TECHNICAL UNIVERSITY OF SOFIA

FACULTY AUTOMATION

DEPARTMENT THEORETICAL ELECTRICAL ENGINEERING

SUMMER SCHOOL

ADVANCED ASPECTS OF THEORETICAL ELECTRICAL ENGINEERING - SOZOPOL'09

in the framework of

THE DAYS OF SCIENCE OF THE TECHNICAL UNIVERSITY OF SOFIA, SOZOPOL, BULGARIA, SEPT. 2009

Part I: Plenary Lectures

Edited by Valeri Mladenov

The main sponsor of the Summer School is:
RESEARCH AND DEVELOPMENT SECTOR,
TECHNICAL UNIVERSITY – SOFIA

Dedicated to prof. Mincho Zlatev's 100 years birthday anniversary

ISSN: 1313-9479

Organizing Committee

Honorary Chairmen:

V. Georgiev – Bulgaria
S. Farchy – Bulgaria
L. Kolev – Bulgaria
S. Papasow – Bulgaria
V. Savov – Bulgaria

Chair:

V. Mladenov, TU-Sofia, Bulgaria

Co Chairs:

S. Guninski, TU-Sofia, Bulgaria
J. Georgiev, TU-Sofia, Bulgaria
I. Iachev, TU-Sofia, Bulgaria

Members:

K. Brandisky, TU-Sofia, Bulgaria
A. Chervenkov, TU-Sofia, Bulgaria
K. Ivanov, TU-Sofia, Bulgaria
N. Petkova, TU-Sofia, Bulgaria
S. Petrakieva, TU-Sofia, Bulgaria
N. Radev, TU-Sofia, Bulgaria
K. Stanchev, TU-Sofia, Bulgaria
R. Stancheva, TU-Sofia, Bulgaria
I. Tabahnev, TU-Sofia, Bulgaria
S. Terzieva, TU-Sofia, Bulgaria
K. Todorova, TU-Sofia, Bulgaria
G. Toshev, TU-Sofia, Bulgaria
S. Vladov, TU-Sofia, Bulgaria
I. Trushev, TU-Sofia, Bulgaria
G. Tsenov, TU-Sofia, Bulgaria
I. Yacheva, TU-Sofia, Bulgaria

International Programme Committee

D. Baldomir, Spain
A. Bossavit, France
A. Brykalski, Poland
H. Butterweck, Netherlands
T. Chady, Poland
K. Demirtchyan, Russia
S. Gratkowski, Poland
K. Hameyer, Germany
W. John, Germany
L. Klinkenbusch, Germany
L. Kolev, Bulgaria
A. Kost, Germany
Z. Leonowicz, Poland
N. Mastorakis, Greece
W. Mathis, Germany
V. Mladenov, Bulgaria
M. Ogorzałek, Poland
S. Osowski, Poland
L. Pichon, France
B. Reljin, Serbia
R. Sikora, Poland (Honorary Chairman)
L. Sumichrast, Slovakia
J. Sykulski, UK
M. Tadeusiewicz, Poland
H. Toepfer, Germany
H. Uhlmann, Germany
R. Weigel, Germany

PREFACE

These Proceedings are organized in two parts and contain the plenary lectures and the regular papers presented at the 7th Summer School *Sozopol'09*, which took place in Sozopol, Bulgaria, between 20 and 23 Sept. 2009 in the framework of the Days of the Science of the Technical University of Sofia. The Summer School covers the advanced aspects of Theoretical Electrical Engineering and it is a platform for postgraduate training of Ph.D. students and young scientists. During the Summer School well-known experts presented some advanced aspects of circuits and systems theory, electromagnetic field theory and their applications. Apart from the educational part of the Summer School a presentation of original authors' papers took place.

The main topics of the Summer School *Sozopol'09* include Circuits and Systems Theory, Signal Processing and Identification Aspects, Electromagnetic Fields, Theoretical Concepts, Applications and Innovative Educational Aspects.

The Summer School *Sozopol'09* has been organized by the Department of Theoretical Electrical Engineering of the Technical University of Sofia with the sponsorship of the Research and Development Sector of the Technical University of Sofia. This has been the seventh edition of the event, after the Summer Schools in 1986, 1988, 2001, 2002, 2005 and 2007. This Summer School is especially dedicated to prof. Mincho Zlatev's 100 years birthday anniversary.

There were 52 participants from 5 different countries at the Summer School this year. There were 14 plenary lectures and 29 regular papers that are published in these Proceedings. Providing the recent advances in Theoretical Electrical Engineering the Proceedings will be of interest to all researchers, educators and Ph.D. students in the area of Electrical Engineering.

Special thanks are due to the Research and Development Sector, Faculty of Automation and the Section of Social Services of the Technical University of Sofia about the overall support of the event. We also want to thank to the World Scientific and Engineering Academy and Society (WSEAS) and company Antipodes Ltd. which partially sponsored the event. We hope to meet again in the following edition of the Summer School to continue the good tradition and collaboration in the field of Theoretical Electrical Engineering.

Organizing Committee
Sofia, Oct. 2009

CONTENTS

IN MEMORIAM TO PROF. MINCHO ZLATEV

Prof. M. P. Zlatev is one of the pioneers who create and develop the engineer education in Bulgaria after the Second World War, particularly in VMEI (the Higher Mechanical and Electrical Engineering Institute) presently the Technical University – Sofia.

As man with remarkable erudition, thoroughness, knowledge and scientific achievements he exerted positive influence on generations of lecturers, engineers and university students both in the field of electrical engineering and their human personalities.

He was born on 11.10.1909 in the town of Sevlievo. He got his secondary school diploma from the Aprilovska gimnazia (Aprilov's High School) in the town of Gabrovo.

In 1931 he graduated from the Higher Technical School in Toulouse, France and received a diploma for an electrical engineer.

He defended the DSc thesis in the Leningrad Polytechnical Institute (USSR) in 1966. His official opponent was academic Neiman.

Starting in 1945 until his retirement (1976) he was the HOD.

His main teaching and scientific subject was theoretical electrical engineering.

Following are his administrative and scientific positions:

- Member of the Academic Council and the Scientific Council of theVMEI from 1954;
- Vice rector on R & D and production activities during periods 1960 – 1962 and 1968 - 1970;
- Doctor Honoris Causa of the Scientific Academy, Toulouse, France;
- Senior member of the IEEE, New York, since 1966;
- A sitting member of the Bulgarian S.I.G.R.E. Delegation since 1956;
- A member of International Association of Informatics since 1974;
- A delegate – expert of Bulgaria in MEK since 1974.

He is a bearer of the following medals:

„Kiril and Metodi", I and II classes (1957, 1959, 1972),

„Cherveno zname na truda" („A red banner of labour") – 1970.

He retired from the VMEI in 1976 and then continued to work in Bulgarian Science Academy.

He passed away on 31.05.1991 in Sofia.

In 1931, after he had finished his engineering education in France, he worked in Electrotechnics Laboratory in the Office of Trams and Light in Sofia and later he was in charge of it.

In 1945 (16.10.1945) he was elected by the Academic Council in the Higher Technical School Sofia (HTS) as an extraordinary professor (the decree № 144 of the Bulgarian Regency). Two days later the HTS - Sofia was renamed as State Polytechnic (State Journal № 248 / 24.10.1945) with two faculties: Mechanical - Electrical - Technological Faculty and Civil Architectural Faculty.

The first curriculums and syllabuses were worked out by both Prof. M. P. Zlatev and Professors Grigor Usunov and Nancho Nanchev. In this respect they are among the pioneers in creation and development the engineering education in Bulgaria after post - war years.

Prof. M. P. Zlatev was an HOD of the first Electrical Department at the State Polytechnic namely Theory of Electricity and Measurement Technics and he remained in this position until his retirement in 1976.

He was nominated as a First Secretary of the section of Electricity at the Technological Faculty of the State Polytechnic. He stayed at this position from 1945 to 1950.

He became a member of the Faculty Council of the new established Faculty of Electrical Engineering at Polytechnic and remained such a mumber till 1974. His Department was renamed as Theoretical and Measurement Technics Department and he continued to be the HOD.

On the 10.06.1953 the State Polytechnic was divided in four Institutes, one of each is Mechanical - Electrical Institute with two faculties. The Theoretical and Measurement Technics Department was included in the structure of the Electrical Faculty at the MEI (26.10.1953) and prof. M. P. Zlatev was elected as a member of the Faculty Council.

The Theoretical and Measurement Technics Department formed two departments: the Theoretical Electrical Department (TED) and the Electrical Measurement Department. Prof. M. P. Zlatev remained the HOD of the TED.

Later on the two departments become a part of the new formed Faculty of Automatics (07.08.1974).

On the 12.12.1945 prof. M. P. Zlatev delivered his first lecture on the Electricity Theory. He also delivered lectures on Electrical Measurement Theory – 1948.

Prof. Zlatev's core courses were the Theoretical Electrical Engineering – part I, II and III. His lectures were remarkable. Prof. Zlatev gave from memory all mathematical expressions and conclusions on black broad in a simple and clear way.

As an examiner he treated every student with respect. Prof. M. P. Zlatev put in his notebook all questions of the exams and later, during the defence of the of a diploma paper, he used to ask same questions.

His personal life is an example to follow and it is a good confirmation of the authority of the VMEI.

Prof. Zlatev is the author of 13 volumes textbooks on Theory of Electricity, Basic Electrotechnics, and Theoretical Electrotechnics.

Some of the titles of the textbooks written by prof. M. P. Zlatev are:

Zlatev, M. Theory of Electricity, part I, Sofia, 1947
Zlatev, M. Theory of Electricity, part II, "Science and Art", 1949
Zlatev, M. Measurement Electrotechnic, 1950
Zlatev, M. Basic Electrical Engineering, vol. I, Sofia, "Tehnica", 1954, 1959
Zlatev, M. Basic Electrical Engineering, vol. II, Sofia "Tehnika", 1957, 1961
Zlatev, M. Basic Electrical Engineering, vol. III, Sofia "Tehnika", 1958, 1962
Zlatev, M. Basic Electrical Engineering, Sofia "Tehnika", 1964
Zlatev, M. Theoretical Electrotechnics, vol. I, Sofia, "Tehnika", 1972
Zlatev, M. Theoretical Electrotechnics, vol. II, Sofia, "Tehnika", 1973
Zlatev, M. Theoretical Electrotechnics, vol. III, Sofia, "Tehnika", 1974

He has written 160 scientific papers and reports. Many of them are published abroad. The first scientific paper of the department abroad is his "Ein neues Zahlen-system zur Berechnung Linear Netzgebilde", published in 1955 in German magazines Dtsch. Electroteschnic.

Prof. Zlatev has also been responsible for many R & D projects and a co-author of 11 patents.

Until now in the Electrical Engineering Department, that was founded by prof. Zlatev, four DSc thesises and more than 35 PhD thesises have been defended. In this manner a staff of many people has been trained and they constitute a base of similar departments in other universities in Bulgaria.

Some 50 monographs, textbooks and manual connected with the training of the electrical engineering students have been published and they are used not only in the TU – Sofia but in other universities in Bulgaria.

Department Theoretical Electrical Engineering,
Technical University of Sofia

ARTIFICIAL INTELLIGENCE FOR SOLVING PARTIAL DIFFERENTIAL EQUATIONS

Nikos E. Mastorakis

WSEAS Research Department, Agiou Ioannou Theologou 17-23, Zografou,
15773, Athens, GREECE
and
Technical University of Sofia, English Language Faculty, Industrial Engineering Department,
Sofia 1000, BULGARIA, mastor@wseas.org

Abstract: *Genetic Algorithms plus Nelder-Mead Optimization for several problems of FEM (Finite Elements Methods) solution of PDEs have been proposed by the author since 1996. Recently the method has been applied on Non-linear Problems in Fluid Mechanics. For example the p-Laplacian operator, or the p-Laplace operator, is a quasilinear elliptic partial differential operator of 2nd order. The p-Laplacian equation is a generalization of the PDE of Laplace Equation and in this paper, we present a way of its solution using Finite Elements. Our method of Finite Elements leads to an Optimization Problem that can be solved by an appropriate combination of Genetic Algorithms and Nelder-Mead. On the other hand recently, the existence of a nontrivial solution to the nonlinear Schrodinger-Maxwell equations in R3, assuming on the nonlinearity the general hypotheses introduced by Berestycki & Lions has been proved. In this paper, the Numerical Solution of the system of PDEs of Schrodinger-Maxwell equations (with a general nonlinear term) via Finite Elements and Genetic Algorithms with Nelder-Mead is proposed. The method of Finite Elements and Genetic Algorithms with Nelder-Mead that has been proposed by the author recently is also used.*

1. INTRODUCTION

Many nonlinear problems in physics and mechanics are formulated in equations that contain the p-Laplacian, (i.e. the p-Laplace operator), where the p-Laplacian operator is defined as follows

$$\Delta_p u := \operatorname{div}\left(|\nabla u|^{p-2}\nabla u\right)$$

In a recent paper, [17], Bognar presented a very interesting numerical and analytic investigation of problems of fluid mechanics that are described with PDEs containing the p-Laplacian operator. Previous publications (also reported in [17]) include reaction-diffusion problems, non-Newtonian fluid flows [18], fluid flows through certain types of porous media ([19], [20], the Lane-Emden equations for equilibrium configurations of spherically symmetric gaseous stellar objects [21], singular solutions for the Emden-Fowler equation [22] and the Einstein-Yang-Mills equations [23], the existence and nonexistence of black hole solutions, nonlinear elasticity [24], glaciology [25] and petroleum extraction [26]. It is clear that for $p=2$: $\Delta p=\Delta$ The study of the p-Laplacian equation started more than thirty years ago. In recent years, rapid development has been achieved for the study of equation involving operator Δp and a

vast literature has appeared on the theory of quasilinear differential equations.). In [27] Strikwerda summarized many Finite Difference Schemes for PDEs. Other relevant studies can be found in [28], [29] and [30].

In [17], Bognar had studied the equation of turbulent filtration in porous media

$$\theta \frac{\partial \rho}{\partial t} = c^{\alpha} \lambda \operatorname{div}\left(\left| \nabla \rho^{n} \right|^{p-2} \nabla \rho^{n} \right), \tag{1}$$

where $\theta > 0$ and the constants $n > 0$ and $p > 1$ satisfy $np > 1$. If we scale out the constants in (1), we derive

$$\frac{\partial u}{\partial t} = \Delta_p \left(u^n \right) \tag{2}$$

where a particular case of (2) is the non-Newtonian filtration equation

$$\frac{\partial u}{\partial t} = \Delta_p u \tag{3}$$

which is also called evolution p-Laplacian equation. The case $p > 1 + \frac{1}{n}$ is called the slow diffusion and the case $p < 1 + \frac{1}{n},$ the fast diffusion.

Also in the paper [17], Bognar studied the equation

$$\frac{\partial u}{\partial t} = \operatorname{div}\left(\left| \nabla u \right|^{p-2} \nabla u \right) + \lambda u^{q}, \tag{4}$$

where $p > 1,\ q > 0$ and λ are some constants, in which the nonlinear term λu^{q} describes the nonlinear source in the diffusion process, called "heat source" if $\lambda > 0$ and "cold source" if $\lambda < 0.$ Just as the Newtonian equation (p=2), the appearance of nonlinear sources will exert a great influence to the properties of solutions and the influence of "heat source" and "cold source" is completely different.

In [31], an attempt is made by the author to solve the equations (2), (3) and (4) using various numerical schemes.

In this paper we will solve the boundary value problem

$$\operatorname{div}\left(\left| \nabla u \right|^{p-2} \nabla u \right) = 0$$

where u is known on the boundary of our domain using Variational Techniques (Finite elements).

The Problem is reduced to an Optimization problem that can be solved by Genetic Algorithms with Nelder-Mead. An early paper of the author with the title “Solving Differential Equations via Genetic Algorithms” was presented in [1]. Actually, the

author presented in 1996 the solution of ODE and PDE using Genetic Algorithms optimization, while the author use the same method to solve various problems in [2]÷[9].

The main Results are given in Section 2 and a numerical example illustrates the method in Section 3.

A discussion for the numerical solution of (2), (3) and (4) by finite elements is also included in Section 4.

2. MAIN RESULTS

We start solving the boundary value problem

$$\operatorname{div}\left(\left|\nabla u\right|^{p-2}\nabla u\right)=0 \tag{4}$$

where *u* is a known function on the boundary of our domain.

As one can see in [32] and [33], the solution of this p-Laplacian equation with Dirichlet boundary conditions in a domain Ω is the minimizer of the energy functional

$$J(u)=\int\left|\nabla u\right|^{p}dv \tag{5}$$

We consider that *u* is written as

$$u=\sum_{n}\lambda_n f_n$$
$$u=\sum_{n} f_n \tag{6}$$

where λ_n have been incorporated in f_n

So, we have the minimization problem

$$\min\int\left|\nabla(\sum_{n} f_n)\right|^{p}dv$$

One can select a triangular mesh and appropriate functions f_n that have non-zero value only in the *n-th* triangle ("finite elements"). So, in a triangular mesh, for example of $\square^2$, we can have $f_n=a_n x+b_n y+c_n$ for the *n-th* triangle. Without loss of generality we consider the case $\square^2$ here *u* in (4).

To avoid to write **continuity conditions** on the common vertices of the triangles of the mesh, one can find that in the *n-th* triangle of the points *s,q,r* (see Figure 1)

$$u_s=a_n x_s+b_n y_s+c_n \tag{7.1}$$
$$u_q=a_n x_q+b_n y_q+c_n \tag{7.2}$$

$$u_r = a_n x_r + b_n y_r + c_n \tag{7.3}$$

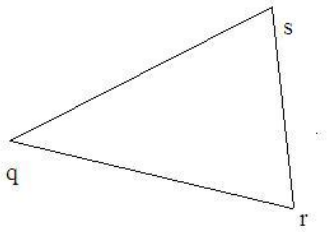


Fig. 1. A triangle in a 2-D mesh

There three equations can be solved with respect to a_n, b_n, c_n and give

$$a_n = \frac{\begin{vmatrix} u_s & y_s & 1 \\ u_q & y_q & 1 \\ u_r & y_r & 1 \end{vmatrix}}{D} \tag{8.1}$$

$$b_n = \frac{\begin{vmatrix} x_s & u_s & 1 \\ x_q & u_q & 1 \\ x_r & u_r & 1 \end{vmatrix}}{D} \tag{8.2}$$

$$c_n = \frac{\begin{vmatrix} x_s & y_s & u_s \\ x_q & y_q & u_q \\ x_r & y_r & u_r \end{vmatrix}}{D} \tag{8.3}$$

$$D = \begin{vmatrix} x_s & y_s & 1 \\ x_q & y_q & 1 \\ x_r & y_r & 1 \end{vmatrix}$$

(which is by the way 2*E where E is the algebraic area of the triangle)

So, from the minimization problem

$$\min \int \left| \nabla \left(\sum_n f_n \right) \right|^p dv$$

we find the equivalent minimization problem

$$\min \int \left| \phi(u_n) \right|^p dv \tag{9}$$

where

$\phi(u_n)$ is the function that we find after replacing

$f_n = a_n x + b_n y + c_n$

In

$\nabla(\sum_n f_n)$

and

a_n, b_n, c_n

are evaluated using (8.1), (8.2), (8.3) for each triangle of the mesh.

Equation (9) can be solved now by a variety of techniques. The author uses Genetic Algorithms with Nelder-Meade for Non-linear Problems as in [2], [3], [4], [5], [6], [7], [8].

The same optimization scheme: Genetic Algorithms with Nelder-Meade will be also applied for (9).

Before proceeding in the solution of the problem, some background on GA (Genetic Algorithms) and Nelder-Mead is necessary. In [4], the author has also proposed a hybrid method that includes a) Genetic Algorithm for finding rather the neiborhood of the global minimum than the global minimu itself and b) Nelder-Mead algorithm to find the exact point of the global minimum itself.

So, with this Hybrid method of Genetic Algorithm + Nelder-Mead we combine the advantages of both methods, that are a) the convergence to the global minimum (genetic algorithm) plus b) the high accuracy of the Nelder-Mead method.

If we use only a Genetic Algorithm then we have the problem of low accuracy.

If we use only Nelder-Mead, then we have the problem of the possible convergence to a local (not to the global) minimum.

These disadvantages are removed in the case of our Hybrid method that combines Genetic Algorithm with Nelder-Mead method. We recall the following definitions from the Genetic Algorithms literature:

Fitness function is the objective function we want to minimize.

Population size specifies how many individuals there are in each generation. We can use various Fitness Scaling Options (rank, proportional, top, shift linear, etc), as well as various Selection Options (like Stochastic uniform, Remainder, Uniform, Roulette, Tournament). Fitness Scaling Options: We can use scaling functions. A Scaling function specifies the function that performs the scaling. A scaling function converts raw fitness scores returned by the fitness function to values in a range that is suitable for the selection function.

We have the following options:

Rank Scaling Option: scales the raw scores based on the rank of each individual, rather than its score. The rank of an individual is its position in the sorted scores. The

rank of the fittest individual is 1, the next fittest is 2 and so on. Rank fitness scaling removes the effect of the spread of the raw scores.

Proportional Scaling Option: The Proportional Scaling makes the expectation proportional to the raw fitness score. This strategy has weaknesses when raw scores are not in a "good" range.

Top Scaling Option: The Top Scaling scales the individuals with the highest fitness values equally.

Shift linear Scaling Option: The shift linear scaling option scales the raw scores so that the expectation of the fittest individual is equal to a constant, which you can specify as Maximum survival rate, multiplied by the average score.

We can have also option in our Reproduction in order to determine how the genetic algorithm creates children at each new generation.

For example,

Elite Counter specifies the number of individuals that are guaranteed to survive to the next generation.

Crossover combines two individuals, or parents, to form a new individual, or child, for the next generation.

Crossover fraction specifies the fraction of the next generation, other than elite individuals, that are produced by crossover.

Scattered Crossover: Scattered Crossover creates a random binary vector. It then selects the genes where the vector is a 1 from the first parent, and the genes where the vector is a 0 from the second parent, and combines the genes to form the child.

Mutation: Mutation makes small random changes in the individuals in the population, which provide genetic diversity and enable the GA to search a broader space. Gaussian Mutation: We call that the Mutation is Gaussian if the Mutation adds a random number to each vector entry of an individual. This random number is taken from a Gaussian distribution centered on zero. The variance of this distribution can be controlled with two parameters. The Scale parameter determines the variance at the first generation. The Shrink parameter controls how variance shrinks as generations go by. If the Shrink parameter is 0, the variance is constant. If the Shrink parameter is 1, the variance shrinks to 0 linearly as the last generation is reached.

Migration is the movement of individuals between subpopulations (the best individuals from one subpopulation replace the worst individuals in another subpopulation). We can control how migration occurs by the following three parameters.

Direction of Migration: Migration can take place in one direction or two. In the so-called “Forward migration” the nth subpopulation migrates into the (n+1)'th subpopulation. while in the so-called “Both directions Migration”, the nth subpopulation migrates into both the (n-1)th and the (n+1)th subpopulation.

Migration wraps at the ends of the subpopulations. That is, the last subpopulation migrates into the first, and the first may migrate into the last. To prevent wrapping, specify a subpopulation of size zero.

Fraction of Migration is the number of the individuals that we move between the subpopulations. So, Fraction of Migration is the fraction of the smaller of the two subpopulations that moves. If individuals migrate from a subpopulation of 50 individuals into a population of 100 individuals and Fraction is 0.1, 5 individuals (0.1 * 50) migrate. Individuals that migrate from one subpopulation to another are copied. They are not removed from the source subpopulation. *Interval of Migration* counts how many generations pass between migrations.

The Nelder-Mead simplex algorithm appeared in 1965 and is now one of the most widely used methods for nonlinear unconstrained optimization [33]÷[35]. The Nelder-Mead method attempts to minimize a scalar-valued nonlinear function of n real variables using only function values, without any derivative information (explicit or implicit).

The Nelder-Mead method thus falls in the general class of direct search methods. The method is described as follows: Let $f(x)$ be the function for minimization.

x is a vector in n real variables. We select $n+1$ initial points for x and we follow the steps:

Step 1. *Order*.

Order the $n+1$ vertices to satisfy $f(x_1) \le f(x_2) \le \ldots \le f(x_{n+1})$, using the tie-breaking rules given below.

Step 2. *Reflect*.

Compute the *reflection point* x_r from $x_r = \bar{x} + \rho(\bar{x} - x_{n+1}) = (1+\rho)\bar{x} - \rho x_{n+1}$, where $\bar{x} = \sum_{i=1}^{n} x_i / n$ is the centroid of the n best points (all vertices except for x_{n+1}). Evaluate $f_r=f(x_r)$. If $f_1 \le f_r < f_n$ accept the reflected point x_r and terminate the iteration.

Step 3. *Expand*.

If $f_r < f_1$, calculate the *expansion point* x_e,

$x_e = \bar{x} + \chi(x_r - \bar{x}) = \bar{x} + \rho\chi(\bar{x} - x_{n+1}) = (1+\rho\chi)\bar{x} - \rho\chi x_{n+1}$

and evaluate $f_e=f(x_e)$. If $f_e < f_r$, accept x_e and terminate the iteration; otherwise (if $f_e \ge f_r$), accept x_r and terminate the iteration.

Step 4. *Contract*.

If $f_r \ge f_n$, perform a *contraction* between $\bar{x}$ and the better of x_{n+1} and x_r.

Outside. If $f_n \le f_r < f_{n+1}$ (i.e. x_r is strictly better than x_{n+1}), perform an *outside contraction*: calculate $x_c = \bar{x} + \gamma(x_r - \bar{x}) = \bar{x} + \gamma\rho(\bar{x} - x_{n+1}) = (1+\rho\gamma)\bar{x} - \rho\gamma x_{n+1}$

and evaluate $f_c = f(x_c)$. If $f_c \le f_r$, accept x_c and terminate the iteration; otherwise, go to step 5 (perform a shrink).

Inside. If $f_r \ge f_{n+1}$, perform an *inside contraction*: calculate

$x_{cc} = \bar{x} - \gamma(\bar{x} - x_{n+1}) = (1-\gamma)\bar{x} + \gamma x_{n+1}$, and evaluate $f_{cc} = f(x_{cc})$. If $f_{cc} < f_{n+1}$, accept x_{cc} and terminate the iteration; otherwise, go to step 5 (perform a shrink).

Step 5. *Perform a shrink step*.

Evaluate f at the n points $v_i = x_1 + \sigma (x_i - x_1)$, $i = 2, \ldots, n+1$. The (unordered) vertices of the simplex at the next iteration consist of $x_1, v_2, \ldots, v_{n+1}$.

After this preparation, we are ready to solve the

$\min \int |\phi(u_n)|^p \, dv$

of (9) as minimization problem.

The minimization is achieved by using Genetic Algorithms (GA) and the method of Nelder-Mead exactly as we described previously. We can use the MATLB software package

(MATLAB, Version 7.0.0, by Math Works).

In the next numerical example (Section 3) our GA has the following Parameters

Population type:
Double Vector Population size: 30
Creation function: Uniform
Fitness scaling: Rank
Selection function: roulette
Reproduction: 6 – Crossover fraction 0.8

Mutation: Gaussian – Scale 1.0,
Shrink 1.0
Crossover: Scattered
Migration: Both – fraction 0.2, interval: 20
Stopping criteria: 50 generations

3. NUMERICAL EXAMPLE

Consider now the following problem (Fig. 2)

$$\operatorname{div}\left(|\nabla u|^{p-2} \nabla u\right) = 0 \tag{4}$$

in the domain

$$u \in [0,2]\times[0,2] - [0,1]\times[0,1]$$

with $u = 0$ in the external boundary:

$x = \pm 2, -2 \le y \le 2$
$y = \pm 2, -2 \le x \le 2$

and $u = 1$ in the internal boundary

$x = \pm 1, -1 \le y \le 1$
$y = \pm 1, -1 \le x \le 1$

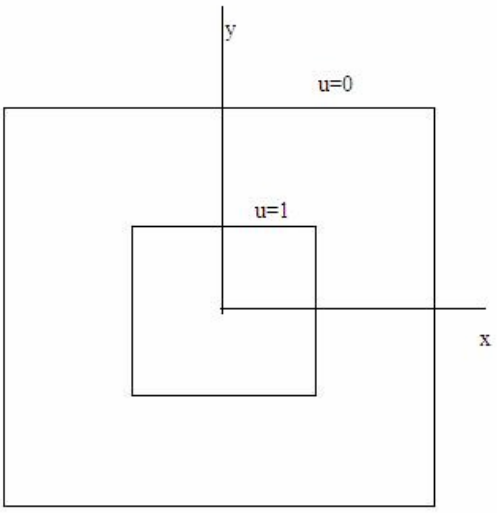


Fig. 2

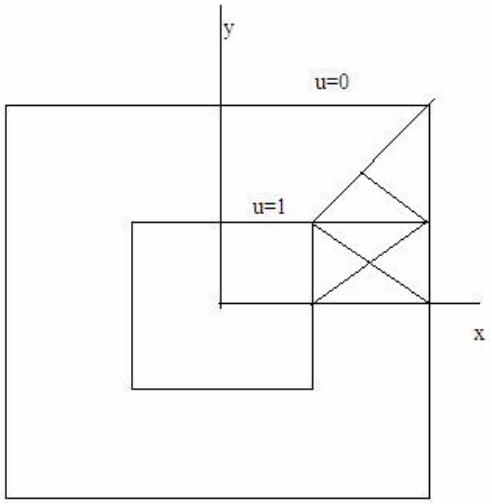


Fig. 3

Due to symmetry, we can split the domain in 8 same trapezoids (trapezia). It is sufficient to solve the problem

$$\operatorname{div}\left(|\nabla u|^{p-2}\nabla u\right) = 0$$

in one of them with the boundary conditions $u = 0$ in the external boundary and $u = 1$ in the internal boundary.

Taking one of these trapezoids and splitting it into 6 triangles like in Fig. 3, we have in some enlargement the following Figure: Fig. 4

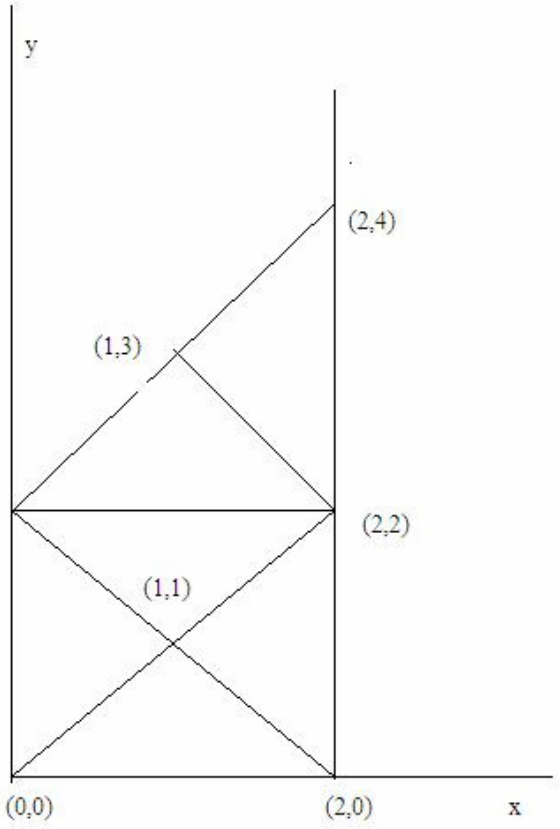


Fig. 4

We consider as $u_1, u_2, u_3, u_4, u_5, u_6, u_7$ the value of the u at the points $(0,0), (2,0), (2,2), (2,4), (1,3), (0,2), (1,1)$
i.e.

$u_1 = u(0,0),$

$u_2 = u(2,0),$

$u_3 = u(2,2),$

$u_4 = u(2,4),$

$u_5 = u(1,3),$

$u_6 = u(0,2),$

$u_7 = u(1,1)$

Then by considering

$$u_p = a_n x_s + b_n y_s + c_n \tag{7.1}$$

$$u_q = a_n x_q + b_n y_q + c_n \tag{7.2}$$

$$u_r = a_n x_r + b_n y_r + c_n \tag{7.3}$$

in every one of the 6 triangles, we solve as in (8.1), (8.2), (8.3) and finally we introduce to

$$J(u) = \int |\nabla u|^p \, dv \tag{5}$$

We have, considering also that $u_1 = u_6 = 1$ and $u_2 = u_3 = u_4 = 0$

So, after some algebraic manipulation we find that we have to minimize the quantity *I*

where

$$I = |2u_5|^p + \left|\sqrt{1^2 + (1-2u_5)^2}\right|^p + \left|\sqrt{1^2 + (1-2u_7)^2}\right|^p + \\ + |(2-2u_7)|^p + \left|\sqrt{1^2 + (1-2u_7)^2}\right|^p + |(2u_7)|^p \tag{10}$$

with respect to u_5, u_7

Now, in order to find the global minimum of *I* we use **GA** *(Population type: Double Vector Population size: 30 / Creation function: Uniform /Fitness scaling: Rank / Selection function: roulette / Reproduction: 6 – Crossover fraction 0.8 / / Mutation: Gaussian – Scale 1.0, Shrink 1.0 / / Crossover: Scattered / Migration: Both – fraction 0.2, interval: 20 /Stopping criteria: 50 generations)*
and continue with **Nelder-Mead**

So we find the following results for different values of *p.*

p	u_5	u_7	*I*
2	0.2500	0.5000	5.5000
3	0.3145	0.5000	5.4623
4	0.3471	0.5000	5.4280
5	0.3678	0.5000	5.3994
6	0.3824	0.5000	5.3754
7	0.3935	0.5000	5.3550
8	0.4024	0.5000	5.3373
10	0.4155	0.5000	5.3082
20	0.4468	0.5000	5.2246
50	0.4721	0.5000	5.1375
200	0.4903	0.5000	5.0582

(The combined method of Genetic Algorithms and Nelder-Mead was proposed by the author in 2005, [2]÷[9] , while the author proposed the solution of ODEs and PDES since July 1996 (See[1])).

4. SOLUTION OF THE EQUATIONS (2), (3) AND (4)

We remind the problems:

$$\theta \frac{\partial \rho}{\partial t} = c^{\alpha} \lambda \operatorname{div}\left(\left| \nabla \rho^{n} \right|^{p-2} \nabla \rho^{n} \right), \tag{1}$$

If we scale out the constants in (1), we derive

$$\frac{\partial u}{\partial t} = \Delta_p \left(u^n \right) \tag{2}$$

where a particular case of (2) is the non-Newtonian filtration equation

$$\frac{\partial u}{\partial t} = \Delta_p u \tag{3}$$

and

$$\frac{\partial u}{\partial t} = \operatorname{div}\left(\left| \nabla u \right|^{p-2} \nabla u \right) + \lambda u^{q}, \tag{4}$$

Consider that u can be written as

$u = \sum_n \lambda_n(t) f_n$ or $u = \sum_n f_n(t)$ where λ_n have been incorporated to $f_n(t)$

In this "dynamic" case, in a triangular mesh of $\square^2$ we can have

$f_n = a_n(t)x + b_n(t)y + c_n(t)$ for the *n-th* triangle.

$$u_s = a_n(t)x_s + b_n(t)y_s + c_n(t) \tag{7.1}$$

$$u_q = a_n(t)x_q + b_n(t)y_q + c_n(t) \tag{7.2}$$

$$u_r = a_n(t)x_r + b_n(t)y_r + c_n(t) \tag{7.3}$$

Of course, we can use higher degree polynomials like quadratic or cubic.
For quadratic:

$$u_s = a_n(t)x_s + b_n(t)y_s + c_n(t) + d_n(t)x_s^2 + e_n(t)d_s^2 + h_n(t)x_s y_s$$

$$u_q = a_n(t)x_q + b_n(t)y_q + c_n(t) + d_n(t)x_q^2 + e_n(t)d_q^2 + h_n(t)x_q y_q$$

$$u_r = a_n(t)x_r + b_n(t)y_r + c_n(t) + d_n(t)x_r^2 + e_n(t)d_r^2 + h_n(t)x_r y_r$$

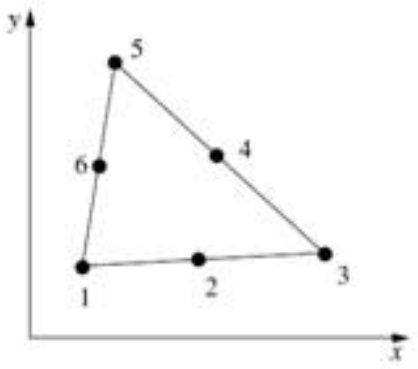


Fig. 5

We express $a_n(t), b_n(t), c_n(t), d_n(t), e_n(t), h_n(t)$ with respect not only u in vertices, but also in a node along the midside of each edge. See Fig. 5.

Finally using the so-called **collocation method** or **least square method** or **the method of moments** ([35]÷[40]) we can obtain a system of non-linear Ordinary Differential Equations that can be solved in a variety of methods (Runge – Kutta etc…).

We have examined the boundary value problem

$$\operatorname{div}\left(\left|\nabla u\right|^{p-2}\nabla u\right)=0$$

where u is a known function on the boundary of our domain using Variational Principle (Finite elements). The Problem is reduced to an Optimization problem that can be solved by Genetic Algorithms plus Nelder-Mead search. An early paper of the author with the title "Solving Differential Equations via Genetic Algorithms" was presented in [1] while the author use the same method to solve various problems in [2]÷[9].

With the Hybrid method of Genetic Algorithm + Nelder-Mead we have combined the advantages of both methods, that are **a)** the convergence to the global minimum (genetic algorithm) plus **b)** the high accuracy of the Nelder-Mead method.

Also, we have discussed briefly the solution of

$$\frac{\partial u}{\partial t}=\Delta_p\left(u^n\right)$$

$$\frac{\partial u}{\partial t}=\Delta_p u$$

and

$$\frac{\partial u}{\partial t}=\operatorname{div}\left(\left|\nabla u\right|^{p-2}\nabla u\right)+\lambda u^q,$$

using the so-called collocation method or least square method or the method of moments.

5. SOLUTION OF THE SCHRODINGER-MAXWELL EQUATIONS

Recently many authors have examined the following system of the non-linear Partial Differential Equations (PDEs) in R3

$$-\nabla^2 u + q\phi u = g(u) \tag{11}$$

$$-\nabla^2 \phi = qu^2 \tag{12}$$

with $g(.)$ being a known function.

The system of (11) and (12) is called:

Schrodinger-Maxwell equations. This system of Equations arises in many mathematical physics contexts, such as in quantum electrodynamics, in nonlinear optics, in nano-mechanics and in plasma physics.

The greatest part of the literature focuses on the study of the previous system for the very special nonlinearity

$$g(u) = -u + |u|^{p-1} u$$

and existence, nonexistence and multiplicity results are provided in many papers for this particular problem (see [18]÷[28]).

In [29], Azzollini, D'Avenia and Pomponio that a solution of a boundary problem of (11) and (2) yields the minimization of some functional.

In this paper solve the problem with the method of finite elements

In this paper we will solve the boundary value problem of

$$-\nabla^2 u + q\phi u = g(u) \tag{11}$$

$$-\nabla^2 \phi = qu^2 \tag{12}$$

where $g(.)$ is known using Variational Techniques (Finite elements). In Section 2, we produce the appropriate functional for minimization. After finding this functional, the solution of (1) and (2) with the nexesary boundary conditions can be easily reduced to an Optimization problem that can be solved by Genetic Algorithms with Nelder-Mead. An early paper of the author with the title "Solving Differential Equations via Genetic Algorithms" was presented in [1].

The author presented in 1996 the solution of ODE and PDE using Genetic Algorithms optimization, while the author use the same method to solve various problems in [2]÷[9].

The main Results are given in Section 2 and a numerical example illustrates the method in Section 3.

6. VARIATIONAL FORMULATION OF (11) AND (12) AND FINITE ELEMENTS APPROACH WITH GA

Consider that our functional is functional of u, ϕ, i.e. $I = I(u, \phi)$

Let the "point" of u_0, ϕ_0 that minimize the $I = I(u, \phi)$

Then for another point u, ϕ

we have

$u = u_0 + \varepsilon_1 u_1, \phi = \phi_0 + \varepsilon_2 \phi_1$

So, we must have the first order conditions

$\frac{\partial I(u,\phi)}{\partial \varepsilon_1} = 0$ and $\frac{\partial I(u,\phi)}{\partial \varepsilon_2} = 0$

Working first for (1) we can formulate:

$$I = \frac{1}{2}\iiint_V (\nabla u)^2 dv + \frac{1}{2}\iiint_V q\phi u^2 dv - \iiint_V G(u)dv + B(\phi) \tag{13}$$

with

$G(u) = \int g(u)du$ and $B(\phi)$ a function in ϕ

It is easy to verify by replacing $u = u_0 + \varepsilon_1 u_1$ that the condition $\frac{\partial I(u,\phi)}{\partial \varepsilon_1} = 0$ leads to

$$\iiint_V (\nabla u_0)(\nabla u_1)dv + \iiint_V q\phi u_0 u_1 dv - \iiint_V g(u_0)u_1 dv = 0$$

Now by applying the Green's first identity we have

$$\iint u_1(\nabla u \cdot \bar{n}) + \iiint_V (-\nabla^2 u_0)u_1 dv +$$

$$+\iiint_V q\phi u_0 u_1 dv - \iiint_V g(u_0)u_1 dv = 0$$

Considering appropriate u_1 we can have $\iint u_1(\nabla u \bar{n}) = 0$ which means

$$\iiint_V (-\nabla^2 u_0)u_1 dv + \iiint_V q\phi u_0 u_1 dv - \iiint_V g(u_0)u_1 dv = 0$$

or

$$\iiint_V (-\nabla^2 u_0 + q\phi u_0 - g(u_0))u_1 dv = 0$$

But u_1 is arbitrary which implies
$-\nabla^2 u_0 + q\phi u_0 - g(u_0) = 0$ i.e. we have (11)

Working analogously with (2) we could have

$$I = \frac{1}{2}\iiint_V (\nabla\phi)^2 dv - \iiint_V q\phi u^2 dv + C(u) \tag{14}$$

with $C(u)$ a function in u.

We must compromise (13) and (14). To this end we multiply the right hand member of (14) with the coefficient -1/2 and finally we propose the functional

$$I = \frac{1}{2}\iiint_V (\nabla u)^2 dv - \frac{1}{4}\iiint_V (\nabla\phi)^2 dv +$$
$$+\frac{1}{2}\iiint_V q\phi u^2 dv - \iiint_V G(u)dv$$

So, the solution of the problem of Schrodinger-Maxwell equations

$$-\nabla^2 u + q\phi u = g(u) \tag{11}$$
$$-\nabla^2 \phi = qu^2 \tag{12}$$

leads to $\min_{u,\phi} I$,
where

$$I = \frac{1}{2}\iiint_V (\nabla u)^2 dv - \frac{1}{4}\iiint_V (\nabla\phi)^2 dv +$$
$$+\frac{1}{2}\iiint_V q\phi u^2 dv - \iiint_V G(u)dv \tag{15}$$

We consider that u is written as $u = \sum_n \lambda_n f_n$ and $\phi = \sum_n \mu_n \tilde{f}_n$
or

$$u = \sum_n f_n \quad , \quad \phi = \sum_n \tilde{f}_n \tag{16}$$

where λ_n have been incorporated in f_n

So, we have the minimization problem of (15).

One can select a triangular mesh and appropriate functions f_n and $\tilde{f}_n$ that have non-zero value only in the n-th triangle ("finite elements"). So, in a triangular mesh, for example of $\square^2$, we can have $f_n = a_n x + b_n y + c_n$ and $\tilde{f}_n = \tilde{a}_n x + \tilde{b}_n y + \tilde{c}_n$ for the n-th triangle. Without loss of generality we consider the case $\square^2$ here u in (14).

To avoid to write continuity conditions on the common vertices of the triangles of the mesh, one can find that in the n-th triangle of the points s,q,r (see Figure 1)

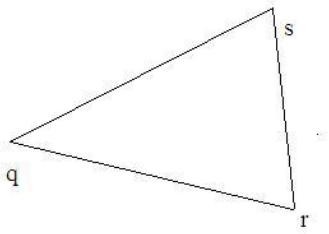


Fig. 6. A triangle in a 2-D mesh

$$u_s = a_n x_s + b_n y_s + c_n \tag{17.1}$$

$$u_q = a_n x_q + b_n y_q + c_n \tag{17.2}$$

$$u_r = a_n x_r + b_n y_r + c_n \tag{17.3}$$

$$\phi_s = \tilde{a}_n x_s + \tilde{b}_n y_s + \tilde{c}_n \tag{17.4}$$

$$\phi_q = \tilde{a}_n x_q + \tilde{b}_n y_q + \tilde{c}_n \tag{17.5}$$

$$\phi_r = \tilde{a}_n x_r + \tilde{b}_n y_r + \tilde{c}_n \tag{17.6}$$

These 6 equations can be solved with respect to a_n, b_n, c_n, $\tilde{a}_n, \tilde{b}_n, \tilde{c}_n$ and give

$$a_n = \frac{\begin{vmatrix} u_s & y_s & 1 \\ u_q & y_q & 1 \\ u_r & y_r & 1 \end{vmatrix}}{D} \tag{18.1}$$

$$b_n = \frac{\begin{vmatrix} x_s & u_s & 1 \\ x_q & u_q & 1 \\ x_r & u_r & 1 \end{vmatrix}}{D} \tag{18.2}$$

$$c_n = \frac{\begin{vmatrix} x_s & y_s & u_s \\ x_q & y_q & u_q \\ x_r & y_r & u_r \end{vmatrix}}{D} \tag{18.3}$$

$$\tilde{a}_n = \frac{\begin{vmatrix} \phi_s & y_s & 1 \\ \phi_q & y_q & 1 \\ \phi_r & y_r & 1 \end{vmatrix}}{D} \tag{18.4}$$

$$\tilde{b}_n = \frac{\begin{vmatrix} x_s & \phi_s & 1 \\ x_q & \phi_q & 1 \\ x_r & \phi_r & 1 \end{vmatrix}}{D} \tag{18.5}$$

$$\tilde{c}_n = \frac{\begin{vmatrix} x_s & y_s & \phi_s \\ x_q & y_q & \phi_q \\ x_r & y_r & \phi_r \end{vmatrix}}{D} \tag{18.6}$$

and

$$D = \begin{vmatrix} x_s & y_s & 1 \\ x_q & y_q & 1 \\ x_r & y_r & 1 \end{vmatrix}$$ (which is by the way 2*E where E is the algebraic area of the triangle)

Hence, from the minimization problem

$$\min_{u,\phi} I$$

where

$$I = \frac{1}{2}\iiint_V (\nabla u)^2 dv - \frac{1}{4}\iiint_V (\nabla \phi)^2 dv +$$

$$+\frac{1}{2}\iiint_V q\phi u^2 dv - \iiint_V G(u) dv$$

CONCLUSION

We find the equivalent minimization problem

$$\min \int W(a_n, b_n, c_n, \tilde{a}_n, \tilde{b}_n, \tilde{c}_n) dv \tag{19}$$

where $W(a_n, b_n, c_n, \tilde{a}_n, \tilde{b}_n, \tilde{c}_n)$ is the function that we find after replacing $f_n = a_n x + b_n y + c_n$ and $\tilde{f}_n = \tilde{a}_n x + \tilde{b}_n y + \tilde{c}_n$ in the above functional and a_n, b_n, c_n, $\tilde{a}_n, \tilde{b}_n, \tilde{c}_n$ will be replaces using (18.1), (18.2), (18.3) (18.4), (18.5), (18.6) for each triangle of the mesh.

Equation (19) can be solved now by a variety of techniques. The author uses Genetic Algorithms with Nelder-Meade for Non-linear Problems as in [2], [3], [4], [5], [6], [7], [8].

The same optimization scheme: Genetic Algorithms with Nelder-Meade is also applied for (19).

REFERENCES:

[1] Nikos E. Mastorakis, "Solving Differential Equations via Genetic Algorithms", Proceedings of the Circuits, Systems and Computers '96, (CSC'96), Piraeus, Greece, July 15-17, 1996, 3rd Volume: Appendix, pp.733-737

[2] Nikos E. Mastorakis, "On the solution of ill-conditioned systems of Linear and Non-Linear Equations via Genetic Algorithms (GAs) and Nelder-Mead Simplex search", 6th WSEAS International Conference on EVOLUTIONARY COMPUTING (EC 2005), Lisbon, Portugal, June 16-18, 2005.

[3] Nikos E. Mastorakis, "Genetic Algorithms and Nelder-Mead Method for the Solution of Boundary Value Problems with the Collocation Method", WSEAS Transactions on Information Science and Applications, Issue 11, Volume 2, 2005, pp. 2016-2020.

[4] Nikos E. Mastorakis, "On the Solution of Ill-Conditioned Systems of Linear and Non-Linear Equations via Genetic Algorithms (GAs) and Nelder-Mead Simplex Search", WSEAS Transactions on Information Science and Applications, Issue 5, Volume 2, 2005, pp. 460-466.

[5] Nikos Mastorakis, "Genetic Algorithms and Nelder-Mead Method for the Solution of Boundary Value Problems with the Collocation Method", 5th WSEAS International Conference on SIMULATION, MODELING AND OPTIMIZATION (SMO '05), Corfu Island, Greece, August 17-19, 2005.

[6] Nikos E. Mastorakis, "Solving Non-linear Equations via Genetic Algorithm", WSEAS Transactions on Information Science and Applications, Issue 5, Volume 2, 2005, pp. 455-459.

[7] Nikos E. Mastorakis, "The Singular Value Decomposition (SVD) in Tensors (Multidimensional Arrays) as an Optimization Problem. Solution via Genetic Algorithms and method of Nelder-Mead", 6th WSEAS Int. Conf. on SYSTEMS THEORY AND SCIENTIFIC COM-

PUTATION (ISTASC'06), Elounda, Agios Nikolaos, Crete Island, Greece, August 21-23, 2006.

[8] Nikos E. Mastorakis, "Unstable Ordinary Differential Equations: Solution via Genetic Algorithms and the method of Nelder-Mead", The 6th WSEAS International Conference on SYSTEMS THEORY AND SCIENTIFIC COMPUTATION, Elounda, Agios Nikolaos, Crete Island, Greece, August 18-20, 2006.

[9] Nikos E. Mastorakis, "Unstable Ordinary Differential Equations: Solution via Genetic Algorithms and the Method of Nelder-Mead", WSEAS TRANSACTIONS on MATHEMATICS, Issue 12, Volume 5, December 2006, pp. 1276-1281.

[10] Nikos E. Mastorakis, "An Extended Crank-Nicholson Method and its Applications in the Solution of Partial Differential Equations: 1-D and 3-D Conduction Equations", WSEAS TRANSACTIONS on MATHEMATICS, Issue 1, Volume 6, January 2007, pp. 215-224.

[11] Saeed-Reza Sabbagh-Yazdi, Behzad Saeedifard, Nikos E. Mastorakis, "Accurate and Efficient Numerical Solution for Trans-Critical Steady Flow in a Channel with Variable Geometry", WSEAS TRANSACTIONS on APPLIED and THEORETICAL MECHANICS, Issue 1, Volume 2, January 2007, pp. 1-10.

[12] Saeed-Reza Sabbagh-Yazdi, Mohammad Zounemat-Kermani, Nikos E. Mastorakis, "Velocity Profile over Spillway by Finite Volume Solution of Slopping Depth Averaged Flow", WSEAS TRANSACTIONS on APPLIED and THEORETICAL MECHANICS, Issue 3, Volume 2, March 2007, pp. 85.

[13] Iurie Caraus and Nikos E. Mastorakis, "Convergence of the Collocation Methods for Singular Integrodifferential Equations in Lebesgue Spaces", WSEAS TRANSACTIONS on MATHEMATICS, Issue 11, Volume 6, November 2007, pp. 859-864.

[14] Iurie Caraus, Nikos E. Mastorakis, "The Stability of Collocation Methods for Approximate Solution of Singular Integro- Differential Equations", WSEAS TRANSACTIONS on MATHEMATICS, Issue 4, Volume 7, April 2008, pp. 121-129.

[15] Xu Gen Qi, Nikos E. Mastorakis, "Spectral distribution of a star-shaped coupled network", WSEAS TRANSACTIONS on APPLIED and THEORETICAL MECHANICS , Issue 4, Volume 3, April 2008.

[16] Iurie Caraus, Nikos E. Mastorakis, "Direct Methods for Numerical Solution of Singular Integro-Differentiale Quations in Classical (case $\gamma \neq 0$)", 10th WSEAS International Conference on MATHEMATICAL and COMPUTATIONAL METHODS in SCIENCE and ENGINEERING (MACMESE'08), Bucharest, Romania, November 7-9, 2008.

[17] Gabriella Bognar, Numerical and Numerical and Analytic Investigation of Some Nonlinear Problems in Fluid Mechanics, COMPUTER and SIMULATION in MODERN SCIENCE, Vol.II, WSEAS Press, pp.172-179, 2008

[18] Astrita G., Marrucci G., Principles of Non-Newtonian Fluid Mechanics, McGraw-Hill, New York, NY, USA, 1974.

[19] Volquer R.E., Nonlinear flow in porous media by finite elements, ASCE Proc., J. Hydraulics Division Proc. Am. Soc. Civil Eng., 95 (1969), 2093-2114

[20] Ahmed N., Sunada D.K., Nonlinear flow in porous media, J. Hydraulics Division Proc. Am. Soc. Civil Eng., 95 (1969), 1847-1857.

[21] Peebles P.J.E., Star distribution near a collapsed object, Astrophysical Journal, Vol. 178, (1972),. 371-376.

[22] Carelman T., Probl`emes math´ematiques dans la th´eorie cin´etique de gas, Almquist-Wiksells, Uppsala, 1957.

[23] Bartnik R., McKinnon J., Particle-like solutions of the Einstein-Yang-Mills equations. Phys. Rev. Lett. 61 (1988), 141-144

[24] Otani M., A remark on certain nonlinear elliptic equations. *Proc. Fac. Sci. Tokai Univ.* 19 (1984), 23--28.

[25] Pelissier, M.-C., Reynaud, L., Étude d'un modèle mathématique d'ecoulement de glacier, C. R. Acad. Sci., Paris, Sér. A 279 (1974), 531-534. (French)

[26] Schoenauer M., A monodimensional model for fracturing, In A. Fasano and M. Primicerio (editors): Free Boundary Problems, Theory Applications, Pitman Research Notes in Mathematics 79, Vol. II., London, 701-711 (1983).

[27] John C. Strikwerda, Finite Difference Schemes and Partial Differential Equations, SIAM, 2004

[28] Hans Petter Langtangen Computational Partial Differential Equations: Numerical Methods and Diffpack Programming, Springer, 2003

[29] W. L. Wood, Introduction to Numerical Methods for Water Resources, Oxford University Press, 1993

[30] Daniel R. Lynch, Numerical Partial Differential Equations for Environmental Scientists and Engineers: A First Practical Course, Springer, 2005

[31] Nikos E. Mastorakis, "Numerical Schemes for Non-linear Problems in Fluid Mechanics", Proceedings of the 4th IASME/WSEAS International Conference on CONTINUUM MECHANICS, Cambridge, UK, February 24-26, 2009, pp.56-61

[32] Evans, Lawrence C. , A New Proof of Local C1,α Regularity for Solutions of Certain Degenerate Elliptic P.D.E.", Journal of Differential Equations 45: 356-373, 1982

[33] Lewis, John L. (1977). Capacitary functions in convex rings, Archive for Rational Mechanics and Analysis 66: 201–224, 1977

[34] Lagarias, J.C., J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions," SIAM Journal of Optimization, Vol. 9 Number 1, pp. 112-147, 1998

[35] J. A. Nelder and R. Mead, "A simplex method for function minimization", Computer Journal, 7, 308-313, 1965

[36] F. H. Walters, L. R. Parker, S. L. Morgan, and S. N. Deming, Sequential Simplex Optimization, CRC Press, Boca Raton, FL, 1991

[37] A. Ern, J.L. Guermond, Theory and practice of finite elements, Springer, 2004, ISBN 0-3872-0574-8

[38] S. Brenner, R. L. Scott, The Mathematical Theory of Finite Element Methods, 2nd edition, Springer, 2005, ISBN 0-3879-5451-1

[39] P. G. Ciarlet, The Finite Element Method for Elliptic Problems, North-Holland, 1978, ISBN 0-4448-5028-7

[40] Y. Saad, Iterative Methods for Sparse Linear Systems, 2nd edition, SIAM, 2003, ISBN 0-8987-1534-2

[41] J.J.Conor and C.A.Brebbia, Finite Element Techniques for Fluid Flow, Butterworth, London, 197685

FRUSTRATION AND CHIRALITY IN ANISOTROPIC LAYERED ANTIFERROMAGNETS

Dimo I. Uzunov

Collective Phenomena Laboratory, Institute of Solid State Physics, Bulgarian Academy of Sciences,
BG – 1784 Sofia, Bulgaria, tel. +359 2 9795821, email: d.i.uzunov@gmail.com

Abstract: *We extend the field theory of antiferromagnets to layered structures on BCT crystal lattices with nearest-neibour (nn) and next-nearest-neighbour (nnn) ferro- and/or antiferromagnetic interactions. For this aim the field theoretical counterpart of a lattice Heisenberg model is derived by the means of standard theoretical methods: Hubbard-Stratonovich transformation and a generalized mean-field approach. It is shown that the inter-layer interactions are a pure thermal fluctuation effect whereas the ground state is characterized by a perfect in-layer antiferromagnetic order and a lack of inter-layer coupling. This is a demonstration of 2D-3D dimensional crossover which is supposed to occur in real antiferromagnets, for example, in the spin-dimer antiferromagnet* $BaCuSi_2O_6$.

Keywords: antiferromagnetism, magnetic order, exchange interaction, field models, magnetism, thermal fluctuations.

1. INTRODUCTORY NOTES

Several recent studies have focused the interest on crystals with complex antiferromagnetic order [1, 2]. The latter is tuned by both temperature T and external magnetic field $H = |\boldsymbol{H}|$. For example, the spin-1/2 dimer compound $BaCuSi_2O_6$ has recently attracted much interest [1]. In this compound the spin dimmers align on a body-centered tetragonal (BCT) lattice. Due to the lattice geometry, this system can be treated as a layered structure consisting of two dimensional (2D) x-y distributed regularly along the z-axis. Loops that include inter-layer hoppings are frustrated and this leads to a cancellation among inter-layer interactions. An interesting behavior related to this inter-layer frustration has recently been reported. This type of systems can be investigated with the help of spin-1/2 Heisenberg models of XXZ type [3].

Here we discuss some basic problems of layered antiferromagnets described by a XXZ Heisenberg model formulated in Ref. [2]. Our task is to derive a field theoretical counterpart of the lattice model introduced in Ref. [2]. Our approach could be used as an alternative of the Monte Carlo simulations [2] in investigations of the possible ordered phase which appear at low temperature T as a result nearest-neighbor (*nn*) and next-nearest-neighbor (*nnn*) exchange interactions. We shall use methods from the modern theory of phase transitions [3], including the particular scheme of derivation of field models of antiferromagnets [4, 5]. In contrast with the standard theoretical approach outlined in Refs. [4, 5] we propose a more detailed derivation of the relations between the microscopic parameters (the exchange constants) and the macroscopic Landau parameters (bare vertex parameters) of the field theory.

Here we shall outline the theoretical approach and report the final result, namely, the field theory describing a wide class of layered antiferromagnets with BCT crystal structure.

2. GENERAL THEORETICAL FRAMEWORK

Consider (pseudo)spin-1/2 anisotropic antiferromagnets with body-centered tetragonal lattice (BCT) of volume $V=L^2L_z$, number of vertices $N= N_0^2N_z \sim 10^{21}-10^{23}$ and lattice constants (a, a, c); N_0^2 is the number of sites in the xy square lattice of area L_0^2, and N_z is the respective number along the z-axis. For concreteness and simplicity, the model of *nn* exchange interactions will be discussed but let us note that some generalizations of the present discussion are straightforward. Our main aim is to deduce some properties of the phase transitions in this type of systems and compare the results with recent Monte Carlo calculations [2].

Starting from a microscopic formulation of the system we derive an effective field Hamiltonian (alias generalized free energy) which describes the thermodynamics, including relevant fluctuation phenomena in a close vicinity of critical points of phase transitions. Having this result, we investigate the role of the external magnetic field on the symmetry or ordering and the shape of the phase diagrams, as well as the interrelationship between the type of the critical behavior and the broken continuous and discrete symmetries of the ordered phases.

The studies of antiferromagnets are performed by dividing the original lattice in two sub-lattices with magnetizations of opposite directions. Within this scheme the actual order parameter is one of the sub-lattice magnetizations, or, which is the same the half of the difference between the two opposite magnetization vectors. Alternatively, for a number of important problems, one may use the so-called alternating magnetization as the actual order parameter. The latter will be used in our investigation.

The comprehensive spin model is the Heisenberg Hamiltonian

$$H = -\frac{1}{2}\sum_{\vec{R},\vec{R}'} J(\vec{R},\vec{R}')\vec{S}(\vec{R}).\vec{S}(\vec{R}') - \sum_{\vec{R}} \vec{h}.\vec{S}(\vec{R}), \qquad (1)$$

where the three-dimensional (3D) spatial discrete vector $\vec{R} = (\vec{r}, z)$ runs the vertices of the lattice [the 2D vector $\vec{r} = (x, y)$ is introduced for a further convenience]. The first term in Eq. (1) represents the exchange interaction of classical spins $\vec{S}(\vec{R})$ located at vertices $\vec{R}$ and $\vec{R}'$. The exchange matrix $J(\vec{R},\vec{R}') \equiv J(|\vec{R}-\vec{R}'|)$ depends only on the distance between the spins. The second term in Eq. (1) represents the Zeeman interaction of the spins with a homogeneous external magnetic field $\vec{H} = \vec{h}/g\mu_B$; g is the gyromagnetic ratio (for electrons, $g \approx 2$), and μ_B is the Bohr magneton. The lat-

ter is assumed to be uniform, i.e., "$\vec{R}$- independent", as it is often in precise experiments (the case of non-uniform magnetic field $\vec{H}(\vec{R})$, or, in certain cases "random" magnetic field is quite complex and leads to another huge area of research in Physics [3].

Now our task is to take into account a spatial anisotropy of type XXZ, i.e., anisotropy along the *z*-direction. It is convenient to consider the lattice as a chain of monoatomic layer parallel to the (*x*, *y*) plane and labeled by the *z*-component of the 3D spatial vector $\vec{R} = (\vec{r}, z)$, whereas the 2D vector $\vec{r}$ indicates the sites in a given layer. We assume that the strength of the exchange interaction between spins located in the same layer (in-layer interaction) is different from the exchange interaction strength between spins at the same distance but located in different layers (inter-layer interaction). For *nn* interactions, the exchange matrix $J(|\vec{R}-\vec{R}'|)$ of such an antiferromagnetic system can be written in the form

$$J(\vec{R}-\vec{R}') = J_{\|}\delta(\vec{r}', \vec{r}+\hat{\delta})\delta(z, z') + J_{\perp}\delta(\vec{r}', \vec{r}+\Delta\vec{r})\delta(z', z+\Delta z), \qquad (2)$$

where $\delta(i, j) \equiv \delta_{ij}$ denotes the Kronecker symbol, $|J_{\|}| > |J_{\perp}|$, the in-layer exchange interaction parameter $J_{II} < 0$ corresponds to antiferromagnetic order, and the inter-layer exchange parameter $J_{\|}$ may take any sign. In Eq. (2), the vector $\hat{\delta} = (a\hat{x} + a\hat{y})$ with $a_1 \equiv a_x = a$ and $a_2 = a_y = a$ describes the in-layer *nn* to the site $(\vec{r}, z)$, whereas $\Delta z = \pm c$ and the inter-layer displacement vector $\Delta r = (b_1\hat{x} + b_2\hat{y})$ points out the inter-layer *nn*. The displacement parameters $b_1 = b_x = b$ and $b_2 = b_y = b$ can be expressed by the lattice constant *a*. For the BCT of $BaCuSi_2O_6$, $b = a/2$; see also the discussion in Ref. [2].

Now the Hamiltonian (1) can be written in the form

$$H = -\frac{J_{\|}}{2}\sum_{z,\vec{r},\vec{r}'}\vec{S}(\vec{r}, z).\vec{S}(\vec{r}', z) - \frac{J_{\perp}}{2}\sum_{z,\vec{r},\Delta\vec{r},z}\vec{S}(\vec{r}, z).\vec{S}(\vec{r}+\Delta\vec{r}, z+\Delta z) - \sum_{\vec{R}}\vec{h}.S(\vec{R}). \qquad (3)$$

3. OUTLINE OF THE DERIVATION OF THE EFFECTIVE FIELD THEORY

Our aim is to derive an effective field theory corresponding to the lattice Hamiltonian (3). This is made by some reliable coarse-graining procedure which makes possible the derivation of the quasi-macroscopic (field) theory from the microscopic spin Hamiltonian. Generally such coarse graining procedures in statistical mechanics are few but two or three of them are very comprehensive and yield a comprehensive description of the thermodynamic properties of both simple and complex condensed matter (many-body) systems [3, 4].

For antiferrimagnets such a theory can be firstly derived on the basis of a lattice field related to the microscopic spin field $\vec{S}(\vec{r},z)$ by introducing two sub-lattices, or, alternatively, by introducing the so-called alternating magnetization $\vec{S}_a(\vec{R}) = \exp(\vec{q}_0.\vec{r})\vec{S}(\vec{R})$ which describes the antiferromagnetic order [5, 6]. The alternating magnetization field $\vec{S}_a(\vec{R})$ is defined with the help of the reciprocal lattice vector $\vec{q}_0 = (\vec{k}_0, q_{0z})$ corresponding to the upper border of the first Brillouin zone: $k \equiv \mid \vec{k}_0 \mid = \pi/a$, and $q_{0z} = \pi/c$.

Now one may use the approach in Refs. [5, 6] and derive the field theory. An improvement of the method available in Refs.[5, 6] can be made with the help of the Hubbard-Stratonovich transformation (see, e.g., Ref. [3]). This transformation and related shift and rotation transformations allows for a correct derivation of the relationship between the microscopic parameters (the exchange interaction parameters) and the quasi-macroscopic parameters ("vertex") parameters of the field theory in tree approximation [3]. Besides, one has to use the long wavelength approximation [3, 4], too. Note, that the coarse graining procedure outlined so far is equivalent to "a generalized self-consistency coarse graining." [6]. Following the latter we have obtained the following result [7] for the field Hamiltonian

$$
\begin{aligned}
H = \int_V d^3R\{ & -\nu_0 \vec{h}.\vec{\sigma}(\vec{R}) \\
& + \frac{r_f}{2}\vec{\sigma}^2 + \frac{r_a}{2}\vec{\varphi}^2 + \frac{c_f}{2}[\nabla_r\vec{\sigma}(\vec{R})]^2 + \frac{c_f'}{2}[\nabla_z\vec{\sigma}(\vec{R})]^2 + \frac{c_a}{2}[\nabla_r\vec{\varphi}(\vec{R})]^2 + \frac{c_a'}{2}[\nabla_x\vec{\varphi}(\vec{R})].[\nabla_y\vec{\varphi}(\vec{R})]\} \\
& + \frac{\beta^3}{6}[\mu^4\vec{\sigma}^4 + \nu^4\vec{\varphi}^4 + 4\mu^2\nu^2(\vec{\sigma}.\vec{\varphi})^2 + 2\mu^2\nu^2\vec{\sigma}^2\vec{\varphi}^2]\}
\end{aligned}
\tag{4}
$$

where $\beta = 1/k_BT$, the parameters of the field theory are given by the microscopic parameters $J_\parallel$ and $J_\perp$ by the algebraic formulae

$$\nu_0 = \frac{J_\parallel + 4J_\perp}{J_\parallel + 2J_\perp}, \quad \mu = 2(J_\parallel + 2J_\perp),\ \nu = 2J, \tag{5a}$$

$$r_a = 2\mid J_\parallel \mid -8\beta \mid J_\parallel \mid^2 \mid, \quad r_f = -(2\mid J_\parallel \mid -8\mid J_\perp \mid) - 8\beta(\mid J_\parallel \mid -2J_\perp)^2, \tag{5b}$$

$$c_a = (4\beta J_\parallel^{\,2} + \frac{J_\parallel}{2})a^2, \quad c_a' = -2J_\perp a^2, \tag{5c}$$

$$c_f = \frac{1}{2}[1 + 8\beta(\mid J_\parallel \mid -J_\perp)](\mid J_\parallel \mid -2J_\perp)a^2 > 0,$$

$$c_f' = -J_\perp[1 + 4\beta(\mid J_\parallel \mid -2J_\perp)]c^2 > 0, \tag{5d}$$

and the vector fields $\vec{\sigma}(\vec{R})$ and $\vec{\varphi}(\vec{R})$ describe the ferromagnetic and antiferromagnetic order, respectively.

These orderings are of in-layer type, i.e. they describe the spontaneous breaking of the symmetry in the layers. The only connection between the layers is given by the gradient term with coefficient c_f'. This is a proof of the fluctuation nature of the interlayer coupling. When thermal fluctuations are not present, the layers are absolutely independent. This situation is possible only at the absolute zero ($T = 0$).

4. FINAL REMARKS

The further analysis of the Hamiltonian (4) follows standard rules [3, 5, 6]. One may show that this theory describes correctly the antiferromagnetic order in layered antiferromagnets of the type mentioned above. For nonzero external magnetic field a spin-flop phase may appear for certain values of the material parameters. The phase transitions in these systems are tuned by the temperature T and the external magnetic field $\boldsymbol{H}$. Without any loss of generality one may perform the thermodynamic analysis of the Hamiltonian (4) by setting a definite direction of the external magnetic field, e.g., $\boldsymbol{H} = (0, 0, H)$.

ACKNOWLEDGEMENTS:

A financial support through grant Ph. 1507 (NFSR-Sofia)) is acknowledged.

REFERENCES

[1] S. E. Sebastian, N. Harrison, C. D. Batista, L. Balicas, M. Jaime, P. A. Sharma, N. Kawashima, and I. R. Fisher, Nature (London) 441, 617 (2006).

[2] Y. Kamyia, N. Kawashima, and S. D. Batista, Preprint Tokyo University (April, 2009).

[3] D. I. Uzunov, "Theory of critical phenomena" (World Scientific, 1993).

[4] D. I. Uzunov, in: "Lectures on Cooperative Phenomena in Condensed Matter," ed. by D. I. Uzunov (Heron Press, Sofia, 1996); pp. 46-110. *Title*: "The Self-consistent Description of Phase Transitions Revisited."

[5] D. R. Nelson, M. E. Fisher, Phys. Rev. Lett. 32, 1350 (1974).

[6] J. M. Kosterlitz, D. R. Nelson, M. E. Fisher, Phys. Rev. B 13, 412 (1976).

[7] D. I. Uzunov, Tokyo University Preprint (July 2009), and seminar report abstract (3th August 2009) (unpublished).

COMPARISON OF EXPERIMENTAL AND THEORETICAL PHASE DIAGRAM OF FERROMAGNETIC UNCONVENTIONAL SUPERCONDUCTORS

Diana V. Shopova, Dimo I. Uzunov

Collective Phenomena Laboratory, Institute of Solid State Physics, Bulgarian Academy of Sciences,
BG-1784 Sofia, Bulgaria, +359 2 9795821, +359 2 9795820, dishopova@gmail.com, d.i.uzunov@gmail.com

Abstract: *The theoretical description of the phase diagrams of ferromagnetic unconventional superconductors with spin-triplet Cooper pairing are explained and described within a general phenomenological theory of Ginzburg-Landau type. The theoretical predictions are in an excellent agreement with the experimental data for certain inter-metallic compounds which a itinerant ferromagnets and exhibit spin-triplet superconductivity at relatively high pressures (~1 GPa) and low temperatures (~1 K). New quantum phase transitions are predicted and proposed for experimental search in real materials. This report is based on recently published results by the present authors.*

Keywords: spin-triplet superconductivity, ferromagnetism, coexistence of superconductivity and ferromagnetism, quantum phase transitions.

1. INTRODUCTION

The spin-triplet pairing allows parallel spin orientation of the fermion Cooper pairs in super-fluid ^{3}He and unconventional superconductors [1]. For this reason the resulting unconventional superconductivity is robust with respect to effects of external magnetic field and spontaneous ferromagnetic order, so it may coexist with the latter. This general argument implies that there could be metallic compounds and alloys, for which the coexistence of spin-triplet superconductivity and ferromagnetism may be observed.

Particularly, both superconductivity and itinerant ferromagnetic orders can be created by the same band electrons in the metal, which means that spin-1 electron Cooper pairs participate in the formation of the itinerant ferromagnetic order. Moreover, under certain conditions the superconductivity is enhanced rather than depressed by the uniform ferromagnetic order that can generate it, even in cases when the superconductivity does not appear in a pure form as a net result of indirect electron-electron coupling.

The coexistence of superconductivity and ferromagnetism as a result of collective behavior of *f*-band electrons has been found experimentally for some Uranium-based inter-metallic compounds as, UGe_2 [2-5], URhGe [6-8], UCoGe [9, 10], and UIr [11, 12]. At low temperature ($T \sim 1$ K) all these compounds exhibit thermodynamically stable phase of coexistence of spin-triplet superconductivity and itinerant (*f*-band) electron ferromagnetism (in short, FS phase). In UGe_2 and UIr the FS phase appears at high pressure ($P \sim 1$GPa), whereas in URhGe and UCoGe, the coexistence phase

persists up to ambient pressure (10^5 Pa = 1 bar). Experiments, carried out in $ZrZn_2$ [13], also indicated the appearance of FS phase at $T < 1$ K in a wide range of pressures ($0 < P \sim 21$ kbar). In Zr-based compounds the ferromagnetism and the *p*-wave superconductivity occur as a result of the collective behavior of the *d*-band electrons. Later experimental results [14, 15] had imposed the conclusion that bulk superconductivity is lacking in $ZrZn_2$, but the occurrence of a surface FS phase at surfaces with higher Zr content than that in $ZrZn_2$ has been reliably demonstrated. Thus the problem for the coexistence of bulk superconductivity with ferromagnetism in $ZrZn_2$ is still unresolved. This raises the question whether the FS phase in $ZrZn_2$ should be studied by surface thermodynamics methods, or it can be investigated by considering that bulk and surface thermodynamic phenomena can be treated on the same footing. Taking in account the mentioned experimental results for $ZrZn_2$ and their interpretation by the experimentalists [13-15] we assume that the unified thermodynamic approach can be applied, although some specific properties of the FS phase in $ZrZn_2$, if its surface nature is confirmed, will certainly need special study by surface thermodynamics.

Here we will investigate the itinerant ferromagnetism and superconductivity of U- and Zr-based inter-metallic compounds within the same general thermodynamic approach. Arguments supporting our point of view are given in several preceding investigations. We should mention that the spin-triplet superconductivity occurs not only in bulk materials, but also in quasi-two-dimensional (2D) systems - thin films and surfaces and quasi-1D wires (see, e.g., [16]). In $ZrZn_2$ both ferromagnetic and superconducting orders vanish at the same critical pressure P_c, a fact implying that the respective order parameter fields strongly depend on each other and should be studied on the same thermodynamic basis [17].

The general thermodynamic treatment does not necessarily specify the system spatial dimensionality D:$D = 3$ describes the bulk properties, and $D = 2$ – very thin films and mono-atomic layers. Within Landau theory of phase transitions (see, e.g., Ref. [18]), the system dimensionality can be distinguished by the values of the Landau parameters. Here we specify the values of these parameters from the experimental data for spin-triplet ferromagnetic superconductors. When the Landau parameters are obtained from microscopic theories, their values depend on the dimension D, only if the respective theory takes into account relevant fluctuation modes of order parameter fields, including long-scale fluctuation modes. We are not aware of well developed theories of this type which may figure out the complex behavior of the mentioned systems. Even in simple theories of band electron magnetism, the Landau parameters are very complex functions of the density of states at the Fermi level and related microscopic parameters. Such complexity does not allow direct comparison between the results from microscopic theory and the experimental data. To make a progress in this situation we assume that the material parameters of our theory are loosely defined and may have values, corresponding to various approximate microscopic theories, as mean-field approximation, spin-fluctuation theory, etc.

For all compounds, cited above, the FS phase occurs only in the ferromagnetic phase domain of the *T-P* diagram. Particularly at equilibrium, and for given *P*, the temperature $T_F(P)$ of the normal-to-ferromagnetic phase (or N-FM) transition is never lower than the temperature $T_{FS}(P)$ of the ferromagnetic-to-FS phase transition (FM-FS transition). This confirms the point of view that the superconductivity in these compounds is triggered by the spontaneous magnetization $\boldsymbol{M}$, in analogy with the well-known triggering of the superfluid phase A_1 in ^{3}He at mK temperatures by the external magnetic field $\boldsymbol{H}$. Such "helium analogy" has been used in some theoretical studies (see, e.g., Refs, [19, 20]), where Ginzburg - Landau (GL) free energy terms, describing the FS phase were derived by symmetry group arguments.

The non-unitary state, with a non-zero value of the Cooper pair magnetic moment, known from the theory of unconventional superconductors and super-fluidity in ^{3}He, has been suggested firstly in Ref. [19], and later confirmed in other studies [8, 20]; recently, the same topic was comprehensively discussed in Ref. [21]. For the spin-triplet ferromagnetic superconductors the trigger mechanism was recently examined in detail [22, 23]. The system main properties are specified by the term in the GL expansion of type $\sim|\boldsymbol{M}||\psi|^2$, which represents the interaction of the magnetization $\boldsymbol{M} = \{M_j;\ j=1, 2, 3\}$ with the complex superconducting vector field $\boldsymbol{\psi} = \{\psi_j\}$. Particularly, this term is responsible for the appearance of $\boldsymbol{\psi}$ (i.e., of superconductivity) for certain *T* and *P* values. A similar trigger mechanism is familiar in the context of improper ferroelectrics [24].

A crucial feature of these systems is the nonzero magnetic moment of the spin-triplet Cooper pairs. As mentioned above, the microscopic theory of magnetism and superconductivity in non-Fermi liquids of strongly interacting heavy electrons (*f*- and *d*- band electrons) is either too complex or insufficiently developed to describe the complicated behavior in itinerant ferromagnetic compounds. Several authors (see [25]) have explored the phenomenological description by a self-consistent mean field theory, and we shall use these results. Besides, we will essentially use our previous theoretical studies results, in particular, the analysis in Refs. [22, 23]. In this paper, we report the main theoretical results achieved in our recent investigations [26, 27] based on the general thermodynamic treatment [22, 23].

We propose a simple, yet comprehensive, modeling of *P* dependence of the free energy parameters, resulting in a very good compliance of our theoretical predictions for the shape the *T-P* phase diagrams with the experimental data. The theoretical analysis is done by the standard methods of phase transition theory. The treatment of fluctuation effects and quantum correlations [28-30] is not included in this report. But the parameters of the generalized GL free energy may be considered either in mean-field approximation as here, or as phenomenological re-normalized parameters which are affected by additional physical phenomena, as for example, the spin fluctuations. The magnetic fluctuations may affect the phase transitions in type-I superconductors [18, 31] and here we shall not dwell on this problem. The reason is that the unconventional ferromagnetic superconductors known so far are of type II.

We demonstrate with the help of present theory that we can outline different possible topologies for the T-P phase diagram, depending on the values of Landau parameters, derived from the existing experimental data. We show that for unconventional (spin-triplet) ferromagnetic superconductors (FSs) there exist two distinct types of behavior, which we denote as Zr-type (or, alternatively, type I) and U-type (or, type II). This classification of the FS, first mentioned in Ref. [26], is based on the reliable interrelationship between a quantitative criterion derived by us and the thermodynamic properties of the spin-triplet FSs. Our approach can be also applied to URhGe, UCoGe, and UIr. Our results shed light on the problems connected with the order of the quantum phase transitions at ultra-low and zero temperatures. They also raise the question for further experimental investigations of the detailed structure of the phase diagrams in the high-*P*/low-*T* region.

2. FREE ENERGY AND PHASE DIAGRAMS

Following Refs. [22, 23] the free energy per unit volume, $F/V = f(\boldsymbol{\psi}, \boldsymbol{M})$, can be written in the form

$$\mathrm{f} = a_s |\boldsymbol{\psi}|^2 + \frac{b_s}{2} |\boldsymbol{\psi}|^4 + \frac{u_s}{2} |\boldsymbol{\psi}^2|^2 + \frac{v_s}{2} \sum_{j=1}^{3} |\psi_j|^4 + a_f \boldsymbol{M}^2 + \frac{b_f}{2} \boldsymbol{M}^4 + i\gamma_0 \boldsymbol{M}.(\boldsymbol{\psi} \times \boldsymbol{\psi}^*) + \delta\, \boldsymbol{M}^2 |\boldsymbol{\psi}|^2. \quad (1)$$

The material parameters satisfy, as in [22, 23], $b_s > 0$, $b_f > 0$, $a_s = \alpha_s (T - T_s)$, and $a_f = \alpha_f [T^n - T_f^n(P)]$, where $n = 1$ gives the standard form of a_f, and $n = 2$ applies for the spin-fluctuation theory (SFT). The terms proportional to u_s and v_s describe, respectively, the anisotropy of the spin-triplet electron Cooper pairs and the crystal anisotropy. Next, $\gamma_0 \sim J$ (with $J > 0$ the ferromagnetic exchange constant) and $\delta > 0$ are parameters of the $\boldsymbol{\psi}$-$\boldsymbol{M}$ interaction terms.

Previous mean-field studies have shown that the anisotropies represented by the u_s and v_s terms in Eq. (1) slightly perturb the size and shape of the stability domains of the phases, while similar effects can be achieved by varying the b_s factor in the $b_s|\boldsymbol{\psi}|^4$ term. For these reasons, in the present analysis we ignore the anisotropy terms, setting $u_s = v_s = 0$, and consider $b_s = b > 0$ as an effective parameter. Then, without loss of generality, we are free to choose the magnetization vector to have the form $\boldsymbol{M} = (0, 0, M)$.

A convenient dimensionless free energy can now be defined by $\tilde{f} = f/(b_f M_0^4)$, where $M_0 = [\alpha_f T_{f0}{}^n / b_f]^{1/2} > 0$ is the value of M corresponding to the pure magnetic subsystem ($\boldsymbol{\psi} \equiv 0$) at $T = P = 0$, and $T_{f0} = T_f(0)$. On scaling the order parameters as $m = M/M_0$ and $\boldsymbol{\varphi} = \boldsymbol{\psi}/[(b_f/b)^{1/4} M_0]$ we obtain

$$\tilde{f} = r\phi^2 + \frac{\phi^4}{2} + tm^2 + \frac{m^4}{2} + 2\gamma m \phi_1 \phi_2 \sin\vartheta + \gamma_1 m^2 \phi^2, \quad (2)$$

where $\phi_j = |\varphi_j|$, $\varphi = |\boldsymbol{\varphi}|$, and ϑ is the phase angle between the complex φ_2 and φ_1. The dimensionless constants are $t = [\widetilde{T}^n - \widetilde{T}_f{}^n(P)]$, $r = \kappa(\widetilde{T} - \widetilde{T}_s)$ with $\kappa = \alpha_s b_f{}^{1/2} / \alpha_f b^{1/2} T_{f0}{}^{n-1}$ $\gamma = \gamma_0 / [\alpha_f T_{f0}{}^n b]^{1/2}$, and $\gamma_1 = \delta / (b b_f)^{1/2}$. The reduced temperatures are $\widetilde{T} = T / T_{f0}$, $\widetilde{T}_f(P) = T_f(P) / T_{f0}$, and $\widetilde{T}_s(P) = T_s(P) / T_{f0}$.

The analysis involves making simple assumptions for the P dependence of the t, r, $\mathcal{5}$, and $\mathcal{5}_1$ parameters in Eq. (2). Specifically, we assume that only T_f has a significant P dependence, described by $T_f(P) = (1 - \widetilde{P})^{1/n}$, where $\widetilde{P} = P / P_0$ and P_0 is a characteristic pressure deduced later. In $ZrZn_2$ and UGe_2 the P_0 values are very close to the critical pressure P_c at which both the ferromagnetic and superconducting orders vanish, but in other systems this is not necessarily the case. As we will discuss, the nonlinearity (n = 2) of $T_f(P)$ in $ZrZn_2$ and UGe_2 is relevant at relatively high P, at which the N-FM transition temperature $T_F(P)$ may not coincide with $T_f(P)$.

The simplified model in Eq. (2) is capable of describing the main thermodynamic properties of spin-triplet ferromagnetic superconductors. There are three stable phases: (i) the normal (N) phase, given by $\varphi = m = 0$; (ii) the pure ferromagnetic (FM) phase, given by $m = (-t)^{1/2} > 0$, $\varphi = 0$; and (iii) the FS phase, given by $\phi_1^2 = \phi_2^2 = (\gamma m - r - \gamma_1 m^2)/2$, $\phi_3 = 0$, where $\sin\vartheta = -1$, and m satisfies

$$(1 - \gamma_1{}^2) m^3 + \frac{3}{2}\gamma\gamma_1 m^2 + (t - \frac{\gamma^2}{2} - \gamma_1 r) m + \frac{\gamma r}{2} = 0. \tag{3}$$

We note that FS is a two-domain phase as discussed in Refs. [22, 23]. Although Eq. (3) is complicated, some analytical results follow, e.g., we find that the second order phase transition line $\widetilde{T}_{FS}(P)$ separating the FM and FS phases is the solution of

$$T_{FS}(P) = T_s + \frac{\gamma_1}{\kappa} t(T_{FS}) + \frac{\gamma}{\kappa}[-t(T_{FS})]^{1/2} \tag{4}$$

Under certain conditions, the $T_{FS}(P)$ curve has a maximum at $\widetilde{T}_m = \widetilde{T}_s + (\gamma^2 / 4\kappa\gamma_1)$ with pressure P_m found by solving $t(T_m, P_m) = -(\gamma^2 / 4\gamma_1{}^2)$. Examples will be given later, but generally this curve extends from ambient P up to a tri-critical point labeled "B", with coordinates (P_B,T_B), where the FM-FS phase transition occurs at a straight line of first order transition up to a critical-end point C. The lines of all three phase transitions (N-FM, N-FS, and FM-FS) terminate at C. For $P > P_C$ the FM-FS phase transition occurs on a rather flat, smooth line of equilibrium transition of first order up to a second tricritical point A with $P_A \sim P_0$ and $T_A \sim 0$. Finally, the third transition line terminating at C describes the second order phase transition N-FM. The temperatures at the three multi-critical points correspond to $\widetilde{T}_A = \widetilde{T}_s$,

$\widetilde{T}_B = \widetilde{T}_s + \gamma^2(2+\gamma_1)/4\kappa(1+\gamma_1)^2$, and $\widetilde{T}_C = \widetilde{T}_s + \gamma^2/4\kappa(1+\gamma_1)$, while the P values can be deduced from the previous equations. These results are valid whenever $T_f(P) > T_s(P)$, which excludes any pure superconducting phase ($|\psi| > 0$, $m = 0$) in accord with the available experimental data. Note that, for any set of material parameters, $T_A < T_C < T_B < T_m$ and $P_m < P_B < P_C$.

A calculation of the T - P diagram from Eq. (2) for any material requires some knowledge of P_0, T_{f0}, T_s, κ, γ, and γ_1. The temperature T_{f0} can be obtained directly from the experimental phase diagrams. The model pressure P_0 is either identical to or very close to the critical pressure P_c at which the N-FM phase transition line terminates at $T \sim 0$ K. The characteristic temperature T_s of the generic superconducting transition is not available from the experiments and thus has to be estimated using general consistency arguments. For $T_f(P) > T_s(P)$ we must have $T_s(P) = 0$ at $P \geq P_c$, where $T_f(P) \leq 0$. For $0 \leq P \leq P_0$, $T_s < T_C$ and therefore for cases where T_C is too small to be observed experimentally, T_s can be ignored. For systems where T_C is measurable this argument does not hold. This is likely to happen for $T_s > 0$ (for $T_s < 0$, T_C is very small). However, in such cases, pure superconducting phase should be observable. To date there are no experimental results reported for such a feature in $ZrZn_2$ or UGe_2, and thus we can put $T_s = 0$. We remark that negative values of T_s are possible, and they describe a phase diagram topology in which the FM-FS transition line terminates at $T = 0$ for $P < P_c$. This might be of relevance for other compounds, e.g., URhGe.

Typically, additional features of the experimental phase diagram must be utilized. For example, in $ZrZn_2$ these are the observed values of $T_{FS}(0)$ and the slop $\rho_0 = [\partial T_{FS}(P)/\partial P]_0$ at $P = 0$. For UGe_2 one may use T_m, P_m, and P_{0c}, where the last quantity denotes the other solution (below P_c) of $T_{FS}(P) = 0$. The ratios γ/κ and γ_1/κ can be deduced using Eq. (4) and the expressions for T_m, P_m, and ρ_0, while κ is chosen by requiring a suitable value of T_C.

Experiments [13] for $ZrZn_2$ indicate $T_{f0} = 28.5$ K, $T_{FS}(0) = 0.29$ K, $P_0 \sim P_c = 21$ kbar, and $T_F(P) \sim T_f(P)$ is almost a straight line, so $n = 1$ describes the P dependence. The slope for $T_{FS}(P)$ at $P = 0$ is harder to estimate; its magnitude should not exceed $T_{f0}/P_c = 0.014$ on the basis of a straight-line assumption, implying $-0.014 < \rho < 0$. However, this ignores the effect of a maximum, although it is unclear experimentally in $ZrZn_2$, at (T_m, P_m). If such a maximum were at $P = 0$ we would have $\rho_0 = 0$, whereas a maximum with $T_m \sim T_{FS}(0)$ and $P_m << P_0$ provides us with an estimated range $0 < \rho_0 < 0.005$. The choice $\rho_0 = 0$ gives $\gamma/\kappa \approx 0.02$ and $\gamma_1/\kappa \approx 0.01$, but similar values hold for any $\rho_0 < 0.003$. The multi-critical points A and C cannot be distinguished experimentally. Since the experimental accuracy [13] is less than ~ 25 mK in the high-P domain ($P \sim 20$-21 kbar), we suppose that $T_C \sim 10$ mK, which corresponds to $\kappa \sim 10$. We employed these parameters to calculate the $T - P$ diagram using $\rho_0 = 0$ and 0.003. The differences obtained in these two cases are negligible, with both phase diagrams being in excellent agreement with experiment. The latter value is used in

Fig. 1, which gives $P_A \sim P_c = 21.10$ kbar, $P_B = 20.68$ kbar, $P_C = 20.99$ kbar, $T_A = T_F(P_c) = T_{FS}(P_c) = 0$ K, $T_B = 0.077$ K, $T_C = 0.038$ K, and $T_{FS}(0) = 0.285$ K. The low-T region is seen in more detail in Fig. 2, where the *A, B, C* points are shown and the order of the FM-FS phase transition changes from second to first order around the critical end-point *C*. The $T_{FS}(P)$ curve has a maximum at $P_m = 6.915$ kbar and $T_m = 0.301$ K. These results account well for the main features of the experimental behavior [13], including the claimed change in the order of the FM-FS phase transition at relatively high *P*. Within the present model the N-FM transition is of second order up *to* $P_C \sim P_c$. Moreover, if the experiments are reliable in their indication of a first order N-FM transition at much lower *P* values, the theory can accommodate this by a change of sign of b_f, leading to a new tricritical point located at a distinct $P_{tr} < P_C$ on the N-FM transition line. Since $T_C > 0$ a direct N-FS phase transition of first order is predicted in accord with conclusions from de Haas-van Alphen experiments and some theoretical studies (see, e.g., the discussion in Refs. [22], 23]). Such a transition may not occur in other cases where $T_C = 0$. In SFT ($n = 2$) the diagram topology remains the same but points *B* and *C* are slightly shifted to higher *P* (typically by about 1 bar).

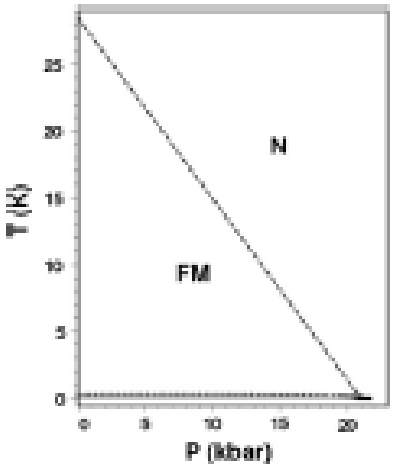


Fig. 1. T - P diagram of $ZrZn_2$ calculated taking $T_s = 0$, $\rho_0 = 0.003$ K/kbar, $T_{f0} = 28.5$ K, $P_0 = 21$ kbar, $\kappa = 10$, and $\gamma / \kappa = 2\gamma_1 / \kappa \approx 0.2$.
The low-T domain of the FS phase is seen more clearly in the following figure

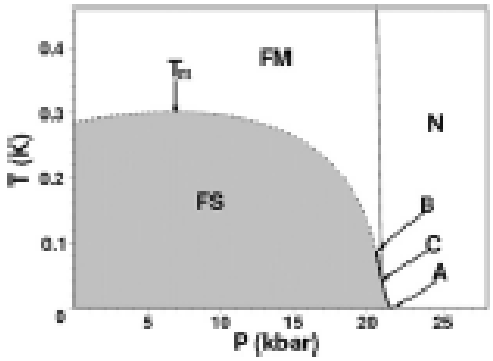


Fig. 2. Detail of Fig. 1 with expanded temperature scale

The experimental data [2, 4, 5] for UGe_2 indicate $T_{f0} = 52$ K, $P_c = 1.6$ GPa (= 16 kbar), $T_m = 0.75$ K, $P_m = 1.15$ GPa, and $P_{0c} = 1.05$ GPa. Using again the variant $n = 1$ for T_f *(P)* and the above values for T_m and P_{0c} we obtain $\gamma/\kappa = 0.098$ and $\gamma_1/\kappa = 0.168$. The temperature $T_C \sim 0.1$ K corresponds to $\kappa \sim 5$. Using these, together with $T_s = 0$, leads to the $T - P$ diagram in Fig. 3, showing only the low-T region of interest. We obtain $T_A = 0$ K, $T_B = 0.481$ K, $T_C = 0.301$ K, $P_A = 1.72$ GPa, $P_B = 1.56$ GPa, and $P_C = 1.59$ GPa. There is agreement with th e main experimental findings, although P_m corresponding to the maximum (found at ~ 1.44 GPa in Fig. 3) is about 0.3 GPa higher than suggested experimentally. If the experimental plots are accurate in this respect, this difference may be attributable to the so-called (T_x) meta-magnetic phase transition in UGe_2, which is related to an abrupt change of the magnetization in the vicinity of (T_m, P_m). Thus, one may suppose that the meta-magnetic effects, which are outside the scope of our current model, significantly affect the shape of the $T_{FS}(P)$ curve by lowering P_m (along with P_B and P_C). It is possible to achieve a lower P_m value (while leaving T_m unchanged), but this has the undesirable effect of modifying P_{c0} to a value that disagrees with experiment. In SFT $(n = 2)$ the multi-critical points are located at slightly higher P (by about 0.01 GPa), as for $ZrZn_2$.

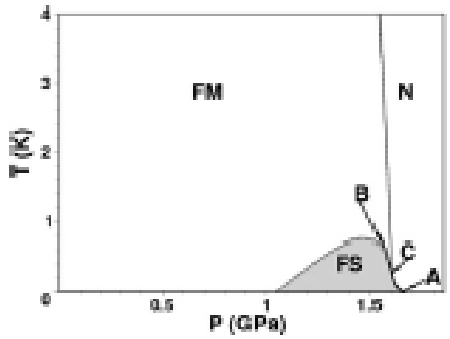


Fig. 3. Low-T part of the T-P diagram of UGe_2 calculated taking $T_s = 0$, $T_f = 52$ K, $P_0 = 1.6$ GPa, $T_m = 0.75$ K, $P_{0c} = 1.05$ GPa ($\kappa = 5$, $\gamma/\kappa = 0.098$, and $\gamma_1/\kappa = 0168$).

The estimates for UGe_2 imply $\gamma_1\kappa \approx 11.9$, so the condition for $T_{FS}(P)$ to have a maximum found from Eq. (4) is satisfied. As we discussed for $ZrZn_2$, the location of this maximum can be hard to fix accurately in experiments. However, P_{c0} can be more easily distinguished, as in the UGe_2 case. Then we have a well-established quantum (zero-temperature) phase transition of second order, i.e., a quantum critical point [22, 23]. From Eq. (4) the existence of this type of solution in systems with $T_s = 0$ (as UGe_2) occurs for $\gamma < \gamma_1$. Such systems (which we label as U-type) are essentially different from those such as $ZrZn_2$ where $\gamma_1 < \gamma$ and hence $T_{FS}(0) > 0$. In this latter case (Zr-type compounds) a maximum $T_m > 0$ may sometimes occur, as discussed earlier. We note that the ratio γ/γ_1 reflects a balance effect between the two ***ψ-M*** interactions. When the trigger interaction (typified by γ) prevails, the Zr-type behavior is found where superconductivity exists at $P = 0$. The same ratio can be expressed as $\gamma_0/\delta.M_0$, which emphasizes that the ground state value of the magnetization at $P = 0$

is also relevant. In general, depending on the ratio of the interaction parameters γ and γ_1, the ferromagnetic superconductors with spin-triplet Cooper fermion pairing can be of two types: type I ($\gamma < \gamma_1$) and type II ($\gamma_1 < \gamma$). The two types are distinguished in their thermodynamic properties.

The quantum phase transition near P_c is of first order. Depending on the system properties, T_C can be either positive (when a direct N-FS first order transition is possible), zero, or negative (when the FM-FS and N-FM phase transition lines terminate at different zero-temperature phase transition points). The last two cases correspond to $T_s < 0$. All these cases are possible in Zr- and U-type compounds. The zero temperature transition at P_{c0} is found to be a quantum critical point, whereas the zero-temperature phase transition at P_c is of first order. As noted, the latter splits into two first order phase transitions.

This classical picture may be changed through quantum fluctuations [22, 23]. An investigation [29, 30] of the quantum fluctuations and the quantum dimensional crossover by renormalization group methods revealed a fluctuation change in the order of this transition to a continuous phase transition belonging to an entirely new class of universality. However, this option exists only for magnetically isotropic order (Heisenberg symmetry) and is unlikely to apply in the known spin-triplet ferromagnetic superconductors.

Even in its simplified form, this theory has been shown to be capable of accounting for a wide variety of experimental behavior. A natural extension to the theory is to add a $\boldsymbol{M}^6$ term which provides a formalism to investigate possible meta-magnetic phase transitions and extend some first order phase transition lines. Another modification of this theory, with regard to applications to other compounds, is to include P-dependence for some of the other GL parameters.

3. FINAL REMARKS

Even in its simplified form, this theory has been shown to be capable of accounting for a wide variety of experimental behavior. A natural extension to the theory is to add a $\boldsymbol{M}^6$ term which provides a formalism to investigate possible meta-magnetic phase transitions and extend some first order phase transition lines. Another modification of this theory, with regard to applications to other compounds, is to include P-dependence for some of the other GL parameters.

The fluctuation and quantum correlation effects can be considered by the respective field-theoretical action of the system, where the order parameters $\boldsymbol{\psi}$ and $\boldsymbol{M}$ are not uniform but rather space and time dependent. The vortex (spatially non-uniform) phase due to the spontaneous magnetization $\boldsymbol{M}$ is another phenomenon which can be investigated by a generalization of the theory by considering non-uniform order parameters fields $\boldsymbol{\psi}$ and $\boldsymbol{M}$. Note, that such theoretical treatments are quite complex and require a number of approximations. As already noted in this paper the magnetic fluctuations stimulate first order phase transitions for both finite and zero phase transition temperatures.

Acknowledgements: Financial support from NFSR-Sofia (through grant Ph. 1507) is gratefully acknowledged.

REFERENCES

[1] D. I. Uzunov, in: "Advances in Theoretical Physics", edited by E. Caianiello (World Scientific, Singapore,1990), p. 96; M. Sigrist and K. Ueda, Rev. Mod. Phys. **63** 239 (1991).

[2] S. S. Saxena, P. Agarwal, K. Ahilan, F. M. Grosche, R. K. W. Haselwimmer, M.J. Steiner, E. Pugh, I. R. Walker, S.R. Julian, P. Monthoux, G. G. Lonzarich, A. Huxley. I. Sheikin, D. Braithwaite, and J. Flouquet, Nature 406, 587 (2000).

[3] A. Huxley, I. Sheikin, E. Ressouche, N. Kernavanois, D. Braithwaite, R. Calemczuk, and J Flouquet, Phys. Rev. **B63**, 144519 (2001).

[4] N. Tateiwa, T. C. Kobayashi, K. Hanazono, A. Amaya, Y. Haga. R. Settai, and Y. Onuki, J. Phys. Condensed Matter **13**, L17 (2001).

[5] A. Harada, S. Kawasaki, H. Mukuda, Y. Kitaoka, Y. Haga, E. Yamamoto, Y. Onuki, K. M. Itoh, E. E. Haller, and H. harima, Phys. Rev. **B75**, 140502 (2007).

[6] D. Aoki, A. Huxley, E. Ressouche, D. Braithwaite, J. Flouquet, J-P.. Brison, E. Lhotel, and C. Paulsen, Nature **413**, 613 (2001).

[7] F. Hard y, A. Huxley, Phys. Rev. Lett. **94**, 247006 (2005).

[8] F. Hardy, A. Huxley, J. Flouquet, B. Salce, G. Knebel, D. Braithwate, D. Aoki, M. Uhlarz, and C. Pfleiderer, Physica B **359-361** 1111 (2005).

[9] N. T. Huy, A. Gasparini, D. E. de Nijs, Y. Huang, J. C. P. Klaasse, T. Gortenmulder, A. de Visser, A. Hamann, T. Goerlach, and H. v. Loehneysen, Phys. Rev. Lett. **99**, 067006 (2007).

[10] N. T. Huy, D. E. de Nijs, Y. K. Huang, and A. de Visser, Phys. Rev. Lett. **100**, 077001 (2008).

[11] T. Akazawa, H. Hidaka, H. Kotegawa, T. C. Kobayashi, T. Fujiwara, E. Yamamoto, Y. Haga, R. Settai, and Y. Onuki, Physica B **359-361**, 1138 (2005).

[12] T. C. Kobayashi,S. Fukushima, H. Hidaka, H. Kotegawa, T. Akazawa, E. Yamamoto, Y. Haga, R. Settai, and Y. Onuki, Physica B **378-361**, 378 (2006).

[13] C. Pfleiderer, M. Uhlatz, S. M. Hayden, R. Vollmer, H. v. Loehneysen, N. R. Berhoeft, and G. G. Lonzarich, Nature **412**, 58 (2001).

[14] E. A. Yelland, S. J. C. Yates, O. Taylor, A. Griffiths, S. M. Hayden, and A. Carrington, Phys. Rev. **B 72**, 184436 (2005).

[15] E. A. Yelland, S. M. Hayden, S. J. C. Yates, C. Pfleiderer, M. Uhlarz, R. Vollmer, H. v Loehneysen, N. R. Bernhoeft, R. P. Smith, S. S. Saxena, and N. Kimura, Phys. Rev. B **72**, 214523 (2005).

[16] C. J. Bolesh and T. Giamarchi, Phys. Rev. Lett. **71**, 024517 (2005); R. D. Duncan, C. Vaccarella, and C. A. S. de Melo, Phys. Rev. B **64**, 172503 (2001).

[17] A. H. Nevidomskyy, Phys. Rev. Lett. **94**, 097003 (2005).

[18] D. I. Uzunov, "Theory of Critical Phenomena" (World Scientific, Singapore, 1993).

[19] K. Machida and T. Ohmi, Phys. Rev. Lett. **86**, 850 (2001).

[20] M. B. Walker and K. V. Samokhin, Phys. Rev. Lett. **88**, 207001 (2002); K. V. Samokhin and M. B. Walker, Phys. Rev. B **66**, 024512 (2002); Phys. Rev. B **66**, 174501 (2002).

[21] J. Linder, A. Sudbo, Phys. Rev. B **76**, 054511 (2007); J. Linder, I. B. Sperstad, A. H. Nevidomskyy, M. Cuoco, and A. Sodbo, Phys. Rev. B **77**, 184511 (2008); J. Linder, T. Yokoyama, and A. Sudbo, Phys. Rev. B **78**, 064520 (2008); J. Linder, A. H. Nevidomskyy, A. Sudbo, Phys. Rev. B **78**, 172502 (2008).

[22] D. V. Shopova and D. I. Uzunov, Phys. Rev. B **72**, 024531 (2005); Phys. Lett. A **313**, 13 (2003).

[23] D. V. Shopova and D. I. Uzunov, in: "Progress in Ferromagnetism Research", ed. by V. N. Murray (Nova Science Publishers, New York, 2006), p. 223; D. V. Shopova and D. I. Uzunov, J. Phys. Studies **4**, 426 (2003) 426; D. V. Shopova and D. I. Uzunov, Compt. Rend Acad. Bulg. Sci. **56**, 35 (2003) 35; D. V. Shopova, T. E. Tsvetkov, and D. I. Uzunov, Cond. Matter Phys. **8**, 181 (2005)181; D. V. Shopova, and D. I. Uzunov, Bulg. J. of Phys. **32**, 81 (2005).

[24] R. A. Cowley, Adv. Phys. **29**,1 (1980); J-C. Toledano and P. Toledano, "The Landau Theory of Phase Transitions" (World Scientific, Singapore, 1987).

[25] K. K. Murata and S. Doniach, Phys. Rev. Lett. **29**, 285 (1972); G. G. Lonzarich and L. Taillefer, J. Phys. C: Solid State Phys. **18**, 4339 (1985); T. Moriya, J. Phys. Soc. Japan **55**, 357 (1986); H. Yamada, Phys. Rev. B **47**, 11211 (1993).

[26] M. G. Cottam, D. V. Shopova, and D. I. Uzunov, Phys. Lett. A **373**, 152 (2008).

[27] D. V. Shopova and D. I. Uzunov, Phys. Rev. B **79**, 064501 (2009).

[28] D. V. Shopova and D. I. Uzunov, Phys. Rep. C **379**, 1 (2003).

[29] D. I. Uzunov, Phys. Rev. B **74**, 134514 (2006).

[30] D. I. Uzunov, Europhys. Lett. **77**, 20008 (2007).

[31] B. I. Halperin, T. C. Lubensky, and S. K. Ma, Phys. Rev. Lett. **32**, 292 (1974); J-H. Chen, T. C. Lubensky, and D. R. Nelson, Phys. Rev. B **17**, 4274 (1978).

PREDICTION OF LIMIT CYCLES IN NONLINEAR DISCRETE SYSTEMS (SIGMA-DELTA MODULATORS)

Valeri Mladenov

Department Theoretical Electrical Engineering, Technical University of Sofia; 8, Kl. Ohridski Str., Sofia 1000, Bulgaria, phone: +0359 965 23 86, e-mail: valerim@tu-sofia.bg

Abstract: *In this paper we present some results about the prediction of limit cycles in discrete nonlinear systems with relay type nonlinearities by using multiple-input describing functions. Sigma-delta modulators (SDMs) are such systems and the approach is applicable for the analysis and design of SDMs. The introduction of Describing Functions (DFs) theory is given, together with the derivations of different type describing functions for the ideal relay nonlinearity. Then the sampled DFs that describe the nonlinear elements in discrete nonlinear systems are introduced and finally the application of this approach for analysis of the limit cycles in SDMs is shown. Several examples and simulation results are presented to demonstrate the applicability of the approach.*

Keywords: Discrete Nonlinear Circuits, Discrete Nonlinear Systems, Sigma- Delta Modulators, Multiple-input describing functions.

1. INTRODUCTION

The Describing Functions method (or short DFs method) is an approximate procedure for investigating the existence of limit cycles in the feedback system shown in the figure below. The concept is one of quasilinearization where a static nonlinear characteristic is represented by a gain depending upon the magnitude of the input signal. The input signal is evaluated on the assumption that it is a sinusoid. In this view the input signal of the nonlinear block NL will be of the following kind: $e=asin(\omega t)$.

The describing function method's philosophy is to replace the nonlinear system part *NL* of the given feedback loop given in Fig. 1 with a linear gain. The gain depends on the amplitude of the input. This is done in such a way that when oscillation is predicted for the system, the amplitude of the oscillation will also be determined. We will use the describing function to predict the value of the amplitude and frequency of limit cycles in these systems.

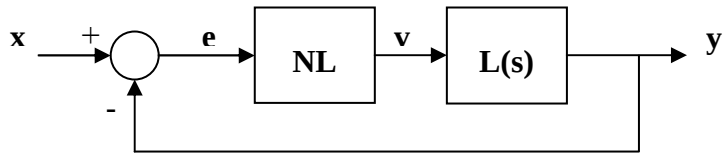


Fig. 1. The feedback system with nonlinearity NL

When a non-linearity is embedded in a system, it may seem odd to use sinusoidal inputs to analyze the response. This procedure is justified whenever the nonlinear dynamics are such that the output of the nonlinear element is filtered in such a way that

the higher frequencies are negligible. Because of the low pass filtering of the transfer function *L(s)* this is justified.

Sigma-Delta modulation has become in recent years an increasingly popular choice for robust and inexpensive analog-to-digital (A/D) and digital-to-analog (D/A) conversion [1], [2], [3]. As a result of this, analog-to-digital and digital-to-analog converters based on 1-bit Sigma-Delta Modulators (SDMs) are widely used in different applications. Sigma delta modulators are the most popular form of analog-to-digital conversion used in audio and wireless applications. They are also commonly used in D/A converters, sample rate converters, and digital power amplifiers. Despite the widespread use of SDMs theoretical understanding of sigma delta concept is still very limited. This is a consequence of the fact that these systems are nonlinear, due to the presence of a discontinuous nonlinearity - the quantizer (relay type nonlinearity). Limit cycles are well known phenomena that often appear in practical SDMs. In frequency domain they correspond to discrete peaks in the frequency spectrum of the modulator. If these peaks are inside the signal base band, the total harmonic distortion increases. Because of this, for data processing applications it is very important to predict and describe possible limit cycles. As will be stressed further, for many practical applications SDMs work on several limit cycles and in [4] this mechanism is investigated in details. In this work a limit cycle model of the SDM operation is created. The limit cycle behavior of the SDM is considered as a result of the interaction of the internal asynchronous limit cycle mechanism with the external sampling clock. The main results in [4] are based on sampled two-sinusoid describing function that has been introduced. The aim of this article is to develop further the approach in [4] by using sampled three-sinusoid describing functions.

The paper is organized as follows. In the next chapter we describe Multiple-Input Describing Functions [5] and focus our investigation on two-sinusoid-input and three-sinusoid input DFs for Ideal relay nonlinearity. Then in chapter 3 we present the sampled describing functions. In chapter 4 we show the use of the describing functions considered for prediction of limit cycles in SDMs and in chapter 5 we demonstrate the use of sampled describing functions for prediction of limit cycles in SDMs. The paper finishes with concluding remarks in the last chapter.

2. MULTIPLE-INPUT DESCRIBING FUNCTIONS

In this chapter we will describe the Multiple-Input Describing Functions [5] regarding different input signals applied to the system from Fig. 1.

If the input signal of the system in Fig. 1 is zero (zero input), then the output of the nonlinear block can be written as a function dependent on $asin(\omega t)$ i.e. $v=f(asin(\omega t))$ where f is the nonlinear function that describes the nonlinear element NE. From the Fourier series expansion of the output signal of the nonlinear block we get the following equation:

$$f(a\sin\omega t)=\sum_{k=0}^{\infty}\left[a_k(a)\sin\omega t+b_k(a)\cos\omega t\right]\approx a_1(a)\sin\omega t+b_1(a)\cos\omega t$$
$$=\sqrt{a_1(a)^2+b_1(a)^2}\,\sin(\omega t+\varphi(a))=\left|N(a)\right|\sin(\omega t+\varphi(a)) \tag{1}$$

where $\left|\mathrm{N(a)}\right|=\sqrt{a_1^2(a)+b_1^2(a)}$ and $\varphi(a)=\tan^{-1}\dfrac{b_1(a)}{a_1(a)}$

The coefficients $a_1(a)$ and $b_1(a)$ are the fundamental-harmonics determined by the Fourier expansion [6], [7]:

$$a_1(a)=\frac{2}{T}\int_0^T f(a\sin(\omega t))\sin(\omega t)\,dt$$
$$b_1(a)=\frac{2}{T}\int_0^T f(a\sin(\omega t))\cos(\omega t)\,dt \tag{2}$$

The describing function $N(a)$ is the transfer function between the input gain of the nonlinear block and the first harmonic in the output:

$$N(a)=\frac{a_1(a)+jb_1(a)}{a}=\frac{\sqrt{a_1^2(a)+b_1^2(a)}}{a}e^{j\varphi(a)}=\frac{\left|N(a)\right|}{a}e^{j\varphi(a)}, \tag{3}$$

where $\varphi(a)=\tan^{-1}\dfrac{b_1(a)}{a_1(a)}$

If the nonlinear element is an ideal relay with characteristic

$$f(e)=\begin{cases} D & for\ \ e\geq 0 \\ -D & for\ \ e<0 \end{cases} \tag{4}$$

shown in Fig. 2, the Fourier coefficients $a_1(a)$ and $b_1(a)$ are

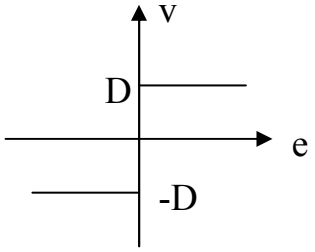


Fig. 2. The nonlinear characteristic of the ideal relay.

$$
\begin{aligned}
a_1 &= \frac{2}{T}\int_0^T f(a\sin\omega t)\sin\omega t\,dt = \frac{4}{T}\int_0^{T/2} D\sin\omega t\,dt \quad \text{(by symmetry)} \\
&= -\frac{4D}{T\omega}\left[\cos\omega t\right]_0^{T/2} = \frac{4D}{\pi} \qquad (\omega = \frac{2\pi}{T}) \\
b_1 &= \frac{2}{T}\int_0^T f(a\sin\omega t)\cos\omega t\,dt = \frac{4}{T}\int_0^{T/2} D\cos\omega t\,dt \quad \text{(by symmetry)} \\
&= -\frac{4D}{T\omega}\left[\sin\omega t\right]_0^{T/2} = 0 \qquad (\omega = \frac{2\pi}{T})
\end{aligned}
\tag{5}
$$

and thus the describing function becomes

$$N(a) = \frac{a_1(a) + jb_1(a)}{a} = \frac{4D}{\pi a} \tag{6}$$

For the input signals x of the system given in Table 1, the corresponding error signals e (input signals for the nonlinear block NL) are also given there.

These input signals for the nonlinear element NL of the feedback system are of course determined under the assumption that the system is in an oscillating state, i.e. the nonlinear block has sinusoidal inputs and therefore a describing function can be obtained.

Input signals	**Error signal (input of NL)**
$x = 0$	$e = a\sin(\omega_a t)$
$x = M \ \ (constant)$	$e = b + a\sin(\omega_a t)$
$x = V\sin(\omega_b t)$	$e = a\sin(\omega_a t) + b\sin(\omega_b t + \theta_b)$
$x = V_1\sin(\omega_b t) + V_2\sin(\omega_c t)$	$e = a\sin(\omega_a t) + b\sin(\omega_b t + \theta_b) + c\sin(\omega_c t + \theta_c)$

Table 1. Reference inputs with corresponding error signals.

If a certain input signal of the nonlinear element can be separated into different parts, then the describing function can also be described in different parts. Consider the reference input $x=M$ and error $e=b+a\sin(\omega t)$. The describing function here has one part, the complex gain $N_a(a,b)$ that relates the sinusoidal input component with the fundamental output component. The second gain $N_b(a,b)$ relates the mean output of the nonlinear block with its mean input. Just as described in the beginning of this chapter these gains can be determined from Fourier expansion. In the case with a bias input M, equations (2) can be rewritten and used to determine $N_a(a,b)$ in the following way:

$$
\begin{aligned}
a_1(a,b) &= \frac{2}{T}\int_0^T f(e)\sin(\omega_a t)\,dt = \frac{1}{\pi}\int_0^{2\pi} f(b + a\sin(\psi_a))\sin(\psi_a)\,d\psi_a \\
b_1(a,b) &= \frac{2}{T}\int_0^T f(e)\cos(\omega_a t)\,dt = \frac{1}{\pi}\int_0^{2\pi} f(b + a\sin(\psi_a))\cos(\psi_a)\,d\psi_a
\end{aligned}
\tag{7}
$$

where $\psi_a = \omega_a t$

If the nonlinear element is an ideal relay shown in Fig. 2 the Fourier coefficients $a_1(a)$ and $b_1(a)$ in this case are [5]:

$$a_1(a,b) = \frac{1}{\pi}\left[\int_0^{\pi+\delta} D\sin(\psi_a)\,d\psi_a + \int_{\pi+\delta}^{2\pi-\delta}(-D)\sin(\psi_a)\,d\psi_a + \int_{2\pi-\delta}^{2\pi} D\sin(\psi_a)\,d\psi_a\right]$$
$$= \frac{4D}{\pi}\cos(\delta) = \frac{4D}{\pi}\cos(\arcsin\frac{b}{a}) = \frac{4D}{\pi}\sqrt{1-\left(\frac{b}{a}\right)^2} \tag{8}$$

and

$$b_1(a,b) = \frac{1}{\pi}\left[\int_0^{\pi+\delta} D\cos(\psi_a)\,d\psi_a + \int_{\pi+\delta}^{2\pi-\delta}(-D)\cos(\psi_a)\,d\psi_a + \int_{2\pi-\delta}^{2\pi} D\cos(\psi_a)\,d\psi_a\right] = 0 \tag{9}$$

Hence

$$N_a(a,b) = \frac{a_1(a,b) + jb_1(a,b)}{a} = \frac{4D}{\pi a}\sqrt{1-\left(\frac{b}{a}\right)^2} \tag{10}$$

and

$$N_b(a,b) = \frac{1}{2\pi b}\int_0^{2\pi} f(b + a\sin(\psi_a))\,d\psi_a = \frac{2D}{\pi b}\arcsin\frac{b}{a} \tag{11}$$

For the calculation of the describing functions for a nonlinear block with two sinusoid-inputs, double Fourier series expansion is used. The input to the nonlinear block is chosen in the following fashion: $e = a\sin(\Psi_a) + b\sin(\Psi_b)$, where $\Psi_b = \omega_b t$ and $\Psi_a = \omega_a t$. The output then becomes $v = f(a\sin(\Psi_a) + b\sin(\Psi_b))$. The Fourier coefficients P_{mn} and Q_{mn} used here are:

$$P_{mn} = \frac{1}{2\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(a\sin(\psi_a) + b\sin(\psi_b))\sin(m\psi_a + n\psi_b)\,d\psi_a d\psi_b$$
$$Q_{mn} = \frac{1}{2\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(a\sin(\psi_a) + b\sin(\psi_b))\cos(m\psi_a + n\psi_b)\,d\psi_a d\psi_b \tag{12}$$

After rewriting the integral for Q_{mn} and some substitutions, it can be proven that $Q_{mn}=0$ for odd $m+n$. This procedure is fully explained in [5]. Further, the only two outputs in terms of frequency we are interested in, are ω_a and ω_b. The output term of frequency ω_a only occurs for $m=1$ and $n=0$ and the output term of frequency ω_b only occurs for $m=0$ and $n=1$. Thus for these two frequencies it follows that $m=0$ and $n=1$. The Two-Sinusoid-Input Describing Functions expressions then become:

$$N_a(a,b) = \frac{P_{10}}{a} = \frac{1}{2\pi^2 a}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(a\sin(\psi_a) + b\sin(\psi_b))\sin(\psi_a)\,d\psi_a d\psi_b$$
$$N_b(a,b) = \frac{P_{01}}{b} = \frac{1}{2\pi^2 b}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(a\sin(\psi_a) + b\sin(\psi_b))\sin(\psi_b)\,d\psi_a d\psi_b \tag{13}$$

Considering the ideal relay characteristic given in Fig. 2 it is proven in [5] that

$$N_a(a,b)=\frac{8D}{\pi^2 a}E(k^2)=\frac{8D}{\pi^2 a}E(m),\quad \text{for } k=\frac{b}{a}<1 \text{ and } m=k^2 \tag{14}$$

and

$$N_b(a,b)=\frac{8D}{\pi^2 k^2 a}\Big[E(k)-(1-k^2)K(k)\Big],\quad \text{for } k=\frac{b}{a} \text{ and } m=k^2 \tag{15}$$

The Three-Sinusoid-Input describing functions for ideal relay are derived in [8], [9]. Here we will describe the approach briefly. The nonlinear element NL is modeled by the Three-Sinusoid-Input describing functions when its periodic input e is considered as a sum of three sinusoids $e= a\sin(\Psi_a)\, a\sin(\Psi_b)+b\sin(\Psi_c)$, where $\Psi_a=\omega_a t$, $\Psi_b=\omega_b t$ and $\Psi_c=\omega_c t$. Thus the output of NL is $v=f(a\sin(\Psi_a)\, a\sin(\Psi_b)+b\sin(\Psi_c))$. In this case the independent input signal x of the system from Fig. 1 consists of two sinusoids with frequencies ω_b and ω_c. The Three-Sinusoid-Input describing functions replaces the nonlinear block and they are valid of course under the assumption of low-pass filtering properties of the linear part of the system $L(s)$. The Fourier coefficients P_{lmn} and Q_{lmn} of the Fourier expansion of f are:

$$\begin{aligned}
P_{lmn} &= \frac{1}{4\pi^3}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(a\sin(\psi_a)+b\sin(\psi_b)+c\sin(\psi_c))\sin(l\psi_a+m\psi_b+n\psi_c)\,d\psi_a d\psi_b d\psi_c \\
Q_{lmn} &= \frac{1}{4\pi^3}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(a\sin(\psi_a)+b\sin(\psi_b)+c\sin(\psi_c))\cos(l\psi_a+m\psi_b+n\psi_c)\,d\psi_a d\psi_b d\psi_c
\end{aligned} \tag{16}$$

From the Fourier series we can see that the output term of frequency ω_a only occurs for $l=1$, $m=0$ and $n=0$. The output term of frequency ω_b only occurs for $l=0$, $m=1$ and $n=0$ and ω_c only occurs for $l=0$, $m=0$ and $n=1$. The three-sinusoid-input describing functions then become [5]

$$\begin{aligned}
N_a(a,b,c)&=\frac{P_{100}}{a}=\frac{1}{4\pi^3 a}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(a\sin(\psi_a)+b\sin(\psi_b)+c\sin(\psi_c))\sin(\psi_a)\,d\psi_a d\psi_b d\psi_c \\
N_b(a,b,c)&=\frac{P_{010}}{b}=\frac{1}{4\pi^3 b}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(a\sin(\psi_a)+b\sin(\psi_b)+c\sin(\psi_c))\sin(\psi_b)\,d\psi_a d\psi_b d\psi_c \\
N_c(a,b,c)&=\frac{P_{001}}{c}=\frac{1}{4\pi^3 c}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(a\sin(\psi_a)+b\sin(\psi_b)+c\sin(\psi_c))\sin(\psi_c)\,d\psi_a d\psi_b d\psi_c
\end{aligned} \tag{17}$$

For the ideal relay given in Fig. 2, $Q_{lmn}=0$ and the describing functions $N_a(a, b, c)$, $N_b(a, b, c)$ and $N_c(a, b, c)$ are obtained in [9] together with a simple polynomial approximation

$$N_a(a,b,c)=\frac{8D}{\pi^3 a}\int_0^{\pi/2}\int_0^{\pi/2}\left[\sqrt{1-\left(\frac{b}{a}\sin\psi_b-\frac{c}{a}\sin\psi_c\right)^2}+\sqrt{1-\left(\frac{b}{a}\sin\psi_b+\frac{c}{a}\sin\psi_c\right)^2}\right]d\psi_b d\psi_c \quad (18)$$

$$N_b(a,b,c)=\frac{8D}{\pi^3 ra}\int_0^{\pi/2}\int_0^{\pi/2}\left[\sqrt{1-\left(\frac{1}{r}\sin\psi_a+\frac{q}{r}\sin\psi_c\right)^2}+\sqrt{1-\left(\frac{1}{r}\sin\psi_a-\frac{q}{r}\sin\psi_c\right)^2}\right]d\psi_a d\psi_c \quad (19)$$

$$N_c(a,b,c)=\frac{8D}{\pi^3 qa}\int_0^{\pi/2}\int_0^{\pi/2}\left[\sqrt{1-\left(\frac{r}{q}\sin\psi_b-\frac{1}{q}\sin\psi_a\right)^2}+\sqrt{1-\left(\frac{1}{q}\sin\psi_a-\frac{r}{q}\sin\psi_b\right)^2}\right]d\psi_a d\psi_b \quad (20)$$

where $r=b/a$ and $q=c/a$.

The corresponding polynomial approximations of $N_a(a, b, c)$, $N_b(a, b, c)$ and $N_c(a, b, c)$ are [9]

$$N_a(a,b,c)=\frac{D}{a}\left[p_{a00}+p_{a10}q+p_{a01}r+p_{a20}q^2+p_{a11}qr+p_{a02}r^2+p_{a30}q^3+p_{a21}q^2r+p_{a12}qr^2+p_{a03}r^3\right] \quad (21)$$

$$N_b(a,b,c)=\frac{D}{ra}\left[p_{b00}+p_{b10}q+p_{b01}r+p_{b20}q^2+p_{b11}qr+p_{b02}r^2+p_{b30}q^3+p_{b21}q^2r+p_{b12}qr^2+p_{b03}r^3\right] \quad (22)$$

$$N_c(a,b,c)=\frac{D}{qa}\left[p_{c00}+p_{c10}q+p_{c01}r+p_{c20}q^2+p_{c11}qr+p_{c02}r^2+p_{c30}q^3+p_{c21}q^2r+p_{c12}qr^2+p_{c03}r^3\right] \quad (23)$$

where the coefficients are

$[p_{a00}, p_{a10}, p_{a01}, p_{a20}, p_{a11}, p_{a02}, p_{a30}, p_{a21}, p_{a12}, p_{a03}]$ = [1.2067, 0.3930, 0.3930, -0.8163, -1.3672, -0.8164, 0.0694, 0.7768, 0.7768, 0.0694]

$[p_{b00}, p_{a10}, p_{b01}, p_{a20}, p_{b11}, p_{b02}, p_{b30}, p_{b21}, p_{b12}, p_{b03}]$ = [0.7103, -0.1785 -0.7732, 0.0588, 1.8302, 1.4378, 0.1168, -0.4359, -1.9687, -0.2804]

$[p_{c00}, p_{c10}, p_{c01}, p_{c20}, p_{c11}, p_{c02}, p_{c30}, p_{c21}, p_{c12}, p_{c03}]$ = [1.5851, 0.9697, 1.4346, -2.0483, 10.6524, -7.7795, 1.7185, -2.2441, -8.7282, 6.5833]

3. SAMPLED DESCRIBING FUNCTIONS

In this chapter a brief description of sampled describing functions approach for analysis of the limit cycles in SDM will be presented. The presentation will follow [4] and it is important for complete understanding the problem. The starting point is the presentation of SDM trough sampled asynchronous SDM. The block diagram of SDM based on sampling with frequency f_s (period $T_s=1/f_s$) the asynchronous one, is given in Fig. 3. In this model a zero order hold (ZOH) is used after the sampling, in order to obtain a continuous signal $y^*(nT_s)$ that is fed back to the continuous input of the system.

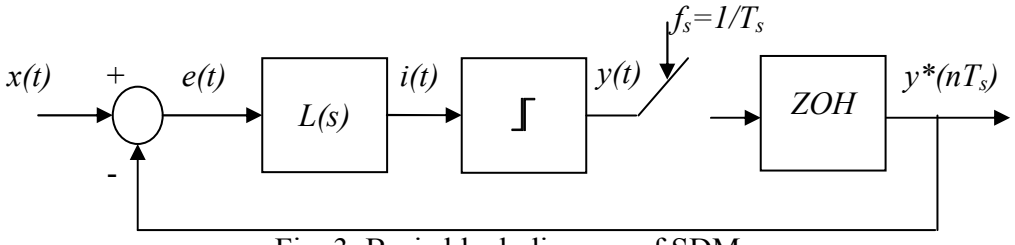


Fig. 3. Basic block diagram of SDM.

The ideal quantizer is memoryless and its position with respect to the sampling operation is of no consequence for the loop operation of the system. To decouple the amplitude quantization from the sampling, the sampling is performed after the quantizer. However, it is important to point out that by placing the sampler after the quantizer, the model gives a better representation for continuous time loop filters. For discrete time loop filters, where the sampling is performed before the quantizer, special attention has to be paid to the possibility for the introduction of aliased components in front of the quantizer. The Limit Cycle Model (LCM) of the SDM assumes a steady state oscillation denoted as limit cycle. The quantizer input signal is approximated with a sinusoid so that the DF theory can be applied. In Fig. 4 the quantizer is modeled with its describing function *N(a)*, where a denotes the amplitude of the quantizers' sinusoidal input signal.

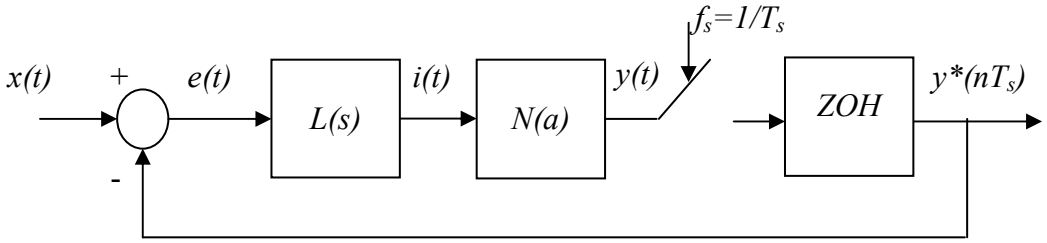


Fig. 4. Describing Function linearized SDM.

The sampling mechanism is explained in details in [4] and the conclusion is that the sampling switch and the ZOH introduce a phase delay φ_s between the quantizer output signal and the SDM output bitstream. This phase delay can take a value between zero and a clock period.

Because in idle mode the SDM is in steady state, the clock introduced phase delay is fixed and depends on the loop filter parameters. The block diagram of the limit cycle model of the SDM is shown in Fig. 5, where the quantizer, the sampling switch and the ZOH are represented by the sampled DF.

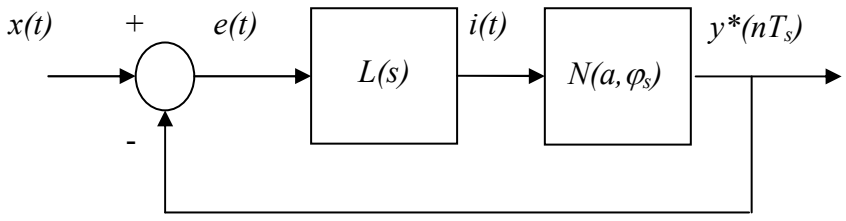


Fig. 5. SDM with sampled DF.

The separation of the quantization in amplitude from the quantization in time allows the introduction of a modified DF that can be denoted as sampled DF and is defined, as follows:

$$N(a,\varphi_s) = N(a)e^{-j\varphi_s} \tag{24}$$

The magnitude of the sampled DF is in fact, the magnitude of the usual DF presented in Chapter 2, while the sampling is incorporated as a phase component in the

overall describing function. In [4], sampled DFs of the binary quantizer are introduced for different combinations of DC or harmonics inputs. Here we will present only the sampled DFs for two sine wave inputs and will introduce the sampled DFs for three sine wave inputs.

The fundamental observation in [4] is that the only idle oscillations that can exist in the SDM loop are even integer multiples *2N* of the sampling period T_s and they are called sub-harmonic modes (because they have a frequency that is an even integer fraction of the applied sampling frequency f_s). Due to sampling the phase delay is different for every sub-harmonic mode. Those modes have frequencies $f_s/2$, $f_s/4$ and $f_s/8$. They are called respectively first, second and fourth.

For limit cycle with frequency $f_s/2$, the clock can introduce in the SDM loop a maximum phase delay of 180 deg. This is true, because for the first sub-harmonic limit cycle, the sampling period corresponds to the half of the period of the limit cycle. For the limit cycle with frequency $f_s/4$, the clock can introduce in the SDM loop a maximum phase delay of 90 deg and for the limit cycle with frequency $f_s/8$, the clock can introduce in the SDM loop a maximum phase delay of 45 deg. Thus for the idle limit cycle oscillations with a lower frequency the impact of sampling on the loop phase behavior decreases proportionally.

When the input signal *x* is nonzero (busy operation), the phase shift that is introduced by the clock depends on the frequency and the amplitude of the signal in front of the quantizer at the particular sampling instance. As this amplitude is continuously varying, the phase shift that is introduced is not fixed. That is why for busy operation the term phase uncertainty is used in [4] for the clock added phase delays. This phase uncertainty remains bounded between the phase limits for each limit cycle mode as determined for idle operation.

According to the above considerations, the idle limit cycle is a stable periodic mode that appears in the SDM for zero input. Let for example the linear loop filter has a second order frequency transfer characteristic

$$L(j\omega) = \frac{G.(j\omega + \omega_z)}{(j\omega + \omega_p)^2} \tag{25}$$

For the SDM that can sustain at least one limit cycle, in analogy to the continuous systems [4] - [9], the relation between the linear part of the loop and the DF representation of the non-linear element is given by:

$$L(j\omega_k) = -\frac{1}{N(a_k, \varphi_{s,k})} \tag{26}$$

where a_k is the *k*-sub-harmonic limit cycle amplitude, $\underline{\omega_k}$ is the *k*-sub-harmonic limit cycle frequency, and $\varphi_{s,k}$ is the clock added loop phase delay for the *k*-sub-harmonic limit cycle mode.

The solutions of equation (26) give the frequency and amplitude of the possible limit cycle oscillations. In the studied SDM this expression reveals the phase/magnitude relation in the closed-loop at each sampling instance. In fact it allows us to investigate the SDM loop for an inherent asynchronous mechanism. The clock operation is treated as superimposed to this internal mechanism and the analysis establishes the interaction of this mechanism with the applied clock frequency.

The observations in [4] are that a finite set of limit cycles can exist in the studied SDM (sampled system), such that in each case the limit cycle frequency is a fraction of the applied clock frequency. For each specific idle limit cycle the clock introduces some fixed delay in the loop. The limit cycle modes can be identified if this delay is taken into account for the evaluation done according to (26). The evaluation can easily be performed graphically as illustrated in Fig. 6 [4].

The clock introduced phase shift for each f_s/N frequency is shown as a horizontal line originating from the discrete frequency points of the locus $L(j\omega_s/N)$ and with length equal to 2π/N as described above. In the asynchronous case an indication for a limit cycle is given by an intersection between $L(j\omega)$ and $-1/N(A)$.

For synchronous (sampled) SDM the limit cycle modes correspond to those fractions of the clock frequency for which the sampling operation adds enough delay in the loop such that the delay lines starting from $L(j\omega_s/N)$ can cross $-1/N(A)$.

For the example given in Fig. 6, two limit cycle modes are possible with frequencies $f_s/2$ and $f_s/4$ (green points). Limit cycles with lower frequencies are not possible because the total phase delay in the loop for lower frequencies does not provide 360 degrees phase shift and thus cannot satisfy the oscillation conditions.

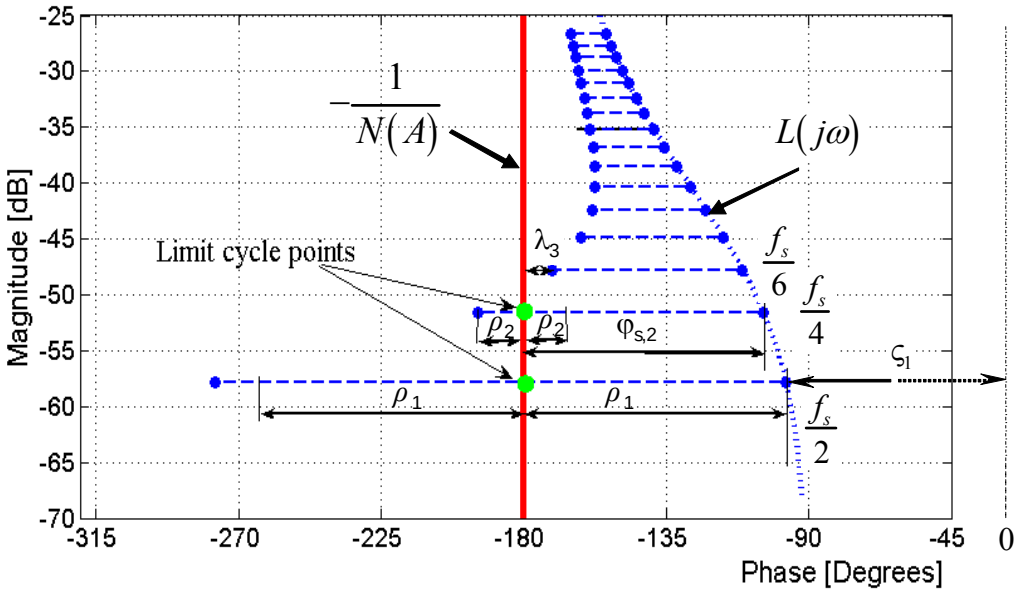


Fig. 5. Graphical description of the limit cycle model of the SDM operation.

The phase margin λ_k for the prevention of lower sub-harmonic modes (see Fig. 6) (where i is an index for the particular sub-harmonic mode) in the loop, is the distance between the maximum phase delay that can be introduced for the particular sub-

harmonic mode, and the -180 deg. line. The phase boundary ρi for each sub-harmonic mode gives a measure for the amount of phase with which the maximum added clock phase delay may exceed the -180 deg. phase line.

Every SDM operates at one or more limit cycles. In fact, the situation when only one limit cycle is possible in the system is a very rare case. In most practical situations, at least two limit cycles should be introduced. As it is stated in [4], the SDM limit cycle state depends entirely on the instantaneous values of the input signal. Here, we state that, in contradiction with the general perception of the SDM operation, the limit cycle with frequency $f_s/2$ is not always the only possible one. Moreover, it may not result in the best performance for the system. In fact, it is shown in [4] that several sub-harmonic modes have to be implemented, in order to get the best trade-off for the system performance.

It should be stressed that DF method is an approximate technique, often useful in engineering practice, which may be used to study the limit cycles in nonlinear systems. The most important requirement to use the approximation of the nonlinear block with DF is that the linear part of the system has to work as a lowpass filter, i.e. to attenuate higher frequencies.

4. PREDICTION OF LIMIT CYCLES IN SDMS

The evaluation in the previous chapter showed that SDMs can operate at more than one limit cycle. In [4], the limit cycle behavior for DC and harmonic inputs is described. The SDM behavior when driven with sinusoidal input signals is evaluated also. For the analysis, the sampled two-sinusoid-input DFs are employed. In this article, we consider the SDM behavior when driven with two sinusoidal input signals. It is inspired by the practical situation when the input signal of the SDM has also very high frequency components. This signal can be modeled as a sum of two sinusoids: one in the baseband and the second one that is very high and sometimes close to the sampling frequency. In order to incorporate the impact of a high frequency signal on the SDM operation, it is necessary to utilize the sampled three-sinusoidal-input DFs for the binary quantizer (ideal relay) described in Chapter 3.

The SDM limit cycle behavior with two sinusoidal inputs is governed by the same mechanisms as discussed in [4] for DC input and one sinusoidal input. The limit cycle behavior is determined by the possible limit cycles in idle mode. The stability region for each possible limit cycle is dependent on the phase boundary for each limit cycle that is possible for the particular SDM. Transitions between the cycles are dependent on the total phase accumulation at each sampling moment due the loop filter, the delay due to the sampling, the input signal and the extra loop delay.

According to DFs method we suppose that the input signal in front of the quantizer consists of three sinusoids, the limit cycle sinusoid with frequency ω_a (even integer fraction of the sampling frequency f_s) and two sinusoids with frequencies ω_b and ω_c due to the signals applied to the SDM input:

$$i(t) = a\sin(\omega_a t) + b\sin(\omega_b t + \theta_b) + c\sin(\omega_c t + \theta_c) = a\sin(\psi_a) + b\sin(\psi_b) + c\sin(\psi_c) \quad (27)$$

where a is the amplitude of the limit cycle and b and c are the amplitudes of the sinusoids due to the signals applied to the SDM input.

Suppose that we have the case depicted in Fig. 6, where only the first and second limit cycle modes with frequencies $\omega_1=\omega_s/2$ and $\omega_2=\omega_s/4$ are possible. When the first limit cycle mode is active, (27) has the form

$$i_1(t) = a_1 \sin((\omega_s / 2)t) + b\sin(\omega_b t + \theta_b) + c\sin(\omega_c t + \theta_c) \tag{28}$$

and when the second limit cycle mode is active, (26) has the form

$$i_2(t) = a_2 \sin((\omega_s / 4)t) + b\sin(\omega_b t + \theta_b) + c\sin(\omega_c t + \theta_c) \tag{29}$$

At each sampling instant, operation at only one limit cycle mode is assumed such that the signal in front of the quantizer can be described by either i_1 or i_2. From (27), when only two idle limit cycles are possible, the SDM operation with two sinusoidal input signals can be described by three groups sampled Three-Sinusoidal-Input DFs defined as follows:

a) Two Three-Sinusoidal-Input DFs, which describe the quantizer response to the limit cycle in the presence of two sinusoidal components:

$N_{a_1}(a_1,b,c)e^{-j\varphi_{s,1}}$ – sampled three-sinusoidal-input DF of the quantizer for limit cycle 1 in the presence of two sinusoidal components with amplitudes b and c;

$N_{a_2}(a_2,b,c)e^{-j\varphi_{s,2}}$ – sampled three-sinusoidal-input DF of the quantizer for limit cycle 2 in the presence of two sinusoidal components with amplitudes b and c;

b) Two Three-Sinusoidal-Input DFs, which describe the quantizer response to the first sinusoidal component in the presence of the limit cycle and the second sinusoidal component:

$N_b(a_1,b,c)e^{-j\varphi_{s,1}}$ – sampled three-sinusoidal-input DF of the quantizer for the sinusoidal component with amplitude b in i_1 in the presence of limit cycle 1 and the sinusoidal component with amplitude c in i_1;

$N_b(a_2,b,c)e^{-j\varphi_{s,2}}$ – sampled three-sinusoidal-input DF of the quantizer for the sinusoidal component with amplitude b in i_2 in the presence of limit cycle 2 and the sinusoidal component with amplitude c in i_2;

c) Two Three-Sinusoidal-Input DFs, which describe the quantizer response to the second sinusoidal component in the presence of the limit cycle and the first sinusoidal component:

$N_c(a_1,b,c)e^{-j\varphi_{s,1}}$ – sampled three-sinusoidal-input DF of the quantizer for the sinusoidal component with amplitude c in i_1 in the presence of limit cycle 1 and the sinusoidal component with amplitude b in i_1;

$N_c(a_2,b,c)e^{-j\varphi_{s,2}}$ – sampled three-sinusoidal-input DF of the quantizer for the sinusoidal component with amplitude c in i_2 in the presence of limit cycle 2 and the sinusoidal component with amplitude b in i_2.

The system behavior at each sampling instant is described by a set of three DFs from the above: one DF for the limit cycle and two for the input signals. In the case when $a_1 > b > c$ and $a_2 > b > c$, based on the current research the analytical description of the magnitude components of the sampled Three-Sinusoidal-Input DF have the following approximate form:

$$N_a(a_1,b,c)=\frac{D}{a_1}\left[p_{a00}+p_{a10}q_1+p_{a01}r_1+p_{a20}q_1^{\,2}+p_{a11}q_1r_1+p_{a02}r_1^{\,2}+p_{a30}q_1^{\,3}+p_{a21}q_1^{\,2}r_1+p_{a12}q_1r_1^{\,2}+p_{a03}r_1^{\,3}\right]$$

$$r_1=\frac{b}{a_1},\ q_1=\frac{c}{a_1}$$

$$N_a(a_2,b,c)=\frac{D}{a_2}\left[p_{a00}+p_{a10}q_2+p_{a01}r_2+p_{a20}q_2^{\,2}+p_{a11}q_2r_2+p_{a02}r_2^{\,2}+p_{a30}q_2^{\,3}+p_{a21}q_2^{\,2}r_2+p_{a12}q_2r_2^{\,2}+p_{a03}r_2^{\,3}\right]$$

$$r_2=\frac{b}{a_2},\ q_2=\frac{c}{a_2}$$

with [p_{a00}, p_{a10}, p_{a01}, p_{a20}, p_{a11}, p_{a02}, p_{a30}, p_{a21}, p_{a12}, p_{a03}] = [*1.2067, 0.3930, 0.3930, -0.8163, -1.3672, -0.8164, 0.0694, 0.7768, 0.7768, 0.0694*]

$$N_b(a_1,b,c)=\frac{D}{ra_1}\left[p_{b00}+p_{b10}q_1+p_{b01}r_1+p_{b20}q_1^{\,2}+p_{b11}q_1r_1+p_{b02}r_1^{\,2}+p_{b30}q_1^{\,3}+p_{b21}q_1^{\,2}r_1+p_{b12}q_1r_1^{\,2}+p_{b03}r_1^{\,3}\right]$$

$$r_1=\frac{b}{a_1},\ q_1=\frac{c}{a_1}$$

$$N_b(a_2,b,c)=\frac{D}{ra_2}\left[p_{b00}+p_{b10}q_2+p_{b01}r_2+p_{b20}q_2^{\,2}+p_{b11}q_2r_2+p_{b02}r_2^{\,2}+p_{b30}q_2^{\,3}+p_{b21}q_2^{\,2}r_2+p_{b12}q_2r_2^{\,2}+p_{b03}r_2^{\,3}\right]$$

$$r_2=\frac{b}{a_2},\ q_2=\frac{c}{a_2}$$

with [p_{b00}, p_{a10}, p_{b01}, p_{a20}, p_{b11}, p_{b02}, p_{b30}, p_{b21}, p_{b12}, p_{b03}] = [*0.7103, -0.1785 -0.7732, 0.0588, 1.8302, 1.4378, 0.1168, -0.4359, -1.9687, -0.2804*]

$$N_c(a_1,b,c)=\frac{D}{qa_1}\left[p_{c00}+p_{c10}q_1+p_{c01}r_1+p_{c20}q_1^{\,2}+p_{c11}q_1r_1+p_{c02}r_1^{\,2}+p_{c30}q_1^{\,3}+p_{c21}q_1^{\,2}r_1+p_{c12}q_1r_1^{\,2}+p_{c03}r_1^{\,3}\right]$$

$$r_1=\frac{b}{a_1},\ q_1=\frac{c}{a_1}$$

$$N_c(a_2,b,c)=\frac{D}{qa_2}\left[p_{c00}+p_{c10}q_2+p_{c01}r_2+p_{c20}q_2^{\,2}+p_{c11}q_2r_2+p_{c02}r_2^{\,2}+p_{c30}q_2^{\,3}+p_{c21}q_2^{\,2}r_2+p_{c12}q_2r_2^{\,2}+p_{c03}r_2^{\,3}\right]$$

$$r_2=\frac{b}{a_2},\ q_2=\frac{c}{a_2}$$

with [p_{c00}, p_{c10}, p_{c01}, p_{c20}, p_{c11}, p_{c02}, p_{c30}, p_{c21}, p_{c12}, p_{c03}] = [*1.5851, 0.9697, 1.4346, -2.0483, 10.6524, -7.7795, 1.7185, -2.2441, -8.7282, 6.5833*]

It should be pointed out that the relation between the amplitudes of the first and second limit cycles modes is $a_1<a_2$.

Let $L(j\omega)=|L(j\omega)|e^{j\zeta(\omega)}e^{-j\rho}$, where ρ is an additional phase delay added to the loop. This phase delay corresponds to a time delay τ and $\rho= \omega\tau$. Taking into account the limit cycle condition (26) [4] – [9] we get

$$L(j\omega_k).N_{a_k}(a_k,b,c,\varphi_{s,k}) = -1,\ \ k = 1,2 \tag{30}$$

where ω_1 and ω_2 are the frequencies of the first and second limit cycle modes and $\rho_1=\omega_1\tau$ and $\rho_2= \omega_2\tau$.

Taking into account (30) the phase conditions for limit cycle oscillation on the first or second limit cycle in the presence of two sinusoidal input signals are given by the solution of the equations

$$\left|L(j\omega_k)\right|.N_{a_k}(a_k,b,c)\left[\sin(\zeta(\omega_k)-\rho_k\cos(\varphi_{s,k})-\cos(\zeta(\omega_k)-\rho_k)\sin(\varphi_{s,k})\right]=0,\quad k=1,2 \tag{31}$$

where $N_{a_k}(a_k,b,c),\ \ k=1,2$ are given by the approximate formulas above.

The magnitude conditions for limit cycle oscillation on the first or second limit cycle are given by the solution of the following equations

$$\left|L(j\omega_k)\right|.N_{a_k}(a_k,b,c)\left[\cos(\zeta(\omega_k)-\rho_k)\cos(\varphi_{s,k})+\sin(\zeta(\omega_k)-\rho_k)\sin(\varphi_{s,k})\right]=-1,\quad k=1,2 \tag{32}$$

To compare the results for the modeling of the limit cycle behavior with more than one input signal with the results obtained in [4], where the dual-input DFs model has been utilized for the analysis of the limit cycle behavior of SDMs and to check the above results in which the model with three-sinusoid-input DFs make sense we consider the following example: a simple first order SDM with sampling frequency f_s=1e^9 *Hz* with linear part described by first order frequency transfer characteristic

$$L(j\omega) = \frac{1e^7}{(j\omega+1e^5)} \tag{33}$$

and additionally added time delay in the loop is chosen to be $\tau=t_s/4$ ($\rho=\omega\ \tau$).

In this case there are two active limit cycle modes illustrated in Fig. 6 by the intersection between the line at -180 *deg*. and the horizontal lines originating from the discrete frequency points $L(j\omega_s/2)$ and $L(j\omega_s/4)$ of the locus $L(j\omega)$ with lengths equal to $2\pi/2$ and $2\pi/4$.

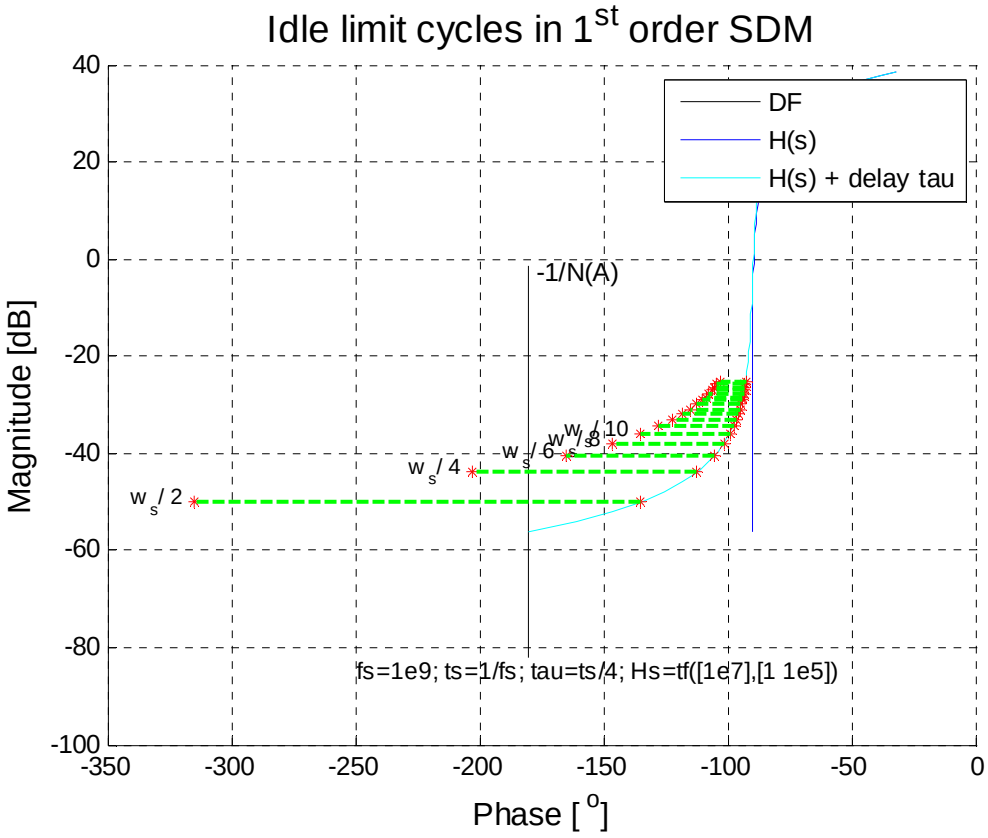


Fig. 6. Graphical description of the limit cycle model of the SDM operation considered in the example

When the input signal of the SDM of Fig. 3 is $x(t)=0.01\sin(2\pi 10^6 t)$ the limit cycle model, developed in [4], based on sampled two-input-sinusoid DFs is used for evaluation of the switching between both limit cycles modes. The results are confirmed by simulations. The Simulink model for this case is given in Fig. 7.

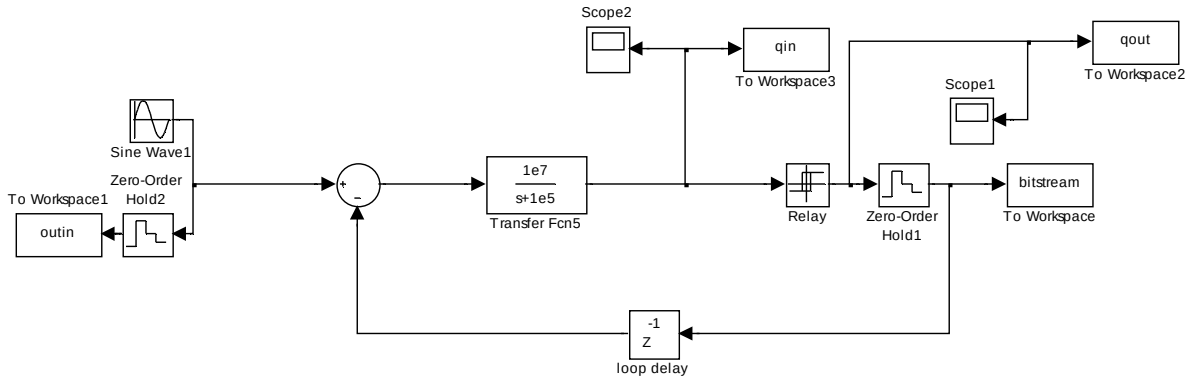


Fig. 7. Simulink model of the SDM with one sin wave input signal considered in the example.

The time domain behavior of the quantizer input is given in Fig 8. It is easy to observe that the SDM works on the first and second limit cycle modes. The switching conditions between both modes are predicted with the limit cycle model developed in [4].

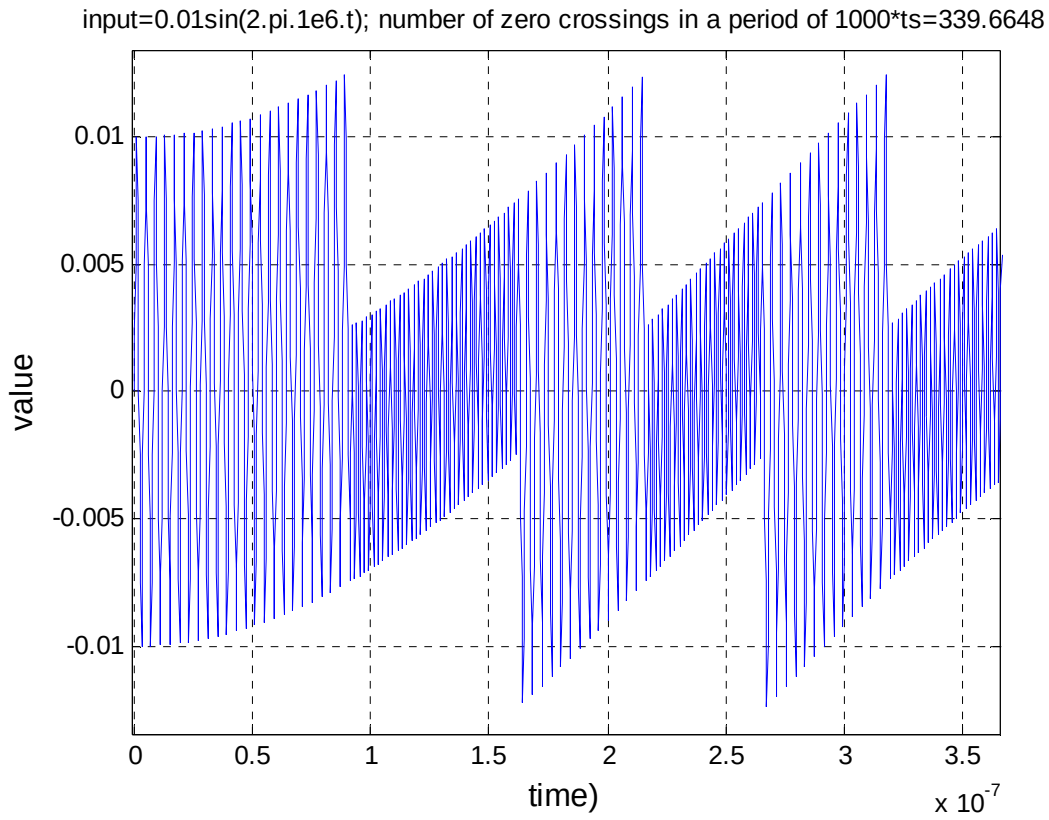


Fig. 8. Quantizer input waveform of the SDM considered in the example

The number of zero crossings in a period of $1000T_s=10^{-6}$ *sec*. is a measure of the time that the SDM works on both limit cycles mode. If the SDM operates only on the first limit cycle mode the number of zero crossings should be 500, whereas if it works on the second limit cycle mode the number of zero crossings should be 250. In the case considered the number of zero crossings is 339.66, i.e. the SDM operates on both limit cycle modes. From the power spectrum of the output bitstream one can observe the peaks at the frequency of the input signal $1e^6$ *Hz* and at the first and second limit cycles modes at $f_s/2$=$0.5e^9$ *Hz* and $f_s/4$=$0.25e^9$ *Hz*. SNR in this case is 18.57 *dB*.

When the input signal of the SDM of Fig. 6 consists of two sinusoids i.e. $x(t)=0.01\sin(2\pi 10^6 t)+0.3\sin(2\pi 10^8 t)$ the developed in [4] limit cycle model using sampled two-input-sinusoid DFs could not predict the switching between both limit cycles modes.

For the analysis of this case, the new sampled three-sinusoid-input DF approach elaborated in this report has been used. The switching conditions between two limit cycle modes are calculated through (32). The Simulink model for this case is given in Fig. 9. The model is similar to this from Fig. 7, but in this case an additional sine wave input signal $0.3\sin(2\pi 10^8 t)$ is used.

The time domain behavior of the quantizer input is given in Fig. 10. It is easy to observe that the SDM also works on the first and second limit cycle modes. The switching conditions between both modes are predicted with the limit cycle model developed in this article. They are different from the case with one sinusoidal signal, but we have to stress again that the SDM operates only on the first and second limit

cycle modes like in the previous case. The switching conditions however are different and can be predicted only with the developed theory based on sampled three-sinusoidal-input DFs. In the case considered the number of zero crossings is 340.00, and this confirms that the SDM operates only on the first and second limit cycle modes. It should be stressed that for this case we used the approximate three-sinusoid-input DFs $N_a(a_1,b,c)$ and $N_a(a_2,b,c)$ given above, for which we obtained the best polynomial approximation with a mean squared error of the approximation 1.0570e-004.

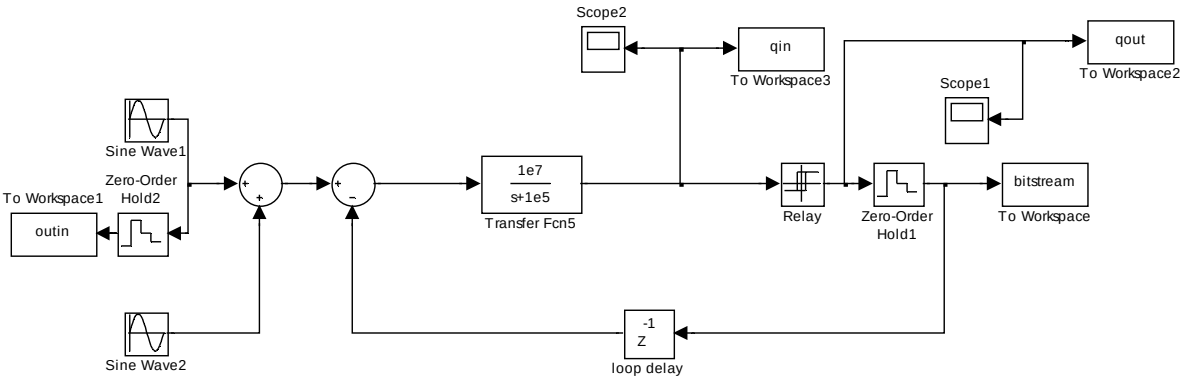


Fig. 9. Simulink model of the SDM considered in the example with two sine waves input signals

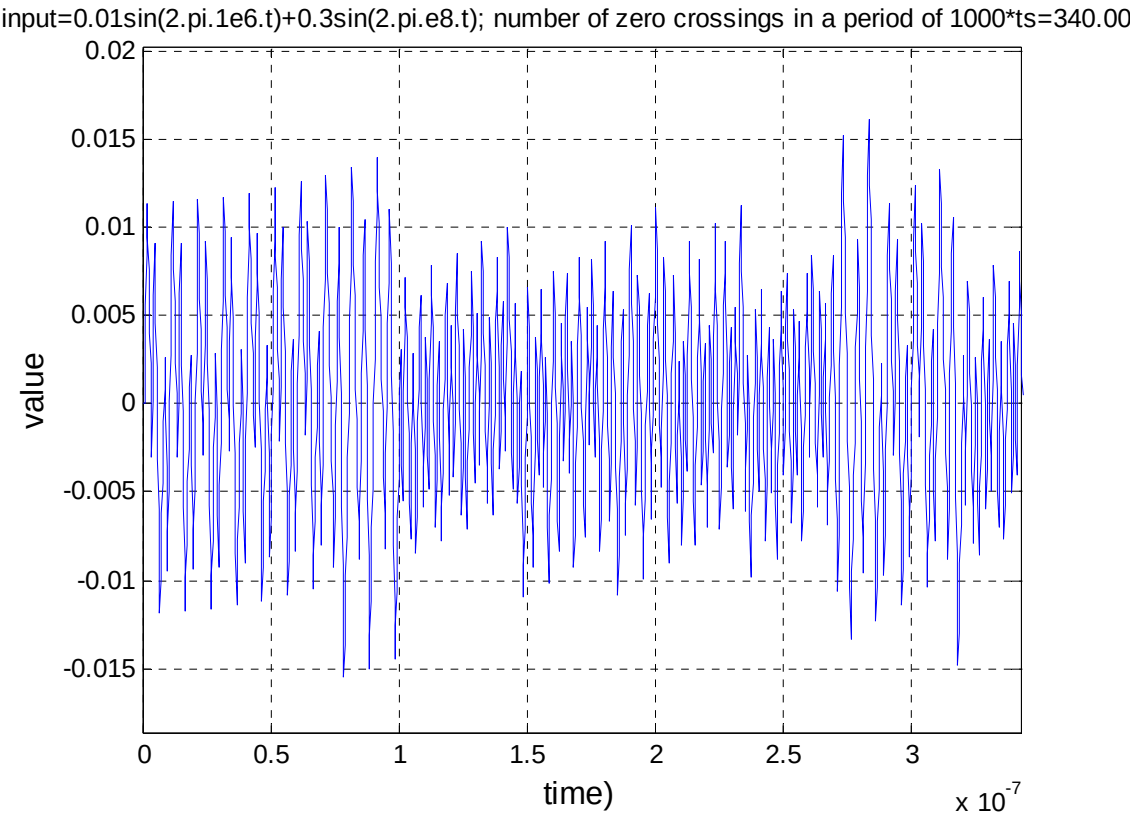


Fig. 10. Quantizer input waveform of the SDM considered in the example with two sine wave input signals

From the power spectrum of the output bitstream one can observe clearly the peaks at the frequencies of the input signal $1e^6$ *Hz* and $1e^8$ *Hz* and at first and second limit cycles modes at $f_s/2$=0.5e9 *Hz* and $f_s/4$=0.25e9 *Hz*. The SNR in this case is 26.56 *dB* i.e. better than in the case with one sinewave input signal. The conclusion of the example with two sinewave input signals is related to the incorporation of a high frequency signal in the developed limit cycle model of the SDM. This incorporation is limited by the actual frequency difference of this signal and the limit cycle mode with the lowest frequency.

On the other hand, the insight of the limit cycle behavior of SDMs helps us to say immediately what the impact of such high frequency signal on the SDM operation would be. It is going to disturb all lower frequency limit cycles. For the example that we consider, there is only one possible low frequency limit cycle and it is disturbed as this can be observed comparing the quantizer input waveforms given in Fig. 8 and Fig. 10. In practice, this situation can be very useful. With a high frequency signal with sufficient amplitude one can actually decrease the loop operation at the low frequency limit cycle and in fact to dither it out. Thus it is actually going to improve the SDM performance.

5. CONCLUSIONS

In this article the application of the multiple-input describing functions approach for adequate modelling of the changes in the limit cycle behaviour of SDMs due to two sinusoidal input signals was discussed. For this purpose a SDM limit cycle model has been investigated by use of sampled DF. Further in the article the sampled three-sinusoid-input DFs for analysis of SDMs have been introduced. The results for the modelling of the limit cycle behaviour with two sine wave input signals have been presented. The switching conditions between the possible limit cycles modes have been given as well. These results could not be obtained based on sampled two-sinusoidal-input DF developed in [4]. Based on elaborated results the effect of the incorporation of a high frequency signal in the developed limit cycle model of the SDMs has been investigated.

It should be stressed that to the best of our knowledge there are no any published results about the analysis of limit cycles based on sampled three-sinusoidal-input DFs and only a few results about general use of three-sinusoidal-input DFs. In this sense the presented results are new and there are many things that have to be elaborated further.

ACKNOWLEDGEMENT

The work was supported by grant from the Technical University - Sofia, Research Project No: 091ni139-08/2009 and N.W.O. grant nr. 040.11.063.

REFERENCES

[1] S. R. Norsworthy, R. Schreier, G. C. Temes (Editors) 1997. *Delta-Sigma Data Converters*, IEEE Press, New York, 1997.

[2] J. C. Candy, G. C. Temes, (Editors) 1992. *Oversampling Delta-Sigma Data Converters*, IEEE Press, New York, 1992.

[3] R. M. Gray, "Oversampled sigma-delta modulation*", IEEE Trans. Commun*., 35, pp.481-489, 1987.

[4] S.F. Ouzounov, *Multi-limit-cycle operation of ΣΔ modulators and efficient decimation, theory and application*, Ph.D. Thesis, Technische Universiteit Eindhoven, the Netherlands, ISBN 978-90-386-1324-6, 2008.

[5] A. Gelb, W. E. Vander Velde, *Multiple-Input Describing Functions and Nonlinear System Design*, McGraw Hill, 1968.

[6] S. S. Sastry, *Nonlinear Systems: Analysis, Stability, and Control*, Springer-Verlag, 1999..

[7] H. K. Khalil. *Nonlinear Systems*, 3rd Edition. Prentice-Hall, 2002.

[8] V. Mladenov, "Prediction of Limit Cycles in nonlinear systems with ideal relay type nonlinearities by using Multiple-input Describing Functions", *Proceedings of the 13th WSEAS International Conference on SYSTEMS*, Rodos, Greece, July 22-24, pp. 39-47, 2009.

[9] V. Mladenov, "Limit cycle prediction using Multiple-input Describing Functions in nonlinear systems with ideal relay type nonlinearities", *Proceedings of the VII-th International Conference "Challenges in Higher Education and Research in the 21st Century*", June 2-5, Sozopol, pp. 82-94, 2009.

STATE-SPACE TRANSIENT ANALYSIS IN THEORETICAL ELECTRICAL ENGINEERING

Kostadin Brandisky, Kantcho Ivanov

Department Theoretical Electrical Engineering, Technical University of Sofia, 8 Kliment Ohridski St., 1000 Sofia, Bulgaria, phone: +35929653809, e-mail: kbran@tu-sofia.bg

Abstract: *The state-space analysis is not very popular in the Theoretical Electrical Engineering - it is used mostly in Control Engineering. However, it has a considerable potential for transient analysis of complex nonlinear circuits, and good pedagogical effect in teaching – the analysis of circuits and systems can be performed using matrix methods to write and solve first-order state-variable differential equations. This paper explain the basics of the method, how it is used it the Theoretical Electrical Engineering and outlines its perspectives. Two examples are solved to demonstrate the method using different numerical platforms.*

Keywords: theoretical electrical engineering, state space, transient analysis

1. INTRODUCTION

Two methods have wide popularity for transient analysis of electric circuits in the subject Theoretical Electrical Engineering:

1. The classical method for integration of differential equations
2. Laplace transforms method

The first method is suitable for simple circuits of 1st or 2nd order. For higher-order circuits it becomes cumbersome, because many dependent initial conditions and integration constants has to be determined. The second method is more suitable for higher order circuits and computer applications, but it is only applicable for linear circuits.

In general, these methods have the following drawbacks:

1. are not applicable to nonlinear circuits;
2. are not easily programmed for computer solution;
3. do not give much information about the circuit, except when full detailed solution for all variables is done.

The missing properties in these methods can be found in another method for transient analysis – *the state-space analysis*, which is mostly used in the Control Engineering for analysis of dynamical systems. It can be easily applied to nonlinear time-dependent systems and gives as a result a system of equations, that can be easily programmed on digital computer [1,2].

The state-space analysis is not very popular in the Theoretical Electrical Engineering. However, it has a big potential for transient analysis of complex nonlinear circuits and a considerable pedagogical effect in teaching - the analysis of circuits and systems can be performed using matrix methods to write and solve first-order state-

variable differential equations. The aim of this paper is to explain the basics of the method, its advantages, how it is used it the Theoretical Electrical Engineering and to outline its perspectives in the future.

2. STATE SPACE ANALYSIS BASICS

2.1. General notions

The essence of the state-space analysis is that it is a technique for determining the state of a system. If the energy stored in one of the system elements changes, the system will be in a different state. For electric circuits, the amount of energy stored at any instant in each type of element is given as

$$W_R = 0 \ , \ W_L = \frac{1}{2} L i^2 , \ W_C = \frac{1}{2} C u^2 \tag{1}$$

These expressions suggest that the current in each inductor and the voltage across each capacitor in the circuit can be chosen as the set of variables to describe the state of the electric circuit.

At any instant of time, a given distribution of stored energy in the system's elements uniquely determines the location of a point in a coordinate system, in which each axis is labeled with a different state variable. As time goes on and the amount of energy stored in each storage element changes, the point moves, describing a path, or *trajectory*, in the coordinate system. This coordinate system is called a *state space* because the instantaneous location of a point in that space specifies exactly how much energy is stored in each element at that instant and thus describes the state of the system. A vector that always reaches from the origin to the point is called the *state vector*. As the point moves in the state space, the state vector changes in both magnitude and direction.

An illustration of the state space and the state vector is shown below. Let's consider the discharge of a capacitor through resistor and inductor. The independent initial conditions are:

$$i_L(0-) = i_L(0+) = 0 \text{ A} ; \ u_C(0-) = u_C(0+) = 1 \text{ V}$$

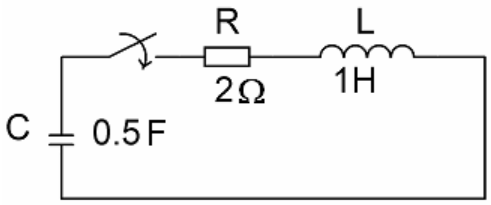


Fig. 1: Simple example for illustration of the state space

The solutions for i_L and u_C are: $u_C(t) = \sqrt{2} e^{-t} \sin(t + \frac{\pi}{4}) ; \ i_L(t) = e^{-t} \sin t$

The $u_C(t)$, $i_L(t)$ and the state space trajectory described by the state variables, are shown on Fig. 2:

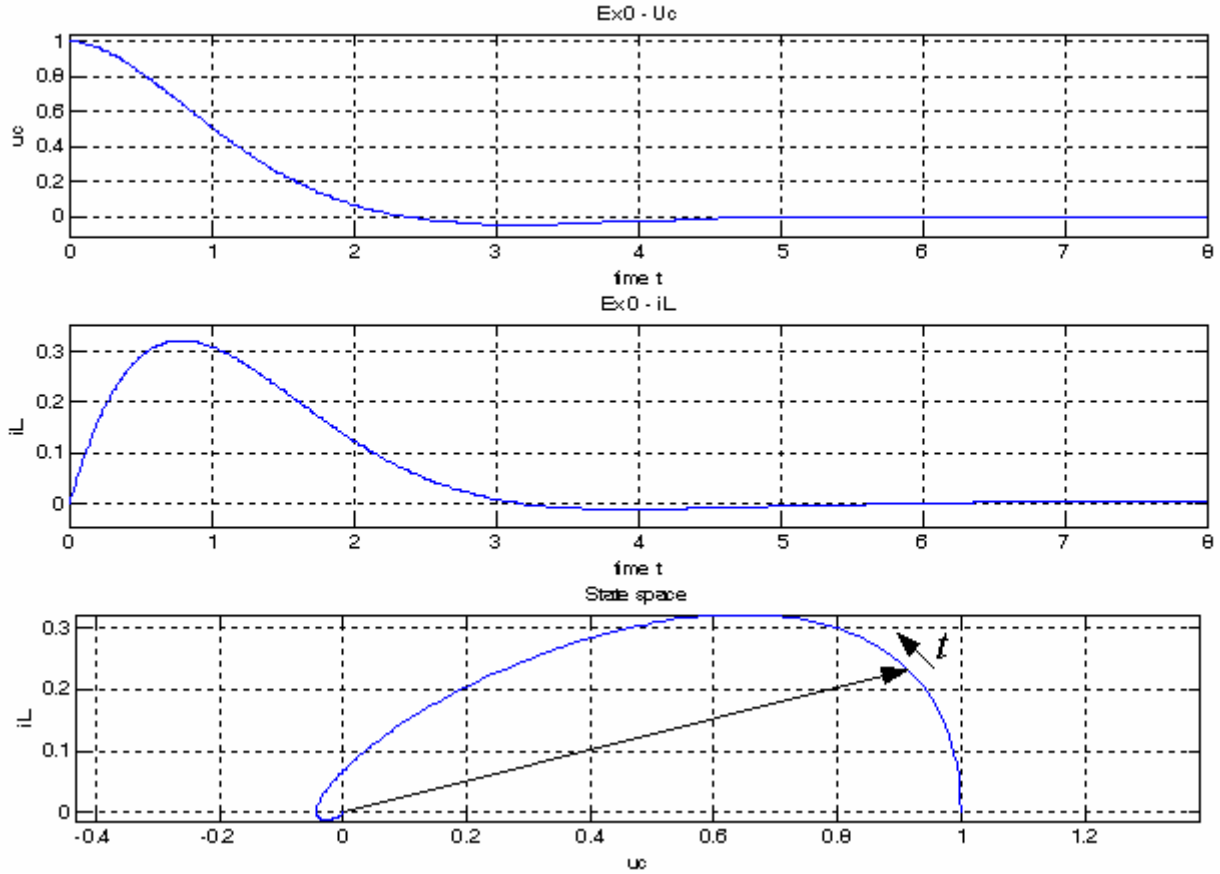


Fig. 2. $u_C(t), i_L(t)$ and the state space trajectory

In Fig. 2, the state trajectory is shown for $t = 0 \div \infty$. The state vector for a specific time moment is also shown. This trajectory has ending point in the beginning of the coordinate system. For a source-free system, the point where the time derivatives of every state variable are simultaneously zero is called *equilibrium point*. An equilibrium point is called *unstable*, if a slight displacement of the system vector away from that point results in the state vector's moving away on a new trajectory. An equilibrium point is called *stable* if the state vector stays in the neighborhood of the equilibrium point after being displaced slightly. In any source-free linear system, the origin is the only equilibrium point. Nonlinear systems can have more than one equilibrium point, the location and description of which are of interest for the researcher.

2.2. State-space analysis algorithm

The main steps in the state-space analysis are [1]:

1. The currents in the inductors and the voltages across the capacitors in the circuit are chosen as state variables. In more general case, the state variables can be the magnetic fluxes and the capacitor charges;
2. All remaining variables - the voltages across the inductors and the currents in the capacitors, are expressed using the state variables;

3. Applying the Kirchhoff's Current Law (KCL) and the Kirchhoff's Voltage Law (KVL), a non-homogeneous system of differential equations is written, which describes the transients in the circuit.
4. The obtained system of equations is reduced to *normal form*, i.e., for each state variable, u_C or i_L, a first-order differential equation is obtained solved in respect to the first time derivative of that state variable.
5. The system in normal form is integrated at specified initial conditions. The initial conditions are determined from the steady state of the circuit before the commutation.
6. After the integration, the required output variables are expressed using the state variables and the available sources.

Some recommendations exist for the choice of normal tree for the circuit [3]:

1. All capacitors and voltage sources must be included in the normal tree. If they are not sufficient, resistive branches are added.
2. All inductors and current sources are taken as links. If they are not sufficient, resistive branches are added.

These rules facilitate the formation of the state equations.

Example of forming the state-space system of equations.

For the circuit on Fig. 3 find the inductor current $i_L(t)$ and the capacitor voltage $u_C(t)$ as functions of time. Find also $i_1(t)$ and $u_{R2}(t)$.

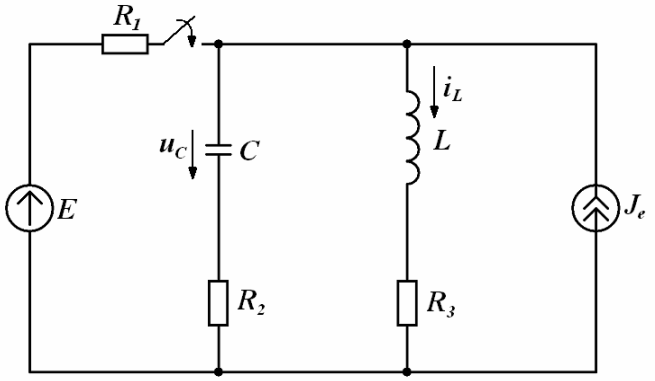


Fig. 3: Example circuit 1

The independent initial conditions, $u_C(0)$ and $i_L(0)$, are found from the circuit before commutation and considered as known.

For the circuit after the commutation (the switch is closed) – Fig. 4, the following equations can be written using KCL and KVL:

$$\left|\begin{array}{c} -i_1 + i_C + i_L - J_e = 0 \\ R_1 i_1 + u_C + R_2 i_C = E \\ -u_C - R_2 i_C + L\dfrac{di_L}{dt} + R_3 i_L = 0 \end{array}\right. \qquad (2)$$

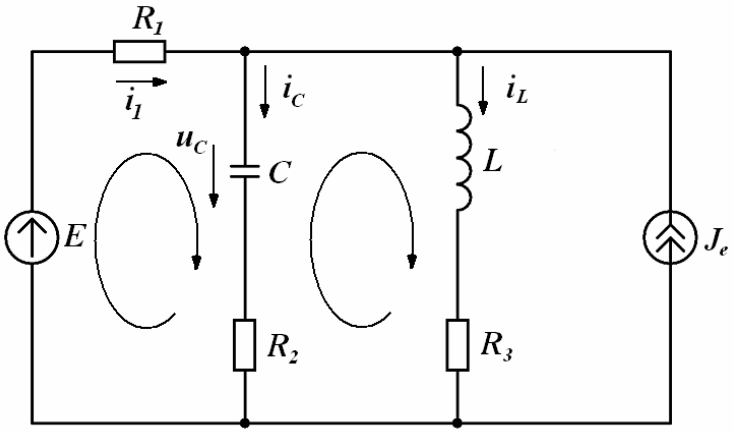


Fig. 4: Example circuit 1 after the commutation

We choose as state variables the inductor current i_L and the voltage across the capacitor u_C. The aim is to transform this system of equations in state-space form - as two first-order differential equations solved with respect to the first time derivatives of i_L and u_C, and the right hand sides being functions only to i_L and u_C.

As the capacitive current is $i_C = C\frac{du_C}{dt}$, after the substitution of i_C the system is:

$$\left|\begin{array}{r} -i_1 + C\frac{du_C}{dt} + i_L - J_e = 0 \\ R_1 i_1 + u_C + R_2 C\frac{du_C}{dt} = E \\ -u_C - R_2 C\frac{du_C}{dt} + L\frac{di_L}{dt} + R_3 i_L = 0 \end{array}\right. \qquad (3)$$

The current i_1 is expressed from the 1st equation

$$i_1 = C\frac{du_C}{dt} + i_L - J_e$$

and substituted into the 2nd equation:

$$R_1 C\frac{du_C}{dt} + R_1 i_L - R_1 J_e + u_C + R_2 C\frac{du_C}{dt} = E$$

After solving for du_C / dt, the first state-space equation will be:

$$\frac{du_C}{dt} = -\frac{1}{(R_1+R_2)C}u_C - \frac{R_1}{(R_1+R_2)C}i_L + \frac{1}{(R_1+R_2)C}E + \frac{R_1}{(R_1+R_2)C}J_e$$

Then, du_C / dt is substituted into the 3rd equation of the system (3) and it is solved for di_L / dt :

$$\frac{di_L}{dt} = \frac{R_1}{(R_1+R_2)L}u_C - \left(R_3 + \frac{R_1R_2}{(R_1+R_2)}\right)\frac{1}{L}i_L + \frac{R_2}{(R_1+R_2)L}E + \frac{R_1R_2}{(R_1+R_2)L}J_e$$

The final state-space system of equations will be:

$$\left|\begin{array}{l} \dfrac{du_C}{dt} = -\dfrac{1}{(R_1+R_2)C}u_C - \dfrac{R_1}{(R_1+R_2)C}i_L + \dfrac{1}{(R_1+R_2)C}E + \dfrac{R_1}{(R_1+R_2)C}J_e \\ \dfrac{di_L}{dt} = \dfrac{R_1}{(R_1+R_2)L}u_C - \left(R_3 + \dfrac{R_1R_2}{(R_1+R_2)}\right)\dfrac{1}{L}i_L + \dfrac{R_2}{(R_1+R_2)L}E + \dfrac{R_1R_2}{(R_1+R_2)L}J_e \end{array}\right. \quad (4)$$

Introducing the vector of the state variables $\boldsymbol{x} = \begin{bmatrix} u_C \\ i_L \end{bmatrix}$, and the vector of external sources $\boldsymbol{f} = \begin{bmatrix} E \\ J_e \end{bmatrix}$ (also called *forcing vector*), the matrix form of the first state-space equation is obtained:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{A}\,\boldsymbol{x} + \boldsymbol{B}\,\boldsymbol{f}(t)\ , \quad (5)$$

where the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are:

$$\mathbf{A} = \begin{bmatrix} -\dfrac{1}{(R_1+R_2)C} & -\dfrac{R_1}{(R_1+R_2)C} \\ \dfrac{R_1}{(R_1+R_2)L} & -\left(R_3 + \dfrac{R_1R_2}{(R_1+R_2)}\right)\dfrac{1}{L} \end{bmatrix};\ \boldsymbol{B} = \begin{bmatrix} \dfrac{1}{(R_1+R_2)C} & \dfrac{R_1}{(R_1+R_2)C} \\ \dfrac{R_2}{(R_1+R_2)L} & \dfrac{R_1R_2}{(R_1+R_2)L} \end{bmatrix} \quad (6)$$

The matrix $\boldsymbol{A}$ is always a square matrix. It is called *characteristic* or *system matrix*. When the lengths of the vectors $\boldsymbol{x}$ and $\boldsymbol{f}(t)$ are not the same, the matrix $\boldsymbol{B}$ is rectangular.

The second equation from the state-space system of equations has the form:

$$\boldsymbol{y} = \boldsymbol{C}\,\boldsymbol{x} + \boldsymbol{D}\,\boldsymbol{f}(t) \quad (7)$$

It is used to obtain the remaining variables in the circuit, except the space variable.

For the example above, the following equations can be written using the Kirchhoff's Voltage Law, in which the additional variables $i_1(t)$ and $u_{R2}(t)$ (not state variables) are present:

$$R_1 i_1 + u_C + R_2 C \frac{du_C}{dt} = E \tag{8}$$

$$u_{R2} = R_2 C \frac{du_C}{dt} \tag{9}$$

Substituting the already found time derivative du_C / dt, the following expressions are obtained for i_1 and u_{R2}:

$$i_1 = -\frac{1}{R_1 + R_2} u_C + \frac{R_2}{R_1 + R_2} i_L + \frac{1}{R_1 + R_2} E - \frac{R_2}{R_1 + R_2} J_e \tag{10}$$

$$u_{R2} = -\frac{R_2}{R_1 + R_2} u_C - \frac{R_1 R_2}{R_1 + R_2} i_L + \frac{R_2}{R_1 + R_2} E + \frac{R_1 R_2}{R_1 + R_2} J_e \tag{11}$$

The matrix $\boldsymbol{C}$ will contain the coefficients before the state variables u_C and i_L and matrix $\boldsymbol{D}$ will contain the coefficients before the sources E and J_e:

$$\boldsymbol{C} = \begin{bmatrix} -\frac{1}{R_1 + R_2} & \frac{R_2}{R_1 + R_2} \\ -\frac{R_2}{R_1 + R_2} & -\frac{R_1 R_2}{R_1 + R_2} \end{bmatrix}; \quad \boldsymbol{D} = \begin{bmatrix} \frac{1}{R_1 + R_2} & -\frac{R_2}{R_1 + R_2} \\ \frac{R_2}{R_1 + R_2} & \frac{R_1 R_2}{R_1 + R_2} \end{bmatrix} \tag{12}$$

Taking into account Eqs. (5) and (7), it can be determined, that the transients in a linear lumped-element circuit are described in the state space transient analysis by the following system matrix equations:

$$\begin{vmatrix} \frac{d\boldsymbol{x}}{dt} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{f}(t) \\ \boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{f}(t) \end{vmatrix} \tag{13}$$

The application of the state space method is connected to the resolution of several basic problems:

1) Finding the number of the state variables n_n.

The quantity n_n is the same as the circuit order n and is less or equal of the total number n_Σ of capacitors and inductors. Similar inequality $(n_n < n_\Sigma)$ arises when inductor cuts or capacitor loops are present.

2) How to solve the matrix Equation 5, representing the state space equations

Several methods exist, and many numerical packages can be used for this aim. These are covered in Chapter 2.3 and 3.

3) How to form the state equations for more complex circuits

There is a method based on the superposition principle, suitable for software implementation. It will be covered in Chapter 2.4.

The solutions of these problems are shortly described below.

2.3. Solution of Eq. (5)

The Equation (5) has *normal form* or *Cauchy form*. It has to be solved at specified initial conditions :

$$\boldsymbol{x}(0+) = \boldsymbol{x_0} \tag{14}$$

There are several approaches to solve this matrix equation.

2.3.1 Analytical solution of Eq. 5:

The matrix equation (5) at initial conditions (14) has general solution in the form [1]:

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x_0} + e^{\mathbf{A}t}\int_0^t e^{-\mathbf{A}\tau}\mathbf{B}\,\mathbf{f}(\tau)d\tau \tag{15}$$

Here, the *matrix exponent* $e^{\mathbf{A}t}$ is used (also called *transition matrix*). It is defined by the absolutely convergent for every t matrix series:

$$e^{\mathbf{A}t} = \mathbf{1} + \mathbf{A}t + \frac{\mathbf{A}^2t^2}{2!} + \frac{\mathbf{A}^3t^3}{3!} + ... = \sum_{k=0}^{\infty}\frac{\mathbf{A}^k t^k}{k!}, \tag{16}$$

In general, the transition matrix is a square matrix. A typical element *a(i,j)* of the transition matrix is the response of the *i*-th state variable due to an initial condition on only the *j*-th state variable.

An example program that integrates the solution (15) can be found in reference [2]. It uses summation of infinite series and numerical integration.

After the state variables are found, the output variables can be obtained using the matrix equation (7).

2.3.2 Solution using numerical integration

The numerical solution of Eq. 5 can be performed using any mathematical package (Matlab, Octave, Mathematica, Derive, Maple) that has functions for numerical integration of ODE. For example, Matlab has 7 functions for initial value problems for ordinary differential equations (ODEs), of different order and accuracy, for general or stiff differential equations (e.g., *ODE45, ODE23, ODE23S*). The application of *ODE45* will be shown later in Example 1.

2.3.3 Solution using symbolic derivation

The symbolic solution of Eq. 5 can be performed using mathematical packages that support symbolic computations (e.g., Matlab, Maple, Mathematica). The big advantage is the possibility to find analytical solution of Eq. 5. For example, Matlab has the function *dsolve* for symbolic solution of ordinary differential equations, which can be used to find analytical solutions of the state space equations. The use of *dsolve* will be shown later in the solution of Example 1.

2.4. Forming the state space equations for complex circuits

A basic question at the application of the state space analysis is the forming of the system of state space equations (5):

The forming of this system can be achieved by application of the superposition principle [1]. The method will be shown for the circuit from Example 1 (Fig. 3). The circuit after the commutation is considered (Fig. 4). The Equation (5) will be formed initially.

The algorithm consists of the following four steps:

1. It is accepted, that i_L and u_C are given;
2. The reactive elements are substituted as follows (Fig. 5):
   - L – with ideal current source having current value i_L and the same direction as i_L;
   - C – with ideal voltage source having voltage value u_C and the opposite direction to u_C.

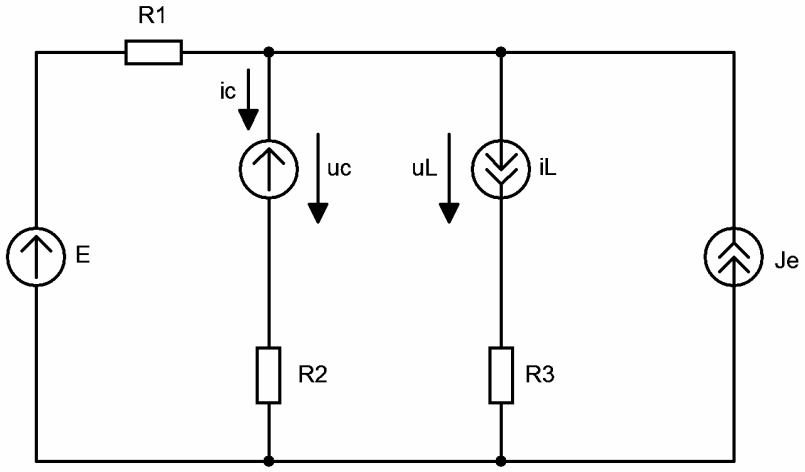
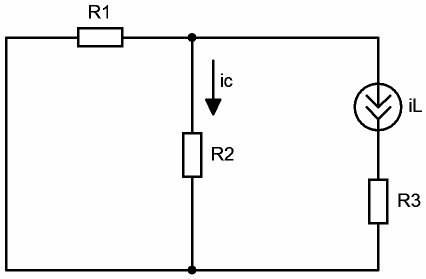


Fig. 5: The circuit after the substitution of u_C and i_L with sources

3. The current i_C and the voltage u_L are found using the superposition principle, leaving only a single source to be active at a time.

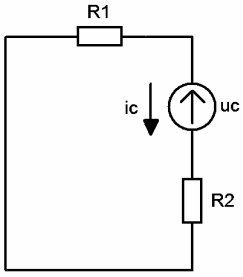
The application of the superposition principle follows the next steps:

Find i_C with only one source to be active at a time:



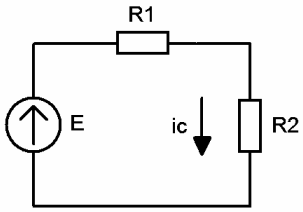
$$i_C = -\frac{R_1}{R_1 + R_2} i_L$$

Fig. 6a: Only i_L acting



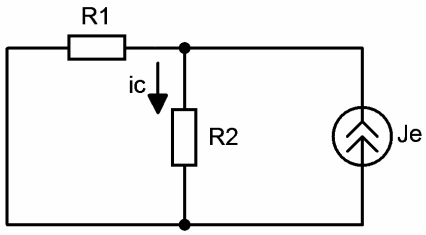
$$i_C = -\frac{u_C}{R_1 + R_2}$$

Fig. 6b: Only u_C acting



$$i_C = \frac{e}{R_1 + R_2}$$

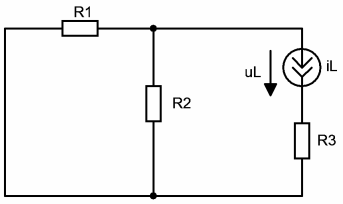
ig. 6c: Only e acting



$$i_C = j_e \frac{R_1}{R_1 + R_2}$$

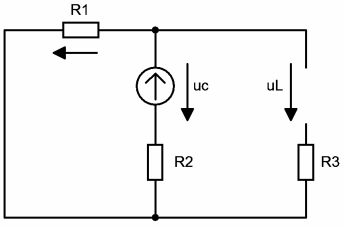
Fig. 6d: Only j_e acting

Find u_L with only one source to be active at a time:



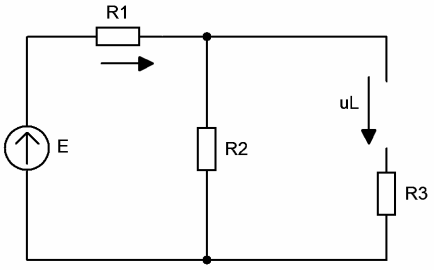
$$u_L = -(R_3 + \frac{R_1 R_2}{R_1 + R_2}) i_L$$

Fig. 7a: Only i_L acting



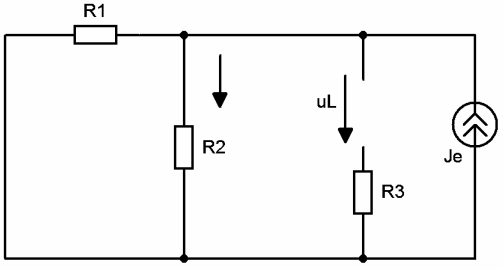
$$u_L = R_1 \frac{u_C}{R_1 + R_2}$$

Fig. 7b: Only u_C acting



$$u_L = \frac{e}{R_1 + R_2} R_2$$

Fig 7c: Only e acting



$$u_L = j_e \frac{R_1 R_2}{R_1 + R_2}$$

Fig. 7d: Only j_e acting

At the end, after superimposing all separate contributions, i_C and u_L will be:

$$i_C = -\frac{R_1}{R_1+R_2} i_L - \frac{u_C}{R_1+R_2} + \frac{e}{R_1+R_2} + j_e \frac{R_1}{R_1+R_2} \tag{17}$$

$$u_L = -(R_3 + \frac{R_1 R_2}{R_1+R_2}) i_L + R_1 \frac{u_C}{R_1+R_2} + \frac{e}{R_1+R_2} R_2 + j_e \frac{R_1 R_2}{R_1+R_2} \tag{18}$$

4. Substituting $i_C = C\frac{du_C}{dt}$ and $u_L = L\frac{di_L}{dt}$ and solving with respect to $\frac{du_C}{dt}$ and $\frac{di_L}{dt}$, we obtain the system of differential equations in normal form:

$$\begin{vmatrix} \frac{du_C}{dt} = -\frac{1}{C(R_1+R_2)} u_C - \frac{R_1}{C(R_1+R_2)} i_L + \frac{1}{C(R_1+R_2)} e + \frac{R_1}{C(R_1+R_2)} j_e \\ \frac{di_L}{dt} = \frac{R_1}{L(R_1+R_2)} u_C - \frac{1}{L}(R_3 + \frac{R_1 R_2}{R_1+R_2}) i_L + \frac{R_2}{L(R_1+R_2)} e + \frac{R_1 R_2}{L(R_1+R_2)} j_e \end{vmatrix} \tag{19}$$

This approach can easily be programmed using an algorithm and program for DC analysis of general circuits. Such a program, written in Fortran, can be found in [2] and can be used to verify the obtained by hand ***A***, ***B***, ***C*** and ***D*** matrices, especially for large, complex circuits, where the manual derivation is difficult and prone to errors.

3. RESULTS OF THE STATE SPACE TRANSIENT ANALYSIS FOR TWO EXAMPLES

3.1. Example 1

Let's find the inductor current $i_L(t)$, the capacitor voltage $u_C(t)$, the current $i_1(t)$ and the voltage $u_{R2}(t)$ for the circuit from Fig. 3. The circuit parameters are: $R_1 = 10\ \Omega$; $R_2 = 20\ \Omega$; $R_3 = 20\ \Omega$; $L = 1$ H; $C = 1000$ μF; $E = 25$ V = *const.*; $J_e = 2$ A = *const.*

First, the state space equations must be determined, as shown in the previous chapter. The already found state space equations (19) will be used here. Then, the equations can be solved using Matlab or PSpice. The advantage of using Matlab and its symbolic toolbox function *dsolve* is that it gives analytical solutions. PSpice gives numerical solution that can be represented as graphs.

3.1.1. Solution with Matlab using the function dsolve:

The function *dsolve('eq1,eq2,...', 'cond1,cond2,...', 'v')* symbolically solves the ordinary differential equations specified by *eq1, eq2,...* using *v* as the independent vari-

able and initial conditions specified by *cond1,cond2,..* The default independent variable is *t*.

The code segment with Matlab solution follows:

```
% EXAMPLE 1 - state space analysis using dsolve
clear all
% declaration of symbolic variables
syms s t x y dx dy X F Y

% input of circuit parameters
R1=10; R2=20; R3=20;
L=1; C=1e-3; E=25; J=2;

% input of independent initial conditions
iL0=J
uC0=R3*J

% create state space equations, with x=uc; y=iL
dx=-1/((R1+R2)*C)*x-R1/((R1+R2)*C)*y+1/((R1+R2)*C)*E+R1/((R1+R2)*C)*J
   dy=R1/((R1+R2)*L)*x-
(R3+R1*R2/(R1+R2))/L*y+R2/((R1+R2)*L)*E+R1*R2/((R1+R2)*L)*J

% converts symbolic equations to character strings
z1=char(dx); z2=char(dy);
s1=strcat('Dx= ',z1,' ,','Dy= ',z2)
in1=num2str(uC0); in2=num2str(iL0);
s2=strcat('x(0)=',in1,' ,','y(0)=',in2)

% call dsolve
[x,y]=dsolve(s1,s2);
x=vpa(x,4); y=vpa(y,4);
x=expand(x); y=expand(y);
x=vpa(x,4)
y=vpa(y,4)

% create output matrices C and D
C=[-1/(R1+R2) R2/(R1+R2); -R2/(R1+R2) -R1*R2/(R1+R2) ]
D=[ 1/(R1+R2) -R2/(R1+R2); R2/(R1+R2) R1*R2/(R1+R2) ]

X=[x; y]
F=[E; J]
% compute output variables
Y=C*X+D*F;
I1 =vpa(Y(1),4)
UR2=vpa(Y(2),4)
```

The results are:

```
x=uc=30.-20.*exp(-30.*t)*sin(10.*t)+10.*exp(-30.*t)*cos(10.*t)
y=iL=1.500+.5000*exp(-30.*t)*sin(10.*t)+.5000*exp(-30.*t)*cos(10.*t)
I1 =1.000*exp(-30.*t)*sin(10.*t)-.5000
UR2 =10.00*exp(-30.*t)*sin(10.*t)-10.00*exp(-30.*t)*cos(10.*t)
```

These results coincide exactly with the results found by the classical method.

3.1.2. Solution with Matlab using numerical integration of the state equations with the function ODE45

The code segment with Matlab numerical solution follows:

```
% main program for using function k50 (Ex 1) and ode45
clear all
[t,y] = ode45(@k50,[0 0.25],[40; 2]);
subplot(3,1,1);
plot(t,y(:,1),'-')
title('Ex 1 - Uc');
xlabel('time t');
ylabel('uc');
subplot(3,1,2);
plot(t,y(:,2),'-')
title('Ex 1 - iL');
xlabel('time t');
ylabel('iL');
subplot(3,1,3);
plot(y(:,1),y(:,2),'-')
title('State space');
xlabel('uc');
ylabel('iL');
```

The function `k50`, that computes the right hand sides of the state equations, must be prepared initially in a separate file `k50.m`

```
   function dydt=k50(t,y)
dydt=[ -100/3*y(1)-1000/3*y(2)+1500; 1/3*y(1)-80/3*y(2)+30 ];
```

The graphs of $i_L(t)$ and $u_C(t)$ are shown on Fig. 8.

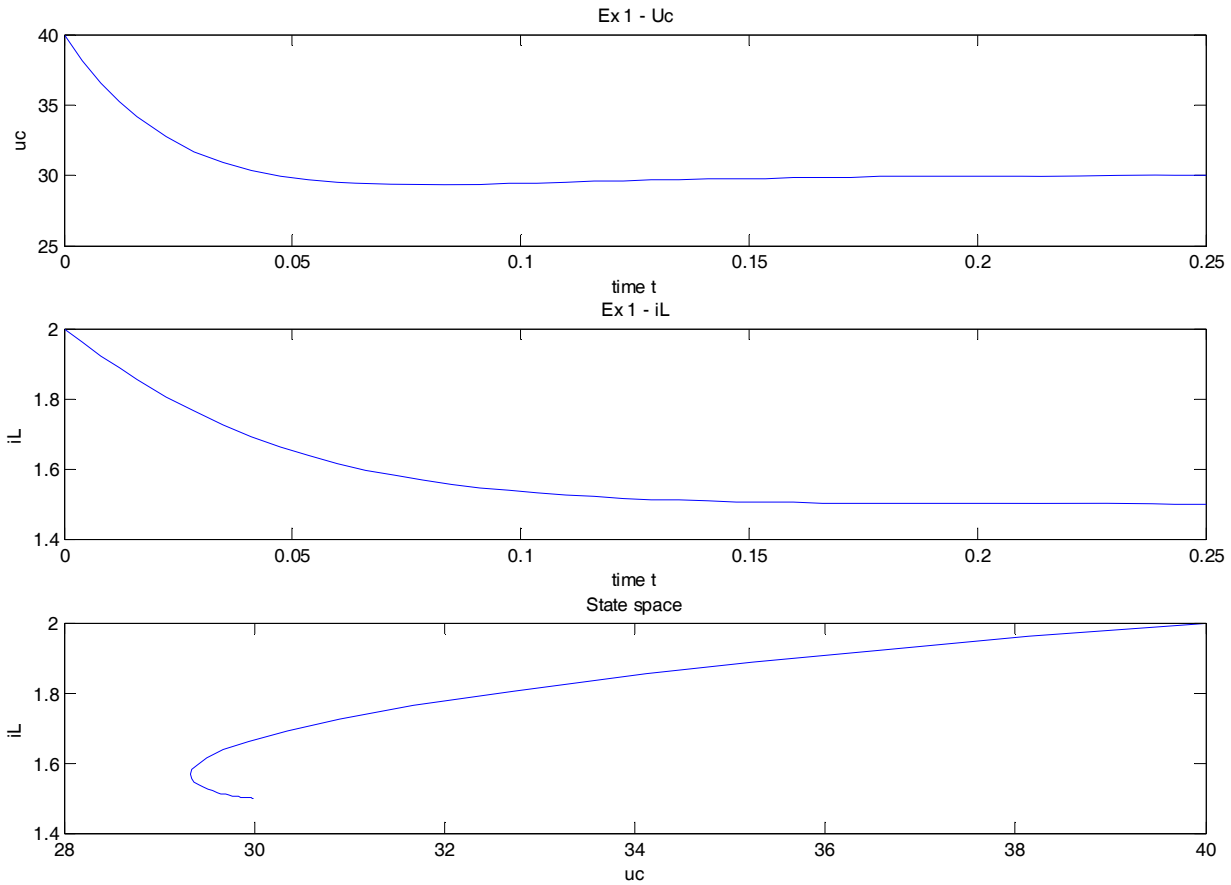


Fig. 8: Graphs of $i_L(t)$, $u_C(t)$ and the state space

ODE45 is a function that solves non-stiff differential equations, with medium order method. The call of ODE45 is

[TOUT,YOUT] = ODE45(ODEFUN,TSPAN,Y0)

with TSPAN = [T0 TFINAL]

This function integrates the system of differential equations $y' = f(t, y)$ from time T0 to TFINAL with initial conditions Y0. For a scalar T and a vector Y, ODEFUN(T,Y) must return a column vector corresponding to $f(t, y)$. Each row in the solution array YOUT corresponds to a time returned in the column vector TOUT.

How to find the final time – TFINAL / TF /.

The duration of the transient in an electric circuit is determined by the eigenvalues of the matrix $\boldsymbol{A}$. These eigenvalues coincide with the roots of characteristic equation of the circuit:

$$\det(p\boldsymbol{1} - \boldsymbol{A}) = \Delta(p) = 0 \tag{20}$$

where $\boldsymbol{1}$ is the unit matrix with the dimension of $\boldsymbol{A}$.

To cover the full duration of the transient (till the steady state is reached), the following condition must be fulfilled [5]:

$$TF > (3 \div 5)\max_k \left(\frac{1}{|p_k|} \right) \tag{21}$$

where TF is the integration interval, and p_k - the eigenvalues of the matrix $\boldsymbol{A}$.

If the eigenvalues are complex conjugate numbers, only their real part is taken.

Matlab has a special function *eig( )* for computing the eigenvalues of a matrix.

Let the matrix $\boldsymbol{A}$ is specified as:

```
A =[ -100/3 -1000/3; 1/3 -80/3]
```

The statement: *E=eig(A)* produces a column vector containing the eigenvalues. For this matrix, the eigenvalues are complex:

```
% E =
% -30.0000 +10.0000i
% -30.0000 -10.0000i
```

For this example the final time TF must be

$$TF > (3 \div 5)\left(\frac{1}{30} \right) = 0.1 \div 0.17\,\text{s}$$

In order to be more conservative, we can choose a longer final time TF=0.25 s.

In the case of stiff systems of differential equations (having big difference between the maximum and minimum eigenvalues) improved numerical integration pro-

grams are available – in Matlab such a function is ODE23S, which solves stiff differential equations with low order method.

3.1.3. Solution using PSpice and ABM (Analog Behavioral Modeling)

The Analogue Behavioral Modeling (ABM) allows the designer to model analogue circuit functions using mathematical equations, tables, and transfer functions. The designer can then simulate systems as a combination of blocks, each of which performs a specific function. PSpice has a library of about 50 ABM functions (blocks), which can be used also for state space transient analysis. Most helpful for this aim are the functions (blocks) INTEG, ABM1, ABM2 and ABM3. In the next solution of Example 1 two kinds of ABM blocks from PSpice are used.

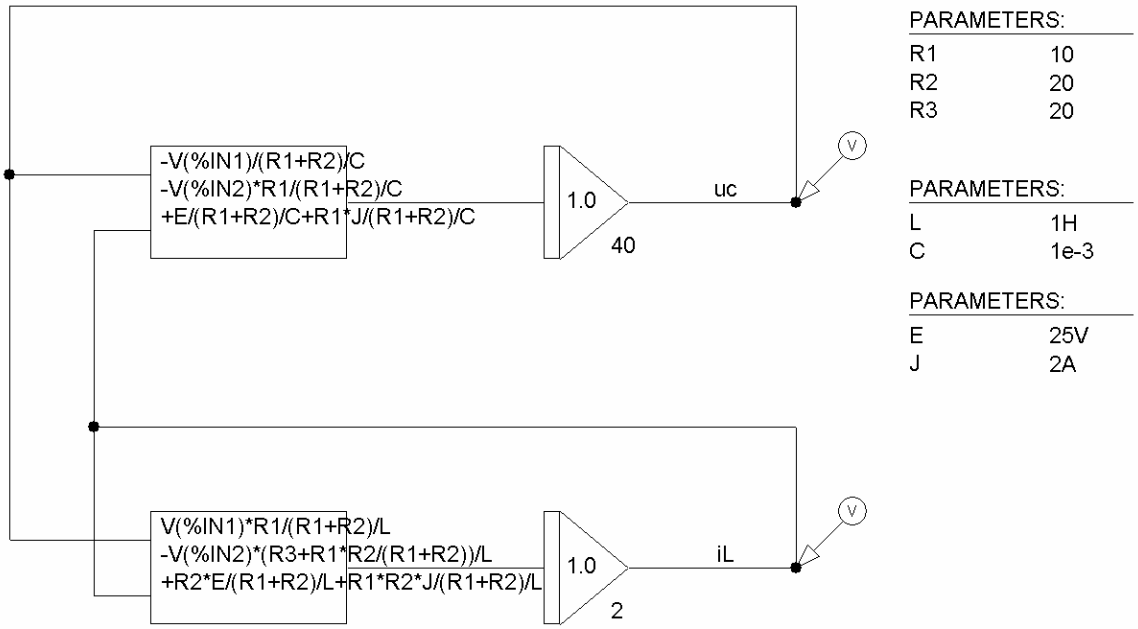


Fig. 9: Using ABM blocks in PSpice

1) INTEG - represents an integrator, one for every differential equation written in normal form. The right hand side of the differential equation in normal form enters as input there. The initial condition for the integration must be specified, as well as the gain, which is usually equal to 1.

2) ABM2 - creates the right hand side of the equation using as input signals the two state variables u_C and i_L, that are obtained at the outputs of the two integrators, and applying mathematical functions over them.

The Transient Analysis parameters are: Print Step: 3ms; Final Time: 300ms; Step ceiling: 0.3ms; Skip Initial transient solution: Y

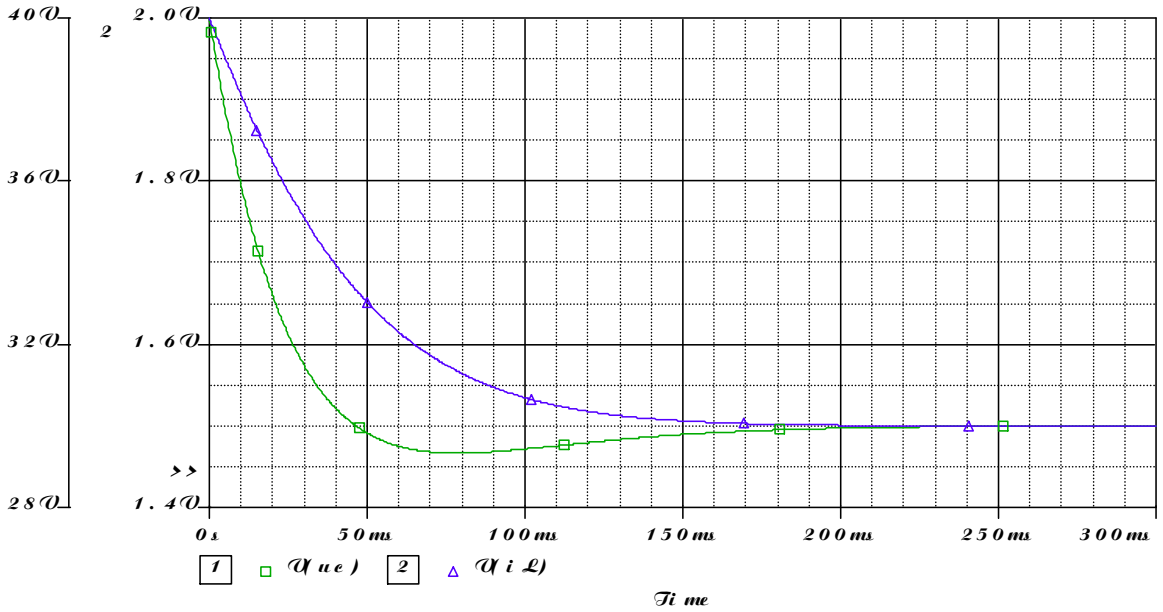


Fig. 10: Graphs of u_C and i_L vs. time

It can be seen that the graphs of u_C and i_L are the same as in Ch. 3.1.2.

3.2. Example 2

A generator circuit with lambda-diode is shown on Fig. 11. The frequency of generation is specified by the LC elements. The circuit parameters are: $R = 900\,\Omega$; $L = 10\,\text{mH}$; $C = 10\,\text{nF}$; $E = 3.2\,\text{V}$. The lambda diode has I-V characteristic with falling part (it has negative differential resistance there, which is used to compensate the losses in the LC-circuit). The aim is to model the transients in this generator and to obtain the waveforms of $u_C(t)$ and $i_L(t)$ and their period and frequency.

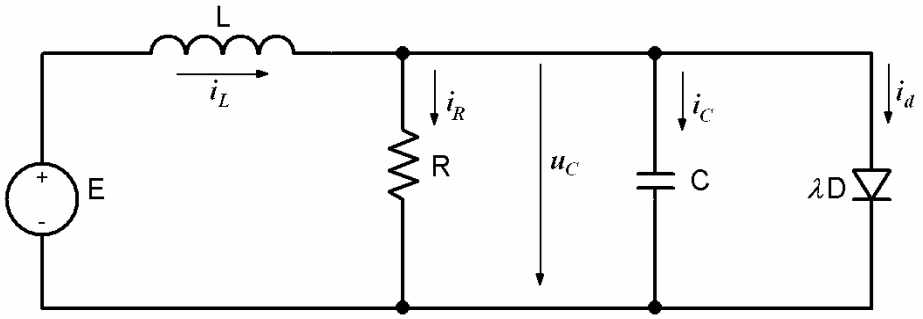


Fig. 11: The circuit for Example 2

Using the Kirchhoff's laws, the following equations that describe the transients in the circuit can be written:

$$\left|\begin{array}{ll} -i_L + i_R + i_C + i_D = 0 & (a) \\ L\dfrac{di_L}{dt} + u_C = E & (b) \\ u_C - Ri_R = 0 & (c) \\ i_d = i_d(u_C) & (d) \end{array}\right. \qquad (22)$$

As state variables, the capacitor voltage and the inductor current are chosen, as they describe the energy status of the circuit. To obtain the state space equations, the system (22) must be expressed in normal form, as a function of the space variables u_C and i_L.

From the Eq. 22c the current i_R is expressed as $i_R = \frac{u_C}{R}$. Also, it is known that $i_C = C\frac{du_C}{dt}$, so these two currents are substituted in Eq. 22a:

$$-i_L + \frac{u_C}{R} + C\frac{du_C}{dt} + i_d(u_C) = 0$$

This equation is solved in respect to du_C / dt to obtain the first state space equation:

$$\frac{du_C}{dt} = -\frac{1}{RC}u_C + \frac{1}{C}i_L - \frac{1}{C}i_d(u_C)$$

The second equation can be found by solving Eq. 22b for di_L / dt:

$$\frac{di_L}{dt} = -\frac{1}{L}u_C + \frac{1}{L}E$$

Thereby, the full state space system of equations will be:

$$\begin{cases} \frac{du_C}{dt} = -\frac{1}{RC}u_C + \frac{1}{C}i_L - \frac{1}{C}i_d(u_C) \\ \frac{di_L}{dt} = -\frac{1}{L}u_C + \frac{1}{L}E \end{cases} \tag{23}$$

The system can be solved numerically using Matlab or PSpice.

3.2.1. Solution with Matlab and ODE45

The following program section uses ODE45 from Matlab to integrate the state-space equations derived above:

```
% Example 2 - main program for using ode45 integration function
% the state space equations are in the function k51
clear all
clc
clf
global R L C E p
R=900;
% R=900 -> sin; R=3000 -> distorted
L=10e-3;
C=10e-9;
```

```
E=3.2;
% polynomial approximation of the i-v characteristic of lambda diode
% u=[V] i=[mA]
p=[ -0.0095 0.1385 -0.5925 0.2601 2.1798 0.0273];
TFINAL=0.4e-3
[t,y] = ode45(@k51,[0 TFINAL],[0.01; 0.0]);
% [0.01; 0.0] are the initial conditions for uc and iL
subplot(3,1,1);
plot(t,y(:,1),'-')
title('Ex 2 - Uc');
xlabel('time t');
ylabel('uc');
subplot(3,1,2);
plot(t,y(:,2),'-')
title('Ex 2 - iL');
xlabel('time t');
ylabel('iL');
subplot(3,1,3);
plot(y(:,1),y(:,2),'-')
title('State space');
xlabel('uc');
ylabel('iL');
```

The right hand side of the state space equations must be programmed in a separate file `k51.m:`

```
function dydt=k51(t,y)
% uc=y(1) ; iL=y(2)
global R L C E p
id=polyval(p,y(1))*1e-3;
dydt=[ -y(1)/(R*C)+y(2)/C-id/C ; -y(1)/L+E/L ];
```

The obtained graphs for $u_C(t)$ and $i_L(t)$ are shown on Figs. 12 and 13, for two cases: $R = 3000\,\Omega$ - distorted sinusoid, and $R = 900\,\Omega$ - nearly perfect sinusoid.

For $R < 800\,\Omega$ and $R > 3400\,\Omega$ generations do not exist. For $R = 900\,\Omega$, the period, found from Fig. 13, is $62.8604\,\mu s$ and the frequency of generation is $f = 15905\,\text{Hz}$. The theoretical frequency of the generated oscillations can be found using the analytical formula for resonant frequency of *LC*-circuit:

$$f = \frac{1}{2\pi\sqrt{LC}} = 15915\,\text{Hz} \tag{24}$$

3.2.2. Solution with PSpice and ABM

The schematics in Fig. 14 uses ABM blocks from PSpice and follows the state space system in Eq. 23. Several ABM blocs are used: two integrators INTEG for calculating the state variables, the blocks ABM2, ABM3 and ABM1. In ABM1, the I-V characteristic of the lambda-diode is specified, approximated by a 5^{th} order polynomial in Horner representation, to form the right hand side of the equations. The results of the simulation are shown on Figs. 15 and 16 for two cases:

$R = 900\,\Omega$ – where pure sinusoids are generated, and

$R = 3000\,\Omega$ – where distorted sinusoids are generated.

The frequency of the generated waveforms for $R = 900\,\Omega$ is $f = 15908\,\text{Hz}$, which is very close to the frequency obtained from the Matlab solution ($f = 15905\,\text{Hz}$).

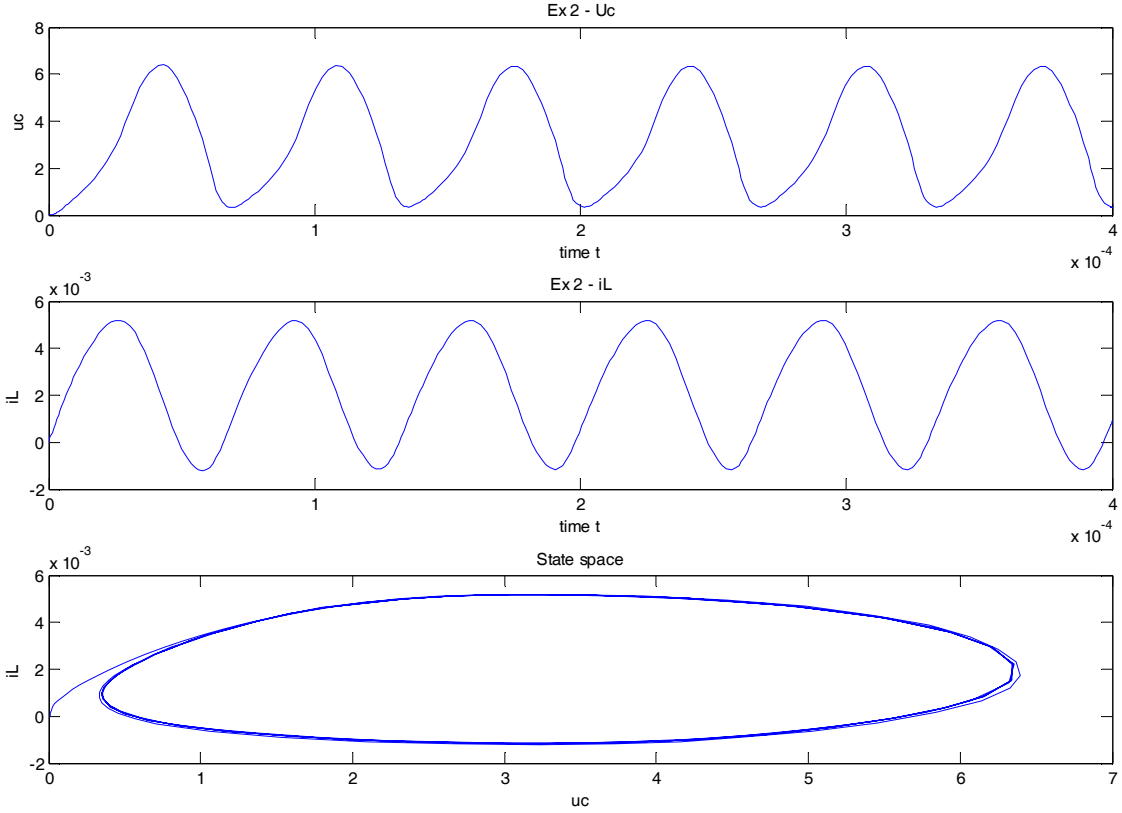


Fig. 12 : $u_C(t)$ and $i_L(t)$ for $R = 3000\,\Omega$ - distorted sinusoid

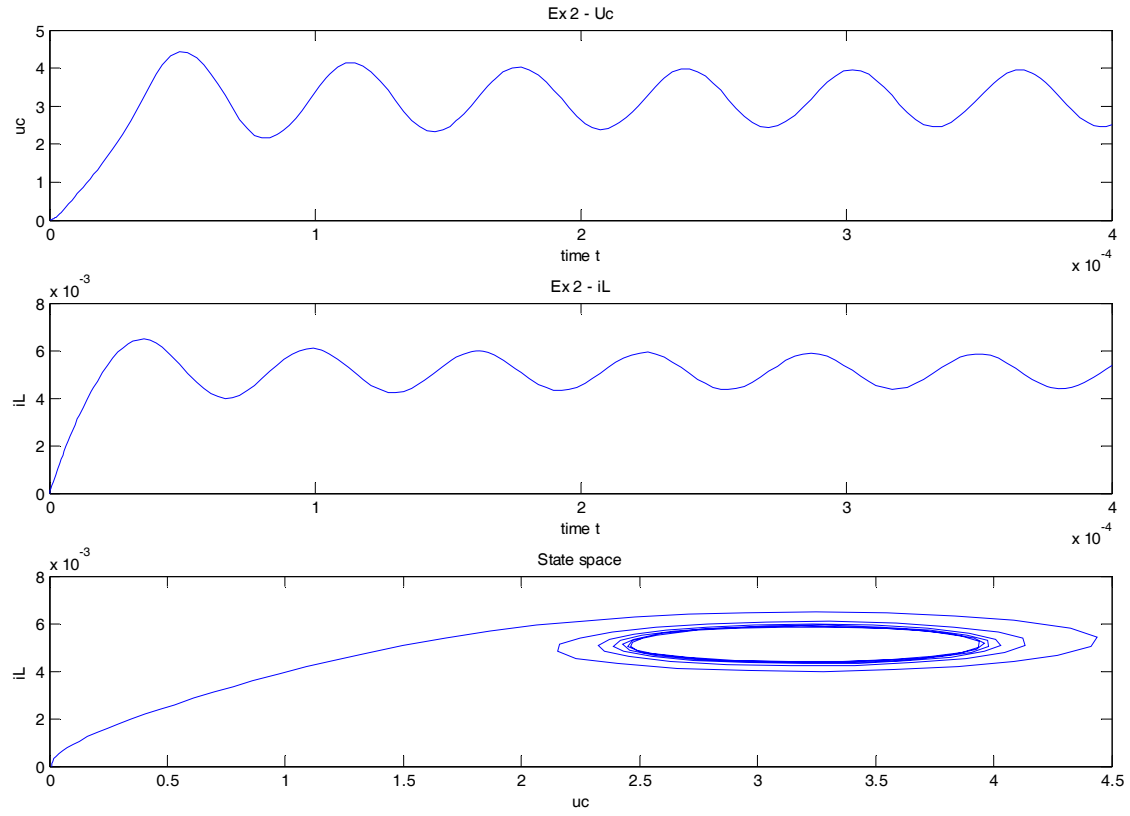


Fig. 13: $u_C(t)$ and $i_L(t)$ for $R = 900\,\Omega$ - nearly perfect sinusoid

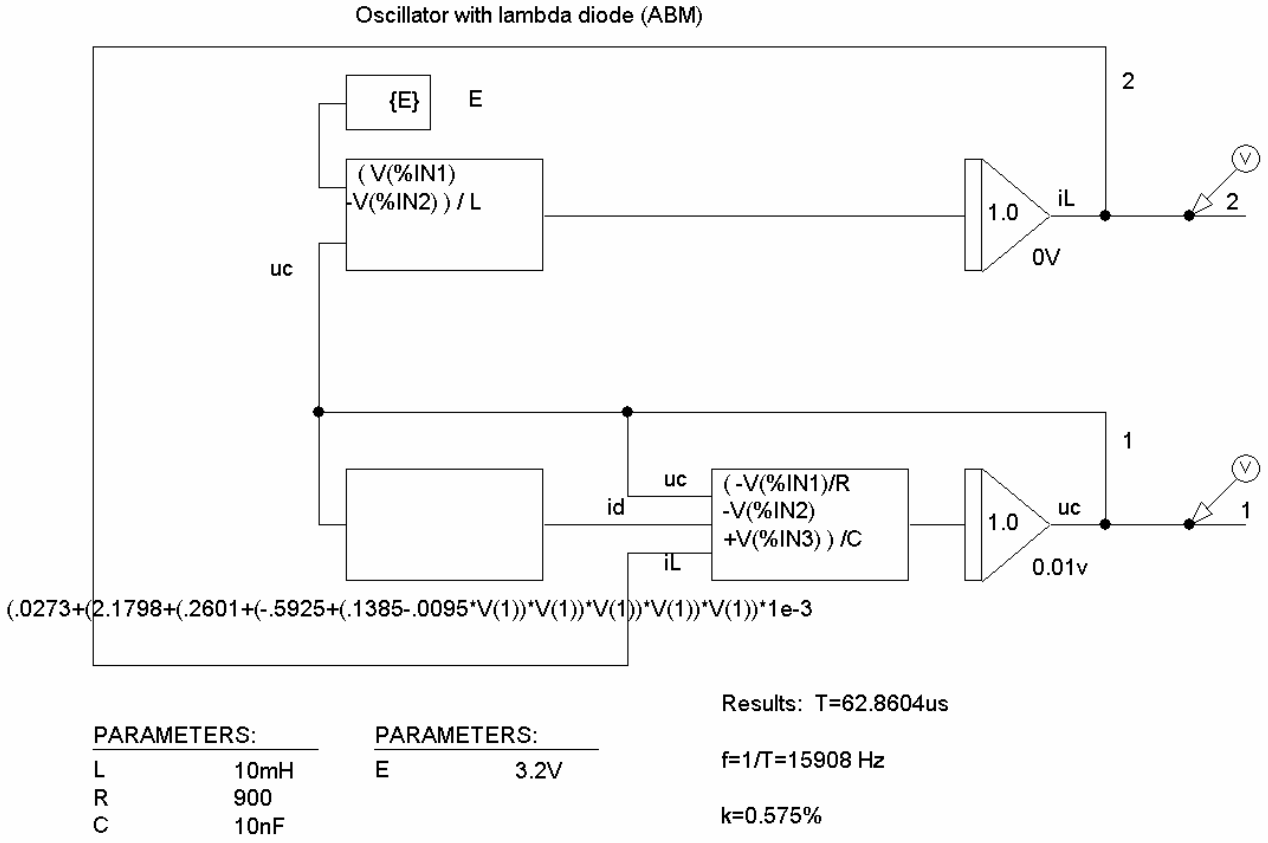


Fig. 14: PSpice schematics for Example 2, using ABM

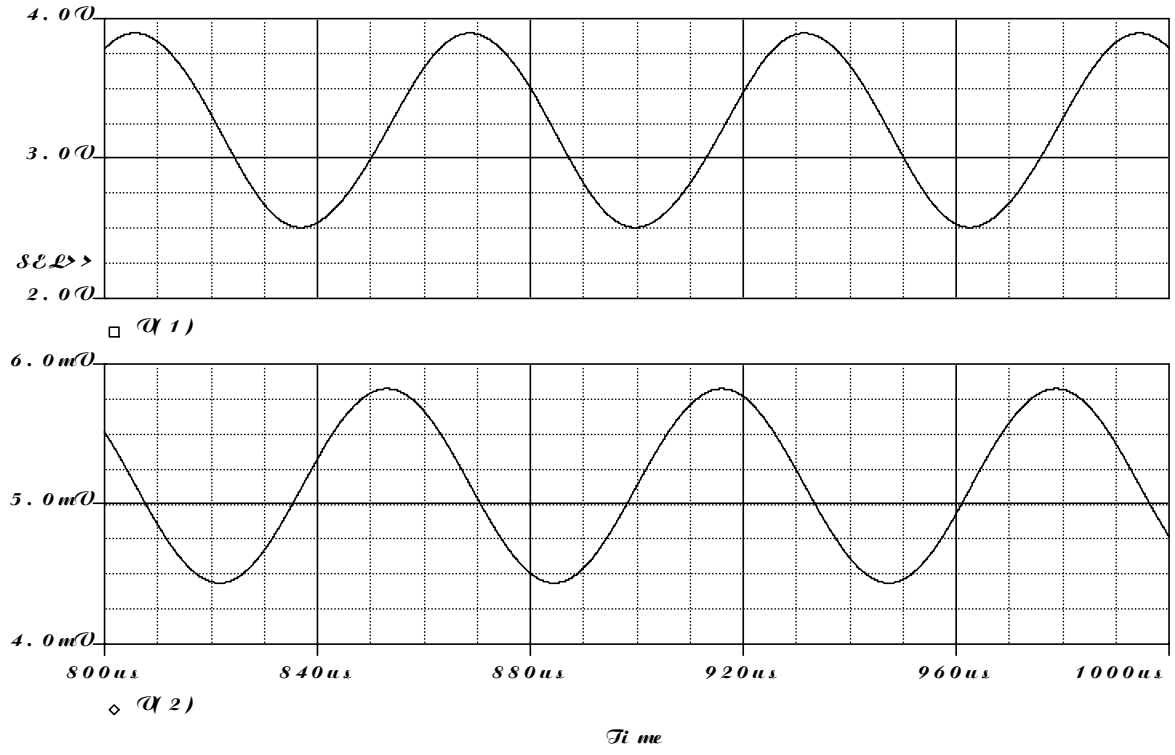


Fig. 15: Graphs of $u_C(t)$ and $i_L(t)$ at $R = 900\Omega$

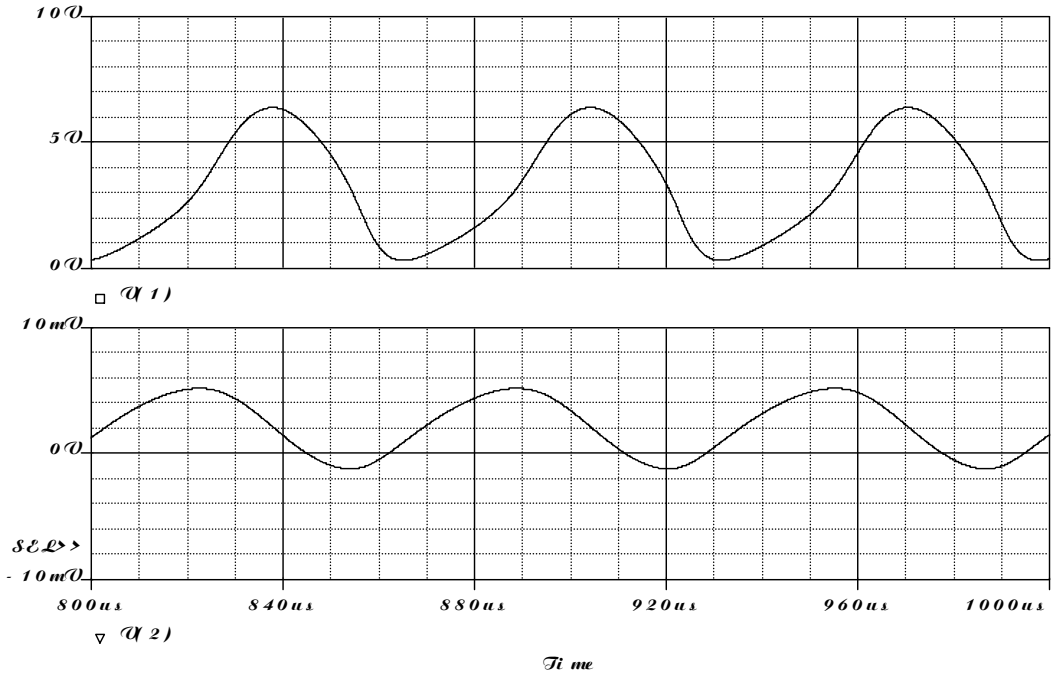


Fig. 16: Graphs of $u_C(t)$ and $i_L(t)$ at $R = 3000\,\Omega$

4. STATE SPACE ANALYSIS OF NONLINEAR CIRCUITS USING LINEARIZATION

The state space analysis method can be used also to analyze nonlinear circuits by linearization [4]. For nonlinear circuits the set of state equations for a source-free system has the general form

$$\begin{aligned} \frac{dx_1}{dt} &= f_1(x_1, x_2, \ldots, x_n) \\ \frac{dx_2}{dt} &= f_2(x_1, x_2, \ldots, x_n) \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \frac{dx_n}{dt} &= f_n(x_1, x_2, \ldots, x_n) \end{aligned} \tag{25}$$

where the x_i are the state variables, and f_i are nonlinear functions of x_i (or at least one of them is nonlinear)

If the fluctuations of the state variables are considered within a small enough region of the state space, the system can be approximated by a set of linear functions. These functions usually can describe the behavior of the state variables, but in that local region only. If the system is source-free, the equations for this linearized system are:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{A}\,\boldsymbol{x} \tag{26}$$

where

$$\boldsymbol{A}=\begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \dfrac{\partial f_1}{\partial x_3} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \cdots & \cdots & \cdots & \dfrac{\partial f_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial f_n}{\partial x_1} & \cdots & \cdots & \cdots & \dfrac{\partial f_n}{\partial x_n} \end{bmatrix} \tag{27}$$

is a matrix called the *Jacobian* matrix.

The partial derivatives in the Jacobian, when evaluated at any given point in the state space, yield linear state equations that are valid at that point.

It is particularly useful when analyzing a nonlinear system to:

(1) find the equilibrium points (where the time derivative of every state variable equals zero);

(2) linearize the system at each such equilibrium point and find the eigenvalues of the linear model that is valid in that neighborhood.

Any eigenvalues having positive real parts will produce natural responses that increase with time, the system thus being unstable in the neighborhood of that equilibrium point.

The stability of a nonlinear system in small region around an equilibrium point can be determined using the linearized in the vicinity of this point state equations. The Lyapunov's theorem is important here: If the eigenvalues of the linearized space state system are:

1) only with real negative part, then this equilibrium point is stable;

2) if at least one eigenvalue has positive real part, the equilibrium point is unstable;

3) if one or more eigenvalues have zero real part, and the remaining – have negative real part, then the stability can not be determined using this linearized system.

This analysis will be applied to the generator circuit with lambda-diode (Example 2 – Fig. 11).

First, the equilibrium points will be found, then the system will be linearized at any such points, the eigenvalues will be found and stability of the circuit will be determined, together with the frequency of generation, if it exists.

The lambda-diode current i_d is approximated as:

$$i_d(u_C)=\left(-0.0095u_C^5+0.1385u_C^4-0.5925u_C^3+0.2601u_C^2+2.1798u_C+0.0273\right)10^{-3}$$

The already found state equations (23) will be used:

$$\begin{cases} \dfrac{du_C}{dt} = -\dfrac{1}{RC}u_C + \dfrac{1}{C}i_L - \dfrac{1}{C}i_d(u_C) \\ \dfrac{di_L}{dt} = -\dfrac{1}{L}u_C + \dfrac{1}{L}E \end{cases}$$

The equilibrium points for u_C and i_L can be found by setting equations (23) to zero. Only one equilibrium point is found for these data:

$$u_C = 3.2\,V\,,\; i_L = 0.0051417$$

Now, let's find the Jacobian matrix of the functions f_1 and f_1 on the right sides of equations (23):

$$f_1 = 10^8 i_L + 950u_C^5 - 13850u_C^4 + 59250u_C^3 - 26010u_C^2 - 329090u_C - 2730$$
$$f_2 = -100u_C + 320$$

$$\boldsymbol{A} = \begin{bmatrix} \dfrac{\partial f_1}{\partial u_C} = 4750u_C^4 - 55400u_C^3 + 177750u_C^2 - 52020u_C - 329090 & \dfrac{\partial f_1}{\partial i_L} = 10^8 \\ \dfrac{\partial f_2}{\partial u_C} = -100 & \dfrac{\partial f_2}{\partial i_L} = 0 \end{bmatrix} \quad (28)$$

Evaluating $\boldsymbol{A}$ at $u_C = 3.2\,V$ and $i_L = 0.0051417$ gives

$$\boldsymbol{A} = \begin{bmatrix} 7332.4 & 10^8 \\ -100 & 0 \end{bmatrix}$$

So, the eigenvalues are found by

$$\det[\lambda\mathbf{I} - \boldsymbol{A}] = 0\;,\; \det\begin{bmatrix} \lambda - 7332.4 & -10^8 \\ 100 & \lambda \end{bmatrix} = 0 \quad (29)$$
$$\lambda^2 - 7332.4\lambda + 10^{10} = 0$$

$$\lambda = \frac{7332.4 \pm \sqrt{7332.4^2 - 4.10^{10}}}{2} = 3666.2 \pm j99933 \quad (30)$$

$$2\pi f = 99933 \qquad f = 15905 \text{ Hz}$$

Since the complex-conjugate eigenvalues have positive real parts, the equilibrium point will be unstable, and there will be generations with $f = 15905$ Hz.

Some conditions can be derived here, giving the relation between the circuit parameter R (loss resistance) and the dynamic resistance R_d of the nonlinear I-V characteristic of the lambda-diode at the bias point.

From the state equations (23), the following 2nd order differential equation can be found for the equilibrium point (where the differential conductance of the I-V characteristic is $\left(\frac{di_d}{du_C}\right)_{u_C=3.2\,\mathrm{V}} = -\frac{1}{R_d} = -\frac{1}{844}\,\mathrm{S}$):

$$\frac{d^2u_C}{dt^2} + \frac{1}{C}\left(\frac{1}{R} - \frac{1}{R_d}\right)\frac{du_C}{dt} + \frac{1}{LC}u_C = \frac{E}{LC} \tag{31}$$

The characteristic equation is:

$$s^2 + \frac{1}{C}\left(\frac{1}{R} - \frac{1}{R_d}\right)s + \frac{1}{LC} = 0 \tag{32}$$

with solutions:

$$s_{1,2} = \frac{-\frac{1}{C}\left(\frac{1}{R} - \frac{1}{R_d}\right) \pm \sqrt{\left[\frac{1}{C}\left(\frac{1}{R} - \frac{1}{R_d}\right)\right]^2 - \frac{4}{LC}}}{2} \tag{33}$$

If the expression under the square root is negative in order to have complex-conjugate roots, the real part of the roots will be positive, if

$$\left(\frac{1}{R} - \frac{1}{R_d}\right) < 0\,,\ \text{or}\ R > R_d = 844\,\Omega \tag{34}$$

The expression under the square root will be negative and there will be complex-conjugate roots, if:

$$\frac{1}{R} < 2\sqrt{\frac{C}{L}} + \frac{1}{R_d},\ \text{or}\ R > 314\,\Omega \tag{35}$$

Both conditions (34) and (35) must be fulfilled in order to have complex conjugate eigenvalues with positive real part, and consequently, to have exponentially growing oscillatory waveform, i.e. $R > R_d = 844\,\Omega$. This growth cannot continue indefinitely. As the diode voltage variation gets large, the end points of the negative slope region of the diode I-V characteristic are attained, and the diode introduces more loss into the system, which acts as a limiting condition and a steady-state oscillation is produced. The numerical experiments prove that stable sinusoidal oscillations arise starting from $R > 850 - 900\,\Omega$.

5. CONCLUSIONS

On the basis of the presented features and the solved problems, the following conclusions can be made regarding the applicability of state space method in the Theoretical Electrical Engineering:

1) The state space method offers easy and systematic way of forming the transient equations.
2) The obtained equations can be solved in several ways, using software packages like Matlab, Mathematica and PSpice, including symbolically, which is very precious for obtaining analytic solutions in the course on TEE.
3) Powerful algorithms and computer programs exist for full computer implementation of the state space transient analysis.
4) The method allows nonlinear circuits to be easily analyzed.
5) The method is very universal and suitable for transient analysis not only in electrical circuits, but in many other kinds of systems: mechanical, thermal, fluid flows, economical systems, etc.
6) The method can be used in combination with the classical method after finding the state equations: the roots of the characteristic equation can be found using Eq. (20), and the dependent initial conditions can be easily calculated from the state equations.
7) The method can be used also in combination with the Laplace transform method: The state space equations can be transformed to Laplace form and symbolically solved using Matlab. Then, inverse Laplace transformation can be applied to find the originals.
8) The state space transient analysis is a valuable method of transient analysis that deserves more attention and space in the curriculum on Theoretical Electrical Engineering and in the Electrical Engineering education as a whole.

6. ACKNOWLEDGEMENT

The work was supported by grant from the Technical University - Sofia, Research Project No: 091ni138-08/2009.

REFERENCES

[1] S. L. Farchy, S. P. Papazow, Theoretical Electrical Engineering - part I, Technika, Sofia, 1987, (in Bulgarian).
[2] John Staudhammer, Circuit Analysis by Digital Computer, Prentice Hall, 1975.
[3] Joseph Edminister, Schaum's Outline of Electric Circuits, 2nd ed., McGraw-Hill, New York, 1994.
[4] Donald Scott, An Introduction to Circuit Analysis-A System Approach, McGraw-Hill, New York, 1987.
[5] V. D. Borchoukova,..., K. G. Brandisky et al., Manual for Computer Solution of Problems in Theoretical Electrical Engineering, 2-nd ed. (Circuit Theory & Electromagnetic Fields), (in Bulgarian), Sofia: Technika, 1992, pp. 82-92.

HAMILTONIAN DYNAMICAL SYSTEMS WITH FINITE DEGREES OF FREEDOM AND OPTICAL SWITCHING

Ivan M. Uzunov

Department of Applied Physics, Technical University – Sofia

Abstract. *Hamiltonian two-degrees of freedom continuous nonlinear dynamical system [4] for description of dual-core nonlinear directional coupler has been studied. The stationary points of this dynamical system are calculated as a function of dimensionless parameter that describes the material and pulse properties. The dimensionless parameter controls the distance to the integrable case [11]. The stability of obtained stationary points is analyzed by means of approach suggested in [17]. Depending on the value of dimensionless parameter, different kinds of motion has been numerically identified. In the case of small value of dimensionless parameter, when the dynamics can be considered as near-integrable, stochastic motion has been observed. Suggestion has been done for a new way of reducing of the original system of nonlinear partial differential equations describing optical switching to dynamical system of finite degrees of freedom by means of one of the group invariant solutions of original system recently found in [15-16].*

Keywords: (nonlinear fiber couplers, Hamiltonian dynamical systems, stochastic motion)

INTRODUCTION

One possibility to study nonlinear guided waves is to transform the corresponding nonlinear partial differential equations with infinite degrees of freedom to dynamical systems with several degrees of freedom [1-4]. For example, recently possibility of application of variational approach in analysis of transmission characteristics of dual-core nonlinear directional coupler was discussed [5]. It was shown that a properly introduced two-degrees of freedom continuous Hamiltonian nonlinear dynamical system can properly describe transmission characteristics of dual-core nonlinear directional coupler. Another type of two-degree of freedom Hamiltonian dynamical system has been introduced in [4]. Two-degree of freedom Hamiltonian nonlinear dynamical system close to this analyzed in [4] has been earlier studied in [6, 7]. Nonlinear dynamical systems close to these considered in [4, 6-7], has been further analyzed in [8-10,12-14]. Question of the integrability of two-degree of freedom continuous nonlinear Hamiltonian dynamical system systems has been solved in [11] by means of the Painleve properties of corresponding equations.

First aim of this paper will be to discuss of the way of simplification of the original system of nonlinear partial differential equations describing optical switching (see Eq.(1) below) in the reduced system of ordinary differential equations. The idea suggested here is that the best choice of trial fuction should be one of the group invariant solutions of the original system found soon in [15-16] (see Eq. (2a-b) below). Main aim of this paper is to investigate the two-degrees of freedom nonlinear Hamiltonian dynamical system (see Eq.(5) below), derived in [4]. The dynamical behavior of this

system is analyzed as function of dimensionless parameter k that describes the material and pulse properties. The stationary (fixed) points of system given by Eq.(5), will be calculated as a function of dimensionless parameter k. The stability of obtained stationary points will be studied.As the system given by Eq.(5) in the case $k=0$, is integrable [11], for small values of parameter k it can be treated as near-integrable system. The conditions for appearance of stochastic behavior will be discussed.

BASIC EQUATIONS

Equations describing a dual-core nonlinear directional coupler

The propagation of pulses in a dual-core nonlinear directional coupler can be described in terms of following system of linearly coupled nonlinear Schrödinger equations [1-4]:

$$\begin{aligned} i\frac{dU}{d\xi}+\frac{1}{2}\frac{\partial^2 U}{\partial\tau^2}+\left|U\right|^2 U+KV=0 \\ i\frac{dV}{d\xi}+\frac{1}{2}\frac{\partial^2 V}{\partial\tau^2}+\left|V\right|^2 V+KU=0 \end{aligned} \tag{1}$$

where $U(\xi,\tau)$ and $V(\xi,\tau)$ are envelope functions of two pulses. The normalized coupling coefficient between the two envelope functions is equal to the linear coupling coefficient between two cores times the dispersion length.

Symmetry group consideration

Important remark is in order. Eq. (1) has been systematically analyzed by means of Lie group techniques in [15-16], where the most general Lie of point symmetries and its Lie algebra has been identified. Eq. (4) (see below) can be derived by means of case C) considered in [15- 16]. (See Eq.(6) in [16]) In this case after substituting the invariant solution of Eq.(1):

$$\begin{aligned} U(\xi,\tau)&=p(\tau)\exp\left[i\left(f(\tau)+\delta\xi\right)\right] \\ V(\xi,\tau)&=q(\tau)\exp\left[i\left(g(\tau)+\varepsilon\xi\right)\right] \end{aligned} \tag{2a}$$

following system of differential equations was obtained (case of $\gamma=0$):

$$\begin{aligned} &2p'f'+pf''+2Kq\sin(g-f)=0 \\ &2q'g'+qg''+2Kp\sin(f-g)=0 \\ &p''-p(f')^2+2p^3-2\delta p+2Kq\cos(g-f)=0 \\ &q''-q(g')^2+2q^3-2\varepsilon q+2Kp\cos(f-g)=0 \end{aligned} \tag{2b}$$

where prime denotes differentiation. The most important feature of Eq. (2b), is the presence of phase functions $f(t)$ and $g(t)$, which describe the time depending phases

of the envelope functions of pulses in dual-core directional coupler. In its general form however, the analysis of solutions of Eq.(1) in the form of Eq. (2), requires consideration of 8-dimensional dynamical system (or nonlinear dynamical system with 4 degrees of freedom).

Basic reduced system

Following [4] stationary pulse solutions are represented in the form:

$$U(\xi,\tau)=u(\tau,q)\exp(iq\xi),\ V(\xi,\tau)=v(\tau,q)\exp(iq\xi), \tag{3}$$

where q is the parameter of soliton state family of solutions and $u(\tau,q)$ and $v(\tau,q)$ are real functions. By rescaling the functions and variables in the resulting system of ordinary equation for $u(\tau,q)$ and $v(\tau,q)$ so that:

$$u=\sqrt{q}\,f, v=\sqrt{q}\,g, \tau=t/\sqrt{q}\,, k=K/q\,,$$

following set of equations for functions f and g has been derived:

$$\begin{aligned}\frac{1}{2}\frac{d^2f}{dt^2}-f+f^3+kg=0\\ \frac{1}{2}\frac{d^2g}{dt^2}-g+g^3+kf=0\end{aligned} \tag{4}$$

The parameter $k=K/q$ describes the material (K) and pulse (q) properties. As we can see, assuming that, $\varepsilon=\delta$, $f(t)=g(t)=0$ from Eq. (2b) immediately follows Eq.(4). We can therefore, conclude that the Eq.(4) used in [4], as well as a trial function used in [5], reflects in a different ways the basic group symmetry properties of the Eq.(1) . In my understanding the best trial function for analyzing of Eq.(1) should be the group invariant solution Eq.(2b) (or Eq. (6) of [15]). Finally the change of variables: $f=(x+y)/\sqrt{2}$, $g=(x-y)/\sqrt{2}$ leads to the system of equations

$$\begin{aligned}\frac{1}{2}\frac{d^2x}{dt^2}-2(1-k)x+x^3+3xy^2=0\\ \frac{1}{2}\frac{d^2y}{dt^2}-2(1+k)y+y^3+3yx^2=0\end{aligned} \tag{5}$$

The magnitude of the parameter k controls the possibility of existing of symmetric $(k>1)$ and antisymmetric soliton states $(k>0)$ [4]. Dynamical system represented by Eq.(5) is the basic reduced system of this analysis.

The Hamiltonian of this system is:

$$H = T + V \tag{6}$$

where $T = 1/2\left(\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2\right)$ is the kinetic energy and

$V = -\left(x^2 + y^2\right) + k\left(x^2 - y^2\right) + 1/2\left(x^2 + y^2\right)^2 - 1/4\left(x^2 - y^2\right)^2$ is the potential energy.

Dynamical system given by Eq. (5) can be considered as governing the motion of a particle in a two-dimensional potential well or as a two nonlinearly coupled harmonic oscillators. In the general case $\left(k \neq 0\right)$, Eq. (5) represents Hamiltonian, but not integrable two degrees of freedom dynamical system. Four integrable Hamiltonian two degree of freedom systems similar to (5) have been found in [11]. One of them 1(ii) (see the notations of [11]) corresponds to the famous Manakov system and recently has been investigated from point of view of dynamical systems in [12].

The other - case 1(i) [11] coincides with Eq.(5) in the case $k = 0$.

$$V = V_{LAKSH} + k\left(x^2 - y^2\right)$$

Of course, $k = 0$ corresponds to the Hamiltonian and integrable system case 1(i) [11]. The second integral of motion of Eq.(5) in case of $k = 0$ is [11]:

$$I = \frac{dx}{dt}\frac{dy}{dt} - 2xy + xy\left(x^2 + y^2\right) \tag{7}$$

As has been shown in [11], dynamical system represented by Eq.(5) ($k = 0$) is not only Hamiltonian and integrable, but also separable. So parameter k controls how far is the potential energy of dynamical system Eq.(5) from the potential energy of integrable variant case 1(i) [11]. This observation allows us to consider dynamical system given by Eq.(5) for small values of parameter k as nearly-integrable system.

Let us rewrite the system Eq.(5) in a following form:

$$\begin{aligned}
&\frac{d}{dt}y_1(t) = y_2(t)\\
&\frac{d}{dt}y_2(t) = 2(1-k)y_1(t) - y_1^3(t) - 3\,y_1(t)y_3^2(t)\\
&\frac{d}{dt}y_3(t) = y_4(t)\\
&\frac{d}{dt}y_4(t) = 2(1+k)y_3(t) - y_3^3(t) - 3\,y_3(t)y_1^2(t)
\end{aligned} \tag{8}$$

where $y_1(t) = x(t),\ y_2(t) = \frac{dx}{dt}, y_3(t) = y(t), y_4(t) = \frac{dy}{dt}$.

STATIONARY POINTS OF BASIC REDUCED SYSTEM

The stationary points (fixed points) of system (8) are the points, where $\dot{y} = 0$, which are the critical points of Hamiltonian H. Therefore, we should solve the following system as a function of the parameter k:

$$DH_y = 0, \tag{9}$$

where DH_y is the derivative of H at y.The stationary points of Eq. (9) have been calculated and results are presented in Appendix A. For convenience, the notation for the different stationary points has been introduced. We have also studied the minima and maxima of potential function in Eq. (6), as a function of y_1, y_3, and parameter k. The obtained results are presented graphically in Appendix B, by means of 3D-dimensional plots of potential function. The obtained results for the coordinates of potential extrema, coincide with the results for stationary points of Eq. (9). The values of potential function at these points have been determined and results are presented in the Appendix C. In the brackets is shown the number of points with a given value of the potential energy. Clearly can be seen how strong is the influence of the change of the parameter k on the number and location of the stationary points. The number of fixed points reduces from 9 in the integrable case $(k = 0)$, till 5 $(k = 0.5)$, and, finally till 3 $(k = 1)$.

STABILITY OF STATIONARY POINTS OF BASIC REDUCED SYSTEM

Linearization about an equilibrium points, is given by:

$$\dot{\delta}y = L\delta y, \tag{10}$$

where $L = JD^2H_0$ is infinitesimally symplectic matrix .

Eigenvalues of linear problem determine the stability of equilibrium points. The general solution of the problem of linear stability of stationary points is found by conditions that all the eigenvalues have negative real part [20]. The marginal case is when the eigenvalues lie on the imaginary axis.

The stationary point is spectrally stable if every solution of the linearization is bounded as $t \to \infty$, that requires that all eigenvalues of L lie on the imaginary axis or in the left half plane. The stationary point is linearly stable, if it is spectrally stable and in addition, in the case of multiple eigenvalues, all Jordan blocks are one-dimensional [17].

L in Eq. (10) is infinitesimally symplectic, therefore, the eigenvalues appear in pairs $\pm\sigma$. Since L is real, eigenvalues must also come in complex conjugate pairs. Eigenvalues for Hamiltonian system may occur in one or more of the following groups: a) pairs on the imaginary axis: $\pm i\omega, \omega \neq 0$; b) pairs on the real axis: $\pm\sigma, \sigma \neq 0$; c) complex quadruplets: $\pm a \pm ib, a, b \neq 0$, and d) zero [17].

As was shown in [17], the degree of the characteristic polynomial that corresponds to Eq.(11) can be reduced by a factor of two. For the case of two degrees of freedom system, in terms of $\tau = -\sigma^2$ corresponding reduced characteristic polynomial is [17]:

$$Q(\tau) = \tau^2 - A\tau + B, \tau = -\sigma^2, \tag{11}$$

where $A = -Tr\left(L^2\right)/2$ is the sum of the principal (2x2) minors, and $B = \det L$. Note that $TrA = 0$. The roots of Eq.(11) are given by: $\tau = \frac{1}{2}A \pm \sqrt{\left(\frac{1}{4}A^2 - B\right)}$.

An stationary point is spectrally stable if all eigenvalues lie on the imaginary axis. Therefore both roots should be non-negative, or

$$A \geq 0,\ 0 \leq B \leq A^2/4. \tag{12a}$$

An stationary point is on the boundary of spectral stability if there is a root at $\tau = 0$:

$$A \geq 0,\ B = 0, \tag{12b}$$

or a multiple root at $\tau > 0$:

$$A \geq 0,\ B = A^2/4. \tag{12c}$$

We have performed linearization about all obtained above determined equilibrium points for all considered values of the parameter k. The corresponding characteristic and reduced characteristic polynomials have been derived and solved. Obtained values for the eigenvalues are presented in Appendix A. As we can see, eigenvalues in agreement with earlier results of [17], occur in one or more of the expected groups: a) pairs on the imaginary axis: $\pm i\omega,\ \omega \neq 0$; b) pairs on the real axis: $\pm\sigma,\ \sigma \neq 0$; c) zero. Many of the obtained eigenvalues are with the zero real parts therefore nonhyperbolic ones [21-22].

Stability of each stationary point has been calculated by means of Eq. (12). Obtained results are also reported in Appendix C, last column. Note that in case of $k = 0$, stationary points M_2, M_4, M_6, M_8, M_9 are unstable, while M_1, M_3, M_5, M_7 are points on the boundary of stability. With increase of the parameter k, from $k = 0$ till $k = 0.4$, the unstable points M_2, M_8 move toward stable point M_1, while unstable points M_4, M_6 move toward stable point M_5. In the case of $k = 0.5$ there are only five stationary points: the points M_2, M_8 collapses with M_1, while M_4, M_6 collapses with M_5. In fact, if we look at the evolution of the eigenvalues of the linearized problem for the points M_1, M_5 for the values of the parameter k between 0.4 and 0.6 we can see that in the beginning $(k = 0.4)$ these are stable points, after these $(k = 0.5)$ one of their pairs of imaginary eigenvalues transformed into double zero, and finally

$(k = 0.6)$ a pair of real on the real axis appear. Such kind of transformation of eigenvalues of linearized problem for Hamiltonian system under perturbation (in our case, change of the parameter k) was earlier described in [19], case (iv). Further increase of the parameter k, leads to the collapses between the unstable points M_1, M_5 and M_9. In the case $k = 1$, configuration of two stable stationary points M_3, M_7 and one unstable point M_9 appears.

GENERIC BEHAVIOR OF BASIC REDUCED SYSTEM

Integrable case $k = 0$

Here, I will perform numerical solution of Eq. (8). Periodic motion appears in the vicinity of stable fixed points (centers), or fixed points with eigenvalues with pure imaginary eigenvalues of linearized problem. Several possibilities to represent the periodic solution will be used. First, time dependences of functions $y_1(t)$ and $y_3(t)$ will be examined. Second, the basic parametric dependences $y_1(y_2)$ and $y_3(y_4)$ will be constructed. In some cases a trajectory of motion will be shown in a three - dimensional space by means of 3D – parametric plots. Finally, Poincare surface of sections ($y_1(y_2)$, or $y_3(y_4)$) will be build.

Initial conditions for the solution of Eq. (8) are determined in the following way. First, we choose the stable fixed point in the vicinity of which we analyze the periodic potion. Second, the value of the Hamiltonian or energy ($H = E$) is fixed, taking into account the value for the extrema of the potential function at this fixed point. Next, the initial values of coordinates $y_1(0), y_3(0)$, are chosen close to the coordinates of the chosen fixed point. The value of $y_2(0)$ is arbitrary, but the value of $y_4(0)$ is calculated using the given value of energy $H = E$.

Let us consider the motion in the vicinity of the fixed point $M_3 : y_1 = 0; y_2 = 0; y_3 = 1.414; y_4 = 0$ (Appendix A, $k = 0$). The value of the energy is $E = -0.9$. 3D ($y_2(t), y_3(t), y_4(t)$)–dependence of the trajectory of motion is shown in the Fig. (1):

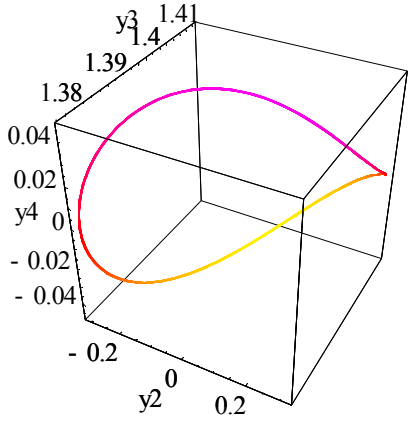


Fig. 1. 3D parametric dependence $y_2(t), y_3(t), y_4(t)$

Time dependences of functions $y_1(t)$ and $y_3(t)$, are shown in Fig. (2a-2b).

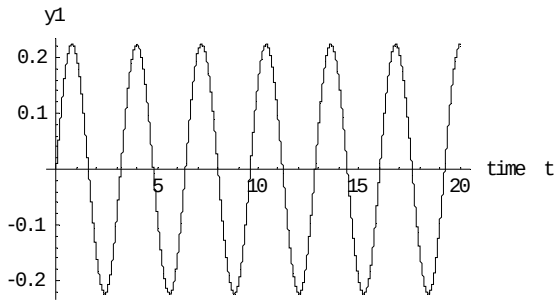


Fig. 2a. Time dependence of $y_1(t)$

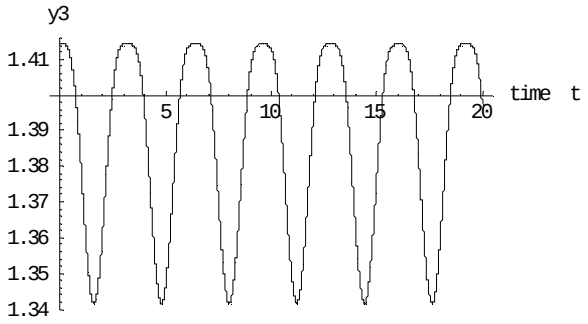


Fig. 2b. Time dependence of $y_3(t)$

The parametric dependences $y_1(y_2)$ and $y_3(y_4)$ are presented in the Fig. (2c-2d).

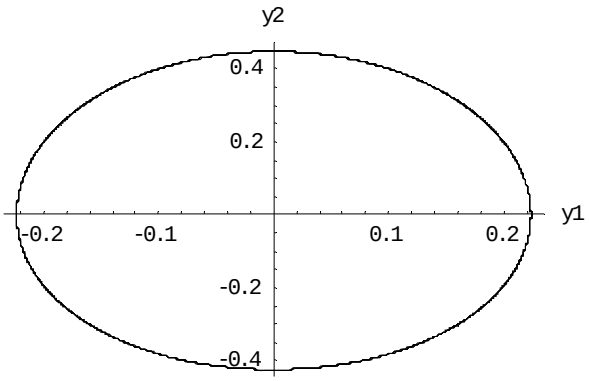


Fig. 2c. Parametric dependence $y_1(y_2)$

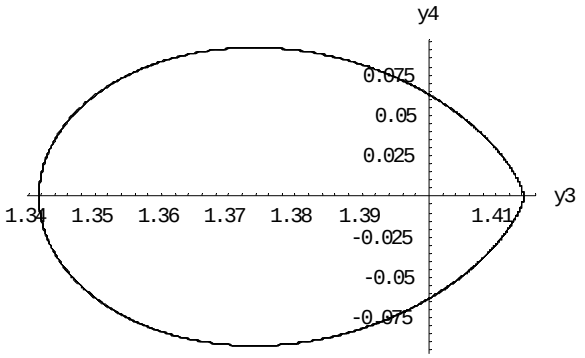


Fig. 2d. Parametric dependence $y_3(y_4)$

In both pairs of figures: Fig. (2a-2b), and Fig. (2c-2d), typical features of periodic motion are observed. In order to investigate the influence on the characteristics of the periodic motion of the value of the energy, following values of energy have been studied: $E = -0.95,-0.85,-0.75,-0.65,-0.55,-0.45$. The parametric dependences $y_1(y_2)$ is shown in the Fig. (3a). With increase of the energy the size of the trajectory increases, but periodicity conserves.

In Fig. (3b) the dependence of the parametric connection $y_3(y_1)$ on energy is shown. As we know, for the case of two linear coupled oscillators parametric dependences like this can be used to estimate the relation between frequencies of both oscillators. Here the relation of frequencies preserves with the increase of the energy. In order to analyze the dependence of the period of the motion on the energy, the time dependences of $y_3(t)$, are compared in Fig. (4a-b), for the following values of energy $E = -0.95,-0.45$.

Fig. 3a. Dependence of the size of trajectory on energy $y_1(y_2)$

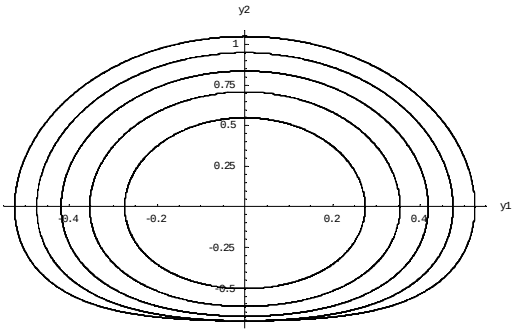
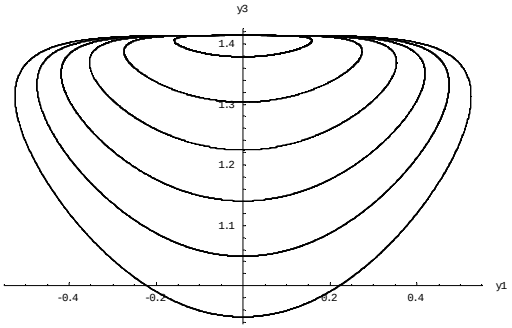


Fig. 3b. Frequency analysis of the motion



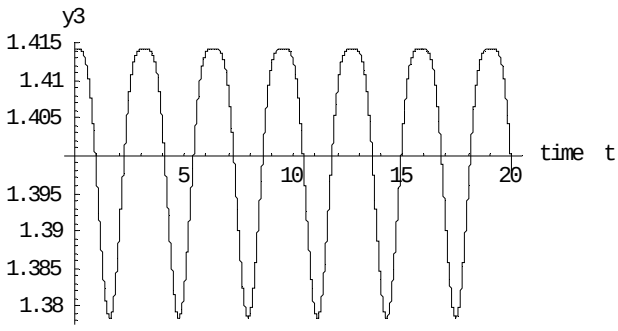


Fig. 4a. $y_3(t)$ for $E = -0.95$

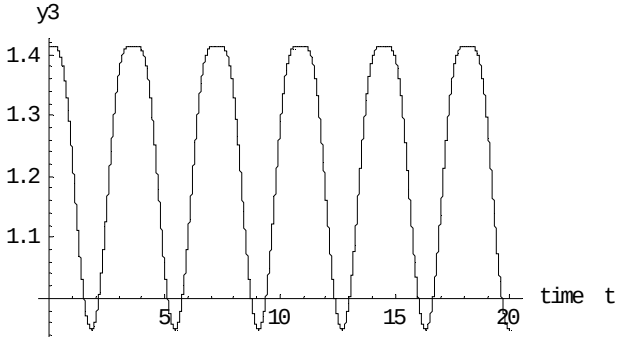


Fig. 4b. $y_3(t)$ for $E = -0.45$

It is clearly observed the increase of the period of the motion with the increase of the energy. Going closer to the separatrice, with her infinitely large period of motion, the period of periodic motion should increase. The increase in the amplitude of the $y_3(t)$, with the increase of the energy can be also seen. Let us consider the motion in the vicinity of the fixed point $M_2 : y_1 = 0.707; y_2 = 0; y_3 = 0.707; y_4 = 0$ (Appendix A, $k = 0$). The value of the energy is $E = -0.48$. 3D ($y_2(t), y_3(t), y_4(t)$)–dependence of the trajectory of motion is shown in the Fig. (5a):

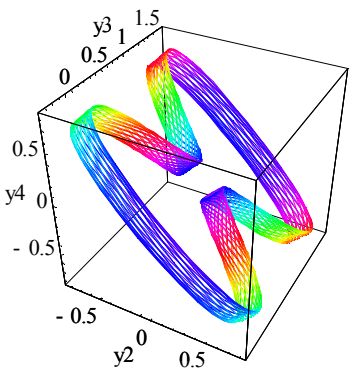


Fig. 5a. 3D dependence of $y_2(t), y_3(t), y_4(t)$

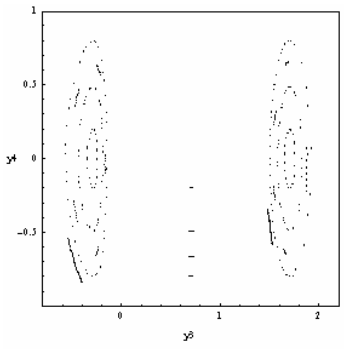


Fig. 5b. Poincare section of trajectory with the plane $y_2 = 0$,

$E = -0.48,-0.38,-0.28,-0.18$

The existence of the tori and their change with the energy:
$E = -0.48,-0.38,-0.28,-0.18$, can be clearly observed by the Poincare section with the plane $y_2 = 0$ presented in Fig.(5b).

Small perturbation $k = 0.1$

The corresponding fixed point for $k = 0.1$ is $y_1 = 0; y_2 = 0; y_3 = 1.483; y_4 = 0$. The value of the energy is $E = -1.21$. In Fig. (6a), the comparison of Poincare sections of trajectory with the plane $y_2 = 0$, calculated for the energies $E = -1.19,-1.09,-1.0,-0.9$ is presented. With increase of the energy the size of the tori gets larger. Quite interesting seems to be behavior of the 3D parametric dependence $y_2(t), y_3(t), y_4(t)$, $E = -1.0$, as shown in Fig. (6b).

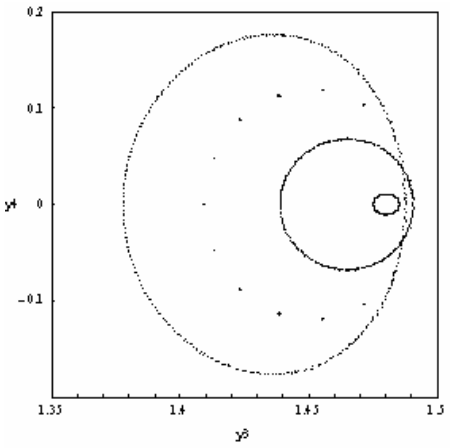


Fig. 6a. Poincare sections of trajectory with $y_2 = 0$, for the cases $E = -1.19,-1.09,-1.0,-0.9$

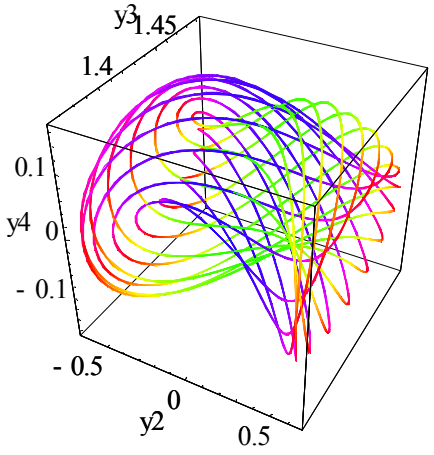


Fig. 6b. 3D dependence $y_2(t), y_3(t), y_4(t)$ for the case $E = -1.0$

Appearance of stochastic motion

We continue our investigation of the dynamical behavior of the basic reduced system for $k = 0.1$ starting with initial condition given by $y_1 = 0;\ y_2 = 0;\ y_3 = 1.483;\ y_4 = 0$, but now increasing the values of energy $E = -0.8,-0.7,-0.6,-0.5,-0.4,-0.3,-0.2,-0.1,0$. Till energy $E = -0.1$ observed dynamics is comparable with described in the previous paragraph. In the case $E = 0$, however stochastic motion has been observed. In the next two figures Fig. (7a-b) are shown the calculated Poincaré sections of trajectory with the plane $y_2 = 0$ for energies $E = -0.1$ and $E = 0$, respectively.

Clearly stochastic behavior is observed in Fig.(7b). Note that in the case of $k = 0.1$, dynamical system given by Eq.(5) is Hamiltonian, but not integrable one. Appearance of stochastic behavior is then possible [21].

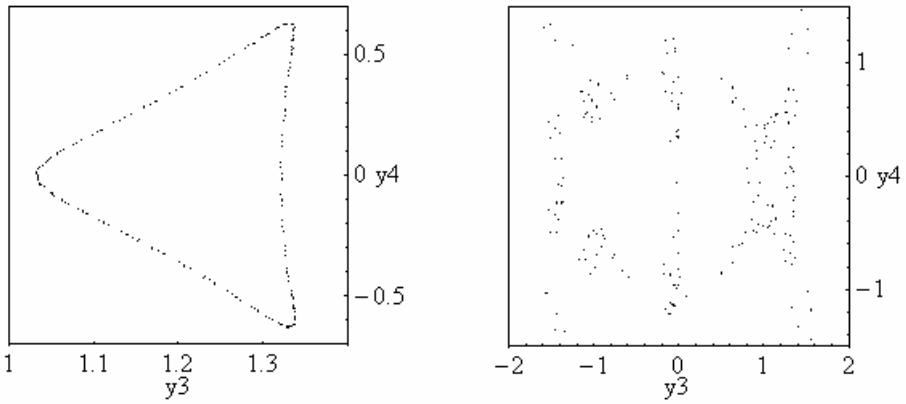


Fig.(7a-b) Poincaré sections of trajectory with the plane $y_2 = 0$, for $E = -0.1$ and $E = 0$, respectively

CONCLUSION

The conjecture has been done here is that the best ansatz to simplify the original system of nonlinear partial differential equations Eq.(1), that describes optical switching is the group invariant solution in the form Eq.(2a), and the corresponding reduced system is Eq.(2b).

The dynamical behavior of two-degrees of freedom Hamiltonian dynamical system Eq. (5), derived in [4] was analyzed as function of dimensionless parameter k that describes the material and pulse properties. The stationary (fixed) points of system Eq. (8), have been calculated as a function of dimensionless parameter k. The stability of obtained stationary points has been studied by means of approach suggested in [17,19]. In the case of small value of dimensionless parameter, when the dynamics can be considered as near-integrable, stochastic motion has been observed.

REFERENCES

[1] Agrawal, G.P. "Applications of Nonlinear Fiber Optics", Academic Press, San Diego, chapter 7, (2001).

[2] Hasegawa A., and Kodama Y., "Solitons in optical communications", Clarendon Press, Oxford, (1995).

[3] Akhmediev N. N., and Ankiewicz A., "Solitons. Nonlinear Pulses and Beams", chapter 8, Chapman and Hall, (1997).

[4] Akhmediev N.N., and Ankiewicz A., "Novel soliton states and bifurcation phenomena in nonlinear couplers", Phys. Rev. Lett., 70, 2395-2398, (1993).

[5] Uzunov, I.M., Muschall, R., Golles, M., Kivshar, Yu.S., Malomed B., and Lederer, F.,"Pulse switching in nonlinear fiber directional coupler", Phys. Rev. E, 51, 2527-2537, (1995).

[6] Akhmediev N.N., Eleonsky V.M., Kulagin N.E., and Shil'nikov L.P., Sov. Tech. Phys. Lett., 15, 587, (1989).

[7] Eleonsky V.M., Korolev V.G., Kulagin N.E., and Shil'nikov L.P., "Bifurcation of vector envelope solutions and integrability of Hamiltonian systems", Sov. Phys. JETP 72, 619,(1992).

[8] Haelterman M., and Shepard A., "Bifurcation phenomena and multiple soliton states in isotropic Kerr media", Phys. Rev. E, 49, 3376-3381, (1994).

[9] Haelterman M., and Shepard A., "Bifurcations of the dark soliton and polarization domain walls in nonlinear dispersive media", Phys. Rev. E, 49, 4512-4518, (1994).

[10] Haelterman M., "Modulational instability, periodic waves and black and white vector solitons in birefringent Kerr media", Opt. Communications, 111, 86-92, (1994).

[11] Lakshmanan, M.,and Sahadevan, R., "Coupled quartic anharmonic oscillators, Painlevé analysis, and integrability", Phys. Rev. A, 31, 861-876, (1985).

[12] Polymilis C., Hizanidis K., and Frantzeskakis D.J., " Phase plane Stäckel potential dynamics in the Manakov system", Phys. Rev. E, 58, 1112-1124, (1998).

[13] Briksin B.B., M.P. Petrov, R.B. Kijan, Zh. Eksp.Teor.Fiz., 107, 732-740, (1995).

[14] Kostov N.A., and Uzunov I.M., "New kinds of periodical waves in birefringent optical fibers", Opt. Communications, 89, 389-392, (1992).

[15] Pulov V., Uzunov I.M., and Chakarov E., "Solitons and laws of conservation for coupled nonlinear Schrödinger equations: Lie group analysisi", Phys. Rev. E, 57, 3468-3477, (1998).

[16] Pulov V., Uzunov I.M., Chakarov E., and Lyutskanov V., "Symmetry classification of solutions to coupled nonlinear Schrödinger equations", Bulg. J. Phys., 34, 241-251, (2007).

[17] Howard, J.E., and Mackay, "Calculation of linear stability boundaries for equilibria of Hamiltonian systems", Phys. Lett. A122, 331-334, (1987).

[18] Гантмахер Ф.Р., „Теория матриц", Москва, Наука, (1966).

[19] Mackay R.S., "Stability of equilibria of Hamiltonian system", Nonlinear Phenomena and Chaos, pp. 254-270, (1986).

[20] Меркин Д. Р., "Введении в теорию устойчивости движения", Наука, Москва, глава 3, (1987).

[21] Guckenheimer J., and Holms P., "Nonlinear oscillations, dynamical systems, bifurcations of vector fields", Springer-Verlag, chapter 3 and chapter 7, (1983).

APPENDIX A: Stationary points of the system, eigenvalues of the corresponding linearized problem, and stability of the corresponding stationary points

k = 0

Point	y_1	y_2	y_3	y_4	λ_1	λ_2	λ_3	λ_4	*Stability*
M_1	+1.414	0	0	0	$-2i$	$2i$	$-2i$	$2i$	*s*
M_2	0.707	0	0.707	0	−1.414	$-2i$	$2i$	1.414	*u*
M_3	0	0	+1.414	0	$-2i$	$2i$	$-2i$	$2i$	*s*
M_4	−0.707	0	+0.707	0	−1.414	$-2i$	$2i$	1.414	*u*
M_5	−1.414	0	0	0	$-2i$	$2i$	$-2i$	$2i$	*s*
M_6	−0.707	0	−0.707	0	−1.414	$-2i$	$2i$	1.414	*u*
M_7	0	0	−1.414	0	$-2i$	$2i$	$-2i$	$2i$	*s*
M_8	0.707	0	−0.707	0	−1.414	$-2i$	$2i$	1.414	*u*
M_9	0	0	0	0	−1.414	−1.414	1.414	1.414	*u*

k=0.1

Point	y_1	y_2	y_3	y_4	λ_1	λ_2	λ_3	λ_4	*Stability*
M_1	1.342	0	0	0	$-1.789i$	$1.789i$	$1.897i$	$-1.897i$	*s*
M_2	0.775	0	0.632	0	-1.359	$-1.986i$	$1.986i$	1.359	*u*
M_3	0	0	1.483	0	$-2.098i$	$2.098i$	$-2.191i$	$2.191i$	*s*
M_4	-0.775	0	0.632	0	-1.359	$-1.986i$	$1.986i$	1.359	*u*
M_5	-1.342	0	0	0	$-1.789i$	$1.789i$	$1.897i$	$-1.897i$	*s*
M_6	-0.775	0	-0.632	0	-1.359	$-1.986i$	$1.986i$	1.359	*u*
M_7	0	0	-1.483	0	$-2.098i$	$2.098i$	$-2.191i$	$2.191i$	*s*
M_8	0.775	0	-0.632	0	-1.359	$-1.986i$	$1.986i$	1.359	*u*
M_9	0	0	0	0	-1.483	-1.342	1.342	1.483	*u*

k=0.4

Point	y_1	y_2	y_3	y_4	λ_1	λ_2	λ_3	λ_4	*Stability*
M_1	1.095	0	0	0	$-0.894i$	$0.894i$	$-1.549i$	$1.549i$	*s*
M_2	0.949	0	0.316	0	-0.985	$-1.723i$	$1.723i$	0.985	*u*
M_3	0	0	1.673	0	$-2.366i$	$2.366i$	$-2.683i$	$2.683i$	*s*
M_4	-0.949	0	0.316	0	-0.985	$-1.723i$	$1.723i$	0.985	*u*
M_5	-1.095	0	0	0	$-0.894i$	$0.894i$	$-1.549i$	$1.549i$	*s*
M_6	-0.949	0	-0.316	0	-0.985	$-1.723i$	$1.723i$	0.985	*u*
M_7	0	0	-1.673	0	$-2.366i$	$2.366i$	$-2.683i$	$2.683i$	*s*
M_8	0.949	0	-0.316	0	-0.985	$-1.723i$	$1.723i$	0.985	*u*
M_9	0	0	0	0	-1.673	-1.095	1.095	1.673	*u*

k=0.5

Point	y_1	y_2	y_3	y_4	λ_1	λ_2	λ_3	λ_4	*Stability*
M_1	1	0	0	0	0	0	$-1.414i$	$1.414i$	
M_3	0	0	1.732	0	$-2.449i$	$2.449i$	$-2.828i$	$2.828i$	*s*
M_5	-1	0	0	0	0	0	$-1.414i$	$1.414i$	
M_7	0	0	-1.732	0	$-2.449i$	$2.449i$	$-2.828i$	$2.828i$	*s*
M_9	0	0	0	0	-1.732	-1	1	1.732	

k=0.6

Point	y_1	y_2	y_3	y_4	λ_1	λ_2	λ_3	λ_4	*Stability*
M_1	0.894	0	0	0	-0.894	$-1.265i$	$1.265i$	0.894	
M_3	0	0	1.789	0	$-2.53i$	$2.53i$	$-2.966i$	$2.966i$	*s*
M_5	-0.894	0	0	0	-0.894	$-1.265i$	$1.265i$	0.894	
M_7	0	0	-1.789	0	$-2.53i$	$2.53i$	$-2.966i$	$2.966i$	*s*
M_9	0	0	0	0	-1.789	-0.894	0.894	1.789	

k=0.8

Point	y_1	y_2	y_3	y_4	λ_1	λ_2	λ_3	λ_4	Stability
M_1	0.632	0	0	0	–1.549	–0.894i	0.894i	1.549	
M_3	0	0	1.897	0	–2.683i	2.683i	–3.225i	3.225i	s
M_5	–0.632	0	0	0	–1.549	–0.894i	0.894i	1.549	
M_7	0	0	–1.897	0	–2.683i	2.683i	–3.225i	3.225i	s
M_9	0	0	0	0	–1.897	–0.632	0.632	1.897	

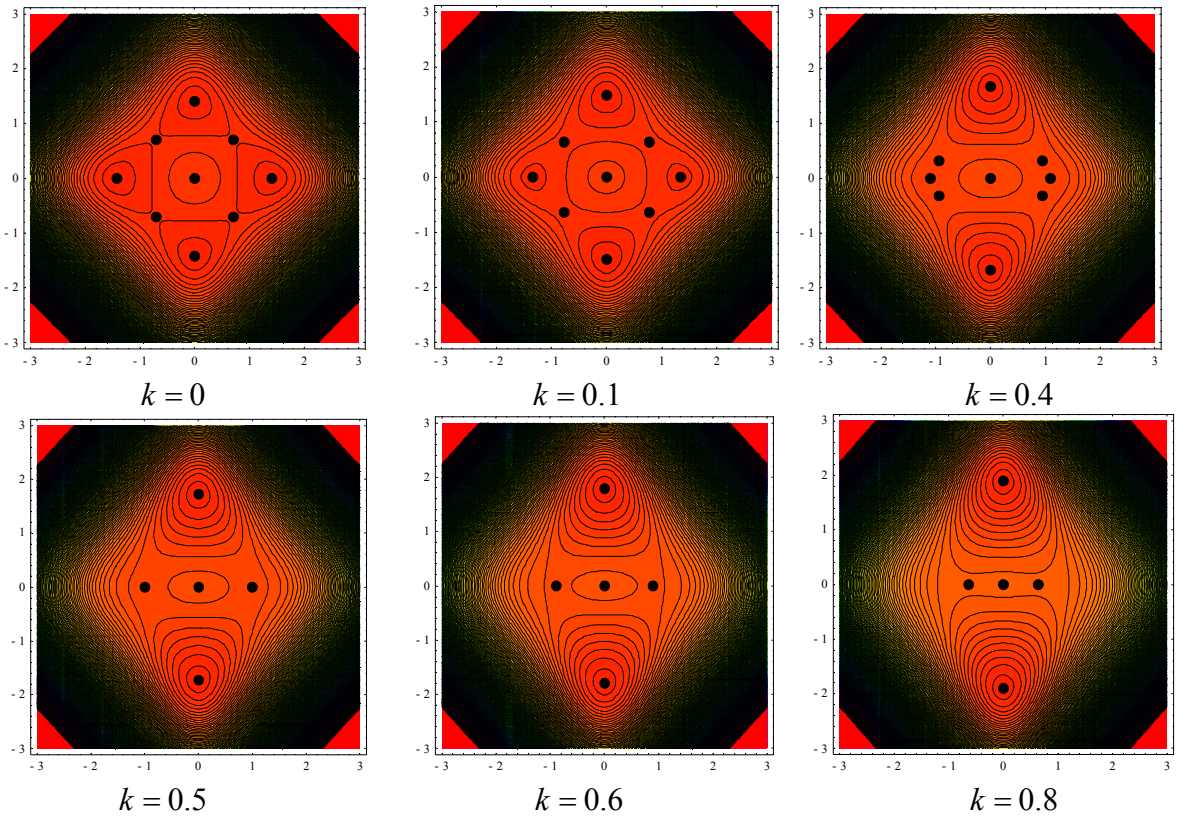
k=0.9

Point	y_1	y_2	y_3	y_4	λ_1	λ_2	λ_3	λ_4	Stability
M_1	0.447	0	0	0	–1.789	–0.632i	0.632i	1.789	
M_3	0	0	1.949	0	–2.757i	2.757i	–3.347i	3.347i	s
M_5	–0.447	0	0	0	–1.789	–0.632i	0.632i	1.789	
M_7	0	0	–1.949	0	–2.757i	2.757i	–3.347i	3.347i	s
M_9	0	0	0	0	–1.949	–0.447	0.447	1.949	

k=1

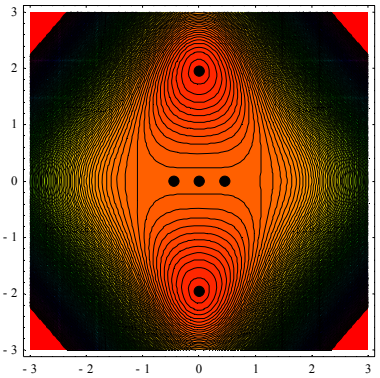
Point	y_1	y_2	y_3	y_4	λ_1	λ_2	λ_3	λ_4	Stability
M_3	0	0	2	0	–2iЦ2	2iЦ2	–2iЦ3	2iЦ3	s
M_7	0	0	–2	0	–2iЦ2	2iЦ2	–2iЦ3	2iЦ3	s
M_9	0	0	0	0	–2	0	0	2	

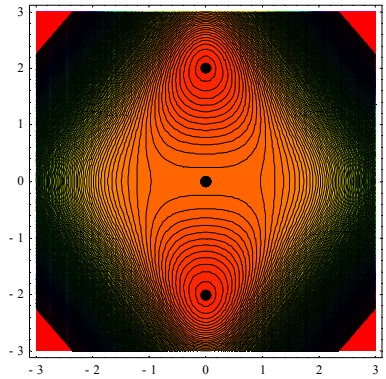
APPENDIX B: Evolution of extremuma of potential of system given by Eq.(6) as a function of the parameter k

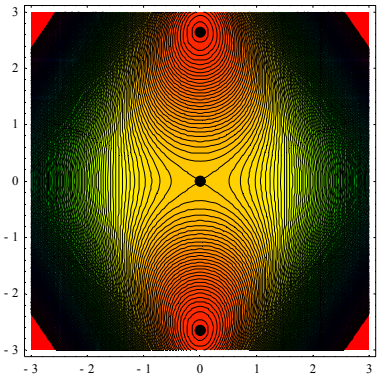


$k = 0$ $k = 0.1$ $k = 0.4$

$k = 0.5$ $k = 0.6$ $k = 0.8$


$k = 0.9$


$k = 1$


$k = 2.5$

APPENDIX C: Values of minima and maxima

K	Number	Values of minima and maxima of V
0	9	-1(4), -0.5(4), 0(1)
0.1	9	-1.21(2), -0.81(2), -0.49(4), 0(1)
0.4	9	-1.96(2), -0.36(2), -0.34(4), 0(1)
0.5	5	-2.25(2), -0.25(2), 0(1)
0.6	5	-2.56(2), -0.16(2), 0(1)
0.7	5	-2.89(2), -0.09(2), 0(1)
0.8	5	-3.24(2), -0.04(2), 0(1)
0.9	5	-3.61(2), -0.01(2), 0(1)
1.0	3	-4.0(2), 0(1)

IONIZATION MODIFICATION OF THE MATERIAL PARAMETERS - A POSSIBLE ROUTE TO GENERATION OF SPATIO-TEMPORAL SOLITONS

I. G. Koprinkov

Department of Applied Physics, Technical University of Sofia, 8 Kl. Ochridski Blvd, 1000 Sofia, Bulgaria,
phone: +02 965 3072, e-mail: igk@tu-sofia.bg

Abstract: *Brief introduction to some important regimes of propagation of optical pulses and beams is presented. The spatio-temporal dynamics of high-intensity femtosecond laser pulses below the ionization "threshold" and generation of optical tsunami pulses within (3+1)-dimensional nonlinear Schrödinger equation is considered. An advanced physical model of propagation of the high-intensity femtosecond laser pulses in presence of ionization is presented. The influence of ionization on the material parameters is studied. Ionization induced inversion of the group velocity dispersion at particular pulse propagation conditions is found. Stabilization of the pulse propagation in space and time is found. A possible pulse stabilization mechanism is discussed.*

Keywords: optical pulses and beams, spatial and temporal solitons, nonlinear Schrödinger equation, nonlinear envelope equation, pulse propagation, pulse compression, spatio-temporal solitons.

1. INTRODUCTION TO SOME BASIC PROPAGATION REGIMES OF OPTICAL PULSES AND BEAMS

The propagation of the light is a complex phenomenon that leads to rearrangement of the light parameters – amplitude, phase, etc. This usually results in spreading of the light in space and time, changing of the spatial and the temporal distribution of the light intensity, evolution of the phase fronts, etc. This takes place for both, the light pulses and the light beams. At some particular conditions, however, stable localized self-reinforced states of the light confined either in space or in time due to a precise balance between the linear and nonlinear processes can be generated. Such light structures are called *solitons.* Depending on whether the light is confined in space or in time, the solitons become spatial or temporal, respectively. Both types of solitons are well investigated, both theoretically and experimentally [1, 2], and only the mechanism of their generation will be briefly reminded here.

The solitons are generated as a result of a nonlinear light-matter interaction. In the case of *the spatial solitons*, the nonlinear effect, *i.e.*, self-focusing, balances the diffraction due to the self-induced change of the refractive index. Changing the refractive index, the electromagnetic field creates an optically written waveguide similar to a graded-index fiber. If the field is also a propagating mode of that waveguide, it will remain confined and it will propagate without changing its shape [1]. Depending on the number of the transversal dimensions the light is confined, such kind of soliton can be reffered to 1D or 2D soliton. In the case of *the temporal soli-*

tons, the nonlinear effect balances the dispersion due to the self-phase modulation. In that case, providing the light is confined in the transversal direction due to some real waveguide structure, *e.g.*, optical fiber, it may propagate long distance at almost no change of its parameters [2]. Such kind of soliton can be refered to (1+1)D soliton.

An extension of the classical soliton concept in conservative systems toward dissipative systems results in a new type of solitary waves, called *dissipative solitons* [3]. In that case, the soliton is generated due to a double balance between the diffraction/dispersion and the nonlinearity, from one side, and loss and gain in the medium, from the other.

In the cases considered above, the generated stable light structures represent partial solitons because the light is confined in one or two dimensions only, whereas in the other dimensions it is supported by some additional structure, *e.g.*, waveguide or optical fiber, or represents an infinite light beam, for which the problem of stability is not actual. In this report, our attention will be focused on the spatio-temporal problems of the light propagation and the generation of stable (3+1)D light structures – spatio-temporal solitons, or "light bullet". The generation of complete (3+1)D solitons in bulk nonlinear medium is a challenging task that is not yet completely understood and solved.

Generation of space-time confined pulse in a given space point of soliton-like intensity distribution, called *optical tsunami*, has been recently predicted by numerical simulations of the (3+1) D nonlinear *Schrödinger equation* [4, 5]. The optical tsunami can be created in a well predictable space point, whose position can be controlled within given degree by means of pulse and medium parameters. Based on such features of the tsunami pulse, a controllable guidance concept as an alternative to the soliton concept has been proposed [5].

Self-compression and stable propagation, along given length, of high-intensity femtosecond laser pulses have been discovered experimentally in a number of atomic and molecular gases and in solid bulk material [6 - 8]. A complete confinement of the light matter in space and in time has been achieved in this way. Self-compression has been also observed in air [9] and ionized noble gasses [10-12]. Although the stability of the pulse is not as perfect as in the case of temporal or special solitons separately, the stabilization of the pulse parameters due to such a self-consistent organization of the light matter is apparent [6-8]. The physical mechanism of the pulse compression below the ionization "threshold was found [4]. The mechanism of the pulse stabilization in the (3+1)D case, however, remains not yet completely understood. In the following, the (3+1)D propagation dynamics of the high-intensity femtosecond laser pulses below and above the ionization "threshold" will be described

2. SPATIO-TEMPORAL DYNAMICS OF HIGH-INTENSITY FEMTOSECOND LASER PULSES BELOW IONIZATION "THRESHOLD" - GENERATION OF OPTICAL TSUNAMI

The ionization strongly and non-instantaneously modifies the material parameters and, in this way, the pulse propagation. It is highly nonlinear process, therefore, it is

very sensitive to the changes of the pulse intensity and its control is very difficult in practice. By a suitable choice of the input pulse energy and/or gas pressure, the peak intensity can be kept below the level at which the ionization causes substantial modification of the material parameters. We consider a physical model of pulse propagation in a centrosymmetric bulk nonlinear medium, which includes the basic set of lowest order optical processes, diffraction, group velocity dispersion (GVD) and Kerr nonlinearity. The pulse propagation will be described by the (3+1)D non-linear Schrödinger equation (NLSE) for the complex field amplitude $\widetilde{E}(r,z,\tau)$

$$\frac{\partial\widetilde{E}}{\partial z} - \frac{i}{2k}\nabla_{\perp}^{2}\widetilde{E} + \frac{i\beta_2}{2}\frac{\partial^2\widetilde{E}}{\partial\tau^2} - \frac{ikn_2}{n_0}\left|\widetilde{E}\right|^2\widetilde{E} = 0 \qquad (1)$$

The field is presented in terms of the envelope-carrier concept [13]

$$E(t) = \widetilde{E}_0(t)\exp(-i\omega_0 t + i\varphi) + c.c. \qquad (2)$$

The lack of irreversible losses in the specified model means the condition for conservation of the pulse energy $W(z)$ holds, *i.e.*,

$$W(z) = \int_{S_\infty}\int_{-\infty}^{+\infty}\left|\widetilde{E}(r,z,\tau)\right|^2 ds\, d\tau = W_0 = const \ , \qquad (3)$$

where the integration is taken over the transversal cross-section *s* and the local time τ. This represents a normalization condition of the complex field amplitude and allows to control the pulse intensity, $I = \left|\widetilde{E}\right|^2$, and the absolute value of the field.

Noble gasses of high ionization potential, argon, neon and helium, have been considered as propagation media. This allows compression of high-intensity/energy pulses while remaining bellow the ionization "threshold". The initial pulse is a linearly polarized chirp-free Gaussian, $\widetilde{E}(r,z=0,\tau) = E_0\exp\left(-r^2/2r_0^2 - \tau^2/2\tau_0^2\right)$, of 100 fs pulse duration (full width at half maximum (FWHM)).

The evolution of the transversal width, peak intensity and the pulse duration of the propagating pulse in neon and helium are shown in Fig. 1(a), (b) and (c), respectively. The absence of ionization ensures a smooth and well controllable pulse propagation dynamics. The latter consists in a single self-compression event at $z = z_{SC}$, Fig. 1(c), in which the pulse duration shortens more than twice and the peak intensity increases more than an order of magnitude, Fig. 1(b). The rapid drop of the pulse intensity after the maximal intensity gain at $z = z_{IG}$ results from pulse splitting, [4, 5]. The peak intensity shows a very specific development during the pulse propagation. Initially, it changes gradually with the propagation distance but as the pulse is approaching $z = z_{IG}$ the pulse intensity starts growing in an "exploding" manner, lead-

ing to a dramatic increase of the peak intensity. This behavior resembles the development of the water tsunami wave and the respective optical pulse has been called *optical tsunami* [5]. The position of maximal peak intensity, the magnitude of the peak intensity, and the minimal pulse duration can be controlled within given limits by means of the pulse energy, Fig. 2, 3 and 4, respectively. Same kind of control can be achieved changing the pressure/number density of the medium.

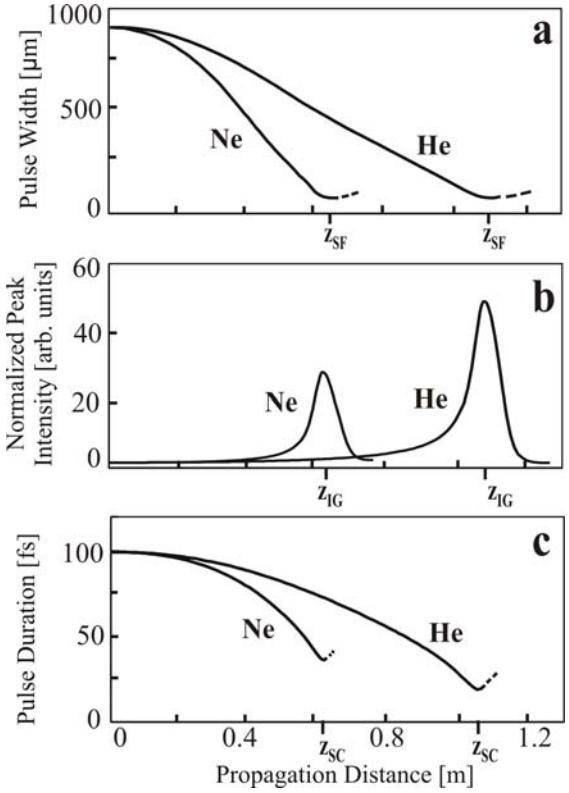


Fig. 1. Evolution of pulse width (a), peak intensity (b) and the pulse duration (c) during the propagation in pressurized neon and helium

The intensity profile of the tsunami pulse in time is very close to that of the sech^2 pulse shape of the fundamental temporal soliton of the (1+1)D NLSE in the case of negative GVD medium [2]. The theoretical and experimental studies show that temporal solitons do not exist within the specified basic set of optical processes and positive GVD, considered here. For many applications, however, a stable pulse along the whole propagation distance is not necessary but only a pulse of well specified parameters in given space point. That is why, the soliton concept (where it is not applicable) can be replaced by a new concept called *controllable guidance concept* [5]. The controllable guidance concept seems to be applicable in broad range of physical conditions, where the pulse propagation dynamics can be unambiguously predicted.

A general explanation of the self-compression of the pulse has been found [4] based on the energy conservation condition, Eq. (3). According to it, the energy integral can be considered as a four-dimensional "volume" (two transversal directions over the cross-section s of the pulse, a longitudinal direction - the local time τ, and a

"vertical" direction - the intensity $|\tilde{E}|^2$), which must be conserved along the propagation coordinate z. The rearrangement of the pulse is triggered by the self-focusing.

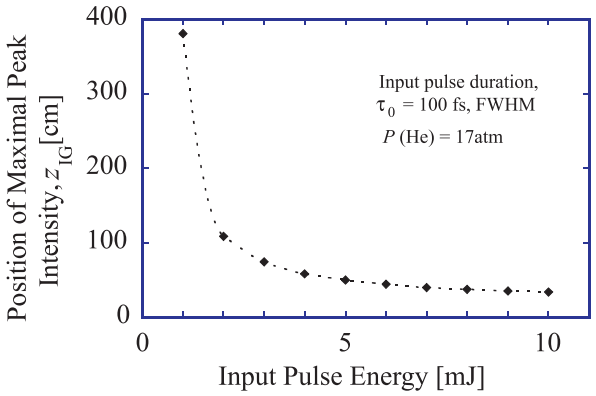


Fig.2. Position of the maximal peak intensity versus pulse energy

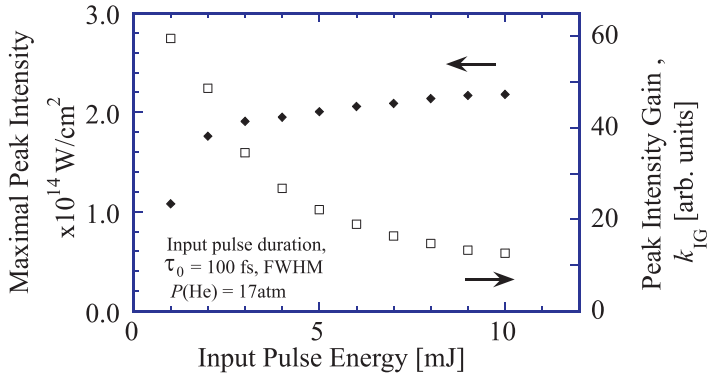


Fig. 3. Magnitude of the maximal peak intensity versus pulse energy

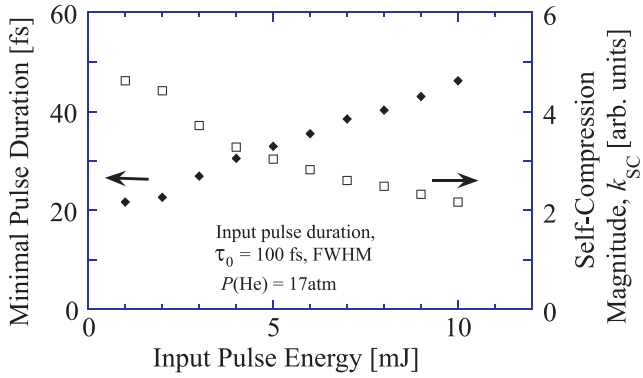


Fig. 4. Minimum pulse duration versus pulse energy

The self-focusing confines the pulse radially from all transversal directions toward the longitudinal axis, Fig.1(a). Due to the low dispersion of the rare gases, the pulse would not expand substantially in time while propagating given distance. At the same time, the confinement of the pulse in all transversal directions results in a strong increase of the pulse intensity, Fig.1(b). The latter leads to an effective (but real) shortening of the pulse (based on the standard FWHM characterization of pulse duration)

until it succeeds to broaden due to the dispersion, Fig.1(c). Such a concept of femtosecond pulse compression, formulated for the first time in [4], has been called self-compression in low dispersion regime. The pulse compression in low dispersion regime is inherently accompanied by a strong increase of the peak intensity at almost no loss of energy. It is substantially related with the low value of GVD and the efficiency of this mechanism increases when the GVD decreases.

3. SPATIO-TEMPORAL DYNAMICS OF HIGH-INTENSITY FEMTOSECOND LASER PULSES ABOVE IONIZATION "THRESHOLD"

3.1. Spatiotemporal dynamics of high-intnesity femtosecond laser pulses in bulk nonlinear media in presence of ionization - *experimental studies*.

The self-compression of high-intensity femtosecond laser pulses has been discovered experimentally in a number of atomic and molecular gasses [6-8]. The evolution of the temporal profile of 150 fs pulse of 1-mJ energy in the case of pressurized argon is shown in Fig.5(a) and the transversal pulse profile is given in Fig.5(b), [7, 8].

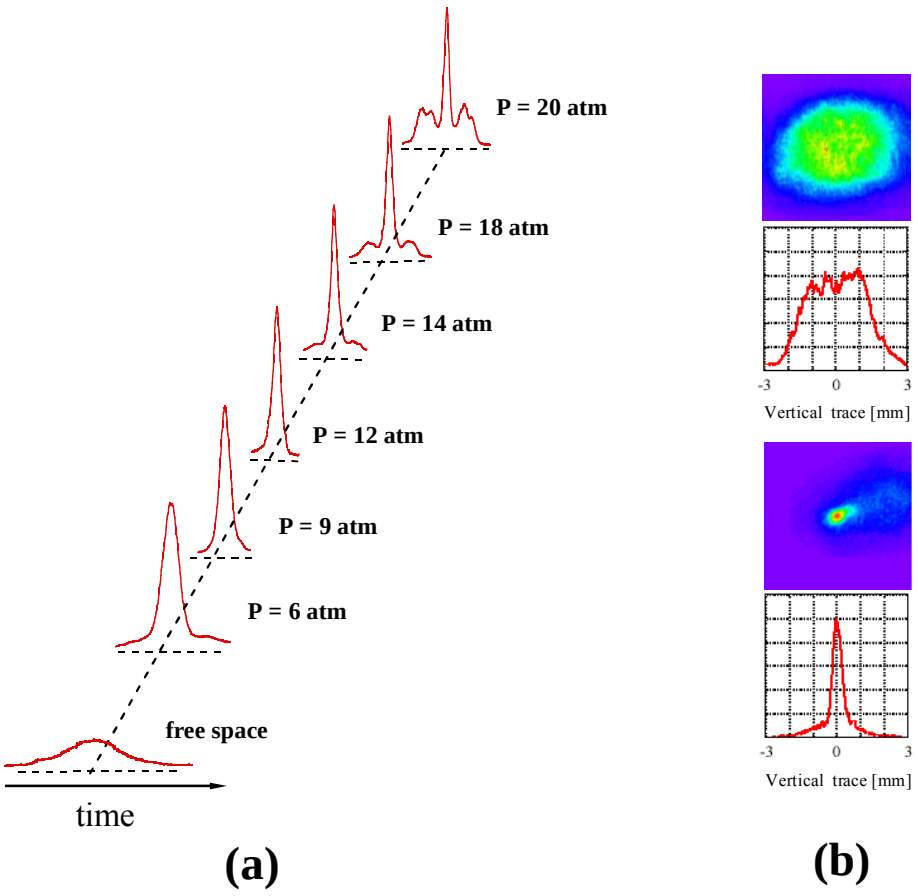


Fig. 5. Temporal (a) and spatial (b) evolution of the high-intensity femtosecond laser pulse in pressurized argon

The main features of the pulse compression found from the experiments [6-8] can be summarized as:

- Self-focusing in space
- Self-compression in time

$\Longrightarrow$ complete space-time trapping of the light matter

- Improvement of the spatio-temporal pulse shape
- Increasing of the peak intensity
- Stable propagation of the pulse along many characteristic lengths (L_{NL}) of the most intense pulse rearrangement factor, in this case – the nonlinearity.

The pulse behavior under consideration cannot be understood within the (3+1)D cubic NLSE because no stable propagation of the pulse in positive GVD medium can be predicted [4, 5, 14, 15]. Above-cubic nonlinearity and ionization have been proposed as possible pulse stabilization factors [7].

3.1. Spatiotemporal dynamics of high-intnesity femtosecond laser pulses in bulk nonlinear media in presence of ionization - *numerical studies*

To describe the specified experimental behavior of the high-intensity femtosecond laser pulses, *an advanced physical model* [16] including the ionization will be used. The propagation equation in this model is (3+1)D nonlinear envelope equation (NEE) (in standard notations [16]), Eq.(4)

$$\begin{aligned}\frac{\partial A}{\partial z} = & \frac{i}{2k_0}T^{-1}\nabla_\perp^2 A + iDA \\ & + i\frac{\omega_0}{c}n_2 T\left[(1-x)\left|A\right|^2 + x\int_{-\infty}^{t} h(t-t')\left|A(t')\right|^2 dt'\right]A - i\frac{\omega_0}{c}n_4 T\left|A\right|^4 A \\ & - i\frac{k_0}{2n_0^2\rho_c}T^{-1}\rho A - \frac{\sigma}{2}\rho A - \frac{\beta_{MPI}(A)}{2}A\end{aligned} \tag{4}$$

Each physical process involved in the model is accounted by a respective term in the NEE. The first and the second terms in the first row of the NEE describe the diffraction and dispersion, respectively. The third and the forth terms in the second row describe the cubic and the quintic nonlinearities that result from the modification of the refractive index of the neutral particles. The cubic term includes both, the instantaneous and the non-instantaneous response of the medium. The fifth, sixth and the seventh terms in the third row of the NEE account for the contribution of the ionization and describe the ionization modification of the refractive index, collisional ionization by inverse bremsstrahlung and multi-photon/tunneling ionization losses, respectively. The dispersion operator $D = \sum_{n\geq 2}^{\infty}\left(\frac{k^{(n)}}{n!}\right)\left(i\partial_t\right)^n$ in our case is truncated after the cubic term, *i.e.*, $D = \frac{k''}{2!}i^2\partial_t^2 + \frac{k'''}{3!}i^3\partial_t^3$.

The propagation of the high-intensity femtosecond pulses is inherently accompanied by ionization. The plasma creation and recombination processes are described by the electron number density kinetic equation (5), [16]

$$\frac{\partial \rho}{\partial t} = W(I)(\rho_n - \rho) + \frac{\sigma(\omega_0)}{I_P}\left|A\right|^2 \rho - f(\rho) \tag{5}$$

The main problem in the description of the electron number density ρ is the correct determination of the ionization rate $W(I)$. Among the general ionization theories, Perelomov-Popov-Terent'ev (PPT) theory [17] gives the best description of the ionization rate at both, the multiphoton and the tunnel ionization regimes. The PPT ionization rate (in standard notations) is given by

$$W_{PPT}(t) = \sqrt{\frac{6}{\pi}}\, I_P \left| C_{n^*l^*} \right|^2 \frac{(2l+1)(l+|m|)!}{2^{|m|}|m|!(l-|m|)!}\left(\frac{2(2I_P)^{3/2}}{E(t)}\right)^{2n^*-|m|-3/2}\left(1+\gamma^2\right)^{|m|/2+3/4} \times A_m(\omega,\gamma)\exp\left(-\frac{2(2I_P)^{3/2}}{3E(t)}g(\gamma)\right) \tag{6}$$

In some cases, however, a direct fit of the experimental data of the ionization of given atom/molecule in terms of multiphoton ionization rate [16] may lead to better result, Eq. (6').

$$W_{MPI} = \sigma_K I^K \tag{6'}$$

Both approaches have been used to account for the ionization rate in our studies.

We have further developed the advanced physical model [16] including the influence of ionization on the GVD, $k'' = \partial^2 k/\partial\omega^2$, recently predicted in [18]. The total GVD can be presented as a sum of two terms, the GVD of the neutrals k_0'' and the GVD of the free electrons k_i''

$$k'' = k_0'' + k_i''. \tag{7}$$

The GVD of the free electrons (standard notations) is [18]

$$k_i'' = -\frac{e^2\lambda^3 N_e}{2\pi^2 m_e c^4\left(1-\frac{e^2\lambda^2 N_e}{\pi m_e c^2}\right)^{3/2}} \tag{8}$$

The high-intensity femtosecond pulse propagation is described solving self-consistently Eqs. (4)-(8). We consider laser pulses of 2mJ pulse energy and 150 fs time duration propagating in pressurized argon of 5atm pressure.

The evolution of the temporal profile of the laser pulses with the propagation at the specified conditions is shown in Fig. 6. The pulse propagation dynamics closely resembles the experimental behavior, shown in Fig. 5. The numerical simulations reproduce the initial pulse compression from $z = 0\,cm$ to $z = 5.9\,cm$, the stable propagation of the compressed pulse from $z = 5.9\,cm$ to $z = 8.2\,cm$, and, finally, the splitting of the pulse at $z = 16.8\,cm$. The degree of the stabilization of the pulse in space and in time can be better illustrated by inspection of the transversal width and the time duration of the pulse. The variation of the transversal width of the pulse, the beam radius, with the propagation distance is shown in Fig. 7. As can be seen, the pulse rapidly collapses down to about $90\mu m$ beam diameter due to the self-focusing and then continues propagating keeping the beam size almost stable. The combined effect of self-focusing in the transversal direction and the slow dispersion expansion in the longitudinal direction (due to the low GVD of the medium) results in a rapid increase of the peak intensity. This is accompanied by a pulse compression, Fig. 8,

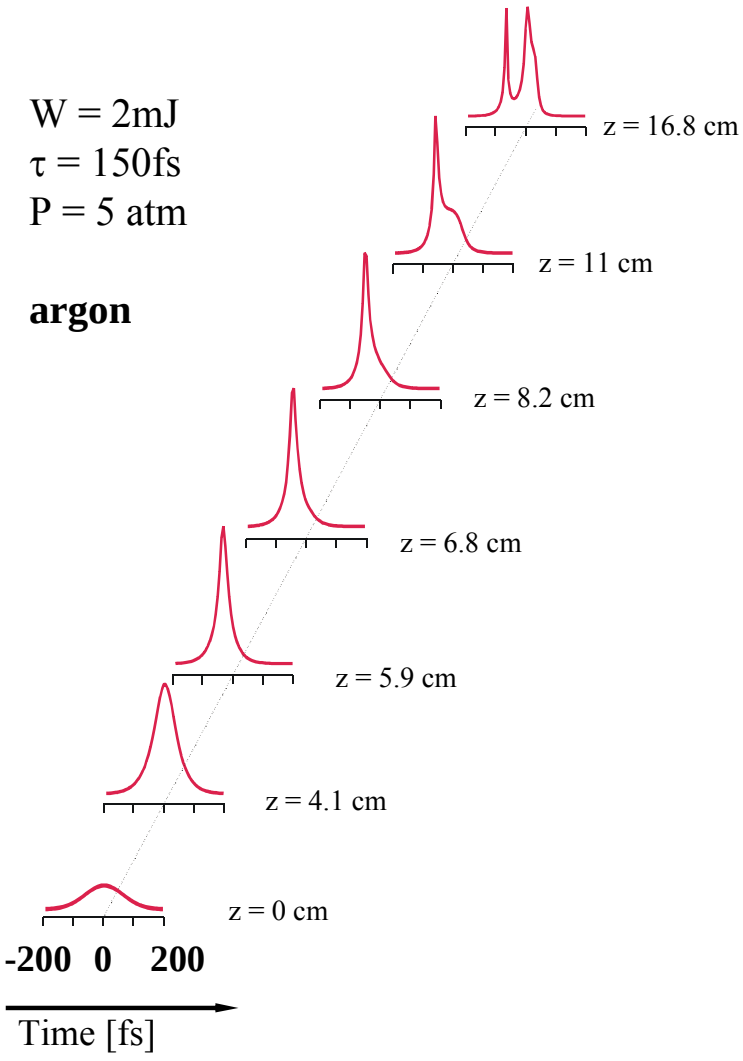


Fig. 6. Propagation dynamics of 150 fs pulses in pressurized argon

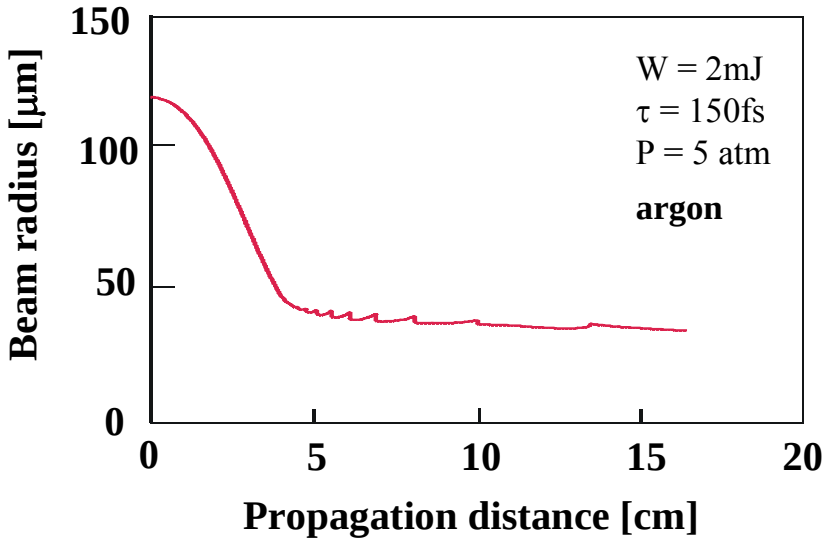


Fig. 7. Evolution of beam radius with the propagation distance

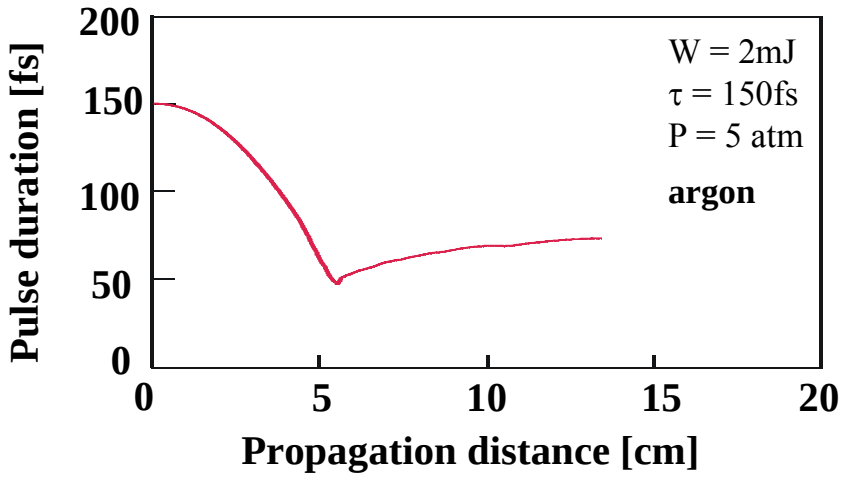


Fig. 8. Evolution of pulse duration with the propagation distance

in agreement with the pulse compression mechanism specified in Sec.2. The present simulations confirm the strong ionization modification of the GVD, predicted in [18], leading even to inversion of the GVD along given propagation distance. This is the first verification of the ionization induced inversion of the GVD in a dynamical propagation regime.

While the stabilization of the beam width is rather good, the stabilization of the pulse duration is not yet quite satisfactory. Nevertheless, including the ionization results in an apparent stabilization of the pulse propagation in comparison to that one below the ionization "threshold", Sec. 2. The stabilization of the pulse propagation observed here results from few balance mechanisms: (*i*) balance of the self-focusing by the ionization modification of the refractive index, (*ii*) balance of the intensity growth by the ionization losses and (*iii*) balance of the dispersion broadening by the ionization contribution to the group velocity dispersion. We believe that the present results will serve as a base of further improvement of the physical conditions for pulse compression and pulse stabilization.

3. CONCLUSIONS

The spatio-temporal dynamics of high-intensity femtosecond laser pulses above the ionization "threshold" has been studied using an advanced physical model. A further development of the physical model has been achieved taking into account the ionization modification of the group velocity dispersion. The spatio-temporal behavior of the laser pulse has been studied. A complete spatio-temporal compression of the pulse and relatively stable pulse propagation has been found from the simulations. A qualitatively agreement between the numerically predicted results and the experimentally observed pulse behavior is found. The present results will allow further improvement of the physical conditions of pulse compression and pulse stabilization and better understanding of the pulse propagation dynamics.

Acknowledgment: The contribution of M. D. Todorov, M. E. Todorova and T. P. Todorov to this report is gratefully acknowledges.

REFERENCES

[1] S. Trillo and W. Torruellas (Eds.), *Spatial solitons* (Springer-Verlag, Berlin 2001)

[2] G. P. Agrawal, *Nonlinear Fiber Optics* (Academic Press, New York, 1994)

[3] N. Akhmediev, A. Ankiewicz (Eds), *Dissipative solitons* (Springer-Verlag, Berlin 2005)

[4] I. G. Koprinkov, M. D. Todorov, M. E. Todorova, T. P. Todorov, J. Phys. B: At. Mol. Opt. Phys. **40,** F231 (2007)

[5] M. D. Todorov, M. E. Todorova, T. P. Todorov, I. G. Koprinkov, Opt. Commun. **281** , 5249 (2008)

[6] I. G. Koprinkov, A. Suda, P. Wang, and K. Midorikawa, Jap. J. Appl. Phys., L 978 (1999).

[7] I. G. Koprinkov, A. Suda, P. Wang, and K. Midorikawa, Phys. Rev. Lett., **84**, 3847 (2000).

[8] I. G. Koprinkov, A. Suda, P. Wang, and K. Midorikawa, "Soliton-Driven Photonics", Kluwer Academic Publishers, Ed. A. D. Boardman and A. P. Sukhorukov, pp. 355-358, (2001).

[9] S. Tzortzakis, B. Lamouroux, A. Chiron, S.D. Moustaizis, D. Anglos, M. Franco, B. Prade, and A. Mysyrowicz, Opt. Commun. **197**, 131 (2001).

[10] N. L. Wagner, E. A. Gibson, T. Popmintchev, I. P. Christov, M. M. Murnane, H. C. Kapteyn, Phys. Rev. Lett., **93**, 173902 (2004).

[11] Couairon, M. Francko, A. Mysyrowics, J. Biegert, and U. Keller, Opt. Lett., **30**, 2657 (2005).

[12] S. Skupin, G. Stibenz, L. Berge, F. Lederer, T. Sokollik, M. Schnürer, N. Zhavoronkov, and G. Steinmeyer, Phys. Rev. E **74**, 056604 (2006).

[13] T. Brabec and F. Krausz, Rev. Mod. Phys., 72, 545 (2000)

[14] J. K. Ranka, R. W. Schirmer, and A. L. Gaeta, Phys. Rev. Lett., Vol. **77**, 3783 (1996)

[15] S. A. Diddans, H. K. Eaton, A. A. Zozulya, and T. S. Clement, IEEE, J. Selected Topics of Quant. Electron., Vol. **4**, 306-316 (1998)

[16] L. Berge, S. Skupin, R. Nuter, J. Casparian, and J-P Wolf, Rep. Prog. Phys.. **70**, 1633 (2007)

[17] A. M. Perelomov, V. S. Popov, M. V. Terent'ev, Sov. Phys. JETP, **23**, 924 (1966)

[18] I. G. Koprinkov, Applied Physics B - Lasers and Optics, **79**, 359 (2004).

INDUCTIVE HEATING OF HYDROGEN PLASMA

Khristo Tarnev[1], Tsanko Tsankov[2]

[1]Department of Applied Physics, Technical University-Sofia, 8, Kliment Ohridski St.,
BG-1000 Sofia, Bulgaria, phone: +359-2-965-3110, e-mail: tarnev@tu-sofia.bg
[2]Faculty of Physics, Sofia University, 5 James Bourchier Blvd.,
BG-1164 Sofia, Bulgaria phone: +359-2-8161-643

Abstract: *The study presents a two dimensional self-consistent model of an inductive discharge in hydrogen. The equations describing the plasma behavior are solved together with Maxwell's equations for the radiofrequency electromagnetic field sustaining the discharge. The stress in the presentation of the results is on the spatial distribution of the electromagnetic field components. The mechanism of power deposition to the plasma is analyzed. The results show that the metal walls of the plasma vessel and the different dimensions in longitudinal and axial direction lead to establishment of a discharge regime with dc current in the plasma – a regime much different from the well known ambipolar diffusion regime*

Keywords: electromagnetic waves in plasma, inductive heating

1. INTRODUCTION

The active research in the field of plasma and gas discharges in the recent years is motivated by their numerous applications. Among some of the most important plasma technologies [1] are thin film deposition, surface processing, ion implantation, plasma-based lighting systems, plasma chemistry, flat-panel displays, environmental and health applications. Another basic direction of the research is the thermonuclear fusion offering the potential of an almost limitless source of energy for future generations.

Depending on the specific technological requirements different types of plasma sources are developed. Efficient plasma sources widely used nowadays – both in the technological reactors and in the fusion plasmas – are the inductive discharges [1, 2]. Mainly two modifications are employed – inductive discharges with cylindrical coil and with a planar coil. In both cases the plasma is heated by the electric field induced by the oscillating magnetic field created by the current in the coil.

Optimization of the sources requires good understanding of the processes in the plasma and of the mechanisms of power deposition leading to fast development of the plasma source modeling [3-7].

The paper presents a 2D model of the driver – the region of power deposition – of hydrogen discharge in diffusion controlled regime.

2. BASIS OF THE MODEL

The particles in the model are electrons e^-, three types of positive hydrogen ions (H^+, H_2^+ and H_3^+), atoms H and molecules H_2. For all the charged particles and for the atoms the continuity equation is solved:

$$\frac{\partial n_j}{\partial t} + \nabla \cdot \vec{\Gamma}_j = \frac{\delta n_j}{\delta t}. \tag{1}$$

Here n_j are the concentrations of the corresponding type of particles, $\Gamma_j = -D\nabla n_j + n_j v_j$ are their fluxes, taking into account the diffusion and the drift in the dc field and $\delta n_j / \delta t$ are the changes in the concentration due to inelastic collisions. The elementary processes involved in the model are

1. Ionization of molecules and atoms

$$\mathrm{e} + \mathrm{H}_2(X\,^1\Sigma_g^+) \rightarrow \mathrm{e} + \mathrm{H}_2^+(\mathrm{v}) + \mathrm{e}$$
$$\mathrm{e} + \mathrm{H}(1s) \rightarrow \mathrm{e} + \mathrm{H}^+ + \mathrm{e}$$

2. Dissociation of molecules and ions

$$\mathrm{e} + \mathrm{H}_2(X\,^1\Sigma_g^+) \rightarrow \mathrm{e} + \mathrm{H}_2^*(b\,^3\Sigma_u^+, a\,^3\Sigma_g^+, c\,^3\Pi_u) \rightarrow \mathrm{e} + 2\mathrm{H}(1s)$$
$$\mathrm{e} + \mathrm{H}_2^+(\mathrm{v}) \rightarrow \mathrm{e} + \mathrm{H}_2^{+*}(2p\sigma_u) \rightarrow \mathrm{e} + \mathrm{H}^+ + \mathrm{H}(1s)$$
$$\mathrm{e} + \mathrm{H}_3^+ \rightarrow \mathrm{e} + \mathrm{H}^+ + 2\mathrm{H}$$

3. Heavy particle collisions

$$\mathrm{H}_2^+ + \mathrm{H}_2 \rightarrow \mathrm{H}_3^+ + \mathrm{H}$$

4. Recombination

$$\mathrm{e} + \mathrm{H}_3^+ \rightarrow \mathrm{H} + \mathrm{H} + \mathrm{H}$$

In the electron-energy balance equation:

$$\frac{3}{2}\frac{\partial}{\partial t}(n_\mathrm{e} T_\mathrm{e}) + \nabla \cdot \vec{J}_\mathrm{e} = Q - P_{(\mathrm{coll})} - e\vec{\Gamma}_\mathrm{e} \cdot \vec{E}_\mathrm{dc} \tag{2}$$

$\vec{J}_\mathrm{e}$ is the energy flux connected to the thermal conductivity and the energy carried by the electron flux, Q is the density of the externally applied power, $P_{(\mathrm{coll})}$ are the energy losses in elastic and inelastic collisions and the term $e\vec{\Gamma}_\mathrm{e} \cdot \vec{E}_\mathrm{dc}$ presents the energy losses or gain due to directed electron motion in the dc field. In the calculation of Q the results for the induced electric field from the electrodynamical part of the model are used.

The gas discharge part of the model is completed by the Poisson equation

$$\Delta\Phi = -\frac{e}{\varepsilon_0}\left(\sum_{j=1}^{3} n_{\mathrm{ij}} - n_\mathrm{e}\right) \tag{3}$$

and the gas-pressure equation

$$p = (n_{\mathrm{H}} + n_{\mathrm{H}_2})\kappa T_{\mathrm{g}}\,. \tag{4}$$

Here Φ is the dc potential ($\vec{E}_{\mathrm{dc}} = -\nabla\Phi$), κ is the Boltzmann constant and T_{g} is the neutral gas temperature.

The boundary conditions are for symmetry on the axis, particle fluxes on the walls and zero potential on the walls. The last condition corresponds to grounded metal walls and is in conformity with the Faraday shields, often used in plasma sources.

In the electrodynamical part of the model the Maxwell's equations results in an equation for the vector potential $\vec{A}$:

$$\frac{1}{\mu_0}\nabla\times(\nabla\times\vec{A}) + (i\omega\sigma - \omega^2\varepsilon_0)\vec{A} = 0\,. \tag{5}$$

In the expression for the conductivity σ the electron concentration from the gas discharge part of the model is used. ε_0 and μ_0 are the electric and the magnetic constants and $\omega = 2\pi f$ is the angular frequency. The components of the electric and magnetic field are calculated from the vector potential:

$$\vec{E} = i\omega\vec{A} \tag{6}$$

$$\vec{B} = \nabla\times\vec{A} \tag{7}$$

In our case (cylindrical geometry with axial symmetry) only the azimuthal components of the vector potential A_φ and electric field E_φ and axial B_z and radial B_r components of the magnetic field are nonzero.

The current in the coil is simulated by a surface current with appropriate axial distribution used as boundary condition in equation (5).

The equations of the model are solved in iterations. First, the gas discharge part is solved with constant external heating. The result for the electron concentration is substituted in the expression for the plasma conductivity and the induced electric field is obtained. At the next step the density of the externally applied power Q is calculated with this electric field. The iterations are repeated until convergence is reached.

3. RESULTS

In this section are presented results for gas discharge in hydrogen at pressure 30 mTorr, externally applied power 500 W and driving frequency 27 MHz.

3.1. Spatial distribution of the plasma parameters

Figure 1 shows the spatial distribution of the some of the plasma parameters and figure 2 – of the electromagnetic characteristics. The external coil, not shown on the figures, is between *z* = 7.5 cm and *z* = 17.5 cm.

Figure 1(a) shows the electron concentration. As it is typically for diffusion controlled regime the maximum of the concentration is on the axis under the coil.

Because of the high thermal conductivity the electron temperature is almost constant in radial direction (Figure 1(b)). The maximum of the temperature is under the coil near the wall. There the electric field has a maximum (Figure 2(a)) while the electron concentration is low. Therefore, the energy gained by an electron from the field is high. This energy is transferred to the axis by the thermal conductivity and to the regions outside the coil by the thermal conductivity and the electron flux (arrows on Figure 1(b)).

The dc potential (Figure 1(c)) has a maximum on the axis like the electron concentration, decreasing in axial and radial direction. Because of the different axial and radial dimensions and the constant potential on the walls ($\Phi = 0$ in the model) the radial dc electric field is stronger than the axial one. The electric field is accelerating for the positive ions and retarding for the electrons. Therefore, for the electrons it is easier to leave the discharge in axial direction while the ions move dominantly in *r*-direction. As a result, there is a dc current in the discharge (Fig 1(d)).

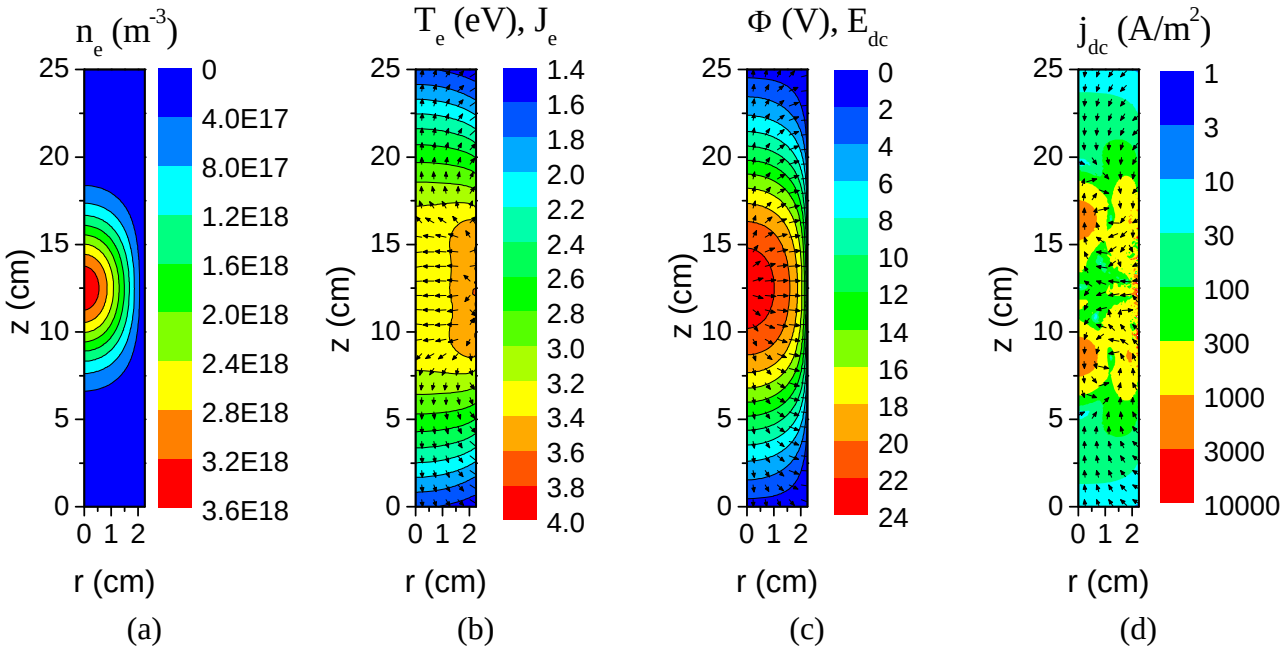


Fig. 1. Spatial distribution of the plasma parameters: electron concentration (a); electron temperature and energy flux (b); dc potential and dc electric field (c) and dc current (d).

3.2. RF field components and induced current

The spatial distribution of the components of the induced electromagnetic field is determined by the profile of the electron concentration and of the current in the coil.

The rf electric field has a maximum under the coil on the wall of the discharge tube (Figure 2(a)). Local maxima corresponding to each turn of the coil could be seen on the figure. Because of the skin effect the electric and magnetic field decrease fast in the plasma. The induced current is shifted to the axis (Figure 2(b)). The explana-

tion of this shift is given on Figure 2(c) with the normalized radial profile at the center of the discharge for the plasma conductivity, HF electric field and the induced current. The plasma conductivity following the electron concentration decreases almost to zero on the wall. In opposite, the HF electric field decreases to zero at $r = 0$. Their product, the induced current $j_\varphi = \sigma E_\varphi$ is zero both on the axis and on the wall and has maximum at approximately $r = 1.5$ cm.

3.3. Propagation of the electromagnetic energy

The energy from the coil propagates towards the axis as a damped transverse electromagnetic wave. The amplitude and the direction of the Poynting vector *S* are shown in Figure 3(a). The amplitude of *S* decreases towards the axis because of the power absorption in the plasma. The time averaged magnitude of *S* is given by $\vec{S} = \vec{E} \times \vec{H} \cos\phi$ with ϕ being the phase difference between the electric and magnetic field. It changes in radial direction (Figure 3(b)) starting from values bigger than 90° in the outer part of discharge to exactly 90° on the axis. Phase difference bigger than 90° means $\cos\varphi < 0$, i.e. the electromagnetic energy flux is directed towards smaller *r*-values to the discharge axis.

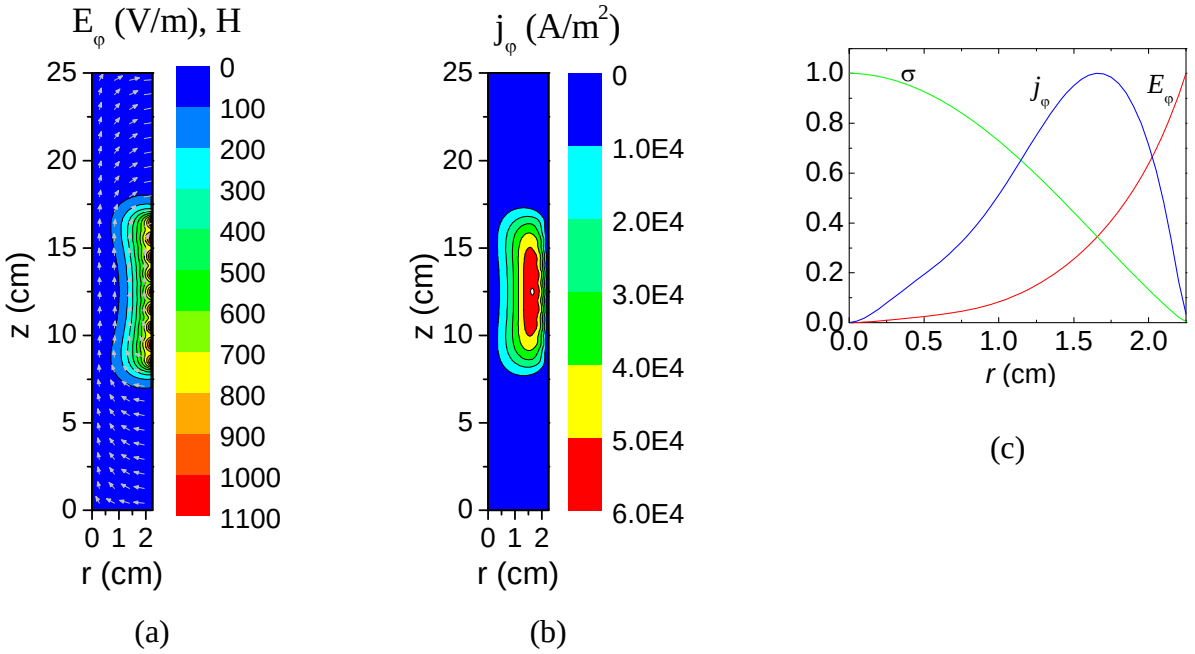
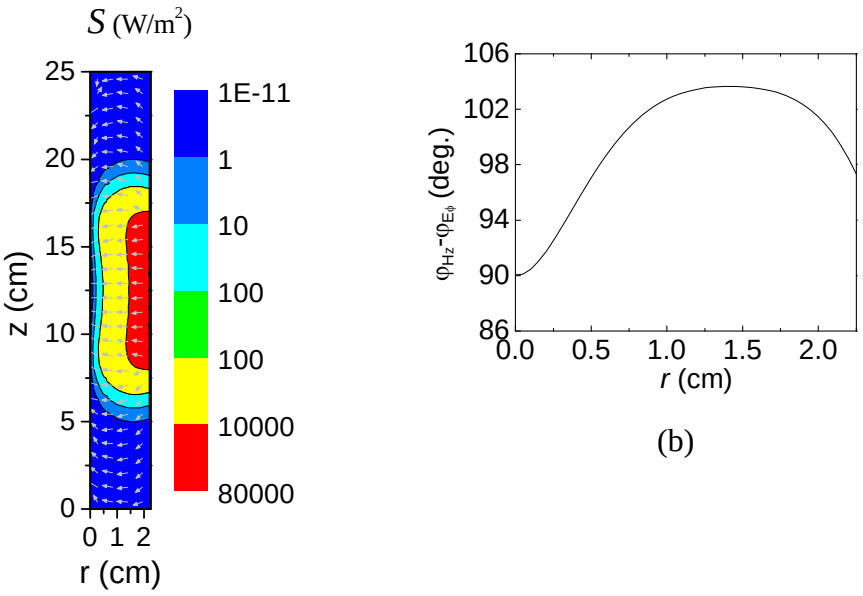


Fig. 2. Spatial distribution of the induced rf electric field intensity (color) and magnetic field intensity (arrows) in (a); induced rf current in (b) and normalized radial profiles at *z* = 12.5 cm of plasma conductivity (green), rf electric field intensity (red) and induced rf current (blue) in (c).



(a) (b)

Fig. 3. Magnitude and direction of the energy flux (Poynting vector) (a) and radial profiles at z = 12.5 of the phase difference between magnetic and electric field components (b).

Figure 4 illustrates the propagation of the electromagnetic energy in the discharge tube.

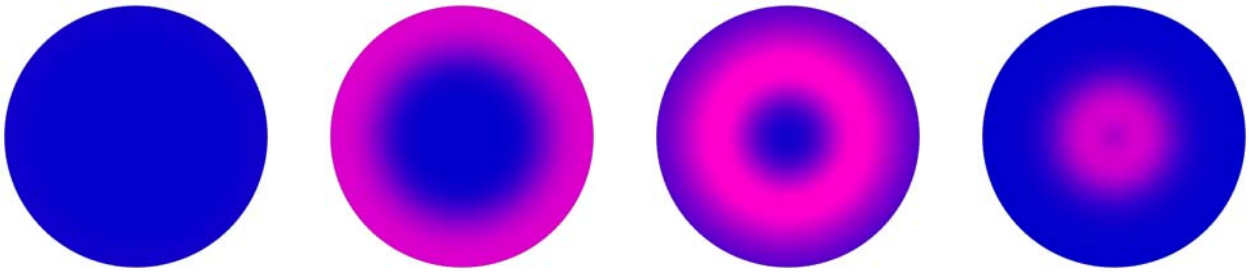


Fig. 4. Normalized spatial distribution of the energy flux at $\omega t = 0^\circ$, $\omega t = 40^\circ$, $\omega t = 80^\circ$ and $\omega t = 120^\circ$.

4. CONCLUSIONS

In this paper a 2D model of the driver region of an inductive gas discharge in hydrogen is presented. The properties of the electromagnetic field sustaining the discharge and of the plasma parameters are self-consistently determined. The results show maximum values of the rf field and of the charged particle concentrations under the external coil. The different dimensions in radial and axial direction and the equipotential (metal) walls of the discharge tube causes appearance of dc current in rf discharge.

ACKNOWLEDGEMENTS

The work is within project ID01/029-DO02-267 supported by the National Science Fund in Bulgaria and it is part of the work on task P2 of the Bulgarian Association EURATOM/INRNE.

REFERENCES

[1] M. A. Lieberman and A. J. Lichtenberg, Principles of Plasma Discharges and Materials Processing, John Wiley and Sons, New York, 2005.

[2] J. Hopwood “Review of inductively coupled plasmas for plasma processing”, Plasma Sources Sci. Technol vol. 1, pp. 109-116, 1992.

[3] P. L. G. Ventzek, R. J. Hoekstra and M. J. Kushner, "2-Dimensional modeling of high plasma density inductively coupled sources for materials processing", J. Vac. Sci. Tech. B. vol. 12, p.461, 1994.

[4] St. Kolev, H. Schlüter, A. Shivarova and Kh. Tarnev, “A diffusion-controlled regime of cylindrical inductive discharges”, Plasma Sources Sci. Technol. vol.15, pp. 744–756, 2006.

[5] Ts. Paunska, H. Schlüter, A. Shivarova and Kh. Tarnev, “Low pressure hydrogen discharges”, Physics of Plasmas, vol. 13, 023504, 2006.

[6] St. Kolev, Ts. Paunska, A. Shivarova and Kh. Tarnev, “Low-pressure inductive discharges”, Journal of Physics: Conference Series, vol. 63, 012006, 2007.

[7] Ts. Paunska, A. Shivarova Kh. Tarnev and Ts. Tsankov, “Spatial distribution of the plasma characteristics of a tandem plasma source”, Proceedings of the 1st International Symposium on Negative Ions, Beams and Sources, Aix-en-Provence, France, 9-12 September 2008, AIP Conference Proceedings , vol. 1097 pp. 99-108, 2009.

TRANSLINEAR CIRCUIT REALIZATIONS OF STATE-SPACE EQUATIONS

Georgi A. Nenov*, Vassia G. Stoyantcheva*

* Higher School of Transport "T. Kableshkov", Sofia
1574 Sofia, 158 Geo Milev Str. BULGARIA,
phone: +359 2 9709 232
gnenov1@gmail.com , vasiags@abv.bg

Abstract. *The paper deals with some aspects and approaches of translinear circuit theory concerning the realization of state-space equations on the base of the static and dynamic translinear principles. The results in this topic lead to a variety circuit realizations of frequency selective translinear circuits – log-domain, square-root-domain, sinh-domain etc.*

Keywords: electrical circuits, translinear circuits, log-domain filters

1. INTRODUCTION

The last three decades mark a significant research interest to the theory and design of translinear circuits. This fact can be explained mainly due the suitability of these circuits for low voltage, low power, large dynamic range, high frequency applications [5]. Commonly the translinear circuits realize linear relationships between their input and output ports, independently of nonlinear relationships between some internal circuit variables. For this purpose one uses the exponential or square-root characteristics of BJT transistors and MOS transistors in sub-threshold region [1, 2]. The basic practical application of translinear circuits is for filter realizations – log-domain, sinh-domain, tanh-domain and square-root domain filters [5].

A very often used method for the synthesis of active filters is based on the state-space presentation of their characteristics but the same approach can be applied to the realization of translinear filters too [2 -6]. The circuit implementation of the common form of well known state-space equation usually uses some simple dynamic translinear subcircuits consisting of BJ transistors and capacitors [3, 4, 6].

2. DYNAMIC TRANSLINEAR PRINCIPLE (DTLP)

Briefly the definition of DTLP is [6]: *The derivative of a BJ transistor collector current can be expressed by a relation of products of currents.* We will prove this statement firstly on the base of the Fig. 1. For the capacitor current it holds

$$i_C(t) = C\frac{dv_C(t)}{dt} = C\frac{du_{BE}(t)}{dt}, \tag{1}$$

for

$$u_{BE}(t) = u_C(t) + E_{const}. \tag{2}$$

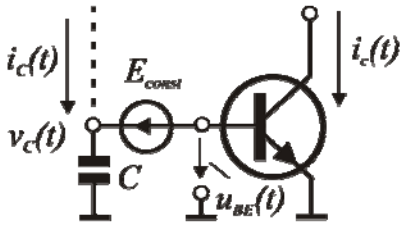


Fig. 1

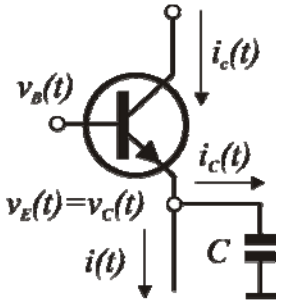


Fig. 2

Having in mind the exponential relationship between the collector current $i_c(t)$ and the voltage $u_{BE}(t)$ [4]

$$i_c(t) = I_s.e^{u_{BE}(t)/U_T}, \tag{3}$$

where I_s is the saturation current and U_T is the thermal voltage of the transistor one obtains

$$u_{BE}(t) = U_T\left(\ln i_c(t) - \ln I_s\right). \tag{4}$$

By substituting the last expression in (1) it follows the first-order nonlinear differential equation [11-13]

$$\frac{di_c(t)}{dt} = \frac{i_C(t)}{CU_T} i_c(t). \tag{5}$$

This expression is a part of the well known scalar form of a state-space equation

$$\frac{dx(t)}{dt} = ax(t) + bf(t), \tag{6}$$

where $x(t)$ is the state variable, a and b are constants and $f(t)$ is an independent variable.

An other example of translinear sub-circuit which satisfies an equation similar to (5) is given in Fig. 2 [4, 6]. Here we have again

$$i_c(t) = I_s e^{\left(v_B(t)-v_E(t)\right)/U_T} \tag{7}$$

and

$$\frac{dv_E(t)}{dt} = \frac{dv_C(t)}{dt} = \frac{1}{C} i_C(t) = \frac{1}{C}\left(i_c(t) - i(t)\right). \tag{8}$$

From (6) and (7) for the first derivative of $I_c(t)$ one obtains

$$\frac{di_c(t)}{dt} = \frac{1}{U_T}\left(\frac{dv_B(t)}{dt} + \frac{i(t)}{C}\right) i_c(t) - \frac{i_c(t)^2}{CU_T}. \tag{9}$$

This expression is a nonlinear differential Bernoulli's equation [10 - 12]. However by the substitution $i_c(t)=1/\tau(t)$ it can be transformed into a linear relationship

$$\frac{d\tau(t)}{dt}+\frac{1}{U_T}\left(\frac{dv_B(t)}{dt}+\frac{i(t)}{C}\right)\tau(t)-\frac{1}{CU_T}=0 \tag{10}$$

Here a problem arises: How to express the variable $\tau(t)$ by some circuit currents, because it is not a current but an inverse current ? The decision is to include the collector current $i_c(t)= 1/\tau(t)$ in a loop, which satisfies the *static translinear principle(STLP)* [2, 10]. Let this loop consists of m clockwise oriented and m counterclockwise oriented identical transistor base-emiter junctions. Then the following relationship between the transistor collector currents holds [10]

$$\prod_{k=1}^{m} i_{c,k}(t)_{clockwise} = \prod_{k=1}^{m} i_{c,k}(t)_{counterclockwise}\,. \tag{11}$$

If for example the current $i_c(t)$ is the clockwise current $i_{c,j}(t)$ of *j-th* junction from (11) one obtains

$$\tau(t)=\frac{1}{i_{c,j}(t)}=\frac{\prod_{\substack{k=1\\k\neq j}}^{m} i_{c,k}(t)_{clockwise}}{\prod_{k=1}^{m} i_{c,k}(t)_{counterclockwise}}. \tag{12}$$

3. STATE VARIABLES CHOICE

The approaches to the realization of translinear filters published in the literature use capacitor voltages $v_{C,i}(t)$ as memories of the circuit or currents as state variables [6]. However the voltages $v_{C,i}(t)$ are nonlinearly related to the state variables $x_i(t)$:

$$x_i(t)=f_i\left(v_{C,i}(t)\right), \tag{13}$$

On the other hand the currents $i_{x,i}(t)$ are linearly related to the state variables $x_i(t)$. Then for the capacitance currents $i_{C,i}(t)$ it holds

$$i_{C,i}(t)=g_i\left(i_{x,i}(t),\frac{di_{x,i}(t)}{dt}\right). \tag{14}$$

In (13) and (14) the functions f_i and g_i are usually *exp, tanh, sinh* or *square-root*.

The generalized form of (14) reflects the dependence of every capacitance current from the vectors of state variables $\mathbf{i}_{x,i}(t)$ and from their first derivatives

$$i_{C,i}(t) = g_i\left(\mathrm{i}_{x,i}(t), \frac{d\mathrm{i}_{x,i}(t)}{dt}\right). \tag{15}$$

In a similar manner the equation (13) gives

$$x_i(t) = f_i\left(\mathrm{V}_{C,i}(t)\right), \tag{16}$$

where $\mathbf{V}_{C,i}(t)$ is the capacitance voltage vector.

4. SOME ASPECTS OF DYNAMIC TRANSLINEAR CIRCUIT REALIZATION

As an example we analyze the circuits in Fig. 3 and Fig. 4 where $i_x(t)$ and $i_{out}(t)$ are state variables [6]. For the first of them according STLP we have

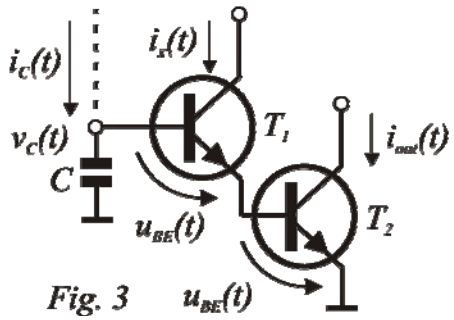


Fig. 3

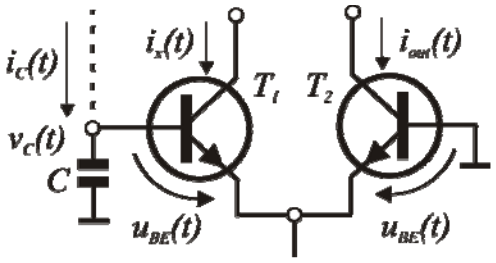


Fig. 4

$$\left.\begin{aligned} i_x(t) &= \frac{I_s^{\,2}}{i_{out}(t)} e^{\frac{v_c(t)}{U_T}}; \\ i_{out}(t) &= \frac{I_s^{\,2}}{i_x(t)} e^{\frac{v_c(t)}{U_T}}; \\ i_C(t) &= CU_T\left(\frac{\frac{di_x(t)}{dt}}{i_x(t)} + \frac{\frac{di_{out}(t)}{dt}}{i_{out}(t)}\right) \end{aligned}\right\}, \tag{17}$$

whereas for the second one the relationships hold

$$\left.\begin{aligned} i_x(t) &= i_{out}(t)e^{\frac{v_C(t)}{U_T}}; \\ i_{out}(t) &= i_x(t)e^{-\frac{v_c(t)}{U_T}}; \\ i_C(t) &= CU_T\left(\frac{\frac{di_x(t)}{dt}}{i_x(t)} - \frac{\frac{di_{out}(t)}{dt}}{i_{out}(t)}\right) \end{aligned}\right\}. \tag{18}$$

Let us constitute the state-space equations of a second-order *LC*- resistive loaded filter given in Fig. 5. Here we have

$$\left|\begin{aligned} &i(t) - C\frac{du_{out}(t)}{dt} - \frac{u_{out}(t)}{R} = 0; \\ &L\frac{di(t)}{dt} + u_{out}(t) - u_{in}(t) = 0 \end{aligned}\right. \tag{19}$$

and after some rearranging the state-space equations follow

$$\left|\begin{aligned} \frac{di(t)}{dt} &= -\frac{1}{L}u_{out}(t) + \frac{1}{L}u_{in}(t); \\ \frac{du_{out}(t)}{dt} &= \frac{1}{C}i(t) - \frac{u_{out}(t)}{RC} \end{aligned}\right., \tag{20}$$

where $i(t)$ and $u_{out}(t)$ are the new state variables.

Because the aim is a current-mode model of the circuit in Fig. 5 to be synthesized one replaces formally all voltages in (21) by corresponding currents and consequently

$$\left|\begin{aligned} \frac{di(t)}{dt} &= a_{12}\left(-i_{out}(t) + i_{in}(t)\right); \\ \frac{di_{out}(t)}{dt} &= a_{21}i(t) - a_{22}i_{out}(t) \end{aligned}\right., \tag{21}$$

where a_{12}, a_{21} and a_{22} are constants and $i(t)$ and $i_{out}(t)$ are the new state variables.

The number of state-variables in (21) must correspond to the number of the capacitors in the searched structure – i.e. two. Next we assume that the synthesis will be based on the circuits in Fig. 1 and Fig. 2. For $C=C_1$ from the third equation in (17) and for $C=C_2$ from the third equation in (18) we express the derivatives $di(t)/dt$ and $di_{out}(t)/dt$ and by substituting them in (21) one obtains

$$\left.\begin{aligned} i(t)\left(i_{C1}(t)+i_{C2}(t)\right) &= I_{12}\left(-i_{out}(t)+i_{in}(t)\right); \\ i_{out}(t)\left(i_{C1}(t)-i_{C2}(t)\right) &= \left(I_{21}i(t)-I_{22}i_{iout}(t)\right) \end{aligned}\right\}. \tag{22}$$

Here

$$\left.\begin{aligned} I_{12} &= 2C_1U_T a_{12}; \\ I_{21} &= 2C_2U_T a_{21}; \\ I_{22} &= 2C_2U_T a_{22} \end{aligned}\right\} \tag{23}$$

denote additional constant currents.

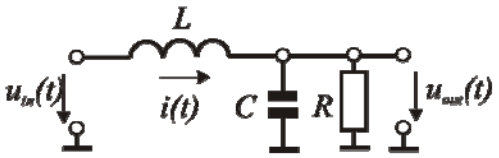


Fig. 5

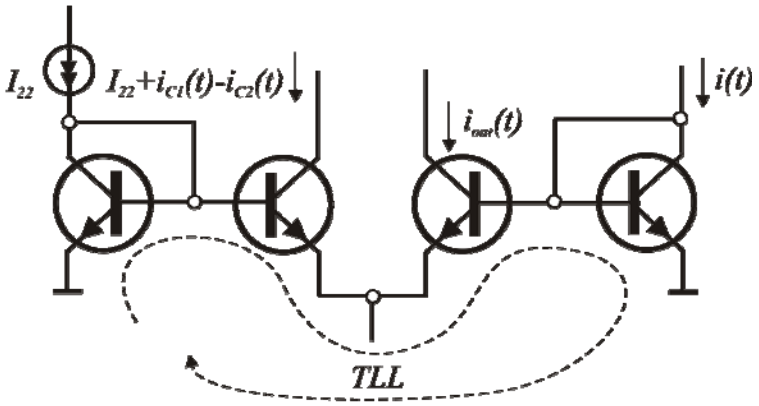


Fig. 6

By adding the term $I_{12}i(t)$ in both sides of the first equation in (22) and the term $I_{22}i_{out}(t)$ in both sides of the second equation we reach to

$$\left.\begin{aligned} i(t)\left(I_{12}+i_{C1}(t)+i_{C2}(t)\right) &= I_{12}\left(-i_{out}(t)+i_{in}(t)+i(t)\right); \\ i_{out}(t)\left(I_{22}+i_{C1}(t)-i_{C2}(t)\right) &= I_{21}i(t) \end{aligned}\right\}. \tag{24}$$

Evidently the last two equations contain products of currents and this fact suggests the use of STLP in the synthesis procedure. The variables that take part in (24) must be positive collector currents of transistors in suitable translinear loops, whereas the currents I_{12}, i_{21} and I_{22} have to be delivered from corresponding current sources. As for every synthesis task here one exists a wide variety of decisions [5, 6]. For example the equation for $i_{C1}(t)$ ($C=C_1$) in (17) corresponds to Fig. 3 and one chooses $i_{c,\,T1}(t)=i(t)$; $i_{c,\,T2}(t)=i_{out}(t)$. Then the second equation in (24) can be realized by a translinear loop (TLL) that contains four transistors (of cause we suppose a properly biasing of them to ensure positive values of all time-depending currents) – Fig. 6. The whole circuit must contain mirrors for current transfer between its parts yet. The circuit realization that follows from the first equation in (24) can be implemented in a similar manner.

Bernoulli cell discussed above is another useful basic subcircuit for the synthesis especially of translinear filters [3, 4] – Fig. 7. Here we have

$$\frac{dv_B(t)}{dt} = U_T\,\frac{1}{i_{in}(t)}\,\frac{di_{in}(t)}{dt} \tag{25}$$

and after substituting this expression in (10) one follows

$$CU_T \frac{d}{dt}\left(\ln\left(\tau(t).i_{in}(t)\right)\right) + i(t) - i_c(t) = 0 \tag{26}$$

This expression reflects KCL for the emiter and consequently the first term in it is the capacitor current $i_C(t)$.

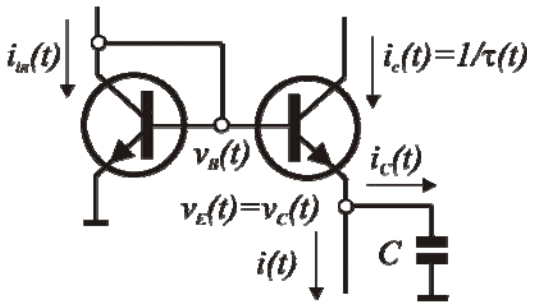


Fig. 7

A specific problem is the interconnection of Bernoulli cells. For this purpose one uses level shifters as it is shown in Fig. 8 where the shifters is implemented by the current sources I_{01} and I_{02}.

Having into account (26) for the first cell in Fig. 8 we have

$$C_1U_T \frac{d}{dt}\left(\ln\left(\tau_1(t).i_{in}(t)\right)\right) + i_1(t) - i_{c1}(t) = 0\,. \tag{27}$$

Further for the second cell from (10) and (26) we can write, respectively

$$\frac{d\tau_2(t)}{dt} + \frac{1}{U_T}\left(\frac{dv_{B2}(t)}{dt} + \frac{i_2(t)}{C_2}\right)\tau_2(t) - \frac{1}{C_2U_T} = 0\,; \tag{28}$$

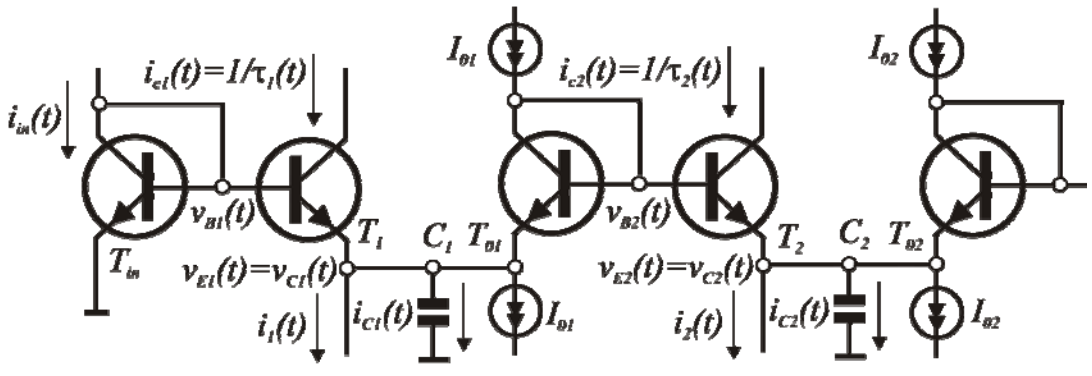


Fig. 8

$$\frac{dv_{C1}(t)}{dt} = U_T \frac{d}{dt}\left(\ln\left(\tau_1(t).i_{in}(t)\right)\right). \tag{29}$$

Because

$$v_{B2}(t) = v_{C1}(t) + U_T \ln\frac{I_{01}}{I_s} \tag{30}$$

it follows

$$\frac{dv_{B2}(t)}{dt} = \frac{dv_{C1}(t)}{dt}. \tag{31}$$

Finally, having in mind (28) ÷ (31) one obtains

$$C_2 U_T \frac{d}{dt}\left(\ln\left(\tau_1(t).\tau_2(t).i_{in}(t)\right)\right) + i_2(t) = \frac{1}{\tau_2(t)} = i_{C2}(t). \tag{32}$$

The last result can be generalized and for the case of an arbitrary number *m* connected Bernoulli cells [4] as follows:

$$i_{C,m}(t) = C_m U_T \frac{d}{dt}\left(\ln\left(i_{in}(t)\prod_{k=1}^{m}\tau_k(t)\right)\right). \tag{33}$$

If we substitute the recurrent relationship

$$w_k(t) = i_{in}(t)\prod_{j=1}^{k}\tau_k(t) = \tau_k(t)w_{k-1}(t) \tag{34}$$

in (33) the KCL applied to (27), (32) and (33) leads to a system of first-order linear differential equations in a state-space form.

$$\left|\begin{array}{l} C_1 U_T \dfrac{dw_1(t)}{dt} + i_1(t)w_1(t) = i_{in}(t); \\ C_2 U_T \dfrac{dw_2(t)}{dt} + i_2(t)w_2(t) = w_1(t); \\ \ldots\ldots\ldots\ldots \\ C_m U_T \dfrac{dw_m(t)}{dt} + i_m(t)w_m(t) = w_{m-1}(t) \end{array}\right. . \tag{35}$$

The defining of the variables $w_k(t)$ can be done by means of the circuit, given in Fig. 9 which is composed by adding transistor pairs to the circuit in Fig. 8. Here one discovers the translinear loops formed by an even number transistor base-emiter junc tions, namely: T_{in}, T_1, T_{q1}, $T_{out,1}$; T_{in}, T_1, T_{01}, T_2, T_{q2}, $T_{out,2}$; T_{in}, T_1, T_{01}, T_2, T_{02}, T_3, T_{q3},

$T_{out,3}$ and so on. Then, according to STLP for m connected Bernoulli cells it follows the generalized expression

$$i_{out,m}(t) = \left(\prod_{k=1}^{m-1} I_{0,m} \right) I_{q,m} . w_m(t) \tag{36}$$

– i.e. every output current $i_{out,k}(t)$ is proportional to the corresponding state variable $w_k(t)$. This circumstance is a precondition for the successful use of Bernoulli cells in translinear filter synthesis [3, 4].

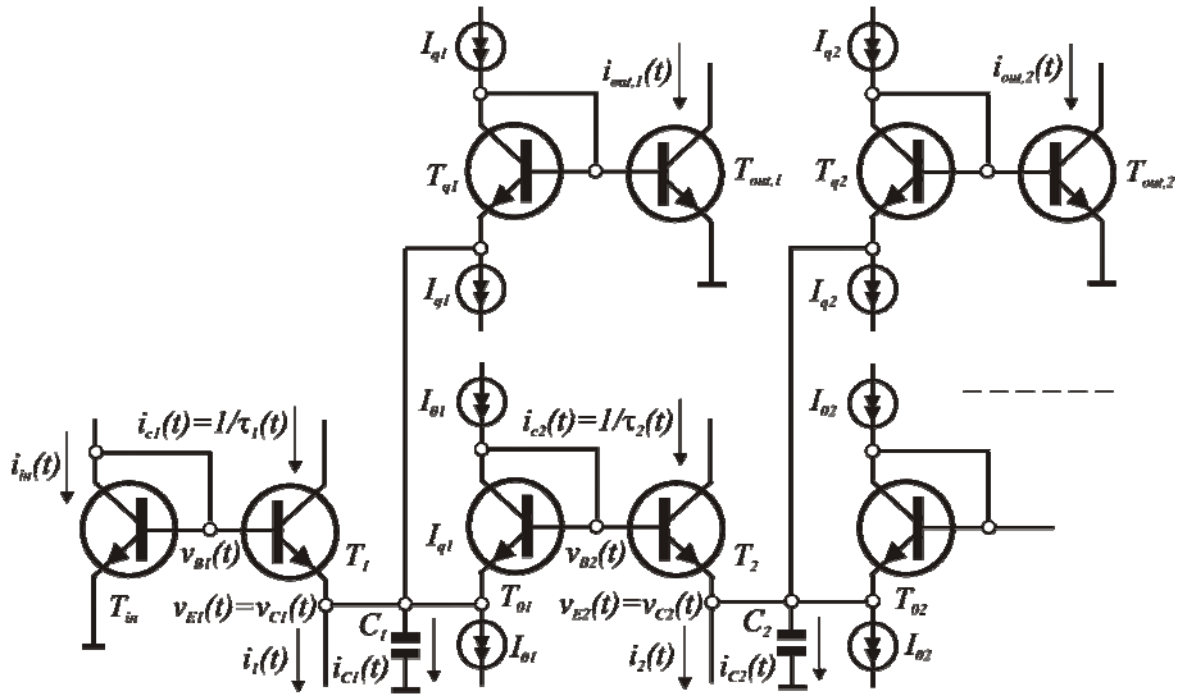


Fig. 9

5. CONCLUSIONS

Several methods for the realization of state-space equations by means of translinear circuits are considered in the paper. Two basic dynamic subcircuits with capacitors as memory elements are often used for the synthesis of log-domain filters with previously determined current-mode transfer functions. All the described methods are based on two principles – *static translinear principle* (STLP) and *dynamic translinear principle* (DTLP). One accessory but important problem in the described in the literature synthesis approaches is the use of great number of additional current sources in order to fix the transistors working points and it is insufficiently explained in the research publications.

REFERENCES

[1] D. R. Frey, “Exponential State Space Filters: A Genetic Current Mode Design Strategy”, IEEE Tr. on Circuits and Systems - I, vol. 43, 1, 1996, pp. 34-42

[2] D. R. Frey, “State-Space Synthesis and Analysis of Log-Domain Filters”, IEEE Tr. on Circuits and Systems - II, vol. 45, 9, 1998, pp. 1205-1211

[3] E. M. Drakakis, A. J. Payne, C. Toumazou, “Log-Domain State-Space”: A Systematic Transistor- Level Approach for Log-Domain Filtering, IEEE Tr. on Circuits and Systems - II, vol. 46, 3, 1999, pp. 290-305

[4] E. M. Drakakis, A. J. Payne, C. Toumazou, Log-Domain Filtering and Bernoulli Cell, IEEE Tr. on Circuits and Systems - I, vol. 46, 5, 1999, pp. 559-571

[5] J. Mulder, W. A. Serdijn, A.C. van der Woerd, A.H.M. van Roermund, Dynamic

[6] J. Mulder, W. A. Serdijn, A.C. van der Woerd, A.H.M. van Roermund, A Generalized Class of Dynamic Translinear Circuits, IEEE Tr. on Circuits and Systems - II, vol. 48, 5, 2001, pp. 501-504

[7] F. M. Diepstraten, F. C. M. Kuijstermans, W. A. Serdijn, P. van der Kloet, A. van Staveren, F. L. Neerhoff, C. J. M. Verhoeven, A.H.M. van Roermund, Dynamic Behavior of Dynamic Translinear Circuits: The Linear Time-Varying Approximation, IEEE Tr. on Circuits and Systems - I, vol. 48, 11, 2001, pp. 1333-1337

[8] S.A.P.Haddad, W.A.Serdijn, High-Frequency Dynamic Translinear and Log-Domain Circuits in CMOS Technology, ISCAS 2002, Scottsdale, AZ, USA, 2002, pp. III 313- III 316

[9] A.Kircay, U.Cam, State-Space Synthesis of Current-Mode First-Order Log-Domain Filters, Turk. J. Elec. Engin, vol. 14, 3, 2006, pp. 399-416

[10] http://www.cliffsnotes.com/WileyCDA/CliffsReviewTopic/Bernoullis-Equation.topicArticleId-19736,articleId-19715.html

[11] http://en.wikipedia.org/wiki/Bernoulli_differential_equation

[12] http://mathworld.wolfram.com/BernoulliDifferentialEquation.html

NUCLEOTIDE GENOMIC SIGNALS: APPLICATION TO THE COMPARATIVE ANALYSIS OF HOMINIDAE MTDNA

Paul Dan Cristea

Biomedical Engineering Center, University "Politehnica" of Bucharest, Spl.Independentei 313, 060042 Bucharest, Romania, phone: +40 21 316 9569, e-mail: pcristea@dsp.pub.ro

Abstract: *The Nucleotide Genomic Signal (NuGS) methodology reveals surprising regularities, both locally and at a global scale, which would be difficult or impossible to identify by using only statistical analysis and pattern matching methods. The approach has been applied for studying large scale features of chromosomes, for detecting mutations in coding and non-coding nucleotide sequences, for analyzing inserts in prokaryote and eukaryote genomes. The lecture briefly presents the NuGS methodology and shows how this approach can be used in the comparative study of hominidae family mitochondrial DNA (mtDNA). The nucleotide imbalance (Nc), the nucleotide pair imbalance (Pu) and the nucleotide path (NuP) of mtNuGSs of both extant (Homo sapiens, Pongo pygmaeus, Pongo pygmaeus abeli, Pan troglodytes, Pan paniscus, Gorilla gorilla) and fosile (Homo sapiens neanderthalensis) hominidae are analyzed. A reference-offset representation of sets of related signals is used.*

Keywords: Nucleotide sequences, Nucleotide complex representation, Nucleotide Genomic Signals, Reference-Offset representation

1. INTRODUCTION: THE NUCLEOTIDE GENOMIC SIGNAL METHODOLOGY

The Nucleotide Genomic Signal (*NuGS*) methodology [1,2] is based on the conversion of nucleotide sequences into digital signals allows to apply signal processing methods for the analysis of genomic data. The method makes straightforward the comparison of nucleotide sequences, allows the fast identification of mutations and inserts, and reveals surprising patterns and regularities in the distribution of nucleotides, pairs of nucleotides and small groups of nucleotides along the sequences. Such features of nucleotide sequences are difficult to find by using only symbolic genomic sequences and standard statistical and pattern matching methods.

We have investigated a large number of mappings of symbolic genomic data to digital genomic signals and we have compared how the structure of the genomic code was highlighted by the various representations and how the features of DNA sequences were revealed by the resulting digital signals. Such a representation has to be both truthful and unbiased. The mapping is truthful if all biologically relevant characteristics of the represented objects are expressed in corresponding mathematical properties of the samples in the resulting digital signal. The mapping is unbiased if the representation does not contain artefacts, i.e., features belonging to the mapping itself, without correspondent in the properties of the initial sequence.

The mapping we have found to be the best in revealing the essential features of nucleotide sequences is a one-to-one unbiased complex representation of nucleotide equivalence classes, which attaches the following complex numbers to adenine, cytosine, guanine and thymine nucleotides:

$$a = \ 1+j, \quad c = -1-j, \quad g = -1+j, \quad t = \ 1-j\,. \tag{1}$$

This mapping, to which corresponds the graphics in Fig.1, can readily be extended to all the other IUPAC [3] nucleotide classes comprising nucleotide pairs, triplets or quadruplets, respectively.

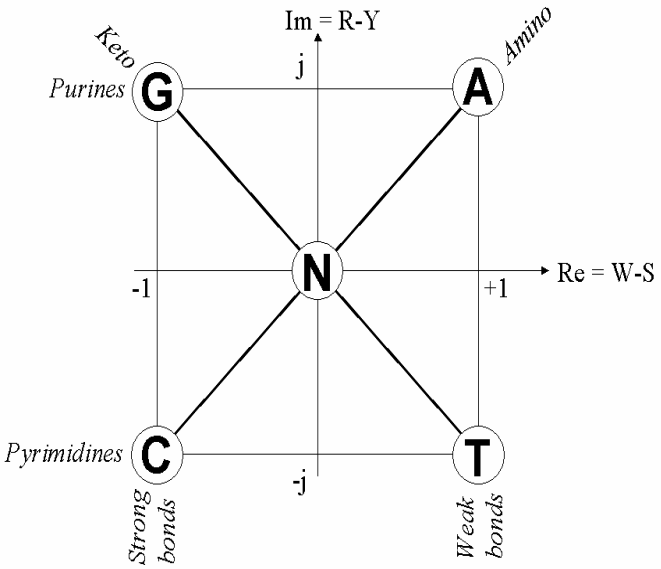


Fig. 1. Complex representation of single nucleotide classes.

Using this representation, each nucleotide $(Nu)_k$ in the sequence ($k = 1,\ldots,n_B$), n_B – the length of the sequence (number of bases), $(Nu)_k \in \{a, c, g, t\}$, is mapped into its complex representation $\mathcal{C}\{(Nu)_k\}$ given in (1). Correspondingly, the symbolic sequence $[(Nu)_k \mid k = 1,\ldots,n_B\,]$ is mapped into the complex signal $[\mathcal{C}\{(Nu)_k\} \mid k = 1,\ldots,n_B\,]$.

The statistics of nucleotides can be described simply by two signatures:

– the *nucleotide imbalance*

$$N_c = 3(n_G\text{-}n_C)+(n_A\text{-}n_T) = (4/\pi)\ \vartheta_c\,, \tag{2}$$

where n_A, n_C, n_G and n_T are the numbers of adenine, cytosine, guanine and thymine nucleotides in the sequence from the first to the current entry, and ϑ_c is the cumulated phase, which is equal to the sum of the phases of all entries of the *NuGS*, from the beginning of the sequence, up to the current position;

– the *nucleotide pair imbalance*

$$P_u = n_+ - n_- = (2/\pi)\ \vartheta_u\,, \tag{3}$$

where n_+ is the number of positive pairs (A→G, G→C, C→T, T→A), and n_- the number of negative pairs (A→T, T→C, C→G, G→A), and ϑ_u is the unwrapped phase of the entries of the *NuGS*.

Another useful representation of a *NuGS* is its complex nucleotide path *NuP*. The n^{th} value of the *NuP* is the sum of the complex representations of the first n nucleotides in the sequence:

$$(NuP)_n = \sum_{k=1}^{n} \boldsymbol{C}\ \{(Nu)_k\}, \quad n \in \{1, n_B\}, \tag{4}$$

The real and the imaginary part of the *NuP* give the *weak–strong* and *purine–pyrimidine* imbalances, respectively:

$$\mathrm{Re}\{NuP\} = n_w - n_s = (n_a + n_t) - (n_c + n_g), \tag{5}$$

$$\mathrm{Im}\{NuP\} = n_r - n_y = (n_a + n_g) - (n_c + n_t), \tag{6}$$

where n_w and n_s are the numbers of weak and strong links, n_r and n_y – the numbers of purines and pyrimidines, n_a, n_c, n_g and n_t – the numbers of adenine, cyanine, guanine and thymine nucleotides in the sequence from the start to the current position.

The *NuGS* approach has been introduced initially to detect the large scale features of DNA sequences that are maintained over distances of 10^6–10^8 base pairs, including both coding and non-coding segments [4]. The methodology proved also adequate for the local analysis of the sequences, such as necessary for the study of pathogen variability [5,6,7,8,9] and the identification of drug resistance [6,8,9]. This is important for fast diagnosis and prompt socio-medical decision in contamination with pathogens such as *Human immunodeficiency virus* (HIV) [5], *Avian* (H5N1) and *Swine* (H1N1) *influenza virus* [6, 7], or *Mycobacterium tuberculosis* (MT) [8,9]. The method has also been used in mtDNA analysis [10, 11]. The regularities in the nucleotide distribution evoke Chargaff's laws [12], but they are more general and refer also to nucleotide pair distribution [1, 2].

Some of the main features of the *NuGSs* are [1, 4]:

(1) A remarkable good linearity of the nucleotide pair imbalance P_u. for most of prokaryotes and eukaryotes. The root mean square error per nucleotide of the linear fitting to P_u is typically very small (*e.g.,* 0.0045 for *MT*), which corresponds to a smooth strait line at large scale.

(2) An approximately piece-wise linear nucleotide imbalance N_c for prokaryotes, and close to zero for eukaryote (using the scale of P_u for N_c).

(3) The extremes of N_c correspond to the origin and the terminus of genome replication.

Re-orienting all exons in a sequence along the same positive direction, reveals some unexpected 'hidden' ancestral features [1, 2]:

(1) A (perfect and predictable) invariance of the nucleotide pair imbalance P_u, resulting from the conservation of the direct (n_+) and inverse (n_-) numbers of pairs when reversing a segment of a DNA double helix, while simultaneously switching its strands, in a process modeling recombination.

(2) An approximately linear shape of N_c after re-orientation, suggests a regular ancestral genomic structure. The current nucleotide longitudinal structure, typical for each species, has evolved from this ancestral structure under selective pressure, to prevent inter-species genetic exchange of material.

The long range regularities show that, from the structural point of view, a genome resembles less to a "plain text", which simply expresses a semantics in accordance to certain grammar rules, but more to a "poem", which also obeys additional rules of symmetry, giving it "rhythm" and "rhyme". The structural constrains of genomic sequences are reflected in the regularities observed in the corresponding genomic signals.

The paper applies the NuGS methodology in the comparative analysis of mitochondrial DNA (mtDNA).

2. MITOCHONDRIAL DNA

The vast majority of the genome of eukaryotic organisms is packaged in the DNA of their nuclear chromosomes. The remaining 10^{-5} of the total genome is stored in the circular DNA of mitochondria (mtDNA) [10,13]. In *hominidae*, a mitochondrion contains between 2 and 10 mtDNAs, each with a length of approximately 16,500 base pairs (16,569 bp, in homo sapiens) and encoding 37 genes. Most of the about 1,500 proteins present in the mitochondria of today extant organisms are encoded by the nuclear DNA. Nevertheless, it is considered that their genes are of bacterial origin and have been transferred to the eukaryotic nucleus during evolution. In mammals, the mt genes encode the 37 proteins shown in Table 1. It must be taken into account that the mitochondrial genetic code table is different from the common genetic code table of prokaryotes and eukaryotes [13].

The mt DNA also comprises a 1122-bp major control segment, the D-loop, which is located between the genes for tRNA-Pro and tRNA-Phe. The D-loop contains the hyper-variable regions HVR1 and HVR2, as well as the main structures responsible for initiation and regulation of mtDNA transcription and replication.

TABLE 1. GENES IN MAMMAL MTDNA

Product		**Genes**
electron trans-	NADH dehydrogenase (complex I)	MT-ND1, MT-ND2, MT-ND3, MT-ND4, MT-ND4L, MT-ND5, MT-ND6
	Coenzyme Q, cytochrome c reductase / cytochrome b (complex III)	MT-CYB
	Cytochrome c oxidase, (complex IV)	MT-CO1, MT-CO2, MT-CO3
	ATP synthase	MT-ATP6, MT-ATP8
2 mt rRNA		MT-RNR1 (12S), MT-RNR2 (16S).
22 mt tRNA		MT-Ala, …, MT-Val (22 genes for 20 for AA)

Nuclear DNA is inherited from both parents, so that its genes are rearranged in the process of recombination. Consequently, it is difficult to use the nuclear DNA for tracing the inheritance lineage. In contrast, mtDNA is normally only maternally inherited, as mitochondria in mammalian sperm are usually destroyed by the egg cell after fertilization. Similarly, the Y chromosome is only paternally inherited, as females do not have it. Therefore, these two types of DNA are largely used for tracking maternal, respectively paternal, lineage in a population, being important in evolutionary studies, medicine and forensic analyses.

3. NUCLEOTIDE GENOMIC SIGNALS OF MITOCHONDRIAL DNA

This section presents comparatively nucleotide genomic signals of the mitochondrial DNA for representatives of the hominidae family.

Fig. 2 shows the nucleotide path (*NuP*) for the hominidae mitochondrial DNA of six related species Homo sapiens (accession number NC 001807 in GenBank [12], length 16571 bp), *Gorilla gorilla* (NC 001645, 16364 bp), *Pan paniscus* (NC 001644, 16563 bp), *Pan troglodytes* (NC 001643, 16554 bp), *Pongo pygmaeus* (NC 002083, 16389 bp). The *NuP* of *Homo sapiens neanderthalensis* is very close of the one for *Homo sapiens*, so that it has not been shown separately in Fig. 2.

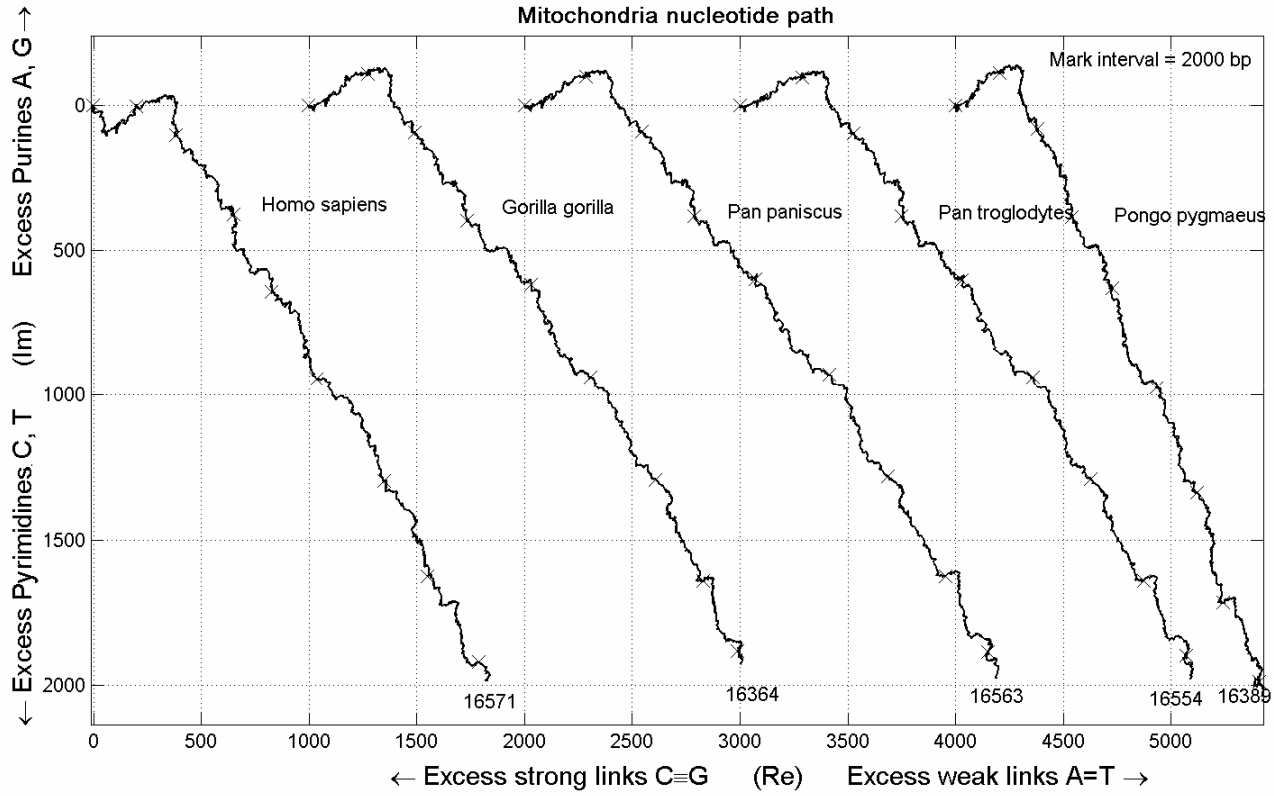


Fig.2. Nucleotide path of the *Hominidae* mitochondrial DNA: for *Homo sapiens* (16571 bp), *Gorilla gorilla* (16364 bp), *Pan paniscus* (16563 bp), *Pan troglodytes* (16554 bp), *Pongo pygmaeus* (16389 bp).

The starting point of the plots have been shifted with 1000 units from each other in the horizontal direction to help the direct comparison of the nucleotide paths. The

mtDNA is quite inhomogeneous, as it contains segments with different structures and functions: 22 genes that encode the mt tRNAs for the 20 amino acids, 2 genes for the ribosomal RNAs and 13 genes encoding proteins involved in the electron transfer chain (see Table 1). The shape of the *NuP* plot reflects this non-homogeneity.

A consistent phylogenetic tree can be built by using separately the various genes and non-coding segments and combining the partial results, as and the histories of the various segments are distinct and there is a marked tendency to compensate the mutations to satisfy the overall regularities of the *NuGS*.

A comparison between the same mitochondrial gene in related species is shown in figures 3, 4, and 5.

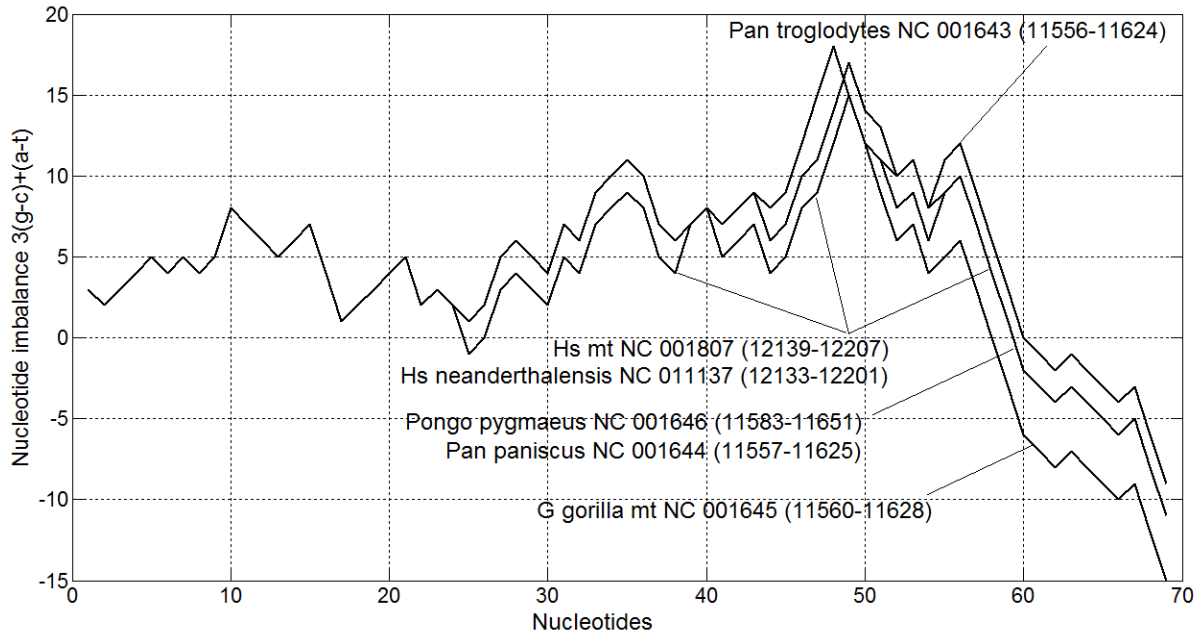


Fig. 3. Nucleotide imbalance signals for the mt gene encoding tRNA-His (the transport RNA of the histidine amino acid) in the *hominidae* family (*H. sapiens, H. sapiens neanderthalensis, Pongo pygmaeus, Pan troglodytes, Pan paniscus*, and *Gorilla gorilla* are considered).

Fig. 3 presents the nucleotide imbalance for the transport RNA of the histidine amino acid (tRNA-His) for six members of the *hominidae* family (*H. sapiens, H. sapiens neanderthalensis, Pongo pygmaeus, Pan troglodytes, Pan paniscus*, and *Gorilla gorilla*). The locations of the gene along the sequences and the accession numbers in the GenBank [13] are given in the figure. The conservative character of the gene shows in its little variability inside the *hominidae* family. Among other well conserved features, the size of the tRNA-His gene is 72 bp for all the considered species. It is remarkable that, for the six distinct members of the *hominidae* family, the number of apparently distinct curves in Fig. 3 remains mainly in the restricted range of one to three.

The resolution of the comparison between the closely related signals corresponding to homologous genes can be increased by using the reference-offset representation. The choice of the central tendency (the trend), which gives the reference, must

be optimize in order to properly extract the common variation of the signals in the considered set, so that the individual offset of each signal really expresses only the features specific to that signal, with minimum cross-talk.

We have used this reference-offset method to analyze the individual variations in a set of genomic signals from related strains of pathogens, or in a pathogen population, especially when identifying mutations resulting in pathogen resistance to drugs. In the cases of stable pathogens, *e.g.*, *Mycobacterium tuberculosis*, the *NuGS* corresponding to the wild type (without mutations) can be used as reference. The offsets express how far form the wild type has evolved each studied strand.

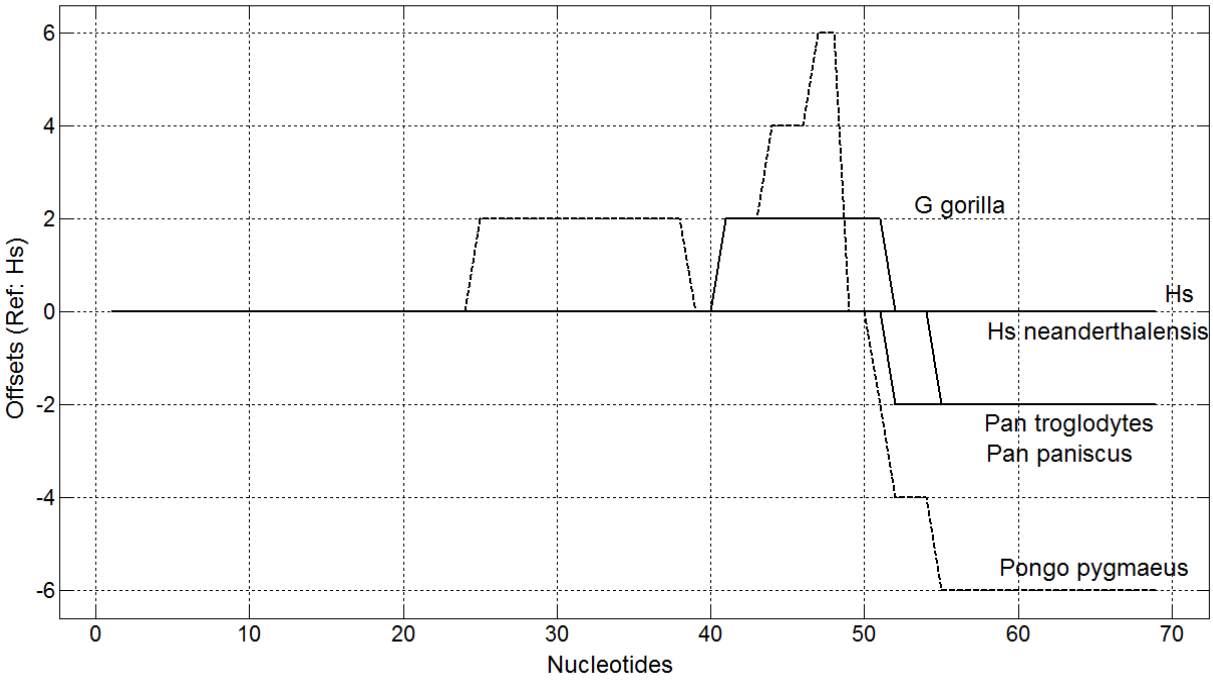


Fig. 4. Offsets of the nucleotide imbalance signals in Fig. with respect to the Homo sapiens reference

In the case of pathogens with a very high variability, *e.g.*, for the *Human Immunodeficiency Virus* (HIV), the choice of the wild type *NuGS* as reference is no longer adequate, because the distance between the wild type and the various instances in the pathogen populations would be too large, and the reference would no longer express the common trend of the set of *NuGS*s. In such cases, the *median* of the set of signals could be used as reference, but better results were obtained when using the *modal step reference* or the *maximum flat reference* []. The offsets express in such cases the divergence inside the set of signals, i.e., the spread of the features of the considered population of pathogens.

For the comparative study of the mtDNA in related species, such as in our case, it is better to take one of the signals as reference. The offsets express, in this case, how far are the other members of the family with respect to the one chosen as reference. Fig. 4 shows the reference-offset representation of the nucleotide imbalance signals in Fig. 3, when using the signal for *H. sapiens* as reference. On most of the tRNA-His

gene length, the offsets are either identical (superposed), or remain constant (parallel to the nucleotide axis), as the signals have the same trend. Abrupt variations occur only in the points in which the genes of the six members of the hominidae family display local mutations, usually single nucleotide polymorphisms (SNPs).

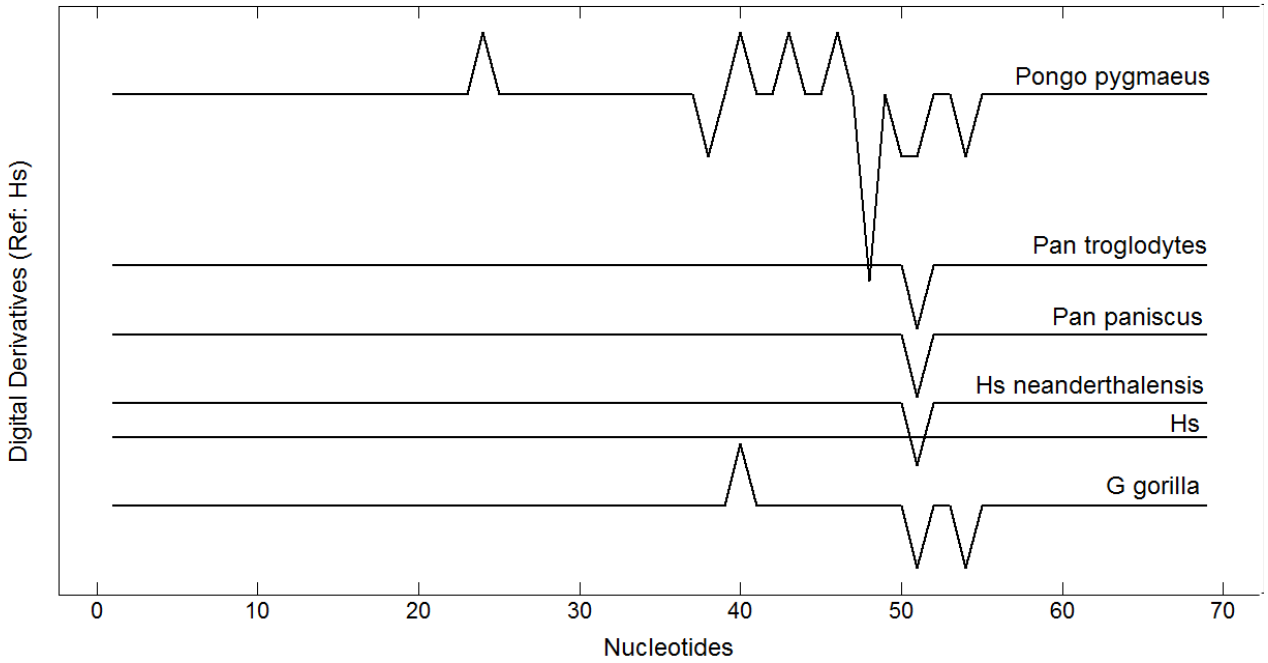


Fig. 5. Digital derivatives of the offsets of in Fig. The distance with respect to the Homo sapiens tRNA-His can be readily estimated

The evolutionary divergence of the species in the *hominidae* family and their distances with respect to the *H. sapiens* reference can easily be established by using the differential signals of the offsets shown in Fig. 5. The distances between species derived from only one gene can be misleading. To build a non-contradictory and plausible phylogenetic tree, it is necessary to combine distances derived from all or a number of homologues genes, and from other non-coding conserved genome regions [1].

The mitochondrial RNA genes have a sequence structure that evokes the nuclear RNA secondary structure and probably has a similar function: the protection of the molecule structure, thus information content, in the cellular environment, with the corresponding increase of the mt tRNA half-time. This feature is revealed by the specific symmetry showed by the nucleotide path (*NP*) of mt tRNA genes.

Fig. 6 presents the *NuP* of the mt tRNA-His gene for six members of the *hominidae* family. The *NuP* contains several loops that reveal a specific complementary palindrome symmetry of the gene. This symmetry allows the single stranded mt tRNA-His to fold on itself and establish hydrogen bonds between complementary nucleotides (A-T, C-G).

The folding of the *NP* on itself can be better followed in Fig. 7 which explicitly shows the *weak–strong* (W-S) and *purine–pyrimidine* (R-Y) imbalances along the nucleotide sequence for the *Homo sapiens* mt tRNA-His gene.

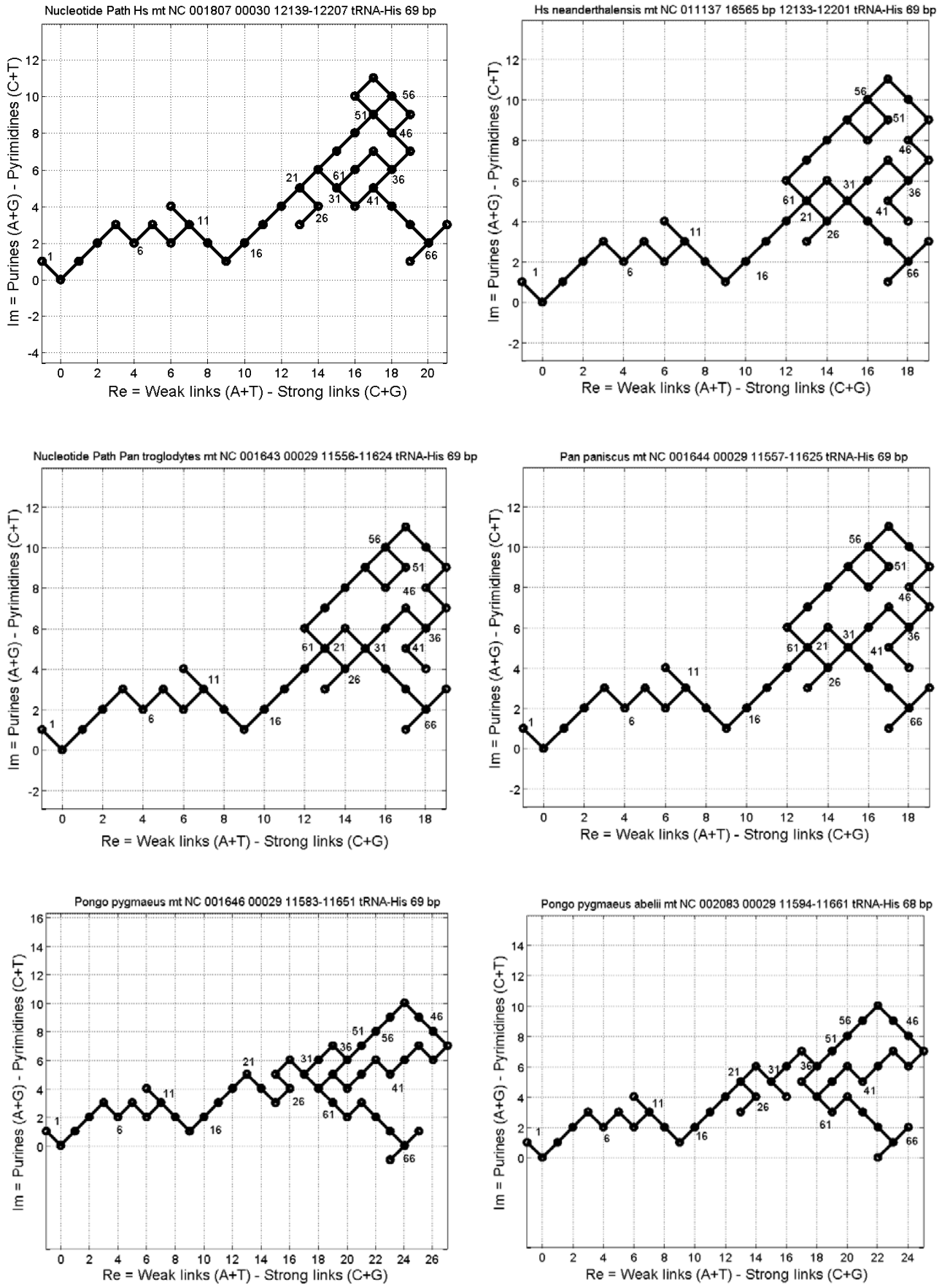


Fig. 6. Nucleotide path of the mt tRNA-His gene for six members of the *hominidae* family. The *NP* contains several loops that reveal the specific symmetry of complementary palindrome.

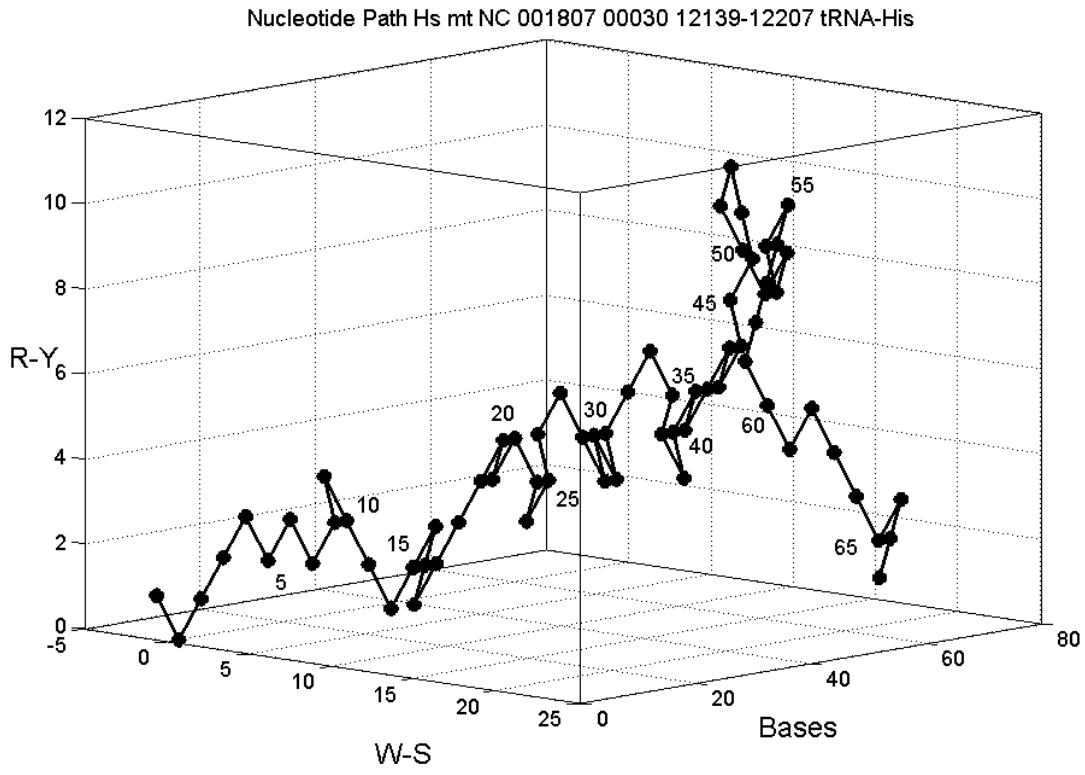


Fig. 7. 3D Nucleotide path of the mt tRNA-His gene for *Homo sapiens*. The return on itself of the *NP* reveals the complementary palindrome symmetry of the sequence.

4. CONCLUSIONS

The paper presents some results in the application of *NuGS* methodology to the comparative analysis of hominoid mitochondrial chromosomes. New approaches based on nucleotide imbalance (cumulated phase), nucleotide pair imbalance (unwrapped phase), 2D and 3D nucleotide path (weak-strong vs purine-pyrimidine imbalance) have been developed to this purpose [3]. An efficient method of representing sets of signals by reference and offsets is used to get a better resolution in the case of related signals. The method has been initially developed for the study of pathogen variability in the context of drug resistance early detection, but showed good results in the analysis of mt DNA.

Regularities in the distribution of nucleotides and pairs of nucleotides have been found, showing that mtDNA has a well ordered structure, closer to the regularity of prokaryote genomes, than to the relative patchy structure of eukaryote genomes.

5. ACKNOWLEDGEMENT

The work was partially supported by the project **142418-LLP-1-2008-1-GR-ERASMUS-ENW** entitled ESTIA-EARTH – "To Sustain the Women Careers as Academics, Researchers and Professionals in Engineering, Computers and Sciences", in the framework of the **Lifelong Learning Programme ERASMUS Network** co-

ordinated by NTUA – the National Technical University of Athens and by the project **214538 – 2008 - BISNES** – "Bio-Inspired Self-assembled Nano-Enabled Surfaces", in the framework of the NMP-2007-1.1-2 Self-assembling and self-organization, NMP-2007-1.1-1 Nano-scale mechanisms of bio/non-bio interactions

REFERENCES

[1] P.D. Cristea, Chapter 1, "Representation and analysis of DNA sequences", in *Genomic Signal Processing and Statistics*, E. Daugherty, I. Shmulevich, J. Chen and Z.J. Wang, Eds., EURASIP Book Series on Signal Processing and Communications, Hindawi Publ. Corp., p. 15–65, 2005

[2] P.D. Cristea, "Conversion of Nitrogenous Base Sequences into Genomic Signals", *Journal of Cellular and Molecular Medicine*, 6(2), p.279–303, 2002.

[3] IUPAC - International Union of Pure and Applied Chemistry, *http://www.iupac.org/, http://www.bioinformatics.org/sms/iupac.html*

[4] P. D. Cristea, "Large Scale Features in DNA Genomic Signals", *Signal Processing, Special Issue on Genomic Signal Processing*, *Elsevier*, **83**, (2003), pp. 871-888.

[5] P.D. Cristea, D. Otelea, R.Tuduce, "Genomic Signal analysis of HIV variability", *SPIE - BIOS 2005, Proc. of SPIE*, **6**, (14), p. 362-372, 2005.

[6] M. Enserink, "Avian Influenza: 'Pandemic Vaccine' Appears to Protect Only at High Doses," *Science*, **309**, pp. 996, 12 Aug. 2005.

[7] P. D. Cristea, "Genomic Signal Analysis of Pathogen Variability", *Progress in Biomedical Optics and Imaging*, Proc. of SPIE, vol. 6088, pp. P1-P12, 2006.

[8] S.T. Cole, R. Brosch, J. Parkhill, et al., "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence", *Nature*, 393 (6685), p. 537-544, 1998.

[9] P. D. Cristea, "Genomic Signal Analysis of Mycobacterium tuberculosis", *Progress in Biomedical Optics and Imaging*, Proc. of SPIE, vol. 6447, p. C-1 – C8, 2007 .

[10] M. Krings et al., "Mitochondrial genome variation and the origin of modern humans", *Nature*, **410** (6828), 2001, pp. 611.

[11] P. D. Cristea, Rodica Tuduce, "Nucleotide Genomic Signal Comparative Analysis of Homo Sapiens and other Hominidae mtDNA**"**, ISSCS 2009 – *International Symposium on Signals, Circuits and Systems,* July 9-10, 2009, Iaşi, Romania, pp. 5-8.

[12] E. Chargaff, "Structure and function of nucleic acids as cell constituents", *Fed. Proc.*, 10, p. 654–659, 1951.

[13] GenBank, National Centre for Biotechnology Information, National Institutes of Health, National Library of Medicine, (NCBI/GenBank), http://www.ncbi.nlm.nih.gov/, 2007.

STRONG STABILITY RADIUS OF LINEAR INTERVAL PARAMETER CIRCUITS

Lubomir Kolev

Department of Theoretical Electrical Engineering, Technical University of Sofia,
bld. Kl. Ohridski 8, 1000 Sofia, Bulgaria, e-mail: lkolev@tu-sofia.bg

Abstract: *A new robust stability concept, the so-called strong stability radius r_{ss}, is introduced for linear circuits whose parameter uncertainties are modeled by an interval matrix* ***A****. Circuits whose parameters range within r_{ss} are guaranteed to have smoother transients in comparison to circuits that are only stable. It is proved that the problem of determining r_{ss} of the original $n \times n$ interval matrix* ***A*** *can be equated to that of finding the robust stability radius r_s of an associated symmetric $2n \times 2n$ interval matrix* ***C****. It is also shown that the stability radius r_s of* ***C*** *can be computed as the maximum eigenvalue of a corresponding interval generalized eigenvalue problem related to* ***C****. The latter problem can be solved with an amount of computational expenses that is a polynomial in the size n of the original matrix.*

Keywords: strong stability radius, generalized eigenvalue problem.

1. INTRODUCTION

As is well known, robust stability analysis of linear time-invariant circuits (systems) whose parameters are not exact but are known to lie within preset intervals can be equated to the problem of estimating the stability margin of a corresponding interval matrix A (e.g. [1]). A more recent trend in the quantitative stability analysis of such circuits is to assess the so-called stability radius r_s of ***A*** [2]. Circuits whose parameter ranges are such that $r_s > 1$ are guaranteed to be robustly stable. However, transients in stable circuits may have considerable overshoots that are sometimes undisirable.

In the present paper, a new robust stability measure, the so-called strong stability radius r_{ss} of an arbitrary $n \times n$ interval matrix ***A*** is introduced. Circuits whose descriptive matrix ***A*** has $r_{ss} > 1$ are guaranteed to exhibit smoother transients as compared to circuits that are just stable.

The problem of determining r_{ss} is evidently *NP*-hard (its numerical complexity is exponential in the size *n* of the matrix) since the simpler problem of computing r_s is known to be *NP*-hard. According to [3], such problems can be solved only for matrices of moderate size ($n \leq 15$). Thus, development of methods for determining r_s or r_{ss} whose numerical complexity is polynomial in *n* is of considerable practical significanse. Such a polynomial complexity method for determining r_s has already been suggested in a recent paper [4].

This paper proposes an approach permitting to equate the original problem of determining the strong stability radius r_{ss} of $\boldsymbol{A}$ to the problem of finding the stability radius r_s of an associated symmetric interval matrix $\boldsymbol{C}$ of size $2n \times 2n$. Next, r_s of $\boldsymbol{C}$ is computed as the maximum eigenvalue of a corresponding interval generalized eigenvalue problem related to **C**. The latter problem can be solved using the polynomial complexity methods from [4] or [5].

2. PROBLEM STATEMENT

As will be seen below, the concept of strong stability is related to the concept of singular values of a real (non-interval) matrix A

$$\sigma_i(A) = \sqrt{\lambda_i(A^T A)} \tag{1}$$

where λ_i is the ith eigenvalue of the matrix involved. Using (1), we define the maximum singular value of A

$$\overline{\sigma_i}(A) = \max(\sigma_i(A) \tag{2}$$

and the maximum singular value of the interval matrix $\boldsymbol{A}$

$$\overline{\sigma_i}(\boldsymbol{A}) = \max(\sigma_i(A), \quad A \in \boldsymbol{A}. \tag{3}$$

The interval matrix $\boldsymbol{A}$ will be called strongly stable if

$$\overline{\sigma}(\boldsymbol{A}) < 1. \tag{4}$$

Thus, we can introduce (by analogy with the standard stability margin) the concept of strong stability margin

$$M_{ss}(\boldsymbol{A}) = 1 - \overline{\sigma}(\boldsymbol{A}). \tag{5}$$

To introduce the concept of strong stability radius $r_{ss}(\boldsymbol{A})$, we need the centered form of the interval matrix.

$$\boldsymbol{A} = A^c + [-R, R] = A^c + \boldsymbol{B}, \quad \boldsymbol{B} = [-R, R]. \tag{6}$$

Then $r_{ss}(\boldsymbol{A})$ is defined as follows

$$r_{ss}(\boldsymbol{A}) = \min\{r \geq 0: \quad \overline{\sigma}(A^c + r\,B) = 1, \ B \in \boldsymbol{B}\}. \tag{7}$$

3. MAIN RESULTS

Let

$$G = \begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix}, \quad \boldsymbol{G} = \begin{bmatrix} 0 & \boldsymbol{A}^T \\ \boldsymbol{A} & 0 \end{bmatrix}. \tag{8}$$

Obviously, G is a symmetric $2n \times 2n$ real matrix belonging to the interval matrix $\boldsymbol{G}$. Rather than using (1), following [6] we have

$$\sigma_i(A) = Pos(\lambda_i(G)) = Pos(\lambda_i(\begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix})) \tag{9}$$

where *Pos*( ) means positive part of ( ). On account of (3), (8) and (9)

$$\overline{\sigma}(\boldsymbol{A}) = \overline{\lambda}(\boldsymbol{G}) = \max\{\lambda_i(G),\ i = 1,...2n,\ G \in \boldsymbol{G}\}. \tag{10}$$

Now using (6), we put $\boldsymbol{G}$ in central form

$$\boldsymbol{G} = G^c + \boldsymbol{H} = \begin{bmatrix} 0 & (A^c)^T \\ A_c & 0 \end{bmatrix} + \boldsymbol{H} \tag{11}$$

with

$$H = \begin{bmatrix} 0 & B^T \\ B & 0 \end{bmatrix}, \quad \boldsymbol{H} = \begin{bmatrix} 0 & \boldsymbol{B}^T \\ \boldsymbol{B} & 0 \end{bmatrix} \tag{11a}$$

($\boldsymbol{B}$ is defined in (6)).

In view of (7), (10) and (11), $r_{ss}(\boldsymbol{A})$ can be defined equivalently as

$$r_{ss}(\boldsymbol{A}) = \min\{r \geq 0:\ \ \overline{\lambda}(G^c + r\,H) = 1,\ H \in \boldsymbol{H}\}. \tag{12}$$

Now recall the stability radius of $\boldsymbol{A}$ [4]

$$r_S(\boldsymbol{A}) = \min\{r \geq 0 : [A^C - r\,R,\ A^C + r\,R]\ \textit{is instable}\} \tag{13}$$

which can be written equivalently as

$$r_s(\boldsymbol{A}) = \min\{r \geq 0:\ \ \det(A^c + r\,B) = 0,\ B \in \boldsymbol{B}\}. \tag{14}$$

Introducing a shift of -1, $\overline{\lambda}(G^c + rH) = 1$ can be replaced with $\overline{\lambda}(G^c - I + rH) = 0$ (I is the identity matrix). Hence, (12) can be written in the form (14) as

$$r_{ss}(\boldsymbol{A}) = 1 + \min\{r \geq 0: \quad \det(G^c - I + r\,H) = 0, \quad H \in \boldsymbol{H}\}. \tag{15}$$

So it is seen that

$$r_{ss}(\boldsymbol{A}) = r_s(G^s - I + r\boldsymbol{H}). \tag{15a}$$

Now introduce the new matrices

$$C^c = G^s - I, \quad C = C^c + \boldsymbol{H}. \tag{16}$$

On account of (12) to (16), it is seen that

$$r_{ss}(\boldsymbol{A}) = r_s(\boldsymbol{C}). \tag{17}$$

Thus, we have proved the following result.

Theorem 1. The strong stability radius $r_{ss}(\boldsymbol{A})$ of the original non-symmetric $n \times n$ interval matrix A can be determined as in (17) by way of the stability radius $r_s(\boldsymbol{C})$ of the associated symmetric $2n \times 2n$ interval matrix $\boldsymbol{C}$ defined in (8), (11) and (16).

This theorem permits us to determine $r_{ss}(\boldsymbol{A})$ using the approach of [4]. Indeed, consider the following interval generalized eigenvalue problem

$$Hx = \mu H^0 x, \quad H \in \boldsymbol{H}, \quad H^0 = -C^c. \tag{18}$$

Since H and H^0 are symmetric, all eigenvalues μ_i of (18) are real. Let μ^* denote the maximum eigenvalue of (18), i.e.

$$\mu^* = \max\left\{\mu: Hx = \mu H^0 x, \quad H \in \boldsymbol{H}\right\}. \tag{19}$$

Then, on account of Lemma 1 in [4] we have the following result.

Theorem 2. The strong stability radius $r_{ss}(\boldsymbol{A})$ of $\boldsymbol{A}$ can be determined using the maximum eigenvalue μ^* of the associated symmetric $2n \times 2n$ interval generalized eigenvalue problem (18) as follows

$$r_{ss}(\boldsymbol{A}) = 1/\mu^*. \tag{20}$$

4. CONCLUSION

It has been shown that the strong stability radius $r_{ss}(\boldsymbol{A})$ of an $n \times n$ interval matrix $\boldsymbol{A}$ can be determining by formula (20). The maximum eigenvalue μ^* of the augmented $2n \times 2n$ generalized eigenvalue problem (18) can be found using the general methods from [4] or [5] (where the interval matrix involved is not necessarily sym-

metric). Since both methods (when applicable) have polynomial complexity, the present approach for determining $r_{ss}(\boldsymbol{A})$ has the same numerical complexity.

It is hoped that the ideas of the present paper can be extended to circuits or systems described by more involved mathematical models (e.g., [7] - [9]).

REFERENCES

[1] L. Kolev, Interval Methods for Circuit Analysis, World Scientific: Singapore, New Jersey, London, 1993.

[2] C. Jansson, and J. Rohn, „An algorithm for checking regularity of interval matrices", SIAM J. Matrix Anal. Appl., 20, pp.756–776, 1999.

[3] B. Polyak, „Robust linear algebra and robust aperiodicity", in Directions in Mathematical Systems Theory and Optimization, A. Rantzer and C. Byrnes, Eds. LNCIS, 2003, pp.249–260.

[4] L. Kolev, "Stability radius of linear interval parameter circuits", in Proceedings of the XXV International Symposium on Theoretical Electrical Engineering (ISTET-09), pp. 87–91, June 2009, Luebeck, Germany.

[5] L. Kolev, "Determining the range of real eigenvalues for the interval generalized eigenvalue problem", COMPEL (The international journal for computation and mathematics in electrical and electronic engineering), vol. 27, pp. 1463 – 1480, 2008.

[6] H. Ahn, and Y. Chen, "Exact maximum singular value calculation of an interval matrix", IEEE Trans. on Automatic Control, vol. 52, pp. 510–514, 2007.

[7] A. Cheremensky, "Geometrical nature of linear system design: Introduction to optimal linear closed-loop system design", *J. Theor. Appl. Mech.* (Sofia), 2, pp. 9-24, 1996.

[8] A.Cheremensky and V. Fomin, Operator Approach to Linear Control Systems, Kluwer Acad.: Dordrecht, Boston, London, 1996.

[9] A.Cheremensky, Newtonian Mechanics and Control, St. Petersburg State University - ICPCM: St. Petersburg, Sofia, 2006.

ANALYSIS AND SYNTHESIS OF PERTURBED SINGLE WELL DUFFING OSCILLATORS

Zhivko Dimitrov Georgiev

Department of Theoretical Electrical Engineering, Technical University of Sofia,
8 Kl. Ohridski Blvd., 1000 Sofia, Bulgaria, e-mail: zhdgeorg@tu-sofia.bg

Abstract: *The paper presents an analysis as well a synthesis of oscillator systems described by single well Duffing equations under polynomial perturbations of fourth degree. It has been proved, that such a system can have unique hyperbolic limit cycle. The analytical condition for the arising of a limit cycle and an equation giving the parameters of this limit cycle are obtained. Moreover, a method for the synthesis of oscillator systems of the considered type, having preliminarily assigned properties, is proposed. The synthesis consists in an appropriate choice of the perturbation coefficients in such a way, that the oscillator equation is to have in advance assigned limit cycle. Both the analysis and synthesis are performed with the aid of the Melnikov function.*

Keywords: Duffing equation, limit cycles, elliptic functions

1. INTRODUCTION

The paper presents an analysis as well a synthesis of oscillator systems described by perturbed single well Duffing equations and admitting self-sustained oscillations. The results obtained are based on the qualitative investigation of dynamical systems, whereupon the self-sustained oscillations are regarded as limit cycles on the phase plane.

The Melnikov theory allows establishing the existence, number, position shape and stability of the limit cycles. Using this theory, a single well Duffing equation under polynomial perturbations of 4^{th} degree, is analyzed. It has been proved that such an equation can have a unique hyperbolic (simple) limit cycle. The conditions for the arising of a limit cycle (or periodic oscillations) and establishing whether this limit cycle is stable, or unstable, have been derived. Moreover, a method for the synthesis of equations of the regarded type, having preliminarily assigned properties, is suggested.

According to the Melnikov theory, the problem of finding the limit cycles is replaced by an equivalent problem of finding the zeros of a given function – the Melnikov function for autonomous systems [2], [5]. In our case the unperturbed solution is presented by Jacobi elliptic functions and thus the expression of the Melnikov function is a function involving complete elliptic integrals. The main point in the analysis of the Melnikov function is to prove that a given ratio of the functions involving complete elliptic integrals is a monotonic function.

Below we give short information concerning the Melnikov theory, which is necessary for the further exposition. It is important to underline, that we analyze only the

case of first order perturbations, respectively first order bifurcations of limit cycles from a "centre" and a first order Melnikov function.

Consider the perturbed Hamiltonian system

$$\left|\begin{array}{l} \dot{x} = \partial H/\partial y + \varepsilon p(x,y,\varepsilon) \\ \dot{y} = -\partial H/\partial x + \varepsilon q(x,y,\varepsilon) \end{array}\right., \tag{1}$$

where the Hamiltonian $H = H(x,y)$ is an analytic function in $\mathbf{R}^2$, the perturbation functions $p(x,y,\varepsilon)$ and $q(x,y,\varepsilon)$ are analytic in $\mathbf{R}^2 \times \mathbf{R}$, ε is a small parameter, $|\varepsilon| << 1$, $(\dot{}) \equiv \mathrm{d}/\mathrm{d}t$ and t is the time.

Denote the solution of system (1) by $(x,y) = (\varphi_\varepsilon(t), \psi_\varepsilon(t))$ and then the solution of the unperturbed system (at $\varepsilon = 0$) is $(x,y) = (\varphi_0(t), \psi_0(t))$.

Further on, we will assume that the unperturbed system has at least one equilibrium point, which is a "centre", surrounded by a continuous one-parameter family of closed trajectories, parameterized by the constant levels set of the Hamiltonian function. Then a given closed trajectory, corresponding to the parameter h, has the following equation on the phase plane

$$\Gamma_0(h): H(x,y) = h \ , \ h \in \Omega \subset \mathbf{R}, \tag{2}$$

where Ω is an open set of values h, for which the trajectories are not degenerated in a point or a separatrix loop.

Denote the period of the unperturbed solution $(x,y) = (\varphi_0(t), \psi_0(t))$ corresponding to the closed trajectory $\Gamma_0(h)$ by $T_0(h)$.

Under these notions, the following function

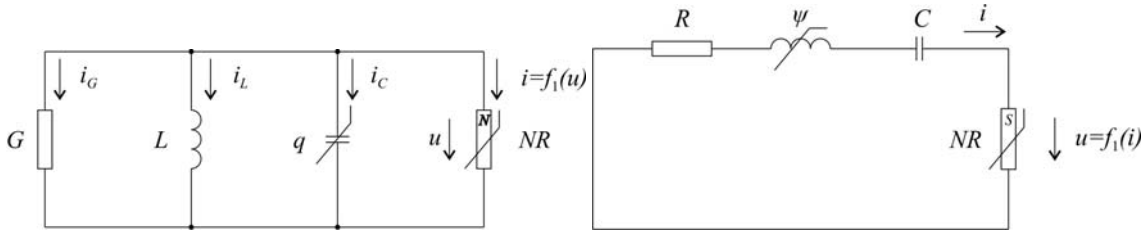
$$M(h) = \int_0^{T_0(h)} [q(\varphi_0, \psi_0, 0)\dot{\varphi}_0 - p(\varphi_0, \psi_0, 0)\dot{\psi}_0]\mathrm{d}t. \tag{3}$$

is called the first order Melnikov function. It plays an important role in the study of limit cycles. Every zero of the function $M(h)$ corresponds to the limit cycle and the stability of this limit cycle is determined by the derivative of the function $M(h)$ at this zero. More details can be found in [5] [6], [7].

In all computations further on, the complete elliptic integrals of the first and second kind, $\mathbf{K}$ and $\mathbf{E}$, are regarded as functions of k^2 (k is a modulus), i.e. $\mathbf{K} = \mathbf{K}(k^2)$ and $\mathbf{E} = \mathbf{E}(k^2)$. Regarding the Jacobi elliptic function sd and the complete elliptic integrals $\mathbf{K}$ and $\mathbf{E}$, some properties and identities that have been used in the presentation are given in the Appendix for the purpose of easy reference. The equations in the Appendix are marked by an A.

2. OSCILLATOR CIRCUITS AND OSCILLATOR EQUATIONS

Examples of AC equivalent circuits of oscillator systems governed by perturbed Duffing equations are shown in Fig. 1. The parallel circuit, shown in Fig. 1a, consists of a resistor with conductance G, an inductive element with inductance L, a nonlinear resistor NR of the type N, having voltage-current characteristic $i = f_1(u)$ and a nonlinear capacitive element having charge-voltage characteristic $u = \alpha q + \beta q^3$, where q is the electric charge associated with the element, and α, β are constants, $\alpha > 0$, $\beta > 0$. The series circuit, shown in Fig. 1b, consists of a resistor with resistance R, a capacitive element with capacitance C, a nonlinear resistor NR of the type S having current-voltage characteristic $u = f_1(i)$ and a nonlinear inductive element having flux-current characteristic $i = \alpha\psi + \beta\psi^3$, where ψ is the magnetic flux associated with the element, and here also α, β are constants, $\alpha > 0$, $\beta > 0$.



a) b)

Fig. 1. Oscillator circuits governed by perturbed Duffing equations

It can be proved that the processes in the oscillator circuits shown in Fig. 1 are described by the following perturbed single well Duffing equation

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} - \varepsilon \frac{\mathrm{d}F(x)}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}t} + ax + bx^3 = 0,\ a > 0,\ b > 0\ . \tag{4}$$

3. MELNIKOV FUNCTION FOR PERTURBED SINGLE WELL DUFFING EQUATION

In our work we assume that the function $F(x)$ is represented by the polynomial

$$F(x) = a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 . \tag{5}$$

In this case equation (4) can be rewritten as the following perturbed Hamiltonian system

$$\left| \begin{array}{l} \dot{x} = y + \varepsilon(a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4) \\ \dot{y} = -ax - bx^3 \end{array} \right. ,\ a > 0,\ b > 0 . \tag{6}$$

Further we will investigate the system (6) from a point of view of the existence of limit cycles.

Denote the solution of the system (6) by $(x,y)=\left(\varphi_\varepsilon(t),\psi_\varepsilon(t)\right)$. The unperturbed system (at $\varepsilon=0$) has a Hamiltonian

$$H(x,y)=\frac{1}{2}y^2+\frac{a}{2}x^2+\frac{b}{4}x^4 \tag{7}$$

and a single equilibrium point $(0,0)$ which is a "centre" surrounded by a continuous one-parameter family of closed trajectories given by

$$\Gamma_0(h)\colon H(x,y)=\frac{1}{2}y^2+\frac{a}{2}x^2+\frac{b}{4}x^4=h\ ,\ h\in(0,\infty)\ . \tag{8}$$

The solution of the unperturbed system $(x,y)=\left(\varphi_0(t),\psi_0(t)\right)$ in the time domain is expressed by the following Jacobi elliptic functions

$$x=\varphi_0(t)=\frac{\sqrt{2}\sqrt{a}k\sqrt{1-k^2}}{\sqrt{b}\sqrt{1-2k^2}}\,\mathrm{sd}\left(\frac{\sqrt{a}}{\sqrt{1-2k^2}}t,k\right), \tag{9a}$$

$$y=\psi_0(t)=\frac{\sqrt{2}ak\sqrt{1-k^2}}{\sqrt{b}(1-2k^2)}\,\mathrm{cd}\left(\frac{\sqrt{a}}{\sqrt{1-2k^2}}t,k\right)\mathrm{nd}\left(\frac{\sqrt{a}}{\sqrt{1-2k^2}}t,k\right), \tag{9b}$$

The period of the unperturbed solution is

$$T_0=T_0(h)=4\mathbf{K}(k^2)\frac{\sqrt{1-2k^2}}{\sqrt{a}}\ . \tag{10}$$

The modulus k and the constant h are related in the following way:

$$k^2=\frac{1}{2}\left(1-\frac{a}{\sqrt{a^2+4bh}}\right),\quad h=\frac{a^2k^2(1-k^2)}{b(1-2k^2)^2}\ , \tag{11a}$$

$$0<k^2<1/2\ ,\ 0<h<\infty\ . \tag{11b}$$

It is easy to prove that $\mathrm{d}h/\mathrm{d}k^2>0$, and $\mathrm{d}k^2/\mathrm{d}h>0$, i.e. the function $h=h(k^2)$ and $k^2=k^2(h)$ are monotone increasing.

According to Equation (3), the Melnikov function for the system (6) is

$$M(h)=-\int_0^{T_0}\left(\sum_{n=1}^{4}(a_n\varphi_0^n)\right)\dot{\psi}_0\,dt=\int_0^{T_0}\left[\left(\sum_{n=1}^{4}a_n\varphi_0^n\right)(a\varphi_0+b\varphi_0^3)\right]dt=$$

$$= \int_0^{T_0} \left[\sum_{n=2}^{7} c_n \left(\frac{\sqrt{2}\sqrt{a}k\sqrt{1-k^2}}{\sqrt{b}\sqrt{1-2k^2}} \right)^n \mathrm{sd}^n \left(\frac{\sqrt{a}}{\sqrt{1-2k^2}} t, k \right) \right] dt \ ,$$

where $c_2 = aa_1$, $c_3 = aa_2$, $c_4 = aa_3 + ba_1$, $c_5 = aa_4 + ba_2$, $c_6 = ba_3$, $c_7 = ba_4$. Introducing a new variable $z = \left(\sqrt{a} / \sqrt{1-2k^2} \right) t$ and taking into account Equation (10) and the properties of the function sd given in the Appendix (Equations (A1)–(A3)), we see that the integrals containing $\mathrm{sd}^{2n+1}(z,k)$, $n = 1,2,3$ vanish and the Melnikov function becomes

$$M(h) = \frac{4\sqrt{1-2k^2}}{\sqrt{a}} \int_0^{\mathbf{K}} \left[\sum_{n=1}^{3} c_{2n} \left(\frac{\sqrt{2}\sqrt{a}k\sqrt{1-k^2}}{\sqrt{b}\sqrt{1-2k^2}} \right)^{2n} \mathrm{sd}^{2n}(z,k) \right] dz \ . \qquad (12)$$

As it is seen from the last result, the Melnikov function and the arising of limit cycles do not depend on the terms with coefficients having odd indices - c_3, c_5, c_7. This means that the terms with even degrees in the polynomial perturbation in the system (6) exert no influence on the arising of limit cycles and further we will assume that $a_2 = a_4 = 0$. This fact is important with the circuit realization of the oscillator equation. In the final reckoning the system (6), which we investigate, is reduced with respect to the limit cycles, to the following system

$$\left| \begin{array}{l} \dot{x} = y + \varepsilon(a_1 x + a_3 x^3) \\ \dot{y} = -ax + bx^3 \end{array} \right| \ . \qquad (13)$$

Under these conditions, after quite long but straightforward computations, using Equations (A4)–(A6) and introducing the notation $m = k^2$, Equation (12) gives the final expression for the Melnikov function, i.e.

$$M(h) = M\big(h(m)\big) = \mathsf{M}\ (m) = \frac{8\lambda\sqrt{a}\sqrt{1-2m}}{(1-2m)^3} \mathsf{B}\,(m; a_1, a_3) \ , \qquad (14)$$

where $\lambda = a/b$ and

$$\mathsf{B}\,(m; a_1, a_3) = a_1 I_1(m) + \frac{2}{5} \lambda a_3 I_2(m) \ , \qquad (15)$$

$$I_1(m) = \frac{1}{3}(1-2m)\big[(1-m)\mathbf{K} - (1-2m)\mathbf{E}\big] \ , \qquad (16)$$

$$I_2(m) = 2(1-m+m^2)\mathbf{E} - (2-m)(1-m)\mathbf{K} \ . \qquad (17)$$

Recall that here and in the equations below, the complete elliptic integrals $\mathbf{K}$ and $\mathbf{E}$ are regarded as functions of m, where $m = k^2$ and k is a modulus, i.e. $\mathbf{K} = \mathbf{K}(m)$ and $\mathbf{E} = \mathbf{E}(m)$ [9].

According to the Melnikov theory, the zeros of $M(h)$ determine the limit cycles emerging from the periodic orbits of the unperturbed Hamiltonian systems. The stability of the limit cycle is determined by the sign of the quantity $\varepsilon(dM/dh)$. Since $h = h(m)$ and $m = m(h)$ are single-valued functions, the zeros of $M(h)$ and $\mathsf{M}(m)$ coincide. In other words, if m_0 is a zero of $\mathsf{M}(m)$, i.e. $\mathsf{M}(m_0) = 0$, then $h_0 = h(m_0)$ is a zero of $M(h)$, i.e. $M(h_0) = 0$. Moreover, from the relations

$$\frac{\mathrm{d}M}{\mathrm{d}h} = \frac{\mathrm{d}\mathsf{M}(m)}{\mathrm{d}h} = \frac{\mathrm{d}\mathsf{M}(m)}{\mathrm{d}m}\frac{\mathrm{d}m}{\mathrm{d}h}, \quad \frac{\mathrm{d}m}{\mathrm{d}h} > 0,$$

it follows that the signs of the quantities $\mathrm{d}M/\mathrm{d}h$ and $\mathrm{d}\mathsf{M}/\mathrm{d}m$ also coincide. Having this in mind, we will look for the zeros of the function $\mathsf{M}(m)$ and the sign of the quantity $\mathrm{d}\mathsf{M}/\mathrm{d}m$. On the other hand, with $m \in (0,1/2)$ the zeros of $\mathsf{M}(m)$ and $\mathsf{B}(m;a_1,a_3)$ coincide. Besides, if m_0 is a zero of $\mathsf{B}(m;a_1,a_3)$, i.e. $\mathsf{B}(m_0;a_1,a_3) = 0$, then the quantities $\mathsf{M}'(m_0)$ and $\mathsf{B}'(m_0;a_1,a_3)$ will have the same signs. Without any loss of generality, we may assume $a_1 = 1$ and put

$$B(m) := \mathsf{B}(m;1,a_3) = I_1(m) + \frac{2}{5}\lambda a_3 I_2(m). \tag{18}$$

Therefore we will look for the zeros of the function $B(m)$, or the roots of the equation $B(m) = 0$, and the signs of the derivative $B'(m)$ evaluated at these roots.

For further analysis it is convenient to rewrite the function $B(m)$ in the following way

$$B(m) = I_2(m)\left[Y(m) + \frac{2}{5}\lambda a_3\right], \tag{19}$$

where

$$Y(m) = \frac{I_1(m)}{I_2(m)} = \frac{\frac{1}{3}(1-2m)\left[(1-m)\mathbf{K} - (1-2m)\mathbf{E}\right]}{2(1-m+m^2)\mathbf{E} - (2-m)(1-m)\mathbf{K}}. \tag{20}$$

The analysis of the function $B(m)$ (respectively the analysis of the Melnikov function), requires the investigation of the function $Y(m)$.

5. ANALYSIS OF THE MELNIKOV FUNCTION AND LIMIT CYCLES

The properties of the function $Y(m)$ necessary for later use are summed up in the following lemma.

Lemma: The following assertions hold:

(a) $Y(0) = \lim_{m \to 0+} Y(m) = +\infty$, $Y(1/2) = 0$ and

$$Y(m) > 0 \text{ for } m \in (0, 1/2). \tag{21}$$

(b) The function $Y(m)$ is strictly monotone decreasing on the interval $(0, 1/2)$, i.e.

$$Y'(m) < 0 \text{ for } m \in (0, 1/2) \text{ .■} \tag{22}$$

The graph of the function $Y(m)$ is shown in Figure 2.

The following theorem gives the main results concerning the limit cycles analysis in perturbed single well Duffing oscillators.

Theorem: The system (13) (with $a_1 = 1$) can have a unique hyperbolic limit cycle. The condition for the arising of a limit cycle is the fulfillment of the following inequality

$$a_3 < 0 \text{ .} \tag{23}$$

The parameters of the limit cycle are determined by the equation

$$Y(m) + \frac{2}{5} \lambda a_3 = 0 \text{ .} \tag{24}$$

In addition, the limit cycle is stable when $\varepsilon > 0$ and unstable when $\varepsilon < 0$.■

The proofs of the formulated above lemma and theorem are ommited because of limited presentation space.

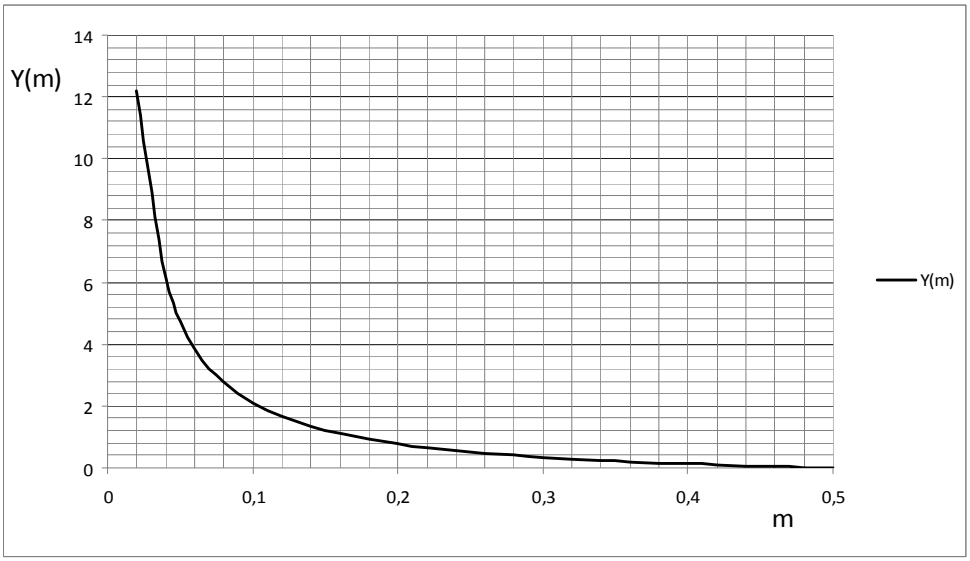


Fig. 2. Graphical plot of the function $Y(m)$

We will give some comments concerning the application of the above theorem for limit cycles analysis. Let m_0 be the solution of the equation (24). The value of m_0

determines the parameters of the limit cycle arising in the system (13). More precisely stated, for a sufficiently small $\varepsilon \neq 0$, there exists h_ε in an $O(\varepsilon)$ neighborhood of $h_0 = h(m_0)$ (see Equations (11a)), such that the perturbed Hamiltonian system (13) has a hyperbolic (simple) limit cycle $\Gamma_\varepsilon(h_\varepsilon)$. The limit cycle $\Gamma_\varepsilon(h_\varepsilon)$ is localized in an $O(\varepsilon)$ neighborhood of the curve $\Gamma_0(h_0)$ and tends to $\Gamma_0(h_0)$ as $\varepsilon \to 0$. The limit cycle $\Gamma_\varepsilon(h_\varepsilon)$ is stable when $\varepsilon > 0$, and unstable when $\varepsilon < 0$.

The limit cycle $\Gamma_\varepsilon(h_\varepsilon)$ in practice coincides with the curve $\Gamma_0(h_0)$ and its first order approximation in the time domain is expressed by the functions (9).

6. SYNTHESIS OF PERTURBED SINGLE WELL DUFFING OSCILLATORS

The presented result can be used to perform a synthesis of the considered oscillators. The relations between the Hamiltonian level h and the elliptic modulus k and $m = k^2$ given in Equation (11a) allow us to compute the value of m_0 corresponding to preliminarily assigned Hamiltonian level h_0. Then Equation (24), for given m_0, yields the coefficient a_3 in such a way, that the system (13) is to have a limit cycle localized in an $O(\varepsilon)$ neighborhood of this assigned Hamiltonian level. The **synthesis procedure** includes the following steps:

1) *Formulation of the synthesis problem*:

Find an oscillator system of the type (13) (with $a_1 = 1$), having a limit cycle localized in an $O(\varepsilon)$ neighborhood of the curve

$$\Gamma_0(h_0): \frac{1}{2}y^2 + \frac{a}{2}x^2 + \frac{b}{4}x^4 = h_0, \tag{25}$$

where the values of a, b, h_0 are given. We note that the preliminarily assigned values of a, b ($\lambda = a/b$) and h_0 form the desired characteristics of the oscillator. In other words, these values determine the shape, position and period of the limit cycle, which will arise in the system (13).

2) *Computation of the values of* m_0 *and modulus* k_0 *of the complete elliptic integrals*:

$$m_0 = \frac{1}{2}\left(1 - \frac{a}{\sqrt{a^2 + 4bh_0}}\right). \; k_0 = \sqrt{m_0}\,; \tag{26}$$

2) *Computation of the value of* $Y(m_0)$:

$$Y(m_0) = \frac{\frac{1}{3}(1-2m_0)\left[(1-m_0)\mathbf{K} - (1-2m_0)\mathbf{E}\right]}{2(1-m_0+m_0^2)\mathbf{E} - (2-m_0)(1-m_0)\mathbf{K}}\,. \tag{27}$$

4) *Computation of the perturbation coefficient* a_3:

$$a_3 = -\frac{Y(m_0)}{(2/5)\lambda} \ . \tag{28}$$

The obtained in this way perturbation coefficient a_3 ensures that the system (13) is to have a simple limit cycle $\Gamma_\varepsilon(h_\varepsilon)$, which is localized in an $O(\varepsilon)$ neighborhood of the curve (25). The limit cycle $\Gamma_\varepsilon(h_\varepsilon)$ is stable when $\varepsilon > 0$, and unstable when $\varepsilon < 0$.

Finally, as an application of the proposed method, we will briefly consider an example of a synthesis of an oscillator system. The computation of the numerical data follows the synthesis procedure.

Example: Find an oscillator system of the type (13) (with $a_1 = 1$), having a simple limit cycle which is localized in an $O(\varepsilon)$ neighborhood of the curve (25) with $a = 1$, $b = 2$ and $h_0 = 3$.

Solution:

1) $a = 1$, $b = 2$, $h_0 = 3$, $\lambda = a/b = 0.5$;

2) $m_0 = 0.4$, $k_0 = 0.632456$;

3) $\mathbf{K}(0.4) = 1.777519$, $\mathbf{E}(0.4) = 1.399392$, $Y(0.4) = 0.124667$;

4) $a_3 = -0.623335 \approx -0.623$.

In this way we find the following system

$$\left| \begin{array}{l} \dot{x} = y + \varepsilon(x - 0.623x^3) \\ \dot{y} = -x - 2x^3 \end{array} \right. . \tag{29}$$

The system (29) has a hyperbolic limit cycle $\Gamma_\varepsilon(h_\varepsilon)$, which in an $O(\varepsilon)$ approximation is presented by the curve $\Gamma_0(3): (1/2)y^2 + (1/2)x^2 + (1/2)x^4 = 3$. The generated by system (29) self-sustained oscillations are expressed practically in the time domain in the following way

$$x(t) = 1.095\,\mathrm{sd}\left(\frac{1}{\sqrt{0.2}}t, 0.632\right) . \tag{30}$$

The phase portrait of system (29) for $\varepsilon = 0.25$ and initial data $(0, 0.5)$ and $(0, 4)$ obtained by a numerical integration is shown in Fig. 3. The numerical computations confirm perfectly the analytical results.

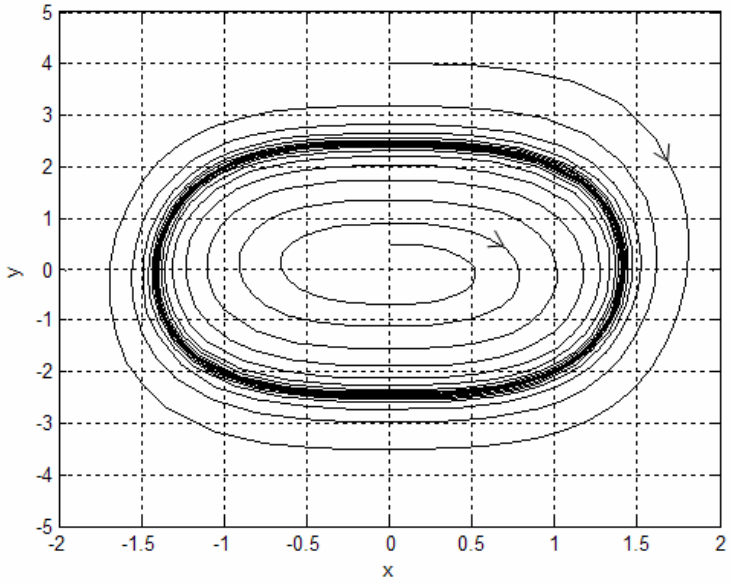


Fig. 3. Phase portrait of system (29)

7. CONCLUSIONS

An analysis, as well a synthesis, of oscillator systems described by single well Duffing equations under polynomial perturbations of 4^{th} degree is presented in this paper. It has been proved that the considered system can have a unique hyperbolic limit cycle. The obtained inequality (23) represents a condition for the arising of a limit cycle and the equation (24) determines the parameters of this limit cycle. Moreover, a method for a synthesis of oscillator systems of the considered type is proposed. The synthesis consists in an appropriate computation of the perturbation coefficients so that the prescribed properties to be fulfilled. Both the analysis and the synthesis are performed with the aid of the Melnikov function.

8. APPENDIX

In this section we give some properties and identities concerning the Jacobi elliptic function sd and the complete elliptic integrals of the first and second kind, $\mathbf{K}$ and $\mathbf{E}$, which are used in the presentation. Recall that the complete elliptic integrals $\mathbf{K}$ and $\mathbf{E}$ are regarded as functions of k^2, where k is a modulus, i.e. $\mathbf{K} = \mathbf{K}(k^2)$ and $\mathbf{E} = \mathbf{E}(k^2)$.

The function $\mathrm{sd} = \mathrm{sd}\, z = \mathrm{sd}(z,k)$ is a periodic and odd function with a period $4\mathbf{K}$. Then the following identities are valid

$$\mathrm{sd}(0,k) = \mathrm{sd}(4\mathbf{K},k) = 0 \ , \ \mathrm{sd}(z,k) = -\mathrm{sd}(z+2\mathbf{K},k) \ , \tag{A1}$$

$$\int_0^{4\mathbf{K}} \mathrm{sd}^{2n+1}(z,k)dz = 0 \ , \ n = 0,1,2,\cdots, \tag{A2}$$

$$\int_0^{4\mathbf{K}} \mathrm{sd}^{2n}(z,k)dz = 4\int_0^{\mathbf{K}} \mathrm{sd}^{2n}(z,k)dz \quad , \; n = 1,2,\cdots, \tag{A3}$$

The following identities are valid

$$\int_0^{\mathbf{K}} \mathrm{sd}^2 \mathrm{d}\, z = \frac{1}{k^2(1-k^2)}\left[\mathbf{E} - (1-k^2)\mathbf{K}\right], \tag{A4}$$

$$\int_0^{\mathbf{K}} \mathrm{sd}^4 \mathrm{d}\, z = \frac{1}{3k^4(1-k^2)^2}\left[2(2k^2-1)\mathbf{E} + (3k^4 - 5k^2 + 2)\mathbf{K}\right], \tag{A5}$$

$$\int_0^{\mathbf{K}} \mathrm{sd}^6 \mathrm{d}\, z = \frac{1}{15k^6(1-k^2)^3}\left[(23k^4 - 23k^2 + 8)\mathbf{E} + (15k^6 - 34k^4 + 27k^2 - 8)\mathbf{K}\right]. \tag{A6}$$

The equation (A4) is given in [10]. The equations (A5) and (A6) are obtained by using a recurrence dependency

REFERENCES

[1] A. A. Andronov, E. A. Leontovich, I. I. Gordon, A. G. Maier, Theory of Bifurcations of Dynamical Systems on the Plane, New York: Wiley, 1973.

[2] L. Perko, Differential Equations and Dynamical Systems, 3rd ed., New York: Springer, 2001.

[3] S. –N. Chow, C. Li, D. Wang, Normal Forms and Bifurcation of Planar Vector Fields, Cambridge: Cambridge University Press, 1994.

[4] C. Christopher, Ch. Li, Limit Cycles of Differential Equations, Basel: Birkhäuser Verlag, 2007.

[5] T. R. Blows, L. M. Perko, "Bifurcation of Limit Cycles from Center and Separatrix Cycles of Planar Analytic Systems", SIAM Review, vol. 36, № 3, pp. 341-376, 1994.

[6] C. Chicone, "On Bifurcation of Limit Cycles from Centers", in Lecture Notes in Mathematics, vol. 1455, New York: Springer, 1991, pp. 20-43.

[7] M. A. F. Sanjuan, "Lienard Systems, Limit Cycles, and Melnikov Theory", Physical Review E, vol. E57(1), pp. 340-344, 1998.

[8] M. Han, T. Zhang, "Some Bifurcation Methods of Finding Limit Cycles", Mathematical Biosciences and Engineering, vol. 3, № 1, pp. 67-77, 2006.

[9] M. Abramowitz, I. A. Stegun (eds), „Handbook of Mathematical Functions", National Bureau of Standards, Applied Mathematics Series, vol. 55, 1964.

[10] P. F. Byrd, M. D. Friedman, "Handbook of Elliptic Integrals for Engineers and Physicists", Berlin: Springer, 1954.

СИНТЕЗ И АНАЛИЗ НА ЛЕНТОВИ АКТИВНИ ФИЛТРИ ОТ ВТОРИ РЕД С ЕДИН ИЗТОЧНИК НА НАПРЕЖЕНИЕ, УПРАВЛЯВАН С НАПРЕЖЕНИЕ (ИНУН), МОДЕЛИРАНИ ПО СХЕМИТЕ НА САЛЕН–КИЙ, С ИЗПОЛЗВАНЕТО НА MATLAB И MICROCAP

Таня Методиева Стоянова[1], Адриана Найденова Бороджиева[2]

[1]Катедра „Теоретична и измервателна електротехника", [2]Катедра „Комуникационна техника и технологии",
Русенски университет „Ангел Кънчев",
България, 7017 Русе, ул. „Студентска" № 8,: [1](00359 82) 888 502, [2](00359 82) 888 734,
e-mail: [1]tstoyanova@ru.acad.bg, [2]aborodjieva@ecs.ru.acad.bg

Резюме: *В тази публикация се синтезират и анализират лентови активни филтри от втори ред с един източник на напрежение, управляван с напрежение, моделирани по схемите на Сален-Кий. Това се извършва по зададена нормализирана предавателна функция по напрежение и при известни нормализиращо съпротивление и средна честота. Синтезът се реализира чрез програмата MATLAB, а анализът на проектираните филтри след избора на стандартни стойности на съпротивленията на резисторите и на капацитетите на кондензаторите в синтезираните вериги, се извършва чрез програмата MicroCAP, предназначена за симулация на аналогови и цифрови вериги. Резултатите ще бъдат използвани в процеса на обучение по дисциплината „Комуникационни вериги", изучавана от студентите от специалност „Комуникационна техника и технологии" от образователно-квалификационната степен „бакалавър".*

Ключови думи: Анализ, синтез, активни филтри, източник на напрежение, управляван с напрежение (ИНУН).

1. ВЪВЕДЕНИЕ

Разглежданите активни филтри от втори ред с източник на напрежение, управляван с напрежение (ИНУН), моделирани по схемите на Сален-Кий [3, 7], са изградени от един операционен усилвател, резистори и кондензатори.

В Таблица 1 са дадени нормализираната предавателна функция по напрежение на активни лентови филтри (ЛФ) от втори ред, както и тяхната предавателна функция, записана чрез коефициента на усилване в лентата на пропускане k_0, кръговата средна честота ω_0 и полюсния качествен фактор Q на тези филтри [1, 2, 4, 5, 6].

Таблица 1. Предавателни функции по напрежение на ЛФ от втори ред

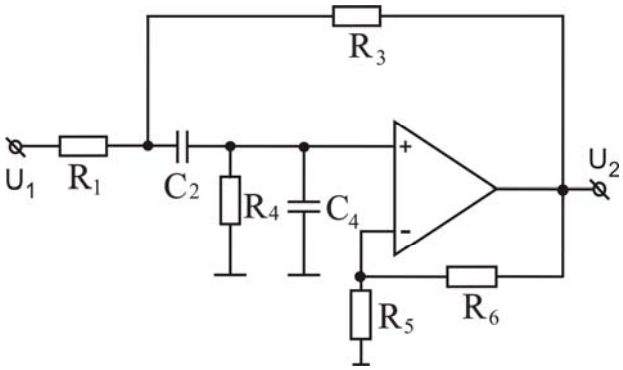
Активни филтри от втори ред		
Лентов филтър (ЛФ)	$T(p)=\frac{ap}{p^2+b_1p+b_0}$	$T(p)=\frac{k_0\frac{\omega_0}{Q}p}{p^2+\frac{\omega_0}{Q}p+\omega_0{}^2}$

2. СЪСТОЯНИЕ НА ПРОБЛЕМА – АНАЛИЗ И СИНТЕЗ НА ЛЕНТОВИ АКТИВНИ ФИЛТРИ ОТ ВТОРИ РЕД С ИНУН, МОДЕЛИРАНИ ПО СХЕМИТЕ НА САЛЕН-КИЙ

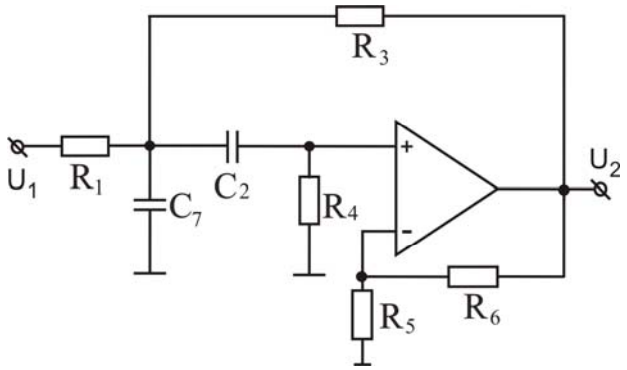
На фиг. 1 и фиг. 2 са дадени две разновидности на лентови активни филтри от втори ред с ИНУН, моделиран по схемите на Сален-Кий, за които операторните проводимости на участъците съответно са:

– за ЛФ – тип I – $Y_1 = G_1,\ Y_2 = pC_2,\ Y_3 = G_3,\ Y_4 = G_4 + pC_4$;

– за ЛФ – тип II – $Y_1 = G_1,\ Y_2 = pC_2,\ Y_3 = G_3,\ Y_4 = G_4, Y_7 = pC_7$.



Фиг. 1. Лентов активен филтър (тип I) от втори ред с ИНУН, моделиран по схемата на Сален-Кий



Фиг. 2. Лентов активен филтър (тип II) от втори ред с ИНУН, моделиран по схемата на Сален-Кий

3. РЕЗУЛТАТИ

Алгоритъмът за синтез и анализ лентов активен филтър от втори ред с ИНУН, моделиран по схемите на Сален-Кий, съдържа следните стъпки:

1. Въвеждане на коефициентите a, b_0 и b_1 в предавателната функция по напрежение $T(p)$ на синтезирания филтър.

2. Изчисляване на нормираните стойности на елементите на синтезирания лентов филтър от I и II тип, чиито схеми са показани на фиг. 2 и фиг. 3.

– за ЛФ – тип I: **Условие за проектиране** – еднакви капацитети в схемата: $C_2 = C_4 = 1$ (като нормирани стойности) и $G_1 = G_3 = G$.

– за ЛФ – тип II: **Условие за проектиране** – еднакви капацитети в схемата: $C_2 = C_7 = 1$ (като нормирани стойности), $G_1 = G_3 = G$.

Параметрите на останалите елементи се получават от решението на следните системи уравнения:

– за ЛФ – тип I: $\left| \begin{array}{l} \mu = \dfrac{a}{G},\ G_4 = \dfrac{b_0}{2G} \\ \dfrac{b_0}{2G} + (4-\mu)G = b_1 \end{array} \right.$,

– за ЛФ – тип II: $\left| \begin{array}{l} \mu \dfrac{G_1}{C_7} = a,\ \dfrac{G_1 + G_4 + (1-\mu)G_3}{C_7} + \dfrac{G_4}{C_2} = b_1 \\ \dfrac{G_4(G_1+G_3)}{C_2 C_7} = b_0 \end{array} \right.$.

Решенията на съответните системи са:

– за ЛФ – тип I:

Случай I:

$$\mu = 4 \Rightarrow G_1 = G_3 = G = \frac{b_0}{2b_1};\ G_4 = b_1;\ \mu = 1 + \frac{R_6}{R_5} = 4 \Rightarrow \frac{R_6}{R_5} = 3 \Rightarrow R_6 = 3R_5.$$

Случай II:

Квадратно уравнение по отношение на G с дискриминанта D: Необходими условия за проектиране:

1) $D = \left[-(b_1 + a)\right]^2 - 8b_0 = (b_1 + a)^2 - 8b_0 \geq 0$ – за наличие на реален корен;

2) $G > 0;\ G^{\mathrm{I,II}} = \dfrac{(b_1 + a) \pm \sqrt{(b_1 + a)^2 - 8b_0}}{8}$, избират се положителните корени (ако има такива); ако няма положителен корен – системата няма решение.

$G_1 = G_3 = G;\ G_4 = \dfrac{b_0}{2G}$.

– за ЛФ – тип II:

Случай I:

$$\mu = 2 \Rightarrow G_1 = G_3 = G = \frac{b_0}{b_1};\ G_4 = \frac{b_1}{2};\ \mu = 1 + \frac{R_6}{R_5} = 2 \Rightarrow \frac{R_6}{R_5} = 1 \Rightarrow R_6 = R_5.$$

Случай II:

Квадратно уравнение по отношение на G с дискриминанта D:

Необходими условия за проектиране:

1) $D = \left[-(b_1 + a)\right]^2 - 4.2.b_0 = (b_1 + a)^2 - 8b_0 \geq 0$ – за наличие на реален корен;

2) $G > 0$; $G^{\text{I,II}} = \dfrac{(b_1 + a) \pm \sqrt{(b_1 + a)^2 - 8b_0}}{4}$, избират се положителните корени (ако има такива); ако няма положителен корен – системата няма решение.

$G_1 = G_3 = G$; $G_4 = \dfrac{b_0}{2G}$.

1. Въвеждане на стойността на нормиращото съпротивление R_N.

2. Въвеждане на средната честота f_0 за лентовия филтър и изчисляване на нормиращата кръгова честота $\omega_N = 2\,\pi f_0$.

3. Изчисляване на денормираните стойности на елементите на двуполюсниците:

– за резисторите – получените стойности за съпротивления $R_k = 1/G_k$ за $k = 1...6$ се умножават с нормиращото съпротивление R_N;

– за кондензаторите – получените стойности за капацитети се разделят на произведението $\omega_N . R_N$.

4. Избор на стандартни стойности на елементите на филтъра.

5. Извеждане на предавателните функции по напрежение $T(p)$ на лентов филтър от I и II тип, съответно от фиг. 2 и фиг. 3.

– за ЛФ – тип I:

$$T(p) = \frac{\mu \dfrac{G_1}{C_4} p}{p^2 + p\left[\dfrac{G_1 + G_4 + (1-\mu)G_3}{C_4} + \dfrac{G_1 + G_3}{C_2}\right] + \dfrac{G_4(G_1 + G_3)}{C_2 C_4}},$$

където: $\left| \begin{array}{l} \mu \dfrac{G_1}{C_4} = a, \quad \dfrac{G_1 + G_4 + (1-\mu)G_3}{C_4} + \dfrac{G_1 + G_3}{C_2} = b_1 \\ \dfrac{G_4(G_1 + G_3)}{C_2 C_4} = b_0 \end{array} \right.$;

– за ЛФ – тип II:

$$T(p) = \frac{\mu \dfrac{G_1}{C_7} p}{p^2 + p\left[\dfrac{G_1 + G_4 + (1-\mu)G_3}{C_7} + \dfrac{G_4}{C_2}\right] + \dfrac{G_4(G_1 + G_3)}{C_2 C_7}},$$

където $\left| \begin{array}{l} \mu \dfrac{G_1}{C_7} = a, \dfrac{G_1 + G_4 + (1-\mu)G_3}{C_7} + \dfrac{G_4}{C_2} = b_1 \\ \dfrac{G_4(G_1 + G_3)}{C_2 C_7} = b_0 \end{array} \right.$.

6. Изчисляване на коефициента на усилване k_0 в лентата на пропускане, на полюсния качествен фактор Q и на средната честота f_0 за лентовия филтър от I и II тип, съответно от фиг. 2 и фиг. 3, след избора на стандартни стойности на елементите.

– за ЛФ – тип I:

$$\omega_0 = \sqrt{\frac{R_1 + R_3}{R_1 R_3 R_4 C_2 C_4}}, rad/s; \qquad f_0 = \frac{1}{2\pi}\sqrt{\frac{R_1 + R_3}{R_1 R_3 R_4 C_2 C_4}}, Hz,$$

$$Q = \frac{\sqrt{\frac{R_1 + R_3}{R_1 R_3 R_4 C_2 C_4}}}{\frac{1}{R_1 C_4} + \frac{1}{R_4 C_4} + (1-\mu)\frac{1}{R_3 C_4} + \frac{1}{R_1 C_2} + \frac{1}{R_3 C_2}};$$

$$k_0 = \frac{\frac{\mu}{R_1 C_4}}{\frac{1}{R_1 C_4} + \frac{1}{R_4 C_4} + (1-\mu)\frac{1}{R_3 C_4} + \frac{1}{R_1 C_2} + \frac{1}{R_3 C_2}}.$$

– за ЛФ – тип II:

$$\omega_0 = \sqrt{\frac{R_1 + R_3}{R_1 R_3 R_4 C_2 C_4}}, rad/s, \qquad f_0 = \frac{1}{2\pi}\sqrt{\frac{R_1 + R_3}{R_1 R_3 R_4 C_2 C_7}}, Hz;$$

$$Q = \frac{\sqrt{\frac{R_1 + R_3}{R_1 R_3 R_4 C_2 C_7}}}{\frac{1}{R_1 C_7} + \frac{1}{R_4 C_7} + (1-\mu)\frac{1}{R_3 C_7} + \frac{1}{R_4 C_2}}; \; k_0 = \frac{\frac{\mu}{R_1 C_7}}{\frac{1}{R_1 C_7} + \frac{1}{R_4 C_7} + (1-\mu)\frac{1}{R_3 C_7} + \frac{1}{R_4 C_2}}.$$

9. Симулационно изследване на синтезирания филтър с използване на програмния продукт MicroCAP – изчертаване на амплитудно-честотната характеристика, определяне на параметрите k_0, Q и f_0 от снетата амплитудно-честотна характеристика.

Разработени са скриптове на MATLAB [8] за изчисляване на нормализираните и денормализираните стойности на компонентите за НЧФ и ВЧФ при зададена нормирана предавателна функция по напрежение.

Пример: Проектиране на лентови филтри с ИНУН (от I и II тип), моделирани по схемата на Сален-Кий, със средна честота $f_0 = 1\,kHz$ и нормирана предавателна функция по напрежение $T(p) = \frac{3p}{p^2 + 0{,}25p + 1}$. Денормализирането по честота и по съпротивление се извършва с нормиращо съпротивление

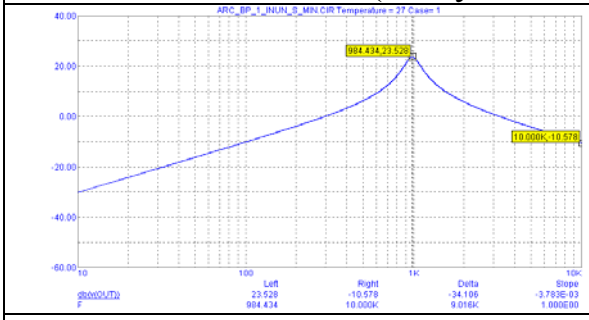
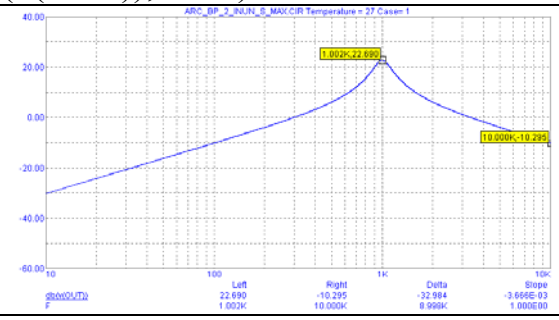
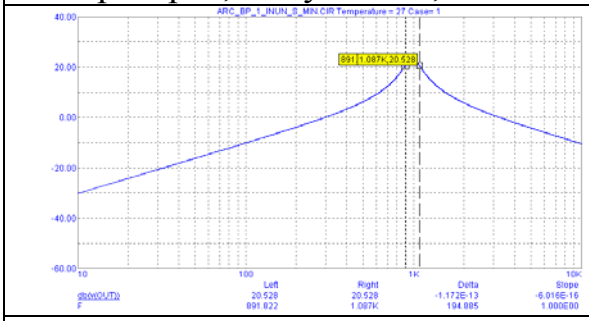
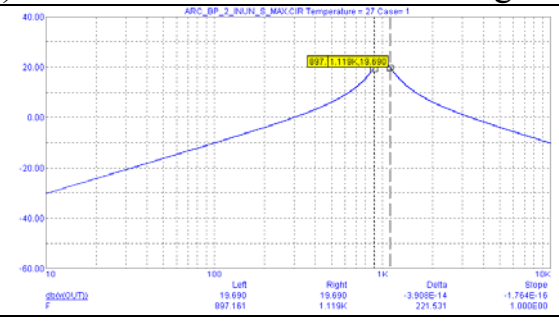
$R_N = 10\ k\Omega$. Резултатите са показани в таблица 2. След това е извършен избор на стандартни стойности по скалата Е-24, които се използват при симулацията с MicroCAP [9] за изчертаване на амплитудно-честотните характеристики (в dB) на проектираните филтри. В последната колона са посочени стойностите на коефициента на усилване k_0 в лентата на пропускане, на полюсния качествен фактор Q и на средната честота f_0 за лентовия филтър, а също и коефициента на усилване на ИНУН μ,след избора на стандартни стойности.

Таблица 2. Резултати от проектирането на ЛФ с ИНУН с MATLAB

Тип на филтъра	Нормализирани стойности	Денормализирани стойности	Стандартни стойности (Е-24)	Забележка
ЛФ I	$R_1 = 1{,}6492$ $C_2 = 1$ $R_3 = 1{,}6492$ $C_4 = 1$ $R_4 = 1{,}2127$ $\mu = 4{,}9477$	$R_1 = 16{,}4922\ k\Omega$ $C_2 = 15{,}9155\ nF$ $R_3 = 16{,}4922\ k\Omega$ $C_4 = 15{,}9155\ nF$ $R_4 = 12{,}1270\ k\Omega$ $R_5 = 10\ k\Omega$ * $R_6 = 39{,}4766\ k\Omega$	$R_1 = 16\ k\Omega$ $C_2 = 16\ nF$ $R_3 = 16\ k\Omega$ $C_4 = 16\ nF$ $R_4 = 12\ k\Omega$ $R_5 = 10\ k\Omega$ $R_6 = 39\ k\Omega$	По-големият корен на квадратното уравнение; $\mu = 4{,}9000$, $f_0 = 1{,}0152 kHz$, $Q = 3{,}7684$, $k_0 = 11{,}3077$.
	$R_1 = 4{,}8508$ $C_2 = 1$ $R_3 = 4{,}8508$ $C_4 = 1$ $R_4 = 0{,}4123$ $\mu = 14{,}5523$	$R_1 = 48{,}5078\ k\Omega$ $C_2 = 15{,}9155\ nF$ $R_3 = 48{,}5078\ k\Omega$ $C_4 = 15{,}9155\ nF$ $R_4 = 4{,}1230\ k\Omega$ $R_5 = 10\ k\Omega$ * $R_6 = 135{,}5234\ k\Omega$	$R_1 = 47\ k\Omega$ $C_2 = 16\ nF$ $R_3 = 47\ k\Omega$ $C_4 = 16\ nF$ $R_4 = 4{,}3\ k\Omega$ $R_5 = 10\ k\Omega$ $R_6 = 130\ k\Omega$	По-малкият корен на квадратното уравнение; $\mu = 14$, $f_0 = 989{,}5369 Hz$, $Q = 5{,}0262$, $k_0 = 15{,}0500$.
ЛФ II	$R_1 = 0{,}8246$ $C_2 = 1$ $R_3 = 0{,}8246$ $R_4 = 2{,}4254$ $C_7 = 1$ $\mu = 2{,}4738$	$R_1 = 8{,}2461\ k\Omega$ $C_2 = 15{,}9155\ nF$ $R_3 = 8{,}2461\ k\Omega$ $R_4 = 24{,}2539\ k\Omega$ $R_5 = 10\ k\Omega$ * $R_6 = 14{,}7383\ k\Omega$ $C_7 = 15{,}9155\ nF$	$R_1 = 8{,}2\ k\Omega$ $C_2 = 16\ nF$ $R_3 = 8{,}2\ k\Omega$ $R_4 = 24\ k\Omega$ $R_5 = 10\ k\Omega$ $R_6 = 15\ k\Omega$ $C_7 = 16\ nF$	По-големият корен на квадратното уравнение; $\mu = 2{,}5000$, $f_0 = 1{,}0028 kHz$, $Q = 4{,}5089$, $k_0 = 13{,}6364$.
	$R_1 = 2{,}4254$ $C_2 = 1$ $R_3 = 2{,}4254$ $R_4 = 0{,}8246$	$R_1 = 24{,}2539\ k\Omega$ $C_2 = 15{,}9155\ nF$ $R_3 = 24{,}2539\ k\Omega$ $R_4 = 8{,}2461\ k\Omega$	$R_1 = 24\ k\Omega$ $C_2 = 16\ nF$ $R_3 = 24\ k\Omega$ $R_4 = 8{,}2\ k\Omega$	По-малкият корен на квадратното уравнение; $\mu = 7{,}2000$,

	$C_7 = 1$ $\mu = 7{,}2762$	$R_5 = 10\ k\Omega$ * $R_6 = 62{,}7617\ k\Omega$, $C_7 = 15{,}9155\ nF$	$R_5 = 10\ k\Omega$ $R_6 = 62\ k\Omega$ $C_7 = 16\ nF$	$f_0 = 1{,}0028 kHz$, $Q = 3{,}7014$, $k_0 = 11{,}0149$.

Таблица 3. Резултати от изследването на ЛФ с ИНУН с MicroCAP

ЛФ I При използване на по-малките корени на системата	ЛФ II При използване на по-големите корени на системата
Определяне на средната честота на лентовия филтър (Analysis→AC→db(v(OUT)), Peak)	
	
$f_0 = 984{,}434\ Hz \quad T_0 = 23{,}528\ dB$	$f_0 = 1{,}002\ kHz \quad T_0 = 22{,}690\ dB$
Определяне на широчината на лентата на пропускане и на качествения фактор на филтъра Go to Y → $T_0 - 3$ (записва се конкретната стойност в появилото се прозорче, в случая: 20,528 или 19,690) → натиска се 2 x Left и 2 x Right	
	
Координати на левия маркер (Left): $f_{01} = 891{,}822\ Hz \quad T_{01} = 20{,}528\ dB$ Координати на десния маркер (Right): $f_{02} = 1{,}087\ kHz \quad T_{02} = 20{,}528\ dB$ Широчина на лентата на пропускане на филтъра: $B = Delta = f_{02} - f_{01} = 194{,}885\ Hz$ Качествен фактор на филтъра: $Q = \frac{f_0}{B} = \frac{891{,}822}{194{,}885} = 5{,}051$	Координати на левия маркер (Left): $f_{01} = 897{,}161\ Hz \quad T_{01} = 19{,}690\ dB$ Координати на десния маркер (Right): $f_{02} = 1{,}119\ kHz \quad T_{02} = 19{,}690\ dB$ Широчина на лентата на пропускане на филтъра: $B = Delta = f_{02} - f_{01} = 221{,}531\ Hz$ Качествен фактор на филтъра: $Q = \frac{f_0}{B} = \frac{1{,}002.10^3}{221{,}531} = 3{,}781$
Определяне на коефициента на усилване в лентата на пропускане на филтъра	
$k_0 = 10^{\frac{T_{0,dB}}{20}} = 10^{\frac{23{,}528}{20}} = 10^{1{,}1764} = 15{,}011$	$k_0 = 10^{\frac{T_{0,dB}}{20}} = 10^{\frac{22{,}690}{20}} = 10^{1{,}1345} = 13{,}630$

4. ИЗВОДИ

1. В публикацията е описан алгоритъм, заложен в програмен модул, с използване на MATLAB, създаден за синтез и анализ на лентови активни филтри от втори ред с ИНУН, моделирани по схемите на Сален-Кий с един операционен усилвател.

2. Изведени са изрази за предавателните функции по напрежение на синтезираните филтри посредством теорията на сигналните графи. Представени са и получените изрази за средната честота, за коефициента на усилване в лентата на пропускане и за полюсния качествен фактор на синтезираните филтри.

3. Разработеният програмен модул ще послужи и за автоматизиране на процеса на генериране на варианти на задания за курсови задачи по дисциплината „3110 Комуникационни вериги", включена като задължителна в новия учебен план на специалността „Комуникационна техника и технологии" за образователно-квалификационна степен „Бакалавър".

ЛИТЕРАТУРА

[1] Huelsman, L.P., P.E. Allen. "Introduction to the Theory and Design of Active Filters", Moscow, 1984 (in Russian).

[2] Manukova, A., A. Borodzhieva. "Communications Circuits – Handbook of Exercises", Rousse, 2002 (in Bulgarian).

[3] Stoyanov, G. "Fundamentals of Communication Engineering", Sofia, Techniques, 1993 (in Bulgarian).

[4] Stoyanova, T., A. Borodzhieva. Synthesis and Analysis of Single-Amplifier Single-Loop Negative Feedback Active Filters Using MATLAB and MicroCAP. Summer School, Advanced Aspects of Theoretical Electrical Engineering, Sozopol '07, 22.09. – 25.09.2007, Sozopol, Bulgaria, Regular papers, Part II, pp. 51 – 57.

[5] Stoyanova, T., A. Borodzhieva. Design and Computer Simulation of Low-Pass and High-Pass Single-Amplifier Multiple-Loop Negative Feedback Active Filters Using MATLAB and MicroCAP. Summer School, Advanced Aspects of Theoretical Electrical Engineering, Sozopol '07, 22.09. – 25.09.2007, Sozopol, Bulgaria, Regular papers, Part II, pp. 58 – 63.

[6] Stoyanova, T., A. Borodzhieva. Analysis and Synthesis of Band-Pass Single-Amplifier Multiple-Loop Negative Feedback Active Filters Using MATLAB and MicroCAP. Summer School, Advanced Aspects of Theoretical Electrical Engineering, Sozopol '07, 22.09. – 25.09.2007, Sozopol, Bulgaria, Regular papers, Part II, pp. 64 – 69.

[7] http://en.wikipedia.org/wiki/Sallen-key.

[8] http://www.mathworks.com.

[9] http://www.spectrum-soft.com/index.shtm.