

# Evolutionary Optimization of Sigma-Delta Modulated Analog Stimulus

Tomasz Golonek

**Abstract**—This paper proposes the evolutionary technique of the stimulus signal optimization for the analog electronic circuit testing purpose. The obtained signal is coded with Sigma-Delta modulation usage that allows to generate it easily by simple microcontrollers without the necessity of expensive D/A peripherals applying. The signal with the controlled impulses density may be obtained on the external output terminal of the typical timer and finally, it defines the analog signal that can be reconstructed after low pass filtering.

**Index Terms**—analog electronic circuits testing; evolutionary optimization; sigma-delta modulation; pulse-density modulation

## I. INTRODUCTION

THE analog electronic circuits (AECs) testing and diagnosing are important research domains [1-5]. The integrated analog circuit quality as well as the final device reliability depend directly on the control regime applied on each stage of the product life cycle. Many researches were concentrated on the cost effective techniques oriented to the AECs production and their build-in self testing procedures. The searching for this kind of concepts of the circuit under test (CUT) specification or/and elements correctness verification is still necessary. They are very useful, especially for the final product quality control, i.e. just after the SMD elements placement and the final device assembling on the production line, as well as for its earlier diagnostic supervising for the desired reliability of the device assurance.

Generally, the analog testing procedures may be sorted as the ones oriented to defected elements detection or to the ones which verify the correctness of the circuit specifications (like a cut-off frequency, a gain level or a signal phase shifting) [1, 2]. They are called fault driven testing (FDT) and specification driven testing approaches respectively (SDT). Of course, the FDT concept may be desired on the prototype testing to the faulty element location, however the SDT idea is suitable for the target specifications of the final product control. Besides, adequately to the moment of the CUT computer aided simulation process, AECs testing techniques may be sorted to concepts with simulations after test (SAT) and with simulations executed before test (SBT). Typically, the SAT methods assure precise diagnose with the fault location and its

identification, but unfortunately they consume much time on the testing stage. However, the SBT algorithms allow to decide about the AEC condition significantly much faster and they are adequate to after prototyping circuit stages.

Due to the phenomena of the produced circuit parameters tolerance dispersion (with Gaussian probability distribution) and the measurements resolution limited by the signal to noise necessary margin, the circuit testing, and especially the analog one, is not a trivial task. This fact leads to the motivation of heuristic computational algorithm usage [4-6]. The exemplary, previously published concepts engage the neural networks and ant colony algorithms as well as evolutionary techniques (genetic algorithm, evolutionary strategies, genetic programming, etc.).

The method described in this paper applies differential evolution (DE) to the optimal testing signal shape determination. The resultant excitation is reconstructed by means of 1-bit digital to analog signal conversion and it allows to avoid the requirements of the DAC usage, i.e. the typical timer built into the microcontroller ( $\mu\text{C}$ ) is enough for this purpose.

Section II contains the introduction to the applied method of pulse density modulation. In Section III the idea and its detailed description are placed. Next, in Section IV the two computational examples are presented. Finally, in Section V there are summarizing conclusions and discussion about the possible future modifications for the proposed approach.

## II. PULSE DENSITY MODULATION

### A. Motivation and General Concept

The Sigma-Delta ( $\Sigma\text{-}\Delta$ ) technique [8-10] is widely used in practice and it codes the analog value by means of pulse density modulation (PDM). The  $\Sigma\text{-}\Delta$  converters need 1-bit quantizer only and they apply high oversampling of the analog signal methodology. The oversampling fact allows to achieve better conditions for anti-aliasing and it gives the possibility of low order analog signal reconstructing filter applying. This concept greatly decreases the complexity and cost of the analog part of AD converter, i.e. inconvenient precise component matching and its trimming is not necessary there, however it is still needed for the alternative kind of converters (e.g. with successive approximation or flash method) for a possibility of achieving good resolution [8].

### B. Introduction to Pulse Density Modulation

The Fig. 1 presents the diagram of  $\Sigma\text{-}\Delta$  signal processing idea. The last obtained pulse is compared to the current momentary value of the analog signal and the resultant from

This work was supported by the Ministry of Science and Higher Education funding for statutory activities (decision no. 8686/E-367/S/2015 of 19 February 2015).

Tomasz Golonek is with the Institute of Electronics, Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, Gliwice, 16, Akademicka Str., 44-100, Poland (e-mail: Tomasz.Golonek@polsl.pl).

subtracting error level is integrated there. Next, the integrator output voltage is quantized by a 1-bit converter that references the accumulated voltage to the zero threshold level. Its output is set to high or low value in case of positive or negative thresholds detections respectively. The PDM digital signal is finally obtained and it allows to easily reconstruct the analog original one by a simple low pass filtering.

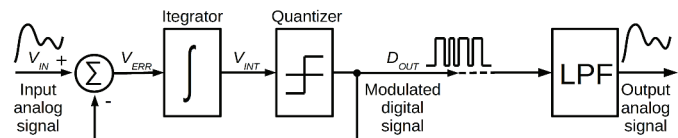


Fig. 1. The first order Sigma-Delta conversion.

The next Fig. 2 illustrates the waveforms for the  $\Sigma$ - $\Delta$  processing track selected points of the system from Fig. 1 and for the exemplary input analog signal voltage value  $V_{IN} = 1V$ . As can be seen there, the system tries to keep  $V_{INT}$  level (accumulated error) possibly near to zero and it leads to the adequate state of the output quant  $D_{OUT}$  synchronous forcing. Finally, the 1-bit output digital stream has impulses with the duration times of high or low level regulated respectively to the target  $V_{IN}$  value compensation. Decomposition of the discrete values of  $D_{OUT}$  signal obtained for exemplary sinusoid coded by  $\Sigma$ - $\Delta$  algorithm is presented in Fig. 3. The  $D_{OUT}$  discrete values density is adequate to the target analog value  $V_{IN}$ . The highest concentrations of positive and negative ones lie near to maximum and minimum areas of the analog signal respectively.

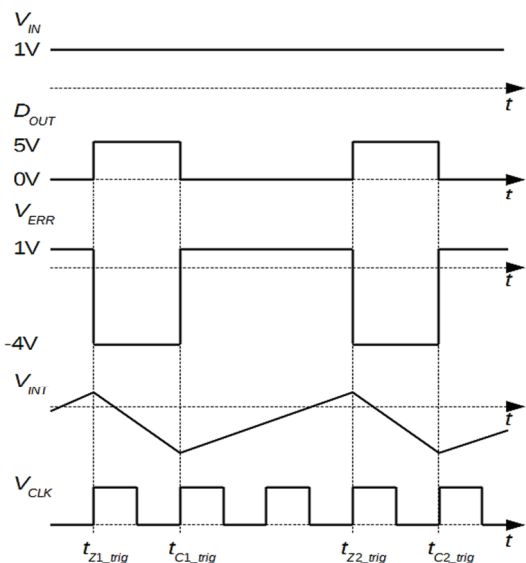


Fig. 2. The Sigma-Delta converter waveforms for the exemplary input signal.

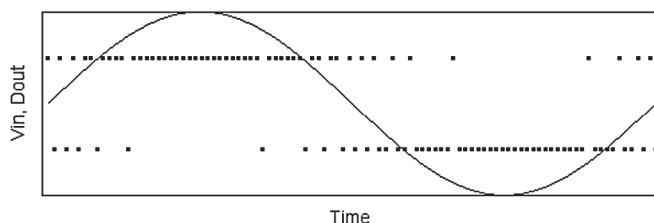


Fig. 3. The Sigma-Delta discrete values decomposition for the sinus signal.

As Fig. 2 shows, each output change of the  $\Sigma$ - $\Delta$  converter is synchronized by clock  $V_{CLK}$  that determines the sampling frequency  $f_s$  of the process. To assure the desired level of SNR for the reconstructed analog signal within the assumed frequency band  $f_c$ , there has to be:

$$f_s \gg f_c, \tag{1}$$

e.g. 64 times of the acoustic signal band is used for the Super Audio Coding standard (SACD). Currently available technologies of the integrated circuit production allow to rise to the high processing speed necessity challenge, additionally the oversampling technique has clear advantages and it decides about this conversion technique usefulness.

### III. TESTING APPROACH DESCRIPTION

#### A. Concept Explanation

The  $\mu$ Cs equipped with the timer connected to the output terminal and a high speed of the external signal generating possibility are widely used today. However, DAC peripheral is not always embedded there because it effects the increase of  $\mu$ C cost. Besides, the flash memory has satisfactory size typically and some part of it may be dedicated to the sequence of bits storing for the specialized testing excitation coding purpose. The mentioned features were the reason of the  $\Sigma$ - $\Delta$  modulation engaging directly for the testing stimulus generating, i.e. without embedded DAC usage. Besides, due to the high oversampling frequency of the original signal, this kind of modulation allows to easily reconstruct the analog excitation by means of a simple RC low pass filtering. The idea of this approach is presented in Fig. 4.

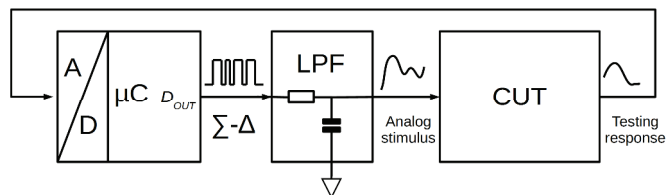


Fig. 4. Testing stimulus generating by means of impulses density modulation.

During the testing stage, the specialized analog excitation is driven to the CUT input and the obtained testing response, after its analog to digital conversion, is analyzed in  $\mu$ C. The shape of this stimulus is determined on before test stage by means of evolutionary system described in the next section.

The final decision about the analog circuit diagnostic state depends on the observed features of the retrieved output signal. In contrary to the first version of this approach from [7], for the system proposed in this paper the absolute error  $E_{TR}$  between the CUT testing response and the ambiguity region (AR) defined for the healthy circuits set was selected for the examination purpose and it is calculated from  $N$  discrete time points of the error function  $e(t)$ :

$$E_{TR} = \sum_{n=0}^{N-1} |e(n)|, \tag{2}$$

The above parameter (2) is easy for calculation, i.e. it needs  $N$  additions only, and its level is function of the error signal  $e(t)$  energy. The number of points  $N$  of discrete signal  $e(n)$  may be calculated from the duration time  $T_D$  of the designed testing stimulus and the frequency  $f_s$  of its sampling:

$$N = T_D \cdot f_s. \quad (3)$$

The general idea of  $E_{TR}$  determination is explained in Fig. 5 where the top and the bottom envelopes of testing responses set obtained from the simulations of healthy CUTs are denoted by squares and circles respectively and they define AR. The level of CUT absolute error (2) depends on the total surface of areas which exceed the allowed AR of the time response  $y(t)$  registered during the testing stage, i.e. it is calculated as integration of the absolute  $e(n)$  values of  $N$  samples placed in undesired regions above and below adequate envelopes (like an exemplary I and II ones in the Fig. 5).

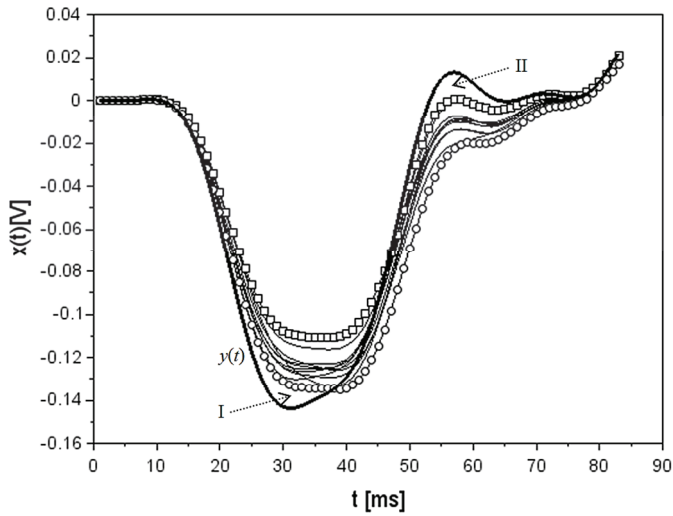


Fig. 5. The exemplary ambiguity region and one response of faulty CUT.

If the value of (2) is higher than the assumed percentage  $\delta_{TH\%}$  of the response signal energy obtained for the unfaulty CUTs averagely then it is classified to faulty group (CF):

$$E_{TR} > \frac{\delta_{TH\%}}{100\%} E_0 \Rightarrow CF, \quad (4)$$

or to the healthy one (CT) otherwise (GO-NO GO testing).

### B. Before Test Stage Procedure

During the before test stage, the shape of the specialized stimulus is optimized and this process needs a few steps. Firstly, the simulations of the circuit considered for the testing have to be started. The proposed technique is dedicated to the selected specifications of AC response examination (e.g. cut-off frequency, attenuation or phase shift values at the selected frequency point(s)), so this kind of analysis should be executed  $I$  times with elements values (CUT parameters) dispersed with the assumed deviations  $\delta_{MC\%}$  (Monte Carlo randomization) and finally it allows to collect the set  $\mathbf{K}_C$  of frequency responses (simulations results determined in

frequency points  $n = 0, \dots, N-1$ ):

$$\mathbf{K}_C = \{K_0(jn\omega), \dots, K_{(I-1)}(jn\omega)\}. \quad (5)$$

Next, the obtained set (5) is grouped into the two subsets of  $J$  elements for each one ( $J < I$ ): with the acceptable values of the observed specifications  $\mathbf{K}_{CH}$  (the healthy CUT patterns) and to the faulty patterns  $\mathbf{K}_{CF}$  otherwise:

$$\begin{aligned} \mathbf{K}_{CH} &= \{K_{CH0}(jn\omega), \dots, K_{CH(J-1)}(jn\omega)\}, \\ \mathbf{K}_{CF} &= \{K_{CFJ}(jn\omega), \dots, K_{CF(2J-1)}(jn\omega)\}, \end{aligned} \quad (6)$$

where  $\{\mathbf{K}_{CH} \cup \mathbf{K}_{CF}\} \subset \mathbf{K}_C$  and  $\{\mathbf{K}_{CH} \cap \mathbf{K}_{CF}\} = \emptyset$ . These subsets are used during the training stage (evolutionary computations) for their isolation level (fault coverage) examination.

Later, to determine the impulse responses  $h_k(n)$  for all frequency characteristics from the training set ( $k=0, \dots, 2J-1$ ), the IDFT operation is made adequately:

$$\begin{aligned} \mathbf{h} &= \{\mathbf{h}_{CH}, \mathbf{h}_{CF}\} = IDFT\{\mathbf{K}_{CH} \cup \mathbf{K}_{CF}\}, \\ \mathbf{h} &= \{h_{CH0}(n), \dots, h_{CH(J-1)}(n), h_{CFJ}(n), \dots, h_{CF(2J-1)}(n)\}, \end{aligned} \quad (7)$$

and it allows to realize all next necessary calculations during the before as well as the after testing stages in the time domain only. Now, for each candidate from population of the testing stimulus, respective  $k$ -th testing response may be determined by means of discrete convolution theorem:

$$y(n) = x(n) * h_k(n), \quad (8)$$

where  $x(n)$  denotes the testing excitation obtained after demodulation of PDM signal coded by the individual from the population and  $h_k(n)$  is the CUT impulse response for its  $k$ -th diagnostic state.

During the  $E_{TR}$  determination stage, bottom  $u_B(n)$  and top  $u_T(n)$  envelopes of healthy CUTs time responses sets are determined firstly and they define the allowed ambiguity area of non-faulty circuit signatures. This AR is caused by the parameters of the real CUT tolerance dispersion and its shape depends on the applied stimulus  $x(n)$ :

$$\begin{aligned} u_B(n) &= \min(x(n) * h_{CH_i}(n)), \\ u_T(n) &= \max(x(n) * h_{CH_i}(n)), \\ i &= 0, \dots, J-1; n = 0, \dots, N-1. \end{aligned} \quad (9)$$

At last, the error signal  $e(n)$  may be calculated for the acquired testing response  $y(n)$  from formulas:

$$\begin{aligned} e(n) &= \frac{f_1(n) + |f_1(n)|}{2} + \frac{f_2(n) + |f_2(n)|}{2}, \\ f_1(n) &= u_B(n) - y(n), \\ f_2(n) &= y(n) - u_T(n), \\ n &= 0, \dots, N-1. \end{aligned} \quad (10)$$

The above equation (10) is implemented in the fitness function procedure and when the observed parameter  $E_{TR}$  calculations (2) process is finished for the all testing responses (for the each of  $K$  patterns), its distribution into population of genotypes is fully defined and the quality  $Q$  factor (11) may be obtained easily for each one.

### C. Differential Evolution

The PDM testing pattern has to be specialized for the correctness of the controlled CUT specifications detection and the differential evolution with DE2 recombination scheme [11] allows to optimize it effectively. This kind of heuristic algorithm imitates natural processes of selection and recombination which proceed in the population of individuals. The genotypes (codes) are delegated for the crossover with the criteria regulated by fitness function  $Q$  that evaluates the phenotypes (solutions) qualities. For the testing usefulness of the PDM waveform evaluation, the weighted equation that contains two components was applied in the described optimization algorithm:

$$Q = \alpha + \frac{\beta}{E_0}, \quad (11)$$

where  $\alpha$ ,  $\beta$  and  $E_0$  are the number of correctly isolated CUT faulty states, minimal energetic distance between the healthy and faulty ones obtained for all patterns from the training set and average energy of healthy circuit testing response respectively. The fitness value  $Q$  is maximized during the evolutionary calculations and its proposed structure allows to control the optimization process with assumed hierarchy, i.e. separation  $\alpha$  maximization has the highest priority and adequately the second component has the lower one.

Each genotype from the population of DE system codes  $\Sigma$ - $\Delta$  digital signal with PDM shape optimized there and its structure is illustrated in Fig. 6. Genes from all positions code the widths (their absolute values are the pulses durations denoted as the integer number of clock periods  $1/f_s$ ) and polarities (their negative or positive signs define low or high levels respectively) of the impulses which create PDM signal finally. The integer, signed values of genes are optimized evolutionarily to achieve the demodulated analog testing signal with the highest possible mark  $Q$ . After finishing the evolutionary cycles, the optimized stimulus is ready and it is coded by the best genotype found globally (i.e. this one from all created  $L$  generations).

Locus ( $m$ )	0	1	3	...	$M-1$
Allel ( $G_{MN}, G_{MN}+1, \dots, G_{MX}-1, G_{MX}$ )	$g_0$	$g_1$	$g_2$	...	$g_{(M-1)}$

Fig. 6. The genotype structure.

## IV. COMPUTATIONAL RESULTS

### A. First Example CUT

The testing technique was examined firstly on the exemplary CUT presented in Fig. 7, i.e. the same one as for the previous version of this presented in [7]. The four parameters of the filter are delegated for their correctness checking:

- DC gain  $A(0)$ ,
- frequency response overshoot  $A_{mx}$ ,
- cut-off frequency  $f_{3dB}$  and
- phase value at cut-off frequency point  $\varphi(f_{3dB})$ .

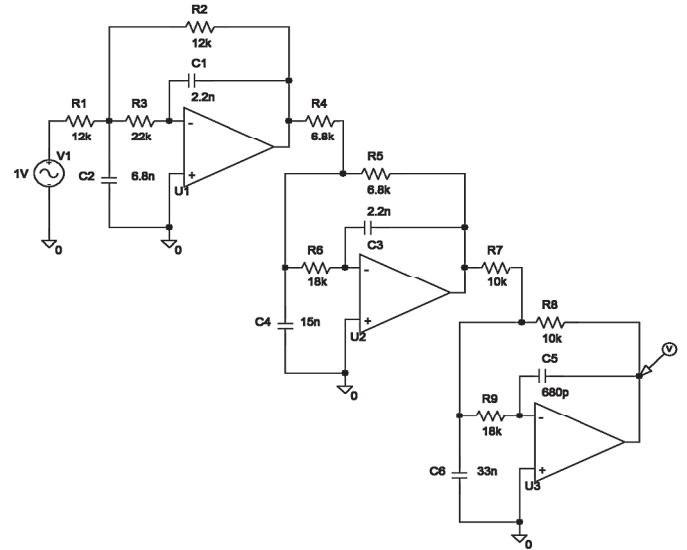


Fig. 7. The first CUT selected for the specifications testing.

Firstly, AC simulations were made and the subset of  $J=100$  patterns for healthy and faulty CUTs were prepared after Monte Carlo analysis ( $K=2 \cdot J = 200$  elements in training set totally). Frequency response of each pattern was defined in range from 1Hz up to 5120Hz at  $N=512$  points. The parameters of all discrete components (resistors and capacitors) were randomized within  $\delta_{MC}=20\%$  dispersion range. Finally, only the patterns with all above the tested specifications equal to the nominal ones or deviated not more than 10% from their nominal levels were included to the healthy group. However, the faulty group consists of patterns for which one or more of these parameters exceed the above threshold value  $\pm 10\%$  maximally, i.e. they are into the range ( $\pm 10\%, \pm 20\%$ ) from nominal value. The proposed structure of the testing patterns allows to evaluate the efficiency of specifications, soft faults isolation especially. However, its larger deviations from the nominal points (hard faults) are much easier to detection and they are not considered here. The Fig. 8 illustrates the exemplary amplitude responses obtained for the healthy and the faulty CUTs.

The evolutionary system of the PDM testing sequences optimization was started with the below settings:

- the size of population  $D = 20$  individuals,
- the last generation number  $L = 100$ ,
- the minimum value of gene  $G_{MN} = -32$ ,
- the maximum value of gene  $G_{MX} = 32$ ,
- the genotype length  $M = 1024$ ,
- the PDM signal clock frequency 2MHz,
- the LPF filter cut-off frequency 1kHz,
- the scheme DE2 weights  $W_1 = 0.5$ ,  $W_2 = 0.3$  (the best and the differential parents usage impacts).



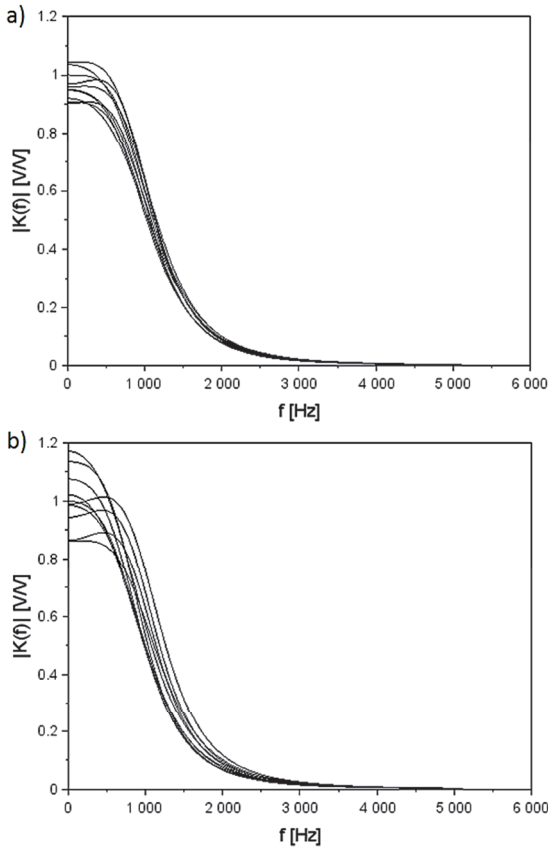


Fig. 8. The exemplary amplitude responses obtained from MC analyses of CUT: a) healthy, with parameters dispersed into tolerance region, b) faulty.

After the system start, the primary population was initialized with  $\Sigma$ - $\Delta$  sequences which code sinusoidal waveforms with random amplitudes, periods and phase shifts. The best genotype for the considered experiment found after 100 iterations codes the specialized analog testing stimulus presented in Fig. 9. This optimized excitation reaches  $\alpha = 98\%$  of the patterns isolation (GO/NO-GO specification driven test detection level) from the training set for the assumed classification threshold  $\delta_{TH}\% = 10\%$ . It proves the method efficiency enhancement obtained in contrary to the previous version of the algorithm [7] for that the fault coverage obtained for the same training patterns was equal to 81%. As can be seen, the parameter (10) engaged for the testing classification is more efficient and it involves minimal computations time.

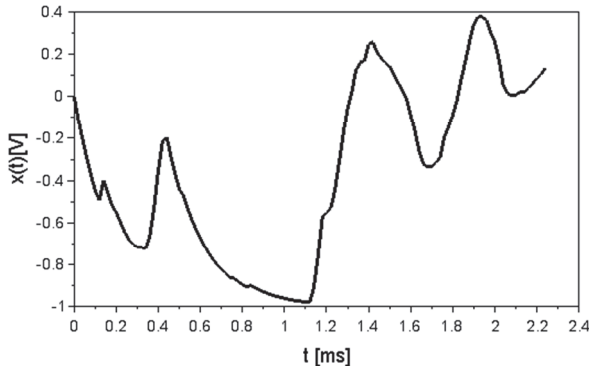


Fig. 9. The waveform of the optimized testing stimulus obtained for the first CUT.

Additionally, to check the generalization level of the approach, the best testing stimulus was examined on the new CUT sets created randomly with changed seed of MC random generator, i.e. another ones than these used during the evolutionary computations stage. Finally, the detection level reaches 95%, so just 3% of efficiency reduction appeared during the new patterns recognition and the GO-NO GO specifications of CUT testing level efficiency is still acceptable there. The detailed results are collected in Tab.1 where the values of quality factors  $Q$  and its components achieved after optimization cycles as well as for the newly created examination set are presented. Additionally, the maximal value of (2) and its variance in the group of results obtained for considered sets of patterns are presented there.

TABLE I  
BEST PHENOTYPES EFFECTIVENESS FOR THE FIRST CUT

Kind of set	$Q$	$\alpha$	$\beta = \min(E_{TR})$	$\max(E_{TR})$	$\text{var}(E_{TR})$
<b>Training</b>	98.114	98	0.056	3.236	58.659
<b>Random</b>	95.104	95	0.049	4.877	82.102

It should be emphasized that the examination set may contain patterns lying very close to unfaulty ones AR and this kind of cases is very difficult to detection. Besides, the additional advantages of this approach are its minimal cost and the fact that the multi-point frequency response measuring (e.g. by means of vobuloscope) and its analyzing is not necessary there. So, this kind of test is especially predisposed to the serial automatic testing on production stage or for the CUT self-testing purpose.

B. Second Example CUT

The second example is presented in Fig. 10. It is an active low pass „Leapfrog” filter, the one from benchmark circuits delegated for the analog testing methods examination and presented in [12]. All CUT specifications delegated for the testing and all parameters of evolutionary system are the same as in the previous one.

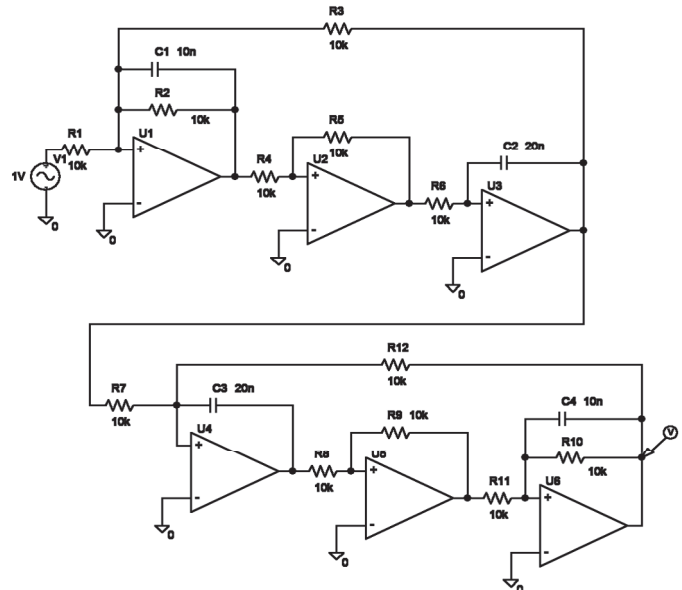


Fig. 10 The second CUT selected for the specifications testing.

The best found phenotype is illustrated in Fig. 11 and it allows to obtain 95% of fault coverage for the training set. The efficiency of the newly created patterns recognition reaches 87%. In this case, the generalization level is worse than previously (i.e. 8% of reduction appeared), but the final result is still good. All the results obtained for the CUT from Fig. 10 testing are collected in Table II.

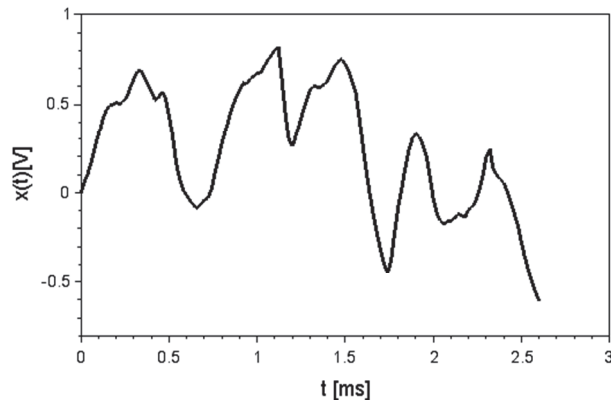


Fig. 11 The waveform of the optimized testing stimulus obtained for the second CUT.

TABLE II  
BEST PHENOTYPES EFFECTIVENESS FOR THE SECOND CUT

Kind of set	$Q$	$\alpha$	$\beta = \min(E_{TR})$	$\max(E_{TR})$	$\text{var}(E_{TR})$
<b>Training</b>	95.148	95	0.052	3.152	50.333
<b>Random</b>	87.101	87	0.039	2.876	44.564

## V. CONCLUSIONS

Considering the described technique simplicity, the obtained results are acceptable. Additionally, the CUT frequency specifications selected for the examination are not easy for fast determination in practice. Of course, a chirp signal may be engaged there, but it is a more expensive solution and it leads to the scanning necessity of the full band of the CUT (e.g. for the frequency response maximal gain location purpose). The future additional improvements of the method are possible there. For instance, the technique may be adapted to the faults identification purpose, e.g. by the parameter (2) decomposition (signatures) analysis. For this kind of faults identification ability improving, the variance of the obtained signatures set may be adequately optimized during the evolutionary computations and then the desired component(s) should be added to the fitness function (11) formula. Of course, the criterion delegated for the CUT diagnostic state classification purpose may be changed to another one and some additional processing procedures of the discrete error signal like a Walsh-Hadamard transformation can be applied, too. Finally, it may result with the technique efficiency improvement, but more advanced calculations engaging involves the method computation time increasing typically, so the testing system structure should be designed respectively to the assumed specific requirements.

## REFERENCES

- [1] L. S. Milor, "A Tutorial Introduction to Research on Analog and Mixed-Signal Circuit Testing", IEEE Transactions on Circuits and Systems-II, Analog and Digital Signal Processing 1998, Vol.45, No.10, pp. 1389-1407.
- [2] W. Toczek., Strategies of Analog Electronic Circuits Testing and Diagnosis, Publication of Gdańsk University of Technology, Poland, 2009 (In Polish).
- [3] T. Golonek, "Analog circuits testing by means of Walsh-Hadamard spectrum of supply current transient state monitoring ", The 6th International Conference on Human System Interaction (HSI), 2013, Sopot, Poland, pp.401-406.
- [4] M. Tadeusiewicz, S. Hałgas, M. Korzybski, "Multiple catastrophic fault diagnosis of analog circuits considering the component tolerances", International Journal of Circuit Theory and Applications, Vol.40, Issue 10, October 2012, pp.1041-1052.
- [5] T. Golonek, P. Jantos, J. Rutkowski, Stimulus with limited band optimization for analog circuits testing, Metrology and Measurement Systems, Vol.19, No.1, 2012, pp. 73-84.
- [6] T. Golonek, P. Jantos, "Memetic Method for Passive Filters Design", Chapter 3 in "Analog Circuits" InTech Book, edited by Dr. Yuping Wu, Prof. at Institute of Microelectronics, Chinese Academy of Sciences, ISBN 978-953-51-0930-3, January 2013, pp.51-68.
- [7] T. Golonek, "Analog testing stimulus optimization for generating by means of Sigma-Delta modulation", The International Conference on Mixed Design of Integrated Circuits & Systems (MIXDES), 2015, Torun, Poland, pp. 521-524.
- [8] Sangil Park, Principles of Sigma-Delta Modulation, Motorola Document APR8/D, Rev.1.
- [9] J. Rudy van de Plassche, "CMOS Integrated Analog-to-Digital and Digital-to-Analog Converters", Kluwer Academic Publishers, 1994.
- [10] S. Soclof, "Applications of analog integrated circuits", Prentice-Hall, 1985.
- [11] Ken Price, Rainer Storn, and Jouni Lampinen, "Differential Evolution - A Practical Approach to Global Optimization," Springer, Berlin, 2005.
- [12] B. Kaminska, K. Arabi, I. Bell, J.L. Huertas, B. Kim, A. Rueda, M. Soma, and P. Goteti, "Analog and Mixed-Signal Benchmark Circuits-First Release", in Proc. International Test Conference, ITC, 1997, pp.183-190.



**Tomasz Golonek** was born in Jaworzno, Poland, in 1973. He received the M.Sc. and Ph.D. degrees in electronics from the Silesian University of Technology, Gliwice, Poland, in 1998, and 2003 respectively. His interest domain are evolutionary computations, analog circuit testing and diagnosing, artificial intelligence algorithms and methods of data mining.