# Informacüje 

Journal of Microelectronics,

## MIDEM

Electronic Components and Materials Vol. 45, No. 2 (2015), June 2015

Revija za mikroelektroniko,
elektronske sestavne dele in materiale letnik 45, številka 2 (2015), Junij 2015

# Informacije MIDEM 2-2015 <br> Journal of Microelectronics, Electronic Components and Materials 

VOLUME 45, NO. 2(154), LJUBLJANA, JUNE 2015 | LETNIK 45, NO. 2(154), LJUBLJANA, JUNIJ 2015

Published quarterly (March, June, September, December) by Society for Microelectronics, Electronic Components and Materials - MIDEM. Copyright © 2014. All rights reserved. | Revija izhaja trimesečno (marec, junij, september, december). Izdaja Strokovno društvo za mikroelektroniko, elektronske sestavne dele in materiale - Društvo MIDEM. Copyright © 2014. Vse pravice pridržane.

Editor in Chief | Glavni in odgovorni urednik<br>Marko Topič, University of Ljubljana (UL), Faculty of Electrical Engineering, Slovenia<br>Editor of Electronic Edition | Urednik elektronske izdaje<br>Kristijan Brecl, UL, Faculty of Electrical Engineering, Slovenia<br>\section*{Associate Editors | Odgovorni področni uredniki}<br>Vanja Ambrožič, UL, Faculty of Electrical Engineering, Slovenia<br>Slavko Amon, UL, Faculty of Electrical Engineering, Slovenia<br>Danjela Kuščer Hrovatin, Jožef Stefan Institute, Slovenia Matjaž Vidmar, UL, Faculty of Electrical Engineering, Slovenia Andrej Žemva, UL, Faculty of Electrical Engineering, Slovenia<br>\section*{Editorial Board | Uredniški odbor}<br>Mohamed Akil, ESIEE PARIS, France Giuseppe Buja, University of Padova, Italy Gian-Franco Dalla Betta, University of Trento, Italy Martyn Fice, University College London, United Kingdom Ciprian Iliescu, Institute of Bioengineering and Nanotechnology, A*STAR, Singapore Malgorzata Jakubowska, Warsaw University of Technology, Poland Marc Lethiecq, University of Tours, France Teresa Orlowska-Kowalska, Wroclaw University of Technology, Poland Luca Palmieri, University of Padova, Italy<br>International Advisory Board |Časopisni svet<br>Janez Trontelj, UL, Faculty of Electrical Engineering, Slovenia - Chairman Cor Claeys, IMEC, Leuven, Belgium<br>Denis Đonlagić, University of Maribor, Faculty of Elec. Eng. and Computer Science, Slovenia<br>Zvonko Fazarinc, CIS, Stanford University, Stanford, USA<br>Leszek J. Golonka, Technical University Wroclaw, Wroclaw, Poland<br>Jean-Marie Haussonne, EIC-LUSAC, Octeville, France<br>Barbara Malič, Jožef Stefan Institute, Slovenia<br>Miran Mozetič, Jožef Stefan Institute, Slovenia<br>Stane Pejovnik, UL, Faculty of Chemistry and Chemical Technology, Slovenia Giorgio Pignatel, University of Perugia, Italy Giovanni Soncini, University of Trento, Trento, Italy Iztok Šorli, MIKROIKS d.o.o., Ljubljana, Slovenia<br>Hong Wang, Xi'an Jiaotong University, China<br>Headquarters | Naslov uredništva Uredništvo Informacije MIDEM MIDEM pri MIKROIKS Stegne 11,1521 Ljubljana, Slovenia T. +386 (0) 15133768 F. +386 (0) 15133771 E. info@midem-drustvo.si www.midem-drustvo.si

Annual subscription rate is 100 EUR, separate issue is 25 EUR. MIDEM members and Society sponsors receive current issues for free. Scientific Council for Technical Sciences of Slovenian Research Agency has recognized Informacije MIDEM as scientific Journal for microelectronics, electronic components and materials. Publishing of the Journal is cofinanced by Slovenian Book Agency and by Society sponsors. Scientific and professional papers published in the journal are indexed and abstracted in COBISS and INSPEC databases. The Journal is indexed by ISI® for Sci Search ${ }^{\oplus}$, Research Alert ${ }^{\oplus}$ and Material Science Citation Index ${ }^{\text {™ }}$.|
Letna naročnina je 100 EUR, cena posamezne številke pa 25 EUR. Člani in sponzorji MIDEM prejemajo posamezne številke brezplačno. Znanstveni svet za tehnične vede je podal pozitivno mnenje o reviji kot znanstveno-strokovni reviji za mikroelektroniko, elektronske sestavne dele in materiale. Izdajo revije sofinancirajo JAKRS in sponzorji društva. Znanstveno-strokovne prispevke objavljene v Informacijah MIDEM zajemamo v podatkovne baze COBISS in INSPEC. Prispevke iz revije zajema ISI ${ }^{\oplus}$ v naslednje svoje produkte: Sci Search ${ }^{\circledR}$, Research Alert ${ }^{\circledR}$ in Materials Science Citation Index ${ }^{\text {™ }}$.
Po mnenju Ministrstva za informiranje št.23/300-92 se šteje glasilo Informacije MIDEM med proizvode informativnega značaja.

Journal of Microelectronics,
Electronic Components and Materials vol. 45, No. 2 (2015)

## Content | Vsebina

| Original scientific paper |  | Izvirni znanstveni članki |
| :---: | :---: | :---: |
| M. E Başak, F. Kaçar:: A New Fully Integrated High Frequency Full-Wave Rectifier Realization | 101 | M. E Başak, F. Kaçar: <br> Nov polno integriran visokofrekvenčni polnovalni usmernik |
| K. Górecki, J. Zarębski, D. Bisewski: An Influence of the Selected Factors on the Transient Thermal Impedance Model of Power MOSFET | 110 | K. Górecki, J. Zarębski, D. Bisewski: <br> Vpliv določenih faktorjev na tranzienten termično impedančen model močnostnega MOSFET |
| P. B. Petrović: <br> Electronically Controllable Current-Mode True RMS to DC Converter | 117 | P. B. Petrović: Elektronsko nadzorovan RMS DC pretvornik v tokovnem načinu |
| B. Wang, Y. Zhuang, X. Li: <br> A Novel Dual Ports Antenna for Handheld RFID Reader Applications | 125 | B. Wang, Y. Zhuang, X. Li: <br> Nova dvovhodna antenna za ročne RFID bralnike |
| S. M. Djurić, N. M. Djurić, M. S. Damnjanović: The Optimal Useful Measurement Range of an Inductive Displacement Sensor | 132 | S. M. Djurić, N. M. Djurić, M. S. Damnjanović: Optimalno uporabno območje induktivnega senzorja premika |
| V. Sklyarov, I. Skliarova, A. Rjabov, A. Sudnitson: <br> Zynq-based System for Extracting Sorted Subsets from Large Data Sets | 142 | V. Sklyarov, I. Skliarova, A. Rjabov, A. Sudnitson: Sistem na osnovi Zynq za izluščitev razvrščenih podsklopov iz obsežnih podatkovnih sklopov |
| A. A. Demidov, O. A. Kalashnikov, A. Y. Nikiforov, A. S. Tararaksin, V. A. Telets: Radiation Behavior and Test Specifics of A-D and D-A Converters | 153 | A. A. Demidov, O. A. Kalashnikov, A. Y. Nikiforov, A. S. Tararaksin, V. A. Telets: Sevalno obnašanje in testne posebnosti A-D in D-A pretvornikov |
| Á. Bűrmen, H. Habal: Computing Worst-Case Performance and Yield of Analog Integrated Circuits by Means of Mesh Adaptive Direct Search | 160 | Á. Bűrmen, H. Habal: <br> Določanje najslabših lastnosti in izplena analognih integriranih vezij $z$ adaptivnim mrežnim direktnim optimizacijskim postopkom |
| Announcement and Call for Papers: 51st International Conference on Microelectronics, Devices and Materials With the Workshop on Terahertz and Microwave Systems | 171 | Napoved in vabilo k udeležbi: <br> 51. Mednarodna konferenca o mikroelektroniki, napravah in materialih z delavnico o teraherznih in mikrovalovnih sistemih |
| Front page: <br> The sensor element, detecting normal displacement (S. Djurić et al.) |  | Naslovnica: <br> Senzorski element za detektiranje normalnega premika (S. Djurić et al.) |

# A New Fully Integrated High Frequency FullWave Rectifier Realization 

Muhammed Emin Başak ${ }^{1}$, Firat Kaçar $^{2}$<br>${ }^{1}$ Yildiz Technical University, Faculty of Naval Architecture and Maritime, Istanbul, Turkey<br>${ }^{2}$ Istambul University, Dept. of Electrical and Electronics, Istanbul, Turkey


#### Abstract

In this paper, a new fully integrated high frequency precise full-wave rectifier which consists of a floating current source (FCS) and four complementary MOS transistors is presented. The presented circuit has an appropriate zero crossing performance, linearity, low component count, and can be adapted to modern IC technologies. It is also suitable for monolithic integrated implementation. Rectifier performance is simulated based on $0.18 \mu \mathrm{~m}$ CMOS technology. The proposed full-wave rectifier circuit provides an operating frequency more than 1 GHz , produces an input operating range from -300 mV to 300 mV and its power consumption is $825 \mu \mathrm{WW}$. LTSPICE simulation results of the circuit are presented which verify the workability of the proposed circuit. Noise analysis is also performed. The equivalent output noise of voltage mode rectifier at the 100 MHz is found as $7.13 \mathrm{nV} / \mathrm{Hz}$. It also exhibits good temperature stability. The presented circuit does not require any passive component; therefore it is suitable for integrated circuit implementation. The proposed circuit exhibits the high frequency operation, the lower power consumption and has the simplest structure compared to all other available works.


Keywords: CMOS; full-wave rectifier; high frequency; floating current source; precision rectifier

# Nov polno integriran visokofrekvenčni polnovalni usmernik 


#### Abstract

Izvleček: V članku je predstavljen nov polno integriran natančen polnovalen usmernik, ki je sestavljen iz plavajočega tokovnega vira in štirih komplementarnih tranzistorjev MOS. Predstavljeno vezje ima primerne lastnosti ničnega prehoda, linearnosti, nizkega števila elementov in se g alahko uporabi v vseh modernih IC tehnologijah. Uporaben je tudi v monolitno integriranih vezjih. Lastnosti usmernika so simulirane v $0.18 \mu \mathrm{~m}$ tehnologiji CMOS. Frekvenca predlaganega usmernika je več kot 1 GHz , vhodno območje od -300 mV do +300 mV , poraba $825 \mu \mathrm{~W}$. Opravljena je bila tudi analiza šuma. Usmernik ima dobro temperaturno stabilnost. Ker ne vsebuje pasivnih elementov je uporaben za integrirana vezja.


Ključne besede: CMOS; polnovalen usmernik; visoka frekvenca; plavajoči tokovni vir; natančen usmernik

* Corresponding Author's e-mail: fkacar@istanbul.edu.tr,


## 1 Introduction

Rectification is essential and demanding aspect of signal processing in instrumentation, measurement and control. Rectifiers have a variety of applications such as: signal processing, signal - polarity detectors, amplitude modulated signal detectors, AC voltmeters and ammeters, watt meters, RF demodulators, function fitting error measurements, RMS to DC conversions, sample and hold circuits, peak value detectors, clipper circuits.

Generally rectifier circuits are employed by using diodes, nevertheless, diodes cannot rectify the incoming signals whose amplitudes are less than their threshold voltages. For this reason, voltage-mode rectifiers containing active element based on operational amplifiers (op-amps),
diodes and resistors, have to be used. However, in consequence of the finite slew-rate and significant distortion during the zero crossing of the input signal effects caused by diode commutation, these circuits operate well only at low frequencies [1-5]. This is a small signal transient problem which cannot be solved by high slewrate op-amps [6]. This problem has been overcome by the use of current mode technique [7-18] thanks to their higher operating frequency, wider bandwidth, larger dynamic range, and lower offset value at the zero crossing area compared with their voltage mode counterparts. However, some proposed rectifier circuits which were improved by the use of current conveyors (CCs) need either grounded or ungrounded resistors or some of them suffer from the limitation of high frequency. The present-
ed circuit in [7] uses a current differencing transconductance amplifier (CDTA) and two diodes at the operating frequency of 5 MHz . CDTA-based precision full-wave rectifier described in [8] exhibits a good performance at a frequency of 5 MHz . The suggested circuits operating at a frequency up to 100 kHz utilize one current conveyor, one voltage conveyor, two diodes and grounded resistors [9-10]. The proposed circuit in [11] employs two differential difference current conveyors (DDCC), but it operates a few MHz . The circuit presented in [12], common-mode two-cell winner-takes-all (WTA) circuits, consisting of 21 transistors and two current sources, can be rectified at signals of frequency over 70 MHz . A single second generation current conveyor (CCCII-) based precision full-wave rectifier circuit is reported in [13]. It employs (CCCII-) with three outputs, two CMOS transistor, and an ungrounded resistor, and has an operating frequency of 100 kHz . The circuit presented in [14] employs three current controlled conveyors and five resistors having a testing frequency of 100 kHz . The reported circuit in [15] utilizes two current conveyors and three NMOS transistors and its operating frequency is up to 100 MHz . The proposed circuit in [16] employing a dual-X current conveyor and three NMOS transistors, has been successfully tested by applying a sinusoidal input voltage with a frequency of 250 kHz . The reported rectifiers in [17-19] have been realized by all CMOS transistors, but they are half wave rectifiers. The proposed circuit in [20] operat-
ing at a few MHz is based on current conveyor and current mirror.

The realization of full-wave rectifier based on an operational transconductance amplifier (OTA) circuits is proposed in [21-27]. However a large number of active and passive components are used in these rectifiers and they have not shown good performance at higher frequencies. In [24], OTAs utilized as the full-wave rectifier are the only active elements, whereas they have been tested at lower frequencies. A three output operational transconductance amplifier with two complementary MOS transistors and a grounded resistor is used to realize non-inverting and inverting full-wave precision rectifiers in [25]. It rectifies high frequencies up to 200 MHz . The circuit presented in [26] is more suitable for IC implementation than previously OTA based circuits and confirms the operation frequency up to 200 MHz . This circuit consists of a dual-output OTA, junction diodes, and a MOS resistor. Another rectifier circuit uses OTA, four CMOS diodes, and a MOS resistor in its realization, providing operating frequency up to 300 MHz as well as good temperature stability in [27]. Table 1 presents the comparison of the proposed precision full-wave rectifier with other designs. The employed full-wave rectifier is superior to the previously proposed full-wave rectifiers in terms of the power consumption, the number of components, and the operating frequency as seen in Table 1.

Table 1: Comparison of the various rectifiers in literature

| Article | DC Supply Voltage Voltage | Technology | Power Compsumption | Operating <br> Frequency | Components | Year |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Proposed | $\pm 2.4 \mathrm{~V}$ | $0.18 \mu \mathrm{~m}$ | $825 \mu \mathrm{~W}$ | 1 GHz | $8 \times$ MOSFET $+2 \times$ current sources | - |
| [3] | $\pm 1 \mathrm{~V}$ | - | - | 100 kHz | OPA1 + OPA $2+2 \times$ diodes $+3 \times$ Resistors | 2007 |
| [4] | $\pm 1 \mathrm{~V}$ | - | - | 1 MHz | AD817 $\times 2+$ AD633 $\times 3+$ AD711 + R | 2010 |
| [5] | $\pm 1 \mathrm{~V}$ | - | - | 1 MHz | $\begin{gathered} \text { AD817 } \times 2+\underset{ }{\text { AD } 633 \times 3} \text { Resistor } \end{gathered}$ | 2011 |
| [7] | $\pm 1 \mathrm{~V}$ | - | - | 5 MHz | CDTA $+2 \times$ Schotty Diodes | 2010 |
| [9] | - | - | - | 500 kHz | Current Conveyor + Voltage Conveyor | 2010 |
| [10] | $\pm 1 \mathrm{~V}$ | - | - | 1 MHz | $2 \times$ CCII +2 diodes or $\mathrm{CCII}+\mathrm{VC}+2$ diodes | 2011 |
| [11] | $\pm 2.5 \mathrm{~V}$ | $0.5 \mu \mathrm{~m}$ | - | 1 MHz | $2 \times$ DDCCI | 2011 |
| [13] | $\pm 2.5 \mathrm{~V}$ | - | - | 5 kHz | $\mathrm{CCCII}+2 \times \mathrm{CMOS}+\mathrm{R}$ | 2007 |
| [15] | $\pm 1.25 \mathrm{~V}$ | $0.25 \mu \mathrm{~m}$ |  | 10 MHz | 23 MOSFET | 2006 |
| [16] | $\pm 1.25 \mathrm{~V}$ | $0.25 \mu \mathrm{~m}$ | - | 1 MHz | DXCCII (20 CMOS) + $3 \times$ NMOS | 2008 |
| [19] | - | - | - | 200 MHz | 26 CMOS + 1 current supply | 2006 |
| [20] | $\pm 1.5 \mathrm{~V}$ | $0.5 \mu \mathrm{~m}$ | - | 10 MHz | 33 MOSFET | 2007 |
| [23] | $\pm 5 \mathrm{~V}$ | - | - | 10 kHz | $4 \times$ OTA or $5 \times$ OTA | 2007 |
| [24] | $\pm 5 \mathrm{~V}$ | $0.5 \mu \mathrm{~m}$ | - | 200 MHz | OTA (24MOSFET) + 2 MOS + Resistor | 2009 |
| [25] | $\pm 5 \mathrm{~V}$ | $0.5 \mu \mathrm{~m}$ | 7.9 mW | 300 MHz | 24 MOSFET | 2010 |
| [27] | $\pm 1.2 \mathrm{~V}$ | $0.5 \mu \mathrm{~m}$ | - | 250 MHz | 31 MOSFET | 2006 |

Floating current source (FCS) was firstly introduced to be used as an output stage for current-mode feedback amplifiers by Arbel and Goldminz in 1992 [28]. Following that, the FCS was used as the output stage of the accurate CCII- proposed in [29-30] to perform the required current conveying action. The FCS has also been used in the realization of fully differential voltage second generation current conveyor [31]. Then, [32] presented two novel floating current source based CMOS negative second generation current conveyor (CCII-).

In this paper, a new circuit for realizing full wave rectifier employing a floating current source, two CMOS diodes, and a MOS resistor, is proposed. The proposed circuit was simulated by LTSPICE simulator with 0.18 $\mu \mathrm{m}$ CMOS model obtained through TSMC (Taiwan Semiconductor Manufacturing Company, Limited). The advantages of the presented structure over the previously presented rectifiers are as follows:

- The presented structure is very compact and consists of an FCS and four CMOS transistors, thus enjoying a simpler structure compared to all other available works [1-27].
- The proposed circuit, verified the operation frequency up to 1 GHz , which is the highest frequency when compared with the previously published rectifiers.
- It does not require any passive component; therefore it is suitable for integrated circuit (IC) implementation.
- It provides high precision voltage rectifying.
- $\quad$ This rectifier has the lowest power consumption $(825 \mu \mathrm{~W})$ in comparison with the hitherto published rectifiers [1-27].


## 2 The Floating Current Source

Floating current source circuit can be viewed as two differential pairs connected in parallel; an NMOS pair and a PMOS pair. It is assumed that $M_{1}-M_{2}$ and $M_{3}-M_{4}$ are matched and operate in the saturation region for the NMOS pair and PMOS pair, respectively. Symbol of the FCS and its MOSFET implementation is shown in Fig. 1 (a) and Fig. 1 (b), respectively. [28] provides two balanced output currents satisfying Kirchhoff's current law. The equations of the output currents are given in below.
$I_{B 2}=I_{B 1}+I_{O 1}+I_{O 2}$
$I_{B 2}=I_{B 1}$
$I_{O 1}=-I_{O 2}$

(a)

(b)

Figure 1: (a) Symbol of floating current source circuit (b) MOSFET implementation of floating current source circuit [28]

It is assumed that $M_{1}-M_{2}$ and $M_{3}-M_{4}$ are equal transistors and so we can say that the transconductance of $M_{1}$ is equal of transconductance of $M_{2}\left(g_{m 1}=g_{m 2}\right)$ and transconductance of $M_{3}$ is equal of transconductance of $M_{4}\left(g_{m 3}=g_{m 4}\right)$. Then the transconductances of the FCS circuit ( $g_{\text {mo1 }}$ and $g_{\text {mo2 }}$ ) are given in Equation (4). The output impedances of the FCS structure are given in Equation (5).
$g_{m o 1}=-g_{m o 2}=\frac{\left(g_{m 1}+g_{m 3}\right)}{2}$
$R_{o 1}=R_{o 2}=\left[\left(\frac{g_{m 3} g_{d s 3}}{g_{m 3}+g_{m 4}}\right)+\left(\frac{g_{m 1} g_{d s 1}}{g_{m 1}+g_{m 2}}\right)\right]^{-1}=$
$=\frac{2}{g_{d s 1}+g_{d s 3}}$
Two balanced output currents are given by [33];

$$
\begin{align*}
& I_{o 1}=-I_{o 2}=-\frac{1}{2} V_{d}\left[\sqrt{k_{n}} \sqrt{I_{B 1}-\frac{k_{n} V_{d}^{2}}{4}}\right]+  \tag{6}\\
& +\sqrt{k_{p}} \sqrt{I_{B 1}-\frac{k_{p} V_{d}^{2}}{4}}
\end{align*}
$$

where $V_{d}=V_{1}-V_{2}$
$V_{1}$ and $V_{2}$ are the voltages applied to $Y_{1}$ and $Y_{2}$, respectively.
$\mathrm{k}_{\mathrm{n}}$ is the NMOS transconductance parameters given by

$$
\begin{equation*}
k_{n}=\mu_{n} C_{O X} \frac{W_{1}}{L_{1}} \tag{7}
\end{equation*}
$$

$k_{p}$ is the PMOS transconductance parameters given by

$$
\begin{equation*}
k_{p}=\mu_{p} C_{O X} \frac{W_{3}}{L_{3}} \tag{8}
\end{equation*}
$$

where
$\mu=$ mobility of carrier;
$C_{o x}=$ gate capacitance per unit area;
W, L = channel width and channel length of the MOS transistor, respectively; $I_{B 1} I_{B 2}=$ bias currents;

## 3 Proposed Full-Wave Rectifier Circuit

The basis of proposed full-wave rectifier is shown in Fig. 2. It is composed of three parts which are FCS, four diodes and resistor. A FCS is used to convert the voltage into two currents through the terminals $p$ and $n$, then the four diodes rectify these currents. Afterwards, resistor converts the rectified current into the output voltage. The voltage source $\mathrm{V}_{\mathrm{b}}$ is approximately equal to the sum of the threshold voltage of $D_{1}$ and $D_{2}$ and keep them ready for conduction [6].

Cross-section of NMOS and PMOS transistors in a psubstrate CMOS process are presented in Fig. 3. The structure of $M_{D 1}$ and $M_{D 2}$ are used to replace diodes $D_{1}$ to $D_{4}$ [26]. The diodes $D_{1}$ and $D_{2}$ are the junction diodes established between p -substrate and $\mathrm{n}+$ diffusion of the drain and source regions and $D_{3}$ and $D_{4}$ are the junction diodes established between n-well and $\mathrm{p}+$ diffusion of the drain and source regions. These diodes operate as a precision rectifier.


Figure 2: The principle of the proposed rectifier


Figure 3: Cross-section of $M_{D 1}$ and $M_{D 2}$ transistors in a p-substrate CMOS process [26].

The proposed high precision full-wave rectifier in detail is shown in Fig. 4. The FCS is consisted of four MOS transistors $\left(M_{1}-M_{4}\right)$ and two bias currents $I_{B 1}$ and $I_{B 2} . M_{D 1}$ and $M_{D 2}$ are used as four diodes as described the above and $M_{R 1}$ and $M_{R 2}$ are operated as a MOS resistor in the saturation region. The resistance value of MOS resistor $\mathrm{R}_{\mathrm{o}}$ can be expressed as;

$$
R_{O}=\frac{1}{2 \mu_{0} C_{O X}(W / L)\left(V_{G S}-V_{T H}\right)}{ }^{(9)}
$$

where, $\mathrm{V}_{\mathrm{TH}}$ is the threshold voltage and $\mathrm{V}_{G S}=\mathrm{V}_{\mathrm{DD}}=\left|\mathrm{V}_{S S}\right|$.

The relativity of the positive and negative polarity input voltage and output voltage can be expressed in below from the definition;
$V_{\text {in }}\left\langle 0 ; I_{\text {out }}=-\frac{g_{m 1}+g_{m 3}}{2} V_{\text {in }}, V_{\text {out }}=I_{\text {out }} \cdot R_{o}\right.$
$\left.V_{\text {in }}\right\rangle 0 ; I_{\text {out }}=\frac{g_{m 1}+g_{m 3}}{2} V_{\text {in }}, V_{\text {out }}=I_{\text {out }} \cdot R_{o}$
The proposed full-wave rectifier is simulated using the schematic implementation shown in Fig. 4. The voltage sources $\mathrm{V}_{\mathrm{DD} 1}$ and $\mathrm{V}_{\mathrm{SS} 1}$ are $\pm 2.4 \mathrm{~V}$ and $\mathrm{V}_{\mathrm{DD} 2}$ and $\mathrm{V}_{\text {SS2 }}$ are $\pm 0.75 \mathrm{~V}$. The TSMC $0.18 \mu \mathrm{~m}$ CMOS model parameters, which are shown in Table 2, are used in the simulations. The W/L parameters of MOS transistors are $1.98 \mu \mathrm{~m} / 0.18 \mu \mathrm{~m}$ for $\mathrm{M}_{1}, \mathrm{M}_{2} ; 1.36 \mu \mathrm{~m} / 1.36 \mu \mathrm{~m}$ for $\mathrm{M}_{3}, \mathrm{M}_{4}$; $0.36 \mu \mathrm{~m} / 0.36 \mu \mathrm{~m}$ for $\mathrm{M}_{\mathrm{D} 1^{\prime}}, \mathrm{M}_{\mathrm{D} 2^{\prime}}$ and $1.36 \mu \mathrm{~m} / 0.72 \mu \mathrm{~m}$ for $M_{R 1}, M_{R 2}$. Bias currents are $165 \mu A$ for both $I_{B 1}$ and $I_{B 2}$ and bias voltage $\mathrm{V}_{\mathrm{b}}$ is 0.7 V .


Figure 4: Proposed full-wave rectifier circuit.

## 4 Simulation Results

The DC transfer characteristic of the proposed full-wave rectifier is shown in Fig. 5, which shows the operating voltage ranging from -300 mV to 300 mV of the input voltage. Diodes and Op-amp based conventional rectifiers are resulting in significant distortion during the zero crossing of the input signal. The zero crossing region of the $D C$ transfer characteristic is shown in Fig. 6. In this figure, the offset is found as $54.8 \mu \mathrm{~V}$ which is the lowest value when compared to all other available works [1-27].

The input $\left(\mathrm{V}_{\text {in }}\right)$ and output signals of inverting full-wave precision rectifier at 100 MHz and 1 GHz frequencies are shown in Fig. 7. The period of the rectified signal ( T ), as seen as the stabilized region in this figure, is equal to 5 ns and 0.5 ns when the input signal's frequency is 100 MHz and 1 GHz , respectively. It can also be clearly observed from the figure that the behavior of the floating current source based full-wave rectifier is very satisfactory. Hence, the proposed rectifier can be used to process the signals at the frequencies up to 1 GHz without causing any major distortions. Applying the 300 $\mathrm{mV}_{\text {peak }}$ sine wave at the input of the proposed fullwave rectifier, the input and output signals at a frequency of 100 MHz is shown in Fig. 8. The power consumption of proposed rectifier is $825 \mu \mathrm{~W}$ which is the lowest value when compared to all other available works [1-27].

We simulate the temperature performance of the DC transfer characteristics of the proposed full-wave rectifier for the varied temperature ranging from $0^{\circ} \mathrm{C}$ to $100^{\circ} \mathrm{C}$ as shown in Fig. 9. The zero crossing regions of the DC transfer characteristic's among different temperatures are shown in Fig. 10. This figure shows that there is only a small amount of difference of the offsets among different temperatures. The offsets for different temperatures are obtained from the LTSPICE simulation results as shown in Fig. 10, given in Table 3.

Table 2: $0.18 \mu \mathrm{~m}$ TSMC model parameters used in the simulation

MODEL NMOS (LEVEL = 7 VERSION $=3.1$ TNOM $=27$ TOX=4.1E-9 XJ $=1 \mathrm{E}-7 \mathrm{NCH}=2.3549 \mathrm{E} 17 \quad$ VTH0 $=0.3725327$ $\mathrm{K} 1=0.5933684 \mathrm{~K} 2=2.050755 \mathrm{E}-3 \quad \mathrm{~K} 3=1 \mathrm{E}-3 \mathrm{~K} 3 \mathrm{~B}=4.5116437 \mathrm{~W} 0=1 \mathrm{E}-7 \quad \mathrm{NLX}=1.870758 \mathrm{E}-7 \quad \mathrm{DVTOW}=0 \quad \mathrm{DVT} 1 \mathrm{~W}=0$ DVT2W=0 DVT0 $=1.3621338$ DVT1 $=0.3845146$ DVT2 $=0.0577255$ U0 $=259.5304169$ UA $=-1.413292 \mathrm{E}-9 \quad$ UB $=$ $2.229959 \mathrm{E}-18$ UC $=4.525942 \mathrm{E}-11 \mathrm{VSAT}=9.411671 \mathrm{E} 4 \quad \mathrm{AO}=1.7572867 \mathrm{AGS}=0.3740333 \mathrm{~B} 0=-7.087476 \mathrm{E}-9 \mathrm{~B} 1=-1 \mathrm{E}-7$ KETA $=-4.331915 \mathrm{E}-3 \mathrm{~A} 1=0 \mathrm{~A} 2=1 \mathrm{RDSW}=111.886044 \mathrm{PRWG}=0.5 \quad \mathrm{PRWB}=-0.2 \mathrm{WR}=1 \mathrm{WINT}=0 \mathrm{LINT}=1.701524 \mathrm{E}-8 \mathrm{XL}=0$ XW=-1E-8 DWG=-1.365589E-8 DWB=1.045599E-8 VOFF=-0.0927546 NFACTOR=2.4494296 CIT=0 CDSC=2.4E-4 CDSCD $=0$ CDSCB $=0$ ETA0 $=3.175457 \mathrm{E}-3$ ETAB $=3.494694 \mathrm{E}-5$ DSUB $=0.0175288$ PCLM $=0.7273497$ PDIBLC1 $=0.1886574$ PDIBLC2 $=2.617136 \mathrm{E}-3 \quad$ PDIBLCB $=-0.1 \quad$ DROUT $=0.7779462 \quad$ PSCBE1 $=3.488238 \mathrm{E} 10 \quad$ PSCBE2 $=6.841553 \mathrm{E}-10$ PVAG=0.0162206 DELTA=0.01 RSH=6.5 MOBMOD=1 PRT=0 UTE=-1.5 KT1=-0.11 KT1L=0 KT2=0.022 UA1=4.31E-9 UB1 $=-7.61 \mathrm{E}-18$ UC $1=-5.6 \mathrm{E}-11 \mathrm{AT}=3.3 \mathrm{E} 4 \mathrm{WL}=0 \mathrm{WLN}=1 \mathrm{WW}=0 \mathrm{WWN}=1 \mathrm{WWL}=0 \quad \mathrm{LL}=0 \quad \mathrm{LLN}=1 \mathrm{LW}=0 \mathrm{LWN}=1 \mathrm{LWL}=0$ CAPMOD $=2$ XPART $=0.5$ CGDO $=8.53 \mathrm{E}-10 \quad \mathrm{CGSO}=8.53 \mathrm{E}-10 \quad \mathrm{CGBO}=1 \mathrm{E}-12 \mathrm{CJ}=9.513993 \mathrm{E}-4 \quad \mathrm{~PB}=0.8 \mathrm{MJ}=0.3773625$ CJSW=2.600853E-10 PBSW=0.8157101 MJSW=0.1004233 CJSWG=3.3E-10 PBSWG=0.8157101 MJSWG=0.1004233 $\mathrm{CF}=0 \quad$ PVTHO $=-8.863347 \mathrm{E}-4 \quad$ PRDSW $=-3.6877287 \quad \mathrm{PK} 2=3.730349 \mathrm{E}-4 \quad$ WKETA $=6.284186 \mathrm{E}-3 \quad$ LKETA $=-0.0106193$ PUO=16.6114107 PUA=6.572846E-11 PUB=0 PVSAT=1.112243E3 PETAO $=1.002968 \mathrm{E}-4$ PKETA $=-2.906037 \mathrm{E}-3$ ) MODEL PMOS ( LEVEL $=7$ VERSION $=3.1 \mathrm{TNOM}=27 \mathrm{TOX}=4.1 \mathrm{E}-9 \mathrm{XJ}=1 \mathrm{E}-7 \mathrm{NCH}=4.1589 \mathrm{E} 17$ VTHO $=-0.3948389$ $\mathrm{K} 1=0.576352 \mathrm{~K} 2=0.0289236 \mathrm{~K} 3=0 \mathrm{~K} 3 \mathrm{~B}=13.8420955 \mathrm{~W} 0=1 \mathrm{E}-6 \mathrm{NLX}=1.337719 \mathrm{E}-7$ DVTOW $=0$ DVT1W $=0$ DVT2W $=0$ DVT0 $=0.5281977$ DVT1 $=0.2185978$ DVT2 $=0.1 \cup 0=109.9762536 \mathrm{UA}=1.325075 \mathrm{E}-9 \mathrm{UB}=1.577494 \mathrm{E}-21 \mathrm{UC}=$ $-1 \mathrm{E}-10$ VSAT $=1.910164 \mathrm{E} 5 \mathrm{~A} 0=1.7233027 \mathrm{AGS}=0.3631032 \mathrm{BO}=2.336565 \mathrm{E}-7 \mathrm{~B} 1=5.517259 \mathrm{E}-7 \mathrm{KETA}=0.0217218 \mathrm{~A} 1$ $=0.3935816 \mathrm{~A} 2=0.401311$ RDSW $=252.7123939 \mathrm{PRWG}=0.5 \mathrm{PRWB}=0.0158894 \mathrm{WR}=1 \mathrm{WINT}=0$ LINT $=2.718137 \mathrm{E}-$ $8 \mathrm{XL}=0 \mathrm{XW}=-1 \mathrm{E}-8 \mathrm{WG}=-4.363993 \mathrm{E}-8 \mathrm{DWB}=8.876273 \mathrm{E}-10 \mathrm{VOFF}=-0.0942201$ NFACTOR $=2 \mathrm{CIT}=0 \mathrm{CDSC}=$ $2.4 \mathrm{E}-4$ CDSCD $=0$ CDSCB $=0$ ETAO $=0.2091053 \mathrm{ETAB}=-0.1097233$ DSUB $=1.2513945$ PCLM $=2.1999615$ PDIBLC1 $=$ $1.238047 \mathrm{E}-3$ PDIBLC2 $=0.0402861$ PDIBLCB $=-1 \mathrm{E}-3$ DROUT $=0$ PSCBE $1=1.034924 \mathrm{E} 10$ PSCBE2 $=2.991339 \mathrm{E}-9$ PVAG $=15$ DELTA $=0.01 \mathrm{RSH}=7.5 \mathrm{MOBMOD}=1 \mathrm{PRT}=0 \mathrm{UTE}=-1.5 \mathrm{KT} 1=-0.11 \mathrm{KT1L}=0 \mathrm{KT} 2=0.022 \mathrm{UA} 1=4.31 \mathrm{E}-9 \mathrm{UB} 1=$ $-7.61 \mathrm{E}-18 \mathrm{UC} 1=-5.6 \mathrm{E}-11 \mathrm{AT}=3.3 \mathrm{E} 4 \mathrm{WL}=0 \mathrm{WLN}=1 \mathrm{WW}=0 \mathrm{WWN}=1 \mathrm{WWL}=0 \mathrm{LL}=0 \mathrm{LLN}=1 \mathrm{LW}=0 \mathrm{LWN}=1 \mathrm{LWL}$ $=0$ CAPMOD $=2$ XPART $=0.5 \mathrm{CGDO}=6.28 \mathrm{E}-10 \mathrm{CGSO}=6.28 \mathrm{E}-10 \mathrm{CGBO}=1 \mathrm{E}-12 \mathrm{CJ}=1.160855 \mathrm{E}-3 \mathrm{~PB}=0.8484374$ $\mathrm{MJ}=0.4079216 \mathrm{CJSW}=2.306564 \mathrm{E}-10 \mathrm{PBSW}=0.842712 \mathrm{MJSW}=0.3673317 \mathrm{CJSWG}=4.22 \mathrm{E}-10$ PBSWG $=0.842712$ MJSWG $=0.3673317$ CF $=0$ PVTHO $=2.619929 \mathrm{E}-3$ PRDSW $=1.0634509$ PK2 $=1.940657 \mathrm{E}-3$ WKETA $=0.0355444$ LKETA $=-3.037019 \mathrm{E}-3 \mathrm{PUO}=-1.0227548 \mathrm{PUA}=-4.36707 \mathrm{E}-11 \mathrm{PUB}=1 \mathrm{E}-21 \mathrm{PVSAT}=-50$ PETA $=1 \mathrm{E}-4$ PKETA $=-5.167295 \mathrm{E}-3)$


Figure 5: DC transfer characteristic of the proposed rectifier.


Figure 6: Simulated results for DC characteristic at zero crossing regions.

The time domain response of the proposed full-wave rectifier of a 100 MHz frequency at different temperatures is shown in Fig. 11 and its details are shown in Fig.12. The offset values of the time domain response at the zero crossing regions are given in Table 3 for various temperatures. The peak outputs $\mathrm{V}_{\text {out }}$ values are also given in Table 3 for various temperatures. This table shows that difference is negligible between the output voltages at zero crossing regions and peak outputs $\mathrm{V}_{\text {out }}$ values for different temperatures. Therefore, we concluded that the proposed full-wave rectifier provides good temperature stability without compensation cir-


Figure 7: Input (sinusoidal) and rectified output waveforms of inverting full-wave precision rectifier of Fig. 4 at 100 MHz and 1 GHz frequencies.


Figure 8: Applying the 300 mV sine wave at the input, input and rectified output waveforms of inverting fullwave precision rectifier at a frequency of 100 MHz .
cuit. Output noise behaviors of the proposed rectifiers with respect to frequency have also been simulated. The equivalent output noise of voltage mode rectifier at 100 MHz frequency is found as $7.13 \mathrm{nV} \sqrt{\mathrm{Hz}}$.


Figure 9: DC transfer characteristic of the proposed full-wave rectifier of a 100 MHz frequency for different temperatures.


Figure 10: Simulated results for DC characteristics at zero crossing regions among different temperatures.

Table 3: The performance of the output waveform of the proposed rectifier at different temperatures.

| Tempera- <br> ture | Offsets <br> (DC <br> characteristics) | Offsets <br> (time <br> domain) | Vout peak <br> (time <br> domain) |
| :--- | :---: | :---: | :---: |
| $0^{\circ} \mathrm{C}$ | $179.81 \mu \mathrm{~V}$ | $304.11 \mu \mathrm{~V}$ | 97.486 mV |
| $27^{\circ} \mathrm{C}$ | $54.8 \mu \mathrm{~V}$ | $41.33 \mu \mathrm{~V}$ | 99.611 mV |
| $50^{\circ} \mathrm{C}$ | $23.5 \mu \mathrm{~V}$ | $-67.03 \mu \mathrm{~V}$ | 100.409 mV |
| $75^{\circ} \mathrm{C}$ | $10.7 \mu \mathrm{~V}$ | $-205.17 \mu \mathrm{~V}$ | 100.496 mV |
| $100^{\circ} \mathrm{C}$ | $5.4 \mu \mathrm{~V}$ | $-302.78 \mu \mathrm{~V}$ | 100.232 mV |



Figure 11: Time domain response of the proposed fullwave rectifier of a 100 MHz frequency at different temperatures.


Figure 12: The detailed time domain response of the proposed full-wave rectifier of a 100 MHz frequency at different temperatures.

## 5 Conclusion

In this paper, a new fully integrated high frequency full-wave rectifier which has been used the least CMOS transistor, has been proposed. Its operating frequency is up to 1 GHz . It does not involve any passive component; thus it is appropriate for integrated circuit implementation. It provides high precision voltage rectifying. The proposed full-wave rectifier provides excellent temperature stability without compensation circuit. LTSPICE simulations confirm the operability of this circuit in a wide frequency range. The fascinating character-
istics of the proposed rectifier are the high frequency operation (up to 1 GHz ), the lowest component count (10 components), the lowest power consumption ( 825 $\mu \mathrm{W}$ ), the lowest offset value ( $54.8 \mu \mathrm{~V}$ ) and suitable for IC fabrication.

## 6 References

1. Gift S. J. G., An improved precision full-wave rectifier, International Journal of Electronics, vol. 89, pp. 259-265, 2002.
2. Gift S. J. G., New precision rectifier circuits with high accuracy and wide bandwidth, International Journal of Electronics, vol. 92, pp. 601-617, 2005.
3. Gift S. J. G., Versatile precision full-wave rectifiers for instrumentation and measurements, IEEE Transactions on Instrumentation and measurements, vol. 56, pp. 1703-1709, 2007.
4. Sahu, P. P., Singh, M., Baishya, A., A novel versatile precision full-wave rectifier, IEEE Transactions on Instrumentation and Measurements, vol. 59, pp. 2742-2746, 2010.
5. Sahu, P. P., Singh, M., Baishya, A., New low-voltage full wave rectification technique without a diode, IET Circuits, Devices and Systems, vol. 5, pp. 33-36, 2011.
6. Toumazou, C., Lidgey F. J. and Chattong S., High frequency current conveyor precision full-wave rectifier, Electronics Letters, vol. 30, No.10, pp. 745746, 1994.
7. Khateb, F., Vavra, J., Biolek, D., A novel currentmode full-wave rectifier based on one CDTA and two diodes, Radioengineering, vol. 19, pp. 437445, 2010.
8. Biolek, D., Hancioglu, E., Keskin, A. U., High-performance current differencing transconductance amplifier and its application in precision currentmode rectification, International Journal of Electronics and Communications, vol. 62, pp. 92-96, 2008.
9. Koton, J., Herencsar, N., Vrba, K. Minimal configuration precision full-wave rectifier using current and voltage conveyors, IEICE Electronics Express, vol. 7, pp. 844-849, 2010.
10. Koton, J., Herencsar, N., Vrba, K., Current and voltage conveyors in current and voltage-mode precision full-wave rectifiers, Radioengineering, vol. 20, pp. 19-24, 2011.
11. Kumngern, M. Precision full-wave rectifier using only two DDCCs. Circuits and Systems, vol. 2, no. 3, pp. 127-132, 2011.
12. Koton, J., Lahiri, A., Herencsar, N., Vrba, K. Currentmode dual-phase precision full-wave rectifier using current-mode two-cell winner-takes-all (WTA)
circuit. Radioengineering, 2011, vol. 20, pp. 428432.
13. Maheshwari, S., Current controlled precision rectifier circuits, Journal of Circuit Systems, and Signal Components, vol. 16, pp. 129-138, 2007.
14. Anuntahirunrat K.,Tangsrirat W.,Riewruja V. and Surakampontorn W., Sinusoidal frequency doubler and full-wave rectifier based on translinear current-controlled current conveyors, International Journal of Electronics, 91, pp. 227-239, 2004.
15. Yuce, E., Minaei, S., Cicekoglu, O., Full-wave rectifier realization using only two CCII+s and NMOS transistors, International Journal of Electronics, vol. 93, pp. 533-541, 2006.
16. Minaei, S., Yuce, E., A new full-wave rectifier circuit employing single dual-X current conveyors, International Journal of Electronics, vol. 95, pp. 777-784, 2008.
17. Chaoui, H., CMOS high-frequency rectifier with unity voltage gain, Electronics Letters, vol. 31, pp. 717-718, 1995.
18. Monpapassorn, A., Dejhan, K., Cheevasuvit, F. CMOS dual output current mode half-wave rectifier, International Journal of Electronics, vol. 88, pp. 1073-1084, 2001.
19. Kumngern, K., Knobnob, B., Dejhan, K., High frequency and high precision CMOS half-wave rectifier, Circuits, Systems and Signal Processing, vol. 29, pp. 815-836, 2010.
20. Kumngern, M., Dejhan, K., Current conveyorbased versatile precision rectifier, WSEAS Transactions on Circuits and Systems, vol. 7, pp. 10701079, 2008.
21. Sanchez-Sineccio, E., Ramirez-Angulo, J., LinaresBarranco, B., Rodriguez-Vazquez, A. Operational transconductance amplifier-based nonlinear function syntheses, IEEE Journal of Solid-State Circuits, vol. 24, pp. 1576-1586, 1989.
22. Heim, P., Krummenacher, F., Vittoz, E. CMOS fullwave operational transconductance rectifier with improved DC transfer characteristic, Electronics Letters, vol. 28, pp. 333-334, 1992.
23. Ramirez-Angulo, J., High frequency, low voltage CMOS diode, Electronics Letters, vol. 28, pp. 298299, 1992.
24. Jongkunstidchai, C., Fongsamut, C. Kumwachara, K., Surakampontorn, W. Full-wave rectifiers based on operational transconductance amplifiers, International Journal Electronics and Communications, vol. 61, pp. 195-201, 2007.
25. Minhaj, N., OTA-based non-inverting and inverting precision full-wave rectifier circuits without diodes, International Journal of Recent Trends in Engineering, vol. 1, pp. 72-75, 2009.
26. Kumngern, M., High frequency and high precision CMOS full-wave rectifier, In Proceedings of IEEE In-
ternational Conference on Communication Systems (ICCS 2010), Singapore, pp. 5-8, 2010.
27. Kumngern, M., Dejhan, K., High frequency and high precision CMOS full-wave rectifier, International Journal of Electronics, vol. 93, pp. 185-199, 2006.
28. Arbel A. F., Goldminz L., Output Stage for CurrentMode Feedback Amplifiers, Theory and Applications, Analog Integrated Circuits and Signal Processing, pp. 234-255, 1992.
29. Arbel A. F., Towards A Perfect CMOS CCII, Analog Integrated Circuits and Signal Processing, pp. 119132, 1997.
30. Awad I. A.,Soliman A. M., New CMOS Realization of the CCII-, IEEE Trans. Circuits, Systems, pp. 460463, 1999.
31. Sobhy E. A., Soliman A. M., Realizations of Fully Differential Voltage Second Generation Current Conveyor with An Application, International J. of Circuit Theory and Applications, vol. 18, Issue 5, 441-452, 2010.
32. Mostafah., and Soliman A. M., Novel Low-Power Accurate Wide-band CMOS Negative Second Generation Current Conveyor Realizations Based on Floating Current Source Building Blocks, Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference, pp. 720 725, 2009.
33. Youssef M. A., Soliman A. M., A Modified CMOS Balanced Output Transconductor with Extended Linearity, Analog Integrated Circuits and Signal Processing, pp. 239-244, 2003.

Arrived: 14.07. 2014
Accepted: 09.02. 2015

# An influence of the selected factors on the transient thermal impedance model of power MOSFET 

Krzysztof Górecki, Janusz Zarębski, Damian Bisewski

Gdynia Maritime University, Department of Marine Electronics, Gdynia, Poland


#### Abstract

The paper presents the results of experimental studies that illustrate the influence of the selected factors, i.e. the size of soldering pads, the PCB copper area, heat-sink size as well as the dimensions and material of the housing on the transient thermal impedance model parameters of MOSFET. Measurements of thermal parameters were performed using the indirect electrical method. Parameters of the transient thermal impedance model were calculated using the estimation procedure elaborated by the authors. The obtained results show an influence of the system cooling parameters on thermal parameters of the semiconductor device.


Keywords: thermal phenomena; transient thermal impedance; compact thermal model

# Vpliv določenih faktorjev na tranzienten termično impedančen model močnostnega MOSFET 

Izvleček: Članek predstavlja eksperimentalne rezultate vplivov določenih faktorjev, kot so velikost spajkalnih površin, področje bakra na PCB, velikost hladilnika in velikost ohišja, na tranzienten termično impedančni model MOSFET. Meritve temperaturnih parametrov so bile opravljene s pomočjo posredne električne metode. Parametri tranzientnega modela so bili določeni na osnovi ocenjevalne procedure avtorjev. Rezultati kažejo vpliv hladilnih parametrov sistema na termične parametre elementa.

Ključne besede: termičen pojav; termična impedanca; kompakten termičen model

[^0]
## 1 Introduction

One of the essential phenomena influencing properties of semiconductor devices is self-heating [1, 2, 3, $4,5]$. It appears with a rise of the semiconductor device internal temperature $T_{j}$ and it is caused by the exchange of electrical energy dissipated in these devices into heat at not ideal cooling conditions. The rise of the device internal temperature causes changes in the course of their characteristics [1, 2, 3] and strongly influences their reliability [6-11]. In order to limit the excess of the device internal temperature as a result of self-heating, proper cooling systems are used [12 14]. Except of classical systems of free space cooling, systems of forced cooling are also applied. This group of cooling systems includes microchannel cooling systems [15], thermoelectric cooling systems [16] or water cooling systems [17].

An increase in the power density dissipated in modern semiconductor chips causes, that a development of non-typical cooling methods are used, for example

- the microchannel heat-sinks [18, 19]. In order to assure efficient electrical insulation between the case of semiconductor device and its heat-sink, different interface materials are applied. Properties of materials are analyzed among others in the paper [20].

Heat removal to the surrounding is realised by three mechanisms [21, 22]: conduction, convection and radiation. The efficiency of these mechanisms depends, among others on the value of the device internal temperature and on the difference between temperature of the device case and the surrounding. In addition, as shown in the papers, e.g. [23, 24], heat transport from the device structure to the surroundings is carried out using a number of paths. Therefore, one should expect this efficiency to undergo some change connected with the changes of the power dissipated in these devices and changes of the manner of their mounting.

Thermal parameters describing the efficiency of removing the heat generated in the semiconductor device to
the surrounding are transient thermal impedance $Z_{\text {th }}(t)$ and thermal resistance $R_{\text {th }}$. The first of the mentioned parameters describes thermal properties of the device during the transient state, whereas the other one - at the steady-state. Transient thermal impedance $Z_{\text {th }}(t)$ of the electronic device is of the form $[1,3,4,25]$

$$
\begin{equation*}
Z_{t h}(t)=R_{t h} \cdot\left[1-\sum_{i=1}^{N} a_{i} \cdot \exp \left(-\frac{t}{\tau_{t h i}}\right)\right] \tag{1}
\end{equation*}
$$

where $\tau_{\tau \mathrm{hi}}$ is the $i$-th thermal time constant, $\mathrm{a}_{\mathrm{i}}$ - weighting factor of thermal time constant, $\mathrm{N}-$ a number of time constants.

Constructors of semiconductor device cases develop new structures characterized by a low value of thermal resistance between the junction and the case of the semiconductor device [24]. This factor of complex thermal resistance junction-to-ambient, however, is not dominant. In turn, the designers of the cooling system of a semiconductor device should include all external parts of the heat flow path. As it was shown, among others in the papers [25-30], the influence of such factors, as the dissipated power, the mounting method and the ambient temperature on the waveform of transient thermal impedance of the device can be important. As it is shown, e.g. in the papers [25, 26,31] the value of thermal parameters can change essentially under the influence of such factors, as e.g. lengths of leads and the area of pads.

On the other hand, the computer analysis of electronic networks needs to use computer models of the devices existing in this network. The accuracy of the obtained results of calculations depends on the accuracy of the used models. In order to take into account self-heating phenomena in computer analyses, the compact thermal models of electronic devices are typically used. Parameters values of such model (see Eq. (1)) are dependent on among others construction of the cooling system as well as location of the semiconductor device bias point [23, 30, 31].

In the paper [31] the influence of the selected factors on thermal resistance of semiconductor devices is considered and some formulas describing this influence are proposed. In turn, in the papers [2,30,32] it is shown, that the waveforms of transient thermal impedance of semiconductor devices depend on power dissipated in these devices. Unfortunately, the results of measurements presented in the cited papers refer to selected semiconductor devices and cooling systems. Therefore, it is justified to carry out systematic research on the influence of the selected factors on both the waveforms of transient thermal impedance and parameters values of the thermal model for a single semiconductor de-
vice at different cooling conditions. SMD Power MOS transistor type IRFR420 contained in DPAK TO-252, was arbitrary chosen for investigations.

In section 2 the used measurement set is described. The considered cooling systems are presented in section 3 . The measurement results of the waveforms of transient thermal impedance of the considered device with the values of parameters describing such waveforms are shown in section 4.

## 2 The measuring set

Transient thermal impedance is determined using the indirect electrical method described in [32]. In this method, the cooling curve [28] of the transistor is measured, whereas the voltage on the forward biased body diode Db of the transistor DUT is used as a thermallysensitive parameter. The measurements performed by the authors show, that the thermometric characteristic $v_{D}(T)$ describing the dependence of the forward voltage of the body diode on temperature of the transistor at the constant body diode current $\mathrm{I}_{\mathrm{M}}=1 \mathrm{~mA}$ is linear in the range of temperature from 25 to $110^{\circ} \mathrm{C}$. Measurements are realized in three steps using the measurement set shown in Fig.1. At first, the calibration of the characteristics $v_{D}(T)$ is carried out. Next, heating of the tested transistor operating in the saturation range is realized. This step of measurement is finished, when the device thermally steady state is achieved. The cooling step starts at time $t=0$, when the transistor is switched off and the body diode is forward biased by the current $I_{\text {M }}$.


Figure 1:The diagram of the measurement set to measure transient thermal impedance of the power MOSFET

In the measuring set, the source $\mathrm{I}_{\mathrm{M}}$ forces the measuring current of the body diode located inside the transistor (DUT). The voltage source $\mathrm{V}_{\mathrm{DD}}$ and the current source $\mathrm{I}_{\mathrm{H}}$ set the dissipated power while the heating. Switches $S_{1}$ and $S_{2}$ are controlled by the PC. As the switches $S_{1}$ and $S_{2}$ the power MOS transistors are used. The position of the switches depends on the measurement step. In calibration and cooling, the switch $\mathrm{S}_{1}$ is open and the
switch $S_{2}$ is in position 2. While heating, the switch $S_{1}$ is closed and the switch $S_{2}$ is in position 1. The values of the voltage and current of the DUT are recorded using a 16-bit A/D converter USB-1608GX-2AO manufactured by Measurement Computing. Maximum sampling rate of the converter is equal to $500 \mathrm{kS} / \mathrm{s}$.

## 3 Cooling systems

Using the measurement system presented in Section 2, transient thermal impedance of the considered power MOS transistor at different set of the cooling system, were measured. The first group of measurements results relates to the transistor mounted on one of the PCBs, where the path mosaic of each PCB is shown in Fig. 2. The black color in Fig. 2 represents areas of PCB covered with the layer of copper. The dimensions of all PCBs are $51 \times 33 \mathrm{~mm}$. As seen, the PCB A1 contains copper in the solder pads and conductive paths, only, whereas the PCBs A2 and A3 are covered with additional layers of copper supporting the heat dissipation process. The PCB A2 differs from the PCB A3 in the path width. The considered transistor was soldered, in turn, to each PCB and the silicon insulating spacer and thermal grease, were used. For comparison, the measurements were also performed for the transistor soldered directly to the wires (without the PCB). This variant of assembly is designated by acronym AO.


Figure 2: PCBs used for investigations
The second group of measurements relates to the same transistor mounted on the aluminum heat-sink (type A-5723 of the length 100 mm ). The measurements were performed for the heat-sink situated in free space (FS) and five types of housings, i.e.: metal housing (ME1) of the dimensions $120 \times 220 \times 195 \mathrm{~mm}$ (volume of about 5 litres), metal housing (ME2) of a dimensions $225 \times 115$ $\times 345 \mathrm{~mm}$ (10 liters), plastic enclosure (PE1) of a dimensions $170 \times 85 \times 70 \mathrm{~mm}$ ( 1 litre), plastic enclosure (PE2) of a dimensions $140 \times 85 \times 170 \mathrm{~mm}$ (2 liters) and plastic enclosure (PE3) of the dimensions $225 \times 210 \times 85 \mathrm{~mm}$ (4 litres).

## 4 Results

The influence of the transistor dissipated power, mounted on the PCB A1, on the waveform of transient thermal impedance is shown in Fig. 3. The values of $\mathrm{Z}_{\text {th }}(\mathrm{t})$ model parameters for 3 arbitrary chosen values of the dissipated power are presented in Table 1. These values were determined using the estimation procedure ESTYM, proposed by the authors [32].


Figure 3: The measured waveforms of transient thermal impedance of the transistor mounted on PCB A1

Table 1: The values of transient thermal impedance model parameters of the transistor mounted on PCB A1

| $\mathrm{P}[\mathrm{W}]$ | 0.12 | 0.51 | 1.18 |
| :--- | :---: | :---: | :---: |
| $\mathrm{R}_{\text {th }}[\mathrm{K} / \mathrm{W}]$ | 95.34 | 95.56 | 87.37 |
| $\mathrm{a}_{1}$ | 0.107 | 0.121 | 0.093 |
| $\tau_{\text {th1 }}[\mathrm{s}]$ | 738.7 | 631.6 | 866 |
| $\mathrm{a}_{2}$ | 0.471 | 0.597 | 0.572 |
| $\tau_{\text {th } 2}[\mathrm{~s}]$ | 99.33 | 90.25 | 103.1 |
| $\mathrm{a}_{3}$ | 0.136 | 0.161 | 0.208 |
| $\tau_{\text {th } 3}[\mathrm{~s}]$ | 17.36 | 15.27 | 21.2 |
| $\mathrm{a}_{4}$ | 0.049 | 0.06 | 0.069 |
| $\tau_{\text {th } 4}[\mathrm{~s}]$ | 2.33 | 2.21 | 2.78 |
| $\mathrm{a}_{5}$ | 0.023 | 0.033 | 0.036 |
| $\tau_{\text {th } 5}[\mathrm{~ms}]$ | 387.2 | 359.1 | 473.2 |
| $\mathrm{a}_{6}$ | 0.17 | 0.015 | 0.011 |
| $\tau_{\text {th } 6}[\mathrm{~ms}]$ | 0.27 | 1.3 | 17.33 |
| $\mathrm{a}_{7}$ | 0.044 | 0.013 | 0.011 |
| $\tau_{\text {th } 7}[\mu \mathrm{~s}]$ | 40 | 40 | 40 |

As seen, thermal resistance is a decreasing function of the dissipated power and has the values of the range from 90 to $100 \mathrm{~K} / \mathrm{W}$. A number of thermal time constants does not depend on the dissipated power and is equal to 7. The thermal time constants have the values of the range from $40 \mu \mathrm{~s}$ to 866 s .

The measurement results of transient thermal impedance of the transistor mounted on various PCBs at the
constant dissipated power equal to 0.9 W are presented in Fig. 4. The values of $Z_{\text {th }}(\mathrm{t})$ model parameters for the measurement results (Fig. 4) are presented in Table 2.

As seen in Fig. 4, the highest value of thermal resistance is obtained for the transistor operating without the PCB (A0), whereas the value of thermal resistance for the transistor operating on the PCB A2 is about twice lower. Comparing the $Z_{\text {th }}(t)$ waveforms of the transistor mounted on the PCB A1 and mounted on the PCBs A2 and A3 it is observed, that the increase of copper area results in a decrease of thermal resistance, whereas the time required to achieve the steady state increases. This is caused by increasing heating capacity due to an increase of the volume of copper. Analysing the contents of Table 2, the correlation between the increase of thermal resistance and the decrease of the longest thermal time constant $\tau_{\text {th1 }}$, is observed. The values of the thermal time constant $\tau_{\text {th1 }}$ vary within the range from more than 120 s to nearly 900 s .


Figure 4: The measured waveforms of transient thermal impedance of the transistor mounted on various PCBs

Table 2: The values of transient thermal impedance model parameters of the transistor mounted on various PCBs

| $\mathrm{P}[\mathrm{W}]$ | A0 | PCB A1 | PCB A2 | PCB A3 |
| :--- | :---: | :---: | :---: | :---: |
| $\mathrm{R}_{\text {th }}[\mathrm{K} / \mathrm{W}]$ | 118 | 89.03 | 62.1 | 68.93 |
| $\mathrm{a}_{1}$ | 0.106 | 0.112 | 0.122 | 0.116 |
| $\tau_{\text {th } 1}[\mathrm{~s}]$ | 129.6 | 682.9 | 899.4 | 889.3 |
| $\mathrm{a}_{2}$ | 0.772 | 0.604 | 0.389 | 0.385 |
| $\tau_{\text {th } 2}[\mathrm{~s}]$ | 29.71 | 92.25 | 127.9 | 127.1 |
| $\mathrm{a}_{3}$ | 0.072 | 0.168 | 0.294 | 0.278 |
| $\tau_{\text {th3 }}[\mathrm{s}]$ | 4.25 | 16.42 | 17.58 | 17.43 |
| $\mathrm{a}_{4}$ | 0.034 | 0.063 | 0.106 | 0.133 |
| $\tau_{\text {th } 14}[\mathrm{~s}]$ | 0.648 | 2.35 | 3 | 2.96 |
| $\mathrm{a}_{5}$ | 0.014 | 0.032 | 0.052 | 0.055 |
| $\tau_{\text {th } 5}[\mathrm{~ms}]$ | 1.13 | 415.3 | 483.4 | 473.8 |
| $\mathrm{a}_{6}$ |  | 0.022 | 0.012 | 0.012 |
| $\tau_{\text {th } 6}[\mathrm{~ms}]$ |  | 12.89 | 24.36 | 16.09 |
| $\mathrm{a}_{7}$ |  | 0.01 | 0.021 | 0.018 |
| $\tau_{\text {th }}[\mathrm{\mu s}]$ |  | 40 | 470 | 380 |

The influence of the dissipated power as well as the mosaic design of the PCB on thermal resistance $R_{t h}$ of the transistor, is presented in Fig. 5.


Figure 5: The measured dependencies of thermal resistance versus the dissipated power for different mounting methods of the transistor

The decreasing dependence of thermal resistance on the dissipated power is observed both for the transistor operating without the PCB (A0) and for the transistor mounted on each PCB (A1, A2, A3), due to an increase of convection efficiency resulting from a case temperature rise of the transistor. The strongest dependence $R_{\text {th }}(p)$ is observed for the transistor operating without any PCB, because an increase of the dissipated power leads to the most significant temperature rise of the transistor. Thermal resistance of the transistor decreases in the considered range of the dissipated power by even dozen percent.


Figure 6: The measured waveforms of transient thermal impedance of the transistor mounted on the heatsink and enclosed in various housings

The commonly used method for reducing thermal resistance of the semiconductor device is mounting the device on the heat-sink. The measurements performed by the authors show [31] that thermal resistance is a decreasing function of the length of the heat-sink and also depends on the spatial orientation of the heatsink [32]. Typically, the semiconductor device, how-
ever, does not operate individually, but is a part of an electronic device that is enclosed in a housing made of metal or plastic of the defined volume. Figure 6 shows the measured transient thermal impedance of the considered MOS transistor mounted on the heat-sink and placed inside different housings. Designations used in Fig. 6 are discussed in detail in section 3 . The values of $\mathrm{Z}_{\mathrm{th}}(\mathrm{t})$ model parameters for the measurements results (Fig. 6) are presented in Table 3.

As seen in Fig. 6, the measured waveforms of transient thermal impedance $Z_{\text {th }}(t)$ of time $<0.1 \mathrm{~s}$ for various housings are generally indistinguishable. This is due to the fact that in the initial phase of the transistor cooling, responsibility for the heat dissipation rests with the physical processes occurring inside the transistor as well as on the border between the transistor case and the heat-sink. The smallest value of thermal resistance is for the transistor operating on the heat-sink without housing. Value of the considered parameter increases with decrease of the housing volume. Apart from this, the use of metal enclosure results in the lower value of $R_{t h}$ in comparison to the plastic one. The differences in the value of $R_{t h}$ for all the considered cooling conditions exceed even $35 \%$. Also, the time $t_{s 5}$ required to obtain the thermally steady state in the transistor increases with an increase of thermal resistance. The time $\mathrm{t}_{\mathrm{ss}}$ for the transistor operating on the heat-sink inside the smallest plastic enclosure (PE1) is about twice grater than for the transistor operating on the heat-sink without the housing.

The influence of the enclosure material and volume on the thermal model parameters is visible due to various conductivity values of the materials used in the construction of enclosure as well as various effectiveness of convection at the surface of the housing.

## 5 Conclusions

The paper presents experimental results concerning the influence of the cooling system construction of SMD MOSFET on the transient thermal impedance model parameters. Decreasing dependence of thermal resistance on the dissipated power and the area of solder pads, known from the previous work of the authors [23, 31], was confirmed. In addition, mounting the transistor inside the housing, even of a large volume, results in an increase of thermal resistance, whereas the influence of the housing volume on time of determining the internal transistor temperature is ambiguous.

Metal housing, which conducts heat between its interior and the surrounding, provides better cooling that the plastic one. Differences in thermal resistance between the housings made of different materials reach $10 \%$. Mounting the transistor on the PCB can improve cooling efficiency even twice. Mounting the transistor on the heat-sink results in even a 15 times decrease of thermal resistance.

Table 3: The values of transient thermal impedance model parameters of the transistor mounted on the heat-sink and enclosed in various housings

| $P[W]$ | FS | ME1 | PE2 | PE3 | ME2 | PE1 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| $\mathrm{R}_{\text {th }}[\mathrm{K} / \mathrm{W}]$ | 4.51 | 5.09 | 5.96 | 5.6 | 4.99 | 6.57 |
| $\mathrm{a}_{1}$ | 0.253 | 0.295 | 0.498 | 0.28 | 0.237 | 0.381 |
| $\tau_{\text {th1 }}[\mathrm{s}]$ | 1212.2 | 1337.3 | 887 | 2178.5 | 1558.2 | 2034.8 |
| $\mathrm{a}_{2}$ | 0.281 | 0.275 | 0.078 | 0.326 | 0.328 | 0.274 |
| $\tau_{\text {th2 }}[\mathrm{s}]$ | 381 | 417 | 186.8 | 554.9 | 451.1 | 626.9 |
| $\mathrm{a}_{3}$ | 0.051 | 0.04 | 0.05 | 0.02 | 0.042 | 0.026 |
| $\tau_{\text {th3 }}[\mathrm{s}]$ | 2.47 | 3.49 | 1.92 | 16.9 | 3 | 6.553 |
| $\mathrm{a}_{4}$ | 0.078 | 0.068 | 0.078 | 0.048 | 0.068 | 0.049 |
| $\tau_{\text {th } 4}[\mathrm{~ms}]$ | 307.5 | 425.5 | 243.5 | 1345 | 370.4 | 660 |
| $\mathrm{a}_{5}$ | 0.131 | 0.12 | 0.113 | 0.082 | 0.122 | 0.087 |
| $\tau_{\text {th } 5}[\mathrm{~ms}]$ | 47.61 | 59.68 | 41.13 | 168.9 | 55.98 | 87.81 |
| $\mathrm{a}_{6}$ | 0.12 | 0.112 | 0.108 | 0.117 | 0.115 | 0.107 |
| $\tau_{\text {th6 }}[\mathrm{ms}]$ | 7.26 | 9.14 | 6.88 | 26.72 | 8.5 | 12.11 |
| $\mathrm{a}_{7}$ | 0,061 | 0,065 | 0,053 | 0,082 | 0,059 | 0,053 |
| $\tau_{\text {th }}[\mathrm{ms}]$ | 1,26 | 1,54 | 1,2 | 3,74 | 1,5 | 1,71 |
| $\mathrm{a}_{8}$ | 0,025 | 0,025 | 0,022 | 0,095 | 0,068 | 0,059 |
| $\tau_{\text {th } 8}[\mathrm{~ms}]$ | 40 | 40 | 40 | 60 | 210 | 170 |

The number of thermal time constants increases with the number of elements in the heat-flow path. The thermal model describing the transistor soldered to the wires contains only 4 thermal constants. The number of thermal constants for the transistor mounted on the heat-sink and operating in the housing, however, is even equal to 8.

As seen from the measurements of $Z_{t h}(t)$ presented in section 4, the influence of the enclosure unit and area of copper on the PCB, results in a visible change in the instantaneous values of $Z_{\text {th }}(t)$ for times exceeding a few seconds. Thus, in construction of the thermal model of a semiconductor device together with its cooling system, it is recommended to use the nonlinear RC Cauer structure [2], wherein each of the heat flow path elements are represented by the two-terminal RCs. Using such structure it is easy to take into account all elements of the heat flow path.

## 6 Acknowledgements

This project is financed from the funds of the National Science Centre which were awarded on the basis of the decision number DEC-2011/01/B/ST7/06740.

## 7 References

1. Zarębski J., Górecki K.: The electrothermal largesignal model of power MOS transistors for SPICE. IEEE Transaction on Power Electronics, Vol. 25 , No. 5-6, 2010, pp. 1265 - 1274.
2. Górecki K., Zarębski J.: Nonlinear compact thermal model of power semiconductor devices. IEEE Transactions on Components and Packaging Technologies, Vol. 33, No. 3, 2010, pp. 643-647.
3. Zarębski J., Górecki K.: SPICE-aided modelling of dc characteristics of power bipolar transistors with selfheating taken into account. International Journal of Numerical Modelling Electronic Networks, Devices and Fields, Vol. 22, No. 6, 2009, pp. 422-433.
4. Székely V., Thermal Testing and Control by Means of Built-in Temperature Sensors. Electronics Cooling, Vol. 4, 1998, No. 3, pp.36-39.
5. Székely V., Rencz M., Courtois B., Thermal Investigations of IC's and Microstructures. Microelectronics Journal, Vol. 28, 1997, No.3, pp. 205-207
6. Castellazzi A., Gerstenmaier Y.C., Kraus R., Wachutka G.K.M.: Reliability analysis and modeling of power MOSFETs in the 42-V-PowerNet, IEEE Transactions on Power Electronics, Vol. 21, 2006, No. 3, pp.603-612
7. Reynolds F.H., Measuring and modeling integrated circuit failure rates. Eurocon'82, Copenhagen: Reliability in Electrical and Electronic Components and Systems. North Holland, Vol. 1, 1982, pp. 36-45
8. Parry J., Rantala J., Lasance C.: Temperature and reliability in electronics systems - the missing link. Electronics Cooling, Vol. 7, No. 4, 2001, pp. 30 - 36
9. Ciappa M., Carbognami F., Cora P., Fichtner W.: A novel thermomechanics-based lifetime prediction model for cycle fatigue failure mechanisms in power semiconductors. Microelectronics Reliability, Vol. 42, 2002, pp.1653-1658
10. Castellazzi A., Kraus R., Seliger N., SchmittLandsiedel D.: Reliability analysis of power MOSFET's with the help of compact models and circuit simulation. Microelectronics Reliability, Vol. 42, 2002, pp.1605-1610
11. Coquery G., Carubelli S., Ousten J.P., Lallemand R.: Power module lifetime estimation from chip temperature direct measurement in an automotive traction inverter. Microelectronics Reliability, Vol. 41, 2001, pp.1695-1700
12. Happer C.A.: Electronic packaging and interconnection handbook McGraw-Hill Handbooks, 2000.
13. Lidow A., Knzer D., Sheridan G., Tam D.: The Semiconductor Roadmap for Power Managment in the New Millennium. Proceedings of the IEEE, Vol. 89, 2001, No. 6, pp. 803-812.
14. Sarno C., Moulin G.: Thermal management of highly integrated electronic packages in avionics applications. Electronics Cooling, Vol. 7, No. 4, 2001, pp. 12-20
15. Raj E., Lisik Z., Fiks W.: Influence of the manufacturing technology on microchannel structure efficiency. Materials Science and Engineering B, Vol. 176, No. 4, 2011, pp. 311-315.
16. Gould C.A., Shammas N.Y.A., Grainger S., Taylor I.: Thermoelectric cooling in microelectronic circuits and waste heat electrical power generation in a desktop personal computer. Materials Science and Engineering B, Vol. 176, No. 4, 2011, pp. 316325.
17. Simons R.E.: Estimating temperatures in a water - to - air hybrid cooling system. Electronics Cooling, Vol. 8, No. 2, 2002, pp. 8-9
18. Garimella S.V., Singhal V., Liu D.: On-chip thermal management with microchannel heat sinks and integrated micropumps. Proceedings of the IEEE, Vol. 94, 2006, No. 8, pp. 1534-1548.
19. Zhang H.Y., Pinjala D., Wong T.N., Toh K.C., Joshi Y.K.: Single-phase liquid cooled microchannel heat sink for electronic packages. Applied Thermal Engineering, Vol. 25, 2005, No. 10, pp. 1472-1487.
20. Prasher R.: Thermal Interface Materials: Historical Perspective, Status and Future Directions. Proceedings of the IEEE, Vol. 94, No. 8, 2006, pp. 1571-1586.
21. Blackburn D.L.: Temperature Measurements of Semiconductor Devices - A Review. $20^{\text {th }}$ IEEE Semiconductor Thermal Measurement and Menagement Symposium SEMI-THERM, 2004, pp. 70-80.
22. Yener Y., Kakac S.: Heat Conduction.Taylor \&Francis, 2008.
23. Górecki K., Zarębski J.: The semiconductor device thermal model taking into account non-linearity and multhipathing of the cooling system. Journal of Physics: Conference Series, Vol. 494, 2014, 012008, doi:10.1088/1742-6596/494/1/012008
24. Carver L.: Innovative packaging design for electronics in extreme enviroments. IEEE Spectrum, No. 5, 2014, pp. S26-S28.
25. Szekely V.: A New Evaluation Method of Thermal Transient Measurement Results. Microelectronic Journal, Vol. 28, No. 3, 1997, pp. 277-292.
26. Górecki K., Rogalska M., Zarębski J.: Parameter estimation of the electrothermal model of the ferromagnetic core. Microelectronics Reliability, Vol. 54, No. 5, 2014, pp. 978-984.
27. Zarębski J., Górecki K.: A New Method for the Measurement of the Thermal Resistance of the Monolithic Switched Regulator LT1073. IEEE Trans. on Instr. and Meas., Vol. 56, No. 5, 2007, pp. 2101-2104.
28. Blackburn D.L., Oettinger F.F., Transient Thermal Response Measurements of Power Transistors. IEEE Transactions on Industrial Electronics and Control Instrum., IECI-22, 1976, No. 2, pp. 134-141
29. Oettinger F. F., Blackburn D. L.: Semiconductor Measurement Technology: Thermal Resistance Measurements, U. S. Department of Commerce, NIST/SP-400/86, 1990.
30. Górecki K., Zarębski J.: The influence of the selected factors on transient thermal impedance of semiconductor devices. Proceedings of the $21^{\text {st }}$ International Conference Mixed Design of Integrated Circuits and Systems MIXDES, 2014, Lublin, pp. 309-314.
31. Górecki K., Zarębski J.: Modeling the influence of selected factors on thermal resistance of semiconductor devices. IEEE Transactions on Components, Packaging and Manufacturing Technology, Vol. 4, No. 3, 2014, pp. 421-428.
32. Górecki K., Zarębski J.: Badanie wpływu wybranych czynników na parametry cieplne tranzystorów mocy MOS. Przegląd Elektrotechniczny, Vol. 85, No. 4, 2009, pp. 159-164.

# Electronically Controllable Current-mode True RMS to DC Converter 

Predrag B. Petrović<br>Faculty of Technical Sciences Čačak, SERBIA


#### Abstract

The paper presents a possible design of an electronically tuneable current-mode RMS-to-DC converter. The circuit consists of a single multiple-output current-controlled current differencing transconductance amplifier (MO-CCCDTA), two current-controlled conveyors (CCCII), and a grounded resistor and capacitor. The errors related to signal processing were investigated and presented in the paper. The PSpice simulation results are depicted, and they agree well with theoretical anticipation. The maximum power consumption of the converter is approximately 5.80 mW , at $\pm 1.2 \mathrm{~V}$ supply voltages.


Keywords: RMS-to-DC converter; current-mode processing; CCCDTA; MOCCCII; simulation

# Elektronsko nadzorovan RMS DC pretvornik v tokovnem načinu 


#### Abstract

Izvleček: Članek opisuje možno obliko elektronsko nastavljivega RMS-DC pretvornika v tokovnem načinu. Vezje je sestavljeno iz enega večizhodnega tokovno krmiljenega tokovno diferencialnega transkonduktančnega ojačevalnika (MO-CCCDTA), dveh tokovno krmičjenih ojačevalnikov (CCCII) in ozemljenega upora ter kondenzatorja. Raziskane in predstavljene so napake v zvezi signalom. Opravljena je bila PSpice analiza, ki se dobro ujema s teorijo. Največja poraba pretvornika je 5.80 mW pri napajalni napetosti $\pm 1.2 \mathrm{~V}$.


KIjučne besede: RMS-DC pretvornik; CCCDTA, MOCCCII, simulacije

[^1]
## 1 Introduction

Magnitude-detection circuits, e.g. envelope detectors, peak detectors and RMS-to-DC converters produce an estimate of a signal's magnitude, and are important elements in communications transceivers, automatic gain control systems and analog spectrum analysers [1].

Mobile communication terminals or handsets have become ubiquitous in the modern society, creating insatiable market demand for and more efficient power supply solutions for the mobile terminal. Accurate monitoring of the transmitting power of the mobile terminal helps to optimise power consumption and performance of the wireless network [2]. Detectors intended for this purpose need to involve wide bandwidth, high input impedance, low loss, low noise, and are expected to be compact and robust in the presence of process-voltage-temperature (PVT) variations. High dynamic range and low power consumption are also desirable.

Different methods have been reported for the precision measuring of the RMS value of an AC voltage, such as sampling [2], Monte Carlo [3] and the wavelet transform [4, 5]. The implicit RMS converter described in [69] has been used for many years. Most of these devices similarly comprise two main parts: a full-wave rectifier (or absolute-value) circuit and a multiplier/divider circuit employing a log-antilog principle. High-frequency performances of these devices are limited to less than 5 MHz due to the band-width and the slow rate of the full-wave rectifier. Design technologies based on bipolar dynamic trans-linear circuits were proposed to implement true RMS-to-DC converters [10, 11]. Although these schemes require only NPN transistors, their circuits are operated in only one quadrant and employ full-wave rectifiers. The new design for RMS-to-DC converter relies on the dual trans-linear-based squarer circuit proposed in [12, 13], where the input current can be a two-quadrant current signal. Given that the full-wave rectifier is not required within this conversion scheme, the circuit exhibits a wide bandwidth, which,
due to the input interference, still appears limited compared to thermal-based or diode-based detectors [14].

Presently, there is a growing interest in synthesising the current-mode circuits because of a number of their potential advantages such as larger dynamic range, higher signal bandwidth, greater linearity, simpler circuitry and lower power consumption. The current differencing transconductance amplifier, CDTA [15], appears to be a versatile component in the realization of a class of analog signal-processing circuits. This device is quite suitable for the synthesis of current-mode circuits with electronically tunable properties. Moreover, the use of the CDTA as an active element provides the circuit implementations with a reduced number of passive elements, thereby leading to compact structures in some applications.

This paper presents the principles of operation, and the detailed circuit design of the new current-mode realisation of the bipolar RMS detector. The proposed detector uses an implicit computation to calculate the RMS value of an input signal, similarly to the translinear principle. The paper looks at a small-signal approach to the problem, i.e. voltage-to-current relations are regarded as linear, while in translinear circuits, the ability of exponential function to convert a sum of signals (voltages) into signal products (currents) is utilised. The fundamental building block is an analog multiplier/ divider realised with one MO-CCCDTA, the anticipated exploitation of the proposed circuit being extended up to 10 MHz , with increased linearity and precision in determining the effective value. Unlike the detector described in $[16,17]$, which was realised using the CMOS technology, the one described in this paper involves simpler and more accurate control structure. Besides, the proposed circuit does not require a more precise bias voltages realization and complex transistor pairing, which was typical of the realisations described in [16, 17]. Additionally, it has fewer active building blocks and allows a faster access to the required feedback - the RMS of the input signal which is the subject of processing. In addition, the circuit comprises two grounded passive components (resistor and capacitor),
rendering it very suitable for the IC implementation. The PSpice simulation results are also shown, and they are in agreement with theoretical analysis.

## 2 Proposed detector circuit

The proposed current-mode RMS detector using the MO-CCCDTA and MO-CCClls [18] is shown in Fig. 1. MO-CCCDTA properties are similar to the conventional CDTA, except that input voltages of MO-CCCDTA are not zero and the MO-CCCDTA has finite input resistances $R_{p}$ and $R_{n}$ at the $p$ and $n$ input terminals, respectively. These parasitic resistances are equal and can be controlled by the bias current $I_{p}$ (Fig. 2).


Figure 1: The proposed RMS circuit
Using the standard notation for the MO-CCCDTA, the circuit can be described by the following constitutive equations:

$$
\begin{equation*}
v_{p}=v_{n}=0 ; i_{z}=i_{p}-i_{n}, \text { and } i_{x}=g_{m} v_{z}=g_{m} Z_{z} i_{z} \tag{1}
\end{equation*}
$$

where $p$ and $n$ are input terminals, $z$ and $\pm x$ are output terminals, $g_{m}$ is the transconductance gain, and $Z_{z}$ is external impedance connected at the terminal z. Based on the expressions above, the current flow out of the terminal $z\left(i_{z}\right)$ is a difference between the input currents through the terminals $p$ and $n\left(i_{p}-i_{n}\right)$. The voltage drop at the terminal $z$ is transferred to the current at the terminal $x\left(i_{x}\right)$ by a transconductance gain $\left(g_{m}\right)$ of the CDTA. These currents, copied to a general number of output current terminals $x$, are equal in magnitude.

While there are several technologies to realize the CDTA, one of the possible bipolar realizations is shown in Fig. 2.


Figure 2: Bipolar realisation of MO-CCCDTA

It mainly comprises a current differencing circuit formed by two current followers, a basic current mirror and a multiple-output transconductance amplifier. Here, the transconductance gain $g_{m}$ of the CDTA is directly proportional to the external bias current $I_{B}$ (three bias currents $I_{B 2} I_{B 3}$ and $I_{B 4}$ in the case of proposed converter circuit, Fig. 1), which can be written by:

$$
\begin{equation*}
g_{m}=\frac{I_{B}}{2 V_{T}} \tag{2}
\end{equation*}
$$

where $V_{T}=26 \mathrm{mV}$ at $27^{\circ} \mathrm{C}$ is the usual thermal voltage given by $k T / q, k=$ Boltzmann's constant $=1.38 \times 10^{-23}$ $\mathrm{J} / \mathrm{K}, T=$ the absolute temperature (in Kelvins), and $q=$ $1.6 \times 10^{-19} \mathrm{C}$.

Generally, a MO-CCCII is a multiple-terminal active building block, as shown in Fig. 1. The port relations of the MO-CCCII can be presented by the following equation [1]:
$i_{y}=0 ; v_{x}=v_{y}+i_{x} R_{x} ; i_{z+}=+i_{x} ; i_{z-}=-i_{x}$
The bipolar realisation of the MO-CCCII is proposed in [18]. In this case, the parasitic resistance $R_{x}$ at the terminal $x$ can be expressed by:
$R_{x}=\frac{V_{T}}{2 I_{B}}$
where $V_{T}$ is the thermal voltage and $I_{B}\left(I_{B 1}\right.$ and $I_{B 5}$ in the proposed realisation, Fig. 1) is the bias current of the conveyor which remains tunable over several decades.

By the routine analysis of the proposed RMS circuit shown in Fig. 1 and using the properties of MO-CCCDTA and MO-CCCII, the output current at $z$ terminal of MO-CCCDTA is obtained by:
$I_{z}=I_{i n}=I_{x 3}$
whereupon the output voltage at $z$ terminal $\left(V_{z}\right)$ of MOCCCDTA equals:

$$
\begin{equation*}
V_{z}=\frac{I_{x 3}}{g_{m 3}}=\frac{2 V_{T} I_{i n}}{I_{r}} \tag{6}
\end{equation*}
$$

Figure 1 infers that $I_{B 2}=I_{\dot{\boldsymbol{n}}}, I_{B 3}=-I_{\dot{\boldsymbol{n}}}$ and $I_{B 4}=I_{r}$. Thus, the $I_{x 1}$ and $I_{x 2}$ can be obtained by:

$$
\begin{align*}
& I_{x 1}=\left\{\begin{array}{c}
g_{m 1} V_{z}=\frac{I_{\text {in }}^{2}}{I_{r}} \text { if } I_{\text {in }}>0 \\
0 \text { if } I_{\text {in }}<0
\end{array}\right. \\
& I_{x 2}=\left\{\begin{array}{c}
0 \text { if } I_{\text {in }}>0 \\
g_{m 2} V_{z}=\frac{I_{\text {in }}^{2}}{I_{r}} \text { if } I_{\text {in }}<0
\end{array}\right. \tag{7}
\end{align*}
$$

The equation above suggests the output current ${ }_{{ }_{\text {out }}}$ as:

$$
\begin{equation*}
I_{o u t}=I_{x 1}+I_{x 2}=\frac{I_{\text {in }}^{2}}{I_{r}}==\frac{R_{x 2}}{R_{\text {in }}^{2}} \frac{V_{\text {in }}^{2}}{V_{\text {out }}} \tag{8}
\end{equation*}
$$

Where $R_{\dot{j}}=R_{x 1}$. The current $I_{\text {out }}$ is then converted to the output voltage, $V_{\text {out }}$, with an implied low-pass filtering function. We can recognise that the output current-to-voltage conversion (with second CCCII) establishes a differential equation relating the current, $l_{\text {out }}$ to the output voltage, $V_{\text {out }}$ i.e.:

$$
\begin{equation*}
\dot{V}_{\text {out }}(t)+\omega_{0} V_{\text {out }}(t)=\frac{1}{C} I_{\text {out }}(t) ; \omega_{0}=\frac{1}{R C} \tag{9}
\end{equation*}
$$

A simple way to obtain this equation is to determine the transfer function relating $I_{\text {out }}$ to $V_{\text {out }}$ and then take this back to the time domain. Equation (9) is the generic time-domain description of a low-pass filter, where the coefficient of the undifferentiated term on the LHS of the equation equals the filter cut-off frequency. Equation (8) can subsequently be combined with the above to obtain:
$\left.\dot{V}_{\text {out }}(t)+\omega_{o} V_{\text {out }}(t)=\frac{1}{R_{1} C} \frac{V_{\text {in }}^{2}(t)}{V_{\text {out }}(t)} ; \quad R_{1}=\frac{R_{\text {in }}^{2}}{R_{x 2}} \quad 10\right)$
We may now multiply both sides of the equation by $2 V_{\text {out }}$ and make a simple observation incorporated into the final result:

$$
\begin{align*}
& 2 V_{\text {out }}(t) \dot{V}_{\text {out }}(t)+2 \omega_{o} V_{\text {out }}^{2}(t)=\frac{2}{R_{1} C} V_{\text {in }}^{2}(t) \Rightarrow  \tag{11}\\
& \frac{d}{d t}\left(V_{\text {out }}^{2}(t)\right)+2 \omega_{o}\left(V_{\text {out }}^{2}(t)\right)=\frac{2}{R_{1} C} V_{\text {in }}^{2}(t)
\end{align*}
$$

Equation (11) is a first-order differential equation relating $\left(V_{\text {out }}\right)^{2}$ and $\left(V_{\text {in }}\right)^{2}$, having the same form as (9). Therefore, the square of the output is a low-pass filtered version of the square of the input. Based on (11), we can assume that:

$$
\begin{equation*}
V_{\text {out }}^{2}(t)=\frac{2}{R_{1} C} \int_{0}^{t} e^{-2 \omega_{0}(t-\tau)} V_{\text {in }}^{2}(\tau) d \tau \tag{12}
\end{equation*}
$$

The equation above (convolution integral) implies that if the square root of both sides is considered, the output is the root mean square of the input voltage, where the integral is assumed to compute mean value function. The implied filtering function is thus given by:

$$
\begin{equation*}
H(s)=\frac{R}{R_{1}} \frac{\omega_{3 d b}}{s+\omega_{3 d b}}, \text { where } \omega_{3 d b}=\frac{2}{R C} \tag{13}
\end{equation*}
$$

The low-pass filter performs averaging of the RMS function and needs to be of a lower corner frequency than the lowest frequency of interest. For line frequency measurements, this filter is simply too large to implement on-chip, but the proposed detector requires only one capacitor on the output to implement the lowpass filter. This capacitor can be selected by the user, depending on frequency range and settling time requirements. Low-pass filtering the square of the input sine functions with some a certain amplitude, frequency and phase shift $\left(V_{\dot{\boldsymbol{n}}}(t)=V \cos (\omega t+\phi)\right.$ ), as suggested by (11), yields a time function $y(t)$ given by:

$$
\begin{equation*}
y(t)=\frac{V^{2}}{2}\left(1+\frac{1}{\sqrt{1+\left(2 \omega / \omega_{3 d b}\right)^{2}}} \cos (2 \omega t)\right)=V_{\text {out }}^{2}(t) \tag{14}
\end{equation*}
$$

The input phase shift, such as the net phase shift, after filtering of the second harmonic, yields zero phase, thereby simplifying the form of $y(t)$ without the loss of generality. $R / R_{1}$ was set to unity for simplicity reasons.

If we assume that the input signal frequency is considerably higher than the filter cut-off frequency, the approximate final output can be rather successfully estimated with just a few terms of a Taylor series. Accordingly, the DC component of the output voltage of the proposed circuit, i.e. the apparent output RMS value of the input and the associated second-harmonic component of the output voltage resulting from the rapidly decreasing magnitudes of higher harmonic terms, such as the ripple (peak-to-peak ripple of the output), is expressed as:

$$
\begin{align*}
& V_{R M S} \approx\left[1-\frac{1 / 16}{1+\left(2 \omega / \omega_{3 d b}\right)^{2}}\right] V_{\text {true }-R M S}  \tag{15}\\
& V_{\text {ripple }} \approx \frac{1 / 2}{\sqrt{1+\left(2 \omega / \omega_{3 d b}\right)^{2}}} V_{\text {true }}-R M S
\end{align*}
$$

The equation above infers the accuracy associated with the proposed circuit for measuring the effective value of the input sine voltage signal. In the case of the input signal described by the Fourier order, the estimation defined by (15) gains in complexity, whereas it also clearly implies that it is possible to filter and single out the effective value of the signal processed in the respective manner.

## 3 Non-ideal system analysis

The effects of MO-CDTA and MO-CCCII non-idealities on the RMS detector performance are to be considered in this section. By considering the non-ideal MO-CCCII characteristics, equation (3) can be rewritten as:

$$
\begin{equation*}
i_{y}=0 ; v_{x}=\alpha v_{y}+i_{x} R_{x} ; i_{z+}=+\beta_{p} i_{x} ; i_{z-}=-\beta_{n} i_{x} \tag{16}
\end{equation*}
$$

where $\alpha=1-\varepsilon_{v}$ and $\varepsilon_{v}\left(\left|\varepsilon_{v}\right| \ll 1\right)$ represents the voltage tracking error from $y$ to $x$ terminal, $\beta_{p}=1-\varepsilon_{p}$ and $\varepsilon_{p}$ $\left(\left|\varepsilon_{p}\right| \ll 1\right)$ denotes the current tracking error from $x$ to $+z$ terminal, while $\beta_{n}=1-\varepsilon_{n}$ and $\varepsilon_{n}\left(\left|\varepsilon_{n}\right| \ll 1\right)$ stands for the current tracking error from $x$ to $-z$ terminal of the MO-CCCII, respectively. Given the non-idealities, currents generated from first and second CCCIIs (first and third circuits of the proposed realization in Fig. 1) can be defined as:

$$
\begin{align*}
& I_{i n}=\frac{\alpha_{1} V_{\text {in }}}{R_{x 1}} ; i_{p}=I_{B 2}=\frac{\beta_{p 1} \alpha_{1} V_{\text {in }}}{R_{x 1}} \\
& I_{B 3}=\frac{\beta_{n 1} \alpha_{1} V_{\text {in }}}{R_{x 1}} ; I_{r}=\frac{\beta_{p 2} \alpha_{2} V_{o u t}}{R_{x 2}} \tag{17}
\end{align*}
$$

In practice, the deviation from the ideal performance of the proposed RMS circuits is mainly due to the nonideal CDTA characteristics, which can be divided into two categories, i.e. parasitic gain effects and parasitic impedance effects. Fig. 3 illustrates the simplified equivalent circuit represented by the behavior of the non-ideal CDTA.


Figure 3: The equivalent circuit of the non-ideal CDTA

A practical CCCDTA device can be modelled as an ideal CCCDTA with finite parasitic resistances and capacitances, as well as non-ideal current transfer gains and a transconductance inaccuracy factor of the CCCDTA. Fig. 3 shows a more sophisticated circuit model to represent the non-ideal CCCDTA device, where $R_{p^{\prime}} R_{n^{\prime}}$ $R_{x^{\prime}}$ and $R_{z}$ are the terminal parasitic resistances. $R_{p}$ and $R_{n}$ are the current-controllable parasitic resistances, where $R_{x}$ and $R_{z^{\prime}}$ as typical values of the parasitic resistances, connected to the terminals $x$ and $z$ respectively, are in the range of several mega-ohms. $C_{x}$ and $C_{z}$ are the terminal parasitic capacitances from terminals $x$ and $z$ to the ground (the shunt output impedances $\left(R_{z} / / C_{z}\right.$ and $\left.R_{\chi} / / C_{x}\right)$ at terminals $z$ and $x$, respectively). Typically, these parasitic capacitances are in the order of several pFs. In Fig. 3, $\alpha_{\mathrm{p}}$ represents the non-ideal current transfer gain from the $p$ terminal to the $z$ terminal of the CCCDTA, $\alpha_{n}$ denotes the non-ideal current transfer gain from the $n$ terminal to the $z$ terminal of the CCCDTA, and $\beta$ is the transconductance inaccuracy factor from the $z$ terminal to the $x$ terminal of the CCCDTA. The typical values of the non-ideal current transfer gains and the transconductance inaccuracy factor $\alpha_{n^{\prime}} \alpha_{p^{\prime}}$ and $\beta$ range from 0.9 to 1 , with an ideal value of 1 .

Based on the circuit representation in Fig. 3 and the proposed RMS detector, and given the non-ideal CDTA characteristics, after applying the non-ideal equivalent circuit mode of the CCCDTA to the proposed circuit, tedious derivations lead to the following modified characteristic equation):
$i_{p}=I_{B 2}=\frac{\alpha_{1} \beta_{p 1} V_{i n}}{R_{x 1}} ; I_{B 3}=\frac{\alpha_{1} \beta_{n 1} V_{\text {in }}}{R_{x 1}} ;$
$I_{r}=\frac{\alpha_{2} \beta_{p 2} V_{\text {out }}}{R_{x 2}}=I_{B 4}$
$V_{z}=\frac{R_{z}}{1+s R_{z} C_{z}}\left(\alpha_{p} i_{p}-\beta g_{m 3} V_{z}\right) \Rightarrow$
$V_{z}=\frac{\alpha_{p} R_{z}}{1+\beta g_{m 3} R_{z}+s R_{z} C_{z}} i p ;$
$g_{m 3}=\frac{I_{r}}{2 V_{T}}=\frac{\alpha_{2} \beta_{p 2} V_{o u t}}{2 V_{T} R_{x 2}}$
The modified output current for the proposed RMS detector can be rewritten as:
$I_{\text {out }}=\left\{\begin{array}{l}\beta g_{m 1} V_{z} \frac{\frac{R_{1 x}}{1+s C_{1 x} R_{1 x}}}{\frac{R_{1 x}}{1+s C_{1 x} R_{1 x}}+\frac{R}{1+s C R}}=\beta g_{m 1} V_{z} \frac{R_{1 x}(1+s C R)}{R_{1 x}(1+s C R)+R\left(1+s C_{1 x} R_{1 x}\right)} ; V_{i n}>0 \\ \beta g_{m 2} V_{z} \frac{\frac{R_{2 x}}{1+s C_{2 x} R_{2 x}}}{\frac{R_{2 x}}{1+s C_{2 x} R_{2 x}}+\frac{R}{1+s C R}}=\beta g_{m 2} V_{z} \frac{R_{2 x}(1+s C R)}{R_{2 x}(1+s C R)+R\left(1+s C_{2 x} R_{2 x}\right)} ; V_{\text {in }}<0\end{array}\right.$

It follows that:

where:

$$
\begin{align*}
& k_{1}(s)=\frac{\alpha_{1}^{2} \beta_{p 1}^{2} \alpha_{p} R_{z}}{\left(2 \frac{V_{T}}{V_{\text {out }}} R_{x 2}+\beta \alpha_{2} \beta_{p 2} R_{z}+2 \frac{V_{T}}{V_{\text {out }}} R_{x 2} s R_{z} C_{z}\right)}  \tag{22}\\
& k_{2}(s)=\frac{\alpha_{1}^{2} \beta_{n 1} \beta_{p 1} \alpha_{p} R_{z}}{\left(2 \frac{V_{T}}{V_{\text {out }}} R_{x 2}+\beta \alpha_{2} \beta_{p 2} R_{z}+2 \frac{V_{T}}{V_{\text {out }}} R_{x 2} s R_{z} C_{z}\right)}
\end{align*}
$$

The expressions above infer that the deviations in the transfer current gains are mainly the result of parasitic gains of the CDTAs. In order to improve the discrepancy to theoretical response, a high-performance CDTA with minor parasitic effects need to be employed. However, easy compensation for these deviations is possible by adjusting the values of $I_{B 1}$ and $I_{B 5^{\prime}}$ respectively. The output voltage of the proposed RMS detector is defined as:
$V_{\text {out }}=\beta g_{m i} V_{z} \frac{R^{\prime}}{1+s C^{\prime} R^{\prime}} ; R^{\prime}=\frac{R R_{i x}}{R+R_{i x}} ;$
$C^{\prime}=C+C_{i x} ; i=1,2$.
Given the non-ideal characteristics of MO-CDTA and CCCIIs, the implied filtering function implies that:
$H^{\prime}(s)=k_{i}(s) \frac{R^{\prime}}{R_{1}} \frac{\omega_{3 d b}^{\prime}}{s+\omega_{3 d b}^{\prime}}$,
where $\omega_{3 d b}^{\prime}=\frac{2}{R^{\prime} C^{\prime}} ; i=1,2$.
Equation (24) suggests that filtering function, represented by the integral operators (equation (12), poses different characteristics in comparison with the ideal situation (equation (13), especially in the operators' behavior at higher frequencies.

## 4 Simulation results

To confirm the given theoretical analysis, the proposed current-mode bipolar RMS circuit in Fig. 1 was simulated using the PSpice program. The CCCDTA and CCClls were realized by the schematic bipolar implementations given in Fig. 2 and [18], with the transistor model parameters of PR200N (PNP) and NP200N (NPN) of the bipolar arrays ALA400 from AT\&T [19]. The supply voltages and the values of the bias currents were $+\mathrm{V}=-\mathrm{V}$ $=1.2 \mathrm{~V}$ and $I_{B 1}=I_{B 5}=100 \mu \mathrm{~A}, I_{P}=300 \mu \mathrm{~A}$ respectively, whereas the input voltage was within the range of $0 \div$ 500 mV .

Fig. 4 shows the wave form of the signal at the output of the circuit shown in Fig. 1 (voltage $V_{\text {out }}(t)$ ), whereby the total power dissipation was 5.80 mW . Small power consumption of the proposed circuits results from the application of low-voltage current mode and transconductance mode integrated circuits, with the use of bipolar transistor technology. Applying the current mode signal processing to solve the issues under consideration is a sensible approach to the problem. However, similar and sometimes lower power consumption can
be achieved using CMOS technology instead of the bipolar one.

The output ripple is always considerably greater than the DC error; therefore, filtering out the ripple can substantially reduce the peak error without applying a long settling-time penalty by simply increasing the averaging capacitor. The rippling of the output voltage generated in this manner is lower than in detector $[16,20]$, followed by the shorter feedback as well. Linearity may seem like an odd property for a device that implements a function involving two very nonlinear processes: squaring and square rooting. However, an RMS-to-DC converter has a transfer function, RMS volts in to DC volts out, that should ideally have a 1:1 transfer function. To the extent that the input to output transfer function does not lie on a straight line, the part is nonlinear. Fig. 5 (a) shows the DC transfer function nearing zero in the proposed circuit. Given that the dynamic range has nonlinearity level lower than 1dB, the dynamic range of the circuit proposed in this paper is around 35 dB . The proposed detector circuit involves higher linearity compared to the ones described in [16, 20, 21].


Figure 4: Time-domain response of the proposed RMS circuit for the sine input signal
$\left(V_{\text {in }}(t)=10 \sin (2 \pi f t)[\mathrm{mV}], f=100 \mathrm{kHz}, R=180 \Omega, C=5 \mu \mathrm{~F}\right)$


Figure 5: a) DC transfer function near zero; b) Performance vs Crest Factor

Crest Factor represents a common method of describing dynamic signal wave shapes. It is the ratio of the peak value relative to the RMS value of a waveform. For example, a signal with a crest factor 4 has a peak four times its RMS value. The proposed circuit performs very well with crest factor 4 or less, and responds with a reduced accuracy to signals with higher crest factors (Fig. 5 (b)). On the Fig. 5 (b) the "SCR waveforms" refers to the ideally chopped sine wave. High performance with crest factors lower than 4 can be directly attributed to the high linearity throughout the proposed solution.

Fig. 4 shows the result for pure sinusoid signal. However, as an RMS detector, the circuit should have consistent response in signals with equal powers but various waveform shapes. Thus, the circuit was simulated using various input waveforms to verify the RMS power
detection function. The simulated detector responded to the single-tone sinusoid, two-tone signals (with frequencies of 1 MHz and 3 MHz , and amplitudes of 100 mV and 50 mV ), square-waves (duty cycle $=50 \%$ ) and triangle waves given in Fig. 6. All the used signals were at 1 MHz . The relative errors were lower than 0.04 $\%$ for $P_{\text {in }} \leqq-20 \mathrm{dBm}$. Given that the dynamic range has nonlinearity level lower than 1 dB , the dynamic range of the circuit proposed in this paper is around 36 dB .

The frequency responses, dynamic range of this bipolar detector, were all comparable and even superior to most diode detectors. The error in computing the effective value of the processed input voltage signal was lower than in [16, 22-26], whereby the circuit of the proposed detector, which includes a wider dynamic range, facilitates the realization more favourably than those described in [21, 23, 26, 27]. Similarly, it does not require a specific compensation procedure.


Figure 6: The simulated response of a single detector to various waveforms.

## 5 Conclusion

This paper reports on a new electronically controllable bipolar translinear RMS-to-DC converter. The proposed circuit employs two CCCIls, one CDTA and two grounded passive elements, which is advantageous for integration point of view. The proposed circuit ensures high precision, wide bandwidth and high accuracy. The PSPICE simulation results were depicted, and they agree well with the theoretical anticipation.

## 6 Acknowledgments

The author wishes to thank Ministry of Education and Science of the Republic of Serbia their support to this work provided within the projects 42009 and OI172057.

## 7 References

1. R. B. Northrop, Analog Electronics Circuits, Reading, MA: Addison-Wesley, 1990.
2. P. Heavey, and C. Whitney, "RMS measuring principles in the application of protective relaying and metering", in Proc. 57th Annu. Conf. Protective Relay Eng.(2004), pp. 469-489.
3. U. Pogliana," "Precision measurement of ac voltage below 20 Hz at IEN", IEEE Trans. Instrum. Meas., vol. 46, no. 2, pp. 369-372, 1997.
4. H. Germer, "High-precision AC measurements using the Monte-Carlo method", IEEE Trans. Instrum. Meas., vol. 50, no. 2, pp. 457-460, 2001.
5. W.-K. Yoon, and M.J. Deveney, "Power measurement using the wavelet transform", IEEE Trans. Instrum. Meas., vol. 47, no. 5, pp. 1205-1210, 1998.
6. M. Novotny, and M. Sedlacek, "RMS value measurement based on classical and modified digital signal processing algorithms", Measurement, vol. 41, no. 3, pp. 236-250, 2008.
7. True RMS' detector, National semiconductor Application Note AN008474, 2002.
8. DSCA33 ISOLATED True RMS Input Module, AN101 Dataforth Corporation, USA 2011.
9. High Precision, Wide-Band RMS-to-DC Converter, Analog Devices Application Note AD637, 2011.
10. J. Mulder, W. A. Serdijn, A. C. Woerd, and A. H. M. Roermund, "Dynamic translinear RMS-DC converter", Electron Lett., vol. 32, pp. 2067-2068, 1996.
11. J. Mulder, W. A. Serdijn, and A. H. M. Roermund, "An RMS-DC converter based on the dynamic translinear principle", IEEE Solid-State Circuits, vol. 32, pp. 1146-1150, 1997.
12. W. Surakampontron and K. Kumwachara, "A dual translinear-based RMS-to-DC converter", IEEE Trans. Instrum. Meas,. vol. 47, pp. 456-464, 1999.
13. R. F. Wasseneaar, E. Seevinck, M. G. van Leeuwen, C. J. Speelman, and E. Holle, "New Techniques for High-Frequency RMS-to-DC Conversion Based on a Multifunctional V-to-I Convertor", IEEE Jour. Sol. Sta. Circ., vol. 23, no. 3, pp. 802-815, 1998.
14. V. Milanović, M. Gaitan, E. D. Bowen, N. H. Tea, and M. E. Zaghlou, "Thermoelectric power sensors for microwave applications by commercial CMOS fabrication", IEEE Elec. Dev. Lett., vol. 18, no. 9, pp. 450-452, 1997.
15. W. Tangsrirat W, T. Dumawipata, and W. Surakampontorn, "Multiple-input single output currentmode multifunction filter using current differencing transconductance amplifiers", Int J Electron Commun (AEU)., vol. 61, pp. 209-214, 2007.
16. P. Petrovic, "RMS Detector of Multiharmonic Signals", ETRI Journal, vol. 35, no. 3, pp. 431-438, 2013.
17. P. Petrovic, and I. Zupunski, "RMS detector of periodic, band-limited signals based on usage of DO-

CClls", Measurement, vol. 46, no. 9, pp. 3073-3083, 2013.
18. W. Tangsrirat, "Current-tunable current-mode multifunction filter based on dual-output cur-rent-controlled conveyors", Int. J. Electron. Commun. (AEÜ), vol. 61, pp. 528-533, 2007.
19. D. R. Frey, "Log-domain filtering: an approach to current mode filtering", IEE Proc Circuit Devices Syst., vol. 140, pp. 406-416, 1993.
20. B. Rumberg, and D. W. Graham, "A Low-Power Magnitude Detector for Analysis of Transient-Rich Signals", IEEE Jour. Sol. Sta. Circ., vol. 47, no. 3, pp. 676-685, 2012.
21. C. Yu, C. L. Wu, S. Kshattry, Y. H. Yun, C. Y. Cha, H. Shichijo, and K. O Kenneth, "Compact, High Impedance and Wide Bandwidth Detectors for Characterization of Millimeter Wave Performance", IEEE Jour. Sol. Sta. Circ., vol. 47, no. 10, pp. 2335-2343, 2012.
22. Y. Zhou, and M. Y. W. Chia, "A Low-Power UltraWideband CMOS True RMS Power Detector", IEEE Trans. on Mic. The. Tec., vol. 56, no. 5, pp. 10521058, 2008.
23. Q. Yin, W. R. Eisenstadt, R. M. Fox, and T. Zhang, "A Translinear RMS Detector for Embedded Test Of RF ICs", IEEE Trans. Instrum. Meas., vol. 54, no. 5, pp. 1708-1714, 2005.
24. K. Kaewdang, K. Kumwachara, and W. Surakampontorn, "A translinear-based true RMS-to-DC converter using only npn BJTs", AEU-Intern. Jour.Elec. Comm., vol. 63, no. 6, pp. 472-477, 2009.
25. E. Farshidi, and H. Asiaban, "A new true RMS-toDC converter using up-down translinear loop in CMOS technology", Analog Integrated Circuits and Signal Processing, vol. 70, no. $\underline{3}$, pp 385-390, 2012.
26. J. Koton, N. Herencsar, and K. Vrba, "Current and Voltage Conveyors in Current and Voltage-Mode Precision Full-Wave Rectifiers", RADIOENGINEERING, vol. 20, no. 1, pp. 19-24, 2011.
27. G. Klahn, "True RMS power detection with high dynamic range", in Proceeding IEEE MTT-S International Microwave Symposium Digest, (1999) vol. 4, pp. 1773-1776.

Arrived: 18. 12. 2014
Accepted: 18.02. 2015

# A Novel Dual Ports Antenna for Handheld RFID Reader Applications 

Bo Wang, Yiqi Zhuang and Xiaoming Li

Xidian University, Xi'an 710071, China


#### Abstract

A compact antenna utilizes two ports to transmit and receive signal separately, different with conventional handheld RFID readers with single port. The proposed antenna can enhance receive sensitivity of handheld RFID readers, since the strong transmitting signal of reader with single port is usually highly coupled with weak receiving backscatter signal of tag. The antenna uses U-shape aperture coupled patch structure that occupies less volume and provides further space-saving efficiency. It is fed by two T-shape microstrip lines with rectangle stubs. The U-shape apertures are used to excite two orthogonal modes for dual polarized operation. The height of the air substrate is reduced to only 4 mm ( 0.032 wavelength) and the volume of antenna is $80 \mathrm{~mm} \times 80 \mathrm{~mm} \times 6.8 \mathrm{~mm}$, which is easy to integrate in Handheld RFID readers. The measured results show -10 dB matching band and -25 dB isolation band from 2.2 GHz to 2.6 GHz and from 2 GHz to 2.6 GHz , respectively. The minimum isolation is -50 dB at 2.48 GHz . The antenna is suitable for applications in handheld RFID readers.


Keywords: handheld RFID reader antenna; two ports; high isolation

# Nova dvovhodna antenna za ročne RFID bralnike 

Izvleček: Kompaktna antena ima dva vhoda za ločeno pošiljanje in sprejemanje signala, kar je različno od običajnih enovhodnih RFID bralnikov. Predlagana antena omogoča večjo sprejemno občutljivost, jas je močen oddajen signal enovhodnih bralnikov večinoma močno sklopljen s šibkim bralnim signalom. U oblika antene porabi manj prostora in omogoča dva ortogonalna načina delovanja za dvopolarizirano delovanje. Napajana je z dvemi T trakastimi linijami T oblike. Višina zračnega substrata je le 4 mm ( 0.032 valovne dolžine), velikost $80 \mathrm{~mm} \times 80 \mathrm{~mm} \times 6.8 \mathrm{~mm}$, kar omogoča enostavno integracijo v ročne RFID bralnike. Meritve izkazujejo ujetost -10 dB in izolativnost pasu -25 dB v območju 2.2 do 2.6 GHz . Najmanjša izolativnost pri 2.48 GHz je -50 dB .

Ključne besede: ročna RFID bralna antena; dva vhoda; visoka izolativnost

* Corresponding Author's e-mail: wangbo_chen@126.com


## 1 Introduction

Recently, the use of radio frequency identification (RFID) systems has become widespread in a variety of applications. Furthermore, handheld RFID readers have become very popular with users, particularly in applications that need to control large and heavy products, which are not easy to move.

Handheld RFID readers reported are most single port with various structures [1-7]. RFID system consists of a tag and reader. The reader transmits a continuous wave (CW) signal and the tag backscatters transmission from the reader to send back data. In a backscatter reader, the transmitted CW signal may be directly coupled to the receiving part of the reader to drastically degrade the receiving sensitivity. The directly coupled CW signal is much larger than the backscatter signal from the tag, and the receiving part of the reader should detect
the weak signal close to such a strong in-band interfere. Therefore, it is essential to separate transmitting and receiving parts with dual ports to achieve high isolation between them.

Over the past years, dual ports reader antenna designs have received considerable attention. Among dual polarized antenna designs, aperture coupled microstrip patch antenna are the most suitable candidates for RFID application [8-17]. Aperture coupling is preferred to other feeding mechanisms of microstrip patch antenna due to its greater design flexibility, easier fabrication and lower cost. The antenna in [8] is utilizes a resonant annular ring slot and a T-shaped microstrip feedline to coupled with radiating patch, thus exciting dual orthogonal linearly polarized mode. The $2 \times 2$ array employing two symmetric dog-bone shaped coupling apertures is proposed to introduce dual linearly
(a)

(b)

(c)

(d)


Figure 1: (a) Side view of the proposed antenna (b) configurations of the proposed antenna (c) top view of the fabricate antenna (d) bottom view of the fabricate antenna
polarized mode [9]. A common method to increase ports isolation is combining branch line with antennas [10-12]. In [13], the antenna is designed with simple microstrip feedline to couple with radiating patch, but performs badly ports isolation with 20 dB . Majority of aperture coupled antennas apply the approach for addressing the requirement for low signal correlation is to increase the height of air substrate to achieve high ports isolation [14-17]. Since, the antenna is to be used with a handheld RFID reader, the size of the antenna in general should be around 100 mm length and width, and around 10 mm in thickness [3]. Therefore, most of open literatures including [8-17] described reader antennas with dual ports are comparable large to be mounted onto a handheld RFID reader, however they are suitable for stationary readers.

The rest sections are arranged as followed: section II presents the detail design and principle of the aperture coupled patch antenna. The measured isolation and impedance matching of the proposed antenna are discussed in section III. In section IV, the parameters of stubs are simulated and analysis. Finally, the conclusions are given in section V .

## 2 Antenna structure and design

In traditional designs of aperture coupled antenna, they are using various shapes of apertures in the ground plane. But these apertures technique requires high air layer in order to reduce the coupling between the two feeding lines, thus increase the volume of antenna inconvenient of integrated in the handheld RFID reader. So this paper applies a novel shape aperture to decrease the air layer.

The configuration of a dual feeding aperture coupled square patch antenna is shown in Figure 1. It consists of two FR4 substrates with dielectric constant of 4.4 and loss tangent of 0.02. A single-layer substrate ( $56 \mathrm{~mm} \times 56 \mathrm{~mm} \times 1.2 \mathrm{~mm}$ ) is suspended $4 \mathrm{~mm}\left(0.032 \lambda_{0}\right.$, $\lambda_{0}$ is free space wavelength) above the double-layer substrate $(80 \mathrm{~mm} \times 80 \mathrm{~mm} \times 1.6 \mathrm{~mm})$. A square patch of $50 \mathrm{~mm} \times 50 \mathrm{~mm}$ is etched on the top side of the singlelayer substrate. The overall volume of proposed antenna is $80 \mathrm{~mm} \times 80 \mathrm{~mm} \times 6.8 \mathrm{~mm}$. Two $50 \Omega$ modified T-shape microstrip lines with width of $\mathrm{W}_{\mathrm{f}}=3 \mathrm{~mm}$ and length of $L_{s}=43 \mathrm{~mm}$ are fed by separate port 1 and port 2 on the bottom side of double-layer substrate. The ground plane with U-slots is etched on the top side. The optimized values of stub width stub-x and length stub-y are 9 mm and 7 mm , respectively. The mathematical equations for calculating the $W_{f}$ and $L_{s}$ are as follows:

$$
\begin{align*}
& \frac{W_{f}}{h}=\left(\frac{1}{8} e^{\mathrm{A}}-\frac{1}{4 \mathrm{e}^{\mathrm{A}}}\right)^{-1}  \tag{1}\\
& \mathrm{~A}=\frac{\mathrm{Z}_{\mathrm{c}} \sqrt{2\left(\varepsilon_{\mathrm{r}}+1\right)}}{119.9}+\frac{1}{2} \frac{\varepsilon_{\mathrm{r}}-1}{\varepsilon_{\mathrm{r}}+1}\left(\ln \frac{\pi}{2}+\frac{1}{\varepsilon_{\mathrm{r}}} \ln \frac{4}{\pi}\right) \tag{2}
\end{align*}
$$

$\mathrm{L}_{\mathrm{s}} \approx \frac{\lambda_{\mathrm{g}}}{4}$
$\lambda_{\mathrm{g}}=\frac{\lambda_{0}}{\sqrt{\varepsilon_{\mathrm{r}}}}$
Where $\lambda_{g}$ is the dielectric wavelength, $\lambda_{0}$ is the air wavelength, $\varepsilon_{r}$ is the effective dielectric constant and $h$ the substrate thickness. The substrate thickness, h , in this paper is 1.6 mm .

The microwave signal is transmitting or receiving through feeding lines. Since the electromagnetic energy is along the feeding lines, the apertures are etched above feeding lines in the ground plane to couple energy to the patch. The square patch is served as a radiator to transmit or receive signals. The feeding line of port 1 excites horizon linear polarization, while that of port 2 excites vertical linear polarization. The two orthogonal polarizations decrease the coupling between two ports. Furthermore, this paper adds two stubs in the end of the feed lines to improve impedance matching and isolation. The current concentrates in the stubs, thus introduces capacitive couple to the square patch. The stubs increase effectively the isolation and impedance matching of proposed antenna. In addition, spurious radiation from the feeding lines is eliminated due to ground plane shielding, resulting in a very low cross polarization level.

An aperture coupled antenna has a narrow bandwidth and poor isolation. Additional stacked patch is utilized to improve the bandwidth. The resonant frequency is mainly determined by the size of the square patch and the amount of coupling is dependent on the aperture length. The advantage of isolating the patch from the feeding line, better radiation pattern symmetry caused by the apertures and impedance matching was obtained through the use of aperture coupled patch antenna. For dual polarization radiation, a square patch is coupled to a pair of microstrip lines through U-shape apertures located beneath the patch, which improves the radiation characteristics of the antenna. The length and width of the aperture have been optimized for acceptable optimum isolation and return loss in the desired band.
(a)

(b)


Figure 2: Surface current distribution of proposed antenna on (a) Feeding lines (b) Ground plane

To investigate the mechanism of mutual coupling between two ports, current distributions in different layers under the patch have been simulated with port 1 excited and port 2 terminated. Thus, we simulate the proposed antenna and get the surface current distribution at 2.45 GHz on the ground plane and feeding lines shown in Figure 2. Figure 2(a) demonstrates that the microwave energy concentrates in junction and stub of T-shape feed line. It can be seen in Figure 2(a) that surface current is flowing along the feeding line from port 1 to port 2 , and gradually attenuated. The current around port 2 is greatly weaker than that around port 1. It is demonstrated an excellent isolation between port 1 and port 2. Figure 2(b) describes that currents concentrate in the specific region of ground plane which is above the stub of feed line and the other end of feed line without stub has less currents. It is concluded that stub has much effects on increasing currents.

Figure 2(b) shows the current around U-slot is decreasing much, attribution to impedance matching.
(a)

(b)

(c)

(d)


Figure 3: Radiating patch surface current distributions for two different phase intervals. (a) $0^{\circ}$ of port 1 (b) $90^{\circ}$ of port 1 (c) $0^{\circ}$ of port $2(\mathrm{~d}) 90^{\circ}$ of port 2

To better understand the excitation behavior of the antenna, Figure 3 only shows the current distributions of phase $0^{\circ}$ and $90^{\circ}$ in port 1 and port 2, respectively, since those of $180^{\circ}$ and $270^{\circ}$ are equal in magnitude and opposite in phase of $0^{\circ}$ and $90^{\circ}$. It is clearly displayed that surface currents cause linear polarization with time and two ports produces orthogonal fields. Due to the symmetrical structure of the proposed antenna, the Tx and Rx port can interchange to create linear polarization. Thus, the proposed antenna has dual linear polarization in one structure, orthogonal polarization improves isolation between two ports.

## 3 Performance of aperture coupled patch antenna

Figure 4(a) shows the simulated and measured return loss of the antenna. The simulated return loss is less than -10 dB over the frequency band of 2.19 GHz to 2.58 GHz , while the measured return loss bandwidth is 400 MHz from 2.2 GHz to 2.6 GHz . It is clearly seen in Figure $4(\mathrm{~b})$ that the measured -25 dB bandwidth of 2-2.6 GHz is obtained with minimum -50 dB at 2.48 GHz , corresponding to the simulated bandwidth of 510 MHz . The simulated and measured peak gain is illustrated in Figure 4(c). The antenna exhibits the measured peak gain from 1.5-3.1 dBi according to the frequency band of $2.4-2.48 \mathrm{GHz}$. The measured and simulated return loss, isolation and peak gain show good agreement. In microwave band, antenna gain is not as critical since active tags are commonly used in many applications.

Figure 5 shows the measured radiation patterns at 2.45 GHz in the orthogonal $\mathrm{XOZ}\left(\mathrm{phi}=0^{\circ}\right)$ and $\mathrm{YOZ}\left(\mathrm{phi}=90^{\circ}\right)$ planes with angular step of $20^{\circ}$. The radiation pattern in YOZ plane is like bow-tie, but that in XOZ plane is unidirectional.

## 4 Parameters simulation and analysis

The parameters simulation is carried out to provide antenna engineers with the information for antenna design and optimization. The length stub-x and stub-y of the stub are the prime parameters that determine the amount of power concentrated in the stub and coupled to the radiating patch such that effect the impedance matching and isolation of proposed antenna. One physical attribute of the antenna is independently varied, while the other parameter is kept unchanged. For clearly visualize, the finial optimized parameters are depicted with red line in each simulation figure. Software High Frequency Structure Simulation (HFSS)


Figure 4: The characters of proposed antenna(a) Return loss (b) Isolation (c) Peak gain
based on finite element method is used in this analysis. The finally optimized values are stub- $x=9 \mathrm{~mm}$, stub$y=7 \mathrm{~mm}$.

The dependencies of the return loss and isolation on stub-x are described in Figure 6. Figure 6(a) and Figure 6(b) describe that the bandwidth of return loss (<-10 dB ) and isolation ( $<-25 \mathrm{~dB}$ ) are expanding with decreasing length of stub-x. Figure 6(a) shows that the return loss bandwidth is decreasing and resonate frequency is shift to lower frequency as the length of stub-x increasing. It is observed that isolation is reducing dramatical-


Figure 5: Measured radiation patterns of proposed antenna at 2.45 GHz
(a)

(b)


Figure 6: Antenna characters for different values of stub-x. (a)Return loss (b)Isolation
ly as the length of stub-x increasing in Figure 6(b). The minimum of isolation with stub-x of 9 mm is -60 dB at 2.45 GHz. The stub length stub-x determines the coupling strength between the feed line and ground. So it has an impact on both of the return loss and isolation.


Figure 7: Antenna characters for different values of stub-y. (a)Return loss (b)Isolation

In Figure 7, the effects of various dimensions of stub$y$ on return loss and isolation are shown. The variation of stub-y affects slightly on return loss, but severely on isolation value. The bandwidths of return loss and isolation are almost same, but optimized value of stub-y of 7 mm shows the best isolation in $2.4 \mathrm{GHz}-2.48 \mathrm{GHz}$.

## 5 Conclusion

A compact antenna with two ports is designed for Handheld RFID reader to enhance receive sensitivity. It is low cost and easy to integrate in the Handheld RFID reader for its height of 6.8 mm . The proposed antenna presents impedance matching of -10 dB and isolation of -35 dB . The -10 dB matching band and -25 dB isolation band cover from 2.2 GHz to 2.6 GHz and from 2 GHz to 2.6 GHz , respectively.

## 6 Acknowledgment

The paper supported by "the Fundamental Research Funds for the Central Universities" (No. JB141107).

## 7 References

1. A.T. Mobashsher and R.W. Aldhaheri, "An improved uniplanar front-directional antenna for dual-band RFID readers," IEEE Antennas and Wireless Propagation Letters, vol. 11, pp. 14381441, 2012.
2. J.H. Bang, B.O. Chinzorig, H.S. Koh, E.J. Cha and B.C. Ahn, "A small and lightweight antenna for handheld RFID reader applications," IEEE Antennas and Wireless Propagation Letters, vol. 11, pp. 10761079, 2012.
3. S.X Ta, H.S. Choo and I. Park, "Planar, lightweight, circularly polarized crossed dipole antenna for handheld UHF RFID reader," Microwave and Optical Technology Letters, vol. 55, no. 8, pp. 18741878, August 2013.
4. W.S. Chen and Y.C. Huang, "A Novel CP Antenna for UHF RFID Handheld Reader", IEEE Antennas and Propagation Magazine, vol. 55, pp. 128-137, 2013.
5. P.V. Nitikin and K.V.S. Rao, "Compact Yagi Antenna for Handheld UHF RFID Reader", IEEE International Symposium Antennas and Propagation an CNC-USNC/URSI Radio Science Meeting, 2010.
6. H.T. Hsu and T.J. Huang, "A Koch-Shaped Log-Periodic Dipole Array (LPDA) Antenna for Universal Ultra-High-Frequency (UHF) Radio Frequency Identification (RFID) Handheld Reader", IEEE Transactions on Antennas and Propagation, vol. 61, pp. 4852-4856, 2013.
7. Y.F. Lin, H.M. Chen, C.H. Chen and C.H. Lee, "Compact shorted inverted-L antenna with circular polarisation for RFID handheld reader", Electronics Letters, vol. 49, pp. 442-444, 2013.
8. C.Y.D. Sim, C.C. Chang and J.S. Row, "Dual-feed dual-polarized patch antenna with low cross polarization and high isolation", IEEE Transaction on Antennas and Propagation, vol. 57, pp. 33213324, 2009.
9. S.K. Padhi, N.C. Karmakar and C.L. Law, "Dual polarized reader antenna array for RFID application", IEEE Antennas and Propagation Society International Symposium, vol. 4, pp. 265-268, 2003.
10. H.W. Son, J.N. Lee and G.Y. Choi, "Design of compact RFID reader antenna with high transmit/receive isolation", Microwave and Optical Technology Letters, vol. 48, pp. 2478-2481, 2006.
11. X.Z Lai, Z.M. Xie, Q.Q. Xie, X.L. Cen, "A dual circularly polarized RFID reader antenna with wideband isolation", Antennas and Wireless Propagation Letters, vol. 12, pp. 1630-1633, 2003.
12. Y.K. Jung and B. Lee, "Dual-Band Circularly Polarized Microstrip RFID Reader Antenna Using Metamaterial Branch-Line Coupler", IEEE Transaction on Antennas and Propagation, vol. 60, pp. 786791, 2013.
13. M.T. Zhang, Y.B. Chen, Y.C. Jiao and F.S. Zhang, "Dual Circularly Polarized Antenna of Compact Structure for RFID Application", Journal of Electromagnetic Waves and Applications, vol. 20, pp.1895-1902, 2006.
14. B. Li, Y.Z. Yin, Y. Zhao, Y. Ding and R. Zou, "Dual-polarised patch antenna with low cross-polarisation and high isolation for WiMAX applications", vo. 47, pp. 952-953, 2011.
15. K. Zhang, F.G. Zhu and S. Gao, "Differential-fed ultra-wideband slot-loaded patch antenna with dual orthogonal polarization", Electronics Letters, vol. 49, pp. 1591-1593, 2013.
16. C.H. Weng, H.W. Liu, C.H. Ku and C.F. Yang, "Dual circular polarisation microstrip array antenna for WLAN/WiMAX applications", Electronics Letters, vol. 46, pp. 609-611, 2010.
17. J.J. Xie, Y.Z. Yin, J.H. Wang and X.L. Liu, "Wideband dual-polarised electromagnetic fed patch antenna with high isolation and low cross-polarisation", Electronics Letters, vol. 49, pp. 171-173, 2013.

Arrived: 29. 12. 2014
Accepted: 09.02. 2015

# The optimal useful measurement range of an inductive displacement sensor 

Snežana M. Djurić, Nikola M. Djurić, Mirjana S. Damnjanović<br>Faculty of Technical Sciences, University of Novi Sad, Serbia


#### Abstract

The purpose of this paper is to find the optimal useful measurement range of an inductive displacement sensor with meander type coils. The optimal useful measurement range was numerically examined using a developed model for impedance calculation. The sensor is composed of two sensor elements, with meander-type inductive coils. Each coil has five turns. With these two sensor elements, it is possible to detect normal displacement (using only one sensor element) and tangential displacement (using both sensor elements). Numerical results showed that the optimal useful measurement range was obtained when the gap of 0.23 mm was inserted in one of the coils of sensor element detecting normal displacement. Experimental results confirmed theoretical predictions. The paper demonstrates developing of a model for impedance calculation of an inductive displacement sensor. With this model, it was possible to determine numerically the optimal useful measurement range of the sensor.


Keywords: inductive coils; inductance calculation; measurement range; displacement

# Optimalno uporabno območje induktivnega senzorja premika 

Izvleček: Namen članka je poiskati uporabno merilno območje induktivnega senzorja premika z meandrasta tuljavami. Optimalno področje je bilo numerično določeno z razvitim modelom za izračune impedanc. Senzor je sestavljen iz dveh senzorskih elementov z meandrasto tuljavo. Vsaka tuljava ima pet zavojev. S tema dvema senzorskima elementoma je mogoče zaznati običajne premike (z uporabo le enega senzorja) in tangencialne premike (pri uporabi obeh senzorjev). Numerični izračuni optimalnega merilnega območja so pri uporabi 0.23 mm reže v enem senzorju pri detekciji normalnega premika. Meritve potrjujejo teoretične izračune. V članku je predstavljen razvoj modela, ki omogoča določitev optimalnega merilnega območja senzorja.

Ključne besede: induktivne tuljave ; izračuni induktivnosti; merilno območje; premik

[^2]
## 1 Introduction

The planar inductive coil sensors have a large scale of application. They can be applied in the inspection of printed circuit boards using eddy-current testing (ECT) technique [1, 2, 3]. The development and comparison of different planar fluxgate magnetic sensor structures realized in PCB technology has been reported in [4]. The planar inductive sensor with planar coil and magnetic core can detect the cracks on nonmagnetic and magnetic specimens [5]. The linear displacement sensor based on the inductive concept using meander coil and pattern guide is used to detect the displacement of moving part on linear machines [6]. The effect of inductive coil shape (meander, square, and circle shape with different turn number of inductive coils) on the sensing performance of a linear displacement sensor has been
analyzed in [7]. A planar inductive coil of circle shape is used in an eddy-current sensor for high resolution displacement detection with reduced temperature coefficient [8]. An eddy current senor with rectangular sensing element, printed by ink-jet technology on a flexible substrate, for displacement application, has been presented in [9]. An inductive sensor for distance measurement employs the principle of magnetic coupling between two coplanar coils [10]. Sensors, fabricated in PCB technology, with planar meander and interdigital coils in series and parallel combination, are used for measurement and monitoring of environmental parameters [11, 12].

In our previous papers [13, 14], design, modeling, and operating principle of an inductive displacement sen-
sor, with meander-type inductive coils, was presented. The sensor is composed of two sensor elements. Each sensor element presents a pair of meander coils. One sensor element detects normal displacement, whereas the other sensor element detects tangential displacement. Sensor element for normal displacement can be used independently, whereas the sensor element detecting tangential displacement is used in combination with the element detecting normal displacement. The sensor element that detects normal displacement, with inserted gap $g$ in the stationary coil, is presented in Figure 1. The width of the segments in the stationary coil is $w_{1}=1.52 \mathrm{~mm}$, in the moving (short-circuited coil) the width of the segments is $w_{2}=0.51 \mathrm{~mm}$. The distance between axes of two neighboring segments is $p=1.78 \mathrm{~mm}$ and the number of turns is five. The gap width influences the useful measurement range of the sensor. The useful measurement range of the sensor is near $y=0$ (zero position - the axes of the segments of the stationary coil are exactly above the axes of the segments of the moving coil.) In this range the input inductance of the sensor element detecting normal displacement is invariant versus tangential displacement ( $y$-direction), thus the element detects only normal displacement. The goal of this paper was to examine the optimal useful measurement range of the sensor.


Figure 1: The sensor element, detecting normal displacement, with inserted gap in the stationary coil.

## 2 Model of the sensor

Each sensor element can be described with its equivalent circuit as it is shown Figure 2, where $R_{1}$ and $R_{2}$ are resistances of the stationary coil (Coil 1) and moving coil (Coil 2), $L_{1}$ and $L_{2}$ are the self-inductances of Coils 1 and 2 , respectively [15].


Figure 2: Equivalent electrical circuit of sensor element.
The input impedance of sensor element is equal to the input impedance of the equivalent circuit:
$\underline{U}_{1}=R_{1} \underline{I}_{1}+j \omega L_{1} \underline{I}_{1}+j \omega M_{12} \underline{I}_{2}$
$R_{2} \underline{I}_{2}+j \omega L_{2} \underline{I}_{2}+j \omega M_{12} \underline{I}_{1}=0$
$\underline{I}_{2}=-\frac{j \omega M_{12}}{R_{2}+j \omega L_{2}} \underline{I}_{1}$
$\underline{U}_{1}=\left(R_{1}+j \omega L_{1}\right) \underline{I}_{1}-j \omega M_{12} \frac{j \omega M_{12}}{R_{2}+j \omega L_{2}} \underline{I}_{1}$
$\underline{Z}_{I N}=R_{I N}+j \omega L_{I N}$
where the total resistance of the impedance is
$R_{I N}=R_{1}+\frac{\omega^{2} R_{2} L_{1} L_{2} k^{2}}{R_{2}^{2}+\omega^{2} L_{2}^{2}}$
and the total reactance of the impedance is
$\omega L_{I N}=\omega L_{1} \frac{R_{2}^{2}+\omega^{2} L_{2}^{2}\left(1-k^{2}\right)}{R_{2}^{2}+\omega^{2} L_{2}^{2}}$
Mutual position between the coils introduces magnetic coupling between coils. The coupling coefficient $k$ is

$$
k=\frac{M_{12}}{\sqrt{L_{1} L_{2}}}
$$

where $M_{12}$ is the mutual inductance between Coils 1 and 2 , for specific mutual position, and $L_{1}$ and $L_{2}$ are the self-inductances of Coils 1 and 2 , respectively. The mutual inductance changes according to displacement between Coils 1 and 2. It can be assumed that the current of conductive segments is uniformly distributed over the whole cross-section because of relatively low working frequency ( 1 MHz ). At this relatively low frequency, the skin and proximity effects are negligible ( $\delta=\sqrt{\rho_{C u} / \pi f \mu_{0}}$, where $\rho_{C u}=1.72 \times 10^{-8} \Omega$ is electrical resistivity of copper $\mathrm{Cu}, f$ the working frequency and $\mu_{0}=4 \pi \times 10^{-7} \mathrm{H} / \mathrm{m}$ the permeability). The concept of the partial inductance was applied as to calculate parameters $L_{1}, L_{2}$ and $M_{12}$.

## 3 Inductance calculation

3.1 Self-inductance calculation of the meander-type coil with inserted gaps

In order to optimize numerically the useful measurement range of the sensor, the mathematical model that describes the influence of the gap on the self-inductance of the coil shown in Figure 3, was developed.


Figure 3: The stationary coil with inserted gap in conductive segments parallel to $x$-axis, $I_{x}$ segments.

The equivalent circuit of the coil with inserted gaps is shown in Figure 4. The gaps were inserted symmetrically in the segments, hence it was assumed that the current in all bars was equal, and that phase shift did not change.


Figure 4: Equivalent circuit of the stationary coil with inserted gap in segments parallel to $x$-axis, $I_{x}$ segments [13].
$\underline{U}_{I N}=\underline{U}_{R x 11}+\underline{U}_{L x 11}+\underline{U}_{R x 21}+\underline{U}_{L x 21}+\ldots$
$+\underline{U}_{R x N 1}+\underline{U}_{L x N 1}+\underline{U}_{R y 1}+\underline{U}_{L y 1}+$
$+\underline{U}_{R y 2}+\underline{U}_{L y 2}+\ldots+\underline{U}_{R y(N-1)}+\underline{U}_{L y(N-1)}$
Because of the coupling between bars, (2) follows:
$\underline{U}_{R x 11}+\underline{U}_{L x 11}=\left(R_{x 11}+j \omega L_{x 11}-j \omega M_{(11)(21)}+\right.$
$+j \omega M_{(11)(31)}-\ldots . .-j \omega M_{(11)(N 1)}+j \omega M_{(11)(12)}-3.2$
$\left.-j \omega M_{(11)(22)}+\ldots . .-j \omega M_{(11)(N 2)}\right) \cdot \frac{\underline{I}_{0}}{2}$
where $R_{x i 1}$ is the resistivity of the left bar in $I_{x i}$ segment, $L_{x i 1}$ is the partial self-inductance of the bar, $M_{(i 1)(j) 1}$ and $M_{(i 1)(2)}$ are the mutual inductances between bars, $N$ is the number of $I_{x}$ segments ( $N=10$ ). The mutual inductance between bars is positive if the current flows through the bars in the same direction and negative if the current flows through the bars in the opposite direction. Summing voltages in all segments, the input voltage $U_{\text {IN }}$ is

$$
\begin{gathered}
\underline{U}_{I N}=\left(R_{x 11}+j \omega L_{x 11}-j \omega M_{(11)(21)}+j \omega M_{(11)(1))}-\ldots-j \omega M_{(11)(N 1)}+j \omega M_{(11)(12)}-\right. \\
\left.-j \omega M_{(11)(22)}+\ldots-j \omega M_{(11)(N 2)}\right) \cdot \frac{I_{I N}}{2}+\left(R_{x 21}+j \omega L_{x 21}-j \omega M_{(21)(11)}-\right. \\
\left.-j \omega M_{(21)(31)}+\ldots+j \omega M_{(21)(N 1)}-j \omega M_{(21)(12)}+j \omega M_{(21)(22)}-\ldots+j \omega M_{(21)(N 2)}\right) \cdot \frac{I_{I N}}{2}+ \\
\vdots \\
+\left(R_{x N 1}+j \omega L_{x N 1}-j \omega M_{(N 1)(11)}+j \omega M_{(N 1)(21)}-\ldots-j \omega M_{(N 1)(N-1) 1)}-j \omega M_{(N 1)(12)}+3.3\right. \\
\left.+j \omega M_{(N 1)(22)}-\ldots+j \omega M_{(N 1)(N 2)}\right) \cdot \frac{I_{I N}}{2}+\left(R_{y 1}+j \omega L_{y 1}+j \omega M_{(y 1)(y 2)}+\right. \\
\left.+j \omega M_{(y 1)(y 3)}+\ldots+j \omega M_{(y 1)(y(N-1))}\right) \cdot \underline{I}_{I N}+\left(R_{y 2}+j \omega L_{y 2}+j \omega M_{(y 2)(11)}+\right. \\
\left.+j \omega M_{(y 2)(3))}+\ldots+j \omega M_{(y 2)(y(N-1))}\right) \cdot \underline{I}_{I N}+ \\
\vdots \\
+\left(R_{y(N-1)}+j \omega L_{y(N-1)}+j \omega M_{(y(N-1))(1))}+j \omega M_{(y(N-1)(y 2)}+\ldots+j \omega M_{(y(N-1))((N-2))}\right) \cdot \underline{I}_{I N}
\end{gathered}
$$

Finally, it is obtained that the input impedance of the stationary coil with inserted gap (Figure 3) is
$\underline{Z}_{I N}=\frac{\underline{U}_{I N}}{\underline{I}_{I N}}=\frac{1}{2} \underline{Z}_{x}+\underline{Z}_{y,}$
where $Z_{x}$ is the impedance of the segments parallel to $x$-axis ( $I_{x}$ segments) and $Z_{y}$ is the impedance of the segments parallel to $y$-axis ( $l_{y}$ segments).

The impedance $Z_{x}$ is given by
$\underline{Z}_{x}=\sum_{i=1}^{N}\left(R_{x i 1}+j \omega\left(L_{x i 1}+\sum_{\substack{j=1 \\ i \neq j}}^{N}(-1)^{i+j} M_{(i 1)(j 1)}+\right.\right.$
$\left.\left.+\sum_{j=1}^{N}(-1)^{i+j} M_{(i 1)(j 2)}\right)\right)$
where $R_{x i 1}$ is the resistivity of the left bar in $I_{x i}$ segment, $L_{x i 1}$ is the partial self-inductance of the bar, $M_{(i 1)(i) 1}$ and $M_{(i 1)(\text { (2) }}$ are the mutual inductances between bars.

The impedance $Z_{y}$ is given as
$\underline{Z}_{y}=\sum_{i=1}^{N-1}\left(R_{y i}+j \omega\left(L_{y i}+\sum_{\substack{j=1 \\ i \neq j}}^{N-1} M_{(y i)(y j)}\right)\right)$
where $R_{y i}$ is the resistivity of $I_{y}$ segments, $M_{(y i)(y)}$ is the mutual inductance between $I_{y}$ segments. The loop envelops the left bars of $I_{x}$ segments, yet the same result is obtained if the loop envelops the right bars of $I_{x}$ segments.

Further extension of the useful measurement range could be achieved with inserting more narrow gaps in each segment parallel to $x$-axis. However, the number of gaps and their width are the consequence of the chosen geometrical parameters and limitations of the chosen technology for sensor prototypes. The geometrical parameters of meander coils were determined as a compromise between the value of the inductance that could be measured by an electrical interface for signal processing and the size of the meander coils. The mathematical model, which describes influence of two gaps inserted in each segment parallel to $x$-axis, has been presented in Figure 5.


Figure 5: Equivalent circuit of the stationary coil with two inserted gaps in segments parallel to $x$-axis, $I_{x}$ segments.

It was assumed that the current intensity in all bars was equal, and that phase shift did not change. The input voltage $U_{\text {IN }}$ is:

$$
\begin{gathered}
\underline{U}_{I N}=\left(R_{x 11}+j \omega L_{x 11}+j \omega M_{(11)(12)}+j \omega M_{((1)(13)}-j \omega M_{(11)(21)}-j \omega M_{(11)(22)}-j \omega M_{(11)(23)}+\right. \\
+j \omega M_{(11)(31)}+j \omega M_{(11)(32)}+j \omega M_{(11)(33)}-j \omega M_{(11)(41)}-j \omega M_{(11)(42)}-j \omega M_{(11)(43)}+ \\
\vdots \\
\left.-j \omega M_{(11)(N 1)}-j \omega M_{(11)(N 2)}-j \omega M_{(11)(N 3)}\right) \cdot \frac{I_{I N}}{3}+\left(R_{y 1}+j \omega L_{y 1}+j \omega M_{(y 1)(22)}+j \omega M_{(y 1)(y 3)}+3.7\right. \\
+\ldots+j \omega M_{(y 1)(y N-1)}+R_{y 2}+j \omega L_{y 2}+j \omega M_{(y 2)(y 1)}+j \omega M_{(y 2)(y 3)}+\ldots+j \omega M_{(y 1)(y N-1)}+ \\
\vdots \\
\left.\quad+R_{y N-1}+j \omega L_{y N-1}+j \omega M_{(y 1)(y N-1)}++j \omega M_{(y 2)(y N-1)}+\ldots+j \omega M_{(y N-1)(y N-2)}\right) \cdot \underline{I}_{I N}
\end{gathered}
$$

Finally, it is obtained that the input impedance of the stationary coil with two inserted gaps is:
$\underline{Z}_{I N}=\frac{\underline{U}_{I N}}{\underline{I}_{I N}}=\frac{1}{3} \underline{Z}_{x}+\underline{Z}_{y}$,
where $Z_{x}$ is the impedance of the segments parallel to $x$-axes ( $I_{x}$ segments) and $Z_{y}$ is the impedance of the segments parallel to $y$-axes ( $l_{y}^{y}$ segments).

The impedance $Z_{x}$ is given as
$\underline{Z}_{x}=\sum_{i=1}^{N}\left(R_{x i 1}+j \omega L_{x i 1}+j \omega \sum_{\substack{j=1 \\ i \neq j}}^{N}(-1)^{i+j} M_{(i 1)(j 1)}\right.$.
$+j \omega \sum_{j=1}^{N}(-1)^{i+j} M_{(i 1)(j 2)}+j \omega \sum_{j=1}^{N}(-1)^{i+j} M_{(i 1)(j 3))}$
$\underline{Z}_{x}=\sum_{i=1}^{N}\left(R_{x i 1}+j \omega L_{x i 1}+j \omega \sum_{j=1}^{N}(-1)^{i+j} M_{(i 1)(j 1)}+\right.$
$+j \omega \sum_{j=1}^{N} \sum_{k=2}^{3}(-1)^{i+j} M_{(i 1)(j k)}$
where $R_{x i 1}$ is the resistivity of a bar in $I_{x}$ segment, $L_{x i 1}$ is the partial self-inductance of the bar, $M_{(i 1)(j)}$ and $M_{(i 1)(j k)}$ are the mutual inductances between bars.

The impedance $Z_{y}$ is given as
$\underline{Z}_{y}=\sum_{i=1}^{N-1}\left(R_{y i}+j \omega L_{y i}+j \omega \sum_{\substack{j=1 \\ i \neq j}}^{N-1} M_{(y i)(y j)}\right) 3.11$
where $R_{y i}$ is the resistivity of $I_{y}$ segments, $M_{(y)(y)}$ is the mutual inductance between $/$ segments.
3.2 Calculation of the self-inductances of the coils $L_{l}$ and $L_{2}$

The sensor was modeled using the concept of the partial inductance. Planar meander-type coils were partitioned into constituent segments. Each partitioned segment was partitioned additionally into a certain number of filaments [13, 14]. Planar meander coils are partitioned into constituent conductive segments as it is shown in Figure 6. There are 19 conductive segments in meander coils.


Figure 6: Arrows show current flow direction in the coil.
The resistances $R_{1}$ and $R_{2}$ of Coils 1 and 2, respectively, are calculated by Equations 3.12 and 3.13
$R_{1}=\sum_{i=1}^{h} R_{i}$
$R_{2}=\sum_{i=1}^{h} R_{i}$
where $h$ is the number of segments in a coil ( $h=19$ ), and $R_{i}$ is the resistance of a segment (parallel to $x$ - or $y$-axes).

The resistance $R_{i}$ is given by Equation 3.14
$R=\rho_{C u} \cdot \frac{l}{w \cdot t}$
where $\rho_{C u}$ is resistivity of copper, $l$ is the length of the segment, $w$ is the width of the segment and $t$ is the thickness of the segment (copper layer).

The self-inductances of meander Coils 1 and 2 ( $L_{1}$ and $L_{2}$ ) can be calculated as a sum:
$L_{2}=\sum_{i=1}^{h} L_{i} \pm \sum_{i=1}^{h} \sum_{\substack{j=1 \\ j \neq i}}^{h}\left|M_{i j}\right|$
$L_{2}=\sum_{i=1}^{h} L_{i} \pm \sum_{i=1}^{h} \sum_{\substack{j=1 \\ j \neq i}}^{h}\left|M_{i j}\right|$
where $L_{i}$ is the partial self-inductance of each straight segment (Figure 6) and $M_{i j}$ is the mutual inductance between each pair of conductive segments (Figure 6), $h$ is the number of partitioned segments in meander coils ( $h=19$ ). The mutual inductance is positive if current vectors in segments $i$ and $j$ are in the same direction or negative if current vectors are in opposite directions.

As it was reported in [16, 17], each partitioned segment was additionally partitioned into a certain number of elementary filaments ( $\left.I_{1 \times i^{\prime}} I_{1 y_{i}} I_{2 \times j^{\prime}} I_{2 y j^{\prime}} \ldots\right)$, having small, rectangular cross sections, as shown in Figure 7. This was done in order to achieve better precision in calculation, because the segment separation dimensions
are not larger than the cross sectional dimensions for all geometries considered. (As it was previously reported, the distance between axis of two neighboring segments is $p=1.78 \mathrm{~mm}$ whereas, in one of the structures, cross sectional dimension is 1.52 mm .) Figure 7 shows the general case of segments partitioning, with overlapping in the corners. In reality, dimensions of the overlapping in the corners are too small to introduce significant error.


Figure 7: A part of sensor element partitioned into filaments. Filaments parallel to $x$-axis are $I_{x}$ long and $d_{x}$ wide, whereas filaments parallel to $y$-axis are $I_{y}$ long and $d_{y}$ wide; $D$ is the width of conductive segments.

The number of filaments is such that dimensions of the cross section of each filament are less than a skin depth $\left(\delta=\sqrt{\rho_{C u} / \pi f \mu_{0}}\right.$, where $\rho_{C u}=1.72 \times 10^{-8} \Omega$ is electrical resistivity of copper $\mathrm{Cu}, f$ the working frequency and $\mu_{0}=4 \pi \times 10^{-7} \mathrm{H} / \mathrm{m}$ the permeability), at the highest frequency of interest [17]. Each segment was partitioned into 24 filaments as to fulfill this condition and as compromise between complexity and accuracy of the model.

The partial self-inductance $L_{i}$ is the sum of the mutual inductances between all pairs of elementary filaments within segment $i$ :


Figure 8: The partial self-inductance calculation of a conductive segment $i$.
$L_{i}=\sum_{\substack{k=1 \\ k=1 \\ l \neq k}}^{n} M_{k l}$
where $n$ is the number of filaments in segment $i$ and $M_{k l}$ is the mutual partial inductance between filaments $k$ and / within segment $i$, as it is shown in Figure 8.

The mutual inductance $M_{i j}$ (Equations 3.15 and 3.16) is the sum of mutual inductances between all pairs of filaments from segments $i$ and $j$ :

$$
M_{i j}=\frac{1}{m_{1} \cdot m_{2}} \sum_{k=1}^{m_{1}} \sum_{l=1}^{m_{2}} M_{k l}
$$

where $m_{1}$ and $m_{2}$ are the number of filaments in segments $i$ and $j$, respectively, and $M_{k l}$ is the mutual inductance between filaments $k$ and $/$ from segments $i$ and $j$, as it is shown in Figure 9. The numbers of filaments in conductive segments are identical, $n=m_{1}=m_{2}=24$.


Figure 9: The mutual inductance calculation between two segments.

### 3.3 The mutual inductance $M_{12}$ calculation

The mutual inductance between Coils 1 and $2\left(M_{12}\right)$ is calculated in a similar manner, as it is presented in Figure 10.


Figure 10: Mutual inductance $M_{12}$ calculation.
The mutual partial inductance is calculated between each pair of elementary filaments, which belong to dif-
ferent coils. $M_{12}$ is the sum of all mutual partial inductances between filaments from Coils 1 and 2. Taking into account, that Coil 2 physically moves with respect to Coil 1, the distance between filaments of Coils 1 and 2 changes regarding the displacement. The distance between filaments is an important parameter for the mutual inductance calculation, as well their mutual position. While calculating the mutual partial inductances between filaments from different coils different equations were applied [18], depending on the mutual position between filaments, as it can be seen in Figure 11. The formula for mutual inductance of two parallel filaments of equal length $(I)$ and distance $(d)$ is
$M=M(l, d)=\frac{\mu}{2 \pi} l\left[\ln \left(\frac{l}{d}+\sqrt{1+\frac{l^{2}}{d^{2}}}\right)-\right.$
$\left.-\sqrt{1+\frac{d^{2}}{l^{2}}}+\frac{d}{l}\right]$
In the model of the sensor, Coil 2 moves with respect to Coil 1 in $y$-z plane and it rotates around $x$ - and $y$ - axis, as well. In the case of rotation, filaments in Coil 2 can be placed in any desired position. Therefore, equations $3.20-3.22$ [18] were applied to calculate the mutual partial inductance between filaments placed in any desired position, as it is shown in Figure12. Based on this model of the sensor, in-house software was specifically developed for resistance, inductance, and impedance calculation of the sensor. The software calculates variation of these parameters versus displacement in $y-z$ plane and versus small rotations of the moving coil around $x$ - and $y$-axes.

$$
\begin{aligned}
& \frac{M}{0.01 \cos \varepsilon}=2\left[(\mu+1) \cdot \operatorname{arctg} \frac{m}{R_{1}+R_{2}}+(v+m) \cdot \operatorname{arctg} \frac{l}{R_{1}+R_{4}}\right. \\
&\left.-\mu \cdot \operatorname{arctg} \frac{m}{R_{3}+R_{4}}-v \cdot \operatorname{arctg} \frac{l}{R_{2}+R_{3}}\right]-\frac{\Omega d}{\sin \varepsilon}
\end{aligned}
$$

in which

$$
\begin{array}{r}
\Omega=\operatorname{arctg}\left\{\frac{d^{2} \cos \varepsilon+(\mu+l)(v+m) \sin ^{2} \varepsilon}{d R_{1} \sin \varepsilon}\right\} \\
-\operatorname{arctg}\left\{\frac{d^{2} \cos \varepsilon+(\mu+l) v \sin ^{2} \varepsilon}{d R_{2} \sin \varepsilon}\right\} \\
+\operatorname{arctg}\left\{\frac{d^{2} \cos \varepsilon+\mu v \sin ^{2} \varepsilon}{d R_{3} \sin \varepsilon}\right\} \\
-\operatorname{arctg}\left\{\frac{d^{2} \cos \varepsilon+\mu(v+m) \sin ^{2} \varepsilon}{d R_{4} \sin \varepsilon}\right\}
\end{array}
$$



Figure 11: Calculation of the mutual inductance between filaments. Calculation depends on the mutual position between filaments.


Figure 12: Two filaments placed in any desired position.

Parameters $I, m, \mu, v$, and $d$ are given. The relations 3.22 calculate the distances $R_{1}, R_{2^{\prime}} R_{3^{\prime}}$ and $R_{4}$

$$
\begin{align*}
& R_{1}^{2}=d^{2}+(\mu+l)^{2}+(v+m)^{2}-2(\mu+l)(v+m) \cos \varepsilon \\
& R_{2}^{2}=d^{2}+(\mu+l)^{2}+v^{2}-2 v(\mu+l) \cos \varepsilon \\
& R_{3}^{2}=d^{2}+\mu^{2}+v^{2}-2 \mu v \cos \varepsilon \\
& R_{4}^{2}=d^{2}+\mu^{2}+(v+m)^{2}-2 \mu(v+m) \cos \varepsilon
\end{align*}
$$

## 4 Results and Discussion

Simulated values of the input inductance versus $y$ displacement for different gap widths $g=[0.15,0.18$,
$0.20,0.23,0.25,0.28,0.30,0.33,0.36,0.38] \mathrm{mm}$ and the most critical normal distance between coils $z=0.1 \mathrm{~mm}$ are presented in Figure 13. It can be observed in Figure 13 , that invariance of the input inductance versus $y$-displacement near $y=0$ is obtained for gaps $g=0.23 \mathrm{~mm}$ red line, $g=0.25 \mathrm{~mm}$ blue line, and $g=0.28 \mathrm{~mm}$ green line. The input inductance changes versus tangential displacement if the gap width is increased above 0.28 mm . If the gap width is decreased below 0.23 mm , then the useful measurement range is narrower, as it can be seen in Figure 13. The variation of the input inductance versus $y$-displacement for gap widths $g=0.23 \mathrm{~mm}$ (red solid line), $g=0.25 \mathrm{~mm}$ (blue dash line), and $g=0.28 \mathrm{~mm}$ (green dot line) is presented in Figure 14. It can be seen that there is a slight variation of the input inductance in the useful measurement range (near $y=0$ ) for gap widths $g=0.25 \mathrm{~mm}$ and $g=0.28 \mathrm{~mm}$. Thus, it was chosen that the optimal useful measurement range was obtained for gap width $g=0.23 \mathrm{~mm}$. The gap width $g=0.23 \mathrm{~mm}$ provides invariance of the input inductance versus tangential displacement in the useful measurement range, thus making sensor element good for detecting normal displacement.

The sensor element for detecting normal displacement with the gap width $g=0.23 \mathrm{~mm}$ was fabricated, characterized, and compared with the sensor element without the gap. Fabricated prototypes are presented in Figure 15. Sensor prototypes were electrically tested by Impedance Analyzer HP4194A, at the working frequency of 1 MHz . Characterization procedure is similar as it was described in [19].


Figure 13: The sensor element that detects normal displacement: Simulated values of the input inductance variation versus y-displacement for different gap widths, and the most critical normal distance between the coils $z=0.1 \mathrm{~mm}$.

The displacement dependence of the input inductance $L_{I N}$ in $y-z$ plane for the sensor element detecting normal displacement without the gap is presented in Figure 16 and with the gap $g=0.23 \mathrm{~mm}$ in Figure 17. The symmetry and periodicity of the input inductance characteristics can be observed in Figures 16 and 17. The difference between local minimums ( $L_{\text {INmin }}$ ) and maximums


Figure 14: The sensor element that detects normal displacement: Simulated values of the input inductance variation versus $y$-displacement for gap widths $g=0.23 \mathrm{~mm}, g=0.25 \mathrm{~mm}$, and $g=0.28 \mathrm{~mm}$, and the most critical normal distance between the coils $z=0.1 \mathrm{~mm}$.

(a)

(b)

(c)

(d)

Figure 15: The sensor element: a) The stationary coil of sensor element that detects normal displacement without the gap, b) The stationary coil with inserted gap $g=0.23 \mathrm{~mm}, \mathrm{c}$ ) The stationary coil of sensor element that detects tangential displacement, and c) The moving (short-circuited) coil.


Figure 16: The sensor element detecting normal displacement: The displacement dependence of the input inductance $L_{N}$ in $y-z$ plane.
$\left(L_{\text {IN max }}\right)$ decreases as the moving coil moves from the stationary coil in $y$-z plane. Eventually, the tendency of the input inductance characteristic is to achieve the self-inductance of the stationary coil. Figures 16 and 17 present displacement-input inductance dependence in the nearly whole $y$-displacement range. However, the useful measurement range is near $y=0$ for sensor element detecting normal displacement.


Figure 17: The sensor element detecting normal displacement with inserted gap $g=0.23 \mathrm{~mm}$ : The displacement dependence of the input inductance $L_{I N}$ in $y$-z plane.


Figure 18: The sensor element detecting normal displacement without the gap and with the gap $g=0.23 \mathrm{~mm}$ : The displacement dependence of the input inductance $L_{I N}$ in $y-z$ plane.

Comparison of the input inductance characteristic $L_{\text {IN }}$ for the sensor element detecting normal displacement without the gap and with the gap, for the most critical normal distance between the coils $z=0.1 \mathrm{~mm}$, is presented in Figure 18. It can be observed that the invariance of the input inductance versus $y$-displacement is achieved in the useful measurement range (near $y=0$ ), as well near other local minimums. Comparison of the useful measurement range between the sensor elements without the gap and with the gap is presented in Figure 19. Displacement step was approximately 0.0635 mm as to accurately analyze the useful measurement range. It can be seen in Figure 19, that the input inductance of the sensor element with the inserted gap is almost invariant near $y=0$ in comparison with the input inductance of the sensor element without the gap.

The smallest change in the position of a moving coil that can be detected depends on the signal processing interface. In case of measuring with Impedance analyzer, which can detect transition in the range of 10 micro ohms, resolution of the sensor can be estimated to 0.1 $\mu \mathrm{m}$. Further, resolution of the sensor is possible to improve with adjusting the design to meet the given applications. The useful measurement range of the sensor element detecting normal displacement, the worst-


Figure 19: Comparison of the useful measurement ranges for the sensor element without the gap and with the gap $g=0.23 \mathrm{~mm}$ when displacement step was 0.0635 mm and for the most critical normal distance between coils $z=0.1 \mathrm{~mm}$.
case results for $g=0$, is nearly 0.31 mm for relative accuracy $\pm 0.5 \%$, and nearly 0.42 mm for relative accuracy $\pm 1 \%$. From Figure 19, it could be observed that these ranges would be wider for the optimal useful measurement range $\mathrm{g}=0.23 \mathrm{~mm}$ and for given accuracies.

## 5 Conclusion

In this paper, concept of the partial inductance was used as to model a planar displacement sensor with inductive coils of meander-type. In addition, modeling of a gap, inserted in a meander coil, was presented as well. In-house software was developed, based on this model, and was used to determine numerically the optimal useful measurement range of the sensor. Results show that for the sensor with specific geometrical parameters, as it is given in the paper, the optimal useful measurement range is obtained if the gap of 0.23 mm is inserted in the stationary coil. Theoretical predictions were confirmed with experimental results.

## 6 Acknowledgment

This work was supported by the Ministry of Education, Science, and Technological Development, Serbia, under Grant TR32016 and Grant III45021.

## 7 References

1. Yamada, S., Nakamura, K., Iwahara, M., Taniguchi, T., and Wakiwaka, H., "Application of ECT technique for inspection of bare PCB", IEEE Transactions on Magnetics, Vol. 39, No. 5, pp. 3325-3327, 2003.
2. Chomsuwan, K., Yamada, S., and Iwahara, M., "Improvement of defect detection performance of PCB
inspection based on ECT technique with multi-SVGMR sensor", Vol. 43, No. 6, pp. 2394-2396, 2007.
3. Bayani, H., Nishino, M., Yamada, S., and Iwahara, M.,"Introduction of a base model for eddy-current testing of printed circuit boards", IEEE Transactions on Magnetics, Vol. 44, No. 11, pp. 4015-4017, 2008.
4. Baschirotto, A., Dallago, E., Malcovati, P., Marchesi, M., and Venchi, G., "Development and comparative analysis of fluxgate magnetic sensor structures in PCB technology", IEEE Transactions on Magnetics, Vol. 42, No. 6, pp. 1670-1680, 2006.
5. Cha, Y.-J., Nam, B., Kim, J., and Kim, K. H., "Evaluation of the planar inductive magnetic field sensors for metallic crack detections", Sensors and Actuators A: Physical, Vol. 162, No. 1, pp. 13-19, 2010.
6. Norhisam, M., Norrimah, A., Wagiran, R., Sidek, R. M., Mariun, N., and Wakiwaka, H.,
7. "Consideration of theoretical equation for output voltage of linear displacement sensor using meander coil and pattern guide", Sensors and Actuators A: Physical, vol. 147, No. 2, pp. 470-473, 2008.
8. Misron, N., Ying, L. Q., Firdaus, R. N., Abdullah, N., Mailah, N. F., and Wakiwaka, H., "Effect of inductive coil shape on sensing performance of linear displacement sensor using thin inductive coil and pattern guide", Sensors, Vol. 11, No. 11, pp. 1052210533, 2011.
9. Wang, H., and Feng, Z., "Ultrastable and highly sensitive eddy current displacement sensor using self-temperature compensation", Sensors and Actuators A: Physical, Vol. 203, pp. 362-368, 2013.
10. Jerance, N., Bednar, N., and Stojanovic, G., "An inkjet eddy current position sensor", Sensors, Vol. 13, No. 4, pp. 5205-5219, 2013.
11. Laskoski, G. T., Pichorim, S. F., and Abatti, P. J., "Distance measurement with inductive coils", IEEE Sensors Journal, Vol. 12, No. 6, pp. 2237-2242, 2012.
12. Yunus, M. A. Md., and Mukhopadhyay S. C., "Novel planar electromagnetic sensors for detection of nitrates and contamination in natural water sources", IEEE Sensors Journal, Vol. 11, No. 6, pp. 1440-1447, 2011.
13. Yunus, M. A. Md., and Mukhopadhyay, S. C., "Development of planar electromagnetic sensors for measurement and monitoring of environmental parameters", Measurement Science and Technology, Vol. 22, No. 2, 025107 (9pp), 2011.
14. Damnjanovic, M. S., Zivanov, Lj. D., Nagy, L. F., Djuric, S. M., and Biberdzic, B. N., "A novel approach to extending the linearity range of displacement inductive sensor", IEEE Transactions on Magnetics, Vol. 44, No. 11, pp. 4123-4126, 2008.
15. Djuric, S. M., Nagy, L. F., Damnjanovic, M. S., Djuric, N. M., and Zivanov, Lj. D., "A novel application
of planar-type meander sensors", Microelectronics International, Vol. 28, No. 1, pp. 41-49, 2011.
16. Wakiwaka, H., Nishizawa, H., Yanase, S., Maehara, O., "Analysis of impedance characteristics of meander coil", IEEE Transactions on Magnetics, Vol. 32, No. 5, pp. 4332-4334, 1996.
17. Ruehli, A. E., "Inductance calculations in a complex integrated circuit environment," IBM Journal of Research and Development, Vol. 16, No. 5, pp. 470-481, 1972.
18. Ruehli A., Paul C., and Garett, J.,"Inductance calculations using partial inductances and macromodels" Proceedings of the International Symposium on EMC, Atlanta, USA, pp.23-27, 1995.
19. Grover, F. W., Inductance calculation, D. Van Nostrand Company, New York, 1946.
20. Djuric, S. M., "Performance analysis of a planar displacement sensor with inductive spiral coils", IEEE Transactions on Magnetics, Vol. 50, No. 4, 4004104 (4pp), 2014.

Arrived: 04.01. 2015
Accepted: 06. 03. 2015

# Zynq-based System for Extracting Sorted Subsets from Large Data Sets 

V. Sklyarov ${ }^{1}$, I. Skliarova ${ }^{1}$, A. Rjabov ${ }^{2}$, A. Sudnitson ${ }^{2}$<br>${ }^{1}$ University of Aveiro / IEETA, Campus Universitário de Santiago, Aveiro, Portugal<br>${ }^{2}$ Tallinn University of Technology, Tallinn, Estonia


#### Abstract

The paper describes hardware/software architecture of a system for extracting the maximum and minimum sorted subsets from large data sets, two methods that enable high-level parallelism to be achieved, and implementation of the system in recently appeared on the market Zynq-7000 microchips incorporating a high-performance processing unit and advanced programmable logic from the Xilinx 7th family. The methods are based on highly parallel and easily scalable sorting networks and the proposed technique enabling sorted subsets to be extracted incrementally with very high speed that is close to the speed of data transfer through highperformance interfaces. The results of implementations and experiments clearly demonstrate significant speed-up of the developed software/hardware system comparing to alternative software implementations.


Keywords: processing system; programmable logic; system-on-chip; sorting networks; hardware/software co-design

# Sistem na osnovi Zynq za izluščitev razvrščenih podsklopov iz obsežnih podatkovnih sklopov 


#### Abstract

Izvleček: Članek predstavlja programsko/strojno zasnovo sistema za izluščitev največjih in najmanjših razvrščenih podsklopov v obsežnih podatkovnih sklopih. Predstavljeni sta dve metodi, ki omogočata visoko stopnjo vzporednosti in implementacijo sistema v tržnem ZYNG-7000 mikročipu na osnovi programabilne logike Xilinx sedme generacije. Metode temeljijo na vzporedni in enostavno razširljivih omrežjih ter omogočajo izluščitev podsklopov s hitrostjo blizu hitrosti prenosa podatkov. Rezultati dokazujejo veliko pohitrenje programsko/strojnih rešitev v primerjavi s programskimi rešitvami.


Ključne besede: processing system; programmable logic; system-on-chip; sorting networks; hardware/software co-design

[^3]
## 1 Introduction

All Programmable Systems-on-Chip (APSoC) from Zynq-7000 family [1,2] combine on the same microchip the dual-core ARM® Cortex ${ }^{\top M}$ MPCore ${ }^{\text {TM }}$-based highperformance processing system (PS) with advanced programmable logic (PL) from the Xilinx $7^{\text {th }}$ family and may be used effectively for the design of hardware accelerators in such areas as hard real-time systems [3], image [4] and data [5] processing, satellite on-board processing [6], programmable logic controllers [7], driver assistance applications [8], wireless networks [9], and many others [2]. Interactions between the PS and PL are supported by different interfaces and other signals through over 3,000 connections [1]. Available four 32/64-bit high-performance (HP) Advanced eXtensible Interfaces (AXI) and a 64-bit AXI Accelerator Coherency

Port (ACP) enable fast data exchange with theoretical bandwidths shown in [1].

Zynq APSoC design flow includes the development of hardware in the PL [10] (supported by available Xilinx IP cores) and software in the PS [11] for different types of applications such as standalone (bare metal) [12], running under an operating system (e.g. Linux) [12] and combined [13]. Hardware implemented in the PL can be the same for standalone and Linux applications but software programs use different functions and interaction mechanisms [12]. Since bare metal projects are generally faster, we will consider them as a base which does not exclude using the results for projects running under operating systems. The latter may benefit from available drivers and other support [12]. Since both
types of projects can run in parallel in different cores [13] they may be combined if required.

Many electronic, environmental, medical, and biological applications need to process data streams produced by sensors and measure external parameters within given upper and lower bounds (thresholds) [14]. Let us consider some examples. Applying the technique [15] in real-time applications requires knowledge acquisition obtained from controlled systems (e.g. plant). For example, signals from sensors may be filtered and analysed to prevent error conditions (see [15] for additional details). To provide more exact and reliable conclusion a combination of different values need to be extracted, ordered, and analysed. Similar tasks appear in monitoring thermal radiation from volcanic products [16], filtering and integration of information from a variety of different sources in medical applications [17] and so on. Since many systems are hard real-time, performance is important and hardware accelerators may provide significant assistance for software products. Similar problems appear in so-called straight selection sorting (in such applications where we need to find a task with the shortest deadline in scheduling algorithms [18]), in statistical data manipulation and data mining (e.g. [19-22]). To describe one of the problems from data mining informally let us consider an example [19] with analogy to a shopping card. A basket is the set of items purchased at one time. A frequent item is an item that often occurs in a database. A frequent set of items often occur together in the same basket. A researcher can request a particular support value and find the items which occur together in a basket either a maximum or a minimum number of times within the database [19]. Similar problems appear to determine frequent inquiries at the Internet, customer transactions, credit card purchases, etc. requiring processing very large volumes of data in the span of a day [19]. Fast extracting the most frequent or the less frequent items from large sets permits data mining algorithms to be simplified and accelerated. Sorting of subsets may be involved in many known methods from this area [e.g. 20-22].

Let us consider a system that collects data produced by some measurements or copies such data from a database. A valuable assistance for applications described above may be provided by fast extraction of the maximum and minimum sorted subsets from the set of collected data, where the maximum/minimum sorted subset contains $L_{\text {max }} / L_{\text {min }}$ data items. This problem can be solved in a software only system. For example, C function qsort permits large data sets to be sorted. After sorting is completed, extracting the maximum and minimum subsets may easily be done collecting them from the top and from the bottom of the sorted set. However, for many practical applications, such as that
are referenced in $[18,19]$, performance of the described above operations is important and software functions need to be accelerated. The paper suggests methods and high-performance implementations for solving the indicated above problem in APSoC from the Xilinx Zynq-7000 family.

The remainder of the paper is organized in five sections. Section 2 presents the proposed system architecture and describes overall functionality. Section 3 suggests two novel methods allowing the maximum and minimum sorted subsets to be extracted from large data sets. Section 4 shows how large subsets (for which hardware resources are not sufficient) can be computed and discusses additional capabilities. Implementation in Zynq microchip and the results of thorough evaluation and comparison of software only and software/hardware solutions with explicit indication of the achievable accelerations are discussed in section 5. Section 6 concludes the paper.

## 2 System Architecture and Functionality

The known results [2,5,12] have shown that software/ hardware solutions may be significantly faster than software only solutions. Let us look at Fig. 1. Clearly, software/hardware system is faster if: $\mathrm{T}_{\mathrm{s}}>\mathrm{T}_{\mathrm{sch}} \leq \mathrm{T}_{\mathrm{sh}}+\mathrm{T}_{\mathrm{h}}$ $+\mathrm{T}_{c^{\prime}}$ where $\mathrm{T}_{s^{\prime}} \mathrm{T}_{\text {sch }}, \mathrm{T}_{\text {sh }} \mathrm{T}_{\mathrm{c}^{\prime}} \mathrm{T}_{\mathrm{h}}$ are time intervals required for different modules. In highly parallel implementations software, hardware and interactions between hardware and software can run concurrently. For example, software may run in parallel with hardware; operations in hardware over previously received data may be done at the same time when new data are being transferred. Thus, $T_{\text {sch }} \leq T_{\text {sh }}+T_{h}+T_{c}$. This paper evaluates and compares software/hardware and software only solutions taking into account all the involved communication overheads and paying special attention to high level of parallelism. For instance we would like communication and application-specific operations to be overlapped in hardware as much as possible (see Fig. 1). Note that while hardware only designs may be the fastest, the complexity of such designs is often limited by the available resources in the PL.


Figure 1: Software only and software/hardware systems

Fig． 2 presents the proposed software／hardware archi－ tecture．Extracting subsets is done in an application－ specific processing block（ASP）which is entirely imple－ mented in the PL．We will discuss the ASP in the next section with all necessary details．There is another block in the PL called communication－specific processing （CSP）which interacts with the PS，i．e．it receives a large set of data items step by step in blocks and transfers the extracted sorted subsets．Besides，CSP is responsible for exchange of control signals between the PS and PL．

The PS is responsible for solving the following tasks：
1．Acquiring data and saving them in either on－chip memory（OCM）or external memory that is DDR．
2．Forming requests to extract subsets in the PL which is done through a set of control signals．
3．Collecting extracted subsets and storing them in OCM or external memory．
4．Verifying the results．
5．Solving exactly the same problem in software．This point is required just for experiments and comparison．
6．Computing the consumed time．
The PL is responsible for solving the following tasks：
1．Processing control signals received from the PS which are：a request（start）to begin data process－ ing；source address in memory of input data（i．e． the address of the set that has to be handled）；desti－


Figure 2：The proposed software／hardware architecture
nation address in memory of output data（i．e．the address to copy the extracted subsets）；the number of blocks Q of input data transferred from the PS to PL；and the number of items in the last block $\mathrm{K}_{\text {last }}$ ． The PL also forms two signals that are sent to the PS which are：an interrupt generated as soon as the job is completed（i．e．the subsets have been extracted and copied to memory）and the num－ ber of clock cycles consumed in the PL which is needed for experiments and comparisons．
2．Extracting subsets on requests from the PS in highly－parallel ASP．
3．Counting clock cycles consumed in the PL from re－ ceiving the request up to generating the interrupt．

| ■－非 processing＿system7＿0 |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| － －Data（32 address bits ：4G） |  |  |  |  |  |
| －maxi＿cdma＿0 | S＿AXI＿LITE | Reg | 0x4E200000 | 64K | －0x4E20FFFF |
| －．．anaxi＿cdma＿1 | S＿AXI＿LITE | Reg | 0x4E210000 | 64K | －0x4E21FFFF |
| －maxi＿cdma＿2 | S＿AXI＿LITE | Reg GPP mapping | 0x4E220000 | 64K | －0x4E22FFFF |
| －$=$ axi＿cdma＿3 | S＿AXI＿LITE | Reg GPP mapping | 0x4E230000 | 64K | －0x4E23FFFF |
| －${ }^{\text {a }}$ axi＿cdma＿4 | S＿AXI＿LITE | Reg | 0x4E240000 | 64K | －0x4E24FFFF |
| －axi＿bram＿ctrl＿0 | S＿AXI | Mem0 | 0x40000000 | 64K 0x4000FFFF |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| －m processing＿system7＿0 | S＿AXI＿HPO | HP0＿DDR＿LOWOCM | 0x00000000 | 512M | －0x1ffrffrF |
| －maxi＿bram＿ctrl＿1 | S＿AXI | Mem0 | 0xC0000000 | 64K | －0xC000FFFF |
| －랴 axi＿cdma＿1 |  | $\text { Mapping of HP AXI port } 1$ |  |  |  |
| Data（ 32 address bits ： 4 G ） |  |  |  |  |  |
| －$=$ processing＿system7＿0 | S＿AXI＿HP1 | HP1＿DDR＿LOWOCM | 0x00000000 | 512M | －0x1FFFFFFF |
| －．．．axaxi＿bram＿ctrl＿2 | S＿AXI | Mem0 | 0xC0000000 | 64K－0xC000FFFF |  |
|  |  |  | of HP A | port 2 |  |
|  |  |  |  |  |  |
| －me processing＿system7＿0 | S＿AXI＿HP2 | HP2＿DDR＿LOWOCM | 0x00000000 | 512M | －0x1FFFFFFF |
| －．．．asaxi＿bram＿ctrl＿3 | S＿AXI | Mem0 | 0xC0000000 | 64K－0xC000FFFF |  |
| （－非 axi＿cdma＿3 ${ }^{\text {E－}}$－Data（ 32 address bits ：4G）Mapping of HP AXI port 3 |  | $\text { Mapping of HPAXI port } 3$ |  |  |  |
|  |  |  |  |  |  |  |  |
| －m processing＿system7＿0 | S＿AXI＿HP3 | HP3＿DDR＿LOWOCM | 0x00000000 | 512M | －0x1FFFFFFF |
| －axa axi＿bram＿ctrl＿4 | S＿AXI | Mem0 | 0xC0000000 | 64K | －0xC000FFFF |
| ®－非 axi＿cdma＿4 ${ }^{\text {a }}$（ Mapping of HP A |  |  |  |  |  |
| Đ－囲 Data（32 address bits ：4G） |  |  |  |  |  |
| －me processing＿system7＿0 | S＿AXI＿ACP | ACP＿DDR＿LOWOCM | 0x00000000 | 512M | －0x1FFFFFFF |
| －ma processing＿system7＿0 | S＿AXI＿ACP | ACP＿QSPI＿LINEAR | $0 \times 5 C 000000$ | 16M | －OxFCFFFFFF |
| －mor pressing＿system7＿0 | S＿AXI＿ACP | ACP＿IOP | 0xE0000000 | 4M | －0xE03FFFFF |
| maxi＿bram＿ctrl＿5 | S＿AXI | Mem0 | 0xC0000000 | 64K | －0xC000FFFF |
| U－Unmapped Slaves（1） |  |  |  |  |  |
| ．．．．．m processing＿system7＿0 | S＿AXI＿ACP | ACP＿M＿AXI＿GP0 |  |  |  |

Figure 3：Address mapping from Vivado 2014.2 block design editor

Note that for experiments and comparisons some additional signals for interactions between the PS and PL may be needed.

There are some generic parameters for which hardware in the PL is statically configured (see Fig. 2). They are:

K - the number of items that are handled in hardware in each block ( $\mathrm{K}_{\text {last }} \leq \mathrm{K}$ );
M - the size of each data item;
$L_{\text {max }}$ - the number of items in the maximum subset; $L_{\text {min }}$ - the number of items in the minimum subset.

Selection of proper AXI ports is very important. Experiments in [23] have shown that for transferring a small number of data items (from 16 to 64 bytes) generalpurpose input/output ports (GPP) are always the best. In Zynq APSoC there are four available 32-bit GPP, two of which are masters and the other two are slaves from the side of the PS. They are optimized for access from the PL to the PS peripherals and from the PS to the PL registers/memories [24]. Since the latter feature is what we need, a master GPP was chosen for transferring control signals shown in Fig. 2. AXI ACP allows cache memory of application processing unit (APU) in the PS to be involved for data transfers and there exists an opportunity to provide either cacheable or non-cacheable data from/to the indicated above memories (i.e. OCM or DDR) [23]. Mapping of memories may be done in computer-aided design software (in our case in Xilinx Vivado block design editor according to addresses given in [1] and shown in Fig. 3, and in Xilinx Software Development Kit - SDK). Experiments in [12,23] have shown that for transferring large volumes of data items AXI ACP is very appropriate. Thus, this port was chosen to receive the source set from memory (OCM or DDR) in the PL and to copy extracted subsets from the PL to memory.

Fig. 4 gives more details about the chosen software/ hardware interactions where: solid arrows $(\rightarrow)$ indicate who is the master (the beginning) and who is the slave (the end); triple compound lines show control flow; and dashed lines indicate directions of data flow (i.e. one direction $-\rightarrow$ or both directions $-\leftrightarrow$ ). Control (and possibly a small number of additional auxiliary) signals are transferred through GPP. An initial (source) set and extracted subsets are copied through AXI ACP. The used memory (OCM or DDR) is indicated by the respective mapping both in hardware (see Fig. 3) and in software, which in our case was described in C language, and the mapping is done like the following:
\#define OCM_ADDRESS
$0 \times 00000000$
\#define DDR_ADDRESS
0x16D84000
0x40000000
OCM_ADDRESS

Note that additional details about mapping with many examples can be found in [12].

The snoop controller [1] in Fig. 4 provides cacheable and non-cacheable access to memories (OCM or DDR) [1]. Cache area can be either disabled or enabled in software with the aid of function Xil_SetTlbAttributes [25]. In particular data received from/copied to memories may be pre-cached, i.e. they can be first saved into faster cache and then transferred with the main goal to increase performance of communications. Note that for standalone programs cache memory is entirely available. For programs running under an operating system (such as Linux) some area in cache memory may be used by programs of the operating system and the size of available cache memory is reduced. Many additional details can be found in [12].


Figure 4: Hardware/software interactions
Initial (source) data set and extracted subsets are accommodated in memory as it is shown in Fig. 5. All necessary details about particular locations and sizes are supplied from the PS to PL through GPP (see Fig. 2).

To extract the maximum and/or minimum sorted subsets the following sequence of operations is executed:

1. The PS prepares source data in memory, calculates the number of blocks $\mathrm{Q}=\lceil\mathrm{N} / \mathrm{K}\rceil$ (the value K is predefined), the number of items in the last block (which can be less than K), and indicates source and destination addresses. Here, N is the total number of data items that have to be processed.

[^4]2. The PS sets the start signal that is permanently tested in the PL.
3. As soon as the signal start is set, the PL transfers blocks of data in burst mode and saves them in a dedicated dual-port embedded block RAM (one port is assigned for transferring data from the PS to PL and another port for copying data from the block RAM to PL registers considered in the next section).


Figure 5: Accommodation of the initial data set and the extracted subsets in memory
4. As soon as the first block is completely transferred to the block RAM through the first port, it is copied through the second port to PL registers that are used as inputs of sorting networks for extracting subsets in ASP.
5. The maximum and minimum subsets are incrementally constructed using methods from the next section and subsequent blocks of source data are transferred from memory to the block RAM in parallel.
6. The block RAM is organized as a circular buffer as it is shown in Fig. 6. If it becomes full data transfer is suspended until space for subsequent block is freed.
7. As soon as all Q blocks are processed the maximum and minimum subsets are ready (the details will be given in the next section).
8. The maximum and minimum subsets are copied from the PL to memory (see Fig. 5).
9. As soon as the previous point is completed, the PL generates a hardware interrupt to the PS indicating that the job has been finished (the details about such interrupts with examples can be found in [12]).
10. Optionally, the PL may count the number of clock cycles for solving the problem in hardware that it supplied to the PS through GPP.
11. PS may solve other problems in parallel with the PL. However, as soon as the interrupt is generated it is handled by the PS. Hence, the extracted subsets may immediately be used, for example, as data needed for projects of higher hierarchical levels.


Figure 6: Block RAM organized as a circular buffer
The circular buffer in Fig. 6 is managed by the PL control unit (see Fig. 4) that is a finite state machine. The buffer is built in the PL block RAM which is written through the first port (used for transfer data from the PS) and read through the second port (used to copy data from the block RAM to PL registers). As soon as the buffer is full, data transfer from the PS to PL is suspended. As soon as some area of the buffer is released (because data have already been read) data transfer is renewed.

## 3 Methods for Extracting Sorted Subsets

Let set $S$ containing N M -bit data items be given. The maximum subset contains $L_{\text {max }}$ largest items in $S$ and the minimum subset contains $L_{\text {min }}$ smallest items in $S$ $\left(\mathrm{L}_{\text {max }} \leq \mathrm{N}\right.$ and $\mathrm{L}_{\text {min }} \leq \mathrm{N}$ ). We mainly consider such tasks for which $\mathrm{L}_{\text {max }} \ll \mathrm{N}$ and $\mathrm{L}_{\text {min }} \ll \mathrm{N}$ which are more common for practical applications. Large and very large subsets may also be extracted and section 4 explains how to compute them. Experiments with such subsets are also reported in section 5 . Sorting will be done in highly parallel networks, such as [26] or [27]. Since N may have very large value (millions of items) it cannot completely be processed in hardware due to unavailability of sufficient resources.

We suggest solving the problem iteratively using hardware architecture of ASP shown in Fig. 7. Data are incrementally received in blocks containing exactly K items and then processed by parallel networks described below. We mentioned above that the last block may contain less than K items. If so, it will be extended up to K items (we will talk about such extension a bit later). Part of sorted items with maximum values will be used to form the maximum subset and part of sorted items with minimum values will be used to form the minimum subset. As soon as all Q blocks have been handled the maximum and/or minimum subsets will be ready to be transferred to the PS.

We suggest two methods enabling the maximum and minimum sorted subsets to be incrementally constructed. The first method is illustrated in Fig. 8.


Figure 7: Basic hardware architecture for ASP


Figure 8: The first method of extracting the maximum and minimum sorted subsets

Sorting networks $\mathrm{SN}_{\text {min }}$ and $\mathrm{SN}_{\text {max }}$ have input registers. The minimum and maximum sorted subsets will be built incrementally in halves of registers indicated at the bottom part of Fig. 8. At initialization step, these parts are pre-loaded with possible maximum and minimum values which data from the source set may have. Such values can be indicated by the PS in additional fields through GPP or calculated in the PL. Then the following steps are executed:

1. The first block containing K M -bit data items is copied from block RAM and becomes available at the inputs of the main SN .
2. The block is sorted in parallel in the main SN which can be done in combinational networks from [26] (such as even-odd merger) or in sequential iterative networks from [27] (such as iterative even-odd transition network). In the last case additional control is provided.
3. $L_{\text {max }}$ sorted items with maximum values are loaded in a half of the $\mathrm{SN}_{\text {max }}$ input register as it is shown in Fig. 8. $\mathrm{L}_{\text {min }}$ sorted items with minimum values are loaded in a half of the $\mathrm{SN}_{\text {min }}$ input register as it is shown in Fig. 8. All the items are resorted by the relevant sorting networks $\mathrm{SN}_{\text {max }}$ and $\mathrm{SN}_{\text {min }}$.
4. A new block is copied from block RAM and becomes available at the inputs of the main SN . Such operations are repeated until all Q-1 blocks are handled.
5. The last block may contain less than $K$ items and it is processed slightly differently. As soon as all Q blocks have been transferred from the PS to the PL block RAM and Q-1 blocks have been handled in ASP, the last block (if it is incomplete) is extended to K items by copying the largest item from the created minimum sorted subset. Thus, the last block becomes complete. Clearly, largest item from the created minimum sorted subset cannot be moved again to the minimum subset and the last block is handled similarly to the previous blocks.

Let as look at an example in Fig. 9.


Figure 9: Example of extracting sorted subsets using the first method

It is assumed that the minimum possible value of data items is 0 and the maximum possible value is 99 (clearly, other values may also be chosen). At the first step (a), shown in left-hand part of Fig. 9, input registers for $\mathrm{SN}_{\text {max }}$ and $\mathrm{SN}_{\text {min }}$ are initialized, and the first block of data becomes available for the main SN. U indicates undefined values. At the next step (b) input registers are updated as it is shown by dashed fragments in Fig. 9. At step (c) a new block of data becomes available. Note that loading the register for the main SN can be done in parallel with copying $\mathrm{L}_{\max } / \mathrm{L}_{\min }$ to $\mathrm{SN}_{\max } / \mathrm{SN}_{\min }$. Items in $\mathrm{SN}_{\text {max }}$ and $\mathrm{SN}_{\text {min }}$ are sorted as soon as the relevant input registers are updated. After executing steps (a) - (g) the maximum and minimum sorted subsets are ready (see the right-hand part of Fig. 9) for the items shown in grey in the main SN. Clearly, this method enables the maximum and minimum sorted subsets to be incrementally constructed for very large sets.

The idea of the second method is illustrated in Fig. 10 on the same example from Fig. 9.


Figure 10: Example of extracting sorted subsets using the second method

Now the size of the networks $\mathrm{SN}_{\text {max }}$ and $\mathrm{SN}_{\text {min }}$ was reduced twice (there are now just 4 M -bit inputs instead of 8 in Fig. 9). Much like Fig. 8 both these networks have input registers ( 4 M -bit registers for our example). At initialization step $\mathrm{SN}_{\text {max }}$ and $\mathrm{SN}_{\text {min }}$ are filled in with the minimum and maximum values which are assumed as before to be 0 and 99. There are two additional fragments in Fig. 10 which contain circuits from [28]. They are composed of comparators shown in Knuth notation [29]. Any comparator converts a two-item input to the two-item output in such a way that the upper value is greater than or equal to the lower value. Let us call circuits from [28] a swapping network. If they are applied to two sorted subsets with equal sizes then it is guaranteed that the upper half outputs of the network con-
tain the largest values from two sorted subsets and the lower half outputs of the network contain the smallest values from two sorted subsets. If we resort separately the upper and the lower parts then two sorted subsets will form a single sorted set. Let us analyse the upper swapping network in Fig. 10. At step (a) inputs of the network are sorted subsets $\{0,0,0,0\}$ and $\{99,92,71,70\}$. Thus, two new subsets $\{70,71,92,99\}$ and $\{0,0,0,0\}$ are created. Sorting them enables the maximum sorted subset $\{99,92,71,70\}$ with four items to be found on outputs of $\mathrm{SN}_{\text {max }}$. At step (c) inputs of the swapping network are sorted subsets $\{99,92,71,70\}$ and $\{98,80,71,69\}$ and two new subsets $\{99,92,80,98\}$ and $\{70,71,71,69\}$ are created. Sorting them enables the maximum sorted subset $\{99,98,92,80\}$ to be built. At step (e) inputs of the swapping network are sorted subsets $\{99,98,92,80\}$ and $\{20,19,18,17\}$ and no swapping is done. Hence, the maximum sorted subset is $\{99,98,92,80\}$ and it is the same as in Fig. 9. The lower swapping network in Fig. 10 functions similarly.

The second method involves an additional delay on the comparators of swapping networks but eliminates copying (through feedbacks in Fig. 8) from the main SN to $\mathrm{SN}_{\text {max }}$ and $\mathrm{SN}_{\text {min }}$. Besides, the sizes of $\mathrm{SN}_{\text {max }}$ and $\mathrm{SN}_{\text {min }}$ are reduced twice.

Let us discuss now an attainable complexity of sorting networks in the PL. It is shown in $[5,27]$ that even in relatively complex field-programmable gate arrays (FPGAs) the size K is limited. For example, for even-odd merge and bitonic merge networks [26] K cannot exceed a few hundreds of 32-bit items even for very advanced FPGAs (such as the largest devices from the Xilinx Vir-tex-7 family [30]). In Zynq devices and circuits from [31] the maximum value of K cannot exceed 100 of 32 -bit items. Iterative even-odd transition networks from [27] permit significantly larger number of items (exceeding thousands of 32 -bit items) to be processed and they may efficiently be used for computing sorted subsets in hardware. Fig. 11 gives an example of the network from [27] which permits up to $K=16$ data items to be sorted.

K M-bit data items that have to be sorted are loaded (from block RAM) to the feedback register (FR). Sorting is executed in a segment of even-odd transition network composed of two linked lines with even and odd comparators. Sorting is completed in K/2 iterations (clock cycles) at most. Note, that almost always the number of iterations is less than $K / 2$ because of the technique [27] according to which if there is no swaps of data on the right-most line of the comparators then sorting is completed. Note that the network [27] possesses significantly smaller combinational delays than networks from [26]. Besides, in the proposed architec-


Figure 11: An example of iterative sorting network from [27] for $K=16$ data items
ture (see Fig. 4) iterations are done at the same time as subsequent data are being received from the PS. Such parallelism enables delays to be optimally adjusted allowing the total performance to be improved.

## 4 Computing Large Subsets and Additional Capabilities

For some practical applications the maximum and minimum subsets may be large and the available hardware resources become insufficient to implement sorting networks. Indeed, in accordance with [12] the largest sorting network that can be implemented in Zynq microchip xc7z020-1clg484c (that will further be used for experiments) is 512 32-bit items. The arising problem can be solved using the following technique. Let $I_{\text {max }}$ and $I_{\text {min }}$ be constraints for the upper $\left(\mathrm{SN}_{\text {max }}\right)$ and bottom ( $\mathrm{SN}_{\text {min }}$ ) parts in Fig. 7, i.e. the circuits $\mathrm{SN}_{\text {max }}$ and $\mathrm{SN}_{\text {min }}$ with larger values (than $\mathrm{I}_{\text {max }}$ and $\mathrm{I}_{\text {min }}$ ) cannot be implemented due to the lack of hardware resources or because of some other reasons. Let the parameters for the maximum and minimum subsets be greater than $I_{\max }$ and $I_{\text {min }}$ i.e. $L_{\text {max }}>I_{\max }$ and $L_{\text {min }}>I_{\text {min }}$. In such case the maximum and minimum subsets can be computed iteratively as follows:

1. At the first iteration, the maximum subset containing $I_{\text {max }}$ items and the minimum subset containing $I_{\text {min }}$ items are computed. The subsets are transferred to the PS (to memories). The PS removes the minimum value from the maximum subset and the maximum value from the minimum subset. Such correction avoids loss of repeated items at subsequent steps. Indeed, the minimum value from the maximum subset (the maximum value from the minimum subset) can appear for subsets to be subsequently constructed in point 3 below and they will be lost because of filtering (see point 3).
2. The minimum value from the corrected in the PS maximum subset is assigned to $B_{u}$. The maximum value from the corrected in the PS minimum subset is assigned to $B_{1}$. The values $B_{u}$ and $B_{1}$ are supplied to the PL through GPP.
3. The same data items (from memory), as in point 1 above, are preliminary filtered in the PL in such a way that only items that are less or equal than $B_{u}$ and greater or equal than $B_{1}$ are allowed to be transferred to block RAM, i.e. computing sorted subsets is done only for the filtered data items. Thus, the second part of the maximum and the minimum subsets will be computed and appended (in the PS) to the previously computed subsets (such as subsets from point 1).
4. The points 2 and 3 above are repeated until the maximum subset with $L_{\text {max }}$ items and the minimum subset with $L_{\text {min }}$ items are computed.

Note, that if the number of repeated items is greater than or equal to $I_{\max } / I_{\text {min }}$, then the method above may generate infinite loops. This situation can easily be recognized. Indeed, if any new subset (that is sent from the PL to the PS) contains the same value repeated K times then an infinite loop will be created. In such case we can use another method based on software/hardware sorters from [12]. In the next section we will present the results of experiments for such sorters.

For some practical applications only the maximum or the minimum subsets need to be extracted. This task can be solved by removing the networks $\mathrm{SN}_{\text {min }}$ (for finding only the maximum subset) or $\mathrm{SN}_{\text {max }}$ (for finding only the minimum subset).

## 5 Implementations, Experiments and Comparisons

Fig. 12 shows the organization of experiments. We have used a multi-level computing system [12]. Initial (source) data are either generated randomly in software of the PS with the aid of C language rand function (see number 1 in Fig. 12) or prepared in the host PC (see number 2 in Fig. 12). In the last case data may be generated by some functions or copied from available benchmarks. Computing subsets in software/hardware systems is done completely in Zynq APSoC xc7z0201clg484c housed on ZedBoard [32] with the aid of the described above software/hardware architecture (see Fig. 4). Computing subsets in software only sorters is completely done in the PS calling C language qsort function which sorts data and after that the maximum and minimum subsets are extracted from the sorted data. The results are verified in software running either
in the PS (see number 3 in Fig. 12) or in the host PC (see number 4 in Fig. 12). Functions for verification of the results are given in [12]. Verification time is not taken into account in the measurements below. Methods that are used for copying files between the PC and APSoCs are explained in [12] with examples.

Synthesis and implementation of hardware modules were done in Xilinx Vivado 2014.2 design environment from specifications in VHDL. Standalone software applications have been created in C language and uploaded to the PS memory from Xilinx SDK (version 2014.2) using methods described in [12]. Interactions with APSoC are done through the SDK console window.


Figure 12: Experimental setup
For all the experiments 64-bit AXI ACP port was used for transferring blocks between the PL and memories. More details about this port can be found in [12,23,33]. The size of each block for burst mode is chosen to be 128 of 64-bit items (two 32-bit items are sent/received in one 64-bit word). Two memories were tested: the OCM and external (on-board) DDR. The OCM is faster because it provides 64-bit data transfers [1], but the size of this memory is limited to 256 KB . The available on ZedBoard 4 Gb DDR provides 32-bit data transfers.

The measurements were based on time units (returned by the function XTime_GetTime [34]) for $L_{\text {max }}=L_{\text {min }}=$ $64, M=32$, and $K=200$. Each unit returned by this function corresponds to 2 clock cycles of the PS [35]. The PS clock frequency is 666 MHz . Thus, any unit corresponds to approximately 3 ns. The PL clock frequency was set to 100 MHz . Fig. 13 shows the time consumed for computing the maximum and minimum subsets for data sets with different sizes in KB (from 2 to 128). Since $\mathrm{M}=32$ the number of processed words ( N ) is equal to the indicated size divided by 4 . Fig. 14 shows the acceleration of software/hardware systems comparing to software only systems. Note that Figs. 13, 14 present diagrams for OCM. If DDR memory is used then communication overheads are slightly increased but acceleration in the software/hardware systems comparing
to software only system is again significant. For $\mathrm{M}=64$ speed-up is increased in almost 2 times.


Figure 13: Computing time in software only and software/hardware systems


Figure 14: Acceleration of software/hardware systems comparing to software only system

If only the maximum or only the minimum subsets have to be computed the acceleration is almost the same, but the occupied hardware resources are reduced.

If the size of the requested subsets is increased in such a way that all data need to be read from memory several times (see section 4) then acceleration is decreased. Table 1 presents the results for extracting larger subsets (containing from 127 to 505 32-bit data items) from 128 KB set.

Table 1: The results for extracting larger subsets from 128 KB set

| $N$ | 127 | 190 | 253 | 316 | 379 | 442 | 505 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Time in $\mu \mathrm{s}$ | 926.4 | $1,393.7$ | $1,856.7$ | $2,320.5$ | $2,780.4$ | $3,245.5$ | $3,708.9$ |

For very large subsets acceleration may even be less than 1, i.e. software only system becomes faster. In such cases software/hardware sorters from [12] can be used directly and they provide acceleration for all potential cases even for $L_{\text {max }}=N$ or $L_{\text {min }}=N$. Such acceleration is not as high as in Fig. 14 and it is equal to 6 for $N=512, K=256$ (now $K$ is the size of blocks sorted in hardware and further merged in software) and 1.4 for $N=33,554,432, K=256$. These results were taken from experiments with data sorters from [12] (in all experiments $M=32$ ). We found that for small and moderate subsets the proposed here methods provide significantly better acceleration.

## 6 Conclusion

The paper suggests hardware/software architecture for fast extraction of minimum and maximum sorted subsets from large data sets and two methods of such extractions based on highly parallel and easily scalable sorting networks. The basic idea of the methods is incremental construction of the subsets that is done concurrently with transfer of initial data (source sets) through advanced high-performance interfaces in burst mode. Thorough experiments were done with entirely implemented on-chip designs in Zynq xc7z020-1clg484c device housed on ZedBoard. The size of initial sets varies from 512 to more than 33 million of 32 -bit words. The results demonstrate significant speed-up comparing to pure software implementations in the same Zynq device, namely performance was increased by 1-2 orders of magnitude for small subsets and by a factor ranging from 1.4 to 6 for very large subsets.

## 7 Acknowledgments

This research was supported by EU through European Regional Development Funds, the institutional research funding IUT 19-1 of the Estonian Ministry of Education and Research, ESF grant 9251, and Portuguese National Funds through FCT - Foundation for Science and Technology, in the context of the project PEst-OE/ EEI/UIO127/2014.

## 8 References

1. Xilinx, Inc. (2014). Zynq-7000 All Programmable SoC Technical Reference Manual. http:// www.xilinx.com/support/documentation/user_ guides/ug585-Zynq-7000-TRM.pdf.
2. Crockett L.H., Elliot R.A., Enderwitz M.A., and Stewart R.W. (2014). The Zynq Book. University of Strathclyde.
3. Hao L. and Stitt G. (2012). Bandwidth-SensitivityAware Arbitration for FPGAs. IEEE Embedded Systems Letters, 4(3), 73-76.
4. Bailey D.G. (2011) Design for Embedded Image Processing on FPGAs. John Wiley and Sons.
5. Sklyarov V., Skliarova I., Barkalov A., and Titarenko L. (2014) Synthesis and Optimization of FPGAbased Systems. Springer.
6. Cristo, A., Fisher, K., Gualtieri, A.J., Pérez, R.M., and Martínez, P. (2013). Optimization of Processor-to-Hardware Module Communications on Spaceborne Hybrid FPGA-based Architectures. IEEE Embedded Systems Letters, 5(4), 77-80.
7. Canedo, A., Ludwig, H., and Al Faruque, M.A. (2014). High Communication Throughput and Low Scan Cycle Time with Multi/Many-Core Programmable Logic Controllers. IEEE Embedded Systems Letters, 6(2), 21-24.
8. Santarini, M. (2013). All Eyes on Zynq SoC for Smart Vision. XCell Journal, 83(2), 8-15.
9. Dick, C. (2013). Xilinx All Programmable Devices Enable Smarter Wireless Networks. XCell Journal, 83(2), 16-23.
10. Xilinx, Inc. (2014) Vivado Design Suite Guides. http://www.xilinx.com/support/index.html/content/xilinx/en/supportNav/design_tools.html.
11. Xilinx, Inc. (2014). Zynq-7000 All Programmable SoC Software Developers Guide. UG821 (v9.0). http://www.xilinx.com/support/documentation/ user_guides/ug821-zynq-7000-swdev.pdf.
12. Sklyarov, V., Skliarova, I., Silva, J., Rjabov, A., Sudnitson, A., and Cardoso, C. (2014) Hardware/Software Co-design for Programmable Systems-onChip. TUT Press.
13. Xilinx, Inc. (2013). Simple AMP Running Linux and Bare-Metal System on Both Zynq SoC Processors. http://www.xilinx.com/support/documentation/ application_notes/xapp1078-amp-linux-baremetal.pdf.
14. Sklyarov, V. and Skliarova, I. (2013). Digital Hamming Weight and Distance Analyzers for Binary Vectors and Matrices. International Journal of Innovative Computing, Information and Control, 9(12), 4825-4849.
15. Zmaranda, D., Silaghi, H., Gabor, G., and Vancea, C. (2013). Issues on Applying Knowledge-Based Techniques in Real-Time Control Systems, International Journal of Computers, Communications and Control, 8(1), 166-175.
16. Field, L., Barnie, T., Blundy, J., Brooker, R.A., Keir, D., Lewi, E., and Saunders, K. (2012) Integrated field, satellite and petrological observations of the No-
vember 2010 eruption of Erta Ale. Bulletin of Volcanology, 74(10), 2251-2271.
17. Zhang, W., Thurow, K., and Stoll, R. (2014). A Knowledge-based Telemonitoring Platform for Application in Remote Healthcare. International Journal of Computers, Communications and Control, 9(5), 644-654.
18. Verber, D. (2011), Hardware implementation of an earliest deadline first task scheduling algorithm. Informacije MIDEM, 41(4), 257-263.
19. Baker, Z.K. and Prasanna, V.K. (2006). An Architecture for Efficient Hardware Data Mining using Reconfigurable Computing Systems. Proc. 14 $4^{\text {th }}$ Annual IEEE Symposium on Field-Programmable Custom Computing Machines, Napa, USA, 67-75.
20. Sun, S. (2011). Analysis and acceleration of data mining algorithms on high performance reconfigurable computing platforms. Ph.D. thesis, lowa State University. http://lib.dr.iastate.edu/cgi/ viewcontent.cgi?article=1421\&context=etd.
21. Wu, X., Kumar, V., Quinlan, J.R., et al. (2014). Top 10 algorithms in data mining. Knowledge and Information Systems, 14(1), 1-37.
22. Firdhous, M.F.M (2010). Automating Legal Research through Data Mining. International Journal of Advanced Computer Science and Applications, 1(6), 9-16.
23. Silva, J., Sklyarov, V., and Skliarova I. (2015) Comparison of On-chip Communications in Zynq7000 All Programmable Systems-on-Chip. IEEE Embedded Systems Letters, 7(1), 31-34.
24. Neuendorffer, S., and Martinez-Vallina, F. (2013). Building Zynq Accelerators with Vivado High Level Synthesis. Proc. ACM/SIGDA Int. Symp. on Field Programmable Gate Arrays, Monterey, CA, USA, 1-2.
25. Xilinx, Inc. (2014). OS and Libraries Document Collection UG647. http://www.xilinx.com/support/documentation/sw_manuals/xilinx2014_2/ oslib_rm.pdf.
26. Baddar, S.W.A.-H., and Batcher, K.E. (2011). Designing Sorting Networks. A New Paradigm. Springer.
27. Sklyarov, V., and Skliarova, I. (2014). High-performance implementation of regular and easily scalable sorting networks on an FPGA. Microprocessors and Microsystems, 38(5), 470-484.
28. Alekseev, V.E. (1969). Sorting Algorithms with Minimum Memory. Kibernetica, 5, 99-103.
29. Knuth, D.E. (2011). The Art of Computer Programming. Sorting and Searching, vol. III. AddisonWesley.
30. Xilinx, Inc. (2014). 7 Series FPGAs Overview. http://www.xilinx.com/support/documentation/ data_sheets/ds180_7Series_Overview.pdf.
31. Mueller, R., Teubner, J., and Alonso, G. (2012) Sorting networks on FPGAs. Int. J. Very Large Data Bases, 21 (1), 1-23.
32. Avnet, Inc. (2014). ZedBoard (Zynq ${ }^{\top M}$ Evaluation and Development) Hardware User's Guide, Version 2.2. http://www.zedboard.org/sites/default/ files/documentations/ZedBoard_HW_UG_v2_2. pdf.
33. Sadri, M., Weis, C., When, N., and Benini, L. (2013). Energy and Performance Exploration of Accelerator Coherency Port Using Xilinx ZYNQ. Proceedings of the $10^{\text {th }}$ FPGAWorld Conference, Copenhagen/Stockholm.
34. Xilinx, Inc. (2013). LogiCORE IP AXI Master Burst v2.0. Product Guide for Vivado Design Suite. http://japan.xilinx.com/support/documentation/ip_documentation/axi_master_burst/v2_0/ pg162-axi-master-burst.pdf.
35. Xilinx, Inc. (2014). Standalone (v.4.1). UG647. http://www.xilinx.com/support/documentation/ sw_manuals/xilinx2014_2/oslib_rm.pdf.

Arrived: 09. 11. 2014
Accepted: 14.04. 2015

# Radiation Behavior and Test Specifics of $A-D$ and $D-A$ Converters 

Alexander A. Demidov, Oleg A. Kalashnikov, Alexander Y. Nikiforov, Alexander S. Tararaksin, Vitaly A. Telets

National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Specialized Electronic Systems» (SPELS), Moscow, Russia


#### Abstract

ADC/DAC radiation failures are mainly due to radiation-induced degradation of precision parameters of the transfer characteristic such as gain, zero offset, full-scale voltage, integral and differential non-linearity, conversion error. ADC/DAC radiation failure specifics is that even a slight deviation of electrical parameter of internal elements (comparator threshold, internal reference voltage, switch leakage, operational amplifier gain, etc.) often leads to significant degradation of ADC/DAC accuracy. ADC/DAC radiation test procedure and facilities are developed and test results are introduced.


Keywords: analog-to -digital converter (ADC); digital-to-analog converter (DAC); radiation; test technique

# Sevalno obnašanje in testne posebnosti A-D in D-A pretvornikov 

Izvleček: ADC/DAC sevalne napake so običajno posledica radiacijsko pogojenega staranja natančnosti parametrov prenosnih karakteristik, kot je ojačenje, ničelni odmik, polna napetost, linearna in diferencialna linearnost in napaka pretvarjanja. Posebnost ADC/ DAC sevalnih napak je, da že majhna sprememba električnih lastnosti elementov (prag primerjalnika, interna referenčna napetost, uhajanje preklopnika, ojačenje...) v veliki meri vpliva na degradacijo natančnosti pretvornika. Predstavljeni so razviti postopki testiranja in rezultati meritev.

Ključne besede: analogni digitalen pretvornik (ADC); digitalno analogen pretvornik (DAC); sevanje; tehnike testiranja

* Corresponding Author's e-mail: oakal@spels.ru


## 1 Introduction

Analog-to-digital and digital-to-analog converters (ADC and DAC) are widely used in space, collider physics, avionics, nuclear power plants, etc. applications, being the essential parts of data pre-processing and control units. Therefore important issues are to analyze radiation behavior particularities and to develop informative radiation test procedures and technique to estimate ADCs and DACs radiation sensitive parameters and characteristics degradation [1], [2].

The most radiation sensitive feature of ADCs and DACs is accuracy which is determined by the transfer characteristic parameters of such as gain, zero offset, fullscale voltage, integral and differential non-linearity, conversion error. ADC and DAC radiation failure specifics as compared with digital integrated circuits (ICs) is that even a slight deviation of a parameter (comparator
threshold, internal reference voltage, switch leakage, operational amplifier gain, etc.) often leads to significant degradation of ADC/DAC accuracy [3].

Total ionizing dose (TID) accumulation in ADC and DAC results in continuous degradation of static and dynamic conversion parameters, while transient irradiation (gamma flesh or single charged particles) may result in ADC output code failures or in DAC output voltage transients. The radiation behavior of various ADCs and DACs is rather complicated and significantly depends on the particular IC architecture and on the bias and operation conditions during irradiation and testing. This should be considered in development of ADC/DAC radiation tests techniques and facilities.

A lot of radiation tests of various ADC and DAC were carried out in the MEPhl-SPELS radiation test labora-
tory (Moscow, Russia) [4]. The analysis of test data demonstrates the critical importance of ADCs and DACs functional tests as compared to other IC groups. ICs dominant failure types (parametric or functional) statistics from our test experience is presented in Fig. 1. One can see the essential prevalence of parametric TID failures for simple logic while other (complex) ICs are characterized by subsequent or even dominant functional failures [5], and ADC/DAC ICs are the leaders.


Figure 1: Relative part (\%) of parametric and functional radiation failures for different ICs classes

Variety of hardware and software solutions has been developed to provide reliable and informative testing of different ADC/DAC ICs directly under irradiation within $-60 \ldots+125$ C temperature range. The system implements both operational modes under irradiation assignment and monitoring of the entire set of static and dynamic parameters which characterize ADC/DAC radiation hardness. We present the structure, the operation principles and the basic technical specifications of the system in this paper.

The used set of original compact radiation test basic facilities is introduced ([2], [4], [6]) including Co-60 and Cs-137 isotopic gamma-sources, electron linear accelerator, flash X-ray machine - all with minimum possible signal cables length (about 1 m only). The used ions cyclotron (in Dubna) and high energy proton synchrocyclotron (in Gatchina) were rather traditional. And we widely used laser and X-ray simulators which give us the unique possibility to measure all ADC/DAC informative parameters and characteristics.

The paper also contains numerous test results of ADCs and DACs which designed by various manufacturers by using various architectures and technologies. We concentrate on TID effects, single event effects (SEE), and transient radiation effects (TRE). The data presented is mostly experimental - the theory of ADC/DAC radiation effects is well known and has been widely presented [7]-[9]. The purpose of this paper is to demonstrate the variety of these effects. We present here the most
typical results, which cause the specific ADC-DAC radiation test technique development.

## 2 Total ionizing dose effects

The typical TID effect in ADC and DAC is transfer function (TF) degradation and the associated degradation of a converter precision parameters (DAC TF - dependence of output voltage/current vs. input code, ADC TF - dependence of output code vs. input voltage). For example, in Fig. 2 TFs of ADC (Fig. 2a) and DAC (Fig. 2b) within the Data Acquisition System (DAS) ADuC812BS (Analog Devices) at different TID values is presented. Fig. 3 shows the TF degradation of ADC AD1671SQ/883B (Analog Devices) with TID accumulation. One can see that TF degradation can be gradual and smooth or sharp and abrupt [10].
(a)

(b)


Figure 2: $A D C$ (a) and DAC (b) within the DAS ADuC812BS - TFs at different TID values: 1 - initial, 2 $12 \operatorname{krad}(\mathrm{Si}), 3$ - $16 \operatorname{krad}(\mathrm{Si})$


Figure 3: ADC AD1671SQ/883B TF degradation with TID accumulation

ADC/DAC TF degradation results in their accuracy parameters degradation - integral nonlinearity (INL), differential nonlinearity (DNL), offset and gain errors. INL is the measure of the deviation values on the actual TF from a straight line. DNL is the difference between an actual step width (for an ADC) or step height (for a DAC) and the ideal value of 1 least significant bit (LSB). Offset error is defined as the difference between the nominal and actual offset points when the digital output (for an ADC) or digital input (for a DAC) is zero. Gain error is defined as the difference between the nominal and actual gain points on TF when the digital output (for an ADC) or digital input (for a DAC) is full scale [11]. As an example, a number of INL and DNL TID-dependencies of DAC within ADuC812BS is shown in Fig. 4. The curves are plotted for several irradiated samples and correspond to average TF changes which are presented in Fig. 2b [12].

It is important to mention that not only the maximum values of ADC/DAC accuracy parameters degrade under irradiation, but the dependencies of these parameters vs. input or output signals (codes) vary too. For example, the dependencies of INL and DNL of ADC PV2 are presented in Fig. 5 and 6 respectively. The different TID behavior of these two parameters may be noted. In the INL graphs there is a rise of "teeth" and general distortion increase (bending). At the same time, the degradation of DNL appears as increase of the spikes amplitude at certain ADC output code. The values of DNL for the rest of the codes do not increase practically [3].

Thus, to determine the radiation behavior of ADCs and DACs with TID accumulation, a set of TFs should be re-
corded during irradiation, which is used to calculate TID dependencies of a converter accuracy parameters.

It should be noted that such "standard" analog and digital parameters of converters as supply current, output voltage, maximum operating frequency etc. also changes under irradiation. However, ADCs and DACs have no specifics when compared with other functional classes of ICs both in these parameters degradation and in their control procedures during testing. Therefore this is not the issue of this paper.


Figure 4: TID dependencies of DAC within ADuC812BS samples DNL (a) and INL (b)

ADC output code on X-axes and DNL (in units of LSB) on $Y$-axes

## 3 Single event effects

Single event effects (SEE) due to single nuclear particles (such as heavy ions and protons) may result in either failures (latch-up, burn-out and so on) or single event upsets (SEU). Failures, as well as the experimental methods of their detection are well known and presented in a large number of publications [13], [14].

initial


84 krad (Si)


120 krad (Si)
 160 krad (Si)


200 krad (Si)
Figure 5: Total ionizing dose degradation of ADC PV2 integral nonlinearity: ADC output code on X-axes and INL (in units of LSB) on $Y$-axes

There is no specifics of SEE failures in ADC and DAC, so in this paper we focus on converters SEU.

There are two types of ADC-DAC SEU. First, DAC SEU may lead to the output voltage (current) spikes during irradiation. Similarly, ADC SEU may occur as the output code pulse (reversible change). Fig. 7 shows the output voltage transients of DAC TLV5638MFKB (Texas Instruments) during irradiation by Xe-ions in the Dubna cyclotron [13].

The second type of SEU is upset of ADC-DAC internal flip-flops and registers as a result of nuclear particles influence. The upsets of data registers can change DAC output voltage (current) or ADC output code while the upsets of control registers can lead to a converter operational mode change. In either case it is usually necessary to restart a converter in order to restore its normal operation.


84 krad (Si)


120 krad (Si)


160 krad (Si)

$200 \mathrm{krad}(\mathrm{Si})$
Figure 6: Total ionizing dose degradation of ADC PV2 differential nonlinearity:


Figure 7: DAC TLV5638MFKB output voltage transients during Xe-ions irradiation.

The purpose of experimental research is to detect ADC and DAC SEU during irradiation at nuclear particle accelerators. Several ions with different Linear Energy Transfers (LET) are usually used. For each ion a SEU cross-section is determined by the equation:
$\sigma_{\text {SEU }}=\mathrm{N}_{\mathrm{SEU}} /\left(\Phi \times \mathrm{N}_{\mathrm{B}}\right)$,
where $\mathrm{N}_{\text {SEU }}$ - number of upsets detected, $\Phi$ - particles fluence at irradiation session, $\mathrm{N}_{\mathrm{B}}$ - number of bits under test. The approximating curve is to be plotted based on these data, and a converter SEU parameters - the threshold LET and the saturation cross-section - are determined. Weibull-function is used for the experimental data approximation. Fig. 8 shows such a curve and SEU parameters for sigma-delta ADC AD7711ASQ (Analog Devices) [14].


Figure 8: ADC AD7711ASQ SEU experimental data, Weibull approximation, and SEU parameters (LET and cross-section)

## 4 Transient radiation effects

Transient radiation effects (TRE) or dose-rate effects are caused by pulsed gamma irradiation. These effects in ADC and DAC are similar to SEE - both failures and upsets are also possible. The difference is that in SEE case only a single circuit element is locally affected by the particle every moment while TRE specific is that all functional elements and parasitic structures are jointly affected by radiation. Upsets are characterized by the threshold level of gamma dose rate and by the recovery time. Moreover, as a rule, there is a clear dependence of an output signal (voltage or current of DAC and code of ADC) pulse response amplitude and duration on the dose rate.

As an example, Fig. 9 shows a set of radiation pulse waveforms of the DAC PA1 output voltage at different dose rates. It is seen the increase of pulses both amplitude and duration. The performance criteria are typically established by the maximum allowable amplitude and duration of the ionization pulses, and are determined by the particular equipment application conditions [15].

Generally, tested DAC and ADC are set to a static operation mode with a certain output level/code, and the
output response at the moment of gamma-ionization pulse is registered. But the upsets may occur in the dynamic operation modes of a converter as well. For example, the waveforms in Fig. 10 illustrate the gamma pulse upset of DAC PA3 when operating in the dynamic mode of sine signal generation [16].


Figure 9: DAC PA1 output voltage pulses at different dose rates


Figure 10: Dose rate upset of DAC PA3 in dynamic mode

## 5 Radiation test technique

As it was already mentioned above, the accuracy parameters of ADC and DAC are determined by the transfer function (TF). For its measurement during a TID test,
full range linearly increasing voltage (code) is to be put on an ADC (DAC) under test inputs, and output ADC code (DAC voltage/current) is to be measured. This procedure should be repeated for all TID values we are interested in, thus it is necessary to carry out the measurements as fast as possible to satisfy the condition [2]:
$\mathrm{T}_{\text {MEAS }}<0.1 \cdot \mathrm{~T}_{\text {RAD }}$,
where $\mathrm{T}_{\text {MEAS }}$ - full TF measurement duration, $\mathrm{T}_{\text {RAD }}$ - time between measurements. The 0.1 factor is normally used in TID test practice to provide relatively short measurement duration as compared to irradiation time. It allows to minimize the influence of annealing during measurement and eliminating test result distortion.

One more test procedure feature is also provided by the timing requirements. According to our experience and data it is very important to measure TF directly during irradiation. Measuring after irradiation would distort the real radiation behavior picture and hardness level because of annealing that can result even in full operation recovery. In Fig. 11 two graphs of CMOS ADC nonlinearity are shown: the first is measured immediately after the $100 \mathrm{krads}(\mathrm{Si})$ irradiation and the second - 12 hours later [3]. One can see that 12 -hours annealing leads to an ADC's operation recovery.


Figure 11: ADC nonlinearity measured immediately after 100 krads(Si) irradiation and 12 hours later (datasheet margins are shown by dashed lines at $\pm 4$ LSB)

Another problem is caused by the TF standard measuring procedure [3], which requires the error of a measuring device (accuracy of the input voltage) should be within $1 / 16$ of a DAC (ADC) LSB value in the range of measurement corresponding to the full range of a DAC (ADC) output (input) voltage.

To satisfy these conditions it is necessary to use the special methods of high accuracy voltage biasing, as well as multiple measurements and averaging the measured values. These procedures should be hard-
ware-implemented to meet ultra hard restrictions on measurement speed. The hardware structure based on a differential amplifier is shown in Fig. 12 [17]. One input of the amplifier is connected to DAC under test voltage output, and another input - to the bias voltage. Direct measurement of the output voltage is replaced by measurement of the adjacent codes output voltages difference.

The specialized ADC and DAC testing system based on the National Instruments hardware, LabView software, and a set of device-under-test boards, adapted to different converters, is developed [18]. The results of radiation tests of several dozen converters carried out using this equipment, have confirmed its effectiveness [19].


Figure 12: Voltage biasing structure for precision DAC TF measuring

## 6 Conclusions

The article presents the typical radiation effects in DACs and ADCs when exposed to different types of ionizing radiation. It can be seen that the converters specifics, which are characterized by both digital and analog parameters, leads to their radiation behavior specifics - the effects are caused by both digital registers and control circuits failures and failures and parameters degradation of analog units.

This feature of ADCs and DACs leads to the fact that the procedure of radiation test has some specific features when compared with test procedure of "pure" digital or analog integrated circuits. It is necessary to use the special control and monitor technique, which combines software control and data processing with precise measurements. The implementation of this technique requires specialized test equipment that should be compatible with the specialized radiation facilities with short signal cables.

## 7 References

1. A. Nikiforov, A. Chumakov, V. Telets, et. al, "IC Space Radiation Effects Experimental Simulation",
// Proc. of Workshop "Space Radiation Environment Modelling New Phenomena and Approaches", Oct. 7-9, 1997, Moscow, Russia, p. 4-11.
2. V. Belyakov, A. Chumakov, A. Nikiforov, V. Pershenkov, "IC's radiation effects modeling and estimation", // Microelectronics Reliability, 1999, v. 40, № 12, pp. 1997-2018. [9] O. Kalashnikov, "Statistical Variations of Integrated Circuits Radiation Hardness", // RADECS Proceedings, 2011, pp. 661-665.
3. O. Kalashnikov, A. Artamonov, A. Demidov, et. al, "ADC/DAC Radiation Test Technique", // Workshop Record 4th European Conf. "Radiations and Their Effects on Devices and Systems" (RADECS 97), 1997, Palm Beach-Cannes, France, pp. 56-60.
4. Compendium of International Irradiation Test Facilities, // RADECS 2011, p. 66.
5. O. Kalashnikov, "Statistical Variations of Integrated Circuits Radiation Hardness", // RADECS Proceedings, 2011, pp. 661-665.
6. A. Chumakov, A. Nikiforov, V. Telets et. al, "IC space radiation effects experimental simulation and estimation methods", // Radiation Measurements, v. 30 (5) , pp. 547-552.
7. T. Turflinger et. al, "Understanding Single Event Phenomena in Complex Analog and Digital Integrated Circuits", // IEEE Trans. Nucl. Sci., 1990, v. 37, pp. 1832-1838.
8. J. Tausch, "Radiation Testing of Mixed-Signal Microelectronics," IEEE NSREC 2000 short course proceedings.
9. T. Turflinger, "Transient Radiation Test Techniques for High-Speed Analog to Digital Converters", // IEEE Trans. Nucl. Sci., 1989, v. 36, pp. 2356-2361.
10. O. Kalashnikov, A. Demidov, A. Nikiforov, et. al, "Integrating analog-to-digital converter radiation hardness test technique and results", // IEEE Trans. Nucl. Sci., 1998, v. 45, pp. 2611-2615.
11. Understanding Data Converters - Application Report, // Texas Instruments, 1995, http://www. ti.com/lit/an/slaa013/slaa013.pdf.
12. O. Kalashnikov, A. Nikiforov, "TID behavior of complex multifunctional VLSI devices", // Proceedings of the International Conference on Microelectronics, ICM, 2014, pp. 455-458.
13. A. Chumakov, A. Pechenkin, A. Egorov, et.al, "Estimating IC susceptibility to single-event latch-up", // Russian Microelectronics, 2008, v. 37 (1), pp. 4146.
14. A. Chumakov, A. Vasil'ev, A. Kozlov, et. al, "Single-event-effect prediction for ICs in a space environment", // Russian Microelectronics, 2010, v. 39 (2), pp. 74-78.
15. T. Agakhanyan, A. Nikiforov, "Predicting the Effect of Pulsed lonizing Radiation on Operational Amplifiers", // Russian Microelectronics, 2002, v. 31 (6), pp. 375-383.
16. A. Nikiforov, A. Sogoyan, "Modeling of high-doserate pulsed radiation effects in the parasitic MOS structures of CMOS LSI circuits", // Russian Microelectronics, 2004, v. 33 (2), pp. 80-91.
17. T.B. Williams, "The calibration of a DAC using differential linearity measurements", //IEEE Trans. on Instr. and Meas., 1982, v.31, №4.
18. D. Bobrovsky, G. Davydov, A.Petrov, et. al, "Realization of electronic component base radiation test methods based on hardware-software complex of National Instruments hardware", // Electronics, 2012, v.5(97), pp. 91-106.
19. A. Sogoyan, A. Artamonov, A. Nikiforov, D. Boychenko, Method for integrated circuits total ionizing dose hardness testing based on combined gamma- and xray- irradiation facilities,// Facta Univesitatis: series Electronics and Energetics, 2014, Vol. 27, No. 3, pp. 329-338.

Arrived: 25. 11. 2014
Accepted: 17.03. 2015

# Computing Worst-Case Performance and Yield of Analog Integrated Circuits by Means of Mesh Adaptive Direct Search 

Árpád Búrmen ${ }^{1}$, Husni Habal ${ }^{2}$<br>${ }^{1}$ University of Ljubljana, Faculty of Electrical Engineering<br>${ }^{2}$ Technical University of Munich


#### Abstract

Estimating the parametric yield of a circuit by means of a Monte Carlo analysis can be slow, particularly when the yield estimate is close to $100 \%$, as a large number of samples are necessary to reach the desired level of confidence. Deterministic numerical algorithms have been successfully used in commercial tools for yield estimation. Many of them are gradient-based. The gradients are estimated numerically using finite differences, because most simulators do not compute sensitivities. In this paper, an approach is proposed based on a derivative-free optimization algorithm from the family of mesh adaptive direct search methods. The basic algorithm is extended with capabilities that speed up the convergence and enable the algorithm to cope with infeasible starting points. The new approach is compared to a commercial tool that uses gradient-based algorithms for worst-case analysis. The results show that the proposed approach is capable of producing accurate results within similar computational budgets.


Keywords: analog circuit design;, design centering; worst-case analysis; yield analysis; optimization; mesh adaptive direct search

# Določanje najslabših lastnosti in izplena analognih integriranih vezij z adaptivnim mrežnim direktnim optimizacijskim postopkom 

Izvleček: Določanje izplena vezja s pomočjo Monte Carlo analize je pogosto zamuden postopek, še posebej, ko se izplen približuje $100 \%$, ker potrebujemo za zanesljive rezultate veliko število vzorcev. Deterministični postopki za določanje izplena so na voljo v komercialnih orodjih. Številni postopki se zanašajo na informacijo o gradientu, ki ga določajo numerično, saj večina simulatorjev ne računa občutljivosti rezultatov. Članek opisuje pristop z uporabo brezgradientnega optimizacijskega postopka iz družine adaptivnih mrežnih direktnih optimizacijskih postopkov. Osnovni postopek je nadgrajen z razširitvami, ki pospešijo konvergenco proti rešitvi problema in omogočajo, da postopek uporabi začetno točko, ki krši omejitve. Predlagani pristop smo primerjali s komercialnih orodjem, ki uporablja gradientne optimizacijske postopke. Rezultati kažejo, da je predlagan pristop sposoben najti pravilne rešitve problemov v primerljivem času.

Ključne besede: načrtovanje analognih vezij; centriranje; določaje najslabših vrednosti lastnosti; analiza izplena; optimizacija; adaptivni mrežni direktni optimizacijski postopki

[^5]
## 1 Introduction

Modern integrated circuits must exhibit adequate performance across a given range of operating conditions, such as supply voltage and temperature, and in the presence of random variations resulting from the manufacturing process [1]. Towards this objective, parametric yield is defined as the fraction of manufactured circuits that meet all performance specifications, such
as minimum gain and maximum power, in consideration of all operating conditions, as well as the statistical distribution of random variations. A prerequisite for designing such a circuit is an efficient means of evaluating the circuit's worst performance.

Manufactured circuits that fail an imposed performance specification must be discarded, such that the
parametric yield is reduced below 100\%. The simplest means to estimate the parametric yield is Monte-Carlo analysis (MCA). Unfortunately, a very large number of performance evaluations are needed for accurate estimation by MCA when the yield is close to 100\% [2]. This is prohibitively inefficient, since each performance evaluation requires a costly circuit simulation. More efficient means to evaluate electrical performance given the worst-case combination of operating conditions and random variations is therefore necessary for robust circuit design; some alternatives have been presented in literature (cf. [2, 3]).

In [2], the worst-case distance (WCD) metric was used to obtain yield estimates with less computation. The WCD method requires the numerical solution of an optimization problem. This problem can be solved in significantly less time than it takes an MCA to obtain similar or more accurate yield estimates. The alternative to yield estimation by WCD is the worst-case performance (WCP) method [2]. In WCP, the worst value of a performance is calculated which corresponds to a predefined parametric yield ( Y ). If this worst value satisfies the performance specification the parametric yield is not smaller than Y .

In general, both WCD and WCP require the solution of a non-linear optimization problem by numerical methods. Deterministic optimizations have been successfully applied to solve the WCD and WCP problems typical in analog integrated circuits -- including academic and commercial tools. These algorithms have been derivative-based, so that the sensitivity of the electrical performances to the value of the operating and statistical parameters was needed (e.g. [2] and the references therein, [3]). In this paper a new deterministic and de-rivative-free method is proposed to solve the WCD and WCP problems. The method is based on mesh adaptive direct search (MADS)[4].

The remainder of the paper is organized as follows: section 2 introduces the mathematical formulation of WCD and WCP. Section 3 gives an overview of MADS and modifications introduced by the proposed approach. The implementation details are the subject of Section 4. Section 5 presents the results and compares them to the results obtained with a derivative-based algorithm implemented in a commercial tool (WiCkeD [5]). The concluding remarks are given in Section 6.

Notation. Inequalities apply to vectors component wise. $\mathbf{O}$ denotes a vector of all-zeros. The $i$-th component of vector $v$ is denoted by $v_{i}$. An element of a matrix $\boldsymbol{A}$ is denoted by $a_{i j}$. The ramp function $\operatorname{ramp}(x)$ is zero for $x<0$ and equal to $x$ otherwise. The group of orthogonal transformations of $R^{n}$ is denoted by $O_{n}$. The $i-$ th orthonormal basis vector is denoted by $\boldsymbol{e}_{i}$.

## 2 Mathematical formulation for worstcase analysis

Let $\boldsymbol{x}_{0}$ denote the vector of $\mathrm{n}_{0}$ parameters describing the circuit's operating condition, also referred to as the operating parameters. The prescribed range of operating conditions within which the circuit must operate is specified by lower and upper bounds on operating parameters given by vectors $\boldsymbol{x}^{L}$ and $\boldsymbol{x}^{H}{ }_{0^{\prime}}$ respectively.

The performance of the circuit is also affected by variations of the manufacturing process which in turn are modeled as mutually dependent random variables. Without loss of generality, the set of dependent process parameters can be mathematically transformed into a set of independent random variables with normal distribution. Let $\boldsymbol{x}_{\mathrm{s}}$ denote a vector representing a realization of these random variables. Components of $\boldsymbol{x}_{5}$ are also referred to as the statistical parameters. By assumption the joint probability density of the statistical parameters can be expressed as

$$
\begin{equation*}
p\left(\boldsymbol{x}_{S}\right)=\frac{1}{(2 \pi)^{\frac{n_{e}}{2}}} e^{-\frac{\left\|\boldsymbol{x}_{s}\right\|^{2}}{2}} \tag{1}
\end{equation*}
$$

Circuit behavior is evaluated by a number of performances, for example power, amplification gain, and bandwidth. The performances are ordered in a vector $\boldsymbol{f}$ with length m . Their value for any specific circuit will depend on the value of the operating parameters $\boldsymbol{x}_{0^{\prime}}$ as well as the statistical parameters $\boldsymbol{x}_{s}$. Component $f_{i}$ of $f$ is the value of a map computed by a simulator. With some abuse of notation one can write $f_{i}\left(\boldsymbol{x}_{\alpha^{\prime}} \boldsymbol{x}_{s}\right)$. For a circuit to behave correctly at ( $\boldsymbol{x}_{d_{d}} \boldsymbol{x}_{s}$ ) it must meet a set of performance specifications of the form $f_{i}\left(\boldsymbol{x}_{\boldsymbol{o}^{\prime}} \boldsymbol{x}_{S}\right) \geq G_{i^{\prime}}$ where $G_{i}$ denotes the target value corresponding to $f_{i}$ Performance specifications of the form $f_{i}\left(\boldsymbol{x}_{\sigma^{\prime}} \boldsymbol{x}_{s}\right) \leq G_{i}$ can be taken into account by replacing $f_{i}\left(\boldsymbol{x}_{\sigma} \boldsymbol{x}_{s}\right)$ and $G_{i}$ with $-f_{i}\left(\boldsymbol{x}_{\boldsymbol{\alpha}} \boldsymbol{x}_{s}\right)$ and $-G_{i}$. A well designed circuit behaves correctly across the given range of operating conditions and for a large percentage of circuits manufactured in the presence of manufacturing process variations (yield). The yield corresponding to performance $f_{i}$ can be computed by integrating (1) over the acceptance region of $f_{i}$ (shaded region in Figure 1) defined as
$f_{i}^{w}\left(\boldsymbol{x}_{S}\right)=f_{i}\left(\boldsymbol{x}_{0}^{w, i}\left(\boldsymbol{x}_{S}\right), \boldsymbol{x}_{S}\right) \geq G_{i}$
where $f_{i}^{w}\left(\boldsymbol{x}_{s}\right)$ is the worst value of $f_{i}$ at $\boldsymbol{x}_{s}$ across the prescribed range of operating parameters and
$\boldsymbol{x}_{0}^{w, i}\left(\boldsymbol{x}_{S}\right)=\arg \min _{\boldsymbol{x}_{0}^{L} \leq x_{0} \leq x_{0}^{H}} f_{i}\left(\boldsymbol{x}_{0}, \boldsymbol{x}_{S}\right)$
This integral cannot be computed analytically and is usually estimated with a Monte Carlo analysis.


Figure 1: The worst-case point $\boldsymbol{x}_{s}{ }^{\text {w,i }}$ and the linearization (dashed line) of $f_{i}^{w}\left(\boldsymbol{x}_{s}\right)=G_{i}$ (thick line) in the space of the statistical parameters. The acceptance region of $f_{i}$ is shaded. Replacing the nonlinear specification with its linearization at $\boldsymbol{x}_{5}{ }^{w, i}$ makes it possible to compute a yield estimate using (6). The approximation introduces an error equal to the integral of (1) over the region shaded in dark grey.

A good yield approximation can be obtained by replacing the performance with its linear model computed at the worst-case point $\left(\boldsymbol{x}_{0}^{w, i}\left(\boldsymbol{x}_{s}^{w, i}\right), \boldsymbol{x}_{s}^{w, i}\right)$ [2]. Figure 1 illustrates the worst-case point in the space of statistical parameters when $n_{s}=2$; the sphere $\left\|\boldsymbol{x}_{s}\right\|=\beta_{i}$ is tangential to the boundary of the acceptance region at $\boldsymbol{x}_{s}{ }^{w, i}$. The statistical parameters corresponding to this point are given by
$\boldsymbol{x}_{S}^{w, i}=\left\{\begin{array}{l}\arg \min _{f_{i}{ }^{w}\left(x_{S}\right) \leq G_{i}}\left\|\boldsymbol{x}_{S}\right\|^{2}, \text { for } f_{i}^{w}(\mathbf{0}) \geq G_{i} \\ \arg \min _{f_{i}^{w}\left(x_{S}\right) \leq G_{i}}\left\|\boldsymbol{x}_{S}\right\|^{2}, \text { otherwise }\end{array}\right.$
The worst-case distance of $f_{i}$ is defined as
$\beta_{i}=\left\{\begin{array}{l}\left\|\boldsymbol{x}_{S}^{w, i}\right\|, \text { for } f_{i}^{w}(\mathbf{0}) \geq G_{i} \\ -\left\|\boldsymbol{x}_{S}^{w, i}\right\|, \text { otherwise }\end{array}\right.$
If $f_{i}^{w}\left(\boldsymbol{x}_{s}\right)$ satisfies the design requirements at $\boldsymbol{x}_{s}=\mathbf{0}$ the worst case distance is positive, otherwise it is negative. By linearizing $f_{i}^{w}\left(\boldsymbol{x}_{s}\right) \geq G_{i}$ at the worst-case point a yield approximation can be computed analytically by integrating (1) over the light grey region in Figure 1. The obtained yield approximation is
$Y_{i}=\frac{1}{2}\left(1+\operatorname{erf}\left(\beta_{i} / \sqrt{2}\right)\right) \approx Y\left(f_{i} \geq G_{i}\right)$
The difference between the actual and estimated yield corresponds to the integral of (1) over the dark grey region in Figure 1.

The computationally intensive component of yield estimation is to find the solution to problem (4). For small yields the computational effort is in the same order of magnitude as that required by a Monte Carlo analysis. For large yields the number of the required Monte Carlo samples grows rapidly as the yield approaches 100\% while the computational effort for solving (4) remains the same. Typically a designer tunes the design parameters until $\beta_{i}$ (and the yield) is maximized.

Problem (4) has a general nonlinear constraint that can only be evaluated by circuit simulation. An alternative approach to yield maximization is often used. Instead of computing the WCD, the WCP corresponding to a given $\beta$ can be computed.

$$
\begin{equation*}
\left(\boldsymbol{x}_{0}^{w, i}, \boldsymbol{x}_{S}^{w, i}\right)=\arg \min _{\substack{x_{0}^{b} \leq \boldsymbol{x}_{0} \leq \boldsymbol{x}_{0}^{H} \\ \|\left. x S\right|^{\prime} \leq \beta^{2}}} f_{i}\left(\boldsymbol{x}_{0}, \boldsymbol{x}_{S}\right) \tag{7}
\end{equation*}
$$

The constraint in the WCP formulation is a convex quadratic function that can be evaluated without circuit simulation. If the $i$-th performance $f_{i}$ satisfies $f_{i}\left(\boldsymbol{x}_{0}^{w, i}\right.$, $\left.\boldsymbol{x}_{s}^{w^{\prime},}\right) \geq G_{i}$, then the WCD $\left(\beta_{i}\right)$ and corresponding yield estimate ( $Y_{i}$ ) will satisfy $\beta_{i} \geq \beta$ and $Y_{i} \geq 1 / 2(1+\operatorname{erf}(\beta / \checkmark 2))$.

Problems (4) and (7) are typical problems for which the initial deterministic method of choice is a gradientbased optimization algorithm, for example a sequential quadratic programming (SQP) or an interior point method [6]. An alternative is to use gradient free optimization methods. Mesh adaptive direct search (MADS) is one of these methods. MADS is capable of handling problems with nonsmooth objective function and constraints. Unfortunately as most derivative-free methods MADS converges slowly to a solution. To accelerate its convergence one can use quadratic models of the objective and of the constraints to compute promising points that speed up the algorithm's progress. The quadratic model can be built gradually by applying an update formula to the current approximation of the Hessian matrix.

## 3 Mesh Adaptive Direct Search

MADS is a family of algorithms where the steps the algorithm takes to explore the search space lie on a grid. In the presented algorithm the grid is defined as

$$
\begin{equation*}
\mathcal{G}_{k}=\left\{\sum_{i=1}^{n} \Delta_{k}^{m} \alpha_{i} \boldsymbol{e}_{i}: \alpha_{i} \in \mathbf{Z}\right\} \tag{8}
\end{equation*}
$$

where $D_{k}{ }^{m}$ denotes the mesh size parameter. The algorithm solves problems of the form

$$
\begin{equation*}
\min _{x \in \Omega \subseteq \mathfrak{K}^{n}} f(\boldsymbol{x}) \tag{9}
\end{equation*}
$$

where $\Omega=\left\{\boldsymbol{x}: \boldsymbol{x}^{L} \leq \boldsymbol{x} \leq \boldsymbol{x}^{H} \wedge c_{i}(\boldsymbol{x}) \leq 0, \mathrm{i}=1, \ldots, n_{c}\right\}$ denotes the set of feasible points. The lower and the upper bounds on the components of $\boldsymbol{x}$ are given by vectors $\boldsymbol{x}^{L}$ and $\boldsymbol{x}^{H}$, respectively. Nonlinear inequality constraints are defined by functions $C_{i}(\boldsymbol{x})$. For convenience the $n_{c}$ functions $c_{i}(\boldsymbol{x})$ are joined in a vector-valued function $c(x)$. The incumbent solution in the $k$-th iteration and the corresponding value of $f$ are denoted by $\boldsymbol{x}_{k}$ and $f_{k^{\prime}}$ respectively. Any point that is considered to be sufficiently good to replace the incumbent solution is referred to as an improving point. The initial point $\boldsymbol{x}_{0} \in \Omega$ corresponds to the first iteration ( $k=0$ ). MADS can handle constraints with the extreme barrier approach by replacing with $+\infty$, whenever $\boldsymbol{x} \notin \Omega$. Unfortunately this also requires the initial point to be feasible. Infeasible initial points can be handled by using a filter [7][8]. The filter approach decides whether a point can replace the incumbent solution by applying a bi-objective criterion based on the values of the objective and constraints at points evaluated in the past.

Algorithm 1: $k$-th iteration of the proposed algorithm based on the MADS framework.

1. Complete the quadratic model by computing the gradient of $f$ and the Jacobian of $\boldsymbol{c}$.
2. Make the model convex by replacing the Hessian $\boldsymbol{H}$ with $\boldsymbol{H}+\in \boldsymbol{I}, \in \geq 0$.
3. Compute $\boldsymbol{s}$ by solving the convex quadratic model and rounding the result to $V_{k}$.
4. Evaluate $f$ and $\boldsymbol{c}$ at $\boldsymbol{x}=\boldsymbol{x}_{k}+\boldsymbol{s}$. If $\boldsymbol{x}$ is an improving point, set $\boldsymbol{x}_{k+1}:=x$ and go to step 8 .
5. Generate the set of poll directions $\mathcal{D}_{k} \subseteq \mathcal{G}_{k}$.
6. Evaluate $f$ and $\boldsymbol{c}$ at $\boldsymbol{x}=\theta\left(\boldsymbol{x}_{k}+d\right)$ for $\boldsymbol{d} \in \mathcal{D}_{k}$. If $\boldsymbol{x}$ is an improving point set $\boldsymbol{x}_{k+1}:=\boldsymbol{x}$.
If the step resulting in an improving point was cut, go to step 8, else go to step 7. When $D_{k}$ is exhausted go to step 8.
7. Evaluate $f$ and $\boldsymbol{c}$ at $\boldsymbol{x}=-\theta\left(\boldsymbol{x}_{k}+2\left(\boldsymbol{x}-\boldsymbol{x}_{k}\right)\right.$ ). If $\boldsymbol{x}$ is an improving point, set $\boldsymbol{x}_{k+1}:=\boldsymbol{x}$.
8. If $\boldsymbol{x}_{k+1}=\boldsymbol{x}_{k^{\prime}}$ set $I_{k+1}:=I_{k}+1$; else if step 7 failed to produce an improving point, set $I_{k+1}:=I_{k^{\prime}}$;
else if $\boldsymbol{x}_{k+1} \neq \boldsymbol{x}_{k^{\prime}}$ the step resulting in $\boldsymbol{x}_{k+1}$ was not cut, and $I_{k}>0$ set $I_{k+1}:=I_{k}-1$; else set $I_{k+1}:=I_{k}$.

Steps 1-4, 5-6, and 7 of Algorithm 1 are also referred to in the MADS literature as the search, the poll, and
the speculative step, respectively. Set $\mathcal{D}_{k}$ is referred to as the set of scaled poll directions. The length of a scaled step is determined by the step size parameter $\Delta_{k}{ }^{\text {p }}$. Function $\theta$ maps points that violate bounds ( $\boldsymbol{x}^{L}$ and $\left.\boldsymbol{x}^{\mathrm{k}}\right)$ to points that satisfy them. A step is cut if $\theta(\boldsymbol{x}) \neq \boldsymbol{x}$. Although the proposed approach uses quadratic models the convergence properties of the MADS framework enable it to find a solution of the optimization problem even when the search step is omitted.

Refining subsequences are sequences of iteration indices $k \in \mathcal{K}$ for which $\Delta_{k}^{p} \rightarrow 0$. The MADS convergence theory applies to refining subsequences. More details can be found in [4] (extreme barrier approach) and [7] (filter-based approach).

Algorithm 1 differs from the basic MADS framework published in the literature ([4][7][9]) in several ways. The normalized poll directions are uniformly distributed on the unit sphere. The algorithm constructs a quadratic model of the objective function using a minimum Frobenius norm-based approach and a linear model of the constraints by means of regression. A quadratic programming solver then uses the model to compute a search step that accelerates the convergence. The point acceptance criterion in the search and the poll step is based on a filter instead on strict descent.

The algorithm that generates the ordered poll steps and the definition of function $\theta$ are the subject of section 3.1. The construction of the quadratic model and a more detailed description of the search step are given in section 3.2. The conditions under which a point is considered to be an improving point are given in section 3.3. The relation between the mesh index $\left(I_{k}\right)$, the mesh size, and the step size is the subject of section 3.4.

### 3.1 The poll step and the set of scaled poll directions

The poll step (steps 4-6 of Algorithm 1) is the one that guarantees the convergence properties of MADS [4][7]. The scaled poll directions $\boldsymbol{d} \in \mathcal{D}_{k}$ are ordered according to the angle they form with the last search (s) or poll (d) direction that resulted in an improving point [9]. The function and the constraints are evaluated at points $\boldsymbol{x}_{k}+\boldsymbol{d}$ corresponding to the ordered scaled poll directions. If $\boldsymbol{x}_{k}+\boldsymbol{d}$ violates any of the bounds imposed by $\boldsymbol{x}^{L}$ and $\boldsymbol{x}^{H}$ it is replaced by $\theta\left(\boldsymbol{x}_{k}+\boldsymbol{d}\right)$. Function $\theta$ modifies the components of $\boldsymbol{x}_{k}+\boldsymbol{d}$ that violate bounds by replacing them with the value of the corresponding violated bound. This has the effect of sliding the point along the violated boundary. The evaluation of points in the poll step is interrupted as soon as an improving point is found (greedy evaluation).

The set of scaled poll directions $\mathcal{D}_{k}$ is generated by applying an orthogonal transformation $\mathbf{0}_{t k} \in \mathcal{O}_{n}$ to $n+1$ vectors forming a regular $n$ simplex $v$ (cf. [10] on how to construct $v$ ) and scaling the resulting vectors with $\Delta_{k}{ }^{p}$. This results in set $\mathcal{U}_{k}=\left\{\Delta_{\mathrm{k}}{ }^{\mathrm{p}} \mathbf{O}_{t k} \boldsymbol{v}: \mathbf{v} \in \mathcal{V}\right\}$ whose members are rounded to the nearest points in $\mathcal{G}_{k}$ to obtain $\mathcal{D}_{k}$. Index $t_{k}$ plays a role in ensuring the convergence properties of the algorithm and will be discussed later.

The sequence of orthogonal transformations $\left\{\boldsymbol{0}_{i}\right\}_{i=0}$ is constructed by Algorithm 2 from a sequence of realizations of a random matrix with independent normally distributed random elements $\left\{\boldsymbol{N}_{i}\right\}_{i=0}$

Algorithm 2: Constructing a sequence of orthogonal transformations [11].

1. Apply QR decomposition to $\boldsymbol{N}_{i}$ resulting in $\boldsymbol{Q}_{i}$ and R.
2. Construct diagonal matrix $\boldsymbol{D}_{i}$ with $d_{i i}=1$ if $r_{i i} \geq 0$ and $d_{i i}=-1$ otherwise.
3. $\boldsymbol{Q}_{i}=\boldsymbol{Q}_{i} \boldsymbol{D}_{i}$.

Sequence $\left\{\boldsymbol{0}_{i}\right\}^{\infty}{ }_{i=0}$ is uniformly distributed (i.e. distributed according to the Haar measure on $\mathcal{O}_{n}$ [11]). Due to this the normalized vectors from the sequence of sets $\left\{\mathcal{U}_{i}\right\}^{\times}{ }_{i=0}$ are uniformly distributed (and consequently dense) on the unit sphere. Furthermore, if the mesh size parameter satisfies $\Delta_{k}{ }^{m} \rightarrow 0$ the union of sets $\mathcal{D}_{k}$ is also dense on the unit sphere (which is required by the MADS convergence theorem [4]) and the distribution of normalized poll directions converges to the uniform distribution on the unit sphere.

### 3.2 The quadratic model and the search step

MADS can be significantly improved if steps 1-4 of AIgorithm 1 examine points obtained by solving a model of the original optimization problem. In the presented method a quadratic model of the objective and a linear model of the constraints are constructed. The model can be formulated as

$$
\begin{align*}
& m_{f}(\boldsymbol{x})=\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{x}_{k}\right)^{T} \boldsymbol{B}\left(\boldsymbol{x}-\boldsymbol{x}_{k}\right)+\boldsymbol{g}^{T}\left(\boldsymbol{x}-\boldsymbol{x}_{k}\right)+f\left(\boldsymbol{x}_{k}\right)  \tag{10}\\
& m_{c}(\boldsymbol{x})=\boldsymbol{J}\left(\boldsymbol{x}-\boldsymbol{x}_{k}\right)+c\left(\boldsymbol{x}_{k}\right) \tag{11}
\end{align*}
$$

Where $\boldsymbol{B}, \boldsymbol{g}$, and $\boldsymbol{J}$ denote the approximate Hessian and the approximate gradient of the objective $f(\boldsymbol{x})$, and the approximate Jacobian of the constraints $\boldsymbol{c}(\boldsymbol{x})$, respectively. The model optimization problem can now be written as

$$
\begin{equation*}
\arg \min _{\substack{m_{c}(\boldsymbol{x}) \leq 0 \\ x^{\leq} \leq x \leq x^{H}}} m_{f}(\boldsymbol{x}) \tag{12}
\end{equation*}
$$

The approximate Hessian matrix is obtained by repeatedly applying an update formula. The initial Hessian approximation is set to an all-zero matrix. Every time the algorithm evaluates three collinear points $\boldsymbol{x}, \boldsymbol{x}+$ $\alpha \boldsymbol{p}$, and $\boldsymbol{x}+\alpha \boldsymbol{p}$ (i.e. after every speculative step that does not violate the bounds) the directional second derivative can be approximated as

$$
\begin{align*}
& \frac{d^{2} f(\boldsymbol{x}+t \boldsymbol{p})}{d t^{2}} \approx f_{p}=\frac{2}{\alpha_{+}-\alpha_{-}}\left(\frac{f\left(\boldsymbol{x}+\alpha_{+} \boldsymbol{p}\right)-f(\boldsymbol{x})}{\alpha_{+}}-\right.  \tag{13}\\
& \left.-\frac{f\left(\boldsymbol{x}+\alpha_{-} \boldsymbol{p}\right)-f(\boldsymbol{x})}{\alpha_{-}}\right)
\end{align*}
$$

Let $\boldsymbol{B}$ and $\boldsymbol{B}_{+}$denote the approximate Hessian and its updated value. When the second directional derivative is available the Hessian update formula from [12] can be used.

$$
\begin{equation*}
\boldsymbol{B}_{+}=\boldsymbol{B}+\left(f_{p}-\boldsymbol{p}^{T} \boldsymbol{B} \boldsymbol{p}\right) \frac{\boldsymbol{p} \boldsymbol{p}^{T}}{\|\boldsymbol{p}\|^{4}} \tag{14}
\end{equation*}
$$

It is more common the points are not collinear. In that case an update technique based on least Frobenius norm updating (LFNU) is used [13]. The proposed algorithm uses is a special case of LFNU for $n+2$ points. It is applied every time a new point is evaluated to update the Hessian of the objective $f$.

The linear part of the model is computed by means of linear regression [14]. Up to $2 n+1$ most recently evaluated points ( $\boldsymbol{x}$ ) satisfying $\left\|\boldsymbol{x}-\boldsymbol{x}_{k}\right\| \leq \rho \Delta_{k}^{p}$ are selected for regression. The regression computes vector $\boldsymbol{g}$ for which $m_{f}(\boldsymbol{x})$ is the closest fit to $f(\boldsymbol{x})$ at the selected points. Similarly the approximate Jacobian $J$ of the constraints is obtained by fitting $\boldsymbol{m}_{\boldsymbol{c}}(\boldsymbol{x})$ to $\boldsymbol{c}(\boldsymbol{x})$.

Whenever a quadratic model of the problem is successfully computed (i.e. the Hessian update and the linear regression are successful) it is used for ordering the scaled poll directions instead of the smallest angle criterion. The primary criterion for model-based direction ordering is the cumulative constraint violation computed as the sum of squares of positive components in vector $\boldsymbol{m}_{\boldsymbol{c}}(\boldsymbol{x})$. The secondary criterion is the value of the quadratic model $\boldsymbol{m}_{f}(\boldsymbol{x})$.

The obtained model is used for computing a trial point for the quadratic search step (step 1 of Algorithm 1). For this purpose problem (12) is solved using a quadratic program solver [15]. The solver can handle only positive definite Hessians matrices. Therefore $\boldsymbol{B}$ is replaced with $\boldsymbol{B}+\in \boldsymbol{I}$ and an additional constraint of the form $\left\|\boldsymbol{x}-\boldsymbol{x}_{k}\right\|_{\infty} \leq \rho_{-} \Delta_{k}^{p}$ is imposed whenever $\boldsymbol{B}$ is not positive definite. The value of $\in>0$ is chosen by repeatedly ap-
plying Cholesky decomposition to $\boldsymbol{B}+\in \boldsymbol{I}$ for increasing values of until the decomposition succeeds [6].

### 3.3 Point acceptance criterion

When the initial point $\boldsymbol{x}_{0}$ is feasible a point $\boldsymbol{x}$ can be considered as improving if it is feasible and $f(\boldsymbol{x})<f\left(\boldsymbol{x}_{k}\right)$. Often the initial point $\boldsymbol{x}_{0}$ is not feasible (i.e. $\boldsymbol{x}_{0} \notin \Omega$ ). Such a situation occurs when one tries to solve (4) to obtain the worst-case distance and chooses $\mathbf{0}$ as the initial value of the statistical parameters. In this case the nonlinear constraints cannot be handled with the extreme barrier approach. A possible alternative is to use a point acceptance criterion based on a filter [8].


Figure 2: The regions with acceptable (light gray) and dominating (dark gray) points for an optimization problem given by $f\left(x_{1}, x_{2}\right)=x^{2}{ }_{1}+x^{2}{ }_{2}$ (dashed contours) and $c\left(x_{1}, x_{2}\right)=-x_{1}-x_{2}+2$ (dotted contours). The points corresponding to the filter entries are marked by dark dots. The white dot marks the solution of the problem at $(1,1) . h_{\max }=2$.

The acceptance criterion based on a filter takes into account an improvement of the objective value, as well as an improvement of the feasibility. For that purpose a function is defined that expresses the constraint violation.
$h(\boldsymbol{x})=\sum_{i=1}^{n_{c}} \operatorname{ramp}\left(c_{1}(\boldsymbol{x})\right)$
For a feasible point the corresponding value of $h(\boldsymbol{x})$ is zero. A filter entry is a tuple of the form ( $f(\boldsymbol{x}), h(\boldsymbol{x})$ ). A filter is a set of mutually non-dominated filter entries. A tuple ( $f_{1}, h_{1}$ ) dominates ( $f_{2^{\prime}}, h_{2}$ ) if $f_{1} \leq f_{2^{\prime}} h_{1} \leq h_{2}$ and the two tuples are not equal. Initially the filter contains only $\left(f\left(\boldsymbol{x}_{0}\right), h\left(\boldsymbol{x}_{0}\right)\right.$ ). A point $\boldsymbol{x}$ is said to be


Figure 3: The filter entries (dark dots) and the solution (white dot) of the problem in Figure 2 in the $f$ - $h$ space. Dark gray and light gray regions correspond to dominating and acceptable points, respectively.

- dominating if the filter is empty or $(f(\boldsymbol{x}), h(\boldsymbol{x}))$ dominates at least one filter entry,
- dominated if at least one filter entry dominates ( $f(\boldsymbol{x}), h(\boldsymbol{x})$ ) or $h(\boldsymbol{x})>h_{\text {max }}$.
- acceptable otherwise.

Figure 2 and Figure 3 illustrate a 2-dimensional problem and a filter with 5 entries. Adding a point to the filter means that the corresponding filter entry ( $f(\mathbf{x}), h(\mathbf{x})$ ) is added to the filter. Dominating points and acceptable points are always added to the filter immediately after they are evaluated. The incumbent solution is always a member of the filter. Adding a dominating point implies that at least one dominated point is removed from the filter so that the filter entries remain mutually non-dominated. An acceptable point does not dominate any of the filter points and thus no points are removed from the filter when the corresponding entry is added. Dominated points are never added to the filter. If parameter $h_{\max }$ is set to 0, MADS behaves as if the extreme barrier approach had been used for handling the constraints.

For every filter point its position is defined by sorting the filter entries according to $h$. The filter entry corresponding to a feasible point is assigned position 0 (rightmost dark dot in Figure 2, leftmost dark dot in Figure 3), while infeasible filter entries are assigned integer positions starting from 1. A point examined by the search step is considered as an improving point if it is not dominated. A point examined in the poll step
and in the speculative step is considered as an improving point if it is not dominated and its position is not higher than the position of the incumbent solution. This effectively requires that the poll and the search step prioritize improving feasibility over improving the objective. When the incumbent solution is feasible these two steps behave as if the extreme barrier approach had been used.

### 3.4 Updating the mesh and the step size

Iterations of Algorithm 1 are assigned a mesh index $I_{k}$ with initial value $I_{0}=0$. The mesh and the step size parameter depend on $I_{k}$.
$\Delta_{k}^{m}=\min \left(1, \Delta^{-2 l_{k}}\right) /\left(\Delta_{0}[1+\gamma\rceil\right)$
$\Delta_{k}^{p}=\Delta^{-l_{k}}$
This strategy (see step 8 of Algorithm 1) refines the mesh and shortens the step when the algorithm is not making progress (i.e. fails to find an improving point). The mesh index is not changed if the speculative step fails to produce an improving point or if the improving point is obtained with a cut step. Otherwise the mesh is coarsened and the step size is increased, but never above its initial value.

Rounding can affect the set of unrounded scaled poll directions $U_{k}$ to such extent that $\mathcal{D}_{k}$ no longer positively spans $\mathbb{R}^{n}$. The effect of rounding is more pronounced when ratio $\Delta^{\mathrm{p}} / \Delta^{\mathrm{m}}{ }_{k}$ is small. Because
$\frac{\Delta_{k}^{p}}{\Delta_{k}^{m}}=\Delta_{0}\lceil 1+\gamma\rceil 2^{\left|l_{k}\right|} \geq \Delta_{0}[1+\gamma\rceil$
the aforementioned situation can be avoided if one chooses a sufficiently large. It can be shown that $\gamma=$ $n^{3 / 2} / 2$ is an appropriate choice for all $\Delta_{0} \geq 1$.

The normalized poll directions from a refining subsequence $\left\{\mathcal{D}_{k}\right\}_{k \in \mathcal{K}}$ must be dense on the unit sphere [4]. This is true if the refining subsequence corresponds to the complete sequence $\left\{\boldsymbol{N}_{i}\right\}^{\propto}$ i=0 ${ }^{\text {. }}$. sen in the following manner.
$t_{k}=\left\{\begin{array}{l}l_{k}, \text { for } l_{k} \geq \max _{i\langle k} l_{i} \\ 1+\max _{i\langle k} t_{i}, \text { othervise }\end{array}\right.$
Index $t_{k}$ increases from iteration to iteration with the exception of iterations that correspond to the finest observed mesh over $0 . . k$. As the mesh index of iterations forming a refining subsequence takes consecutive values from $\{0,1,2, \ldots\}$ the same values are also as-
signed to thereby causing $\left\{\mathcal{D}_{\mathcal{K}}\right\}_{\mathrm{k} \in \mathcal{K}}$ to correspond to the complete sequence $\left\{\boldsymbol{N}_{j}\right\}_{i=\sigma^{\infty}}$

## 4 Finding the worst-case point

The outline of the proposed approach comprises the same steps as [5]. The SQP-based optimization algorithm is replaced with the proposed version of MADS. The initial point in the space of statistical parameters is computed from the linearized optimization problem. An extended stopping criterion is proposed based on the approximate gradient of the circuit's performance.

In the beginning $\boldsymbol{x}_{s}=\mathbf{0}, \boldsymbol{x}_{0}$ is set to the nominal value of the operating parameters $\boldsymbol{x}_{0}^{\text {nom }}$, and the set of relevant statistical parameters is empty. The procedure for solving problem (4) and problem (7) is given by Algorithm 3.

Algorithm 3: One pass of the main algorithm for solving problem (4) / problem (7)
Solve $\boldsymbol{x}_{0}^{\omega 2}=\arg \min _{x_{0}^{L} \leq x_{0} \leq x_{0}^{H}} f_{i}\left(\boldsymbol{x}_{0}, \boldsymbol{x}_{S}\right)$. If $f_{i}\left(\boldsymbol{x}_{0}, \boldsymbol{x}_{0}^{w 2}\right)$, set $\boldsymbol{x}_{0}:=\boldsymbol{x}_{0}^{\omega 2}$.

1. Compute the approximate sensitivity of $f_{i}$ to statistical parameters.
2. Update the set of relevant statistical parameters and compute the initial point for step 4.
3. Solve (4) or (7) in the space of relevant statistical parameters to obtain the new value of $\boldsymbol{x}_{s}$.

In step 1 of Algorithm 3 the set of worst operating parameters is determined. The performance corresponding to $\left(\boldsymbol{x}_{0}, \boldsymbol{x}_{s}\right)$ is evaluated. Every operating parameter is perturbed to its respective upper and lower value resulting in the need to evaluate $2 n_{0}$ points by circuit simulation. The results are used for constructing the initial vector of operating parameters ( $\boldsymbol{x}_{0}{ }^{w 1}$ ). Every component of this vector is equal to the nominal value, the upper bound, or the lower bound of the corresponding operating parameter, whichever produced the worst value of $f_{i}$. The optimization in step 1 of Algorithm 3 is completed with the MADS algorithm as proposed in section 3 starting from $\boldsymbol{x}_{0}{ }^{w 1}$ and using the extreme barrier approach. Steps taken by the optimizer are scaled in such manner that a step of length 1 in direction of any operating parameter corresponds to $1 / 16$ of the difference between the upper and the lower bound.

The sensitivity to the statistical parameters (step 2 of Algorithm 3) is computed at ( $\boldsymbol{x}_{\alpha^{\prime}} \boldsymbol{x}_{s}$ ) using forward differences. The parameters are perturbed by $1 / 64$ of the difference between the lower and the upper bound (-10 and 10 , respectively). Let $\Delta x$ and $\Delta f_{i}$ denote the parameter perturbation and the corresponding difference in
the performance, respectively. The components of the gradient with respect to the statistical parameters $\left(\nabla_{s} f_{i}\right)$ can then be approximated as $\Delta f_{i} / \Delta x$.

The obtained sensitivity information is used for eliminating the statistical parameters that contribute little to the behavior of $f_{i}$ (step 3 of Algorithm 3). For this purpose the absolute performance differences $\left|\Delta f_{i}\right|$ are ordered and all parameters that contribute less than $1 \%$ to the total change of $f_{i}$ are removed in increasing order of contribution until the cumulative contribution of the removed parameters reaches $25 \%$ or there are no statistical parameters left. The remaining statistical parameters are added to the set of relevant statistical parameters.

Let $\boldsymbol{g}_{S}$ denote the estimated gradient in the space of statistical parameters. Components of the gradient not corresponding to relevant parameters are set to 0 . For problem (4) the initial point is obtained by updating $\boldsymbol{x}_{5}$ to
$\boldsymbol{x}_{S}-\frac{f_{i}\left(\boldsymbol{x}_{0}, \mathbf{0}\right)-G_{i}}{\left\|\boldsymbol{g}_{S}\right\|^{2}} \boldsymbol{g}_{S}$
For problem (7) $x_{s}$ is replaced with
$-\frac{\boldsymbol{g}_{S}}{\left\|\boldsymbol{g}_{S}\right\|^{2}} \beta$

MADS is then used for solving the main optimization problem (step 4 of Algorithm 3) in the space of statistical parameters. The value of $h_{\text {max }}$ is chosen as max (100, $\left.h\left(\boldsymbol{x}_{0}\right)\right)$ so that the initial point is always added to the filter. The scaling of parameters is the same as in step 1 of Algorithm 3. The main optimization in case of problem (7) is stopped if the constraint satisfaction condition $|c(\boldsymbol{x})|<\beta \varepsilon_{c}$ and the gradient angle condition $\angle\left(\nabla_{s} f_{\text {}}\right.$ $\left.\nabla_{s} c\right)<\varepsilon_{a}$ are satisfied (note that $c(\boldsymbol{x})$ is a scalar because the problem has only one nonlinear constraint). These two stopping conditions are applied only if the step satisfies $\Delta^{\mathrm{p}}<0.5$. The constraint satisfaction condition for problem (4) is formulated somewhat differently as $|c(\boldsymbol{x})|<3| | \boldsymbol{g}_{s}| | \varepsilon_{c}$. In the presented examples $\varepsilon_{c}=10^{-2}$ and $\varepsilon_{a}=15^{\circ}$ are used. Regardless of these stopping conditions MADS is stopped when $\Delta^{\mathrm{p}}$ drops below 0.01.

Algorithm 3 is repeated in multiple passes until the set of relevant statistical parameters remains unchanged in step 3 and the accepted solution in step 1 of Algorithm 3 does not change $f_{i}\left(\boldsymbol{x}_{\alpha^{\prime}} \boldsymbol{x}_{s}\right)$ by more than $1 \%$ compared to the difference between $f_{i}\left(\boldsymbol{x}_{0}^{\text {nom }}, \mathbf{0}\right)$ and $f_{i}$ $\left(\boldsymbol{x}_{\sigma^{\prime}} \mathbf{0}\right)$ from the first pass.

The following values of optimizer parameters were used: $\Delta=4, \Delta_{0}=2^{20}, \rho=\Delta^{2}, \rho_{-}=1$. For problem (7) the gradient of the constraint with respect to the statistical parameters can be expressed explicitly as $2 \boldsymbol{x}_{5}$ and is not computed numerically. Similarly for problem (4) the gradient and the Hessian of the objective can be expressed as $2 \boldsymbol{x}_{s}$ and $\boldsymbol{I}$ (identity matrix), respectively.

## 5 Application and verification of the approach

The proposed approach was implemented in the PyOPUS framework [16] and its performance was compared to that of a commercial worst-case analysis tool WiCkeD [5]. Both algorithms were tested on two circuits: a Miller operating transconductance amplifier (OTA) in Figure 4 and a folded cascode operating transconductance amplifier (FCOTA) in Figure 5.


Figure 4: Miller transconductance amplifier.


Figure 5: Folded cascode transconductance amplifier.
Both circuits have 3 operating parameters (temperature, supply voltage, and bias current). A mismatch model with two statistical parameters per transistor was used. Global variations of the manufacturing process were modeled with 10 statistical parameters. The circuits in Figure 4 (Figure 5) have 26 (42) statistical parameters. The results are listed in Table 1. The first and the second column list the names of the performances and their types (i.e. $f_{i}>G_{i}$ or $f_{i}<G_{j}$ ). The worst-case performances at $\beta=3$ obtained by solving problem (7) are listed in columns titled WC. The number of circuit eval-
uations and the number of algorithm passes are listed in the columns to the right of the WC column.

Problem (4) is solved with $G_{i}$ set to the WC value at $\beta=$ 3 obtained by WiCkeD (third column). The worst-case point obtained by solving this problem lies at $\left\|\boldsymbol{x}_{s}\right\|=3$. The obtained value of $b$ is listed in columns titled WCD and the number of circuit evaluations and algorithm passes is listed in the columns to the right of the WCD column.

The results in Table 1 show that the proposed approach is capable of finding the solution of problem (7) within 5\% accuracy. The two cases where the accuracy was worse than $5 \%$ are marked with an asterisk in the

WC column. The settling time (rise) of the Miller OTA was different due to the noise in the performance. In case of the PSRR VSS performance of the FCOTA circuit MADS converged to a different local minimizer. A more pessimistic worst case value was found by MADS in one case (shaded cell in the table). The number of circuit evaluations required by MADS was in 7 cases (marked with an asterisk) significantly worse than that required by WiCkeD. On the other hand in two cases MADS was significantly faster than WiCkeD (shaded cells in the table). On the remaining cases both algorithms exhibited similar performance.

Solving problem (4) is somewhat more challenging. The proposed approach found the same solution within $5 \%$

Table 1: Summary of the results obtained with WiCkeD and the proposed MADS-based algorithm. A WC/WCD value (the number of evaluations) that is more than $5 \%(20 \%)$ worse than the corresponding result obtained by WiCkeD is denoted by an asterisk.

|  |  | WiCkeD |  |  |  | MADS |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Circuit / Performance | type | WC | Evals | WCD | Evals | WC | Evals | Passes | WCD | Evals | Passes |
| Miller OTA |  |  |  |  |  |  |  |  |  |  |  |
| Swing [V] | > | 1.43 | 139 | 2.99 | 145 | 1.43 | 147 | 2 | 3.00 | *176 | 2 |
| Gain [dB] | > | 68.0 | 88 | 3.00 | 94 | 68.0 | 98 | 1 | 3.00 | 106 | 1 |
| UGBW [MHz] | $>$ | 1.61 | 93 | 3.00 | 100 | 1.61 | 98 | 1 | 3.02 | 116 | 1 |
| Phase margin [ ${ }^{\circ}$ ] | $>$ | 67.3 | 129 | 3.00 | 123 | 67.3 | *299 | 2 | 3.04 | *438 | 2 |
| CMRR [dB] | $>$ | 65.2 | 98 | 3.00 | 104 | 65.3 | *150 | 2 | 3.00 | *166 | 2 |
| PSRR VDD [dB] | $>$ | 85.0 | 124 | 3.00 | 112 | 85.3 | *396 | 3 | *3.21 | *398 | 3 |
| PSRR VSS [dB] | > | 61.0 | 92 | 3.00 | 98 | 61.0 | 98 | 1 | 3.00 | 106 | 1 |
| Settling $\downarrow$ [ $\mu \mathrm{s}$ ] | $<$ | 0.892 | 134 | 3.00 | 151 | 0.892 | 145 | 2 | 3.01 | 165 | 2 |
| Settling $\uparrow$ [ $\mu \mathrm{s}$ ] | $<$ | 1.04 | 108 | 3.00 | 116 | *1.03 | 102 | 1 | 3.00 | *195 | 2 |
| Slew $\downarrow$ [ $\mathrm{V} / \mu \mathrm{s}]$ | > | 1.10 | 94 | 3.00 | 96 | 1.10 | 99 | 1 | 3.00 | 115 | 1 |
| Slew $\uparrow[\mathrm{V} / \mu \mathrm{s}]$ | > | 0.953 | 461 | 3.06 | 196 | 0.960 | 101 | 1 | 3.09 | *260 | 2 |
| FCOTA |  |  |  |  |  |  |  |  |  |  |  |
| Offset (high) [mV] | < | 11.2 | 124 | 3.02 | 194 | 11.2 | *202 | 2 | 3.00 | 211 | 2 |
| Offset (low) [mV] | $>$ | -11.9 | 124 | 3.03 | 194 | 11.9 | *200 | 2 | 3.00 | 231 | 2 |
| Swing [V] | $>$ | 0.478 | 122 | 3.00 | 127 | 0.476 | 130 | 1 | 2.95 | 130 | 1 |
| Gain [dB] | $>$ | 70.7 | 125 | 3.03 | 131 | 70.8 | *291 | 2 | 3.02 | *332 | 2 |
| UGBW [MHz] | $>$ | 6.28 | 130 | 3.00 | 137 | 6.28 | 133 | 1 | 3.00 | 149 | 1 |
| Phase margin [ ${ }^{\circ}$ ] | $>$ | 85.6 | 222 | 3.00 | 227 | 85.6 | *368 | 2 | 3.01 | *493 | 3 |
| CMRR [dB] | $>$ | 60.8 | 290 | 3.06 | 265 | 60.8 | *460 | 2 | 3.00 | *456 | 2 |
| PSRR VDD [dB] | $>$ | 55.8 | 236 | 3.00 | 207 | 58.0 | 282 | 2 | 3.12 | *642 | 2 |
| PSRR VSS [dB] | > | 54.6 | 315 | 3.00 | 227 | *58.3 | 249 | 2 | *3.17 | *315 | 2 |
| IRN@100Hz [ $\mu \mathrm{V} / \sqrt{ } \mathrm{Hz}]$ | < | 3.17 | 126 | 3.00 | 133 | 3.17 | 133 | 1 | 3.01 | 151 | 1 |
| IRN@10kHz [ $\mu \mathrm{V} / \sqrt{ } \mathrm{Hz}]$ | < | 0.321 | 126 | 3.00 | 133 | 0.321 | 133 | 1 | 3.02 | 154 | 1 |
| IRN@1MHz [ $\mu \mathrm{V} / \mathrm{V} \mathrm{Hz}]$ | < | 59.4 | 126 | 3.03 | 132 | 59.3 | 133 | 1 | 3.00 | 144 | 1 |
| 1/f corner [kHz] | < | 437 | 122 | 3.00 | 128 | 436 | 143 | 1 | 3.01 | 147 | 1 |
| Settling $\downarrow$ [ $\mu \mathrm{s}]$ | $<$ | 0.131 | 142 | 3.00 | 148 | 0.131 | 135 | 1 | 3.00 | 146 | 1 |
| Settling $\uparrow$ [ $\mu \mathrm{s}$ ] | $<$ | 0.127 | 148 | 3.00 | 154 | 0.127 | 134 | 1 | 3.01 | 180 | 1 |
| Slew $\downarrow$ [ $\mu \mathrm{s}$ ] | > | 4.17 | 207 | 3.00 | 202 | 4.17 | 203 | 2 | 3.00 | *250 | 2 |
| Slew $\uparrow[\mu \mathrm{s}]$ | $>$ | 4.28 | 196 | 3.00 | 213 | 4.28 | 205 | 2 | 3.00 | 255 | 2 |

accuracy in all but two cases marked with an asterisk in the WCD column of Table 1. Both of them (as well as the PSRR VDD performance of FCOTA) are the result of convergence to a different local minimizer. In such cases a fair comparison with WiCkeD is not possible. When the number of circuit evaluations is considered both approaches exhibit similar performance on more than half of the performances. The cases where the proposed approach is significantly slower than WiCkeD are marked with an asterisk.

All optimization problems except for two are solved in one or two algorithm passes. Both MADS and WiCkeD face the same disadvantage originating from the local nature of the underlying optimization algorithms. Due to it the obtained solution can be a local minimizer and not the actual solution of problem (4) or (7) because the outcome greatly depends on the choice of the initial point.

MADS performs best on noisy nonlinear problems for which points exist where the function or the constraints cannot be evaluated (i.e. the simulator fails to converge or the circuit's performance cannot be evaluated). For such problems the finite difference approximation of the gradient cannot be computed and classical optimization methods like SQP used in commercial tools exhibit slow progress or fail. On these problems we expect MADS to outperform commercial gradientbased tools.

## 6 Conclusion

Finding the worst performance and the worst-case distance of a circuit's performance are important subproblems that arise in the process of automated integrated circuit sizing. The solution to these problems enables the designer to verify the satisfaction of the minimum yield requirement. This is an accurate and less costly alternative to yield estimation by Monte-Carlo analysis. An approach for solving both problems by means of MADS was presented. Several extensions were implemented in the general MADS framework that make it possible for the algorithm to rapidly close in on the solution of the optimization problem. The proposed algorithm was tested on two real world integrated circuit design problems. The results were compared to the results obtained with a commercial worst-case analysis tool (WiCkeD) that uses a gradient-based optimization algorithm. The results show the proposed approach is competitive with the approach used in WiCkeD.

## 7 Acknowledgements

The research was co-funded by the Ministry of Education, Science, and Sport (Ministrstvo za Šolstvo, Znanost in Šport) of the Republic of Slovenia through the programme P2-0246 Algorithms and optimization methods in telecommunications.

## 8 References

1. K. Papathanasiou, "A designer's approach to device mismatch: Theory, modeling, simulation techniques, scripting, applications and examples", Analog Integrated Circuits and Signal Processing, vol. 48, no. 2, pp. 95-106, 2006.
2. H. E. Graeb, "Analog design centering and sizing", Springer, 2007.
3. A. Singhee, R. A. Rutenbar (eds.), "Extreme Statistics in Nanoscale Memory Design", Springer, 2010.
4. C. Audet, J. E. Dennis, Jr., "Mesh adaptive direct search algorithms for constrained optimization", SIAM Journal on Optimization, vol. 17, no. 1, pp. 188-217, 2006.
5. MunEDA inc, "WiCkeD, a tool suite for nominal and statistical custom IC design", available at http://www.muneda.com/Products/, 2014.
6. J. Nocedal, S.Wright, "Numerical optimization", Springer, 2006.
7. C. Audet, J. E. Dennis, Jr., "A progressive barrier for derivative-free nonlinear programming", SIAM Journal on Optimization, vol. 20, no. 1, pp. 445472, 2009.
8. R. Fletcher, S. Leyffer, "Nonlinear programming without a penalty function", Mathematical Programming, vol. 91, no. 2, pp. 239-270, 2002.
9. C. Audet, A. Ianni, S. Le Digabel, C. Tribes, "Reducing the Number of Function Evaluations in Mesh Adaptive Direct Search Algorithms", SIAM Journal on Optimization, vol. 24, no. 2, pp. 621-642, 2014.
10. A. R. Conn, K. Scheinberg, L. N . Vincente, "Introduction to derivative-free optimization", SIAM, 2009.
11. G. W. Stewart, "The efficient generation of random orthogonal matrices with an application to condition estimators", SIAM Journal on Numerical Analysis, vol. 17, no. 3, pp. 403-409, 1980.
12. D. Leventhal, A. S. Lewis, "Randomized Hessian estimation and directional search", Optimization: A Journal of Mathematical Programming and Operations Research, vol. 60, no. 3, pp. 329-345, 2011.
13. M. J. D. Powell, "Least Frobenius norm updating of quadratic models that satisfy interpolation conditions", Mathematical Programming, vol. 100, no. 1, pp. 183-215, 2003.
14. A. L. Custódio, L. N. Vincente, "Using sampling and simplex derivatives in patters search methods", SIAM Journal on Optimization, vol. 18, no. 2, pp. 537-555, 2007.
15. M.S.Andersen, J. Dahl, L.Vandenberghe," ${ }^{\text {CVXOPT, }}$ Release 1.1.6", available at http://cvxopt.org/userguide/index.html, 2014.
16. "PyOPUS - Circuit Simulation and Optimization", available at http://fides.fe.uni-lj.si/pyopus/, 2014.

Arrived: 12.02. 2015
Accepted: 09.05. 2015

## MIDEM 2015

## 51s INTERNATIONAL CONFERENCE ON MICROELECTRONICS, DEIICES AND MATERIALS WITH THE WORKSHOP ON TERAHERTZ AND MICROWAVE SYSTEMS

## Announcement and Call for Papers

September $23^{\text {th }}-25^{\text {th }}, 2015$
Hotel Golf, Bled, Slovenia
ORGANIZER: MIDEM Society - Society for Microelectronics, Electronic Components and Materials, Ljubljana, Slovenia

CONFERENCE SPONSORS: Slovenian Research Agency; IMAPS, Slovenian Chapter; IEEE, Slovenian Section; Zavod TC SEMTO.

## GENERAL INFORMATION

The $51^{\text {th }}$ International Conference on Microelectronics, Electronic Components and Devices with the Workshop on Terahertz and Microwave Systems continues a successful tradition of the annual international conferences organised by the MIDEM Society, the Society for Microelectronics, Electronic Components and Materials. The conference will be held at Hotel Golf, Bled, Slovenia, wellknown resort and conference centre, from SEPTEMBER $\mathbf{2 3}^{\text {th }}-\mathbf{2 5}^{\text {th }}, 2015$.

## Topics of interest include but are not limited to:

- Workshop focus: Terahertz and Microwave Systems
- Novel monolithic and hybrid circuit processing techniques,
- New device and circuit design,
- Process and device modelling,
- Semiconductor physics,
- Sensors and actuators,
- Electromechanical devices, Microsystems and nano-


## systems,

- Nanoelectronics
- Optoelectronics,
- Photonics,
- Photovoltaic devices,
- New electronic materials and applications,
- Electronic materials science and technology,
- Materials characterization techniques,
- Reliability and failure analysis,
- Education in microelectronics, devices and materials.


## ABSTRACT AND PAPER SUBMISSION:

Prospective authors are cordially invited to submit up to 1 page abstract before May $\mathbf{1 s t}^{\text {st, 2015 }}$. Please, identify the contact author with complete mailing address, phone and fax numbers and e-mail address.
After notification of acceptance (June 15 ${ }^{\text {th }}$, 2015), the authors are asked to prepare a full paper version of six pages maximum. Papers should be in black and white. Full paper deadline in PDF and DOC electronic format is: August 31 ${ }^{\text {st }}, 2015$.

## IMPORTANT DATES:

- Abstract deadline: May 1 ${ }^{\text {st, }} 2015$ (1 page abstract or full paper)
- $\quad$ Notification of acceptance: June $\mathbf{1 5}^{\text {th }}, 2015$
- Deadline for final version of manuscript: August 31 ${ }^{\text {st }}, 2015$

Invited and accepted papers will be published in the conference proceedings.

Deatailed and updated information about the MIDEM Conferences is available at
http://www.midem-drustvo.si/ under Conferences.


# Boards of MIDEM Society Organi društva MIDEM 

MIDEM Executive Board | Izvršilni odbor MIDEM<br>President of the MIDEM Society | Predsednik društva MIDEM<br>Prof. Dr. Marko Topič, University of Ljubljana, Faculty of Electrical Engineering, Slovenia<br>Vice-presidents | Podpredsednika<br>Prof. Dr. Barbara Malič, Jožef Stefan Institute, Ljubljana, Slovenia<br>Dr. Iztok Šorli, MIKROIKS, d. o. o., Ljubljana, Slovenija<br>\section*{Secretary|Tajnik}<br>Olga Zakrajšek, UL, Faculty of Electrical Engineering, Ljubljana, Slovenija<br>MIDEM Executive Board Members | Člani izvršilnega odbora MIDEM<br>Prof. Dr. Slavko Amon, UL, Faculty of Electrical Engineering, Ljubljana, Slovenia Darko Belavič, In.Medica, d.o.o., Šentjernej, Slovenia<br>Prof. Dr. Bruno Cvikl, UM, Faculty of Civil Engineering, Maribor, Slovenia<br>Prof. DDr. Denis Đonlagič, UM, Faculty of Electrical Engineering and Computer Science, Maribor, Slovenia Prof. Dr. Leszek J. Golonka, Technical University Wroclaw, Poland Leopold Knez, Iskra TELA d.d., Ljubljana, Slovenia<br>Dr. Miloš Komac, UL, Faculty of Chemistry and Chemical Technology, Ljubljana, Slovenia<br>Prof. Dr. Miran Mozetič, Jožef Stefan Institute, Ljubljana, Slovenia Jožef Perne, Zavod TC SEMTO, Ljubljana, Slovenia<br>Prof. Dr. Giorgio Pignatel, University of Perugia, Italia<br>Prof. Dr. Janez Trontelj, UL, Faculty of Electrical Engineering, Ljubljana, Slovenia

Supervisory Board | Nadzorni odbor
Prof. Dr. Franc Smole, UL, Faculty of Electrical Engineering, Ljubljana, Slovenia Mag. Andrej Pirih, Iskra-Zaščite, d. o. o. , Ljubljana, Slovenia Dr. Slavko Bernik, Jožef Stefan Institute, Ljubljana, Slovenia

## Court of honour | Častno razsodišče

Emer. Prof. Dr. Jože Furlan, UL, Faculty of Electrical Engineering, Slovenia Prof. Dr. Radko Osredkar, UL, Faculty of Computer and Information Science, Slovenia Franc Jan, Kranj, Slovenia

Informacije MIDEM
Journal of Microelectronics, Electronic Components and Materials
ISSN 0352-9045

Publisher / Založnik:
MIDEM Society / Društvo MIDEM
Society for Microelectronics, Electronic Components and Materials, Ljubljana, Slovenia Strokovno društvo za mikroelektroniko, elektronske sestavne dele in materiale, Ljubljana, Slovenija


[^0]:    * Corresponding Author's e-mail: gorecki@am.gdynia.pl

[^1]:    * Corresponding Author's e-mail: predrag.petrovic@ftn.kg.ac.rs

[^2]:    * Corresponding Author's e-mail: snesko@uns.ac.rs

[^3]:    * Corresponding Author's e-mail: skl@ua.pt

[^4]:    // OCM address (see [1] for details)
    // DDR address (see [1] for details)
    // GPP address (see [1] for details)
    // for this example OCM address is chosen

[^5]:    * Corresponding Author's e-mail: arpadb@fides.fe.uni-l.j.si

