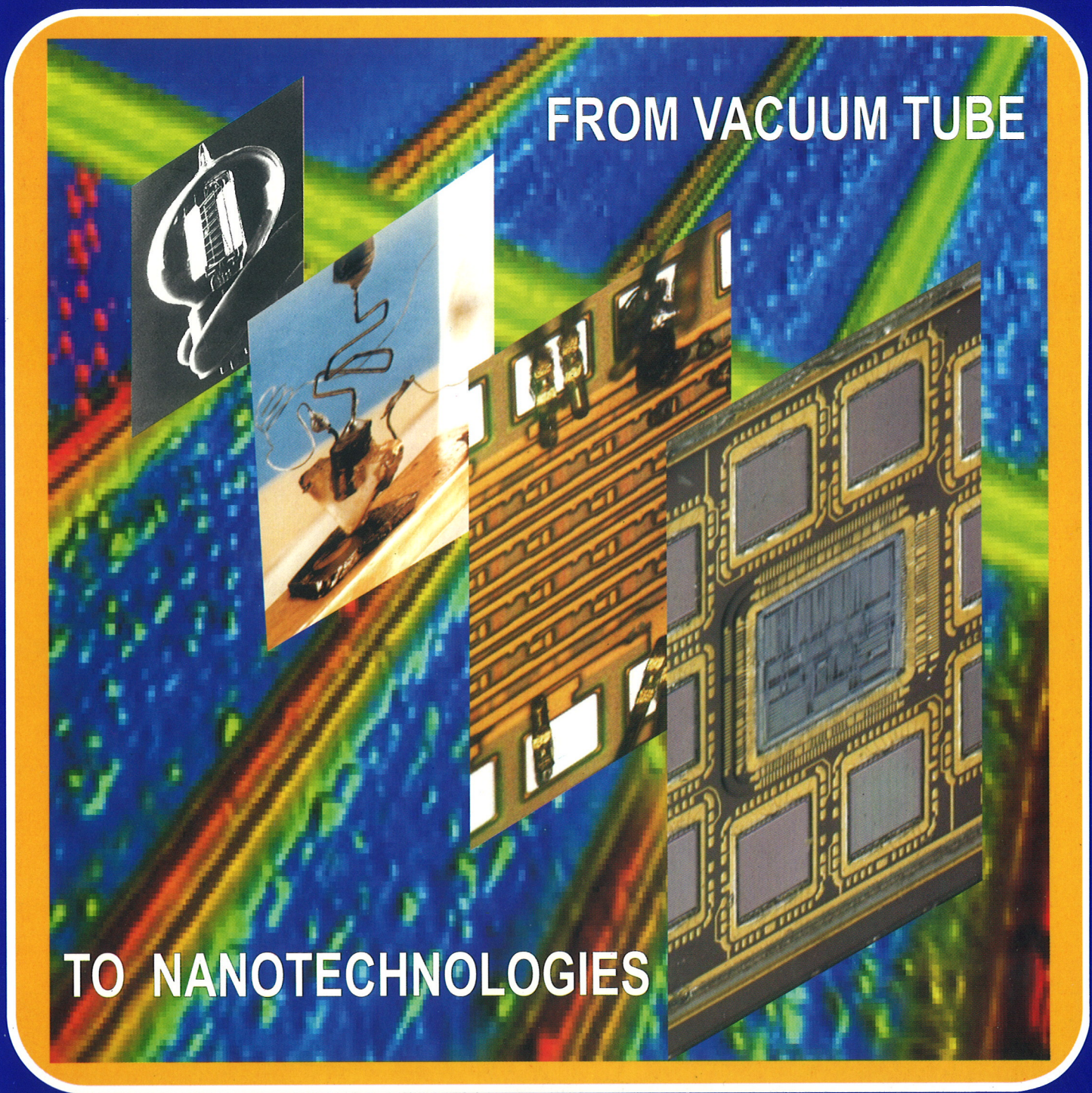
UDK 621.3:(53+54+621+66)(05)(497.1)=00 ISSN 0352-9045

INFORMACIJE MIDEM 1·2007

Strokovno društvo za mikroelektroniko elektronske sestavne dele in materiale

Strokovna revija za mikroelektroniko, elektronske sestavne dele in materiale
Journal of Microelectronics, Electronic Components and Materials

INFORMACIJE MIDEM, LETNIK 37, ŠT. 1(121), LJUBLJANA, marec 2007



UDK 621.3:(53+54+621+66)(05)(497.1)=00 ISSN 0352-9045

INFORMACIJE MIDEM 1 ◦ 2007

INFORMACIJE MIDEM LETNIK 37, ŠT. 1(121), LJUBLJANA, MAREC 2007
INFORMACIJE MIDEM VOLUME 37, NO. 1(121), LJUBLJANA, MARCH 2007

Revija izhaja trimesečno (marec, junij, september, december). Izdaja strokovno društvo za mikroelektroniko, elektronske sestavne dele in materiale - MIDEM.
Published quarterly (march, june, september, december) by Society for Microelectronics, Electronic Components and Materials - MIDEM.

Naslov uredništva
Headquarters
Uredništvo Informacije MIDEM
MIDEM pri MIKROIKS
Stegne 11, 1521 Ljubljana, Slovenija
tel.: + 386 (0)1 51 33 768
faks: + 386 (0)1 51 33 771
e-pošta: Iztok.Sorli@guest.arnes.si
http://www.midem-drustvo.si/

Letna naročnina je 100 EUR, cena posamezne številke pa 25 EUR. Člani in sponzorji MIDEM prejemajo Informacije MIDEM brezplačno.
Annual subscription rate is EUR 100, separate issue is EUR 25. MIDEM members and Society sponsors receive Informacije MIDEM for free.

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 ARRS in sponzorji društva.
Scientific Council for Technical Sciences of Slovene Research Agency has recognized Informacije MIDEM as scientific Journal for microelectronics, electronic components and materials.
Publishing of the Journal is financed by Slovene Research Agency and by Society sponsors.

Znanstveno-strokovne prispevke objavljene v Informacijah MIDEM zajemamo v podatkovne baze COBISS in INSPEC.

Prispevke iz revije zajema ISI® v naslednje svoje produkte: Sci Search®, Research Alert® in Materials Science Citation Index™

Scientific and professional papers published in Informacije MIDEM are assessed into COBISS and INSPEC databases.

The Journal is indexed by ISI® for Sci Search®, Research Alert® and Material Science Citation Index™

Po mnenju Ministrstva za informiranje št.23/300-92 šteje glasilo Informacije MIDEM med proizvode informativnega značaja.

Grafična priprava in tisk
Printed by
BIRO M, Ljubljana

Naklada
Circulation
1000 izvodov
1000 issues

Poštnina plačana pri pošti 1102 Ljubljana
Slovenia Taxe Percue

UDK621.3:(53+54+621+66), ISSN0352-9045 Informacije MIDEM 37(2007)1, Ljubljana

VSEBINA **CONTENT**

Obnovitev članstva v strokovnem društvu MIDEM in iz tega izhajajoče ugodnosti in obveznosti

Spoštovani,

V svojem več desetletij dolgem obstoju in delovanju smo si prizadevali narediti društvo privlačno in koristno vsem članom.Z delovanjem društva ste se srečali tudi vi in se odločili, da se v društvo včlanite. Življenske poti, zaposlitev in strokovno zanimanje pa se z leti spreminjajo, najrazličnejši dogodki, izzivi in odločitve so vas morda usmerili v povsem druga področja in vaš interes za delovanje ali članstvo v društvu se je z leti močno spremenil, morda izginil. Morda pa vas aktivnosti društva kljub temu še vedno zanimajo, če ne drugače, kot spomin na prijetne čase, ki smo jih skupaj preživeli. Spremenili so se tudi naslovi in način komuniciranja.

Ker je seznam članstva postal dolg, očitno pa je, da mnogi nekdanji člani nimajo več interesa za sodelovanje v društvu, se je Izvršilni odbor društva odločil, da stanje članstva uredi in **vas zato prosi, da izpolnite in nam pošljete obrazec priložen na koncu revije.**

Naj vas ponovno spomnimo na ugodnosti, ki izhajajo iz vašega članstva. Kot član strokovnega društva prejemate revijo »Informacije MIDEM«, povabljeni ste na strokovne konference, kjer lahko predstavite svoje raziskovalne in razvojne dosežke ali srečate stare znance in nove, povabljene predavatelje s področja, ki vas zanima. O svojih dosežkih in problemih lahko poročate v strokovni reviji, ki ima ugleden IMPACT faktor.S svojimi predlogi lahko usmerjate delovanje društva.

Vaša obveza je plačilo članarine 25 EUR na leto. Članarino lahko plačate na transakcijski račun društva pri A-banki : 051008010631192. Pri nakazilu ne pozabite navesti svojega imena!

Upamo, da vas delovanje društva še vedno zanima in da boste članstvo obnovili. Žal pa bomo morali dosedanje člane, ki članstva ne boste obnovili do konca leta 2006, brisati iz seznama članstva.

Prijavnice pošljite na naslov:

MIDEM pri MIKROIKS

Stegne 11

1521 Ljubljana

Ljubljana, marec 2007

Izvršilni odbor društva

ON-RESISTANCE OF POWER MOSFETS

Janusz Zarębski

Gdynia Maritime University, Department of Marine Electronics

Key words: MOSFET, silicon carbide parameters, R_{ON} resistance.

Abstract: The paper concerns the problem of modelling of the drain-to-source ON-Resistance (R_{ON}) of power MOSFETs. Two kinds of the transistor structures: VDMOS and CoolMOS made of a silicon and a silicon-carbide are considered in the paper.

Upornost močnostnih MOSFET tranzistorjev v prevodnem stanju

Kjučne besede: MOSFET, parametri silicijevega karbida, upornost R_{ON}

Izvleček: V prispevku prikažemo probleme modeliranja upornosti izvor – ponor, R_{ON}, močnostnih MOSFET tranzistorjev v prevodnem stanju. Opišemo dve tranzistorski strukturi: VDMOS in CoolMOS izvedeni v siliciju, oz. silicijevem karbidu.

1. Introduction

Currently, silicon power MOSFETs are one of the most intensive developed and modified devices intended mainly for power converters.

The substantial drawback of the considered class of power devices is their relatively high value of the drain-to-source ON-Resistance (R_{ON}), what often results in unacceptable values of the energy losses at the high values of the drain current. The value of R_{ON} depends on both the values of the device breakdown voltage and the semiconductor parameters. The decreasing of R_{ON}, especially in the high-voltage power MOSFETs, is the very important challenge for the producers of these devices. This task can be reached by developing new structures (e.g. CoolMOS transistors), as well as by using advanced (wide band-gap) semiconductor materials, e.g. silicon-carbide (SiC).

The paper presents the estimation of influence of the impurity doping of the epitaxial-layer of the selected power MOS transistors (VDMOS, CoolMOS) on their drain-to-source ON-Resistance. The considerations are performed for two semiconductors: silicon and the most popular polytypes of silicon carbide.

2. The theoretical dependences

Power silicon MOSFETs (Fig. 1) are commonly used in electronics and energoelectronics in the voltage range from about fifteen up to one thousand volts. The main component of the switch-ON resistance (R_{ON}) of high-voltage VDMOSTs is the resistance represented by the epitaxial layer. The fundamental relation of R_{ON} on the breakdown voltage (U_{BR}) is /1/.

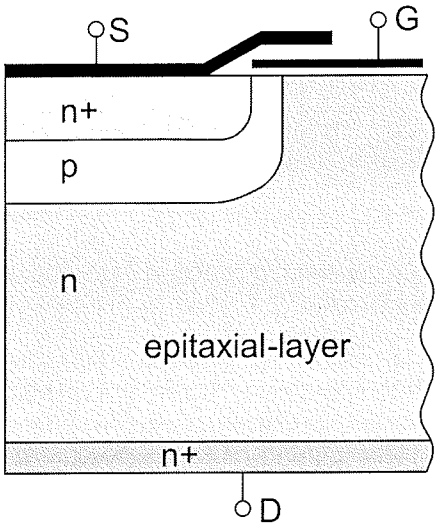


Fig. 1. The considered structure of the VDMOS transistor

$$R_{ON} = \frac{4 \cdot U_{BR}^2}{\varepsilon_0 \cdot \varepsilon_r \cdot \mu \cdot E_c^3} \qquad (1)$$

Where: μ - electron mobility, $\acute{I}_0$ - the permittivity of free space, $\acute{I}_r$ - relative permittivity, E_C - critical electric field are a semiconductor parameters.

It is seen, that R_{ON} increases strongly with increasing of U_{BR} (to the second power) whereas it decreases for semiconductors characterized by higher values of μ and E_C (third power).

On the other hand, at a fixed value of the epitaxial-layer thickness, the higher voltage VDMOS transistor have to be lightly doped (it concerns the epitaxial layer) according to the dependence /1/.

$$U_{BR} = \frac{\varepsilon_0 \cdot \varepsilon_r \cdot E_c^2}{2 \cdot q \cdot N} \qquad (2)$$

where N - doping concentration, q - elementary (electron) charge. In turn, lower doping concentration results in higher value of R_{ON}.

The new generation of power MOSFETs - named CoolMOS transistors (Fig. 2) described for the first time in /2, 3/, have been offered by Infineon Technologies since 1998.

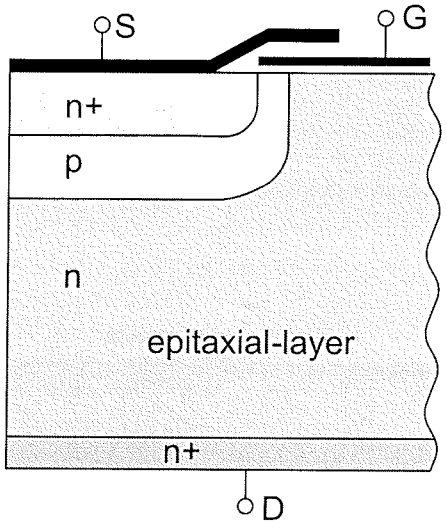


Fig. 2 The considered structure of the CoolMOS transistor

CoolMOS transistors (Fig. 2) belong to the class of superjunction devices presented in /4/. In CoolMOS transistors the epitaxial layer consists of parallely connected heavily doped n and p pillars of the width equal to d. At the drain-to-source voltage, typically higher than 30 ÷ 50 volts, the pillars become fully depleted by lateral extension of the depletion region from the p-pillar/n-pillar junction. On the other hand, the pillar-doping concentration depends on the pillar width. The thinner epitaxial-layer, the higher doped one.

The resistance R_{ON} and the breakdown voltage U_{BR} of the CoolMOS transistor are related by the following dependence /4/.

$$R_{ON} = 4 \cdot d \cdot \frac{U_{BR}}{\varepsilon_0 \cdot \varepsilon_r \cdot \mu \cdot E_c^2} \qquad (3)$$

As seen, R_{ON} resistance is proportional to U_{BR}, what means that CoolMOS transistors are especially attractive for high-voltage applications. On the other hand, narrower pillars, the lower R_{ON}, what have to result in the heavily dopped ones.

3. The semiconductors parameters

As seen from Eqs. 1 and 3, two semiconductor parameters: charge mobility and critical electric field are of the fundamental importance to determine the R_{ON}- resistance. In Table 1 the values of these parameters for silicon (Si) and three of the most popular polytypes of silicon carbide: 4HSiC, 6H-SiC and 3C-SiC are presented.

Parametr	Si	3C-SiC	4H-SiC	6H-SiC
$\mu\left[\frac{cm^2}{V \cdot s}\right]$	1400	900	700	400
$E_C\left[\frac{V}{cm}\right]$	3.10^5	$1{,}5.10^6$	$2.6.10^6$	$2{,}1.10^6$
ε_r	11,9	9,7	9,6	10

The values in Table 1 concern lightly doped semiconductors. As results from a lot of publications, assuming the fixed values of the considered parameters is a great simplification, because their values are a function of temperature, doping concentration, pressure, etc. Fig. 3 presents the dependences of $\mu(N)$ and $E_C(N)$ for the silicon and silicon carbide, respectively.

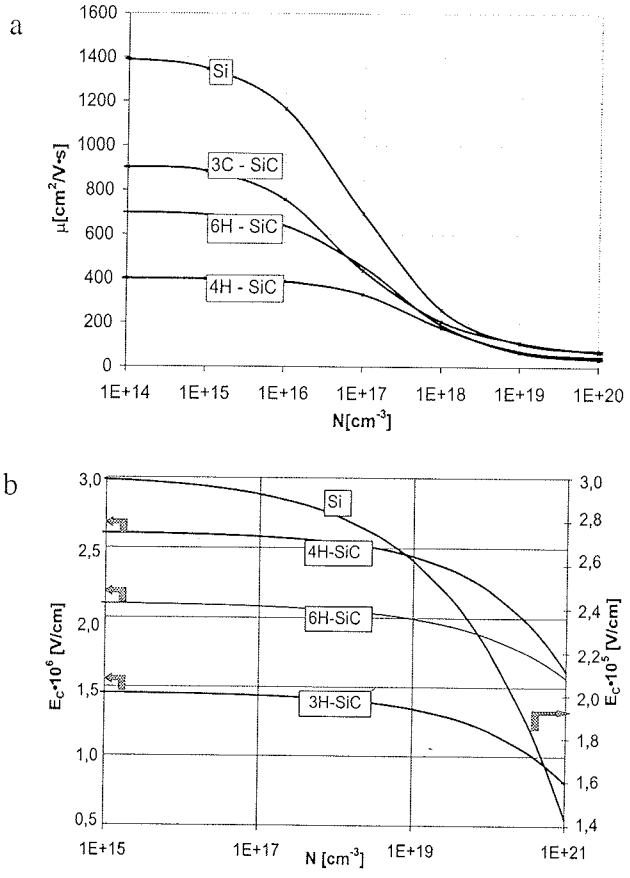


Fig. 3. Dependencies μ(N) and EC(N)

As seen from Fig. 3 the value of the electron mobility is nearly the proper low-doped value for the doping concentration less than any critical value N_{Cr}, which for silicon (N_{CrSi}) is equal to 10^{14} cm^{-3}, whereas for silicon-carbide is much higher, e.g. for 4H-SiC we have $N_{Cr4HSiC} = 10^{16}$ cm^{-3}. The analogical observation (Fig. 3b) concerns the critical field, for which $N_{CrSi} = 10^{15}$ cm^{-3} and $N_{Cr6HSiC}$ H" 10^{18} cm^{-3}, respectively.

4. Results

On the parameter values from Table 1, the dependences of $U_{BR}(N)$ for VDMOS transistors made of silicon (Si) and silicon-carbide (SiC) were calculated from Eq. 2 (Fig. 4). In turn, the dependence of $R_{ON}(U_{BR})$ for silicon and silicon-carbide VDMOS and CoolMOS transistors at three various values of the pillars width, corresponding to the same materials parameter values are presented in Fig. 5.

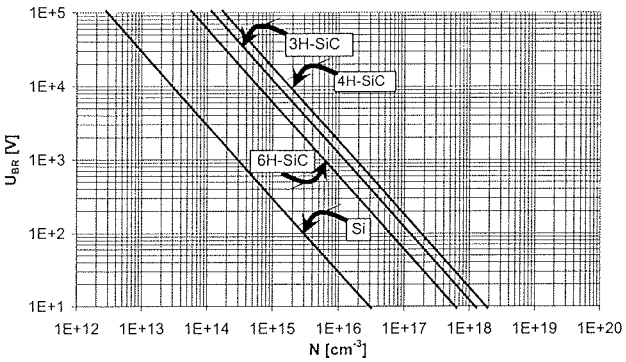


Fig. 4. Dependencies of UBR on N

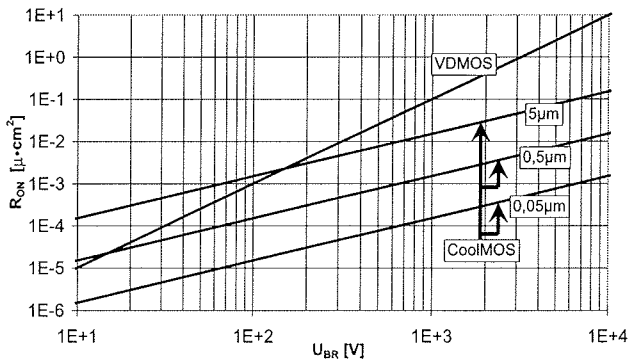
It is seen from $U_{BR}(N)$ characteristics that for silicon transistors of the breakdown voltage higher than 20V, the doping concentration not exceed 10^{16} cm^{-3}, whereas the SiC transistors of the same U_{BR} value can be even one hundred times more heavy doped (N = 10^{18}).

The current technology allows to manufacture the silicon CoolMOS transistors of the pillar width equal to 5 μm. For examples, for the 1 kV CoolMOS transistor, assuming that $d/2 \sim 1/\sqrt{N}$ and d = 5 μm, the doping concentration can increase 10^5 times. Note, that the tenfold degreasing of the pillar with (d = 500 nm) results in the further hundredfold increasing of N doping until the value equal to 10^{21} cm^{-3} (for Si). As it is seen from Figs 4,5 the values of E_C and μ decrease considerably at the doping concentration $N > N_{Cr}$. So, this phenomenon should be taken into account, when the R_{ON} resulting from Eqs. (1,3) is calculated. For instance, in Fig 6 the dependence of $\Delta R_{ON}/R_{ON}$ as a function of U_{BR} of the form

$$\frac{\Delta R_{ON}}{R_{ON}} = \frac{\left[R_{ON}^{const.} - R_{ON}^{var}\right]}{R_{ON}^{const.}} \cdot 100\% = f(U_{BR}) \qquad (4)$$

for Si and SiC VDMOS transistors is presented. In this equation the resistances $R_{ON}^{const.}$ and R_{ON}^{var} are calculated with the

a Si



b 4H-SiC

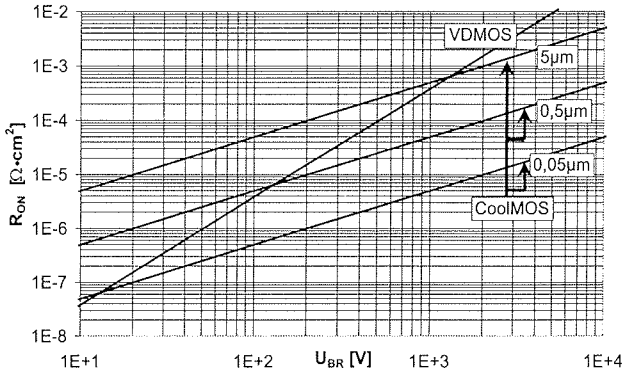


Fig. 5. Dependencies of RON on UBR

use of the semiconductor parameters of the constant values (table I) and the values obtained from Fig. 3, respectively. To get the proper value of the doping concentration N corresponding to the selected value of the device breakdown voltage, the diagram shown in Fig.4 was used.

As seen, using the dependence $E_C(N)$ and μ(N) results in the higher values of the ON-resistance, what is of the great importance, especially for silicon-carbide devices.

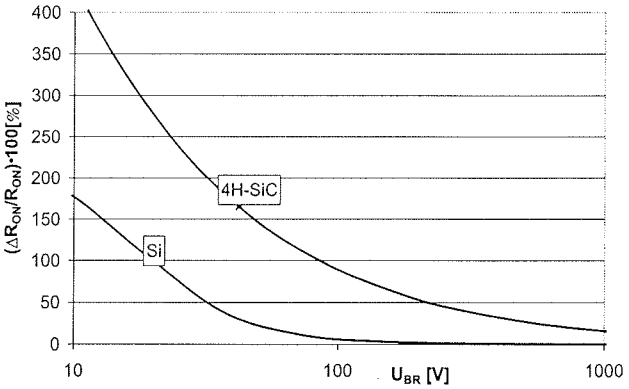


Fig. 6 Dependencies of ÄRON on UBR for VDMOS transistors

5. Conclusion

In the paper the problem of modelling of the drain-to-source ON-Resistance of silicon and silicon carbide power MOSFETs is discussed. Two kinds of transistors: VDMOS and CoolMOS, made of silicon and silicon carbide are considered. It was shown that ě(N) and $E_C(N)$ dependences should be taken into consideration while the R_{ON} resistance of the investigated power MOS transistors is calculated.

References

/1/ Baliga J.: "Power Semiconductor Device Figure of Merit for High-frequency Applications, IEEE Electr. Dev. Lett., Vol. 10, 1989, p. 455.

/2/ Lorenz L., Deboy G., Marz M., Stengel J., Backhofner A.: "Drastic Reduction ON-Resistance with CoolMOS", PCIM Europe ISSUE 5/1998, pp 250-258.

/3/ Deboy G., Marz M., Stengel J.P, Strack, Tihanyi J., Webber H.: "A New Generation of High Voltage MOSFETs Breaks the Limits of Silicon", IEDM Tech, Dig., p. 683.

/4/ Fujihira T.: „Theory of Semiconductor Super-junction Devices", Jpn, J. Appl. Phys., Vol. 36, pp. 6254-626

Prof. Janusz Zarębski
Gdynia Maritime University
Department of Marine Electronics
Morska 83, 81-225 Gdynia, POLAND,
Tel. ++48 58 6901599, fax ++48 58 6217353
E-mail: zarebski@am.gdynia.pl

Prispelo (Arrived): 10.08.2006 *Sprejeto (Accepted):* 30.03.2007

LOW-POWER, HIGH-SPEED SRAM DESIGN: A REVIEW

Tan Soon-Hwei, Loh Poh-Yee, Mohd-Shahiman Sulaiman, Zubaida Yusoff

Faculty of Engineering, Multimedia University, 63100 Cyberjaya, Selangor, Malaysia

Key words: CMOS, SRAM, Low-Power, High-Speed, Dual-Port, Asynchronous

Abstract: This article describes the challenges in the design of low-power SRAM and difficulties in designing a high-speed static RAM. Following an overview of general issues, approaches/techniques in achieving low-power SRAM, as well as techniques to increase SRAM operating frequency are described. Following the overview of each approach, the challenges and trade-offs are outlined.

Načrtovanje hitrih vezij SRAM z majhno porabo - pregled

Kjučne besede: CMOS, SRAM, nizka poraba, velika hitrost, dvojni vhod, asinhronski

Izvleček: V prispevku opišemo izzive pri načrtovanju vezij SRAM z nizko porabo in težave pri načrtovanju hitrih pomnilnikov RAM. Po uvodnem pregledu splošnih pojmov opišemo tehnike in pristope k načrtovanju vezij SRAM z nizko porabo, kakor tudi tehnike za povečanje frekvence delovanja vezij SRAM. Po opisu obeh pristopov naštejemo njune izzive in kompromise.

1 Introduction

Static Random Access Memory (SRAM) is mainly used as an embedded block (EBB) memory circuitry such as level 1 and level 2 caches in a microprocessor operating based on the Principle of Locality or used as a data buffer at a chip's interface. SRAM occupies a large portion of the modern digital chips and its capacity is forecasted to further increase in the new era of System on Chip (SoC). Increased SRAM density results in larger internal capacitance and thus limits the operating frequency and power consumption. Besides that, leakage current is now one of the major contributors to chip's overall power consumption. As static power continues to dominate, appropriate measures have to be taken to minimize the effects of leakage currents, especially in short channel devices.

This article summarizes power savings and high-speed techniques that can be implemented into SRAM design to overcome speed degradation and high power dissipation issues. The next section of this article addresses general issues in designing SRAM with large memory capacity, with example referring to the implementation of a 1-Mbit SRAM.

2 General Considerations

The general power consumption equation for SRAM is /1/:

$$P = V_{dd}[mI_{active} + m(n-1)I_{hold}] + V_{dd}[(n+m)C_{de}V_{\text{int}}f] + V_{dd}[C_{pt}V_{\text{int}}f + I_{dcp}] \quad (1)$$

First term of the equation dominates SRAM power consumption nowadays. I_{active} is active current of m selected memory cells while I_{hold} is the data retention current of others unselected memory cells. Second term represents power consumption of decoders and is the power needed for switching internal node capacitance. Third term represents power consumption of peripheral circuitries. It consists of dc current and current needed for switching capacitance. In general, active power depends on the switching capacitance and direct-path current while data retention power depends on the leakage current (sub-threshold leakage current - main contributor and reverse-biased current) and size of memory.

SRAM's memory operation delay usually consists of word-line decoding delay, data sensing or data writing delay and delay for resetting circuitry to initial condition. These delays are mostly proportional to the size of memory and also depend on the techniques that adopted into the design.

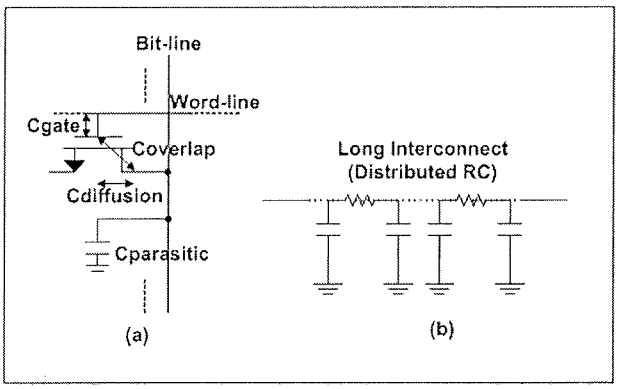


Figure 1 (a) Transistor's Parasitic
(b) Transmission line's Parasitic

Technology scaling and larger memory capacity contributes to larger bit-line capacitance and resistance as longer interconnect is required and more transistors are connected to a common line (Refer to Figure 1). Therefore, charging or discharging the line capacitance to transmit voltage signal causes huge active power consumption and signal propagation delay. In addition, larger number of memory cells raises the leakage power and decoding de-

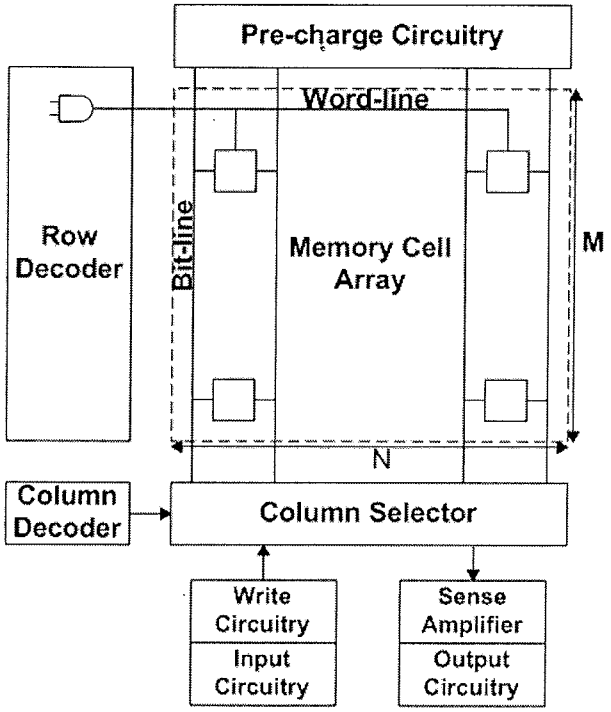


Figure 2 General SRAM Architecture

lay as well. The leakage power has been foreseen to dominate SRAM power in the Nano-range process technology as threshold voltage is scaled down along with the transistor's channel length. Other than that, designers have to make sure that the power race conditions in circuit are handled well during design phase to diminish the crowbar current.

2.1 Low-Power and High-Speed SRAM Architecture

Figure 2 is the general architecture that utilizes two way decoding method. If a 1-Mbit SRAM is going to be designed with 32-bit word size and 16:1 column muxing, it will have physical size of 2048 rows x 512 columns. The huge capacitive load on bit-lines and word-lines (large C_{de} and C_{pt}) causes high power and speed degradation problem. Instead, memory partitioning and bit-line partitioning techniques can be employed during the architecture-level design phase.

The principle in memory partitioning technique is to partition memory array into several portions and to map these portions to different physical memory banks that can be selected or deselected independently. The word-lines are divided into several sub levels and only certain sub word-lines are activated during operation. As a result, the number of transistors that are connected to the bit-lines and word-lines is lesser and hence smaller capacitance switched during operation. Also, it reduces active current I_{active} by shutting down portions that are not accessed potentially. Indirectly, it reduces the RC delay of decoding stage and word-line as well. Divided Word Line (DWL) architecture (Refer to Figure 3a) and Hierarchical Word Decoding (HWD) architecture (Refer to Figure 3b) are two of the most famous memory partitioning techniques that are proposed by Masahiko Yoshimoto et al. /2/ and Toshihiko Hirose et al. /3/ respectively.

DWL architecture requires lesser number of decoding stages and hence lesser number of transistors and smaller decoding delay. Furthermore, it is easier to be adopted into the design especially for layout design as routing and placement is simpler than the HWD architecture. However, the active power and decoding delay of DWL architecture exceeds HWD architecture when the memory capacity is above 256-Kbit due to larger number of multi-divided blocks that raises the load capacitance on global word-lines while keeping the number of selected memory cells small /3/. When the size of SRAM exceeds 1-Mbit, HWD architecture should be adopted into the design instead /4/. However, the number of portions should be constrained as excessively large number of portions imposes area overhead and wiring overhead that tends to increase power dissipation. Hence, HWD architecture might not be able to help in the near future as memory capacity is kept increasing. Divides the word-line into higher number of levels is no more a feasible solution as the decoding delay is increased as well.

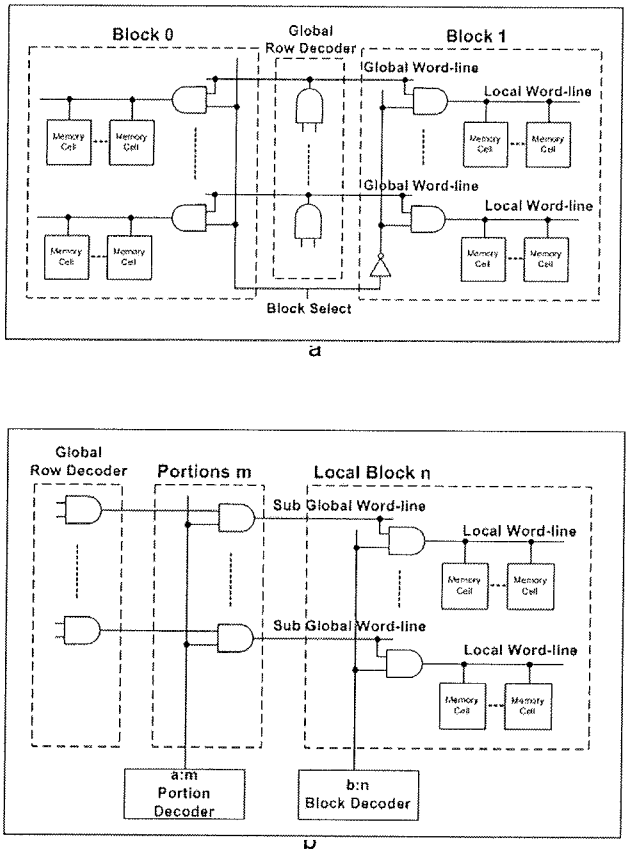


Figure 3 (a) Divided Word-Line (DWL) Architecture (b) Hierarchical Word Decoding (HWD) Architecture

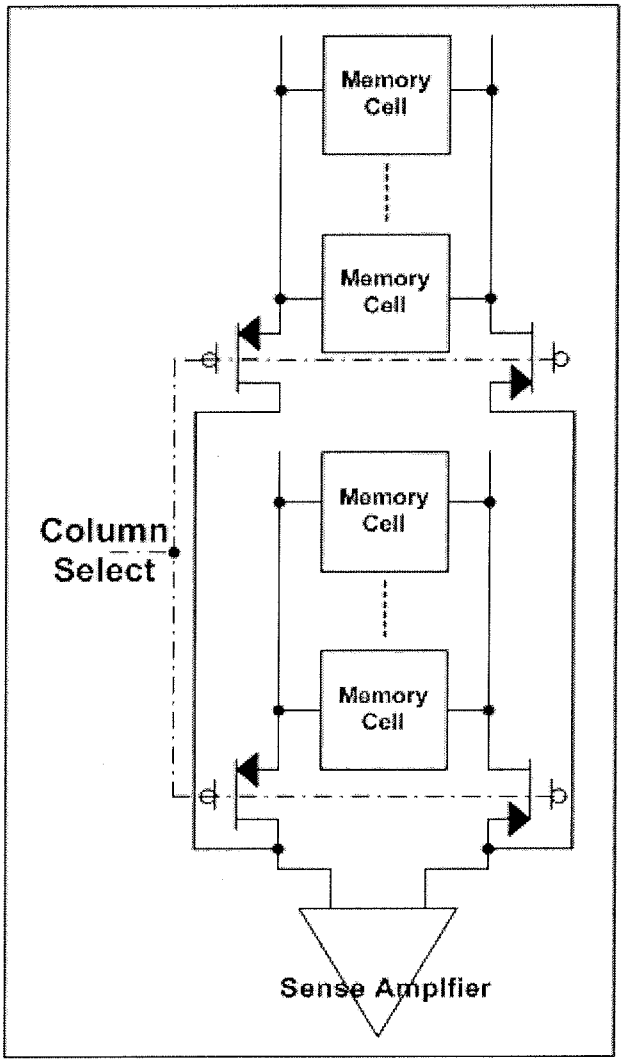
Bit-line partitioning technique is used to reduce bit-line capacitance and hence lower active power and access delay at the expenses of slightly larger number of transistors. It can be implemented in two ways, either using multiple-level muxing (Refer to Figure 4a) or divided bit-line approach (Refer to Figure 4b) proposed by Karandikar Ashish et al. /5/. Multiple-level muxing method increases node capacitance at the output of pass-transistors and hence number of the bit-line partition level should be limited. Besides, the load on column select signal line becomes larger due to longer wire routing and more transistors are connected. Divided bit-line approach reduces the bit-line capacitance to achieve lower active power and smaller RC signal delay. Besides that, decoding stage can be made smaller for this approach and thus reduces the decoding delay. Although the total current driving force of memory cell becomes smaller due to additional pass transistor in between the global bit-line and local bit-line, it can achieve better performance as long as the number of divided blocks and number of memory cells per a divided block are constrained. Theoretically, the optimal number of memory cells per a divided block is around 8. Sizing of the global bit-line pass transistor has to be large enough to pull down the bit-line voltage. It also has to be smaller than the size when the drain capacitance added to global bit-line exceeds the current driving capability.

3 Low Power and High Speed SRAM Circuit Design Techniques

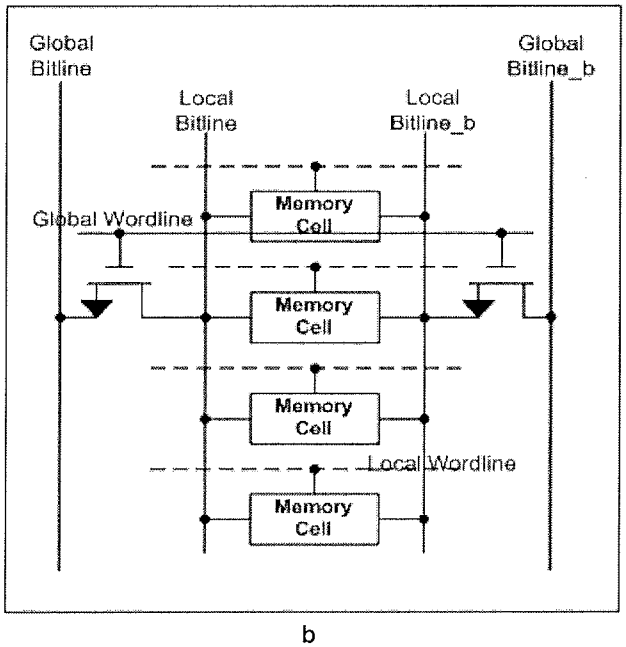
ISSCC2005 mentioned that semiconductor industry begins to shift from synchronous SRAM design to asynchronous design to solve power consumption problem. If external clock signal is used to derive internal control signals, pulse width of the derived signals might be too large for an operation to complete. For instance, bit-lines voltage can be pulled down to a level more than sufficient for a memory operation (Refer to Figure 5a). Besides, some internal circuits are turned on every cycle yet no operation is requested. As a result, asynchronous SRAM design using Address Transition Detection (ATD) circuitry is invented to reduce dc current, active current and switching capacitance by supplying pulse signals to the internal circuitries /6/. A simple ATD circuit together with the pulsed-signal waveform is shown in Figure 5b.

A pulse is generated whenever there is a transition in the input signals. The delay element circuit is designed such that pulse duration will be just enough for an SRAM operation to complete. Modified Schmitt trigger delay circuit that has lesser delay variation and hot-electron effect can be used /7/. However, this technique is still not widely used by the industry due to large variation of generated pulse across process, voltage and temperature (PVT) corners.

Voltage-mode sense amplifier and write circuitry usually cause large bit-line voltage swing and hence power dissipation problem as $P_{discharge} = C_{bl}\left[dV_{swing} / dt_{operation}\right] \cdot V_{dd}$.



a



b

Figure 4 (a) Multiple-level MUXING Method (b) Divided Bit-line Approach

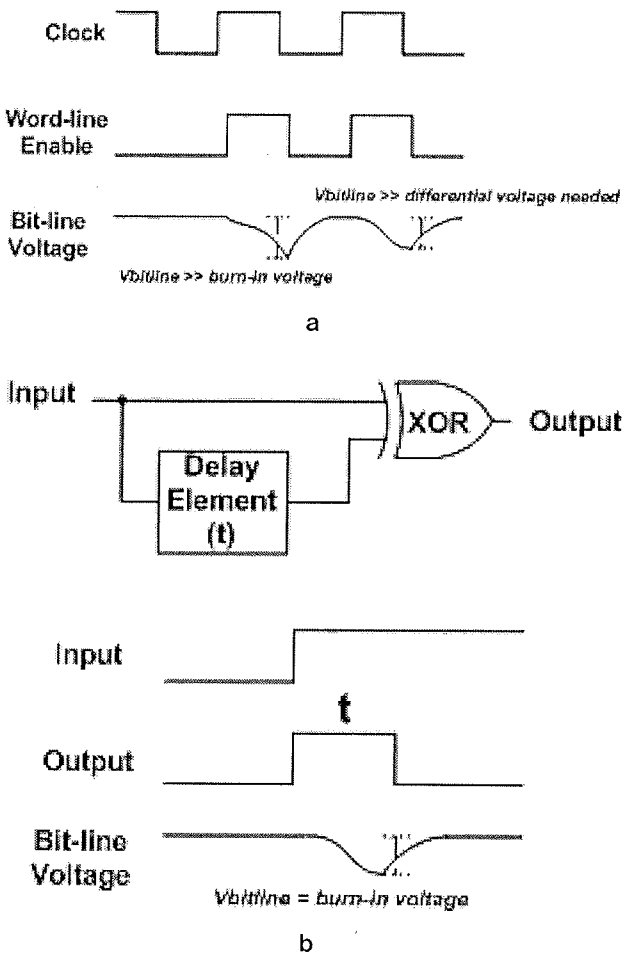


Figure 5 (a) Bit-line Over-discharged
(b) ATD and Output Pulsed-Signal

Also, speed degradation is another problem when the bit-line parasitic is large. Usually SRAM cell current is small as it is designed to have minimum-sized transistors and thus bit-line development time is long. Other than that, another long period is required to restore bit-lines voltage level after every memory operation. Current-mode technique can be implemented to avoid these problems. It was first proposed by Evert Seevinck et al. in 1991 /8/. Its main idea is to use a low-resistance current-signal circuit to reduce voltage swing on the long transmission line and to avoid speed degradation problem caused by signal delay in the high capacitive interconnect. Current-mode technique can be implemented for both sense amplifier and write circuitry. Figure 6a shows the bit-line peripheral circuit that has implemented full current-mode technique. Figure 6b and 6c illustrate the current-mode read (Clamped Data-line Sense Amplifier) and write circuitry /9/. It is found that the delay of sensing and writing process is insensitive to the bit-line or data-line capacitance /4/,/10/. In addition, it helps to reduce coupling noise to the adjacent bit-lines as bit-lines voltage swing is minimized. However, the current-mode circuit itself causes large crowbar current and direct-current path also exist between pre-charge circuit and sense amplifier during operation. Thus the circuit has to

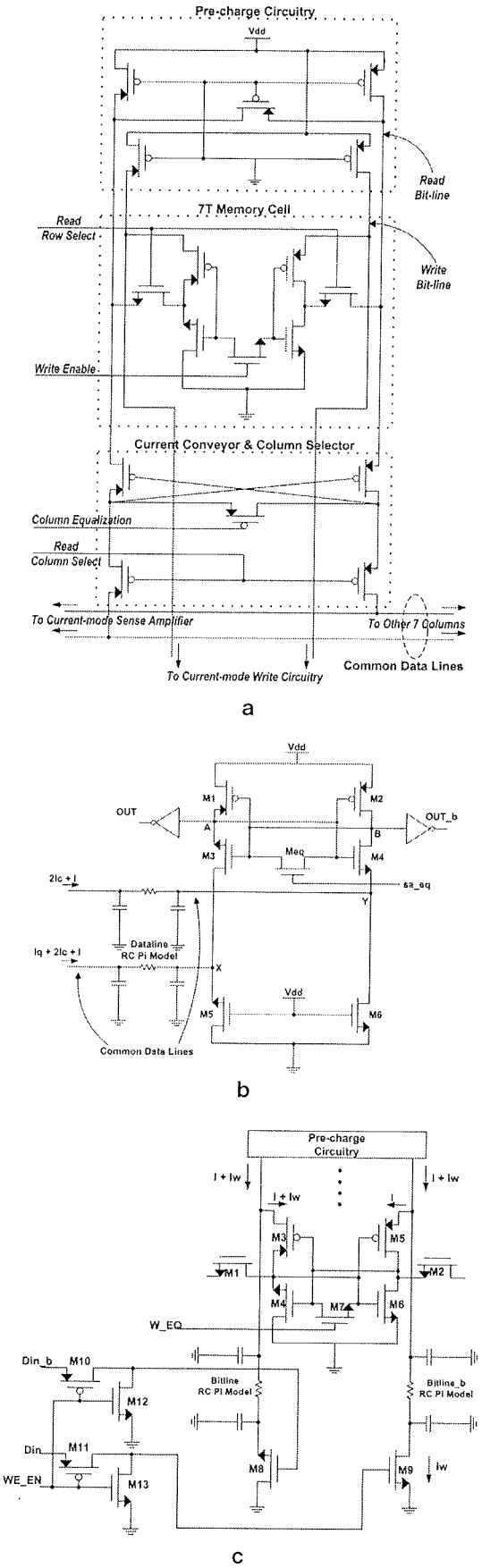


Figure 6 (a) Bit-line Peripheral Circuit
(b) Clamped Data-line Sense Amplifier
(c) Current-mode Write Circuitry

be properly designed or else it will offset the advantage of reducing capacitance switched during operation. It can be minimized by using dynamic load controlling circuit which has lesser effects to the speed of current-mode circuit. The load circuit is turned when there's no operation to reduce the current flowing. Besides that, it also requires additional transistors for memory cells that increase the die size.

Charge transfer sense amplifier (CTSA) allows low power sensing without sacrificing the speed (Refer to Figure 7). It operates based on charge re-distribution mechanism between very high bit-line capacitance and low output capacitance of the sense amplifier /11/. According to the charge conservation equation: $\Delta V_{bitline} \cdot C_{bitline} = C_{out} \cdot V_{out}$, bit-line swing depends on the ratio between bit-line and output capacitance where larger bit-line capacitance leads to lower bit-line swing and hence lower power. Besides, this circuit diminishes the direct-path current as the pre-charge circuit can be turned off after the word-line is enabled. Unlike current-mode sense amplifier that requires the pre-charge circuit to be turned on.

There're also several techniques proposed purposely to reduce writing delay and write power. There is a low power write technique proposed by Alowersson that biases the bit-lines voltage closed to low voltage level /12/. It first builds up a small differential voltage for the bit-lines and utilizes the regenerative effect of cross-coupled PMOSs to amplify small differential storage nodes voltage of the memory cell. However, this technique causes slow read operation as word-lines voltage are required to be reduced to prevent spurious discharge of memory cell storage nodes. Another low swing write technique proposed by Amrutur /13/ uses boosted voltage for word-line signal and virtual ground line (Refer to Figure 8). After the virtual ground line is driven high to clear memory cell's content, small differential voltage between the bit-lines is built up and this small differential signal is transferred into the storage node. The memory cell amplifies this small differential signal after virtual ground line is driven low. This technique has several problems such as voltage drop of virtual ground line, boosted voltage generation circuit is required and memory block width is required to be the same as input data size.

In the decoder design, designer is required to choose suitable gate logic style and then finds the optimal sizing for each gate and adding appropriate buffers to reduce decoding delay. Logical effort is one of the popular techniques in finding out optimal sizing for a chain of logic gates. Either NMOSs or PMOSs in the logic gates can be sized such that only one of the signal's transitions is enhanced and the opposite transition is weakened /14/. This helps to reduce the input capacitance of logic gates and thus overall address decoding delay. However, such implementation needs separate reset devices to prevent slow reset transition. These reset devices uses self-resetting logic (SR-CMOS) technique and delayed reset logic (DRCMOS) technique.

Another low power and high speed approach - half-swing pulse-mode technique uses half-swing signals to reduce the power in decoder and write circuitry /15/-/16/. Half swing technique limits the signal swing on the high capacitance bit-lines and I/O lines. In addition, the charge cycling between positive and negative pulses further reduces the power. The power dissipation is $P_{half-swing} = C_{bitline} \cdot (V_{dd}^2/2) \cdot f$ compares to the power of full swing voltage-mode circuits which is $P_{full-swing} = 2C_{bitline} \cdot V_{dd}^2 \cdot f$ and hence it can save 75% power theoretically. The problem with this technique is it uses half-swing pulse-mode gate family that requires dual-threshold or multi-threshold voltage transistors which might not be supported by certain process technology.

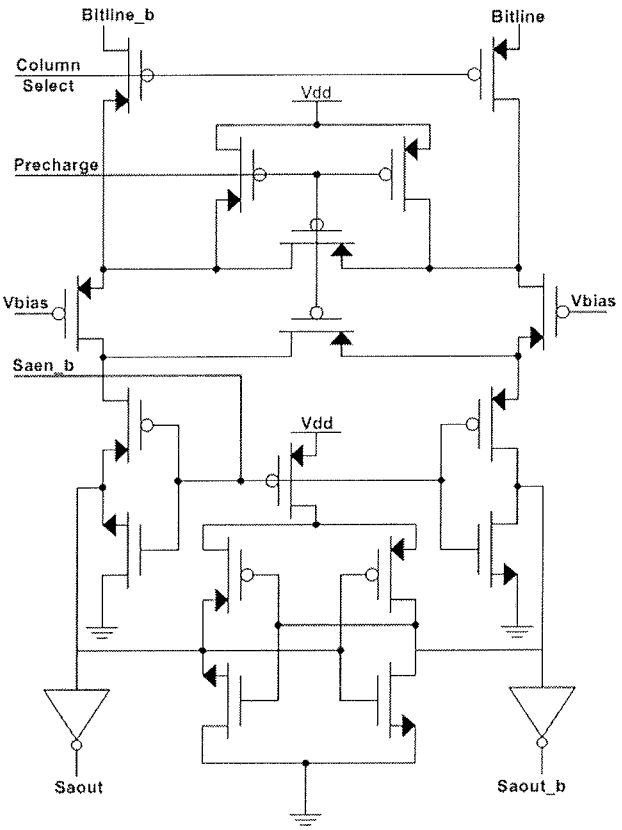


Figure 7 Charge Transfer Sense Amplifier (CTSA)

Supply voltage and threshold voltage are both lowered along with the technology scaling. The immediate effect is higher static power dissipation as the main contributor of leakage current - sub-threshold leakage is increased exponentially (Refer to the sub-threshold drain current I_{Dsub} equation in BSIM3v3 MOSFET model: $I_{Dsub} = I_{s0} \cdot \left[1 - e^{-V_{ds}/V_t}\right] \cdot \left[e^{\left[V_{gs} - V_T - V_{off}\right]/nV_t}\right]$). A new driving scheme which lowers the word-line voltage and raises the ground line voltage is proposed to reduce leakage current in standby mode /17/. Also, Gated-Vdd/Gated-Ground circuit techniques are proposed by Michael Powell et al. and Amit Agarwal et al. to reduce leakage power during standby mode /18/-/19/. The main idea is to disconnect

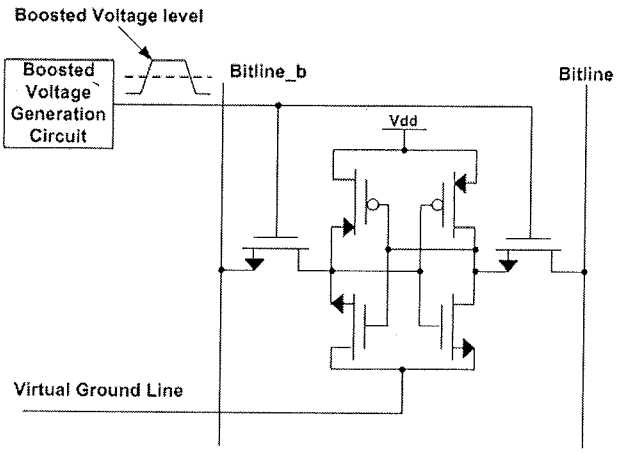


Figure 8 Low Swing Write Technique Proposed by Amrutur

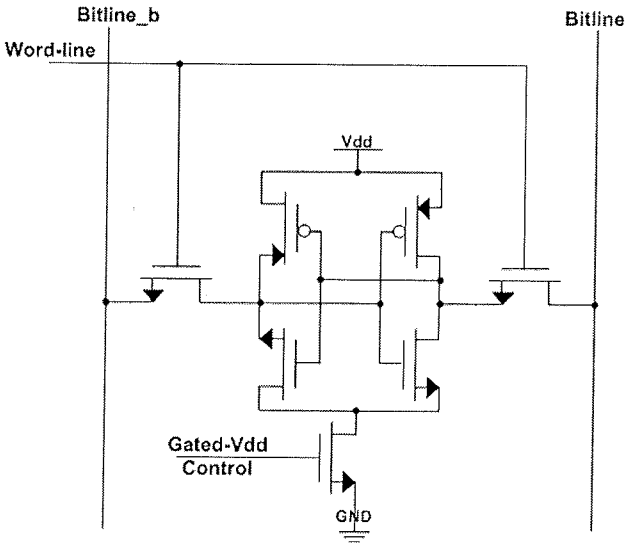


Figure 9 Gated-Vdd/Gated-Gnd Technique for Reducing Leakage

the SRAM cell circuit from supply voltage during standby mode and thus diminish the leakage current. However, an additional NMOS is needed to be put in the leakage path and this increases the layout of memory cell (Refer to Figure 9). Although gated-Vdd NMOS can be shared by multiple memory cells, but the transistor size has to be large enough to sink the current during active mode. Larger size improves the operation speed but worsen the power and area. Other than that, this technique needs no extra circuitry as the decoder itself can be used to control the gated-Vdd NMOS. On-chip Voltage Down Converter (VDC) circuit can be adopted to step down the supply voltage during standby mode /20/. This can greatly reduce the data retention power at the expenses of extra efforts to design the additional circuits and current consumption of the VDC.

4 Conclusion

Low-power high-speed SRAM design must deal with many circuit and architecture issues, including process limitations. The article has focused on difficulties in the design of low-power SRAM as well as the design of high-speed SRAM, presenting a number of power reduction techniques design approaches to increase SRAM's operating frequency.

References

/1/ Kiyoo Itoh et al., "Trends in Low Power RAM Circuit Technologies", *IEEE Symp. Low Power Elec. Dig. of Tech. Papers*, pp. 84-87, Oct. 1994.

/2/ Masahiko Yoshimoto et al., "A Divided Word-Line Structure in the Static RAM and Its Application to a 64K Full CMOS RAM", *IEEE J. Solid-State Circuits*, vol. 18, pp. 479-484, Oct 1983.

/3/ Toshihiko Hirose et al., "A 20-ns 4-Mb CMOS SRAM with hierarchical word decoding architecture", *IEEE J. Solid-State Circuits*, vol. 25, pp. 1068-1074, Oct 1990.

/4/ Tan Soon Hwei, "A 160-MHz, 45-mW, Asynchronous Dual-Port 1-Mb CMOS SRAM", *Final Year Project Report - FOE Multimedia University*, April 2005.

/5/ Karandikar Ashish et al., "Low Power SRAM Design using Hierarchical Divided Bit-Line Approach", *Proc. International Conference on Computer Design*, pp. 82-88, Oct 1998.

/6/ Martin Margala, "Low-Power SRAM Circuit Design", *IEEE In. Workshop on Memory Tech., Design and Testing*, pp. 115-122, Aug. 1999.

/7/ Akinori Sekiyama et. al, "A 1-V Operating 256-kb Full CMOS SRAM", *IEEE J. Solid-State Circuits*, vol. 27, pp. 776782, May 1992.

/8/ Evert Seevinck et al., "Current-mode Techniques for High-speed VLSI Circuits with Application to Current Sense Amplifier for CMOS SRAMs", *IEEE J. Solid-State Circuits*, vol. 26, pp. 525-536, April 1991.

/9/ Jinn-Shyan Wang et al., "Low-Power Embedded SRAM with the Current-Mode Write Technique", *IEEE J. Solid-State Circuits*, vol. 35, pp. 119-124, Jan 2000.

/10/ H. Wang et al., "A Low Power Current Sensing Scheme for CMOS SRAM", *IEEE In. Workshop on Memory Tech. Design & Testing*, pp. 37-45, Aug. 1996.

/11/ Manoj Sinha et al., "High-Performance and Low-Voltage Sense-Amplifier Techniques for sub-90nm SRAM", *Proc. International Conference on SOC*, pp. 113-116, Sept. 2003.

/12/ Jonas Alowersson et al., "SRAM Cells for Low-Power Write in Buffer Memories", *IEEE Symposium on Low Power Electronics*, pp. 60-61, Oct. 1995.

/13/ Bharadwaj S. Amrutur, "Design and Analysis of Fast Low Power SRAM", *Thesis for the Degree of Doctor of Philosophy - Stanford University*, Aug. 1999.

/14/ Bharadwaj S. Amrutur et al., "Fast Low-Power Decoders for RAMs", *IEEE J. Solid-State Circuits*, vol. 36, pp. 1506-1515, Oct 2001.

/15/ Kenneth W.Mai et al. "Low Power SRAM Design Using Half-Swing Pulse-Mode Techniques", *IEEE J. Solid-State Circuits*, vol. 33, pp. 1659-1670, Nov 1998.

/16/ T. Mori et al., "A 1V 0.9mW at 100Mhz 2kx16b SRAM utilizing a Half-Swing Pulsed-Decoder and Write Bus Architecture in 0.25um Dual Vt CMOS", *IEEE International Solid-State Circuit Conference*, pp. 354-355, 1998.

/17/ K. Osada et al., "16.7fA/Cell Tunnel-Leakage-Suppressed 16-Mbit SRAM Based on Electric-Field-Relaxed Scheme and Alter-

nate ECC for Handling Cosmic-Ray-Induced Multi-Errors", *ISSCC Digest of Technical Papers*, pp. 302-303, Feb 2003.

/18/ Michael Powell et al., "Gated-Vdd: A Circuit Technique to Reduce Leakage in Deep-Submicron Cache Memories", *Proc. of the 2000 Int. Symp. On LPED*, pp. 90-95, 2000.

/19/ Amit Agarwal et al., "DRG-Cache: A Data Retention Gated-Ground Cache for Low Power", *Proc. Conference on Design Automation*, pp. 473-478, June 2002.

/20/ Koichiro Ishibashi et al., "A Voltage Down Converter with Submicroampere Standby Current for Low-Power Static RAMs", *IEEE J. Solid-State Circuits*, pp. 920-926, June 1992.

Tan Soon-Hwei, Loh Poh-Yee, Mohd-Shahiman Sulaiman, Zubaida Yusoff
Faculty of Engineering, Multimedia University,
63100 Cyberjaya, Selangor, Malaysia

Prispelo (Arrived): 20.07.2006 *Sprejeto (Accepted):* 30.03.2007

UDK621.3:(53+54+621+66), ISSN0352-9045 Informacije MIDEM 37(2007)1, Ljubljana

AUTOMATICALLY ADJUSTABLE SUPPLY SYSTEM

M. Atanasijević-Kunc, V. Kunc*

Faculty for Electrical Engineering, Ljubljana, Slovenia
*IDS, Ljubljana, Slovenia

Key words: power-supply control, control design

Abstract: In the context of complex integrated circuits we are often faced with the problem of supply crosstalk among different parts of the system and requirements for power supply rejection well beyond the level inherently offered by the building blocks of the system. In such cases the only solution is to integrate into the system also the supply controller(s). In case the system is supplied with defined and stable supply voltage the integration of supply regulator(s) is a straightforward task. But if the system is required to operate in a wide range of external supply voltages and if the system performance is dependant on supply voltage (what is often the case) than the supply regulator(s) design becomes a crucial point in the overall system design. The presented paper offers a solution for automatic adjustment of supply regulator(s) to the applied supply voltage. This ensures that the optimum between power supply rejection demand and the demand for as high as possible internal supply voltage is reached for any external supply voltage applied, thus greatly enhancing the overall system performance.

Avtomatsko nastavljivi napajalni sistem

Kjučne besede: regulacija napajalnega sistema, načrtovanje vodenja

Izvleček: V primeru kompleksnih integriranih vezij se pogosto srečujemo s problemom presluha med različnimi deli sistema in zahtevo po odpornosti sistema na motnje na napajanju, ki presega nivo, ki jo nudijo sestavni deli vezja. V takšnih primerih predstavlja edino uspešno pot reševanja dopolnitev sistema z notranjim regulatorjem napajanja. V primeru, ko sistem napaja stabilna in točno določena napajalna napetost, je integracija tovrstnega regulatorja relativno preprosta. Če pa se soočamo z zahtevo po delovanju sistema v zelo širokem področju napajalnih napetosti in so pri tem lastnosti vezja odvisne od napetostnega napajanja (kar običajno drži), potem ustrezno načrtovanje napajalnega regulatorja postane ključnega pomena za načrtovanje celotnega vezja. V prispevku smo predstavili rešitev, ki omogoča avtomatsko uglaševanje napajalnih regulatorjev glede na zunanje napajanje sistema. Takšna rešitev ponuja optimalni kompromis med zahtevo po učinkovitem zmanjševanju neželenih vplivov spremenljivega zunanjega napajanja in zahtevo po visokem napetostnem nivoju notranjega napajanja. Na takšen način učinkovito razširjamo možnost prilagajanja zunanjemu napajanju v zelo širokem področju in s tem tudi bistveno izboljšamo lastnosti delovanja celotnega sistema.

1. Introduction

The presented automatic supply module was developed for systems which demand supply regulators to achieve desired power supply rejection but in the same time require a maximum possible supply voltage level for best performance. A typical system of this kind is an RF transmitter, where the supply voltage level directly defines the maximum output power available (for defined antenna impedance). An RF transmitter is also a system which often needs a supply regulator to ensure adequate PSRR.

This means the supply module must find an optimum between two contradicting system requirements. The demand for high PSRR inherently requires a relatively large voltage drop across the supply regulator, thus significantly lowering the regulator output voltage which is the supply voltage for the rest of the system. This is in direct contradiction with the demand for as high supply voltage (for the rest of the system) as possible to ensure high output power.

The best solution of this conflict is supply system which sets the voltage drop across the regulator to a minimum level which still guarantees the required PSRR /1/. This implies that the supply regulator output is set to a voltage relative to the external supply voltage (regulator input voltage). Assuming that the external supply voltage is not fixed and the system is required to operate at different external supply levels, an automatic adjustment system is needed to reach an optimal voltage drop across the regulator, which offers the best compromise between the contradictory demands described above.

2. System level-solution

The system presented in Fig. 1 is typical application of the system solution described. The block diagram can be divided in two parts. The basic functions of the system are presented in the main system block. The nature of the functions in this block (for instance RF output amplifier stage) requires a supply regulator to ensure proper power supply rejection (PSRR) and as high as possible supply voltage. The rest of the block diagram in Fig. 1 are bocks associated with supply voltage regulator providing the supply voltage for the main block which ensures the optimum compromise between the demand for high PSRR and the wish for highest possible supply voltage.

The operation of supply voltage regulator system can be divided into three steps /1/. In the first step the system detects the start of operation. This can be done by observing the input signals of the system (enable, power-down and similar) and/or by observing the voltage on the input

and output of the voltage regulator. The start of operation is detected when the input voltage is settled after the initial transient. The additional condition is that also the output voltage of regulator settles, thus taking into account the impact of buffer capacitor typically connected to the voltage regulator output. When the start condition is recognized the adjustment step begins.

The adjustment phase starts with the maximum setting of the voltage regulator reference /3/. This results in the output voltage almost identical to the input voltage (minimum voltage drop on across the regulator). At the same time the control block sends a signal to the main system to put the main system in the state of maximum current consumption. This ensures that the automatic adjustment is done in the worst case situation for the supply regulator. The system waits the settling transients to finish and than measures the voltage drop across the regulator. If the voltage drop is lower than the optimum value stored in the control block the control block reduces the adjustable reference for one step thus reducing the output voltage for 100mV. The system waits again for the transient to settle and repeats the measurement of the voltage drop across the regulator. If the voltage is still lower than the optimum value another voltage reduction step is started. If the voltage drop is higher the adjustment phase is finished.

The simulation of the system operation is presented in Fig. 2 and the measurements of actual implementation are shown in Figs. 6, 7 and 8.

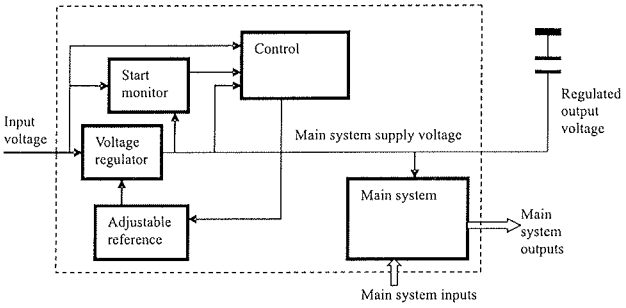


Fig. 1. Schematic system representation

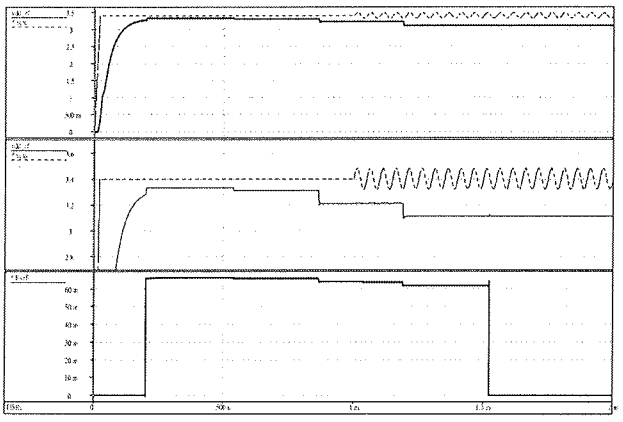


Fig. 2. Simulation results of system operation

Canvas one in Fig. 2 shows the input voltage (external supply voltage) as dashed line and the regulated output voltage as solid line. The input voltage has also an AC component (160mVpp) to high lighten the power supply rejection properties of the regulator. When the regulator input and output voltages settle (at t=200μs), start condition is recognized and the adjustment phase starts. The adjustment step was set to 100mV and a new adjustment is done every 300μs. After the third adjustment the voltage difference between the regulator input and output voltage exceeded the target value of 250mV and the adjustment phase was finished. A voltage zoom of canvas one is shown in canvas two. As described before the main system is put in state of maximum consumption during the adjustment phase. This can be seen in canvas three, presenting the current from external supply.

3. Transistor-level design

Besides the system design as described in section 2, a lot of care was devoted to the design of the supply regulator cell /2/. The demands for this cell were very high. It has to operate with minimum voltage drop (typically 200mV), it has to be stable and provide power supply rejection in a large range of load currents (from 200μA to 100μA) and it has to have adjustable output voltage in the range from 2.7V to 5V. Through the design it was assumed that the external supply blocking capacitor would be typically 2.2μF but the regulator has to operate for any supply blocking capacitor bigger than 1μF.

The dominant pole in the supply regulator loop has to be the output pole defined by the blocking capacitor to ensure fastest regulator response. The position of this pole changes drastically with the choice of blocking capacitor and the load current /2/. Since the second pole, which is typically the result of the gate capacitance of the output transistor (node A1), has to be placed at last an order of magnitude from the dominant pole, this implies the need to have the second pole as high as possible. The position of this pole also limits the minimum value of supply blocking capacitor. This means that the main design goal was to place the second pole as high as possible considering the allowed current budget of 200μA.

The required maximum current capability (100mA) at given minimum supply voltage and requested regulator voltage drop define the size of output transistor (M_out). So the capacitance of the node g_out (Fig. 3) is given and can not be significantly changed. The position of the second pole is thus mainly dependant on the output resistance of the electronic driving the gate of the output transistor (transistors M3 and M4). The chosen topology for the driver was trans-conductance amplifier with short channel devices to additionally reduce the output impedance. The simplified schematic is presented in Fig. 3.

The inputs **rvs1**, **rvs2** and **rvs3** in Fig. 3 are used to adjust the regulator output voltage in 100mV steps. The **Vref**

input is the 1.6V reference voltage and the **I10u** is the current input for bias current.

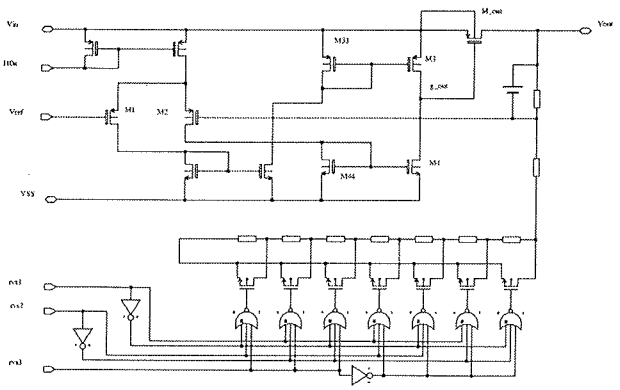


Fig. 3. The simplified schematic of the regulator cell

As discussed the main goal of optimization design is the positioning of the pole on the gate of output transistor (g_out). The optimization parameters are the current in driver transistors (M3 and M4), the length of those transistors and the current mirror ration of M3/M33 and M4/M44 pairs. The main capacitance in this node is the gate to source capacitance of the output transistor M_out. Since this capacitance is between the input voltage Vin and the node g_out it practically acts also as feed-forward system improving the regulator response to rapid changes in input voltage. The simulated response of the regulator to input voltage change (200mV step) and output current change (10mA - 60mA -1mA) is shown in Fig. 4 and 5. Fig. 4 presents a simulation response with 1μF blocking capacitor and Fig. 5 a simulation result with 100mF blocking capacitor. Canvas 1 presents the input voltage Vin, canvas 2 the output voltage Vout, canvas 3 the gate voltage of output transistor M_out (node g_out) and the bottom canvas the current from external source.

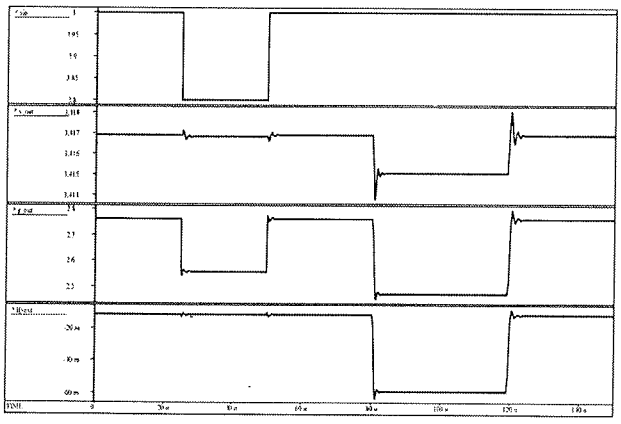


Fig. 4. Simulated reponse of regulator in the case of 1μF blocking capacitor

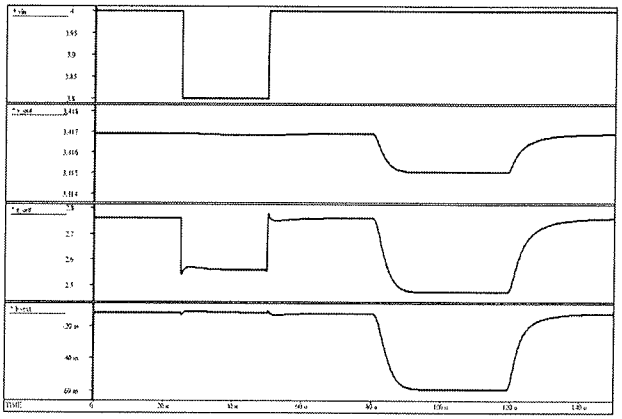


Fig. 5. Simulated response of regulator in the case of 100μF blocking capacitor

4. Implementation

The actual implementation of the described supply system was in a RF chip comprising 200mW RF output stage and receiver stage with AGC /4/. Such system needs supply voltage regulation to ensure PSRR and minimum voltage drop on the regulator to deliver the maximum possible RF output power. This was clearly an ideal case to test the automatic supply adjustment system.

The system was designed with a reasonable level of flexibility which allows the setting (using configuration register) of the desired voltage difference between the input and output voltage of the regulator.

The system performs as designed and ensures the best compromise for any supply voltage used for the system. The measurement of supply regulator system in different conditions is presented in three oscilloscope plots in Figs. 6, 7 and 8.

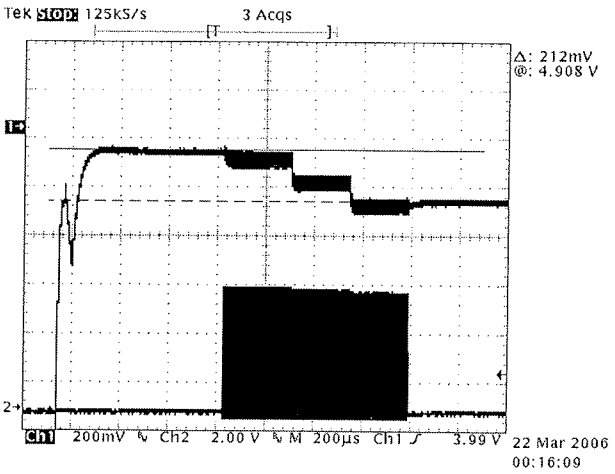


Fig. 6. System response data in the neighbourhood of the ideal conditions

Fig. 6 shows the system operation in the neighbourhood of the ideal conditions. The external supply is clean and

has negligible output resistance. Trace one shows the regulator output voltage. After switching on the input voltage the regulator output is charged to the highest level. This is slightly lower than the input voltage and in any case lower than the maximum allowed voltage for the rest of the system. During adjustment phase the regulator output voltage is lowered for 100mV every 300µs till the voltage difference exceeds the set level (200mV). During the adjustment the RF output (lower trace) is switched on since the rest of the system must be in an operation mode with the highest possible consumption.

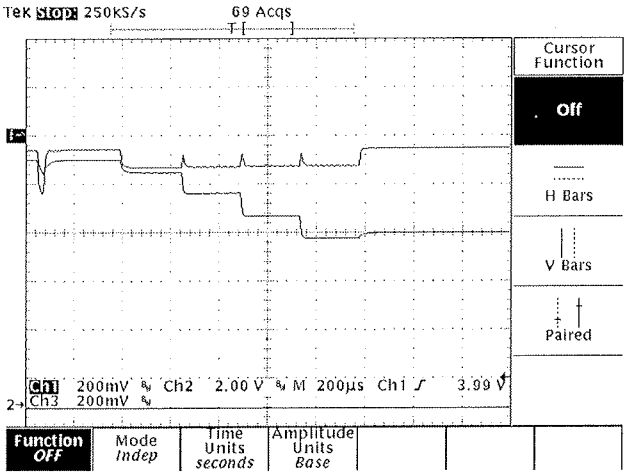


Fig. 7. System response data in case of external supply with a significant output resistance

The oscilloscope shot in Fig. 7 presents the system operation in case of external supply with a significant output resistance (2Ω). The upper trace is the external supply voltage (input voltage) and the lower trace is the regulator output voltage. At the moment the adjustment phase starts the external supply level drops (for approximately 100mV) due to increased power consumption. The automatic supply regulator system adjusts the regulator output voltage taking into account the voltage drop of input voltage in case of highest possible current consumption of the system. A spike seen on the input voltage at each lowering of output voltage results from the reduced current consumption as the system consumption current is supplied from the output buffer capacitor during the transient.

In Fig. 8 the external supply has 2 Ω output resistance and is polluted with 80mVpp 20kHz sine-wave signal. Again the input voltage is the upper trace and the output voltage is the lower trace. We can see the starting point of adjustment phase when the input voltage drops for 100mV as on Figs. 4 and 5. Each adjustment step lowering the output voltage improves the rejection of AC signal in the regulator output signal. It clearly demonstrates the conflict of the demands for high PSRR and low voltage drop across the regulator. This measurement is a good example how the described system searches and finds the optimum compromise between the two conflicting demands.

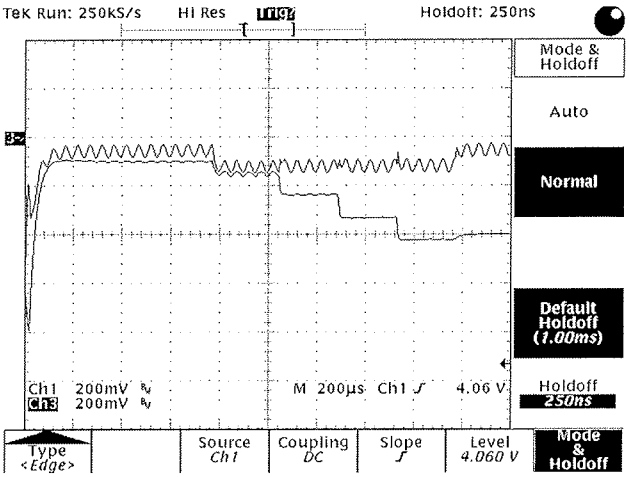


Fig. 8. System response data in case of external supply with 2 Ω output resistance and polluted with sine-wave signal

5. Conclusion

During the design of integrated medium power RF output stages we were faced with the problem how to ensure required PSRR in case of wide range of possible supply voltages without sacrificing more of the supply voltage available as it is absolutely necessary. The presented solution is a system which automatically searches for the best compromise between the conflicting requirements of high PSRR and low drop of supply regulator system.

The system was implemented as a vital part of a RF integrated system and was proved to operate as desired, thus providing intelligent supply system for RF power application. The solution is also patent pending.

References

/1/ V. Kunc, M. Atanasijević-Kunc, A. Vodopivec, "Postopek za reguliranje napajalne napetosti", Patentna prijava, številka prijave P-200500319, datum prijave 22. 11. 2005. Ljubljana: Patentna pisarna, 2005.

/2/ G. A. Rincon-Mora, P. E. Allen, "A low-voltage, low quiescent current, low drop-out regulator", IEEE Journal of Solid-State Circuits, vol. 33, pp. 36-44, January 1998.

/3/ A. Pleteršek, A compensated bandgap voltage reference with Sub-1-V supply voltage, Springer publisher: Analog integr. circuits signal process, vol. 44, pp. 5-15, 2005.

/4/ V. Kunc, M. Atanasijević-Kunc, Automatic gain adjustment in contactless communication systems, Inf. MIDEM, letn. 33, št. 2, pp. 115-117, 2003.

M. Atanasijević-Kunc,
Faculty for Electrical Engineering, Ljubljana, Slovenia

V. Kunc
IDS, Ljubljana, Slovenia

Prispelo (Arrived): 31.08.2006 *Sprejeto (Accepted):* 30.03.2007

UDK621.3:(53+54+621+66), ISSN0352-9045 Informacije MIDEM 37(2007)1, Ljubljana

KAOTIČNI KRIPTOGRAFSKI SISTEM Z UPORABO VEZIJ FPGA

Bojan Jarc, Matej Šalamon

Univerza v Mariboru, Fakulteta za elektrotehniko računalništvo in informatiko, Maribor, Slovenija

Kjučne besede: kriptografski sistemi, kaos, digitalni filtri, vezja FPGA.

Izvleček: *V naslednjem prispevku predstavljamo kaotični kriptografski sistem in možnost njegove strojne izvedbe. Jedro kriptografskega sistema predstavljata multipomična šifrirna oz. dešifrirna funkcija (enačbe 2, 3, 4) in generator psevdo-kaotične sekvence (slika 3), realiziran z digitalnim sitom drugega reda. Razmeroma enostavna struktura in izvajanje preprostih matematičnih operacij (seštevanje, odštevanje, množenje s skalarjem) dopuščajo strojno realizacijo sistema, ki omogoča hitro delovanje. Kriptografski sistem sestavlja šifrirno in dešifrirno vezje (slika 5). Pri njegovi realizaciji smo uporabili 16 bitno aritmetiko s stalno vejico in vezje FPGA XC3S500E družine Spartan-3E. Enota odprtega sporočila je bila 16 bitna. Pri načrtovanju smo uporabili kombinirani grafični opis in opis z visokonivojskim jezikom VHDL (slika 6). Maksimalna frekvenca ure, pri kateri je bila inverznost med operacijo šifriranja in dešifriranja še zagotovljena, je znašala 25 MHz (sliki 8 a, b). Pri 7-kratni (N=7) ponovitvi šifrirne funkcije je bila maksimalna frekvenca šifriranja 3,125 MHz. Ocenili smo tudi hitrost delovanja sistema pri uporabi drugih vezij FPGA (tabela 1). Ugotovili smo, da bi lahko z vezji družine Virtex 4 dosegli hitrost šifriranja 77,71 MHz pri N=1.*

Chaotic Cryptographic System Using FPGA Circuits

Key words: cryptographic systems, chaos, digital filters, FPGA circuits.

Abstract: *In this paper we present a chaotic cryptographic system and its hardware realization. The core of chaotic cryptographic system is a multi-shift cipher (eq. 2, 3) or decipher (eq. 4) and pseudo-chaotic sequence generator (fig. 3), realized with second order digital filter. Relatively simple structure and executions of simple mathematical functions (addition, subtraction, scalar multiplication) allows us to use a hardware realization, suitable for high frequencies. A chaotic cryptographic system is composed of cipher and decipher circuit (fig. 5). For its realization we have used 16 bit fixed point arithmetic and FPGA XC3S500E circuit of Spartan 3E family. Plaintext digits was 16 bit long. At designing stage we have used combined schematic and language VHDL description approach (fig. 6). Maximum clock frequency at which cipher and decipher were inversive was 25 MHz (fig. 8 a, b). Choosing the multi-shift function repetition number N = 7, allows us to encrypt the plaintext with frequency 3,125 MHz. Performance was estimated for other FPGA circuits (table 1). For Virtex 4 FPGAs we've achieved maximal clock frequency 77,71 MHz at N = 1..*

1. Uvod

Nekoč zgolj zanimiv fenomen deterističnega kaosa, je postal v zadnjem desetletju tudi praktično uporaben. Na področju digitalnih komunikacijskih sistemov je apliciran v različnih kompresijskih, šifrirnih in modulacijskih gradnikih, potrebnih pri prenosu informacij. Znani so npr. različni načini pretvorbe informacijskega signala v kaotičnega na oddajni strani in izločitev informacijskega signala iz kaotičnega na sprejemni strani. Med najpomembnejše sodijo: kaotično maskiranje, kaotično preklapljanje in kaotična modulacija.

Med leti 1990 in 1995 je bil cilj številnih raziskav, razviti sistem, ki bo omogočal modulacijo in šifriranje s pomočjo enega samega kaotičnega sistema. Ob tem sta se formirali dve različni raziskovalni področji: področje kaotične modulacije in kaotična kriptografija /1/.

Kriptografija je matematična disciplina, ki s ukvarja z varnostjo informacij, šifriranjem, avtentičnostjo in avtorizacijo. Z njo so običajno povezani: simetrični blokovni šifrirni sistemi, generatorji naključnih števil, tokovni šifrirni sistemi in asimetrični šifrirni sistemi oziroma šifrirni sistemi z javnim ključem /4/.

Apliciranje kaosa v kriptografiji oziroma pojav ti. kaotičnih kriptografskih sistemov se je pojavilo z odkritjem možnosti sinhronizacije kaotičnih oscilacij identičnih kaotičnih sistemov /3/. Sinhronizacija dveh identičnih kaotičnih sistemov je uspešna, če je mogoče, kljub njuni hiperobčutljivosti na začetne pogoje, zagotoviti njuno identično kaotično oscilira-nje. Sinhroniziranost kaotičnih sistemov na šifrirni in dešifrirni strani je namreč potreben pogoj za reverzibilnost oddajne in sprejemne strani kaotičnega kriptografskega sistema.

Kaotični sistemi se v kriptografiji uporabljajo kot generatorji naključnih števil za generiranje kriptografskih ključev in za naključno inicializacijo posameznih spremenljivk v kriptografskih algoritmih. Z odkritjem obstoja kaosa v digitalnih sistemih se v kriptografiji pojavlja vse več psevdo-kaotičnih sistemov. Ti sistemi imajo končno število različnih stanj, zato lahko generirajo le navidezno naključne oziroma psevdo-kaotične sekvence števil, ki so sicer periodične, periode ponavljanja pa so običajno zelo velike.

V prispevku predstavljamo kaotični kriptografski sistem, v katerem smo za generiranje naključnih sekvenc uporabili kaotično digitalno sito II. reda, za šifrirno in dešifrirno funkcijo pa posebno multipomično funkcijo. Opis digitalnega sita kot generatorja psevdo-naključnih sekvenc je opisan v drugem poglavju. Tretje poglavje je namenjeno opisu predlaganega šifrirnega in dešifrirnega algoritma, ki omogoča ti. tokovno šifriranje in dešifriranje. Implementacija celotnega

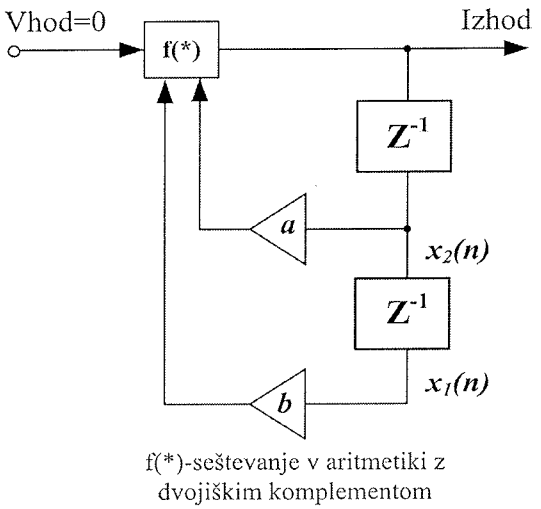
kriptografskega sistema v vezju FPGA je opisana v četrtem poglavju, peto poglavje pa opisuje rezultate meritev.

2. Generator psevdo-kaotične sekvence

Generator naključnih sekvenc je zraven šifrirne in dešifrirne funkcije najpomembnejši del vsakega kriptografskega sistema. Resnično naključne sekvence je mogoče dobiti le s pomočjo naključnih fizikalnih procesov kot je npr. metanje kovanca, termični šum na uporu ali Zener diodi itd. Med nje se pogosto uvrščajo tudi kaotični procesi, z značilno deterministično naključnostjo. Ker naključni procesi niso ponovljivi, z njimi ni mogoče zagotoviti potrebne reverzibilnosti šifrirnega in dešifrirnega algoritma, saj sta za njo potrebna dva popolnoma enaka vira naključnih sekvenc.

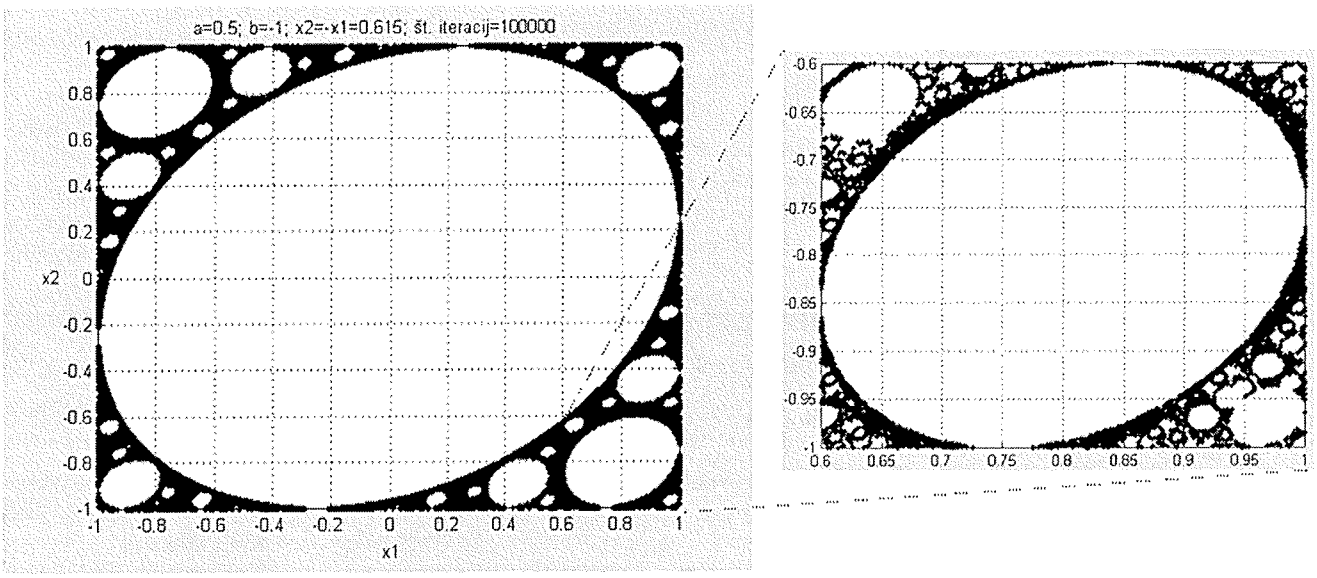
Večina današnjih kriptografskih sistemov uporablja generatorje psevdo-naključnih sekvenc. Ti lahko zaradi končnega števila različnih stanj, generirajo le navidez naključna oziroma psevdo-naključna zaporedja števil. Takšna zaporedja so sicer periodična, njihova perioda pa zelo velika[1] in običajno vnaprej izračunljiva. Psevdo-naključne sekvence z majhno periodo za šifrirne namene niso primerne, saj omogočajo hitro razkritje tajnega ključa. Zelo znan je preprost način generiranja psevdo-naključnih sekvenc s pomičnimi registri in povratno zanko LFSR[2] /4/.

Med psevdo-naključne sekvence se velikokrat uvrščajo tudi psevdo-kaotične sekvence /5/, čeprav je med njimi kar nekaj razlik /6/. V generatorju psevdokaotičnega signala poteka hiperobčutljiv iterativni proces, na osnovi katerega se generira psevdokaotična sekvenca števil, ki je paradoksalno urejeno neurejena. Kaotično naključni proces je namreč v primerjavi s povsem naključnim procesom bolj urejen oziroma manj naključen. Urejeno neurejenost potrjujejo urejene fraktalne podobe, ki pri resnično naključnih procesih ne obstajajo.



Slika 1: Nelinearni model digitalnega sita drugega reda.
Fig. 1: Second order digital filter nonlinear model.

Eden izmed sistemov, ki se lahko obnaša kaotično, je tudi digitalno sito II. reda. Čeprav je njegovo obnašanje natančno obravnavano v prispevkih /7/, na tem mestu omenimo le nekatere pomembnejše ugotovitve, ki bodo omogočile lažje razumevanje, v nadaljevanju predstavljenega, kaotičnega kriptografskega sistema. Slika 1 prikazuje nelinearni model digitalnega sita II. reda, ki se lahko ob določenih pogojih obnaša kaotično.



Slika 2: Kaotična trajektorija.
Fig. 2: Chaotic trajectory.

1 *Periode psevdo-naključnih sekvenc, uporabljenih v kriptografskih sistemih znašajo 2^{256} in več.*

2 *LFSR- Linear Feedback Shift Registers.*

Za to morajo biti izpolnjeni naslednji pogoji /7/:

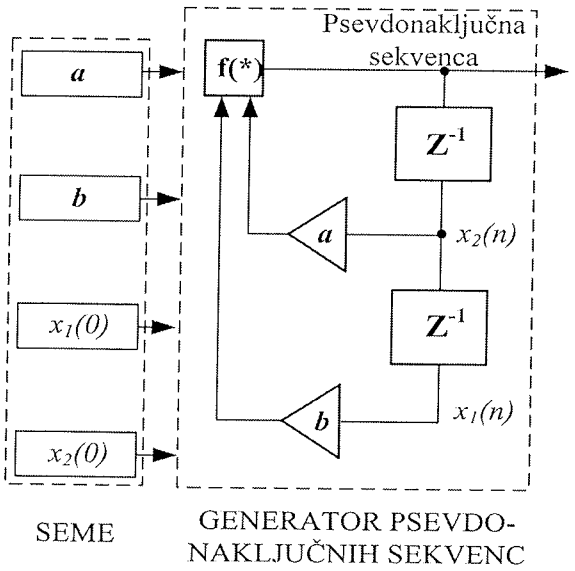
- na vhodu mora biti ničelni signal;
- koeficienta *a* in *b* morata biti izbrana tako, da sistem deluje na robu stabilnosti;
- za začetni stanji $x_1(0)$ in $x_2(0)$ morata veljati:

$$x_1(0) = -x_2(0) = x_0 > \frac{1}{2}\cdot(2-a)^{1/2} = 0{,}612372435695... \quad (1)$$

Kaotično obnašanje sita predstavlja trajektorija, ki v ravnini x_1-x_2 opisuje fraktalno geometrijo elips. Takšna trajektorija je prikazana na sliki 2. Dobili smo jo na osnovi simulacije 16-bitne strukture pri naslednjih parametrih sita: a=0,5; b=-1; $x_2(0)$= -$x_1(0)$=0,615.

V primeru 16-bitne strukture je število vseh različnih stanj oziroma vrednosti spremenljivk x_1 in x_2 2^{16} = 65536. Te vrednosti se pojavljajo psevdo-naključno. Velikost period je spremenljiva in odvisna od začetnega stanja spremenljivk stanj x_1 in x_2 ter števila bitov, s katerimi spremenljivki predstavimo.

Predstavljeno digitalno sito lahko v kaotičnem režimu delovanja uporabimo kot generator psevdo-naključne sekvence (slika 3), ki je generirana na osnovi začetnih stanj spremenljivk x_1 in x_2 ter koeficientov *a* in *b*. Te vrednosti predstavljajo seme oziroma začetno stanje za algoritem, ki opisuje delovanje sita.



Slika 3: Generator psevdo-kaotičnih sekvenc.
Fig. 3: Pseudo-chaotic stream generator.

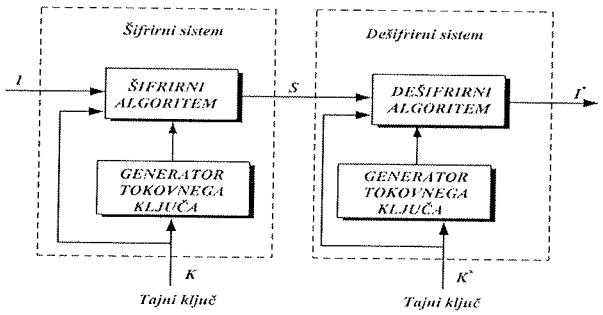
3. Kaotični kriptografski sistem

Klasičnim kriptografskim sistemom se v zadnjih nekaj letih pridružujejo novi kaotični kriptografski sistemi, ki temeljijo na fenomenu determinističnega kaosa oziroma na zelo preprostih, vendar hiperobčutljivih iterativnih postopkih. Ker je področje determinističnega kaosa še precej neraziskano, je razumljivo, da je tudi t.i. kaotična kriptografija relativno nov subjekt raziskav, vezana na metode in standarde, ki so šele v razvojni fazi. Danes zasledimo tako tokovne kot blokovne kaotične kriptografske sisteme.

V tokovnih simetričnih šifrirnih sistemih so generatorji psevdonaključne sekvence uporabljeni kot generatorji tokovnega ključa, kar pomeni, da se šifriranje odprtega sporočila izvaja bit za bitom, in sicer tako, da se tokovni ključ kombinira z biti odprtega sporočila. Ti sistemi so zelo hitri in izvajajo šifriranje manjših enot odprtega sporočila, kot je npr. bit ali znak (8-bitov) v datoteki, točka v digitalni sliki itd.

Druga vrsta šifrirnih sistemov so blokovni. Pri teh se odprto sporočilo razdeli na tako dolge bloke, kot jih zahteva algoritem[3], nato pa se biti posameznega bloka, v skladu s šifrirnim algoritmom, premeščajo in kombinirajo s tajnim ključem, generiranim z generatorjem naključnih sekvenc.

V nadaljevanju obravnavamo tokovni kaotični kriptografski sistem, ki temelji na multipomični šifrirni/dešifrirni funkciji. Tokovni ključ je generiran s pomočjo prej predstavljenega kaotičnega digitalnega sita. Blokovna shema celotnega kriptografskega sistema je prikazana na sliki 4.



Slika 4: Blokovna shema kriptografskega sistema.
Fig. 4: Cryptographic system block scheme.

Šifriranje odprtega sporočila *I* se izvaja postopoma po korakih (slika 5). *N*-ti vzorec odprtega sporočila *i*(*n*) in psevdonaključne vrednosti *k*(*n*), se s pomočjo tajnega ključa *K* in posebne nelinearne, rekurzivne, multipomične funkcije, pretvori v šifriran vzorec tajnopisa *s*(*n*). Multipomično šifrirno funkcijo /8/ opisuje enačba:

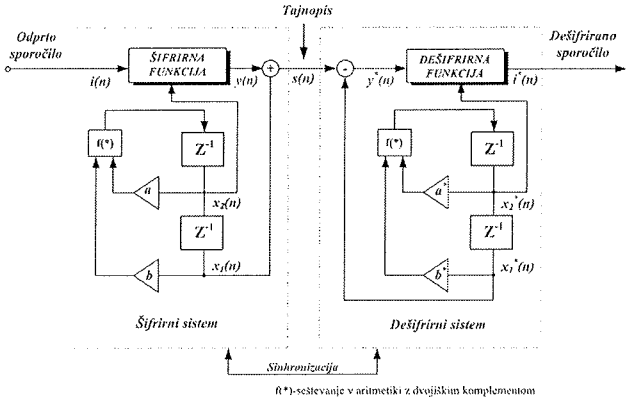
$$s(n) = \underbrace{f_1(...f_1(f_1}_{N}(i(n), \underbrace{k(n)), k(n)),...k(n))}_{N} + k(n+1) \quad (2)$$

pri čemer je *N* poljubno število iteracij nelinearne funkcije f_1:

$$f_1(x,k) = \begin{cases} (x+k)+2\cdot h & -2\cdot h \le (x+k) \le -h \\ (x+k) & -h \le (x+k) \le h \\ (x+k)-2\cdot h & h \le (x+k) \le 2\cdot h \end{cases} \quad (3)$$

Tajni ključ sestavlja več vrednosti: koeficienta sita *a* in *b*, začetni stanji $x_1(0)$ in $x_2(0)$, število iteracij multipomične šifrirne funkcije *N* in parameter *h*.

3 *Več sto kilo bajtov.*



Slika 5: Podrobnejša shema kriptografskega sistema.
Fig. 5: Detail scheme of cryptographic system.

Šifrirna funkcija $y(i(n))$ bo bijektivna, če bo vrednost spremenljivke h izbrana tako, da bosta x in k vedno znotraj intervala $(-h, h)$. Samo v tem primeru bo obstajal tudi inverzni - dešifrirni algoritem, ki ga lahko v skladu z oznakami na sliki 5 opišemo z naslednjo dešifrirno funkcijo:

$$i^*(n) = \underbrace{f_1(...f_1(f_1(y^*(n), \underbrace{-x_2^*(n)), -x_2^*(n)), ..., -x_2^*(n)}_{N})}_{N} \quad (4)$$

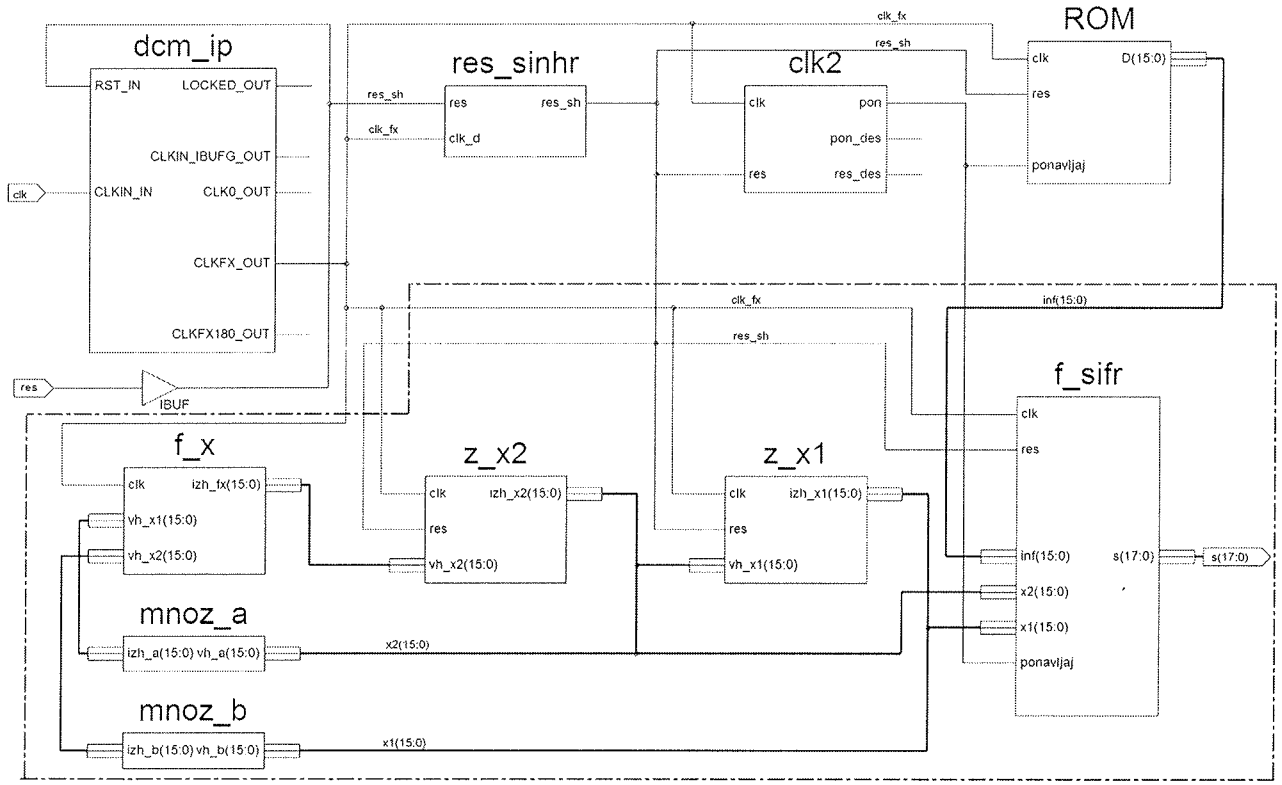
Ker sta šifrirna in dešifrirna funkcija rekurzivni, bo za izračun posamezne vrednosti $y(n)$ oziroma $i^*(n)$ potreben določen čas, ki bo odvisen od izbranega števila iteracij N.

4. Implementacija kriptografskega sistema

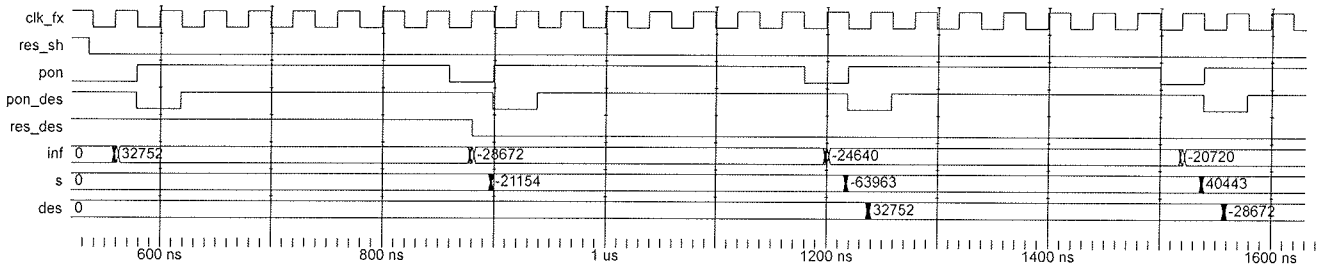
Kriptografski sistem je mogoče implementirati s pomočjo programa, ki se izvaja na računalniku ali s strojno opremo. V našem primeru smo se odločili za realizacijo s polji programirljivih logičnih vezij (vezja FPGA) in aritmetiko s fiksno vejico. Implementirali smo 16-bitno strukturo.

Današnja vezja FPGA ponujajo, razen osnovnih logičnih elementov, bistveno kompleksnejše gradnike: množilnike, bloke za digitalno procesiranje signalov pa tudi procesorje. Iz slike 1 je razvidno, da sta najkompleksnejši enoti generatorja psevdo-kaotičnih sekvenc, množilnika. Ker pa sta multiplikatorja, a = 0,5 in b = -1, je realizacija obeh množenj enostavna, in sicer z vezjem za pomik binarne besede v desno in vezjem za tvorjenje dvojiškega komplementa. Tako realiziran generator psevdo-kaotičnih sekvenc bo sposoben v eni periodi urinega signala generirati novo psevdo-kaotično število.

Čim krajše razvojno obdobje, možnost enostavne nadgradnje sistema in spremembe tajnega ključa narekujejo načrtovanje z uporabo visokonivojskega jezika. Podjetje XILINX predstavlja enega izmed vodilnih proizvajalcev in ponuja širok spekter vezij FPGA. Razen visoko zmogljivih (družina Virtex), ponuja tudi srednje zmogljiva vezja (družina Spartan). Hkrati je prosto dostopna osnovna verzija načrtovalskega orodja ISE. Orodje ISE podpira opis vezij z vi-



Slika 6: Izvedbena struktura šifrirnega sistema v vezju FPGA.
Fig. 6: FPGA realization block scheme of cipher circuit.



Slika 7: Simulirani časovni odzivi kriptografskega sistema izvedenega v vezju FPGA.
Fig. 7: Simulated waveforms of cryptographic system realized in FPGA circuit.

sokonivojskim jezikom (VHDL, Verilog, ...), shematski opis in opis s pomočjo diagrama stanj. Skupaj z logičnim simulatorjem ModelSim, omogoča osnovno funkcionalno preverjanje delovanja vezja in časovno simulacijo.

Kriptografski sistem smo realizirali s pomočjo razvojne plošče Spartan-3E /9/. Jedro razvojne plošče predstavlja vezje XC3S500E družine Spartan-3E. Poleg klasičnih gradnikov (logična vrata, logične celice, pomnilnik RAM, ...), so v XC3S500E na voljo tudi množilniki 16×16 bitov, ki pa jih naše vezje ne izkorišča. Za opis vezja smo uporabili kombiniran pristop z visokonivojskim jezikom VHDL in shematskim opisom. Zaradi enostavnejše primerjave z osnovno blokovno shemo (glej sliko 5), smo vezje razdelili na osnovne enote, kot to prikazuje slika 6.

Na sliki 6 je šifrirni sistem obkrožen s prekinjeno črto. Posamezni bloki so predstavljeni z naslednjimi imeni:

- *mnoz_a*, *mnoz_b*: moženje s konstantama *a* in *b*, izvedeno s pomikom vsebine za bit v desno (*mnoz_a*) in s tvorjenjem dvojiškega komplementa (*mnoz_b*);
- *z_x1*, *z_x2*: zakasnilna elementa;
- *f_x*: seštevanje v aritmetiki z dvojiškim komplementom;
- *f_sifr*: multipomična šifrirna funkcija;
- *dcm_ip*: generator signala ure z delitvijo signala zunanje ure *clk*;
- *res_sinhr*: vezje za reset;
- *clk2*: generator krmilnih signalov;
- *ROM*: pomnilnik z odprtim sporočilom.

Vodila znotraj šifrirnega sistema, kot tudi enota (beseda) odprtega sporočila *inf* (slika 6), so 16 bitna. Enota tajnopisa *s* je zapisana z 18 biti. Da hitrosti delovanja šifrirnega sistema ne bi omejevale vhodno/izhodne enote razvojne plošče, je odprto sporočilo *inf* dolžine 5500 besed, shranjeno v pomnilniku tipa ROM (slika 6) in se ponavljajoče pošilja šifrirnemu sistemu. Splošno gledano, sta lahko začetni stanji $x_1(0)$ in $x_2(0)$ zakasnilnih elementov *z_x1* in *z_x2* poljubni, z enačbo (1) določeni vrednosti. V našem primeru smo ju zapisali v dvojiškem komplementu s 16 bitno besedo in sicer z $B179_{16}$ (-0,6135) in $4E87_{16}$ (0,6135). Število iteracij *N* šifrirne funkcije (enačbi 2 in 3) določa blok *clk2*, s pomočjo kontrolnega signala *pon*. Posamezna iteracija šifriranja se izvede v eni periodi ure *clk_fx*. Šifriranje z nelinearno funkcijo (3) se ponavlja dokler signal *pon* ostaja na nivoju logične enice. S postavitvijo signala *pon* na nivo logične ničle se beseda tajnopisa *s* prenese na izhod in zajame novo vrednost odprtega sporočila *inf*.

Pravilnost delovanja šifrirnega sistema smo preverili s pomočjo dešifrirnega sistema, ki smo ga realizirali v istem vezju FPGA. Njegova struktura je podobna strukturi šifrirnega sistema na sliki 6, zato je podrobneje ne bomo opisovali. V realnem okolju sta šifrirni in dešifrirni sitem običajno dislocirani napravi. Da kriptografski sistem deluje pravilno, je potrebno pred pošiljanjem tajnopisa poskrbeti za sinhronizacijo generatorjev psevdo-kaotičnih sekvenc ter varno izmenjavo tajnega ključa. Ključ sestavlja podatek o številu iteracij multipomične šifrirne funkcije ter vrednosti začetnih stanj $x_1(0)$ in $x_2(0)$. Načrtovanje vezja za sinhronizacijo in varno izmenjavo tajnega ključa ni predmet naših raziskav, zato smo sinhronizacijo zagotovili kar s pomočjo signalov *res_des*, *pon_des* in *clkfx180_out* (slika 6). Navedeni signali predstavljajo reset, kontrolni signal iteracij in uro dešifrirnega sistema. Izbrali smo tajni ključ in ga uporabili na šifrirnem in dešifrirnem sistemu. Simulirane časovne poteke signalov, z upoštevanimi zakasnitvami povezav in elementov, prikazuje slika 7.

Označbe signalov so naslednje:

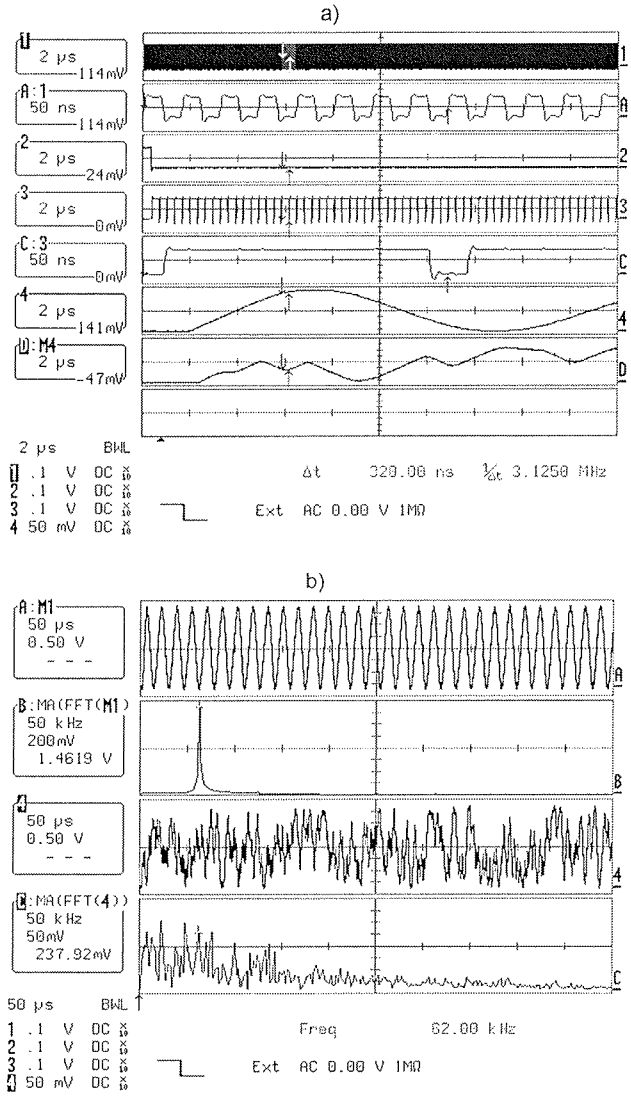
- *clk_fx*: signal ure, s pomočjo fazne zanke sintetiziran iz *clk*;
- *res_sh*: z uro *clk_fx* sinhroniziran asinhroni reset *res*;
- *pon*, *pon_des*: krmilna signala šifrirnega in dešifrirnega sistema, ki določata število ponovitev multipomične funkcije;
- *inf*: oprto sporočilo;
- *s*: tajnopis;
- *des*: dešifrirano odprto sporočilo.

Za izbrano vezje FPGA XC3S500E-4 smo uspeli zagotoviti inverznost med operacijo šifriranja in dešifriranja do frekvence ure f_{clk_fx} = 25MHz. Frekvenco ure omejujejo časovne zakasnitve elementov in povezav v vezju FPGA. Če je število ponovitev multipomične šifrirne funkcije $N > 1$, se frekvenca šifriranja odprtega sporočila ustrezno zmanjša. Na sliki 7 je prikazan primer za $N = 7$, kjer je šifriranje besede odprtega sporočila trajalo 320 ns oz. s frekvenco

3,125 MHz. Enak čas je potreben tudi za dešifriranje (signal *des* na sliki 7).

5. Rezultati meritev

Pravilno delovanje vezja potrjujejo rezultati meritev. Čeprav lahko šifriramo poljubne podatke kot so tekst, digitalizirani avdio ali video signal, smo delovanje kriptografskega sistema preverili s pomočjo šifriranja in dešifriranja sinusnega signala. Digitalizirane vrednosti smo zapisali v pomnilnik ROM (slika 6). Primerjavo med odprtim sporočilom in tajnopisom smo izvedli s pomočjo 12 bitnega serijskega DA pretvornika LTC2624, ki je na voljo na razvojni plošči. Sliki 8 a) in b) prikazujeta izmerjene odzive.



Slika 8: Rezultati meritev: a) Časovni odzivi šifrirnega sistema. b) Časovna poteka tajnopisa in dešifriranega odprtega sporočila, ter njuna amplitudna spektra.

Fig. 8: Measured waveforms: a) Cypher circuit waveforms. b) Waveforms Simulated waveforms of cryptographic system realized in FPGA circuit.

Slika 8 a) prikazuje: časovni potek signala ure *clk_fx*, izsek signala *clk_fx*, signal *res_sh*, signal *pon*, izsek signala *pon*, odprto sporočilo *inf* (sinusni signal) in tajnopis *s*. Slika 8 b) prikazuje dešifrirano sporočilo (potek A), njegov amplitudni spekter (potek B), časovni potek šifriranega signala (potek 4) in njegov amplitudni spekter (potek C). V amplitudnem spektru šifriranega signala ni mogoče prepoznati oprtega sporočila.

Maksimalna dopustna frekvenca delovanja kriptografskega sistema je veliki meri odvisna od razmestitve elementov, dolžine povezav v vezju FPGA in od izbranega vezja FPGA oz. tipičnih zakasnitev logičnih blokov ter povezav. V procesu načrtovanja običajno uporabljamo postopke avtomatskega razmeščanja in povezovanja na katere lahko vplivamo. S vhodnimi parametri in z določitvijo časovnih omejitev, ki jih morajo izpolnjevati elementi vezja in povezave, lahko vplivamo na doseženo maksimalno frekvenco delovanja. Tako smo s postopkom Fmax /10/ poiskali vhodne nastavitve algoritmov avtomatskega razmeščanja in povezovanja, ki maksimirajo hitrost delovanja šifrirnih, vezij realiziranih na petih različnih platformah. Rezultate prikazuje tabela 1.

Družina	Vezje FPGA	Frekv. [MHz]
Spartan 3E	XC3S500E-4	39,44
Spartan 3E	XC3S400-5	39,83
Virtex 2P	XC2VP2-7	57,62
Virtex 4	XC4VLX25-12	77,23
Virtex 4	XC4VFX12-12	77,71

Tabela 1. Predvidene maksimalne frekvence delovanja šifrirnega vezja glede na izbrano vezje FPGA.
Table 1. Expected maximum clock frequency of cipher circuit depending on chosen FPGA circuit.

Maksimalne frekvence delovanja smo le ocenili, saj dejanskega delovanja nismo preizkusili. Najvišji hitrosti delovanja lahko pričakujemo v primeru uporabe vezij družine Virtex 4. Najnovejše družine Virtex 5 nismo preizkusili, saj je razvojno orodje ISE WebPack ne podpira.

6. Zaključek

V prispevku smo predstavili kaotični kriptografski sistem in možnost njegove strojne realizacije. Šifrirni in dešifrirni sistem sestavljata multipomična šifrirna oz. dešifrirna funkcija in generator psevdo-kaotičnih sekvenc, realiziran s kaotičnim digitalnim sitom. Oba sistema smo izvedli v istem vezju FPGA družine Spartan 3E. Struktura šifrirnega sistema je bila 16 bitna.

Bistvena prednost predlagane strojne izvedbe pred programsko, je hitrost izvajanja šifrirnega in dešifrirnega algoritma. Za generiranje psevdo-kaotične vrednosti je potrebna le ena perioda urinega signala. V našem primeru sta bili

operaciji šifriranja in dešifriranja še inverzni, če je bila frekvenca ure 25 MHz. Maksimalna frekvenca šifriranja je odvisna predvsem od števila ponovitev multipomične šifrirne oziroma dešifrirne funkcije. S pomočjo postopka Fmax pa smo ocenili, da bi lahko pri uporabi vezij družine Virtex 4, frekvenco ure povečali na 77,71MHz.

Programirljivost vezij FPGA omogoča enostavno spremembo števila bitov strukture kaotičnega sistema in dolžino tajnega ključa. Uporaba daljših enot šifriranja ne bi vplivala na hitrost šifriranja, ampak le na kompleksnost sistema, saj bi se povečal uporabljen delež razpoložljivih logičnih blokov v vezju FPGA.

7. Literatura

/1/ N. Masuda, G. Jakimoski, K. Aihara, L. Kocarev: *Chaotic Block Ciphers: From Theory to Practical Algorithms*, IEEE Transaction Circuits and Systems-I: letnik 53, št. 6, junij 2006, str. 1341-1352.

/2/ T. Stojanovski, L. Kocarev: *Chaos-Based Random Number Generators-Part I: Analysis,* IEEE Transaction Circuits and Systems-I: Fundamental Theory and Applications, letnik 48, št. 3, marec 2001, str. 281-288.

/3/ L. M. Pecora, T. L. Carroll: *Synchronization in Chaotic Systems,* Physical Review Letters, letnik 64, št. 8, 1990, str. 821-824.

/4/ B. Schneier: Applied *Cryptography, Second Edition: Protocols, Algorthms, and Source Code in C*, John Wiley & Sons, New York, 1996.

/5/ Greg M. Bernstein, Michael A. Lieberman: *Method and apparatus for generating secure random numbers using chaos*, US Patent No. 5007087, april 1991.

/6/ Gernot Kubin: *What Is a Chaotic Signal ?*, IEEE Workshop on Nonlinear Signal and Image Processing, Greece, 20 - 22. junij, 1995, http://poseidon.csd.auth.gr/Workshop/papers/.

/7/ M. Šalamon, T. Dogša: *Nevarnost kaosa v digitalnem situ drugega reda*, Inf. MIDEM, letnik 30, št. 1, marec 2000, str. 37-42.

/8/ T. Yang, C. W. Wu, L. O. Chua: *Cryptography Based on Chaotic Systems*, IEEE Transaction Circuits and systems-I: Fundamental theory and applications, letnik 44, št.5, maj 1997, str. 469-472.

/9/ Xilinx, Inc.: *Spartan-3E Starter Kit Board User Guide*, UG230 (v1.0) March 9, 2006.

/10/ H. Patel: *Synthesis and Implementation Strategies to Accelerate Design Performance*, Xilinx, Inc., July, 2005.

Doc. dr. Bojan Jarc
Doc. dr. Matej Šalamon

oba UNIVERZA V MARIBORU
FAKULTETA ZA ELEKTROTEHNIKO,
RAČUNALNIŠTVO IN INFORMATIKO
2000 Maribor, Smetanova 17, Slovenija
e-mail: *bojan.jarc@uni-mb.si, matej.salamon@uni-mb.si*

Prispelo (Arrived): 20.12.2006 *Sprejeto (Accepted):* 30.03.2007

IMPLEMENTATION OF NON-INTRUSIVE FAULT DETECTION IN EMBEDDED CONTROL SYSTEMS

Domen Verber, Matej Šprogar[1], Matjaž Colnarič

University of Maribor Faculty of Electrical Engineering and Computer Science, Maribor, Slovenia

Key words: embedded control systems, fault management, fault detection, monitoring cells, evolutionary computing.

Abstract: Paper presents fault detection in embedded control systems by the so-called monitoring cells. The basic idea is to monitor input/output variables and internal states of systems, processes or sub-processes by using acquired and built-in knowledge about the normal behavior in order to detect abnormalities. Paper gives the detailed architecture and the operation of the monitoring cells. The concept is applicable even if only a limited knowledge about the control system is available. In such cases the proposed automated learning of the monitoring function can be used. In the second part, two different implementations of the monitoring cell are presented. The first one uses discrete analogue devices and a field programmable gate array. The second is based on the programmable system-on-a-chip devices.

Izvedba neintruzivne detekcije napak v vgrajenih krmilnih sistemih

Kjučne besede: vgrajeni krmilni sistemi, ravnanje z napakami, detekcijanapak, nadzorne celice, evolucijsko učenje.

Izvleček: Članek predstavlja detekcijo napak v vgrajenih sistemih s tako imenovanimi nadzornimi celicami. Osnovna ideja je nadzor vhodno/izhodnih spremenljivk in notranjih stanj sistemov, procesov in podprocesov z uporabo pridobljenega in vgrajenega znanja o normalnem obnašanju in z namenom prepoznati nepravilnosti. Članek podrobno predstavlja arhitekturo in delovanje nadzornih celic. Koncept je uporaben tudi, kadar je na razpolago le omejeno poznavanje krmilnega sistema. V takšnih primerih se lahko uporabi predlagano avtomatsko učenje nadzorne funkcije. V drugem delu sta predstavljeni dve različni izvedbi nadzorne celice. Prva uporablja diskretne analogne enote in programirljiva FPGA vezja, druga pa temelji na programirljivih sistemih na čipu (pSoC).

Introduction

Embedded control systems are rapidly becoming the invisible mind behind most modern appliances. Their major spread could be observed even in safety critical environments, where failures can have serious or even fatal consequences. However, highly dependable programmable embedded systems for safety critical applications still lack proper scientific treatment. Controllers must be (besides being dependable) flexible in order to cut production costs. Flexibility (achieved mainly by programmability), however, is in conflict with dependability.

Faults in programmable control systems are unavoidable. Fault management as a discipline is embracing four types of techniques /7/: (a) fault avoidance is preventing faults in the design phase; 1(b) fault removal is attempting to find faults before the system enters service (testing); (c) fault detection is finding faults during system service and minimizing their effects, and (d) fault tolerance is allowing the system to operate correctly even in the presence of faults. A number of competent authors elaborated this area, for example /5, 1/.

Failure in a system can be handled for example by redundancy, diversity, reconfiguration etc. Firstly, however, it must be detected. For detection some sort of a dependable monitoring subsystem must be used that detects abnormalities and triggers appropriate corrective actions. Because of complexity, safety related issues of the system should be designed, evaluated and implemented independently and in parallel with the functional part; the same is with fault detection - to achieve it, a monitoring component called monitoring cell (MC) is introduced to supervise the control function. Early in the development of the system each MC is considered as an abstract object. Later on, the MCs are implemented as hardware and/or software components. Paper presents a MC concept for embedded control systems implemented in hardware either using discrete components with field programmable gate array (FPGA) or programmable system-on-chip (PSoC).

The MC must detect run-time faults, which are the most difficult to discover because they are a consequence of an unpredictable event or chain of events. One way to detect them is to observe whether the system's states are within reasonable limits at all times. To recognize what is

1 *Partially supported by Ministry of Higher Education, Science and Technology of Republic of Slovenia by the post-doctoral project grant no. Z2-6398-0796-04*

"normal" we propose to observe the system during normal operation by recording all inputs, outputs and internal states which are believed to affect system's future behavior. Based on these recordings a machine learning technique can be used to learn the normal behavior. Evolutionary algorithms (EA), for example, are one suitable paradigm for this task /3/.

Section 2 explains the concept of the fault detection by monitoring cells. The detailed description of their operation is given in the Section 3. A method for establishing the monitoring function using machine learning is described in Section 4. Finally, Section 5 gives two case-studies of MC implementation.

Detecting faults by monitoring cells

A typical present day control system consists of physical process components, sensors, actuators, distributed computers with communication networks, and a lot of software. Examples are control systems in industrial plants, nuclear reactors, avionics, etc. The size and complexity of such systems increases every year and so does the probability of a fault in one of the many components.

A control system can usually be divided into a set of well-defined sub-processes or tasks, running on a processing resource called *control cell*, each performing a specific *control function*. A task takes its inputs from sensors and/or from the results of other processes, produces results that can be used by other tasks, and controls the controlled system through actuators. Tasks are triggered by synchronous or asynchronous events.

Causes for faults in such systems are either hardware or software related; additionally the input and output signals may not comply with functional specifications of the system also causing faulty behavior. In hard real-time systems, improper temporal behavior is also considered a fault. Another, not so obvious reason for errors are temporal inconsistencies of the signals; e.g. the signals' changes can be too steep, frequency of events can be to high, etc.

The basic task of a monitoring cell is to monitor the validity of input $\boldsymbol{x}$ and output $\boldsymbol{y}$ values and (possibly) the internal states $\boldsymbol{m}$ of the control process. As a result, the correctness c is reported to a higher fault management layer that will handle the detected fault. If feasible, additional diagnostics parameter d can also be provided.

The control cell under surveillance must have physically accessible input and output signals. Since it can implement any (sub) process, it is important to identify control functions with clear and explicit relations between the input and output signals. The monitored control cell is considered a *gray* box with defined external functional behavior and with at least partly known internal structure, which is observable through its obtainable internal states. There are several reasons for that decision:

- The monitored component is not necessarily a black box and this knowledge can reveal additional and more accurate information about any faulty behavior.
- White box can be too complex. If chosen, it would be necessary to implement the monitor cell physically inside the original software and hardware to limit the communication problems. That, however, would be too intrusive and would increase the complexity and reduce the performance of the control part, what is counterproductive.
- White box implementation on the same resources would introduce another central point of failure.
- Clear separation between the control and monitoring function simplifies the design; both functions can be done separately and by different designers with diverse competence enhancing the dependability to some extent.

If the control function is a component with a well defined behavior but unknown internal structure, it can as well be taken as a black box. We propose to physically separate the control part and the monitoring cell, i.e. to employ separate hardware. While it may be more expensive, it provides much more competent implementation of the above guidelines. Also, complexity is kept lower by partitioning of the functions.

To allow for the gray box implementation, both input and output digital and analogue signals, as well as the internal states need to be observable. There is a number of possibilities for the latter: either they are made accessible via standard parallel or serial interfaces at the control cell, or another feature of contemporary processors is made use of, like JTAG boundary scan testing and in-system programming /4/ or a concept similar to "background debugging mode" high-speed clocked serial debugger interface by the CPU32 Motorola family.

Monitoring cell should be by orders of magnitude less complex than the control cell and should produce as little interference with the control environment as possible. It should be built from simple and robust components with low probability of failure. The consequence could be that the complexity of monitoring functions may be limited (e.g. no floating point arithmetic, etc.). This limitation, however, can be of advantage: because of the simplicity such a system could be formally verified and possibly certified by a certification authority.

When an abnormality is detected by a monitoring cell, the diagnostic mechanism will attempt to acquire as many details as possible. The error or a failure together with possible description will be coded in the diagnostic signal and mediated to the upper layers of fault management system where appropriate actions will be taken by, e.g., re-configuration manager. This, however, is beyond the scope of this paper.

Evaluation of the monitored signals

As shown in Figure 1, the control function F operates on input values ***x*** and accessible internal states ***m*** (if any) to produce output values ***y***. It is assumed that it operates in discrete time in a PLC-like fashion. It acquires its inputs at a specific point in time *t* and after some delay produces the results and updates the internal states. During this time no new inputs can be evaluated.

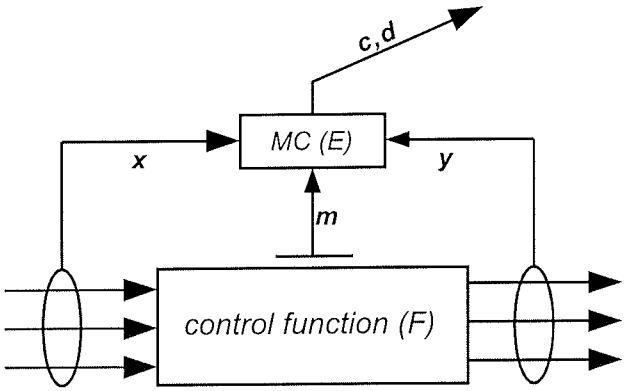


Fig. 1. Concept of the Monitoring cell.

By enumerating the time intervals, the notion of time can be reduced into discrete (integer) values as shown in (1), where F represents the control function.

$$(y(t+1), m(t+1)) = F(x(t), m(t)); t \in IN \qquad (1)$$

The MC acquires the current inputs at the beginning of the cycle at the instant *t*; at the same time the inputs x are read by the control function. At the end of the cycle (what is also the start of the next cycle) at instant *t + 1* (when outputs of the system are produced) the output ***y*** and internal state ***m*** of control function are read and the final evaluation of the subsystem is performed by MC. The MC effectively performs the function *E* defined by (2).

$$E(x(t), m(t), x(t+1), m(t+1), y(t+1)) \qquad (2)$$

Note that the function *E* can operate on both new and old instances of ***x*** and ***m***, what allows for early detection of discrepancies on the inputs and internal states at the instant *t + 1*.

MC operates on three groups of inputs – ***x***, ***m***, ***y***, the former two in instances *t* and *t + 1*. These are the values actually available to MC for evaluation. To simplify further discussion all inputs are called *signals* s_i; thus, function (2) can be rewritten as $E(\mathbf{s}(t))$.

To simplify the analysis and the implementation, the evaluation function *E* can be decomposed into a set of simpler evaluation tests E_{part} that evaluate parts of the system. At the end the partial evaluations are combined to produce the final result. If any of the sub-evaluations E_{part} indicates an illegal state of the system, the final result should also be noted as illegal.

A minimalist evaluation function E checks the integrity of individual signals. The information on their basic properties is acquired from the system specifications, technical documentation or similar sources. This way at least information on the data ranges – valid and invalid values of different signals – is extracted. In certain cases this is the only step that can be performed.

Using this information a simple classification of a signal into a legal or illegal class can be made. However, it is not always possible to find a sharp boundary between valid and invalid states. Additional buffer zones should be introduced where the validity of the signals can not be determined. This is the foundation of the partial evaluation function E_i that determines correctness c_i of the signal $\mathbf{s}_i$.

$$c_i = E_i(\mathbf{s}_i)$$

$$c_i \in C=\{\text{valid, invalid, undetermined}\}$$

This enables a simple validation of each signal. For example, any invalid input value can be detected (possibly from a faulty sensor) and/or illegal output can be detected and, consequently, prevented.

However, a more thorough validation of the system should be performed to make sure that outputs are consistent with the inputs. For this, the properties of transformation function *F* of the system must be known. If there is enough knowledge about the system behavior, analytical methods can be used.

Theoretically, it is possible to observe all possible combinations of signals, although such analysis is very complex. Instead, a partial signal dependency analysis can be performed, where the correlation between different pairs of signals $(\mathbf{s}_i, \mathbf{s}_j)$ is determined. For correlated pairs further analysis and simplification is plausible. Based on this, function E_{ij} can be constructed.

$$c_{ij} = E_{ij}(\mathbf{s}_i, \mathbf{s})_{,j} \; c_{ij} \in C$$

If the correlation of pairs of signals is unsatisfactory, a combination of three or more signals can be attempted.

Establishing the monitoring function

The monitoring function is sometimes impossible to create analytically (for example when the details of the control system's operation are unknown). One solution is to use machine learning (ML) to find it. Unlike analytical construction, ML is based on the recorded operation of the control cell – the control cell is observed during an interval of valid operation to produce a learning set *L* of instances (points in the search space *S*), every instance representing the input signals ***s*** to the MC. The decision models created by both ML and analytical approach are in turn used to determine the correctness *c* of any signal ***s***. MC must determine whether a currently occurring signal instance ***s*** belongs to the space of known valid instances in *L* or not – a classification / clustering problem. The MC's processing

limitations prevent the use of complex models to discover the overlapping of current instance with *L*.

The decision model must divide the search space *S* into disjoint sub-spaces – clusters. Cluster is a small group of signals of the same type. The created clusters should hold the samples from the learning set and represent the *valid* sub-space; the sub-space out of cluster(s) is holding *invalid* points. The main problem is that clusters are computationally difficult to create and use. Consider the learning set in Fig. 2: the clustering is easily done by a human. However, significant computing power is needed to create *and* use this oddly looking region as a classification model. Moreover, the number of clusters is also unknown making the clustering process even more difficult.

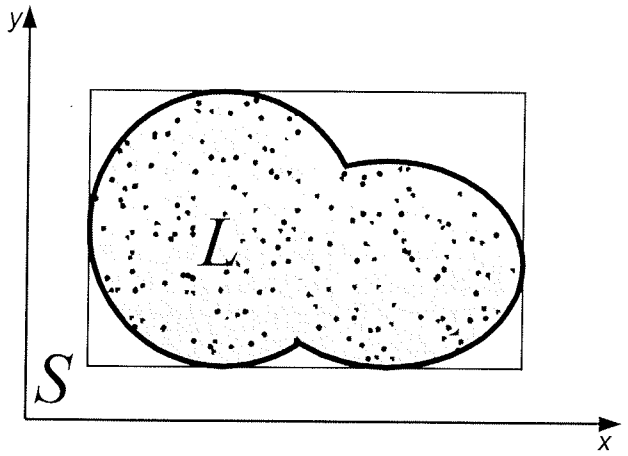


Fig. 2. Learning set enclosed in a human-drawn clustering region together with a bounding hyper-cube.

Hyper-cubes

The simplest and fastest monitoring function for the MC implemented in primitive hardware uses clusters in the shape of orthogonal hyper-cubes; a hyper-cube is limited by two opposing hyper-planes in every dimension. To avoid the troublesome creation and optimization of *n*-hyper-cubes a simple two-value (s_i, s_j) analysis is preferred. Since each component pair (s_i, s_j) is evaluated separately, a separate discretization of s_i is possible in the context of s_j and vice-versa. Such partitioning exploits the correlations between the parameters. The bounding hyper-cube (dash-dot rectangle in Fig. 2) is defined by the ranges of the values – the lower and upper limit are the simplest validation standards.

A completed model partitions the search space into disjoint sub-spaces with an outer bounding hyper-cube. The configuration of hyper-cubes is done off-line prior to employing the MC because sufficient processing power must be available to determine their size and position. The maximum number of hyper-cubes available for testing is defined by the MC's hardware. The whole process is an optimization task of maximizing the hyper-cubes while keeping the error to a minimum. Figure 3 shows one possible partitioning of the learning samples using three hyper-cubes $H_{1\text{-}3}$ covering *all* learning samples in *L*; the dashed regions inside *Hi* yet outside *L* are this model's error.

Constructing the hyper-cubes

The construction of hyper-cubes is based solely on the learning set. Unfortunately the learning set does not include all possible signals of the control cell. Even more, it does not contain a single *invalid* signal. It is impossible to collect all valid/invalid signals in advance or else the fault-detection itself would be unnecessary.

Invalid points are the only measure of error when creating hyper-cubes inside the bounding hyper-cube. If the algorithm is not given a human-drawn clustering region, it has to establish the cluster(s) by itself. The trivial hyper-cube which includes all points in *L* is the bounding hyper-cube itself. If the criteria are advanced to exclude invalid space, this space must first be recognized. In general it is impossible to assert whether a point in space is valid or not; if it is "close enough" to any existing valid point it could be valid, too. The question is how far is close enough.

First solution is to optimize for the *smallest* sub-space which includes whole *L*. Effectively, this method will create several *positive* hyper-cubes that together constitute a positive sub-space $V^+ = \cap H_i$ inside the bounding cube (see Fig. 3).

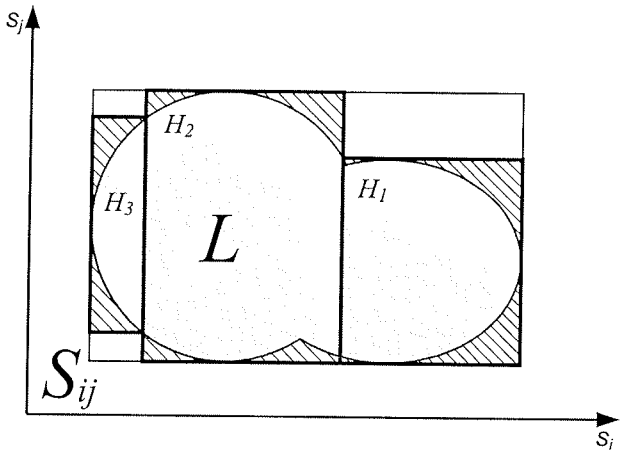


Fig. 3. One possible partitioning of the search space.

Second solution is the inverse of the first: *negative* hyper-cubes inside the bounding hyper-cube yet without common points with *L* are created. This scenario is shown in Fig. 4, where the three hyper-cubes X_{123} are positioned at the corners of the bounding hyper-cube to constitute the negative hyper-space *V*.

Any signal instance out of the bounding cube is always classified invalid. Also invalid are all points inside *V*, all other points *can* be valid. The optimization goal of both V^+ and *V* is to minimize the errors made by the clustering model.

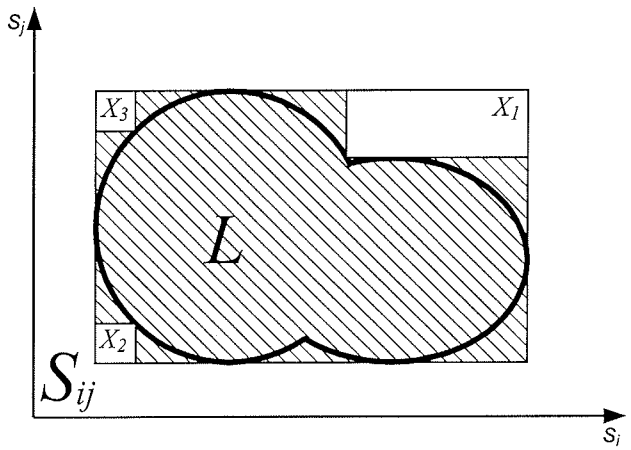


Fig. 4. Learning set fully outside of the negative hyper-space V⁻=)"Xi.

This single criterion optimization can be solved using a number of ML techniques, for example with evolutionary algorithms.

Evolutionary algorithms

The process of biological evolution by natural selection can be viewed as procedure for finding better solutions to some externally imposed problem of fitness[2]. Given a set of solutions (the initial population of individuals), selection reduces that set according to fitness, so that solutions with higher fitness are over-represented. A new population of solutions is then generated based on variations (mutation) and combinations (recombination) of the reduced population. Sometimes the new population will contain better solutions than the original. When this sequence of evaluation, selection, and recombination is repeated many times, the set of solutions (the population) will generally evolve toward greater fitness /2/.

Usually the evolutionary algorithms can be outperformed (EAs are rather slow) by the field-specific algorithms. For the purpose of optimization, however, EAs are excellent. Similar EA techniques differ in the implementation details and the nature of the particular problem. To optimize the MC's hyper-cubes all EA techniques are applicable; the most straightforward are genetic algorithms (GA, /3/) and differential evolution (DE, /6/).

For the negative model V^-, EAs need a fitness function that favors larger hyper-spaces V^- without valid signals. The simplest raw fitness function (5) divides the size of the hyper-space $|V^-|$ with the error $\epsilon(V^-)$, which is simply the count[3] of signals inside V^-.

$$f(V^-) = \frac{|V^-|}{\varepsilon(V^-)+1} \qquad (5)$$

For the positive model V^+, however, the fitness function must favor smaller V^+ and penalize any (valid) signals outside V^+. Again, higher values are better.

$$f(V^+) = \frac{|S|-|V^+|}{|L|-\varepsilon(V^+)+1} \qquad (6)$$

Example

An example result is shown in figure 5. For the simplicity only two variables (an input and an output) are used. The observed control cell has non-linear logarithmic characteristic. The learning algorithm was differential evolution (DE) employing positive fitness model V^+ (6), 30.000 iterations and default DE values for other settings. DE was used off-line to produce four hyper-cubes that included 42 distinct two-dimensional signals with a 12 bit resolution. In this example the depicted positive model V^+ is better than the V^- model because the points are relatively well aligned.

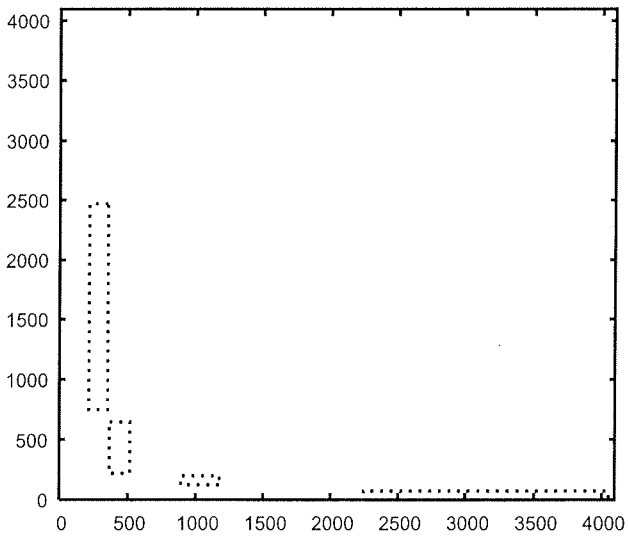


Fig. 5. A positive model of four hyper-cubes for 2 non-linear signals.

Implementations of the MC

MC can be implemented in several ways. If original application code is available and can be modified, the MC can be implemented as a set of monitoring routines that run together with the control software on the same processing resources. After the input is acquired, the pre-evaluation routine is called. It evaluates the individual signals, quantifies them and stores them for later assessment. Similarly, post-evaluation routine is called before output is produced.

2 *An equivalent term utility can be used in place of fitness.*

3 *∈ (X) is the cardinal number |F| of set F, where F={s; s ∈ L ∧ s ∈ X}*

Additionally it also checks for proper mappings between inputs and outputs. The main disadvantage of this approach is that modifying the original code may distort the temporal characteristic of the system. Also changing of the original application is rarely possible. Because of this, this approach was not taken in the research.

Higher degree of dependability and agility can be achieved by using dedicated hardware solutions. Continuous reduction of prices for the hardware makes this approach economically feasible. The monitoring device should run in two different modes of operation. In the first mode it measures and records the input and output signals. These measurements are consequently used as a learning base for the off-line construction of the monitoring function. In the second mode of operation it actually monitors the control system. By using the same device for both sampling and evaluation any differences between measurement and monitoring hardware can be eliminated.

The conceptual diagram of MC monitoring hardware implementation is shown in Figure 6. Signals from the control system ***s*** are connected to a set of registers (implemented as two-stage FIFO buffers) that hold their current value ***s(t)*** and the previous values ***s(t-1)***. Analogue and numeric values are processed by the quantization blocks (labeled Q in the Figure 6). Each quantization block transforms the input into corresponding discrete value in order to simplify the evaluation. These transformations are determined by the off-line learning phase (see 4). Signals ***s*** may also contain some internal states ***m*** of the control system. For the hardware implementation of MC, however, special access points must provide current values of internal states.

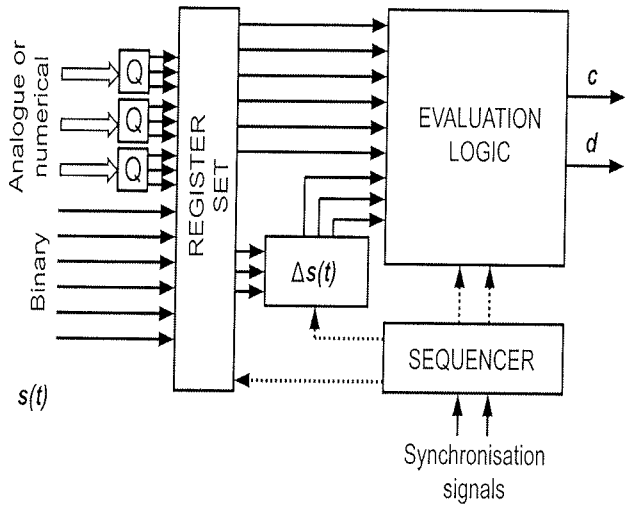


Fig. 6. Implementation of the monitoring cell.

To observe the dynamics of the signal, the block $\Delta s_i(t)$ compares new and previous values and calculates the difference producing additional information for the evaluation.

Parallel to the execution of the monitoring function certain physical characteristics of the control cell hardware can be observed allowing for detection of any abnormalities (e.g. increased temperature or current consumption). It is also possible to validate temporal properties of the observed process using simple watchdog timers.

Components of the MC are synchronized with a simple sequencer according to external synchronization signals that indicate the beginning of each execution cycle. If no such explicit signal exists, it can be derived from other signals (e.g., from control lines of I/O devices). Signals like reset, power-save, etc. should also be considered to determine the system's current mode of operation – the monitoring logic is not applicable in all modes.

The outputs *c* and *d* are generated by the evaluation logic. Mostly, *c* is just a status that alerts for the (potential) fault in the system. In most cases *c* can be determined by using pure digital logic. For diagnostic output *d*, which gives more detailed information about the cause of the fault, more complex logic may be required.

Implementation of MC with discrete components

In the fist case study, a solution with discrete analogue devices, FPGA chip and a simple microcontroller is explained. For the A/D conversion 12 bit A/D converters ADS7841 with serial communication interface were used. The FPGA was Spartan-IIe with 3072 programmable slices and 8 KB of dedicated memory blocks. This hardware was used in other experiments /8/.

The monitoring logic for one of the analogue signals is shown in Figure 7.

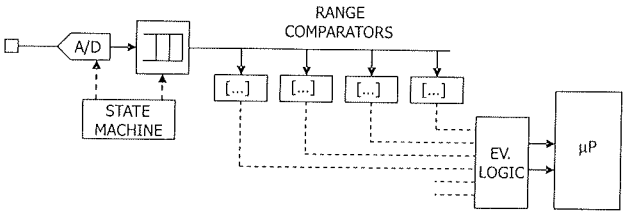


Fig. 7. Discrete monitoring logic for a single analogue signal.

First, the analogue signal is acquired by A/D converter and captured using simple state machine automaton. The state machine periodically generates control signals that trigger A/D conversion and, after conversion is done, reads the result. By utilizing FPGA device, it was possible to implement a number of state machines that are working in parallel, each serving a single A/D interface. To limit the number of wires needed for the communication, the A/D converters with serial communication protocol were used. From the 12 bits provided by the A/D device only the upper six were used. Similarly, the discrete digital signals are acquired periodically by other state machines.

To compensate for the delays between input and outputs, and to evaluate the dynamic characteristic of the signals, each sample goes through a simple two-stage FIFO buffer. Later, both stages are available for the evaluation.

The basic evaluation is performed by a set of range comparators, each testing if the input value is within the predetermined range. The results from this and other evaluation channels are then combined with simple Boolean evaluation logic, which is implemented as a truth table inside the dedicated memory blocks of the FPGA – the status signals are interpreted as a memory address containing the appropriate output. The output *c* states if all signals are in the valid ranges; output *d* is a vector designating validity of individual inputs. All constants for the range comparators and the content of the memory blocks are generated off-line and may be changed only by the full reprogramming of the FPGA device. For the implementation of evaluation logic for a single evaluation channel with a four range comparators approximately 150 slices are used.

For the communication with the fault management system, a simple 8 bit microcontroller is used. It is also needed for the initialization of the MC at start-up, for initial data acquisition, and later for the debugging and diagnostic. It is possible to transfer the microcontroller into the FPGA; however, some functionality (e.g. enhanced debugging) is lost.

Implementation of MC with programmable SoC

The above solution requires separate analogue circuits for the analogue signal manipulation. Nowadays, more compact and low cost novel technology with programmable system-on-chip is available. Those chips integrate a microcontroller and configurable blocks of analogue and digital logic. In the case study the PSoC CY8C29466 Mixed Signal Array is used. PSoC is a trademark for a family of programmable SoC devices from Cypres. PSoC devices include configurable blocks of analogue and digital logic, as well as programmable interconnects. Additionally, a fast CPU, Flash program memory, SRAM data memory, and configurable I/O ports are included. The analogue part is composed of dozen configurable blocks, each allowing the creation of complex analogue functions like A/D and D/A converters, comparators, filters, amplifiers, etc. More complex functions are implemented by combining several primitive cells. The digital part is composed of several digital blocks. Each block is an 8-bit resource that can be used alone or combined with other blocks to form 8, 16, 24, and 32-bit peripherals. The capabilities of those blocks are greater than its counterpart in typical FPGA device and may be configured as counters and timers, PWMs, different serial communication interfaces etc. CPU has full control over the configuration of the analogue and digital blocks.

The conceptual diagram of the MC evaluation logic with PSoC is shown in Figure 8. Each channel consists of a Programmable Gain Amplifier (PGA) and a six bit Successive Approximation Register (SAR) A/D converter. The PGA allows for adapting to different signal levels. This way low-voltage signals can be observed. To observe quantities larger than 5V, an off-chip voltage divider is required. By using the six bit A/D converter, the quantification was included in the conversion. Other possible implementations of A/D conversion with the PSoC device exists, however, they are either slower or consume more analogue cells. The hardware provides four analogue data acquisition channels that can be expanded to eight by using the also provided two-way analogue multiplexers. The digital part of the device allows up to sixteen bit of discrete data acquisition. Although it was possible to implement buffering and preliminary signal evaluation with digital cells on the device, they were implemented in software because of various limitations of the digital part.

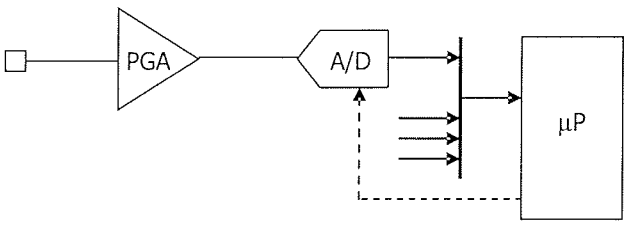


Fig. 8. Logical organization of a single analogue acquisition channel with PSoC.

The evaluation logic is executed by the microprocessor. It is implemented as a series of tests that check the inclusion of a variable in hyper-cubes. The hyper-cubes are prepared off-line and are loaded to the device during the initialization phase. The solution used by FPGA (i.e., the implementation with truth tables in memory blocks) would use more memory than is available.

Comparison of different implementations of MC

The solution with discrete analogue devices is somewhat more flexible and robust. It can use different kinds of A/D converters to accommodate various kinds of signals. In addition, the external A/D converters are usually much less sensitive to the voltage overloads. The main benefit of using this approach is the speed because multiple monitoring channels and evaluation logic in the same device can be constructed. The sampling, quantization and evaluation for different signals occur in parallel. If the evaluation logic is simple enough, the execution cycle can be in range of several micro-seconds. This is much shorter than the time needed for the A/D conversion. Therefore, it is possible to evaluate the signals from one execution cycle during the acquisition of the signals from the next one. The drawback of this solution is the price. The estimated price for the parts used in the experiments is more than 30 euros and it increases with each additional A/D converter. In contrast the programmable SoC solution can cost (for up to the 4 analogue channels) less than 5 euros.

The solution with programmable SoC devices is much more compact and requires less supporting components than the previous one. However, apart from the A/D conversion, all of the processing of data is performed with the microprocessor. This impacts the execution time because all evaluations must be done sequentially. For example, for the configuration where two analogue signals were observed and evaluated with four regions (like in the figure 6), the execution time of the evaluation was 83 microseconds (with 24MHz system clock). On the other hand the typical A/D conversion takes only 25 microseconds. Therefore the evaluation can not be performed in parallel with the conversion. If more dimensions are needed, the execution time increases accordingly. This is in contrast to the first approach where additional variable have almost no impact to the execution time due the parallel nature of the execution.

The case studies use only simple evaluation functions although more complex regions than the hyper-cubes could be used that require higher mathematical operations. The FPGA is unsuitable for such calculations, the PSoC however, is powerful enough to implement them (e.g., it has two dedicated fast 8 by 8 bit multipliers).

Conclusions and future work

In the paper a concept of monitoring cell for supervision of the plausibility of a control function is presented. Knowing (or having learned) the normal behavior of the control function, one can detect abnormal behavior of input and output signals, internal states, dynamics, and coarse behavior of the output function with respect to the inputs in the previous time instance. For situations where the control function is not known in details, the automatic process based on the machine-learning principles is proposed. The samples of signals acquired from the longer period of time are analyzed and categorized into a set of multi dimensional regions. Two different solutions for MC implementation are described. The proposed devices can be relatively easily applied in variety of existing control systems. The first solution is appropriate when short response times are required. Due to the multiprocessing nature of evaluation logic used, the execution time is almost independent of the number of signals observed. However, the solution is relatively expensive, harder to implement and can be used only with simple evaluation functions. The approach with programmable SoC devices is more compact, simpler to construct and more appropriate for more sophisticated evaluation functions. However, because the main processing is done in software, the reaction times are much longer. The ideal solution would be the combination of both approaches: programmable mixed signal SoC device with large FPGA on a single chip. To our knowledge, there is no commercially available device of such kind on the market yet.

In the future work, other classification models will be considered, possibly more powerful yet of similar complexity as the hyper-cubes. Part of the research is focused on utilizing genetic programming (GP) because it is capable of directly producing the control function of the monitoring cell. The idea is to find appropriate set of functions that are easy enough to be implemented on simple hardware and yet good enough to appropriately describe the valid control signals. The other part of the research is focused on the implementation of those functions. The final goal is to include a machine-learning algorithm inside the monitoring device where evaluation function can be determined dynamically and might adapt to different operation modes of the system.

References

/1/ L.H. Chiang, E.L. Russell, and R.D.Braatz. Fault Detection and Diagnosis in Industrial Systems. Springer, 2001.

/2/ D.W. Hillis. Co-evolving parasites improve simulated evolution as an optimization procedure. Physica D, 42:228-234, 1990.

/3/ J.H. Holland. Adaptation in natural and artificial systems. The University of Michigan Press, Ann Arbor, MI, 1975.

/4/ 1149.1 IEEE. Test Access Port & Boundary Scan Architecture. IEEE, New York, 1990.

/5/ R. Isermann. Fault Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance. Springer, 2005.

/6/ K. Price and R. Storn. Differential evolution: A simple evolution strategy for fast optimization. Dr. Dobbs Journal of Software Tools, 22(4):18-24, 1997.

/7/ Neil Storey. Safety-Critical Computer Systems. Addison-Wesley Longman, 1996.

/8/ D. Verber, B. Lent, and W.A. Halang. Firmware support for disjunctive dataflow driven distributed control applications. In Z. Bradač, F. Zezulka, M. Polanski, and V. Jirsik, editors, Proceedings of IFAC workshop on programmable devices and embedded systems PDeS 2006, pages 84-89, 2006.

Domen Verber, Matej Šprogar, Matjaž Colnarič
University of Maribor Faculty of Electrical Engineering and Computer Science, Smetanova 17, 2000 Maribor, Slovenia
domen.verber@uni-mb.si

Prispelo (Arrived): 10.08.2006 *Sprejeto (Accepted):* 30.03.2007

MICROBAROMETER IN A VIRTUAL SYSTEM

Žarko Gorup

University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia

Key words: Virtual instruments, microbarometer, data acquisition systems, signal processing

Abstract: A new virtual system for the detection of small pressure variations in the infrasound band of the characteristic Bora wind from the North Adriatic coast has been developed. It is composed of a precision pressure transducer (PPT), standard data acquisition (DAQ) card and virtual instruments (VIs) in the LabVIEW environment. Data collection and analysis of small atmospheric pressure variations during the Bora need a unique pressure sensor and signal conditioning electronics. We have fulfilled several major requirements of the measuring system: sufficient frequency bandwidth (0 Hz to 20 Hz), the resolution of the PPT response and low-noise design, a changeable and adjustable graphic user interface, configurability of the system, and analytical tools for data elaboration. Sufficient linearity is obtained by appropriate selection of a PPT with a signal conditioning circuit. Several efficient methods for the reduction of noise in electronic circuits have been used. The best effect in signal conditioning is achieved with low-noise electronic components, an appropriate circuit layout, and correct cabling and ground connection. Signals generated in the PPT are conditioned, acquired and analyzed under the control of a particular virtual instrument (VI) and subinstruments (subVI). DAQ and analysis are supervised from the front panel of a particular VI. The principal aim of the data analysis is the power density function (PDF) obtained with certain complex virtual VIs. The custom-designed Power Spectrum Analyzer (PSA) VI offers many useful options: graphical presentation of time and frequency, data records selection, changeable windowing and measurement of noise level.

Mikrobarometer v virtualnem sistemu

Kjučne besede: Virtualni instrumenti, mikrobarometer, sistemi za zajemanje podatkov, procesiranje signalov

Izvleček: Delo opisuje virtualni sistem za zaznavanje in merjenje majhnih sprememb zračnega tlaka. Merjene spremembe povzroča burja, ki je značilen veter severnega Jadrana. Merilno napravo sestavlja precizen senzor tlaka, standardna večfunkcijska kartica za zajemanje podatkov in virtualni instrumenti v programskem okolju LabVIEW. Zbiranje in analiza podatkov o spremembah zračnega tlaka, ki jih povzroča burja, zahteva poseben senzor tlaka in elektronsko vezje za prilagajanje signalov. Realizirani sistem izpolnjuje vse glavne zahteve merilnega sistema: ustrezen frekvenčni pas (0 Hz do 20 Hz), ločljivost preciznega senzorja tlaka in malošumna izvedba vhodnega vezja. Poleg tega načrtovani sistem omogoča spremenljiv in prilagodljiv uporabniški vmesnik, fleksibilno zgradbo sistema in analitična orodja za obdelavo podatkov. Linearnost odziva smo dosegli s primernim senzorjem in vezjem za prilagajanje signalov. Zmanjšanje šuma smo dosegli z uporabo metod načrtovanja malošumnih vezij. V vezju za prilagajanje signalov smo uporabili malošumne elektronske komponente, primerno načrtovano tiskano vezje ter pravilne medsebojne povezave in povezavo mas. Signali, ki jih generira senzor se prilagajajo, zajemajo in analizirajo pod nadzorom načrtovanega virtualnega instrumenta (VI) in virtualnih podinstrumentov. Zajemanje podatkov in analizo lahko spremljamo prek čelne plošče določenega VI. Glavni cilj analize podatkov je izračun in prikaz funkcije gostote močnostnega spektra, do katere pridemo z gradnjo kompleksnih VI. Zgradili smo uporabniško orientiran VI "analizator močnostnega spektra", ki nudi več uporabnih funkcij: grafično predstavitev spektra v časovnem in frekvenčnem prostoru, izbiro podatkovnega niza, izbiro okenske funkcije in meritev šumnega nivoja.

1 Introduction

Infrasound is inaudible sound with frequencies below the human hearing threshold of 20 Hz. The lower frequency cut-off of infrasound is limited by the thickness of the atmosphere. In general, infrasound is measured within a frequency range of 0.005 (T=200 s) to 20 Hz (T=50 ms). Within this frequency band, many sources of both known and unknown origin generate infrasound. Sources that can often be detected for seconds, minutes or hours are: winds, volcanoes, sea waves, explosions, meteors, mountain-associated waves and aurora. Infrasound can be measured with either a low-frequency microphone or a high-frequency barometer. In the project we preferred the use of a microbarometer because of its robustness with respect to field application and durability. Furthermore, microbarometers can measure much longer periods than microphones.

Measurements of small and rapid variations of atmospheric pressure have historically been made by a variety of methods. All of them are enabled by means of different types of pressure transducers /1/ connected to data acquisition, elaboration and storage systems. Standard analog barometers have resolutions in the range of hPa (or mb). Precise measurements of atmospheric pressure changes are made with microbarometers with resolution in the range of 0.1 hPa to 0.002 hPa /2/, /3/. Such instruments are suitable for observation of fluctuations of the pressure of the wind. The Bora in the Karst region (Kras, Carso) is a particular type of wind in which the energy peaks are pulsations. They are of particular interest because they disappear and reappear within an episode /4/. Spectral analysis of the Bora reveals information about energy content, time distribution and local characteristics.

A better understanding of the principal physical characteristics of the Bora provides a key to solving local problems in the areas of construction, architecture, environmental planning, agriculture, traffic control, etc. This can lead to cost reductions and an increase in the quality of life. A preliminary spectrum analysis was carried out on patterns of atmospheric pressure oscillations generated by the Bora, as recorded in Trieste (Italy) /5/.

Bora wind gusts produce immediate changes of pressure that the pressure transducer has to convert proportionally to a voltage. A microbarometer with a very accurate, sta-

ble and sensitive built-in pressure transducer is needed for such purposes. Some commercially available pressure transducers match the specifications for accurate continuous measurements of instantaneous atmospheric pressure changes. The microbarometer employed uses a high-accuracy pressure transducer system made by Setra Systems Inc. /6/and /7/. The PPT is connected to a custom-designed circuit which performs conditioning of the analog signal.

All measurements of low-level signals are subject to noise. Various methods describe the reduction of noise in signal conditioning circuits, but they are difficult to implement in practice /8/. With regard to noise, there are some significant factors, among which the most influential are grounding, shielding, proper ground connection and suitable circuit layout strategy. On the other hand, there are several methods for lowering noise in the measuring signal. The most well known is analog filtering, which is always used in antialiasing circuits. Another method is averaging, which can be performed in hardware or in the software of the measuring system.

The analog signal terminates in a DAQ system. Many standard PC-compatible plug-in DAQ cards offer many advantages in comparison with classic measuring instruments. Today's multifunction cards are reliable, accurate and cost effective. Moreover, they use plug-and-play technology and are compatible with the Windows operating system and with most software packages designed for building VIs. General purpose VIs, such as multimeters, function generators and oscilloscopes, are frequently offered, but special measuring VIs are rarely available on the market.

Choosing the right DAQ card is very important for the design of a measuring system. The selection should be made on the basis of electronic specifications, with close correlation of the card's characteristics and measurement needs. Full hardware compatibility with LabVIEW plays an important role in the final choice of a DAQ card.

A virtual instrument (VI) can be created in the LabVIEW environment. Using VIs, a virtual measuring system can be set up which enables the user to carry out measurements on various subassemblies under specific program control. This kind of instrument has some advantages compared with classical or traditional measuring instruments /9/.

The challenge of the project was to design a VI-based microbarometer for spectral analysis of the Bora. The system must be capable of acquiring samples from the PPT, performing measurements using an appropriate standard DAQ card /10/, and processing and storing the acquired signal. The solution was combining a high-performance plug-in data acquisition card and signal conditioning in order to obtain a precise measurement of wind parameters. The VI enables the user to acquire data from the PPT (located outdoor) and process the data, calculate performance results, log data and generate reports /11/.

2 Transducer unit

Precision pressure transducer

The Precision Pressure Transducer (PPT) is a compact sensor device that generates a very sensitive analog output signal proportional to the pressure being measured. The PPT unit has an analog output and is individually calibrated at the factory for temperature variations over the full scale pressure span across a -40 to 85°C range.

The PPT used is a Setra Systems Inc. Model 270 analog sensor which measures absolute (atmospheric) pressure. It contains a capacitor as a sensing element, which results in a simple, durable and fundamentally stable device. The sensing scheme is presented in Figure 1.

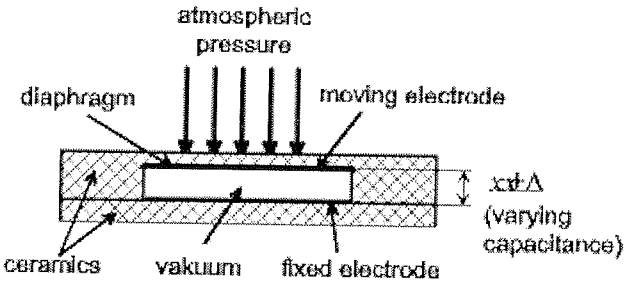


Fig. 1: Cross-section of a capacitive absolute pressure sensor.

In a typical configuration, a compact housing encloses two closely-spaced, parallel, electrically-isolated metallic surfaces. One surface is essentially a diaphragm capable of slight flexing under applied air pressure. The diaphragm is constructed of a low-hysteresis material, such as 17-4 PH SS or a proprietary compound of fused glass and ceramic: Setraceram. These firmly secured surfaces (or plates) are mounted so that a slight mechanical flexing of the assembly Dx, caused by a change in applied pressure, alters the gap x between them to $x \pm Dx$ (creating, in effect, a variable capacitor), as depicted in Figure 1.

Capacitive pressure transducers are not common but do have higher sensitivity to pressure changes than do other pressure-sensing devices (typically 10 to 100 times). They are much less sensitive to thermal stresses and local diaphragm stresses since capacitive transducers integrate the movement of the entire surface of the diaphragm, while piezoresistive (PR) and piezoelectric (PE) devices use localized strain measurements. Capacitive transducers commonly have small capacities and generally are more expensive and larger than other devices because they usually carry their signal conditioning circuitry on the same board as the sensor.

The sensor alone is very sensitive to environmental coupling, so it must be mounted in a mechanically and electrically protected space. In the case of Setra's PPT, protection is afforded by a stainless steel housing to cover the sensor and functional electronics (excitation circuit, amplifier, power supply, analog buffer, etc.). To reduce noise

and increase the signal-to-noise ratio, the amplifier unit must be located as close as possible to the sensor element.

The capacitance of the sensor is a pure measuring value, so it must not be changed in the remaining circuitry. Signals containing the measurement are conducted to the signal conditioning circuit. The sensitivity of the signal-conducting wires is such that all other conductors need a shield.

3 Data acquisition

The data acquisition (DAQ) unit is an important subassembly with the main functions of conditioning and multiplexing input channels, digitalization of input analog signals, timing control, synchronization and providing a digital I/O port. Usually we consider the DAQ card as a DAQ system, but in this case we must distinguish between the DAQ card and DAQ process. The DAQ process includes the conditioning part of the PPT and relevant software control.

The data acquisition system (DAS) contains two main subunits: signal conditioning and signal conversion. Signal conditioning is partially accomplished by the PPT unit, while the conversion is on the input stage of the DAQ multifunction card. Analog-to-digital (ADC) signal conversion with all the necessary logic and data path circuitry is part of the multifunction DAQ card.

3.1 Signal and data flow

Once an analog signal from the sensor output is achieved, signal conditioning is applied. The main premise of the signal conditioning is linearity and response of the analog signal. After that, signal conversion follows and the data become digital information. The basic scheme of signal and data flow is presented in Figure 2.

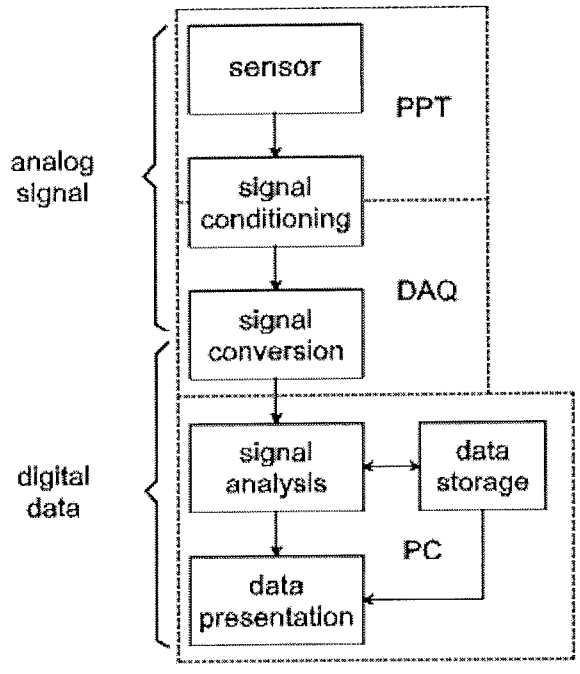


Fig. 2: Signal and data flow from PPT to data storage.

Analog signal flow begins with the PPT as the generator and ends with the ADC. Because of the very low-level signals of the first part of the circuit, they are very sensitive to noise and other disturbances from the surrounding environment.

The signal path is followed by temporary data storage and data elaboration. Under the control of software, suitable data records are created and signal analysis is performed. The virtual system permits the user to observe the results graphically in quasi real time. At the same time, data are stored on two independent memory media for later processing and analysis.

3.2 Signal conditioning

A linear transformation between signal excitation and response must be achieved by the signal conditioning circuit. Components that can be driven to the limits of the supply voltage (rail-to-rail) make the maximum dynamic range available. Signal conditioning is performed in two successive phases: the first phase is accomplished on the sensor unit and the second is performed on the standard DAQ multifunction board.

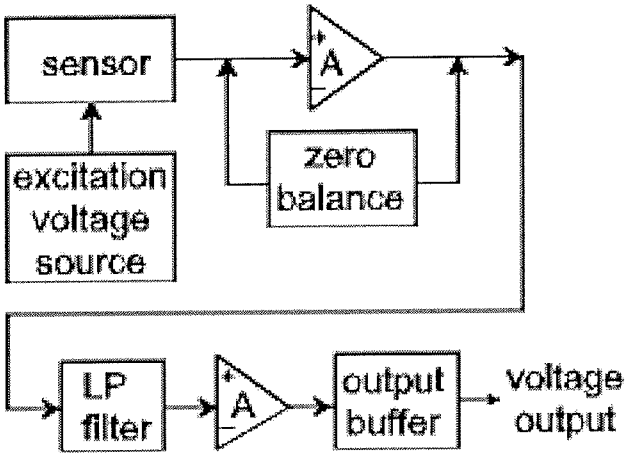


Fig. 3: Signal conditioning circuit.

The main task of the signal conditioning circuit is adequate signal propagation, during which the signal is set to fit the ADC input voltage swing. Signal amplification reduces noise and protects signals from other disturbances. To increase dynamic range in this low-voltage environment, the first operational amplifier A in Figure 3 processes rail-to-rail input signals and drives rail-to-rail output signals. A circuit that maintains constant transconductance g_m over the input common mode ranges reduces distortion. The bandwidth of the signal conditioning circuit must provide a frequency response within the range of interest for the fastest rate of change of the signal.

The antialiasing filter is one of the most important elements of the DAQ system /12/. It is impossible to differentiate between noise with frequencies in the band and out of the

band of interest. Only an analog filter can preserve signal integrity and extract real frequencies from folded frequency components. The cut-off frequency of the antialiasing filter must lie below the Nyquist frequency. The method for determining and implementing the appropriate analog filter parameters is controlled using Microchip's "FilterLab" software. The proposed circuit has been exported and simulated by PSPICE (Cadence Inc.) and the result is presented in Figure 4.

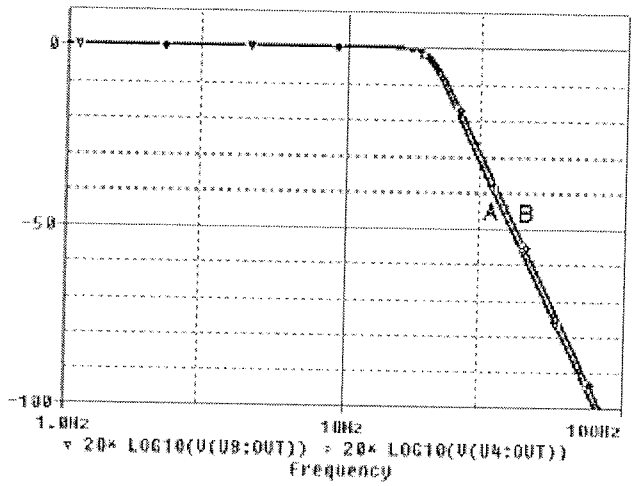


Fig. 4: Calculated (A) and realized (B) transfer function of the antialiasing filter.

Implementation with a Butterworth filter design has been carried out. A Butterworth filter is used in the filter implementation of the antialiasing filter. For this circuit, an 8^{th}-order filter is used with a cut-off frequency of 10 Hz. Four active Sallen-Key filters are used. This filter attenuates the pass band signal to 80dB at a frequency of 20 Hz.

The effect of the filter is clearly visible in Figure 5. The output signal with noise is measured directly in the illustration above and passing the filter stage in the illustration below.

3.3 Signal Conversion

The ADC is an Analog Devices successive approximation register type with a maximum 200 kS/s conversion rate. It operates from a single 5 V power supply. The resolution of the ADC is 16-bit, or 1-in-65536. The converter's integrated circuit contains a high-speed 16-bit sampling ADC, an internal conversion clock, internal reference, error correction circuits, and both serial and parallel system interface ports. The ADC is fabricated using a high-performance, 0.6 mm CMOS processor and is operable from -40°C to +85°C.

The single 5 V supply of the ADC typically dissipates only 35 mW. Its power dissipation decreases with throughput to, for instance, only 15 µW at a 100 SPS throughput. It consumes 7 µW maximum when in power-down mode. The circuit has superior integral nonlinearity (INL). A maximum

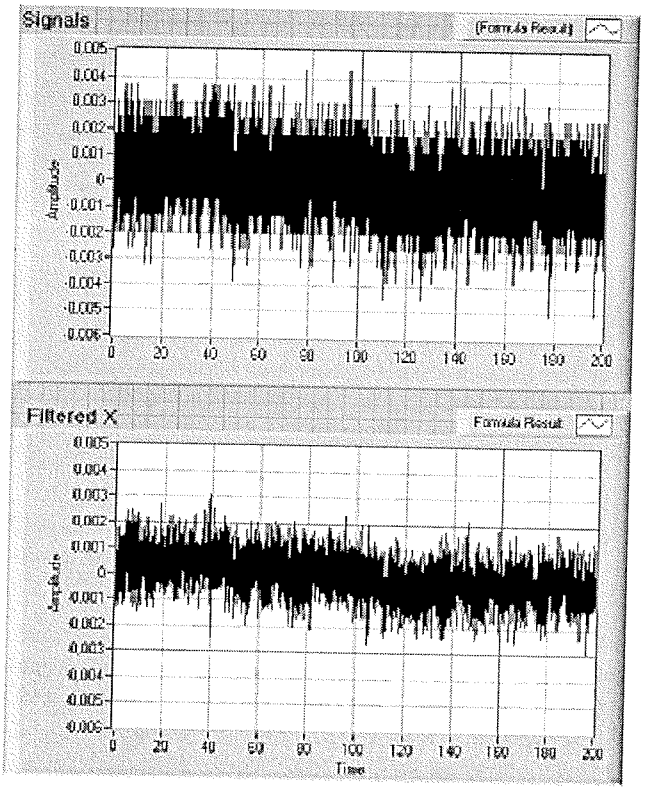


Fig. 5: Output noise level of the output signal without (above) and with filtering (below).

INL of 3 LSB with no missing 16-bit code is achievable. Serial or versatile parallel interfaces (8 bits or 16 bits) or a 2-wire serial interface arrangement compatible with both 3 V and 5 V logic are available.

Conversion is also possible in burst mode as a software-selectable option at a burst rate of T = 5 µs. The ADC uses a RAM buffer which holds 8 K samples. Data transfer can be programmed input-output (I/O) or direct memory access (DMA). DMA modes are "demand" or "non-demand" using scatter/gather operations. Configuration memory contains up to 8 K elements, which can be stored data for channel programming, gain, and offset of the on-board signal conditioning circuit. Maximum sampling rate reaches 200 kS/s with single channel operation and a 15-minute warm-up. Accuracy is assured by internal calibration. Measurements are valid for operational temperatures within ±1 °C of internal calibration temperature and ±10 °C of factory calibration temperature.

All hardware configuration options on the multifunction card are software controlled. Some configuration options, such as digital channel configuration (input or output), have been configured with installation software (InstaCal). Once selected, any program that uses the Universal Library can initialize the hardware according to these settings.

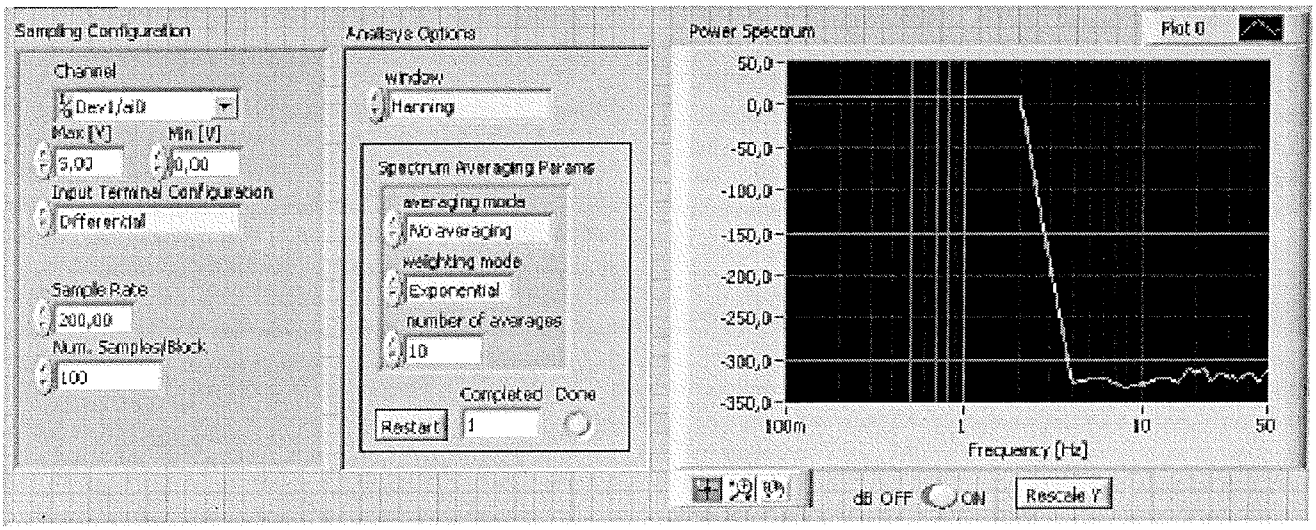


Fig. 6: Front panel of power spectrum VI.

4 Virtual instrument for data acquisition and analysis

4.1 Configurability and design of the system

One of the goals of the project was the design of a virtual system controlled by a VI. This permits a very flexible configuration with possibilities of choosing or changing various parameters and settings. Combining of different subVI hardware and software operations can also be performed, plus complex mathematical operations involved in signal analysis.

With the VI, the average power spectrum of an input signal in the time domain can be computed. Various averaging modes for measurement have been tested, such as RMS averaging, vector averaging, or peak hold, as well as the number of averages. The number of averages influences the noise floor. Vector averaging requires the use of a trigger in order to lower the noise floor without lowering the fundamental signal along with it. VI has various filter selections, where the type of window used in the particular measurement can be selected.

Once all properties of the system were known and defined, two-pass top-down software design started. Initially, a front panel was created. Functions were added and tested successively, continually being aware of the main goal. An example of the front panel of power spectrum VI is presented in Figure 6.

The supervisory part of the software was built in the second pass of the design process. The block diagram responsible for program execution contains primitives and subVIs. The block diagram for wind analysis data is conceptually created with DAQ control and an analytical part.

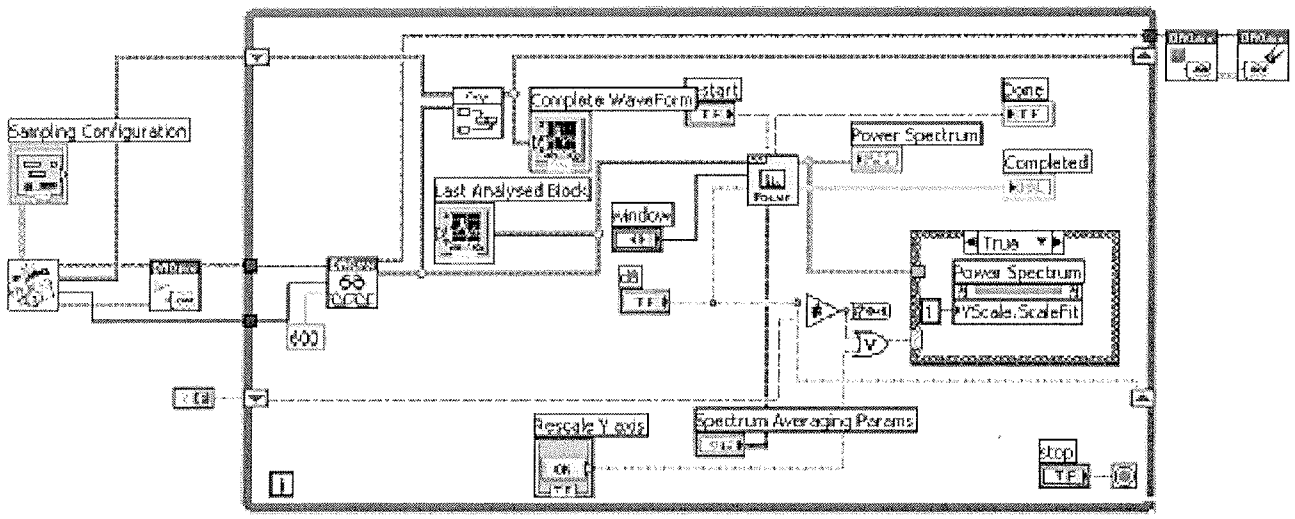


Fig. 7: Partial block diagram of the power spectrum VI.

The control part manages the system hardware. It contains various settings to optimize acquisition parameters and visualization of data. The Functions menu of LabVIEW's analysis module offers all the useful operations and procedures needed to assemble a signal analysis VI. We used several suitable subVIs to facilitate and speed up the design process. A view of a part of the block diagram created as a power spectrum VI is presented in Figure 7.

4.2 Signal analysis and data presentation

The main objective of the analysis is to determine the spectral composition of the observed signals of pressure fluctuations in order to determine characteristic features as a function of weather conditions, in particular of wind direction and speed.

It has been assumed that the observed signals result from a superposition of low-frequency waves (with a range of time periods extending from several minutes to several hours) and higher-frequency waves. Overall power is the sum of the contribution of each component. PSD was estimated using a built-in Fast Fourier Transform (FFT) algorithm. Despite the limitations implicit in the limited frequency resolution and the power "leakage" which results from data windowing, this method is computationally efficient and obtains reliable results for a wide range of processes.

A LabVIEW-based computer program (VI) was developed for spectral analysis of large data samples. Data records were segmented into time intervals corresponding to the duration of the analyzed phenomenon. The most interesting data samples can also be presented on a graphic interface over time. Our analysis succeeded in the following steps:

a) trend removal,

b) low and high frequency limit,

c) estimation of power spectral density.

For the proper data presentation one has to know and understand the physical phenomena and the origin of the acquired data. There are several sources which determine the micro-fluctuations of atmospheric pressure extending over a wide range of frequencies (which may involve several orders of magnitude). Among them are frequently observed global and local configurations of the earth's surface. The Bora is a typical dry wind from the Julian Alps, characterized by strong gusts (over 50 m/s) blowing over the Karst and the Gulf of Trieste, predominantly in winter. It also reaches the surrounding countryside, but in a weakened form. The analysis refers to samples of Bora wind recorded in Trieste on a windy day over a time span of six hours.

An example for the verification of system operation is the data acquisition system registering pressure fluctuations on the PPT with sampling rate of 100 samples/s. Upon removal of linear and seasonal trends, a low-pass filter (by means of a moving-average model) was applied to the data.

Data elaboration was carried out by various LabVIEW sub-VI's found in the Mathematics & Statistics module.

On windy days, thunderstorms can also very frequently occur. Such cases present an additional source of noise, i.e. atmospheric noise. Atmospheric noise is a non-stationary process that introduces non-regular disturbances in the measuring signal. Detection and elimination of this kind of noise will be the task of our further work.

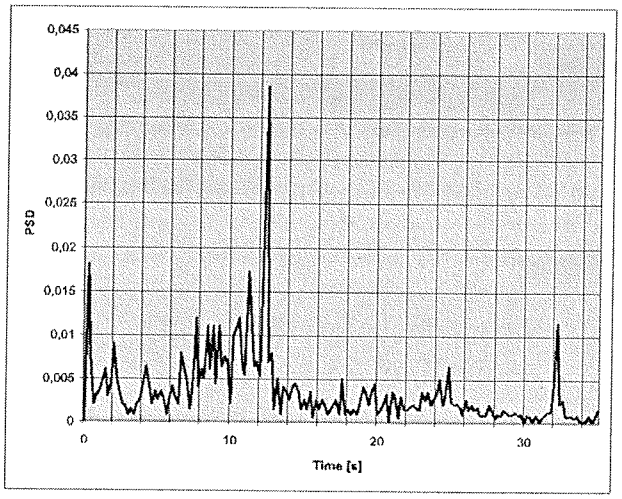


Fig. 8: Power spectral density of the Bora in the Gulf of Trieste.

The results of spectrum analysis carried out using the FFT method is presented in Figure 8. Frequency-to-time transformation of the PSD was used to facilitate physical interpretation of the phenomena. The analysis demonstrates the existence of notable features during peak force of the wind (a period of approximately 12 seconds). These results also agree with recent measurements made by Belusic et al. /4/ as well as observations conducted over fifty years ago by Mosetti /5/using a classical mechanical microbarograph.

5 Conclusions

The characteristics of the Bora wind can be observed using a microbarometer. The availability of powerful, low-cost microprocessors and memory has eased the task of designing the signal processing necessary for high-performance air data collection. Even with the sophisticated digital signal processing provided by microprocessors, the performance of air data instruments is still strongly dictated by pressure transducer characteristics: basic accuracy, resolution, long-term stability and reliability over the spectrum of environmental conditions encountered in ecological investigations.

Digital-output pressure transducers have on-board ADCs, which, all else being equal, provide higher performance and easier integration of air data measurements into PC analysis and the presentation environment.

The usage of a suitable sensor and a particular electronic arrangement leads to substantial noise reduction and offers the ability to determine source characteristics such as apparent PSD and RMS noise. A VI with a graphical user interface enables the configuration of important features of acquisition and analysis, so that various types of winds, which all cause instantaneous variations of pressure, can be distinguished. This offers a new, simple, intuitive front-panel user interface with versatile functionality. The data acquisition of very low pressure changes is applied as a monitoring technique for determination of wind characteristics. The processing and interpretation of pressure micro-variation recordings has been illustrated with data from the Gulf of Trieste. Current research efforts are concentrated on hardware and software improvements: integration of micro-controlled signal conditioning and conversion electronics on board, further reduction of noise on the analog signal path, wireless digital data transmission and web integration of the virtual system.

With the LabVIEW program users can create a PPT-interface VI with custom capability for both single PPT and use with a DAQ card. Users can easily set up an interface to the PPT for establishing communication, issuing commands, configuring the device, charting data in real time and storing data to text files. This project also demonstrated the usage of the LabVIEW program for creation and utilization of more complex programs. Software development in a graphical environment can be quickly understood and it is easy to use. VI proves to be the optimal solution for projects where a nonstandard instrument or a particular measuring environment is needed.

By employing VIs, many advantages have been achieved. Firstly, the cost of the system, using various subunits, is very low. Secondly, development time is short, enabling immediate use of the system. Thirdly, system maintenance expenses are lower, involving the PPT and DAQ card only. A very simple hardware configuration also elevates the quality of the research. We have achieved excellent accuracy, reliability and stability in a virtual measuring system.

6 References

/1/ The Pressure, Strain and Force Handbook, Vol. 28, *Omega Engineering, Inc.*, Stamford, 1992.

/2/ Richiardone, R. and Manfrin, M. (2002): A Rain Episode Related to a Mesoscale Gravity Wave during MAP. *Preprints volume of the 10th Conference on Mountain Meteorology American Met. Soc.*, Boston, 213-214.

/3/ Digiquartz Pressure Instrumentation: Features, Quality and Standards, *Paroscientific Inc.*, 4500 148th Av. N.E. Redmond, WA 98052 USA, www.paroscientific.com, Date: Aug. 11th, 2006

/4/ Danijel Belušić, Miroslava Pasarić, Zoran Pasarić and Mirko Orlić, The Dynamics of Pulsations in the Bora Flow, *University of Zagreb*, Zagreb, Croatia

/5/ M. Ceschia, G. Iob, Spectral estimates of the surface pressure fluctuations associated with the "bora" wind recorded in Udine and Trieste, *Proc of the 6th International Meeting on Statistical Climatology*, 19-23 June, 1995, University College, Galway, Ireland, pp. 621-624

/6/ SETRA Model 270 Specifications, SSP270 Rev.C 04/19/01, 159 Swanson Road, Boxborough, MA, USA, www.setra.com, Date: August 23rd, 2006

/7/ Operating Instructions Model 270 Pressure Transducer, Rev. D 6/98, SETRA, 159 Swanson Road, Boxborough, MA, USA, www.setra.com, Date: August 23rd, 2006

/8/ Henry W. Ott, Noise Reduction Techniques In Electronic Systems, 2nd Ed., *John Wiley & Sons*, New York, 1988

/9/ Using LabVIEW with Honeywell's Precision Pressure Transducer (PPT), AN-105 (Preliminary), *Honeywell Inc.*, www.ssec.honeywell.com, Date: July 20th, 2006

/10/ PCI-6013E/6014E User Manual, *National Instruments Corporation*, Part Number 322072C-01, December 2000

/11/ Robert H. Bishop, LabVIEW Student Edition 7i, *Prentice Hall*, Upper Saddle River, NJ 07458, 2001

/12/ Bonnie C. Baker, Anti-Aliasing, Analog Filters for Data Acquisition Systems, AN-699, *Microchip Technology Inc.*, 1999

Doc. dr. Žarko Gorup
Faculty of Electrical Engineering, University of Ljubljana
Tržaška 25, SI-1000 Ljubljana, Slovenia
E-mail: zarko.gorup@fe.uni-lj.si
tel +386 1 4768 325; fax +386 1 4768 725

Prispelo (Arrived): 29.09.2006 *Sprejeto (Accepted):* 30.03.2007

UDK621.3:(53+54+621+66), ISSN0352-9045 Informacije MIDEM 37(2007)1, Ljubljana

HIBRIDNI OSCILATOR ZA ANALOGNE IN DIGITALNE PRIZEMELJSKE TELEVIZIJSKE ODDAJNIKE

Andrej Kosi, Mitja Solar

Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko, Maribor, Slovenija

Kjučne besede: analogna prizemeljska televizija, digitalna prizemeljska televizija, sintezator z neposredno sintezo signala, fazno sklenjena zanka.

Izvleček: Trenutno poteka ali se pripravlja prehod oddajanja analognega televizijskega signala v oddajanje digitalnega televizijskega signala. Časovno je ta prehod v evropski uniji okvirno določen, od posameznih držav članic pa je odvisno, kdaj ga bodo izvedle. Prehod je povezan z velikimi finančnimi investicijami, zato so zanimivi oddajniki ali sklopi oddajnika, ki jih lahko uporabimo tako v analognem kot digitalnem načinu oddajanja. Eden od sklopov oddajnika, ki ga je treba v tem primeru prilagoditi ostrejšim zahtevam, je oscilator. V prispevku je predstavljen hibridni oscilator, primeren za analogne in digitalne prizemeljske televizijske oddajnike in kanalne pretvornike. Analogna omrežja so v osnovi večfrekvenčna, kar pomeni, da so sosednje oddajne celice na različnih frekvenčnih kanalih. Digitalna omrežja pa so lahko večfrekvenčna, enofrekvenčna ali kombinacija obeh. To ima za posledico, da so zahteve za oscilator v digitalnem oddajanju zahtevnejše. Zadani cilj je bil razviti oscilator, ki bo izpolnjeval zahteve na tržišču glede lastnosti takšnega oscilatorja.

Predlagano rešitev smo poimenovali hibridni oscilator. Hibridni oscilator združuje dva koncepta, ki se uporabljata za generiranje frekvence. En koncept je fazno sklenjena zanka, drugi pa sintezator z neposredno sintezo signala (v nadaljevanju DDS-oscilator). Z nobeno od obeh metod ni možno enostavno izpolniti vseh zahtev za oscilator. Bistveni zahtevi sta predvsem korak nastavljanja izhodne frekvence in fazni šum oscilatorja. Z združitvijo dobrih lastnosti obeh metod je mogoče narediti oscilator z želenimi lastnostmi. DDS-oscilator je uporabljen kot vir signala z referenčno frekvenco za fazno sklenjeno zanko. Z majhnim spreminjanjem frekvence DDS-oscilatorja dosežemo večjo frekvenčno ločljivost fazno sklenjene zanke, s katero pa lahko pokrijemo frekvenčna področja, ki jih ni mogoče neposredno pokriti z DDS-oscilatorjem.

Hybrid Oscillator for Analog and Digital Terrestrial Television Transmitters

Key words: analog terrestrial television, digital terrestrial television, DDS (Direct Digital Synthesizer), PLL (Phase Lock Loop).

Abstract: Currently, the transition or preparation for the changeover from analog television to digital television is in progress. Consequences of this transition are modifications that will have to occur in the distribution, transmission and reception of television signals. This article will mainly focus on television transmitters, or more precisely to their oscillator. The demands for oscillators used in digital television are different from those used in analog television and for this reason, simple modification of the latter is not possible. The major differences are in the smallest frequency step and phase noise demand. The smallest frequency step for digital television is 1 Hz, which is far less than the few kHz step required for analog television. The phase noise requirement is specified in the standard for DVB-T (Digital Video Broadcasting - Terrestrial) and is also higher than that for analog television.

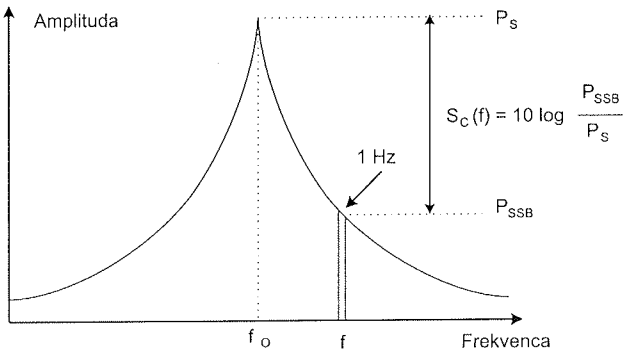
This article will present the hybrid oscillator, which is suitable for use in both analog and digital transmitters. The term hybrid oscillator refers to an oscillator that combines two or more different technologies together. In this case, the hybrid oscillator combines direct digital synthesizer (DDS) and phase locked loop (PLL). The advantages and disadvantages of DDS and PLL are shortly presented. DDS output is limited with frequency range, frequency spurs and harmonics. On the other hand, DDS has a fast settling time and a very fine frequency step (μHz). PLL can be used for higher frequencies with a wider frequency range, however it is hard to achieve a small frequency step, low phase noise and fast settling time at the same time. This is because the improvement of one parameter degrades the others. For our hybrid oscillator, DDS is used as a source for the reference frequency of the PLL. In so doing, fine frequency step is achieved and the wanted phase noise is preserved. For best results it is very important to use high quality electronic parts that contribute little to phase noise degradation.

On the basis of conclusions, we have made a prototype of this hybrid oscillator for UHF (Ultra High Frequency) frequency band. In addition, a number of different measurements of the hybrid oscillator were taken to confirm theoretical calculations and expectations. The final measurement results have shown that the hybrid oscillator is suitable for use in both analog and digital television transmitter.

1 Uvod

Prehod prenosa televizijskih vsebin iz analognega v digitalni način je po svetu v polnem teku. Najprej se je ta prehod začel na satelitskem oddajanju. Sčasoma so temu trendu sledili še kabelski operaterji in prizemeljsko oddajanje. Vrstni red je posledica zahtevnosti in stroškov uvedbe oddajanja v digitalnem načinu. Prehod prizemeljskega oddajanja v digitalni način spremljajo tudi številne spremembe v oddajniški infrastrukturi. Tako nekateri oddajniški sklopi nimajo več ustreznih parametrov in jih je potrebno zamenjati z drugimi. Eden od takih sklopov je tudi oscilator za televizijski oddajnik. Oscilator, ki ustreza zahtevam za analogno oddajanje v večini primerov ni primeren za digitalno oddajanje. Osnovna razlika je v ločljivosti nastavljanja izhodne frekvence in zahtevanem nivoju faznega šuma za oscilator. Ločljivost oscilatorja podaja najmanjšo razliko med dvema frekvencami, ki jo je moč nastaviti. Fazni šum pa

predstavlja razliko nivoja amplitude pri nekem danem odmiku od nosilne frekvence v primerjavi z nivojem amplitude nosilne frekvence. Običajno se podaja v enoti dBc/Hz za dani odmik od nosilne frekvence /1/. Za lažjo predstavo o izrazu fazni šum si fazni šum lahko predstavljamo kot razmerja energije v 1 Hz pasovnem področju za dani odmik od nosilne frekvence proti energiji nosilne frekvence. Primer določitve faznega šuma je prikazan na sliki 1.1. Določanje faznega šuma se v praksi izvaja s posebnimi instrumenti, ki omogočajo takšne meritve.



Slika 1.1: Definicija faznega šuma.

Oscilatorja za analogno oddajanje ni moč na enostaven način prilagoditi novim zahtevam. Večina oscilatorjev za analogno oddajanje temelji na principu fazno sklenjene zanke (Phase Lock Loop - PLL) /2/, /3/. Fazni šum za analogno televizijo ni tako zelo pomemben in zato obstoječi oscilatorji odstopajo od novih zahtev. Ustrezna frekvenčna ločljivost za analogno televiziji je v rangu nekaj kHz. Frekvenčna ločljivost za digitalno televizijo pa znaša 1 Hz. Zahteve za oscilator za digitalno oddajanje se torej v teh dveh pogledih bistveno razlikujejo od zahtev za analogno oddajanje.

Ena od alternativ fazno sklenjene zanke je sintezator z neposredno sintezo signala (ang. DDS - Direct Digital Synthesizer) /4/, /5/. Prednost DDS-oscilatorja pred fazno sklenjeno zanko je bistveno hitrejše nastavljanje izhodne frekvence ter zelo velika točnost nastavljene frekvence. Ima pa DDS-oscilator nekaj pomankljivosti kot so harmonske frekvence, nezaželene frekvence (ang. spurious frequencies) in omejeno frekvenčno področje uporabe. Združitev dobrih lastnosti obeh metod nam omogoči izdelavo hibridnega oscilatorja. V našem primeru smo uporabili DDS-oscilator kot vir signala z referenčno frekvenco za fazno sklenjeno zanko. Z majhnim spreminjanjem frekvence DDS-oscilatorja dosežemo večjo frekvenčno ločljivost fazno sklenjene zanke, s katero pa lahko pokrijemo frekvenčna področja, ki jih ni mogoče neposredno pokriti z DDS-oscilatorjem.

Za RF (ang. Radio Frequency) projekte so zelo pomembne izkušnje, ki smo si jih pridobivali na prejšnjih projektih /6/. Elektronske komponente smo izbirali na osnovi spoznanj in meritev predhodnih izdelkov. Izbira najbolj ustreznih kvalitetnih komponent, ki so trenutno na voljo na tržišču, je predpogoj za doseganje dobrih rezultatov.

2 Zahtevane karakteristike in koncept hibridnega oscilatorja

Oscilator mora tako za analogno kot za digitalno televizijo izpolnjevati določene zahteve. Pri predstavitvi smo se osredotočili na dve bistveni zahtevi. To sta kot že omenjeno fazni šum in frekvenčna ločljivost. Kot že omenjeno je zahteva za frekvenčno ločljivost 1 Hz. Zahteva izhaja iz dejstva, da lahko imamo v digitalnem načinu oddajanja sosednje oddajnike na istem frekvenčnem kanalu (SFN - Single Frequency Network). Oddajniki morajo oddajati na točno isti frekvenci ob istem času isto vsebino, zato morajo biti vsi oddajniki v primeru SFN omrežja sinhronizirani. Običajno je sinhronizacija izvedena s pomočjo globalnega pozicionirnega sistema (GPS - Global Positioning system).

Drugi kriterij nivo faznega šuma je podan v tabeli 2.1. Ta kriterij ni natančno določen, zato veljajo za fazni šum priporočila. Vrednosti, ki so se uveljavile kot zahtevane v veliki večini primerov so nastale na osnovi praktičnih preizkusov in raznih projektnih skupinah. Eden izmed takih projektov je bil poznan pod oznako AC106 VALIDATE (Verification And Launch of Integrated Digital Advanced Television in Europe). Osnova za priporočila so tudi standardi, ki so nastali pod okriljem evropskega inštituta za telekomunikacije (ETSI - European telecommunications standards institute) /7/, /8/, /9/.

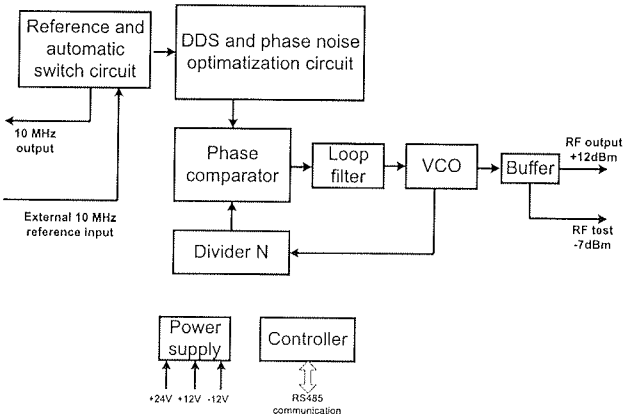
Fazni šum oscilatorja mora tako biti enak ali boljši vrednostim podanim v tabeli 2.1, da zadosti priporočilom za digitalno televizijo. V posameznih državah se zahteve v posameznih točkah tudi razlikuje od priporočil, zato smo si zadali nalogo zadovoljiti tudi te zahteve in smo svoje ciljne vrednosti še dodatno zaostrili.

Tabela 2.1: Nivo faznega šuma za digitalno televizijo.

Odmik od nosilca	Zahtevene vrednosti	Ciljne vrednosti
10 Hz	<-55 dBc/Hz	<-55 dBc/Hz
100 Hz	<-85 dBc/Hz	<-90 dBc/Hz
3 kHz	<-85 dBc/Hz	<-90 dBc/Hz
10 kHz	<-95 dBc/Hz	<-95 dBc/Hz
100 kHz	<-112 dBc/Hz	<-115 dBc/Hz
1 MHz	<-130 dBc/Hz	<-130 dBc/Hz

Poenostavljeno blokovno shemo hibridnega oscilatorja prikazuje slika 2.1. Za vsak oscilator je zelo pomemben izvor referenčnega signala, ki vpliva na več parametrov celotnega oscilatorja. Običajno se uporabljajo OCXO (Oven Controlled Crystal Oscillator) oscilatorji s posebej izbranimi parametri. Zaradi sinhronizacije oddajnikov v SFN omrežjih, s pomočjo GPS signalov, je izbrana vhodna frekvenca za digitalne oddajnike 10 MHz. Ker je frekvenco reference določena, se zelo zmanjša maneverski prostor za načrtovanje oscilatorja. Pri izbiri primernega oscilatorja moramo veliko pozornosti posvetiti poteku faznega šuma, saj ta direktno določa maksimalni možen končni fazni šum hibridnega oscilatorja. Pomemben parameter je tudi stabilnost referenčne frekvence. Če referenčna frekvenca ni stabilna se to neposredno pozna na odstopanju izhodne

frekvence. Primerna temperaturna stabilnost referenčne frekvence za digitalno televizijo je 10^{-8} ali manj. Primerna kratkoročna stabilnost na sekundo je 10^{-11} ali manj. OCXO oscilator predstavlja cenovno gledano večji del materialnih stroškov celotnega izdelka, zato je zasnovan tako, da je možna uporaba OCXO oscilatorjev drugih proizvajalcev in drugih fizičnih dimenzij. Tako je mogoče uporabiti trenutno cenovno bolj ugoden OCXO v aplikacijah, ki ne zahtevajo tako velike stabilnosti frekvence.

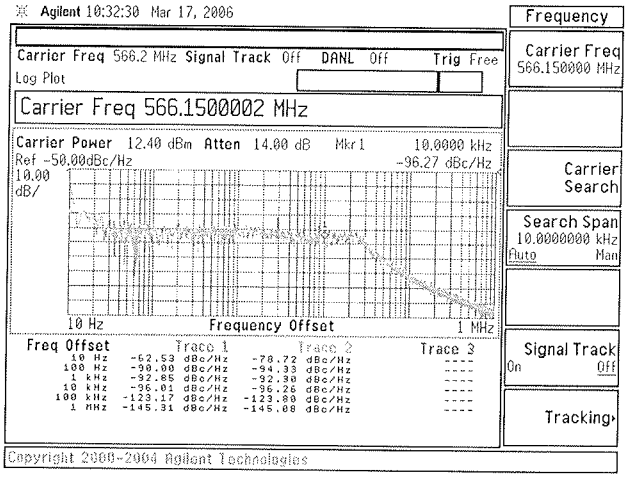


Slika 2.1: Blokovna shema hibridnega oscilatorja.

Referenčni signal na vhodu v DDS-oscilator je signal iz OCXO oscilatorja. Pri izbiri DDS-oscilatorja moramo biti pozorni predvsem nato, da ima le-ta ustrezno karakteristiko poteka faznega šuma. Ta je podana kot mejna vrednost, pri kateri je možno doseči določene vrednosti drugih izbranih parametrov. Podatek o faznem šumu najdemo v tabelah in grafih specifikacij DDS-oscilatorja. Z upoštevanjem, da se bo fazni šum v fazno sklenjeni zanki poslabšal za faktor 20*log(Fout/Fin) lahko izločimo neprimerne DDS-oscilatorje za takšno aplikacijo. Pomembna pri izbiri DDS-oscilatorja je tudi velikost faznega registra, ker določa največjo ločljivost DDS-oscilatorja in izhodne frekvence. Odvisno od izbire drugih parametrov je dovolj 32-bitni oziroma 48-bitni register. Optimizacijsko vezje služi za optimalen potek faznega šuma in zmanjševanje vpliva nezaželenih frekvenc. Naslednji sklop na shemi, ki sledi optimizacijskemu vezju je v bistvu klasična fazno sklenjena zanka. Izbira komponent zanke je podrejena zagotavljanju ustreznega faznega šuma. Napetostno krmiljeni oscilator (VCO - Voltage controlled oscillator) ima neposreden vpliv na potek faznega šuma izven pasovne širine zančnega filtra (ang. loop filter). Zato je izbira določena s želenimi vrednostmi faznega šuma podanimi v tabeli 2.1. Pomembni podatki za napetostno krmiljeni oscilator so vrednosti za odmik od nosilca 10 kHz in več. Fazni primerjalnik (ang. phase comparator) in delilnik z N (ang. Divider N) sta del integriranega vezja. Pomembni parametri integriranega vezja so: - fazni šum, - vhodno frekvenčno področje in - področje izbire parametra N. Zančni filter določa vpliv poteka DDS-oscilatorja kot reference v fazno sklenjeno zanko in vpliv poteka faznega šuma napetostno krmiljenega oscilatorja. Z ustrezno izbiro pasovne širine zančnega filtra dobimo optimalen potek glede na oba posamezna poteka. Pri izbiri elementov in načrtovanju fazno sklenjene zanke si lahko pomagamo s različnimi programi in literaturo /10/, /11/.

3 Meritve in rezultati

Na podlagi rezultatov analiz in simulacij smo izdelali prototip hibridnega oscilatorja. Na prototipu smo naredili več meritev in nekaj najpomembnejših rezultatov je prikazanih v nadaljevanju. Na sliki 3.1 je meritev faznega šuma za frekvenco 566.150000 MHz opravljena z merilnikom faznega šuma Agilent E4443A. Vidimo, da je meritev faznega šuma znotraj zastavljenih ciljev. Odstopanje frekvence je v primeru zunanje sinhronizacije z GPS sistemom enako 0,2 Hz (<1 Hz).



Slika 3.1: Primer meritve faznega šuma.

Rezultati drugih meritev in tehnični podatki so prikazani v tabeli 3.1.

Tabela 3.1: Tehnične lastnosti oscilatorja.

Podatek	Vrednost
Frekvenčno področje delovanja (UHF)	510 MHz – 900 MHz
Frekvenčna ločljivost	<1 Hz
Absolutno odstopanje frekvence brez GPS signala	<150 Hz
Absolutno odstopanje frekvence z GPS signalom	<1 Hz
Izhodna impedanca	50 Ω
Fazni šum	10 Hz: <-55 dBc 100 Hz: <-90 dBc 1 kHz: <-90 dBc 10 kHz: <-95 dBc 100 kHz: <-115 dBc 1 MHz: <-130 dBc
Izhodni nivo signala	+12 dBm ± 2 dB
Nivo signala na testnem izhodu	-7 dBm ± 2 dB
Nivo signala v primeru neujetja zanka	< -10 dBm
Nivo nezaželenih signalov: - harmonske frekvence - neharmosnke frekvence	 < -20 dBc < -65 dBc
Zunanja referenca	10 MHz
Vhodni nivo zunanje reference	-10 dBm to +13 dBm /50 Ω
Interna referenca	10 MHz
Komunikacija	RS485 ASCII protokol
Fizične dimenzije	60 x 30 x 205 mm
Temperaturno območje delovanja	0°C to 50°C

4 Zaključek

V članku smo prikazali koncept in merilne rezultate hibridnega oscilatorja. Če povzamemo merilne rezultate lahko zaključimo, da je izdelan hibridni oscilator ustrezen za uporabo v analognih in digitalnih prizemeljskih oddajniških sistemih. Predstavljen hibridni oscilator je možno uporabiti tudi za druge aplikacije, kjer je potrebna velika frekvenčna ločljivost, stabilnost in ustrezen potek faznega šuma. Izdelan prototip sicer pokriva UHF prizemeljske frekvence, vendar je možno enak koncept uporabiti tudi za druga frekvenčna območja. Treba je le ustrezno izbrati posamezne komponente, ki so primerne za določeno frekvenčno območje. Druga frekvenčna področja lahko pokrijemo tudi tako, da obstoječi hibridni oscilator uporabimo skupaj z delilniki oziroma množilniki frekvence.

5 Literatura

/1/ Article Phase Noise and Jitter-A primer for Digital Designers by Neil Roberts, High-Performance Analog Designer, Published in EE Design on July 14, 2003, http://timing.zarlink.com/assets/Phase_Noise_and_Jitter_Article.pdf.

/2/ Phase-Locked Loops for High-Frequency Receivers and Transmitters-Part 1 by Mark Curtin and Paul O'Brien Analog Dialogue 33-3 (© 1999 Analog Devices), http://www.analog.com/library/analogDialogue/archives/33-03/phase/phase_locked.pdf.

/3/ Phase-Locked Loops for High-Frequency Receivers and Transmitters-Part 2 by Mark Curtin and Paul O'Brien Analog Dialogue 33-5 (© 1999 Analog Devices), http://www.analog.com/library/analogDialogue/archives/33-05/phase_locked/PLLs_2.pdf.

/4/ Analog Devices, A Technical Tuturial on Digital Signal Synthesis, 1999, www.ieee.li/pdf/essay_dds.pdf.

/5/ Jeffrey H. Reed, Brian D. Woerner, Direct Digital Synthesis A Tutorial, The 9th International Symposium on Personal, Indoor, and Mobile Radio Communications, September 1999, www.mprg.org/Tech_xfer/ppt/Jeff-PIMRC-Part1.pdf.

/6/ Andrej Kosi, Hibridni oscilator za digitalno zemeljsko televizijo, magistrska naloga, april 2005, Fakulteta za elektrotehniko, računalništvo in informatiko v Mariboru.

/7/ European Broadcasting Union, ETSI EN 300 744 v1.5.1, Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television, November 2004.

/8/ European Broadcasting Union, ETSI TR 101 190 v1.2.1, Digital Video Broadcasting (DVB); Implementation guidelines for DVB terrestrial services; Transmission aspects, November 2004.

/9/ European Broadcasting Union, ETSI ETR 290, Digital Video Broadcasting (DVB); Measurement guidelines for DVB systems, May 1997.

/10/ Analog Devices, ADIsimPLL Software, Version 2.7, http://www.analog.com/pll.

/11/ Dean Banerjee, PLL Performace, Simulation, and Design, Second Edition, 2001, www.national.com/appinfo/wireless/files/DeansBook_4_01.pdf.

mag. Andrej Kosi univ. dipl. inž. je zaposlen v podjetju ELTI d.o.o Gornja Radgona,

docent dr. Mitja Solar je predavatelj na Fakulteti za elektrotehniko, računalništvo in informatiko v Mariboru.

Fakulteta za elektrotehniko, računalništvo in informatiko v Mariboru
Smetanova 17
2000 Maribor

Prispelo (Arrived): 21.09.2006 *Sprejeto (Accepted):* 30.03.2007

ESTIMATION OF LINEAR ELECTRICAL MOTOR PERFORMANCE USING FINITE ELEMENT METHOD

Miroslav Bugeza[1], Rastko Fišer[2]

[1]ELEK Svetovanje d.o.o. Ljubljana, Slovenia
[2]University of Ljubljana, Faculty of Electrical Engineering, Slovenia

Key words: linear induction motor, linear synchronous motor, permanent magnet, magnetic field, finite element method, mathematical modeling, propulsion force, attraction force, high-dynamic drive.

Abstract: The paper presents the approach to estimation of operational characteristics of linear induction motor and permanent magnet linear synchronous motor using finite element method. A detailed insight into magnetic field distribution and produced output force is a basic step for motor performance evaluation. A 2D non-linear magnetostatic and eddy-current solver were used for magnetic field and force calculations. The obtained force vectors were decoupled into the propulsion and the attraction force components. The former is vital for determination of motor steady-state and transient characteristics, while the latter is important in mechanical designing of the drive base.

For both types of linear motor calculations were carried out at several excitation currents and different air-gap lengths to optimize the drive performance. The results obtained from the magnetic field calculations could also be applied for determination of motor parameters (e.g. inductances), which are necessary in drive control system simulations. Several laboratory tests were performed to verify the results of the presented method.

Vrednotenje obratovalnih lastnosti linearnih električnih motorjev s pomočjo metode končnih elementov

Ključne besede: linearni asinhronski motor, linearni sinhronski motor, trajni magneti, magnetno polje, metoda končnih elementov, matematično modeliranje, potisna sila, privlačna sila, visoko dinamični pogon.

Izvleček: Članek obravnava postopek določitve obratovalnih karakteristik dveh posebnih tipov električnih motorjev - linearnega asinhronskega motorja in linearnega sinhronskega motorja s trajnimi magneti, ki temelji na uporabi metode končnih elementov. Le-ta omogoča natančen izračun porazdelitve magnetnega polja in vrednosti elektromagnetnih sil, kar je ključnega pomena za zanesljivo oceno pogonskih lastnosti. Glede na različen fizikalni princip delovanja obeh obravnavanih motorjev smo modelirali spremenljivo magnetno polje oziroma uporabili nelinearni magnetostatični izračun. Vektor elektromagnetne sile vsebuje potisno komponento, ki določa zmogljivost pogona ter pritezno komponento, ki jo moramo upoštevati pri dimenzioniranju mehanske konstrukcije pogona.

Za obe vrsti linearnih motorjev so prikazani rezultati izračunov v obliki odvisnosti elektromagnetnih sil od obratovalnih pogojev pri različnih vzbujalnih tokovih in za različne širine zračne reže med gibljivim primarjem in nepremičnim sekundarjem. Iz znanega magnetnega polja je možno določiti še vrsto dodatnih parametrov linearnih motorjev (npr. induktivnosti), ki so pomembni pri modeliranju, simulacijah ali optimiranju nadzorno krmilnega sistema pogona. Realizirani laboratorijski testni pogon dolžine 4 m omogoča meritve električnih in mehanskih parametrov obeh vrst motorjev. Pri doseženi hitrost 2.5 m/s in pospešku 12 m/s^2 je ponovljiva natančnost pozicioniranja 10 μm.

1. Introduction

A machinery construction applying linear electrical motors offers many challenges for engineers and brings up a new philosophy of machine design. In the case of rotational electrical motors it is often required to transform rotation into linear movement, which is unnecessary when using linear motors. The main advantage of linear over the usual rotational motors is outstanding accuracy at high dynamics due to the absence of intermediate gears, screws and crank shafts for linear moving loads. Linear drives are recently used in machine and textile tools, linear tables, saws, shears, separators, transportation systems and many others /1/. As an example, a manufacturer of CNC universal lathes introduced the highly dynamic linear drive (velocity of 100 m/min and acceleration values up to 15 m/s^2), which reduces idle times to a minimum and increases productivity up to 30% by utilization of the latest linear drive technology /2/. At the same time the quality of products increased considerably due to maximum positioning accuracy and precision, which is generally in the contradiction to top-rate dynamics. In addition the contactfree power transfer is almost wear-free, which has a positive effect on the reliability of the machine and reduces maintenance costs.

2. Test drive with linear electrical motors

In *Laboratory of electrical drives* the linear drive was realized as a model of a flying shear, which is a common industrial application for cutting continuous product to a set length. The cutting tools moving in the perpendicular direction are typically mounted on a carriage that moves parallel to the product flow. As it presents linear motion,

usage of the linear motor is preferable /3/. The high cutting speed and the accuracy are major demands for industrial flying shears. The speed depends mainly on the drive dynamic capability and the corresponding construction. Applying the linear motor instead of two rotational ones reduces moving masses thus dynamic performance of the device is higher at the same or even better accuracy of positioning.

Fig. 1 shows the test linear drive, which has the overall length of 4 m. Essential built-in components like linear guideways, bearings, and position sensor guarantee the moving trolley velocity of 2.5 m/s at the acceleration up to 12 m/s^2. The accuracy of the repeated positioning is limited to 10 mm. For achieving even better dynamic and accuracy performance capabilities more sophisticated and expensive components should be selected.

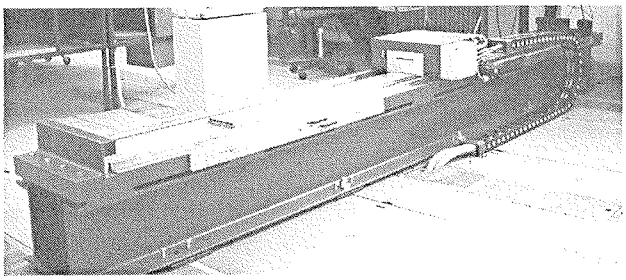


Fig. 1. Test drive with linear electrical motor

The base of linear drive (Fig. 2) was designed for optional application of two linear motor types, the linear induction motor (LIM) and permanent magnet linear synchronous motor (PMLSM), respectively. Basic parameters for both types of linear motors are stated in Table 1.

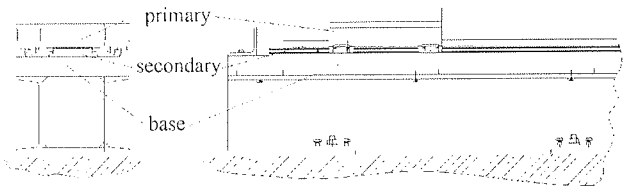


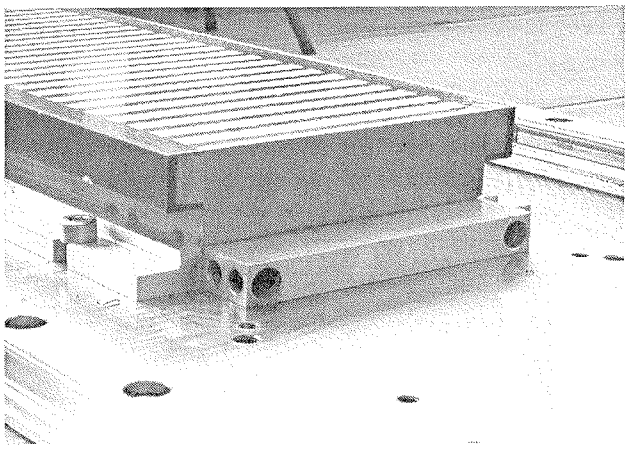
Fig. 2. Main components of the linear drive: base with main plate, primary and secondary part of linear electrical motor

Table 1. Basic constructional and nominal parameters of the linear motors

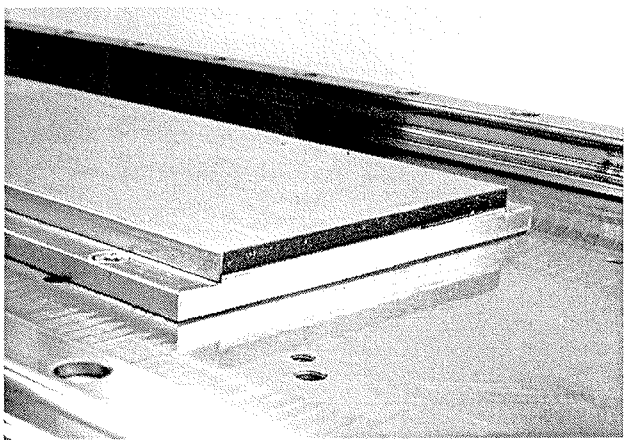
	Number of pole-pairs	Length of primary (mm)	Length of secondary (mm)	Rated force (N)	Rated current (A)	Rated speed (m/s)
LIM	2	390	3000	1000	19	1
PMLSM	4	390	3000	1500	17	2.5

Both primary parts are relatively similar with slightly distinction in laminations and winding distribution. The main difference is in secondary parts of linear motors. The LIM has the squirrel cage secondary with copper as a conductive material (Fig. 3.a). For induction motors of this size the aluminium squirrel cage would be more common, but since linear motors are not yet manufactured in mass-production such technology is more convenient. Copper bars have also better conductivity, which enables the higher secondary induced currents and improved performance characteristics of the LIM /4/.

The PMLSM has the secondary part (Fig. 3.b) made of rear earth permanent magnets (NdFeB) mounted on soft magnetic backplate, which represents secondary yoke. Permanent magnets are contributing to high magnetic field densities, and consequently to very high energy density of the PMLSM type of linear motor /5/.



(a)



(b)

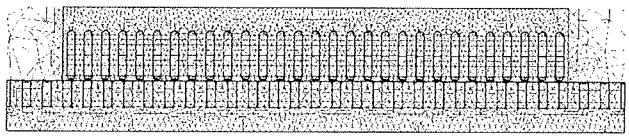
Fig. 3. Squirrel cage secondary of LIM (a), and PMLSM's secondary part with NdFeB magnets (b)

Although a 4 m drive base plate is relatively long, the final length of the linear drive disables steady-state operation of linear motors. During transients between acceleration and braking, measurements of electrical and mechanical parameters are difficult to perform. Nevertheless for optimal operation of the drive the detailed insight into linear motor performance characteristics is very important /4/, /5/, /

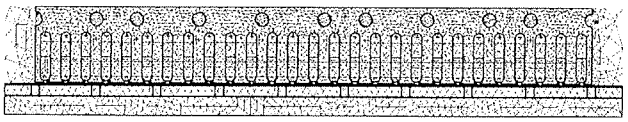
6/. To simulate the steady-state force components characteristics of the LIM and the PMLSM, computation of magnetic field distribution was performed using finite element method (FEM).

3. FEM analyses of linear motor performance

In order to perform a 2D FEM calculations of magnetic field, the motor cross-section was divided into triangular finite elements. To ensure the accurate calculation results the mesh with 16962 triangle elements was generated in the case of LIM (Fig. 4.a), and 13772 triangle elements in the case of PMLSM (Fig. 4.b), respectively. In order to reduce the complexity of the geometry and the number of nodes of finite element mesh, the calculations are usually performed over the smallest symmetrical part of the rotational motor model. This simplification cannot be applied in the case of linear motors due to the asymmetrical magnetic field distribution and longitudinal end-effects /6/. In Fig. 5 the details of finite element mesh in one magnetic pole of the LIM and PMLSM are shown.

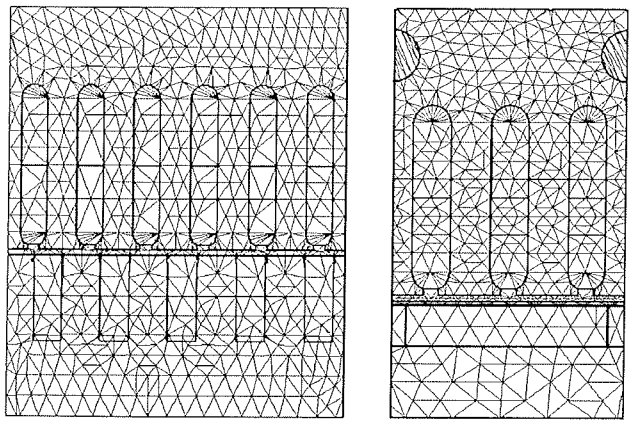


(a)



(b)

Fig. 4. Finite element mesh in the cross-section of LIM (a), and PMLSM (b)



(a) (b)

Fig. 5. Finite element mesh in detail of one magnetic pole of LIM (a), and PMLSM (b)

3.1. FEM calculation of LIM magnetic field

When the velocity of the moving primary part differs from the velocity of the longitudinal traveling magnetic field, eddy-currents of the slip frequency are induced in the squirrel cage of the LIM's secondary part. In such case the magnetic field is described with Helmholtz differential equation

$$\vec{\nabla}\times(\frac{1}{\mu}\cdot\vec{\nabla}\times\vec{A})=\sigma\cdot\vec{\nabla}\phi-\sigma\cdot\frac{\partial\vec{A}}{\partial t} \qquad (1)$$

where m is permeability, $\vec{A}$ is magnetic vector potential, s is conductivity and f is electric scalar potential. Magnetic vector potential $\vec{A}$ is calculated by the implementation of a 2D eddy-current solver.

Due to 2D field calculation, which disables the secondary loops configuration, the secondary cage was modelled as passive conductors that means they have eddy and displacement currents flowing through them, but have no component of source current. All secondary conductors (bars) are defined to be connected parallel in short circuit, which in fact represents two end-rings /7/. In this procedure variations of material properties and excitations can be considered. Except in the case of locked secondary condition, the currents and flux density in secondary and primary part are pulsating at different frequencies. Since the computation is done using phasor analysis, all calculated variables must be of the same frequency. To resolve this difficulty we can define primary excitation currents as ampere-turns with frequency corresponding to instantaneous slip value by parametric calculation for chosen steady-state points. Such approach enables simulation of running conditions at many secondary speed values from start-up to no-load condition using very simple and unexpensive FEM software.

3.2. FEM calculation of PMLSM magnetic field

In the case of PMLSM type of linear motor, the primary part velocity is synchronized with longitudinal magnetic field, and the magnetic field is described with the Poisson differential equation

$$\vec{\nabla}\times(\frac{1}{\mu}\cdot\vec{\nabla}\times\vec{A})=\vec{J} \qquad (2)$$

where $\vec{J}$ represents the current density, which consists of excitation currents in primary three-phase winding and equivalent magnetizing current of permanent magnets / 8/, /9/.

Permanent magnets used in PMLSM secondary part are NdFeB - Vacodim 633. The equivalent magnetizing current used in FEM model was determined from demagnetizing characteristics shown in Fig. 6.

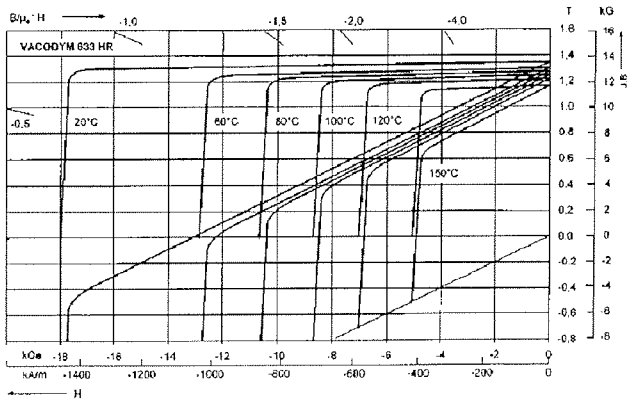


Fig. 6. Demagnetizing characteristics of Vacodym 633

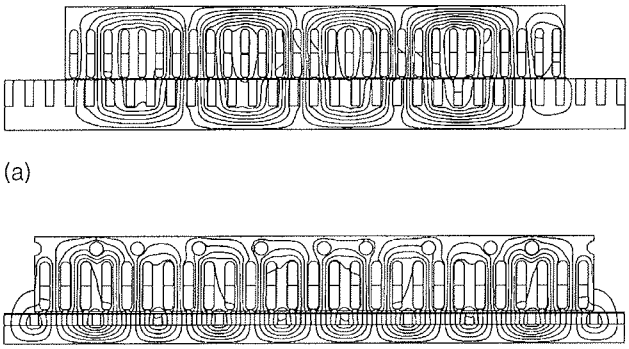
For mechanical protection of permanent magnets the protective plate made of stainless steel is applied (Fig. 3b). At steady-state the velocity of the PMLSM equals to the synchronous speed, thus the conductive protective plate has no influence on propulsion force. Contrary at transient states, induced currents in the protective plate contribute the additional component to the propulsion force thus reduces oscillations, while the velocity of the moving primary increases or decreases /10/. To estimate this phenomenon the eddy-current solver was used like in the case of LIM type of linear motor. The comparison of the calculated total propulsion force and components due to the induced currents in the protective plate shows that the influence of induced currents in the protective damper plate can be neglected, especially at higher loads /11/.

3.3. Magnetic field distribution

The primary parts of both types of linear motors have a two-layer winding with half-filled end slots, which is typical topology /1/. Because of this constructional particularity the number of poles is odd. Contrary to the magnetic field in rotational motors, which is always symmetrically distributed along the air-gap, the magnetic conditions in linear motors are regularly asymmetrical.

Fig. 7a shows the magnetic field distribution of LIM at maximal force at slip 0.12. The same case of maximal force for PMLSM at maximal load angle 90° is presented in Fig. 7b. The boreholes of the water-cooling system are placed in the yoke of the PMLSM primary lamination (Fig. 7b). They obviously have an additional effect on asymmetrical magnetic flux distribution. The LIM also has a water-cooling system, however aluminium units are mounted above the primary and under the secondary parts and have no influence on the flux lines. The magnetic field distribution in a detail of one magnetic pole of the LIM and PMLSM are presented in Fig. 8a and Fig. 8b, respectively.

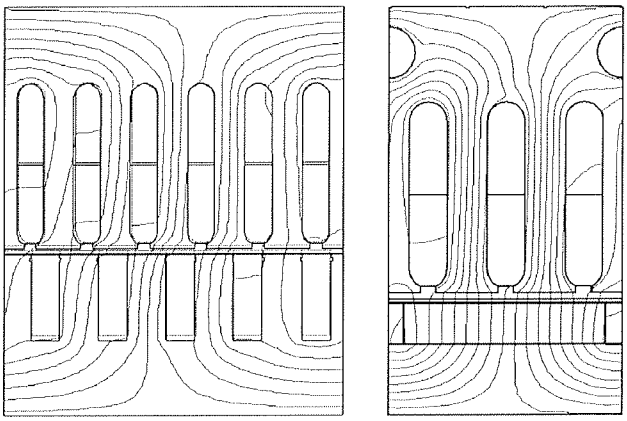
The magnetic field in the LIM is excited by currents in primary winding. Due to the half-filled end slots winding the magnetic field distribution in the air-gap is unproportional regarding magnetic pole division (Fig. 9.a). In the PMLSM the magnetic field is excited by currents in primary winding



(a)

(b)

Fig. 7. Magnetic field distribution in the cross-section of LIM (a), and PMLSM (b)

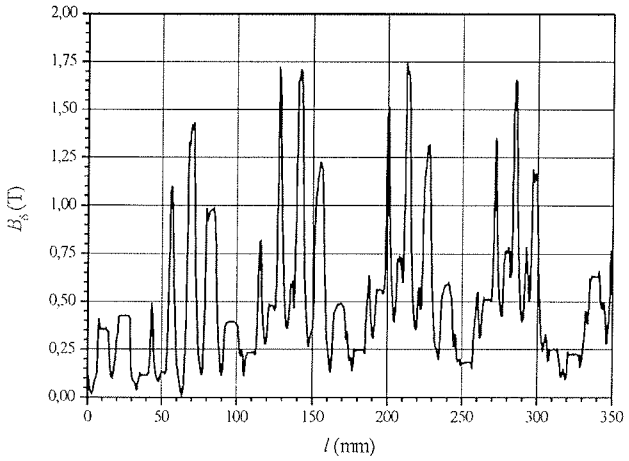


(a) (b)

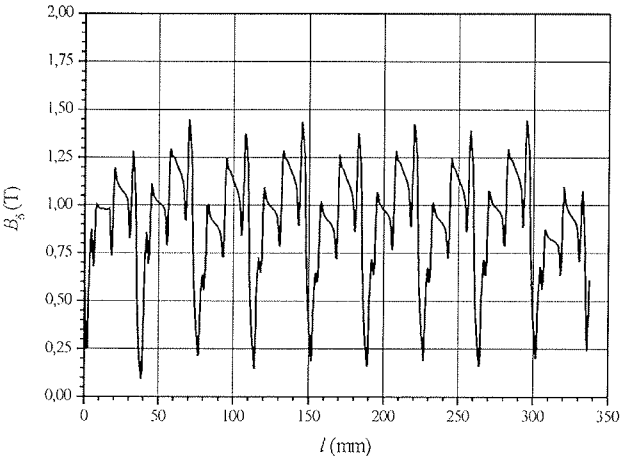
Fig. 8. Magnetic field distribution in a detail of one magnetic pole of LIM (a), and PMLSM (b)

and NdFeB permanent magnets, thus the PMLSM's air-gap field distribution is more proportional (Fig. 9b).

The NdFeB permanent magnets used in PMLSM are capable of producing strong magnetic field, therefore the average air-gap flux density is much higher (0.91 T) in comparison to the LIM's average air-gap flux density (0.49 T), which represents a 54 % difference in benefit of the PMLSM. This distinction between both described types of linear motors is obvious also from the nominal data presented in Table 1. For the same physical size and similar primary excitation currents is the nominal propulsion force of the LIM 1000 N, while for the PMLSM is denoted to 1500 N. Such higher power density of the PMLSM in comparison to the LIM is the main reason, that the PMLSM type of linear motor is much more often used in machine tool applications. Nevertheless for more robust, less dynamic and cheaper applications (e.g. longer transportation tracks) the choice of the LIM type of linear motor would be more appropriate solution.



(a)



(b)

Fig. 9. Air-gap magnetic flux density distribution of LIM (a), and PMLSM (b)

4. Simulation of steady-state force characteristics

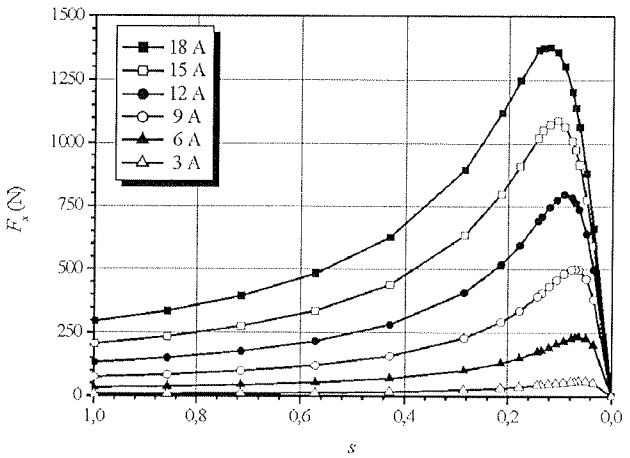
The electromagnetic force of the LIM was calculated from magnetic field results by Maxwell's stress tensor. The differential equation for the force $\vec{F}$ produced on a surface is:

$$d\vec{F} = \frac{1}{2}(\vec{H}\cdot(\vec{B}\cdot\vec{n}) + \vec{B}\cdot(\vec{H}\cdot\vec{n}) + (\vec{H}\cdot\vec{B})\cdot\vec{n}) \qquad (3)$$

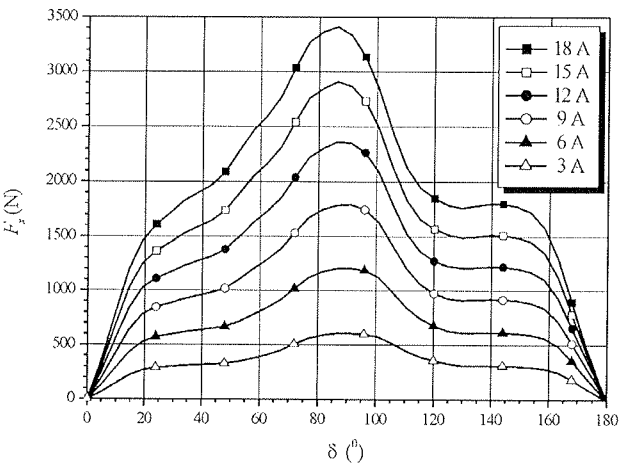
where $\vec{H}$ is magnetic field intensity, $\vec{B}$ is magnetic flux density and $\vec{n}$ is normal vector on the surface. When force components for the whole range of movement are computed, the steady-state force characteristic can be determined. The calculated force vector consists of propulsion force F_x and attraction force F_y. The propulsion force determines the performance characteristics of the linear motor, while the attraction force between primary and secondary part is in fact parasitic and thus it has to be taken into account when designing the construction of the drive set-up.

For both motors the propulsion force characteristics were calculated at different excitation currents. Fig. 10a shows steady-state force characteristics of the LIM and Fig. 10b the ones of the PMLSM. There is obvious influence of detent force in PMLSM characteristics, which could be reduced in a design phase with an appropriate choice of secondary configuration and permanent magnet length / 12/. As expected approximately two times higher excitation primary current is reqired in the LIM to achieve the same propulsion force as in the case of the PMLSM.

Characteristics of the attractive force between primary and secondary part are shown in Fig. 11. At the same excitation primary current the ratio of attractive to maximal propulsion force is about 5 in case of the LIM and 4.2 in the case of the PMLSM.

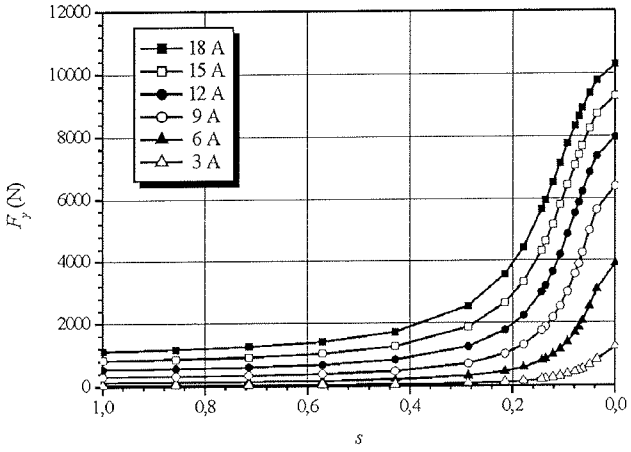


(a)

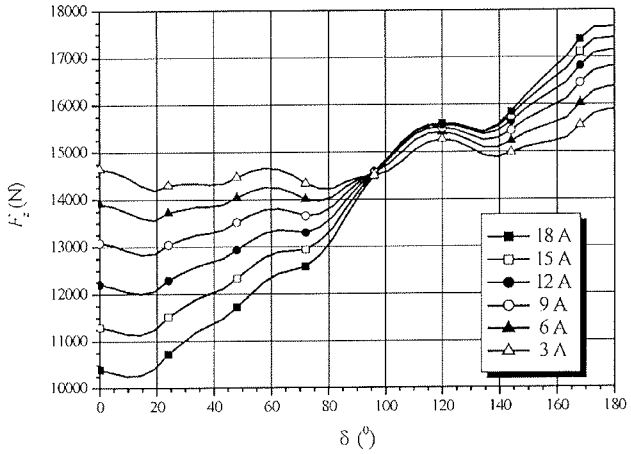


(b)

Fig. 10. Propulsion force characteristics at different excitation currents of LIM (a), and PMLSM (b)

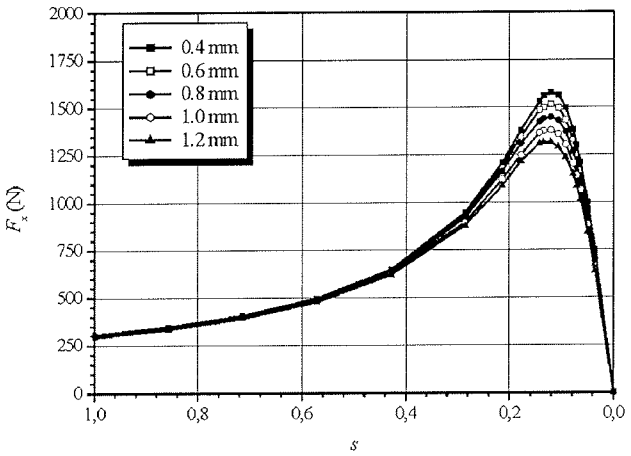


(a)

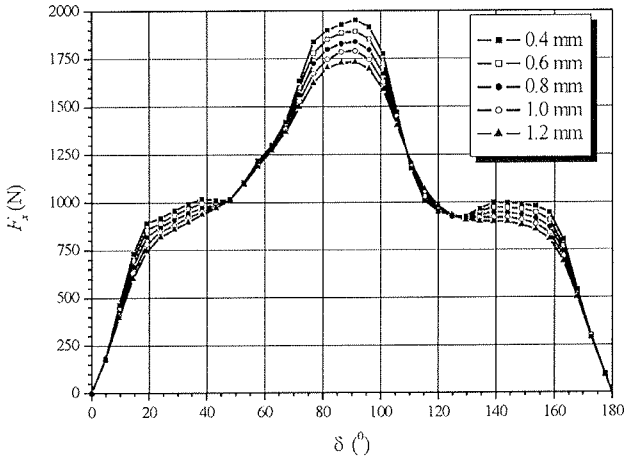


(b)

Fig. 11. Attraction force characteristics at different excitation currents of LIM (a), and PMLSM (b)



(a)



(b)

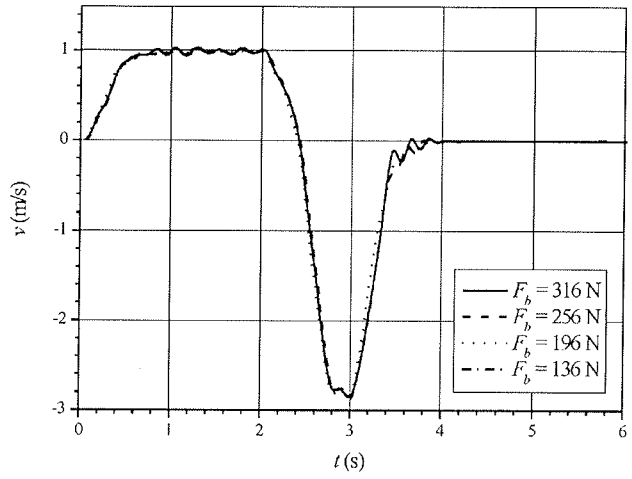
Fig. 12. Propulsion force characteristics for different air-gap lengths of LIM (a), and PMLSM (b)

The PMLSM produces the same propulsion force at just half of the current of the LIM. For adequate comparison the force characteristics for different air-gap lengths have to be calculated at different excitation currents: for example 18 A for the LIM and 9 A for the PMLSM. Comparison of such characteristics is presented in Fig. 12. While the air-gap was changing from 0.4 mm to 1.2 mm, the propulsion force declined for 18.8 % in the case of LIM and for 12.8 % in the case of PMLSM.

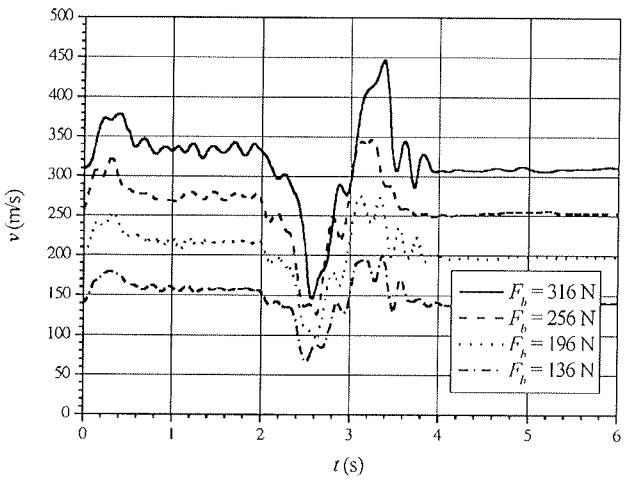
5. Experimental results

Since the geometry and material properties can be fully taken into account, the magnetic field results also enable precise determination of several others variable motor parameters (e.g. inductances), which are used in further research of LIM's and PMLSM's operational performance. For study of transient operation, which is in fact the case when dealing with limited drive lenght of linear motors, nonlinear dynamic simulation models were developed. Detailed explanations of such models are reported in /13/ for the LIM and in /14/ for the case of PMLSM.

During research several measurements were performed on laboratory models of the LIM and the PMLSM. Fig. 13a shows transient velocity response (start-up, steady-state, reversing, stand-still), while Fig.13b presents propulsion force of the LIM at several increasing load levels. The active load of the test drive is connected to the primary of the linear motor via a steel cable. Oscillations occur due to the strain of the steel cable. The same test for the PMLSM type of motor is shown in Fig. 14. Both motors are supplied and controlled by the same inverter thus demonstrate similar results although it can be noticed that PMLSM has smoother performance due to more rigid force characteristic.

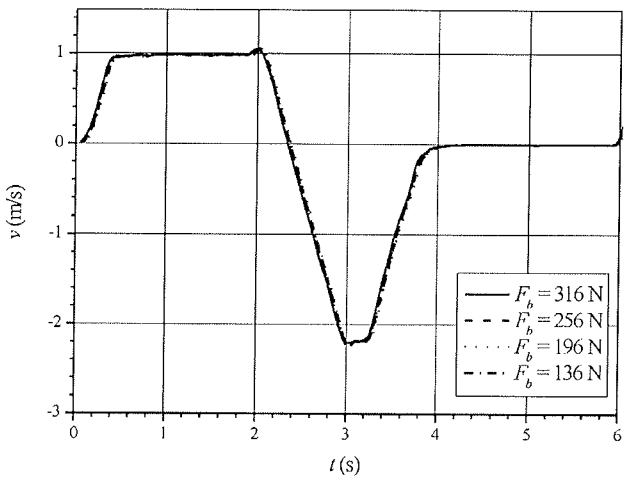


(a)

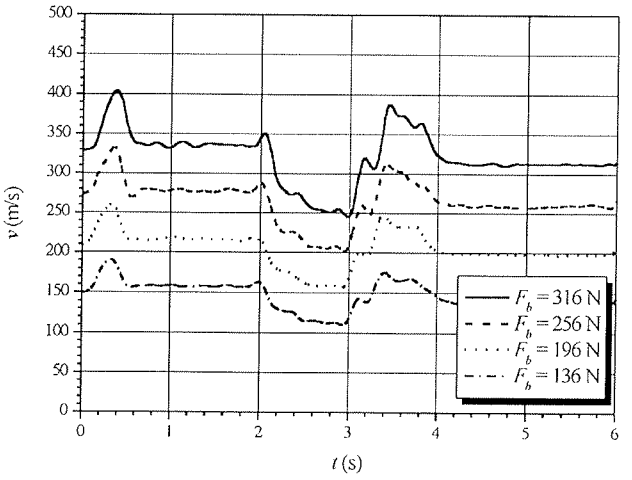


(b)

Fig. 13. Measured velocity and propulsion force transients of LIM at increasing load



(a)



(b)

Fig. 14. Measured velocity and propulsion force transients of PMLSM at increasing load

6. Conclusions

Both types of linear motors are designed to achieve the best possible performance characteristics - LIM's squirrel cage made of copper, and NdFeB permanent magnets in the PMLSM' secondary part. Intense electrical and mechanical transients cause difficulties during measurements of motor parameters, therefore FEM magnetic field computations are of great importance for detailed insight into the motors operational characteristics.

From force calculation results it is obvious, that the PMLSM has higher propulsion force at the same excitation currents, furthermore the ratio between the attraction and the propulsion force components is lower than in the case of the LIM. As expected, the overall comparison of the LIM and the PMLSM type of linear motors reveals, that the PMLSM has better performance characteristics for high-dynamic applications. Employing permanent magnets in motor design, the dimensions of the PMLSM are smaller, thus more appropriate for servo drives. Nevertheless the LIM still has the advantage, when lower price or simpler and more robust construction are required.

7. References

/1/ I. Boldea, S.A. Nasar, *Linear Motion Electromagnetic Devices*, Taylor&Francis, 2001.

/2/ www.gildemeister.com.

/3/ S. Ceferin, M. Bugeza, R. Fišer, "Linear Synchronous Drive in flying Shear Application", *Electromotion Journal*, Vol. 12,No. 2/3, pp. 161-165, 2005.

/4/ J.F. Gieras, *Linear Induction Drives*, Oxford, Claredon Press, 1994.

/5/ J.F. Gieras, *Linear Synchronous Motor: Transportation and Automation Systems*, London, CRC Press, 2000.

/6/ M. Poloujadoff, *The Teory of Linear Induction Machinery*, Oxford, Claredon Press, 1980.

/7/ R. Fišer, S. Ferkolj, "Application of a Finite Element Method to Predict Damaged Induction Motor Performance", *IEEE Transaction on Magnetics*, Vol. 37, No. 5, September 2001, pp. 3635-3639.

/8/ C. M. Ong, *Dynamic Simulation of Electrical Machinery*, Prentice Hall, New Jersey, 1998.

/9/ M. Bugeza, D. Makuc, R. Fišer, "Estimation of Nonlinear Equivalent Circuit Elements Using Finite Element Method", EPNC 2004, Poznań, Poland, June 28-30, 2004, *Conference Proceedings*, pp. 37-38.

/10/ D. Miljavec, "Finite Element Method in Evaluation of Torque Ripple in Synchronous Reluctance Motor", *Compel*, Vol. 17, Nr. 1/2/3, 1998, pp. 369-373.

/11/ M. Bugeza, R. Fišer, "Estimation o Permanent Magnet Protective Plate Effect on Dynamic Characteristics of PMLSM", *Przeglad Electrotechniczny*, Nr. 5, 2006, pp. 26-29.

/12/ S.M. Jang, S.H. Lee, I.K. Yoon, "Design Criteria for Detent Force Reduction of Permanent-Magnet Linear Synchronous Motors With Halbach Array", *IEEE Transactions on Magnetics*, Vol. 38, No. 5, September 2002.

/13/ M. Bugeza, R. Fišer, "Dynamic model of linear induction motor", ISEF 2005, Baiona, Spain, September 15-17, 2005, *Conference Proceedings*, pp. 1-5.

/14/ M. Bugeza, D. Makuc, R. Fišer, "Dynamic model of permanent magnet linear synchronous motor", ICEM 2004, Cracow, Poland, September 5-8, 2004, *Conference proceedings*, pp. 1-6.

Asst. Miroslav Bugeza, M.Sc.
ELEK svetovanje d.o.o. Ljubljana
Marice Kovačeve 7, SI - 1000 Ljubljana, Slovenia
Phone: +386 (0)1 5654 050; Fax: +386 (0)1 5654 055
E-mail: miro@elek.si

Asst. Prof. Dr. Rastko Fišer, univ. dipl. ing.
University of Ljubljana, Faculty of Electrical Engineering
Tržaška 25, SI - 1000 Ljubljana, Slovenia
Phone: +386 (0)1 4768 280; Fax: +386 (0)1 4264 630
E-mail: rastof@fe.uni-lj.si

Prispelo (Arrived): 08.11.2006 *Sprejeto (Accepted):* 30.03.2007

ESTABLISHMENT OF NMR MAGNETOMETER FREQUENCY TRACEABILITY WITH THE LONG WAVE TRANSMITTER DCF77

Samo Beguš, Dušan Fefer

Faculty of Electrical Engineering, University of Ljubljana, Slovenia

Key words:NMR magnetometer, DDS, sampling detector, DCF77 transmitter

Abstract: A simple, low power and cost-effective solution for the establishment of nuclear magnetic resonance magnetometer reference oscillator frequency traceability with the use of the long-wave DCF77 transmitter is described. The system assures a relative oscillator uncertainty of less than 10^{-7}. A low-cost direct digital synthesizer (DDS) is used to divide the reference signal and a binary counter is used to further divide the reference signal to obtain the signal for phase comparison with the DCF77 signal by means of a sampling detector. The signal from the sampling detector is amplified, digitized, filtered and analyzed by a microcontroller and a 12-bit analog-to-digital converter. The reference oscillator is not phase locked to the DCF77 signal, but only the difference of the two frequencies is measured as time of the period of the demodulated and filtered signal from the sampling detector. The frequency of the reference oscillator is calculated with the DDS frequency tuning word setting where the measured time of the period of the demodulated signal reaches the maximum value.

Vzpostavitev frekvenčne sledljivosti JMR magnetometra s pomočjo dolgovalovnega oddajnika DCF77

Kjučne besede: JMR magnetometer, DDS, vzorčevalni detektor, DCF77 oddajnik

Izvleček: V prispevku je predstavljena enostavna, energetsko varčna in cenovno ugodna rešitev za zagotavljanje sledljivosti referenčnega oscilatorja magnetometra, ki deluje na principu jedrske magnetne resonance, s pomočjo dolgovalovnega oddajnika DCF77. Sistem zagotavlja relativno negotovost oscilatorja manjšo od 10^{-7}. Signal referenčnega oscilatorja se deli s cenenim direktnim digitalnim sintetizatorjem (DDS) in binarnim števcem, dobljeni signal pa se fazno primerja z DCF77 signalom s pomočjo vzorčevalnega detektorja. Signal vzorčevalnega detektorja se ojača, digitalizira, filtrira in analizira z mikrokrmilnikom in 12-bitnim analogno-digitalnim pretvornikom. Referenčni oscilator ne deluje v fazno sklenjeni zanki z DCF77 signalom, ampak se meri razlika frekvenc obeh signalov, kot čas periode demoduliranega in filtriranega signala vzorčevalnega detektorja. Frekvenca referenčnega oscilatorja se izračuna z uporabo nastavitve DDS-a, pri kateri doseže merjeni čas periode demoduliranega signala največjo vrednost.

Introduction

The most accurate standards for magnetic flux density unit are based on magnetometers using nuclear magnetic resonance (NMR). The basic principle for measuring magnetic flux density with a NMR magnetometer is based on the conversion of magnetic flux density to frequency measurement. The magnetometer uses a reference oscillator as a reference source for frequency measurement.

When a NMR magnetometer is used as a part of the magnetic flux standard in an accredited calibration laboratory, its reference oscillator is periodically calibrated to assure the traceability of measurements (to the primary frequency standard).

The Laboratory for magnetic measurements uses an absorption type magnetometer /1/ for measurements of magnetic flux density. Its reference oscillator uncertainty should be in the order of 10^{-7} to have a minimal effect on the magnetometer measurement uncertainty.

A simple and cost-effective solution for assuring the traceability of the NMR magnetometer reference oscillator using the long-wave transmitter DCF77 is presented.

The frequency traceability of the NMR magnetometer is assured by means of the comparison of the divided NMR magnetometer reference oscillator signal with the signal from the DCF77 long-wave transmitter. There are also other ways of establishing frequency traceability of a reference oscillator such as the GPS signal /2, 3/ and short-wave transmitters /4/, but by using the long-wave transmitter the circuit implementation is relatively simple and no special receivers are required.

The DCF77 transmitter is transmitting the reference time and frequency signal at 77.5 KHz. The transmitter uses both amplitude and phase modulation of the carrier. The modulation does not influence the use of the carrier frequency for frequency comparison because the amplitude modulation is carried out by sinking the output power to 25 % of its maximum power (and not by interrupting the transmission) and the phase modulation is a pseudo-random phase modulation with a zero mean value /5/.

The DCF77 receiver

A block diagram of the DCF77 receiver is presented in Figure 1. The reference oscillator is a temperature compensated crystal oscillator (TCXO) operating at 40 MHz. Its frequency could be tuned electronically.

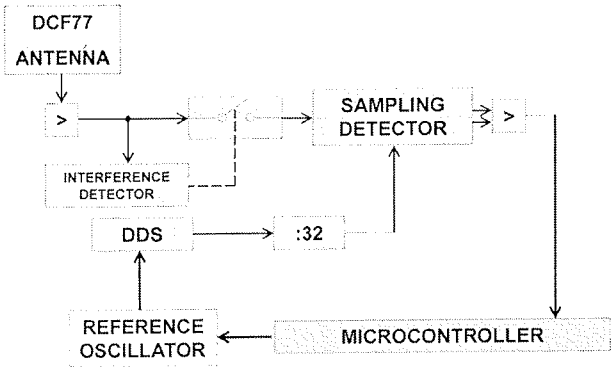


Fig. 1: Block diagram of the DCF77 receiver.

The reception of the DCF77 signal is ensured with a ferrite bar antenna, a three stage amplifier and a band-pass filter. The low-cost DDS used has a 10-bit DAC and a 28-bit frequency tuning word (FTW) to generate the reference signal for the demodulator based on the sampling detector /6/.

A DDS operating with a 40 MHz reference oscillator and with either a 28-bit, 32-bit or 48-bit FTW has a relative frequency resolution of $1.9 \cdot 10^{-6}$, $1.2 \cdot 10^{-7}$ or $1.8 \cdot 10^{-12}$ respectively. The calculations represent DDS output frequency resolutions relative to the DCF77 frequency. Without the binary frequency divider, the use of a DDS with a larger FTW would be necessary (more expensive, higher power requirements).

The DDS signal is filtered with a low-pass filter and further divided by 32 with a binary counter in order to increase frequency resolution of the DDS signal. In this way a square wave signal with a duty cycle of 50 % and with a relative frequency resolution of $6 \cdot 10^{-8}$ at 77.5 kHz is obtained. The square wave output signal drives the sampling detector based on a quad bilateral high-speed complementary metal oxide semiconductor analog switch. Two switches are used as a single pole double throw (SPDT) switch and are driven by a complementary 77.5 kHz signal. The other two switches are inserted between the SPDT switch and the capacitor of the sampling detector and are controlled by the interference detector circuit (noise blanker).

Instead of phase locking the NMR TCXO to the DCF77 signal, only the difference of the two frequencies is measured and displayed as time of the half-period (time between two reference crossings of the demodulated signal). It is also advantageous to measure time because the amplitude of the DCF77 received signal could be monitored at the same time. If the relative frequency uncertainty of the reference NMR oscillator is in the order of 10^{-7}, it can be neglected, compared to other NMR magnetometer measurement uncertainty contributions. This value of relative frequency uncertainty can be assured without phase locking the NMR TCXO.

The FTW of the DDS used in the DCF77 receiver is changed and the time of the period t_p of the demodulated signal is measured. The FTW at which the t_p reaches the largest value is used in calculation of the TCXO frequency.

The DDS is controlled by the microcontroller of the NMR magnetometer.

Experimental results

The prototype of the presented receiver was built and tested.

Figure 2 shows the schematic diagram of the DCF77 antenna circuit. Coil L1 is wound on a ferrite bar and together with the fixed capacitor C1 and trimmer capacitor C2 forms a resonant circuit tuned to 77.5 KHz. Transistor Q1 forms the first amplifier stage. Its output is fed via a coaxial cable to the DCF77 receiver.

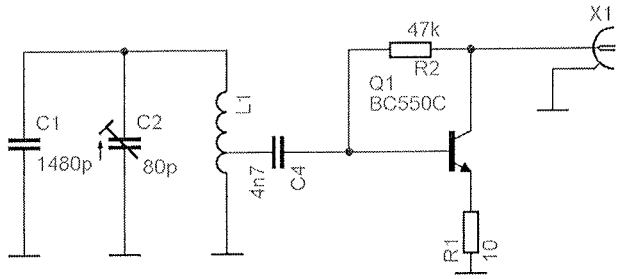


Fig. 2: Schematic diagram of the DCF77 antenna circuit.

Figure 3 shows the DDS circuit. The reference oscillator output is buffered and fed to the DDS (IC1) and to the NMR magnetometer (REF CLK). The control signals are generated by the microcontroller of the NMR magnetometer.

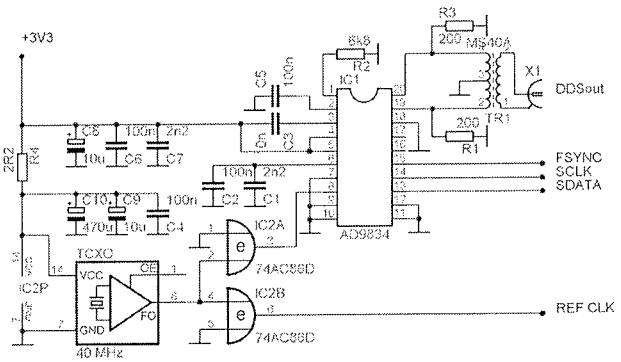


Fig. 3: Schematic diagram of the DDS circuit.

The DCF77 receiver schematic is shown in Figure 4. The constant current source for DCF77 antenna supply is built with a field-effect transistor Q2. The DCF77 signal amplitude at the source of Q2 is approximately 0.07 mV. The DCF77 signal is amplified in a two stage amplifier to approximately 8 mV and band-pass filtered with a resonant circuit tuned to 77.5 KHz (C2, TR1). The amplified signal is fed to a sampling detector built with a quad analog switch (IC1). Two switches operate at the receiving frequency,

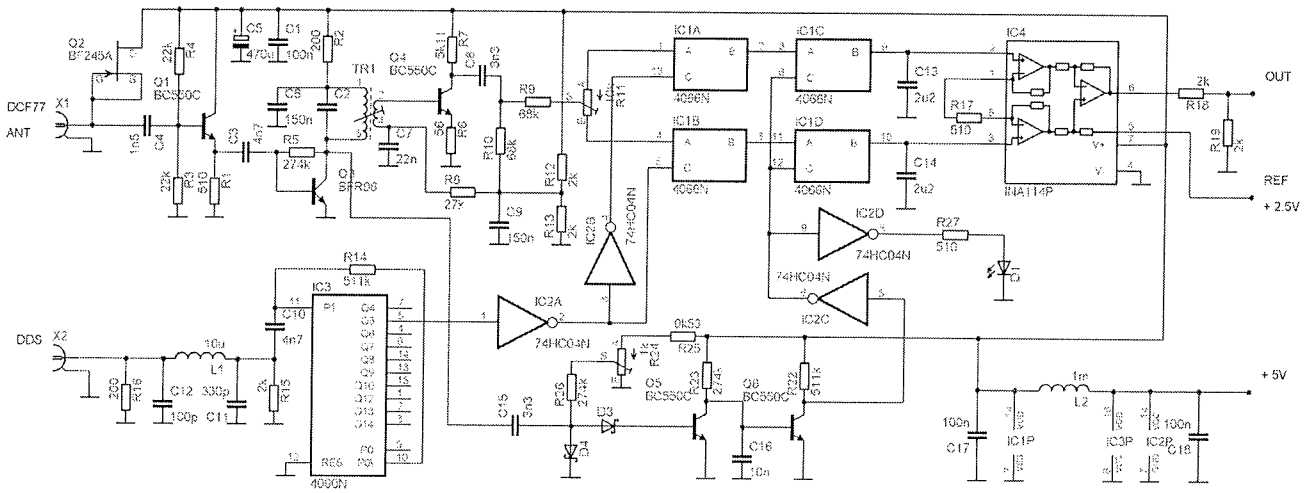


Fig. 4: Schematic diagram of the DCF77 receiver.

while the other two disconnect the sampling capacitors C13 and C14 in case of strong interferences and act as a switch of the noise blanker circuit. The direct current offset of the demodulated signal is adjusted to the mid-level with R11. The demodulated signal is amplified (40 dB) with the instrumentation amplifier IC4 and fed to the analog-to-digital converter (ADC) of the microcontroller via a resistive divider to adapt to the ADC input voltage range.

The reference signal for the sampling detector from the DDS is low-pass filtered (C12, L1, C11), converted to the square-wave and divided by 32 with the binary counter IC3. The divided output and the inverted output are used to control the analog switches of the sampling detector (IC1A and IC1B).

The noise blanker consists of a two-diode detector circuit (D3, D4), DC amplifier (Q5, Q6) and a signal shaper (IC2C). The level at which the noise-blanker is activated is adjusted with the potentiometer R24.

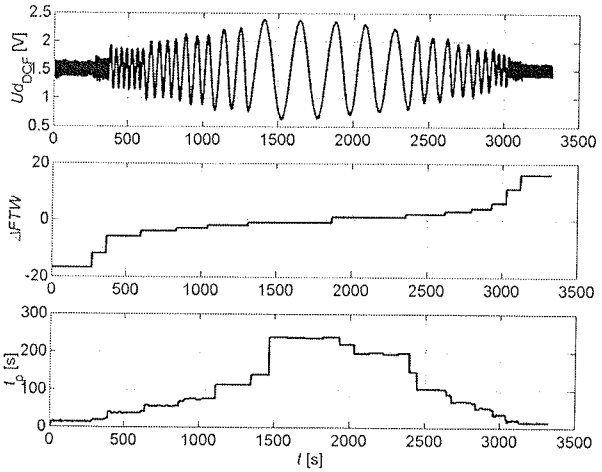


Fig. 5: Demodulated signal from the DCF77 transmitter Ud_{DCF} *and the time of one periode at different settings of the DDS FTW of the DCF77 receiver* ΔFTW, ΔFTW = FTW - *16643112.*

The DCF77 receiver uses 3.3 V and 5 V power supply. The current consumption of the 3.3 V power supply (used for the reference oscillator and the DDS) is 20 mA, the current consumption of the 5 V power supply (used for other circuity of the DCF77 receiver) is 23 mA.

The digitized demodulated signal from the DCF77 receiver and the FTW value were sent to the personal computer and stored on the hard disc for later processing and presentation.

The output frequency of the DDS is set to 77.5 kHz · 32 = 2.480 MHz. Assuming a TCXO frequency of 40 MHz, the calculated *FTW* is:

$$FTW = \frac{f_{OUT} \cdot 2^{28}}{f_{REF}} = 16642998 \qquad (1)$$

where f_{OUT} is the output frequency of the DDS and f_{REF} is the frequency of the reference oscillator (TCXO). Figure 5 shows the filtered demodulated signal from the DCF77 transmitter Ud_{DCF} and the time of one period t_p at different settings of the DDS FTW of the DCF77 receiver *DFTW*. The amplitude of the demodulated signal is changed because of the low-pass filter. The filter is a simple moving average filter, with the equation

$$Ud_{DCF} = (1 - kf_{DCF}) \cdot Ud_{DCF} + kf_{DCF} \cdot PSDdcf_{ADC}, \qquad (2)$$

where the $PSDdcf_{ADC}$ is the signal form the ADC and kf_{DCF} is an adjustable parameter of the filter. The t_p has the largest value when *FTW* = 16643112. This *FTW* is used in calculation of the TCVCXO frequency f_{REF}:

$$f_{REF} = \frac{f_{DCF77} \cdot 32 \cdot 2^{28}}{FTW} = 39.999726 \text{ MHz}. \qquad (3)$$

Figure 6 shows the spectrum of the demodulated and filtered signal Ud_{DCF} from the DCF77 transmitter. The peak at 0.088 Hz corresponds to t_p of 114 s. The peak at 1 Hz

and its harmonic components are due to the amplitude modulation of the DCF77 signal used for transmission of the time code.

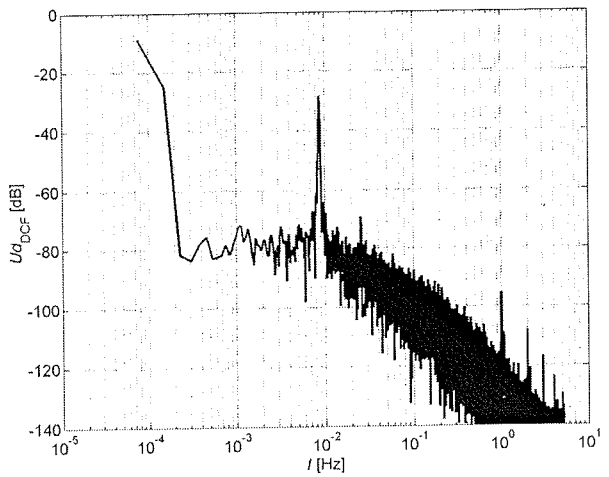


Fig. 6: Spectrum of the demodulated signal from the DCF77 transmitter.

The spectrum was obtained from the recorded signals during the measurement with the NMR magnetometer lasting over 3 hours and a half. The *FTW* was changed to 16643110 intentionally to obtain a stable periodic signal Ud_{DCF}. The time of one period t_p should not be less than 75 s in order to obtain a relative uncertainty of the reference crystal oscillator of less than 10^{-7}.

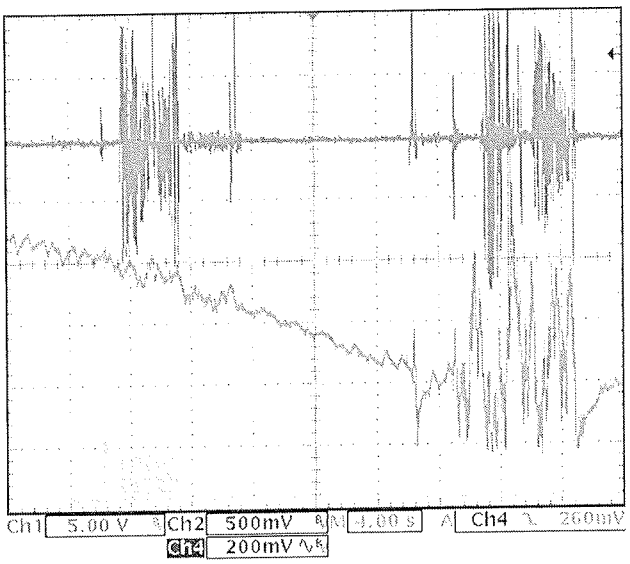


Fig. 7: Demodulated signal with enabled and disabled automatic interference filter.

The noise-blanker efficiently suppresses the interferences, as can be seen in Figure 7. The left half of the figure was recorded with the noise blanker enabled. The upper trace is the radio frequency signal present at the resonant circuit formed by TR1 and C2, the middle signal is the demodulated signal fed to the ADC and the lower signal is taken from the buffer IC2D and has high level when the analog switches disconnect the sampling capacitors due to the interferences received. On the right half of the figure the noise-blanker has been disabled. The received interferences cause high noise level superimposed on the demodulated signal. Low level variations of the demodulated signal are due to the amplitude modulation of the DCF77 signal.

Conclusion

Frequency traceability of the NMR magnetometer reference oscillator is established by means of the DCF77 long-wave transmitter with a relative uncertainty of less than 10^{-7}. Beside the DDS integrated circuit the DCF77 receiver uses only standard elements. The microcontroller used is part of the NMR magnetometer. The software is simple to implement as only a moving average filter and a counter to measure the time of one period of the demodulated signal are needed.

The presented system could be used also with other instruments that need reference oscillator traceability.

Acknowledgment

The Laboratory of Magnetic Measurements would like to thank the Metrology Institute of the Republic of Slovenia and the Ministry of Education and Sport for financial support with maintenance of the national standard for magnetic flux density.

References

/1/ Begus S and Fefer D, A combined digital-analog absorption type proton NMR magnetometer for measuring low magnetic fields, (to appear in Measurement Science and Technology).

/2/ Lombardi M. A., Nelson L. M., Novick A. N. and Zhang V. S., Time and Frequency Measurements Using the Global Positioning System, Cal Lab, Jul. Avg. Sept., 2001.

/3/ Shera B., A GPS-Based Frequency Standard, QST, July, 1998.

/4/ Kamas G and Lombardi M A 1990, Time and Frequency Users Manual NIST special publication 559.

/5/ Piester D, Hetzel P, Bauch A, Zeit- und Normalfrequenzverbreitung mit DCF77, PTB-Mitteilungen 114, Heft 4, 2004.

/6/ Youngblood G, A Software-Defined Radio for the Masses, Part 1, QEX Jul-Aug, 2002.

Samo Beguš
E-mail: samo.begus@fe.uni-lj.si
Dušan Fefer
E-mail: dusan.fefer@fe.uni-lj.si

Faculty of Electrical Engineering
University of Ljubljana
Tržaška 25, 1000 Ljubljana, Slovenia
Tel: +386 1 47 68 216, Fax: +386 1 47 68 214

Prispelo (Arrived): 27.10.2006 *Sprejeto (Accepted):* 30.03.2007

Informacije MIDEM
Strokovna revija za mikroelektroniko, elektronske sestavine dele in materiale

NAVODILA AVTORJEM

Informacije MIDEM je znanstveno-strokovno-društvena publikacija Strokovnega društva za mikroelektroniko, elektronske sestavne dele in materiale - MIDEM. Revija objavlja prispevke s področja mikroelektronike, elektronskih sestavnih delov in materialov. Ob oddaji člankov morajo avtorji predlagati uredništvu razvrstitev dela v skladu s tipologijo za vodenje bibliografij v okviru sistema COBISS.
Znanstveni in strokovni prispevki bodo recenzirani.

Znanstveno-strokovni prispevki morajo biti pripravljeni na naslednji način:

1. Naslov dela, imena in priimki avtorjev brez titul, imena institucij in firm
2. Ključne besede in povzetek (največ 250 besed).
3. Naslov dela v angleščini.
4. Ključne besede v angleščini (Key words) in podaljšani povzetek (Extended Abstract) v angleščini, če je članek napisan v slovenščini
5. Uvod, glavni del, zaključek, zahvale, dodatki in literatura v skladu z IMRAD shemo (Introduction, Methods, Results And Discsussion).
6. Polna imena in priimki avtorjev s titulami, naslovi institucij in firm, v katerih so zaposleni ter tel./Fax/Email podatki.
7. Prispevki naj bodo oblikovani enostransko na A4 straneh v enem stolpcu z dvojnim razmikom, velikost črk namanj 12pt. Priporočena dolžina članka je 12-15 strani brez slik.

Ostali prispevki, kot so poljudni cčlanki, aplikacijski članki, novice iz stroke, vesti iz delovnih organizacij, inštitutov in fakultet, obvestila o akcijah društva MIDEM in njegovih članov ter drugi prispevki so dobrodošli.

Ostala splošna navodila

1. V članku je potrebno uporabljati SI sistem enot oz. v oklepaju navesti alternativne enote.
2. Risbe je potrebno izdelati ali iztiskati na belem papirju. Širina risb naj bo do 7.5 oz.15 cm. Vsaka risba, tabela ali fotografija naj ima številko in podnapis, ki označuje njeno vsebino. Risb, tabel in fotografij ni potrebno lepiti med tekst, ampak jih je potrebno ločeno priložiti članku. V tekstu je treba označiti mesto, kjer jih je potrebno vstaviti.
3. Delo je lahko napisano in bo objavljeno v slovenščini ali v angleščini.
4. Uredniški odbor ne bo sprejel strokovnih prispevkov, ki ne bodo poslani v dveh izvodih skupaj z elektronsko verzijo prispevka na disketi ali zgoščenki v formatih ASCII ali Word for Windows. Grafične datoteke naj bodo priložene ločeno in so lahko v formatu TIFF, EPS, JPEG, VMF ali GIF.
5. Avtorji so v celoti odgovorni za vsebino objavljenega sestavka.

Rokopisov ne vračamo. Rokopise pošljite na spodnji naslov.

Uredništvo Informacije MIDEM
MIDEM pri MIKROIKS
Stegne 11, 1521 Ljubljana, Slovenia
Email: Iztok.Sorli@guest.arnes.si
tel. (01) 5133 768, fax. (01) 5133 771

Informacije MIDEM
Journal of Microelectronics, Electronic Components and Materials

INSTRUCTIONS FOR AUTHORS

Informacije MIDEM is a scientific-professional-social publication of Professional Society for Microelectronics, Electronic Components and Materials - MIDEM. In the Journal, scientific and professional contributions are published covering the field of microelectronics, electronic components and materials.
Authors should suggest to the Editorial board the classification of their contribution such as : original scientific paper, review scientific paper, professional paper...
Scientific and professional papers are subject to review.

Each scientific contribution should include the following:

1. Title of the paper, authors' names, name of the institution/company.
2. Key Words (5-10 words) and Abstract (200-250 words), stating how the work advances state of the art in the field.
3. Introduction, main text, conclusion, acknowledgements, appendix and references following the IMRAD scheme (Introduction, Methods, Results And Discsussion).
4. Full authors' names, titles and complete company/institution address, including Tel./Fax/Email.
5. Manuscripts should be typed double-spaced on one side of A4 page format in font size 12pt. Recommended length of manuscript (figures not included) is 12-15 pages
6. Slovene authors writing in English language must submit title, key words and abstract also in Slovene language.
7. Authors writing in Slovene language must submit title, key words and extended abstract (500-700 words) also in English language.

Other types of contributions such as popular papers, application papers, scientific news, news from companies, institutes and universities, reports on actions of MIDEM Society and its members as well as other relevant contributions, of appropriate length , are also welcome.

General informations

1. Authors should use SI units and provide alternative units in parentheses wherever necessary.
2. Illustrations should be in black on white paper. Their width should be up to 7.5 or 15 cm. Each illustration, table or photograph should be numbered and with legend added. Illustrations, tables and photographs must not be included in the text but added separately. However, their position in the text should be clearly marked.
3. Contributions may be written and will be published in Slovene or English language.
4. Authors must send two hard copies of the complete contributon, together with all files on diskette or CD, in ASCII or Word for Windows format. Graphic files must be added separately and may be in TIFF, EPS, JPEG, VMF or GIF format.
5. Authors are fully responsible for the content of the paper.

Contributions are to be sent to the address below.

Uredništvo Informacije MIDEM
MIDEM pri MIKROIKS
Stegne 11, 1521 Ljubljana, Slovenia
Email: Iztok.Sorli@guest.arnes.si
tel.+386 1 5133 768, fax.+386 1 5133 771