centrifugal compressor. *American Journal of Engineering and Applied Sciences*, vol. 02 no. 04, pp. 676-682.
 10. Hafaifa Ahmed, Ferhat Laaouad and Kouider Laroussi, A new engineering method for fuzzy reliability analysis of surge control in centrifugal compressor. Proceeding of the 2nd International Conference on Control, Instrumentation and Mechatronic Engineering (CIM2009), Melaka, Malaysia. 2-3 June 2009.
 11. Isabel Sartori, Edilson M. de Assis, Adilton L. da Silva, Rosana L.F. Vieira de Melo, Ernesto P. Borges, e Silvio A.B. Vieira de Melo, (2009). Reliability Modeling of a Natural Gas Recovery Plant Using q-Weibull Distribution. *Computer Aided Chemical Engineering*, vol. 27, pp. 1797-1802.
 12. Kucheryavyi V. I., Mil'kov S. N., (2009). Statistical modeling of the lifetime of face seals in an oil-transfer pump. *Journal of Machinery Manufacture and Reliability*, vol. 38, no. 1, pp 94-99.
 13. Li Zhengmei, Zhou Qiong, An Qi, Tang Jianping, (2011). Influences of eccentric unbalances on loads and life of auto water pump bearing. *Mechanism and Machine Theory*, vol. 46, no. 3, pp. 253-263.
 14. Mansour Aghababaei Jazi, Chin-Diew Lai, Mohammad Hossein Alamatsaz, (2010). A discrete inverse Weibull distribution and estimation of its parameters. *Statistical Methodology*, vol. 7, no. 2, pp. 121-132.
 15. Mokronosov E. D., Bogdanov V. V. and Eltyshev M. N., (2007). Increase in the operating reliability of a well sucker-rod pump. *Chemical and Petroleum Engineering*, vol. 43, no. 11-12, pp. 756-758.
 16. Mouloud Guemana, Slimane Aissani and Ahmed Hafaifa, (2011). Use a new calibration method for gas pipelines: An advanced method improves calibrating orifice flowmeters while reducing maintenance costs. *Hydrocarbon Processing Journal*, vol. 90, no. 8, pp. 63-68.
 17. Qing Zhang, Cheng Hua, Guanghua Xu, (2014). A mixture Weibull proportional hazard model for mechanical system failure prediction utilising lifetime and monitoring data. *Mechanical Systems and Signal Processing*, vol. 43, no. 1-2, 3, pp. 103-112.
 18. Raif Sakin, İrfan Ay, (2008). Statistical analysis of bending fatigue life data using Weibull distribution in glass-fiber reinforced polyester composites. *Materials & Design*, vol. 29, no. 6, pp. 1170-1181.
 19. Saad J. Almalki, Saralees Nadarajah, (2014). Modifications of the Weibull distribution: A review. *Reliability Engineering & System Safety*, vol. 124, pp. 32-55.
 20. Saad J. Almalki, Jingsong Yuan, (2013). A new modified Weibull distribution. *Reliability Engineering & System Safety*, vol. 111, pp. 164-170.
 21. Schijve J., (2005). Statistical distribution functions and fatigue of structures.

- International Journal of Fatigue*, vol. 27, no. 9, pp. 1031-1039.
22. Tieling Zhang, Min Xie, (2011). On the upper truncated Weibull distribution and its reliability implications. *Reliability Engineering & System Safety*, vol. 96, no. 1, pp. 194-200.
 23. Vitkovskii V., Kirillov I. R., Chaika P. Yu., Kryuchkov E. A., Poplavskii V. M., Nosov Yu. V., Oshkanov N. N., (2007). Performance-based reliability assessment of electromagnetic pumps. *Atomic Energy*, vol. 102, no. 2, pp. 122-128.
 24. Zhejun Gong, (2006). Estimation of mixed Weibull distribution parameters using the SCEM-UA algorithm: Application and comparison with MLE in automotive reliability analysis. *Reliability Engineering & System Safety*, vol. 91, no. 8, pp. 915-922.

Authors' contacts

**Ahmed Hafaifa⁽¹⁾, Kouzou Abdellah⁽¹⁾,
Mouloud Guemana⁽²⁾ and Nadji
Hadroug⁽¹⁾**

⁽¹⁾ Applied Automation and Industrial Diagnostics Laboratory, Faculty of Science and Technology, University of Djelfa 17000 DZ, Algeria

⁽²⁾ Faculty of Science and Technology, University of Medea, Algeria

Ahmed Hafaifa⁽¹⁾

E-mail: hafaifa.ahmed.dz@iecc.org

Kouzou Abdellah⁽¹⁾

E-mail: kouzouabdellah@iecc.org

Mouloud Guemana⁽²⁾

E-mail: guemana_m@yahoo.fr

Nadji Hadroug⁽¹⁾

E-mail: N_Hadroug@univ-djelfa.dz



ICT TECHNOLOGIES IN WILDLAND FIRES MODELLING FOR TEST CASES (BULGARIA)

NINA DOBRINKOVA

Abstract: *In this paper we will provide an overview of the used ICT tools for wildland fire propagation on the Bulgarian territory from a team working in IICT-BAS. Most of the data is received by the forest departments in Zlatograd and Harmanly where the main simulated wildland fires are located. The summary of the results is giving the conclusion that the fires in Bulgaria need more detailed burning materials classification, but with the ICT tools used the achieved results are quite promising.*

Key words: *wildland fires, simulations with ICT tools for wildland fires, Harmanly test case, Zlatograd test case*

1. Introduction

Due to climate change and very fast developing of the modern society forest fires are growing problem for the southern countries of Europe. Statistics concerning the development of forest fires in the last years are published in reports by the centres in the EU which monitor the forest conditions [1]. Observations in the member states of EU in the period 1980–2005 showed an increasing number of areas affected by the fires. Bulgaria was not included in the study because it was not a member of EU until 2007, but independent research done on the topic describes an increase of wildland fires since 1995 on the Bulgarian territory until now [2, 3]. The problem is not only in Bulgaria, but it is well known worldwide and research teams in the US have begun their research in the field in the beginning of last century to investigate the tools and methods to simulate wildland fire propagation. The Missoula Fire Laboratory in Rocky Mountains is the pioneer in the research with its scientist Richard Rothermel [4]. In fact, Rothermel and his team in Missoula are considered as founders of fire modelling research. The first working wildland fire behaviour model was created by them in 1972 and since then it has been improved several times. Every subsequent model's modifications included consideration of further parameters that could make the simulations obtained by scientists more realistic and accurate.

In the early 80s of the last century, M. Grishin, professor at the University of Tomsk, Russia, has worked and developed a model [5] which uses data on the types of burning materials in the taiga (mostly conifers) and takes into account

that the main combustion intensity happens in the crowns of trees.

Approximately at the same time attempts at wildland simulations were made in laboratories in Sydney, Australia. These models did not become as popular as the US ones, because of some computational issues at that time, concerning PC power, data coverage, etc. Nowadays many laboratories are involved in modelling of wildland fire behavior propagation. It is easier to find such teams in the affected zones, simply because this knowledge saves time and efforts to the incident commanders on the field.

The development of modern information and communication tools allows application of cutting-edge technologies to solve problems related to the forest and field fires. The use of these tools not only allows early detection of fires but also enables the prediction of the dynamics related to fire spread of the fire line and the extent of the possible damages of the environment and citizens' property.

The main types of wildland fires described in the specialized literature are classified as follows: type 1—surface fires [4,6]; type 2—crown fires [5,7]; type 3—spotting fires as modified crown fires and type 4—fire acceleration, when the terrain has steep slopes.

In our paper we will present how through different ICT tools the knowledge about local Digital Elevation Model (DEM), land cover (burning materials) and meteorological conditions can give good enough predictions for fire spread according to the different scenarios elaborated by the fire analysts working with the data.

2. First simulations with Harmanli test site

The first elaborated data was for a test case of a fire near by a village next to Harmanli town. The data collected there was processed by Geographical Information Systems (GIS) in order to have visual representation for the final burning scars of the simulated fires. That was needed for calibration purposes. The used technology was based on the parallel computations where semi-empirical approach for fire spread implies that the fire spreads toward the normal to the front line, which is expressed by a modified Rothermel's formula. The burning area for time t has been presented as area Ω , which is a set of points in the (x, y) plane. Let:

$$\tilde{S} = \min \{B_0, R_0 + \phi_w + \phi_s\}, \quad (1)$$

where B_0 is the fire spread against the wind direction, R_0 is the fire spread in absence of wind, $\phi_w = a (\bar{v} \cdot \bar{n})^b$ is the wind correction and $\phi_s = d \nabla z \cdot \bar{n}$ is the terrain correction, \bar{v} is the wind, ∇z is a terrain component of the normal \bar{n} of the fire line, a , b and d are certain constants. In this case the WRF-Fire model (renamed in 2011 SFire) describes the maximal fire spread with the formula:

$$S = \begin{cases} 0, & \text{ako } \tilde{S} < 0 \\ S_{\max}, & \text{ako } \tilde{S} > S_{\max} \\ \tilde{S}, & \text{ako } 0 \leq \tilde{S} \leq S_{\max} \end{cases}, \quad (2)$$

where S_{\max} is the maximal spread of the fire. After the ignition of the combustible materials its amount of combustible material in point (x, y) decreases exponentially and is given by the formula:

$$F(x, y, t) = F_0(x, y) e^{-(t-t_i)/(x,y)/W(x,y)}, \quad (3)$$

where t is the time, t_i is the ignition time, F_0 is the initial quantity of the burning materials (before their ignition) and $W(x,y)$ does not depend on the time, but on the burning materials.

Heat flow released from the fire is presented in the atmospheric model as a layer above the ground deployed in height [8]. The flow depends on the amount of material burned and can be presented by the formula:

$$\Phi = -A(x, y) \frac{\partial}{\partial t} F(x, y, t). \quad (4)$$

This representation is needed because the atmospheric model from WRF does not support limits on heat flow. The coefficients and functions B_0 , R_0 , S_{\max} , a , b , d , W and A are determined by laboratory experiments.

For each point in the plane the coefficients of combustible materials are given as representatives of one of Anderson's 13 categories [9], valid for US burning materials and applied to Bulgarian conditions. Different altitude values for the wind are also used. WRF-Fire has internal representation for each of the categories, which gives options for determining when the simulated area is outside US. WRF-Fire uses a different modelling of the spread of fire through so called level-set functions [10]. In this approach a function $\psi = \psi(x, y, t)$ is set, which defines subareas of Ω by the rule:

$$\Omega(t) = \{(x, y) \in \Omega : \psi(x, y, t) < 0\}. \quad (5)$$

These areas are considered to be burned, and the line of fire is set by the curve:

$$\Gamma(t) = \{(x, y) \in \Omega : \psi(x, y, t) = 0\}. \quad (6)$$

The function $\psi(x, y, t)$ satisfies the equation:

$$\frac{\partial \psi}{\partial t} + S(x, y) |\nabla \psi| = 0, \quad (7)$$

which can be solved numerically.

The formulas (1–7) are sufficient for the general description of the simulation of the development of fire. At the beginning the atmospheric model interpolates the wind to match the size of the smaller area of the fire. Then a numerical method is applied for evaluating the level set function. The next step applies quadrature formulas for evaluation of the amount of the burning material. Simultaneously the heat flow released to the atmosphere is estimated. The last usually forces changes in the weather and therefore the steps need to be repeated again.

WRF-Fire is tested for first time with real wildland fire in Bulgaria. The fire happened near Leshnikovo village (Harmanli) in August 14–17, 2009. Unfortunately, unlike in the US, not all of the necessary data is available for free in Bulgaria. Thus we develop and present procedures for manual production of some data needed for the land cover and land use by applying GIS tools on available orthophoto pictures for the area of interest. The method

uses information about the terrain and the meteorological data (for the corresponding period) as an input file. The most successful simulation was done on the supercomputer at the University of Denver by distant connection. In Table 1 the simulation results are presented according to the number of the cores used.

Table 1: The time required for the simulation presented in seconds depending on the number of processors running the parallel execution of processes showing that in 120 cores the simulations run as real time. Everything above is faster than real fire propagation.

Cores	6	12	24	36	60	120	240	360	480	720	960	1200
Fire line propagation in km.	1.91	1.08	0.50	0.34	0.22	0.13	0.08	0.06	0.06	0.04	0.10	0.04
Region 1	6.76	7.05	2.90	2.06	1.20	0.73	0.45	0.32	0.26	0.23	0.24	0.17
Region 2	0.00	0.00	0.00	0.02	0.02	0.04	0.04	0.06	0.06	0.08	0.07	0.15
Total sec. which is the coeff. for real time	10.59	9.21	3.91	2.75	1.64	0.99	0.61	0.44	0.37	0.31	0.44	0.26

With this first simulations of our team in the test site nearby Harmanli town has been elaborated a methodology for collection, procession and implementation of real data for test sites on Bulgarian territory. The selected model was having as input meteorological data, DEM and only 13 burning classes which led to the idea that we can experiment also with different models like BEHAVE Plus and FARSITE for our next tests.

3. BEHAVE Plus and FARSITE simulations for the test sites in Zlatograd forestry department

In the framework of bilateral cooperation program between Greece and Bulgaria 2007-2013 our team was having the opportunity to work in the Zlatograd forestry department. The study area was the territorial state-owned forestry department in Zlatograd, which covers an area of 33,532 ha, where 31,856 ha are state forests. Most forests are in early to mid-seral successional stages, with only small amounts of mature to old forest. Stand age is highly variable, ranging from 20 to 80 yrs; most stands range between 35 to 50 yrs with the average being 46 yrs. Average stem stock is 140 m³ ha⁻¹. The average forest canopy cover is 81%.

In terms of climate, the region is part of the continental-Mediterranean climatic region, south-Bulgarian climatic sub-region and East Rodopi mountain low climate region. The average annual temperature is 10.8°C, with a maximum temperature

in July of 20.6°C and minimum temperature in January of -0.8°C, indicating moderate summers and relatively mild winters. Extreme values of annual average maximum and minimum temperature are respectively 17.1°C and 4.9°C, the monthly maximum is in August (28.9°C) and the average monthly minimum occurs in January (-3.9°C). Average annual rainfall reaches 1000 mm. Maximum precipitation amounts for the period from April to October range from 10.0 mm for 5 min to 46.3 mm for 60 min and 59.7 mm for more than 60 min. The average annual relative humidity is 75% which is an indication of good growing conditions; maximum relative humidity values of 85% occur in November. However, approximately 13-15 days per year have relative humidity less than or equal to 30%, during which time wildfires may be of higher concern.

The data we were working on was about fifteen wildfires that occurred in 2011 to 2012 within the Zlatograd municipal territory were provided by the Zlatograd forestry department; this data included vegetation type, area burned (in decares where 10 decares = 1 hectare), date, and start and end hours of the fire event (Table 2). These wildfires burned in a variety of vegetation types and were more than likely started by humans to clear agricultural debris or prepare fields, based on the proximity to villages. Paper maps from the forestry department identified the ignition location and final fire shape; this data was digitized in a GIS which allowed each ignition point to be viewed with background orthophotos and the spatial Zlatograd vegetation classification showing pre-fire vegetation (Table 2)

Table 2: Fire information provided by the Zlatograd Forestry Department for the period 2011-2012

Fire No.	Vegetation type	Burned area in decares	Date of occurrence	Hour of start	Hour of end
1	Durmast	3.0	25 March 2012	1330	1530
2	Beechwood	5.0	29 March 2012	1400	1800
3	Scotch pine	1.0	16 June 2012	1500	1700
4	Scotch pine	7.0	6Aug. 2012	1640	1950
5	Scotch pine	5.0	6 Aug. 2012	1710	2130
6	European black pine	4.0	27Aug. 2012	1200	1600
7	Scotch pine	3.0	5 Sept. 2012	1400	2030
8	Scotch pine	6.0	6 Sept. 2012	1400	1930
9	Scotch pine	2.0	6 Oct. 2012	1600	2320
10	Scotch pine	1.0	16 March 2011	1310	1400
11	Scotch pine	1.0	5 April 2011	1715	1900
12	Scotch pine	1.0	10 April 2011	1130	1530
13	Grassland	3.0	30 Aug. 2011	1400	1800
14	Scotch pine	4.0	12 Sept. 2011	1230	1900
15	Scotch pine	1.0	15 Sept. 2011	1600	1830

After we collected and located the forest fires we did runs with BehavePlus point based prediction system in order to analyze fire growth and behavior for homogeneous vegetation with static weather data. We used standard fuel models developed for US and we evaluated which fuel models were best able to produce estimates of fire behavior and growth in BehavePlus similar to those observed on each of the fifteen fires. In addition to fuel model, BehavePlus requires inputs for weather, fuel moisture, slope, and duration of the burning period. We obtained weather data for each fire from TV Met, a private company in Bulgaria, which provided the ability to calculate fine dead fuel moisture values. Due to the paucity of available weather data in Bulgaria, we had to assume that weather recorded for the weather station closest to each particular fire is consistent with weather experienced on the wildfire. We estimated live herbaceous and live woody fuel moisture values based on the expected phenological stage for the time of year that the fire occurred. To estimate slope, we first acquired a 30 m resolution digital elevation model (DEM) from the National Institute of Geophysics, Geodesy, and Geography in Bulgaria, then subsequently calculated the average slope for each fire using standard geospatial processing in ArcGIS (ESRI 2010). Burn period length for each fire was obtained from the Zlatograd forestry department data (Table 2).

Based on initial BehavePlus results using standard fuel models, custom fuel models were developed for some vegetation types not well represented by the US fuel models. Custom fuel models were developed for native durmast oak and grass as well as one of the Scotch pine sites by modifying fuel loading parameters to better match local vegetation and reflect the lack of woody debris in the understory.

Following evaluation of fuel models with BehavePlus, we then performed analyses in FARSITE, a spatial fire growth system that integrates fire spread models with a suite of spatial data and tabular weather, wind and fuel moisture data to project fire growth and behavior across a landscape. We defined our test landscapes using a 500 m buffer zone around each of the fifteen Zlatograd fires.

Input for FARSITE consists of spatial topographic, vegetation, and fuels parameters compiled into a multi-layered "landscape file" format. Topographic data required to run FARSITE include elevation, slope, and aspect. Using the aforementioned 30 m DEM, we calculated an aspect layer, and then clipped elevation, aspect, and slope rasters to the extent of our fifteen test landscapes. Required vegetation data include fuel model and

canopy cover. Fuel models within the 500 m buffered analysis area for each individual fire were assigned based on our BehavePlus analyses; fuel model assignments were tied to the dominant vegetation for each polygon based on the Zlatograd forestry department's vegetation data. Canopy cover values were visually estimated from orthophoto images and verified with stand data from the Zlatograd forestry department. Additional canopy variables (canopy base height, canopy bulk density, and canopy height) that may be included in the landscape file were omitted, as these variables are most important for calculating crown fire spread or the potential for a surface fire to transition to a crown fire. None of the fifteen fires analyzed experienced crown fire.

Tabular weather and wind files for FARSITE were compiled using the weather and wind data from TV Met, Bulgarian meteorological company that included hourly records. Tabular fuel moisture files were created using the fine dead fuel moisture values calculated for the BehavePlus analyses for 1-hr timelag fuels. The 10-hr fuel moisture value was estimated by adding 1% to the 1-hr fuel moisture and the 100-hr fuel moisture was generally calculated by adding 3% to the 1-hr fuel moisture. The live fuel moisture values previously estimated for BehavePlus analyses were used to populate live herbaceous and live woody moisture values.

All simulations performed in FARSITE used metric data for inputs and outputs. An adjustment value was not used to alter rate of spread for standard fuel models, rather custom fuel models were created. Crown fire, embers from torching trees, and growth from spot fires were not enabled.

As an example of one of our successful FARSITE runs, we present the results from a single wildfire that burned in grassland vegetation, for which we developed custom fuel models. This fire occurred on August 30, 2011, starting at 1400 and ending around 1800, and burned a total area of 0.3 ha. We used the following input parameters to model this small grassland fire in FARSITE:

Fuel moisture values: 6% (1-hr), 7% (10-hr), 9% (100-hr), 45% (live herbaceous), and 75% (live woody);

Daily maximum temperatures: 17-21°C;

Daily minimum relative humidity: 24-50%;

Winds: generally from the west-southwest at 1-2 k h⁻¹

The fire size as calculated using FARSITE was 0.5 ha, which seems reasonable considering the modeled size would not have included the suppression actions that most likely occurred given the close proximity of a village to this fire figure 1.

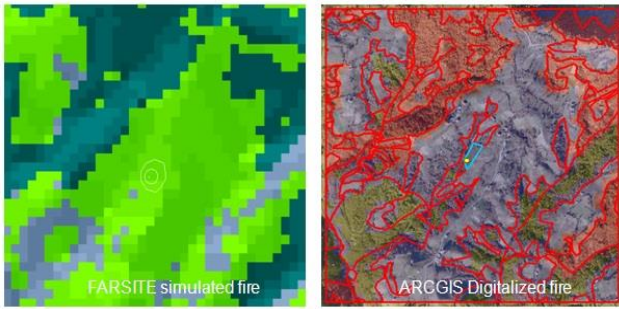


Figure 1: *FARSITE* run for a grassland fire, where size of the fire is very close to the real one, but the shape is different, because of wind information discrepancies

An example of another fire we modeled in FARSITE using standard fuel models was a fire that occurred on March 29, 2012 in a beechwood forest. This fire burned for a total of four hours, starting at 1400 and ending around 1800, and burned a total area of 0.5 ha. Wind speeds were variable throughout the burning period as they were quite high during the early afternoon but tapered off throughout the day. In this case we used the following input parameters in FARSITE:

Fuel moisture values: 3% (1-hr), 4% (10-hr), 5% (100-hr), 40% (live herbaceous) and 70% (live woody);

Daily maximum temperatures: 7-10°C;

Daily minimum relative humidity: 36-40%;

Winds: generally from the north-northeast at 10-2 k h⁻¹

The projected fire size from FARSITE was 0.9 ha. Based on the close proximity of a village to the fire location (figure 2) it is quite reasonable to assume that local residents responded to the fire in a volunteer capacity; these suppression actions are not accounted for in the FARSITE analysis. Decreasing winds through the afternoon may have significantly helped suppression activities.

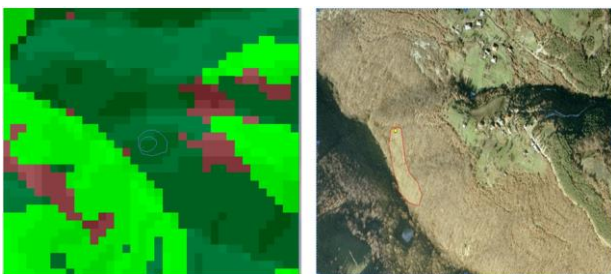


Figure 2: *FARSITE* run for wildfire that burned in a beechwood forest, where size of the fire is very close to the real one, but the shape is different, because of wind information discrepancies

From this two modeled fires we were able to estimate that the standard fuel models established for US can be run also for Bulgarian cases, but only

after a good calibration with real case studies where vegetation and past events are well observed. However FARSITE and BehavePlus provided reasonable outputs for future work in the field of fire behavior fuel modeling on the Bulgarian territory. The work done in more details is described in [11,12,13,14].

4. Conclusion

This paper was an overview paper presenting the test case studies referring to wildland fire propagation and all related simulations. Our main goal was to show the used tools so far on Bulgarian territory and open the ground for a discussion in the field is similar or better tools can be used by having different ICT application than the ones described.

5. Acknowledgement

This work has been supported by the Bulgarian Academy of Sciences Program for support of young researchers No: ДФНП-95-А1 and by the National Science Fund of the Bulgarian Ministry of Education, Youth and Science under Grant FNI I02/20.

REFERENCES

1. Report 6, Forest Fires in Europe 2005, Joint Research Centre—European Commission, EUR 22312 EN, European Communities 2006.
2. Internet resources of the National Fire Safety and Civil Protection Service of Bulgaria.
3. Ecopolis, bulletin 48 (2001), Forest fires reach catastrophic scales (In Bulgarian) http://www.bluelink.net/bg/bulletins/ecopolis12/1_os_1.htm.
4. ROTHERMEL R. C. A mathematical model for predicting fire spread in wildland fuels. Research Paper INT-115. Ogden, UT, US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, 1972, 1–40.
5. GRISHIN A., A. GRUZIN, V. ZVEREV. Mathematical modelling of the spreading of high-level forest fires. *Soviet Physics Doklady*, 28 1983, No 4, 328-330.
6. ALBINI FRANK A. Estimating wildfire behavior and effects. Gen. Tech. Rep. INT-30, Ogden, UT, US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, 1976.
7. PERMINOV V. Mathematical Modeling of Crown Forest Fires Initiation. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture

- Notes in Bioinformatics), Vol. **2667**, 2003, 549-557.
8. Patton, E.G., Coen, J.L.: WRF-Fire: A coupled atmosphere-fire module for WRF. In: Preprints of Joint MM5/Weather Research and Forecasting Model Users Workshop, Boulder, CO, June 22-25. NCAR(2004)221223.
 9. Anderson, H.E.: Aids to determining fuel models for estimating fire behavior. USDA Forest Service, Intermountain Forest and Range Experiment Station, Research Report INT-122 (1982) <http://www.fs.fed.us/rm/pubsint/intgr122.html>.
 10. http://ccm.ucdenver.edu/wiki/Jan_Mandel/Blog/2010_Dec_2011_Jan
 11. Dobrinkov G., **Dobrinkova N.**, “Input Data Preparation for Fire Behavior Fuel Modeling of Bulgarian Test Cases (Main Focus on Zlatograd Test Case).”, 10th International Conference on "Large-Scale Scientific Computations" LSSC'15, Sozopol 8-12 June 2015, Lecture Notes in Computer Science 9374, ISSN 0302-9743, ISSN 1611-3349 (electronic), ISBN: 978-3-319-26519-3, DOI 10.1007/978-3-319-26520-9, Springer Germany, pp. 335–342, **2015**.
 12. **Dobrinkova N.**, Hollingsworth L., Heinsch F.A., Dillon G., Dobrinkov G., “Bulgarian fuel models developed for implementation in FARSITE simulations for test cases in Zlatograd area”. (E-proceeding: <http://www.treesearch.fs.fed.us/pubs/46778>) Wade DD & Fox RL (Eds), Robinson ML (Comp) (**2014**) ‘Proceedings of 4th Fire Behavior and Fuels Conference’, 18-22 February 2013, Raleigh, NC and 1-4 July 2013, St. Petersburg, Russia. (International Association of Wildland Fire: Missoula, MT), p.513 - p.521.
 13. Dobrinkov G., **Dobrinkova N.**, “Wildfire behavior modeling data preparation for FARSITE simulations in Bulgarian test cases”, 5th International Conference on Cartography & GIS & Seminar with EU cooperation on Early Warning and Disaster/Crisis Management 15-21 June **2014**, Proceedings Vol.2, ISSN:1314-0604, **2014**, Riviera, Bulgaria, p.763- p. 770.
 14. **N.Dobrinkova**, G. Dobrinkov, “Farsite and WRF-Fire models, Pros and Cons For Bulgarian Cases”, 9th International Conference on "Large-Scale Scientific Computations" LSSC'13, Sozopol 3-7 June 2013, Lecture Notes in Computer Science 8353, ISBN: 978-3-662-43879-4, Springer Germany, pp. 382–389, **2014**.

Authors' contacts: Institute of Information and Communication Technologies – Bulgarian Academy of Sciences (IICT-BAS)
 Address: Acad. Georgi Bonchev str. Bl. 2
 1113, Sofia, Bulgaria
 E-mail: ninabox2002@gmail.com
 E-mail: nido@math.bas.bg



EXPERIMENTAL SYSTEM FOR WIND MODELLING

GEORGI RUZHEKOV, TEOFANA PULEVA, TS. SLAVOV

Abstract: *The paper describes the developed airflow generation system which characteristics are closed to the real wind. It is a part of wind turbine laboratory test bench "Wind generator" and it will be used in research work to explore advanced control algorithms of wind turbine generator operation in different regimes and control strategies.*

Key words: *Wind Power Spectral Density, Wind Generator, Laboratory test bench, Programmable Logical Controller*

1. Introduction

The utilization of wind energy to perform mechanical work dates back thousands of years. Due to the oil crisis in the early 70s, the technology for the production of wind energy is undergoing strong development, from experimental technology for charging batteries became one of the main methods of generating electricity.

Wind energy is free, clean and infinite. Nowadays, the production of electricity from wind energy is the fastest growing method for obtaining electricity from a renewable energy source.

At the end of 2014 the worldwide installed capacity of wind turbine generators (WTG) is 369,6 GW, thereby only in 2014 is prevented the release of more than 600 tons of carbon emissions.

The main problem in using WTG is the unpredictable and continuously changing behavior of the primary energy source - the wind (a non-stationary stochastic process, with turbulence and extreme variations of the wind speed) and the high quality requirements of supplied electricity.

The conversion of wind energy in electric power and the technical quality standards make the utilization of wind turbines and their connection to the electric power system (EPS) impossible without application of advanced automatic control.

The presented system for experimental modeling of the wind is part of the control system of laboratory test bench for exploring control strategies and algorithms for wind turbine generator control [1, 2, 3 and 4]. It is designed to create an air flow with characteristics similar to those of real wind. This modeling approach allows research work to be carried out of wind turbine control system

with air flow characteristics similar to those of the wind.

2. Laboratory test bench description

The laboratory test bench "Wind generator" (Fig.1) consists of the following components:

- Subsystem for air flow generation – (2): axis ventilator with diameter 1m, which is driven by an induction motor 4 kW (2). The ventilator is mounted in the wind tunnel (3). The induction motor is controlled by the frequency inverter ELDI/ B-4.0 kW (4) and a braking resistor FR-ABR-H5.5K – 110 Ω (4a).
- Sub-system for measuring the air flow speed - anemometer (5) (the output signal is an impulse sequence), speed measuring tube and differential pressure transmitter (6) (output signal 4-20 mA). These devices are mounted on guide rods, which allow measuring the wind speed at any point. The profile of the generated airflow has been explored which allows the position of the sensor to be determined, where it measures the airflow rate close to the average one.
- Subsystem „Wind generator“, which consists of the wind wheel (7), induction sensor for rotational speed measurement (8), sensor for torque measurement (9), three-phase alternator (10), three-phase rectifier (11). The generated power is supplied in the module for the load control of the generator (12). This module allows voltage and current measurement of the generator. Using the output signal from the controller the generator load can be controlled.
- Sub-system for technological process control (13) based on PLC produced by Siemens, CPU

1215 and input/output signal modules. The input signals are: the anemometer signal (impulse sequence), pressure differential transmitter signal – analog signal 0-10 V, turbine rotational speed (impulse sequence), signal from the torque sensor, generator voltage and the load current. The output signals are: control signal of the ventilator motor speed – frequency inverter reference signal 0-10 V and reference signal of the module for generator load control.

- SCADA system (14) – it provides a graphical interface which visualizes the plant operation,

the measured signals, as well as setting the air flow rate and the load of the generator.

3. Wind dynamics

The dynamic characteristics of wind are a combination of weather conditions and characteristics of geographical location. In literature, the wind is modeled as a non-stationary stochastic process. It is a result of superposition of two components

$$v(t) = v_s(t) + v_i(t), \quad (1)$$

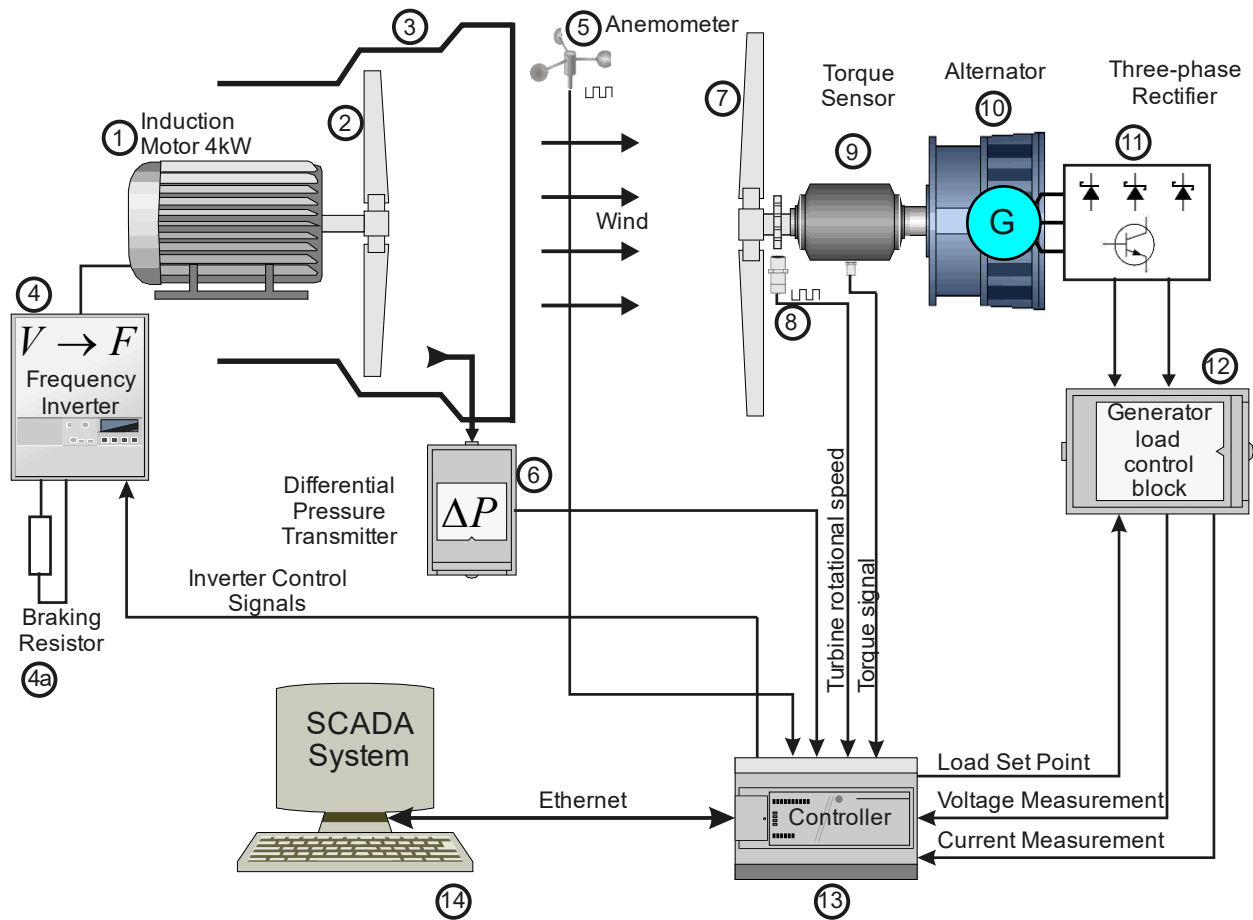


Fig. 1 General structure of the laboratory bench

where $v_s(t)$ is a low frequency component describing the slow fluctuations of the wind speed, $v_i(t)$ is a high frequency component which is a

model of the turbulence (high rate wind speed variations). In Fig. 2 a wind model described by the Van der Hoven spectrum is shown [5].

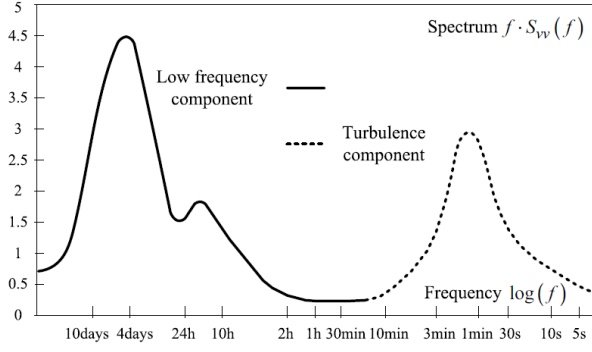


Fig. 2 Van der Hoven spectrum of the wind speed

The low frequency component is a model of the low rate variations of the wind speed. It depends on the characteristics of the geographical location. It can be represented by the Weibul distribution [5]:

$$v_s = ave^{-1/2a\bar{v}^{-2}}, \quad (2)$$

where \bar{v} is the mean value of the wind speed for a period of one hour, and a is a parameter that depends on the mean value of the wind speed for a long time interval.

By definition the turbulence includes all wind speed variations with frequencies above the spectral gap. It is described by the power spectrum. Widespread models are the spectra of Karman and Kaimal [5, 6]:

$$\frac{f * S_{vv}(f)}{\sigma^2} = \frac{4f \frac{L_t}{v_s}}{\left(1 + 70.8 \left(f \frac{L_t}{v_s}\right)^2\right)^{5/6}}, \quad (3)$$

$$\frac{f * S_{vv}(f)}{\sigma^2} = \frac{4f \frac{L_t}{v_s}}{\left(1 + 6f \frac{L_t}{v_s}\right)^{5/3}}, \quad (4)$$

where $S_{vv}(f)$ a spectral density, L_t is the range of the turbulence, which depends on the topology of the terrain, and f is the frequency. Details on the calculation of these coefficients can be found in [5]. The normalized spectra of Karman (line) and Kaimal (dash line) are given in Fig.3.

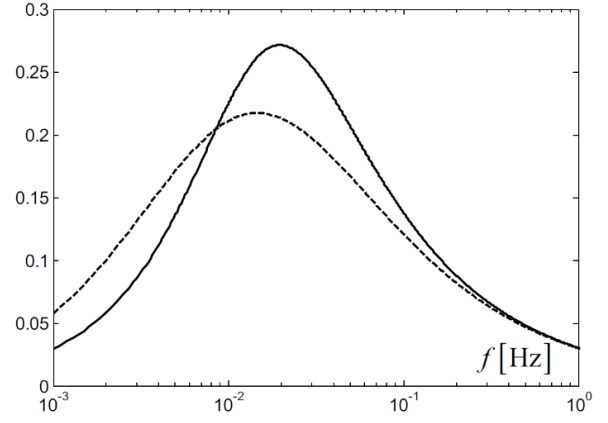


Fig.3. Normalized spectra of Karman and Kaimal

4. Air flow generation

The generation of the airflow is accomplished based on the spectrum density model (*Karman or Kaimal*) $S_{vv}(f)$ Eq. (3) and Eq. (4). We need a discrete representation of the spectrum with respect to the frequency and the magnitude. Let $f(i)$, $i=1, N$ are discrete values with respect to the frequency. The maximum frequency value in the signal spectrum is chosen to be $f(N) = 0.35 \text{ Hz}$, and the minimal one - 10^{-4} Hz , number of discrete points - $N/2 = 20$, $N = 40$ [4]. A logarithmically spaced frequency is used which corresponds to the frequency range (0.1 mHz - 0.35Hz). In accordance with the Kotelnikov - Shannon sampling theorem the sample frequency is determined $F_s = 1 \text{ Hz}$, which corresponds to the sample time $T_s = 1 \text{ s}$. The spectral density magnitude quantization is performed using the trapezium rule – the energy between two neighbor points of frequency is calculated in accordance with the expression:

$$S_{vv}(f_i) = \frac{1}{2}(S_{vv}(i) + S_{vv}(i+1))(f(i+1) - f(i)). \quad (5)$$

The signal generation is performed by Discrete Fourier Transform (DFT). The signal has the following form

$$s(k) = \sum_{i=1}^{N/2} A_i \cdot \cos(2\pi \cdot f(i) \cdot k + \varphi(i)), \quad k = 1 \dots N, \quad (6)$$

where:

A_i is the magnitude of the i^{th} harmonic,
 $f(i)$, $i = 1, N/2$ are the frequency discrete points;
 $\varphi(i)$ - initial phase value of each harmonic.

The values A_i и $\varphi(i)$ have to be calculated. The harmonics magnitude calculation is performed by using the Parseval's theorem.

$$\sum_{k=0}^{N-1} |s(k)|^2 = \frac{1}{N} \sum_{n=0}^{N-1} |S_{vv}(n)|^2, \quad (7)$$

where the left-hand side of the equation describes the signal energy, and the right-hand side – the sum of the energy of all harmonics in this signal realization with a length N .

The energy of each harmonic signal (sin, cos) is

$$E_i = \frac{A_i^2}{2} N. \quad (8)$$

In accordance with the Parseval's theorem for each of the harmonics is held

$$\frac{A_i^2}{2} N = \frac{2}{N} |S_{vv}(f_i)|^2, \quad (9)$$

therefore

$$A_i = \frac{2}{N} |S_{vv}(f_i)|. \quad (10)$$

The spectral density (Karman or Kaimal) $S_{vv}(f)$ represents the signal energy distribution in a frequency range, but it doesn't contain information about the initial phase of each of the harmonics. In the signal modelling the initial phase of the harmonics is given as a random number in the interval

$$\varphi(i) = [0 \div 2\pi] \text{ or } \varphi(i) = \text{rand}.2\pi,$$

where *rand* is a function, that generates an uniform distributed number in the interval $[0 \div 1]$.

Based on the above described theoretical procedure the signal that models the wind speed is described by the expression:

$$V(k) = V_s + \sum_{i=1}^{N-1} A_i \sin(2\pi f(i)kT_s + \phi_i), \quad (11)$$

where k is the index of the current value of the wind speed.

The high frequency component can be obtained as an output signal of a shaping filter assuming a white noise as an input signal. The transfer function of the shaping filter is [5]

$$H_t(j\omega) = \frac{K_F}{(1 + j\omega T_F)^{5/6}}, \quad (12)$$

where the parameters K_F and T_F depend on the low frequency component v_s of the wind model. One of the approaches to obtain the values of these parameters is [5]

$$T_F = \frac{L_t}{v_s} \quad (13)$$

$$K_F = \sqrt{\frac{2\pi}{B(1/2, 1/3)} \cdot \frac{T_F}{T_s}}, \quad (14)$$

where the notation B is a beta-function.

The block-scheme of the algorithm for wind modeling is given in Fig. 4.

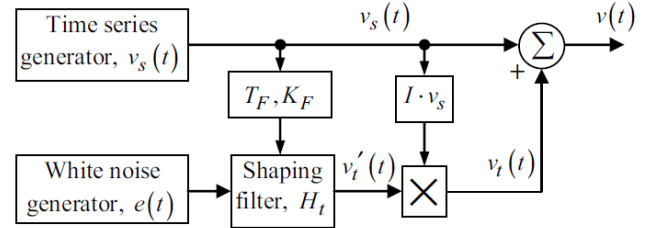


Fig. 4 Block-scheme of the algorithm for wind modeling

The implementation of the shaping filter described by non-rational transfer function Eq. (12) leads to computational problems. In this case an approximation of Eq. (12) by the transfer function Eq. (15) is applied [5]

$$H_t(s) = K_F \cdot \frac{m_1 T_F s + 1}{(T_F s + 1)(m_2 T_F s + 1)}, \quad (15)$$

$$m_1 = 0.4, \quad m_2 = 0.25.$$

In the shaping filter implementation by PLC it is necessary to obtain a discrete representation of Eq. (15). For example the Euler approximation can be applied

$$s = \frac{z-1}{T_s},$$

where the sample time $T_s = 0.1$ s.

As a result the discrete transfer function of the shaping filter is obtained

$$H_t(z) = \frac{4K_F T_F z^{-1} - K_F (4T_F - 1)z^{-2}}{25T_F^2 - (50T_F^2 - 12, 5T_F) \cdot z^{-1} + (25T_F^2 - 12, 5T_F + 1)z^{-2}}. \quad (16)$$

The computational procedure is implemented in the PLC environment as a difference equation.

5. Implementation of the algorithm for modeling of the wind speed

A real time computation of the current value of wind speed using a sample time $T_s = 0.1s$ is performed in the PLC environment. The block scheme shown in Fig. 4 is applied. The algorithm is performed in following steps:

1. Computation of the current value of the low-frequency component in accordance of Eq. (11).
2. Computation of the shaping filter parameters K_F and T_F .
3. Generation of uniform distributed random number in each sample time. A program for random number generator is developed in the PLC environment. It is based on a linear congruential algorithm [4].
4. Computation of the difference equation which represents the shaping filter implementation Eq. (16).
5. Computation of the turbulence component $v_t(k)$.
6. Computation of the resulting component as it is shown in Fig. 4.

On the base of the presented above algorithm the wind speed is performed. It has to be implemented by the system shown in Fig. 1. As can be seen the airflow is created by the ventilator, the induction motor of which is controlled by a frequency inverter. The airflow speed is determined by the motor rotational speed which depends on the frequency of the motor supply voltage.

$$V_{wind} = f(InvFreq), \quad (15)$$

where V_{wind} is the airflow speed which is measured by the pressure differential transmitter and tube for airflow speed measurement, $InvFreq$ - is the frequency of the inverter output voltage that supplies the ventilator. Hence the setting of the airflow speed is accomplished by the inverter frequency reference

$$InvF = f^{-1}(V_{wind}). \quad (16)$$

The implementation of Eq. (16) requires experiments to obtain the relationship described by Eq. (15) in look-up table format. The wind speed is measured for ten values of the inverter frequency. For this purpose an HMI application is developed which allows to set the inverter frequency and to record the values of the airflow velocity. Then Eq.

(16) is implemented as an approximation of the collected data in the look-up table format.

Software for PLC implementation is developed. It comprises the following components:

- Setting the shape of the wind speed variation (constant or sinusoidal). A noise with a normal distribution can be added to these reference signals. The noise variance can be set by the user;
- Transforming the airflow speed into the frequency (inverter reference).
- Computation the inverter control signal;
- Measuring and data storage of the airflow, rotational speed of the wind turbine and the voltage of the generator terminals.

Experiments are carried out using an airflow modelled by:

- constant component and additive noise;
- sinusoidal component with or without additive noise;
- signal which characteristics are similar to the wind model.

The processes visualization and the data records are accomplished by SCADA system, the main window of which is shown in Fig. 5. The data records are in CSV format. This file format is compatible with other software systems. For this purpose a MATLAB function is developed which allows reading the file, visualization the data, selection of a certain interval and data storage for next processing.

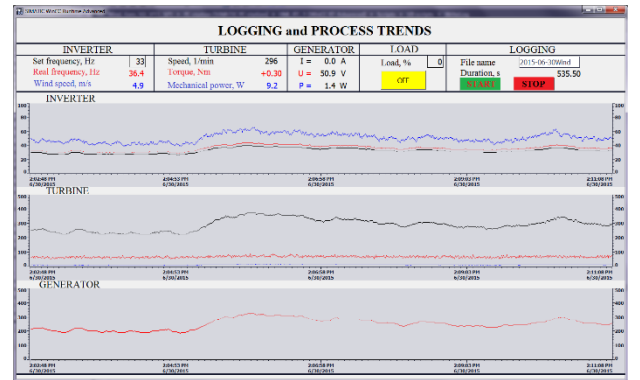


Fig. 5 The main window of the SCADA system

The visualization of real experimental results obtained by the laboratory test bench is shown in Fig. 5. They are

- Inverter frequency (Inv Frequency, Hz);
- Airflow speed (Wind speed, m/s);
- Wind turbine rotational speed (Turbine speed, rpm).

The correlation of these three signals can be noticed as well as the stochastic nature of the airflow speed. The generated airflow can be used in investigation

of the wind turbine generator control system performance. In Fig. 6 the main control window is shown. It is possible to control the system from this screen in manual mode of operation. For example, it is possible to switch on or off the fan motor, to increase or to decrease the fan speed using a slider, to control the generator load or to visualize the main measured signals.

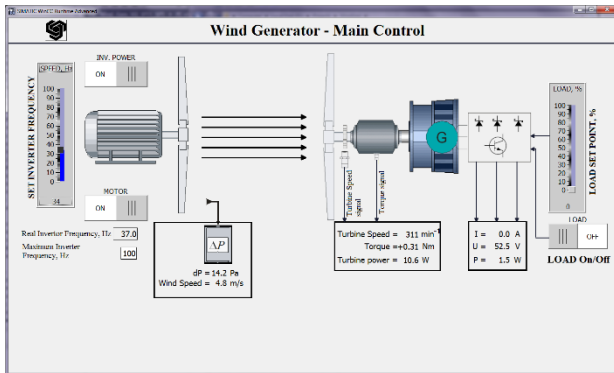


Fig. 6 The main control window of the SCADA system

6. Experimental results

Based on the above described approach for physical modeling of the airflow a software subsystem is implemented. It allows entering the main parameters of the simulated airflow:

- the magnitude of the spectrum harmonics A_i and the corresponding frequencies f_i ;
- the initial phase $\phi(i) = [0 \div 2\pi]$ generated by a developed program generator of normally distributed random number;
- Setting of the wind speed mean value, as well as the minimum and the maximum value of the wind speed;
- Setting the height of the point z where the airflow model is calculated. Setting the coefficient z_0 , that takes into account the roughness of the ground.

In Fig. 7 experimental data obtained by the physical model of the airflow is shown. It represents the random nature of the airflow with or without turbulence.

7. Conclusion

A multifunctional laboratory test bench is developed. It allows being explored different control strategies of wind generators. The presented subsystem for airflow generator offers convenient way different experiments to be carried out in laboratory conditions which are similar to the real applications. This system can be applied to study the effect of mechanical efforts to the system and to

investigate different strategies and control algorithms in the field of renewable energy sources operation in power engineering.

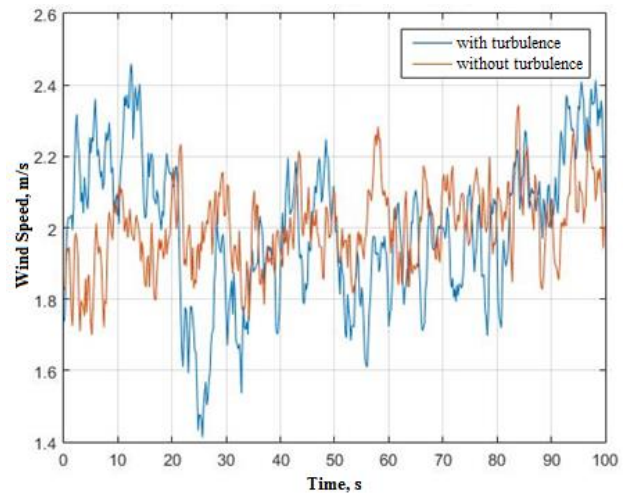


Fig. 7 Airflow with and without turbulence

References

1. Ruzhekov G., I. Tiyanov, Laboratory Bench "Wind Generator" – control system, Symposium with International participation "Control of Energy, Industrial and Ecological Systems" 2013, 7-8 Nov. 2013, Bankya, Proceedings, pp. 73-76 (in Bulgarian)
2. Ruzhekov G., I. Tiyanov, Laboratory Bench "Wind Generator" – control system, Stage 1, „Automatics and Informatics“, №3, 2013, pp. 46-50. (in Bulgarian)
3. Tiyanov I., G. Ruzhekov, Multidisciplinary Wind Stand Control System "Wind Power Generator, NSSTC 2013, Sozopol, pp. 15-19.
4. Ruzhekov, G., T. Puleva, Experimental Test Bench for Wind Modelling, International Conference „Automatics and Informatics'15" 4-7 Nov., 2015, Sofia. (in Bulgarian)
5. Munteanu I., A. Bractu, N. Cutululis, *Optimal Control of Wind Energy Systems, Towards a Global Approach*, Springer, 2008.
6. Vihiala, H. Control of Variable Speed Wind Turbines, PhD Thesis, Tampere University of Technology, 2002.
7. Lubosny, Z. Wind Turbine Operation in Electric Power Systems, Springer 2003.

**Technical University of Sofia
Department of Systems and Control**

E-mail: g_ruzhekov@tu-sofia.bg
tpuleva@tu-sofia.bg
ts_slavov@tu-sofia.bg



PERFORMANCES STUDY OF MULTI-LAYER LINEAR GRADIENT INDEX ANTIREFLECTIVE BASED ON OXYNITRIDE SILICON

KADDOURI AMEUR MILOUD¹, KOUZOU ABDELLAH², HAFIFA AHMED²

Abstract: *This paper deals with the study of the performances of a multi-layers linear gradient index (MLLGI) antireflective coating (ARC) based on oxynitride silicon in comparison of the well known conventional single layer (SL) antireflective coating (silicon or silicon nitride). this study allows to show clearly the improvement brought by the linear gradient index layers AR from point of view of the sensitivity, the thickness of the layers and the angle of incidence. On the other side, this study allows to determine the optimal values of sensitive parameters (Thickness, index...) for the realization of gradient index ARC. It is proved in this study that multi-layers antireflective coatings with better efficiency under oblique incidence can be obtained with affordable methods of implementation.*

Key words: *antireflective coating (ARC), silicon oxynitride, linear gradient index, oblique incidence.*

1. Introduction

The materials used for manufacturing solar cells (Si, GaAs, CdTe...) have high refractive indexes. Hence, the reflection losses are high. For silicon more than 35% of the incident light flux is reflected, which affects greatly the efficiency of photovoltaic devices [1]. To improve the efficiency of solar cells, anti-reflective layers (ARL) are generally used, they are deposited by different methods and following various designs. To achieve an antireflective coating (ARC) with graded index (GI), a deposit of a dielectric layer possessing a refractive index gradually decreasing from the substrate to the ambient medium can be used. Theoretically, if the index decreases gradually from n_s (index of the substrate) to n_0 (index of the ambient medium) the reflection will be null. Unfortunately, this cannot be achieved practically because the high refractive index materials are generally absorbent [2]. Indeed, it was proved previously in some published works that the use of oxynitride layers with graded index allows significantly to reduce the reflectance in the range of wavelengths where the monocrystalline silicon cell is sensitive [15-19]. In this paper the antireflection layers with linear gradient index that are achieved by PECVD silicon oxynitride (SiN_xO_y) are presented. Measures of reflectance

factor were recorded. A theoretical model was developed to study the behavior of these AR layers at normal and oblique incidence [3]. The comparison between the theoretical curves and the experimental measurements of the reflectance factor has allowed to validate the model [4,14]. The use of this model permits the study of the parameters (thickness, optical indices) where the main aim is the optimization of their performances. the present paper contributes to the performance optimization of the linear gradient index antireflective layers. On the other side, the improvement that can be brought to the photovoltaic parameters of a conventional silicon solar cell including the photocurrent and energy efficiency has been studied.

2. Antireflective coating optimization (ARCO)

The use of antireflection coatings (ARC) on solar cells allows reducing the reflection losses and increasing the photocurrent (percentages of almost 50% [5,6]) and therefore, the short-circuit current. The open circuit voltage and the shape factor remain practically unchanged. Consequently, the energetic efficiency is improved in the same proportions as the photocurrent. For this reason, the ARC has to be optimized taking as a reference the illumination current, practically it is nearly equal to the short-circuit current [7,8]. To calculate the photocurrent, it is assumed that the ARC is

deposited on a P + N monocrystalline silicon cell having a surface recombination velocity $S = 106 \text{ cm/s}$, using the solar spectrum AM1.5, which is used for this kind of calculation. The photocurrent is then expressed as follows:

$$J_{ph} = \int_{\lambda_1}^{\lambda_2} [1 - R(\lambda)] E(\lambda) S(\lambda) d\lambda \quad (1)$$

$S(\lambda)$: spectral sensitivity (values published by M. Ogeret [7]).

$R(\lambda)$: Reflection factor (calculated by the software achieved by the authors).

$E(\lambda)$: illumination spectrum AM1.5 [7].

For the integral calculation the Monte Carlo method is used [9]. The photocurrent depends on the domain of wavelengths used for the integration. However, for a silicon cell, generally the wavelength range of 300nm - 1100nm is used [7,8]. Preliminary calculations of the photocurrent of linear gradient ARC for a thickness of 260 nm, $n = 2.0$ to 1.5 and the photocurrent of a solar cell without ARC have showed an improvement of photocurrent of 35%. The parameters optimization has allowed certainly to obtain better results [12,13].

2.1 The thickness optimization

Number The improvement in relative value of the photocurrent is due to the presence of the ARC which is presented as follows:

$$\frac{\Delta I}{I_0} = \frac{I_{PH} - I_0}{I_0} \quad (2)$$

I_{ph} : The photocurrent with ARC.

I_0 : The photocurrent without ARC.

The two currents are calculated from equation (1). The curves below show the relative increase of the photocurrent depending on the thickness of the multi-layer linear gradient ARC compared with traditional single-layer SiO₂ and Si₃N₄. It is important to note that the photocurrent depends on the integration domain. However, the choice of 300nm to 1100nm range is more reasonable since it includes the entire sensitivity range of the silicon solar cell test. For single layer ARC it was verified that the maximum improvement of the photocurrent is obtained for thicknesses equal to the quarter of the wave length centered at $\lambda = 600 \text{ nm}$ (75 nm Si₃N₄ for $n = 2.1$ and 100 nm for SiO₂ $n = 1.5$).

It is noted that the single-layer ARC Si₃N₄ gives good results for $n = 2.1$ which is closer to the optimum value (NONS) $1/2 = 1.95$. On the other side, it is noted that the photocurrent obtained with

linear gradient ARC and the ARC Si₃N₄ are comparable. The photocurrent is improved by 40% which is a satisfactory result in comparison to the published results [10,11]. However, the gradient ARCs are less sensitive to an eventual error on the deposited thickness and/ or to a non-homogeneity on the surface of the deposited layers. This can be observed clearly on the very narrow peaks for the single layer ARC compared to the gradient ARC. It is also observed that the improvement of 40% remains effectively between 80nm and 160nm Fig. 1.

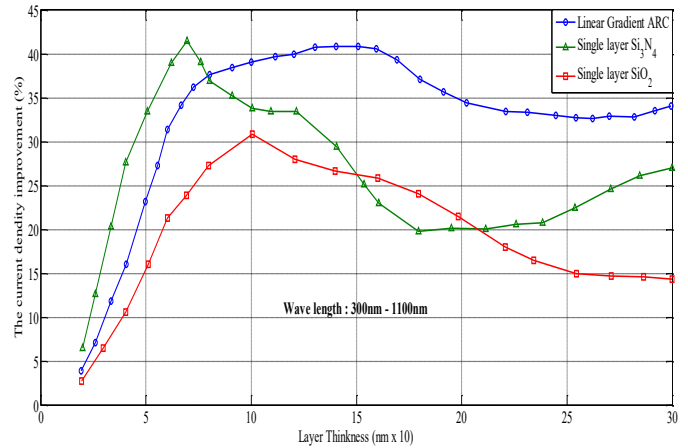


Fig. 1. The variation of the current density versus the thickness value

2.2 Effect of the gradient

Figure 2. shows that greater is the variation of the gradient, the improvement of the photocurrent density is more better. This amounts is due to the fact that the reflectance factor is decreased.

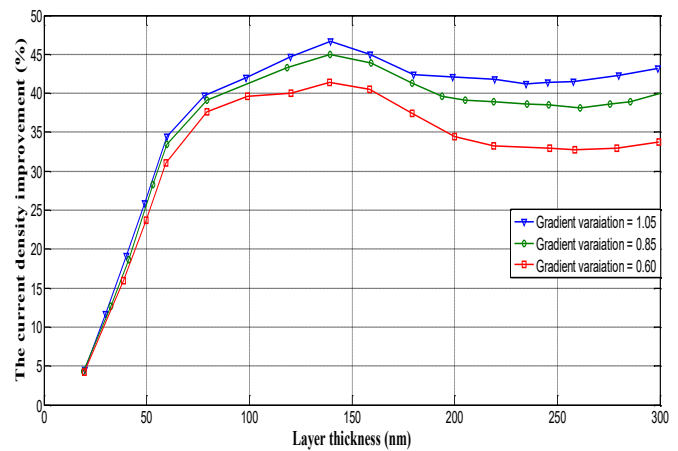


Fig. 2. The variation of the current density versus the thickness value

2.3 The behavior under an oblique incidence

Taking the optimal values for each type of layers, their behavior face to the oblique incidence Fig 3. It appears that the gradient ARCs behave better under oblique incidence than the conventional single-layer ARC. For the ARCs Si₃N₄ and SiO₂ the improvement decreases by more than 10% at 50 ° of incidence, while for the linear gradient ARC it decreases by only 3%.

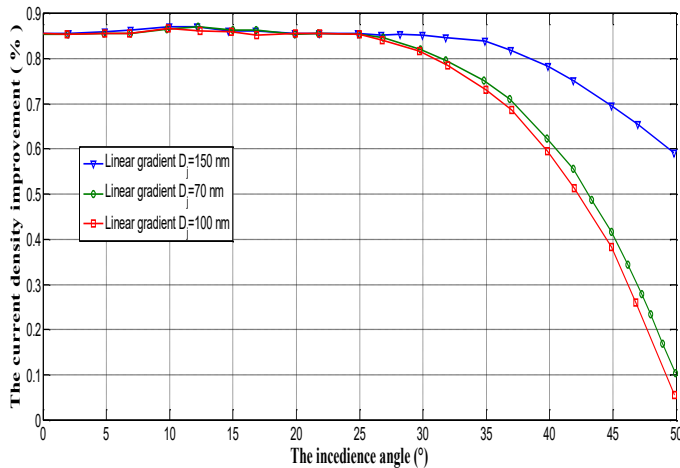


Fig. 3. The variation of the current density versus the incidence angle

3. Conclusion

The study of the influence of each sensitive parameter of the linear gradient index ARC (including thickness, Δn gradient and angle of incidence) has been presented, where the main aim is to ensure their optimization design. The single layer linear gradient ARCs were compared to the widely used single layer ARC of type Si₃N₄ and SiO₂ with quarter of the wave length. Based on the calculations performed at normal and oblique incidence leads to the following conclusions:

- The antireflective coating with graded index are less sensitive to a possible surface non-homogeneity and/or a gap on the deposited thickness. It was proved that the thicknesses between 100 nm and 200 nm gives nearly an over gain of 45% of the photocurrent in comparison a cell without antireflective coating.
- The linear graded index layers are less sensitive to the variation of the incidence angle which is generally the case of the solar cell practical operating conditions. This implies a considerable energy gain if it is used on the whole day. On the other hand, at an incidence angle of 50, the photocurrent lose is only 2.5% for the gradient antireflective coating compared to the photocurrent

lose value of 10% of the conventional antireflective coating.

REFERENCES

1. K. CHOPRA, "Solar selective coatings," 3 rd workshop on thin films physics and technology, pp.306-311, 1999.
2. K. CHOPRA, "Solar selective coatings," 3 rd workshop on thin films physics and technology, pp.306-311, 1999.
3. E. PELLETIER, "La réalisation des traitements optiques de surface," Systèmes optiques, Ecole d'été – institut d'études scientifiques de Cargèse, Vol.I, PP 399-440, 1991.
4. J. VIARD, "Caractérisation des couches minces d'oxynitrides élaborées par PECVD," Thèse, Université de Montpellier. Science et Technologie, Février 1996.
5. A. M. KADDOURI, A. MAHDJOUR, S.CALLARD et J. JOSEPH, "Etude des couches antireflets à base d'oxynitride de silicium," 1st EEEEC Laghouat 2000.
6. A.S. CALLARD, A. GAGNAIRE, J. JOSEPH, "Fabrication and characterisation of graded refractive index silicon oxynitrid thin films," J.Vac.Sci.Technol. A, Vol.15, N°4, PP 2088-2094, 1997.
7. J. ZHAO, M.A. GREEN, "Optimized antireflection coatings for high-efficiency solar cells," IEEE Trans.Electron Devices, Vol.38, N° 8, PP 1925-1934, 1991.
8. M. ORGERET, "Les Piles solaires : le composant et ses applications," Ed Masson, 1985.
9. Z. CHAN, P. SONA, J. SALAMI, A. ROHAUGI, "A Novel and effective PECVD SiO₂/SiN antireflection coatings for solar cells," IEEE Trans.Elec.Devi., Vol.40, n° 6, PP 1161-1165,1993.
10. M. EMINYON, K.RUBIN, "Introduction à la simulation des systèmes physiques," (Inter-editions, 1994.)
11. A.W. BLAKERS, M.J. STOCKS, "Modelling the measurements of silicon solar cells under natural sunlight," 14th European P.V.Solar ener. Conf, Spain, 1997.
12. A. LAUGIER, "Matériaux pour la conversion photovoltaïque," Institut National des Sciences Appliquées de Lyon, 1983.
13. A.M.Kaddouri, A.Belaidi, "Optimisation des couches antireflets à gradient d'indice-comportement en incidence oblique", Conférence Internationale sur la Physique et

- ses Applications, USTOMB Oran, Décembre 2003.
14. A.M.Kaddouri, A.Belaidi, "Optimisation des performances des couches antireflets à gradient d'indice linéaire a base d'oxynitruure de silicium", International Conference on Electrical/Electronics Engineering Communication & Physical systems, C U Saida, Mai 2004.
 15. A.M.Kaddouri, A.Kouzou, A.Belaidi, "The study of the graded index anti-reflective coating based on silicon oxynitride", Journal of Optoelectronics and Advanced Materials, Vol.17, No.1-2, January-February.2015, p.50-61.
 16. M.Beye, F.Ndiaye, A.S.Maiga, "Antireflection Coatings Combining Silicon Nitride with Silicon Nanoparticule", Journal of Engineering and Technology, Vol.1, Issue.2, 2014.
 17. A.Alemu, A.Freundlich, N.Badi, C.Boney, A.Ben-saoula, "Low temperature deposited boron nitride thin filmsfor a robust anti-reflection coating of solar cells", Solar Energy Materials and Solar Cells, Vol.94, No.5, pp.921-923, 2010.
 18. B.G.Lee, I.Skarp, V.Malinen, S.Li, S.Choi, H.M.Branz, in Proveeding of the IEEE Photovoltaic Specialists Conference, Austin, Tex, USA, 2012.
 19. J.Kim, J.Park, J.H.Hong and all, "Double antireflection coating layer with silicon nitride and silicon oxide for polycrystalline silicon solar cell", Journal of Electroceramics, Vol.30, No.1-2, pp.41-45, 2013.
 20. M.Li, L.Zeng, Y.Chen, L.Zhuang, X.Wang, H.Chen, " Realization of colored multicrystalline silicon solar cells with SiO₂/SiN_x :H double layer antireflection coatings", International Journal of Photoenergy, Vol.30, Article ID 352473, 8 pages, 2013.
 - 21.

Authors' contacts:

Address:

¹Laboratoire des Science et Informatique des Matériaux (MSIL), ST Faculty, Djelfa University, Algeria

²Applied Automation and Industrial Diagnostics Laboratory (LAADI), Djelfa Universitydepartments of science and technology, Algeria

² University Amar Telidji of Laghouat

E-mail: amkaddouri@yahoo.fr

E-mail: kouzouabdellah@ieee.org

E-mail:Hafaifa@hotmail.com



VALIDATION OF A MODEL FOR GENERATION OF HOURLY AMBIENT TEMPERATURE TIME SERIES

DIMITAR DIMITROV, ATANAS ILIEV, NEVENKA KITEVA - ROGLEVA

Abstract: *Simultaneous solar radiation data and the ambient temperature hourly data time series are essential for photovoltaic systems analyses, as well as for the analyses for other power systems equipment. In this paper, a model for generation of hourly ambient temperature data for locations, where exist only the monthly mean temperature values and the corresponding hourly data of solar radiation on a horizontal surface. The validation of the model is carried out for the region of Skopje, R. Macedonia through statistical parameters comparison. Slight change of conversion factors is proposed.*

Key words: *ambient temperature, solar radiation, time series, statistical validation*

1. Introduction

The ambient temperature data series are important input for simulation of photovoltaic systems, load consumption, etc. [1-4]. Temperature variations are not so rapid, thus hourly temperature values are usually sufficiently short intervals for estimation for power consumption, performance analyses of renewable systems, etc. For that purpose, for more profound analyses, long-term hourly temperature time series are often required.

Time series of the ambient temperature are measured and recorded at many meteo-stations throughout the world. However, sets of monthly mean temperature data are available more often. On the other hand, for analyses of renewable energy systems simultaneous hourly data for solar radiation and the temperature are required as input

Having in mind the above, for that purpose, models for synthetic generation of hourly time series had been developed. The models for generation of hourly temperature time series had been reviewed in [5]. In [6] was concluded that the sufficient models should not only take into account the statistical and sequential properties of the temperature, but they should also involve the apparent cross-correlation between the solar radiation and the temperature. These models differ by their exactness, complexity, universality, required input data, etc.

The most known software package for generation of meteorological data, METEONORM (versions 2-4) [7], uses an updated model [8]. The model is based on the assumption that the temperature variation amplitudes during daytime are proportional to the corresponding daily global

solar radiation amplitudes. Inputs in the model are hourly data of global solar radiation on a horizontal surface and monthly mean ambient temperature. Apart of these, the model uses conversion factors that depend on the meteorological and geographical characteristics of the location.

In this paper, validation of the updated Scartezzini's model is tested for the region of Skopje, R. Macedonia. The inputs are the measured data of global solar radiation on a horizontal surface and the calculated monthly mean ambient temperature for the same sequence of time. Version 5 and above of METEONORM [9], use another model, developed within the Project "SoDa" [10], but as explicitly reported, the quality of generated data is very similar to the data from the model used in the previous versions [7].

2. Generation of temperature time series

The ambient temperature data series contain deterministic and stochastic components. In the updated Scartezzini's model [7], the deterministic component is extracted from the hourly solar radiation time series. The solar radiation profile for each day is smoothed, stretched and time displaced using specific conversion factors that are typical for the location and the month. The stochastic component is a result of other (then solar radiation) influencing factors to the temperature.

The generated time series are composed of successively chained daily profiles. Each profile begins at sunrise hour and ends at sunrise hour in the following day. The profiles are joined in a way that the temperature at sunrise of ending profile is equalized to the temperature in the sunrise hour at

the beginning of incoming profile. At the end of each month the generated time series are checked if they satisfy the allowed limits and constraints. If the conditions are satisfied the hourly time series are adjusted to the monthly mean, otherwise the generation process starts over again.

The hourly time series of solar radiation are also rarely available, thus their generation is required. In [11, 12] models for generation of these time series are presented.

2.1. Deterministic variations

The variation of the temperature, mainly, is determined by the variation of the solar radiation. However, due to the heating capacities of air masses, surrounding earth surface and objects etc., the thermal effects have greater inertia than the radiation effects. For that reason the temperature variations can't follow the immediate variations of the solar radiation (clouds covering the sun, etc). Also, as a consequence of greater inertia, the heating effect of solar radiation is felt with a delay. Namely, if the maximal solar radiation is at 12:00, the maximal temperature is (for example) at 15:00. Having in mind these, the modelled daily solar radiation profile is smoothed and time displaced. The smoothing is done by averaging the hourly solar radiation data for a given time interval (of several hours), centered in the considered hour. Its values may vary from 6 hours in winter up to 9 hours in summer months. Then the smoothed profile is displaced with a time lag of 2 or 3 hours. Both, the values of the smoothing interval and the time lag are site and monthly dependent conversion factors. The smoothed and time displaced daily solar radiation profile is then converted to a temperature by a stretching factor (another site and monthly dependent conversion factor).

The obtained profile gives the solar radiation dependent temperature component, but only from the sunrise the sunset hours. After the sunset, the temperature decreases depending on the time lag value. When the time lag is 2 hours, the temperature for the further hour continues to be derived from the daily temperature profile, and time lag of 3 hours – from the temperature profile for the next 2 hours after the sunset hour.

For each subsequent hour up to midnight, the temperature decrement is being reduced by a constant factor, which is between 1.15 and 1.5. After the midnight, a linear temperature decrement factor is assumed. This factor depends on the total solar radiation in the following day. When it exceeds 25 % of the daily extraterrestrial solar radiation (maximal possible radiation), the decrement factor is 1/2 of the temperature decrement at last hour before midnight, and when is

below 25 % it is 1/3. However, from the measured data and a large number of temperature generation runs, it was concluded that the decrement factor of 1/1.5 is more suitable for the whole year when the total solar radiation exceeds 25 % of the daily extraterrestrial solar radiation.

2.2. Stochastic variations

The temperature variations that are not a direct consequence of the solar radiation (wind, rain, snow, dew, fog, etc.), are represented with stochastic processes [8]. There are two main components of the stochastic processes: hourly and daily perturbations.

The hourly perturbations ΔT_{AR} are generated by the following autoregressive process:

$$\Delta T_{AR}(h) = 0.7 \cdot \Delta T_{AR}(h-1) + 0.385 \cdot r \quad (1)$$

where: $\Delta T_{AR}(h)$ is the perturbation of the temperature at hour h , r is a normally distributed random variable with mean equal to 0 and standard deviation equal to 1.

The daily perturbations are represented by temperature differences from one midnight to the next $\Delta T_{mm}(n_d)$. The midnight-to-midnight temperature differences are obtained by a random process, which values have Gaussian distribution with the following mean value:

$$\begin{aligned} \Delta \bar{T}_{mm}(n_d) = & \\ = & \begin{cases} -0.01 + 0.05 \sin \left[(n_d - 288) \frac{360}{365} \right], & H < 0.4H_0 \\ 0, & 0.4 \leq H \leq 0.8H_0 \\ 0.02 - 0.06 \sin \left[(n_d - 288) \frac{360}{365} \right], & H > 0.8H_0 \end{cases} \quad (2) \end{aligned}$$

and the following standard deviation:

$$\begin{aligned} \sigma_{T_{mm}}(n_d) = & \\ = & 1 + \left\{ 2 + 2 \sin \left[(n_d - 288) \frac{360}{365} \right] \right\} \sin^2(2\varphi) \quad (3) \end{aligned}$$

where: n_d is the number of the day in the year, H is the total daily global solar radiation on a horizontal surface, H_0 is the total daily extraterrestrial solar radiation on a horizontal surface, φ is geographical latitude of the location.

The midnight-to-midnight temperature difference is assumed that linearly changes over the course of a day. Thus, the hourly temperature change for hour h is:

$$\Delta T_{mm}(n_d, h) = \frac{h}{24} \Delta T_{mm}(n_d) \quad (3)$$

2.3. Constraints and adjustments

Having in mind, that the generation process of hourly temperature time series depends on random numbers, it should be kept stable and realistic under all possible conditions. This is done by application of few constraints and adjustments.

The first is month-to-month transition constrain. It is applied by superposing, to the generated temperature values, a supplementary midnight-to-midnight temperature trend $\Delta T_{\text{sup}}(h)$ during the first day of the month. With this trend is avoided, the midnight temperature difference between the first and the second day in the month $T_{\text{mid}}(1)$, to be higher than 4 °C from the monthly average temperature \bar{T}_m , i.e.:

$$\Delta T_{\text{sup}}(h) = \begin{cases} \frac{h}{24}(\bar{T}_m - 4 - T_{\text{mid}}(1)), & T_{\text{mid}}(1) < \bar{T}_m - 4 \\ 0, & \bar{T}_m - 4 \leq T_{\text{mid}}(1) \leq \bar{T}_m + 4 \\ -\frac{h}{24}(T_{\text{mid}}(1) - \bar{T}_m + 4), & T_{\text{mid}}(1) > \bar{T}_m + 4 \end{cases} \quad (4)$$

Other constrain relates to the average daily temperature values. These values should not deviate more than 10 °C from the monthly average temperature. If such value occurs, the process of generation starts over from the beginning of the month.

Also, the hourly temperature values are limited with minimum and maximum values around the monthly average temperature. These values are site dependent and they were extracted from long-term (more than 10 years) analyses of worldwide range of weather stations. Also, after occurring such value, the generation process starts over from the beginning of the month.

In temperate and cold climatic zones, in January and February, the generated minimum temperature data had been observed to be higher from the measured data. For that reason, so called "cold weather outbreaks", in the generation process within these months, are implemented. The outbreaks are applied by superposing a supplementary daily trend. The duration of the outbreaks is set to 4 days. The daily trend for the first day is another conversion factor, which has a negative value for the first day. In the second day, the daily trend is assumed to be 1/2 from the trend in the first day, and the third and fourth day the supplementary trend is -1/2 from the first day (it has a positive value). The appearance of the cold weather outbreaks is implemented as random event. For any single day, its probability is set to 5 %, so that in the first two months in the year the

probability of appearance of cold weather outbreak is about 95 %. If no outbreaks are to happen up to the end of February, the generation process introduces one within these months, with a randomly chosen starting day.

The average value of the generated monthly temperature profiles should not deviated more than 4 °C than the monthly mean temperature. If this happen the generation process starts over from the beginning of the month. Otherwise all temperatures of the monthly profile are adjusted, so that the monthly average of generated temperature profile have exactly the same value as the monthly mean of the measured data. The adjustment is done in a way that all generated values are increased or decreased by the difference between the generated and measured monthly mean value. This may lead to a discontinuity at the beginning of the month. If the discontinuity is greater than 4 °C, the last two hours of the preceding month and the first two hours of the considered month are smoothed.

2.4. Location specific conversion factors and constraints

The generation process of hourly temperature time series, requires a number of conversion factors. They depend on the climatic sub-zone characteristics of the location and the period in the year. In [7] is given a review of climatic sub-zones. According to the meteorological and geographical characteristics the locations with similar characteristics belong to a same climatic sub-zone. In [7] are also given the conversion factors and some constraints for each corresponding climate sub-zone. Their values were extracted from long-term meteorological data analyses for a wide range of locations.

3. Validation of the model

The hourly time series of the global solar radiation on a horizontal surface and the ambient temperature, are measured for a period of 20 consecutive months. The measurements of the hourly solar radiation and the ambient temperature data, used for this paper, were measured at the Solar Energy Laboratory at the Faculty of Electrical Engineering and Information Technologies, University "Ss. Cyril and Methodius" in Skopje [14]. Within this period, due to technical problems, in the measured time series appeared several gaps. Since a continuum for at least the whole month is required for generation of the ambient temperature time series, there were successfully extracted months with uninterrupted time series. Hence, the generation of temperature time series was possible for the whole year.

Table 1. Conversion factors and constraints for a temperate elevated continental climate [7]

Month	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Min. from average (°C)	-26	-23	-24	-18	-15	-15	-13	-14	-16	-18	-20	-25
Max. from average (°C)	16	20	20	20	18	17	15	16	17	17(20)	15	16
Smoothing interval (h)	7	7	8	9	9	9	9	9	9	8 (9)	8	7
Time displacing (h)	2	2	3	3	3	3	3	3	3	3	2	2
Stretching factor ($^{\circ}\text{C}/\text{Wm}^2$) (from the climatic sub-zone)	0.028	0.026	0.024	0.021	0.017	0.016	0.015	0.018	0.022	0.025	0.027	0.028
Stretching factor ($^{\circ}\text{C}/\text{Wm}^2$) (from the measured data)	0.023	0.027	0.019	0.020	0.019	0.019	0.019	0.022	0.023	0.025	0.028	0.027

Skopje with its latitude of 42°N, longitude of 21.2°E, altitude of 250 m and other characteristics belongs to a temperate elevated continental climate sub-zone. The conversion factors and extreme temperature constraints for this climatic sub-zone are given in Table 1.

The measured data show a minor deviation from the characteristic values for the mentioned sub-zone. Namely, in October the maximal hourly temperature appeared to be 20 °C higher than the average. Also, for this month, the generation process shows that the smoothing interval is suitable to be 9. The reduction factor is taken to 1.15. The values of the stretching factor are given at the Tab. 1. Their values are close to those declared for the climatic sub-zone. The conversion factors with these corrections were used in further testing of the model's validation. A supplementary daily trend for cold weather outbreaks in January and February, for this climatic sub-zone, is -14.4 °C.

For the model validation, the generated and the measured temperature time series were compared. Despite the strong correlation between the solar radiation and the temperature, the ambient temperature has also a stochastic component. For that reason, the validation was done by comparing of some important parameters of time series. Since there were no long-term measured hourly time series of temperature and corresponding hourly time series of solar radiation, deeper statistical analyses could not have been done. Hence, some issues are subjectively assessed.

The monthly average temperature values are model's input. As was explained, the model adjusts the generated time series to have these monthly mean values. So, this parameter is automatically correct.

In Fig. 1, the hourly profiles of the measured and the generated time series for two months (June and January) are shown. As it can be seen, the generated time series fairly good follow the hourly variations of the temperature. Still, it is evident that the deviation of generated from the

measured data is larger in January than in June. However, there is a good matching of the average hourly values from the generated and measured time series, which can be noted from the average daily temperature profiles given in Fig. 2. Nevertheless, there is a discrepancy between the average extreme (minimal and maximal) hourly values of the generated and measured data. Namely, the minimal hourly average of generated data is higher than the measured for 4 % for January and December, around 11 % for May, June, July and August, up to 16 % for March, from the measured average temperature amplitude.

The average daily temperature values of the measured (the thick line) and the generated values from five generating runs (other lines), for January and July, are given in Fig. 3. It can be seen that generating runs may differ, but still they are in reasonable limits. Larger deviations from the measured values are observed in winter months.

The distribution of measured and generated (5 runs) hourly temperature values are given in Fig. 4. There were also analyzed the generated (5 runs) minimal and the maximal temperature values (Fig. 5). The solid line represents the constraints imposed by the model. These values are obtained by appending the minimal and maximal temperatures (Table 2) to the monthly average temperature the values. Slight exceeding from these values is allowed after the adjustment to the monthly average temperature. The differences in minimal and maximal generated and measured exist because to the generation process are imposed the limits from the climatic sub-zone.

There were analyzed midnight-to-midnight temperature differences (Fig. 6). The average temperature difference is well reproduced for all months, while the standard deviation is acceptable for most months. The difference between standard deviation from the measured and the generated time series is higher for January, February, November and December.

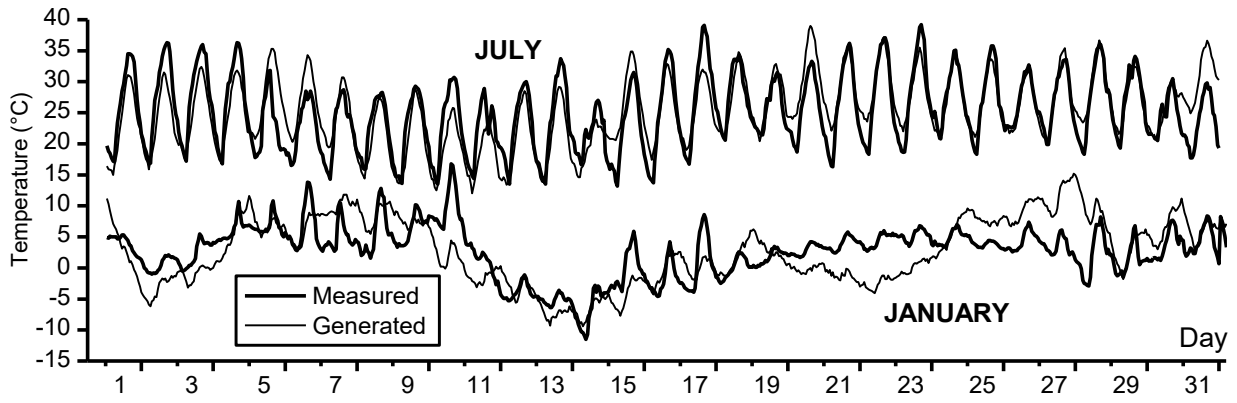


Fig. 1. Measured and generated hourly temperature values for June and January

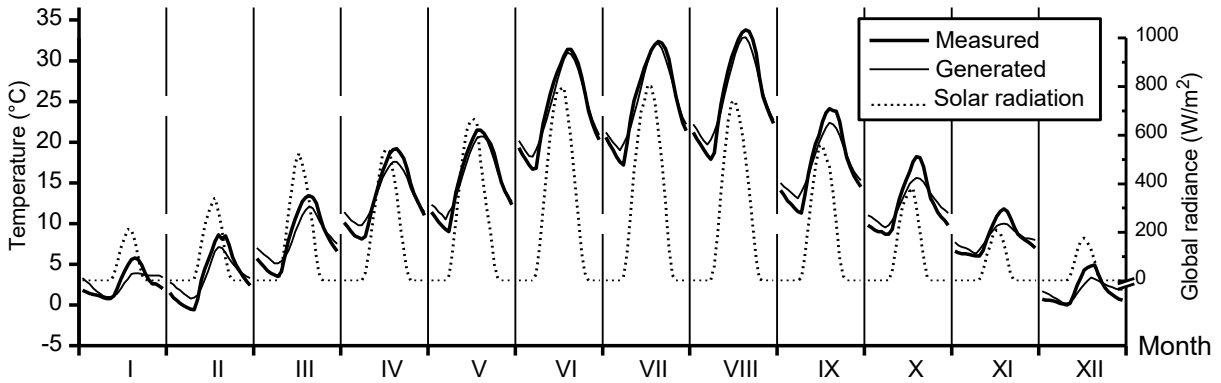


Fig. 2. Measured and generated daily average temperature daily profiles

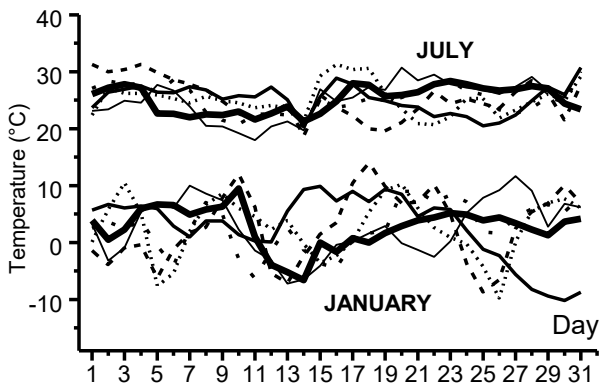


Fig. 3. Measured and generated average daily temperature values for Jul and January

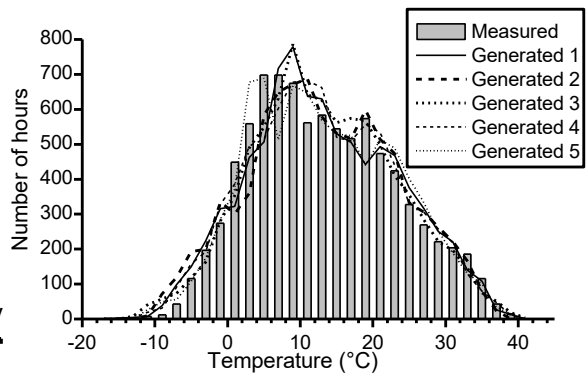


Fig. 4. Distribution of hourly values of measured and generated time series

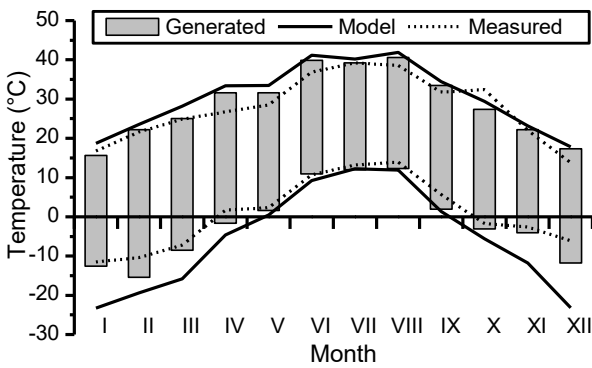


Fig. 5. Minimal and maximal hourly temperature values

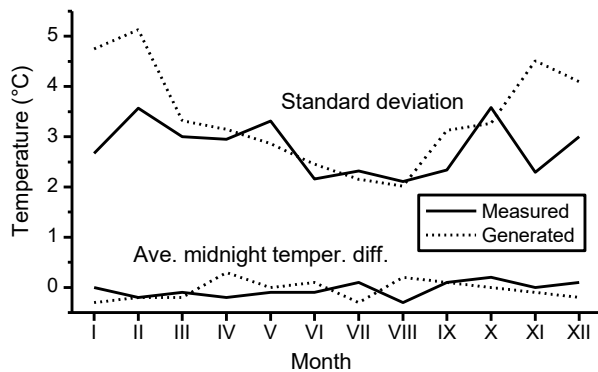


Fig. 6. Average and standard deviation of the midnight-to-midnight temperature differences

4. Conclusion

In this paper the updated Scartezzini's model for generation of the hourly temperature values was elaborated. It was tested its validation for the region of Skopje. There were analyzed a number of parameters that are relevant for its assessment. The conclusions and remarks from this paper are very likely to be valid for locations with similar climatic characteristics as Skopje.

The comparing of the generated and measured temperature time series shows that the shapes of the generated temperature daily profiles were well represented. The deviations from the measured values were expected due to use of random components. However, it was evident that the amplitudes of the average hourly profiles of the generated time series are lower than the generated. In winter this deviation is unacceptably high and this is the weakest point of the model. This is likely because the generation process in winter months is very unstable and it is easy to get out of control. This probably may be improved by adding additional constraints, which likely can be identified from long-term measured data.

The distribution of generated hourly temperature values was generally acceptable. The hourly values of generated time series were within allowed minimal and maximal limits. The average midnight-to-midnight temperatures were well represented. The standard deviation of these values for summer months is good, but in winter months these values differ from the measured time series.

REFERENCES

1. C. S. Chen et al., (1996). Determination of Customer Load Characteristics by Load Survey System at Taipower. *IEEE Transactions on Power Delivery*, Vol. 11, No. 3, pp. 1430-1436,
2. D Dimitrov, V Glamocanin, J Pop-Jordanov, D Andonov, (1997). Simulated evaluation of grid-support photovoltaic systems regarding distribution loss reduction. Part 1: Contributions. CIRED. *14th International Conference and Exhibition on Electricity Distribution*, IEE Conf. Publ. No. 438.
3. D. Dimitrov, K. I. Kocev. (2002). An Approach to Assessment of the Power Loss Reduction in Distribution Networks with a Grid-Connected Photovoltaic System. *3rd Mediterranean Conference and Exhibition on Power Generation, Transmission, Distribution and Energy Conversion – MEDPOWER 2002, Athens, Greece.*
4. T. Markvart. (1994). *Solar Electricity*, John Wiley and Sons.
5. K. M. Knight, S. A. Klein, J. A. Duffie. (1991). A Methodology for Synthesis of Hourly Weather Data. *Solar Energy*, Vol. 46, No. 2, pp. 109-120, Elsevier Science.
6. D. Heinman, J. Schumacher. (1996). Synthesis of Hourly Ambient Temperature Time Series Correlated with Solar Radiation, *2nd ISES Europe Solar Congress, EuroSun'96*, Freiburg, Germany,
7. J. Remund, S. Kunz. (1997). *METEONORM, Global Meteorological Database for Solar Energy and Applied Climatology Version 3*. METEOTEST, Bern, Switzerland,
8. J.-L. Scartezzini, M. N. Ferguson and F. Bochud. (1990). *Compression of multi-year meteorological data, Final Report*. OFEN Project EF-REN (90)009, Solar Energy and Building Physics Laboratory, Department of Architecture, Swiss Federal Institute of Technology, Lausanne, Switzerland.
9. J. Remund, S. Kunz. (2003). *METEONORM: Global Meteorological Database for Engineers, Planners, and Education, Version 5*. METEOTEST, Bern, Switzerland,
10. Project, "SoDa", DG "INFSO" IST-1999-12245, <http://www.soda-is.com/>. (last visited February 2016)
11. R. Aguiar, M. Collares-Pereira. (1992). TAG: A Time-Dependent, Autoregressive, Gaussian Model for Generating Synthetic Hourly Radiation. *Solar Energy*, Vol. 49, No. 3, pp. 167-174, Elsevier Science.
12. V. A. Graham, K. G. T. Hollands, (1990), "A Method to Generate Synthetic Hourly Solar Radiation Globally", *Solar Energy*, Vol. 44, No. 6, pp. 333-341, Elsevier Science,
13. "Providing information on Solar Radiation", <http://www.helioclim.org/>. (last visited February 2016)
14. D. Dimitrov, V, Stoilkov, A. Iliev. (2012): Solar Energy Laboratory – Summary of Ten Years Continuous Experience, *Renewable Energy & Power Quality Journal*, No.10, ISSN 2172-038X.

University "Ss. Cyril and Methodius"
 Faculty of Electrical Engineering & IT
 Karpos bb, P.O. Box 574
 1000 Skopje
 Macedonia
 E-mail: ddimitar@feit.ukim.edu.mk
 E-mail: ailiev@feit.ukim.edu.mk
 E-mail: ailiev@feit.ukim.edu.mk



A MODEL OF A SINGLE-PHASE TWO-ROTOR AXIAL GENERATOR WITH RARE EARTH MAGNETS

NIKOLA GEORGIEV

Abstract: *This paper makes a study of a single-phase synchronous 16-pole generator with axial magnetic field and rare earth magnets. The generator consists of two rotors with sixteen rare earth magnets each and one stator with twelve windings. An equivalent magnetic circuit of the generator has been given. A simulation has been realized in the application Schematics (OrCAD), the main magnetic flux and the magnetic fluxes of dissipation have been calculated. The induced phase e.m.f. in one winding has been calculated by means of the main magnetic flux, taking into account the number of windings, the pole pairs of the generator, its revolutions per minute and the maximum value of the magnetic flux. Both the phase voltages of the generator at idle running and the power consumption at active load have been calculated with the help of the obtained expression for the r.m.s. value of the phase e.m.f. Experimental studies of the output electrical parameters of the considered single-phase generator at idle running and active load have been conducted. The output characteristics, obtained from the model, have been compared to those, from the experimental measurements.*

Key words: *single-phase synchronous generator, axial magnetic field, rare earth magnets, model*

1. Introduction

Generators with NdFeB rare earth magnets have a simple and reliable construction and high power per unit of weight. The absence of both an excitation winding and current results in high efficiency in operation [1].

Initially the generators were made mostly with ferrite and alnico magnets, having significantly lower energy. The basic structure of the generators with axial magnetic field was unilateral: with a rotor and a stator and the bilateral structures appeared only recently – either with one rotor and two stators or with two rotors and a stator [2].

Paper [3] considers a single-phase generator, which is bilateral with two rotors and a stator, for which the flux linkage has been calculated by the

finite elements method and then the r.m.s. value of the phase e.m.f. has been defined. Similar are the models in [4], where the magnetic flux and the instant value of the e.m.f. have been calculated; and in [5], where the r.m.s. value of the stator e.m.f. has been defined.

This paper studies and models in Schematics (OrCAD) environment a low-power single-phase synchronous generator with axial magnetic field and rare earth magnets, consisting of two rotors and a stator.

2. Exposition

The considered single-phase axial generator consists of two rotors with sixteen rare earth magnets each and a stator with twelve windings connected in series.

Fig. 1 shows the construction scheme of the considered two-rotor single-phase generator with a view from the left, in which the following notations are used: by 1 the aluminum stator is denoted; 2 is the stator winding, 3 – the steel rotors; 4 – the rare earth magnets; and 5 is the shaft, driving the rotors.

In order to make an equivalent magnetic circuit it is necessary to define the e.m.f. F and the equivalent magnetic reluctance $R_{\mu M}$ for one of the rare earth magnets [1]

$$F = R_{\mu M} \cdot \Phi_r, \quad R_{\mu M} = \frac{l_M}{\mu_0 \cdot \mu_r \cdot S_M}, \quad (1)$$

where l_M is the length of the average magnetic line of force through the magnet; μ_r and μ_0 are the relative magnetic permeability and the air permeability, respectively, while S_M is the cross-section of the magnet.

The magnetic flux in case of a short magnetic circuit can be calculated by means of the residual magnetic induction B_r [1]

$$\Phi_r = B_r \cdot S_M. \quad (2)$$

Thus the equivalent circuit of the rare earth magnet is obtained.

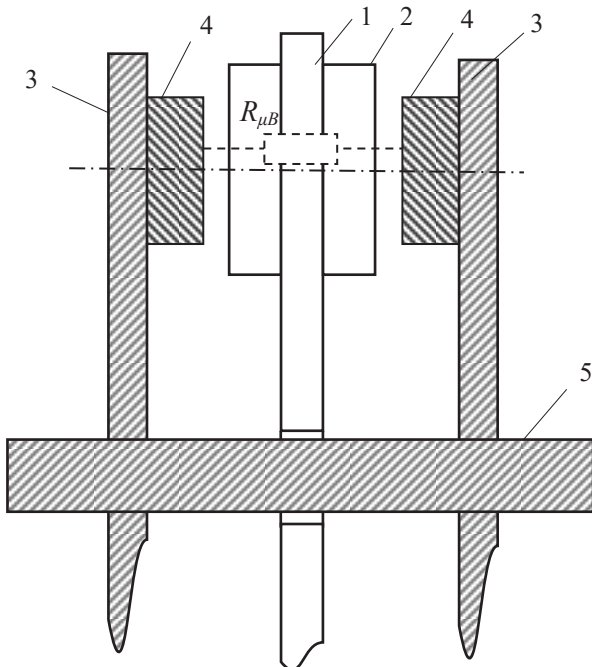


Fig. 1. Construction scheme of the two-rotor single-phase generator, a view from the left

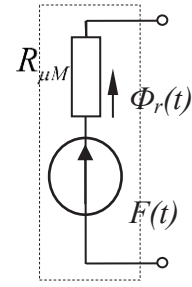


Fig. 2. Equivalent circuit of the rare earth magnet

Fig. 3 presents the equivalent magnetic circuit of the single-phase generator. $R_{\mu B}$ and $R_{\mu CT}$ represent the magnetic reluctances of the air gap and of the steel rotors. By means of the magnetic reluctance $R'_{\mu B}$ the losses from the dissipation of the magnetic flux between two neighboring magnets are defined, while $R''_{\mu B}$ shows the losses from the dissipation between the corresponding magnet and rotor

$$R_{\mu B} = \frac{l_B}{\mu_0 \cdot S_B}, \quad R_{\mu CT} = \frac{l_{CT}}{\mu_0 \cdot \mu_r \cdot S_{CT}}, \quad (3)$$

$$R'_{\mu B} = \frac{l'_B}{\mu_0 \cdot S'_B}, \quad R''_{\mu B} = \frac{l''_B}{\mu_0 \cdot S''_B},$$

where l_B , l'_B , l''_B , l_{CT} are the lengths of the average magnetic lines of force through the air gaps and of the steel rotors, while S_{CT} , S_B , S'_B , S''_B are the cross-sections of the steel rotors and the air gaps.

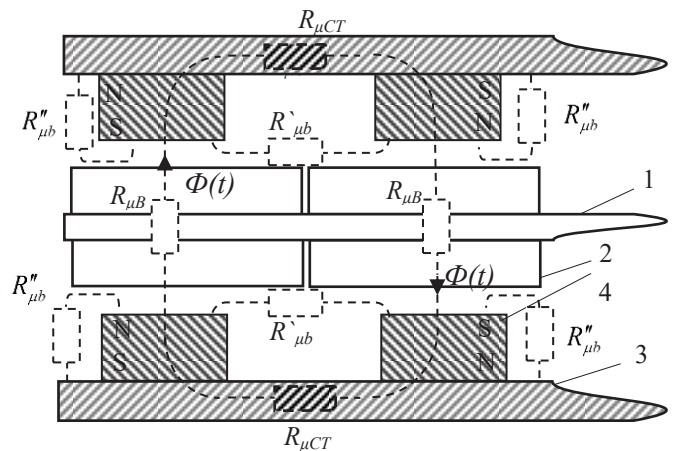


Fig. 3. Equivalent magnetic circuit

By means of the equivalent magnetic circuit from Fig. 3, the full equivalent magnetic circuit of the single-phase generator is obtained - Fig. 4.

After calculating the magnetomotive voltages and the magnetic reluctances by (1), (2) and (3) and realizing a simulation in the application Schematics (OrCAD), the magnetic fluxes in the generator are obtained. In this case electromagnetic analogy is used, in which the sought magnetic fluxes correspond to the currents in the electric circuits. Therefore the dimensions of the calculated magnetic fluxes are in microamperes, not in microwebers - Fig. 5.

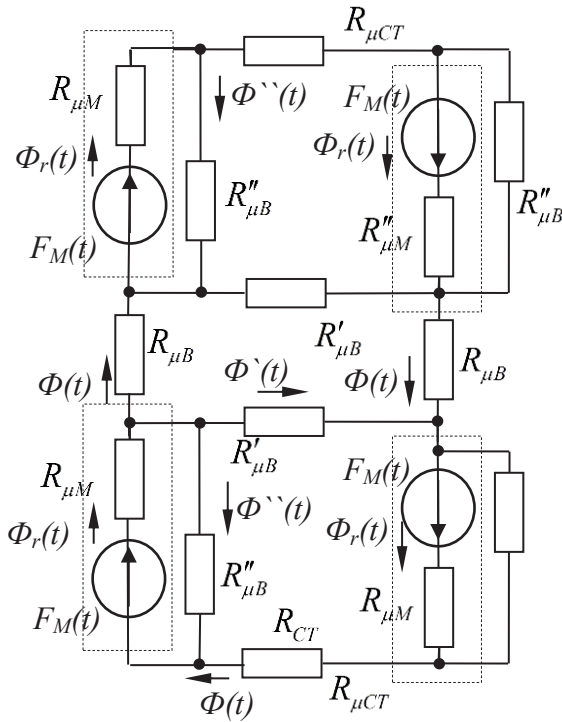


Fig. 4. Full equivalent magnetic circuit

The performed simulation shows that 68% of the magnetic flux, created by the permanent magnets $\Phi_r(t)$ goes through the two coils of the single-phase generator $\Phi(t)$, and the remaining 32% are dissipation fluxes. The bigger dissipation flux is the one between two neighboring magnets $\Phi^{\prime}(t)$, which is 59% from the losses. The smaller $\Phi^{\prime\prime}(t)$ (41%) is the magnetic flux of dissipation between the corresponding magnet and rotor.

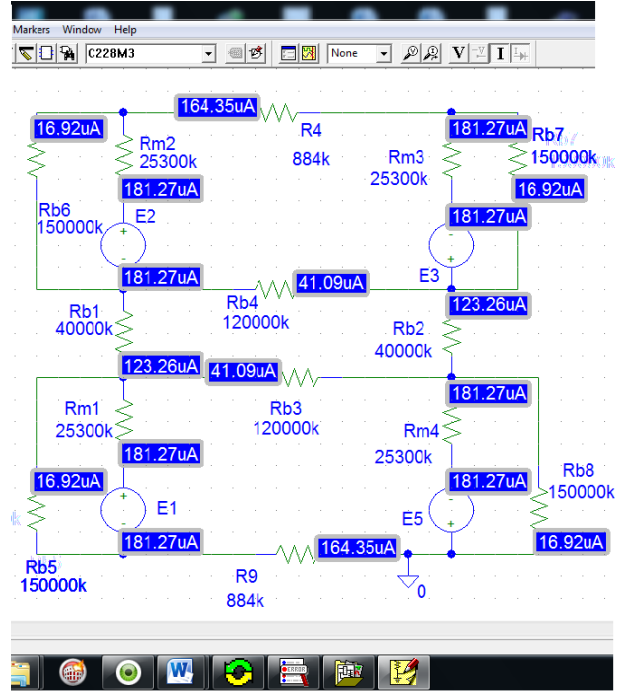


Fig. 5. Simulation scheme in Schematics

By means of the calculated magnetic flux, going through the two windings of the single-phase generator $\Phi_m=0,123$ mWb, the r.m.s. value of the induced e.m.f. in one stator winding of the generator can be calculated

$$E_1 = \frac{\omega W}{\sqrt{2}} \Phi_m \quad (4)$$

After expressing the circular frequency by means of the revolutions per minute and the number of pole pairs, it is obtained

$$\omega = \frac{2\pi \cdot p}{60} n \quad (5)$$

$$E_1 = \frac{W}{\sqrt{2}} \frac{2\pi \cdot p}{60} n \cdot \Phi_m \quad (6)$$

Since the single-phase generator has 12 stator windings, connected in series, the full r.m.s. value of the phase e.m.f. is equal to

$$E_{12} = \frac{12 \cdot W}{\sqrt{2}} \frac{2\pi \cdot p}{60} n \cdot \Phi_m \quad (7)$$

The formula, obtained this way, takes into account the number of turns in the stator winding, the number of generator poles and of the stator windings, the revolutions per minute and the maximum magnetic flux, going through one of the stator windings of the generator with rare earth magnets.

After using (7), the phase e.m.f. of the generator at idle running can be calculated. The

equivalent electric circuit at idle running in a complex effective form is presented in Fig. 6, where by R_{12} and L_{12} the active resistance and the inductivity of the 12 stator windings are denoted respectively.

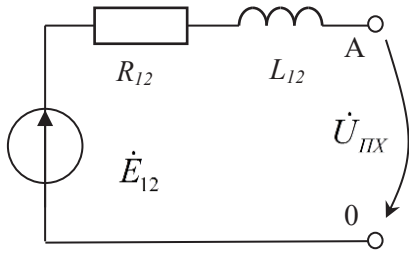


Fig. 6. Equivalent electric circuit at idle running

In this case the r.m.s. value of the phase voltage at idle running is equal to the phase e.m.f. of the generator

$$U_{PIX} = E_{12} \quad (8)$$

Table 1 and Fig. 6 present the r.m.s. values of the calculated by the model phase voltage at idle running of the generator, and the measured phase voltage within a range of revolutions per minute from 0 до 1000 min^{-1} .

Table 1. Phase voltage, calculated by means of the model and measured experimentally

n, min^{-1}	0	200	400	600	800	1000
U_{mod}, V	0	23,1	42,4	68,3	89,2	122,2
U_{meas}, V	0	24,6	49,2	73,8	98,4	123

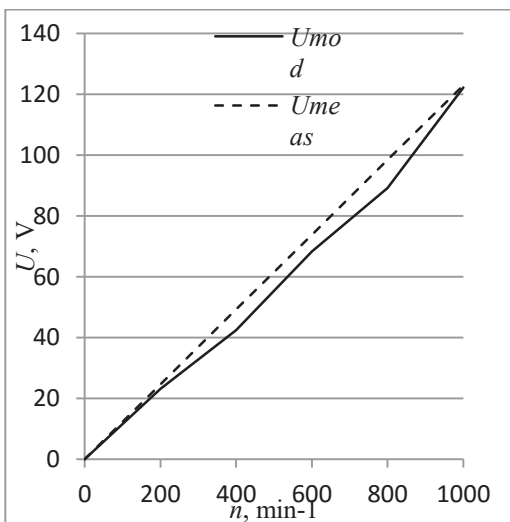


Fig. 7. Calculated by the model and measured in experiments phase voltage at idle running

The maximum relative and fiducial errors for the voltages at idle running are correspondingly $\delta_{max}=9,3\%$ and $\gamma_{max}=7,5\%$.

By using (7), the phase e.m.f. at active load can be calculated. In this case the active load resistance R_T is also added to the equivalent electric circuit in a complex effective form - Fig. 8.

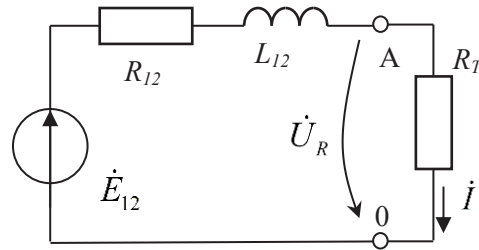


Fig. 8. Equivalent circuit at active load

The r.m.s. value of the phase current is obtained by means of Kirchoff's second law

$$I = \frac{E_{12}}{\sqrt{(R_{12} + R_T)^2 + (\omega L_{12})^2}} \quad (9)$$

Thus the r.m.s. values of the voltage on the active load can be defined, as well as the active power, consumed by it

$$U_R = R_T I, \quad P = U_R I \quad (10)$$

Table 2 and Fig. 6 give the r.m.s. values of both the calculated by the model and measured active power at active load of 5Ω for revolutions per minute from 0 до 1000 min^{-1} .

Table 2. Calculated by means of the model and measured in experiments active power

n, min^{-1}	0	200	400	600	800	1000
P_{mod}, W	0	21,2	72,4	169,5	262,1	377,2
P_{meas}, W	0	19,2	67,7	161,3	253,5	363,1

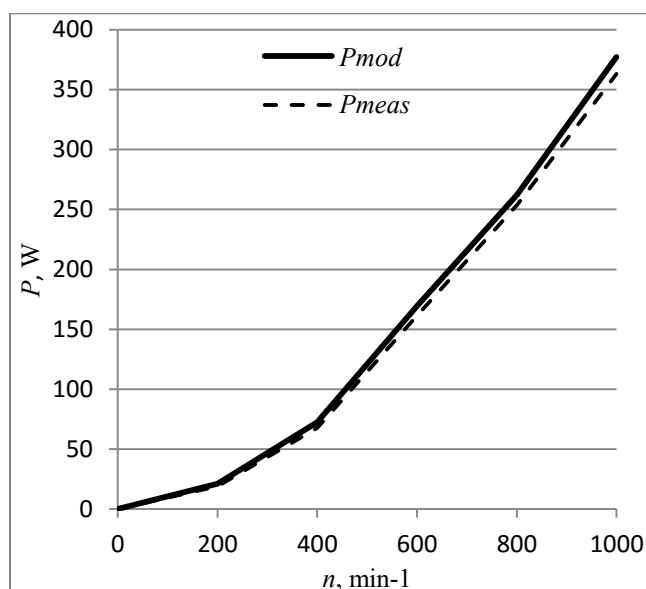


Fig. 9. Calculated and measured active power at active load $R_T = 5\Omega$

The maximum relative and fiducial errors for the voltages at active load $R_T = 5\Omega$ are correspondingly $\delta_{\max} = 3,9\%$ and $\gamma_{\max} = 3,9\%$.

3. Conclusion

The following conclusions can be drawn based on the presented theoretical derivation, software simulation and experimental studies:

1. An equivalent magnetic circuit of the single-phase generator has been made, where the equivalent circuits of the rare earth magnets, the magnetic reluctances of the air gaps, the steel stator and the shaft have been taken into account. The magnetic reluctances, taking into consideration the losses from the dissipation of the magnetic flux between two neighboring magnets and between the corresponding magnet and the rotor have also been included in this circuit.
2. Electromagnetic analogy has been used and a simulation in the software application Schematics (OrCAD) has been realized, wherein the basic magnetic flux and the magnetic fluxes of dissipation have been calculated.
3. By means of the calculated basic magnetic flux, the induced e.m.f. in one winding has been calculated and then the phase voltage at idle running has been obtained.
4. An expression for the r.m.s. value of the phase e.m.f. has been obtained, in which the number of windings, of the pole pairs and of stator windings of the generator have been taken into account, as well as the number of its revolutions per minute and

the maximum value of the magnetic flux in the middle of the corresponding winding.

5. The obtained model simulates with good precision the work of the studied single-phase generator with rare earth magnets, wherein the maximum relative and fiducial errors for the phase voltages at idle running are respectively $\delta_{\max} = 9,3\%$ and $\gamma_{\max} = 7,5\%$, while for the power at active load they are correspondingly $\delta_{\max} = 3,9\%$ and $\gamma_{\max} = 3,9\%$.
6. In order to reduce the maximum relative and fiducial errors, it is necessary to take into consideration both the asymmetries in the process of making the prototype of the generator, and the chatter from the rotation of the rotor.

REFERENCES:

- [1] Mohammad, M. Sc. and Widyan, S. (2006). *Optimization, Construction and Test of Rare-Earth Permanent-Magnet Electrical Machines with New Topology for Wind Energy Applications*, 17-22. Fakultät IV –Elektrotechnik und Informatik der Technischen Universität, Berlin.
- [2] Aritz, E. (2013). *Overview of Axial Flux Machines*. Electric Energy Magazine №4, Spain.
- [3] Wang, R. and Kamper, M. (2005). *Optimal Design of a Coreless Stator Axial Flux Permanent-Magnet Generator*. IEEE Transactions on magnetics, vol. 41.
- [4] Bolund, B. and Bernhoff, H. (2005). *Flywheel energy and power storage systems*. Renewable and Sustainable Energy Reviews.
- [5] Wannakarn, P. and Kinnares, V. (2011). *Microcontroller based Grid Connected Inverter for Axial Flux Permanent Magnet Generator*. IEEE PEDS 2011, Singapore.

Department of Electrical Engineering
 Technical University–Sofia, Branch Plovdiv
 25 Tsanko Dystabanov St.
 4000 Plovdiv
 BULGARIA
 E-mail: geotek@abv.bg



CALCULATING MAGNETIC INDUCTION IN THE AIR GAP OF A THREE-PHASE TWO- ROTOR AXIAL GENERATOR WITH RARE EARTH MAGNETS

NIKOLA GEORGIEV

Abstract: *A formula for calculating the magnetic induction in a definite point of the air gap of a two-rotor axial three-phase generator with rare earth magnets has been derived. By means of the obtained formula the main magnetic flux, going through the windings of the stator, has been found. Thus the induced e.m.f. in one winding has also been calculated, from where the phase e.m.f. and the phase voltages of the three-phase generator with axial magnetic field have been obtained for the mode of idle running. Experimental studies of the phase voltages at idle running have been conducted and the results have been compared with the calculated values. The calculated by the proposed formula maximum magnetic induction has been compared with the result, obtained by means of a computer program, also calculating the maximum induction in a point of the air gap of a two-rotor generator.*

Key words: *magnetic induction, three-phase synchronous generator, axial magnetic field, rare earth magnets, idle running*

1. Introduction

Axial generators with rare earth /NdFeB/ magnets, having power of up to 3 kW, are frequently used in practice due to: their simple, reliable and easy to produce construction, the lack of excitation winding and current, resulting in high efficiency in operation [1].

The magnetic attraction between the rotor and the stator is avoided in these generators and thus the resistive moment is reduced. These machines are much more efficient than the ones with radial magnetic field due to the lack of losses in the core [2].

In order to obtain the basic electrical parameters of the axial generators with permanent magnets, it is necessary to calculate the magnetic induction in the air gap. Under "air gap" the distance between the magnets of the rotor and the stator is understood for one-rotor type of generators, and between the magnets of the two rotors for two-rotor axial generators.

Paper [4] presents a formula for calculation of magnetic induction in a definite point of the air gap for a two-rotor axial generator with permanent

magnets, in which the residual induction for the two magnets, as well as their dimensions, have been taken into account.

This paper derives a formula for calculating the magnetic induction in a particular point of the air gap in a two-rotor axial three-phase generator with rare earth magnets, taking into consideration the residual magnetic induction of the two magnets, their geometrical dimensions and the magnetic permeability of the air gap.

By means of the obtained formula for calculating the magnetic induction in a definite point of the air gap, the main magnetic flux, going through the stator windings is also calculated. Thus the induced e.m.f in one winding is found and thence the phase voltage of a three-phase, two-rotor, synchronous, 16-pole generator with axial magnetic field at idle running is obtained.

2. Exposition

The considered three-phase axial generator consists of two rotors with 16 rare earth magnets

and 12 windings each, connected in a star - four for a phase.

Fig.1 shows the construction scheme with a view from the left and a magnetic circuit of the considered two-rotor, three-phase generator, in which the following notations are used: 1 – the aluminum stator; 2 – a winding; 3 – the steel rotors; 4 – the rare earth magnets; and 5 – the shaft, driving the rotors.

The magnetic reluctance of the air gap $R_{\mu B}$, the magnetic reluctance, which takes into account the dissipation $R'_{\mu B}$, the magnetic reluctance of the rare earth magnets $R_{\mu M}$, and the magnetic reluctance of the steel stators and the shaft $R_{\mu CT}$ are presented in Fig. 1.

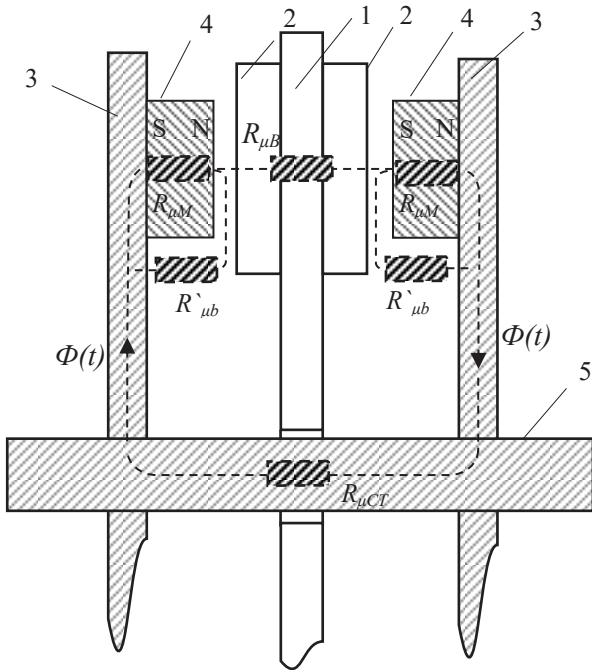


Fig. 1. Construction scheme, a view from the left and magnetic circuit of the two-rotor three-phase generator

For calculating the magnetic flux in the air gap, i.e., the one, going through one of the generator windings, the magnetomotive voltage F and the equivalent magnetic reluctance $R_{\mu M}$ of one of the rare earth magnets are defined [1]

$$F = R_{\mu M} \cdot \Phi_r, \quad R_{\mu M} = \frac{l_M}{\mu_0 \cdot \mu_r \cdot S_M} \quad (1)$$

The magnetic flux in case of a short magnetic circuit can be found [1] in the following way

$$\Phi_r = B_r \cdot S_M, \quad (2)$$

and the magnetic reluctances of the air gap with the winding, of the steel and of the magnetic reluctance, which takes into account the dissipation, are

$$R_{\mu B} = \frac{l_B}{\mu_0 \cdot S_B},$$

$$R_{\mu CT} = \frac{l_{CT}}{\mu_0 \cdot \mu_r \cdot S_{CT}}, \quad (3)$$

$$R'_{\mu B} = \frac{l'_B}{\mu_0 \cdot S'_B},$$

where:

l_B, l_{CT}, l'_B are the lengths of the average magnetic lines of force through the air gap, the steel (the rotors and the shaft) and the fluxes of dissipation;

μ_r - the relative magnetic permeability;

S_{CT}, S_B, S'_B - the cross-sections of the steel, of the air gap and the dissipation;

B_r - the residual magnetic induction.

The equivalent magnetic circuit in an instant form at closing the magnetic flux, going through the shaft of the generator, is of the type.

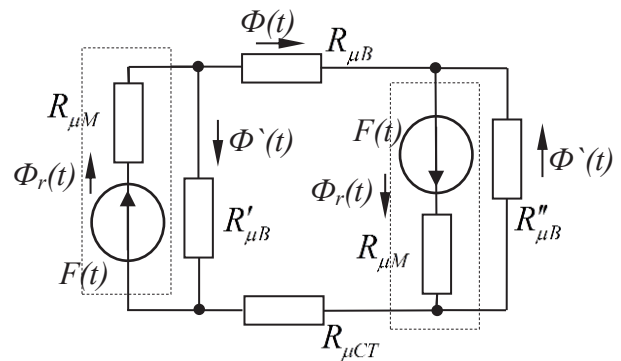


Fig.2. Equivalent magnetic circuit

By means of the magnetic reluctances $R'_{\mu B}$, the losses from the dissipation of the magnetic flux, created by the two magnets, are taken into consideration, but for a generator with two rotors and a small air gap they are much bigger and can be neglected. The magnetic reluctance $R_{\mu CT}$ can also be neglected since it is hundred times less than the one of the air gap in which the stator winding is located. Then the equivalent magnetic circuit is simplified in the following way:

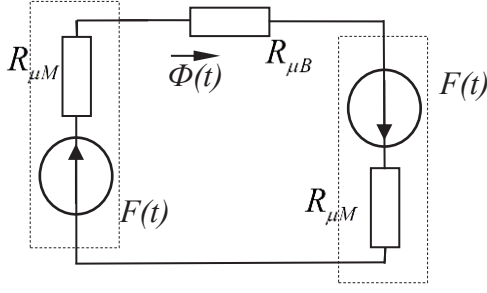


Fig.3. Simplified magnetic circuit

Using the obtained magnetic circuit – Fig.3 – and Kirchoff's second law for magnetic circuits, it is found

$$\Phi(t) = \frac{2F(t)}{2R_{\mu M} + R_{\mu B}} \quad (4)$$

From the above expression the maximum value of the magnetic flux can be obtained

$$\Phi_m = \frac{2F_m}{2R_{\mu M} + R_{\mu B}} \quad (5)$$

The maximum value of the magnetic induction is equal to

$$B_m = \frac{\Phi_m}{S_B} = \frac{2F_m}{S_B(2R_{\mu M} + R_{\mu B})} \quad (6)$$

By substituting (1), (2) into (6), it is obtained

$$B_m = \frac{2R_{\mu M} B_r S_M}{S_B(2R_{\mu M} + R_{\mu B})} \quad (7)$$

$$B_m = \frac{2B_r}{\frac{S_B}{S_M} \left(2 + \frac{R_{\mu B}}{R_{\mu M}} \right)}$$

After expressing the magnetic reluctances from (1) and (3), for the maximum value of the magnetic induction it is obtained

$$B_m = \frac{2.B_r}{2 \cdot \frac{S_B}{S_M} + \frac{l_B}{l_M} \mu_r} \quad (8)$$

In order to find the maximum value of the magnetic induction in any point x of the air gap in the range from the surface of the magnet 0 to the

middle of the air gap $l_B/2$, the above expression is changed in the following way

$$B_m = \frac{2.B_r}{2 \cdot \frac{S_B}{S_M} + \mu_r \cdot \frac{l_B}{l_M} \left(1 + \frac{x}{l_B} \right)} \quad (9)$$

The residual magnetic induction for the two magnets, the geometric dimensions of the magnets, their magnetic permeability and the magnetic reluctance of the air gap are taken into consideration by the obtained formula.

For the considered generator $l_B=18$ mm, $l_M=10$ mm, $x=9$ mm and $B_r=1,2$ T.

The table below gives the maximum induction for three points in the air gap of the generator, calculated both by the proposed formula and with the help of a computer simulation named "field calculator" [5].

Table 1. Maximum induction in three points of the air gap

x , mm	5	7	9
$B(x)$, mT (Model)	374	362	351
$B(x)$, mT (Field Calc.)	395	360	348

From Table 1 it can be seen that the maximum relative error is $\delta_{max}=5,32\%$, and it is even smaller for the air gap $\delta_{max}=0,86\%$ of the generator $x=9$. This confirms the adequacy of the proposed formula for calculating the maximum induction in any point x in the air gap of a generator with permanent magnets.

By means of (9) the maximum value of the magnetic flux in the middle of the air gap can be found

$$\Phi_m = B_m \cdot S_B = 139 \mu Wb \quad (10)$$

The r.m.s. value of the induced e.m.f. in one winding with W turns can be calculated by the following expression

$$E_1 = \frac{\omega W}{\sqrt{2}} \Phi_m \quad (11)$$

After expressing the circular frequency ω by means of the revolutions per minute n and the number of pole pairs p , it is obtained

$$\omega = \frac{2\pi \cdot p}{60} n, \quad (12)$$

$$E_1 = \frac{W}{\sqrt{2}} \frac{2\pi \cdot p}{60} n \cdot \Phi_m. \quad (13)$$

Since each phase consists of four stator windings, connected in series, then the full r.m.s. value of the phase e.m.f. is equal to

$$E_A = E_4 = \frac{4W}{\sqrt{2}} \frac{2\pi \cdot p}{60} n \cdot \Phi_m. \quad (14)$$

Thus the final expression for the full r.m.s. value of the e.m.f. for the corresponding phase is obtained, in which the number of windings and of the pole pairs, as well as the maximum value of the magnetic flux in the middle of the corresponding stator winding have been taken into account.

Using (14), the phase voltages at idle running can be calculated (Fig. 4).

The r.m.s. value of the output voltage at idle running in this case is equal to the phase e.m.f. of the generator

$$U_{IX} = U_{AO} = E_A. \quad (15)$$

Fig. 4 presents the electric circuit of the stator for phase A, wherein the phase active resistance R_A and the inductive resistance X_{LA} (expressed by the inductivity L_A) have not been taken into account at idle running.

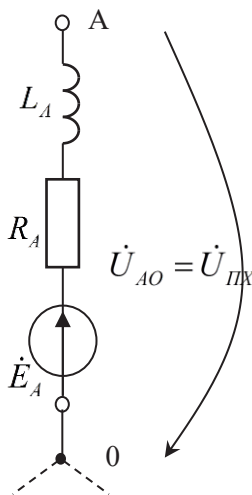


Fig. 4. Electric circuit of phase A of the stator

Table 2 and Fig. 5 present the r.m.s. values of both the calculated by the model phase voltage at idle running and the voltage, measured for the

prototype of the generator for phase A, at a range of revolutions from 0 до 1000 min^{-1} .

Table 2. Measured and calculated by the model phase voltage

n, min^{-1}	0	200	400	600	800	1000
$U_A = U_4, \text{V}$ Meas.	0	8,7	17,3	26,1	33,9	43
$U_A = U_4, \text{V}$ Mod.	0	9,4	18,8	28,5	37,6	47

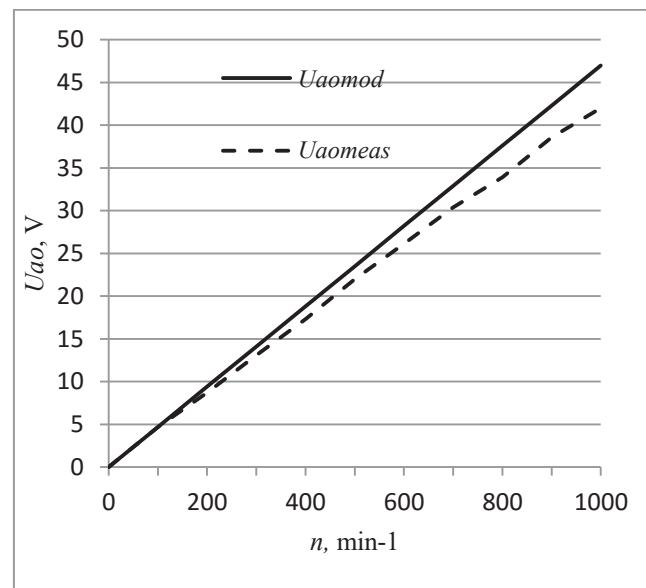


Fig. 5. Measured and calculated by the model phase voltage

The maximum relative and fiducial errors at idle running are correspondingly $\delta_{max}=8,5\%$ and $\gamma_{max}=8,5\%$. The errors have the same values due to the fact that the biggest differences between the measured and the calculated voltage at idle running are for the highest number of revolutions per minute at 1000 min^{-1} .

3. Conclusion

The following conclusions can be drawn based on the theoretical derivation and the experimental studies:

1. A formula has been obtained for calculating the magnetic induction in a certain point of the air gap in a two-rotor axial three-phase generator with rare earth magnets.

2. The derived formula takes into account the magnetic induction for the two magnets, their geometrical dimensions, their magnetic permeability and the magnetic reluctance of the air gap.
3. The adequacy of the proposed formula has been verified; the maximum relative error is $\delta_{max}=5,32\%$, while for the air gap of the generator $x=9$ mm, where the middle of the stator winding is, it is even smaller $\delta_{max}=0,86\%$.
4. The maximum value of the basic magnetic flux, going through the stator windings can be defined by means of the proposed formula. Thus the induced e.m.f. in one winding is also calculated and thence the phase voltage at idle running is found.
5. The obtained formula simulates with good precision the work of the studied prototype of the three-phase generator, wherein the maximum relative and fiducial errors for the phase voltages at idle running are respectively $\delta_{max}=8,5\%$ and $\gamma_{max}=8,5\%$.
6. In order to reduce the maximum relative and fiducial errors the asymmetries in the process of making the prototype have to be taken into consideration, as well as the chatter from the rotation of the rotor.
7. An expression for the full r.m.s. value of the e.m.f. for the corresponding phase has been obtained, in which the number of windings and of the pole pairs of the generator, together with its revolutions and the maximum value of the magnetic flux in the

middle of the corresponding winding have been taken into consideration.

REFERENCES

1. Mohammad, M. Sc. and Widyan, S. (2006). *Optimization, Construction and Test of Rare-Earth Permanent-Magnet Electrical Machines with New Topology for Wind Energy Applications*, 17-22. Fakultät IV – Elektrotechnik und Informatik der Technischen Universität, Berlin.
2. Gieras, J. F. and Wang, R. (2004). *Axial flux permanent magnet brushless machines*. Dordrecht, The Netherlands: Kluwer.
3. Kamper M.J. (2009). *Analysis and design of Axial flux permanent magnet wind generator system for direct battery charging application*. Stellenbosch University, Matieland, South Afrika.
4. Schroeler, K.H. (2010), *Permanent Magnets – materials and magnet system*, 5-6. IBS Magnet, vol. 13, Berlin.
5. [www//Magnetic2Field Calculators and MagnetCalculationsMagnetSolutionsfromDexterMagneticTechnologies.html](http://www.Magnetic2FieldCalculatorsandMagnetCalculationsMagnetSolutionsfromDexterMagneticTechnologies.html)

Department of Electrical Engineering
 Technical University–Sofia, Branch Plovdiv
 25 Tsanko Dystabanov St.
 4000 Plovdiv
 BULGARIA
 E-mail: geotek@abv.bg



FLUORESCENT SPECTROSCOPY APPLICATIONS IN VETERINARY MEDICINE AND HEALTHCARE

VANYA PLACHKOVA

Abstract: *For testing liquid samples from large and small sick ruminants was improved field method for fluorescence assay. In addition to samples from sick animals method is applicable for testing of raw milk that was taken directly from the manufacturer instead the commercially available. The specific method of fluorescent analysis described in this report is characterized by precision, fast response and easy feasibility. To confirm the efficacy of the fluorescence analysis for the detection of bacteria in ruminants and recognition of the different types of raw milk fluorescence analysis carried out with the help of a veterinarian. An essential point in the fluorescent analysis of samples from sick animals proved the fact of the companies establishing the presence of bacteria cause disease, which had no external symptoms in the animal. On the other hand important point in the testing of raw milk has proved that by fluorescence analysis can be obtained from samples of milk from one animal strain or mixed milk from more breeds. Fluorescence spectroscopy is used for analysis of olive oils and fruit juice too. It is shown that fluorescence spectroscopy is an efficient and fast method for analysis of purity and composition-native olive oil and juice, which are an integral part of healthy lifestyle. With fluorescent analysis we can easily ascertain the difference between pure olive oil and extra virgin olive oil. By fluorescence analysis is made easy and fast correlation between the purity of freshly squeezed fruit juice compared to those sold in markets*

Key words: *Fluorescence analysis, emission wavelengths, veterinary medicine, healthy food*

1. Introduction

The lack of methods for early diagnosis of animals is a major problem for farmers, which results in a reduction in population of animals. In order to prevent the serious disease spread an early diagnosis of the animal is required. Most drugs of the veterinary pharmaceutical industry are expensive for small and medium companies. Besides that the drugs are expensive, they do not guarantee that the animal will recover. It turns out that despite the expensive drugs used to treat the animals they do not guarantee their rescue and livestock should be 'scrapped'. Therefore perfecting methods for early diagnosis of cattle is extremely important field of study. Early diagnosis of the problem animal diseases will allow increasing the benefits of livestock and reducing morbidity.

More and more often fluorescence spectroscopy gets used for analysis of milk products. In case of pre-prepared library of data, this method is fast and efficient. Unlike approved chemical tests for certification of agricultural products under fluorescent spectroscopy does not require additional supplies in the analysis of samples. In recent years there have been more yogurts, which are a mix example of cow and sheep's milk or sheep's and goat's milk or cow's and goat's milk. Therefore, attention was paid to an examination by fluorescence method of mixed milks.

Increasingly often are used impurities and substitutes for natural products in the food industry. By fluorescence analysis can be done easy diagnosis of the purity of the food product and its composition.

2. Material and methods

2.1. Field fluorescence measurements

Pattern of fluorescence spectroscopy is shown in Figure 1. Since fluorescence is often very weak, and in addition, in all directions, then in order not to saturate the receiver, the useful fluorescence signal is measured in a direction which is less than 45° relative to the excitation radiation.

Preferably the light source of the scheme is laser diode (LD) as the spectral width is very small. LED used in the experiment has a relatively wide spectral width of radiation from 30-40 nm and angular distribution of the radiation is in the range of large corner $\pm 30^\circ$. It was selected to work with two laser diodes with one wavelength of 245 nm, which is used for the samples from infected animals, and the other 255 nm, which is used for the analysis of raw milk. Source analysis of samples from sick animals is chosen after preliminary investigations it was found that the bacteria in the body of cattle have low emission wavelength. Source analysis of samples of raw milk was selected after extensive tests were conducted and it was found that only at a wavelength of 255 nm and detects the difference in emission wavelengths between sheep and goat milk.

LEDs irradiate the sample and its emission wavelength is transmitted through the optical fiber to a CMOS detector. The sensitivity of the CMOS detector is in the range of 200 nm to 1100 nm. Its resolution is about $\delta\lambda = 5$ nm.

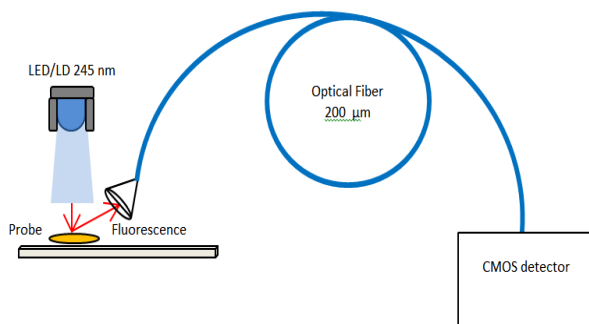


Fig.1. Schematic of Field Fluorescence measurements.

Fluorescence analysis was carried out at a dairy farm. It was decided to make field measurements, in order to avoid damage to the samples in transit and thus to ensure the more reliable fluorescence assay. With the above described production were studied various liquid samples from sick cows such as vaginal secretions, blood, nasal discharge and abscess, fluid samples from dead sheep sick of *Pasteurella*, a sample of the small intestine, colon sample. It was performed a

fluorescence analysis of raw milk from sheep, goat, cow and mixed milk from several breeds of cows

2.2. Classical fluorescence spectroscopy

Fluorescence spectra were recorded using a Spectrometer - FluoroLog3 - TCSPC - Horiba Scientific. The Steady-state of Spectrometer is Broadband 450-W xenon arc lamp from UV to near-IR. The Resolution of FluoroLog3 is 0.2 nm. Of all the modern time-domain methods, including boxcar integration and streak cameras, TCSPC has by far the best dynamic range. Fixed-wavelength "Plug and Play" interchangeable NanoLED pulsed laser-diodes and LEDs. Wavelengths of 280, 340, 375, 405, 440, 473, 635, 650, 785 and 830 nm are available for laser-diodes. 370, 455, 465, 485, 560, 590, 605, 625 for LEDs.

Standard optical pulse durations are <200 ps (<100 ps typical) for laser-diodes, <1.5 ns for LEDs.

2.3 Experimental set-up for the observation of fluorescence spectra

UV fluorescence spectroscopy was used to analyze and olive juice. The experimental set up is presented on Figure.2. Fluorescence spectra are recorded using fiber-optic spectrometer *Avantes 2048* with a spectral sensitivity within the 250 - 1100 nm range. The resolution of the spectrometer is $\delta\lambda = 5$ nm. An optical fiber with a core diameter of 200 μm is used to bring light to the probe and to measure the scattered and fluorescent light. A collimator with a lens of an aperture $D = 5$ mm is used to gather more light and focus it into a receiving fiber connected to the spectrometer. To block stray light, the cuvette holder is supplied with a cap.

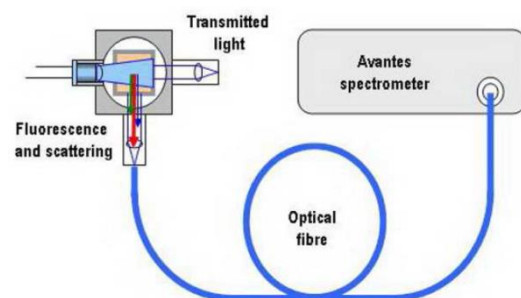


Fig.2. Experimental set-up for the observation of fluorescence spectra.

3. Results and discussions

3.1 Fluorescence spectroscopy of sheeps and cows

The excitation spectra for all studied samples of sick cows (blood sample, sample of the abscess, a sample of vaginal and nasal discharge) is 245 nm. Bacteria in the body of small ruminants have low emission wavelength. Recorded with the emission wavelengths set at: 590 nm for *Fusobacterium necrophorum*, 338 nm for *Corynebacterium pyogenes*, 579 nm for *E. coli*, 330 for *S. Aureus*, 700 for β -Hemolytic *Stafilococcus*, 350 for *Streptococcus pyogenes* and 448 for *Mannheimia haemolytica*

The excitation spectra for all studied samples of dead sheep (sample of the abscess open about lung cancer cient sample, a sample of colon) is 245 nm. Bacteria and nutrients in the body of small ruminants have low emission wavelength. Recorded with the emission wavelengths set at: 337 nm for *Pasteurella multocida*, 357 nm for sheep milk, 452 nm for β -defensin, 510 for lipopigment lipid, 400 for infusoria, 342 for pepsin and 475 for P-glycoprotein.

The aim of the study with sheep is bacterium *Pasteurella multocida*, which is responsible for the death of a sheep, other substances which are registered characterize the composition of the body of small ruminants.

The samples of sick cows were analyzed by veterinarian out at the dairy farm where the system for fluorescent analysis was positioned.

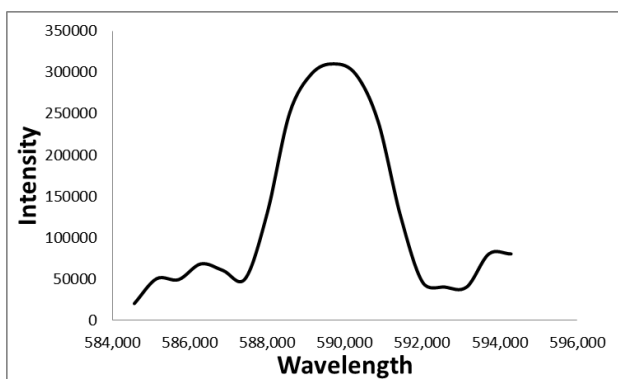


Fig.3. Fluorescence spectroscopy of *Fusobacterium necrophorum* with emission wavelength of 590 nm at length of excitation $\lambda = 245$ nm.

On Fig.3 and Fig.4 are clearly visible emission wavelengths of *Fusobacterium necrophorum* and *Corynebacterium pyogenes*, which were registered

in the blood sample of one of the cows. Thus the diagnosis was confirmed by a veterinary specialist in emerging hoofed disease.

Sick sheep field autopsy was conducted by a veterinarian thus samples were procured for fluorescence spectroscopy.

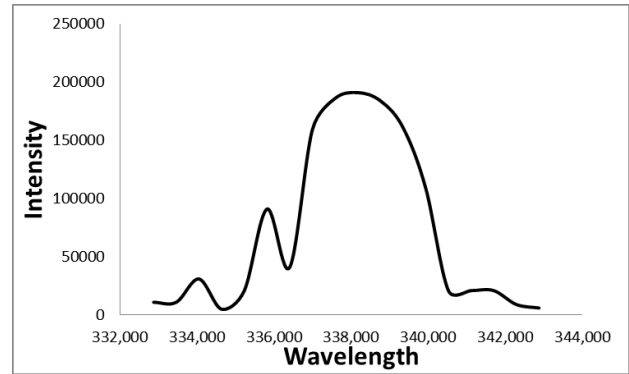


Fig.4. Fluorescence spectroscopy of *Corynebacterium pyogenes* with the emission wavelength of 338 nm at length of excitation $\lambda = 245$ nm.

On Fig.5 is clearly visible emission wavelength of *Pasteurella multocida*, which was recorded in a sample of abscess around the open areas of the lung of the animal. Thus the diagnosis was confirmed by a veterinary specialist for deadly fast growing Pasteurellosis disease, which is contagious to humans.

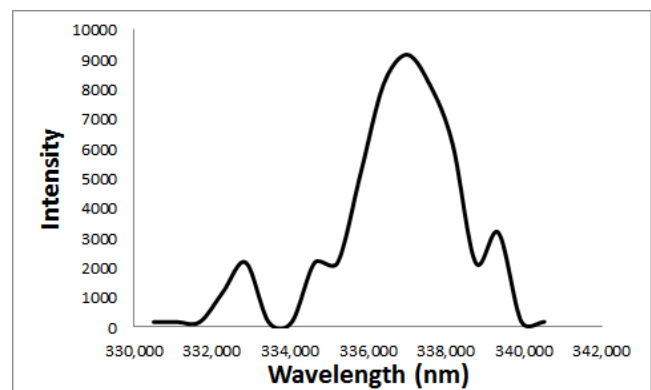


Fig.5. Fluorescence spectroscopy of *Pasteurella multocida* with emission wavelength of 347nm at length of excitation $\lambda = 245$ nm.

Pasteurella multocida was discovered by fast-acting fluorescent analysis as its emission wavelength have been previously studied by reference in specialized sources.

Other bacteria except *Pasteurella multocida* were not found, which means that the only disease in terms of fluorescence analysis is pasteurellosis. Besides samples of abscess was made and a

fluorescent analysis of liquid samples taken from small and large intestines.

3.2 Fluorescence spectroscopy of milk

For all studied samples of milk (cow milk, mixture cow milk, goat milk and sheep milk) the excitation wavelength is 255 nm. This is due to the fact that it only has a small difference in the peaks of fluorescence between goat and sheep milk. Records were made at the emission wavelengths set to: 343 nm for raw cow milk, 357 nm for sheep milk, 355 nm for goat milk and 366 nm for mixture cow milk.

Samples of small ruminants vary considerably in the emission wavelengths with that of the cow, which belongs to a class in cattle.

Consideration was given to a certain fluorescence spectroscopy test of mixed milks. The proportion of samples was 1:1, it is not currently performed fluorescence assay at various percentages of one or another type of milk. The excitation wavelength for all studied samples of mixture milk is 255 nm. The records with the emission wavelengths are set at: sheep + cow milk - 364 nm, sheep + goat milk - 360 nm, goat + cow milk - 354 nm.

Considerable attention was paid to an examination by fluorescence method of mixed milks presented on Fig.6 and Fig.7. Mixes a mixture of milk from large ruminant (cow) and retail (respectively goat and sheep) ratio of the samples was 1:1.

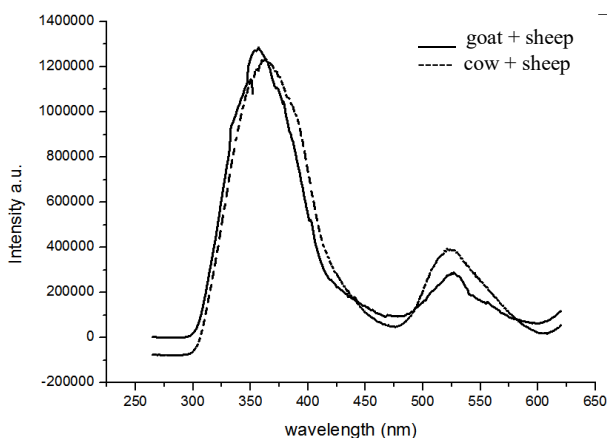


Fig.6. Difference between the emission wavelength of goat + sheep and cow + sheep when irradiated at 255 nm.

Currently made fluorescence assay at different rates contents of one or other type of milk. It turned out that the difference between the emission wavelength at various mixes is not so great as fluctuation between milk from a breed of cow milk from cows of several breeds. This means that in a blend of equal portions of the milk of the large and small ruminants is obtained more similar performance than mixing the milk from different breeds of cattle.

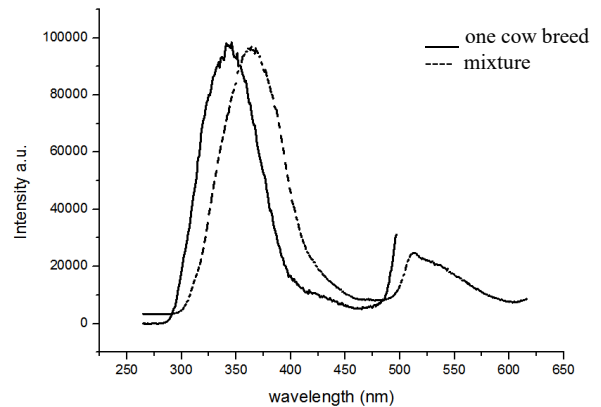


Fig.7. Emission wavelength difference between one cow breed and cow mix when irradiated at 255 nm.

Essential is the moment on Fig.7. In fluorescence diagnosis with respect to testing cow's milk is that the milk taken from a cow breed differs in wavelength of fluorescence with that which is a mixture of several breeds. This means that the method of fluorescence is suitable for determining the quality of milk for a selected breed.

3.2 Fluorescence spectroscopy of apple juice

Thirty two samples from fruit juices were investigated. They were from four different fruits and nine different well known brands that are widely distributed in the shop industry. The juices were drained using a juicer for 4 different fruits as it does not add additional impurities. In Fig. 8 we can see a difference between emission wavelength of freshly squeezed apple juice and emission wavelength of apple juice of market. At apple juice of market have two relative emission wavelengths, which are located exactly on both sides of emission wavelength of freshly squeezed apple juice, this fact shows that the impurities that are added to the juice, which are commercially available are those which are close to the composition of freshly squeezed juice.

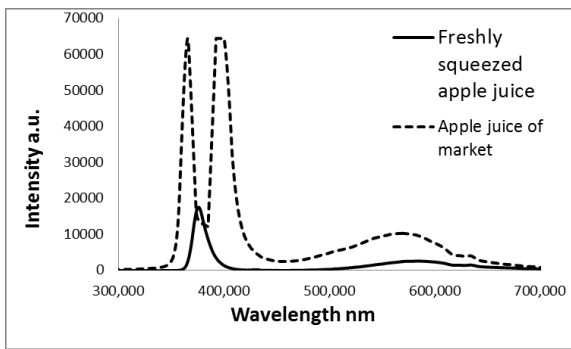


Fig.8. Fluorescence spectroscopy apple juice for excitation at $\lambda = 250$ nm.

3.4 Fluorescence spectroscopy of olive oil

Twenty four samples of different brands Spanish and Greek olive oils sold in supermarkets were investigated. The fluorescence of the samples was studied by exciting them with LEDs emitting at 370 nm, 395 nm, 425 nm and 450 nm. A 90° geometry of light detection in a 10×10 mm cuvette was used. Samples were studied without any preliminary solution.

The conclusion of the fluorescence analysis Pure olive oil there would be emission wavelength of colza. In physicochemical analyzes are difficult to distinguish between olive oil, which has impurities of colza and virgin olive oil. In Fig.9 is presented the difference between Pure olive oil and Extra virgin olive oil. There is a difference between the emission wavelength both types of olive oil.

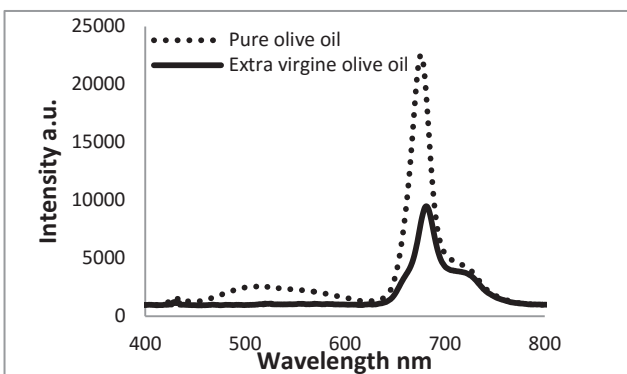


Fig.9. Fluorescence spectroscopy olive oil for excitation at $\lambda = 425$ nm.

4. Conclusions

Advanced is fast and easily accessible method for field fluorescence spectroscopy of liquid samples from cattle. The optimal wavelength for fluorescence analysis of diseased ruminants is 245 nm. The fluorescence method is suitable for early diagnosis of samples for the presence of bacteria in

cattle and before the presence of symptoms of the disease, which causes the respective bacteria.

Bacterium *Pasteurella multocida* was detected by fluorescence analysis in the body of a sheep, which is dangerous for the human organism. By fluorescence analysis autopsies of sick animals can be avoided with harmless to the human organism method as most autopsies of animals are harmful to humans.

Optimal wavelength for fluorescence analysis of pure milk from ruminant animals is 255 nm. The fluorescence method is suitable for analysis of both individual samples of milk, and mixing samples (from different animals cow, goat or sheep). The fluorescence method is suitable for detection if milk from a cow breed or mix of several breeds (this would represent the interests for producers who select the breeds).

Optimal wavelength for fluorescence analysis of apple juice is 245 nm. At this wavelength differences are detected most easily in the composition of freshly squeezed juices and those that are available in stores. The method of fluorescent analysis of juices is fast and efficient. With a rich enough library can be done effective analysis of the purity of the natural juices offered by many companies. Samples of their products compared with freshly squeezed juice from the respective plate.

Optimal wavelength for fluorescence analysis of olive oil is 425 nm at this wavelength is most clear the difference between the composition of Pure olive oil and Extra virgin olive oil. Fluorescence analysis of olive oils is a fast and efficient way to establish the purity of olive oil.

Fluorescence analysis is a quick and efficient method for accurate analysis both in veterinary medicine and food industry. This method can successfully replace so far used chemical and physical methods. This is an easy and relatively inexpensive method of analysis, since it does not require additional supplies and one unit can check multiple samples. In most cases it has been shown that the fluorescence assay is more accurate than many of the physics-chemical assay, as impurities that are added in a product is similar in chemical composition to the natural, and only the fluorescence analysis can perform accurate and adequate contrast between the samples.

REFERENCES

1. Nora E. Sobero' n, Virginia S. Lioy, Florencia Pratto, Andrea Volante and Juan C. Alonso (2011) Molecular anatomy of the *Streptococcus pyogenes* pSM19035 partition and segrosome complexes Vol. 39, No. 7, 2624–2637 Nucleic Acids Research

2. Zhiyi Liu, Heng Shi 1, Le Liu, Sunan Deng, Yanhong Ji, Suihua Ma, Hui Ma and Yonghong He Line-Monitoring, Hyperspectral Fluorescence Setup for Simultaneous Multi-Analyte Biosensing *Sensors* 2011, 11, 10038-10047
3. Pratto,F., Suzuki,Y., Takeyasu,K. and Alonso,J.C. (2009) Single-molecule analysis of protein·DNA complexes formed during partition of newly replicated plasmid molecules in *Streptococcus pyogenes*. *J. Biol. Chem.*, 284, 30298–30306.
4. D. J. D'Amico 1 and C. W. Donnelly. Microbiological quality of raw milk used for small-scale artisan cheesenproduction in Vermont: Effect of farm characteristics and practices (2010) *J. Dairy Sci.* 93 :134–147
5. Mabmouri, O., Rouissi, H., Dridi, S., Kammoun, M., De Baerdemaeker, J., & Karoui, R. (2008). Mid infrared attenuated total reflection spectroscopy as a rapid tool to assess the quality of Sicilo–Sarde ewe's milk during the lactation period after replacing soybean meal with scotch bean in the feed ration. *Food Chemistry*, 106, 361–368.
6. Karoui, R., Dufour, E., & De Baerdemaeker, J. (2007). Monitoring the molecular changes by front face fluorescence spectroscopy throughout ripening of a semihard cheese. *Food Chemistry*, 104, 409–420.
7. Zhiyi Liu, Heng Shi 1, Le Liu, Sunan Deng, Yanhong Ji, Suihua Ma, Hui Ma and Yonghong He 8. Line-Monitoring, (2011) Hyperspectral Setup for Simultaneous Multi-Analyte Biosensing *Sensors* 11, 10038-10047
9. Pratto,F., Suzuki,Y., Takeyasu,K. and Alonso,J.C. (2009) Single-molecule analysis of protein·DNA complexes formed during partition of newly replicated plasmid molecules in *Streptococcus pyogenes*. *J. Biol. Chem.*, 284, 30298–30306.
10. Dhammika N. Atapattu, Nicole A. Aulik, Darrell R. McCaslin, and Charles J. Czuprynski, Brief heat treatment increases cytotoxicity of *Microb Pathog.* (2009) *Mannheimia haemolytica* leukotoxin in an LFA-1 independent manner *March*, 46(3): 159–165

"Dept. Optoelectronics & Lasers", Techn. Univ.-
Sofia, Branch Plovdiv and Scientific Laboratory
"Quantum and Optoelectronics", R&D
Devison, Techn. Univ.-Sofia
Phone (032/ 659 752):
E-mail: vania_plachkova@abv.bg



ELECTRONIC CONVERTER FOR FIBER OPTIC THERMOMETER

SVETOSLAV IVANOV¹, YANKA IVANOVA²

Abstract: *Fiber optic sensors working on the reflectometric principle are used to measure temperature, liquid level and concentration of solutions. This article shows the results of the tests of an especially constructed fiber optic thermometer for measuring temperature. The transient's processes in the pulse optoelectronic transducer, converting temperature into an analog signal, have been studied. The optical receiver amplifies the difference between the amplitudes of the signals, received from the fiber optical sensor and from the optical fiber, transmitting the optical signal directly from the source. A method is proposed for digital linearization of the transfer characteristic of the converter. The constructed fiber-optic thermometer allows measurement of temperatures ranging from 20 °C to 150 °C with measurement accuracy of 1 °C.*

Key words: *fiber optics, sensors, electronic transducers, temperature sensors.*

1. Introduction

Fibers have many uses in remote sensing. Depending on the application, fiber may be used because of its small size, or because no electrical power is needed at the remote location, or because many sensors can be multiplexed along the length of a fiber by using light wavelength shift for each sensor, or by sensing the time delay as light passes along the fiber through each sensor. Lower temperature targets say from 100°C to 400°C can be measured by activating various sensing materials such as phosphors, semiconductors or liquid crystals with fiber optic links offering the environmental and remoteness advantages. Fiber-optic sensors, working on the reflectometric principle with external modulation of the luminous flux, can be used for measuring temperature. When such sensors are used, the amplitude of the intensity of the light pulses passing along the fiber most frequently changes. [1,2]. Based on a U-shaped bending, sensors for measuring liquid levels, temperature and concentration of solutions have been created, following the reflectometric principle. Schemes for optical registration have been created, in which conversion of light pulses into electrical occurs [3,4]. After that their demodulation by an amplitude detector follows. The resulting analog signal from the output of the detector enters the input of the analog-to-digital converter (ADC) [5,6]. Additional linearization of the parameters of the sensors is typically needed for increasing measurement accuracy.

The purpose of the studies carried out and the obtained results is to create an analog converter with a DC output signal, proportional to the amplitude value of both the electrical and, respectively, of the light pulses. Provided that the transmission characteristic of the sensor is a monotonously increasing function, linearization of the measurement results is possible in the temperature range of measurement.

2. Schematic diagram of the electronic converter

The schematic diagram of the created converter is shown in Figure 1. The generator G generates rectangular electrical impulses with constant frequency and constant duty cycle. These impulses control both the LED and the amplitude detector. The light pulses, which have passed through the optical fiber and have been modulated by the sensitive element S, go to the converter, which converts current into voltage. A PIN photodiode is used as an optical receiver. The electrical impulses from the output of the converter A go to the input of the detector. The input of the detector is an integrating circuit - RI , CI , whose output is connected to the input of the first circuit sample/hold S/H I. The output of the detector is an integrating circuit - $R2$, $C2$. The control of both schemes sample/hold (SH) is carried by the

generator of electrical impulses. The first scheme “remembers” or holds the maximum value of the output voltage from the output of the integrating circuit during the pulse, received at its input. The second scheme (S/H II) is designed to maintain this peak value as a permanent signal at its output.

The LED on the transmitter is connected via a coupling with two quartz optical fibers FO1 and FO2. The optical fibers are with a gradient refractive index with a cross section 50/125 μm .. One of the fibers is connected to the sensor for measuring temperature, and the other carries light directly.

By the use of the two fibers the influence of the change in intensity of the light, coming out of the LEDs during the process of measuring, is eliminated. The electronic converter amplifies the difference between the received signals from the two photodiodes.

The optical fiber with the sensing element is a fiber with a U-shaped bending, contained within the sensor. Part of the surface of the U-shaped bending is with a removed optical cover, the remaining optical core is placed in a thermal optical medium.

The used thermal optical medium is a semitransparent polymer, whose refractive index is changed by the temperature.

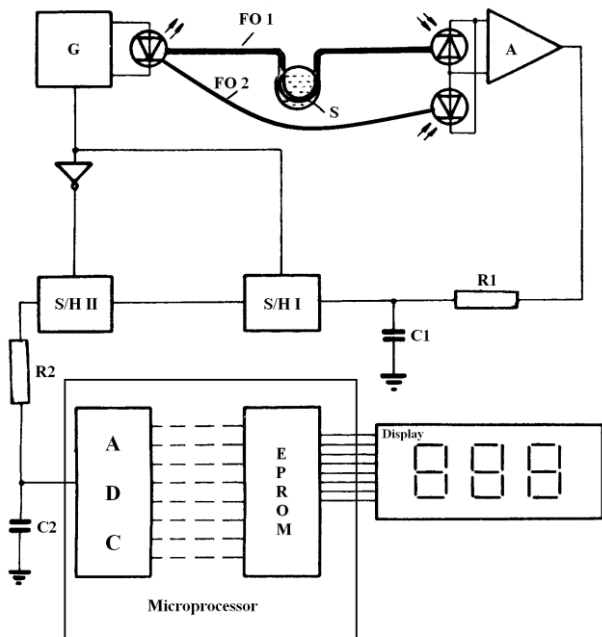


Fig. 1. Schematic diagram of the studied fiber-optic thermometer

The input of the optical receiver is differentiating. It consists of two PIN photodiodes, connected both counter and in parallel. The two optical fibers, coming from the transmitter, and the sensor, are connected to the respective photodiodes.

3. Transient processes in the integrating circuit

Upon receipt of rectangular pulses with a constant frequency and constant duty cycle at the input of the integrating circuit R_I, C_I , a transition process with permanent extreme values is set at its output. The maximum value of the voltage U_{max} while charging the capacitor and the minimum value during its discharge U_{min} are proportional to the amplitude of the input pulses. The maximum value of the voltage to which the capacitor is charged U_{max} is the value which is maintained constant by the schemes sample/hold at the output of the amplitude detector. This value is equal to:

$$U_{max} = K \cdot U_i, \quad (1)$$

where: - U_i is the amplitude value of the output voltage of the electronic converter

$$\text{- coefficient } K = (1 - e^{-t_A}) \quad (2)$$

The coefficient t_A is expressed by the parameters of the received input pulses at the input of the amplitude detector in the following way:

$$t_A = \frac{\alpha}{f \cdot T_c} + \ln \left(\frac{1 - e^{\frac{1}{f \cdot T_c}}}{1 - e^{\frac{1-\alpha}{f \cdot T_c}}} \right), \quad (3)$$

where: α - duty cycle of the pulses;

f - the pulse frequency;

T_c - the time constant of the integrating circuit R_I, C_I .

The coefficient K is the coefficient of transmission of the amplitude detector. Figure 2 shows a graph of the dependence of the coefficient K on the transmission frequency of the input pulses. The calculations have been made at constant duty cycle of the input pulse $\alpha = 0,3$ and time constant of the integrating circuit equal to $T_c = 0,027\text{s}$.

Figure 3 graphically shows the dependence of the coefficient of transmission K on the duty cycle of the received pulses at constant pulse frequency equal to 70Hz. The time constant of the integrating circuit is equal to $T_c = 0,027\text{s}$.

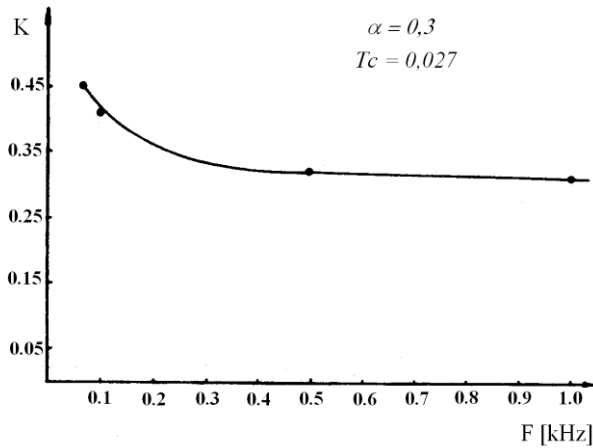


Fig. 2. Variation of the coefficient of transmission of the amplitude detector on the frequency of control impulses

An advantage of the integrating circuit in its role of an input for the detector is that it greatly reduces the high-frequency noise accompanying the signal at the input of the scheme.

The analyses show that the coefficient of pulsation of the output voltage is equal to $K_p=0,09\%$, with variation of the output voltage in the range from 0.5 to 3,5V.

An analogue-to-digital converter is a part of the microprocessor used for the linearization of the transfer characteristic of the sensor and the control of the digital indication.

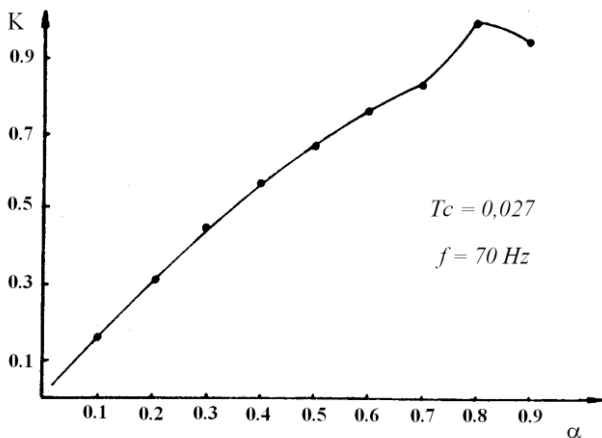


Fig. 3. Amending the coefficient of transmission of the duty cycle of the control pulses

3. Block for digital linearization of the characteristic

The digital code from the output of the analog-to-digital converter (ADC) is delivered to the address buses of the permanent memory

(EPROM). Thus, by means of the generated address a memory cell is addressed, in which the real value of the measured quantity has been recorded in advance. If we denote the value of the measured quantity by θ , and the readings from the output of the ADC by X , the graph of the relationship $X = F(\theta)$ for a fiber-optic temperature sensor has the type, shown in Figure 4. With the help of EPROM the inverse function $\theta = F^{-1}(X)$ is realized. During a memory write each value of the measured quantity θ_n is recorded in a range of cells in the memory with addresses for θ_n

$$\text{from address } \frac{x_n + x_{n-1}}{2} \text{ to } \frac{x_n + x_{n+1}}{2},$$

where $x_n = F(\theta_n)$, $x_{n-1} = F(\theta_{n-1})$ и $x_{n+1} = F(\theta_{n+1})$

The graph of the inverse function with the intervals, in which $\theta_n = const$ is shown in Figure 5. The determination of the addresses, in which the values of the measured quantity will be recorded, can also be done analytically.

$$x_n - \frac{1}{2 \cdot \theta' \cdot (x_n)} < x < x_n + \frac{1}{2 \cdot \theta' \cdot (x_n)} \quad (4)$$

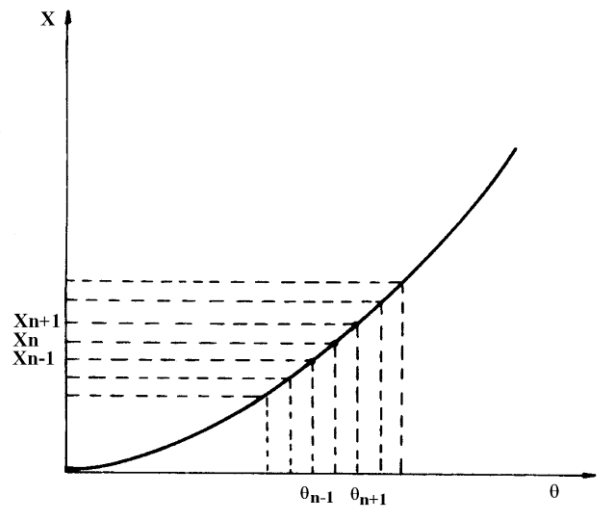


Fig. 4. Amending the output data of the ADC as a function of temperature

The latter inequality determines the range of address values x , in which the value θ_n value θ_n for respectively x_n must be recorded. Thus the maximum value of the absolute error in the entire measurement interval will be constant and equal to the half of the smallest significant digit in the process of measurement.

The constructed fiber-optic thermometer permits measuring temperatures ranging from 20 ° C to 150 ° C. The accuracy of measurement is 1 ° C. The scheme of linearization of the sensor characteristic allows

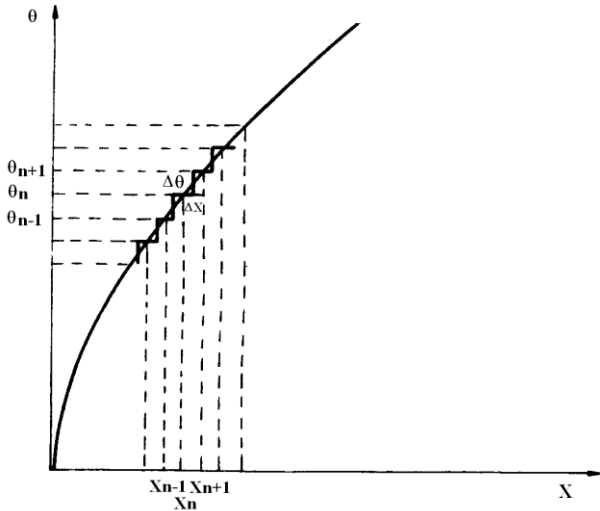


Fig. 5. Graph of the inverse function of measuring temperature

for achieving an absolute error of $\pm 0,5$ ° C in the entire range of measuring.

The fiducially percentage error at a maximum value of the measured quantity 150 ° C, is equal to 0.033%.

4. Conclusion

In constructing a detector for amplitude modulated pulses extreme values of the transient processes in an integrating circuit can be used with success. The maximum value of the output voltage from the integrating circuit can be established as a permanent signal by using two circuits connected in series - "sample /hold". The theoretical calculations and the experimental results prove the constant coefficient of transmission of the detector and the low pulsation of the output voltage while detecting pulses with a constant frequency and constant duty cycle.

In circuits for digital linearization of sensor characteristics it is appropriate to use EPROM memory. If the characteristic is a strictly increasing or a strictly decreasing function, linearization is

possible, in which a constant absolute error is obtained for the entire range of measurement.

The constructed opto-electronic transducer has been implemented to a fiber-optic thermometer, measuring temperatures ranging from 20 ° C to 150 ° C. The characteristic of the sensor is non-linear and a strictly growing function. The resultant absolute error is $\pm 0,5$ ° C in the entire range of measurement.

REFERENCES

1. K.T. Grattan, T. Sun, Dr (2000), Fiber optic sensor technology: an overview, *Sensors and Actuators A: Physical*, Volume 82, Issues 1–3, 15 May 2000, Pages 40–61.
2. Feng, X., Zhang, X., Sun, C., Motamedi, M., and Ansari, F., Stationary Wavelet Transform Method for Distributed Detection of Damage by Fiber-Optic Sensors (2014), *Journal of Engineering Mechanics*, April 2014, Vol. 140, No. 4
3. International Electrotechnical commission. *Calibration of Optical Time-Domain Reflectometers (OTDR)-Part 2: OTDR for Multimode Fibres*; Nr. 61746-2; International Electrotechnical commission: Geneva, Switzerland, 2010.
4. Koyamada, Y.; Eda, Y.; Hirose, S.; Nakamura, S.; Hogari, K. Novel fiber-optic distributed strain and temperature sensor with very high resolution. *IEICE Trans. Commun.* **2006**, E89-B, 1722–1725.
5. Nazarathy, M.; Newton, S.A.; Giffard, R.P.; Moberly, D.S.; Sischka, F.; Trutna, W.R., Jr.; Foster, S. Real-time long range complementary correlation optical time domain reflectometer. *J. Lightw. Technol.* **1989**, 7, 24–38.
6. Anderson, D.; Bell, F. *Optical Time Domain Reflectometry*; Tectronix: Beaverton, OR, USA, 1997; p. 61.

Authors' contacts Address:

1) E-mail: etehsv@gmail.com

2) E-mail: yankakiss777@abv.bg



ON ADVANTAGEOUSLY USE OF GAUSSIAN BEAM END DIODE LASER PUMPING OF ND:YAG LASER IN TWO-WAVELENGTH COAXIAL ARCHITECTURE

MARGARITA DENEVA

Abstract: *We propose to take advantage of the specific light intensity distribution in the pump beam when the cylindrical active Nd:YAG rod laser is end-pumped by a Gaussian beam diode laser to create a new effective two-wavelength laser. The essentially high pump intensity in the central part of the rod increases the pump power density there ~ 3-4 times (depending of chosen radius) in comparison with the peripheral rod part. The increasing strongly favors the generation of weaker lasing lines in the axial part. This advantage is utilized by employing coaxial two-channel laser architecture, where the optically separated axial and peripheral parts of the rod generate each in its own spectrally selective resonator. Beside the competition-less generation at two wavelengths (lines), other specific and essential advantages of the proposed laser solution are: i) to produce emissions at a weaker line (generated in the axial part) and at a stronger line (in peripheral part), simultaneously and with equalized energy without loss of pump energy for the equalisation; ii) the two emissions are produced and emitted naturally in coaxial beams using the full rod volume. Except the argued proposal, we considered in details theoretically the action of such laser on the example of generation of the pair lines - of the weaker at 1.36 μm and the strongest at 1.06 μm respectively and show the expected advantages. Such laser can be applied effectively for generation of pair of Nd:YAG lines from 1.06, 1.32, 1.34, 1.36 and 1.44 μm and also for the generation in other materials (including Q-switching).*

Keywords: *Gaussian beam diode pumped laser, two-wavelength generation, Nd:YAG laser, coaxial geometry laser architecture, independent generation control*

1. Basics

In spectroscopic and optical instrumentations laser light at two or more wavelengths, in particular, at two different emission lines (or colors) and with tunability are employed. Such light is of essential interest in space composition determination by the technique of Differential Absorption Spectroscopy (DAS), including distant monitoring of gas pollutants and aerosols in the atmosphere (LIDAR) [1-5], in spectroscopy, in nonlinear optics [6-8], in metrology [9], in bio-medical investigations and treatments [10]. The superposition of the beams of two lasers is a simplest way to obtain two-wavelength laser light [11]. However, such a realization need two lasers, special filters for neighboring wavelengths and is difficult for synchronization of two sub-systems for operation using two nanosecond-timed Q-switchers.

A simpler and cheaper way to obtain controlled two-wavelength laser emission is to use a single laser that generates the two wavelengths in a single exciting pulse or in CW emission – i.e. so-called two-wavelength (dual line, dual color) laser with a flashlamp (for high energy output) or diode pumping. The detailed review of such lasers is given in [12]. The diode pumping is of essential interest assuring compact technical solutions and high energetical effectiveness for CW and pulsed regime [13-15]. Especially, high simplicity, compactness and effectiveness are for the case of so-called diode laser end-pumped Nd:YAG cylindrical laser. Such pumping consists of introduction of the entire pump light into the crystal and along its axial rod direction.

One effective way to produce the two emissions is to use a single laser rod, in which each wavelength is generated in separated pumped parts of the rod volume. The dual-wavelength generation

in separated parts has the advantage to avoid so-called “wavelength competition effect” that is typical and leads to a problem when the two wavelengths are generated in common parts of the rod [16, 17, 14]. Our scientific group, on the base of new development of our patented principle, has developed and published an effective, coaxial geometry, flash-lamp pumped two-wavelength Nd:YAG laser [12]. We have used the focalizing effect of the cylindrical rod of the side pumped flash-lamp light and have shown the effectiveness of application of the coaxial two-channel architecture for the case.

The aim of this work is to introduce and show the attractiveness and usefulness of the application of the coaxial geometry laser solution for the case of the diode end-pumped beam with Gaussian distribution. The combination of this two circumstances strongly favors the creation of effective laser of the noted type as we shall show in the work.

2. Principle and scheme of the proposed laser

The basic optical scheme of the proposed Nd:YAG two-wavelength coaxial-geometry laser pumped by diode laser with Gaussian beam is given in Fig. 1.

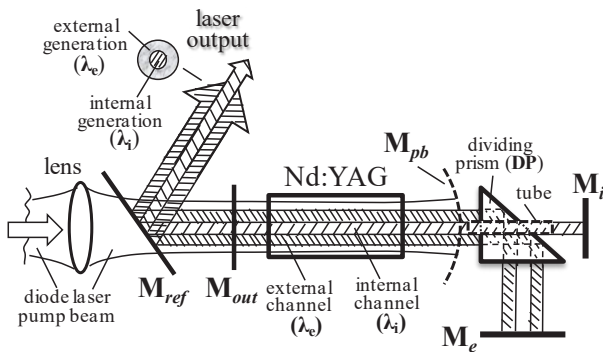


Fig. 1. Optical scheme of the laser with channel separator prism *SP* with hole.

In the description of the principle and in the consideration, given below, we use the following combination of elements for the Nd:YAG diode pumped set-up: Nd:YAG cylindrical rod with diameter $d = 0.5$ cm, length $l = 0.5$ cm and concentration of Nd^{3+} -ions of 0.6 at %. (in the scheme in the figure - Nd:YAG). Note that in the consideration we will use the parameters of our available specialized laser crystal designed for the diode pumping. The pumping is by the pulses of diode laser at $0.808 \mu\text{m}$ with Gaussian intensity distribution and energy of 120 mJ and pulse duration of $\sim 200 \mu\text{s}$. The input pump beam passes through lens with suitable focal length to form the spot at

the entrance of the Nd:YAG rod with diameter of ~ 0.5 cm, containing $\sim 99\%$ of the incident beam power P_{tot} (i.e. 1.6 times higher than the Gaussian radius in this cross-section following, e.g. [13]). The length of the lens assures the Rayleigh length essentially larger than the rod length l . Thus we can consider the beam ray propagation into the crystal as near parallel to the crystal axis. Before to enter into the rod, the pump beam passes through a dichroic mirror M_{ref} (with reflectivity 0.86 for $1 - 1.4 \mu\text{m}$, and ~ 0.1 for $0.808 \mu\text{m}$). The Nd:YAG laser output mirror M_{out} is also dichroic (reflectivity 0.86 for $1.32 - 1.36 \mu\text{m}$ and 0.1 for $1.06 \mu\text{m}$ and $R \sim 0.1$ for $0.808 \mu\text{m}$). At the opposite end of the crystal is suitably adjusted near-plane dichroic mirror M_{pb} with radius of curvature few meters (reflectivity ~ 0.1 for $1 - 1.4 \mu\text{m}$ and ~ 0.86 for $0.808 \mu\text{m}$). The generation at the two wavelengths (scheme in Fig. 1) is produced in two coaxially disposed and optically separated parts of the described Nd:YAG laser rod, for each part being formed its own spectrally-selective resonator with length of each $L \approx 10$ cm. As simplest optical separation of the two rod parts with their resonators (hatched differently in Fig. 1) is used a dividing rectangular prism (DP). It is with 0.38 cm diameter bore hole and length of 0.4 cm through the cathetus and the axis, coinciding with the rod axis. In the hole a very thin (~ 0.1 mm wall) metal tube with 2 cm length is introduced, which detail assures very good separation of both channels as we have shown in our previous work [12]. The prism is declined at $\sim 5^\circ$ with respect to the crystal axis to avoid the back Fresnel's reflection. In the external, peripheral part of the active rod and its resonator (external channel), the light at the first wavelength λ_e is generated. The second wavelength λ_i is obtained in the coaxial, round cross section, internal rod part, and its corresponding resonator (internal channel). We will consider here the scheme in simplest manner with two flat-flat Selective Mirror Resonators (SMRs) using, as the common output coupler, the dichroic mirror M_{out} . The end mirror M_e of the external resonator for generation at $\lambda_e = 1.06 \mu\text{m}$ is with $R_e = 0.99$ for $1.06 \mu\text{m}$ and the end mirror M_i of the internal resonator for generation at $\lambda_i = 1.36 \mu\text{m}$ with $R_i = 0.99$ for $1.36 \mu\text{m}$. The last mirror is with less than 0.70 reflectivity for the other wavelength and AR ends coated for the $1.06 \mu\text{m}$. By replacing both end mirrors with suitable reflectivity we can change the generated line in the corresponding channel (if the pumping is sufficient). The ends of the crystal are specially AR coated for the wavelengths in the range $0.9 - 1.4 \mu\text{m}$. As we have realized for the case of flashlamp pumping coaxial geometry laser [12], the end mirrors M_i and M_e can be replaced

with suitable prism-selective blocks and for sufficiently pumping energetic characteristic may obtain in each channel tunable generation.

3. Theoretical modeling and treatment of the two-wavelength operation of the proposed laser

Due to essential difference for the spectroscopic data for Nd:YAG absorption cross section at $0.808 \mu\text{m}$, we have realized our proper estimation of our crystal absorption. We measured the absorption of the crystal for $0.594 \mu\text{m}$ wavelength of our yellow He-Ne laser. Taking the measured absorption and the graphs of the absorption of Nd:YAG as a function of the wavelength, given in the representative literature [13, 14], we have evaluated that for our crystal the absorption coefficient for pumping wavelength of $0.808 \mu\text{m}$ is 3.9 cm^{-1} . Having all necessary data, given already, we have used the following approach. Firstly, we have calculated, in our point of view - correctly, the pumped power distribution in the laser rod.

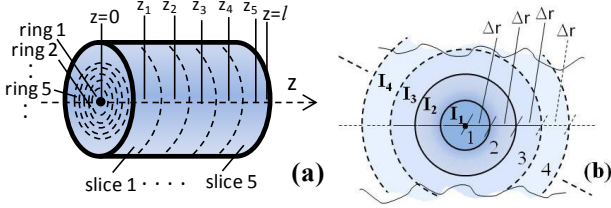


Fig. 2. Schematic Nd:YAG rod division: (a) longitudinal in slices (b) and radially in rings (b); $z=0$ and $z=l$ are the front and rear surfaces of the crystal, respectively.

Schematically for the theoretical treatment, the Nd:YAG rod is divided radially at rings and longitudinally in slices. This division is clarified in Fig. 2. The cross-section of the rod is divided at 8 rings with equal increasing radii for each sequent by Δr (see Fig. 2b). Along the axis Z of the crystal the division is at 5 slices each with equal length $\Delta l=l/5$. For our treatment we consider the cylindrical Nd:YAG crystal with concentration $\sim 0.6\%$ Nd^{3+} ions, length of 0.5 cm and cross-section diameter of 0.5 cm . Respectively we have $\Delta l=0.1 \text{ cm}$ and $\Delta r=0.0625 \text{ cm}$. In considered by us case, the pumping diode laser power is 120 mJ in a near rectangular pump pulse with duration of $\sim 200 \mu\text{s}$. The pump beam is focalized with appropriate lens in manner that the focalized total beam power drops at the ends of the Nd:YAG rod down to 2% (ratio of the rod diameter and pump Gaussian beam diameter is ~ 1.6). Thus, in practice the total pump power is used to pump the crystal.

We will provide a theoretical consideration for the noted already data. The aim of our consideration will be to find the limits of the ring around the optical axis, where the output energy at the wavelength $1.36 \mu\text{m}$ is comparable with this one generated at $1.06 \mu\text{m}$ in the external rings. We expect that it is possible due to the essentially different cross-section of the stimulated emission for the two wavelengths and the difference in the pumping power (intensity) in the central axial part and peripheral part of the crystal.

Let's first consider the distribution of the pumping for each cylinder formed by border lines parallel to the axis of the crystal and closed by the pair of rings at the front and rear sections of the crystal (partial cylinders).

The maximal intensity I_0 in the center of the beam at the entry plane of the rod was calculated on the base of well known relation:

$$I_0 = \frac{2P^{\max}}{\pi \cdot \omega^2} \quad (1)$$

where P^{\max} is the maximal pump power in the pulse and ω is the Gaussian radius of the pump beam. The intensity for each partial cylinder is calculated for the middle of the ring radius ($r_i = (2 \cdot i - 1) \cdot (\Delta r / 2)$, $r = 1, 2 \dots 8$) and is assumed to be homogeneous for the ring. The points on the graphics in Fig. 3 present the values for the corresponding rings.

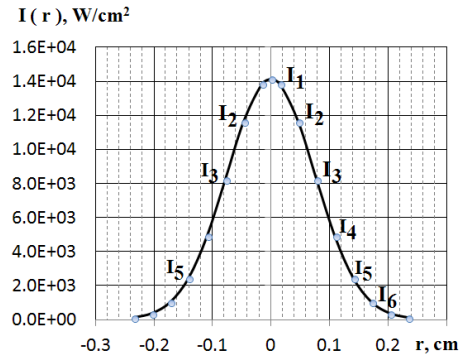


Fig. 3. Calculated distribution of the initial intensity (I_i) in cross-section at the entrance for the corresponding ring (i - see Fig. 2), on the axis X is given the distance r from the center ($r=0$) of the beam cross-section.

The distribution of the pump power for each ring, the area of the corresponding ring and pump intensity are presented in Fig. 4.

In the calculations for propagation along the Z axis, we have used the Bouguer's law, taken into account the effect of absorption saturation. We have evaluated the change of the pump intensity correspondingly for each ring, slice by slice - forward and backward after reflecting by the mirror M_{pb} . For each slice we sum forward and backward values, accepting that for any ring of the slice the

intensity is constant, equal to this one at the half length of the slice. We assume a total input in the

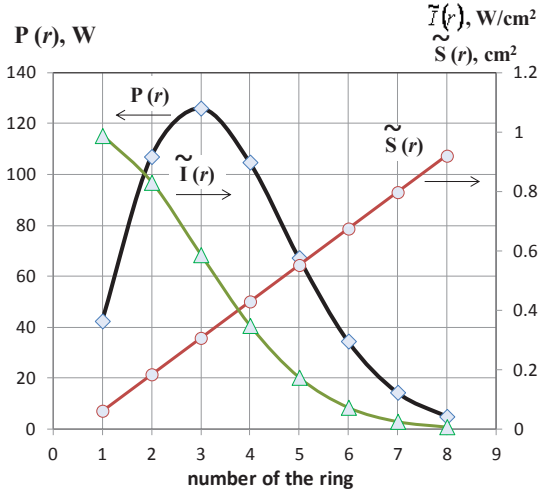


Fig. 4. The pumping power $P(r)$ corresponding to each ring. The corresponding starting pump intensity $I(r)$ and the area $S(r)$ of the ring are also given. Here $\tilde{I}(r) = I(r)/1.4 \cdot 10^4$ and $\tilde{S}(r) = S(r) \cdot 20$.

crystal pump power $P_{tot} \approx 100$ mJ / 200 μ s, Gaussian beam intensity distribution. In Fig. 2(a) z_i ($i = 1, 2, \dots, 5$) corresponds to the middle of the length's slice, where the result intensity is calculated. The values of the obtained intensity I_{r,Z_i} for the first five rings, where generations exist, are presented in Fig.5.

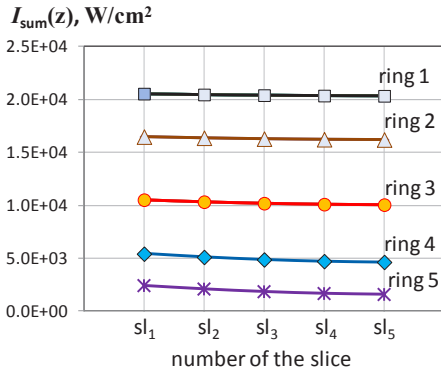


Fig. 5. The result pump intensity for the first five rings obtained after summing the forward and backward values, calculated slice by slice. On axis X are plotted the number of the slice sl_i , $i = 1, 2, \dots, 5$ (in the middle in which the results are obtained).

The calculation graphics given in Fig.5 show for each partial cylinder practically the same sum pump power (and energy) - uniformity both in slices and in the cross section - i.e. in the corresponding ring. This uniformity determines the fact that we can accept homogeneity of the pumping with different energy density for the partial cylinders (the change in the intensities from the first to the last slice for the corresponding cylinder is of order of some percentages and practically could be

accepted as constant for each ring). On this base, we can calculate the generation for each partial cylinder applying the set of rate differential equations [14] adapted for the case, presented in next point. The considerable difference in the pump power for the rod parts around the axis and in the peripheral areas, enabling us to expect generation at weak lines (with smaller values of the emission cross-section) comparable to that at the stronger lines, including the main 1.06 μ m, when it is generated in peripheral part of the crystal. In the analysis below, we show this possibility in the case of the line 1.36 μ m, generated in near axial part and the line 1.06 μ m generated in the peripheral part of the crystal.

4. Analysis of the laser generation

Taking the obtained results for the approximately constant value of the pump energy into slices along the axis of the crystal - i.e. possibility to assume homogeneity of the excitation (with different energy density for the corresponding ring), we apply the adapted set of differential rate equations [14] for the analysis. The adapted system that describes our case and assumed parameters is:

$$\begin{cases} \frac{dN_2}{dt} = R_p^{(e,i)}(t) - B^{(e,i)} \cdot q^{(e,i)} \cdot N_2 - \frac{N_2}{\tau_c^{(e,i)}} \\ \frac{dq^{(e,i)}}{dt} = V_a^{(e,i)} \cdot B^{(e,i)} \cdot q^{(e,i)} \cdot N - \frac{q^{(e,i)}}{\tau_c^{(e,i)}} \end{cases} \quad (2)$$

$$\text{with } P_{out}^{(e,i)}(t) = (\gamma_{out}^{(e,i)} \cdot c / 2L') \cdot h\nu^{(e,i)} \cdot q^{(e,i)}(t) \quad (3)$$

Here $B^{(e,i)} = \sigma_{21}^{(e,i)} \cdot l \cdot c / V_a^{(e,i)} \cdot L'$, s^{-1} ;
 $\sigma_{21}^{(e)} = 2.8 \times 10^{-19}$ cm²; $\sigma_{21}^{(i)} = 0.6 \times 10^{-19}$ cm² - emission cross-section of the Nd:YAG laser crystal for $\lambda_e = 1.06$ μ m and $\lambda_i = 1.36$ μ m [12], l - crystal length; $c = 3 \times 10^{10}$ cm/s is the light velocity; $L' = 10.42$ cm is the optical length of the resonator. The lifetime of the upper laser level is $\tau = 230$ μ s. The term in (3) $h\nu^{(e,i)} \sim 10^{-18}$ J is the energy of the generated photons for the corresponding wavelength. The dumping time of the photon in the resonator is $\tau_c^{(e,i)} = L'/c \cdot \gamma^{(e,i)}$, where $\gamma^{(e,i)}$ describes the loss in the corresponding resonator following Ref. [14] accounting the corresponding considered ring (partial cylinder). The calculations are for the laser and pumping parameters, given already. The total number of active Nd³⁺ ions/cm³ used in the calculations is 6.9×10^{19} cm⁻³ (for 0.6 % Nd³⁺). The pump rate $R_p^{(e,i)}(t) = P_p^{(e,i)}(t) / (h \cdot \nu^{(e,i)}) \cdot V_a^{(e,i)}$ is defined on the base of the part of the optical pumping power

that corresponds to the considered ring and wavelength. According our data, the obtained pump power for each ring is plotted in graphics in Fig.5. The pump power is a function of the area and the intensity in the ring. $R_p^{(e,i)}$ is related to the temporal shape of pump energy in combination with the diode pump pulse shape. Our oscilloscope observation shows that the pump pulse can be well approximated by a trapezoid shape with a rise time of 100 ns, near-plateau part of 200 μs and fall time of 100 ns. $V_a^{(e,i)}$ is the active volume of the considered ring for the generation of the corresponding wavelength. In the calculations we have taken the geometrical data for each ring derived from the description of the laser rod separation given above.

The system is solved using the Runge-Kutta method. For the solutions we take, as initial conditions $N = 0$ and $q^{(e,i)} = 1$. Due to the practically rectangular pump pulse, the pump power in each cylindrical part during the pumping is constant. The system (2-3) is solved independently for the generation in the other considered coaxial parts. From the solution of the system, we obtain the temporal shape of the laser output power and its integration gives the output energy [14]. For the used parameters, described above, generation in five rings in total is obtained – in the first three rings the 1.36 μm wavelength is generated, and for the next two rings a generation for the 1.06 μm wavelength is produced. The type of the generations for two of the rings is presented in Fig.6.

The output energy for the wavelengths and for the corresponding partial cylinder is plotted in Fig.7.

The total output energy produced for the 1.36 μm wavelength is ~ 17 mJ and for the 1.06 μm wavelength is ~ 15 mJ. Thus, as we have shown, it is really possible to obtain practically equal energetic characteristics of both generations.

5. Discussion and Conclusion

In the work we have shown the possibility to use the specific light intensity pump beam distribution in the Gaussian beam in combination with coaxial geometry to obtain a new effective two-wavelength laser. Following our theoretical modeling and computer analysis, we have shown the potentiality for efficient two-wavelength laser operation in the Nd:YAG crystal at an weak and the strongest wavelengths with practically equal output energies. The weak line is generated in near axial part, where naturally the pump energy density is higher than the one in the peripheral part, where the

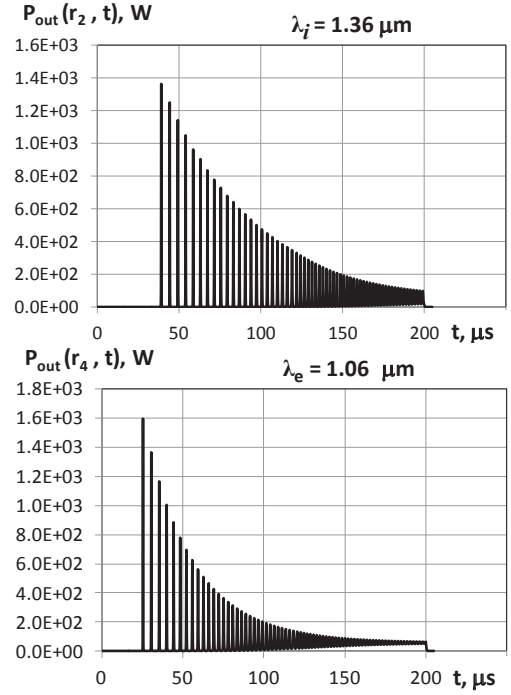


Fig. 6. The temporal generation in the second and in the fourth ring of the active medium for both wavelengths – 1.36 μm and 1.06 μm , respectively.

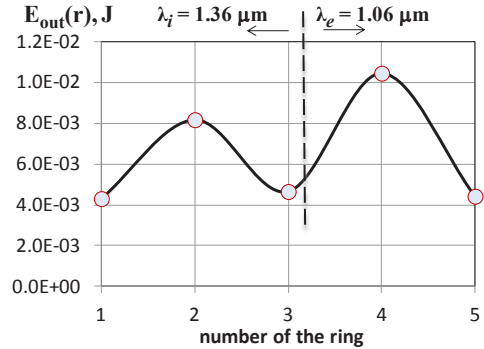


Fig. 7. Distribution of the calculated output energy for each partial cylinder for both wavelengths.

strong line is generated (difference of the pump energies $\sim 3-4$ times). On the base of the applied approach for calculation, it was obtained the limits of the size in the laser crystal where the two noted outputs are energetically equal. The case considered as an example is for the lines 1.36 μm and 1.06 μm . The analysis shows that equality of the energies for both lines can be achieved for the axial part with a diameter 0.38 cm (generation at 1.36 μm) for the whole diameter of the crystal of 0.5 cm and the other line – strongest 1.06 μm line, is generated in the remaining peripheral part. This advantage is utilized employing coaxial two-channel laser architecture, where the optically separated axial and peripheral parts of the rod generate each in its own spectrally selective resonator. Beside the absence of wavelength-competition effect [16,17], which strongly worses the two-wavelength generation, the practically equal energy emission at the weaker

and the strongest line can be produced simultaneously without loss of pump energy for the equalisation. The two emissions are produced and emitted naturally in coaxial beams using the entire rod volume. Such Gaussian beam diode pumped laser can be applied effectively for generation of different pair of Nd:YAG lines with noted advantages. The technique developed in the work can be applied also for other materials used as active medium, also for other type of suitable beam distribution for longitudinal pumping.

Acknowledgements

This work was supported partially by the contracts DNTS/Austria 01/3 (Project No. BG 07/2013) and DRILA 01/7-19 (TR PHC RILA 2011 projet No. 25197 VB France) and the Technical University of Sofia, Branch Plovdiv - R&D Laboratory "Quantum and optoelectronics". M.D. thanks to Photonics Institute, Vienna University of Technology for the possibility to carry out joint work in their laboratories.

REFERENCES

1. Spinhirne, J.D., Chudamani, S., Cavanaugh, J.F., Buffon, J.L. (1997). Aerosol and cloud backscatter at 1.06, 1.54, and 0.53 μm by airborne hard-target-calibrated Nd:YAG/methane Raman lidar. *Applied Optics*. 36(15), 3475-3490
2. Zanzottera, E. (1990). Differential absorption lidar techniques in the determination of trace pollutants and physical parameters of the atmosphere. *Crit. Rev. Anal. Chem.* 21, 279-319
3. Vaicikauskasa, V., Kuprionis, Z., Kaucikasa, M., Svedasa, V., Kabelkaa, V. (2006). Mid-infrared all solid state DIAL for remote sensing of hazardous chemical agents, Proc. of SPIE. 6214, 62140E
4. Farley, R.W., Dao, P.D. (1995). Development of an intracavity summed multiple-wavelength Nd:YAG laser for a rugged, solid-state sodium lidar system. *Appl. Opt.* 34, 4269-4273
5. Demtröder, W. (2003). *Laser spectroscopy: basic concept and instrumentation*. 3th ed. Springer, Germany
6. Rico, M.L., Valdes, J.L., Martinez-pastor, J., Capmany J. (2009). Continuous-wave dual-wavelength operation at 1062 and 1338 nm in $\text{Nd}^{3+}:\text{YAl}_3(\text{PO}_3)_4$ and observation of yellow laser light generation at 592 nm their self-sum-frequency-mixing. *Optics Commun.* 282(8), 1619-1621
7. Saha, A., Ray, A., Mukhopandhay, S., Datta, P.K., Saltiel, S.M. (2007). Littrow-type discretely tunable, Q-switched Nd:YAG laser around 1.3 μm . *Appl. Phys. B* 87, 431-436
8. Zhao, L.N., Su, J., Hu, X.P., Lv, X.J., Xie, Z.D., Zhao, G., Xu, P., Zhu, S.N. (2010). Single-pass sum-frequency-generation of 589-nm yellow light based on dual-wavelength Nd:YAG laser with periodically-poled LiTaO_3 crystal. *Opt. Express*. 18(13), 13331-13336
9. Dandliker, R., Thalmann, R., Prongue, D. (1988). Two-wavelength laser interferometry using superheterodyne detection. *Optics Letters*. 13(5), 339-341
10. Devaux, B.C., Roux, F.X. (1996). Experimental and Clinical Standards and Evolution of Lasers in Neurosurgery. *Acta Neurochirurgia, the Europ. J. Neurosurgery*. 138, 1135-1147 & the literature cited there in.
11. Tankovich, N., Lukashev, A. (2003). United states patent, Pat.No. US 6,613,042 B1, Sep. 2
12. Deneva, M., Nenchev, M., Wintner, E., Topcu S. (2015). Coaxial-geometry tunable dual-wavelength flashlamp-pumped Nd:YAG laser. *Opt. Quant. Electron.* 47:3253-3271; DOI 10.1007/s11082-015-0205-3
13. Koechner, W. (2005). *Solid-State Laser Engineering*, 6 revised and updated edition, Springer, Berlin
14. Svelto O. (2010). *Principles of lasers*, 5th ed. Springer Science+Business Media, LLC 2010, ISBN 978-1-4419-1301-2 (e-ISBN 978-1-4419-1302-9)
15. Ifflander, R. (2001). *Solid-State Lasers for Materials Processing*, Springer-Verlag Berlin. Heidelberg, New York, series in Optical Sciences. ISSN0342-4111, ISBN3-540-66980-9
16. Louyer, Y., Wallerand, J.-P., Himbert, M., Deneva, M., Nenchev, M.: Two-wavelength passive self injection controlled operation of a diode pumped cw Yb:doped crystal lasers. *Appl. Opt.* 42(27), 5463-5476 (2003).
17. Deneva, M., Uzunova, P., Nenchev, M.: Tunable subnanosecond laser pulse generation using an active mirror concept. *Opt Quant Electron.* 39, 193-212 (2007).

Author's contact

Address:

Technical University of Sofia, Branch Plovdiv

Dept. "Optoelectronics and Laser Engineering" and R&D Laboratory "Quantum and Optoelectronics"

25 Tsanko Diustabanov St.

4000 Plovdiv, Bulgaria

E-mail: mdeneva@yahoo.com
deneva@tu-plovdiv.bg



LIGHT BASED TECHNIQUES AS POSSIBILITY TO SOLVE THE POLLUTION PROBLEMS IN LASER IGNITION COMBUSTION OF GASEOUS MATER. A PILOT STUDY

MARGARITA DENEVA¹ MARIN NENCHEV¹, NADEJDA KAYMAKANOVA¹, ERNST WINTNER²,
MARK HIMBERT³

Abstract: *The work concerns the physics and techniques of laser ignition combustion of gaseous matter, in particular oriented to an important practical case – application in internal combustion engines. The problem, in the general case, is related to the formed micro-particles by the combustion that can pollute the entry windows for the laser ignition beam and also, to absorb and disperse the last in its trajectory. On the base of our new experimental results (completed with theoretical estimations) we propose and develop, with necessary argumentation, the principle of two techniques for solution of the noted problem (as a minimum – for the case of relatively low polluting fuel). The developed principle is based on the use of complex laser pulses, each pulse is with specially formed in cross-section two beams with convenient temporal, spatial and energy control. As we have shown on the base of our investigation, applying such pulses, we can realize that their light pressure is able to drive out the polluting particles away from the acting laser beam trajectory and the window. We present also some properties of the developed special type of Nd:YAG laser (on the base of our patented approach) suitable for the discussed applications.*

Keywords: *laser ignition combustion of gaseous matter, internal combustion engines, micro-particles pollution, laser light pressure*

1. Introduction.

General. The work concerns the physics and techniques of laser ignition combustion of gaseous matter, in particular oriented to an important practical case – application in internal combustion engines [1-3]. In this type ignition systems, a powerful laser beam is focused in the gas-fuel mixture in the engine cylinder and initiates an electric discharge and respectively an combustion, analogously to the use of the traditional electric sparking plugs [1,2]. This technique, in present, is from resurgent interest, due to its specific advantages and the essential progress of laser techniques.

At the present time, such laser ignition technique is with real practical applicability in combustion engines using gas mixtures that provide negligible polluting products by the explosion. In the literature are described practical applications of the laser ignition in high dimension engines with large (tens or more liters) working volumes, which employ gases, such as methane, the last producing negligible pollutions as fuel [3].

As it is discussed in the literature [e.g. 1], the laser ignition presents some useful advantages and potential. This justifies the work for its development. Note some of them such as: to affects advantageously the ignition of very lean mixtures – the COV-pi value for NO_x emission level < 200 mg/Nm³ is definitely below the curve for a corresponding electrical ignition system; generally, laser ignition tends to yield better smoothness of engine operation, also - for easier correction the moment and the point of the ignition spark, shorter ignition delays leading to shorter burning times and reducing losses (as a potential benefit of laser ignition); light weight of the ignition system and its long life, and especially - the ability to change and regulate the precise and technically highly simplified depth of occurrence of the ignition spark by changing the focus of the beam. The influences of the water vapor and humidity is decreased.

Problem. To extend the applicability of the method for laser ignition, common and essential problem that must be solved is the cleaning or better – to preserve the light input windows from the

deposition of the products, formed from the fuel combustion in the cylinders. Also, the problem with the pollutants is additionally complicated due to the partially decreasing of the intensity of the laser beam (end part) in the cylinder related with their absorption and scattering. The existence of residual pollutions in the formation of the gas-fuel in the subsequent engine cycle, as addition, decreases the intensity of the igniting laser beam. The use of fuels that produce undesired pollutants is typical case for the traditional combusting engines in the cars.

In practice, however, the decisions of the noted problems with pollutions, for the general case, are far from trivial. The simplest idea to clean the window by directly use of simple ignition laser beam is not convenient due to its drastically no suitable parameters for such purpose, as we also will show below.

Aim of the work. The present work gives a pilot study - principle, experiment, theoretical evaluations, concerning the development of general technique for solution of the discussed already problem of the pollution by the products of the fuel burning. The developed principle is to use for the ignition the laser pulses, wherein each pulse is specially formed in cross-section as a coaxial or parallel combination of two light beams with appropriate intensity, temporary and energetically parameters and directed in convenient manner to the window. As we show mainly on the experimental study (and partially, in the explications - by the theoretical estimation), such pulses are suitable to offer wanted effect. The potential for real application, in any case, is for the internal combustion engines that work with fuel, producing relatively low pollutants (e.g. without mixing of benzine and engine oil etc.).

2. Experimental results and theoretical evaluations – base of the proposition.

Generalized schematic picture of the working cylinder of combustion engine with laser ignition system is given in Fig.1. Schematically on the picture is presented also the pollutant particles of the combustion after electric discharge, the last being produced by the focalized laser beam (the notations are given in the picture). In the focal region the ignition beam light intensity must be of order 10^{11} - 10^{13} W/cm² ($\sim 10^8$ - 10^9 W/cm² on the entrance window; partially focalization).

Due to that here we plan to present a general study for the discussed in point I question, the consideration will be on the modeling, using the clearly evident and simplified examples.

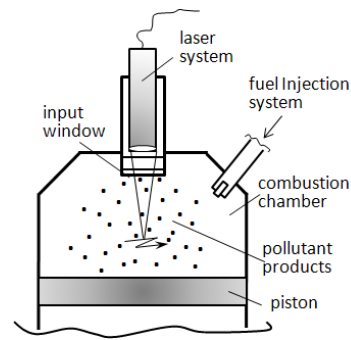
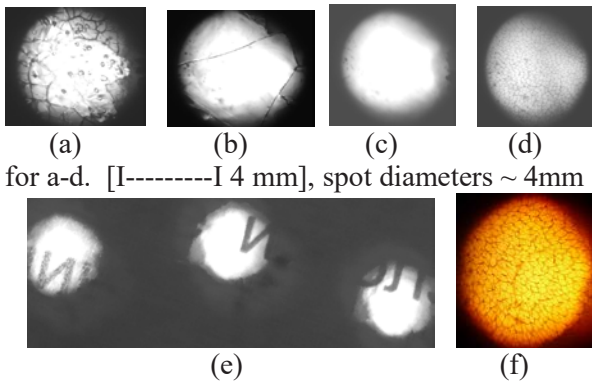


Fig. 1. Schematic picture of the working cylinder of combusting engine with laser ignition system.

The results, on the one hand give new specific scientific knowledge, on the other hand they can be used as a starting base-principle for further more detailed and concrete technical development of pollutant problem solutions. For this study we have modeled in the experiments the impurity with black thin layers on glass or quartz plates. The layers were produced by a smoke of burning paraffin candle (note that the paraffin frequently can be one of the components of the fuel). The thickness of the layer using microscope measurement is evaluated to be ~ 30 - 70 μm . As it is expected, the acting of the light depends strongly on the power density and energy density; however we have found the quantitatively such dependences.

Typical results - examples of this study are shown in Fig.2. The beam of Nd:YAG laser, needed for combustion, illuminates the layer with high power density light (on the window of $\sim 10^7$ W/cm²; 50-100 mJ/ 70 ns; 0.4 cm illuminated area). Such illumination strongly changes the layer and the support and they become definitely nontransparent. We relate this with fusion and crystallization and besides these factors with the thermo-ablation. The damage of the glass or quartz support surface is relatively low, however no reversible. For the clean support the beam with such power density passes without damages. The typical microscope picture of the plate surface in this case is shown in Fig. 2(a). Thus, the self-cleaning of the strongly polluted input window by the ignition pulses themselves is unrealizable.

In the series of the experimental studies, by varying the illumination laser beam energy, we have found that for the parameters, corresponding to the free lasing, i.e. ~ 0.7 J/350 μs (~ 5.5 J/cm²; ~ 16 kW/cm²) can be estimated parameters allowing good cleaning of the windows. The series of the microscope (low magnification) photographs are shown in Fig.2 (a), (b), (c) with given in the figure



for e – [I-----I 4mm], the spot diameter ~ 4 mm.

Fig. 2. Photographs of the cleaned areas of a layer for different illuminating beam parameters. In (a)- for high power Q -switched beam ($\sim 10 \text{ MW/cm}^2$; $\sim 70 \text{ ns}$); (b,c,d) - for free lasing ($\sim 0.3-1.5 \text{ J}$, $\sim 350 \mu\text{s}$) with illumination: (b) $\sim 35 \text{ kW/cm}^2$; $\sim 12 \text{ J/cm}^2$ (c) $\sim 16 \text{ kW/cm}^2$; $\sim 5.5 \text{ J/cm}^2$; (d) $\sim 7 \text{ kW/cm}^2$; $\sim 2.5 \text{ J/cm}^2$. In Fig.2 is shown the cases around the optimum of the cleaning, being best for the central spot where the conditions are as this one for Fig.2(c). Below the glass is disposed text with letters. Fig.2(f) shows the case of very regular crystallization (good picture, however worst cleaning).

text values. The good cleaning is shown in Fig.2(c). For energy and power density of $\sim 10 \text{ J/cm}^2$ and $\sim 30 \text{ kW/cm}^2$ we obtain very clear picture of crystallization (f).

Thus, for high power densities are typically highly prominent disturbances of the polluted window (power density of $\sim 10^8-10^{10} \text{ W/cm}^2$ on the transparent material, especially with layer of absorption impurities). To considerably lower values ($\sim 30 \text{ kW/cm}^2$, respectively close to the free generation at $\sim 0.3-1 \text{ J} / 350 \mu\text{s}$) the destruction is strongly reduced. Also, for suitable pulses in this order of the beam parameters, an acceptable cleaning of the impurities can be achieved, as it is shown in Fig.2(c). Measured light losses for the optimal case after the cleaning are $\sim 5\%$ and are related with some remaining dots on the layer. The question of damage of the windows, covered by pollutants (also in minimal quantities) is specific to each case and information in the literature is missing or presents individual situation. In any case, the obvious requirement is that the power density and energy of the passing radiation through the window must not exceed the passport data threshold of laser destruction. For the pure glass (quartz) window by catalog data typical threshold is $2-5 \text{ J/cm}^2$ for 10 ns ($\sim 10^9 \text{ W/cm}^2$) pulse, in particular for standard quality optical glass B7. Note that the laser ignition power in the focal region is $\sim 10^{12} \text{ W/cm}^2$ inside the working gas [3] and is essentially smaller (not focalized) at the window. The dependence of

the damage threshold by the light power density is related with square root of the pulse length. Our experiments have as aim outlining the general nature of the processes and expected advancement of the idea to propose approach for solution of the pollution problem. From the presented results we make the important conclusion that the cleaning of the impurities in the considered by us case need a relatively low power and energy density ($\sim 20 \text{ kW/cm}^2$; $\sim \text{few joules J/cm}^2$) light. This power correspond well to the free lasing operation of the widely used for the combustion motors ignition Nd:YAG lasers, especially in compact realization with diode pumping [5].

On the base of the presented above results, it seems that can be applied a simplest solution for laser ignition using the formed pulse that is composed by preliminary free generation and following ignition pulse. Both generations are with parameters, mentioned already and propagate in common beam. However, the cleaned particles are directed along ignition beam trajectory, thus providing energy losses for it. We propose other solution that avoids this problem. The solution is based on the experimental results given above in this point and these ones - below.

The ideas and the results of the continuation of the experiments are clarified by the series of the pictures and photographs shown in Fig.3 and Fig.4. During our experiments, described in Fig.2, we have observed other important peculiarity of the layer cleaning. Schematically our observation is illustrated in Fig.3. In Fig.3(a) we present the case of illumination of glass-plate with thin black paraffin layer and for the case of low power incident laser beam ($\sim \text{kW}$), acting relatively long time - of hundred microseconds. The illuminated layer area (with $\varnothing \sim 0.1 \text{ cm}$ - diameter) for such power is mainly fused and vaporized and is also accompanied with poorly developed photo-ablation (following the microscope observations). The particles are dispersed in all directions – in near spherical diagram of motion. The essentially most interesting are the cases of illumination with laser beam ($\varnothing \sim 0.1 \text{ cm}$) having moderately high energy and power of free lasing: $E \sim 0.5 - 1.5 \text{ J}$, $\tau \sim 100-400 \mu\text{s}$ (Fig.3-b,c). In the last case the cloud of the thrown out particles is transformed into a sharp torch around the axis of the incident beam.

The results of the detailed experimental investigations are illustrated in the Fig. 4. We have used combination of layered glass plate (GPPL in the Figure) and clean parallel indicated (marking) glass plate (MGP). The two plates are at distance of 0.7 cm each other. For the illumination with laser beams with parameters as these ones, described for Fig.3 (b) and (c), we obtained experimental results,

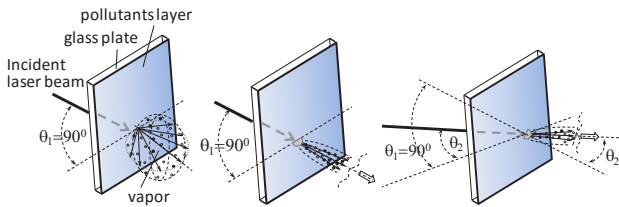


Fig. 3. The qualitative schematic of the trowing out particles motion, following the preliminary experiments in Fig.2 .For illumination with free generation beam heaving: (a) low power ($\sim 1-2 \text{ kW/cm}^2$) and energy ; (b) for moderate power ($\sim 20 \text{ kW/cm}^2$) and falling at 90° and (c) with moderate power ($\sim 20 \text{ kW/cm}^2$) and energy at angle $\theta_2 \neq 90^\circ$.

corresponding qualitatively to the estimated one (the discussion to Fig.3). Here and below we note the formed vapors and thermo-ablation micro-particles by the illuminating beam for simplicity with “vapor” or “pollutant particles”.

The most important experimental results is the observation that the motion of the vapor is really in the sharp torch close to the direction of the laser beam. This is illustrated in the photograph in Fig. 4(b)-(c) which shows the cleaned spots on the layer and the formed spots on the marking plate. The acting beam is from Nd:YAG, $1.06 \mu\text{m}$ wavelength, free lasing laser, traced by pilot red He-Ne laser beam. The energetic parameters of the illumination is $\sim 0.5\text{J}/350 \mu\text{s}$, $\sim 30 \text{ kW/cm}^2$, $\sim 15 \text{ J/cm}^2$, spot diameter on the layer $\sim 0.1 \text{ mm}$. The photograph in Fig.4(b) illustrates the situation when the glass-plates composition is tilted at angle of $\sim 20^\circ$ with respect to the incident laser beams. The red color of the spot on the indicated glass plate is formed by the dispersed beam red light of the pilot He-Ne laser. Note that both beams propagate in completely superimposed trajectories. It can be seen that on the indicated glass plate the spot, formed by the vapors, coincides with the incident beam place (make evident by the diffusion of He-Ne beam).

To be more clear the last situation, the ensemble of both plates is turn to be perpendicular to the incident beams as it is shown in Fig.4(c) .The Nd:YAG laser beam is stopped and illumination is only by the pilot He-Ne laser beam. The deviation of the spot formed by the vapors from the direction of the He-Ne laser beam and the measurement show that really the vapors follow the Nd:YAG beam direction. We have investigated the obtained spots of vapors using appropriate microscope. The microscope photograph of part of the spot is given in Fig.5. The evaluation of the averaged dimension of the particles in the spots shows manly the presence of particles with diameters of $\sim 0.5 \mu\text{m}$, also the particles with smaller than these diameters (difficult to be measure with our technique) and

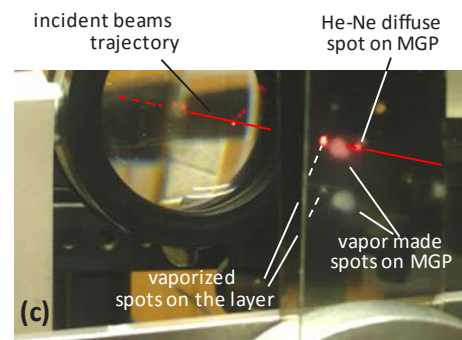
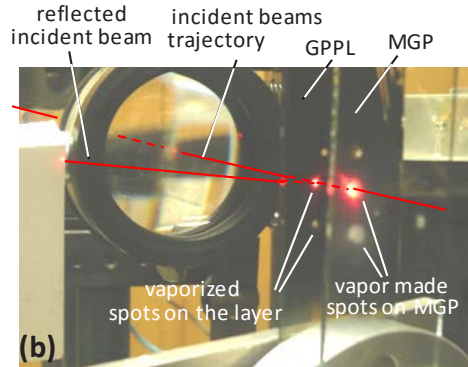
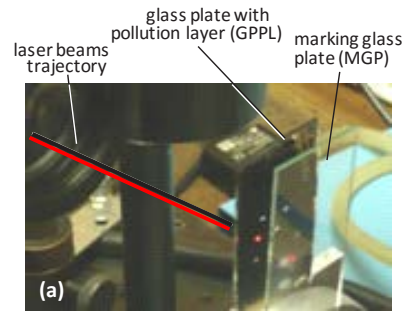


Fig.4. Schematic of the determination of the particles motion. (a) actual photograph of the experimental arrangement; (b) case of beams at incident angle $\neq 0$ on the plates (He-Ne and Nd:YAG); (c) case of perpendicularly tilted to the beams plates system, Nd:YAG stopped (see the text).

negligible percentage of particles of $\sim 1-2 \mu\text{m}$. This gives us the idea that the pollutant particles are obtained by vaporization in combination with thermo-ablation (noted existence of small number of particles with higher diameter). However, following our investigation, practically we can accept that the average particle diameter is $0.5 \mu\text{m}$.

Discussion: The realistic physical idea for the motion of the “vapors” can be related with the light pressure of the Nd:YAG acting beam. Using the estimated value of particle diameters, we will consider the particles as spheres with diameter of $0.5 \mu\text{m}$ and we will analyze the effect of the light pressure. Let’s estimate the momentum, created on each particle by the light photons momentums that act on this particle. For the laser spot on the layer we have measured the diameter of $\sim 0.1 \text{ cm}$

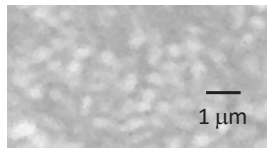


Fig.5. Photograph of the produced particle, thrown out by the beam illumination of the layer.

(partially focalized illuminating beam). The laser wavelength is $1.06 \mu\text{m}$, the case that we consider is for the energy in the incident 1 J and the pulse duration $350 \mu\text{s}$. This corresponds to energy density in the illuminating area $1.3 \times 10^6 \text{ J/m}^2$. Thus, the number of the incident photons of m^2 , each with energy of $3 \times 10^{-19} \text{ J}$ is $0.4 \times 10^{25} \text{ photons/m}^2$. The volume of the particle-sphere is $0.65 \times 10^{-19} \text{ m}^3$. We take into account that our experimental observation shows both the produced particles float in the air and also the specific mass of the paraffin is $\sim (0.8-0.9) \times 10^3 \text{ kg/m}^3$. Taken into account this observation we can accept reasonably that the specific mass of the pollution particle space is $\sim 2-5 \cdot 10^3 \text{ kg/m}^3$. As below we will see, the relatively large variation of this value ($1-7 \text{ kg/m}^3$) is not critical for the proposal. Thus, taken the mass density to be $5 \cdot 10^3 \text{ kg/m}^3$ the mass of the particle is $3.25 \times 10^{-16} \text{ kg}$. The cross-section of the particle for the illuminated photons can be accepted to be equal of the main circle of the sphere and respectively with area of $0.2 \times 10^{-12} \text{ m}^2$. Thus, any particle accepts, during the illumination, portion of 8×10^{11} photons. The momentum of each photon is $p = h/\lambda$ equal to $6.6 \times 10^{28} \text{ kg.m/s}$ and, accepting correctly that the all illuminating the particle photons move in the same direction and also are absorbed by this particle, we obtain that the sum momentum received by the particle by the laser illumination is $\sim 0.53 \times 10^{-15} \text{ kg.m/s}$. The average speed, can be calculated to be $\sim 2 \text{ m/s}$, using $m.v = 0.53 \times 10^{-15} \text{ kg.m/s}$ and with the calculated above mass of the particle. If the energy in the area of the layer illumination is 3 J , the speed will be 6 m/s and in acceptance for the specific mass of $\sim 2 \cdot 10^3 \text{ kg/m}^3$, the speed must be $\sim 16 \text{ m/s}$. Also, what is very important, the obtained pulse momentum is directed toward to the direction of the beam propagation, and the particles move in direction of the beam. From our discussion we can conclude that the light pressure is the main factor of their motion. Possible influences can have also the pressure of more intensively formed drops in direction of the incident powerful beam (thermo-ablation, vaporization). The particles motion is impeded with the increasing of the gas pressure; however this is favorable condition for our aim. In any case, the described above experimental result is really obtainable. It can be considered as an

important observation and base of useful technical solution.

3. Principles of the schemes of light cleaning and preserving the input window.

Two schemes, proposed on the base of the described experimental results and the discussion in the previous section are shown in Fig.6 and Fig.7.

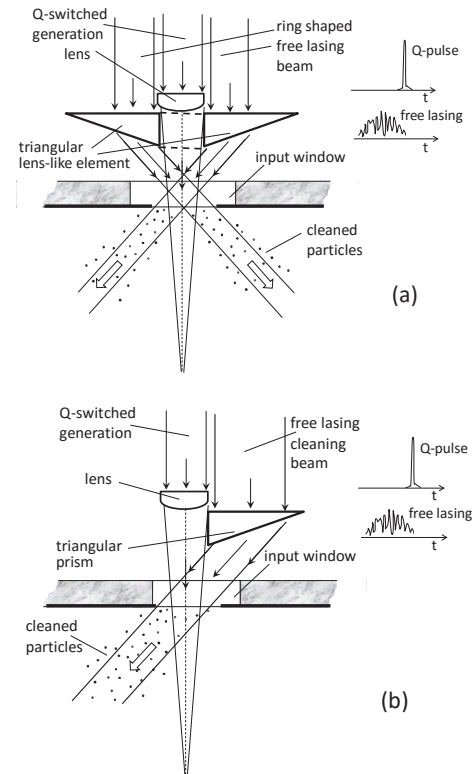


Fig.6. Schematic of light cleaning and preservation arrangement with (a) coaxial beams and (b) with parallel beams.

In both schemes a single, two channels laser beam with independently controlled light in each channel (temporally, energetically) is used. The light in the external channel is with suitable parameters (free lasing) for cleaning the window and in the other – internal – for fuel combustion ignition (Q-switching); the two emissions are with suitable temporal disposition. Fig. 6(a) clarifies the principle of the laser ignition engine with two coaxially disposed beams - the external-cleaning (from the external free lasing) and the other - Q switched by the internal axial channel. The Fig.6(b) gives the case with closely parallel beams. The optical focalizing system on the $\sim 3-4 \text{ mm}$ window in Fig.6 consist of prismatic lens-like element that directs the external beam on the window and pushes the evaporated pollution particles outside the projection of the window. The flat-concave lens focalizes the igniting beam inside the fuel. Thus, the vapor particles are pushed outside the ignition beam

trajectory. The estimations take into account that the normal combustion ignition forms the wave of the combustion that moves with $\sim 20\text{-}40$ m/s. Applying theoretical evaluation as the given one above in this point, it can be predicted the reality of the possibility for the two momentum to can quench each other or its sum to be momentum, directed outside the window projection (see the Fig.7).

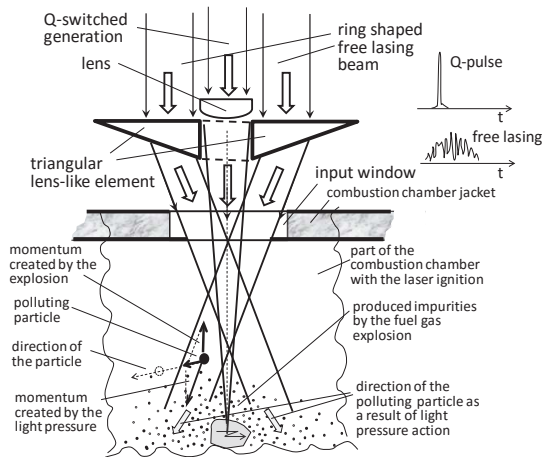


Fig.7. Principle with light-pressure window preservation

Such lasers with two-channel geometry for coaxial [3] and parallel [4] output beams (from single rod) are realized by us on the example of Nd:YAG rod. A solution, using the specific properties of the Gaussian beam in combination with coaxial laser geometry, is proposed for diode end-pumped Nd:YAG laser [5]. Each channel generates and is controlled independently. One example, of the laser spots with coaxial beams, is shown in Fig.8(a). In Fig.8(b) are shown the oscilloscope traces ($50\mu\text{s}/\text{div}$) of two convenient temporally emissions for realization of the schemes.

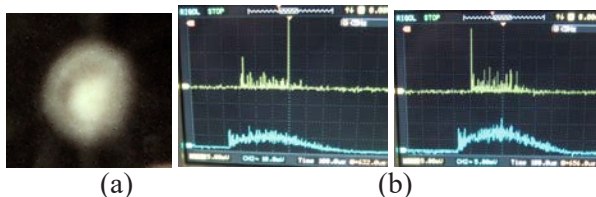


Fig.8. (a) laser spots in coaxial beams; (b) oscilloscope traces of the convenient emissions.

For both photographs, the bottom oscilloscope traces are for the free lasing and the top one –for the Q-switched ignition pulse, the last accompanied with free lasing that help for cleaning or suppression of the formed particle. It is easy to change the type of emission of each of the channels. The preliminary (laboratory modeling) tests, using these lasers, prove the expected possibility to surmount the discussed problem.

4. Conclusion

In the work we have presented the experimental results, accompanied with theoretical evaluation, related with the controlled cleaning by the laser light of thin layers of pollutants on the transparent supports. The work is especially to look to the laser ignition in internal combust engines. From our study we have made the essential conclusion about the possibility to use the laser light pressure in needed manner for cleaning – with deviation of the formed pollutant particles that the last do not disturb the entry of the ignition laser beam. On the base of the results we have proposed with the argumentation two types of schemes for suitable laser-optical arrangement for laser ignition of the combustion in gaseous matter, including the laser ignition in engines.

REFERENCES

1. Wintner, E., Deneva, M., Nenchev, M. (2014). Laser Ignition of Engines – A Contribution to Environmental Protection and a Challenge to Laser Technology”, *Ann. J. Electron.*, 1-8 ISSN 1314-0078ET and the literature cited therein.
2. Dale, J.D., Smy, P.R., Clements, R.M. (1978). *Laser ignited internal combustion engine – an experimental study*; SAE Congress, paper 780329, Detroit
3. Deneva M., Nenchev M., Wintner E., Topcu S. (2015). Coaxial-geometry tunable dual-wavelength flashlamp-pumped Nd:YAG laser *Opt. Quant. Electron.*, DOI 10.1007/s11082-015-0205-5
4. Kissov H., Deneva, M., Nenchev, M. (2013). Development of a tunable, competition less flash-lamp pumped Nd:YAG laser generated of a chosen pair of two lines, *Proc. SPIE*, 8770-50, 1-7
5. Deneva M., On advantageously use of Gaussian beam end diode laser pumping of Nd:YAG laser in two-wavelength coaxial architecture, *Techsys'2016*, submitted for publication.

M. Deneva, M. Nenchev, N. Kaymakanova are with “QOE” Laboratory, R&D Dept., Technical University of Sofia and OELE, Plovdiv, BG;

e-mail: mdeneva@yahoo.com

E.Wintner is with Vienna University of Technology, Photonics Institute, Gusshausstr. 25-29, 1040 Vienna, Austria;

e-mail: ernst.wintner@tuwien.ac.at, Austria;

M.Himber is from CNAM, Université CNAM, Institut de métrologie, Paris, France.



DEVELOPMENT OF THE THERMO-SENSITIVE PAPER APPLICATION AS METHOD FOR LASER BEAM SPOT STUDY

MARGARITA DENEVA, MARIN NENCHEV, VALKO KAZAKOV, NADEJDA KAYMAKANOVA

Abstract: *We present a detailed investigation and development of thermo-sensitive paper application as a method, which in strongly defined conditions can be successively used for laser beam spot study. This old technique is applied in the literature and in the laboratory practice only for visual illustration of the laser beam spot by the whitening of the incident area of black thermo-sensitive paper, with accidentally taken sensitive materials. However, looking in details, in competition with the modern electronic beam scope techniques, it presents some essential advantages. As it can be expected and we have marked in our experiments – it cannot be spectrally sensitive in very large range, as example - from UV to IR; also the registration cannot be disturbed by different noise sources; as addition, it is extremely cheap and accessible. In our knowledge in the literature has not systematic study of the possibility of this simplest method. This work presents results of the detailed and systematic study of the noted technique-method. We have shown that using appropriate, well defined and reproducible as production sensible materials in combination with their precise known reaction to the laser light with specific parameters and using standard computer treatment, the noted advantages of this technique can be successfully employed for laser spot parameters investigation.*

Key words: *laser spot study, thermo-sensitive paper*

I. Introduction. General aim of the work.

The aim of the work is detailed investigation and development of the practical application of Thermo-Sensitive Paper Registration Technique (TeSPeRT) for laser spot imaging. This method, as we show, in strongly defined conditions, can be successfully used for laser beam energetic parameter study. The base of the technique, as a rule, is the whitening in the laser beam incident area of the black thermo-sensitive paper. This old technique [1,2] is used in the literature and in the laboratory practice only for visual illustration (including also successfully - for the interesting cases) of the laser spot. The reported applications are used accidentally taken sensitive materials and without conclusion about energetic characteristics of the spot. However, looking in details, this technique presents some essential advantages in competition with the modern electronic beam-scope apparatus. First of all, as it can be expected and we have marked in our experiments with two-wavelength lasers [3], few properties can be marked: a correct registration cannot be spectrally sensitive in very large range, as example - from UV to IR (e.g. 0.3-3 μm and longer); secondly, the

registration cannot be disturbed by the electromagnetic noise and third - as addition it is extremely cheap. In our knowledge, (according to our research) no systematic and detailed study of this technique is given in the literature. It was neglected especially due to the facilitated accessibility of the modern electronic beam-scope apparatus. Thus we formulate the aim of the work to study in detail the technique of beam profile registration using a black thermo-sensitive paper.

II. Experimental investigations and results.

Following our preliminary observations we have planned to: 1) study the reaction between different strongly defined blacked-out paper sheets and the laser beam illumination and to choose more convenient and widely obtainable combination; 2) study the reaction of the registration on the illuminating beam energetic and temporal parameters to find the range of the linear widening; 3) study the dependence from the beam wavelength.

First point of our investigation is related to testing of series of blacked-out materials on the illuminating beam reaction. We have found that very suitable material for the marking is the tracing

paper blacked one (1Xe) or two (2Xe) times through the copy machine. The copy machine used was standard xerography type action Konica Minolta Dialfa Di 5510. The blacking-out was by copying of the black paper at normal blacked position for the machine option. We do not observe noticeable difference comparing few tracing papers of different producers and also the blacking by the different copiers of the noted before types (Sharp-MX-3500, Toshiba 2500c) for normal blacking operation. The transmission of the pure tracing paper is measured to be $\approx 50\%$ when the paper is placed closely to the entrance of the light power meter (Thorlabs SN6050506); in this manner the measured with flat and large entrance of 2 cm diameter accept all the diffused light. He-Ne yellow ($0.595 \mu\text{m}$) laser beam with incident diameter on the paper of $\sim 2 \text{ mm}$ and power of 3.6 mW is used. For the 1Xe and 2Xe blacked paper the transmissions are $\approx 9\%$ and $\approx 6\%$ respectively. As we show below, the use of 2Xe paper is an optimal case. The illumination by the laser beam was from the not blacked layered side of the sheet. When it was from the layered side, the spot image frequently was not pure, having ‘moustaches’, probably produced by the reinjection of the part of the vaporizing (or/and thermo-ablating) micro-particle by the incident beam light pressure.

We combine this registration with tracing of the whitening spots from the paper by computer scanner and computer treatment of the image. The treatment includes two type of manipulation with adapted standard programs. The first treatment was the tracing of the whitening along the beam spot diameter in 2D image. The other - second treatment was to create 3D images of the whitening in any point of the spot.

In the study we have used our two homemade flash-lamp pumped lasers – Nd:YAG and Nd:Glass. The first laser (I) produces a controllable energetically emission at $1.06 \mu\text{m}$ or at $1.36 \mu\text{m}$. The output energy can be varied from 0.1 J to 1.5 J in pulse duration approximately $350 \mu\text{s}$. The spot in cross section of the emitted beam is with good approximation in circle with diameter at the laser output of $\sim 0.7 \text{ cm}$. The second laser (II) produces a beam with cross section spot of 1 cm at the laser output variable between 8 and 10 J with pulse duration of 3 ms and the laser oscillate at the wavelength $1.06 \mu\text{m}$. Simultaneously with the spot registration we have measured the laser output energy (FIELDMAX energy meter, Coherent, USA) is used, furthermore a two-channel storage 200 MHz oscilloscope RIGOL DS1202CA (USA) and light detectors with resolution better than 5 ns (R108BG). Both lasers are multi-mode and operate in plane-plane mirrors resonators.

The typical case of spot registration and computer treatment are presented in Fig.1. The spots are formed by the beam from Nd:YAG laser (I, $1.06 \mu\text{m}$) at the 2Xe paper. The illuminating energies that are varied by the pumping are given in the figure. The averaged energy densities for the spots are in the limits from 0.4 to 0.65 J/cm^2 . The ratio (measured energy)/(calculated volume in arbitrary units) for the spots are:1-(0.71J)/(1x); 2-(0.43J)/(0.61x);3-(0.45)/(0.62x); 4-(0.63J)/(0.89); 5-(0.72J)/(1.03);6-(0.49)/(0.73). The volumes are calculated using corresponding for this operation computer program. The analytic calculation of the volume, using the tracing curve in Fig.1(c) and

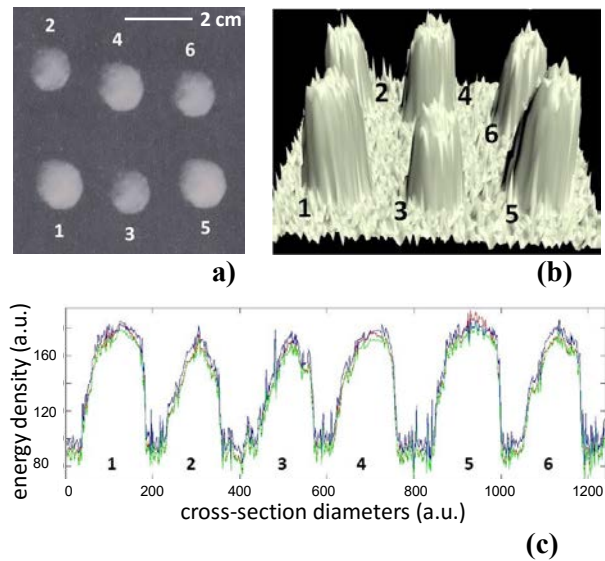


Fig.1. Photographs of the beam spots on the paper (2Xe) (a); the 3D computer imaging (b), and 2D tracing curves along the spot diameter (c).

accepting the symmetry in rotation, gives very close results as these ones for Fig.1(b). The ratios energy/volumes for any spots pairs can be marked to be equal. This signifies linearity of the whitening with the laser energy density in the considered energy range $0.4\text{--}0.8 \text{ J}$ (energy density ~ 0.4 to 0.65 J/cm^2).

From the experimental results as those shown in Fig.1 and from general reasons, we conclude that it is correct to relate the distribution of the whitening in the spots with corresponding energy density distribution. Let's take the schematic for the analysis as this one given in Fig.2. Note the energy density distribution with $w_E(x,y)$ where X,Y are the axis in the plane of the spot and $w_E(x,y)$ is in the Z axis. Let's accept that the whitening $w_h(x,y)$ in relative units is proportional to the $w_E(x,y)$, i.e. $w_h(x,y) = k_1 \cdot w_E(x,y)$, where k_1 is the coefficient of the proportionality. The volume in a.u., limited by the function $w_h(x,y)$, is:

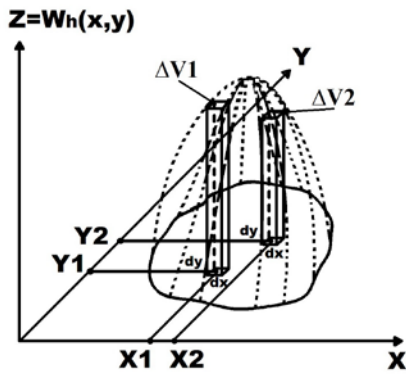


Fig.2. To the calculation of the relation beam illumination energy - volume of the spatial figure, limited by the whitening

$$V = \sum \Delta V_i \text{ and } V = \iint_{x,y} k_1 \cdot w_E(x,y) dx dy = k_1 \cdot E_{jb}, \quad (1)$$

as the volume under surface $w_h(x,y)$ is accepted to be proportional of the energy E_{ib} in the illuminating beam. This is the case in Fig.1, for the given limits the linearity presents.

Following the above explanation, one of the questions concerning the use of the TeSPeRT for laser beam energetic parameters determination is to investigate the linearity of the whitening of the paper as a function of the laser energy density in the spot. For this purpose, we investigate the formed laser spot on the 2Xe paper using the Nd:YAG laser II - with output 8-10 J. The multimode operation gives a relatively uniform illumination and we can use an averaged energy distribution in the spot.

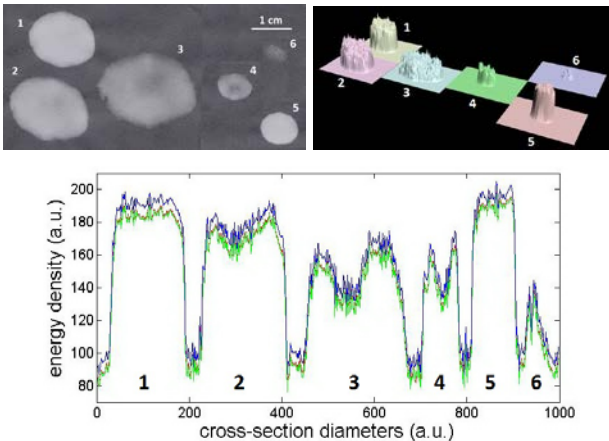


Fig.3. To the determination of the averaged whitened function $w_h(x,y)$ as averaged energy density in the spot. The spot photographs (a), 3D image (b) and 2D tracing curves along the spot diameters (c).

The illuminating average energy density is for the spot noted in (1) - 4.3 J/cm²; (2) - 3.3J/cm²; (3) - 2.2 J/cm²; (4) - 1.5 J/cm²; (5) - 6.2J/cm². For the spots 1:2 – the measured energy ratio 8J/ 8J = 1;

computer treatment obtained volume ratio is 0.94. Respectively, for the spots 1:3 - 8J/9.2J = 0.87; for 2:3 = 8J/ 9.2J = 0.87. In other words, the difference between the measured energy ratio and this one obtained by the spot treatment is ~ 10% (probably due to the precision of our home made thermo-electric bridge meter, convenient for the energy measurement of long pulses ~ few ms).

As it is seen from Fig.2(b) and Fig.2(c) – the saturation effect starts for the spot (1) and is clearly evident for the spot (5). Thus the energy density of ~ 4 J/cm² is upper limit of the correct registration for this paper and pulse length (~350μs). As a low limit for the reliable registration we accept ~ 0.2 J; (~0.5 J/cm²), that is illustrated in Fig.3 (spot 5). A typical transmission of the saturation whitened spot area is ~18 %. The testing experience has shown us that the transmission of the whitened spot (i.e. the bleaching) cannot be applied for energy density evaluation; there is no proportionality as this one for the whitening that was observed. We relate the explanation with the shown below whitening by the appearance of white products by the burning of the black layer by laser irradiation. Note that, for the 1Xe and 2Xe paper the limits are not essentially different, being a little smaller for 1X. The saturation effect for 3Xe paper is high, however the registrations of the low energy density spots (0.3-0.4 J) are not reliable; also such layering is practically very difficult. Thus, our observation shows that for moderate energy density 2Xe paper is well suitable. If the pulse is with high energy density, the latter can be decreased by filter. Other question is the dependence of the pulse power density. For this purpose we have used the passive Q-switched operation of the laser (I) with output (and illuminating) energy of 0.5 J and pulse length of ~ 1 μs. The energy density was ~ 2 J/cm² and power density ~4 MW/cm². We compare the spot parameters on the paper 2Xe with this one for the free lasing of 2 J/cm² and pulse of ~350 μs.

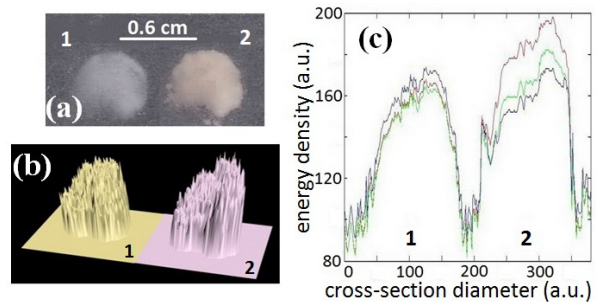


Fig.4. Spot photographs (a) and theirs computer 3D (b) and 2D (c) treating for the Illumination by the free lasing (1) and Q-switched pulse (2).

The difference of the reaction is in the little coloring of the Q-switched marked spot in red, probably related with additional products of paper burning (see below) and the increased transmission of the paper area at the spot (~ 2 times, in general, also low $\approx 25\%$). However, the ratio (energy)/(volume of bleaching) is not essentially different that this one, obtained for the free generation (Fig.3). Thus, the TeSPeRT, as we have sowed, is applicable also for the case of relatively long Q-switched pulses ($\sim 1-10\ \mu\text{s}$; $\sim 1\ \text{MW}$). For shorter and high power pulses the additional investigations are needed. The described above observation shows that the main factor for the whitening is the energy density (for the moderate pulses duration of our investigation ($\sim 1\ \mu\text{s} - 3\ \text{ms}$)).

Other important investigation is the dependence of the wavelength of the illuminated beam. Using our laser (I) we have produced free-lasing generation at the wavelength $\lambda = 1.06\ \mu\text{m}$ and at $\lambda = 1.36\ \mu\text{m}$ and have studied the paper recording reaction. Fig.5 illustrates the spots and their diametric tracing, for the illumination with 1J energy beam at $\lambda = 1.06\ \mu\text{m}$ (left spot and trace) and for 0.56 J at $\lambda = 1.36\ \mu\text{m}$. The study shows good coincidence of the ratios energy/energy and volume/volume for the two wavelengths, that shows the expected wavelength independence (investigation for the range $0.53\ \mu\text{m} - 1.36\ \mu\text{m}$).

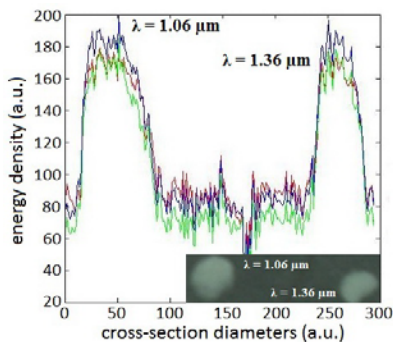


Fig.5. Spot photographs (inside) and their computer 2D treating (curves) for the illumination by the $\lambda = 1.06\ \mu\text{m}$ wavelength laser pulse with energy 1.1J and for $\lambda = 1.36\ \mu\text{m}$ with pulse energy 0.56 J.

The investigation shows, as already mentioned that the whitening results from changing the structure of the illuminated area of the black-out xerox layer. The small black composite of the layer (dimension of $\sim 10-20\ \mu\text{m}$) is transformed to be white in colour. A little decrease of its dimension is observed. Also some part of black material of the layer was knocked out under thermo-bleaching by the incident beam. The main reason of the whitening is the first mechanism. This is proved by our experiment with treatment of the whitened spot in condition of photographing (our computer

scanning) the spot on black and white background. Not essential difference was observed.

In Fig.6, as illustration of our observations, is plotted the microscope photograph of the black layer (2Xe) on the tracing paper before and after the laser beam illumination. The similar picture is observed also for 1Xe paper. The discussed bleaching effect for the composites of the layer can be observed.

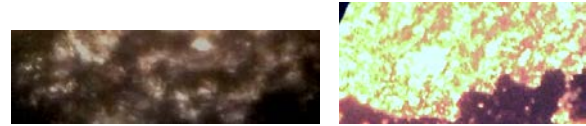


Fig.6. Microscope photograph of the black layer on the tracing paper (2Xe) before and after the laser beam illumination.

III. Conclusion

In the work we present results from detailed and systematic study of the of thermo-sensitive paper registration (TeSPeRT) in laser technique. We have shown that using appropriate, well defined and reproducible as production sensible materials in combination with their precise known reaction to the laser light with specific parameters and using standard computer treatment, this technique can be successfully employed for laser spot parameters investigation. Its essential advantages that can be marked by our study are: determination of the energy distribution in the laser spot; comparison of the energies of two pulses; having an etalon spots in order to evaluate correctly the laser pulse energy; no spectral and noise sensitivity. In addition, this technique is extremely cheap and accessible.

References:

1. Ifflander, R. (2001). *Solid-State Lasers for Materials Processing*. Springer-Verlag Berlin. Heidelberg, New York, series in Optical Sciences. ISSN0342-4111, ISBN 3-540-66980-9 (2001).
2. Carlos B. Roundy, Spiricon, Inc., <http://aries.ucsd.edu/LMI/TUTORIALS/profile-tutorial.pdf>, Beam Profile Measurements and the literature therein.
3. Deneva M., Nenchev M., Wintner E., Topcu S. (2015). Coaxial-geometry tunable dual-wavelength flashlamp-pumped Nd:YAG laser. *Opt. Quant. Electronics*, DOI 10.1007/s11082-015-0205-5 online publ.

M. Deneva, M. Nenchev, V. Kazakov, N. Kaymakanova are with "QOE" Laboratory, R&D Dept., Technical University of Sofia and OELE, Plovdiv, BG; e-mail: mdeneva@yahoo.com

PERFORMANCE SENSING SOLUTIONS FOR A GREEN ENVIRONMENT DESIGNED IN BULGARIA

PETAR MITSEV, ANGEL TANEV, TSVETELINA LAZAROVA

Abstract: *This paper presents novel green automotive platinum sensing technology. In recent years, worldwide emissions legislation has been introduced and is rapidly becoming more stringent. With alternative vehicular propulsion methods far from becoming mainstream reality, leading automotive providers have intensified efforts in the direction of reducing the harmful footprint of their products. This is being accomplished via smaller, more optimally designed internal combustion engines, necessitating an increased and higher-performance sensor content per vehicle. This paper elaborates on temperature sensor application in automotive exhaust gas performance sensing.*

Key words: *automotive exhaust gas, platinum temperature sensor, materials technology*

1. Introduction

The growing environmental awareness within society has necessitated an immediate and effective solution towards air pollution reduction and control [6]. A large portion of the latter is produced every day by the global vehicle fleet, responding to the ever-growing transportation needs of an expanding world population, with automotive being among the three major energy consumer sectors. The above requires a greater degree of sensing, operating under harsher application conditions, prompting the need for novel sensor technology.



Fig. 1. *Automotive sensors: general view [11]*

2. Theoretical Background

In practice, there are 3 sensor technologies that are viable in vehicle exhaust: thermocouples, thermistors, and resistance temperature detectors (RTD). Each of them have their own advantages and disadvantages, and will be considered in brief detail.

Thermocouples are formed by welding two dissimilar metals together forming a bimetallic junction that produces a temperature-dependent voltage as a result of the thermoelectric effect [17].

For a vehicle application, a type K (chromel-alumel), type R or S (platinum-rhodium) or type N (Nicrosil-Nisil) would be used for the range of temperatures previously mentioned. Thermocouples can be relatively low cost sensors (for example, using type K) as compared with thermistors or RTDs. However, the electrical system is significantly more expensive since there must be introduced a compensation for voltages produced whenever there is a change in wire material (often called cold-junction compensation).

Thermocouples can be made with very little mass which allows for a fast response with changing temperature. In order for the sensor to minimize drift and be reliable in a vehicle exhaust environment however, the thermocouple must be protected by a sheath, and therefore they are made thicker. Thus much of the fast response advantage is reduced.

Thermistors are made from various nonmetallic conductors (i.e. metal oxides). The types of thermistors found in a vehicle exhaust environment will typically produce a negative temperature coefficient (NTC), meaning the resistance will decrease with increasing temperature. Thermistors offer a high sensitivity over a smaller range in

temperature than either thermocouples or RTDs [17].

At 0°C the resistance can be over 100000 Ω , at 200°C: 200 to 500 Ω , and at 800°C 50 Ω . Therefore, thermistors can achieve very high sensitivities over a particular range of temperatures. However, achieving nearly the same accuracy over a large range in temperatures is not possible (unless several pull up resistors are used) due to the highly nonlinear characteristic response. Thermistors can be made with a very small thickness for quick response. However, they are not able to withstand even mild vehicle exhaust environments without being protected by a metal or ceramic insulated sheath thus causing the sensor response to be relatively slow.

Tolerance of a thermistor depends on its intended range of use. For example, thermistor tolerances in manifold air temperature sensors (MAT) or coolant sensors are very tight over the relatively narrow range of measurement (ex. 0.6°C from 0° to 100°C). However in a vehicle exhaust that can vary between -40°C and 1000°C, thermistors have a fairly poor tolerance depending on the temperature range (2% to 6% of temperature). As previously described, thermistors typically have a very high resistance below $T=100^\circ\text{C}$. This makes it difficult to meet requirements of being able to read the sensor at temperature -40°C, or being able to perform OBD (On-Board Diagnostics) start up diagnostics at 20°C.

Table 1 shows a summarized comparison of the three different sensor types. The RTD is the most accurate of the three sensor types over the entire measurement range over time. It is important to note that zero hour accuracy is only one portion of total accuracy; how the sensor performs in the application over time at temperature is vital [17]. Figure 2 [13,14,15,16] shows two thermistors and a typical RTD designed for vehicle exhaust use (thermocouples are not shown since they generate an output voltage instead of measuring a resistance). As can be seen from the graph, the thermistors show a very non-linear curve as compared to the RTD. Further, there is a significant difference between the two different thermistor manufacturers, with one not being practical for measuring temperatures below 100°C. The tolerances indicated on the graph show that the RTD has a much tighter tolerance than thermistors over the measurement range.

Table 1. Comparison of three sensor technologies

Characteristic	Thermocouple	Thermistor	RTD
Linearity	Poor	Poor	Good
Accuracy	Fair	Fair to poor	Excellent
Sensitivity			
<600°C	Poor	Excellent	Good
>600°C	Poor	Fair	Good
Signal Level			
<600°C	Poor	Excellent	Good
>600°C	Poor	Fair	Good
Response time	Poor (3 mm dia)	Poor	Good (No sheath)
System Complexity	Poor	Good	Excellent
Standardized output signal	Excellent	Poor	Good
-40°C measurement	Excellent	Poor	Excellent
20°C OBD II measurement	Excellent	Poor	Excellent

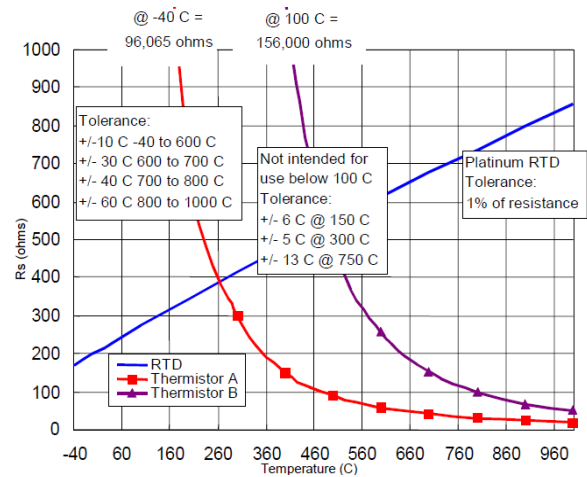


Fig. 2. Comparison of three sensors technologies.

Resistance Temperature Detectors (RTDs) operate through the principle of electrical resistance change in metal thermistors. Platinum is the most widely specified RTD element type although nickel, copper, and Balco (nickel-iron) alloys are also used. Platinum is popular due to its wide temperature range, accuracy, stability, as well as the degree of standardization among manufacturers, thus interchangeability. RTDs are characterized by an almost linear positive change in resistance with respect to temperature. They exhibit the most linear signal with respect to temperature of any electronic sensing device [1].

Pt. Resistivity of alloys. 3.

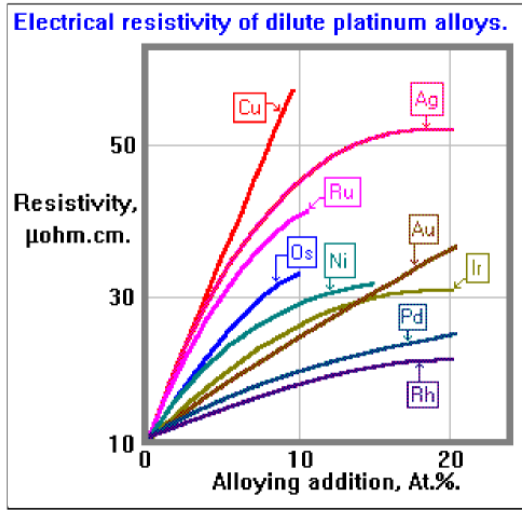


Fig. 3. Resistivity of various Pt alloys [6].

The platinum and copper RTDs are most commonly used [1]. In general, RTDs are employed over the wide temperature range from -260°C up to 850°C , and some of these RTDs are applied up to 1200°C [1].

Pt - Rh. Resistivity & TCR. At.%. Pt - Rh alloys, resistivity and T.C.R.

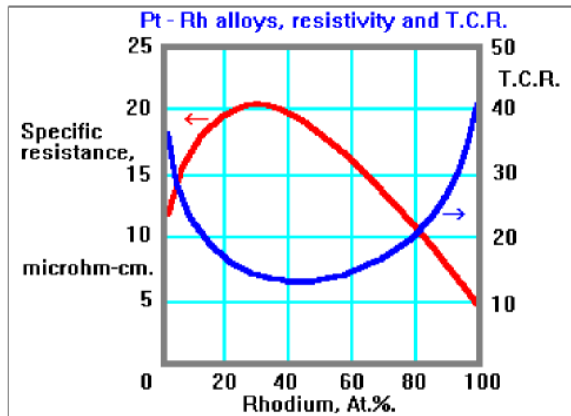


Fig. 4. Effect of Rh addition on Pt TCR [12].

The effect of adding rhodium over the platinum's temperature coefficient of resistance (TCR) is given on Fig.4. The quantity of rhodium (Rh) in the Pt-Rh alloy is less than 1%. This is commonly used to control the TCR value.

The resistance vs. temperature equation is shown below, which is represented by the best-fit quadratic equation of the measured data [2]:

$$R(T) = R_0(1 + \alpha T + \beta T^2) \quad (1)$$

where:

$R(T)$ - sensor resistance $[\Omega]$ vs Temp $(^{\circ}\text{C})$,
 α is a slope of R/R_0 vs T line,

β is a non-linear (quadratic) term.

For a standard Pt RTDs, the coefficients α (slope) and β (quadratic term) are summarized in Table 2:

Table 2. Pt RTD international standards.

Standard	Temperature coefficient of resistance, (TCR)	α	β
	$[^{\circ}\text{C}]^{-1}$	$[^{\circ}\text{C}]^{-1}$	$[^{\circ}\text{C}]^{-2}$
DIN 43760	0.003850	$3.9080 \cdot 10^{-3}$	$-5.8019 \cdot 10^{-7}$
American	0.003911	$3.9692 \cdot 10^{-3}$	$-5.8495 \cdot 10^{-7}$
ITS-90	0.003926	$3.9848 \cdot 10^{-3}$	$-5.870 \cdot 10^{-7}$
IEC 60751:2008	0.003851	$3.9083 \cdot 10^{-3}$	$-5.775 \cdot 10^{-7}$

The standard IEC 751 sets two tolerance classes for the interchangeability of platinum Pt RTDs [1]:

Class A: Tolerance $(^{\circ}\text{C}) = \pm(0.15 + 0.002|t|)$

Class B: Tolerance $(^{\circ}\text{C}) = \pm(0.3 + 0.005|t|)$

where: $|t|$ - absolute temperature in $^{\circ}\text{C}$.

Class A applies to temperatures from -200°C to 650°C , and only for RTDs with three or four-wire configurations. Class B covers the entire range from -200°C to 850°C .

The Platinum resistance as a function of measured temperature is shown on Fig.5. The function is almost linear over a wide range of temperatures (up to 850°C). The resistance at 0°C is $R_0 = 200\Omega$.

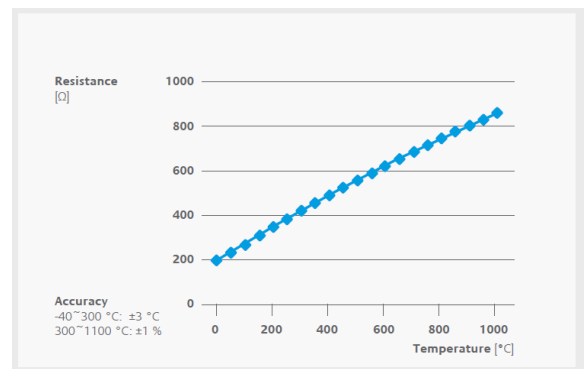


Fig. 5. $R=f(T)$ function [7]

To accomplish the above, design engineering is involved along the scale axis, via different disciplines, shown below on Fig.6 [3,4,5,18]:

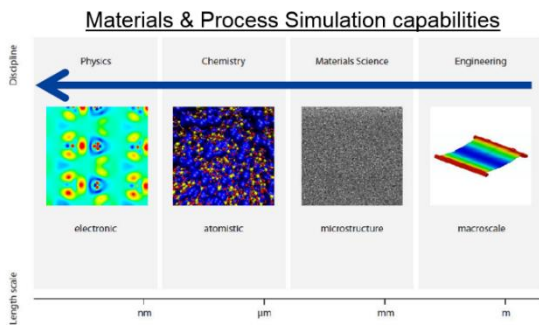


Fig. 6. Physical scale versus scientific disciplines, used in sensor design [8].

3. Application

The technology, described in this paper, is used to satisfy the world's growing need for safety, energy efficiency and a clean environment. Incorporation examples are provided below (see Fig.7):



Fig. 7. Areas of application [9]

Due to the increased automotive systems introduction together with their complexity, sensor content per vehicle is increasing (Fig.8, top):

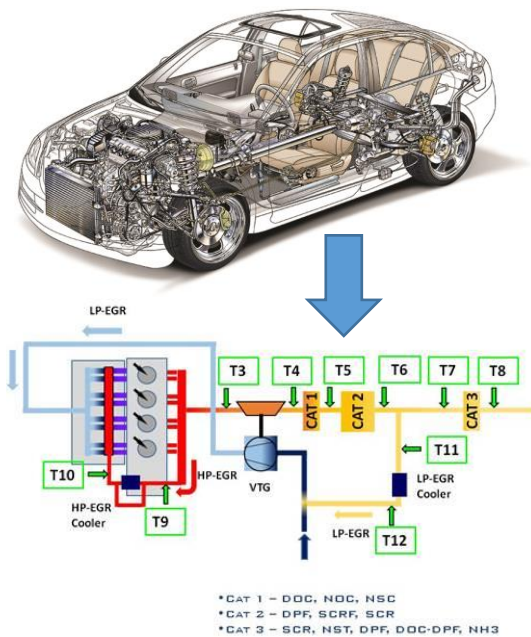


Fig. 8. Temperature sensor positions along the exhaust tract [10]

Typically, several exhaust gas temperature sensors are positioned along the thermodynamic chain (Fig.8, bottom). For example, diesel trucks- starting with the 2007 model year- used for on highway applications are required to have an exhaust treatment system to reduce emissions, specifically large particulate matter. The diesel particulate filter (DPF) is a component of the system. DPFs are designed to physically filter particulate matter (soot). A regeneration process removes the accumulated soot from the filter.

4. Conclusion

In this paper, the application of automotive exhaust gas temperature sensing, along the length of the thermodynamic tract, has been described. The influence of metal alloy composites on resistivity and the temperature coefficient of resistivity, thus sensor signal, has been elaborated. The physical scale to scientific discipline design implications has been discussed, providing for performance sensors of increased stability and reliability.

REFERENCES

- Adler B., *Reliability Aspects of Temperature Measurement*, Moore Industries-International, Inc., North Hills, CA 91343.
- <http://www.ni.com/tutorial/7115/en/>
- Fischer et al. (2001). Practical Experience with New Oxide Dispersion Hardened Platinum Materials, Tucson, Arizona, USA, June 9-12, 2001.
- Jose' S. Moya, Sonia Lopez-Esteban, Carlos Pecharroma'n (2007). The challenge of ceramic/metal microcomposites and nanocomposites, *Progress in Materials Science* 52, 1017-1090.
- Ke Fan, Min Liu, Tianyou Peng, Liang Maa, Ke Dai (2010). Effects of paste components on the properties of screen-printed porous TiO₂ film for dye-sensitized solar cells, *Renewable Energy* 35, 555-561.
- Savitski E. et al (1978). Physical Metallurgy of Platinum Metals.
- Kyocera Pt 200 RTD product datasheet
- Materials Design Medea product presentation
- Sensata Technologies Inc
- Sensata Technologies Inc
- Sensata Technologies Inc
- The PGM database, www.pgmdatabase.com
- Desmarais R., et al. (2001), "How to Select and Use the Right Temperature Sensor", *Sensors Magazine*.
- LeGare J., et al. (1995), "Temperature Sensors for On-board Diagnosis of LEV/ULEV Systems", *Automotive Engineering*.

15. Mathews D., (2000), "Choosing and Using a Temperature Sensor", Sensors Magazine.
16. Volbrecht A., (1998), "Temperature Measurement: Making Sense of it All", Sensors Magazine.
17. Nelson C., et al., (2004), The Development of a RTD Temperature Sensor for Exhaust Applications, Delphi Corporation, SAE International.
18. Petrov, N. I., (2007), Reliability's investigations of risk technical systems. Yambol, Bulgaria: Publishing House "Jelio Uchkov".

Authors' Contacts:

Address: Sensata Technologies, Sofia.

Petar Mitsev, e-mail: pmm@sensata.com

Angel Tanev, e-mail: a-tanev@sensata.com

Tsvetelina Lazarova, e-mail:

tlazarova@sensata.com



GENERATOR OF NETWORK DoS ATTACKS

VENETA ALEKSIEVA, IVAYLO ZHELIAZKOV

Abstract: *This paper presents a system for generation of network DoS attacks, using protocols of TCP / IP suite - UDP, TCP and ICMP. The purpose of the developed system is for education and it offers versatile tools for simulating different types of attacks. DoS attacks with various parameters are implemented. There are presented results of experimental research.*

Key words: *DoS attack, UDP attack, TCP attack, ICMP attack, Syn flood attack*

1. Introduction

Network Denial of Service (DoS) attacks do not aim penetration, establishing control over data or eavesdrop on communication. Their task is to cause a collapse in network devices or servers, thus compromising the whole system or only a range of services, making them unavailable to the legitimate users.

Statistics provided by Kaspersky Lab [1] present that DDoS in 2015 has decreased by 11% in comparison to 2014. But during the first quarter of 2015 they are 23,095 and most of the attacks are carried out on the territory of China, the US and Canada. The longest attack was 140 hours. Financial institutions, banks and large companies in the energy sector have registered loss of over \$ 12 million annually. It directs insurers to offer insurances against hacker attacks. In 2015 the company Loyd British Insurance [2] has registered an increase in the amount of revenue as a result of the conclusion of insurance policies with companies to cover the risks of hacking, worth 2.5 billion dollars. Globally, the cost of cyber security is 77 billion dollars in 2015 and to 2020 worldwide companies expect to invest about 170 billion dollars to protect their data.

There are many tools and technologies, whose task is to identify and block various types of attacks [3,4] - antivirus programs, firewalls, static ARP tables, physical firewall devices, expensive specialized equipment like Cisco Guard XT 5650, FortiDDoS-400B, FortiDDoS -2000B and others. To respond effectively to these suspicious activities to computer or network resources, it is needed to know the nature of the attacks.

The proposed generator simulates various network attacks with different parameters. This allows to study the effectiveness of various protection tools against these threats.

2. Related Works

Most of the realizations of network DoS attacks use high-level languages like C and C++, because they have tools to work with sockets (in particular raw sockets). With these tools the programmer constructs own TCP/IP headers of messages that will attack a system. The operating system Windows and many modern Linux distributions impose a number of restrictions on the use of raw sockets on kernel level. For example, the raw sockets are prohibited in versions of Windows after XP, and Server-oriented versions after Server 2008R2, as in many of the latest versions of popular Linux distributions such as Ubuntu, Debian, Fedora. The idea is similar at all - do not allow the operating system to become the conductor of this type of action or impose a number of restrictions against malware to fulfill its purpose. The weaknesses and shortcomings in the TCP / IP stack and the freedom that has the system programmer on the form and content of IPv4 packets is a precondition for abuse. The flexibility of programming languages is large enough to allow the development of algorithms that dynamically change the parameters of the IPv4 packets, which are the attacked victims. This obstructs prevention and localization of beginning stages of the DoS attack.

There are many solutions to generate DoS attacks for testing purposes under Windows, for example:

- LOIC (implements attacks on TCP/ HTTP/ UDP/ ICMP) [5]
- XOIC (implements attacks on TCP/ HTTP/ UDP/ ICMP) [6]
- PyLoris (implements attacks HTTP, which used SSL, SOCK Proxy) [7]

They have some disadvantages, as low efficiency, inability to replace addresses of packets (limits after Windows XP SP2 and Server 2008R2).

In most decisions is absent an option for simulation of attack from more than one source simultaneously.

The generators of network attacks under Linux offer more variety and efficiency than generators under Windows and they offer the possibility to replace addresses, for example:

- hping3 (implements attacks on TCP/ HTTP/ UDP/ ICMP, pack dumping, port scanning, transporting files regardless of the aggressive policy of firewalls) [8]
- GoldenEye (implements attacks on HTTP) [9]
- PyLoris (implements attacks on TCP/ HTTP/ UDP/ ICMP) [10]

They generate packets with static content that is easily recognizable from firewalls. Other disadvantage is that they are so tightly specialized.

3. DoS attacks implemented in the proposed generator

For realization of the proposed generator of DoS attacks is selected operation system - Kali Linux 2.0 sana, because it is optimized for performing penetration tests and many of the restrictions imposed by other popular Linux distributions and in most versions of Windows (except Windows XP / 2000 / Server 2000/2008), with respect to the RAW socket, are absent. The selected programming environment is Code Blocks 10.05, with the compiler GNU GCC.

To prevent possible abuse with this proposed noncommercial generator many restrictions are implemented as:

- Prohibition of multiple instantiation of software – starting it in parallel processes;
- Limit the maximum number of simultaneously working threads to 500. Each thread after 500 will wait in the FIFO queue to be started;
- The number of implementations of various types of attacks is reduced to four most popular attacks - flood with UDP packets, flood with TCP packets, flood with ICMP packets (there are two realizations - Ping of Death and broadcast attacks), flood with TCP SYN messages;
- The maximum total number of bytes sent to the "victim" will be limited to a maximum value of type unsigned long, regardless of the duration of the attack;
- Maximum size of a single ICMP packets is 64 bytes;
- The data buffer of UDP and TCP packets is limited to 27 bytes.

The proposed generator of DoS attacks provides the following features:

- set of popular DoS attacks to perform simulation tests;

- set of tools for customizing the parameters of the attack;
- tools for automated tests in order to re-simulation of various scenarios with minimal user intervention;
- good performance and high level of control on the performed tests;
- low system requirements;
- the most simple and intuitive set of commands;
- creation of log files for analysis.

Figure 1 presents the general algorithm for generating of each attack. The particular specificity of each attack is described below.

3.1. UDP/ TCP flood attack

These attacks use protocol 17 (UDP - User Datagram Protocol) or protocol 6 (TCP - Transmission Control Protocol). They generate packets, whose data content consists from a random combination of letters and numbers. The length of payload data may vary between 10 and 27 characters. The source IP address and the source port number in the IP header may be user defined or randomly generated.

3.2. ICMP flood attack

This attack uses protocol 1 (ICMP- Internet Control Message Protocol) with generation of echo-packets. The size of a regular ICMP Echo-Request packet is 32 bytes. It seems as ICMP Echo-Request packet of ping communication.

Most of the computer systems cannot process a packet that exceeds the maximum size accepted for IPv4 protocol - 65535 bytes due to the small size of echo packets used in the attack was fixed on 64 bytes as payload part will contain randomly generated data. The algorithm of conducting of this attack is similar to the previous one, but before check the *timeout* is added *sockopt broadcast*.

3.3. SYN flood attack

This attack uses protocol 6 (TCP - Transmission Control Protocol). It generates SYN (synchronization) packets to user-defined destination port of attacked server or to its random port. The attack uses the weaknesses in the three way handshake communication mechanism. When the victim received SYN message, it responds with a SYN/ACK (synchronization acknowledged) message for each open port and it waits for ACK (acknowledged) packets from bogus clients.

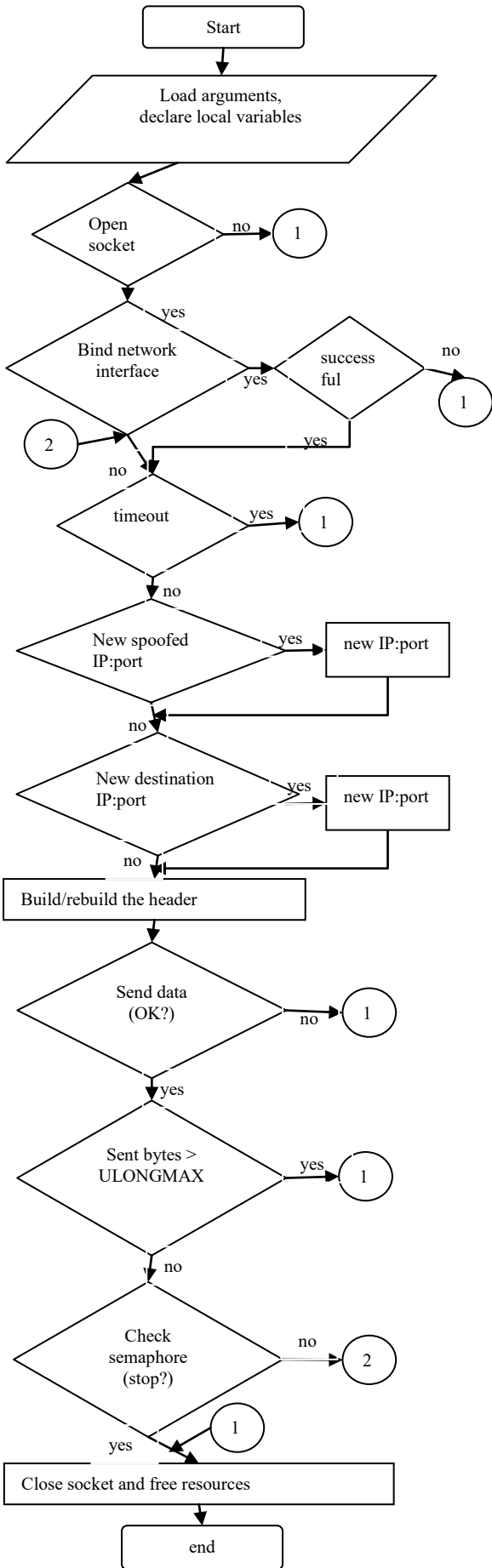


Fig. 1. UDP/TCP Flood Attack

Thus the attacker forces the server to allocate a large number of resources for each of the open connections for nonexistent clients in the extremely short period of time. So, the attacked host cannot service their regular customers during the *timeout* and deprives them of the provided service.

Figure 2 presents the console application of the proposed generator of DoS attacks. There are presented different options for generating of different attacks.

```
limits
-----
1) root access required
2) max thread pool workers: 500
3) max cmd line length: 1024
4) available devices: lo, eth0, wlan0
5) linux distribution allowing raw socket usage
-----

command      |description
-----
'pause'      |all threads will be suspended
'resume'     |all threads will be resumed
'stop'       |all threads will be interrupted and stoped
'exit'       |exit from the application
CTRL+C      |terminates the process immediately
-----

attack command |description
-----
'config [filepath]' |attack from configuration file
'tcp'              |attack using tcp packets
'udp'              |attack using udp datagrams
'icmp'             |attack 'ping of death'
'synf'             |attack flooding host with fictive tcp connections
'all'              |start all attacks at the same time
-----

option      |usage      |description
-----
[-s src_addr:src_port] |optional |spoofed source address/port
[-d dest_addr:dest_port] |mandatory |victim address/port
[-t threads] |optional |number of threads
[-tm seconds] |optional |timeout in seconds(0 = infinity)
[-i interface] |optional |network interface name
```

Fig. 2. The Parameters for Generating an Attack

```
in-app command observer started
[20:46:02 cmd] > icmp -s yahoo.com:5645 -d google.com:6543 -tm 5 -t 3 -i wlan0
> icmp -s yahoo.com:5645 -d google.com:6543 -tm 5 -t 3 -i wlan0

Parsed:
type: 3
src address: yahoo.com
src port: 5645
dest address: google.com
dest port: 6543
threads: 3
timeout: 5 sec
device: wlan0
resolved: 'yahoo.com' to '206.190.36.45'
resolved: 'google.com' to '185.4.81.115'
type: 3, 3 threads added to task-queue

[20:46:39 cmd] > udp -d facebook.com:80 -tm 3 -t 100 -i wlan0
> udp -d facebook.com:80 -tm 3 -t 100 -i wlan0

Parsed:
type: 1
src address:
src port: 0
dest address: facebook.com
dest port: 80
threads: 100
timeout: 3 sec
device: wlan0
resolved: 'facebook.com' to '69.171.230.68'
type: 1, 100 threads added to task-queue

[20:47:13 cmd] > stop
> stop
```

Fig. 3. Attacks Examples

The executed attack can be stopped for a certain period ('pause'- blocks all running threads in the pool), to be recovered ('resume'- restarts all blocked threads in the pool), fully stopped ('stop'- terminates to execute of all threads in the pool and releases used resources). There are two ways to finish work with the application – choice 'exit' or immediate close process with a keyboard shortcut

'CTRL + C'. The attacks can be run separately, but it is possible to activate them simultaneously.

Figure 3 presents the screenshot from ICMP flood attack and UDP attack to popular sites.

4. Experimental results

Tests are provided in two similar experimental situations - the attacked machine at the first situation is running under Linux Slackware, and at the second - under Windows 7. Tests are provided in the local Ethernet network.

4.1. Test 1

The victim (the host under attacks): CPU Intel Core2Duo T7700 2x2.40 GHz; RAM(physical) 2.00GB ddr2 ; OS Slackware 32-bit; interface eth0 192.168.0.213;

The attacker (the host who execute attacks): CPU Intel Core i7 vPro 4x2.13 GHz; RAM(physical) 4.00GB ddr3; OS Kali Linux 2.0 sana 64-bit; interface eth0 192.168.0.233.

4.1.1. Attack 1. UDP

The attacker generates UDP packets with replaced IP/port in the UDP header. The aim of the attack is to check for a short period of time with what quantity of packets can be overload communication channel.

The result is presented on Fig.4. The source is only one with IP 2.5.11.17 and number of port 29686. Average 7538560 byte/s (or 188464 packets) reach to victim during 1 second on communication channel.

```
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on eth0, link-type EN10MB (Ethernet), capture size 65535 bytes
16:29:35.537016 IP 2.5.11.17:29686 > 192.168.0.213:12345: Flags [S], seq
0:22, win 5840, length 22
16:29:35.537116 IP 192.168.0.213:37001 > 43.124.5.23.1234: Flags [R.], seq 0,
ack 23, win 0, length 0
16:29:35.537118 IP 43.124.5.23.1234 > 192.168.0.213:37001: Flags [S], seq
0:22, win 5840, length 22
16:29:35.537220 IP 192.168.0.213:37001 > 43.124.5.23.1234: Flags [R.], seq 0,
ack 1, win 0, length 0
```

Fig. 4. Screenshot from *tcpdump -i eth0*, started on the host under UDP flood attack

4.1.2. Attack 2. TCP

The attacker generates TCP packets with SYN = 1, with replaced IP/port in the TCP header. The aim of the attack is to check for a short period of time with what quantity of packets can be overload communication channel. In addition it checks whether the attacked machine attempts to return ACK = 1.

The result is presented on Fig.5. The source is only one with IP 43.124.5.23 and number of port 1234. Average 21976971 byte/s (or 430921packets) reach to victim during 1 second on communication channel.

```
tcpdump: verbose output suppressed, use -v or -vv for full protocol
decode
listening on eth0, link-type EN10MB (Ethernet), capture size 65535
bytes 15:23:52.758066 IP 43.124.5.23.1234 > 192.168.0.213.37001: Flags
[S], seq 0:22, win 5840, length 22
15:23:52.758187 IP 192.168.0.213.37001 > 43.124.5.23.1234: Flags [R.],
seq 0, ack 23, win 0, length 0
15:23:52.758284 IP 43.124.5.23.1234 > 192.168.0.213.37001: Flags [S],
seq 0:22, win 5840, length 22
15:23:52.758320 IP 192.168.0.213.37001 > 43.124.5.23.1234: Flags [R.],
seq 0, ack 1, win 0, length 0
```

Fig. 5. Screenshot from *tcpdump -i eth0*, started on the host under TCP flood attack

4.1.3. Attack 3. ICMP

The attacker generates ICMP packets with replaced IP in the ICMP header. The aim of the attack is to check for a short period of time with what quantity of packets can be overload communication channel.

The result is presented on Fig.6. The source is only one with IP 43.24.52.40. Average 69200 byte/s reach to victim during 1 second on communication channel.

```
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on eth0, link-type EN10MB (Ethernet), capture size 65535 bytes
15:31:45.385035 IP 43.24.52.40 > 192.168.0.213: ICMP echo request, id 0, seq
0, length 320
15:31:45.385123 IP 43.24.52.40 > 192.168.0.213: icmp
15:31:45.385158 IP 43.24.52.40 > 192.168.0.213: icmp
15:31:45.385189 IP 43.24.52.40 > 192.168.0.213: icmp
```

Fig. 6. Screenshot from *tcpdump -i eth0*, started on the host under ICMP flood attack

4.1.4. Attack 4. SYN Flood

The attacker generates TCP packets without payload, with SYN = 1 and with replaced IP/port in the TCP header. The aim of the attack is to check for a short period of time with what quantity of SYN packets can be overload communication channel. In addition it checks whether the attacked machine attempts to return ACK = 1.

The result is presented on Fig.7. Less than 1 second 756 TCP SYN packets reach to victim, each of them is from different source.

```
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on eth0, link-type EN10MB (Ethernet), capture size 65535 bytes
15:35:23.732130 ARP, Request who-has 192.168.0.5 tell 192.168.0.254, length
46
15:35:23.762543 IP 192.168.0.213.56588 > 194.141.24.15.domain: 44679+ PTR?
5.0.168.192.in-addr.arpa. (42)
15:35:24.375846 IP 132.173.65.17.46757 > 192.168.0.213.37876: Flags [S], seq
0, win 5840, length 0
15:35:24.388798 IP 129.7.180.18.39154 > 192.168.0.213.37876: Flags [S], seq
0, win 5840, length 0
15:35:24.397984 IP 152.167.96.5.38370 > 192.168.0.213.37876: Flags [S], seq
0, win 5840, length 0
15:35:24.406207 IP 8.97.147.89.35694 > 192.168.0.213.37876: Flags [S], seq 0,
win 5840, length 0
```

Fig. 7. Screenshot from *tcpdump -i eth0*, started on the host under SYN flood

4.2. Test 2

The victim (the host under attacks): CPU Intel Core2Duo T7700 2x2.40 GHz; RAM(physical) 2.00GB ddr2RAM(total virtual)

4.00GB; OS Windows 7 Enterprise 32-bit; interface eth0 192.168.0.16;

The attacker (the host who execute attacks):
CPU Intel Core i7 vPro 4x2.13 GHz;
RAM(physical) 4.00GB ddr3; OS Kali Linux 2.0
sana 64-bit; interface eth0 192.168.0.233.

4.2.1. Attack 1. UDP

The result is presented on Fig.8. The source is only one with IP 12.5.32.56:45464. During 1 second to the victim reach average 7461467 byte/s (or 181987 packets).

404691	17.577137	12.5.32.56	192.168.0.16	UDP	60	45464 + 34354	Len=13
404692	17.577137	24.13.1.61	192.168.0.16	UDP	60	29619 + 34354	Len=15
404693	17.577137	62.31.15.47	192.168.0.16	UDP	61	46672 + 34354	Len=19
404694	17.577138	12.5.32.56	192.168.0.16	UDP	60	45464 + 34354	Len=13
404695	17.577138	62.31.15.47	192.168.0.16	UDP	61	46672 + 34354	Len=19
404696	17.577138	24.13.1.61	192.168.0.16	UDP	60	29619 + 34354	Len=15
404697	17.577138	12.5.32.56	192.168.0.16	UDP	60	45464 + 34354	Len=15

Fig. 8. Screenshot from Wireshark, started on the host under UDP flood attack

4.2.2. Attack 2. TCP

The result is presented on Fig.9. The source is only one with IP 76.41.16.45:46672. During 1 second to the victim reach average 10386417 byte/s (or 185471 packets).

7908	65.	71.9.66.72	192.168.0.16	TCP	69	[TCP Retransmission] 1234 → 37001 [SYN] Seq=0 Win=5840 Len=15
7909	65.	71.9.66.72	192.168.0.16	TCP	69	[TCP Retransmission] 1234 → 37001 [SYN] Seq=0 Win=5840 Len=15
7910	65.	76.41.16.45	192.168.0.16	TCP	70	[TCP Retransmission] 1234 → 37001 [SYN] Seq=0 Win=5840 Len=16
7911	65.	71.9.66.72	192.168.0.16	TCP	69	[TCP Retransmission] 1234 → 37001 [SYN] Seq=0 Win=5840 Len=15
7912	65.	76.41.16.45	192.168.0.16	TCP	70	[TCP Retransmission] 1234 → 37001 [SYN] Seq=0 Win=5840 Len=16
7913	65.	76.41.16.45	192.168.0.16	TCP	70	[TCP Retransmission] 1234 → 37001 [SYN] Seq=0 Win=5840 Len=16
7914	65.	76.41.16.45	192.168.0.16	TCP	70	[TCP Retransmission] 1234 → 37001 [SYN] Seq=0 Win=5840 Len=16
7915	65.	71.9.66.72	192.168.0.16	TCP	69	[TCP Retransmission] 1234 → 37001 [SYN] Seq=0 Win=5840 Len=15
7916	65.	76.41.16.45	192.168.0.16	TCP	70	[TCP Retransmission] 1234 → 37001 [SYN] Seq=0 Win=5840 Len=16
7917	65.	76.41.16.45	192.168.0.16	TCP	70	[TCP Retransmission] 1234 → 37001 [SYN] Seq=0 Win=5840 Len=16
7918	65.	76.41.16.45	192.168.0.16	TCP	70	[TCP Retransmission] 1234 → 37001 [SYN] Seq=0 Win=5840 Len=16

Fig. 9. Screenshot from Wireshark, started on the host under TCP flood attack

4.2.3. Attack 3. ICMP

The result is presented on Fig.10. The source is one with IP 4.19.25.41. Less than 1 second to the victim reach 69200 bytes.

18417	33.	4.19.25.41	192.168.0.16	IPv4	414	Fragmented IP protocol (proto=ICMP 1, off=9840, ID=10e1)
18418	33.	82.3.69.254	192.168.0.16	IPv4	414	Fragmented IP protocol (proto=ICMP 1, off=49016, ID=10e1)
18419	33.	82.72.55.141	192.168.0.16	IPv4	414	Fragmented IP protocol (proto=ICMP 1, off=760, ID=10e1)
18420	33.	4.19.25.41	192.168.0.16	IPv4	414	Fragmented IP protocol (proto=ICMP 1, off=3416, ID=10e1)
18421	33.	4.19.25.41	192.168.0.16	IPv4	414	Fragmented IP protocol (proto=ICMP 1, off=9120, ID=10e1)
18422	33.	82.72.55.141	192.168.0.16	IPv4	414	Fragmented IP protocol (proto=ICMP 1, off=5696, ID=10e1)
18423	33.	4.19.25.41	192.168.0.16	IPv4	414	Fragmented IP protocol (proto=ICMP 1, off=9696, ID=10e1)
18424	33.	82.72.55.141	192.168.0.16	IPv4	414	Fragmented IP protocol (proto=ICMP 1, off=6080, ID=10e1)
18425	33.	4.19.25.41	192.168.0.16	IPv4	414	Fragmented IP protocol (proto=ICMP 1, off=9880, ID=10e1)
18426	33.	82.3.69.254	192.168.0.16	IPv4	414	Fragmented IP protocol (proto=ICMP 1, off=5886, ID=10e1)
18427	33.	82.72.55.141	192.168.0.16	IPv4	414	Fragmented IP protocol (proto=ICMP 1, off=6486, ID=10e1)

Fig. 10. Screenshot from wireshark, started on the host under ICMP flood attack

4.2.4. Attack 4. SYN Flood

The result is presented on Fig.11. Less than 1 second to the victim reach 432 TCP SYN packets, each of them is from different source.

3	2.9218.	57.48.47.56	192.168.0.16	TCP	60	46672 + 12345 [SYN] Seq=0 Win=5840 Len=0
4	2.9290.	9.10.7.6	192.168.0.16	TCP	60	28610 + 12345 [SYN] Seq=0 Win=5840 Len=0
5	2.9473.	37.12.15.28	192.168.0.16	TCP	60	28420 + 12345 [SYN] Seq=0 Win=5840 Len=0
6	2.9541.	5.26.55.45	192.168.0.16	TCP	60	28610 + 12345 [SYN] Seq=0 Win=5840 Len=0
7	2.9712.	33.36.52.42	192.168.0.16	TCP	60	35671 + 12345 [SYN] Seq=0 Win=5840 Len=0
8	2.9741.	29.32.27.49	192.168.0.16	TCP	60	41236 + 12345 [SYN] Seq=0 Win=5840 Len=0
9	2.9808.	23.17.50.39	192.168.0.16	TCP	60	28610 + 12345 [SYN] Seq=0 Win=5840 Len=0
10	2.9971.	41.25.19.54	192.168.0.16	TCP	60	41453 + 12345 [SYN] Seq=0 Win=5840 Len=0
11	3.0037.	16.35.254.46	192.168.0.16	TCP	60	28610 + 12345 [SYN] Seq=0 Win=5840 Len=0
12	3.0105.	3.38.40.53	192.168.0.16	TCP	60	28610 + 12345 [SYN] Seq=0 Win=5840 Len=0
13	3.0130.	30.51.24.21	192.168.0.16	TCP	60	41236 + 12345 [SYN] Seq=0 Win=5840 Len=0
14	3.0205.	34.20.22.43	192.168.0.16	TCP	60	28672 + 12345 [SYN] Seq=0 Win=5840 Len=0
15	3.0309.	13.18.44.14	192.168.0.16	TCP	60	41236 + 12345 [SYN] Seq=0 Win=5840 Len=0
16	3.0334.	11.2.31.4	192.168.0.16	TCP	60	28389 + 12345 [SYN] Seq=0 Win=5840 Len=0

Fig. 11. Screenshot from wireshark, started on the host under SYN flood attack

The results for both tests are similar - in UDP and ICMP attacks the observed total number of B/s are very close. However during TCP Flood and SYN Flood attacks, where the used protocol is TCP, it is observed discrepancy approximately twice the volume of records. It is important to point that during the tests the firewall of victim is "on". The probable cause for the discrepancy in results is precisely defined restrictive policy that is imposed over the incoming data flow.

5. Conclusion

This paper presents a generator of DoS attacks. There are implemented and tested four attacks - UDP attack, TCP attack, ICMP attack, SYN flood attack.

The main goal of the future work is to extend functionality, providing this program with more popular types of attacks as HTTP flooding, which received popularity in recent years with attacks on various public websites and web services, TCP Session Hijacking and ARP poisoning, because they are in the root of different *man in the middle* attacks – rerouting, change / replacement of packages or sniffing, etc.

The possibility of multiple instantiation of the program on single machine would also be a good step for future work. Multiprocessing start of attacking program would allow more realistic simulation of DDoS attack, launched from a single machine, but in the same time this would allow to control synchronization of execution of DoS attacks from several machines, which work in parallel in master-slave mode (DDoS attack).

REFERENCES

1. Kaspersky Lab, Kaspersky DDoS Intelligence Report for Q4 2015, (2016), <https://securelist.com/analysis/quarterly-malware-reports/73414/kaspersky-ddos-intelligence-report-for-q4-2015/>
2. LLOYD'S, (2016), www.lloyds.com/
3. Mirkovic J, S. Dietrich, D. Dittrich, P. Reiher, "Internet Denial of Service: Attack and Defense Mechanisms", Prentice Hall 2004, ISBN: 0-13-147573-8

4. Zargar S.T., J. Joshi, D. Tipper, "A Survey of Defense Mechanisms Against DDoS Flooding Attacks", IEEE Communications Surveys & Tutorials, Vol. 15, Issue: 4, 2013, pp. 2046 – 2069, ISSN:1553-877X.
5. <https://sourceforge.net/projects/loic/>
6. <https://sourceforge.net/projects/xoic/>
7. <https://sourceforge.net/projects/pyloris/>
8. www.hping.org/download.html
9. golden-eye.soft112.com
10. pyloris.soft112.com

Address: Technical University of Varna,
Str. „Studentska”1
Phone: 052/383 439
E-mail: valeksieva@tu-varna.bg
E-mail: ijelqzkov@live.com



DEVELOPING OF A LIBCLOUD DRIVER FOR OWNCLOUD

SIMEON TSVETANOV

Abstract: *Apache Libcloud [1] is a Python library that abstracts the differences between cloud providers and allows unified access to different cloud resources through an easy to use API. Although Libcloud supports impressive number of providers and methods [3], it doesn't support the popular storage platform ownCloud [2] and that was the motivation for developing a Libcloud driver for ownCloud.*

Key words: *Cloud Technologies, Libcloud, owncloud, Webdav.*

1. Introduction

Cloud storages are kind of online storage sites where the data is stored in data centers. The hosting companies operate with large data centers, so people who need to store their data, purchase or rent a storage capacity of them. People or organizations hire or purchase cloud space in order to store customers, user or corporate data. On the other hand, cloud customers can use the storage to store their provided applications. Data in the cloud space can be accessed through APIs or through a web-based system for managing content [4].

Grid and cloud technologies provide an abstraction of the computational resources and make the access to them much easier, but users still face the problem of interacting with the different interfaces of the different service providers. Libcloud aims to solve this problem by providing unified interface for most of the cloud service providers, fig 1.

Driven by increasing demand for standards in cloud technologies [5] Apache Libcloud supports more than 30 cloud providers including Amazon EC2 (Elastic Cloud Compute), Rackspace Cloud, Terremark, IBM Cloud, GoGrid Linode, etc.

Libcloud acts as a driver in order to abstracts the differences between cloud providers and to allow unified access to different cloud resources. Provided services are divided into the following categories:

- Compute – allows the user to manage cloud and virtual servers offered by different providers;
- Cloud Storage - services such as Amazon S3 and Rackspace CloudFiles;
- Load Balancers as a Service - services such as Amazon Elastic Load Balancer and GoGrid Load Balancers;

- DNS as a Service - services such as Amazon Route 53 and Zerigo;
- Container - to install and deploy containers onto container based virtualization platforms;
- Backup - to manage Backup as a Service and services such as EBS Snaps, GCE volume snap and dimension data backup.

Some frequently used term in Libcloud:

Node - represents a cloud or virtual server.

NodeSize - represents node hardware configuration. Usually this is amount of the available RAM, bandwidth, CPU speed and disk size.

NodeImage - represents an operating system image.

NodeLocation - represents a physical location where a server can be.

NodeState - represents a node state. Standard states are: running, stopped, rebooting, terminated, pending, and unknown.

KeyPair - represents an SSH key pair object.

StorageVolume - represents a storage volume

VolumeSnapshot - represents a point in time snapshot of a StorageVolume

Most users have long been looking carefully which is the largest, free and qualitative package with cloud services that can be used. The right choice of provided services is important because of subsequent difficulties in migration from one cloud provider arise the need to restructure files and there are enormous effort needed for migration.

OwnCloud [6] is a suite of client-server software for creating file hosting services and using them. OwnCloud is functionally very similar to the widely used Dropbox, with the primary functional difference being that OwnCloud is free and open-

source, and thereby allowing anyone to install and operate it without charge on a private server.

Its open architecture and APIs allow IT to easily add corporate branding, and extend core functionality as needed to meet the company's evolving needs. From enterprise directories to SAML, SharePoint or Windows network drives, ownCloud allows you to leave data where it lives while exposing secure access to users on any device.



Fig. 1. Accessing different cloud providers through Libcloud API

OwnCloud is in the top of any competitive analysis and has been preferred for self-hosted storage platform, but that opens the issue about the integration with the rest of the system, fig 2.

2. Design of the Libcloud driver

Before starting with the implementation of the new driver, some investigation about the possible ways to interact with ownCloud is needed.

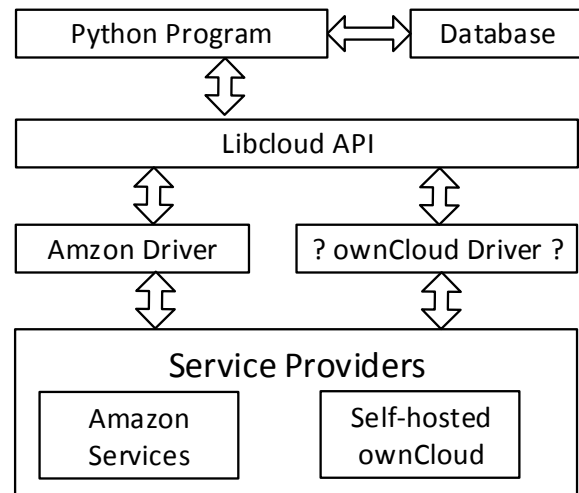


Fig. 2. Global view of the system is showing where the new ownCloud driver is required

OwnCloud is written in PHP and has a local database for storing the whole file structure and the application.

The platform support WebDAV [7] protocol for file transfer that could be used for desired interaction. Web Distributed Authoring and Versioning (WebDAV) is an extension of the Hypertext Transfer Protocol (HTTP) that allows clients to perform remote Web content authoring operations. It provides a framework for users to create, change and move documents on a server; typically a web server or web share. The most important features of the WebDAV protocol include the maintenance of properties about an author or modification date, namespace management, collections, and overwrite protection.

To use its features, a new Libcloud module for WebDav communication is needed. Then the developed driver could use that WebDAV module in order to provide the required functionality.

This approach allows Libcloud to interact with any ownCloud server and with minor modification to any server supporting WebDav protocol.

3. Development of the new driver

The first step of developing the new driver is to install Libcloud. The library is available on PyPi [8] and could be easily installed using pip.

```
pip install apache-libcloud
```

Then a new module for WebDAV interaction (presented in webdav.py file) has to be added in libcloud/utills folder.

Next step is to create a driver which uses the already created WebDav module in order to send requests to the server. The driver file is called

owncloud.py and has to be added in libcloud/storage/drivers folder.

Then is time to bind the driver with the appropriate provider by adding the following lines in libcloud/storage/providers.py

```
Provider.OWNCLOUD:
('libcloud.storage.drivers.owncloud',
 'OwnCloudStorageDriver'),
```

Here is the list of affected files.

New files:

webdav.py file in libcloud/utils folder

owncloud.py file in libcloud/storage/drivers folder

Modified:

libcloud/storage/**providers.py** added entry for owncloud as a provider

The code is available on Github <https://github.com/hpcbg/libcloud> under GNU GPL.

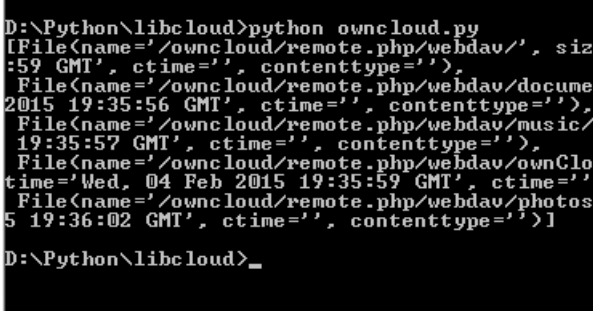
This is how to create new Libcloud driver in general and it particular to create one for interaction with any server that supports WebDAV. The new driver could be used in a standard manner and users just need to provide the appropriate credentials to get access to the files.

4. Testing and experiments

OwnCloud could be installed on a standard LAMP [9] server in order to use its embedded WebDAV functionality.

First test is just to list the files on the server WebDAV directory.

```
from pprint import pprint
# Obtain reference to the provider driver
from libcloud.storage.types import Provider
from libcloud.storage.providers import get_driver
OwnCloudFiles=get_driver(Provider.OWNCLOUD)
# Instantiate the driver with your provider
credentials
driver=OwnCloudFiles('example.com', 'username',
'password')
# Start using the driver
basePath='/owncloud/remote.php/webdav'
containers=driver.list_all(basePath)
pprint(containers)
```

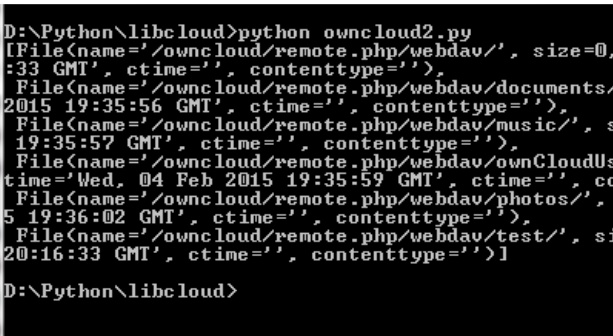


```
D:\Python\libcloud>python owncloud.py
[File(name='/owncloud/remote.php/webdav/', size
:59 GMT', cttime='', contenttype=''),
 File(name='/owncloud/remote.php/webdav/docume
2015 19:35:56 GMT', cttime='', contenttype=''),
 File(name='/owncloud/remote.php/webdav/music/
19:35:57 GMT', cttime='', contenttype=''),
 File(name='/owncloud/remote.php/webdav/ownClo
time='Wed, 04 Feb 2015 19:35:59 GMT', cttime='
File(name='/owncloud/remote.php/webdav/photos
5 19:36:02 GMT', cttime='', contenttype='')]
D:\Python\libcloud>_
```

Fig. 3. Listing the files on the server

Second test is to create new directory.

```
containers=driver.create_directory(basePath+'/test')
containers=driver.list_all(basePath)
pprint(containers)
```

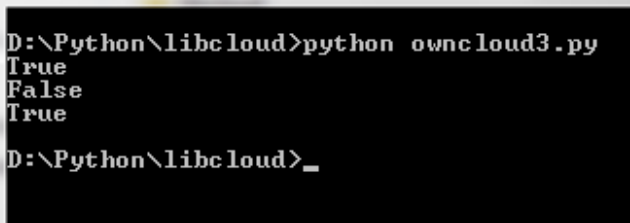


```
D:\Python\libcloud>python owncloud2.py
[File(name='/owncloud/remote.php/webdav/', size=0,
:33 GMT', cttime='', contenttype=''),
 File(name='/owncloud/remote.php/webdav/documents/
2015 19:35:56 GMT', cttime='', contenttype=''),
 File(name='/owncloud/remote.php/webdav/music/', s
19:35:57 GMT', cttime='', contenttype=''),
 File(name='/owncloud/remote.php/webdav/ownCloudUs
time='Wed, 04 Feb 2015 19:35:59 GMT', cttime='', cc
File(name='/owncloud/remote.php/webdav/photos/',
5 19:36:02 GMT', cttime='', contenttype=''),
 File(name='/owncloud/remote.php/webdav/test/', si
20:16:33 GMT', cttime='', contenttype='')]
D:\Python\libcloud>
```

Fig. 4. Creating new test directory

Third test is to check the existence of a file.

```
exists = driver.check_exists(basePath+'/music')
pprint(exists)
exists = driver.check_exists(basePath+'/music2')
pprint(exists)
exists=driver.check_exists(basePath+'/UserManual
.pdf')
pprint(exists)
```



```
D:\Python\libcloud>python owncloud3.py
True
False
True
D:\Python\libcloud>_
```

Fig. 5. Check the existence of a file.

As is shown on the picture above the program check the existence of a file and return True or False if it exist or not.

Those were just simple examples of some basic features in order to present the idea of using the unified interface to access remote information resources.

The driver is used in standard way with make it easy to write less code using Python language and Libcloud library to access file from the remote server.

5. Conclusion

The paper presents the development of a new driver which extends Libcloud library in order to access files on ownCloud server.

This approach extends the Libcloud functionality allowing unified access to different cloud resources through an easy to use API. Users might manipulate impressive number of providers and methods, but also can access the popular storage platform ownCloud using developed driver.

Experiments proved the correct work of the basic functionality for manipulating files like: create, rename, delete, upload, etc.

The developed driver allows Libcloud to interact with any ownCloud server and with minor modification to any server supporting WebDav protocol. The new driver could be used in a standard manner and users just need to provide the appropriate credentials to get access to the files.

REFERENCES

1. <https://libcloud.apache.org>
2. <https://owncloud.org/>
3. https://libcloud.readthedocs.io/en/latest/supported_providers.html
4. Tsvetanov, S., Doundakova, I., "Cloud Storage Integration Platform", International Scientific Conference Computer Science'2015, September 8-10, Durres, Albania.
5. Edmonds, A., Metschm, T., Papaspyrou, A., Richardson, A., Toward an Open Cloud Standard, IEEE Internet Computing, Issue No.04 - July-Aug. (2012 vol.16), pp: 15-25,
6. <https://owncloud.com/owncloud-architecture-overview/>
7. <http://www.webdav.org/specs/rfc2518.html>
8. <https://pip.pypa.io/en/stable/>
9. https://doc.owncloud.org/server/8.0/admin_manual/installation/source_installation.html

Author's contacts

Address:

Sofia 1756, 8 Kliment Ohridski blvd.

E-mail: s_tsvetanov@tu-sofia.bg



GENERATING 2D GAUSSIAN KERNELS FOR THE PURPOSE OF STUDYING FPGA-BASED LINEAR SPATIAL FILTERS

IVAN KANEV, PETYA PAVLOVA, DIMITRE KROMICHEV, JORDAN GENOFF

Abstract This paper presents one approach to generating normalized 2D Gaussian kernels. Defined are the conditions for determining the coefficients of the kernels. Selected is an approach through which the kernel coefficients are represented as positive integers. This approach aims at speeding up the time to calculate the linear spatial filters (LSF) and reducing the number of logic elements. An algorithm for generating 2D Gaussian kernels is proposed. The studies are conducted using the software tool "Gaussian Kernel Generator (GKG)". As a result of the studies, the Gaussian kernel coefficients, which are represented as real and integer numbers, are determined on the basis of the preset criteria. The obtained results can be employed in designing and studying LSF through using FPGA platforms.

Keywords: linear spatial filters, LSF, weighted average, normalized Gaussian 2D Kernel, VHDL, FPGA

1. Introduction

Filtering of images by employing linear spatial filters (LSF) is one of the basic tools used in image processing. [1]. The development of Field Programmable Gate Arrays (FPGA) and ASIC technologies suggests a new modern approach to the hardware realization of LSF [4]. The filtering of an input image $p(x, y)$ of raster size $M \times N$.

Results in computing new values for the gradations of each pixel - $g(x, y)$ depending on the gradations of the pixels around it. The filtering effect is determined by the mask of weighted coefficients $w(a, b)$ defined within a window of neighbourhood of size $S \times S$, where S is an odd number and $S \geq 3$, and the mask contains finite number of elements [2]. One of the most common LSF is the weighted average filter which is defined by the following expression [3]:

$$g(x, y) = \frac{\sum_{a=-k}^k \sum_{b=-k}^k w(a, b) \cdot p(x+a, y+b)}{\sum_{a=-k}^k \sum_{b=-k}^k w(a, b)}, \quad (1)$$

where

$$k = \frac{S-1}{2}, \quad k = \{1, 2, \dots\}; \quad (2)$$

$$x = 0, 1, 2, \dots, M-1; \quad (3)$$

$$y = 0, 1, 2, \dots, N-1. \quad (4)$$

In (1) x, y are used to denote the global coordinates of pixels in the input image $p(x, y)$, and a, b denote the relative coordinates of the same image within the window of size $S \times S$.

Most often the realization of $w(a, b)$ is based on a normalized Gaussian 2D operator, also designated as Gaussian kernel. [3].

The normalized Gaussian 2D operator is defined by the following expression:

$$w(a, b) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{a^2 + b^2}{2\sigma^2}\right). \quad (5)$$

where

$$-k \leq a, b \leq k. \quad (6)$$

Designing LSF by employing VHDL or Verilog demands the generation of Gaussian kernels to be used in the study of components utilized in the realization of FPGA platform based filters. This way, the selection of conditions for generating Gaussian kernels

focuses on the tool set to be used in defining the parameters of components through which the filters will be realized.

To achieve this objective, the following tasks are to be performed:

- to conduct studies focused on the Gaussian kernel generation; the aim is defining basic parameters of the components;
- to define an algorithm for generating test Gaussian kernels;
- to conduct studies utilizing tools for Gaussian kernel generation

2. Gaussian kernel generation

The normalized Gaussian 2D operator being used, the Gaussian kernel weighted coefficients are a function of two arguments - k and σ . Taking into account that k are calculated in advance (2), for a certain k the function (5) depends only on σ . Let us analyze two specific cases in defining the values of weighted coefficients in (5) within the limits of a and b :

1. Let us assume that

$a, b = k$ (or a and $b = -k$), then

$$w(k, k) = \frac{1}{2\pi\sigma^2} \exp - \frac{k^2}{\sigma^2}, \quad (7)$$

and the function (5) has a minimum.

2. Let us assume that

$a, b = 0$, then

$$w(0, 0) = \frac{1}{2\pi\sigma^2} \quad (8)$$

function (5) has a maximum.

From (6) and (7) a conclusion can be easily arrived that the discrete values of $w(a, b)$ (5) for $\sigma > 0$ are confined within the interval

$$0 < w(a, b) < 1 \quad (9)$$

and are positive real numbers.

In this study, the Gaussian kernel generation is to be realized under the following conditions:

#1. For the purpose to avoid calculating $w(a, b)$ (5) using components represented by floating point numbers, Gaussian kernel coefficients have to be represented as positive integers. This will speed up the computation time and reduce the number of logical elements being used.

#2. Filtering through the use of weighted average filter requires that all the pixels of the input image should take part into the computation of the image being filtered. This condition implies that all the coefficients of the Gaussian kernels are different from zero.

Let us use $w_{int}(k, k)$ to denote the minimum value of the function (5), represented as a positive integer.

Let us assume that the following dependence is valid:

$$w_{int}(k, k) = \lfloor w(k, k) \cdot V_{int} \rfloor = C_{min}, \quad (10)$$

where:

$$V_{int},$$

is a coefficient which calculates the function (5) down to positive integers;

$$C_{min},$$

$C_{min} = \{1, 2, \dots\}$ is a constant defining the concrete value of $w_{int}(k, k)$.

Then for the conditions #1 and #2 to be fulfilled, for a certain C_{min} , it is enough to define such a value of V_{int} , that is capable of satisfying the equality (9).

This statement can be supported by the following arguments:

Let us use $w_{int}(a, b)$ to denote the Gaussian kernel coefficients calculated down to positive integers. Then

$$w_{int}(a, b) = \lfloor w(a, b) \cdot V_{int} \rfloor \geq C_{min}, \quad (11)$$

By which the conditions #1 and #2 are satisfied due to the fact that all the Gaussian kernel weighted coefficients are positive integers, different from zero.

3. Algorithm for test Gaussian kernel generation

VHDL and Verilog require that the filter description should contain the definition of the signal width in bits which allows for its components to be parameterized. Let us use $2^m, 2^{m+1}$ to define the interval of possible solutions, where $m, m+1$ denote the number of binary digits within the interval containing the possible solution.

Start of the algorithm

Input parameters: k (2); C_{min} (10).

Step #1. Defining the initial value of m .

For a k , m can be defined by comparing the number of binary combinations within the interval

$$2^m, 2^{m+1} \quad (12)$$

to the number of the elements in the Gaussian kernel (5).

Let us have

$$2^m \geq (2k+1)^2, \quad (13)$$

then

$$m = \left\lceil \log_2 \left((2k+1)^2 \right) \right\rceil. \quad (14)$$

Step #2. Defining V_{int} and V_R

V_{int} can be defined on the basis of the upper limit of the interval which can contain the possible solution:

$$V_{int} = 10^{\left\lceil \log(2^{m+1}+1) \right\rceil} \quad (15)$$

Let us use V_R to denote the reciprocal value of V_{int} normalized to V_{int}

$$V_R = \frac{C_{min}}{V_{int}}. \quad (16)$$

V_R is introduced for the purpose of computing the Gaussian kernels within the

interval denoted by (13). This approach allows an evaluation of the error resulting from calculating the Gaussian kernel coefficients from real numbers down to integers.

Step #3. Defining σ_L for the lower limit of the interval.

Let us use $w_L(0,0)$ to denote the maximum value of the Gaussian function for the lower limit of the interval assuming that:

$$w_L(0,0) = 2^{m+1} V_R. \quad (17)$$

Then from (7) σ_L for the lower limit of the interval can be defined:

$$\sigma_L = \frac{1}{\sqrt{2\pi \cdot 2^{m+1} V_R}}. \quad (18)$$

Step #4. Defining σ_H for the upper limit of the interval.

Let us use $w_H(0,0)$ to denote the maximum value of the Gaussian function for the upper limit of the interval assuming that:

$$w_H(0,0) = 2^m V_R, \quad (19)$$

Then from (7) σ_L for the upper limit of the interval can be defined:

$$\sigma_H = \frac{1}{\sqrt{2\pi \cdot 2^m V_R}} \quad (20)$$

Step #5. Defining $w(a,b)$ and $w_{int}(a,b)$.

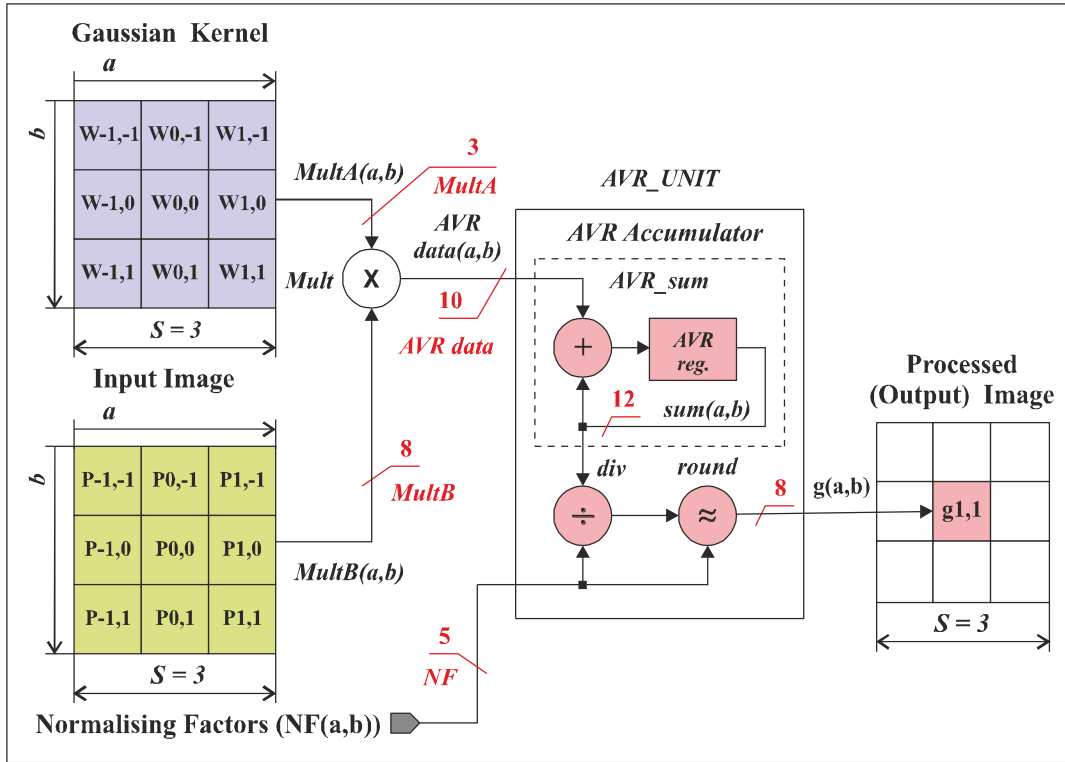


Fig. 1. Conceptual model of FPGA-based weighted average filter $k=1$

Within the interval

$$\sigma_L \leq \sigma \leq \sigma_H \quad (21)$$

a value of σ for which

$$w(k,k) = V_R \quad (22)$$

is sought.

If the equality is not satisfied, the algorithm is repeated from *step* #2 for

$m = m + 1$. Otherwise the calculated value for σ is used to compute $w(a,b)$ (5) and the Gaussian kernel integer coefficients:

$$w_{\text{int}}(a,b) = \lfloor w(a,b) \cdot C_{\text{int}} \rfloor. \quad (23)$$

End of algorithm

4. Conceptual model and operations for the hardware realization of weighted average filters

Defining the conceptual module of the components used in the hardware realization of weighted average filters is the first very important step for their complete study and development. The conceptual model (Figure 1) has to suggest solutions to the following problems: the basic operations by which the components will be realized; the

subcomponents that will be used in the realization of these operations; definitions of the signals that connect different subcomponents.

The hardware realization of the weighted average filters can be achieved through employing the following operations:

Multiplying the elements characterized by the same coordinates in the Gaussian Kernel and Input Image matrices. Generally, this operation is determined by the following dependence:

$$AVRdata(a,b) = w(a,b) \cdot p(a,b). \quad (24)$$

Operation (24) can be realized in terms of subcomponent for hardware multiplication *Mult*, embedded in FPGA. If we assume that the coefficients of weights $w(a,b)$ and the current pixels $p(a,b)$, are stored in memories, then (24) can be presented in the following way:

$$AVRdata(a,b) = MultA(a,b) \cdot MultB(a,b). \quad (25)$$

The products $AVRdata(a,b)$ define an input data flow which is used in the component AVR_UNIT computations that are to follow.

1. Summation of products $AVRdata(a,b)$. The operation $sum(a,b)$ in component AVR_UNIT computes the dividend of the filter (1):

$$sum(a,b) = \sum_{a=-k}^k \sum_{b=-k}^k AVRdata(a,b). \quad (26)$$

Operation $sum(a,b)$ is realized by subcomponent $AVR_Accumulator$ which includes adder AVR_sum and register AVR_reg . The adder receives the current value of $AVRdata(a,b)$ and the current value of AVR_reg as operands. All the values $AVRdata(a,b)$ having been added up, the output value of $sum(a,b)$ is stored in the register AVR_reg .

2. Normalization of filter pixels $g(a,b)$. This operation is aimed at equalizing the pixel width in the input image and the pixel width in the filtered image. Generally, this operation is determined by the following expression:

$$g(a,b) = div + round, \quad (27)$$

where :

$$div = \frac{sum(a,b)}{NF(a,b)}, \quad (28)$$

is an operation by which the filtered pixels are normalized through employing division;

$NF(a,b)$ is normalization factor, and

$$NF(a,b) = \sum_{a=-k}^k \sum_{b=-k}^k w(a,b); \quad (29)$$

$round$ is a factor used to perform rounding over the result of operation div .

5. Rules for calculating the width of signals and subcomponents

The integer coefficients of Gaussian kernel having been defined, it is possible for the width of signals: $MultA(a,b)$; $MultB(a,b)$; $AVRdata(a,b)$; NF ; $sum(a,b)$, and the normalization factor of the output function $NF_{int}(a,b)$ to be calculated (Fig.1). Defining the width of signals is required in the designing of weighted average LSF with VHDL or Verilog.

#1. Let us use $MultA$ to denote the width of saignals $MultA(a,b)$ (25), and w_{max} to

denote the central element value in the weighted coefficients matrix:

$$w_{max} = \lfloor w(0,0) \cdot C_{int} \rfloor, \quad (30)$$

then:

$$MultA = \lceil \log_2(w_{max} + 1) \rceil. \quad (31)$$

#2. Let us use $MultB$ to denote the width of saignals $MultB(a,b)$ (25), and p_{max} to denote maximum value of pixels in the input image.

Then:

$$MultB = \lceil \log_2(p_{max} + 1) \rceil. \quad (32)$$

For the purposes of this study, it is assumed that the input image pixels are confined to the interval:

$$0 \leq p(a,b) \leq p_{max}, \quad (33)$$

where $p_{max} = 255$.

#3. Let us use $AVRdata$ to denote the width of signals $AVRdata(a,b)$,

then:

$$AVRdata = \lceil \log_2((p_{max} + 1) \cdot (w_{max} + 1)) \rceil. \quad (34)$$

#4. Let us use NF to denote the width of signals $NF_{int}(a,b)$, where:

$$NF_{int}(a,b) = \sum_{a=-k}^k \sum_{b=-k}^k w_{int}(a,b), \quad (35)$$

is the value of normalization factor of the filtered pixel (29), computed by utilizing the Gaussian kernel integer coefficients.

Then:

$$NF = \lceil \log_2(NF_{int}(a,b)) \rceil. \quad (36)$$

#5. Let us use sum to denote the width of saignals $sum(a,b)$ (26).

Then:

$$sum = \lceil \log_2(NF_{int}(a,b) \cdot (p_{max} + 1)) \rceil. \quad (37)$$

6. Results

On the basis of the algorithm for generating test Gaussian kernels, a Gaussian kernel generator (GKG) has been proposed.

For certain k and C_{\min} , the employing of GKG presents the following results:

- The value of σ , for which $w(k,k)=V_R$, is computed

- The weighted coefficients of test Gaussian kernels $w(a,b)$, represented by integers, are generated.

- The weighted coefficients of test Gaussian kernels $w_{\text{int}}(a,b)$, calculated down to positive integers, are generated. For the graphical representation of $w_{\text{int}}(a,b)$, GKG generates a 3D figure.

- The normalization factor integer value $NF_{\text{int}}(a,b)$ is computed.

- The width of signals *MultA* (32); *MultB* (33); *AVRdata* (35); *NF* (36); *sum* (38) is calculated. The normalization factor of the filtered pixel $NF_{\text{int}}(a,b)$ is computed.

- A file containing the weighted coefficients of Gaussian kernels $w_{\text{int}}(a,b)$ is generated. This file is used by FSM ROMizer [6] for the purpose of designing of ROM-based FSMs that control components involved in the realization of weighted average filters.

The following tables and figures present the results, obtained by employing GKG for $k = \{1, 2, 3\}$ and $C_{\min} = 1$:

6.1. Test kernels:

$$k = 1, NF_{\text{int}}(a,b) = \{9, 10, 16\}$$

Table 1

Test kernel Box Filter- $k=1$;
 $NF_{\text{int}}(a,b) = 9$

a/b	-1	0	1
-1	1	1	1
0	1	1	1
1	1	1	1

Table 2

Test kernel - $k = 1, NF_{\text{int}}(a,b) = 10$.

a/b	-1	0	1
-1	1	1	1
0	1	2	1
1	1	1	1

Table 3

Test kernel $k = 1; NF_{\text{int}}(a,b) = 16$.

a/b	-1	0	1
-1	1	2	1
0	2	4	2
1	1	2	1

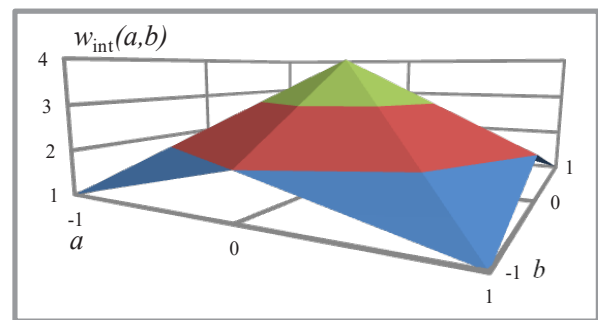


Fig. 2. Graphical representation of function $w_{\text{int}}(a,b) : k = 1$;

6.2. Generating Gaussian kernels with GKG – $k = 2$

Table 4

Conditions for Generating test Gaussian kernels – $k = 2$

k	C_{\min}	m	V_{int}	σ
2	1	7	1000	0,86447891

Table 5

Test Gaussian kernels for $k = 2$ – discrete values of function $w(a, b)$ (5) represented by real numbers

a/b	-2	-1	0	1	2
-2	0,001008740	0,007507216	0,014657011	0,007507216	0,001008740
-1	0,007507216	0,055869961	0,109079929	0,055869961	0,007507216
0	0,014657011	0,109079929	0,212966514	0,109079929	0,014657011
1	0,007507216	0,055869961	0,109079929	0,055869961	0,007507216
2	0,001008740	0,007507216	0,014657011	0,007507216	0,001008740

Table 6

Test Gaussian kernels for $k = 2$ – discrete values of function $w(a, b)$ (5) represented by integers.

a/b	-2	-1	0	1	2
-2	1	8	15	8	1
-1	8	56	109	56	8
0	15	109	213	109	15
1	8	56	109	56	8
2	1	8	15	8	1

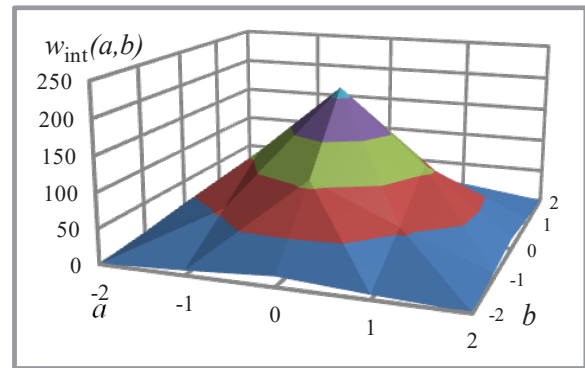


Fig. 3. Graphical representation of function $w_{\text{int}}(a, b) - k = 2$

6.3. Generating Gaussian kernels with GKG – $k = 3$

Table 7

Conditions for Generating test Gaussian kernels - $k = 3$

k	C_{\min}	m	V_{int}	σ
3	1	9	10000	1,12344622

Table 8

Test Gaussian kernels for $k = 3$ – discrete values of function $w(a, b)$ represented by real numbers .

a/b	-3	-2	-1	0	-1	-2	-3
-3	0,000100890	0,000731292	0,002400131	0,003566836	0,002400131	0,000731292	0,000100890
-2	0,000731292	0,005300675	0,017397039	0,025853740	0,017397039	0,005300675	0,000731292
-1	0,002400131	0,017397039	0,057097812	0,084853060	0,057097812	0,017397039	0,002400131
0	0,003566836	0,025853740	0,084853060	0,126100135	0,084853060	0,025853740	0,003566836
-1	0,002400131	0,017397039	0,057097812	0,084853060	0,057097812	0,017397039	0,002400131
-2	0,000731292	0,005300675	0,017397039	0,025853740	0,017397039	0,005300675	0,000731292
-3	0,000100890	0,000731292	0,002400131	0,003566836	0,002400131	0,000731292	0,000100890

Table 9

Test Gaussian kernels for $k = 3$ - discrete values of function $w(a,b)$ represented by integers

a/b	-3	-2	-1	0	-1	-2	-3
-3	1	7	24	36	24	7	1
-2	7	53	174	259	174	53	7
-1	24	174	571	849	571	174	24
0	36	259	849	1261	849	259	36
-1	24	174	571	849	571	174	24
-2	7	53	174	259	174	53	7
-3	1	7	24	36	24	7	1

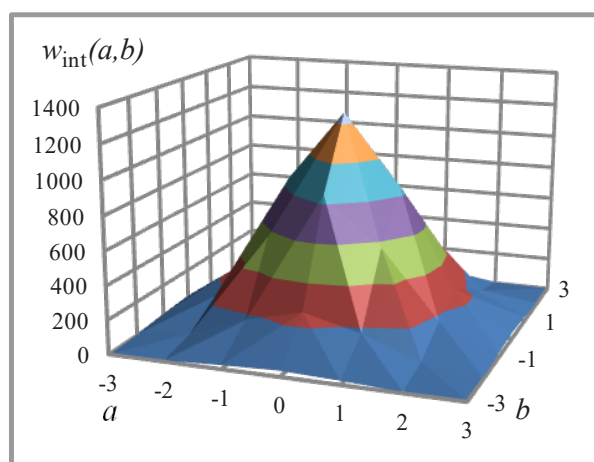


Fig. 4. Graphical representation of function $w_{\text{int}}(a,b) - k = 3$

Table 10

Signal width and values of the filtered pixel normalization factor

k	1*	1	2	3
<i>MultA</i>	1	3	8	11
<i>MultB</i>	8	8	8	8
<i>AVRdata</i>	8	10	16	19
<i>Sum</i>	12	12	18	22
<i>NF</i>	4	5	10	14
$NF_{\text{int}}(a,b)$	9	16	993	9974
<i>Remark*: Box mask</i>				

An example for using the data in Table 10 for $k=1$ are presented in Fig. 1.

7. Conclusion

The studies conducted in this paper are focused on generating Gaussian kernels for the purpose of determining basic parameters of the components used in the realization of FPGA-based LSF. An original algorithm for generating Gaussian kernels is

proposed. A conceptual model of FPGA-based LSF is defined. Defined are the signals and components needed for the hardware realization of LSF. A Gaussian kernel generator (GKG) is proposed. GKG is utilized to calculate the width of the signals defined in the conceptual model. The results obtained are employed in the designing and studying of FPGA-based linear spatial filters [6]

References

- [1] Ganguly S. et al., Automatic Analysis of Smoothing Techniques by Simulation model Based Real-Time System for Processing 3D Human Faces, International Journal of Embedded Systems and Applications (IJESA) Vol.4,No.4, 2014
- [2] Gonzalez R., Woods R., Digital Image Processing, Prentice Hall, 2008
- [3] Hanumantharaju M. C. et al, Design of Novel Algorithm and Architecture for Gaussian Based Color Image Enhancement System for Real Time Applications Advances in Computing, Communications and Informatics (ICACCI) International Conference, 2013
- [4] Uwe Meyer-Baese, Digital Signal Processing with Field Programmable Gate Arrays (FPGA), 4th Edition, ISBN 978-3642453083, Springer, 2014
- [5] Bailey D. G., Design for Embedded Image Processing on FPGAs, First Edition. John Wiley & Sons (Asia) Pte Ltd. Published by John Wiley & Sons (Asia) Pte Ltd. ISBN: 978-0-470-82849-6, 2011
- [6] Kanev I., An Approach to FPGA- based Design of Weighted Average Linear Spatial Filters, Journal of the Technical University – Sofia Plovdiv branch, Bulgaria “Fundamental Sciences and Applications” Vol. 21, pp 263-268 ISSN1310-8271, 2015.

Department of Computer Systems and Technologies, Technical University - Sofia, Branch Plovdiv 25 Tsanko Dyustabanov Str.
 Email: itkanev@yahoo.com
 Email: p_pavlova@tu-plovdiv.bg
 Email: jgenoff@tu-plovdiv.bg
 University of Plovdiv “Paisii Hilendarski”, Tzar Assen str. 24
 Email: dkromichev@yahoo.com



LOOK-UP TABLE-BASED MICROPROCESSOR ENERGY MODEL

LUBOMIR BOGDANOV

Abstract: *The following paper discusses the creation of an energy model of a single-core, single-threaded, deeply-embedded microprocessor for the needs of an Instruction Set Simulator (ISS) called Powot Simulator. The modeled architecture is ARM Cortex-M4 but the measurement methods used are universal and can be applied to others as well. The main goal of the research is simplicity and therefore a look-up table of the target ISA with energy costs is implemented.*

Key words: *energy model, microprocessor, embedded systems, look-up table*

1. Introduction

Estimation of the consumed energy at early stages of the development is important in embedded systems with battery supplies and eco-friendly equipment. Developers use various kinds of simulators like gem5, SimplePower, Watch, SESAME, etc. Some of them allow performance simulations, others – energy, and sometimes we see both. In [1] we proposed an Instruction Set Simulator (ISS) called Powot Simulator for estimating the energy consumption of a microcontroller that executes single-threaded, OS-less C program. The model used there is fairly simple – an energy value is assigned to each instruction's mnemonic. The only parameters that are varied are:

- Microprocessor frequency
- Number of instruction operands.

Those parameters were called model domains. They reflect the set of possible energy consumptions depending on the user-defined system frequency and compiler generated assembler instructions with different number of operands (respectively – different addressing modes).

For a first version of the ISS, the model complexity was enough. The ISS estimated the energy consumption on a function call basis. However, with the release of Powot Simulator version 2 the energy evaluation is statement-based. In other words, the analysis takes place inside each C function. The modeling detail has to be increased with additional information about Dynamic Frequency Scaling (DFS) and Dynamic Voltage Scaling (DVS) APIs that separate code in different frequency and voltage regions (or domains). To make the simulator better, other domains are introduced as well – memory, temperature and

instruction type. This leads to rethinking the model concept – the new simulator version requires a new energy model.

2. Related work

Energy modeling in embedded systems is done in different levels of abstraction [2] – starting from the highest, ESL, and going down to the lowest transistor-level. In [3] the energy simulation takes place at the highest level. The environment used is called SESAME and the purpose of the research is to speed-up energy simulation. This would allow faster Design Space Exploration (DSE) of the developed system. The technique used is called signature-based power modeling. The application, architecture and mapping are separated. Application events such as process execution and process communication are reflected in the architecture specification with event tables containing timing values. Those tables are expanded to support energy simulation as well. However if the power/energy measurements are done for a specific hardware configuration, the energy model would be valid only for that hardware.

To work around this problem an Abstract Instruction Set (AIS) is used that is later tied to a wide range of microprocessors. In a case study, examples with two microprocessors were given – PowerPC and ARM. Interconnects and dedicated hardware components were not considered at the time of writing of that paper. The microprocessor energy model included information for a dynamic pipeline, read/write accesses to memory, statistics about ALU/register file/cache accesses during application events, leakage power, buses (represented as a set of wires and their respective capacitances), clock and respective buffers (using

analytical methods). The main memory was not modeled. Due to the bigger set of parameters used, we could say that this is fine-grained, high-level energy model of a microprocessor. The model results yielded 5 – 10% simulation error on average to the simulator Wattch. Wattch is an ISS simulator and is at lower level of abstraction which means – it would take greater time for simulation.

In [4] a pipeline-aware ISS is developed for VLIW processors. Due to the specific type of core the following parameters were taken into account: instruction ordering, pipeline stall probability, instruction cache miss probability, power consumption of different stages of the pipeline. It is an analytical model that makes the link between the instructions and the different functional units of the core. The simulator that uses this model is cycle-accurate and gives average error of 4.8% compared to a gate-level simulator.

In [5] an interesting case of a temperature model of a microprocessor is presented. The model is analytical. It is implemented with a C structure. The structure contains parameters such as object geometry, material properties, boundary conditions, etc. The model is used to measure the different effects of the temperature over time in a microprocessor package. To do this, space and time convolutions of plane-generated temperature and power density are used with the help of FFT to speed up the computations. As a conclusion, the temperature consists of two components – total power dissipation of the chip and power density at certain point. The power dissipation on the other hand depends on leakage currents and user-program activity.

A very good guide to designers of microprocessor models is [6]. All of the parameters that could add up to the energy consumption are explained: analytical power models (that include formulae for sub-threshold and gate leakage), thermal models (steady-state, dynamic and adaptive analysis), leakage-temperature model, reliability models (including the effects of electromigration, thermal cycling, time-dependent dielectric breakdown, stress migration), system frequency models and chip technology variation.

A pipelined microprocessor model is described in [7]. In it the core is separated in two parts – datapath and control path. The datapath is represented with high-level, multi-bit modules and buses which yields fast simulation times. The control path is modeled with behavioral HDL code. According to [7] the pipeline must take into account effects such as stalling, squashing and bypassing.

In [8] an entire modeling and validation environment, called MMV, is proposed. The microprocessor (and even the whole system) is

modeled using UML diagrams. On architectural level the processor consists of core (described in RTL), main memory and ISA. The ISA model includes information about each instruction. The fields that describe it are: group (or instruction category that could be data transfer, arithmetic, etc), if-else-block (describes a conditional behavior of the instruction), operands (to hold the data to be processed) and addressing mode type of the operands.

3. Energy measurement setup

The energy model proposed in this paper is far less complex than the models described in the previous section. No analytical description of any part of the system is needed. The microprocessor is considered a black box. The energy consumed by each instruction is measured as the average current I_{dd} , multiplied by the time for its execution, multiplied by the voltage V_{dd} (which is held constant in our case). I_{dd} is measured with a shunt resistor and differential amplifier that will introduce error since there will be a voltage decrease of the microcontroller supply. The time window is measured with the help of a free GPIO pin that is toggled before and after the execution of the instruction. The microprocessor of interest is ARM Cortex-M4 implemented on a Texas Instruments' Tiva TM4C123G deeply-embedded microcontroller. A development board with the needed setup is available from TI and is shown in Fig. 1. The compiler used is arm-none-eabi-gcc.

There are, however, some considerations that are worth noting. First of all, the energy consumed in the period $0 \div T$ is an integral of the power consumption during this period. On the other hand, with this setup, we are measuring the average current I_{dd} . This will introduce some error. True energy measurements are possible with equipment

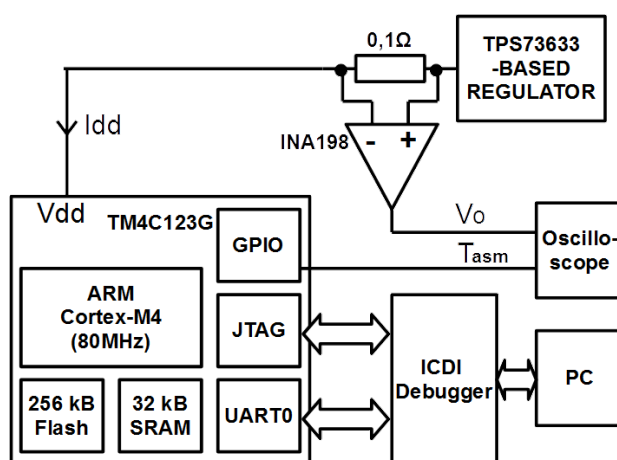


Fig. 1. Experimental Setup for ISA Energy Measurements

that is cycle-accurate which requires fast differential amplifier, fast GPIO pin and good oscilloscope. To simplify the setup, no such equipment is used.

The settle time of the output for input change from 10 mV to 100 mV of the sense amplifier INA198 is provided in the datasheet as:

$$T_{SETTLE} = 2\mu\text{s} \quad (1)$$

meaning that the microprocessor current consumption should remain unchanged for time greater than 2 μs , otherwise the I_{dd} measurements won't be correct. In our case we have chosen the code:

```
while(1){
    set_gpio(1);
    for(...){
        asm(...);
    }
    set_gpio(0);
    delay_ms(50);
}
```

for the base energy estimation, where the highlighted code should be greater than T_{SETTLE} . We have chosen this time to be equal to or greater than 100 μs which was done experimentally and yielded the number 1326 loop iterations. Adding the instruction under test with the inline assembly macro `__asm()`; will increase this time even more.

In order to reduce the overhead of the `for()` loop, the latter was substituted with code in inline assembler. The code of interest was:

```
__asm("    mov    r10, #1326 \n");
__asm("loop: subs  r10, r10, #1 \n");
//Instruction under test
asm("    add    r9, r7, r8 \n");
__asm("    subs  r10, #0 \n");
__asm("    bne   loop \n");
```

This, however, led to a problem – the time window of the tested instruction is almost always smaller or comparable with the other 4 loop instructions. Even though it is possible to make such measurements, the chosen approach is different. The instruction under test (the high-lighted line in the example above) is duplicated 20 times. This way the time window of interest is much bigger than the time window of the auxiliary code. A graph of the execution time of an ARM Cortex-M4 “add” instruction is shown in Fig. 2.

After setting up all hardware and software, the energy measurements were performed for most of the ARM Cortex-M4 ISA. Pseudo instructions were also measured because the object file of a program may contain some of them.

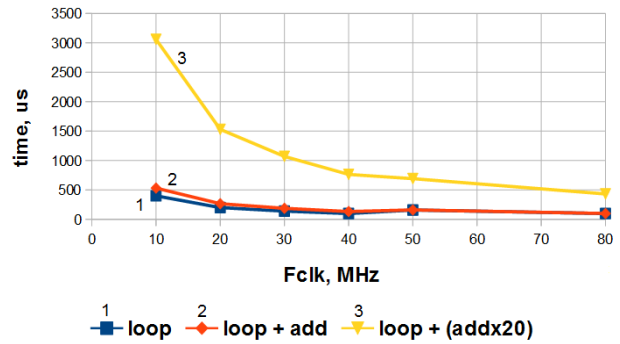


Fig. 2. Execution time of an empty loop, loop + add instruction, and loop + 20 add instructions

4. Energy model file format

The energy model file format is intended to be simple in order to help improve the proliferation of Powot Simulator in industry. It is a human-readable text file that contains a look-up table for each instruction from the target ISA. Therefore any description language is not necessary – there is no need for conditional and functional description. All the information is available to the simulator.

The text file can contain ‘#’ for line comments, like many other configuration files. The comment span is up to the end of line. Any line that contains ‘#’ is considered a comment, therefore this symbol should not be used for model data. Users can use as much comments as they want – this won’t affect the parsing of the file (which is done once at simulator start up).

The Powot Simulator model file is a TAB-separated file that consists of two columns – a tag and a parameter:

```
[TAG][‘\t’][PARAM]
```

Each parameter can be expanded to multiple parameters by using TAB:

```
[TAG][‘\t’][PARAM1][‘\t’][PARAM2]...
```

There are special tags that describe set of domains and therefore require a respective terminating tag. The syntax is:

```
[DOMAIN_START]
```

```
.....
```

```
[DOMAIN_END]
```

The model file begins with a tag [METRICS] that tells the simulator in what units is the data in the look-up table. If, for example, the current is in mA, time is in ms and voltage is in V, then the base energy will be in μJ .

Following is the [DIVISOR] tag. It tells the simulator to divide the energy units by a specific number. The idea is to make the model more flexible. The energy representation can be changed easily without the need to rewrite the whole model file.

Next is a domain description header that contains informative data to the user. It can be accessed through a special command line parameter passed to the simulator. It is advisable to always include such a header because the first entry of each domain is considered a default entry for the code analysis, if no specific parameters are set from the command line.

Currently the supported domains are the following. We start with address map description of the target microcontroller:

```
[ADDR][‘\t’][MEM_NAME][‘\t’]
[ADDR_START][‘\t’][ADDR_STOP]
```

where the memory name [MEM_NAME] could be FLASH, SRAM, FRAM, EEPROM, etc. Then the start and end address of this region is specified with [ADDR_START] and [ADDR_STOP]. Special function register could also be placed here but peripheral simulation, up to this date, is not possible.

Various other domains are declared with a single tag and one or more parameters following it. The supported domains are:

- [TEMPERATURES]
- [VOLTAGES]
- [FREQUENCIES]
- [OPERANDS]

where the first three are fully supported, while the operands tag can only specify the number of instruction operands and not the addressing mode.

After the declaration of the domains contained in this file, a declaration of DFS and DVS supported APIs is placed. This is very important improvement in the model – it allows us to detect parts of code with dynamic frequency-voltage scaling. The tags are named:

- [DFS_API_START] and [DFS_API_END]
- [DVS_API_START] and [DVS_API_END]

In between the tags there should be a declaration of the change in state that the API forces and the energy cost of the API itself in the following format:

- [FREQUENCY][‘\t’][TIME][‘\t’]
[ENERGY]
- [VOLTAGE][‘\t’][TIME][‘\t’] [ENERGY]

where the “time” parameter shows the latency of the API call and is included in case we want to turn the energy simulator into a performance simulator in the future.

Next the main part of the model file is included – the look-up energy table of each instruction. The format is the following:

- [INSTR] – the instruction mnemonic, for example add, nor, movs, etc
- [TYPE] – the type of instruction. Currently this is just a string of type: mem (memory access), datap (data processing), sat

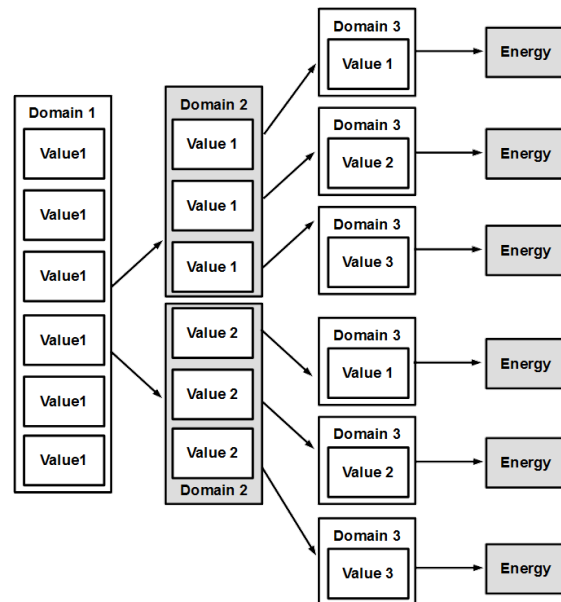


Fig. 3. Structure of the look-up energy table of an instruction

(saturation), contr (control and branch), misc (miscellaneous). In the future with the help of this tag we would be able to tell whether the code accesses peripheral and if so, will invoke separate simulation for that peripheral

- Domain data – energy data for each domain.

The energy data is entered in a tree-like manner as shown in Fig. 3. The more domains we add, the more energy measurements have to be performed. Alternatively the process will be automated in the future with the help of specialized energy profiling equipment.

5. Instruction types

Measuring different types of instructions requires augmenting the original software described in the experimental setup section. Some of the instructions simply cannot be put inside a loop. Such examples are branch and table instructions, as well as instructions that change processor state. There are instructions that were not modeled due to their dependence on external (to the microprocessor) events such as WFI and WFE that put the microprocessor in a sleep state.

The registers that are used by the instruction under test are initialized prior to entering the measurement loop. Due to the opulent number of instruction types, the initialization procedure differs.

According to [9] the Cortex-M4 instruction set can be divided in 9 groups: memory access, data processing, multiply and divide, saturating, packing and unpacking, bitfield, branch and control, floating-point and miscellaneous instructions. Not all types are presented in this paper. Only the special cases are shown here.

5.1. Branch instructions

Measuring branches like “b” and “bl” was implemented with an additional “nop” instruction at the branch address. This means that from the measured energy we have to subtract the energy of the “nop” (measured separately) and that would give us the base energy cost of the branch instruction itself. The chosen code was:

```
__asm("          bl test_loop1");
__asm("test_loop1:  nop");
__asm("          bl test_loop2");
__asm("test_loop2:  nop");
...
__asm("          bl test_loop20");
__asm("test_loop20: nop");
```

The same scenario was implemented with the table instruction “tbb”. The only difference is that the initialization must include loading of the first table label and specifying zero offset:

```
__asm("ldr r7, #Table1");
__asm("mov r8, 0x00");
```

5.2. Conditional instructions

The Cortex-M4 implements conditional instruction execution in a so-called IT block. This is done for better high-level language support. Modeling instruction blocks with look-up tables is not possible because the number of instructions inside the block, as well as their type, can vary and this would introduce too much combinations. That part of the model is not well defined – only one IT block was measured. The users will have to either provide their own measurements or just copy the available block and change the instructions. The only measured block is:

```
__asm("itte  ne");
__asm("andne r7,r7,r8");
__asm("addsne r9,r9,#1");
__asm("moveq r9,r10");
```

5.3. Stack-related instructions

An interesting case are the stack-related “push” and “pop” instructions. Since they place and retrieve data from the stack, the latter has to be increased drastically due to the chosen loop repetition number. The number 1326 means that 1326 32-bit values will be pushed on the stack, which is approximately 5 kB. For this type of microcontrollers this value is big but since no other complex code is being executed, that is no problem and the stack is resized to this value.

In order to make multiple loop measurements the “push” and “pop” instructions are used in pairs. The GPIO pin toggling is set to measure only one of the two instructions at a time. Due to memory constraints, only one instance is placed inside the loop.

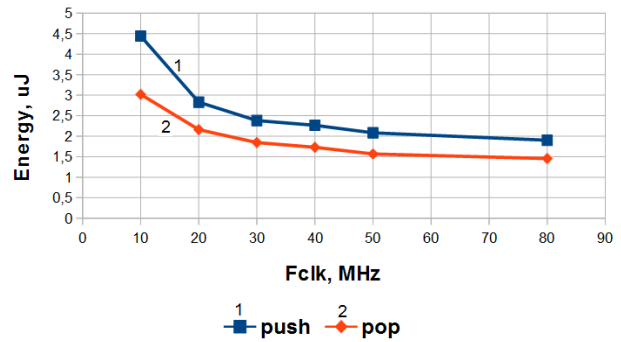


Fig. 4. Energy consumption of push and pop instructions (executed 1326 times)

The base energy cost measurements showed that there is a noticeable difference between those two instructions. This effect is shown in Fig. 4.

5.4. Supervisor call instruction

The “svc” instruction is used with embedded OS whenever there is need to request privileged operations or system resources. An effect of this instruction is that an interrupt is generated. To cope with it an empty interrupt handler “void SVCIntHandler(void) { }” was added. All of the handler’s prologue and epilogue instructions were extracted from the C code and their base energy values (previously measured) were assigned. This way we can subtract the handler’s energy overhead and find the real energy cost of the “svc” instruction.

6. Experimental results

The final energy file of a Tiva TM4C123G microcontroller consists of 116 measured instructions. A total of 6 values were considered for the frequency domain. One value is for temperature and voltage domains. Most of the instructions have one value for the memory domain but some selected ones were measured with two values – FLASH and SRAM. Six DFS APIs were included.

Fig. 5 presents frequency domain data of 25 Cortex-M4 instructions. Most of the instructions exhibit same energy curves but some, like the “isb”, have nonlinearities between 40 and 50 MHz. Those effects are due to the presence of Flash buffers that turn on at system clock of 40 MHz or higher. Such details can aid energy optimizations of code.

7. Conclusion

With the help of the ISS energy model simulations with relative error of 5 % is achievable. However absolute error of 20 – 30 % and more was measured due to the simple and basic method used for the measurements. Future research area would be to automate the process and make cycle-accurate

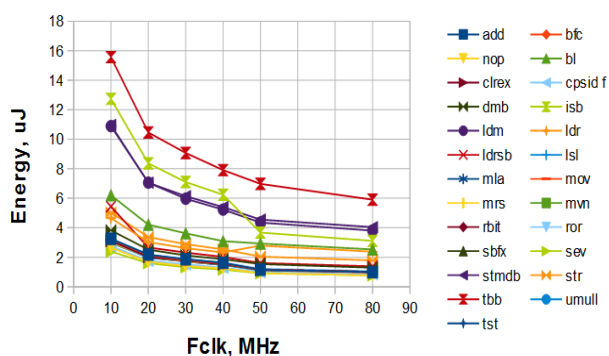


Fig. 5. Energy consumption of 25 ARM Cortex-M4 instructions (executed 1326 times from flash, on a TM4C123G)

energy measurements that should increase accuracy and speed-up model development.

REFERENCES

1. Bogdanov, L. V. and Ivanov, R. M. (2013). Microprocessor Power Modeling and Simulator Implementation, *IX Национална Студентска Научно-техническа Конференция: Сборник с доклади*, pp. 67 – 73, Technical University – Sofia, Sofia.
2. Young-Hwan, P., Pasricha, S., Kurdahi, F., Dutt, N. (2011). A Multi-Granularity Power Modeling Methodology for Embedded Processors, *IEEE Transactions on very large scale integration (VLSI) systems, Vol. 19*, pp. 668 – 681, *IEEE Xplore*, 2011.
3. Stralen, P. and Pimentel, A. (2010). A High-level Microprocessor Power Modeling Technique Based on Event Signatures. *Journal of Signal Processing Systems*, Volume 60, pp. 239 - 250.
4. Zaccaria, V., and Other(2004). *Power Estimation and Optimization Methodologies for VLIW-based Embedded Systems*, Chapter 4. Kluwer Academic Publishers, Dordrecht.
5. Michaud, P. and Sazeides, Y. (2007). ATMI: Analytical Model of Temperature in Microprocessors. *Proc. MoBS*, pp. 1 – 10.
6. Brooks, D., Dick, R., Joseph, R. Shang, L (2007). *Power, Thermal, and Reliability Modeling in Nanometer-Scale Microprocessors*, Volume 27, Issue 3, pp. 49 – 62, IEEE, Micro, IEEE.
7. Campenhout, D. (1999). *Functional Design Verification for Microprocessors by Error Modeling*, Chapter 4, pp. 83 – 89. Dissertation, The University of Michigan.
8. Sreekumar, V. K., and Other(2007). Model Based Test Generation for Microprocessor Architecture Validation. *VLSI Design*, pp. 465 – 472.
9. *Cortex-M4: Generic User Guide*, ARM Limited, (2010).

Contacts:

Address: Boul. St. Kliment Ohridski №8,
 Technical University – Sofia, Rectorate,
 lab. 1362, Sofia, Bulgaria
 Phone: +359 2 9653362
 E-mail: lbogdanov@tu-sofia.bg



BIG DATA IN BIOINFORMATICS

DOBRINKA PETROVA

Abstract: *The paper is primarily dedicated to big data challenge in bioinformatics. The term 'big data' is defined, considering its characteristics, problems that arise and general solutions for them. The growth of big biological data is discussed in details, providing digits and outlining trends. Big data problems in bioinformatics are discussed, as well as solutions for them, concerning storage, processing, analysis and knowledge extraction.*

Key words: *bioinformatics, big data, biological databases, high performance computing*

1. Big data

1.1. Definitions and characteristics

According to a formal definition, proposed soon in [1], big data represents the information assets characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value. This statement is based upon some widely accepted former explanations of big data, described in terms of its characteristics and the challenges it presents. In 2001, Doug Laney defined data growth and management challenges as being three-dimensional, addressing volume, velocity and variety of data [2]. Volume refers to the size of the data. Velocity addresses the speed at which data can be produced, transferred and analyzed. It also refers to the rate of change of data. Variety is determined by the range of data types and sources. This concept, known as the "3 V's" has been updated [3] and extended to 4 or even 5 V's [4], adding Value and Veracity to big data characteristics. Including the last 'V' – veracity, introduces the question of trust and uncertainty of data and the outcome of its analysis.

No matter which definition of the term "big data" is used, it generally refers to data that is too big or too complex to process on a single machine [5]. Living in an age where data is growing orders of magnitude faster than ever before, such "too big" and "too complex" data does not represent just an abstract problem. According to International Data Corporation, the digital universe — the data we create and manipulate, is doubling in size every two years, and by 2020 will reach 44 zettabytes, or 44 trillion Gbytes [6]. The digital universe is diverse — created by the millions of digital cameras, sensors and devices connected over the Internet, as well as the Large Hadron Collider (LHC) at CERN [7] for example. The volume of data produced at the LHC presents a considerable processing challenge –

approximately 600 million times per second particles collide within the LHC, which equals to 30 Petabytes (PB) of data for a year.

1.2. Problems, initiatives and solutions

The primary goal while collecting, storing, processing and analyzing big data is to extract significant information from it. According to Cox and Ellsworth [8], two distinct problems may arise here based on big data collections or big data objects. Big data collections are aggregates of many data sets that are individually manageable, but as a group are too large to fit on disk. Typically they are multisource, often in disparate formats, distributed among multiple physical sites, and stored in different types of data repositories. Big data objects are single source objects that are too large to be processed by standard algorithms and software on available hardware.

Aiming to make the most of the fast-growing volume of digital data and to solve the problems appeared, USA Administration announced \$200 million in a "Big Data Research and Development Initiative" [9] in 2012. In the same year Research Councils UK presented their strategic framework "Investing for Growth: Capital Infrastructure for the 21st Century" [10]. Vital part of it is the innovation "Big Data and Energy Efficient Computing" with a total sum of £189 million to invest. Universities [11], [12], [13] and associations [14] also launched their research initiatives focused on big data problem.

Diverse solutions for storing, processing and analyzing big data are proposed as a result from the initiatives and investigations. Cloud computing as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources [15] is a promising solution for both storage and analysis of

big data. There are proposals – such as Hadoop Distributed File System (HDFS) [16], which concern the storage and management of huge amount of information at the file system level. HDFS is a Java-based file system that provides scalable and reliable data storage of large files across multiple machines. A solution at a higher level for storage and management of unstructured big data are NoSQL (non-relational) databases. They are preferred in some cases, aiming to overcome the limitations of relational databases.

Massive parallelism, both in storage and computation, is searched and investigated to solve the problems, associated with big data. Proposed by Google, MapReduce [17] is a programming model for processing and generating large data sets. Programs written in this functional style are automatically parallelized and executed on a large cluster of commodity machines. HDFS along with MapReduce are modules in Apache Hadoop – open source framework for distributed processing of large data sets across clusters of computers. Apache Spark [18] is another open source cluster computing framework for large-scale data processing, based on DAG (Direct acyclic graph) execution engine.

Along with open source solutions, some of the major vendors as Oracle [19], IBM [20] and Microsoft [21] propose business-oriented solutions for big data storage and manipulation.

Machine learning, as well as cloud computing, takes place in big data analysis. Supervised, unsupervised, and hybrid machine learning approaches are widely used tools for descriptive and predictive analysis of big data [22]. Systems for distributed machine learning are proposed, such as PaDDMAS [23], MLbase [24] and GraphLab [25] – implemented on the cloud.

Maybe one of the best strategies for management of big data is used at CERN. The huge amount of data, produced by the LHC, is reduced with more than 99% in a two-stage process before the real storing and analysis to take place. Then, the data is recorded and processed onto the world's largest computing grid, whose mission is to provide computing resources to store, distribute and analyse the data generated by the LHC.

Next sections of the paper are dedicated to big data challenge in the field of bioinformatics. Section 2.1 provides background on the problems that arise; while in section 2.2 existing solutions are examined and discussed.

2. Big data in bioinformatics

2.1. Bio data in digits and trends

As a result from many projects and initiatives in bioinformatics, the volume of biological data is increasing extremely fast – it is

doubling every 9 months [26]. According to the 2016 Database Issue of Nucleic Acids Research [27] Molecular Biology Database Collection consists of 1685 databases.

European Bioinformatics Institute (EMBL-EBI) [26] maintains the world's most comprehensive range of freely available, up-to-date biological databases – from raw DNA sequences to curated proteins, chemicals, structures, systems, pathways, ontologies, etc. Total disc storage at EMBL-EBI is 75 PB as of December 2015, where the actual total volume of all data resources is about 30%. The largest part of data, accumulated at EMBL-EBI are sequences. European Nucleotide Archive [28], respectively, is the resource at EMBL-EBI with the largest storage capacity, required to store the collection of more than 724 million DNA and RNA sequences. 54,3% of them are assembled sequences from Whole Genome Shotgun projects (WGS) [29] and more than 25% arrive from the partners in the International Nucleotide Sequence Database Collaboration [30] – GenBank [31], the European Molecular Biology Laboratory [32] and DNA DataBank of Japan [33] which exchange data on a daily basis.

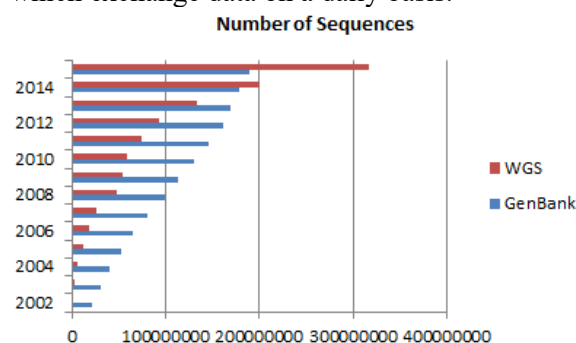


Fig. 1. Number of sequences in GenBank and WGS

Fig. 1 and fig. 2 show the rates, at which entries of WGS and GenBank (an annotated collection of all publicly available DNA sequences) are submitted since 2002 till present [34]. The number of sequences (fig. 1) from WGS is increasing with an exponential rate and reaches 317.1 million at the end of 2015, which comprises 58,3% growth per last year.

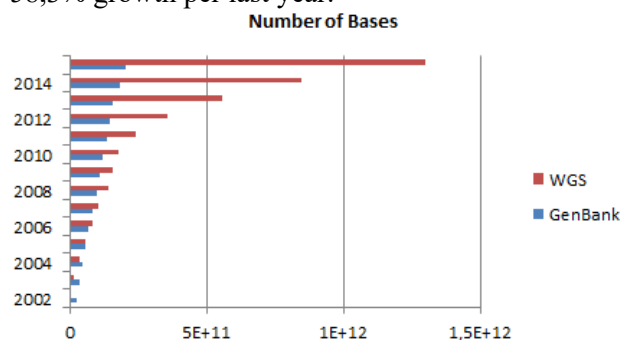


Fig. 2. Number of bases in GenBank and WGS

The number of bases (fig. 2) from WGS follows the same dependences – it is increased with more than 53% for 2015. The growth of GenBank for the last year is 10.3% for bases and 5.5% for sequences, corresponding to 204 billion bases and more than 190 million sequences, stored in this data repository. According to [34], from 1982 till present, the number of bases in GenBank has doubled approximately every 18 months. In table 1, along with GenBank, are presented selected databases – representatives in five categories with publicly available statistics. The number and type of stored entries are given for each database, and growth per last year is calculated.

Table 1. Biological databases statistics

Database name	Entries	Growth
Nucleotide Sequence Database		
GenBank [31]	204 billion bases	10.3 %
	190 million sequences	5.5 %
RNA sequence database		
RNAcentral [35]	9321083 sequences	14 %
Protein sequence databases		
UniProtKB [36]		
• TrEMBL	59718159 sequences	10% for last month
• Swiss-Prot	550299 sequences	
Protein structure database		
PDB [37]	115918 structures	8.8%
Microarray data and other Gene Expression databases		
• Gene Expression Omnibus [38]	1727633 samples	26.5%
• ArrayExpress [39]	1911539 assays	18%

The most significant growth of deposited entries can be seen for UniProtKB/TrEMBL – for the first month of 2016 year it is 10% and the number of protein sequences reaches almost 60 million [36]. The increasing rate of GEO [38] is

also noticeable – 26.5% for the last year and more than 1.7 million samples from 65500 series and 15443 platforms. The other microarray database, mentioned in table 1 – ArrayExpress [39] achieves 18% growth with 63530 experiments up to date and 40.34 TB of archived data.

According to [40] the cost of generating data is falling and it enables researches and corresponding projects to produce more data. So, one of the major bottlenecks in life science research today is not generation, but storage of biological big data sets. Solutions are also required for transforming big data stored in heterogeneous and different-in-nature sources into a structured well-interpretable format for further manipulation. Applications are needed that will process in parallel large and often distributed data sets for target data analysis and knowledge extraction.

2.2. Solutions for big data problem

General solutions for ‘big data’ problem are applicable in the field of bioinformatics too. In this section are briefly discussed bioinformatics applications and projects and the approaches, which they use to effectively store, process and analyze big biological data.

2.2.1. Storing big data in bioinformatics

Cloud computing is one of the most popular solutions, which promises to address not only storage, but also acquisition and analysis issues of big biological data. By storing on the cloud large data sets, which previously required hours to locate, download and analyze, they become quickly and easily accessible (with their last updates) to a wide range of devices connected over the Internet and can be seamlessly integrated into cloud-based applications for data analysis. The world's largest set of data on human genetic variation, produced by the international 1000 Genomes Project [41], is publicly available on the Amazon Web Services (AWS) cloud [42] since 2012. AWS provides a centralized repository of many other public biological data, including archives of GenBank, Ensembl project [43], Unigene [44], etc.

Another approach to manage the storing of huge amount of data, growing faster than the storage available is to compress the data. There is a variety of standard data compression algorithms, but they are not efficient enough in compressing biological data. In 2011, the Pistoia Alliance [45], supported by AWS, initiated a competition to award US\$15,000 to the developer of the best novel open-source compression algorithm for Next Generation Sequencing (NGS) data. The winner proposed a cluster of algorithms [46] that all delivered high performance in terms of compression ratio,

compress and decompress time. In [47] for large-scale compression of genomic sequences the Burrows–Wheeler transform is applied, in [48] the compression is based on cascading Bloom Filters.

The European Bioinformatics Institute uses CRAM [49] as a compression technique to handle nucleotide data on a very large scale. But, to address the challenge of storing exponentially growing biological data in a storage capacity, growing in linear fashion, at the beginning of 2016, EMBL-EBI defines the task for developing novel compression methods as an important part of institute’s work [50].

2.2.2. Processing and analysis of big data in bioinformatics

Big biological data are challenge not only for storing, but also for processing and analyzing in order to turn them into real knowledge. Various high-performance computing techniques are applied in dealing with the huge amount of data. Examples include cluster computing ([51], [52]), grid [53] and on-chip supercomputing [54]. Cloud computing, often in conjunction with Hadoop and MapReduce, is also a widely used solution for big data analysis in bioinformatics. There are a variety of software tools in bioinformatics, developed as Software as a service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS).

SaaS is a software model, in which the capability provided to the consumer is to use the provider’s applications running on a cloud infrastructure. In this way, SaaS eliminates the need for local installation and facilitates remote access to cloud-based software services. Among the examples of SaaS are: Hadoop implementation of BLAST – CloudBLAST [55], scalable software pipeline for whole genome resequencing analysis – Crossbow [56], cloud computing tool for calculating gene expression in large RNA-seq datasets – Myrna [57] and highly sensitive read mapping with MapReduce – CloudBurst [58].

PaaS provides the capability for users to deploy applications onto the cloud infrastructure, where computer resources scale automatically and dynamically to match application demand. Eoulsan [59] is such a framework (based on Hadoop) for high throughput sequencing analyses. BioPig [60] is also Hadoop-based framework for genomic data analysis. Galaxy Cloud [61] is a cloud-based Galaxy for large scale data analyses. CloudGene [62] is a platform for MapReduce programs on private and public clouds.

IaaS as a model offers a full computer infrastructure – physical or virtual machines, and provisions processing, storage, networks, and other fundamental computing resources. Cloud BioLinux

[63] is a publicly accessible virtual machine that enables scientists to quickly provision on-demand infrastructures for high-performance bioinformatics computing. Cloud Virtual Resource (CloVR) [64] is also implemented as a single portable virtual machine. It provides several automated analysis pipelines for microbial genomics, whole genome and metagenome sequence analysis. Bionimbus [65] is a cloud-based infrastructure for managing, analyzing and sharing genomics datasets.

The Embassy Cloud [50], launched in 2013, is an IaaS that enables groups to work in private, secure, virtual-machine-based workspaces hosted within EMBL-EBI’s data centres. Virtual machines and analysis pipelines are accessible from anywhere with an Internet connection for external groups that are collaborating with EMBL-EBI teams, but the aim is to expand service to be widely available.

Along with analysis, supported by HPC techniques, Semantic Web technologies and ontologies are applied in order to extract real knowledge from diverse and heterogeneous big biological data. The development of metadata for biological information on the basis of Semantic Web standards is regarded as a promising approach for a semantic-based integration of biological information [66]. Ontologies, on the other hand, organize the information in a hierarchical structure, thus dealing with its heterogeneity and integration [67], and providing suitable representation of biological data for effective quering and analysis.

3. Conclusion

As a result from many projects and initiatives, and due to the falling cost of the experiments, biological data have grown in a steady exponential fashion, thus defining the big data challenge in bioinformatics. The huge amounts of biological data – heterogeneous and often distributed, bring up the questions of their storage, processing, analysis and knowledge extraction.

Compression is one of the approaches, used to resolve the important issue of the storage of data on a very large scale. Cloud computing is a popular solution, which addresses not only storage, but also acquisition and analysis issues of distributed big biological data. Other high-performance computing techniques – such as cluster computing, grid and on-chip supercomputing are also applied in dealing with processing and analysis of biological data. Semantic Web technologies and ontologies are used for knowledge extraction, addressing the diversity and heterogeneity of big biological data.

Considering the rapid changes in biological field due to technological advances, which produce more data in diverse formats, faster than before, the problems of storing, processing and analyzing them

are still important and awaiting their diverse solutions.

REFERENCES

1. De Mauro, A., Greco, M. and Grimaldi, M. (2014). What is Big Data? A Consensual Definition and a Review of Key Research Topics. *AIP Conference Proceedings*, 1644, p. 97-104.
2. Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*.
3. Beyer, M. A. and Laney, D. (2012). The Importance of 'Big Data': A Definition. *Gartner*.
4. Demchenko, Y., Grosso, P., de Laat, C. and Membrey, P. (2013). Addressing big data issues in scientific data infrastructure. In: *2013 International Conference on Collaboration Technologies and Systems (CTS)*, San Diego, IEEE, p 48–55.
5. Landset, S., Khoshgoftaar, T. M., Richter, A. N. and Hasanin T. (2015). A survey of open source tools for machine learning with big data in the Hadoop ecosystem. *Journal of Big Data*.
6. Turner, V., Reinsel, D., Gantz, J. F. and Minton, S. (2014). The Digital University of Opportunities: Rich Data and the Increasing Value of the Internet of Things. *IDC White Paper*, <http://idcdocserv.com/1678>
7. <http://home.cern/topics/large-hadron-collider>
8. Cox, M. and Ellsworth, D. (1997) Managing big data for scientific visualization. *ACM Siggraph '97 course #4 exploring gigabyte datasets in real-time: algorithms, data management, and time-critical design*.
9. https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf
10. <http://www.rcuk.ac.uk/research/infrastructure/big-data/>
11. <http://www.bigdata.cam.ac.uk/>
12. <http://bigdata.csail.mit.edu/>
13. <http://vcresearch.berkeley.edu/>
14. <http://bigdata.ieee.org/>
15. Mell, P. and Grance, T. (2011). The NIST Definition of Cloud Computing. *National Institute of Standard and Technology, U. S. Department of Commerce*.
16. HDFS - https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html
17. Dean, J., and Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Commun. ACM* 51, 1, 107-113.
18. Apache Spark - <http://spark.apache.org/>
19. Oracle: Big Data for the Enterprise. (2013).
20. <http://www03.ibm.com/software/products/en/category/bigdata>
21. <http://enterprise.microsoft.com/en-us/>
22. Kashyap, H., Ahmed, H.A., Hoque, N., Roy, S., and Bhattacharyya, D.K. (2014). Big Data Analytics in bioinformatics: A Machine Learning Perspective. [arXiv:1506.05101v1](https://arxiv.org/abs/1506.05101v1)
23. Rana, O., Walker, D., Li, M., Lynden, S., and Ward, M. (2000). PaDDMAS: parallel and distributed data mining application suite. *Parallel and Distributed Processing Symposium, 2000. IPDPS 2000. Proceedings. 14th International*, Cancun, 2000, 387-392.
24. Kraska, T., Talwalkar, A., Duchi, J. C., Griffith, R., Franklin, M. J., and Jordan, M. I. (2013). MLbase: A distributed machine-learning system. *CIDR 2013*.
25. Low, Y., Bickson, D., Gonzalez, J., Guestrin, C., Kyrola, A., and Hellerstein, J.M. (2012) Distributed GraphLab: a framework for machine learning and data mining in the cloud. *Proc. VLDB Endow.* 5, 8, 716-727.
26. <http://www.ebi.ac.uk>
27. Rigden, D. J., Fernandez-Suarez, X. M. and Galperin, M. Y. (2015). The 2016 database issue of Nucleic Acids Research and an updated molecular biology database collection. *Nucleic Acids Research*, 44 (D1): D1-D6.
28. Gibson,R., Alako,B., Amid,C., Cerdeno-Trarraga,A., Cleland,I., Goodgame,N., ten Hoopen,P., Jayathilaka,S., Kay,S., Leinonen,R. et al. (2016). Biocuration of functional annotation at the European Nucleotide Archive. *Nucleic Acids Research*, 44.
29. <http://www.ncbi.nlm.nih.gov/genbank/wgs>
30. INSDC - <http://www.insdc.org/>
31. Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Sayers, E.W. (2013). GenBank. *Nucleic Acids Research*, 41(D1):D36-42.
32. <http://www.embl.org/>
33. <http://www.ddbj.nig.ac.jp/>
34. <http://www.ncbi.nlm.nih.gov/genbank/statistics>
35. <http://rnacentral.org/about-us#growth-chart>
36. <http://www.ebi.ac.uk/uniprot/TrEMBLstats>
37. <http://www.rcsb.org/pdb>
38. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S. and Soboleva A. (2013). NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Research*, 41: D991-5.
39. Kolesnikov, N. et al. (2015). ArrayExpress update-simplifying data submissions. *Nucleic Acids Research*, doi:10.1093/nar/gku1057

40. <http://www.ebi.ac.uk/about/brochures>
41. McVean et al. (2012). An integrated map of genetic variation from 1092 human genomes. *Nature* 491, pp.56-65.
42. <http://aws.amazon.com/public-data-sets/>
43. Flicek, P., et al. (2014). Ensembl 2014, *Nucleic Acids Research* 42 Database issue: D749-D755.
44. Pontius, J.U., Wagner, L., and Schuler, G.D. (2003). UniGene: a unified view of the transcriptome. In: *The NCBI Handbook*. Bethesda (MD): NCBI.
45. <http://www.pistoiaalliance.org/projects/sequence-squeeze/>
46. Bonfield, J.K., and Mahoney, M.V. (2013). Compression of FASTQ and SAM Format Sequencing Data. *PLoS ONE* 8(3): e59190.
47. Cox, A.J., Bauer, M.J., Jakobi, T., and Rosone, G. (2012). Large-scale compression of genomic sequence databases with the Burrows-Wheeler transform. *Bioinformatics*, 28(11):1-6.
48. Rozov, R., Shamir, R., and Halperin, E. (2014). Fast lossless compression via cascading Bloom filters. *BMC Bioinformatics* 15(Suppl 9):S7
49. Hsi-Yang Fritz, M., Leinonen, R., Cochrane, G. and Birney, E. (2011) Efficient storage of high throughput DNA sequencing data using reference-based compression. *Genome Research*, 21, 734–740.
50. Cook, C. E., Bergman, M. T., Finn, R. D., Cochrane, G., Birney, E., and Apweiler, R. (2016). The European Bioinformatics Institute in 2016: Data growth and integration. *Nucleic Acids Research* 44 (D1): D20-D26.
51. BioHPC - http://biohpc.org/info_arch.aspx
52. Elixir - <http://www.elixir-europe.org/>
53. Kertesz, A., Otvos, F., Kacsuk, P. (2013) A Case Study for Biochemical Application Porting in European Grids. *Concurrency and Computation: Practice and Experience*, 26(10):1730–1743.
54. http://www.nvidia.com/object/bio_info_life_sciences.html
55. Matsunaga, A., Tsugawa, M. and Fortes, J. (2008). CloudBLAST: Combining MapReduce and Virtualization on Distributed Resources for Bioinformatics Applications. In *eScience, 2008, IEEE Fourth International Conference on eScience*, pp.222-229.
56. Langmead, B., Schatz, M.C., Lin, J., Pop, M., Salzberg, S.L. Searching for SNPs with cloud computing. *Genome Biol* 10:R134
57. Langmead, B., Hansen, K., and Leek, J. Cloud-scale RNA-sequencing differential expression analysis with Myrna. *Genome Biology* 11:R83.
58. Michael C. Schatz. (2009). CloudBurst: highly sensitive read mapping with MapReduce. *Bioinformatics*, 25 (11): 1363-1369.
59. Jourdren, L., Bernard, M., Dillies M.A., and Le Crom, S. (2012). Eoulsan: A Cloud Computing-Based Framework Facilitating High Throughput Sequencing Analyses. *Bioinformatics* 28 (11): 1542-1543.
60. Nordberg, H., Bhatia, K., Wang, K., and Wang, Z. (2013). BioPig: a Hadoop-based analytic toolkit for large-scale sequence data. *Bioinformatics* 29(23):3014–3019.
61. Afgan, E., Baker, D., Coraor, N., Goto, H., Paul, I. M., Makova, K. D., Nekrutenko, A., and Taylor, J. (2011). Harnessing cloud-computing for biomedical research with Galaxy Cloud. *Nature Biotechnology*, 29(11), 972–974.
62. Schonherr, S., Forer, L., Weißensteiner, H., Kronenberg, F., Specht, G., and Kloss-Brandstätter, A. (2012). Cloudgene: a graphical execution platform for MapReduce programs on private and public clouds. *BMC Bioinformatics* 13:200.
63. Krampis, K, Booth, T., Chapman, B., Tiwari, B., Bicak, M., Field, D., and Nelson, K. (2012). Cloud BioLinux: pre-configured and on-demand bioinformatics computing for the genomics community. *BMC Bioinformatics* 13:42.
64. Angiuoli, S.V., Matalaka, M., Gussman, A., Galens, K., Vangala, M., Riley, D.R., Arze, C., White, J.R., White, O., and Fricke, W.F. (2011). CloVR: a virtual machine for automated and portable sequence analysis from the desktop using cloud computing. *BMC Bioinformatics*, 12:356.
65. Heath, A. P., Greenway, M., Powell, R., Spring, J., Suarez, R., Hanley, D., and Grossman, R. L. (2014). Bionimbus: a cloud for managing, analyzing and sharing large genomics datasets. *Journal of the American Medical Informatics Association: JAMIA*, 21(6), 969–975.
66. Cannata, N., Schröder, M., Marangoni, R., and Romano, P. (2008). A Semantic Web for bioinformatics: goals, tools, systems, applications. *BMC Bioinformatics*, 9 (4):S1.
67. Blake, A., Bult, C. (2006). Beyond the data deluge: data integration and bio-ontologies. *Journal of Biomedical Informatics*, 39 (3), 314–320.

Authors' contacts:

Dobrinka Petrova

63 Sankt Peterburg Blvd, Plovdiv, Bulgaria

Phone: 032/ 659 727

E-mail: dpetrova@tu-plovdiv.bg



COMMUNICATION SYSTEM FOR LAN SERCHING

HRISTO VALCHANOV

Abstract: *The resource sharing in local networks allows efficient use of information. The built in Windows 7 search system does not have possibility to search within the local network. This paper presents a communication system for distributed indexing and searching in a local Windows network. The indexing process can be performed both in operational memory and in disk storage. Access to the system's functionality is organized on the basis of user groups. The experimental studies and comparative analysis are presented.*

Key words: *Windows Search, Distributed Search, Indexed Search*

1. Introduction

One of the advantages of local networks is the sharing of resources between users. The development of computer technologies and systems creates the need for development of systems for searching and information retrieval [1].

The Windows operating system is widely used in the world. One of its service, which is a system component after Windows Vista, is an indexing service (Windows Search Indexer). It speeds up the built-in file search system on the local machine. Despite the improved functionality this system has a number of disadvantages. One major limitation is that it does not support indexing and searching in a local network of Windows-based computers. Partially possibility is provided through the use of obvious configured by the user shared resources. For indexing them, however, they must be previously turned in offline mode, which excludes them from sharing mechanism during this period. Excluding the sharing service from a single user (Network Discovery and File and Printer Sharing) eliminates the ability other users to search for the previously shared resources.

There are a number of free alternative tools of indexing and searching in Windows. Some tools, like grepWin, MasterSeeker, FileSeek and Listry [2], despite their enhanced functionality, are intended solely for desktop search. There are also tools that provide indexing and search in shared drives on a network. Accelerated indexing and search is provided by Locate32 [2]. The disadvantage of the software is the lack of search in the contents of the file. LanHunt [3] is a Java-based application using the database index, which speeds up the search process, but the search is provided

only by single words. Other free and flexible applications are LanSearch Pro [3] and Everything [4]. Despite their rich functionality, the searching is limited to only the name or file type. A characteristic of search LAN applications is that they are based entirely on the sharing system of Windows. Excluding the sharing service limits their functionality only to a desktop indexing and search.

This paper presents the architecture of the communications subsystem of a distributed environment for indexing and searching in a local Windows network.

2. System for indexing and searching

The system for indexing and searching has distributed architecture. It consists of multiple search engines (SE) that are installed as a service on all machines in the local network involved in the process of searching for information. Each SE operates independently of the others and indexes the information on the local machine.

On Figure 1 is shown the structure of a SE.

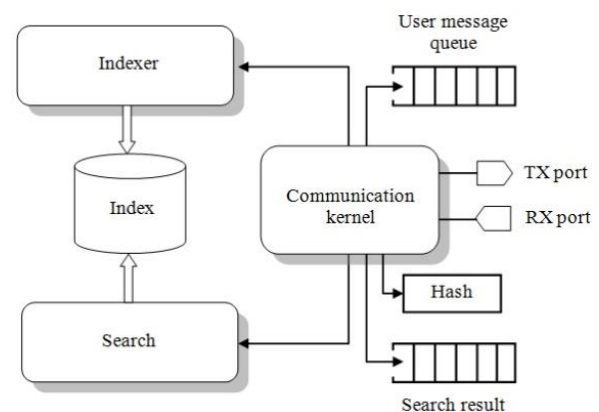


Fig.1. Structure of a SE

With indexes is accelerated the speed of searching for information in single or in multiple documents. The indexing is based on the inverted index concept [5]. It is a list (dictionary) of found terms (words) and for each word is built a list of file identifiers (postings), which contain this term. For each word is maintained its position and frequency of occurrence in each file. Thus allows searching by phrases and sorting the results.

In order to effectively use of the resources of the machine, the construction of indexes was implemented in two ways: in the RAM memory and on the disc.

When the index is building in RAM, the dictionary is implemented as a modified binary tree. Each node has subnodes for the symbols of the alphabet. The goal of this method is to accelerate the process of indexing.

The index building in a disk memory is effective on large collections. The difference lies in the use of buffer with a fixed size, which stores the posting lists. When the buffer is full, it is stored on the disk and the indexing process continues. After completing the indexing process the dictionary must be sorted. Due to its inability to load in the computer memory, an external sorting is applied. To speed up the search process after completion of the sorting, a map of sorted indexes is created. It contains all the initial letters of the dictionary and their start and end positions in the index file.

The search module processes the received queries, which can be submitted both locally or received over the network from other machines. Upon receipt of a query it is filtered and checked for the presence of logical operators, mask terms or phrases. Then each term of the query is checked for presence in the list with built indexes.

The system allows execution of several types of queries: searching by a word, by a part of a word (searching with a mask), searching by a phrase and complex logical expressions.

3. Communication subsystem

The communication subsystem ensures exchange of information between started on separate machines SE. This exchange involves the transmission of search queries to all machines and sending the resulting search results. These actions suggest a high load on the network. In terms of network performance an important issue is the choice of suitable transport protocol. One way to achieve this is to send the search query for processing to all SE using multicast delivery.

The current implementation of the proposed communication system is based on Winsock2 multicast [8]. The nature of the exchanged data (words, phrases, search results in the form of paths

to files and files themselves) requires to be delivered reliably. The integrated in Winsock2 multicast protocol for reliable delivery is Pragmatic General Multicast protocol (PGM) [9]. The PGM is a transport layer protocol for reliable multicast for applications that require ordered or disordered, without duplicates delivery of data from multiple sources to multiple recipients. The PGM ensures that a group recipient will receive any data packets or it is able to detect unrecoverable data packet loss.

The communication subsystem (Figure 1) consists of several basic components:

- Communication kernel - the core that provides basic functionality.
- User message queue - a synchronized queue of messages that are intended to send or process.
- Search result – a buffer to store search results.
- Hash – a hash buffer used in the phases of the authentication process.
- RX, TX port – PGM sockets for message receiving and sending through the network.

3.1. Users and groups

The used for communication multicast address is fixed and common to all instances of the application. In this form each user could participate in searches. In a real environment, this is insufficient. In order to divide users into subsets and limit searches to within a given subset the mechanism of the groups was introduced. Each user can be member of one or several groups.

3.2. Security of the communications

The concept of group allows flexibility in limiting the perimeter of searching, but by itself does not provide protection from reading the data which are transmitted from machines outside the group. For providing such a protection the concept for multicast encryption is used [10]. Unlike classical scheme instead of having one key for the overall multicast group, now is introduced one key for each of group, called a group key. Any user who wants to do a search, is obliged to use appropriate group key to encrypt messages sent.

Maintaining the group keys and information about membership in groups is implemented by a special centralized component - authenticator (Figure 2). It is located in the global system space – it is not a member of any group and accepts all messages in the multicast group but handles only those that are related to its role.

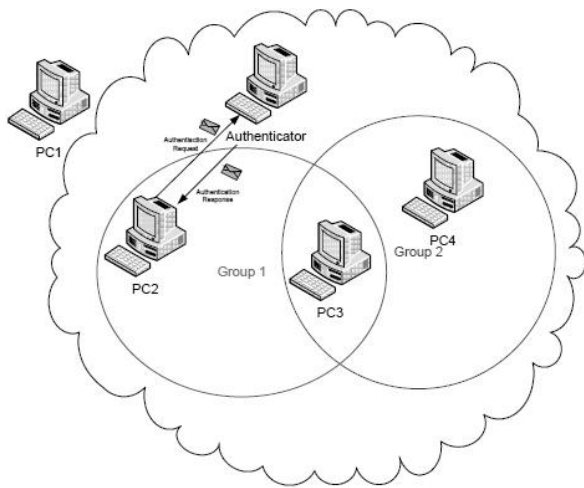


Fig. 2. Groups and authentication

3.3. The encryption system

The encryption system consists of a library that provides cryptographic functions and a module for generating the key pairs. It is implemented on the basis of Cryptography API Next Generation (CNG) [11]. In the current implementation are used integrated in the program code keys. As a future development is provided keys generation as a random sequence of symbols in the authenticator and verifying them that are unique among the current set of generated group keys.

4. Experimental study and results

The goal of the study is to explore the functionality of the proposed system in a real local network and to compare its performance with that of the built-in Windows system for indexing and searching. The experiments were carried out in a local network, including four separate clients and an authenticator machine. The computers are running Windows 7 Enterprise operating system.

Two test groups are performed - indexing and search queries.

- Indexing. Three types of indexing are tested: indexing in RAM, indexing in disc memory and Windows Indexing.

- Processing of search queries. Four query types are tested:

- Searching by a term – *term*;
- Searching by a mask – *term**;
- Searching by a phrase – *term1 term2*;
- Logical searching in the form *term1 AND term2 OR term3*.

The system load is very high for Microsoft indexing and indexing in RAM. For RAM indexing all indexes are stored in memory, leading to high load, while for Microsoft has no information how the indexes are built.

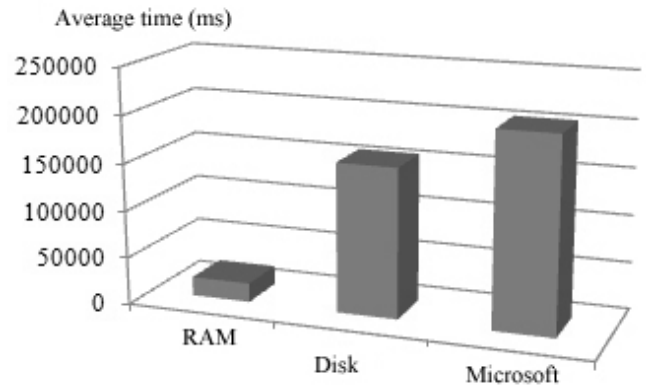


Fig.3. Average indexing time

On Figure 4 are shown the results of the processing of queries. When comparing the averages of the results it is observed that the processing of queries with a mask in the operating memory (a tree) is performed two times faster than in disk memory (by repeatedly loading the vector buffer) and the greatest time has the Microsoft search.

For searching for a single word and for logical expressions again the best results gives the search in RAM. The algorithm with the vector buffer is slower, while Microsoft implementation is two times slower than the disk search. When searching a phrase the search algorithm with tree answers twelve times faster than searching with vector buffer. The Microsoft search again has the highest average time.

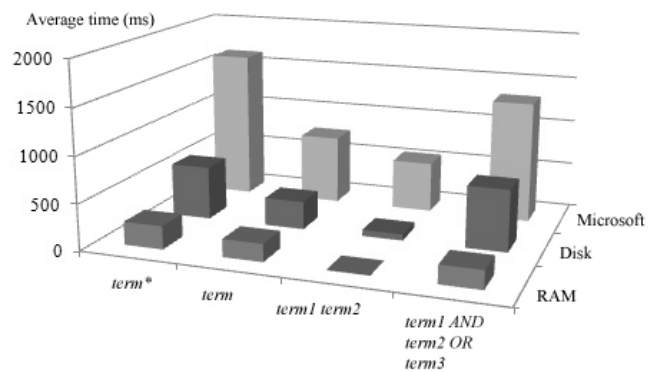


Fig. 4. Average time for query processing

5. Conclusions

This paper presents a communication subsystem of a distributed environment for indexed search in a local Windows network. The process of indexing can be performed both in operational and in disk memory. An experimental comparisons and evaluations of the two methods for indexing are performed. The proposed search system was compared with the built into the Windows 7 system

for indexing and searching. The results indicate that it has reached a performance boost from the existing implementation in Windows.

Goal of a future work is to develop a hybrid indexing algorithm and automatically to generate keys for authentication.

REFERENCES

1. Baeza R., Ribeiro B. (2011). *Modern Information Retrieval: The Concepts and Technology behind Search*. Addison-Wesley.
2. Five Alternative Search Tools for Windows. (2015). <http://www.thewindowsclub.com/windows-search-alternative-tools>.
3. Tools to Search Any Files in LAN. (2016). <https://www.raymond.cc/blog/search-find-and-locate-any-files-on-local-area-network-shared-folders/>.
4. Everything Search Engine. (2016). <http://www.voidtools.com/>.
5. Buettcher S., C. Charles. (2010). *Information Retrieval: Implementing and Evaluating Search Engines*. The MIT Press.
6. Flex Lexycal Analyzer. (2016). <http://flex.sourceforge.net>.
7. BulStem. (2016). <http://lml.bas.bg/~nakov/bulstem/>.
8. Microsoft. Windows Socket 2 Architecture. (2016). <https://msdn.microsoft.com/en-us/library/windows/desktop/ms740650%28v=vs.85%29.aspx>.
9. PGM Reliable Transport Protocol. (2016). <https://tools.ietf.org/html/rfc3208>.
10. Shoufan A. (2012). *High Performance Group Key Management*. Akademikerverlag.
11. Cryptography API Next Generation. (2016). <https://msdn.microsoft.com/>.

Authors' contacts:
 Technical University of Varna
 Varna, 1, Studentska Str.
 E-mail: hristo@tu-varna.bg



TECHNOLOGY OF COMPUTATIONS IN CANNY EDGE DETECTION FOR FPGA IMPLEMENTATION

DIMITRE KROMICHEV

Abstract: *Precise and effective, Canny edge detection is computationally complicated and difficult for a fast hardware realization. This paper is focused on the technology of computations in John Canny's algorithm aimed at a speed optimized FPGA implementation. Each computational module is scrutinized in terms of memory utilization, organization of computations, and integer arithmetic speed. Proposed are new techniques and algorithms to accomplish the goal of speeding up Canny edge detection execution with respect to FPGA characteristics and functionalities.*

Key words: *Canny edge detection, FPGA, speed, technique, algorithm*

1. Introduction

John Canny's algorithm [6] has gained the recognition of being the standard gauge for the edge detection segment of digital image processing. It is a low level filtering technology intended to be used in software. Field Programmable Gate Array (FPGA) has been increasing its market share over the last two decades. One of the basic requirements for an FPGA-based hardware implementation of Canny edge detection is speed.

Described in the literature are several approaches addressing the Canny algorithm's being difficult for a fast execution on FPGA: 1) Split Canny [10]. It uses input image fragmentation for fast Gaussian smoothing and pipelining effectiveness. In terms of the speed parameter, the design flaws are: both on-chip and external memory data transfers; a substantial quantity of repeated computations for fragment boundary pixels; recursive computations in hysteresis thresholding. 2) Distributed Canny [9]. It attempts to spare the on-chip memory at the expense of a constant throughput through performing the computations at block level. Design flaws: parallel computations speed is eroded by excessive processed image defragmentation calculations; recursive computations. 3) Area optimized Canny [8]. It is mainly concentrated on a new architecture of the non-maximum suppression and hysteresis thresholding modules. Design flaws: sequential multiplications; recursive computations. 4) Modified Canny [7]. The image is partitioned into q

sub-images and each sub-image is further divided into $p \times m \times m$ blocks. The architecture consists of q processing units. Design flaws: heavy demands on memory; recursive computations.

The objective of this paper is to present the technology of fast Canny edge detection computations in compliance with FPGA characteristics. The task is to scrutinize the Canny algorithm speed optimization capabilities proposing new computational techniques and algorithms as well as an effective utilization of the fastest FPGA memory. The targeted hardware is Altera FPGAs. Relevant to the analyses and conclusions arrived at in this paper are only gray scale images.

2. Computations in the Canny algorithm Modules

2.1. Gaussian smoothing

Gaussian weighted average filtering is a representative of linear spatial filtering

$$v(m, n) = F[u(m, n)] \quad , \quad (1)$$

where

$u(m, n)$ is the input image,
 $v(m, n)$ is the output (filtered) image,
 F is the filtering function

If F is defined as a $Z \times Z$ neighborhood operation (Z is an odd number and $Z \geq 3$), Gaussian

smoothing is:

$$v(m, n) = \frac{1}{s} \sum_{k=-w}^w \sum_{l=-w}^w g(k, l)u(m + k, n + l) \quad (2)$$

where

$u(m,n)$ is the input image,
 $v(m,n)$ is the filtered image,
 $w = \frac{Z-1}{2}$, $w = \{1,2,..\}$,
 $g(k,l)$ is the Gaussian filter ($k = Z, l = Z$),
 s is the sum of all coefficients in the mask (*normalization factor*); the coefficients are integers based on the binominal series.

2. 1. 1. Memory utilization

For the FPGA-based Canny algorithm to start being executed, an exact number of 8-bit values need to be available for smoothing in at least one of the memories on the FPGA board. On-chip memory is the fastest memory on the Altera FPGAs [1],[2],[3],[4]. Consequently, the highest frequency this memory can be read from presents the upper speed limit the execution of computations can eventually approximate to. The image matrix pixels are loaded into the memory sequentially, i.e. row by row. From (2), and taking into account the technical characteristics of the Altera FPGA on-chip memory operations, if the Gaussian filter is of size $Z \times Z$, and the image is of size $M \times N$, the exact number of available pixels required for the Gaussian smoothing module to commence the linear spatial filtering operations should be:

$$(Z-1)*N + (Z+1) \quad (3)$$

Parallelism in the execution of instructions, which, in terms of speed, is a great advantage of the FPGA-based hardware implementation over the software implementation, provides the capability of all the pixels under the Gaussian mask being accessible within a single clock cycle.

On-chip memory on Altera FPGAs is limited in quantity, and Canny edge detection is a rather complicated algorithm, requiring a serious amount of storage space. Consequently, it is quite inefficient for the whole image to be stored into memory prior to starting Gaussian filtering. The continuous execution of smoothing is guaranteed, if the memory contains only those rows of the image matrix that are currently involved in the computations. The image rows needed for the Gaussian smoothing are loaded in Dual-port RAMs. The depth of a single Dual-port RAM depends on

the restrictions imposed on the memory organization by Altera. If a Dual-port RAM's depth is equal or bigger than the number of columns in the image matrix, then the number of all the Dual-port RAMs providing a single clock access to all the pixels under the Gaussian mask of size $Z \times Z$ is:

$$(Z*Z+1)/2 \quad (4)$$

The condition expressed by (3) being satisfied, and the filtering of the pixel with image matrix coordinates $((Z-1)/2, (Z-1)/2)$ having been started, the next 8-bit value belonging to the row of highest ordinal number that is currently under the Gaussian mask is simultaneously loaded into Dual-port RAM. In this way, the next pixel is immediately available for smoothing, and consequently the Gaussian filtering computations are guaranteed to be continuously executed. All the pixels in the image matrix row of number $(Z-1)/2$ having been filtered, the smoothing continues with the first subject to being filtered pixel from the image matrix row of number $(Z+1)/2$. For the computational process to be continuous, while the pixels from the image matrix row of number $(Z-1)/2$ are being filtered the pixels belonging to image matrix row of number Z need to be sequentially stored in Dual-port RAMs whose exact number is presented by the expression

$$(Z+1)/2 \quad (5)$$

When image matrix row of number $(Z+1)/2$ starts being filtered the Dual-port RAMs containing the input image matrix row of number 0 are not relevant to the Gaussian smoothing computations any more. So, every clock cycle these pixels are sequentially overwritten by the pixel values of the input image matrix row of number $(Z*2+2)/2$. For one thing, this technique of revolving Dual-port RAMs and storing one image row through replacing another that is not usable spares the fastest memory on the Altera FPGAs, thus providing the opportunity images of different sizes to be processed by the FPGA implemented Canny edge detection. The total number of Dual-port RAMs required for the Gaussian smoothing module is:

$$(Z*Z+1)/2 + (Z+1)/2 \quad (6)$$

2. 1. 2. Organization of computations

With all the pixels under the Gaussian mask being accessible within a single clock cycle, the available built-in hard multipliers providing the fastest multiplication on the Altera FPGAs (Altera LPM_MULT function [5]) are employed for the

execution of convolution. The next step consists of each multiplication result being divided by the normalization factor. The quotients obtained are 8-bit in width which is an important fact with respect to the characteristics of computations on the Altera FPGAs. The last step is addition of the division results to calculate the final value of the Gaussian smoothed image pixel. For the purpose of execution speed optimization, the parallel adder is used (Altera PARALLEL_ADD function [5]), taking the advantage of maximum speed of arithmetic operations on the Altera FPGAs being guaranteed for input data width of 8 bits [5]. The output of the parallel adder is the filtered pixel.

2. 1. 3. Integer arithmetic speed

While addition and multiplication are the two fastest representatives of integer arithmetic on the Altera FPGAs [1],[2],[3],[4], in terms of speed, division is definitely a weak point. Using a divider by employing the Altera LPM_DIVIDE function [5] will immensely erode the speed parameter of Canny's algorithm execution. Consequently, the conventional integer division has to be replaced with a technique capable of reducing the computational time and guaranteeing the same accuracy of results. The algorithm of this technique encompasses the following sequence of steps: **1)** Defining the positive integer that serves as divisor in the conventional division. In this case, it is the normalization factor of the Gaussian filter. **2)** Experimentally, an appropriate value equal to a power of 2 is selected. This value is divided by the integer number from *step 1*). **3)** Only the integer part of the division result calculated in *step 2*) is taken. **4)** The number that will be divided is multiplied by the integer number from *step 3*). **5)** From the result calculated in *step 4*), a sequence of bits (starting with LSB and going leftwards) is dropped out. The number of the dropped out bits is equal to the number of the power of 2 selected in *step 2*). **6)** The value of MSB of the bits that are dropped out in *step 5*) is added to the value represented by the remaining bits in the multiplication result calculated in *step 4*) after the sequence of bits is dropped out in *step 5*). The positional number of this bit in the multiplication result calculated in *step 4*) is equal to: [the number of the power of 2 selected in *step 2*)] - 1. **7)** The result of addition in *step 6*) is the quotient.

The normalization factor is a filter specific constant. It has to be different from a power of 2. If it is equal to some power of 2, only the operation bit slicing (*steps 5*), *6*), and *7*) of the algorithm) is relevant to the calculations.

To apply this technique, all the coefficients in the Gaussian filter should have been multiplied by the integer number calculated in *step 3*) prior to starting the Canny edge detection computations. Thus, a Gaussian mask is modified with respect to coefficients' values, remaining the normalization factor unchanged.

Being based on two of the five fastest operations on Altera FPGAs [1],[2],[3],[4], the presented technique guarantees the utmost speed. Combined with the other two most advanced in terms of speed arithmetic operations – addition and multiplication, this technique provides for a maximum speed of Gaussian smoothing computations.

2. 2. Computation of orthogonal gradients

2. 2. 1. Memory utilization

Gaussian filtering having been applied, the total number of pixels in the input image is reduced, and the total number of filtered image pixels is :

$$M*N - \{[M - (Z-1)] * (Z-1) + N * (Z-1)\}. \quad (7)$$

The exact number of available Gaussian smoothed pixels required for this module to commence the parallel computations of the vertical and horizontal gradients applying the two Sobel filters (Fig. 1.)

-1	0	1	-1	-2	-1
-2	0	2	0	0	0
-1	0	1	1	2	1

Gx

Gy

Fig. 1. Sobel masks for x and y

should be:

$$2*(N-(Z-1)) + 4. \quad (8)$$

As with Gaussian smoothing module, the filtered fixels are stored in revolving Dual-port RAMs for the purposes of guaranteeing the access to all the pixels under the Sobel masks within a single clock cycle, thus employing pipelining of computations, and saving memory. The total number of Dual-port RAMs utilized in this module is 7.

2. 2. 2. Organization of computations

As gradient filters, the two Sobel masks have positive and negative coefficients. If in a 3x3 neighborhood $C(x,y)$ is the central pixel, and the numbers of neighboring pixels are as shown (Fig. 2.),

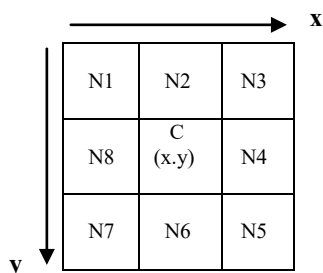


Fig. 2. 3x3 neighborhood pixels

the equations used to calculate the x gradient and the y gradient for $C(x,y)$ are:

$$G_x = \frac{(N_3 + 2N_4 + N_5) - (N_1 + 2N_8 + N_7)}{4} \quad (9)$$

$$G_y = \frac{(N_7 + 2N_6 + N_5) - (N_1 + 2N_2 + N_3)}{4} \quad (10)$$

There are two important facts here:

1) Division by the sum of positive coefficients in the Sobel masks avoids the need for scaling; 2) The difference in signs of the values calculated for both gradients is applied in the speed optimization of gradient direction computations.

2. 2. 3. Integer arithmetic speed

The multipliers and the divisors being a power of 2, integer arithmetic guarantees a maximum speed of computations in this module.

2. 3. Computation of gradient magnitude and gradient direction

2. 3. 1. Memory utilization

The orthogonal gradients having been computed, the total amount of pixels to be used in calculating the magnitude and direction of the gradient is:

$$M*N - \{[M-(Z+1)] * (Z+1) + N * (Z+1)\}. \quad (11)$$

Mathematically, gradient magnitude G_M is computed using

$$G_M = \sqrt{G_x^2 + G_y^2}, \quad (12)$$

and gradient direction G_D is computed using:

$$G_D = \tan^{-1} \left(\frac{G_y}{G_x} \right) \quad (13)$$

There is no need for memory storage of a certain number of pixels processed in the previous computational module for the computations in (12) and (13) to start being executed. The pairs of pixel values computed in (9) and (10) are immediately pipelined and processed in (12) and (13). Consequently, no Dual-port RAMs are utilized in this module.

2. 3. 2. Organization of computations

Altera provides a specialized integer square root function – ALTSQRT [5]. Although the 8-bit input data width is advantageous to the speed parameter, the maximum clock rate the function can be executed at [5] is not compatible with the accomplishment of the speed optimization goal. That is why, (12) is replaced by the approximation

$$|G_M| = |G_x| + |G_y|, \quad (14)$$

the latter being functional inasmuch as it does not distort the local maximum calculations in the computational module to follow.

The hardware implementation of (13) erodes very seriously the speed of Canny edge detection. To overcome this hurdle, (13) is replaced by computing the ratio between G_y and G_x . The ratio is calculated on the basis of two reference points and the signs of G_y and G_x . One of the reference points is used to represent an angle of 22.5° , and the other - 67.5° . The most appropriate

approximation to angle 22.5° is the fraction $\frac{2}{5}$ ($\tan^{-1} \left(\frac{2}{5} \right) = 21.8014094^\circ$), and to angle 67.5° - the fraction $\frac{5}{2}$ ($\tan^{-1} \left(\frac{5}{2} \right) = 68.1985905^\circ$). In this way, the gradient direction calculation is based on:

- 1) $G_y > 0$ & $G_x > 0$
 $G_y < 0$ & $G_x < 0$
 If $|G_x|*2 \geq |G_y|*5$ $G_D = 0$
 If $|G_x|*2 < |G_y|*5$ & $|G_x|*5 > |G_y|*2$ $G_D = 45$
 If $|G_x|*5 \leq |G_y|*2$ $G_D = 90$
- 2) $G_y > 0$ & $G_x < 0$
 $G_y < 0$ & $G_x > 0$
 If $|G_x|*2 \geq |G_y|*5$ $G_D = 0$

$$\begin{aligned} \text{If } |G_x|^2 < |G_y|^5 \ \& \ |G_x|^5 > |G_y|^2 \quad G_D = \mathbf{135} \\ \text{If } |G_x|^5 \ |G_y|^2 \quad G_D = \mathbf{90}. \end{aligned} \quad (18)$$

2.3.3. Integer arithmetic speed

Employing (14) in computing the gradient magnitude guarantees a maximum speed of execution. Although (15) is much faster than (13), the comparison functions on the Altera FPGAs cannot match addition and multiplication in terms of speed. Consequently, there is a speed incongruity between the two parallelly executed computations. It is the speed of (15) that to a great extent determines the general speed characteristics of this module. There is an input image size dependable compensation for the comparison functions being slower – the essence of FPGA computations in the next module demands that the number of values representing the magnitudes should exceed the number of values representing the directions by a whole image row. Consequently, the larger the number of columns in the input image matrix, the narrower the gap between the comparative speeds of gradient magnitude and direction calculations.

2.4. Non-maximum suppression

2.4.1. Memory utilization

Computing the local maximum being based on comparing one magnitude with two adjacent magnitudes positioned along the direction of the gradient, the exact number of calculated magnitudes required for this module to start executing is

$$2*(N-(Z+1)) + 4, \quad (16)$$

while the exact number of calculated directions is estimated to be

$$(N-(Z+1)) + 4. \quad (17)$$

They are to be stored in separate revolving Dual-port RAMs whose total number in this module is 10.

2.4.2. Organization of computations

Non-maximum suppression is based on:

$$\begin{aligned} M(x,y) &= M(x,y) \quad \text{only for} \\ [M(x,y) > M_1(x,y) \ \& \ M(x,y) > M_2(x,y)]; \\ M(x,y) &= 0 \quad \text{in all the other cases,} \end{aligned}$$

where

$M(x,y)$ is a pixel possibly belonging to the edge, $M_1(x_1,y_1)$ and $M_2(x_2,y_2)$ are the magnitudes of the two adjacent pixels positioned along the gradient direction.

2.4.3. Integer arithmetic speed

The overall speed performance in this module is influenced by two factors: 1) the rate of loading the calculated in the previous module gradient direction values in the Dual-port RAM; 2) comparison functions are the indispensable integer arithmetic here, and taking into account that the total number of their execution is a constant, it is the maximum clock rate for input data width of 8 bits that is of crucial importance.

2.5. Hysteresis thresholding

2.5.1. Memory utilization

The result of this module computations being the final edge detected image, it is required that there be available on-chip memory to store

$$M*N - \{[M - (Z+3)] * (Z+3) + N * (Z+3)\} \quad (19)$$

pixels. So far in the computational process, the appropriate pipelining techniques based on the revolving Dual-port RAMs, has left enough storage space for all these pixels to be stored into the fastest memory on Altera FPGAs.

2.5.2. Organization of computations

Each pipelined magnitude being compared to the two calculated in advance threshold values (high - $T1$, and low - $T2$), the essential concentration here is the determining of connectivity/non-connectivity with respect to $T1$. The algorithm focuses on several steps:

1) A 3x3 window centered around each pixel satisfying the condition $T1 > GM > T2$ is used. All the eight neighbouring magnitudes are tested with respect to $T1$. **2)** If there is a value bigger than $T1$, this pixel is defined as an edge and its coordinates in the image matrix are stored in a separate memory buffer to avoid redundant double check of five magnitudes when the window moves on to the next pixel. **3)** If none of the neighbouring magnitudes is bigger than $T1$, but at least one falls between $T1$ and $T2$, then only the peripheral magnitudes of a 5x5

window are checked for values bigger than $T1$. **4)** If such a value is available, then the pixel is an edge. **5)** If one or more of these sixteen peripheral magnitudes in the 5×5 window falls between $T1$ and $T2$, then its (their) coordinates are stored in another memory buffer. If such a value is not available, then the pixel is not an edge. **6)** The described procedure is repeated for the next pixel in the image row, taking into account that from the second pixel downwards the contents of the two memory buffers are constantly checked. The second buffer contains those pixels which can eventually be nominated for edge pixels on the basis of indirect connectivity.

In terms of speed, two advantages are achieved: total avoidance of recursive computations, and immense reduction of the calculations based on the comparison functions to define the edge/non-edge status of a single pixel.

2. 5. 3. Integer arithmetic speed

In this module, too, the comparison functions are the indispensable integer arithmetic. For the purpose of speed optimization, the accomplishable goal here is their reduction to the bare minimum on the platform of the appropriate computational algorithm implementing the hysteresis thresholding.

3. Conclusion

The technology of Canny edge detection computations for implementation on Altera FPGAs has been analyzed in this paper. The computational modules of John Canny's algorithm are scrutinized on the platform of memory utilization, organization of computations, and integer arithmetic for the purpose of optimizing the speed parameter. This goal is accomplished by employing a set of approaches: optimized memory utilization in terms of revolving Dual-port RAMs; fast technology for Gaussian smoothing employing a novel organization of computations combined with a new advanced integer division replacing technique; avoiding additional scaling calculations to prevent eroding the speed in the Sobel filtering module; defining a new algorithm for computing the gradient direction; new computational mechanism for the hysteresis thresholding algorithm aimed at total avoidance of recursion as

well as maximum reduction of the number of times the comparison functions, which are much slower than the other Altera FPGA integer arithmetic, are used. The proposed technology of computations is a solid basis for an advanced computational architecture of FPGA-based Canny algorithm focused on speed.

REFERENCES

1. Altera Corporation. Cyclone IV Device Handbook, Volume 1, 2014
2. Altera Corporation. Cyclone V Device Handbook, Volume 1, 2014
3. Altera Corporation. Stratix III Device Handbook, Volume 1, 2014
4. Altera Corporation. Stratix IV Device Handbook, Volume 1, 2014
5. Altera Corporation. Integer Arithmetic IP Cores User Guide, 2014
6. Canny, J. F. (1986). A computational approach to edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8, pp. 679-698
7. Chandrashekar N.S. and Nataraj K. R. (2013). Design and Implementation of a Modified Canny Edge Detector based on FPGA, *International Journal of Advanced Electrical and Electronics Engineering*, (IJAE), Vol.2, (1), pp. 17-21
8. Chandrashekar N.S. and Nataraj K.R. (2013). NMS and Thresholding Architecture used for FPGA based Canny Edge Detector for Area Optimization, *Proceeding. of International Conference on Control, Communication and Power Engineering*, pp. 80-84
9. Divya. D. and Sushma P. S. (2013). FPGA Implementation of a Distributed Canny Edge Detector, *International Journal of Advanced Computational Engineering and Networking*, Vol. 1, (5), pp. 46-51
10. Shamlee V. and Jeyamani (2014). A Split Canny Edge Detection: Algorithm and its FPGA Implementation. *International Journal of Science and Research (IJSR)*, Vol. 3 (11), pp. 1198-1205

University of Plovdiv
Plovdiv 4000
24 Tzar Asen Street
e-mail: dkromichev@yahoo.com



MECHANICAL AND TRIBOLOGICAL PROPERTIES OF AlN HARD FILMS DEPOSITED BY REACTIVE MAGNETRON SPUTTERING

NIKOLAY PETKOV¹, TOTKA BAKALOVA², TETIANA CHOLAKOVA¹,
LUKÁŠ VOLESKÝ², PETR LOUDA²

Abstract: Aluminium nitride films with a thickness of 2 μm are grown on high speed steel and Si (100) substrates by DC magnetron sputtering at 200°C in a gas mixture of argon and nitrogen, using a pure aluminium target. The influence of process parameters on the mechanical and tribological properties of deposited AlN films was studied. The surface topography of the AlN is investigated by atomic force microscopy in contact mode and an average surface roughness of 52nm is obtained. Mechanical properties of the films are examined by nanoindentation and microscratch techniques. The best samples showed the nanohardness of 28 GPa, elastic modulus of 437 GPa and good adhesion to the high speed steel substrates. Using a Multi-Specimen Test System in ball-on-disc mode at the room temperature are evaluated the friction coefficient and the wear rate of AlN. The calculated wear rate of the AlN films is $1.22 \times 10^{-3} \text{ mm}^3 \text{ N}^{-1} \text{ m}^{-1}$ and the coefficient of friction is 0.31.

Key words: reactive magnetron sputtering; AlN thin films; mechanical properties; tribological properties.

1. Introduction

Aluminum nitride (AlN) has many attractive properties including high thermal and chemical stability, high electrical resistivity, low dielectric contact, high thermal conductivity, high hardness, good mechanical strength, high melting point, high acoustic velocity, good dielectric properties and transparency in the visual and infrared region [1-4]. Because of these properties AlN film have application in various industries as microelectronics, optoelectronics, mechanical engineering, in passivation of thin films, insulator layers, optical sensors in the ultraviolet and hard coatings [1-3].

This paper presents some results of the mechanical and tribological properties of AlN thin films obtained by reactive magnetron sputtering deposition.

2. Experimental details

AlN coatings were grown onto steel (ISO 4957, HS6-5-2C) and Si (100) substrates by reactive DC magnetron sputtering technique. The substrate holder was located at 10 cm from the target. Prior to deposition process, the samples were cleaned in ultrasonic bath at the temperature of 60°C in acetone (8 min) followed by 8 min in isopropanol. After this, the samples were mounted in the PVD

chamber, in order to suffer a sputter-cleaning process. Before each deposition, the substrates were sputtered-cleaned at an argon atmosphere for 15 min. The bias voltage with the values of 900V (sample1) and 600 V (sample 2) was applied by a r.f. source (13.56 MHz) for two different series of samples in order to investigate its influence on the layer properties.

After the cleaning, samples were submitted to the coating process. The AlN films were obtained by sputtering of pure Al (99.99%) target in Ar/N₂ atmosphere. Deposition chamber was evacuated at $4.6 \times 10^{-3} \text{ Pa}$ before the film growth process by a turbomolecular pump.

Technological conditions of the deposition process for the two series of samples presented the best properties were: working pressure of 0.8 Pa, the ratio of working gases in the chamber Ar:N₂/7:3; magnetron power supply of 600 W and negative substrate bias voltage of 75V. The samples temperature was not higher than 200°C.

The influence of some the sputtering conditions, such as the N₂ concentration in the sputtering atmosphere, the bias voltage and the distance between the target and the substrate over the AlN films properties are investigated.

Properties of the deposited films are examined by the next techniques:

Thickness of the AlN films is confirmed using a “stylus” profilometer by measuring the difference in height of the step between an uncoated part and a coated part on Si substrate, provided by the profile scan.

The surface topography of the AlN is investigated by atomic force microscopy in contact mode. The surface area used for the AFM analysis was $10\mu\text{m}\times 10\mu\text{m}$. Two main parameters were analysed: average roughness (Ra) and root mean square roughness (Rms).

The hardness and the elastic modulus were measured by CSM Instruments nanoindentation technique with a trigonal shaped Berkovitch indenter and calculated from the loading–unloading curves. The Berkovitch tip area function was estimated with the Oliver–Pharr method.

The adherence strength evaluation is carried out with a Micro Scratch tester equipped with a diamond tip (Rockwell, radius 120°).

The tribological properties (wear and friction) are studied using CETR UMI Multi-Specimen Test System in ball-on-disc mode at a room temperature in dry friction conditions. The wear scars of the balls and tracks on the samples are observed with light optical microscope (LM) Carl ZEISS Axiolmager M2.

3. Results and discussion

3.1. Surface morphology

Surface topography of coating is examined with an atomic force microscopy (AFM) in contact mode. Fig. 1 shows the result from scan area $10 \times 10 \mu\text{m}^2$. The AFM analysis indicated that the coatings are dense, with an average surface roughness (Ra) of 52 nm and the root mean square (Rms) roughness of 71 nm.

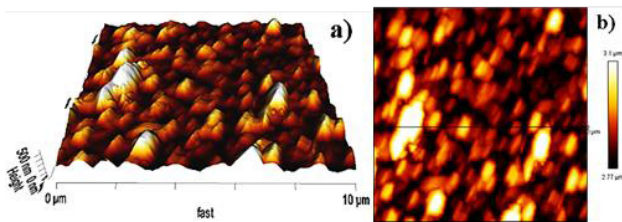


Fig. 1. (a) 2-D and (b) 3-D AFM image of surface topography of AlN thin films.

3.2. Nanoindentation

The CSM Instruments Indentation Tester for hardness estimation of the AlN coating surface is used. Measurement of nanohardness is made with a diamond indenter prism of Berkovitch, with loads in the range of 1-500 mN. Maximum penetration depth during measurement is not more than 10% from the coating thickness. Table 1 present the measurement results of nanohardness (H), Young modulus (E), and calculated result of plasticity

index [11] H/E and coating's resistance to plastic deformation [11] H^3/E^2 . The measured hardness value of AlN coatings is agree with previous reports, between 13 and 37 GPa [1, 5-9].

Table 1. Mechanical properties of the best AlN samples.

Sample, No	H[GPa]	E [GPa]	H^3/E^2	H/E
#1	28.25	437	0.118	0.065
#2	22.33	249	0.180	0.090

3.3. Scratch test analysis

Scratch tests were carried out in order to evaluate the adhesion between the film and the substrate. The adhesion properties of AlN coatings can be characterized by the following two stages: the first, L_{C1} , with the lower critical load, defined as the load where first cracks occurred (cohesive failure); and the second, L_{C2} , with the upper critical load, which shows the first delamination when the scratch track occurred (adhesion failure). The failure mechanisms were analyzed by optical microscopy, observing carefully the scratch grooves. Each failure mechanism was identified and measured the corresponding distance to the start point, in order to observe the corresponding normal load in the range of 0–100 N. Figure 2 presents the scratch test results for the #2.

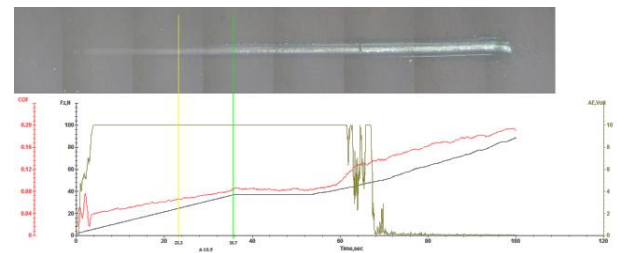


Fig. 2. Image of the micro-scratch test results of the #2 with marked L_{C1} and L_{C2} .

The first cohesive and adhesive failures were registered at 9.6 N and 21.6 N, respectively for the #1 and at 23.3 N and 35.7 N, respectively for the #2. The investigation of mechanical properties of the AlN revealed, that the value of bias voltage on the stage “substrate cleaning” have influence over the film hardness and adhesion.

3.4. Tribology

The tribological properties (wear and friction) were studied using a ball-on-disk rotating tribometer. The ball material was steel by AISI 420C with diameter of 6.35 mm. Used test parameters are: load force of 5N, sliding circumference radius of 4 mm and sliding speed of 60 rpm. The surface of sample after tribology test was scanned with profilometer. Recording of images from a light optical microscope is obtained

by CCD camera. The scanned data can then archive, analyze and modify by the software. The goal of data software analysis is the processing of visual information, elimination of random errors and extraction of typical features characterizing the image. [12]

3.4.1. Coefficient of friction

Coefficient of friction is a value that shows the relationship between the force of friction between two objects and the normal force between the objects. Therefore, its value is correct only for the friction pair for which is determinate. Here the friction pair are: (1) the uncoated steel substrate and steel ball, and (2) the AlN coating, deposited on the steel substrate, and steel ball. Fig. 3 shows the evolution of the friction coefficient results during the ball-on-disk tribology test in the case of dry friction. Curve 1 present the friction evolution between the AlN coating and ball. The coefficient of friction is around 0.31, and the coating endurance is 40 cycles (40 s) of dry friction in these circumstances. For comparison, Wu and co-authors are reported a friction coefficient value of 0.3235 for single AlN layer, in the case of 3 mm GCr15 steel ball counter body[7]. Curve 2 (Fig. 3) present friction between uncoated sample and ball. The coefficient of friction between those two parts is around 0.15, and the surface endures 130 cycles (130 s) of dry-friction in these circumstances.

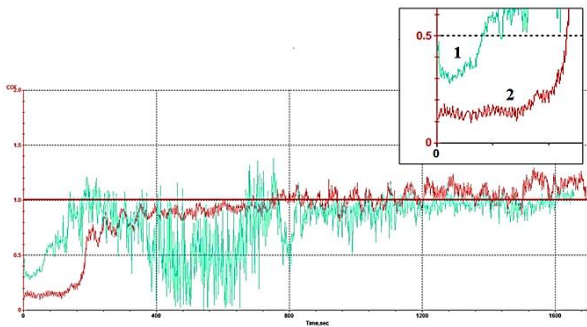


Fig. 3. Friction curves of the AlN coating sample 1 (1) and steel substrate (2) sliding against steel ball at 25°C.

3.4.2. Wear rate

Transferred material from ball to the coated surface and from coating to the ball surface, during tribology test ball-on-disc, makes difficult to measured wear of the coating. Therefore, wear of the counter body (ball) is measured and is being reported. The estimation of ball wear is made by Eq. (1) according standard EN 1071-13:2010.

$$V = \frac{\pi A^3 B}{32D} \quad (1)$$

where V is wear volume in mm^3 , A is measuring wear scar diameter in direction of erasure in mm, B is measuring wear scar diameter in direction perpendicular of the erasure in mm, and D is the ball diameter in mm.

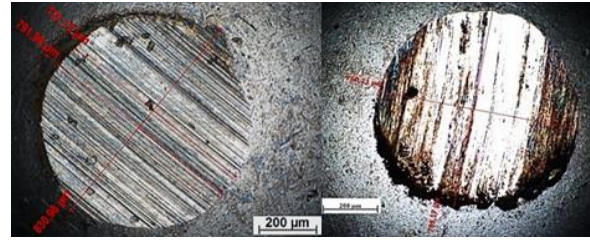


Fig. 4. Wear scar of the ball used for ball-on-disc test on sample with AlN coating (left) and on the uncoated sample (right).

Diameter values on the wear scar are measured by LM. Fig. 4 (left) presents wear scar photography of the ball from ball-on-disc test made over the AlN coated sample. The average value of the measured wear scar diameter in direction of erasure is $A = 756.56 \mu\text{m}$, and the measured wear scar diameter in direction perpendicular of the erasure is $B = 830.98 \mu\text{m}$. Therefore, the calculated value of ball wear by is $V = 5.56 \times 10^{-3} \text{mm}^3$. Fig.4 (right) presents wear scar photography of the ball from ball-on-disc test made over the uncoated sample. The measured values of the diameters are $A = 750.71 \mu\text{m}$ and $B = 786.17 \mu\text{m}$. Therefore, the calculated value of ball wear by is $V = 5.14 \times 10^{-3} \text{mm}^3$.

Wear rate estimation of the coating is made by the follow equation [10]:

$$\text{wear rate} = \frac{\text{wear volume}}{\text{load} \times \text{sliding distance}} \quad (4)$$

where sliding distance is in m, load in N, and wear volume (in mm^3) of the coating is calculated as follow:

$$V = Sh \quad (5)$$

$$S = 2\pi R_s d_s \quad (6)$$

where h is the penetration depth (or layer thickness), S is the track area, R_s is the sliding circumference radius, and d_s is the averagetrack width.

Thickness of AlN coating is $2 \mu\text{m}$. Average track width is 0.61mm (Fig. 5). Therefore, the calculated wear volume of the AlN coating is $30.662 \times 10^{-3} \text{mm}^3$. Sliding distance is equivalent of 200 cycles, consequently, the calculated wear rate of the AlN coating is $1.22 \times 10^{-3} \text{mm}^3 \text{N}^{-1} \text{m}^{-1}$.

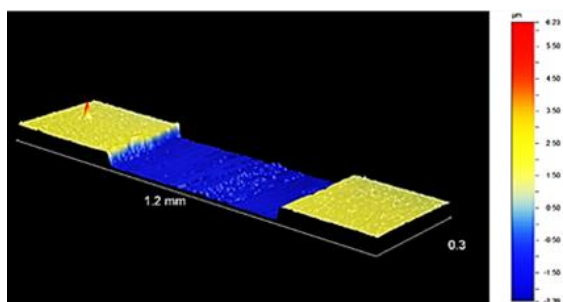


Fig. 5. The wear track profile of the sample with AlN coating (sliding with steel ball), 3Dimage.

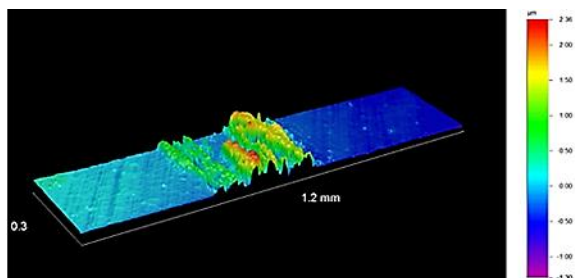


Fig. 6. The wear track profile from the ball-on-disc test on the sample without coating, 3D image.

Fig. 5 presents the profile of the track on the sample with AlN coating after ball-on-disc test with measured track width value by mechanical profilometer. Fig. 6 presents the profile of the track on uncoated sample after ball-on-disc test with measured tracks width value by mechanical profilometer. High speed steel (ISO 4957) has a significantly higher wear resistance compared with steel AISI420, due to the high percentage of carbon (C) and a high proportion of tungsten (W). This effect is observed in Fig. 6.

4. Conclusions

AlN thin films were deposited on HSS and silicon substrates by DC reactive magnetron sputtering in a gas mixture of argon and nitrogen. The AlN coatings are dense, with an average surface roughness of 52 nm. The samples have shown good adhesion to the high speed steel substrates, high value of hardness, low coefficient of friction and good adhesion properties. An in-depth study of the tribological properties of these coatings revealed that AlN films have higher wear resistance in comparison to high speed steel substrates. It has been found that the bias voltage applied to the substrate during the cleaning affects the properties of the layer.

ACKNOWLEDGEMENT

The paper was supported in part by the project LO1201 through the financial support of the Ministry of Education, Youth and Sports in the framework of the targeted support of the “National

Programme for Sustainability I” and the OPR&DI project “Centre for Nanomaterials, Advanced Technologies and Innovation” registration number CZ.1.05/2.1.00/01.0005.

REFERENCES

1. L. Yate, J.C. Caicedo, A. Hurtado Macias, F.J. Espinoza-Beltran, G. Zambrano, J. Munoz-Saldana, P.Prieto, Surf. & Coat. Technology **203**, 1904 (2009)
2. K. Tonisch, V. Cimalla, Ch. Foerster, H. Romanus, O. Ambacher, D.Dontsov, Sensors and Actuators A **132**, 658 (2006)
3. S. Venkataraj, D. Severin, R. Drese, F. Koerfer, M. Wuttig, Thin Solid Films **502**, 235 (2006)
4. F. Hajakbari, A. Hojabri, M.M. Larijani, Journal of Applied Chemical Research, 8, **3**, 45 (2014)
5. T.A. Rawdanowicz, V. Godbole, J. Narayan, J. Sankar, A. Sharma, Compos. Part B **30**, 657 (1999)
6. A.L. Ji, L.B. Ma, C. Liu, C.R. Li, Z.X. Cao, Diamond Relat. Mater. **14**, 1348 (2005)
7. Z.G.Wu, G.A. Zhang, M.X.Wang, X.Y. Fan, P.X. Yan, T. Xu, Appl. Surf. Sci., **253**, 2733 (2006)
8. X.Wang, A. Kolitsch, F. Prokert, W. Möller, Surf. Coat. Technol. **103–104** 334 (1998)
9. J.H. Lee, W.M. Kim, T.S. Lee, M.K. Chung, B. Cheong, S.G. Kim, Surf. Coat. Technol. **133 - 134**, 220 (2000)
10. K. Singh, P.K. Limaye, N.L. Soni, A.K. Grover, R.G Agrawal, A.K. Suri, Wear, **258**, 1813 (2005)
11. A. Leyland, A. Matthews, Wear, **246**, 1 (2000)
12. T. Bakalova, T., L. Volesky, P. Louda and Z. Andrsova. The Use of Optical Microscopy to Evaluate the Tribological Properties. Manufacturing technology. Vol. 1, N. 3. pp. 256 – 261. ISSN 1213-2489 (2014).

Central Laboratory of Applied Physics,
Bulgarian Academy of Sciences, 61, St.
Peterburg Blvd. 4000 Plovdiv, Bulgaria,
E- mail: petkovnik@gmail.com
E- mail: ipfban-dve@mbox.digsys.bg

Institute for Nanomaterials, Advanced
Technologies and Innovation, Technical
university of Liberec, Studentská 1402/2,
461 17 Liberec 1, Czech Republic,
E-mail: totka.bakalova@tul.cz
E-mail: lukas.volesky@tul.cz
E-mail: petr.louda@tul.cz



DEPOSITION OF AICN THIN FILMS BY REACTIVE DC MAGNETRON SPUTTERING AND TRIBOLOGY

NIKOLAY PETKOV¹, TOTKA BAKALOVA²

Abstract: *The process of friction between two solid surfaces is inevitably accompanied by the occurrence and energy losses. If the surfaces are separated by a layer of dissimilar material, the friction process also takes place within the layer, and the wear of the surface and possibly the energy losses due to friction are reduced considerably. AICN films with differing composition are deposited on the samples from high-speed steel by reactive DC magnetron sputtering of pure Al (99.99%) with various different gas ratios of N₂ and CH₄. The deposition chamber, before films growth processes, is evacuated up to approximately 5.2×10^{-3} Pa by a turbomolecular pump. During the deposition process and Ar plasma cleaning process the bias voltage is applied with r.f. source (13.56 MHz). Tribological testing (EN1071-13:2010) is conducted with a constant load of 5 N at room temperature and humidity of 40 ± 2 %, by using a ball made from steel ISO 683/13 (X105CrMo17) with a diameter of 6.350 mm.*

Key words: *thin films, magnetron sputtering, tribology, coefficient of friction, wear*

1. Introduction

Covalent compounds comprising light elements have excellent chemical and mechanical stability, due to the small atomic radius and strong interatomic bonding [1, 2]. Carbides, nitrides and aluminium have attracted considerable interest from multidisciplinary researchers and have found extensive applications in various industry branches [1-6]. In recent years, the research work has been shifting gradually, but steadily, from the binary materials to ternary ones [1, 2, 6].

Tribology is the study of the thermal, mechanical and chemical interactions that occur in mutually acting solid surfaces and their surroundings. Tribology should be understood as an interdisciplinary field of study. An explanation of the various scientific disciplines is often essential for an understanding of tribological problems. [7-9]

During relative movement between solid bodies, friction and wear are an inevitable occurrence. Both a low friction coefficient and low wear are desirable for most systems; this is associated with ensuring the reliable working of the system and less energy consumption, which leads to an extension of lifespan. Some systems require a high friction coefficient (e.g. a tyre on the road) and some systems require even higher wear (e.g. pencils), but most effort is focused on developing application systems with a low a friction coefficient and minimal wear. [7- 9] Wear is usually defined as

“damage to the hard surface comprising a progressive loss of material due to relative movement between the surface and the surface of another material or substance” [10].

Two solid bodies mutually deform each other if they come into contact. Deformation in the first body occurs as a result of mechanical resistance of the second body and vice versa. An important factor is whether or not the solid body comes into contact with the entire surface, partially or the surfaces are in a non-conforming contact. Contact interactions are further complicated by surface roughness. The development of structures of various mechanical systems is associated with the selection of suitable materials with specific properties. Friction and wear depend on many parameters and conditions, including: material pairing, contact geometry, applied normal load, contact pressures, relative sliding speed, material surface topography and roughness, environment, temperature, chemical interactions, sliding direction (unidirectional, reciprocating, random, etc.) [7-9]

Many mechanical systems need to work in various different environments. Some components must provide reliable wear and friction against several different types of materials. [10]

In this work are studied the mechanical and tribological properties of the AICN thin films, with different compositions, deposited by the reactive

DC magnetron sputtering as bias polarization voltage is applied by r.f. source.

2. Experimental information

Experiments are carried out in cooperation between the Central Laboratory of Applied Physics, Bulgarian Academy of Sciences and the Institute for Nanomaterials, Advanced Technology and Innovation, Technical University of Liberec (TUL).

Polished discs (\varnothing 20 mm \times 5 mm thick) of EN ISO HS 6-5-2 tool steel are used as the substrate material. It is a high-speed Mo-W steel with high toughness. The chemical composition of the EN ISO HS 6-5-2 steel [wt %] is: C 0.80–0.90; Mn max. 0.45; Si max. 0.45; P max 0.035; S max 0.035; Cr 3.80–4.60; Mo 4.50–5.50; W 5.50–7.00; V 1.50–2.20.

AICN films with differing composition are deposited on the samples from high-speed steel by reactive DC magnetron sputtering of pure Al (99.99%) with various different gas ratios of N_2 and CH_4 . Fig. 1 shows the substrate holder located at 10 cm from the magnetron target (left) and shape of the plasma during deposition process (right).

Vacuum chamber, before films growth processes, is evacuated up to approximately 5.2×10^{-3} Pa by a turbomolecular pump. During the deposition process and Ar plasma cleaning process the bias voltage is applied with r.f. source (13.56 MHz).



Fig. 1. The substrate holder location and plasma in vacuum chamber during deposition process.

Table 1 shows the gas flow ratio between the reactive gasses. Gas flow ration between Ar and the reactive gasses during deposition processes is 7/3. To achieve good adhesion to the substrate cleaning in a glow discharge of argon for 30 min at chamber pressure of 4 Pa and negative bias voltage of 900 V is performed, followed by deposition of a contact transition layer of AlN. Parameters of the deposition processes are: working chamber pressure of 0.7 Pa, magnetron power of 600W, negative bias voltage of 75V, and deposition time of 40 min. Sample temperature, during deposition is not higher than 200°C.

Prior to deposition process, the samples were cleaned in ultrasonic bath at the temperature of

60°C in acetone (8 min) followed by 8 min in isopropanol.

Table 1. Gas flow ration

Type of coatings	AICN-1	AICN-2	AICN-3
N_2/CN_4	5/1	2/1	1/1

Using CSM Instruments nanoindentation technique with a trigonal shaped Berkovitch indenter, the hardness and the elastic modulus were measured and calculated from the loading–unloading curves. The Berkovitch tip area function was estimated with the Oliver–Pharr method.

The adherence strength evaluation is carried out with a Micro Scratch tester equipped with a diamond tip (Rockwell, radius 120°).

Using CETR UMI Multi-Specimen Test System in ball-on-disc mode at a room temperature in dry friction conditions are studied the layers tribological properties (wear and friction). The wear scars of the balls and tracks on the samples are observed with light optical microscope (LM) Carl ZEISS Axiolmager M2.

3. Results and Discussion

3.1. Mechanical properties

Evaluation of the thin films thicknesses d is made by calottes. The diameter of used ball for this purpose is 30 mm, and the thickness results are given in Table 2.

The CSM Instruments Indentation Tester for hardness estimation of the AICN thin films surfaces is used. Measurement of nanohardness is made with a diamond indenter prism of Bercovich and loads in the range 1-500 mN. Maximum penetration depth during measurement is not more than 10 % from coating thickness. Table 2 shows the results from measurement of nanohardness H , elastic module E , and calculated ratio of plastic deformation H/E between them. This ration gives information for wear resistance of the thin films [11].

Table 2. Thickness d , Nanohardness H , elastic module E , and plastic deformation H/E

Type of coatings	d [μ m]	H [GPa]	E [GPa]	H/E
AICN-1	1.236	14	125	0.112
AICN-2	1.225	12	499	0.024
AICN-3	1.115	7.7	197	0.039

High H/E ratio is often a reliable indicator of good wear resistance in a coating, with the caveat that super- or ultra-hardness should not be pursued at the cost of extreme elastic property mismatch

between coating and substrate, since this will restrict the practical applicability of such coatings. Regardless of such factors, the reliability of the data employed to calculate H/E values needs to be assured, if the theoretical benefits of coating materials are to be translated into 'real' industrial applications. [12]

3.2. Scratch test analysis

Bruker's UMT-1 micro-mechanical test instrument platform provides ultimate flexibility in indentation and scratch testing.

Progressive Load Scratch Test: A stylus is moved over a specimen surface with a linearly increasing load until failure occurs at critical loads (L_{ci}). Normal force and tangential force are recorded. The failure events are examined by an optical microscope. L_c is a function of coating-substrate adhesion, stylus-tip radius, loading rate, mechanical properties of substrate and coating, coating thickness, internal stress in coating, flaw size distribution at substrate-coating interface, and friction between stylus-tip and coating. Figure 2 presents the scratch test results for the AICN-1.

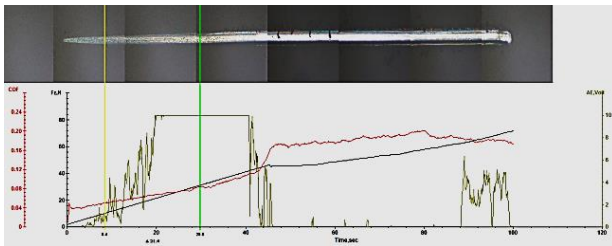


Fig. 2. Image of the scratch test results of the AICN-1 with marked L_{C1} and L_{C2}

Scratch test technique is carried out to characterize film adherence strength. The adhesion properties of AICN thin films is characterized by the following two stages: (1) L_{C1} , with the lower critical load, defined as the load where first cracks occurred (cohesive failure); and (2) L_{C2} , with the upper critical load, which shows the first delamination when the scratch track occurred (adhesion failure). The thin film adhesion scratch tests disclose the cohesion and adhesion properties of the tested coating. In virtue of the tests carried out, it is found that the critical load L_{C1} fitted within the range 5 – 11 N, and L_{C2} fitted within the range 18 – 31 N (Table 3).

3.3. Tribology measurement

Tribological testing is carried out in the Laboratory of Preparation and Analysis of Nanostructures at the Technical University of Liberec. Tribological testing (EN1071-13:2010) is conducted with a constant load of 5 N at room temperature and humidity of $40 \pm 2\%$, by using

a ball made from steel ISO 683/13 (X105CrMo17) with a diameter of 6.350 mm, and follow chemical composition [wt %]: C 0.95 - 1.20; Si max 1.0; Mn max 1.0; P max 0.040; S max 0.015; Cr 16.0 – 18.0; Mo 0.40 – 0.80 (57–60 HRC). Radius, of the circle along the "ball" body is moving, is 4 mm.

The basis of the tribological measurements is the "ball-on-disc" testing method. Fig. 3 shows schematically "ball-on-disc" configuration. The measurement involves the injection of fixed attachments of the test piece ("ball") in the form of balls of the chosen material defined as the force to drive (the test sample). An essential part of the test measurement is a friction sensor. The coefficient of friction (CoF) between the unit and the disc is determined during the test measurement.

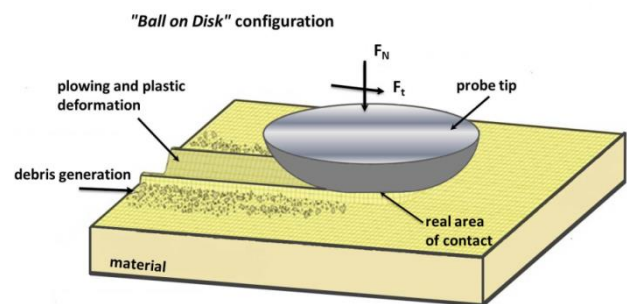


Fig. 3. Schematic representation of "ball-on-disc" configuration [7]

Fig. 4 presents graphically the measured dry friction coefficients, and their numerical values are given in Table 3.

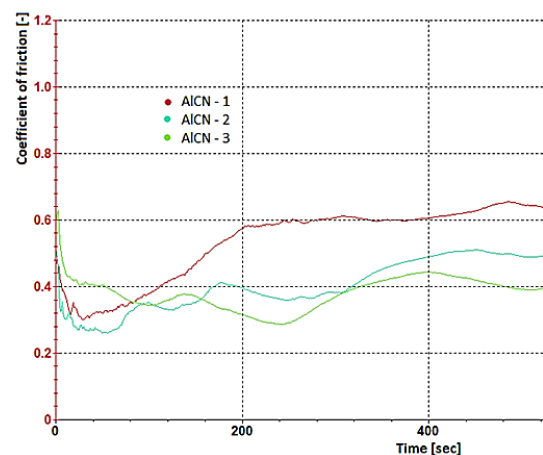


Fig. 4. Dry friction coefficients of AICN thin films

The optical microscope Zeiss AXIO Imager M2 is employed to evaluate the scar size on the steel balls. Attrition can be defined as a progressive loss of material due to relative movement between both surfaces. The result is a specific amount of ploughed material, in some cases also plastic deformation or displacement of the disc material due to side forces acting on the ball.

Evaluation of wear of the friction pairs of the ball against the steel is performed according to standard EN 1071-13:2010, as calculation of wear volume V_{pin} [mm³] of the balls is made by Eq. (1). Where A [mm] is measuring wear scar diameter in direction of erasure, B [mm] is measuring wear scar diameter in direction perpendicular of the erasure, and D [mm] is the ball diameter.

$$V_{pin} = \frac{\pi A^3 B}{32D} \quad (1)$$

Table 3. CoF and adhesion properties of AICN thin films, and wear volume of the balls

Type of coatings	CoF [-]	$V_{pin} \times 10^{-3}$ [mm ³]	L_{C1} [N]	L_{C2} [N]
AICN-1	0.354 ± 0.038	1.105	11	31
AICN-2	0.317 ± 0.040	0.585	5	18
AICN-3	0.336 ± 0.081	0.870	6.5	26

4. Conclusions

The experiments describe the mechanism of the process of interaction of the friction elements and their wear. The described experiments simulate the processes during machine cutting and enable us to predict the behaviour of materials during their practical application.

In this material are presented and studied AICN thin films deposited by reactive DC magnetron sputtering as the bias polarization voltage is applied by r.f. source. The measured nanohardness of this layers are commensurate with result presented in Ref. [6], but are too far from reported 53.4 GPa in Ref. [1].

The study shows that the changes in gas flow mixture at the deposition process have influence over mechanical and tribological properties of the thin films, as decrease of CoF, hardness and value of ratio H/E . Notably, that the observed differences in the thickness are inessential.

ACKNOWLEDGEMENT

The paper was supported in part by the project LO1201 through the financial support of the Ministry of Education, Youth and Sports in the framework of the targeted support of the “National Programme for Sustainability I” and the OPR&DI project “Centre for Nanomaterials, Advanced Technologies and Innovation” registration number CZ.1.05/2.1.00/01.0005.

REFERENCES

1. A.L. Ji., L.B. Ma, C. Liu, C.R. Li, Z.X. Cao, *Diamond & Related Materials* 14, 1348 (2005)
2. A.L. Ji., L.B. Ma, C. Liu, P. Zheng, C.R. Li, Z.X. Cao, *Applied Physics Letters* 86, 021918 (2005)
3. A. Lousa, S. Gimeno, *J. Vac. Sci. Technol., A, Vac. Surf. Films* 15 (1) (1997) 62.
4. C.M. Zettering, M. Ostling, K. Wongchotigul, M.G. Spencer, X. Tang, *J. Appl. Phys.* 82 (6) (1997) 2990.
5. J.H. Xu, L.H. Yu, I. Kojima, *J. Appl. Phys.* 94 (10) (2003) 6827.
6. L. Yate, J.C. Caicedo, A. Hurtado Macias, F.J. Espinoza-Beltran, G. Zambrano, J. Munoz-Saldana, P. Prieto, *Surf. & Coat. Technology* 203, 1904 (2009)
7. Bakalova, T., *aj. Nanoadditives SiO₂ and TiO₂ in Process Fluids. Manufacturing technology.* 1. vyd. Ústí n. L.: J. E. Purkyne University in Usti nad Labem, 2015, vol. 4. pp. 502 – 508, ISSN 1213-2489.
8. K. Holmberg, A. Matthews: *Coatings Tribology: Properties, Mechanisms, Techniques and Applications in Surface Engineering.* 2nd ed. Boston: Elsevier Science, xv, 560 p., [3] p. of plates. Tribology and interface engineering series, 56. ISBN 9780444527509, 2009.
9. Burakowki, T., Wierzchon, T. (1999). *Surface engineering of metals*, CRC Press LLC, ISBN 0-8493-8225-4, New York.
10. Shaji, S., Radhakrishnan, V., (2003). *Application of solid lubricants in grinding: investigations on graphite sandwiched grinding wheels.* *Mach. Sci. Technol.* 7, 137-155.
11. J. Musil, *RSC Adv.*, 2015, 5, 60482
12. A. Leyland, A. Matthews, *Wear*, 246, 1 (2000)

Central Laboratory of Applied Physics,
Bulgarian Academy of Sciences, 61, St.
Peterburg Blvd. 4000 Plovdiv, Bulgaria, E-
mail: petkovnik@gmail.com

Institute for Nanomaterials, Advanced
Technologies and Innovation, Technical
university of Liberec
Studentská 1402/2, 461 17 Liberec 1, Czech
Republic, Phone: +420 485353818, E-mail:
totka.bakalova@tul.cz.

EXPERIMENTAL DETERMINING OF THE ELASTICITY MODULE OF POLYMER CONCRETE COMPOSITES

ILIA POPOV, PEPO YORDANOV, SABI SABEV

Abstract: *This study reviews the test method of bending intended for quantitative determination of the module of linear deformation – E of gamma polymer concrete composites. The experimental samples have the form of rectangular parallelepiped (beam type). A testing table for bending is described.*

Keywords: *polymer concrete composites, module of linear deformation, regression analysis.*

1. Introduction

The higher requirements towards the accuracy of dimensions and forms of the details processed on metal cutting machines (MCM), and especially of these on MCM with CPU are major reason for increasing the quality parameters, efficiency and reliability of the metal cutting machines. In this line of thought, the dynamic processes are considered as priority in the structural design and calculation of the MCM.

The conventional approach used until recently for analysis of the static and dynamic behavior of the body parts of the support system of the MCM, in which the deformation, pressures and dynamic parameters in them are determined, always included the preparation of a kind prototype and its testing as well. The physical prototyping and testing require lots of time and resources. All this reflects on the productivity, quality, accuracy, and cost price of the production.

Due to the widespread application of the computer equipment and automation of the engineering labour in machine building sector, virtual prototyping became possible. In it, the product is designed immediately through a 3D (three-dimensional) prototype with high level of complexity. For the different virtual tests, of specific importance are the quantitative values of the material constants of the structural material of the product. These values are necessary as input data for the various software products for engineering analysis. In this regard, the **intention** of this study is: To determine the modules of elasticity

of gamma polymer concrete composites, made in the Test and Research Laboratory of Technical University – Sofia, Plovdiv Branch. The same will be used as structural material for body and body parts in production equipment.

2. Theoretic prerequisites

One of the approaches in testing the technical polymer composites is through bending test [1]. The simplest case of bending is the pure bending. In this instance, prevailing normal stresses from bending moment is applied at the cross-sections of the sample type "beam".

Fig. 1 shows the three-point loading of the beam, with bending force P and distance between the supports – l .

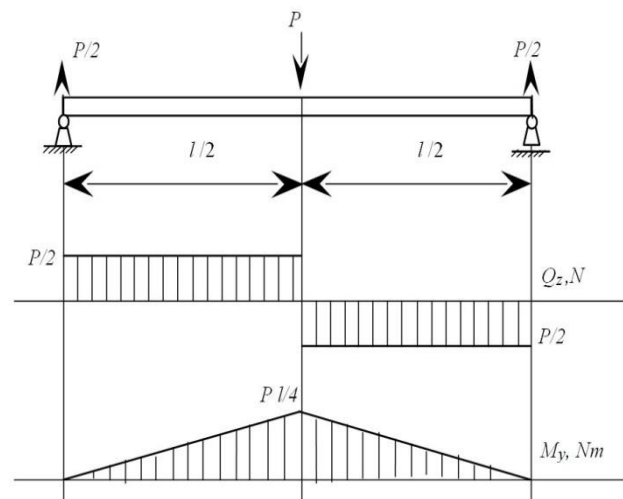


Fig. 1. Diagrams in the cases of bending

From the theory of bending [2, 3] it is known that the deformation f of a beam lying on two supports and loaded with focused static force P is equal to:

$$f = \frac{Pl^3}{48EJ_y} \quad (1)$$

On this grounds we express the force P :

$$P = \frac{48EJ_y}{l^3} \cdot f, \quad (2)$$

where:

EJ_y - is the linear hardness of bending;

E - module of linear deformations;

J_y - moment of inertia versus the central axis of inertia y . For the rectangular cross section it is equal to:

$$J_y = \frac{b \cdot h^3}{12} \quad (3)$$

After substituting (3) in (1) and (2), the following result is obtained:

$$f = \frac{Pl^3}{4Ebh^3}, \quad P = \frac{4Ebh^3}{l^3} \cdot f. \quad (4)$$

The diagram of the functional dependence between the loading force P and the bending flexure (deformation) of the sample body f is shown on Fig.2.

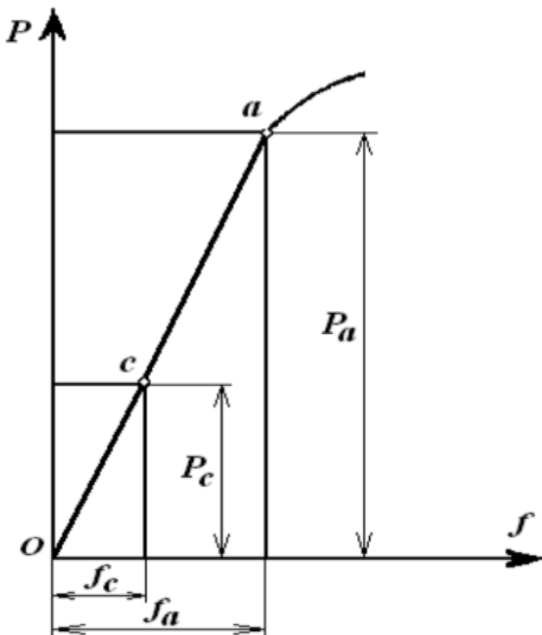


Fig. 2. Functional dependence between P and f

Regarding the boundary of proportionality, where the Hooke's law applies, we can correctly write down that:

$$P_a = \frac{4Ebh^3}{l^3} \cdot f_a, \quad (5)$$

$$P_c = \frac{4Ebh^3}{l^3} \cdot f_c \quad (6)$$

$$\text{and } P_a - P_c = \frac{4Ebh^3}{l^3} \cdot (f_a - f_c) \quad (7)$$

On the grounds of (7) we draw the dependency for the module of linear deformation - E :

$$E = \frac{(P_a - P_c)l^3}{4bh^3(f_a - f_c)} \quad (8)$$

3. Methodics of the experiment

On the grounds of the bending test standard [1, 4] and the above clarified algorithm for determining the module of linear deformation a laboratory table is designed and made. The sample bodies will be loaded with fixed forces as follows: 200 N, 300 N, 400 N and 500 N. Thus three force differences are formed. Respectively, at each loading, the bending flexure of the respective sample will be measured. The obtained results will be processed with the help of the statistical program Minitab 17.

The sample bodies for experimental determining of the module of the linear deformations of polymer concrete composites (15 composites) have the form of rectangular parallelepiped (beam), with dimensions 30x30x350 mm. The form and dimensions are consistent with the generally accepted standardization regulations.

The table for determining the module of linear deformation in bending is shown on Fig.3 and Fig.4. The tested polymer concrete sample 4 is placed freely on the supports 2. They are fixed between the sides 1 through bolts 21. The supports are made of steel (ANSI 304) with diameter $\phi 30$, consistent with the requirements of the standard [1], and they are positioned at $l = 320 \text{ mm}$ one from the other. The loading is done through ball screw and nut transmission 7-14, transmitting the loading to

the sample with the help of the steel (ANSI 304) support 6, with diameter $\phi 30$.,

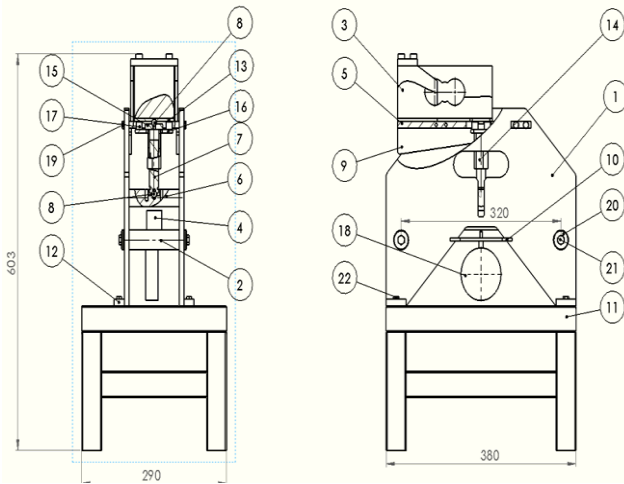


Fig.3. Testing table

The loading force is measured by a tensiometer (sensor 3, model "1-PW12BC3/300k-C" by the American company "HBM", with sensitivity $2mV/V$ and maximum loading up to $3kN$. It is connected to a standard terminal ST150 of the German company BIZERBA, Fig.4.

The deformations of the samples are measured by a micrometer 18 (range $1mm$ and separating capacity $1\mu m$), mounted on a stand 10.



Fig.4. Testing table

Table 1 Bending flexure f in μm

№ specimen	Value E						
	Load N				Received $E \times 10^9 Pa$		
	200	300	400	500	E_1	E_2	E_3
1.1	57	95	135	174	25,50	24,22	24,84
1.2	55	97	138	180	23,07	23,63	23,07
1.3	60	105	145	184	21,53	24,22	24,84
2.1	56	105	155	205	19,77	19,38	19,38
2.2	68	118	170	218	19,38	18,63	20,19
2.3	75	122	175	232	20,61	18,28	17,00
3.1	25	44	57	77	50,99	74,53	48,44
3.2	34	53	75	105	50,99	44,04	32,30
3.3	58	95	134	168	26,19	24,84	28,50
4.1	62	112	160	216	19,38	20,19	17,30
4.2	58	100	155	207	23,07	17,62	18,63
4.3	72	121	172	220	19,77	19,00	20,19
5.1	75	130	194	252	17,62	15,14	16,70
5.2	67	125	184	234	16,70	16,42	19,38
5.3	65	125	184	245	16,15	16,42	15,88
6.1	88	152	235	310	15,14	11,67	12,92
6.2	95	178	257	342	11,67	12,26	11,40
6.3	98	170	257	350	13,46	11,14	10,42
7.1	70	122	195	258	18,63	13,27	15,38
7.2	80	140	195	275	16,15	17,62	12,11
7.3	76	140	210	280	15,14	13,84	13,84
8.1	90	158	227	296	14,25	14,04	14,04
8.2	100	170	240	305	13,84	13,84	14,91
8.3	80	155	222	295	12,92	14,46	13,27
9.1	74	122	173	237	20,19	19,00	15,14
9.2	72	131	185	241	16,42	17,94	17,30
9.3	68	126	185	240	16,70	16,42	17,62
10.1	71	141	200	263	13,84	16,42	15,38
10.2	57	118	170	232	15,88	18,63	15,63
10.3	65	118	185	240	18,28	14,46	17,62
11.1	68	127	186	241	16,42	16,42	17,62
11.2	75	135	200	265	16,15	14,91	14,91
11.3	75	138	204	265	15,38	14,68	15,88
12.1	74	138	201	275	15,14	15,38	13,09
12.2	98	160	225	310	15,63	14,91	11,40
12.3	80	155	227	308	12,92	13,46	11,96
13.1	57	115	155	232	16,70	24,22	12,58
13.2	66	113	167	225	20,61	17,94	16,70
13.2	60	110	160	215	19,38	19,38	17,62
14.1	60	96	138	174	26,91	23,07	26,91
14.2	61	98	132	170	26,19	28,50	25,50
14.3	52	92	130	168	24,22	25,50	25,50
15.1	52	112	154	221	16,15	23,07	14,46
15.2	68	123	195	256	17,62	13,46	15,88
15.3	70	132	195	260	15,63	15,38	14,91

4. Experimental results

The obtained experimental results for the modules of linear deformations of the polymer concrete composites are grouped in Table 1, where $E1, E2, E3$, are respectively the calculated values on the boundary of proportionality for the three force differences.

5. Processing of results

The experimental polymer concrete composites are multicomponent systems with interdependent and bilaterally limited components due to which the regression models represent the so called aligned (canonical) polynomials.

The mathematical and statistical processing is made with the software product MINITAB 17. For the mathematical description of the target function \hat{y} - (the module of linear deformation E) an aligned model of third power of the following type is used:

$$\hat{y} = \sum_{i=1}^q b_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q b_{ij} x_i x_j + \sum_{i=1}^{q-2} \sum_{j=i+1}^{q-1} \sum_{l=j+1}^q b_{ijl} x_i x_j x_l$$

The data in Table 1 is processed and the following regression model was obtained:

$$\hat{y} = 9724x_1 + 6724x_2 + 42976x_3 + 298x_4 - 67885x_1x_2 - 296035x_1x_3 - 13601x_1x_4 - 30676x_2x_3 - 8560x_2x_4 - 47173x_3x_4 + 538594x_1x_2x_3 + 76319x_1x_2x_4 + 319844x_1x_3x_4 - 81458x_2x_3x_4$$

	Sum of the squares	Degrees of freedom	Dispersion assessment
Regression	$Q_R = 2280,9$	$\nu_R = 13$	$s_R^2 = 175,454$
Error	$Q_{rez} = 515,21$	$\nu_{rez} = 31$	$s_{rez}^2 = 16,620$
Total	$Q_o = 2796,1$	$\nu_o = 44$	

Model adequacy:

The coefficient of determination is calculated, namely $R^2 = 81,57\%$, while the corrected coefficient of determination has the following value $R_{adj}^2 = 73,85\%$. In order to check the significance of R^2 a value has been calculated, having Fisher's distribution:

$$F = \frac{s_R^2}{s_{rez}^2} = \frac{175,454}{16,620} = 10,556 \quad (9)$$

In Table [5] with Fisher's distribution F_T was calculated with significance level $\alpha = 0,05$ and degrees of freedom $\nu_R = 13 \nu_{rez} = 31$:

$$F_T = F(0,05;13;31) = 2,04 \quad (10)$$

As $F > F_T$ ($10,556 > 2,04$) a conclusion can be made that R^2 is significant.

In order the coefficient to be significant, the following have to be valid $|t_i| \geq t_T(\alpha; \nu_{rez})$. In Table [6] with Student's distribution, with level of significance $\alpha = 0,05$ the following was reported $t_T = (0,05;31) = 1,695$. On the basis of a comparison of t_T with the values of Table 2, the conclusion can be made that almost all coefficients are insignificant with the exception of one.

Table 2 Quote from the report

Term	Coef	SE Coef	T	P	VIF
C2	9724	7026	*	*	2655742
C3	6724	10240	*	*	3605541
C4	42976	23993	*	*	1039105
C5	298	361	*	*	188941
C2*C3	-67885	77682	-0,87	0,389	4122959
C2*C4	-296035	189561	-1,56	0,129	1289979
C2*C5	-13601	10693	-1,27	0,213	3250748
C3*C4	-30676	79938	-0,38	0,704	147043
C3*C5	-8560	15280	-0,56	0,579	4166049
C4*C5	-47173	29342	-1,61	0,118	793799
C2*C3*C4	538594	179541	3,00	0,005	14755
C2*C3*C5	76319	106558	0,72	0,479	3982846
C2*C4*C5	319844	239754	1,33	0,192	1042078
C3*C4*C5	-81458	106558	-0,76	0,450	129269

The analysis of the residuals is made by means of the charts of standardized residuals, Fig.5.

Fig. 5 clearly shows the presence of two mistakes: Observation no. 7 and no. 9. The experimentally measured values of the controllable parameter are: $y_7 = 57,99$ and $y_9 = 26,51$, while the forecasted values of the model are: $y_{7,9} = 42,265$. The difference is significant and respectively the standardized residual is $d_7 = 15,725$ and $d_9 = -15,755$. This provides us grounds to eliminate observations no.7 and no. 9 and to process the data again.

After their elimination the following regression model is obtained:

$$\hat{y} = 9728x_1 + 6711x_2 + 43061x_3 + 297x_4 - 67802x_1x_2 - 296703x_1x_3 - 13605x_1x_4 - 30688x_2x_3 - 8537x_2x_4 - 47272x_3x_4 + 539390x_1x_2x_3 + 76168x_1x_2x_4 + 320640x_1x_3x_4 - 81610x_2x_3x_4$$

	Sum of the squares	Degrees of freedom	Dispersion assessment
Regression	$Q_R = 1158,94$	$\nu_R = 13$	$s_R^2 = 89,149$
Error	$Q_{rez} = 19,72$	$\nu_{rez} = 29$	$s_{rez}^2 = 0,680$
Total	$Q_o = 1178,66$	$\nu_o = 42$	

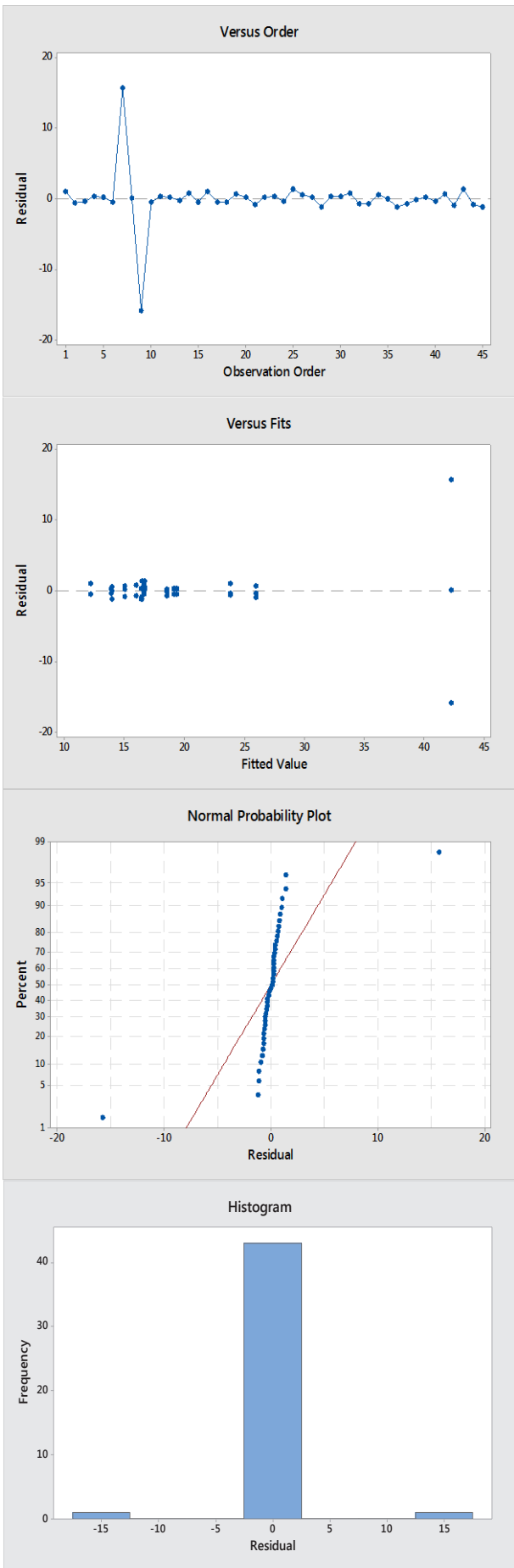


Fig. 5. Graphs residues

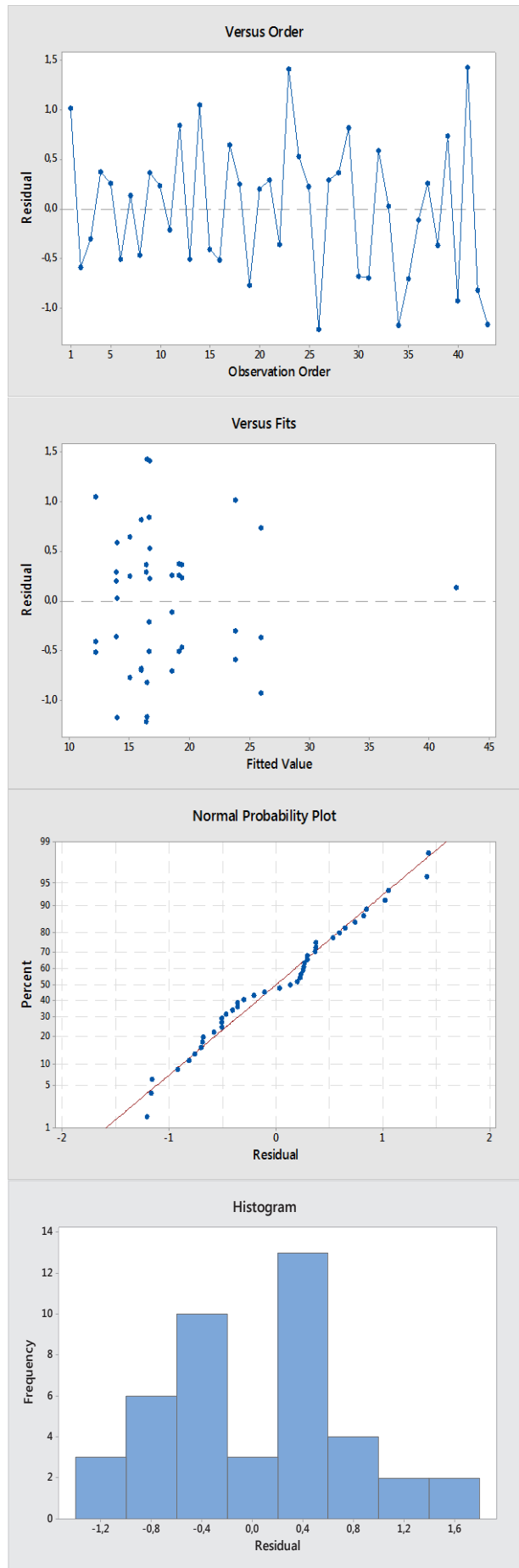


Fig. 6. Graphs residues

Model adequacy:

The coefficient of determination is calculated $R^2 = 98,33\%$, while the corrected coefficient of determination has the following value $R_{adj}^2 = 97,58\%$. In order to check the significance of R^2 a values was calculated having Fisher's distribution:

$$F = \frac{s_R^2}{s_{rez}^2} = \frac{89,149}{0,680} = 131,1 \quad (11)$$

In Table [5] with Fisher's distribution F_T was calculated with level of significance $\alpha = 0,05$ and degrees of freedom $\nu_R = 13$; $\nu_{rez} = 29$:

$$F_T = F(0,05;13;29) = 2,04 \quad (12)$$

As $F > F_T$ ($131,1 > 2,04$) a conclusion can be made that R^2 is significant.

In order the coefficient to be significant, the following should be true $|t_i| \geq t_T(\alpha; \nu_{rez})$. In [6] with Student's distribution, at level of significance $\alpha = 0,05$, $t_T = (0,05;29) = 1,699$ was calculated. On the basis of the comparison of t_T with the values in Table 3, the conclusion can be made that all coefficients are significant, therefore the model is adequate and preserves itself as it is.

Table 3 *Quote from the report*

Term	Коеф.	SE Coef	T	P	VIF
C2	9728	1425,3	*	*	2585076
C3	6711	2092,7	*	*	3428745
C4	43061	5232,1	*	*	1207901
C5	297	73,5	*	*	182452
C2*C3	-67802	15827,6	-4,28	0,000	3975959
C2*C4	-296703	41316,6	-7,18	0,000	1498033
C2*C5	-13605	2164,0	-6,29	0,000	3146107
C3*C4	-30688	16170,8	-1,90	0,068	147091
C3*C5	-8537	3133,7	-2,72	0,011	3973054
C4*C5	-47272	6357,1	-7,44	0,000	910835
C2*C3*C4	539390	40690,5	13,26	0,000	18526
C2*C3*C5	76168	21834,0	3,49	0,002	3870952
C2*C4*C5	320640	51851,2	6,18	0,000	1191430
C3*C4*C5	-81610	21834,0	-3,74	0,001	132672

The analysis of the residuals is made by means of the charts of the standardized residuals, Fig.5.

The analysis of the residuals does not show disruption of the prerequisites for the regression analysis. On Fig. 5 it can be seen that all residuals are within the range ± 2 . Therefore a conclusion can be made that there are no gross errors.

6. Conclusion

The results from the current study can be summarized as follows:

- Experimentally, the quantitative values have been obtained of the modules of elasticity of 15 different polymer concrete composites.
- The opportunity for obtaining true information about the linear deformations of the viscous-elastic materials of the type of polymer concrete composite, with this set of measurement equipment is feasible and adequate.
- A table was designed and produced for determining the module of elasticity through bending tests of the samples.

REFERENCES

1. ASTM-D-790. *Standard Test Methods for Flexural Properties of Unreinforced and Reinforced Plastics and Electrical Insulating Materials.*
2. КОЗЛОВСКИЙ А.Э., В.В. Бойцова. (2007). *Механические свойства материалов. Методы испытаний.* ГОУ ВПО «ИГХТУ» . ИВАНОВО.
3. Фудзий, Т., М. Дзак. (1982) *Механика разрушения композитных материалов.* Перев. с японского. М., "Мир".
4. ASTM C580 - 02(2008) - *Standard Test Method for Flexural Strength and Modulus of Elasticity of Chemical-Resistant Mortars, Grouts, Monolithic Surfacing, and Polymer Concretes*
5. http://www.socr.ucla.edu/applets.dir/f_table.html#FTable0.05
6. <http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf>.

The publication is related to work on a research project № 152PD0048-24/2015 on "Research and optimization of range polymer concrete composites in terms of their physical constants."

Contact:

Technical University of Sofia, branch
Plovdiv

E-mail: ilgpo@abv.bg

E-mail: piyordanov@tu-plovdiv.bg

E-mail: sabi_sabev@abv.bg



AN EXPERIMENTAL STUDY OF THE INFLUENCE OF ADVANCE PROCESSOR ON SOME OPERATING PERFORMANCE INDICATORS OF SPARK IGNITION ENGINE WORKING WITH LPG

KRASIMIR AMBAREV

Abstract: *This article presents the results of experimental study of the effects of embedding of advance processor A.E.B. PUMA on some operating performance indicators of spark ignition engine Honda D15Z8 in his work with LPG. There are also analyzed data for acceleration time from 0 to 100 km / h with the engine working respectively with: gasoline, LPG with switched on advance processor and LPG with switched off advance processor. The times were compared with the acceleration time specified by the factory. There are also analyzed the results of measurements of power and torque of the driving wheels with the engine running on gasoline, LPG by working with switched on and working with switched off advance processor.*

Key words: *experimental study, SI engine performance, LPG, advance processor.*

1. Introduction

The depletion of the oil reserves globally and the increasingly strict requirements towards environmental pollution are a prerequisite for the use of gas fuels both in the newly manufactured, and for conversion of the existing motor vehicles in order to work with such fuels. The use of gas fuels in the internal combustion engines is related to a number of other advantages as well [1]. In the commonly integrated gas systems intended for the engine to work with gas fuel, the operational times of the gas injectors are defined through adjustment of the times of the gasoline injectors by means of correction coefficients for every operational mode of the engine. That is, an operating card for the alternative fuels is created, on the basis of the gasoline fuel card of the respective engine. In the commonly integrated LPG systems of fourth generation, belonging to the low or average price class, there is no option for correction of the angle of ignition timing, defined on the basis of the factory set card, in the instances when the engines work with LPG. Having in mind the fact that the speed of combustion of the fuel-air mixture is different when the engine works with gasoline and with LPG, an additional correction is required of the angle of ignition timing in the different operational modes. In the engine control systems of the automobiles without knock sensor, there is no option for correction of the angle of ignition timing

when working with fuels with different properties. In such instances, in order to improve the performance and the cost effectiveness of the engine when working with LPG, it is recommended to install an external advance processor, correcting the angle of spark ignition timing in the different operational modes of the engine. Different types of advance processors exist. In some of them the angle of spark ignition timing is corrected depending on the rotation frequency, while in others there is a further correction due to the engine loading, which is detected by the signal of the MAP sensor or the flow meter.

There are different methods of measuring the engine power. In some of them the engine is fixed on a special stand with loading device, while in the others the power of the driving wheels is measured. In this instance the motor vehicle is positioned on a roll-stand. On the grounds of respective relationships the engine power is calculated on the basis of the driving wheels, i.e. in this instance it is measured indirectly. Other variant for indirect measuring of the engine power is by means of road tests related to the determination of the dynamic characteristics of the automobile. They include measuring the time necessary for the motor vehicle to reach the preliminary set speed, which is usually 100 km/h, measurement of the distance necessary to reach the set speed.

It has been found out that the time for steady motion of the cars in urban traffic is just 15-25%, while the time for acceleration is 30-45% of the total time. Therefore an important dynamic characteristic of the automobiles is to accelerate fast, which impacts their average speed of movement and is in direct relation with the power performance of the engine.

The acceleration is achieved though consecutive transition from the slower to the faster gears, and it is characterized with the following major indicators: linear acceleration, time and distance for acceleration from the set initial to the end speed.

In the theoretical analysis of the acceleration the initial time is not taken into account, during which time in the case of starting from stationary position the angular velocity of the crankshaft of the engine and the primary shaft of the gear box are equaled by smooth switching on the clutch. This time is short and it does not influence significantly the total time for the acceleration. In addition to this, when pressing the clutch usually the accelerator pedal is pressed to its maximum, which corresponds to the engine operation as per external speed characteristic. Under these conditions the greatest accelerations are achieved.

The following equation is used for calculating the acceleration:

$$a = \frac{D - \psi}{\delta_a} g. \quad (1)$$

Therefore the acceleration depends on the difference that expresses the reserve of dynamic factor after overcoming the driving resistance, and on the influence of the rotating masses, expressed by the coefficient δ_a .

The acceleration time within a certain range is calculated analytically with the equation

$$t_y = \int_0^{t_y} dt = \int_{V_u}^{V_y} \frac{1}{a} dV. \quad (2)$$

The acceleration distance for the same speed ranges calculated with the equation

$$S_y = \int_0^{S_y} dS = \int_0^{t_y} V dt. \quad (3)$$

2. LPG operating system for Internal combustion engines

In a light passenger vehicle make Honda Civic Aerodeck, supplied with 16-valve, 4-stroke engine with spark ignition, with code D15Z8,

which is with gasoline injection fuel system PGMFI, developed by Honda, an LPG system is installed, consisting of a computer supplied with electric installation AC STAG 300 ISA2 [2], a fuel injector for LPG – Alex Barracuda BR124P, reducer-evaporator [3] – Tomasetto AT09 Alaska, euro-gauge and gas container, homogenized according to standard ECE 67R-01 for operation with alternative fuels, as well as the necessary pipes.



Fig. 1. Front kit of LPG system

The LPG/CNG injection rail Valtek 3Ω , included in the standard front kit offered on the market – Fig.1 is replaced with a gas injector by the company Alex, model Barracuda BR124 [4] – Fig. 2.



Fig. 2. Injection rail Alex Barracuda

In addition, an advance processor is installed, produced by the Italian company A.E.B., model 515N Puma [5], shown on Fig. 3.

This type of advance processor is integrated in series in the electric chain between the electronic control unit of the engine and the commutator, which in this engine is positioned in the housing of the ignition distributor of the engine. The purpose of the advance processor is to dephase over time the electric impulse to the commutator generated by the electronic control unit of the engine. The introduced correction depends on the rotation frequency of the crankshaft according to a preliminary set algorithm

in its software. There is an option to switch off the advance processor in instances of idle running of the engine. For this purpose it is connected to the throttle position sensor.



Fig. 3. Reprogrammable timing advance processor A.E.B. Puma

The advance processor is supplied with the software required for the respective engine – version 515N-07. A basic value is set of the angle of correction at 6 degrees, through the switches on the device PCB.

During the tests the motor vehicle is equipped with standard sized tires 185/60R14, product of the company Kleber Krisalp HP2, mounted on metal rims. The pneumatic tires are with pressure according to the instructions of the factory producer.

The measurements are made on a horizontal asphalt road, without wind.

For the measurements a board computer „Infoboard“ is used, „touch screen“ version [6]. „Infoboard“ is built on the basis of a 16 bits processor, supplied with colour sensor display. It has own operational system, powered by DC constant voltage by the diagnostics 5 pin connector of the car. It works with electronic control units of the engine compatible with the standard OBD2A, which is the original control unit of the engine with code D15Z8. For making the connection between „Infoboard“ and the electronic control unit of the car, a standard communication protocol ISO 9141-2 is used, where the connection is through the k-line through the diagnostics 5 pin connector of the car.

The board computer offers options such as: measuring the fuel consumption – in a certain moment and average for the given distance, average for the exploitation process of the vehicle, detailed data for each separate trip; provides an option related to the cost effective driving of the vehicle, reading and visualization of live data, clearing of errors, software for time and distance of

acceleration, service intervals, expenditures related to the exploitation of the car, allowed speed assistant, option for recording the test results, and others.



Fig. 4. Board computer „Infoboard“, „touch screen“ version

A speed test was used for the measurements-“Speed Test” from the menu “Dyno”. The measurements include the time and distance of acceleration of the car from 0 to 100 km/h. The data for the rotation frequency and speed of motion are obtained from the respective sensors of the engine control system. The time is measured by the integrated timer.

It is necessary to start the test when the car is moving, before stopping at the start position. As a result of the measurement the relationship is visualized between the speed and time, as well as between the rotation frequency of the crankshaft and the time. The acceleration time and the acceleration distance are calculated necessary for reaching a certain speed.

The board computer allows also measurements of the power and the torque of the driving wheels, for first gear when the engine works with gasoline, with LPG with switched on or switched off advance processor. The programme “Dyno Run” from the menu “Dyno” is used. As a result of the measurements, the graphic relationship is obtained between the power and torque of the driving wheels and the rotation frequency. The display shows the maximum value of each of the two variables and the rotation frequency at which they are reached.

3. Experimental research

In order to determine the results of the operation of the advance processor on the engine performance, consecutive measurements were made of the acceleration time of the car from speed 0 to 100 km/h [7]. The measurements are done at initial angle of displacement of the correction curve – 6

degrees and the engine operating with: 1) gasoline; 2) LPG with switched on advance processor; 3) LPG with switched off advance processor. The plane of the correction coefficients for the gas injectors during the measurements is presented in Fig.5.

When the engine operating with LPG and the advance processor is switched on and switched off, the multiplier map is the same.

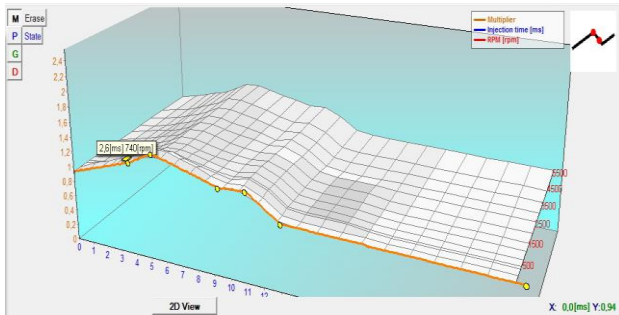


Fig. 5. 3D multiplier map

The results obtained from three consecutive measurements of each one of the three cases are average weighted and visualized on Fig. 6, 7 and 8.



Fig. 6. Vehicle speed and engine speed with gasoline

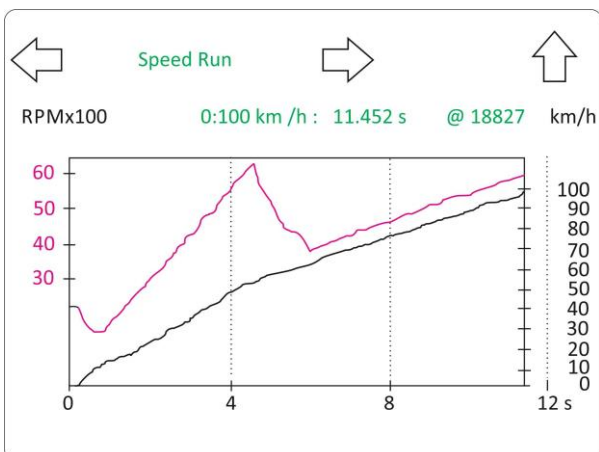


Fig. 7. Vehicle speed and engine speed with LPG and switched on advance processor.

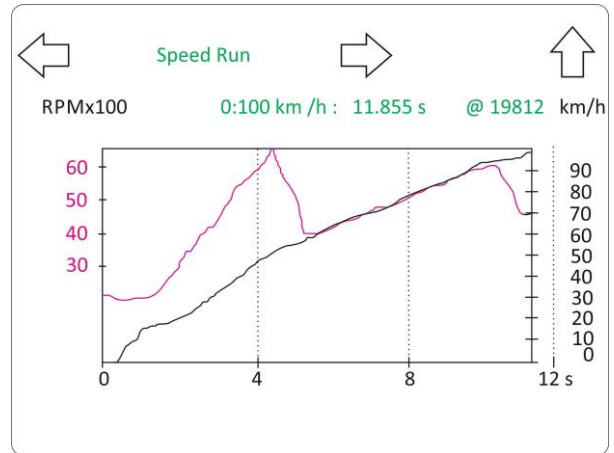


Fig. 8. Vehicle speed and engine speed with LPG and switched on advance processor.

The results from the measurements are summarized in Table 1.

Table 1. Time and distance of acceleration from 0 to 100 km/h

Fuel	Time of acceleration, [s]	Distance of acceleration, [m]
Gasoline Petrol 96 Extra Force®	11,055	185
LPG and switched on advance processor	11,452	188
LPG and switched off advance processor	11,855	198
Gasoline [8]	11,1	-

On the basis of the results obtained in the measurements, the following more significant conclusions can be made:

1) When the engine works with LPG and switched on advance processor, the acceleration time increases with 3,59% compared to the variant when the engine is working with gasoline, while the acceleration distance increases with 1.62%.

2) When the engine works with LPG and switched off advance processor, the acceleration time increases with 7.24% compared to the variant when the engine is working with gasoline, while the distance increases with 7.02%.

When measuring the power of the driving wheels and the torque, three measurements of the engine were made, while working with: 1) gasoline; 2) LPG with switched on advance processor; 3) LPG with switched off advance processor. The measurements were made on a horizontal, asphalt

road, without wind [9]. The results from the measurements are presented on Fig. 9, 10 and 11.

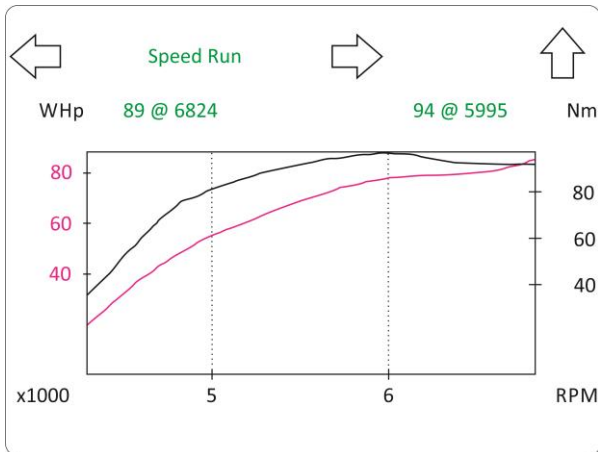


Fig. 9. Power and torque of the spark ignition engine operating with gasoline

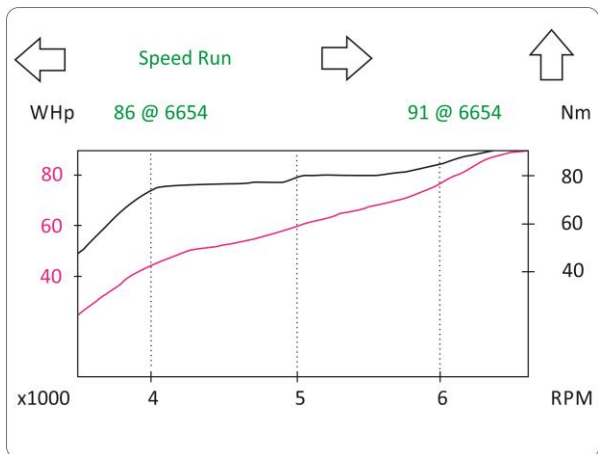


Fig. 10. Power and torque of the spark ignition engine operating with LPG and switched on advance processor.



Fig. 11. Power and torque of the spark ignition engine operating with LPG and switched off advance processor.

The results from the measurements are summarized and presented in Table 2.

Table 2. Power and torque at driving wheels

Fuel	Power, [WHP]/ [rpm]	Torque, [Nm]/ [rpm]
Gasoline Petrol 96 Extra Force®	89/6824	94/5995
LPG and switched on advance processor	86/6654	91/6654
LPG and switched off advance processor	82/6793	87/5084
Gasoline [8]	114/6500*	134/5400**

* Engine power, [HP]

** Engine torque, [Nm]

The performed measurements show the following:

1) When the engine works with LPG and switched on advance processor, the measured power of the driving wheels is lower with 3,37%, the torque – with 3,19%, compared to operating the gasoline.

2) When the engine works with LPG and switched off advance processor the acceleration time increases with 7,87% compared to the engine operating with gasoline, while the distance is increased with 7,44%.

If we take that the transmission performance is 0,85, then the maximum measured power of the driving wheel corresponds to 105 HP power of the engine.

4. Conclusion

The use of advance processor in cars with spark ignition engines without the option for automatic correction of the angle of ignition timing improves the dynamic performance of the car, and in particular the power performance of the engine. The best results with regards to time and distance of acceleration from 0 to 100 km/h are achieved when the engine operates with gasoline. The highest power and torque of the internal combustion engine is observed when it is operating with gasoline. The performed measurements show that better results are achieved of the spark-ignition engine with LPG with switched on advance processor, compared to the variant of switched off advance processor. The relatively small differences in the measured results

gives us grounds to assume that the function of the advance processor does not influence significantly the exploitation performance of the engine with LPG systems of fourth generation subjected to testing.

REFERENCES

1. Ivanov, Z. и Dimitrov, A. (2009), *LPG systems for cars*, Tehnika, Sofia.
2. <http://www.ac.com.pl/en/produkt/362/stag-300-isa2>
3. <http://www.tomasetto.com/en/products/lpg-reducers/model-at09>
4. http://www.autogaz-alex.pl/index.php?option=com_content&view=article&id=190:05-barracudaen
5. <http://www.aeb.it/Products/Products/ProdID/24/CatID/8&catid=21:injectors&Itemid=133&lang=en>
6. <http://www.infobordbg.com/>
7. Meywerk M. (2015), *Vehicle dynamics*, Publisher, John Wiley & Sons Ltd.
8. http://www.automobile-catalog.com/auta_perfl.php
9. Neykov S., Yordanov P., Katsov D., Penchev S. (2006), *Handbook for laboratory works in "Automotive Engineering part I"*, Publishing - Printing unit of the Technical University of Sofia, Branch Plovdiv, Plovdiv.

Technical University of Sofia,
Branch Plovdiv
25 Tsanko Diustabanov St,
4000 Plovdiv
Bulgaria
e-mail: kambarev@tu-plovdiv.bg



$L_p(k)$ - SOLUTION OF NONLINEAR VOLTERRA-HAMMERSTEIN INTEGRAL EQUATIONS

ATANASKA GEORGIEVA, ALBENA PAVLOVA

Abstract: In this paper we study the existence of $L_p(k)$ -solution over compact intervals of nonlinear Volterra-Hammerstein integral equations. The existence of such a solutions is based on well-known Schauder's fixed point theorem in Banach space. We give example which illustrates main result.

Key words: nonlinear Volterra-Hammerstein integral equations, Schauder's fixed point theorem, compact operator, $L_p(k)$ -solution

1. Introduction

Integral equations play a very important role in nonlinear analysis and find numerous applications in engineering, mathematical physics, economics and so on (see [1], [4], [8]). For instance, the nonlinear Volterra-Hammerstein integral equations appear in nonlinear physical phenomena such as electro-magnetic fluid dynamics, reformulation of boundary value problems with a nonlinear boundary condition (see [4]).

The existence of solutions for nonlinear Volterra-Hammerstein integral equations has been considered in many paper and books [3], [4], [7]. In this paper we find sufficient conditions for the existence of $L_p(k)$ -solutions of nonlinear Volterra-Hammerstein integral equations. In [5] are found sufficient conditions for the existence of solutions in spaces $L_p(k)$ ($1 \leq p < \infty$) of linear nonhomogeneous impulsive differential equations with an unbounded linear operator.

2. Preliminaries

Let X be a Banach spaces. By $C(X, X)$ we denote the space of all continuous operators acting in X . Let $\mathbb{R}_+ = [0, +\infty)$.

We consider the nonlinear Volterra-Hammerstein integral equation

$$x(t) = f(t) + \int_a^t V(t,s)g(s, x(s))ds \quad (1)$$

where the functions $f(\cdot) : [a, b] \rightarrow \mathbb{R}$, $g(\cdot, \cdot) : [a, b] \times \mathbb{R} \rightarrow \mathbb{R}$ and the kernel $V(\cdot, \cdot) : [a, b] \times [a, b] \rightarrow \mathbb{R}$.

We introduce the following conditions:

(H1) There exists a function $k(\cdot, \cdot) : [a, b] \times [a, b] \rightarrow \mathbb{R}_+$, positive constant M_1 such that

$$\sup_{t \in [a, b]} \int_a^t k(t,s)ds < M_1, \quad \inf_{t \in [a, b]} k(t,s) \neq 0 \quad \text{and} \quad |V(t,s)| \leq k(t,s), \quad \text{where } a \leq s \leq t \leq b.$$

(H2) There exist functions $k_1(\cdot), k_2(\cdot) : [a, b] \rightarrow \mathbb{R}_+$ for all $s \in [a, b]$ and positive constants M_2, M_3

$$\text{such that } \sup_{t \in [a, b]} k_1(t) < M_2, \quad \sup_{t \in [a, b]} \int_a^t k_2(s)ds < M_3 \quad \text{and} \quad |V(t,s)| \leq k_1(t)k_2(s), \quad a \leq s \leq t \leq b.$$

We introduce the following space

$$L_p(k) = \left\{ x(\cdot) : [a, b] \rightarrow \mathbb{R}; \sup_{t \in [a, b]} \int_a^t k(t,s)|x(s)|^p ds < \infty \right\} \quad \text{for } 1 \leq p < \infty$$

with norm

$$\|x\|_{L_p(k)} = \sup_{t \in [a,b]} \left(\int_a^t k(t,s) |x(s)|^p ds \right)^{\frac{1}{p}}.$$

Definition. The solution x of the nonlinear Volterra-Hammerstein integral equation (1) is said to be $L_p(k)$ -solution if $x \in L_p(k)$.

By using Schauder's fixed point theorem, we prove the existence of a $L_p(k)$ -solution of nonlinear Volterra-Hammerstein integral equation (1).

Theorem 1. [6] Let K be a closed and convex subset of a Banach space X . The operator $T \in C(K, K)$ and compact. Then T has a fixed point in K .

3. Existence of $L_p(k)$ -solution

Theorem 2. Let the following conditions are fulfilled:

1. The kernel $V(t, s)$ satisfies the following conditions:

1.1. The condition (H1) hold.

1.2. $\lim_{h \rightarrow 0} \sup_{t \in [a,b]} \int_a^t k(t,s) \left(\sup_{\tau \in [a,b]} |V(s+h, \tau) - V(s, \tau)|^p \right) ds = 0$ and $\sup_{\tau \in [a,b]} |V(s, \tau)| \in L_p(k)$.

2. The function $g(s, x)$ satisfies the following conditions:

2.1. There exist the functions

$G(\cdot) : [a, b] \rightarrow \mathbb{R}_+$, $\Phi(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that $|g(s, x)| \leq G(s)\Phi(|x|)$ for all $s \in [a, b]$, $x \in \mathbb{R}$, where $G \in L_q(k)$, $\frac{G}{k_3} \in L_q(k)$, $k_3(s) = \frac{1}{\inf_{t \in [a,b]} k(t, s)}$, $\Phi(|x|) \leq L|x|$ for all $x \in \mathbb{R}$ and some positive constant L , $\frac{1}{p} + \frac{1}{q} = 1$.

Moreover, assume that the function G satisfies the condition $2^{\frac{1}{q}} L M_1 \cdot \|G\|_{L_q(k)} < 1$.

2.2. There exists positive constant L_1 such that $\left| \frac{\partial g}{\partial x}(s, x) \right| \leq L_1$ for each $s \in [a, b]$, $x \in \mathbb{R}$.

3. The function $f \in L_p(k)$ and $\lim_{h \rightarrow 0} \sup_{t \in [a,b]} \int_a^t k(t,s) |f(s+h) - f(s)|^p ds = 0$.

Then the nonlinear Volterra-Hammerstein integral equation (1) has a $L_p(k)$ -solution.

Proof: Let $r > 0$ be a positive real number to be fixed later and define the closed ball B_r of $L_p(k)$ by $B_r = \{x \in L_p(k) : \|x\|_{L_p(k)} \leq r\}$. It is clear that B_r is a closed and convex subset of $L_p(k)$.

Let T be the operator defined on $L_p(k)$ by

$$Tx(t) = f(t) + \int_a^t V(t,s)g(s, x(s))ds, \quad t \in [a, b]. \quad (2)$$

First, we will prove that the operator $T : B_r \rightarrow B_r$ is continuous.

Let $x \in B_r$. Then from conditions 1.1, 2.1 and Hölder's inequality for the operator T we have

$$\begin{aligned} |Tx(s)| &\leq |f(s)| + \int_a^s |V(s, \tau)| |g(\tau, x(\tau))| d\tau \leq |f(s)| + \int_a^s k(s, \tau) G(\tau) \Phi(|x(\tau)|) d\tau \leq \\ &\leq |f(s)| + \left(\sup_{s \in [a,b]} \int_a^s k(s, \tau) |x(\tau)|^p d\tau \right)^{\frac{1}{p}} \cdot \left(\sup_{s \in [a,b]} \int_a^s k(s, \tau) G^q(\tau) L^q d\tau \right)^{\frac{1}{q}} = \\ &= |f(s)| + L \cdot \|x\|_{L_p(k)} \cdot \|G\|_{L_q(k)} \leq |f(s)| + r \cdot L \cdot \|G\|_{L_q(k)}. \end{aligned}$$

Hence

$$\begin{aligned} \|Tx\|_{L_p(k)} &= \left(\sup_{t \in [a,b]} \int_a^t k(t,s) |Tx(s)|^p ds \right)^{\frac{1}{p}} \leq \left(\sup_{t \in [a,b]} \int_a^t k(t,s) (|f(s)| + r.L.\|G\|_{L_q(k)})^p ds \right)^{\frac{1}{p}} \leq \\ &\leq 2^{\frac{p-1}{p}} \left(\left(\sup_{t \in [a,b]} \int_a^t k(t,s) |f(s)|^p ds \right)^{\frac{1}{p}} + \left(\sup_{t \in [a,b]} \int_a^t k(t,s) r^p . L^p . \|G\|_{L_p(k)}^p ds \right)^{\frac{1}{p}} \right) = \\ &= 2^{\frac{1}{q}} \left(\|f\|_{L_p(k)} + r.L.\|G\|_{L_q(k)} \left(\sup_{t \in [a,b]} \int_a^t k(t,s) ds \right)^{\frac{1}{p}} \right) \leq 2^{\frac{1}{q}} \left(\|f\|_{L_p(k)} + r.L.\|G\|_{L_q(k)} . M_1^{\frac{1}{p}} \right). \end{aligned}$$

Consequently

$$\|Tx\|_{L_p(k)} \leq 2^{\frac{1}{q}} \left(\|f\|_{L_p(k)} + r.L.M_1^{\frac{1}{p}} . \|G\|_{L_q(k)} \right). \quad (3)$$

Hence the condition $T(B_r) \subset B_r$ is satisfied for any positive real number r satisfying

$$r \geq \frac{2^{\frac{1}{q}} \|f\|_{L_p(k)}}{1 - 2^{\frac{1}{q}} L.M_1^{\frac{1}{p}} . \|G\|_{L_q(k)}}.$$

Next we prove that the operator T is continuous over B_r .

Let $\{x_n\}$ be a sequence of B_r converging to x in the norm $\|\cdot\|_{L_p(k)}$. Since $(B_r, \|\cdot\|_{L_p(k)})$ is complete, then $x \in B_r$.

From condition 2.2 and Hölder's inequality we have

$$\begin{aligned} |Tx_n(s) - Tx(s)| &\leq \int_a^s |V(s,\tau)| |g(\tau, x_n(\tau)) - g(\tau, x(\tau))| d\tau \leq \\ &\leq \int_a^s k(s,\tau) \left| \frac{\partial g}{\partial x}(s, \theta_s x_n(\tau) + (1 - \theta_s)x(\tau)) \right| |x_n(\tau) - x(\tau)| d\tau \leq \\ &\leq \int_a^s k(s,\tau) L_1 |x_n(\tau) - x(\tau)| d\tau \leq \left(\sup_{s \in [a,b]} \int_a^s k(s,\tau) |x_n(\tau) - x(\tau)|^p d\tau \right)^{\frac{1}{p}} \left(\sup_{s \in [a,b]} \int_a^s k(s,\tau) . L_1^q d\tau \right)^{\frac{1}{q}} = \\ &= L_1 . M_1^{\frac{1}{q}} . \|x_n - x\|_{L_p(k)}, \text{ where } \theta_s \in (0,1). \end{aligned}$$

Hence

$$\begin{aligned} \|Tx_n - Tx\|_{L_p(k)} &= \left(\sup_{t \in [a,b]} \int_a^t k(t,s) |Tx_n(s) - Tx(s)|^p ds \right)^{\frac{1}{p}} \leq \left(\sup_{t \in [a,b]} \int_a^t k(t,s) L_1^p . M_1^{\frac{p}{q}} . \|x_n - x\|_{L_p(k)}^p ds \right)^{\frac{1}{p}} = \\ &= L_1 . M_1^{\frac{1}{q}} . M_1^{\frac{1}{p}} . \|x_n - x\|_{L_p(k)} = L_1 . M_1 . \|x_n - x\|_{L_p(k)}. \end{aligned}$$

Then $\lim_{n \rightarrow \infty} \|Tx_n - Tx\|_{L_p(k)} = 0$ and hence the operator T is continuous over B_r .

We will prove that the operator T is compact in B_r .

We denote the set $T(B_r) = \{Tx : x \in B_r\}$. We will prove that $T(B_r)$ is a compact subset in B_r .

By using compactness criterion in the spaces $L_p(k)$ from [2] the compactness of the set $T(B_r)$ is ensured:

1. There exists positive constant C , such that $\|Tx\|_{L_p(k)} \leq C$, for all $x \in B_r$.

$$2. \lim_{h \rightarrow 0} \left(\sup_{t \in [a, b]} \int_a^t k(t, s) |Tx(s+h) - Tx(s)|^p ds \right)^{\frac{1}{p}} = 0. \quad (4)$$

From (3) we obtain that the constant $C = 2^{\frac{1}{q}} \left(\|f\|_{L_p(k)} + r.L.M_1^{\frac{1}{p}} \cdot \|G\|_{L_q(k)} \right)$.

Consequently $T(B_r)$ is uniformly bounded.

We will prove the condition (4).

Let $x \in B_r$. From condition 2.1 and Hölder's inequality we have

$$\begin{aligned} & |Tx(s+h) - Tx(s)|^p \leq \\ & \leq 2^{p-1} |f(s+h) - f(s)|^p + 2^{2(p-1)} \left(\int_a^s |V(s+h, \tau) - V(s, \tau)| |g(\tau, x(\tau))| d\tau \right)^p + \\ & + 2^{2(p-1)} \left(\int_s^{s+h} |V(s+h, \tau)| |g(\tau, x(\tau))| d\tau \right)^p \leq \\ & \leq 2^{p-1} |f(s+h) - f(s)|^p + 2^{2(p-1)} \left(\int_a^s \frac{k^{\frac{1}{p} + \frac{1}{q}}(s, \tau)}{k_3^q(\tau)} |V(s+h, \tau) - V(s, \tau)| G(\tau) L|x(\tau)| d\tau \right)^p + \\ & + 2^{2(p-1)} \left(\int_s^{s+h} k(s+h, \tau) G(\tau) L|x(\tau)| d\tau \right)^p \leq \\ & \leq 2^{p-1} |f(s+h) - f(s)|^p + \\ & + 2^{2(p-1)} \left(\int_a^s k(s, \tau) |V(s+h, \tau) - V(s, \tau)|^q \frac{G^q(\tau)}{k_3^q(\tau)} L^q d\tau \right)^{\frac{p}{q}} \cdot \left(\sup_{s \in [a, b]} \int_a^s k(s, \tau) |x(\tau)|^p d\tau \right)^{\frac{1}{p} \cdot p} + \\ & + 2^{2(p-1)} \left(\int_s^{s+h} k(s+h, \tau) G^q(\tau) L^q d\tau \right)^{\frac{p}{q}} \cdot \left(\int_s^{s+h} k(s+h, \tau) |x(\tau)|^p d\tau \right)^{\frac{1}{p} \cdot p} \leq \\ & \leq 2^{p-1} |f(s+h) - f(s)|^p + \\ & + 2^{2(p-1)} L^p \cdot \sup_{\tau \in [a, b]} |V(s+h, \tau) - V(s, \tau)|^p \cdot \|x\|_{L_p(k)}^p \cdot \left\| \frac{G}{k_3} \right\|_{L_q(k)}^p + \\ & + 2^{2(p-1)} L^p \cdot \left(\int_s^{s+h} k(s+h, \tau) G^q(\tau) d\tau \right)^{\frac{p}{q}} \cdot \|x\|_{L_p(k)}^p. \end{aligned}$$

Then

$$\begin{aligned} & \left(\sup_{t \in [a, b]} \int_a^t k(t, s) |Tx(s+h) - Tx(s)|^p ds \right)^{\frac{1}{p}} \leq \\ & \leq \left(\sup_{t \in [a, b]} \int_a^t k(t, s) \left[2^{p-1} |f(s+h) - f(s)|^p + 2^{2(p-1)} L^p \cdot \sup_{\tau \in [a, b]} |V(s+h, \tau) - V(s, \tau)|^p \cdot \|x\|_{L_p(k)}^p \cdot \left\| \frac{G}{k_3} \right\|_{L_q(k)}^p + \right. \right. \\ & \quad \left. \left. + 2^{2(p-1)} L^p \cdot \left(\int_s^{s+h} k(s+h, \tau) G^q(\tau) d\tau \right)^{\frac{p}{q}} \cdot \|x\|_{L_p(k)}^p \right] ds \right)^{\frac{1}{p}} \leq \\ & \leq 2^{\frac{p-1}{p}} \left(\sup_{t \in [a, b]} \int_a^t k(t, s) |f(s+h) - f(s)|^p ds \right)^{\frac{1}{p}} + \end{aligned}$$

$$\begin{aligned}
 &+2^{\frac{2}{q}}L.\|x\|_{L_p(k)} \cdot \left\| \frac{G}{k_3} \right\|_{L_q(k)} \left(\sup_{t \in [a,b]} \int_a^t k(t,s) \sup_{\tau \in [a,b]} |V(s+h,\tau) - V(s,\tau)|^p ds \right)^{\frac{1}{p}} + \\
 &+2^{\frac{2}{q}}L.\|x\|_{L_p(k)} \cdot \left(\sup_{t \in [a,b]} \int_a^t k(t,s) \left(\int_s^{s+h} k(s+h,\tau) G^q(\tau) d\tau \right)^{\frac{p}{q}} ds \right)^{\frac{1}{p}}.
 \end{aligned}$$

From conditions 1.2, 3 and $G \in L_q(k)$ it follows

$$\lim_{h \rightarrow 0} \left(\sup_{t \in [a,b]} \int_a^t k(t,s) |Tx(s+h) - Tx(s)|^p ds \right)^{\frac{1}{p}} = 0.$$

Hence the operator T defined by (2) is compact in B_r .

From Schauder's fixed point theorem it follows the nonlinear Volterra-Hammerstein integral equation (1) has a $L_p(k)$ -solution in B_r . \square

Theorem 3. *Let the following conditions are fulfilled:*

1. *The kernel $V(t,s)$ satisfies the following conditions:*

1.1. *The condition (H2) holds.*

1.2. $\lim_{h \rightarrow 0} \sup_{t \in [a,b]} \int_a^t k_2(s) \left(\sup_{\tau \in [a,b]} |V(s+h,\tau) - V(s,\tau)|^p \right) ds = 0$ and $\sup_{\tau \in [a,b]} |V(s,\tau)| \in L_p(k_2)$.

2. *The function $g(s,x)$ satisfies the following conditions:*

2.1. *There exist the functions*

$G(\cdot) : [a,b] \rightarrow \mathbb{R}_+$, $\Phi(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, *such that $|g(s,x)| \leq G(s)\Phi(|x|)$ for all $s \in [a,b]$,*

$x \in \mathbb{R}$, *where $G \in L_q(k_2)$, $\frac{G}{k_2} \in L_q(k_2)$, $\Phi(|x|) \leq L|x|$ for all $x \in \mathbb{R}$ and some positive constant L ,*

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Moreover, assume that the function G satisfies the condition $2^{\frac{1}{q}}L.M_2.M_3.\|G\|_{L_q(k_2)} < 1$.

2.2. *There exists positive constant L_1 such that $\left| \frac{\partial g}{\partial x}(s,x) \right| \leq L_1$ for all $s \in [a,b], x \in \mathbb{R}$.*

3. *The function $f \in L_p(k_2)$ and $\lim_{h \rightarrow 0} \sup_{t \in [a,b]} \int_a^t k_2(s) |f(s+h) - f(s)|^p ds = 0$.*

Then the nonlinear Volterra-Hammerstein integral equation (1) has a $L_p(k_2)$ -solution.

The Theorem 3 is proof analogously of Theorem 2.

4. Examples

We will give an example, which illustrate conditions of Theorem 2 and Theorem 3.

Example 1.

Consider the nonlinear Volterra-Hammerstein integral equation

$$x(t) = \frac{e^{-t}}{t+1} + \int_0^t \frac{\sin \frac{1}{8}(t+2s+1)}{(s^2+1)} e^{-s} \ln(sx(s)+1) ds, \quad t \in [0,1].$$

In this case the functions $f(t) = \frac{e^{-t}}{t+1}$, $t \in [0,1]$, kernel $V(t,s) = \frac{\sin \frac{1}{8}(t+2s+1)}{s^2+1}$ and

$$g(s,x) = e^{-s} \ln(sx+1).$$

For the functions $k(\cdot, \cdot)$, $G(\cdot)$ and $\Phi(\cdot)$ we obtain $k(t, s) = \frac{t}{8} + \frac{s}{4} + \frac{1}{8}$, $G(s) = e^{-s}$ and $\Phi(|x|) = |x|$.

For the constants M_1 , L and L_1 we obtain $M_1 = \frac{3}{8}$, $L = 1$ and $L_1 = 1$.

It is easily seen that the functions satisfy the conditions of Theorem 2.

Example 2.

Consider the nonlinear Volterra-Hammerstein integral equation

$$x(t) = t^2 e^{-t} + \int_0^t \frac{(s+t)}{4} \ln(1+t) s e^{-sx(s)} \arctg x(s) ds, \quad t \in [0, 1].$$

In this case the functions $f(t) = t^2 e^{-t}$, kernel $V(t, s) = \frac{(s+t)}{4} \ln(1+t)$ and $g(s, x) = s e^{-sx} \arctg x$.

For the functions $k_1(\cdot)$, $k_2(\cdot)$, $G(\cdot)$ and $\Phi(\cdot)$ we obtain $k_1(t) = t$, $k_2(s) = \frac{s+1}{4}$, $G(s) = s$ and $\Phi(|x|) = |x|$.

For the constants M_2 , M_3 , L and L_1 we obtain $M_2 = 1$, $M_3 = \frac{1}{2}$, $L = 1$ and $L_1 = 1 + \frac{\pi}{2}$.

It is easily seen that the functions satisfy the conditions of Theorem 3.

REFERENCES

1. Agarwal R.P., O'Regan D., Wong P.J. Y. (1999). *Positive solutions of differential, difference and integral equations*. Kluwer Academic, Dordrecht.
2. Antonevich A.B., Radyno Y.V., (1984). *Functional analysis and integral equations*. Minsk Higher School, 351 p.
3. Banas J., Darwish M. A., Olszowy L., Sadarangani K. (2015). *Existence and asymptotic behaviour of solutions of differential and integral equations in some functional spaces*. Hindawi Publishing Corporation, Journal of Functional Spaces, ID 235043.
4. Burton T.A. (1983). *Volterra integral and differential equations*, Academic Press, New York.
5. Georgieva A., *Properties of $L_p(k)$ -solutions of linear nonhomogeneous impulsive differential equations with unbounded linear operator*, Anniversary International Conference REMIA 2010, 129-135 p.
6. Karoni A. (2005). *On the existence of continuous solutions of nonlinear integral equations*. Applied Mathematics Letters (18), 299-305 p.
7. Mateknejad K., Nouri K., Mollapourasi R. (2009). *Existence of solutions for some nonlinear integral equations*, Communications in nonlinear science and numerical simulations vol.14, issue 6, 2559-2564 p.
8. O'Regan D., Meehan M. (1998). *Existence theory for nonlinear integral and integro-differential equations*, Kluwer Academic, Dordrecht.

Department of
Mathematics, Physics and Chemistry
Technical University – Sofia,
Plovdiv Branch
25 Tsanko Dyustabanov Str.
4000 Plovdiv
BULGARIA
E-mail: akosseva@gmail.com

Faculty of Mathematics and Informatics
Plovdiv University
24 Tsar Asen Str.
4000 Plovdiv
BULGARIA
E-mail: atanaska@uni-plovdiv.bg



ODD-DIMENSIONAL SPACES WITH SYMMETRIC AFFINE CONNECTION

IVAN BADEV, GEORGI ZLATANOV

Abstract: Studies odd-dimensional spaces with symmetric affine connection equipped with additional structures. Riemann spaces are determined which contain such structures.

Key words: spaces with affine connection, coordinate net, Riemann spaces, composition.

Mathematics Subject Classification (2010): 53B05, 53B99.

1. Introduction

Spaces with symmetric affine connection, Weyl space and Riemann spaces with additional structures are studied in [1, 3, 4, 5, 8]. In the studied of these spaces additional structures are define with tensors of type (1, 1). By help of n independent vector fields in [10, 11] we introduce an apparatus for the study of spaces symmetric affine connection which contain special compositions or special nets.

This apparatus is used for learning triples of compositions [2] almost paracontact structures, almost complex structures [6] and structures defined with tensors of any kind [7].

In the present work, by means of the same apparatus we study special odd-dimensional spaces endowed with a symmetric affine connection and a covariantly constant affnor. This affnor is defined by $n+1$ independent vector fields and their reciprocal covectors. The obtained results are applied to $2n+1$ dimensional Riemann spaces.

2. Basic notions

Let A_{2n+1} be a space endowed with a symmetric affine connection ∇ . The Christoffel symbols of ∇ are denoted by $\Gamma_{\alpha\beta}^{\nu}$.

Let us introduce the following notations

$$\alpha, \beta, \gamma, \sigma, \nu, \delta = 1, 2, \dots, 2n+1$$

$$i, j, s, k = 1, 2, \dots, n \quad (1)$$

$$\bar{i}, \bar{j}, \bar{s}, \bar{k} = n+1, n+2, \dots, 2n.$$

Let V_{α}^{β} be independent vector fields. The net defined by V_{α}^{β} is denoted by $\{V\}$. The

reciprocal covector fields V_{β}^{α} of the vector fields V_{α}^{β} are determined by

$$V_{\alpha}^{\beta} V_{\beta}^{\nu} = \delta_{\alpha}^{\nu} \Leftrightarrow V_{\beta}^{\nu} V_{\nu}^{\alpha} = \delta_{\beta}^{\alpha}, \quad (2)$$

where δ_{α}^{ν} is the identity affnor.

Let $\{V\}$ is the coordinate net. Then we have

[10, 11]

$$\begin{matrix} V_1^{\alpha}(1,0,0,\dots,0), V_2^{\alpha}(0,1,0,\dots,0), \dots, V_{2n+1}^{\alpha}(0,0,\dots,0,1), \\ V_{\alpha 1}(1,0,0,\dots,0), V_{\alpha 2}(0,1,0,\dots,0), \dots, V_{\alpha 2n+1}(0,0,\dots,0,1). \end{matrix} \quad (3)$$

In addition to the usual coordinates $x^{\alpha} (\alpha=1,2,\dots, n+1)$, in A_{2n+1} we introduce coordinates with respect to the coordinate net $\{V\}_{\alpha}$

which will be denoted by u^{α} .

The following derivative equations are known to be valid [9, 10]

$$\nabla_{\sigma} V_{\alpha}^{\beta} = \Gamma_{\alpha}^{\nu} \sigma V_{\nu}^{\beta}, \nabla_{\sigma} V_{\beta}^{\alpha} = -\Gamma_{\nu}^{\alpha} \sigma V_{\beta}^{\nu}. \quad (4)$$

In accordance to [10, 11], in the parameters of the coordinate net $\{V\}_{\alpha}$ we have

$$\Gamma_{\alpha}^{\beta} \sigma = \Gamma_{\alpha\alpha}^{\beta}. \quad (5)$$

In the space A_{2n+1} , we consider a composition $X_p \times X_q$ of two basic manifolds X_p and X_q , ($p+q=2n+1$), i.e. their topological product. Two positions (tangent spaces) denoted by $P(X_p)$ and $P(X_q)$ pass thought each point of the

space A_{2n+1} [3, 4]. According to [3] the coordinates u^α are adapted to the composition $X_p \times X_q$.

It is known that a composition is completely defined by the affiner field a_α^β , satisfying the condition [3, 4]

$$a_\alpha^\beta a_\beta^\nu = \delta_\alpha^\nu. \quad (6)$$

The composition $X_p \times X_q$ is of the type (c, c), or of the type (c,-), or of the type (-,c) if the positions $P(X_p)$ and $P(X_q)$, or the positions $P(X_p)$, or the positions $P(X_q)$ are translated parallelly along any line of the space A_{2n+1} [4].

Let us consider the following affiners

$$\begin{aligned} a_\alpha^\beta &= V_i^\beta V_\alpha^i + V_i^\beta V_\alpha^{\bar{i}} - V_{2n+1}^\beta V_\alpha^{2n+1} \\ b_\alpha^\beta &= V_i^\beta V_\alpha^i + V_{2n+1}^\beta V_\alpha^{2n+1} - V_{\bar{i}}^\beta V_\alpha^{\bar{i}} \\ c_\alpha^\beta &= V_{\bar{i}}^\beta V_\alpha^{\bar{i}} + V_{2n+1}^\beta V_\alpha^{2n+1} - V_i^\beta V_\alpha^i. \end{aligned} \quad (7)$$

From (2) and (7) follow the equations $a_\alpha^\beta a_\beta^\nu = b_\alpha^\beta b_\beta^\nu = c_\alpha^\beta c_\beta^\nu = \delta_\alpha^\nu$. Hence the affiners (7) are defined compositions in A_{2n+1} . Let the affiner a_α^β define the composition $X_{2n} \times X_1$ with positions $P(X_{2n})$ and $P(X_1)$. The projecting affiners of this composition are given by:

$${}^1 a_\alpha^\beta = V_i^\beta V_\alpha^i + V_{\bar{i}}^\beta V_\alpha^{\bar{i}} \quad \text{and} \quad {}^2 a_\alpha^\beta = V_{2n+1}^\beta V_\alpha^{2n+1}. \quad \text{Let the}$$

affiner b_α^β define the composition $Y_{n+1} \times Y_n$ with positions $P(Y_{n+1})$ and $P(Y_n)$. The projecting affiners of this composition are given by:

$${}^1 b_\alpha^\beta = V_i^\beta V_\alpha^i + V_{2n+1}^\beta V_\alpha^{2n+1} \quad \text{and} \quad {}^2 b_\alpha^\beta = V_{\bar{i}}^\beta V_\alpha^{\bar{i}}. \quad \text{And let}$$

the affiner c_α^β define the composition $Z_{n+1} \times Z_n$ with positions $P(Z_{n+1})$ and $P(Z_n)$. The projecting affiners of this composition are given by:

$${}^1 c_\alpha^\beta = V_{\bar{i}}^\beta V_\alpha^{\bar{i}} + V_{2n+1}^\beta V_\alpha^{2n+1} \quad \text{and} \quad {}^2 c_\alpha^\beta = V_i^\beta V_\alpha^i.$$

Obviously;

$$V_i^\alpha, V_{\bar{i}}^\alpha \in P(X_{2n}), V_{2n+1}^\alpha \in P(X_1),$$

$$V_i^\alpha, V_{2n+1}^\alpha \in P(Y_{n+1}), V_{\bar{i}}^\alpha \in P(Y_n),$$

$$V_{\bar{i}}^\alpha, V_{2n+1}^\alpha \in P(Z_{n+1}), V_i^\alpha \in P(Z_n).$$

3. Space A_{2n+1} with additional structures

Let A_{2n+1} be a space endowed with a symmetric affine connection. Consider the following affiner field

$$J_\alpha^\beta = V_i^\beta V_\alpha^i - V_{2n+1}^\beta V_\alpha^{2n+1}. \quad (8)$$

Theorem. The equation

$$\nabla_\sigma J_\alpha^\beta = 0 \quad (9)$$

it fulfilled if and only if the coefficients of derivative equations (4) satisfy

$$\bar{T}_s^i = T_s^{\bar{i}} = 0, \quad T_s^{2n+1} = \bar{T}_s^{2n+1} = 0, \quad T_{2n+1}^i = \bar{T}_{2n+1}^{\bar{i}} = 0. \quad (10)$$

Proof. According to (4) equality (9) take the form

$$\begin{aligned} T_i^\nu \sigma V_\nu^\beta V_\alpha^i - T_i^\nu \sigma V_\nu^\beta V_\alpha^{\bar{i}} - T_{2n+1}^\nu \sigma V_\nu^\beta V_\alpha^{2n+1} \\ T_{2n+1}^\nu \sigma V_\nu^\beta V_\alpha^i = 0. \end{aligned} \quad (11)$$

After contacting of (11) with $V_s^\alpha, V_{\bar{s}}^\alpha$ and V_{2n+1}^α

successively we obtain

$$\begin{aligned} T_s^\nu \sigma V_\nu^\beta - T_s^{\bar{i}} \sigma V_i^\beta + T_s^{2n+1} \sigma V_{2n+1}^\beta = 0 \\ - T_{\bar{s}}^i \sigma V_i^\beta + T_{\bar{s}}^{2n+1} \sigma V_{2n+1}^\beta = 0 \\ - T_{2n+1}^i \sigma V_i^\beta - T_{2n+1}^\nu \sigma V_\nu^\beta + T_{2n+1}^{\bar{i}} \sigma V_{\bar{i}}^\beta = 0. \end{aligned} \quad (12)$$

The last equations are equivalent of (11). Since the vector fields V_α^β are independent, equalities (12) are equivalent to conditions (10).

Corollary 1. The equation (9) it fulfilled if and only if in the parameters of the coordinate net $\{V_\alpha^\beta\}$ the

Christoffel symbols satisfy

$$\Gamma_{\sigma s}^i = \Gamma_{\sigma \bar{s}}^{\bar{i}} = 0, \Gamma_{\sigma \bar{s}}^{2n+1} = \Gamma_{\sigma s}^{2n+1} = 0, \quad (13)$$

$$\Gamma_{\sigma 2n+1}^i = \Gamma_{\sigma 2n+1}^{\bar{i}} = 0.$$

Proof. In the parameters of the coordinate net $\{V_\alpha^\beta\}$ we have (5). Hence they (10) and (13) are equivalent.

Corollary 2. The space A_{2n+1} is a space of (c, c) composition $X_{2n} \times X_1$ and (c,-) composition $Y_{n+1} \times Y_n$ and (-, c) composition $Z_n \times Z_{n+1}$ if a is satisfied (9).

Proof. Follows from [3, 4] and (13).

Definition. The one-dimensional generalized composition $Y_{n+1} \times Z_{n+1}$ will be called of type

(c, c) if the positions $P(Y_{n+1})$ and $P(Z_{n+1})$ are translated parallelly along any line of space A_{2n+1} .

Corollary 3. The one-dimensional generalized composition $Y_{n+1} \times Z_{n+1}$ is of type (c, c) if is fulfilled (9).

Proof. Followed of [3, 4] and (13).

4. Riemann space V_{2n+1} with additional structures

Let V_{2n+1} be a Riemann space with metric tensor $g_{\alpha\beta}$ and connection of Levi-Civita. The fundamental equation is $\nabla_{\sigma} g_{\alpha\beta} = 0$.

Let vector fields V_{α}^{β} satisfies the equations $g_{\alpha\beta} V_{i}^{\alpha} V^{\beta} = 0$, $g_{\alpha\beta} V_{i}^{\alpha} V^{\beta} = 0$ and

$g_{\alpha\beta} V_{i}^{\alpha} V^{\beta} = 0$. In parameters of coordinate net

$\{V_{\alpha}\}$, the matrix of tensor $g_{\alpha\beta}$ is

$$(g_{\alpha\beta}) = \begin{pmatrix} g_{ij} & 0 & 0 \\ 0 & g_{ij} & 0 \\ 0 & 0 & g_{2n+1, 2n+1} \end{pmatrix}.$$

Let the affinor (8) satisfy the equation (9). The coefficients of connection of Levi-Civita, in parameters of coordinate net $\{V_{\alpha}\}$ satisfy (13).

According to [3, 4] the compositions $Y_{n+1} \times Y_n$ and $Z_{n+1} \times Z_n$ will be of type (c, c) . From equalities

$$\nabla_s g_{ij} = 0, \quad \nabla_{2n+1} g_{ij} = 0, \quad \nabla_s g_{ij} = 0,$$

$$\nabla_{2n+1} g_{ij} = 0, \quad \nabla_s g_{2n+1, 2n+1} = 0, \quad \nabla_s g_{2n+1, 2n+1} = 0$$

and (13) we obtain, that the linear element of the Riemann space V_{2n+1} in coordinate net $\{V_{\alpha}\}$ is

$$ds^2 = g_{ij}(u^s) du^i du^j + g_{ij}(u^s) du^i du^j +$$

$$g_{2n+1, 2n+1}(u^{2n+1})(du^{2n+1})^2. \quad (15)$$

Therefore the Riemann space V_{2n+1} is a space of $Y_n \times Z_n \times X_1$ compositions and the positions $P(Y_n)$, $P(Z_n)$ and $P(X_1)$ are mutually orthogonal. The positions $P(Y_n)$, $P(Z_n)$ and $P(X_1)$ are translated parallelly along any line of V_{2n+1} , then $Y_n \times Z_n \times X_1$ is a Cartesian composition.

The Riemann space V_{2n+1} , which allows Cartesian compositions $Y_n \times Z_n \times X_1$ with mutually perpendicular positions $P(Y_n)$, $P(Z_n)$ and $P(X_1)$

is compatible enough. His linear element is the sum of the line elements that determine the internal geometries of these positions. Between parallelly positions exist isometric accordance.

The projecting affinors of composition $Y_n \times Z_n \times X_1$ are given by: ${}^1d_{\alpha}^{\beta} = V_{i}^{\beta} V_{\alpha}^i$,

$${}^2d_{\alpha}^{\beta} = V_{i}^{\beta} V_{\alpha}^i \text{ and } {}^3d_{\alpha}^{\beta} = V_{2n+1}^{\beta} V_{\alpha}^{2n+1}.$$

Obviously $V_{i}^{\alpha} \in P(Y_n)$, $V_{i}^{\alpha} \in P(Z_n)$ and

$$V_{2n+1}^{\alpha} \in P(X_1).$$

For arbitrary vector V^{α} we have

$$V^{\alpha} = {}^1d_{\beta}^{\alpha} V^{\beta} + {}^2d_{\beta}^{\alpha} V^{\beta} + {}^3d_{\beta}^{\alpha} V^{\beta} =$$

$${}^1V^{\alpha} + {}^2V^{\alpha} + {}^3V^{\alpha}, \text{ when } {}^1d_{\beta}^{\alpha} V^{\beta} = {}^1V^{\alpha} \in P(Y_n),$$

$${}^2d_{\beta}^{\alpha} V^{\beta} = {}^2V^{\alpha} \in P(Z_n), {}^3d_{\beta}^{\alpha} V^{\beta} = {}^3V^{\alpha} \in P(X_1).$$

REFERENCES

1. Adati T., Miyaza T., (1977), On paracontact Riemann manifolds. *TRU Math.*, 13(2), 27-39.
2. Ajeti M., Teofilova M., Zlatanov G., (2011), Triads of compositions in an even-dimensional space with a symmetric affine connection. *Tensor, N.S.*, 73(2), 171-187.
3. Norden A., P., (1963), Spaces of Cartesian compositions. *Izv. Vyssh. Uchebn. Zaved. Math.*, 4, 117-125 (in Russian).
4. Norden A., P., Timofeev G. N., (1972), Invariant criteria for special compositions of multidimensional space. *Izv. Vyssh. Uchebn. Zaved. Math.*, 8, 81-89 (in Russian).
5. Sasaki S., (1980), On paracontact Riemann manifolds *TRU Math.*, 16(2), 75-86.
6. Teofilova M., Zlatanov G., (2013), Odd-dimensional Riemann spaces with almost contact and almost paracontact structures. *Adv. Math. Sci. J.* 2, 25-35.
7. Teofilova M., Zlatanov G., (2014), Special spaces with a symmetric affine connection. *Tensor, N.S.*, 75(2), 154-165.
8. Timofeev G. N., (1976), Invariant characteristics of special compositions in Weyl spaces. *Izv. Vyssh. Uchebn. Zaved. Math.*, 1, 87-95 (in Russian).
9. Tsareva B., Zlatanov G., (1996), Geometry of the nets in equiaffine spaces. *J. Geom.*, 55, 192-201.
10. Zlatanov G. (2002), Compositions generated by special nets in affinely connected spaces. *Serdica Math. J.* 1001-1012.
11. Zlatanov G. (2011), Special compositions in affinely connected spaces without a torsion. *Serdica Math. J.* 37, 211-220.

Badev Ivan
Technical University-Sofia branch Plovdiv
25 Tsanko Djustabanov str
4000 Plovdiv Bulgaria
E-mail: ivanbadev@abv.bg

Zlatanov Georgi
Faculty of Mathematics and Informatics,
Plovdiv University,
236 Bulgaria Blvd., 4003 Plovdiv Bulgaria
E-mail: zlatanovg@gmail.com



CARTESIAN GENERALIZED COMPOSITIONS IN SPACES WITH A SYMMETRIC AFFINE CONNECTION

IVAN BADEV, MARTA TEOFILOVA, GEORGI ZLATANOV

Abstract: An approach for generalization of the term composition is introduced. Cartesian generalized compositions in spaces with a symmetric affine connection are studied. Special characteristics of the spaces containing such compositions are obtained. Riemannian spaces endowed with special Cartesian generalized compositions are examined.

MSC (2010): 53B05, 53B99

Key words: spaces with affine connections, Riemannian spaces, Cartesian compositions, generalized compositions

1. Introduction

In [7, 8], an apparatus for the study of spaces with a symmetric affine connection containing special compositions or nets is introduced. This apparatus is constructed by help of independent vector fields. The same apparatus is applied to the study of triples of compositions in [1], four-dimensional manifolds with additional structures in [2], and other tensor structures in [5].

In this paper, we introduce a generalization of the term composition. By help of the apparatus given in [7, 8], we explore Cartesian generalized compositions in spaces with a symmetric affine connection. In the parameters of the special coordinate net we obtain the metric tensor of a Riemannian space endowed with special Cartesian generalized compositions.

2. Basic Notions

Let A_N be a space with a symmetric affine connection. Consider r differentiable manifolds $X_{n_a} \subset A_N$, where by n_a we denote the dimension of the manifold X_{n_a} . Further, consider the manifolds $Y_{m_1} = X_{n_1} \cup X_{n_r}$, $Y_{m_2} = X_{n_2} \cup X_{n_r}, \dots$, $Y_{m_{r-1}} = X_{n_{r-1}} \cup X_{n_r}$, where $n_1 + n_2 + \dots + n_r \leq N$. We will say that the manifolds $Y_{m_1}, Y_{m_2}, \dots, Y_{m_{r-1}}$ define an m_{r-1} dimensional generalized composition of order n_r which will be denoted by $Y_{m_1} \cup Y_{m_2} \cup \dots \cup Y_{m_{r-1}}$. If $n_r = 0$ the manifolds

$Y_{m_1}, Y_{m_2}, \dots, Y_{m_{r-1}}$ define a usual composition in A_N . By $P(Y_{m_1}), P(Y_{m_2}), \dots, P(Y_{m_{r-1}})$ we will denote the positions (tangent spaces) of the manifolds $Y_{m_1}, Y_{m_2}, \dots, Y_{m_{r-1}}$, respectively.

Definition 1. The generalized composition $Y_{m_1} \cup Y_{m_2} \cup \dots \cup Y_{m_{r-1}}$ will be called *Cartesian* if the positions $P(Y_{m_1}), P(Y_{m_2}), \dots, P(Y_{m_{r-1}})$ are translated parallelly along any line in the space A_N .

Further we will use the following notations:

$$\alpha, \beta, \gamma, \delta, \sigma, \nu = 1, 2, \dots, N; \quad (1)$$

$$i, j, k, s = 1, 2, \dots, m; \quad (2)$$

$$\bar{i}, \bar{j}, \bar{k}, \bar{s} = m+1, m+2, \dots, m+n; \quad (3)$$

$$a, b, c, d = m+n+1, m+n+2, \dots, m+n+p = N.$$

Let v_σ^α ($\alpha, \sigma = 1, 2, \dots, N$) be independent vector fields. The reciprocal covector fields to v_σ^α are denoted by v_α^σ and are determined by the following equations

$$v_\sigma^\alpha v_\beta^\sigma = \delta_\beta^\alpha \Leftrightarrow v_\beta^\sigma v_\sigma^\alpha = \delta_\beta^\alpha, \quad (4)$$

where δ_β^α is the identity affinor. The independent vector fields v_σ^α determine one-parameter families of lines v_α which define a net $\{v\}$ in A_N .

Let us choose the net $\{v\}$ for the coordinate net. Then, according to [7, 8] and (4), we have

$$\begin{aligned} &v_1^\alpha(1, 0, \dots, 0), v_2^\alpha(0, 1, \dots, 0), \dots, v_N^\alpha(0, 0, \dots, 1); \\ &v_\alpha^1(1, 0, \dots, 0), v_\alpha^2(0, 1, \dots, 0), \dots, v_\alpha^N(0, 0, \dots, 1). \end{aligned} \quad (5)$$

In addition to the arbitrary coordinate net x^α ($\alpha=1, 2, \dots, N$) in A_N we introduce the coordinates u^α with respect to the net $\{v\}$.

The coefficients of the affine connection in A_N and its rule of covariant differentiation are denoted by $\Gamma_{\alpha\beta}^\sigma$ and ∇ , respectively. According to [6, 7, 8], the following derivative equations hold

$$\nabla_\sigma v_\alpha^\beta = \overset{v}{T}_{\alpha v}^\sigma v_\alpha^\beta, \quad \nabla_\sigma v_\beta^\alpha = -\overset{v}{T}_{\sigma v}^\alpha v_\beta^\alpha. \quad (6)$$

In the parameters of the coordinate net $\{v\}$ we have [6, 7, 8]

$$\overset{\beta}{T}_\sigma = \Gamma_{\sigma\alpha}^\beta. \quad (7)$$

3. Two-dimensional Cartesian generalized compositions of order p

Let us consider the following affinors

$$\begin{aligned} a_\alpha^\beta &= v_i^\beta v_\alpha^i + v_a^\beta v_\alpha^a - v_{\bar{i}}^\beta v_\alpha^{\bar{i}}, \\ b_\alpha^\beta &= v_{\bar{i}}^\beta v_\alpha^{\bar{i}} + v_a^\beta v_\alpha^a - v_i^\beta v_\alpha^i. \end{aligned} \quad (8)$$

Since $a_\alpha^\beta a_\beta^\sigma = \delta_\alpha^\sigma$, according to [3, 4], the affinor a_α^β defines the composition $X_{m+p} \times X_n$. Let $P(X_{m+p})$ and $P(X_n)$ be the positions of basic manifolds X_{m+p} and X_n , respectively. According to [7, 8], the projecting affinors of the composition $X_{m+p} \times X_n$ are given by

$${}^1a_\alpha^\beta = v_i^\beta v_\alpha^i + v_a^\beta v_\alpha^a, \quad {}^2a_\alpha^\beta = v_{\bar{i}}^\beta v_\alpha^{\bar{i}}. \quad (9)$$

Obviously $v_i^\alpha, v_a^\alpha \in P(X_{m+p}), v_{\bar{i}}^\alpha \in P(X_n)$. If v^α is an arbitrary vector field then

$$v^\alpha = {}^1a_\beta^\alpha v^\beta + {}^2a_\beta^\alpha v^\beta = V^{\alpha 1} + V^{\alpha 2}, \quad \text{where}$$

$$V^{\alpha 1} = {}^1a_\beta^\alpha v^\beta \in P(X_{m+p}) \quad \text{and} \quad V^{\alpha 2} = {}^2a_\beta^\alpha v^\beta \in P(X_n).$$

Analogously, the affinor b_α^β defines the composition $X_{n+p} \times X_m$ with positions $P(X_{n+p})$ and $P(X_m)$ and projecting affinors

$${}^1b_\alpha^\beta = v_{\bar{i}}^\beta v_\alpha^{\bar{i}} + v_a^\beta v_\alpha^a, \quad {}^2b_\alpha^\beta = v_i^\beta v_\alpha^i. \quad (10)$$

Obviously $v_{\bar{i}}^\alpha, v_a^\alpha \in P(X_{n+p}), v_i^\alpha \in P(X_m)$.

Theorem 1. *The two-dimensional generalized composition $X_{m+p} \cup X_{n+p}$ of order p is Cartesian if only if the coefficients of the derivative equations (6) satisfy the following conditions*

$$\overset{\bar{i}}{T}_j \sigma = \overset{\bar{i}}{T}_a \sigma = \overset{i}{T}_j \sigma = \overset{i}{T}_a \sigma = 0. \quad (11)$$

Proof. The generalized composition $X_{m+p} \cup X_{n+p}$ is Cartesian if only if the positions $P(X_{m+p})$ and $P(X_{n+p})$ are translated parallelly along any line in the space A_N . According to [3, 4, 7, 8], the last condition holds if only if the projecting affinors ${}^1a_\alpha^\beta$ and ${}^1b_\alpha^\beta$ satisfy the following equalities

$${}^1a_\delta^\alpha \nabla_\sigma {}^1a_\alpha^\beta = 0, \quad {}^1b_\delta^\alpha \nabla_\sigma {}^1b_\alpha^\beta = 0. \quad (12)$$

By (6) and (9), the first equality of (12) takes the form

$$\begin{aligned} &(v_j^\alpha v_\delta^j + v_b^\alpha v_\delta^b) (\overset{v}{T}_i^\sigma v_\alpha^{\beta i} - \overset{i}{T}_\sigma v_\alpha^{\beta i} v_\alpha^v) \\ &+ \overset{v}{T}_\sigma v_\alpha^{\beta i} v_\alpha^a - \overset{a}{T}_\sigma v_\alpha^{\beta i} v_\alpha^v = 0. \end{aligned} \quad (13)$$

Hence, by considering (4), from (13) we get

$$\begin{aligned} &\overset{v}{T}_j^\sigma v_\alpha^{\beta j} v_\delta^j - \overset{i}{T}_\sigma v_\alpha^{\beta j} v_\delta^j - \overset{a}{T}_j^\sigma v_\alpha^{\beta j} v_\delta^j \\ &- \overset{i}{T}_\sigma v_\alpha^{\beta b} v_\delta^b + \overset{v}{T}_\sigma v_\alpha^{\beta b} v_\delta^b - \overset{a}{T}_b^\sigma v_\alpha^{\beta b} v_\delta^b = 0. \end{aligned} \quad (14)$$

Since the covector fields v_β^α are independent (14) is equivalent to the equations:

$$\begin{aligned} \overset{v}{T}_j \sigma v^\beta - \overset{i}{T}_j \sigma v^\beta - \overset{a}{T}_j \sigma v^\beta &= 0, \\ \overset{i}{T}_b \sigma v^\beta - \overset{v}{T}_b \sigma v^\beta + \overset{a}{T}_b \sigma v^\beta &= 0. \end{aligned} \quad (15)$$

The vector fields v^β are independent too.

Therefore, equalities (15) are equivalent to $\overset{\bar{i}}{T}_j \sigma = \overset{\bar{i}}{T}_a \sigma = 0$.

Reasoning analogously, the second equation of (12) is equivalent to $\overset{i}{T}_{\bar{j}} \sigma = \overset{i}{T}_a \sigma = 0$ which proves the statement. ■

From (7) and (11) we obtain

Corollary 1. *In the parameters of the coordinate net equations (11) take the form*

$$\Gamma_{\sigma j}^{\bar{i}} = \Gamma_{\sigma a}^{\bar{i}} = \Gamma_{\sigma \bar{j}}^i = \Gamma_{\sigma a}^i = 0. \quad (16)$$

Application 1. Let us consider a Riemannian space V_N with metric tensor $g_{\alpha\beta}$ and Levi-Civita connection ∇ . Let the generalized composition $X_{m+p} \cup X_{n+p}$ be Cartesian. The vector fields v^β are chosen so that $g_{\alpha\beta} v_i^\alpha v_j^\beta = g_{\alpha\beta} v_i^\alpha v_a^\beta = g_{\alpha\beta} v_{\bar{i}}^\alpha v_a^\beta = 0$. In the parameters of the coordinate net $\{v\}$ the matrix of the tensor $g_{\alpha\beta}$ is given by

$$(g_{\alpha\beta}) = \begin{pmatrix} g_{ij} & 0 & 0 \\ 0 & g_{\bar{i}\bar{j}} & 0 \\ 0 & 0 & g_{ab} \end{pmatrix}. \quad (17)$$

By (16), (17) and the fundamental equation

$$\nabla_\sigma g_{\alpha\beta} = 0 \quad (18)$$

in the parameters of the coordinate net $\{v\}$ we

obtain $g_{ij} = g_{ij}(u^s)$, $g_{\bar{i}\bar{j}} = g_{\bar{i}\bar{j}}(u^{\bar{s}})$, $g_{ab} = g_{ab}(u^\alpha)$.

Further, let us use the notations (1), (2) and

$$\begin{aligned} a, b, c, d &= m+n+1, m+n+2, \dots, \\ m+n+p &= N-1. \end{aligned} \quad (19)$$

Consider the affinors

$$\begin{aligned} c_\alpha^\beta &= v_i^\beta v_\alpha^i + v_a^\beta v_\alpha^a - v_{\bar{i}}^\beta v_\alpha^{\bar{i}} - v_N^\beta v_\alpha^N, \\ d_\alpha^\beta &= v_{\bar{i}}^\beta v_\alpha^{\bar{i}} + v_a^\beta v_\alpha^a - v_i^\beta v_\alpha^i - v_N^\beta v_\alpha^N. \end{aligned} \quad (20)$$

As $c_\alpha^\beta c_\beta^\sigma = \delta_\alpha^\sigma$ the affiner c_α^β defines the composition $X_{m+p} \times X_{n+1}$ [3, 4]. Let $P(X_{m+p})$ and $P(X_{n+1})$ be the positions of the basic manifolds X_{m+p} and X_{n+1} , respectively. According to [7, 8] and (9), the projecting affinors of the composition $X_{m+p} \times X_{n+1}$ are given by

$${}^1c_\alpha^\beta = {}^1a_\alpha^\beta, \quad {}^2c_\alpha^\beta = {}^2a_\alpha^\beta + v_N^\beta v_\alpha^N. \quad (21)$$

Analogously, the affiner d_α^β defines the composition $X_{n+p} \times X_{m+1}$ with positions $P(X_{n+p})$ and $P(X_{m+1})$ and projecting affinors given by

$${}^1d_\alpha^\beta = {}^1b_\alpha^\beta, \quad {}^2d_\alpha^\beta = {}^2b_\alpha^\beta + v_N^\beta v_\alpha^N. \quad (22)$$

Let us prove the following

Theorem 2. *The two-dimensional generalized composition $X_{m+p} \cup X_{n+p}$ of order p is Cartesian if only if the coefficients of the derivative equations (6) satisfy the following conditions*

$$\overset{\bar{i}}{T}_j \sigma = \overset{\bar{i}}{T}_a \sigma = \overset{i}{T}_{\bar{j}} \sigma = \overset{i}{T}_a \sigma = \overset{N}{T}_j \sigma = \overset{N}{T}_{\bar{j}} \sigma = \overset{N}{T}_a \sigma = 0. \quad (23)$$

Proof. According to [3, 4, 7, 8], the positions $P(X_{m+p})$ and $P(X_{n+p})$ are translated parallelly along any line in the space A_N if only if the projecting affinors ${}^1c_\alpha^\beta$ and ${}^1d_\alpha^\beta$ satisfy the conditions

$${}^1c_\delta^\alpha \nabla_\sigma {}^1c_\alpha^\beta = 0, \quad {}^1d_\delta^\alpha \nabla_\sigma {}^1d_\alpha^\beta = 0. \quad (24)$$

By (6), (9) and (21) the first equality of (24) takes the form (14). Since the covector fields v_β^α are independent, equation (14) is equivalent to equations (15). As the vector fields v_α^β are independent it follows that equalities (15) are equivalent to the following conditions $\overset{\bar{i}}{T}_j \sigma = \overset{\bar{i}}{T}_a \sigma = \overset{N}{T}_j \sigma = \overset{N}{T}_a \sigma = 0$. Reasoning analogously, the second equation of (24) is equivalent to the conditions $\overset{i}{T}_{\bar{j}} \sigma = \overset{i}{T}_a \sigma = \overset{N}{T}_{\bar{j}} \sigma = \overset{N}{T}_a \sigma = 0$ which completes the proof of the statement. ■

By (7) and (23) we obtain the following

Corollary 2. In the parameters of coordinate net $\{v\}_\alpha$ equations (23) take the form

$$\begin{aligned}\Gamma_{\sigma j}^{\bar{i}} &= \Gamma_{\sigma a}^{\bar{i}} = \Gamma_{\sigma \bar{j}}^i = \Gamma_{\sigma a}^i = \Gamma_{\sigma j}^N \\ &= \Gamma_{\sigma \bar{j}}^N = \Gamma_{\sigma a}^N = 0.\end{aligned}\quad (25)$$

Application 2. We consider the Riemannian space $V_N(g_{\alpha\beta})$ and the Cartesian generalized composition $X_{m+p} \cup X_{n+p}$. The vector fields v_α^β are chosen so that

$$\begin{aligned}g_{\alpha\beta} v_i^\alpha v_{\bar{i}}^\beta &= g_{\alpha\beta} v_i^\alpha v_a^\beta = g_{\alpha\beta} v_{\bar{i}}^\alpha v_a^\beta \\ &= g_{\alpha\beta} v_i^\alpha v_N^\beta = g_{\alpha\beta} v_{\bar{i}}^\alpha v_N^\beta = g_{\alpha\beta} v_a^\alpha v_N^\beta = 0.\end{aligned}$$

In the parameters of the coordinate net $\{v\}_\alpha$ the matrix of the tensor $g_{\alpha\beta}$ is given by

$$(g_{\alpha\beta}) = \begin{pmatrix} g_{ij} & 0 & 0 & 0 \\ 0 & g_{\bar{i}\bar{j}} & 0 & 0 \\ 0 & 0 & g_{ab} & 0 \\ 0 & 0 & 0 & g_{NN} \end{pmatrix}. \quad (26)$$

Having in mind (18), (25) and (26), in the parameters of coordinate net $\{v\}_\alpha$ we obtain

$$g_{ij} = g_{ij}(u^s, u^N), \quad g_{\bar{i}\bar{j}} = g_{\bar{i}\bar{j}}(u^{\bar{s}}, u^N), \quad g_{ab} = g_{ab}(u^\alpha)$$

and $g_{NN} = g_{NN}(u^N)$.

4. Four-dimensional Cartesian generalized compositions of first order

Let A_{2n+1} be a space with a symmetric affine connection. Further we will use the notations

$$\begin{aligned}i, j, s, k &= 1, 2, \dots, m; \\ \bar{i}, \bar{j}, \bar{k}, \bar{s} &= m+1, m+2, \dots, m+n; \\ a, b, c &= m+n+1, \dots, m+n+p; \\ \bar{a}, \bar{b}, \bar{c} &= m+n+p+1, \dots, m+n+p+q = 2n.\end{aligned}\quad (27)$$

Let us consider the affinors

$$\begin{aligned}{}^1J_\alpha^\beta &= v_i^\beta v_\alpha + v_{2n+1}^\beta v_\alpha - v_{\bar{i}}^\beta v_\alpha - v_a^\beta v_\alpha - v_{\bar{a}}^\beta v_\alpha, \\ {}^2J_\alpha^\beta &= v_{\bar{i}}^\beta v_\alpha + v_{2n+1}^\beta v_\alpha - v_i^\beta v_\alpha - v_a^\beta v_\alpha - v_{\bar{a}}^\beta v_\alpha, \\ {}^3J_\alpha^\beta &= v_a^\beta v_\alpha + v_{2n+1}^\beta v_\alpha - v_i^\beta v_\alpha - v_{\bar{i}}^\beta v_\alpha - v_{\bar{a}}^\beta v_\alpha, \\ {}^4J_\alpha^\beta &= v_{\bar{a}}^\beta v_\alpha + v_{2n+1}^\beta v_\alpha - v_i^\beta v_\alpha - v_{\bar{i}}^\beta v_\alpha - v_a^\beta v_\alpha.\end{aligned}\quad (28)$$

Since ${}^1J_\alpha^\beta {}^1J_\beta^\sigma = \delta_\alpha^\sigma$, according to [3, 4], the affinor ${}^1J_\alpha^\beta$ defines the composition $X_{m+1} \times X_{n+p+q}$. Let $P(X_{m+1})$ and $P(X_{n+p+q})$ be the positions of the basic manifolds X_{m+1} and X_{n+p+q} , respectively. Obviously $v_i^\alpha, v_{2n+1}^\alpha \in P(X_{m+1})$ and $v_{\bar{i}}^\alpha, v_a^\alpha, v_{\bar{a}}^\alpha \in P(X_{n+p+q})$. According to [7, 8], the projecting affinors have the form

$$\begin{aligned}{}^1J_\alpha^\beta &= v_i^\beta v_\alpha + v_{2n+1}^\beta v_\alpha, \\ {}^2J_\alpha^\beta &= v_{\bar{i}}^\beta v_\alpha + v_a^\beta v_\alpha + v_{\bar{a}}^\beta v_\alpha.\end{aligned}\quad (29)$$

Similarly, the affinors ${}^2J_\alpha^\beta, {}^3J_\alpha^\beta, {}^4J_\alpha^\beta$ define the compositions $X_{n+1} \times X_{m+p+q}, X_{p+1} \times X_{m+n+q}, X_{q+1} \times X_{m+n+p}$, respectively, with positions $P(X_{n+1}), P(X_{m+p+q}); P(X_{p+1}), P(X_{m+n+q}); P(X_{q+1}), P(X_{m+n+p})$. The projecting affinors of these compositions are given by

$$\begin{aligned}{}^2J_\alpha^\beta &= v_{\bar{i}}^\beta v_\alpha + v_{2n+1}^\beta v_\alpha, \\ {}^2J_\alpha^\beta &= v_i^\beta v_\alpha + v_a^\beta v_\alpha + v_{\bar{a}}^\beta v_\alpha, \\ {}^3J_\alpha^\beta &= v_a^\beta v_\alpha + v_{2n+1}^\beta v_\alpha, \\ {}^3J_\alpha^\beta &= v_i^\beta v_\alpha + v_{\bar{i}}^\beta v_\alpha + v_{\bar{a}}^\beta v_\alpha, \\ {}^4J_\alpha^\beta &= v_{\bar{a}}^\beta v_\alpha + v_{2n+1}^\beta v_\alpha, \\ {}^4J_\alpha^\beta &= v_i^\beta v_\alpha + v_{\bar{i}}^\beta v_\alpha + v_a^\beta v_\alpha.\end{aligned}\quad (30)$$

Let us consider the four-dimensional generalized composition of first order $X_{m+1} \cup X_{n+1} \cup X_{p+1} \cup X_{q+1}$ with positions $P(X_{m+1}), P(X_{n+1}), P(X_{p+1}), P(X_{q+1})$. We prove the following

Theorem 3. The four-dimensional generalized composition $X_{m+1} \cup X_{n+1} \cup X_{p+1} \cup X_{q+1}$ is Cartesian if only if the coefficients of the derivative equations (6) satisfy the conditions

$$\begin{aligned}
 \bar{i} \Gamma_j^\sigma &= \bar{i} \Gamma_j^\sigma = \bar{i} \Gamma_j^\sigma = 0, \\
 \bar{i} \Gamma_j^\sigma &= \bar{i} \Gamma_j^\sigma = \bar{i} \Gamma_j^\sigma = 0, \\
 \bar{i} \Gamma_j^\sigma &= \bar{i} \Gamma_j^\sigma = \bar{i} \Gamma_j^\sigma = 0, \\
 \bar{i} \Gamma_j^\sigma &= \bar{i} \Gamma_j^\sigma = \bar{i} \Gamma_j^\sigma = 0, \\
 \bar{i} \Gamma_j^\sigma &= \bar{i} \Gamma_j^\sigma = \bar{i} \Gamma_j^\sigma = 0, \\
 \bar{i} \Gamma_j^\sigma &= \bar{i} \Gamma_j^\sigma = \bar{i} \Gamma_j^\sigma = 0.
 \end{aligned} \tag{31}$$

Proof. The generalized composition $X_{m+1} \cup X_{n+1} \cup X_{p+1} \cup X_{q+1}$ is Cartesian if only if the positions $P(X_{m+1})$, $P(X_{n+1})$, $P(X_{p+1})$ and $P(X_{q+1})$ are translated parallelly along any line in the space A_{2n+1} . According to [3, 4, 7, 8], the last condition holds if only if the projecting affinors ${}^1J_\alpha^\beta$, ${}^2J_\alpha^\beta$, ${}^3J_\alpha^\beta$ and ${}^4J_\alpha^\beta$ satisfy

$$\begin{aligned}
 {}^1J_\delta^\alpha \nabla_\sigma {}^1J_\alpha^\beta &= 0, & {}^2J_\delta^\alpha \nabla_\sigma {}^2J_\alpha^\beta &= 0, \\
 {}^3J_\delta^\alpha \nabla_\sigma {}^3J_\alpha^\beta &= 0, & {}^4J_\delta^\alpha \nabla_\sigma {}^4J_\alpha^\beta &= 0.
 \end{aligned} \tag{32}$$

According to (6) and (29), the first equality of (32) takes the form

$$\begin{aligned}
 (v^\alpha v_\delta + v_{2n+1}^\alpha v_\delta) (\bar{i} \Gamma_j^\sigma v^\beta v_\alpha - \bar{i} \Gamma_j^\sigma v^\beta v_\alpha) \\
 + \bar{i} \Gamma_{2n+1}^\sigma v^\beta v_\alpha - \bar{i} \Gamma_{2n+1}^\sigma v^\beta v_\alpha = 0.
 \end{aligned} \tag{33}$$

Since the covector fields v_β are independent (33) is equivalent to the equations

$$\begin{aligned}
 \bar{i} \Gamma_j^\sigma v^\beta - \bar{i} \Gamma_j^\sigma v^\beta - \bar{i} \Gamma_j^\sigma v^\beta &= 0, \\
 \bar{i} \Gamma_{2n+1}^\sigma v^\beta - \bar{i} \Gamma_{2n+1}^\sigma v^\beta + \bar{i} \Gamma_{2n+1}^\sigma v^\beta &= 0.
 \end{aligned} \tag{34}$$

Since the vector fields v^β are independent it follows that equations (34) are equivalent to

$$\bar{i} \Gamma_j^\sigma = \bar{i} \Gamma_j^\sigma = \bar{i} \Gamma_j^\sigma = 0, \quad \bar{i} \Gamma_{2n+1}^\sigma = \bar{i} \Gamma_{2n+1}^\sigma = \bar{i} \Gamma_{2n+1}^\sigma = 0.$$

Analogously, one can prove that the rest of equations in (32) are equivalent to the rest of the equations (31).

Corollary 3. *In the parameters of the coordinate net $\{v\}_\alpha$ equalities (31) take the form*

$$\begin{aligned}
 \Gamma_{\sigma j}^{\bar{i}} &= \Gamma_{\sigma j}^a = \Gamma_{\sigma j}^{\bar{a}} = 0, \\
 \Gamma_{\sigma \bar{j}}^i &= \Gamma_{\sigma \bar{j}}^a = \Gamma_{\sigma \bar{j}}^{\bar{a}} = 0, \\
 \Gamma_{\sigma a}^i &= \Gamma_{\sigma a}^{\bar{i}} = \Gamma_{\sigma a}^{\bar{a}} = 0, \\
 \Gamma_{\sigma \bar{a}}^i &= \Gamma_{\sigma \bar{a}}^{\bar{i}} = \Gamma_{\sigma \bar{a}}^i = 0, \\
 \Gamma_{\sigma 2n+1}^i &= \Gamma_{\sigma 2n+1}^{\bar{i}} = \Gamma_{\sigma 2n+1}^a = \Gamma_{\sigma 2n+1}^{\bar{a}} = 0.
 \end{aligned} \tag{35}$$

Application 3. Let $V_{2n+1}(g_{\alpha\beta})$ be a Riemannian space and the generalized composition $X_{m+1} \cup X_{n+1} \cup X_{p+1} \cup X_{q+1}$ be Cartesian. In the parameters of the coordinate net $\{v\}_\alpha$ the matrix of $g_{\alpha\beta}$ is given by

$$(g_{\alpha\beta}) = \begin{pmatrix} g_{ij} & 0 & 0 & 0 & 0 \\ 0 & g_{\bar{j}\bar{j}} & 0 & 0 & 0 \\ 0 & 0 & g_{ab} & 0 & 0 \\ 0 & 0 & 0 & g_{\bar{a}\bar{b}} & 0 \\ 0 & 0 & 0 & 0 & g_{2n+1 2n+1} \end{pmatrix}. \tag{36}$$

By (18), (35) and (36) in the parameters of coordinate net $\{v\}_\alpha$ we obtain $g_{ij} = g_{ij}(u^s)$, $g_{\bar{j}\bar{j}} = g_{\bar{j}\bar{j}}(u^{\bar{s}})$, $g_{ab} = g_{ab}(u^c)$, $g_{\bar{a}\bar{b}} = g_{\bar{a}\bar{b}}(u^{\bar{c}})$ and $g_{2n+1 2n+1} = g_{2n+1 2n+1}(u^\alpha)$.

Acknowledgement. The second author would like to thank project NI15-FMI-004 of the Scientific Research Fund, Plovdiv University, Bulgaria.

REFERENCES

1. Ajeti, M., Teofilova, M. and Zlatanov G. (2011). Triads of compositions in an even-dimensional space with a symmetric affine connection. *Tensor, N. S.*, 73(3), 171-187.
2. Ajeti, M. (2014). Four dimensional spaces with a symmetric affine connection and additional structures. *Tensor, N. S.*, 75(1), 145-153.
3. Norden, A.P. (1963). Spaces of Cartesian composition, *Izv. Vyssh. Uchebn. Zaved. Math.*, 4, 117-128 (in Russian).
4. Norden, A.P. and Timoffev, G.N. (1972). Invariant criteria for special compositions of multidimensional spaces, *Izv. Vyssh. Uchebn. Zaved. Math.*, 8, 81-89 (in Russian).

5. Teofilova, M. and Zlatanov, G. (2014). Special spaces with a symmetric affine connection, *Tensor, N. S.*, 75(2), 154-165.
6. Zlatanov, G. and Tsareva, B. (1996). Geometry of the nets in equaffine spaces, *J. Geom.*, 55, 192-201.
7. Zlatanov, G. (2002). Compositions generated by special nets in affinely connected spaces, *Serdica, Math., J.*, 28, 1001-1012.
8. Zlatanov, G. (2011). Special compositions in affinely connected spaces without a torsion, *Serdica, Math., J.*, 37, 211-220.

Ivan Badev
Technical University-Sofia, Plovdiv Branch,
25 Tsanko Diustabanov Str., 4000 Plovdiv.
E-mail: ivanbadev@tu-plovdiv.bg

Marta Teofilova, Georgi Zlatanov
Plovdiv University,
236 Bulgaria Blvd., 4003 Plovdiv
E-mail: marta@uni-plovdiv.bg
E-mail: zlatanovg@gmail.com



AN ALGORITHM FOR DETERMINING AND COMPARING THE LEVELS OF ECONOMIC DEVELOPMENT IN THE DISTRICTS IN BULGARIA

DELYANA DIMOVA

Abstract: *This article presents an algorithm for determining and comparing the levels of economic development in the districts in Bulgaria. The necessary information is extracted from the created relational database. These data concerning gross domestic product, gross value added, employment rate and salaries are studied, compared and analysed for the relevant time interval. The developed algorithm can be also used in analyzing other economic and financial indicators.*

Key words: *algorithm, data processing, relational database, economic development*

1. Introduction

There is a huge quantity of text, audio, video, and other documents available on the Internet, on about any subject. Users need to be able to find relevant information to satisfy their particular information needs. There are two ways of searching for information: to use a search engines or to browse directories organized by categories. There is still a large part of the Internet that is not accessible (for example private databases and intranets) [1].

As the Web keeps expanding, it is increasingly difficult for the Information Retrieval (IR) systems to find relevant information which can satisfy users' needs based on simple search queries. The users express their need using queries [2]. The problem for search engines is not only to find topic relevant results, but results consistent with the user's information need [3, 4]. IR is different from data retrieval, which is about finding precise data in databases with a given structure [1]. The quality of information depends on the accuracy, integrity and consistency of the data in a database [5]. Ideally, this can be achieved if one item of data is stored only once. In reality this may not be possible as some duplication of data may be necessary in certain database architectures, e.g. relational databases make use of tables to store data and establish relationships between the tables by common fields [5].

Structured information in various guises is becoming ubiquitous on today's computers. This may include a user's contacts list, their calendar of events, other structured files such as spreadsheets or

databases. In addition to this, unstructured textual content may refer or add information to entities from a user's structured information space [6].

The current paper considers some of these exposed problems. National Statistical Institute provides economic and financial information on its website [7]. These data on gross domestic product, gross value added, employment rate, salaries, employees and etc. are saved in separate files. The information from the files is structured and stored in a created relational database [8]. This database should be expanded. In this context three new relational schemes are designed. The built database contains the following eleven tables (fig. 1):

- Countries (id_country, name);
- Value income (id, year, value-avg-household, id_source, value-avg-person, structure);
- Income sources (id_source, sources, id_country);
- Provinces (id_pro, province, id_country);
- Expenditure groups (id-g, expenditure group, id-country);
- Employees (id3, number, year, id_pro, employment rate);
- Salaries (id2, salary, year, id_pro);
- Value groups (id-v, year, avg- household, avg-person, structure, id-g).
- Households expenditure (id, share-home, share-food, total expenditure per person, year, id_pro);
- Household income (id1, share-salary, share-pensions, share-other social transfers, total income per person, year, id_pro);

- Gdp_Gva (id, years, Gva, Gdp, Gdp per capita, Id_pro)

The aim of this article is to present an algorithm for determining and comparing the levels of economic development in the districts in Bulgaria. The necessary information is extracted from the relational database. These data concerning gross domestic product, gross value added, employment rate and salaries are studied, compared and analysed for the relevant time interval.

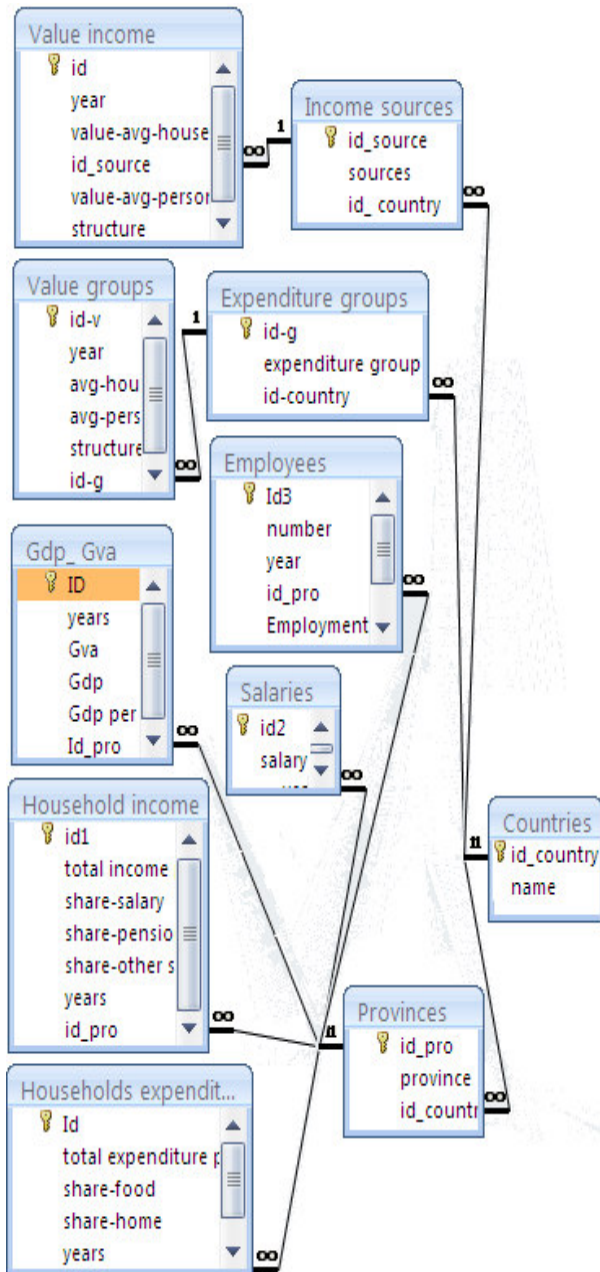


Fig. 1. General scheme of the relational database

2. Description of the algorithm

The scheme of the developed algorithm is presented in fig. 2. The input data for the algorithm are stored in the relational database. The information concerning the studied indicators (gross domestic product, gross value added, employment

rate and salaries) in the relevant districts is searched in several fields located in the following tables:

- Countries (id_country, name);
- Provinces (id_pro, province, id_country);
- Salaries (id2, salary, year, id_pro);
- Employees (id3, number, year, id_pro, employment rate);
- Gdp_Gva (id, years, Gva, Gdp, Gdp per capita, Id_pro).

Users could select the time interval. In this study, the research period covers generally 14 years from 2000 to 2013. The algorithm also allows the user to choose any smaller or larger period.

Practically, the choice of tables from the database depends on the choice of the relevant indicators. Data for the first two indicators (gross domestic product (Gdp), gross value added (Gva)) are extracted from the tables Countries, Provinces and Gdp_Gva. Information about the next two indicators (employment rate and salaries) respectively, is extracted from the tables Countries, Provinces, Employees and Countries, Provinces; Salaries. The obtained data from the database are also presented in tabular form. They should be compared and analysed. For this purpose it is necessary to find:

- The highest and the lowest value of the studied indicators for the given years;
- The intervals within which the values of the indicators change;
- The indicators that have insignificant change for the defined period;
- Percentage change of the salaries in the considered districts for the given years;
- The change of the employment rate for selected time interval;

These data for each of the indicators (gross domestic product, gross domestic product per capita, gross value added, employment rate and salaries by districts) also need to be arranged in ascending order of values for each year of the time interval. The obtained information could be summarized as a result of this sorting process. There are cases in which users could choose studying of data only for some of the listed indicators. It is also possible to include the data for all 28 districts in Bulgaria or part of them.

The levels of economic development in the relevant districts are determined on the basis of the performed comparisons and analysis. The results of data processing enable the user to present the obtained conclusions for given time period. The studied information about the listed indicators could be provided in graphical form. This allows to track much more quickly the dynamics of change of indicators for arbitrary selected interval. The developed algorithm can be also used in analyzing

other economic and financial indicators.

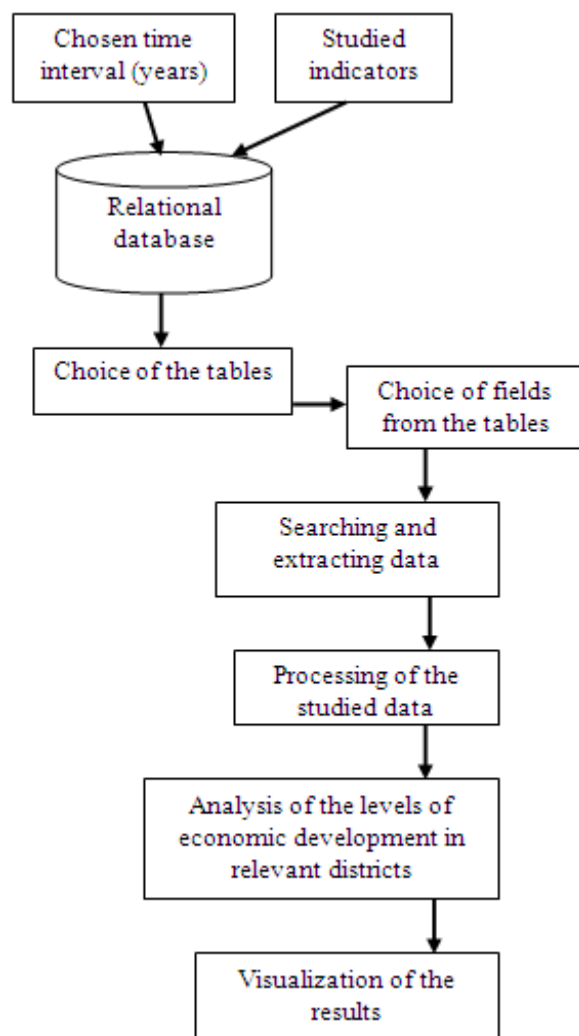


Fig. 2. Scheme of the algorithm

3. Results and discussion

The choice of indicators determines these tables in the relational database from which the studied information could be searched. The developed software is used for finding and extracting the data concerning the above listed economic indicators for 14-years time interval. Subsequently, the obtained information is sorted by pre-defined criteria. As a result, for the selected 28 districts, the values of the gross domestic product for each year from the included period are arranged in ascending order. Invariably, the first five positions in the investigated time interval are occupied from the following districts:

- Sofia (capital);
- Plovdiv;
- Varna;
- Burgas;
- Stara Zagora.

In this case, as can be seen from table 1 the values of the gross domestic product in the district of Sofia

(capital) are highest. The software calculates that their increase is more than 4,5 times for the period from 2000 to 2013. The second position is ranked from the district of Plovdiv. The other three districts change their positions (from third to fifth) within the examined 14 year period. It should be noted an interesting fact. Quite naturally, the values of the mentioned indicator are significantly higher in the district of Sofia (capital) in comparison with those in the other districts.

The analysis shows that the values of the gross domestic product for the years between 2000-2004 are the lowest in two districts:

- Vidin;
- Targovishte.

In the period 2005-2013 the values of the studied indicator are lowest in the following districts:

- Vidin;
- Silistra.

Table 1. Presentation of the sorted data for the investigated districts

Year	District	GDP (million BGN)	Year	District	GDP (million BGN)
2000	Stara Zagora	1 487	2007	Stara Zagora	2 715
2000	Varna	1 742	2007	Burgas	3 351
2000	Burgas	1 829	2007	Varna	4 312
2000	Plovdiv	2 058	2007	Plovdiv	4 659
2000	Sofia (capital)	7 148	2007	Sofia (capital)	23 484
2001	Burgas	1 539	2008	Stara Zagora	3 254
2001	Stara Zagora	1 710	2008	Burgas	3 818
2001	Varna	1 928	2008	Varna	4 993
2001	Plovdiv	2 333	2008	Plovdiv	5 058
2001	Sofia (capital)	8 554	2008	Sofia (capital)	27 433
2002	Burgas	1 654	2009	Stara Zagora	3 421
2002	Stara Zagora	1 699	2009	Burgas	3 701
2002	Varna	2 111	2009	Varna	4 816
2002	Plovdiv	2 480	2009	Plovdiv	5 481
2002	Sofia (capital)	9 945	2009	Sofia (capital)	27 989
2003	Burgas	1 869	2010	Stara Zagora	3 322
2003	Stara Zagora	1 901	2010	Burgas	3 680
2003	Varna	2 280	2010	Varna	4 630
2003	Plovdiv	2 755	2010	Plovdiv	5 539
2003	Sofia (capital)	10 799	2010	Sofia (capital)	29 316
2004	Stara Zagora	2 103	2011	Stara Zagora	3 536
2004	Burgas	2 136	2011	Burgas	3 963
2004	Varna	2 616	2011	Varna	5 178
2004	Plovdiv	3 148	2011	Plovdiv	6 062
2004	Sofia (capital)	12 402	2011	Sofia (capital)	31 754
2005	Stara Zagora	2 367	2012	Stara Zagora	3 913
2005	Burgas	2 685	2012	Burgas	3 976
2005	Varna	2 958	2012	Varna	5 400
2005	Plovdiv	3 663	2012	Plovdiv	6 178
2005	Sofia (capital)	14 602	2012	Sofia (capital)	31 726
2006	Stara Zagora	2 476	2013	Burgas	4 063
2006	Burgas	2 949	2013	Stara Zagora	4 071
2006	Varna	3 573	2013	Varna	5 323
2006	Plovdiv	4 017	2013	Plovdiv	6 374
2006	Sofia (capital)	18 173	2013	Sofia (capital)	32 232

Source: Based on data from National Statistical Institute

Invariably, the district of Vidin ranks the last position in the 14 year time interval.

In addition, it should be noted that similar results are also obtained from data processing for the second studied indicator (gross value added).

A little different is the situation for the examined data concerning gross domestic product per capita. As can be expected, the values of this indicator are again the highest in the district of Sofia (capital) and they significantly exceed those in other districts. At the beginning of the period (about 4 years) the values of the gross domestic product per capita in the district of Burgas are lower in comparison with those in the districts of Vratsa, Kyustendil, Gabrovo. Another fact is of an interest. More than ten districts have higher value of the studied indicator than this one in the district of Plovdiv in the first three years of the time interval. In the next years, this indicator increase gradually. The district of Plovdiv already occupied the 6th position in the last year from the examined period.

Specially, the data on the employment rate of population aged 15 years and over are investigated in the interval between 2008-2013. The values of this indicator have varied in quite wide range (from approximately 33% to 62%) for the chosen period. The developed software displays two districts where the employment rate is highest. In this case, as can be seen from table 2 they are the following:

- Sofia (capital);
- Blagoevgrad.

They occupy relevant the first two positions in the selected 6 years interval.

The number of districts, in which the employment rate is less than 40%, constantly increases. It already is seven in 2013 compared with 2008, where the district is only one. Eight of the twenty eight districts have had employment rate above 50% at the beginning of the period. This number decreased to four in the next year. Subsequently, the number of districts is only two in the last 4 years of the examined interval. Generally more are the districts in which the employment rate varies in range between 40%-45,9% in the period 2008-2013. A small part are those ones (from five to eight) in which this indicator reaches values between 46%-50%, during three consecutive years from 2010 to 2012 (table 2). The change of the values of the employment rate is relatively small (about 3,6%) in the investigated period in the following districts:

- Pernik;
- Razgrad;
- Stara Zagora.

At the same time, the change of this indicator is quite large (over 10%) in the 6 year period in five

districts:

- Kardzhali;
- Lovech;
- Pazardzhik;
- Smolyan;
- Varna.

It should be noted that the employment rate declined steadily, only in the district of Lovech in throughout the studied time interval. While in the other districts, periods are alternated in which the indicator decreases and then increases or vice versa. The values of the employment rate are the lowest in the district of Vidin in the years between 2008-2012. The situation is quite different in the last year of the examined time interval. The lowest values are observed in the following districts:

- Silistra;
- Lovech;
- Targovishte.

Table 2. Visualization of the results for the employment rate

employment rate -%				
year	min	max		
2008	39,1	61,5		
2009	36,4	60,7		
2010	35	57,4		
2011	33,6	55,7		
2012	37	55,7		
2013	37,1	56,1		
year	district	district	district	district
2008	Blagoevgrad	57,20	Vidin	39,10
2008	Sofia (capital)	61,50		
2009	Blagoevgrad	55,40	Vidin	36,40
2009	Sofia (capital)	60,70		
2010	Blagoevgrad	53,60	Vidin	35,00
2010	Sofia (capital)	57,40		
2011	Blagoevgrad	53,80	Vidin	33,60
2011	Sofia (capital)	55,70		
2012	Blagoevgrad	53,10	Vidin	37,00
2012	Sofia (capital)	55,70		
2013	Blagoevgrad	51,4		
2013	Sofia (capital)	56,1		
number of the districts				
year	under 40%	40%-45,9%	46%-50%	over 50%
2008	1	8	11	8
2009	2	11	11	4
2010	7	11	8	2
2011	7	14	5	2
2012	5	13	8	2
2013	7	9	10	2

Source: Own calculation on the basis of data from National Statistical Institute

In this study the data on the average annual salary are also investigated. The analysis shows that the values of this indicator are the highest in the period from 2000 to 2013 in six districts:

- Sofia (capital);
- Stara Zagora;
- Vratsa;
- Sofia;
- Varna;
- Burgas.

The average annual salary grows at the fastest pace in the district of Sofia (capital), as is shown in table 3. The developed software calculates that the increase is about 4 times in 14-years interval. The calculations also show that the average annual salary in the listed five districts is about 1,3 times lower in comparison with this one in the district of Sofia (capital). Another interesting fact should be noted. The values of the indicator in the district of Vratsa are higher in 2002 than those in 2003. A similar situation is obtained for the examined indicator in the district of Smolyan in the period 2010-2011.

Invariably, the average annual salary in the district of Smolyan is the lowest in the first 6 years of the considered time interval. Its values increased gradually over the next five years. Then this indicator decreased again at the end of the period. The district of Vidin has ranked last by average annual salary in the years between 2008-2013. The rate of growth of this indicator is the slowest in the mentioned district. The values of the studied indicator are very similar in the second half of the period in the following districts:

- Blagoevgrad;
- Haskovo.

The percentage change of the average annual salary is calculated in 2013 compared to 2002. The software displays that it varies in wide range from 1,6% to 9,75% (table 3). It should be noted that the number of the districts in which the percentage increase is about 6,57% to 6,93% is the largest. In the case, these districts are:

- Gabrovo;
- Razgrad;
- Lovech;
- Kyustendil;
- Sofia (capital);
- Plovdiv.

As a result of this research could be done the following conclusion: The level of economic development in the district of Sofia (capital) is significantly higher in the examined period. Then rank the districts of Varna, Burgas, Stara Zagora and Plovdiv. The level of economic development in the districts of Vidin, Silistra, Targovishte is lower

in comparison with this one in the other districts in the considered years.

Table 3. Analysis of the data for average annual salary

year	districts	%	districts	%	
2013	Vidin	5,81	Burgas	5,44	
	Vratsa	1,65	Sliven	6,22	
	Lovech	6,62	Stara Zagora	3,26	
	Montana	6,44	Yambol	6,47	
	Pleven	6,02	Blagoevgrad	4,7	
	Veliko Tarnovo	7,71	Kyustendil	6,77	
	Gabrovo	6,93	Pernik	5,12	
	Razgrad	6,62	Sofia	3,65	
	Ruse	6,07	Sofia (capital)	6,57	
	Silistra	7,67	Kardzhali	9,19	
	Varna	4,2	Pazardzhik	5,87	
	Dobrich	6,32	Plovdiv	6,71	
	Targovishte	9,75	Smolyan	4,36	
	Shumen	4,77	Haskovo	5,01	
	period	districts	period	districts	
	2000	2014 Sofia (capital)	2000	2014 Stara Zagora	
	2000	2014 Sofia	2000	2014 Burgas	
	2000	2014 Varna	2000	2014 Vratsa	
	increase	increase			
	4 times	1.3 times			

Source: Own calculation on the basis of data from National Statistical Institute

4. Conclusion

This article presents an algorithm for determining and comparing the levels of economic development in the districts in Bulgaria. The necessary information is extracted from a created relational database. These data concerning gross domestic product, gross value added, employment rate and salaries are studied, compared and analysed for the relevant time interval. The examined period includes the years from 2000 to 2013. The developed algorithm can be also used in analyzing other economic and financial indicators. The software has been developed for data processing. It significantly helps users - economists and financiers in decision making in their practice.

REFERENCES

1. Inkpen, D. Information Retrieval on the Internet. Available at: http://www.site.uottawa.ca/~diana/csi4107/IR_draft.pdf, [2016-02-23]
2. Suma, D., Acharya, U. D., Geetha, M., Raviraja Holla, M. (2014). XML Information Retrieval: An overview. *International Global Journal for Engineering Research*, volume 10, issue 1, pp. 26-32
3. Mamoon, H., El-Bakry, H., Slamaa, A. (2014). Visualization for Information Retrieval in Regional Distributed Environment. *IJCSNS International Journal of Computer Science and Network Security*, vol.14, no.9, pp. 12-27
4. Mamoon, H., El-Bakry, H., Salama, A. (2013). Interactive Visualization of Retrieved Information. *International Journal of Knowledge Engineering and Research*, vol. 2, issue 4, pp. 217-231
5. Ram, S. (2008). Teaching Data Normalization: Traditional Classroom Methods versus Online Visual Methods – a Literature Review. *Proceedings of the 21st Annual Conference of the National Advisory Committee on Computing Qualifications (NACCCQ)*, Auckland, New Zealand, pp. 327-330
6. Tablan, V., Damljanovic, D., Bontcheva, K. (2008). A Natural Language Query Interface to Structured Information. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M. (eds.) *ESWC 2008. LNCS*, vol. 5021, pp. 361–375. Springer, Heidelberg
7. National Statistical Institute: <http://www.nsi.bg>
8. Dimova, D. (2013). Software for Extraction and Assessment of Data for Household Budget in Bulgaria. *Proceedings of the international scientific conference Unitech'13*, Gabrovo, Bulgaria, volume II, pp. II-245-II-248

Department „Mathematics, Informatics and Physics” Agricultural University,
12 Mendeleev blvd, Plovdiv, Bulgaria,
E-mail: delyanadimova@abv.bg



PARAMETRIC REGRESSION MODELS FOR CUBR LASER

ILIYCHO ILIEV

Abstract: *A comparison of the developed polynomial regression models of first, second and third level of the output laser power of CuBr laser is presented. It is found that regression equations involving all 10 independent variables (predictors) show better statistical indicators and more faithfully describe the influence of the input laser characteristics on the output laser power.*

Key words: *Copper bromide vapor laser, laser output power, parametric model, polynomial model, regression analysis, regression model*

1. Introduction

The object of study in this paper is a low-temperature copper bromide vapor (CuBr) laser, with a wavelength of 510.6 nm and 578.2 nm. It is accepted that metal vapor and metal compound vapor lasers have been studied in great detail through experiments and in theory. During the last few years as a result of active experiment development, a significant increase (up to 120-150 W) of the output power was achieved. Copper bromide (CuBr) vapor lasers continue to be the most powerful laser source in the visible spectrum. This is the reason for its active application in medicine, physics, chemistry, ecology, technology, and scientific research. There are over 2000 scientific publications, papers, books, and monographs in renowned science journals in this field, as well as papers presented at prestigious science conferences. For this reason, they are subject to particular interest and rank among the 12 most commercialized lasers in the world. They are manufactured in many countries throughout the world - Russia, USA, China, England, Bulgaria, etc. The use of copper halides gives significant advantages to this type of lasers when compared to pure copper lasers. The first of these advantages is that the operating temperature of the active laser tube is reduced by around 1000 °C as compared to pure copper lasers working at temperatures of about 1500 °C. This allows the use of cheaper material for the manufacture of the laser tube - usually quartz. Another advantage is that CuBr vapor lasers are cooled by air and do not require additional water- or another type of cooling.

2. Subject of study

The subject of investigation is a copper bromide vapor laser which is an original Bulgarian design developed at the Laboratory of Metal Vapor Lasers at the Georgi Nadjakov Institute of Solid State Physics of the Bulgarian Academy of Sciences, Sofia. The first patent related to this laser is that of Sabotinov, N.V. et al., Bulg. patent No.:28674, 1975.

A conceptual schematic of the laser source is given in Fig. 1.

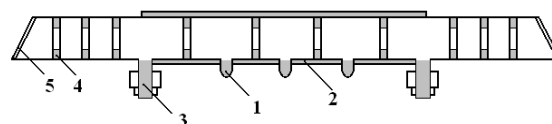


Fig. 1. *Construction of CuBr laser tube: 1- reservoirs with copper bromide, 2- heat insulation of the active volume, 3- copper electrodes, 4- inner diaphragms, 5- tube windows*

Neon is used as a buffer gas. In order to improve efficiency, small quantities of hydrogen are added. Unlike the high-temperature pure copper vapor laser, the copper bromide vapor laser is a low-temperature one, with an active zone temperature of 500 °C. The laser tube is made out of quartz glass without high-temperature ceramics as a result of which it is significantly cheaper and easier to manufacture. The discharge is heated by electric current (self-heating). It produces light impulses tens of nanoseconds long. Its main advantages are: short initial heating period, stable laser generation, relatively long service life, high values of output power and laser efficiency.

On the base of a huge amount of experimental data different statistical regression

models have been obtained in [1-3]. In this study the dependence of basic laser operational physical characteristics, such as the inner diameter of the laser tube, the length of the active laser zone, the input supplied electric power, pressure of the additional hydrogen gas and more on the output laser power is considered. The aim of the study is to summarize, compare and analyze the obtained linear and nonlinear models in order to identify the more reliable results, which describe the investigated dependency.

3. Data description

We use experimental data for copper bromide laser with wavelengths 510.6 nm and 578.2 nm, published in [1]. The initial data consist of more than 300 experiments including eleven laser parameters. The ten independent initial variable parameters (laser characteristics) are: D (mm) – inside diameter of the laser tube; dr (mm) – inside diameter of the rings; L (cm) – length of the active area (electrode separation); Pin (kW) – input electrical power; PL (kWm^{-1}) – input electrical power per unit length, $PH2$ (Torr) – hydrogen gas pressure, Prf (kHz) – pulse repetition rate; PNe (Torr) – neon gas pressure, C (pF) – equivalent capacity of the capacitor bank, Tr ($^{\circ}\text{C}$) – temperature of the CuBr reservoirs. The main dependant variable we observe here is the output laser power P_{out} (W).

As it was previously established in [2], for our data, only the first six of these variables show statistically significant influence on the output laser power P_{out} . To this reason in this paper we use only the data about the first six variables, i.e. D , dr , L , Pin , PL and $PH2$. These six variables will be referred to as basic variables. While they correlate strongly each other, a direct construction of the regression models is not applicable [1]. So, we need to combine them mathematically in uncorrelated artificial group variables, called factors. This procedure is performed by the multivariate factor analysis. In our case we use three factors, which explain about 93-94% of any one of selected samples of examined data.

Table 1. Rotated component matrix with three factors^(a)

Variables	Component (factor)		
	F1	F2	F3
Pin	0.935		
Dr	0.909		
L	0.780		
D	0.767		
PL		-0.881	
PH2			0.915

^(a)Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 5 iterations.

There exist many different methods for extraction and rotation of factors. We use the more popular methods: principal component analysis for extraction of the selected number of factors and rotation by the Varimax method with Kaiser normalization [1]. For the initial data sample the obtained rotated matrix is shown in Table 1. For clarity, the factor loadings less than 0.5 are skipped.

4. Parametric regression models for CuBr laser

From the previously derived models we will present linear and non-linear parametric models for CuBr laser to analyze their statistical indicators. In these models the obtained factor variables are used as independent uncorrelated variables. This approach is also known as principal component regression [1].

4.1. General relationship between factors and laser output power for CuBr laser

Based on [2], in an exploratory manner, we will later on consider regression models using three groups of variables as predictors: first, second and third groups as follows:

$$F_1, F_2, F_3 \quad (1)$$

$$F_1, F_2, F_3, F_1^2, F_2^2, F_3^2, F_1.F_2, F_1.F_3, F_2.F_3 \quad (2)$$

$$\begin{aligned} &F_1, F_2, F_3, F_1^2, F_2^2, F_3^2, F_1.F_2, F_1.F_3, \\ &F_2.F_3, F_1^3, F_2^3, F_3^3, F_1^2.F_2, F_1^2.F_3, \\ &F_2^2.F_3, F_1.F_2^2, F_1.F_3^2, F_2.F_3^2, F_1F_2F_3 \end{aligned} \quad (3)$$

The corresponding models will be noted as 0, 1, and 2 order models, respectively.

The results from the modeling have been presented in this section. For parametric methods it is assumed that data and population distribution are nearly normal. All calculations and analyses have been carried out at level of significance 0.05. The comparison between models has been conducted via the commonly used indices, such as multiple correlation coefficient R , coefficient of multiple determination R^2 (RSquare), and adjusted R^2 .

With the help of the three orthogonal PCA factors F_1, F_2, F_3 (first group of predictors (1)) and the stepwise or linear procedure we obtain the MLR (Multiple Linear Regression)-0th order models for estimation of the dependent variable P_{out} :

$$\hat{P}_{out} = 40.589 + 29.717F_1 + 4.155F_2 + 12.941F_3 \quad (4)$$

To note, that in fact, all coefficients as well as all subsequent estimates have been obtained at a lower significance level of 0.000 for model MLR-0th order, with no interactions.

Using the nine predictors (2) a more precise regression model was constructed, marked as MLR-1st order. The corresponding equation is:

$$\hat{P}_{out} = 38.283 + 28.090F_1 + 4.866F_2 + 11.992F_3 + 2.336F_1^2 \quad (5)$$

For the second-order model the obtained equations are, respectively:

$$\hat{P}_{out} = 38.270 + 39.511F_1 + 8.243F_3^2 + 8.718F_3^3 - 2.294F_1^3 + 3.053F_2^2F_3 - 0.976F_2^3 \quad (6)$$

This means that (2) and (3) describe 94.6% of the sample

The basic statistics of these models are given in Table 2. In this table we include the commonly used indices, such as coefficient of multiple determination R^2 , adjusted R^2 , and standard error of the estimate.

Table 2. Results from constructed parametric regression models for estimation of output laser power P_{out} using MLR (Multiple Linear Regression)

Model	R^2	R^2 Adj.	Std. Err. of the Estimate	Number of Predictors
(4)	0.946	0.944	7.92540	3
(5)	0.950	0.948	7.63382	4
(6)	0.967	0.963	6.27075	7

4.2. General relationship between ten independent initial variable parameters and laser output power for CuBr laser

It must be noted, that some disadvantage of the obtained principle regression models (4)-(6) is that the actual physical variables are "hidden" behind the factor variables. Also the prediction of laser power requires the application of some specialized software and its skilled use. This makes the approach difficult for immediate use. For this reason, the issue of developing polynomial parametric models containing explicitly the 10 independent variables is topical.

The initial approach requires finding a linear regression relationship of the type:

$$\hat{P}_{out} = a_0 + a_1D + a_2dr + a_3L + a_4Pin + a_5PH2 + a_6PL + a_7PRF + a_8PNe + a_9C + a_{10}Tr \quad (7)$$

where $a_i, i = 0, 1, \dots, 10$ are the regression coefficients (parameters) to be determined.

The factor and regression analysis performed showed that there is a high degree of correlation between the predictors and that an equation of type (7) is not statistically significant [1]. For this reason we will look polynomial models from the second, third and higher level. In [3] has developed a non-linear parametric model of the second degree, which is as follows:

$$\hat{P}_{out} = -18.308 + 0.686dr.Pin - 7.567Pin.C + 27.569Pin - 4.362Pin^2 - 0.006L.PRf + 0.274PRf.C - 0.612dr.C + 0.259D.PH2 + 18.093C \quad (8)$$

Model equation (8) contains 8 of the 10 independent variables: $dr, D, L, Pin, PL, PRF, PH2,$ and C . It does not include variables PNe and Tr .

Non-linear parametric model of the third degree [4] has the form:

$$\hat{P}_{out} = 10.243 + 64.453Pin - 0.143PL + 0.023D.PRf + 0.303D.C - 0.061dr.PL - 0.011L.PRf - 0.419L.C - 2.009Pin.PL + 0.437Pin.PRf + 2.341PL.PH2 - 2.791PL.C - 0.579PRf.C + 0.165C.Tr - 0.0121D^2 + 0.025.dr^2 - 41.015PH2^2 - 11.333C^2 + 0.003PL^3 - 2.1 \cdot 10^{-5} PRF^3 + 1.613C^3 \quad (9)$$

It is observed that equation (9) contains only three third degree terms: PL^3, PRF^3 and C^3 . Model equation (9) contains all 10 independent variables: $dr, D, L, Pin, PL, PRF, PH2, C, PNe$ and Tr . The basic statistics of the models (8) and (9) are given in Table 3.

Table 3. Main statistical indicators of nonlinear models from second and third degree of the output laser power P_{out} using MLR

Model	R^2	R^2 Adj.	Std. Err. of the Estimate	Number of Predictors
(8)	0.967	0.962	6.56416	10
(9)	0.976	0.972	5.67721	21

5. Discussion and conclusion

Comparing the basic statistical parameters in Tables 2 and 3 show that the non-linear patterns, included in equations (8) and (9) give higher statistical indices in terms of coefficients of determination R^2 and adjusted R^2 than those of type (5) and (6). This fact follows from statistical theory, i.e. models of higher degree give higher coefficient of determination. However, in our case we additionally obtained lower errors of estimates and better agreement of equations (8) and (9) with the known experimental data [1]. So, as a statistical inference we have to expect more accurate prediction of laser power when constructing new laser sources with high performance. The main reason in our opinion is that in equations (5) and (6) involved only 6 of all 10 independent physical parameters. Equation (8) contains 8 independent parameters, and equation (9) includes all the 10 variables.

From the previous analysis we can conclude that equation (9) has the best statistics. It describes more adequately the complex processes in the laser medium and can be accurately used in planning and creating new laser sources.

ACKNOWLEDGMENTS

This work was supported in part by Paisii Hilendarski University of Plovdiv NPD under Grant NI15-FMI-004.

REFERENCES

1. Gocheva-Ilieva, S.G. and Iliev, I.P. (2011). *Statistical models of characteristics of metal vapor lasers*. Nova Science Publ. Inc., New York.
2. Gocheva-Ilieva, S.G. and Iliev, I.P. (2010). Parametric and nonparametric empirical regression models: case study of copper bromide laser generation. *Mathematical Problems in Engineering*, volume 2010 (Article ID 697687), pp. 1-15.
3. Denev, N., Iliev, I. and Gocheva-Ilieva S. (2013). Second-degree polynomial model of laser generation for a GuBr laser. *WSEAS Transactions on Circuits and Systems*, volume 12(4), pp.129-139.
4. Iliev, I. and Denev, N. (2014). Parametric models of the third degree for output parameters of a CuBr laser. *International Journal of Scientific & Technology Research*, volume 3(11), pp. 28-32.

Department of Mathematics, Physics and
Chemistry
Technical University–Sofia, Branch Plovdiv
25 Tsanko Diustabanov St.
4000 Plovdiv
BULGARIA
E-mail: iliev55@abv.bg



PRINCIPLES AND INSTRUMENTS OF LEAN METHODOLOGY

TONI MIHOVA, VALENTINA NIKOLOVA - ALEXIEVA

Abstract: *The report referred to the essence, principles and tools of LEAN methodology. Based on research conducted in fifty large Bulgarian enterprises gives the results on their willingness to make this management methodology in real practice.*

Key words: *LEAN thinking Just in Time, 5S, TPL*

1. Introduction

Industrial globalization that has taken place in recent decades has forced manufacturers to produce high quality products in the most efficient manner. Striving to improve manufacturing operations has become a key issue for the industry. There was a need for Bulgarian producers to be flexible and competitive. Businesses put increasing pressure on production, companies seeking to improve their operational management through the implementation of new initiatives, methodologies and tools. One of them is LEAN methodology. Bulgarian companies have shown interest in it from the 70s. It is a set of tools and techniques developed by industrial giant "Toyota" from its suppliers, but also by its competitors in the automobile sector. Long since, the ideas that create left the borders of Japan and are widespread in the USA, Western Europe, China, India, Turkey and others. Adapted to Bulgarian reality, they are applicable and useful for the modern Bulgarian industry. Their logic, simplicity, precision and focus on continuous improvement help enterprises to deal with most of the production problems.

The aim of the report is revealing the essence of this management methodology, its principles and instruments, an indication of the results of the survey on attitudes to it and on this basis, formulate conclusions and recommendations for the use of LEAN philosophy.

2. Essence of the LEAN methodology

LEAN concept was born from the giants of thought Adam Smith, Frederick Taylor, Henry Ford and Edward Deming. It was put into practice by the creator of Toyota factories - Kiishiro Toyoda. Taiji Ono approach developed under the name TPS (Toyota Production System), subsequently became known as Lean.

2.1. Definitions

Will point out three definitions, we think most clearly reflect the essence of LEAN methodology. - LEAN is a management approach that directs and limits the cost of inputs only in achieving goals that create added value for the end customer. [1] - Methodology that aims to build a production system reduces the time between order and delivery by eliminating surplus [2]. - Overall concept of corporate governance, including the knowledge and tools that organizations use to get rid of all activities and procedures to production processes, time-consuming, but not creating added value for the users of its products. The application of LEAN tools, especially in combination with this concept of Six Sigma, has a remarkable impact on the results of each organization adopted it as its philosophy. [3]

If we summarize, we could look LEAN methodology as an approach for effective business management and human resources.

2.2. Lean Principles

In their book "The Machine That Changed the World" Womack and Jones define 5 principles of Lean Manufacturing. These Lean Principles define the philosophy of LEAN.

These principles are:

1. Specify Value as perceived by the Customer
2. Identify the Value Stream
3. Make the Value Flow through the Value Stream
4. Pull the Value from the Value Stream
5. Strive for Perfection.

These LEAN principles can be used to define the aim of any lean system which is to "Clearly specify value in order to line up all the activities for a specific product (family) along a value stream and

make the value flow smoothly at the pull of the customer in pursuit of perfection” [4],[5]

2.2.1. The Principle of Value

The value changes over time and depends on the customer. In creating its operating and use resources that create value except "wasteful" - any activity that consumes resources but no added value. [5],[6]. The customer is willing to pay for value, but not for "the wastes". In the manufacture, as mentioned below most frequently observed kinds of surplus are known as 7 losses. Lean thinking allows us to find these losses and consequently reduce them.



Fig.1. LEAN Principles

Overproduction - producing more than is possible to sell such products or for which there is no demand. In both cases we produce a product without being sure that there is a customer who wants it and will buy it. So we expend resources that are not redeemed. Overproduction is considered the largest of the seven losses.

Movement - unnecessary movement of materials, semi-finished products, people, tools, documents and more. Excess movement impedes the work, makes it more tiring, time consuming and so on. But the customer is unwilling to pay for unnecessary movement.

Transport - cross-flows, imbricated or counterclaims flows in production, unnecessary transportation of products from one process to the next, unnecessary transportation to and from treatment, or something still not in place or not located near the site of its usage. The customer is unwilling to pay for unnecessary transportation.

Downtimes - big losses from freezing working capital in production due to stays and a periods of waiting - before and after processing operations between the departments and interoperational stocks, a man waiting to release the machine, machine waiting to be loaded by the operator, a station waits another station to submit his work, workers wait for their boss to order what to do, production stays due to lack of materials or for

repairing damaged machines. The customer does not want to pay for unnecessary downtime. [6] **Inventory** - machines, tools and gadgets that rarely or never use, plus it positioned along the process that impede him and cluttering. The process can be clogged by the accumulation of the working areas of machinery and transport corridors waiting for treatment products, stopped batches unnecessary at present tools and devices or overcrowded wastage isolators. It does not bring value to the customer and he does not want to pay. [5]

Overproduction - in production may have excessive auxiliary, elaboration or corrective operations due to the thoughtlessness of the technology. Or the technology is not stabilized, leading to scrap and repeated operations. The customer does not want to pay that we have not yet learned to do things well from the first time.

Wastage - every product come out of production, which is of low quality and has remained unfit for use. Often wastage entails loss during processing and therefore reuse of materials. Wastage is usually the result of the implementation of the remaining losses.

Production surpluses impede adaptability, increase permanent and direct costs, lower productivity and quality, create discomfort in the workplace. With more expensive and poor quality product can not achieve upward customer satisfaction. As a result, sharply worsened overall picture of production and commercial efficiency.

LEAN approach facilitates production and functional managers and professionals to focus on such decisions that lead to increased value added for the customer.

2.2.2. The Lean Principle of Value Stream

The value stream is the series of interrelated processes that produce the value; from raw material through to use. Just the steps that add the value, not those steps that add no value. Value adding and non-value adding steps make up every process. Within lean we are looking to remove or minimize those non-value adding steps.

We are looking at the value stream from the point of view of the product (or value), not looking at individual departments or even companies. This can be mapped using tools such as value stream mapping to create current and future state maps of the overall flow.

Think about the actual processes when you define the process stream; not in terms of your companies departments. [6]. Many lean implementations end up creating value streams where there are no "departments" as such but a pure flow through the company for the value itself.

2.2.3. The Principle of Making Value Flow

The ideal flow would be “one piece flow”, however this is often not feasible due to machine setups and the need to flow multiple product streams through individual machines or cells. But what we are trying to achieve is the “flow” of products or value from one step to the next, each step being a value adding step. Never delay a value adding step by a non-value adding step, where possible have these done in parallel to the value adding one. This flow is achieved through a whole host of ideas and tools from Kanban, to small machines, through cell design and so on.

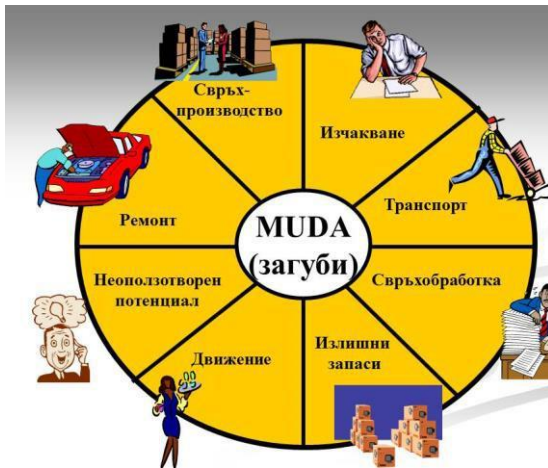


Fig.2. Wastes of Lean Manufacturing

2.2.4. Pull systems a Lean Principle

One of the biggest wastes in any system is that of inventory; inventory hides all of the other problems in your system and causes so many other wastes. Inventory takes up space, it requires stacking, storing and transporting, it eats your capital!

The ideal system is one where the customer makes the order and you manufacture that product only when ordered. This is an ideal situation and many industries do not believe that they can achieve it. But lean is a never ending journey and with evolving technology and customer requirements it is possible to achieve it.

Consider a book; if you go into a book store there are shelves that are stacked high with thousands of books, some which will never be purchased and will either be sent to “bargain clearance outlets” or will be pulped. The reason being that the huge presses that produce these books have to be run for “economic” batch sizes. But with today’s technology people no longer need the printed book, they can download the text to their kindle or laptop and read it on demand. For those that still want the feel of the printed word in their hands many bookstores now have machines which will print and bind your individual book while you

wait! Single piece flow at the demand of the customer in an industry that would have laughed at you if you suggested it a few years ago!

Pull production rather than pushing production through a factory using scheduling can be achieved very simply in many cases. By using kanban systems and “supermarkets” it is possible to set up your production so that customer orders can be quickly satisfied when they are made and the parts either manufactured to order from standard components or drawn from a small buffer stock and then replenished in much the same way a super market will refill the shelves as customers buy product. In fact it is a super market that gave the idea for this method of production.

2.2.4.Perfection

If you achieve the first four steps you will have already prevented a huge amount of waste (seven wastes) from appearing within your processes. However; with the help and support of all of your employees you need to strive towards perfection; delivering exactly what the customer wants, when they want it at an acceptable (Minimum) price with zero waste.

Don’t go out and benchmark your competitors and try to match them or beat them by a little; the aim is zero waste and the ability to deliver your customers value.

Involve every employee within your company in implementing lean tools such as Kaizen to drive continuous improvement of each and every aspect of your company. Lean is not just about improving a production cell; it also about improving every other process from order processing to invoicing and customer service.

2.3. LEAN tools

We will briefly examine some of the basic LEAN tools developed several decades by leading companies to achieve excellence in terms of efficiency, quality of organization.

2.3.1.LEAN thinking

LEAN thinking is a business methodology that aims to provide a new way to think about how to organize human activities to deliver more benefits to society and value to individuals while eliminating waste. The term lean thinking was coined by James P. Womack and Daniel T. Jonesto capture the essence of their in-depth study of Toyota’s fabled Toyota Production System.

The aim of lean thinking is to create a lean enterprise, one that sustains growth by aligning customer satisfaction with employee satisfaction, and that offers innovative products or services profitably while minimizing unnecessary over-costs to customers, suppliers and the environment. The basic insight of lean thinking is that if you train every person to identify wasted time and effort in

their own job and to better work together to improve processes by eliminating such waste, the resulting enterprise will deliver more value at less expense while developing every employee's confidence, competence and ability to work with others.

2.3.2. Just In Time

"Just in time" is to produce only what is needed in the quantity you need and when the need of the customer. JIT is a symbol of: short production cycle; minimized stocks of materials and unfinished production; balanced production load; minimum transportation, higher labor productivity and timely delivery is simply the result of all this. Summarized JIT works to reduce most of the groups of "surplus" (identified through the creation of VSM).

So specify those goals can be achieved only if used well-established methodology. JIT methodology is based on several instruments here we mention some of them.

One of the main tools in the JIT is the creation of a suitable processing scheme (Layout).

2.3.3. „5 S“

In brief „5 S“ is a system in which through various practical methods applied by a team effort, the people maintenance neat and clean workplaces, workstations and departments.

Good layout in the workshop itself is a prerequisite for efficient processes as it saves time for operators. On the other hand it is also a barometer of quality: in how workers relate to instruments used, how maintain the purity of their workplace can judge their attitude towards work itself and to quality. In summary “5 S” gives an overall assessment of the commitment of employees and management to the company and its condition.

Five steps are: "Sort" "Systematic arrangement"; "Shining purity"; "Standardisation"; "Self-discipline". The steps in 5S are built logically and consistently. The proper study of the system leads to so desired goals - clean and stacked workplaces, convenient and pleasant working environment.

2.3.4. TPM

TPM is a system created and developed in Japan for several decades. With an initial focus on the production and machines, it evolved into a program covering almost all activities in the organization. TPM aims to:

creating a corporate culture that maximizes the efficiency of the production system;

- practical organization of production to prevent losses before they emerged throughout the life cycle of the production system in order to achieve zero accidents, zero defects and zero failures;

- full involvement of all departments of the company including document production, maintenance, research and developed marketing, etc;

- engaging all employees from top management to machine operators;

- achieving zero losses through activities "overlapping small teams."

2.3.5. Kaizen

Kaizen – a Japanese word that literally translated means a change for the "better". The most common perception of the word is related to continuous improvement. Kaizen ensure that we continue to develop and improve. In most plant managers think that if you put a box in which employees put written proposals from them, it will begin to fill. Of course in business practice there are different situations and many companies tried this approach are not convinced that it works. When it undertake an effective system for generating Kaizen must first consider what causes people not to put proposals in the box, namely:

1. The level of the majority of employees in our factories is so low that they can not work up the guts to take pen and paper and write a proposal.

2. They are afraid lest someone scolded them for what they wrote or simply laughs on them.

3. Even if there is anyone who dares to write something, the process of giving feedback is so long that one forgets what he wrote or it is no longer valid.

4. The senior and middle level management are not involved in the process.

How to eliminate causes and to create a working system? For this purpose it is necessary to build an entirely new structure of the organization in which production is divided into teams of about 10 people and each team is responsible unless his leader and his coach. All representatives of the administration become coaches and give their direct contribution to production. It is they who encourage and motivate operators to participate in the process of continuous improvement - Kaizen. In so created structure the language with the most Kaizen and his coach have a chance to be the best after a fair competition with clear criteria. Through this approach, we encourage the competitive spirit in people, but not greedy, pechelbarski sense, but as a desire to outdo the other [2].

3. Results of the survey

The study was conducted in 50 large companies in Bulgaria on knowledge of the methodology LEAN, attitudes about the use and effects of adopting some of LEAN techniques and instruments. The methods we used are questionnaires and interviews with managers of

enterprises. The survey was conducted between October 2015 - February 2016.

What are the concrete results concerning the examination of LEAN practices?

80% of participants indicated in questionnaires that are highly aware of the essence, principles and tools of LEAN methodology. The main sources of information mentioned by them are seminars, conferences and various forms of training.

15% of them know to a lesser extent managerial approach, 3% said they did not know and 2 percent did not respond.

In response to the second question related to attitudes about using some of the LEAN principles and tools, 70% think they are not adequately prepared. The reasons most often found answers: "Not applicable to us," "We do not have the necessary financial resources" "I will meet resistance from the workers," "Too radical change will be," "I do not think it is necessary at this stage."

And only 10% are reflected in polls that have implemented some of LEAN tools.

Interesting is the fact that they are companies with foreign participation. From interviews with them about the results of LEAN practices, we can systematize the following benefits:

1. After the implementation of Just in Time in four enterprises reported:

- a significant reduction in unnecessary movements and balance of work which leads to increased productivity;
- terminated a large number of cross back and forth movements of raw materials and blanks;
- reduced volume of work in progress blocking valuable financial resources;
- clarity of material flow in production.

2. Following the introduction of KANBAN system in 2 enterprises have improved significantly: -

- communication and coordination between the units "Planning" - "Logistics" - "Production";
- places for stocking of raw materials;
- operator efficiency - largely in order and vision of stocks overcome chronic chaos and scratching materials;
- machine downtime due to organizational problems with the materials.

3. Following the introduction of the "5S" in 4 companies:

- a system layout for all instruments;
- the working environment is visualized;
- remove unnecessary items.

The conclusions from the survey include:

1. The highest percentage of respondents are familiar with the essence of LEAN methodology, its tools and principles.
2. Attitudes of the majority of study participants were skeptical about turning LEAN approach in actual practice.
3. A relatively small part implemented some major LEAN tools, actually. The results have led to increased productivity and efficiency of the activities of these companies.

4. Conclusion

The report made an attempt to mark some important aspects of the LEAN philosophy for effective business management. The survey results show that there is still a small part of Bulgarian companies, which mainly those with foreign participation reap the benefits at issue management methodology.

Requirements for enhancing the competitiveness of Bulgarian enterprises make it necessary to increase the efficiency of their operations. In this connection it is necessary to use the opportunities offered by LEAN methodology.

REFERENCES

1. Lean THINKING, <http://lean.bg/wcm/lean-thinking/>
2. Principles and tools in the Lean methodology, teaching materials of "LEAN Bulgaria" Ltd, 2016
3. Jorgova, Julia A., Framework for applications of Lean tools for visualization and analysis in the services, Scientific University of Rousse - 2011, Volume 50, Series 5.1, Lean Thinking | Lean
4. Principles, <http://leanmanufacturingtools.org/>
5. Yordanova-Dinova, S. Georgieva, Study the economic profitability of milk processing companies in the country, ISSN 1314 -1880, (www.hst.bg/bulgaria/cnference.htm).
6. Yordanova-Dinova, P., S. Georgieva, Analysis of the effectiveness in the "milk processing" companies (analysis of turnover of current assets) Scientific papers, Series B " Volume XVII, Plovdiv, 2015, pp. 261-265.

Author's contacts:

Assoc.prof. Toni Mihova, Ph.D.
 Technical University – Sofia, branch of Plovdiv
 E-mail: expert2009@abv.bg

Assoc.prof. Valentina Nikolova-Alexieva, Ph.D.
 University of food technologies,
 E-mail: valentina_nikolova@abv.bg

CAUSES, PREVENTION AND CORRECTION OF PROJECT FAILURES AND BUDGET OVERRUNS

THOMAS-MICHAEL MOSER

Abstract: *Many large ICT projects do not complete in quality, time and budget or fail completely. The influence of greater efforts and diminishing results as economic rule is highlighted. The necessity of firm negotiation procedures between the contract parties is described. Project methodologies are discussed in their relevance and contribution to project success in large and complex environments. A pragmatic flexible portfolio of methods tailored to the specific project scenario seems advisable. It is shown why social communication and clear contractual commitments must be fully integrated into the engineering process.*

Key words: *Project management methodologies, Information Communication Technologies, Software engineering, Social interactions, Project failures, Risk management, Organizational turnarounds, Negotiation procedures and contracts, Portfolio management, Agile work, HERMES, V-model, SCRUM, Team work*

1. The Situation

The situation of the Information Communication Technology (ICT) markets is characterized by increased complexity and bigger capital requirements than ever before. The failure rate of ICT projects is high. A study of 5,400 large ICT projects shows:

17% percent fail to such an extent that they pose a risk to the entity to get out of business.

45% cost more than planned; 56% do not meet full specifications and are of lower value than projected. About 7% are not ready in time [1].

The causes may appear as resulting only from reasons of economics and management. But, the causes symbolize an underlying web of interdependencies reaching above the common scope of monetary and organizational considerations [2]. They link with social interaction issues, powers of groups and roles, and political factors.

2. Purpose of the contribution

This paper should provide a short insight into the causes of project failures, how to prevent them, and what corrective measures to apply. The methodology follows a scientific approach that arguments and conclusions must be consistent, repeatable and measurable. Out of scope are deep reaching explorations because of the summary exposition of the contribution.

3. Greater efforts and diminishing results

There seems to be a paradox that greater efforts do not result in the same expected amount of

greater results. One can see that projects fail at a considerable rate even when supplied with more money or resources. This seems to be a contradiction.

The "laws" of microeconomics supply a possible answer. Economics and management do not represent a true science like physics and mathematics; hence, they must be evaluated by other means as by tests, controls and audits. Therefore, this "law" is being described by a socio-economic scientific approach. "Economics is the science which studies human behavior as a relationship between given ends and scarce means which have alternative uses"[3].

The "greater efforts and diminishing results" rule is according to Stigler "As equal increments of one input are added; the inputs of other productive services being held, constant, beyond a certain point the resulting increments of product will decrease, i.e., the marginal product will diminish" [4].

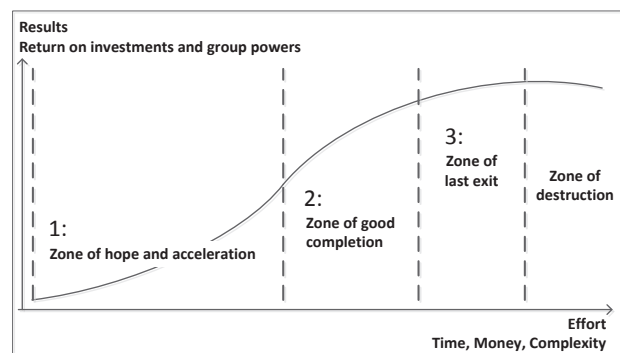


Fig. 1 Diminishing returns of results

The x-axis (right) shows the effort under the course of time, money (investments), and complexity of

interrelated matters of the project. The y-axis (up) demonstrates the outcome as of project results. The results show four zones:

Zone 1 represents “hope and acceleration”. The project kick-off took place. The team is full of spirits. During short time frames results are being achieved by an accelerating pace. Administrative overhead is at a minimum or ignored to a large extent.

Zone 2 starts with the “turning point”. More and more efforts are required to get increasing results. It may present leads to the fulfillment of project results to a large extent and gives time for consideration to finish the project under a positive close. This demands that there is acceptance that project results may be fulfilled to only 80% or similar, but are working after deployment (fit for use).

Zone 3 is described as the opportunity of the “last exit”. It culminates to the saturation point of a complete project fulfillment. Compared to the preceding two zones, only small additional increments of results are achievable compared to the efforts of time and money by increasing complexity. Under managerial human jargon there is the talk of “better more 80% results than costly striving after the one 100% illusion”. The zone 3 represents the last chance of the exit out of the project before decline to retirement/death. The exit shows four principal choices:

If agreed by the sponsors and stakeholders of the project, a reduced set of features will conclude the project, working if not totally “100% perfect”, but being operational.

The project will be terminated to be migrated to a new kick-off with lessons learned to get things better and faster.

The project will be integrated/migrated to a similar environment of specifications and scope.

The project will be closed as not being successful. All effort commitments must be written off. The motto prevails: “Better an end with horror than a horror without end [5]”.

Zone 4 is the area behind the optimum. All efforts are increasingly greater than the results being anticipated. This leads to retirement of the project by orderly closure or sudden death of activities and funding. The longer this phase lasts, the more results will be destroyed. Therefore, zone 4 can be described as “zone of destruction”.

4. Negotiation procedures to handle diminishing results and project failures

There exist inherent factors combined with efforts to achieve results which are persistent and must be considered.

The feeling and experience is widespread that commonly applied methods cannot find an answer how to better master failures and collapses of

large and complex undertakings. There is the famous widespread quote by Albert Einstein: You can never solve a problem on the level on which it was created. “Probleme kann man niemals mit derselben Denkweise lösen, durch die sie entstanden sind“ [6].

Many project management engineering methods base on the classic opinion of the hierarchical organizational model. This was driven to perfection by the German constructional context of the “Aufbau- und Ablauforganisation”, meaning defining the structural set-up (Aufbau) and the process flows (Ablauf) between the units within the organization [7].

The organization followed a strict top-down methodology. This model worked quite well during war times and the progressive build of economies after destruction. Human resources had been considered as productive factors (human “capital”, de facto an investment opportunity), but not with their own personal factors as creativity and motivational participation [8].

We are in the evolution of the information age. Information technology and communication by the internet span the whole globe; reaching from rich to poor, from small to big. The classic central organizational formula of governing from the top to all dispersed units on a global scale does not find an economic and manageable answer to rising complexity and fragmentation by inflicting different systems and cultures.

A different methodology compared to classic organizational concepts must be considered to overcome project failures and budget overruns [9].

On the stipulation that new methods must be applied, this provokes the question how to go (Что делать?) [10]. It points to a decision on whether there should be an “elite” decision to implement new organizational structures and procedures; or should this better follow a spontaneous “open” organizational pattern [11].

The necessary change in methodology presents those two alternatives, as an elite driven project or an open society (community) approach.

It would be too easy to make a strict decision for the one way or another. Mostly, one does not have green fields to start entirely from scratch with activities.

Large complex project endeavors do rarely grow from the sudden. They are based in large systems environments with traits in corporate history, where culture and interest groups of sometimes diverging goals, strategies, and tactics exist. This can be resolved by a negotiating (bargaining) process between the interest groups and their stakeholders [12].

This process aims to define the methods to handle diminishing results and projects failures. It is characterized by the following five general phases:

Phase 1: Prenegotiation: This important first phase sets the conditional framework and prepares management about vision, mission, strategies, and the tactical operational measures necessary. The result is either to proceed with opening the project, or finish the preliminary undertakings as not being feasible for going into realization.

Phase 2: Negotiators: On approval that the project will be undertaken, the negotiators will be named. They will represent the major interest groups and might form the steering committee later. The negotiators typically come from sides of the general management or others as sponsors, technical bodies, the operations department, from finance, legal and others being involved. Normally, a team approach will be followed to arrive at constructive results.

Phase 3: Negotiations: After having named the negotiators and established the body for work, the negotiation process will start. Normally, the process opens with demands from the operations department bringing a comprehensive (wish) list of requirements on the desk. This list might be extreme in demands and constitutes the first negotiation topic to start with. This might be countered by the technical department, finance, legal and motions of the sponsor. The counter offer will be very low in order to reduce the requirements list and keep spending and risks at a minimum.

Both sides will prioritize their demands and reduce them to a level that an agreement will be possible. Once, they have arrived at an agreement, an informal preliminary contract will be made. If the parties cannot reach an agreement, the situation might be at deadlock, which is up for decision by management or goes into a new round for negotiations; possibly by an appointed conciliator and/or exchange of named negotiators. The negotiation can be put on hold or break down in case.

Phase 4: Contract: Once, the informal preliminary contract has been made, the written final binding contract will be worked out. It will be sent to management and sponsor for review, ratification, and approval. It contains the purpose of the project, vision, mission and project charter; the financing, quality management; functions and roles involved, risk management and controlling principles. Attached, there will be the operational project plan with work break-down structure, timeline, milestones and controlling/updating provisions and naming of the steering committee if necessary. After signature by management, the project goes live. One of the leading standards to follow is the "PMBOK® Guide & Standards" [13], giving a suitable project management body of knowledge.

Phase 5: Project Implementation: After ratification and approval, the project will be imple-

mented as agreed in the project plan. If there are discrepancies during execution, they will be rated whether there are new features which will be handled by change management; or severe disputes which need handling by issue management.

Those five phases are mandatory for any project situation whether new, restart, migration or turnaround. Most project failures and budget overruns have their origins in missed or vague results of the negotiation process. The major roots and causes result from:

Poor or no prenegotiations have been executed. Lacking vision and managerial commitment provoke a failing project from the beginnings.

The panel of negotiators show no even representation of interests, or have important stakeholders missing. One sided contract parties may pressure decisions which carry high risks for implementation success.

The negotiations will not meet the purpose of mutual benefits. Either the operational side poses requirements which are unrealistic or vague. Or the technical side blocks changes in order not getting bothered. Finally, finance or legal might construct blocking factors without fully knowing neither technical requirements nor understanding necessities by operations.

The final contract has not been carefully worked out. Especially, the points of escalation management, risk management and sanctions must be clearly written down and agreed upon by all parties involved. Often, the project plan has not passed quality control in reference to standards and best practices incorporated. In order to "reduce" cost, the budget might be too low and does not secure enough risk coverage provisions. This is one of the major factors for budget overruns and delays or the complete project break down.

After substantial failures at the Department for Works and Pensions with huge losses of around £663m [14] and other deficiencies, the UK government has enforced a guidance which permits a valuable orientation on how to plan, evaluate and control spending [15].

5. Methods and feasibilities

The above framework of negotiations, phases and work results showed the foundation on which successful projects must be started, modified, repaired or finished.

After establishment of this framework foundation, the operational deployment of methods and metrics can be established. Often, methodology will be put in front of all activities instead of the fundamentals for adequate negotiations to enable the participating units to achieve project success [16]. Without a stable organizational framework - integrating

creative and motivated people - methodologies have got no suitable foundation.

First, an agreement (contract) must be made, and then the methodology can come into effect. Methodology is like a tool, enabling a set project situation to run most efficiently.

The methodology should secure that the work is conducted not ad-hoc driven; but planned, consistent and repeatable. Several methodologies serve this purpose.

It is not in the scope of this paper to discuss the pros and cons of several methodologies in detail; but to mention the importance on how to handle methodologies under operational deployment requirements. There is no “perfect” methodology as “one size fits all” under common jargon. Methods depend on the environment where they will have to be applied to. Economics and management can be explained from a social perspective; they represent a social science. Hence, methods in social environments are subordinated under the value judgments of parties involved, i.e. human evaluations, beliefs and trends in industry. Therefore, no “real” or “true” methodology will exist in this context.

One can only find a pragmatic answer from experience. There is the dilemma between choices and personal goals and values of decision makers, and demands by organizational bodies. When large and complex project situations are expected to suffer from severe failures or are in danger to break down, it may be best to implement a portfolio of methodologies.

The decision for only one - “the one” - methodology carries high risks of being wrong as the decision is based on personal judgments of a decision body. It seems advisable to run a portfolio approach. A portfolio is a collection of objects of a certain commonly defined type. Portfolio management has been successfully applied to the management of ICT project departments [17].

Hence, in many cases there cannot be a single-sided methodology seem appropriate. One should build a portfolio of methodologies to prevent or repair problematic project situation of failures and budget overruns.

The common project methodologies can be clustered into traditional, agile and impulse driven. The following summary list of clusters of methodologies has been derived from large project exposure over 25 years by the author. It does neither constitute a binding list nor complete overview, attributed to the requirements of an aggregated view in this paper. The focus applies to major methodologies used in software engineering methods.

Traditional: The beginnings are with the waterfall model, a methodology to have a top-down se-

quential flow with fall back options in case of failures. This may be regarded as old-fashioned today, but is still suitable for clearly defined projects over a longer timeline. Many variations exist as the spiral model among others [18].

The V-model [19], (V means “Vorgehensmodell”, translated to “procedural methodology”) has its origin in large German public projects, designed after the successful economic build-up period, dating back to 1992. Like the analogy to a V-letter shape it combines requirements analysis (left V-letter side with tests against the right opposite V-letter horizontal side on user acceptance and similar below). Each development step is mirrored by symmetrical verification and validation. The V-model has got many variations which have been successfully transferred to industry, especially in the tools and machinery industry, also to the premium car manufacturers and suppliers. It became the reference model to leading German technology export oriented conglomerates. The latest version of the V-model has advanced to the “XT” suffix which means “eXtreme Tailoring.”

The Swiss HERMES methodology [20] is developed and maintained by Swiss government authorities. It derives best practices as there are from the German V-model and CMMI® [21] and others with the focus on work flow perfection and extensive document management under strict auditing requirements. HERMES shows perfection to detail and seems applicable to complex projects as in e-government and big data environments.

Agile: The agile method is an answer to the common dilemma that undertakings seem to take too long and errors get detected too late. A solution was provided to cut ways short and get better results. The basics of agile are to leave the sequential process like in the classic waterfall methodology and intensify communication and coordination. Teams work closely together, ideally all people are in one large room. Developers and testers collaborate on the workplace together on the screen and in group meetings. Builds are done in short intervals and may contain many parallel development branches merged [22]. One method is Extreme Programming (XP) which puts teamwork and communication as top priority. The industry uses the methodology of Simultaneous Engineering [23] (or Concurrent Engineering in American English usage). The traditional separation of the construction process and the following phases of production planning is broken. Both phases are overlapping to shorten the cycles. Again, teamwork and communication are the core components in parallel and integrated workflows. It is frequently used in mechatronics which demand the integration of mechanics, electronics, electric works and software engineering.

Impulse driven: The end user (customer) is in the focus. The “end user” is understood internally as product owner as in production, marketing, sales, finance and others.

The “user stories” are in plain common language. They build with their requirements scenarios the foundation from which the development process starts. There will be short development cycles (Sprints) of only several weeks or months, or even shorter. Constant interactions between the end user and the development teams are monitored and coordinated so that the results can be better adapted to vague, changing and volatile requirements. Presently, the SCRUM framework [24] is used to a large extent under those circumstances.

6. Conclusions

The findings show that large and complex projects fail at a considerable rate. The financial losses can be very high and may endanger companies or institutions.

It seems that the causes are routed less in the wrong application of methodologies or tools. Social communication, a defined negotiation process and clear contracts are at least equally important.

One might conclude that a good team with committed management but having only average methodologies and tools might reach more than an inferior team with diverging stakeholder groups and uncommitted management even when supplied with best tools and latest methodologies.

Common practice still emphasizes the formal side of methodologies and show strong exposure to find the right tools and frameworks. Nevertheless, this attitude seems not to warrant the road to success as cases of large failed projects demonstrate.

Due to overall rapid changes in requirements, shorter time-to-market cycles and increasing process complexities, a pragmatic flexible portfolio of methods tailored to the specific situation seems advisable.

Social interactions and contractual firmness, with the full commitment of all parties involved, must be regarded as equal parts and have to be fully integrated with the formal engineering process.

REFERENCES

1. Michael Bloch, Sven Blumberg, Jürgen Laartz (2012). *Delivering large-scale IT projects on time, on budget, and on value*. Research paper. McKinsey & Company. New York.
2. Natalie Disque (2016). *Project Management Collection: Failed IT Project Business Cases, A Career Guide to Lessons Learned: Boxset of Business Cases from IT, Government and Healthcare Industries*. CreateSpace. North Charleston.
3. Lionel Charles Robbins, Baron Robbins (1932). *An Essay on the Nature and Significance of Economic Science*. Macmillan. London.
4. George J. Stigler (1987). *The Theory of Price*. Macmillan. New York.
5. Ian Slater (1987). *Orwell: The road to airstrip one*. Second Edition. P. 216. McGill-Queen's University Press. Montreal & Kingston, London, Ithaca.
6. Albert Einstein commonly quoted; nevertheless represents an unsourced attribution.
7. Erich Gutenberg (2008). *Theorie der Unternehmung*. Gabler. Wiesbaden.
8. Erich Gutenberg (2000). *Theory of the firm*. Springer. Berlin.
9. Alan G. Robinson, Dean M. Schroeder (2014). *The Idea-Driven Organization: Unlocking the Power in Bottom-Up*. Ideas Berrett-Koehler Publishers. Oakland.
10. Warren Parry (2015). *Big Change, Best Path: Successfully Managing Organizational Change with Wisdom, Analytics and Insight*. Kogan Page. London, Philadelphia.
11. Николай Гаврилович Чернышевский (1975). *Что делать?* Серия "Литературные памятники" Л. Наука. Moscow.
12. Karl R. Popper (author), Alan Ryan (introduction), E.H. Gombrich, contributor (2013). *The Open Society and Its Enemies*. Princeton University Press. Princeton.
13. Carrie J Menkel-Meadow, Andrea Kupfer Schneider, Lela Porter Love (2014). *Negotiation: Processes for Problem Solving (Aspen Casebook)* 2nd Edition. Aspen Publishers. New York.
14. Project Management Institute (2015). *A guide to the Project Management Body of Knowledge (PMBOK Guide)*. 5th edition. Project Management Institute (PMI). Newton Square, PA.
15. Charlotte Jee (2014). *HM Treasury admits to potential £663m of Universal Credit IT cost write-off*. COMPUTERWORLDUK. London.
16. HM Treasury (2015). *Guidance. Early financial cost estimates of infrastructure programmes and projects and the treatment of uncertainty and risk*. HM Treasury, Infra-structure UK. London.
17. Robert K. Wysocki (2013). *Effective Project Management: Traditional, Agile, Extreme (2013)*. 7th edition. Wiley. New York.
18. F. Warren McFarlan (1981). *Portfolio Approach to Information Systems*. Harvard Business Review. Harvard Business Publishing. Watertown, Massachusetts.

18. Richard Hall Thayer (Author), Merlin Dorfman (Author), Steve Tockey (Foreword, Contributor), Friedrich L. Bauer (Foreword), Norma E Fenton (Contributor), & 3 more (2013). *SOFTWARE ENGINEERING ESSENTIALS, Volume III: The Engineering Fundamentals (Volume 3); 4th edition*. Software Management Training Press. Carmichael, California.
19. Der Beauftragte der Bundesregierung für Informationstechnik (BIT) 2016. *V-Modell XT®*. Bundesministerium des Innern (BMI). Berlin.
20. Schweizerische Eidgenossenschaft, Bundesverwaltung (2016). Informationssteuerungsorgan des Bundes ISB. *Projektmanagement-Methode HERMES*. Schweizerische Eidgenossenschaft. Bern.
21. CMMI® Institute (2016). *CMMI® for Development (CMMI-DEV), Version 1.3*. CMII Institute. Pittsburgh, PA, US.
22. Robert C. Martin (2013). *Agile Software Development, Principles, Patterns, and Practices*. Pearson Education Limited. Harlow, United Kingdom.
23. Gerhard Pahl, W. Beitz, Jörg Feldhusen, Karl-Heinrich Grote (2007). *Engineering Design: A Systematic Approach 3rd ed.* Springer. Heidelberg.
24. Kenneth S. Rubin (2012). *Essential Scrum: A Practical Guide to the Most Popular Agile Process*. Addison-Wesley. Boston.

Authors' contacts

Thomas-Michael Moser
 ul. Hashovska Nr. 21
 4000 Plovdiv
 Bulgaria

E-mail: thomas.moser@bulcor.com



WILL AGILE METHODS RESOLVE THE COMPLEXITY CRISIS IN SOFTWARE ENGINEERING?

THOMAS-MICHAEL MOSER

Abstract: *The accelerating growth of technologies exercises increasing pressure on people and system. This complexity crisis does not only cover software engineering but extends to products, markets, companies and users. Agile work seems to provide a solution. A case of classic versus agile realization is discussed. The culture of an organization and cross functional socio-economic relations seem to be more important than monetary funds and the application of state-of-the-art technologies. It is proposed to apply a portfolio mix of classic and agile methods to become faster at less cost with better quality.*

Key words: *Project management methodologies, Information Communication Technologies, Social interactions, Project failures, Risk management, Organizational turnarounds, Agile work, HERMES, V-model, SCRUM, Team learning, Cross-functional communication*

1. Present challenges for engineering

We live in a world where we experience accelerating pressure on people and systems. The business side of software engineering enforces shorter development cycles at less cost and with higher quality results. The global market economies favor those who are faster and better on features, tailor-made flexibility and price [1].

2. Purpose and scope of this paper

This paper describes what "agile" software engineering means and what it is good for, based on extensive project experience by the author. The scope of this contribution aims to show how to use the agile engineering methodology, and how to prevent and cure project failures. Out of scope is an in-depth evaluation of methodologies due to the summary character of this paper.

3. The roads to "agile"

The decline of the classic mainframe started with the arrival of the personal computer, the decline continued when distributed processing was established. We got a "complexity crisis" which covers not only the software engineering, but extends to products, markets, companies and users [2].

What does "agile" mean? Agile is the ability to move quickly and easily ahead. It comes from the Latin verb «agere» which means to do, to make, to act. This demands to cooperate, communicate and coordinate with others to achieve a joint objective.

"Complex systems are networks made of a number of component that interact with each other, typically in a nonlinear fashion. Complex systems

may arise and evolve through self-organization, such that they are neither completely regular nor completely random, permitting the development of emergent behavior of macroscopic scales [3]".

Agile work seems to provide solutions for the situation that software engineering has to be done faster with better quality and at lower cost responding to global competitive requirements. This has been further elaborated by the <agilemanifesto.org> with the "Principles behind the Agile Manifesto" [4].

3.1. Case of classic versus agile realization

This paper will speak from a practical standpoint, trying to avoid to take a position for one or another methodology.

The comparison is a pragmatic summary list, not a theoretical essay. It serves as sketch and is not intended as proposition for doing projects in some exclusive way. The following case is fiction but based on real world work aggregated from numerous projects to present practical explanations in order to support insight and possible lessons learned.

3.2. Case scenario of development program

A large international company wants to replace the present line of a diagnostics machine. The application for this new machine must be developed as a mechatronics undertaking, i.e. for the mechanical parts by steering the sensors, actors, bus systems, embedded systems, step motors etc. The resulting data will be collected in a commercial database for which an application must be written to run the database as data warehouse with ETL-procedures.

Organization of the program: One large room for 3 teams with a basic machine prototype. One team has technical engineers for the mechanical/electrical design. The other team takes care of the mechatronics and embedded systems. The third (team 3) develops the commercial application, i.e. database, warehouse, web application and the interface to internal commercial requirements.

It is assumed that one team has a headcount from 12 to 20 people. The program consists of one program director, one project manager for each one team with up to three team leads. Programmers and team leads must be hired as contractors. The same applies to the build and configuration manager.

In order to keep this short case simple, we further assume that all other factors stay the same (*ceteris paribus* clause). This means, we assume that supporting factors like the staff function of quality management, systems architect; back office of work package planning, program office etc. are set equal to both: the classic versus agile engineering work.

We only watch the team 3 for the commercial, web based application.

3.3. Development infrastructure

Let's start with the basic software development infrastructure by mentioning the major components only.

The classic case uses classic methodology and works in a Microsoft Windows environment, following the leading German software engineering methodology, the V-Model [5] and the Swiss HERMES methodology for large and complex environments [6]. The agile side works under Linux.

Technical environment of the classic case:

Windows Server 2012R2, Framework: Visual Studio 2015 Enterprise with Team Foundation Server (TFS), HP Quality Center for testing, SQL Server 2014. NoSQL will not be being considered as it does not fit into the one vendor industrial standard SQL family line. File hosting and general project management collaboration are possible on Azure.

General program and project management are executed with MS Project. Development tracking of work packages, builds and releases will be operated with TFS [7] and Visual studio.

Technical environment of the agile case:

LAMP development environment. MariaDB version 10. For web services horizontal scaling and sharding of vague data loads, NoSQL as MongoDB, Couchbase or similar may be considered. Framework deployed: Eclipse IDE platform [8]. Open to use the following frameworks: Zend Framework, Symfony, CodeIgniter, CakePHP, Smarty, Laravel and others. The LAMP stack will run in virtual machines. File

hosting and team collaboration run on GitHub. General program and project management are conducted with Open Project.

3.4. Process flow of classic phases

3.4.1. Preliminary project work to do

List of business requirements
Development cost estimate, risk provisions
Timeline estimate
Participants, stakeholders, contracts
Draft of project plan
Proposal of alternatives to management for decision: Make - rework - on hold, or cancel

“Go” by management for the classic way **Program starts**

Program and project plans established
Monitoring, controlling and updating
Change and issue management active
Sourcing of manpower/contracting

Discussion and opinion:

MS Project gives good enterprise choice. Database and office suite in exchange with Active Directory work well. GANNT charts with resources are fully deployed.

There is easy to find certified manpower used to work in clearly defined organizational structures.

As requirements are vague in the beginnings, manpower might lack creativity and flexibility to adapt to uncertainty as development is still open and might experience early frustration by the unknown.

High costs exist for licensing frameworks and tools. Team Foundation Server and HP Quality Center are tough to handle; they show more of an “old school” forms’ orientation. One cannot live creativity, but must follow quite predefined rigid structures and workflows. The systems do not adapt to people - but people must adapt to the system.

Visual Studio is not too much different to Eclipse on the agile side, but based on latest programming languages as C#. It is preconditioned to define a unified platform. This seems better than Open Source, fragmented by fluctuations among market players and possible lack of support.

3.4.2. Concept of operations

Design of the operations of software engineering program and projects, laid down by leading project management standards as in the Project Management Body of Knowledge [9]. Technically, Visual Studio provides excellent control by unified build processes as with continuous integration [10]. Branching and merging during development are fully controllable and can be automated to a large extend.

3.4.3. Requirements and architecture

Solicitation of requirements are done with the business side. Modelling and coding of requirements with unique IDs are established. The Pflichtenheft (duty book) with requirements of business and formalized with UML use cases are compiled; which will be symmetrically mirrored by the Lastenheft (list of specifications) as programming instructions according to platform, architecture, languages and tools.

The top level design of the architecture of systems structure and flows, drawn by hardware, software, communication paths, will be worked out.

3.4.4. Detailed Design

The refinements of the architecture with the requirements will fully be signed-off by business and technical staff. The design is now ready for the programming phase.

3.4.5. IMPLEMENTATION

This covers the execution of the programming work according to the detailed design specifications and the final list of requirements being approved by all stakeholders.

3.4.6. Tests, integration, operation and maintenance

Integration, Test, Verification

Integration of modules programmed. Unit and module tests by programmers and test department. Verification that modules work under integration by quality management and quality assurance.

Tests are conducted to cases developed during the detailed design phase.

System verification and validation

Quality management and quality assurance check the working function of the complete system against the test cases derived from the requirements and architecture phase. Verification checks that the formal coding works in input, output, processing and communication. Validation confirms the systems usability whether graphical user interfaces, ergonomics of handling etc. match business requirements.

Operation and maintenance

This final phase covers end user testing against the concept operations from the side of business. It will be concluded by the sign-off by business and the management of the end user. The system goes live into production. Development and ICT hand the system over to maintenance; either in-house, at client site (on premise) or outsourced.

Discussion and opinion:

The classic way shows a clearly defined process. It can be easily monitored and controlled over all process phases. In contrast it requires a high level

of disciplines by all participants among all phases. During quite early phases, requirements and designs must be defined with their according test cases. Our time is characterized by rapid changes in end-user needs, shortage of specialists and tight budgets under increasing time pressure.

The classic methodology seems to work best if there is a very structured environment allowing for extended development cycles to provide excellent quality just in time or better.

The classic way is less feasible if business cannot clearly define use cases under rapidly changing market demands and insecure outcome of endeavors in risky global markets' environments.

If business scales down in a hierarchical way under clearly defined scope of business, the classic methodology works best.

The classic way has a less successful position to answer unpredictable changes as there are demand spikes in horizontal markets as in the information industry and online business, where flexibility and sudden changes must be served instantly.

Next, there will be the agile methodology as case sketch, considering the phases under the requirements of the development infrastructure according to 3.3. mentioned above. This is oriented on SRUM [11] as one of the leading methodology.

3.5. Process flow of agile phases

3.5.1. Preliminary project work to do

This is similar to the classic way as in 3.4.1.

“Go” by management for the agile way **The program starts**

Program and project plans established
Monitoring, controlling and updating
Change and issue management active
Sourcing of manpower/contracting

Discussion and opinion:

Open source project planning tools barely match industry standards. An own solution must be made. It is difficult to deploy GANNT charts with resources in large programs. There are third party offers available, but this works against the mandate of a unified software landscape.

Manpower seems to be abundant on a global scale, but to evaluate some kind of standardized certification, qualification and experience seems difficult if not impossible sometimes. High risk provision must be made for people not matching qualifications, detected by daily work after sourcing and being hired. High personnel role changes and fluctuation in the beginnings must be anticipated. New hires and training require additional funds and cause delays in meeting deadlines of projects.

There is a relatively small cost for the development infrastructure to start with - as Open Source has fairly low cost (but is not entirely free under commercial usage). Enterprise support requires budgets.

People in the open source community are used to work under stress; they share, are flexible and creative. They do not get frustrated on uncommon things and chaotic start-up environments and work conditions as under time and budget pressure.

Many open source tools are flexible and joyful to operate. One can live creativity. But, one needs deep practical knowledge to work efficiently with scripting by command lines during daily routines.

The Eclipse IDE framework is not too much different compared to Visual Studio. But it is heavy-weight on JAVA, coming of age [12]. Nevertheless, Open Source provides many intelligent choices and flexibility in design and application.

3.5.2. Concept of operations

The concept of operations by the agile methodology is completely different compared to the classic one. There is the assumption that requirements in the beginnings will be vague and can only be resolved by communication and coordination between the stakeholder groups [13].

Agile is a role driven development model. Three main roles exist. First, there is the Product owner. He is the communicator who works with the stakeholders and the agile team. He can transmit requirements from the business side to the technical side of the team and vice versa.

Next is the Scrum master. He is responsible for getting the team to work efficiently by removing impediments. He takes care that the work packages are on track by coordinating and supervising the backlog. The Scrum master must not have to be considered as project lead or similar. He has got no management function. His role is better seen as team facilitator to enable smooth and efficient work.

The third role is represented by the Development team. They realize the engineering work in steps of certain short work phases, called Sprints, normally within cycles of four weeks, or even shorter or longer.

3.5.3. Work flow principles

The team is self-organizing. Meetings to coordinate are held daily, mostly done as 20 minutes' stand-ups late in the morning. Work packages are pasted with sticky notes on a large poster or wall as task board. They follow a swim lane principle as planned, under work, pending, done, or similar.

When the sprint finishes, two meetings will be held. One is the sprint review to show what has been done and what points require clarification or rework.

The other meeting at the end of a sprint is the sprint retrospective where participants and stakeholder discuss the status, lessons learned and improvements.

The work packages are tracked by the work backlog showing what is planned, under work, finished; pending or needing resolution. The backlog is graphically shown as "burn-down chart", depicting the "ideal curve" of projected time of deadline, and milestones, against the curve of the actual backlog with prognosis implied. This is frequently shown in real-time reporting on monitors in the development room to give immediate feedback. The feedback comes from TFS or other systems where the development engineering work is captured. There are two types of charts. One is the sprint burn-down chart, the other one constitutes the release burn-down chart.

Discussion and opinion:

The above mentions major topics of the SCRUM methodology as an agile way to demonstrate the completely different character compared with classic software engineering methods.

Requirements under agile are derived from "user stories". This enables the end-user (business side) to formulate their wishes, demands and stipulations in common language. The requirements are then transformed into work packages, being further analyzed, formalized, and discussed by the team under coordination of the product owner and facilitation of the scrum master.

Compared to classic ways, this agile procedure shows better ways when requirements are vague and there is time pressure. This is often the way, when business says: "Just make us, ..., do this ..." without further specifying what exactly to do when and how with what concrete results. This attitude is quite typical in large environments. Agile helps to refine this and produce results during the sprint period.

Therefore, agile allows for a fast head start and can produce tangible results quickly. It may resolve parts by the "complexity crisis" mentioned at the beginnings of this paper. Vague, unknown or "tricky" requirements, demands, and expectations can be divided and put into smaller pieces to work on, to resolve; and to present working solutions within a relatively short time frame.

On the other hand, "agile" might be misunderstood as "liberty", like free style programming. This is not appropriate. Agile requires a constant personal involvement of creative and motivated participation. This is tougher to follow than a common "ordered to execute" management scheme.

Normally, organizations are still bound to a hierarchical organization set-up. Agile on the contrary is cross functional, similar to a matrix organization. Agile trades speed against clearly defined

routes and outcome. The product owner and scrum master must coordinate and balance between diverging stakeholder and group interests, but have no real decision power. This leaves a lot of responsibility to the team members.

This is less of an engineering matter but shows that a cultural, socio-economic phenomenon exists. Agile teams must be “social” people, but also excellent in technically complex matters to resolve them quickly in a fast and creative way.

Most large complex projects are either outsourced or work in environments with many specialists from different cultures and nations. They first must find a common basis – the “team building” [14] has to take place. They must understand each other and being able to communicate freely in order to falsify work results, share findings; and have to change course quickly. It seems uncommon for some people to work self-directed in a team coordinated pattern, instead of executing clear and direct work orders. And others who will act freely, might be misguided to see “their” code results as something of their “own” not sharing with others easily.

Both the product owner and the scrum master must be personalities to guide the team by coordination and protection through the ups and downs of complex projects.

With the right mix of people and a flexible organizational environment, agile may beat classical project methodologies when complexity must be quickly broken up under time and budget pressure.

4. What is the outcome of the case?

Naturally, one must follow-up what happened to the outcome of the case sketch depicted in chapter 3.1. in this paper. Here is how it turned out.

Prior projects were done in the classic way, only. Later, management demanded shorter development cycles with less budgets over time. Inspired by the trend to agile work, two large project were switched to agile methodology completely.

The first project went for two years and got stuck by internal reasons. Insufficient funds or the lack of qualified specialists did not play a major role. They could afford an excellent technical development environment and hired the best specialists they could acquire. The team finding was difficult from the start and continued. The corporate environment remained conservative and did not open up to cross functional work. The agile teams started to retreat from open work to some kind of “splendid isolation”. As there was no true cross functional support, many of the team members became frustrated, worked on their own. They participated in team communication in a formal way, and lost empathy for others.

The development process got out of control. Because “flexibility” existed as main agile principle,

individual development branches grew to more than 200 during one sprint. To make a build and release, those more than 200 branches had to be merged, causing numerous errors and exceptions during the consolidation and merge process. The build manager grew his team to five assistants writing scripts and algorithms to handle the chaotic growth wave of merges during the build and release process at the end of each sprint. He “advanced” to some “authority” of sole power owner, and became the “single point of failure” in the process. The flood of errors and break-downs of builds grew from month to month.

The project had to be stopped. Nobody could detect all error causes and repair them under reasonable economic effort anymore. As consequence, the project was outsourced and redone later. A substantial write-off and time delay had to be accepted.

The second project started two years later. The organization had acquired deeper awareness of the roots and causes. In order to get things faster and better, the headcount was doubled. The rule of diminishing results was not obeyed. Applied to the situation it means that just doubling the specialist workers does not necessarily mean to get more work done faster. Similar deficiencies as in the prior project started to build up. Wasted funds began to accumulate, deadlines of sprints and major releases were delayed more than six months, the product quality dropped from sprint to sprint.

But, the project was turned around and saved. Other departments like systems integration were bound into the development process early, participating in Scrum meetings and sending testers from systems integration to the desk of the developers. Testing was jointly done on a day by day basis. The project finalized successfully. Still under more cost and later than projected, but it worked, and the diagnostics machinery could be completed. Again, the cross functional open workstyle was the success factor.

5. Lessons learned and conclusions

It seems to be too easy to conclude that agile work saves the engineering efforts from the complexity crisis and realizes things better, faster, and at less cost.

Agile is embedded into the environment of the organization where it has got the working area. If the environment is not matching agile organization culture and set-up; then, agile concepts cannot succeed as expected.

The prerequisite is that the organizational structure must be permissible for cross functional agile work based on open communication and cooperation. If the organization is not “mature” enough [15] to provide this capable environment of creative self-

organization; projects of agile nature may end in isolated spaces, doomed for failure.

As mentioned above, agile work is less a matter of funds and technically state-of-the-art methodology. Agile will prosper when compatible cultural and socio-economic conditions exist.

Therefore, the pre-project organizational situation in a large environment must be analyzed carefully whether mature, capable prerequisites are given, before a commitment for or against agile projects can be made.

If this procedure is too difficult to decide, one might consider to follow a portfolio mix. This means to enrich the classic methodology with agile parts – as in the web technologies engineering sections for example – to better master vague requirements, hard to predict environments or challenges of horizontal big data scaling, just to mention a few scenarios.

On the other hand, during agile projects, “conservative” classic process flows can be implemented into certain cycles, mainly to include traditionally working departments in the agile work as there are mechanical engineering or systems integration to name some as an example. This gives support to narrow cultural and socio-economic distances and to bridge organizational maturity disparities.

REFERENCES

1. Kate Gillespie, H. David Hennessey. (2016). *Global Marketing*. 4th Edition. Routledge. New York, NY.
2. John L Mariotti. (2008). *The Complexity Crisis: Why too many products, markets, and customers are crippling your company--and what to do about it*. Published by Adams Media, an F+W Publications Company. Avon, Massachusetts.
3. Hiroki Sayama. (2015). *Introduction to the Modeling and Analysis of Complex Systems*. Page: 3. Open SYNY Textbooks, Milne Library. State University of New York at Genesco. Genesco, New York.
4. Website on one of the leading agile initiatives: <http://www.agilemanifesto.org/>
5. Der Beauftragte der Bundesregierung für Informationstechnik (BIT) 2016. *V-Modell XT®*. Bundesministerium des Innern (BMI). Berlin.
6. Schweizerische Eidgenossenschaft, Bundesverwaltung (2016). *Informationssteuerungsorgan des Bundes ISB. Projektmanagement-Methode HERMES*. Schweizerische Eidgenossenschaft. Bern.
7. Gary Gauvin. (2015). *Essentials of Administering Team Foundation Server 2015: Using TFS 2015 to accelerate your software development*. 1st ed. 2015 Edition. Springer Science+Business Media. New York.
8. Ram Kulkarni. (2015). *Java EE Development with Eclipse*. 2nd Edition. Packt Publishing. Birmingham, United Kingdom.
9. Project Management Institute (2015). *A guide to the Project Management Body of Knowledge (PMBOK Guide)*. 5th edition. Project Management Institute (PMI). Newton Square, PA.
10. Jez Humble. David Farley. (2011). *Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation*. Pearson Education, Boston, MA.
11. Robert C. Martin. (2013). *Agile Software Development, Principles, Patterns, and Practices*. Pearson Education Limited. Harlow, United Kingdom.
12. Jaroslav Tulach. (2008). *Practical API Design: Confessions of a Java Framework Architect*. Springer-Verlag. New York, NY.
13. Dean Leffingwell. (2011). *Agile Software Requirements: Lean Requirements Practices for Teams, Programs, and the Enterprise (Agile Software Development Series)*. Addison-Wesley, Pearson Education, Inc. Boston, MA.
14. Paul Gustavson. Stewart Liff. (2014). *A Team of Leaders: Empowering Every Member to Take Ownership, Demonstrate Initiative, and Deliver Results*. American Management Association. New York, NY.
15. CMMI® Institute (2016). *CMMI® for Development (CMMI-DEV), Version 1.3*. CMII Institute. Pittsburgh, PA.

Authors' contacts

Thomas-Michael Moser
ul. Hashovska Nr. 21
4000 Plovdiv
Bulgaria

E-mail: thomas.moser@bulcor.com



JOURNAL

**OF THE TECHNICAL UNIVERSITY - SOFIA
PLOVDIV BRANCH, BULGARIA
“FUNDAMENTAL SCIENCES AND APPLICATIONS”**

publishes new and original results in the fields of MATHEMATICS, MECHANICS, PHYSICS, CHEMISTRY, ECONOMICS and their applications in technical sciences. They can further be expanded to be published elsewhere.

The manuscript should be sent to the Journal Editor-in-Chief

Prof. Michail Petrov, PhD

E-Mail: journal@tu-plovdiv.bg

Acceptance for publication will be based on a positive recommendation from the Journal Editorial Board.

Technical University - Sofia, Plovdiv Branch
25, Tsanko Dyustabanov St., Plovdiv 4000, BULGARIA

<http://www.tu-plovdiv.bg>

ISSN 1310-8271