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

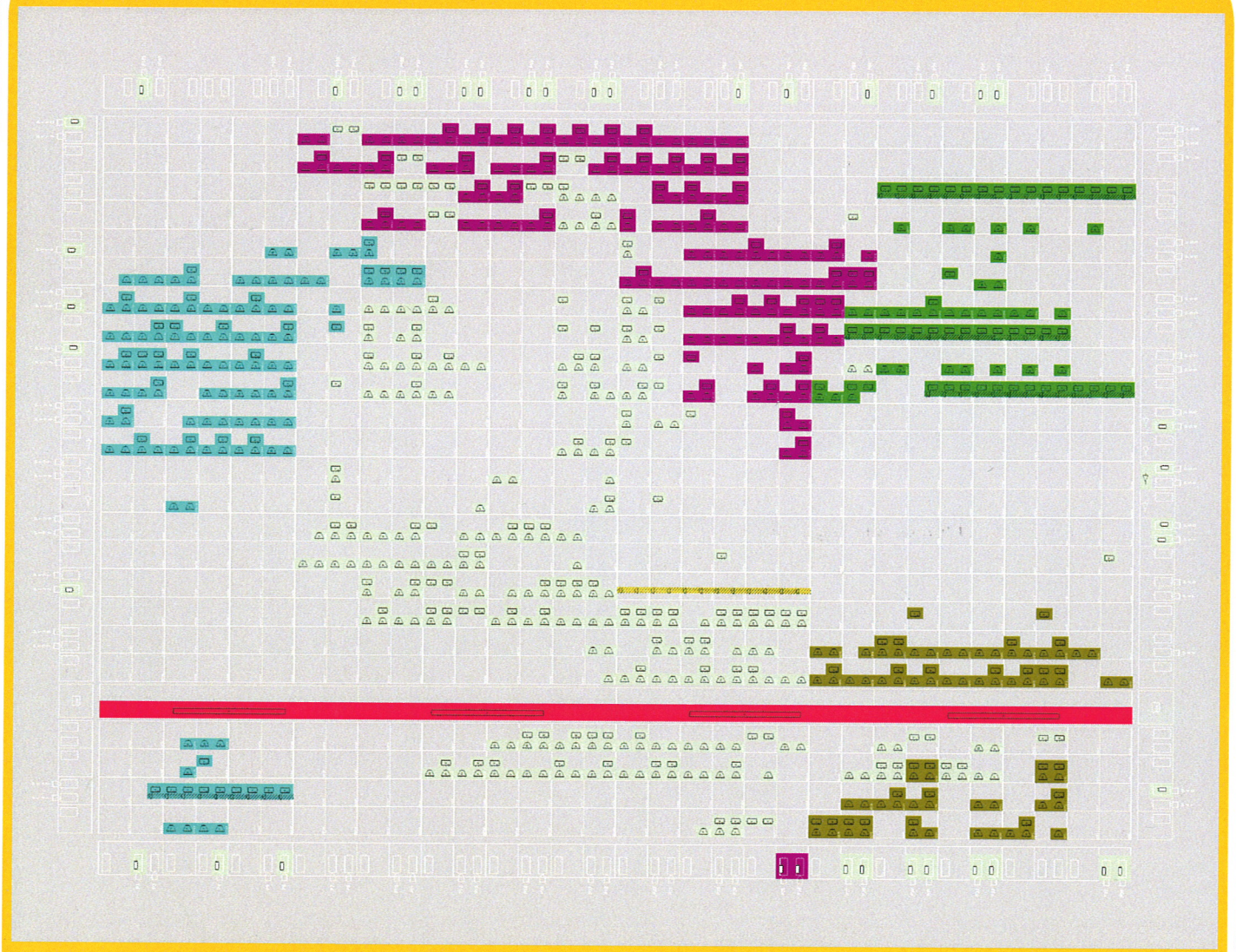
INFORMACIJE MIDEM

Strokovno društvo za mikroelektroniko elektronske sestavne dele in materiale

1·2009

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

INFORMACIJE MIDEM, LETNIK 39, ŠT. 1(129), LJUBLJANA, marec 2009



LNIV Laboratorij za načrtovanje integriranih vezij

FAKULTETA ZA ELEKTROTEHNIKO
UNIVERZA V LJUBLJANI

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

INFORMACIJE MIDEM 1 ◦ 2009

INFORMACIJE MIDEM LETNIK 39, ŠT. 1(129), LJUBLJANA, MAREC 2009
INFORMACIJE MIDEM VOLUME 39, NO. 1(129), LJUBLJANA, MAREC 2009

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 39(2009)1, Ljubljana

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 2009, brisati iz seznama članstva.

Prijavnice pošljite na naslov:

MIDEM pri MIKROIKS

Stegne 11

1521 Ljubljana

Ljubljana, marec 2009

Izvršilni odbor društva

FAST MOS TRANSISTOR MISMATCH OPTIMIZATION – A COMPARISON BETWEEN DIFFERENT APPROACHES

Gregor Cijan[1], Tadej Tuma[2], Sašo Tomažič[3], Árpád Bűrmen[4]

[1] Regional Development Agency of Northern Primorska, Šempeter pri Gorici, Slovenia
[2,3,4] University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia

Key words: MOS transistor mismatch, optimization, mismatch simulation, integrated circuits

Abstract: In this paper two different approaches for calculating the standard deviation of circuit performance measures caused by MOS transistor mismatch are presented. The short CPU time needed for mismatch evaluation makes it possible to include the proposed approaches in a circuit optimization loop as a criterion subject to optimization. Both mismatch evaluation methods were tested on four different circuits. The optimized circuits were compared to the circuits obtained from an optimization run where the list of criteria did not include mismatch. The results show that a significant reduction of standard deviations is obtained when mismatch evaluation is included in the optimization loop.

Hitra optimizacija neujemanja MOS tranzistorjev – primerjava različnih pristopov

Kjučne besede: neujemanje MOS tranzistorjev, optimizacija, simulacija neujemanja, integrirana vezja

Izvleček: V članku sta predstavljena dva različna pristopa za izračun standardnih deviacij lastnosti vezja, ki jih povzroča neujemanje identično načrtovanih MOS tranzistorjev. Glavna prednost opisanih pristopov je hiter izračun standardnih deviacij lastnosti vezja, ki so posledica neujemanja. To je ključnega pomena, če želimo posledice neujemanja vključiti v kriterijsko funkcijo optimizacijskega postopka. Oba pristopa sta bila preizkušena z optimizacijo štirih različnih vezij. Lastnosti tako dobljenih vezij smo primerjali z lastnostmi vezja dobljenega z optimizacijskim postopkom, ki ni vključeval učinkov neujemanja. Primerjava je pokazala, da je tovrstna vključitev neujemanja v optimizacijsko zanko smiselna, saj se standardna devicija lastnosti vezja občutno zmanjša.

1 Introduction

Mismatch is an effect that arises in IC fabrication and is a limiting factor of the accuracy and reliability of many analog integrated circuits. The main reason for mismatch is the stochastic nature of the fabrication process. Due to mismatch two equally designed transistors exhibit different electrical behaviour. Consequently the operating point and other circuit characteristics differ from their desired values. Mismatch can be divided into a systematic and a stochastic component. The systematic component is not considered in this paper because it can be reduced to great extend with proper layout /1/, /2/. The stochastic component is caused by random microscopic device architecture fluctuations. It can be reduced with better process control and larger transistor areas /3/, /4/. Most often the Gaussian distribution is used for modelling the stochastic variations of model parameters. The amount of mismatch can be expressed with standard deviation (σ) of transistor model parameters.

Mismatch can be modelled in many different ways /3/-/6/. Because of the limited availability of mismatch model parameters only some of them can be used for general purpose. One of the simplest models is the Pelgrom model (1) /3/.

$$\sigma(\Delta P) = \frac{A_P}{\sqrt{WL}} \tag{1}$$

In this model the standard deviation (σ) of the parameter difference (ΔP) between two identically drawn transistors depends on parameter A_P (which in turn is technology-dependent) and effective channel dimensions W and L. In the optimization runs presented in this paper we used (1) because it is simple and the technology-dependent parameters are available in the literature /7/. In /8/ it is shown that the model (1) is suitable for the 0,18µm technology. Due to the limited availability of mismatch parameters, this model is still frequently used for mismatch evaluation. In this paper two different methods of mismatch simulation are presented and tested on four different circuits.

Most commonly used transistor parameters in mismatch modelling (mismatch parameters) are threshold voltage (V_T) and current factor (β). The standard deviation of V_T and β can be expressed as

$$\sigma(\Delta V_T) = \frac{A_{V_T}}{\sqrt{WL}} \tag{2}$$

$$\sigma\left(\frac{\Delta\beta}{\beta}\right) = \frac{A_\beta}{\sqrt{WL}} \tag{3}$$

The technology dependant parameters A_{Vt} and A_β for different types of technologies are available in /7/.

2 Mismatch optimization

A robust circuit exhibits adequate performance in all corners. A corner defines a group of different process variations. The performance of the circuit is expressed with the cost function which is a sum of penalties /9/. Each measure has a goal and if the measured value deviates from this

goal, a penalty which is proportional to the violation, is added to the cost function. The goal is to minimize the cost function taking into account all corners. For this purpose the Constrained simplex /9/ optimization method has been used, which performs remarkably well on circuit optimization problems /10/.

To include mismatch in the optimization as yet another criterion, it has to be simulated first. The goal of mismatch simulation is to obtain a standard deviation of circuit properties caused by the stochastic nature of transistor model parameters. This standard deviation can be included in the cost function. In this paper two different approaches for mismatch simulation are presented. The first one is the sensitivity-based approach and the second one is the min-max approach. In both approaches a design of an operational amplifier will be used for better understanding. Consider an operational amplifier where a designer is interested in the standard deviation of the output voltage caused by the stochastic nature of the transistor model parameters. Beside the offset voltage performance measures like swing at gain, bandwidth, phase margin, etc. are also important in the design process. All these performance measures are circuit properties but only offset voltage is relevant for mismatch analysis.

2.1 Sensitivity-based approach

The sensitivity-based approach assumes that mismatch parameters are not correlated and that the changes caused by the stochastic nature of model parameters are within the bounds where the circuit behaves linearly. The evaluation of the standard deviations of the circuit properties can be divided in three major steps:

Step 1: Calculate the standard deviation of every relevant transistor parameter (mismatch parameter).

Step 2: Calculate the sensitivity of circuit properties to all mismatch parameters.

Step 3: Calculate the approximated standard deviation of the circuit property.

In a circuit composed of k MOS transistors only $n \leq k$ MOS transistors are relevant to the mismatch analysis. $m=2 \cdot n$ standard deviations must be calculated (n standard deviations for the threshold voltages and n standard deviation for the current factors). The remains k-n MOS transistors belong usually to the start up circuit or power down control.

In step 2 the sensitivity (α) of a circuit property P_X (in our example this is the output offset voltage) to every mismatch parameter is calculated. The sensitivity indicates how much the variation of a mismatch parameter affects circuit property P_X. The sensitivity is calculated using the perturbation approach (4)

$$\alpha_i = \frac{P_X(\delta_i) - P_X(0)}{\delta_i} \qquad i = 1,\ldots,m \qquad (4)$$

Where $P_X(0)$ is the value of the circuit property when all mismatch parameters are set to their nominal values while the $P_X(\delta_i)$ is the value of the circuit property when one mismatch parameter is perturbed. The perturbed value (for example V_T) is the sum of the nominal value and one standard deviation ($\delta_i = V_T + \sigma(\Delta V_T)$) of the respective mismatch parameter.

Assuming that mismatch parameters are uncorrelated the standard deviation of a circuit property can be expressed as:

$$\sigma^2(P_X) = \sum_{i=1}^{m} \alpha_i^2 \cdot \sigma_i^2 \qquad (5)$$

The sensitivity-based approach requires $m+1$ circuit simulations to calculate the sensitivities. One simulation is needed for the nominal mismatch parameter values and m simulations are needed for the perturbed circuits.

2.2 Min-Max approach

With this approach we estimate the extreme value of a circuit property P. We assume P has an extreme when all mismatch parameters are at their extreme values. This is true if P is a monotonic function of the mismatch parameters. Which extreme value ($+\sigma$ or $-\sigma$) a mismatch parameter should take in order for P to take its extreme value depend on the sensitivity. Just like with the sensitivity-based approach we assume that mismatch parameters are not correlated. Once the upper (max) and lower (min) extreme of P are obtained, the upper bound on the standard deviation of the circuit property can be calculated. The min-max approach can be divided in 4 steps.

Step 1: Calculate the standard deviation of every mismatch parameter.

Step 2: Obtain the signs of the sensitivities to all mismatch parameters.

Step 3: Measure the extreme (min and max) values.

Step 4: Calculate the upper bound on the standard deviation.

Step 1 and step 2 are very similar to the corresponding steps in the sensitivity-based approach. To obtain the signs of the sensitivities $m+1$ simulations are necessary (m is the number of mismatch parameters). In step 3 the extreme values of the circuit properties are calculated. This is done using two simulations per circuit property (one for the upper and one for the lower extreme). To measure the upper extreme (P_{max}) we increase (or decrease) the value of every mismatch parameter by one standard deviation if the sensitivity is positive (if the sensitivity is negative). The same is done for the lower extreme (P_{min}), except that the reasoning is opposite. For positive sensitivity the value of the mismatch parameter is decreased and for negative sensitivity it is increased. In step 4 the upper bound on the standard deviation can now be calculated (6).

$$\sigma(P_X) \leq M_\sigma(P_X) = \left| \frac{P_{\max} - P_{\min}}{2} \right| \qquad (6)$$

3 Examples

The proposed approaches for mismatch evaluation were included in the optimization loop and the obtained results were compared with the results of the optimization run without mismatch evaluation. The comparison was done on four different circuits:

- Bandgap reference circuit (BGR)
- Operational amplifier (OPA)
- Beta multiplier reference circuit (BMR)
- Comparator (COMP)

The first optimization run (A) considers only performance measures, while the second and the third run (B, C) include mismatch. In the second run (B) the sensitivity-based approach is used for evaluating the mismatch and in the third run (C) the min-max approach is used. All circuits in this paper have been simulated using the SPICE OPUS simulator and the BSIM3 model of a 0,18 µm process technology. To obtain robust circuits every circuit has been simulated in three different corners. Every corner was described with the corresponding temperature, supply voltage, MOS transconductance, etc.

3.1 Bandgap reference (BGR)

A stable voltage reference is very important in many circuits. A bandgap voltage reference (BGR) is capable of providing a voltage almost independent of temperature and supply voltage fluctuations. In this paper we optimized a 1-V low power CMOS bandgap reference based on resistive subdivision (Figure 1) which is in detail described in /11/.

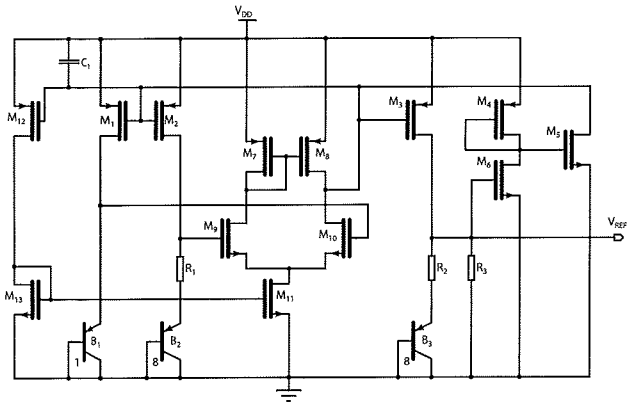


Fig 1: Bandgap voltage reference (BGR)

It is known that sample to sample variations (mismatch) are larger than the variations caused by the temperature or supply voltage fluctuation. Therefore mismatch is a dominating factor determining the absolute accuracy of the bandgap reference circuit /11/.

The optimization parameters were the three resistances (R_1-R_3) and the channel dimensions (width and length) of all MOS transistor except transistors M_4-M_6 which constitute the start up circuit. The list of performance measures and parameters was the same for all of the three optimization runs. The cost function was composed of the following performance measures:

- output voltage change when the temperature varies from -20°C to 50°C,
- output voltage change when the supply voltage varies from 1V to 1,6V,
- circuit area,
- standard deviation of the output voltage caused by mismatch.

Table 1 lists the results of the optimization. The first and the second column contains the names and the desired values of the performance measures while the others list the results of optimization runs A, B and C. The standard deviation σ (V_{REF}) is calculated from 1000 Monte Carlo simulations.

Table 1: Comparison of three different optimization runs (BGR-circuit)

		Desired value	Optimization processes		
			BGR-A	BGR-B	BGR-C
Perf. Meas.	Max (dV_{REF}/dV_{DD}) [mV/V]	< 4	1,80	3,04	3,39
	Max (dV_{REF}/dT) [mV/°]	< 0,15	0,15	0,15	0,14
	Area [µm²]	< 6000	3521	5996	5985
	σ (V_{REF}) [mV]	< 7	**16,2**	**7,05**	**6,67**

It can clearly be seen that including mismatch effects in the cost function results in the enlargement of the transistors area and the reduction of the reference voltage variation caused by mismatch. The maximal reference voltage slope with respect to the supply voltage increases, while the maximal slope with respects to the temperature remains almost unchanged. Despite the large reduction of the standard deviation (more than 2 times) all the results are still within the specified bounds. Figure 2 shows 30 Monte Carlo simulations of the circuit obtained from run A where mismatch was neglected. If we compare this figure to figure 3 where the results of run B are plotted the reduction of the standard deviation is clearly visible.

3.2 Operational amplifier (OPA)

The operational amplifier is one of the fundamental building blocks of analog integrated circuits. Due to mismatch an operational amplifier exhibits a random offset voltage. In this paper the operational amplifier from figure 4 was optimized.

The optimization parameters were the capacitance of the capacitor and the channel dimensions (width and length) of all transistors except M_{N1S} and M_{P1S}, which are used for shutting down the amplifier. All performance measures and their goals (desired value) are listed in table 2. The standard deviation $\sigma(V_{OUT})$ is calculated from 1000 Monte-Carlo simulations.

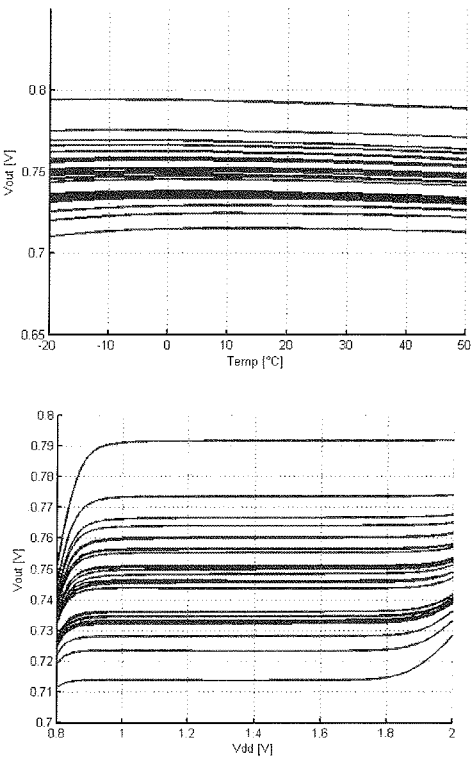


Fig 2: Variation of the BGR output voltage with respect to the temperature and the supply voltage when mismatch is not included in the optimization loop (30 samples).

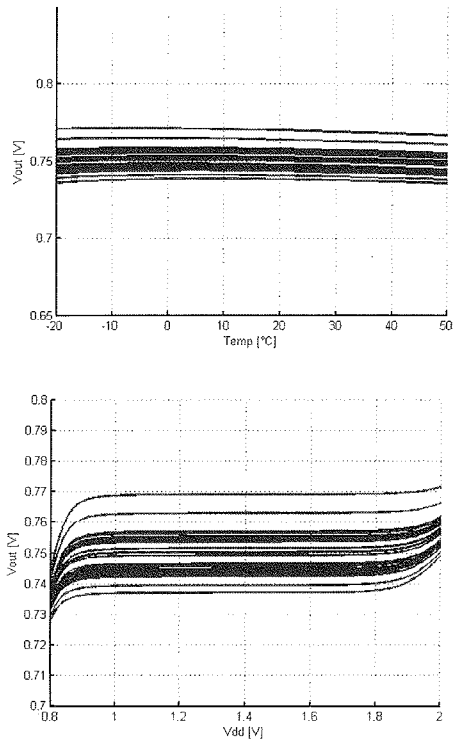


Fig 3: Variation of the BGR output voltage with respect to the temperature and the supply voltage when mismatch is included in the optimization loop (30 samples).

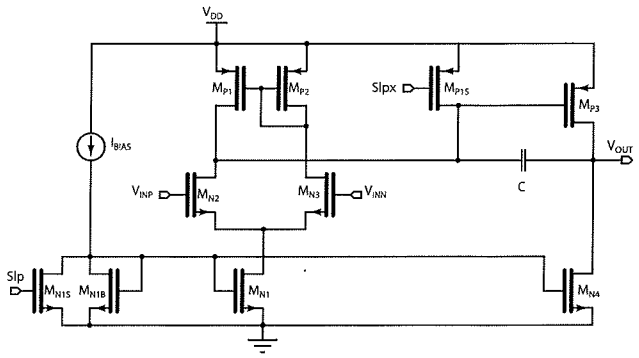


Fig 4: Operational amplifer (OPA)

Table 2: Comparison of three different optimization runs (OPA-circuit)

		Desired value	Optimization processes		
			OPA-A	OPA-B	OPA-C
Perf. measures	Swing at gain [V]	> 2	2,58	2,24	2,18
	Phase margin [°]	> 45	61,8	65,7	67,2
	Unity gain b.w. [MHz]	>18	32,7	20,3	43,1
	Gain [dB]	> 70	72,1	83,5	85,3
	Area [μm²]	< 250	233,8	249,7	249,8
	σ (V_{OUT}) [mV]	< 1,8	**4,44**	**1,87**	**1,70**

The results of the three optimization runs are listed in table 2. We can see that for the circuit obtained from the first run (without mismatch) the output voltage has a standard deviation of 4,44 mV (offset). Both optimization runs that included mismatch produced better results. In runs B and C the standard deviation of the output voltage was reduced by a factor of 2,3 or more. Most of the remaining performance measures stayed within the desired range.

3.3 Beta-multiplier reference (BMR)

The Beta-multiplier circuit is used for providing a stable and temperature independent current reference for a whole range of circuits like operational amplifiers, comparators, etc. It can also be used as a voltage reference circuit.

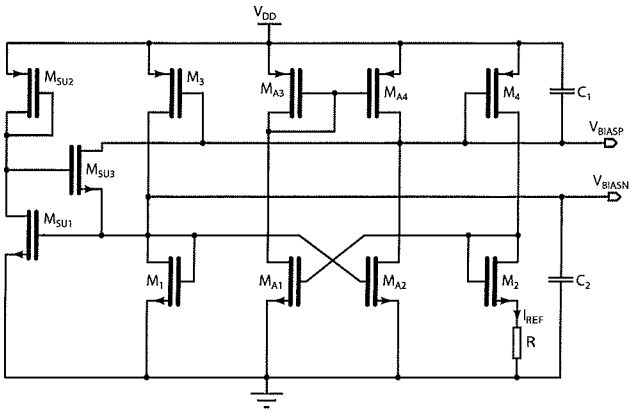


Fig 5: Beta multiplier reference (BMR)

The circuit in figure 5 can provide a stable current (I_{REF}) that flows through resistance R. This current is fairly stable with respect to temperature and supply voltage variations.

One can mirror the reference current using gate voltages V_{BIASP} and V_{BIASN}. The optimization goals were to minimize the variation of the current caused by the change of temperature and supply voltage. The optimization parameters were the resistance R and channel dimensions of all transistors except transistors that constitute the start up circuit (M_{SU1}, M_{SU2}, M_{SU3}). The results of the three optimization runs are listed in table 3. The standard deviation $\sigma(I_{REF})$ is calculated from 1000 Monte Carlo simulations.

Table 3: Comparison of three different optimization runs (BMR-circuit)

		Desired value	Optimization processes		
			BMR-A	BMR-B	BMR-C
Perf Meas	Max (dI_{REG}/dV_{DD}) [µA /V]	< 0,4	0,31	0,30	0,49
	Max (dI_{REG}/dT) [µA /°]	< 0,06	0,047	0,053	0,060
	I_{REG} [µA]	= 20	19,9	20,0	20,0
	Area [µm^2]	< 500	490	497	513
	σ (I_{REG}) [µA]	< 1	**3,45**	**0,99**	**0,53**

The first two performance measures (dI_{REG}/dV_{DD} and dI_{REG}/dT) provide the information on the output current maximum variation with respect to the supply voltage and temperature variations. The third measure is the value of the reference current while the last two are the circuit area and the standard deviation of the reference current caused by mismatch. From table 3 it can clearly be seen that the standard deviation ($\sigma(I_{REG})$) from runs B and C (where mismatch was included in the cost function) is more than 3 times smaller than the one obtained in run A. All the design requirements are fulfilled except the area in run C where a small goal violation occurs. The standard deviation obtained in run C is nearly two times smaller then the goal. The reason is that min-max approach calculates the maximum value of the standard deviation, which is in this example 2 to 3 times bigger than the real standard deviation (see table 5). Due to this the weight of this performance measure is effectively bigger than the weight of the circuit area.

3.4 Comparator (COMP)

The comparator from Figure 6 is a decision-making circuit. The output voltage (V_{out}) switches from 0V to V_{dd} when V_{INP} is greater than V_{INN}. The output switches back to 0V when V_{INP} becomes smaller than V_{INN}. The output does not switch instantly when the difference V_{INP}-V_{INN} changes sign meaning that there is some hysteresis present in the circuit. Mismatch causes that the width of the hysteresis to vary randomly. The optimization run which includes mismatch attempts to remove the variation of hysteresis.

The optimization parameters were the channel dimensions of all transistors. The results of the optimization runs are listed in table 4.

The first performance measure (Delay time LH) measures the time form the moment when V_{INP} crosses V_{INN} and the moment when V_{OUT} reaches 90% of the difference between the initial and the final value. The second measure (Delay time HL) is the same as previous with the difference that

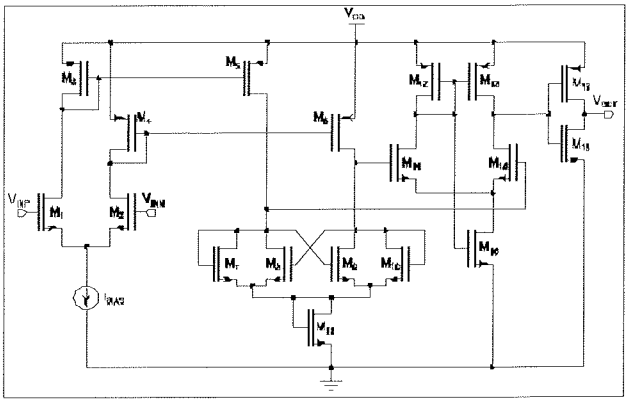


Fig 6: Comparator (COMP)

Table 4: Comparison of three different optimization runs (COMP-circuit)

		Desired value	Optimization processes		
			COMP-A	COMP-B	COMP-C
Perf. measures	Delay time LH [ns]	< 5	3,68	4,94	5,80
	Delay time HL [ns]	< 5	3,87	3,38	4,66
	Rise time [ns]	< 0,3	0,17	0,13	0,11
	Fall time [ns]	< 0,3	0,18	0,18	0,26
	Hysteresis [mV]	< 1	0,83	0,56	1,95
	Positive slope [mV]	< 1	0,93	0,46	0,40
	Negative slope [mV]	< 1	0,51	0,46	0,40
	Overshoot [mV]	< 50	13,7	6,18	8,26
	Undershoot [mV]	< 50	6,63	8,52	4,18
	Area [µm]	< 90	85,4	89,5	88,1
	σ (V_{HIST}) [mV]	< 5	**10,2**	**5,29**	**4,50**

the falling edge of V_{OUT} is measured (when V_{OUT} reaches 10% of the difference between the initial and final value). The rise time is the time needed for the output to rise from 10% to 90% of the difference and the fall time is measured between the points where the output crosses the 90% and 10% level of the difference. The fifth measure is the width of the hysteresis, while the last two measures provide the slope of the hysteresis. From the results it can be seen that almost all the design requirements are fulfilled and that the standard deviation of the hysteresis is reduced to half if mismatch is considered in the optimization run.

3.5 Comparison of the approaches for mismatch evaluation

In table 5 the standard deviations of the circuit properties affected by mismatch are listed for all four circuits. For every circuit the computational effort of the mismatch evaluation has been calculated for all three methods: Monte-Carlo, sensitivity-based approach and min-max approach.

Table 5 shows that the mismatch effect calculated with the sensitivity-based approach is close to the value obtained from 1000 Monte Carlo simulations. The min-max approach overestimates the mismatch. This can also be seen from table 5. Typically the min-max approach results in 2 to 4 times larger values than Monte-Carlo analysis. The main difference between the two presented approaches and Monte-Carlo approach is the number of simulations need-

Table 5: Comparison between of different approaches for mismatch evaluation

	Monte Carlo Approach	Sensitivity-Based Approach	Min-Max Approach
BGR-A [mV]	16,2	16,3	43,6
BGR-B [mV]	7,05	7,04	19,3
BGR-C [mV]	6,67	6,71	17,5
OPA-A [mV]	4,45	4,40	9,59
OPA-B [mV]	1,87	1,87	4,49
OPA-C [mV]	1,70	1,71	4,03
BMR-A [µA]	3,45	3,39	7,53
BMR-B [µA]	0,99	0,99	2,71
BMR-C [µA]	0,53	0,52	1,64
COMP-A [mV]	10,2	9,59	34,8
COMP-B [mV]	5,29	4,99	19,9
COMP-C [mV]	4,50	5,57	12,5

ed to evaluate the standard deviation of a circuit property. To obtain the actual value of the standard deviation of a circuit property 1000 or more Monte-Carlo simulations are needed. The sensitivity-based approach is significantly faster since it needs only $m+1$ simulations (where m is the number of mismatch parameters) to obtain similar values as Monte-Carlo approach. The min-max approach also gives satisfying results with only $m+3$ simulations.

4 Conclusion

With the reduction of transistor dimensions the mismatch is becoming the dominating factor of the accuracy of many analog circuits. In the examples it was shown how mismatch can be included in circuit optimization. Two different ways of mismatch evaluation were presented. The sensitivity-based approach returns more realistic values while on the other hand the min-max approach results in the upper (lower) bound of a circuit performance measures. Optimization runs using these two methods have been conducted on four different circuits and the results were compared to the results of an optimization run where mismatch was neglected. The comparison shows that significant improvements of circuit performance can be achieved. Both optimization runs where mismatch was included resulted in circuits that exhibited similar performance.

5 Acknowledgment

The research has been supported by the Ministry of Higher Education, Science and Technology of Republic of Slovenia within programme P2-0246 - Algorithms and optimization methods in telecommunications.

6 References

/1/ M.F. Lan, A. Tammineedi, R. Geiger, Current mirror layout strategies for enhancing matching performance, Analog Integrated Circuits and Signal Processing, vol. 28, pp. 9-26, 2001

/2/ C. He, X. Dai, H. Xing, D. Chen, New layout strategies with improved matching performance, Analog Integrated Circuits and Signal Processing, vol. 49, pp. 281-289, 2006

/3/ M.J.M. Pelgrom, A.C.J. Duinmaijer, A.P.G. Welbers, Matching properties of MOS transistors, IEEE Journal of Solid-State Circuits, vol. 24, pp. 1433 -1439, 1989

/4/ J. Bastos, M. Steyaert, A. Pergoot, W. Sensen, Mismatch characterization of submicron MOS transistors, Analog Integrated Circuits and Signal Processing, vol. 12, pp. 95-106, 1997

/5/ M. Conti, P. Crippa, S. Orcioni, C. Turchetti, Parametric yield formulation of MOS IC's affected by mismatch effect, IEEE Transaction on Computer-Aided Design of Integrated Circuits and systems, vol. 18, pp. 582-596, 1999

/6/ U. Grünebaum, J. Oehm, K. Schumacher, Mismatch modelling and simulation - a comprehensive approach, Analog Integrated Circuits and Signal Processing, vol. 29, pp. 165-171, 2001

/7/ Peter R. Kinget., Device mismatch and tradeoffs in the design of analog circuits, IEEE Journal of Solid-State Circuits, vol. 40, pp. 1212-1224, 2005

/8/ R. Difrenza, P. Llinares, G. Ghibaudo, The impact of short channel and quantum effects on the MOS transistor mismatch, Solid-State Electronics, vol. 47, pp. 1161-1165, 1997

/9/ A. Buermen, D. Strle, F. Bratkovic, J. Puhan, I. Fajfar, T. Tuma, Automated robust design and optimization of integrated circuit by means of penalty functions, Aeu-International Journal of Electronics and Comunications, vol. 57, pp. 47 - 56, 2003

/10/ J. Puhan, T. Tuma, I. Fajfar, Optimization methods in SPICE: a comparison, ECCTD '99: proceedings, eds. C. Beccarti et. al. (Stresa, Italy, 1999), pp. 1279-1282.

/11/ K. Lasanen, V. Korkala, E. Räisänen-Ruotsalainen, J. Kostamovaara, Design of a 1-V low power bandgap reference based on resistive subdivision, Circuit and Systems, vol. 3, pp. 564 - 567, 2002

Gregor Cijan
RRA severne Primorske d.o.o. Nova Gorica
Mednarodni prehod 6, SI-5290 Šempeter pri Gorici
E-mail: gregor.cijan@rra-sp.si
Telefon: (01) 4768 322

izr. prof. dr. Tadej Tuma
Univerza v Ljubljani, Fakulteta za elektrotehniko
Tržaška 25, SI-1000 Ljubljana
E-mail: tadej.tuma@fe.uni-lj.si
Telefon: (01) 4768 329

prof. dr. Sašo Tomažič
Univerza v Ljubljani, Fakulteta za elektrotehniko
Tržaška 25, SI-1000 Ljubljana
E-mail: saso.tomazic@fe.uni-lj.si
Telefon: (01) 4768 432

doc. dr. Árpád Bűrmen
Univerza v Ljubljani, Fakulteta za elektrotehniko
Tržaška 25, SI-1000 Ljubljana
E-mail: arpadb@fides.fe.uni-lj.si
Telefon: (01) 4768 322

Prispelo (Arrived): 27.11.2007 *Sprejeto (Accepted): 19.03.2009*

SMD FILM CAPACITORS FOR INTEGRATING A/D CONVERTERS

Andrej Levstek, Matija Pirc

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

Key words: integrating A/D converter, SMD film capacitor, dielectric absorption, humidity, surface resistance, polymer dielectric

Abstract: Lead free technology has significantly influenced the choice of commercially available capacitors, especially those intended for surface mount. A case study of the appropriate SMD capacitor selection for a high accuracy integrating A/D converter is presented. The converter is part of a smart sensor that encompasses a simple microcontroller and an analog transducer, which is in this case a platinum resistor. An overview of traditional and new polymer dielectric materials is given with the emphasis on the commercial selection of SMD capacitors. Trends of the film capacitor industry in the recent years are explained through the physical properties of the materials used and the imposed legislative restrictions. The often overlooked pitfalls of capacitor selection are sequentially described. The effect of the dielectric absorption of the charge on the conversion error is theoretically analyzed for the intermittent mode of operation. Inaccuracy due to the recovery of the absorbed charge is eliminated by the use of a polyphenylene sulfide (PPS) film capacitor; however, at high humidity some capacitive sensors exhibit abnormal deviations.

Based on some additional experiments, we determine the influence of the capacitor's parallel resistance on the conversion result. The synergic impact of high humidity and temperatures, and flux residues on the surface resistance of stacked capacitors is proven by experimental measurements carried out in a climatic chamber. The experimental measurements of SMD capacitor insulation resistance show that naked stacked capacitor construction is not suitable for a relative humidity above 80%. In such cases, slightly larger encapsulated SMD capacitors must be used to maintain the desired high accuracy.

SMD filmski kondenzatorji za integracijske A/D pretvornike

Kjučne besede: integracijski A/D pretvornik, folijski SMD kondenzator, dielektrična absorpcija, površinska upornost, polimerni dielektrik

Izvleček: Prepoved uporabe svinca v tehnologiji izdelave tiskanih vezij je pomembno vplivala na tržno ponudbo kondenzatorjev, namenjenih za površinsko montažo. V prispevku je podan potek izbire ustreznega kondenzatorja za integracijski A/D pretvornik z visoko ločljivostjo. Sam D/A pretvornik je del inteligentnega senzorja, ki ga sestavlja mikrokrmilnik in ustrezen analogni merilni pretvornik, v opisanem primeru je to platinski upor. V delu je najprej podan pregled polimernih dielektrikov s poudarkom na njihovi rabi za SMD komponente. Trendi zadnjih let v industriji kondenzatorjev za elektroniko, so opisani s stališča fizikalnih lastnosti dielektričnih filmov in prepovedi rabe svinca v elektronskih napravah. Spregledane pomanjkljivosti izbranega tipa, so podane zaporedno, kot so se pojavljale pri razvoju. Izpeljali smo analitično zvezo med dielektrično absorpcijo kondenzatorja in pogreškom posamične pretvorbe. Merilno napako zaradi sproščanja absorbiranega naboja v času pretvorbe, smo odpravili z uporabo polifenil sulfidnega (PPS) kondenzatorja.

Z eksperimenti v klimatski komori smo ugotovili, da imajo nekatera vezja nenormalno velika odstopanja, ki nastanejo zaradi vpliva vlage na gole PPS kondenzatorje. Izpeljali smo analitično zvezo med velikostjo skupne upornosti med sponkama kondenzatorja in merilno napako. Ugotovili smo sinergični vpliv vlage, temperature in ostankov fluksa na površinsko upornost nezaščitenih kondenzatorjev, ki so izdelani z zlaganjem metaliziranega filma. Meritve izolacijske upornosti so pokazale, da takšni kondenzatorji niso primerni za vlažnosti zraka nad 80 %. Za precizni integrator so primernejši dimenzijsko nekoliko večji SMD kondenzatorji v plastičnem ohišju.

1. Introduction

Electronic integrators offer a simple solution for achieving an accurate A/D conversion of low voltage levels whenever the speed of conversion doesn't play a significant role. Dual slope integrating A/D converters can be implemented by low cost digital microcontroller and a simple additional analog circuit. Such A/D converters can be used for accurate conversion voltages that are proportional to slowly varying physical quantities like temperature or atmospheric humidity.

Conversion errors due to the non-idealities of the analog components, i.e., the input offset voltage of the operational amplifier (opamp), are minimized by a simple solution, so that low-cost analog components can be used. The choice of utilized capacitor seems unimportant since the value of capacitance does not appear in the conversion equations at all /1/. Such an approach is over-simplified but this fact does not become obvious until experimental measurements of prototypes under various climatic conditions are performed. The physical dimensions of electronic circuits are steadily decreasing, which is a consequence of the growing demand for hand-held devices. The introduction of surface-mount technology (SMT), which became widely used around the year 1990, engendered important changes in the field of high performance metallized film capacitors. The small size of SMD capacitors has raised many problems in their construction, because of the intense heat transfer from the metallized soldering pads to the plastic dielectric film during the reflow soldering process.

The first widely used SMD capacitors were multilayer ceramic chips (MLCC). This construction and the high relative dielectric constant of the ceramics result in such capacitors having a high capacitance packing density and relatively low equivalent serial resistance (ESR). The inorganic nature of the dielectric used in MLCCs minimizes the impact of the thermal stress during soldering. Polymer

film capacitors are much more affected by the raised temperature levels because of the thermoplastic nature of the dielectric film. The choice of SMD polymer film capacitors on the electronic components market is predominated by trough-hole film types and by ceramic SMD chips. As a consequence, film capacitors are rather expensive, therefore cheaper alternatives are used wherever it is possible.

In the next chapter an overview of the important polymer dielectric materials is given, followed by a case study of the capacitor selection for an integrating A/D converter. Finally, the results of practical experiments are presented. The comments on the outcomes and some practical hints for the selection of appropriate SMD polymer film capacitors conclude the paper.

2. Materials for film capacitors

The traditionally used plastic materials for dielectric films in capacitors are polystyrene (PS), polyester (PE), polycarbonate (PC) and polypropylene (PP). These materials are used for film capacitors with low loss and stable capacitance in the range from 1 nF up to 10 µF. Film capacitors below 1 nF are offered only by a small number of producers; and especially SMD types are very rare. The range of capacitances below 100 pF is almost exclusively covered by C0G ceramic capacitors (Table 2), which are featured with a low capacitance temperature coefficient α_C and the dissipation factor $\tan\delta$.

2.1 Polystyrene capacitors

For years, polystyrene (PS) capacitors were the best choice for critical analog applications. In the middle of the 1990s the production of PS capacitors slowly ceased. There were several reasons that caused polystyrene capacitors to disappear from production. The maximal operating temperature of PS film and capacitors is very low, only 85°C (see Table 1). Additionally, the low heat resistance of PS film allows neither the construction of SMD components nor the vacuum-deposition of aluminum, hence only film/foil capacitors were (are) produced. This construction lacks the self-healing capability, i.e., the ability to clear faults (such as pores or impurities in the film) under the influence of voltage. Although PS capacitors have low absorption of moisture, they can be easily damaged by printed board cleaning solvents. PS capacitors that are still available from old stocks are not intended for new designs. New materials like polyphenylene sulfide (PPS) should be used instead of PS.

2.2 Polycarbonate capacitors

Polycarbonate (PC) metallized film and film/foil capacitors were traditionally the logical choice for high performance applications for operation at elevated temperatures. This material is featured with a negligible temperature coefficient α_C for temperatures in the range of 20°C ÷ 40°C, which is the common operating temperature range of precision electronic equipment. In spite of the higher operating temperatures of this film, commercially available PC capacitors for surface mount were never produced. In the year 2000, the major producer of PC capacitors WIMA from Germany /2/, ceased their production after finding it unprofitable. This decision caused the major producer of capacitor grade PC film, Bayer AG, to stop its production upon completion of the final order. Nevertheless, PC film capacitors are still available and produced at least by Electronic Concepts Inc. form USA with its own in-house produced dielectric film /3/. Polycarbonate film is almost the perfect material for high performance capacitors but is very sensitive to moisture absorption, thus good encapsulation is required to protect the dielectric film against humidity. Hermetically sealed PC capacitors are available only with wired trough-hole terminals and are primarily intended for military applications.

2.3 Polyester capacitors

Polyester films have become the standard dielectric for capacitors in electronic applications. Polyester film for capacitors is biaxially oriented polyethylene terephthalate (PET) developed by DuPont in the mid-1950s and is well-known under the trade name Mylar. This material has good mechanical and electrical properties for temperatures in the range from -55°C to +125°C. Their high dielectric strength, and the highest dielectric constant among commercially used dielectric films, make PET capacitors cheap and volume-effective. Metallized PET film capacitors are produced in any combination of the construction alternatives given in Table 1.

Table 1: Manufacturing and construction alternatives of PET capacitors

Parameter	Alternative	
Environmental protection	Naked	Protected
Mounting terminals	SMD	Trough hole
Construction	Stacked	Wound
RoHS compliance	Yes	No

These capacitors are the most frequently used type of plastic film capacitors in electronic circuits - primarily for DC or low frequency purposes - because the dissipation factor $\tan\delta$ of polyester is the highest among contemporary film materials. Even though it is not a high quality material, in many respects, PET films perform much better than multilayer ceramic capacitors (MLCC) using X7R or Z5U dielectric ceramics. Some producers, e.g., AVX /4/, offer PET-HT capacitors with an improved temperature range of up to +125°C with a nominal voltage derating factor of 1.25 %/°C above T_R = 105°C, which represents an increase of 20°C with respect to the standard types.

2.4 Polypropylene capacitors

Polypropylene (PP) film has, for many years, been used for high performance applications, especially for medium and high power electronic circuits where high impulse current capability is required. This material has very low dielectric absorption *DA* making a PP capacitor the best choice for the charge-storing device in precision integrators, sample and hold amplifiers and other electronic circuits that retain analog signals in the form of electric charge. Additionally, PP capacitors are characterized by a constant temperature coefficient α_C = -200 ppm/°C and the second highest volume resistivity among dielectric film materials (Table 2). The main deficiency of PP is its somewhat limited temperature range, which prevents the construction of PP as an SMD component. Standard PP capacitors use metallized film and film/metal foil construction for self-healing and high impulse current capability, respectively.

2.5 Polyphenylene sulfide capacitors

Polyphenylene sulfide (PPS), a dielectric material with excellent electrical and thermal properties, was invented by Toray/Japan. This chemical company started the production of capacitor grade PPS film in 1988 under the trade name of Torelina®, and is still the only producer. In the same year PPS capacitors were made commercially available by WIMA/Germany, but their production was plagued with many difficulties. In 2001 WIMA /2/ temporarily ceased their production due to problems connected with inconsistent film quality and availability. A detailed examination of self-healing of different metallized polymer films by Walgenwitz et. al. /5/ has shown only insignificant distinctions among PET, PEN and PPS. In any case, achieving self-healing in PPS film is not a particular problem. The problematic availability of PC capacitors, the European Council Directive on the Restriction of Hazardous Substances (RoHS Directive, 2002/95/EC) /6/, and good the mechanical properties of biaxially oriented PPS film at higher temperatures have prompted numerous activities for the reliable production of PPS capacitors. The construction of SMD film capacitors has become very demanding due to the elevated melting point of leadless soldering compounds. After the break in 2000 WIMA restored the production of PPS capacitors, encouraged by a mixture of technological and commercial factors. After 2001 both trough hole and SMD types of PPS metallized capacitors were made generally available by several manufacturers.

PPS has excellent electrical properties that exceed PC in many aspects. Almost no sensitivity to humidity and a far higher operating temperature range are the most important attributes of PPS when compared to PC. Despite the advantages of PPS film (expressed by the figures in Table 2), this material has two shortcomings. Firstly, it is expensive and secondly, it is produced only by Toray/Japan, which may cause hitches in its supply.

2.6 Polyethylene naphtalate capacitors

Polyethylene naphtalate (PEN) was not used for film capacitors until 2000. The RoHS directive adopted by the EC in 2003 caused producers of electronic components to utilize substitutes for the existing materials as these could not stand the thermal stress generated by elevated soldering temperatures without significant degradation. The electrical properties of PEN film are very similar to those of PET, but the overall performance of PEN is inferior - provided that the maximum operating temperature is not taken into account /7/. PEN capacitors are larger than the corresponding PET types because the dielectric constant ε_r and the dielectric strength E_B of PEN are lower. The ratio of PEN capacitor size to the corresponding PET capacitor is between 1.5 and 2. PEN film SMD capacitors are in compliance with the RoHS directive, and are suitable for IR or vapor phase reflow processes. PEN capacitors are

Table 2: Properties of capacitor dielectric materials

Parameter	Unit	Dielectric								
		PS	PC	PET	PEN	PP	PPS	PTFE	C0G	X7R
Dielectric constant ε_r		2.2	2.9	3.3	3.0	2.2	3.0	2.0	12...40	700...2000
Dielectric strength	V/µm	100	200	400	300	600	250	150	200	
C temperature coefficient α_C	Ppm/°C	-120	±80	+600	+200	-300	-150	-80	±30	±1000
Dissipation factor tanδ (1 kHz)	10^{-4}	5	15	80	80	5	20	1	15	350
Volume resistivity ρ	Ωcm	10^{18}	10^{16}	10^{17}	10^{17}	10^{18}	10^{17}	10^{19}	10^{17}	10^{16}
Dielectric absorption *DA*	%	0.01	0.1	0.5	1.2	0.02	0.05	0.01	0.6	2.5
Operating temperature T_{min}/T_{max}	°C	-55 125	-55 100	-55 105	-55 125	-55 100	-55 140	-55 200	-55 125	-55 125
Self-healing		no	yes	yes	yes	yes	yes	no	no	no
SMD configuration		no	no	yes	yes	no	yes	no	yes	yes

also featured with improved temperature stability with respect to PET. The capacitance temperature coefficient α_C of PEN is approximately only one third of the α_C of PET. The dielectric absorption *DA* of PEN is the biggest among the polymer film dielectrics. Its value of approximately 1 % is the order of magnitude of *DA* specified for MLCCs using X7R ceramics.

2.7 Other dielectric materials

Three types of dielectric materials, that have not been mentioned previously, are also listed in Table 2. Polytetrafluoroethylene (PTFE) better known under DuPont Company's trade name Teflon®, is an excellent insulating material, but PTFE film capacitors are very rare. Proper metallization of PTFE film is very difficult, this material is very expensive and films of a thickness < 6 µm are not commercially available /7/. PTFE capacitors are used in high power applications where their high operating temperature range and low dissipation factor justify their high price.

The data on ceramic dielectric materials in Table 2 are given for comparison, since MLCC chips are very popular and cheap. In fact only C0G ceramics, also known as NP0, are a real match for polymer films as far as stable high performance capacitors go. In addition to the materials discussed, X7R ceramics offer a cost and room efficient solution when large capacitances in small packages with low equivalent serial resistance (ESR) are required.

3. Capacitor for integrating A/D converter

3.1 Four slope integration

Dual slope integration is a well-known method for accurate A/D conversion /1/. Accuracy and resolution are two distinctive features of such A/D converters. The resolution of integrating converters is determined by the ratio between the period of the clock that is used for counting and the time of integration, which is measured in clock periods. Arbitrary resolution can be achieved by appropriate selection of these two parameters, but at high resolutions conversion times can become unacceptably long, since maximal counter frequencies are limited. On the other hand, the accuracy of the result is determined by the used reference, if everything else is done ideally.

Integrating A/D conversion is very useful for measuring slowly varying quantities, e.g., strain, temperature, humidity, illumination etc. Furthermore, smart sensors with digital output can be designed as a combination of a standard microcontroller, an integrating A/D converter, and an analog sensor of a physical quantity. Low cost uncalibrated sensor devices may be used, without compromising the accuracy of the final result because the deficiencies of the analog sensing device are compensated numerically. The required signal conditioning data are obtained by calibration in the final stage of production and consequently stored in the nonvolatile portion of memory.

The pitfalls of capacitor selection are illustrated by the case of the small resistive temperature sensor. The important, i.e., the analog part of the smart sensor is shown by the simplified schematic diagram in Figure 1.

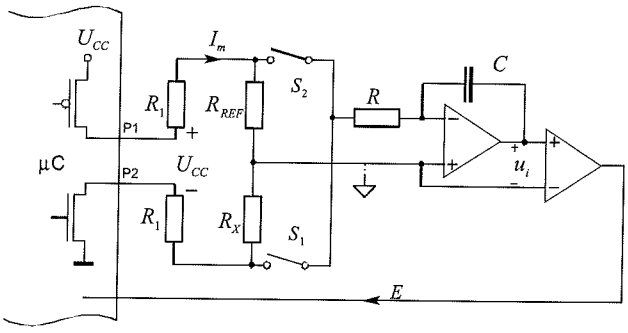


Fig. 1: Integrating A/D converter

The conversion is initiated by closing switch S_1 for fixed time t_0 determined by a certain number of clock periods. The integrator output voltage u_i ramps up, reaching a maximum value that is proportional to the voltage across the sensor resistance R_X. At the end S_1 is opened and S_2 is closed. The output ramps down with a slope that is proportional to the voltage of a very stable resistor R_{REF}. When the integrator voltage becomes negative with respect to analog ground, the timer inside the microcontroller is stopped by the negative edge of the comparator output *E*. The plots of main converter signals are shown in Figure 2.

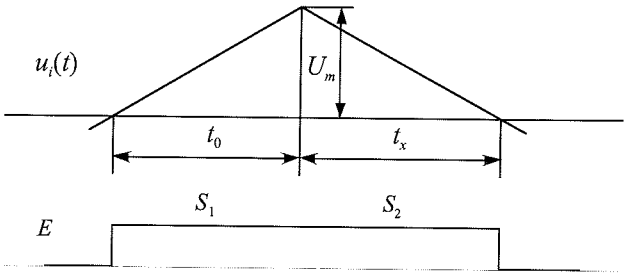


Fig. 2: Time diagram: u_i integrator output, E comparator output

The peak integrator voltage can be expressed by

$$U_m = I_m R_x \cdot \frac{t_0}{RC} = I_m R_{REF} \cdot \frac{t_x}{RC} \qquad (1)$$

where I_m represents the measuring current through the sensor R_X and reference resistor R_{REF}, respectively. From the unknown resistance of the sensor is given by

$$R_x = R_{REF} \frac{t_x}{t_0} \qquad (2)$$

meaning that only the stability of R_{REF} has influence on the result accuracy. This would be true if the opamp and comparator were ideal. The dual slope principle is not sensi-

tive to the instability of the integrator time constant *RC* as long as the constant remains unaltered during conversion time $t_0 + t_x$. The integrator peak voltage U_m given by remains unaffected by the comparator input offset voltage, since the counting of both times, charging t_0 and discharging t_x, are started and stopped at the same integrator voltage, respectively. The actual conversion is started by closing S_2 until the integrator output u_i becomes negative then both switches (S_1 and S_2) are toggled. The charging time t_0 counter is not triggered until the rising edge of the comparator output *E*.

The input offset voltage U_0 of the opamp in the integrator induces an error that can be expressed as

$$\Delta R_X = \frac{2U_0}{I_m} \tag{3}$$

with I_m denoting the measuring current (Figure 1). As it is shown in /8/ this error is compensated by reversing the polarity of the measuring current I_m, which is done by negation of the logic outputs P1 an P2 (Figure 1). The accurate result is the mean of the results obtained with both polarities of the current I_m

$$R_x = R_{REF}\frac{t_{x1} + t_{x2}}{2t_0} \tag{4}$$

The procedure with four slopes of integration, shown in Figure 3, doubles the required conversion time, but low cost opamps with offset voltages $|U_0| \leq 1$ mV may be used.

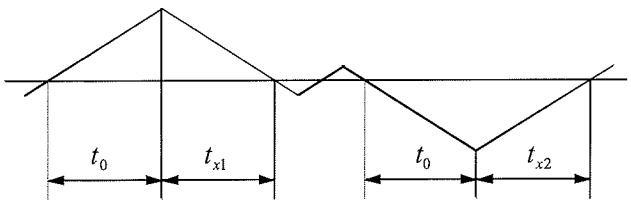


Fig. 3: Plot of the integrator output $u_i(t)$ in the four slope A/D converter

3.2 Dielectric absorption

The analyzed integrator is part of an intelligent resistive sensor of small physical dimensions (25 × 9 mm), therefore small passive components are used. The long integration time, which is necessary to achieve the prescribed resolution, and the low supply voltage require relatively large capacitance *C* = 100 nF that prevents the integrator output from reaching saturation. The first logical choice was an X7R ceramic chip capacitor, characterized by its small dimensions and SMD package. The value of capacitance appears neither in eqn. nor in , therefore the temperature coefficient and tolerance are not important for this purpose.

Experimental tests have shown poor accuracy in the intermittent mode of operation. The sensor was designed for battery powered systems, so the analog part of the circuit is powered only when the conversion takes place. Dielectric absorption of the capacitor has been overlooked, and the effect of the absorbed charge has not been noticed in continuous mode since the capacitor mean voltage is zero.

For the great majority of capacitor applications the dielectric absorption coefficient *DA* is not an important parameter. It matters only in some sample and hold circuits, and as is obvious, in integrators that operate once in a while and have long integrating times. This phenomenon can be measured as a small voltage that reappears across the open capacitor terminals after a charged capacitor has been thoroughly discharged. When voltage is applied to the capacitor plates a certain small part of the stored charge becomes bound on the surface of the dielectric. The process of charge recovery is governed by pretty long time constants that depend merely on the used dielectric material. Measurement of the absorption coefficient *DA* according to the standard MIL-C-19978 D /9/ is depicted in Figure 4.

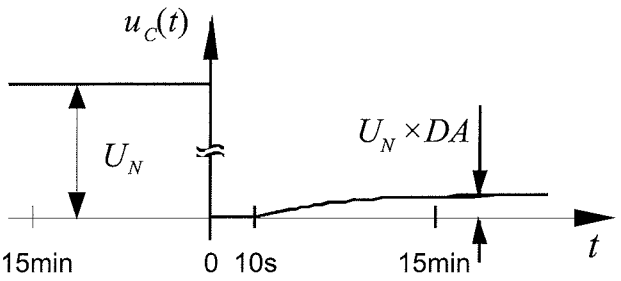


Fig. 4: Timing and definition of voltages associated with the measuring of the dielectric absorption (U_N denotes nominal voltage).

The effects of dielectric absorption in electric circuits are studied by suitable models that replace the capacitor in question. These models /10/, /11/ can be quite complex but in most cases a simple model shown in Figure 5 is sufficient for basic understanding. For commonly used dielectrics, 50% of the final voltage is recovered in 1 to 10 seconds, whereas it can take as much as 15 minutes to reach within 5% of the final value.

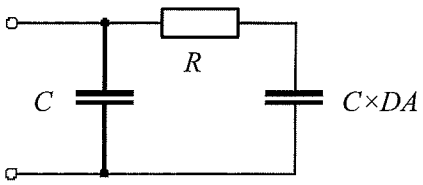


Fig. 5: The basic model of the dielectric absorption in capacitors

The resistance in the model of Figure 6 is given by

$$R = \frac{\tau}{DAC} \tag{5}$$

where *t* denotes the dominant recovery time constant and *DA* is the absorption coefficient (Table 2). The values of τ for particular materials are usually not specified and have

to be determined by experimental measurements if a greater accuracy than the generally presumed range from 1 to 10 s is desired.

Detailed analysis has shown that the integrating capacitor was charged almost to supply voltage (U_{cc} = 3.3 V) when the negative supply pin of the amplifier chip was not tied to ground to reduce supply current. The error caused by recovered charge is drastically reduced by the four slope integration method. The error of n-th successive measurement after the amplifier is turned on is given by

$$\frac{\Delta R_X}{R_X} = \frac{U_{C0}\, DA}{2U_m}\left(1 - e^{-\frac{T}{\tau}}\right)^2 e^{-\frac{2T}{\tau}(n-1)} \qquad (6)$$

where U_{C0} is the integrating capacitor initial voltage, $T = t_0 + t_X$ is the conversion time of one polarity, τ is the dominant time constant of the absorbed charge recovery, and U_m is the peak voltage of the integrator (Figure 2). The used ceramic chip capacitor has turned out to be completely inadequate for accurate temperature measurements on the basis of platinum resistors. Errors due to the dielectric absorption of the consecutive resistance measurements of the platinum resistor R_x (Pt 1000) are shown in Table 3. The results are expressed as temperature errors in °C using the temperature coefficient of platinum 3850 ppm/°C. The values in Table 3 are calculated using for two different absorption coefficients, whereas the other parameters are the same: U_{C0} = 3.3 V, U_m = 1V, τ = 3s, T = 1s. The errors calculated for X7R ceramics are in good agreement with the measurements, which have initiated more detailed analyses.

Table 3: Error of consecutive A/D conversions expressed in temperature

Consecutive measurement no.	ΔT [°C]	
	DA = 2.5% (X7R)	*DA* = 0.05% (PPS)
1	0.860	17×10^{-3}
2	0.442	8.84×10^{-3}
3	0.227	4.54×10^{-3}
4	0.116	2.33×10^{-3}
5	0.059	1.19×10^{-3}

The theoretical error caused by the absorption of PPS film SMD capacitor is smaller than the desired resolution of the design, therefore raw conversion data, i.e., timer counts that measure time t_x, have been observed. The raw results of consecutive conversions (after power up) are within plus minus one count, provided that the temperature is constant.

4. Tests in climatic chamber

4.1 Naked PPS SMD capacitor

Encouraged by the theoretical results (Table 3) an adequate substitution for the X7R MLCC has been found in the form of the stacked PPS film capacitor. These capacitors are almost a perfect choice and feature a very low absorption coefficient *DA* and very low dissipation $\tan\delta < 5\cdot10^{-4}$ in the temperature range from -25°C to 80°C. The data for PPS shown in Table 2 are rather misleading because the worst values over the whole temperature range are given. In addition, PPS capacitors are available as small SMD components that save space on the PCB and fit on solder pads provided for the former ceramic capacitor. The construction of naked stacked film capacitors is shown in Figure 7. The lateral side is usually left without any coating /4/.

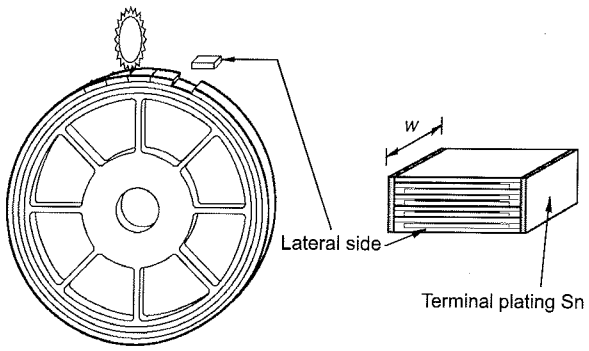


Fig. 6: Production and construction of stacked naked PPS film chip capacitors

In stacked-film production technology, large rings of metallized film are wound onto core wheels with diameters up to 60 cm. Then the rings are sawed apart obtaining well defined width (dimension W in Figure 6). In this way, capacitances with very low tolerances are obtained, since the active body is very homogenous, without the air pockets which are typical of flattened wound bodies. Actual measurements have proved that PPS capacitors have negligible dielectric absorption; hence no differences have been detected between the results of the first and immediately repeated measurements.

Upon the verification of the calibrated sensors, some of them returned values that were up to 2°C lower than the actual temperature in the chamber, which was 12°C. Among the 120 devices under test, about 10% were bad, i.e., $|\Delta T| > 0.1$°C. It turned out that humidity inside the chamber had run out of control. At temperatures around 12°C the humidity had exceeded 90%. The deviation of certain circuits was obviously influenced by humidity, so protection against moisture should improve the performance of the PCB in humid environments. Polyurethane coating applied to the assembled PCBs did not help. The deviations of the bad circuits remained unacceptable.

Since it was not clear which part of the circuit was affected by humidity, several experiments were carried out. The results in Figure 8 show that high humidity and temperature affect the four slope A/D converter. In this experiment the temperature dependent resistors (Pt1000) of three sensors were kept outside the climatic chamber at a constant temperature, while the PCBs were exposed to temperatures increasing in increments at high humidity and decreasing at low humidity, respectively. As one can note from the plots in Figure 7, the impact of the temperature increments in the presence of high humidity is not the same for all circuits, but when humidity is low the circuits are left virtually unaffected.

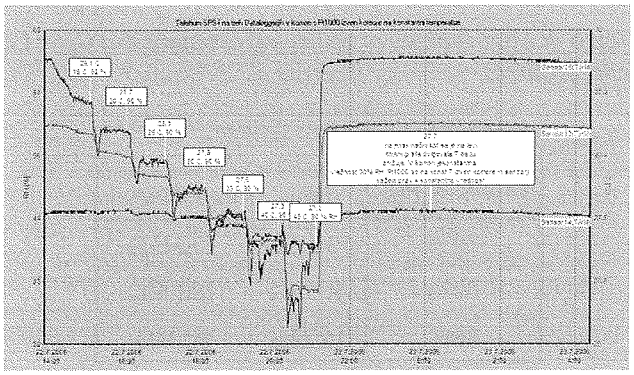


Fig. 7: Measurement results of three sensors with naked PPS capacitors: at high humidity RH = 90% the chamber temperature was increased in 5°C increments from 15°C to 45°C, then the air was dried to RH = 30% and the temperature was decreased to 15°C, again in decrements of 5°C. Each step of this temperature reduction lasted 1hour.

Next, the integrating PPS chip capacitors of the three tested samples were replaced by encapsulated PET capacitors with wire terminals and then the same experiment was repeated. The plots that are shown in Figure 8 reliably indicate that certain PPS chip capacitors were affected by high humidity and not the PCB itself. The results registered by the third sensor are meaningless since a fault occurred during the replacement of the capacitor. The steps in the upper and middle plot are a consequence of the temperature coefficients of each reference resistor R_{REF}, because the sensors were kept at different, i.e., constant temperatures during the test.

4.2 The influence of humidity on insulation resistance

Obviously, some of the used PPS capacitors were influenced by moisture. The influence of absorbed moisture on capacitance was excluded by empirical immersion tests, so only the surface resistance between capacitor terminals R_P remains as the possible cause of inaccurate conversion, if this resistance is decreased due to air humidity.

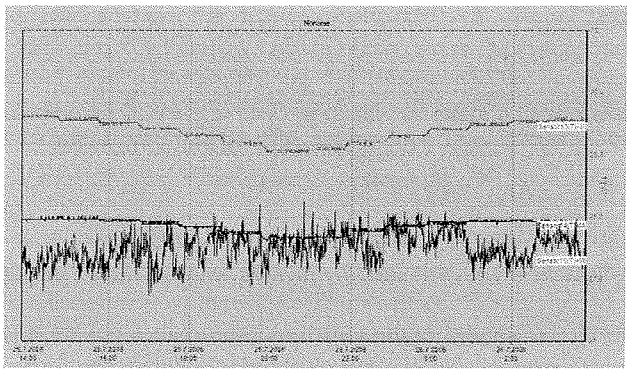


Fig. 8: Measurement results of three sensors with encapsulated PET capacitors: at RH=90% the chamber temperature was increased in increments of 5°C from 15°C to 45°C (the left half of the diagram), then the air was dried to RH=30% and the temperature was decrementally reduced down to 15°C.

Firstly, we have to estimate the order of magnitude of such a decrease of resistance that could cause the observed inaccuracies. As is shown in Figure 9, the parallel resistance represents a leak for the charge stored in the integrating capacitor. The sensitivity can either be derived from the exact analytic solution, or a few simple approximations can be used. The latter alternative is presented as follows.

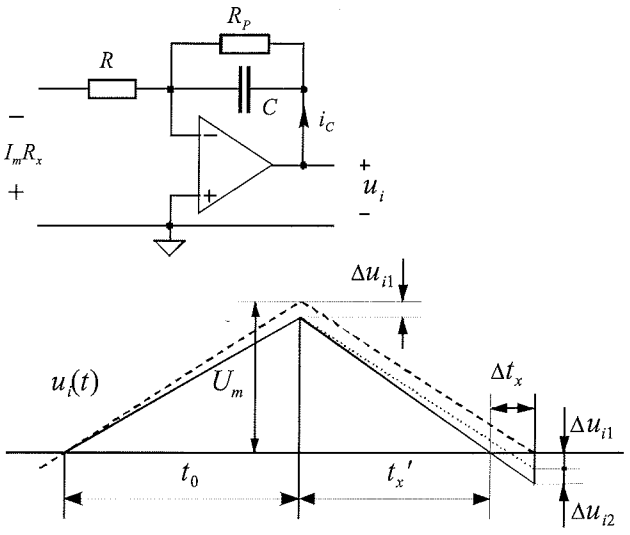


Fig. 9: Integrator with insulation resistance R_P (above), plot of integrator output voltage u_i with (——) and without R_P (- - -) (bellow)

At the end of the first step of conversion the peak integrator output is reduced by

$$\Delta u_{i1} = \frac{\Delta q}{C} = -\frac{1}{C}\int_{t}^{t+t_0} \frac{u_i(t)}{R_P}\,dt = -\frac{1}{2}\cdot\frac{U_m\, t_0}{R_P C} \qquad (7)$$

with Δq denoting the charge that leaks through R_P and U_m is the voltage that would be reached if there were no leak-

age, i.e., $R_P \rightarrow \infty$. The shape of $u_i(t)$ in is considered straight and $\Delta u_{i1} = U_m$ is neglected in the integral. During the second step of conversion, $u_i(t)$ decreases faster than in the ideal case (Figure 9) and after discharging for t_0 it would become negative. The relative error of the conversion

$$\frac{\Delta t_x}{t_x} = \frac{\Delta u_{i1} + \Delta u_{i2}}{U_m} = \frac{2\Delta u_{i1}}{U_m} = -\frac{t_0}{R_P C} \qquad (8)$$

where t_x is approximated by t_0, hence $\Delta u_{i2} = \Delta u_{i1}$. In our special case where the actual input voltage is proportional to the resistance of the platinum temperature sensor it is convenient to express the difference between the measured and the actual temperature

$$\Delta T = \frac{\Delta R_x}{R_x} \cdot \frac{1}{\alpha_R} = -\frac{t_0}{R_P C \cdot \alpha_R} \qquad (9)$$

where α_R denotes temperature coefficient of platinum. The plot of this temperature deviation for C = 100 nF, t_0 = 0.5 s and α_R = 3850 ppm/°C is shown in Figure 10.

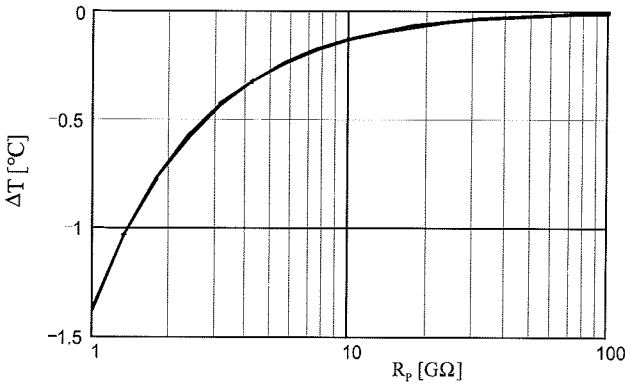


Fig. 10: Deviation of measured temperature vs. insulation resistance R_P

The derivation of eqn. is carried out only for the first phase of A/D conversion because both phases of the actually implemented four slope conversion are equally influenced by the charge leak. According to capacitor producer AVX /4/ the insulation resistance of a 100 nF capacitor is >60 GΩ for T < 75°C. Under normal conditions such values of R_P cannot engender noticeable deviations. Moreover, the constant resistance that appears between the capacitor terminals is taken into account during calibration, so that the conversion results in operation are influenced only if this resistance is decreased owing to environmental influences.

The theoretical plot in Figure 10 shows that the parallel resistance has to decrease to about 1GΩ when the result fails for approximately -1.5 °C. Such errors can be noted in the experimental results shown in Figure 7, where it is obvious that in the presence of damp air the parallel resistance of some capacitors is reduced to 50% or less by rising the temperature for 5°C.

4.3 Insulation resistance measurements

Both our theoretical and experimental findings have been verified by measurements of two sets of PPS film capacitors (100 nF/16 V), i.e., brand new capacitors and the desoldered ones from the circuits that turned out as bad. The new capacitors were immersed for 24h in a 20% isopropyl alcohol (IPA) water solution. The immersion caused no measurable differences in the capacitance and dissipation factors. The measured capacitance tolerance of all samples was les than 1%. Furthermore, the insulation resistance of the devices was measured at different air humidities. The leakage current at the applied voltage 1V was measured by a precise picoampermeter. The results are summarized by mean values in Table 4.

Table 4: Mean values of insulation resistance at U_{DC}=1V and T= 25°C,

Relative Humidity [%]	R_i	
	new	*desoldered*
35%	100 GΩ	100 GΩ
60%	40 GΩ	100 MΩ
90%	10 GΩ	30 MΩ

The results in Table 4 clearly indicate that the origin of the noticed inaccuracies lies in the resistance between the capacitor terminals. PPS capacitors have very low moisture absorption, but the naked types are affected by the side effects of PCB assembly, because the new devices perform much better when leakage is involved. The work of Hunt and Zou /12/ has shown the importance of appropriate selection of the flux in the soldering paste. The residues of soldering fluxes contribute to the surface conductivity as the humidity is increased. This effect is especially pronounced for weak organic acid (WOA) based fluxes at high humidity (> 85%). It has been already mentioned that the PCB's were protected against moisture but that these efforts turned out to be unsuccessful. The tested sensors were washed in deionized water, dried and protected with a thin polyurethane coating (Urethane 71) but the surface of some PPS chip capacitors remained contaminated by flux residues. The applied coating should be substantially thicker, which is a rather unpractical solution.

The weak point of the naked chips is the exposed lateral sides (Figure 7) on which the metal atoms that form the capacitor plates can be found. These lateral sides are additionally vulnerable due to the small gaps between the clusters of stacked dielectric film. It is almost impossible to clean these gaps once they get contaminated. The first prototypes were soldered by hand using ordinary soldering wire filled with resin. In this case, increasing the humidity does not reduce the surface resistivity /12/, therefore the high humidity effects were not noticed until a different technology of PCB assembly was used. It is important to note the fast response of the observed leakage

current to the changes in humidity, which obviously points to the fact that only the surface of the capacitor is involved in this process.

5. Conclusion

The choice of SMD film capacitors on the market has gone through significant changes that were initiated by the EC RoHS directive. New high temperature dielectric materials have been introduced in production. PET, PEN and PPS films are used for plastic film SMD capacitors. Special attention must be paid during the assembly of PET capacitors with regard to the reflow soldering process. PEN and PPS capacitors tolerate slightly higher temperatures in the reflow soldering process which in turn is beneficial for the reliability and quality of the leadless solder contacts. PEN capacitors should be avoided if dielectric absorption is important. PPS capacitors are now commonly available from various producers, but are the most expensive.

The construction of the capacitor should be carefully selected for each particular application. Stacked film capacitors have tight tolerances and require the least space on the PCB. As described in this paper, their naked construction is vulnerable to humidity, which reduces the parallel resistance due to surface conductivity. Wound capacitors are made by individually rolling the metallized film ribbons into cylindrical rolls which are then flattened to a prismatic shape. Wound capacitors are less sensitive to humidity, since the outer layers of the roll protect the capacitor core inside. Wound types are available naked or encapsulated in plastic boxes that provide additional protection against the environmental influences. Both variants of wound capacitors require more space on the PCB than the stacked one.

The described case study shows the importance of careful component selection from amongst the variety that is offered on the market. Besides choosing the right dielectric material, it is equally important to consider the construction of the capacitor. Of course, it is almost impossible to anticipate the behavior and interactions of real components that are exposed to harsh climatic conditions. Intensive computer-controlled testing of prototypes in a climatic chamber is an important step in the good design of demanding electronic products.

6. References

/1/ S. Tumański, Principles of electrical measurement, CRC Press, New York, 2006, ISBN 9780750310383.

/2/ WIMA Main Catalog 2001 Edition, accompanying notice,

/3/ http://www.ecicaps.com

/4/ http://www.avx.com, AVX Film chip capacitors, Product's master catalog version 7.1,

/5/ B. Walgenwitz, J-H. Tortai, N. Bonifaci, A. Denat, Self healing of metallized polymer films of different nature, *Proc. of 2004 IEEE Int. Conf. on Solid Dielectrics*, Toulouse, France, 5-9 July 2004, Vol.1, pp. 29- 32, ISBN 0-7803-8348-6

/6/ http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0095:EN:HTML

/7/ K.Saarinen, M. Niskala, E. Matero, J. Perälä, SMD plastic film capacitors for high temperature applications, http://www.evoxrifa.com

/8/ A. Levstek, Pretvornik s štirikratno integracijo za mikrokrmilniški senzor temperature, Proc. of 16th ERK 2007, Portorož, Slovenija,September 24-26, 2007, Vol. A, pp. 42-45

/9/ http://www.wima.de/EN/absorption.htm

/10/ C. Iorga, Compartmental analysis of dielectric absorption in capacitors, *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 7, no. 2, Apr. 2000, pp. 187-192

/11/ K. Kundert, Modeling dielectric absorption in capacitors, *The designer's guide* 2006, http://www.designers-guide.org/Modeling/

/12/ C. Hunt, L. Zou, The impact of temperature and humidity conditions on surface insulation resistance values for various fluxes, *Soldering and surface mount technology,* Vol. 11, No. 1, 1999 MCB University Press, pp 36-43

Dr. Andrej Levstek, univ. dipl.ing.el.
Asst. Matija Pirc, univ.dipl. ing. el.
University of Ljubljana
Faculty of Electrical Engineering
Laboratory of Microsensor Structures and Electronics
Tržaška cesta 25, SI-1000 Ljubljana, Slovenia
E-mail: andrej.levstek@fe.uni-lj.si

Prispelo (Arrived): 07.03.2008 Sprejeto (Accepted): 19.03.2009

ELECTROMAGNETIC COMPATIBILITY IN INTEGRATED CIRCUITS: A REVIEW

Jure Koselj*, Drago Strle**

*Kolektor Nanotesla Institut, 1521 Ljubljana
**University of Ljubljana, Faculty of Electrical Engineering

Key words: electromagnetic compatibility, electromagnetic interference, measurement methods, measurement setup.

Abstract: Electromagnetic compatibility (EMC) studies the unintentional generation, propagation and reception of electromagnetic energy with reference to the unwanted effects. Its goal is to avoid any interference between different equipment, which uses electromagnetic phenomena, in the same electromagnetic environment. This paper is focused on EMC of integrated circuits (ICs) and presents a review of EMC history, IEC standards for EMC in ICs, measurement methods and measurement setups.

Elektromagnetna združljivost v integriranih vezjih: pregled

Kjučne besede: elektromagnetna združljivost, elektromagnetna interferenca, merilne metode, merilni sistemi.

Izvleček: Elektromagnetna združljivost preučuje nenamerno generiranje, prenašanje in dovzetnost elektromagnetne energije v odnosu z neželenimi efekti. Njen cilj je izogniti se interferencam med različnimi napravami, ki delujejo na elektromagnetnem pojavu, v skupnem elektromagnetnem okolju. Članek je osredotočen na elektromagnetno združljivost v integriranih vezjih in obravnava pregled zgodovine elektromagnetne združljivosti, IEC standarde za elektromagnetno združljivost v integriranih vezjih, merilne metode in merilne sisteme.

1. Introduction

Electromagnetic compatibility is the branch of electrical sciences which studies the unintentional generation, propagation and reception of electromagnetic energy with reference to the unwanted effects (Electromagnetic Interference or EMI) that such energy may induce. The goal of EMC is the correct operation, in the same electromagnetic environment, of different equipment which uses electromagnetic phenomena, and the avoidance of any interference effects.

In order to achieve this, EMC pursues two different kinds of issues. Emission issues are related to the unwanted generation of electromagnetic energy, and to the countermeasures which should be taken in order to reduce such generation and to avoid the escape of any remaining energies into the external environment. Susceptibility or immunity issues, in contrast, refer to the correct operation of electrical equipment in the presence of unplanned electromagnetic disturbances.

Electrostatic discharge (ESD) and latchup effect are also connected with EMC phenomena. *Latchup* may be defined as the creation of a low-impedance path between power supply rails as a result of triggering a parasitic device. In this condition, excessive current flow is possible, and a potentially destructive situation exists. After even a very short period of time in this condition, the device in which it occurs can be destroyed or weakened; and potential damage can occur to other components in the system. Latchup may be caused by a number of triggering factors including overvoltage spikes or transients, exceeding maximum ratings, and incorrect power sequencing.

ESD is the sudden and momentary electric current that flows between two objects at different electrical potentials and is a serious issue in solid state electronics. Integrated circuits are made from semiconductor materials such as silicon and insulating materials such as silicon dioxide. Either of these materials can suffer permanent damage when subjected to high voltages, as a result there are a number of antistatic devices that help prevent static build up.

Interference, or noise, mitigation and hence electromagnetic compatibility is achieved by addressing both emission and susceptibility issues, i.e., quieting the sources of interference, making the coupling path between source and victim less efficient, and making the potential victim systems less vulnerable.

2. A brief historical review

It started more than forty year ago, when in 1965, Gordon Moore co-founder of Intel Corporation, published his long term vision of the evolution of integrated circuit. He point out a trend in integrated circuit complexity and predicted an exponential growth in the available memory and calculation speed of microprocessors, doubling every year /1/. With minor correction (doubling every 18 months), Moore's law still holds today.

The American army was a pioneer in the field of integrated circuit EMC. In 1965 they studied the effects of the electromagnetic fields due to nuclear explosions on electronic devices used in launch missile sites. The result was simulating software SCEPTRE /2/ that was developed at IBM, for simulating the effect of nuclear radiation on electronic

components and it was used to correlate simulations and experimental measurements.

One of the earliest academic publication was the simulation of the 741 integrated operational amplifier published by Wooley in 1971 /3/. The author simulated the different stages of this IC with the simulating software CANCER (an ancestor of today's simulator SPICE).

In 1975, Whalen published studies on the radio-frequency pulse susceptibility of discrete transistors /4/. He also edited special issue about interference between electromagnetic sources in the very high frequency band (VHF 30 MHz-300 MHz), ultra high frequency band (UHF 300 MHz-3 GHz) and even extremely high frequencies with radars (EHF 3 GHz-30 GHz) /5/. Two years later in 1981 Roach characterized the sensitivity of 1 Kbyte NMOS memories /6/. Some years later, a study by Tront was published /7/. The topic of the study was about the behavior of the 8085 processor in presence of 100 and 220 MHz radio-frequency interference. By using simulation software SPICE, he reproduced some phenomena observed during measurements.

In 1990 H. Bakoglu published a book /8/ witch comprehend a summary of the parasitic effects in integrated circuits, packaging and printed circuit boards (PCB's). His book describes different problems linked to transient current consumption at active edges of the clock and detailed basic mechanisms for integrated circuit resonance. In the same year Kenneally et al. published measurement results for simple integrated circuit in CMOS and TTL technologies /9/. His results showed that the sensitivity of circuit decrease when interfered frequency (1-200 MHz) increases. He also discovered significant differences between fabrication technologies. CMOS circuits tend to be less robust than TTL circuits.

Synchronous switching noise is most significant chip-level concerns for EMC and signal integrity engineers. In 1993, R. Downing et al. published a paper with a topic on the characterization of decoupling capacitance effects including on-chip decoupling and decoupling in proximity to the integrated circuit.

The research on susceptibility of integrated circuits continued in 1995 with published paper by Laurin et al. The research focused on the effects of an electromagnetic wave coupling to PCB traces and the consequences of this coupling on simple circuit behavior /10/. The researchers observed no perturbations on the component with field strengths of 200 V/m, which is specified field strength in MIL-STD-461B. By adding long metal wire that was a half-wavelength long at the interference frequency, the field was reduced to 2 V/m. This low field caused severe malfunctions due to erroneous switching. The research also comprehends the difference between a static and transient regime. In the static regime, only perturbations with high en-

Table 1: Standards for EMC emission, susceptibility and impulse immunity measurement of integrated circuit

Standard	Description	Status in 2008
IEC 62215 Measurement of impulse immunity up to 1 GHz		
IEC 62215 1	Definitions	
IEC 62215 2	Synchronous transient injection method	Published in 2007
IEC 62215 3	Electrical fast transient (EFT), ESD immunity	New proposal
IEC 62132 Measurement of electromagnetic immunity up to 1 GHz		
IEC 62132 1	Definitions	Published in 2006
IEC 62132 2	TEM/GTEM Cell method	Committee draft for voting
IEC 62132 3	Bulk current injection (BCI)	Published in 2007
IEC 62132-4	Direct RF power injection (DPI)	Published in 2006
IEC 62132-5	Workbench Faraday cage	Published in 2005
IEC 61967 Measurement of electromagnetic emission up to 1 GHz		
IEC 61967 1	Definitions	Published in 2002
IEC 61967 2	TEM/GTEM Cell method	Published in 2005
IEC 61967 3	Surface scan method	Published in 2005
IEC 61967-4-ed 1.1	1 Ω/150 Ω direct coupling method	Published in 2006
IEC 61967 4 am1	1 Ω/150 Ω direct coupling method	Published in 2006
IEC 61967-4-1	1 Ω/150 Ω direct coupling method	Published in 2005
IEC 61967-5	Workbench Faraday cage	Published in 2003
IEC 61967 6	Magnetic probe method	Published in 2002
IEC 61967 6 am1	Magnetic probe method	Published in 2008

ergy affected logic levels were in the transient regime, even weak perturbations could affect switching delays.

Two years later J. F. Chappel investigated and proposed a new technique that raised the immunity level of ICs from less than 1,5 V to over 5 V in the frequency range 1 to 10 MHz. Other circuits have been proposed because of their high immunity to radio frequency interference (RFI) including Schmidt triggers, low-voltage differential swing circuits and delay-insensitive structures.

In year 2000, NASA published an updated version of the handbook on EMI immunity /11/, which gave valuable information on the immunity of integrated circuits with frequency up to 10 GHz. The author presents measurement results concerning simple components and comparison with measurements taken in the 70's.

Fiori published a study of RFI effects on analog amplifier up to 2 GHz /12/. The measurement equipment involved microwave probes, positioned directly on chip, to maintain 50-ohm impedance from the measurement equipment to the integrated circuit. The author observed that DC shift of the amplifier offset increases with RFI amplitude but it remained almost constant from 100 MHz to 2 GHz.

3. EMC measurement methods

3.1. EMC Standards

A set of EMC regulations which fixed the maximum limits for parasitic emission and immunity levels for electronic devices were established in Europe in 1996. The probable result of these regulations was revived interest of the researchers and engineers for this subject.

The International Electrotechnical Commission (IEC) is one of the standard bodies that are addressing the need for standardized IC EMC measurement methods. At the component level, the most important standards were developed under the supervision of the Technical Committee 47A, which is focused on integrated circuits.

This committee prepares international standards with a focus specifically on logic circuits, memory, converters and hybrid modules.

3.2. Measurement of IC emission

The measurement of the conducted or radiated emissions from IC, under specified conditions, can give useful information about the potential and severity of emissions in a tested application. IEC 61967 standard defines measuring method, measurement conditions, test equipment, specific PCB requirements, consistent test procedures as well as contest of the test report.

3.3. Radiated emission

There are two basic methods for evaluating ICs of being the source of radiated EM emissions: TEM/GTEM cell test and near-magnetic field scan. The transverse electromagnetic mode (TEM) cell and the gigahertz TEM (GTEM) cell are commonly used for measuring EM emissions radiated by an IC, but can be also used for measuring IC immunity to EM fields. EM radiated emission measurement method by TEM/GTEM is standardized as IEC 62967-2 (Table 1). The TEM cell (Fig. 1) is an expanded rectangular waveguide with an inner conductor called the septum, whose characteristic impedance is set to 50 Ω. It is terminated by two tapered ends to connect 50 Ω - adapted coaxial cables.

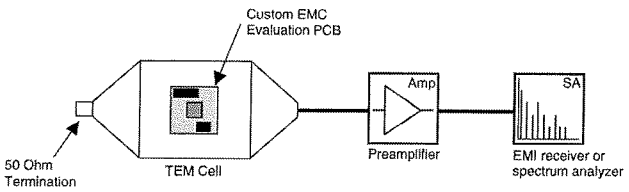


Fig. 1: IEC 62967-2 TEM cell radiated emissions test setup /13/

The GTEM cell (Fig. 2) was designed to overcome the TEM cell frequency limitation. It consists of a tapered section of rectangular transmission line, with an offset and sloping septum plate. The septum is tapered so that the characteristic impedance is maintained to 50 Ω along the length of the cell. The wide extremity is terminated by a distributed resistive load that operates as a 50 Ω load circuit at low frequency and by pyramidal foam absorbers that attenuate the EM wave at high frequency.

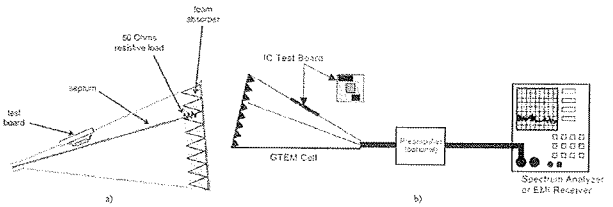


Fig. 2: a) GTEM cell, b) IEC 62967-2 GTEM cell radiated emission test setup /14/

Measuring the radiation from the same IC in TEM or GTEM cells can be compared using the correlation factor between two cells /14/.

The near-field scanner was introduced to the EMC problem of the IC by K. Slattery /15/. Measurement methods which determine the EM field by using RF receiver can be classified in two main techniques. The direct technique involves a coaxial cable who connects the probe to the receiver (Fig. 3).

The second technique creates a perturbation by introducing a scatterer at the desired observation point, to enhance the spatial resolution and to reduce the parasitic coupling between the probe and device under test (DUT).

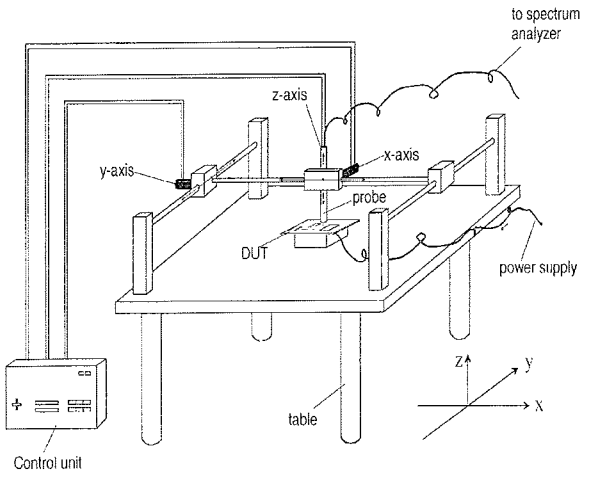


Fig. 3: IEC 61967-3 near-field radiated emission test setup /14/

3.4. Conducted emission

Another method for evaluating ICs is to measure the conducted noise currents on each pin.

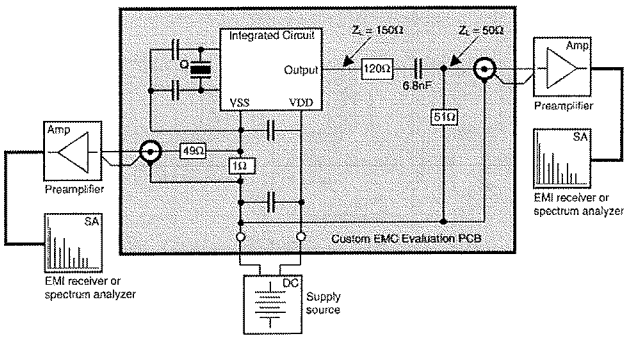


Fig. 4: IEC 61967-4 conducted emission test setup /14/

IEC 61967-4 defines a method of determining an IC's conducted electromagnetic emissions by measuring the RF voltage developed across a standardized load as shown in Fig. 4. For high power ICs, the 1 Ω probe is replaced by a 0.1 Ω resistance, or split among the return current paths /14/.

The Workbench Faraday Cage (WBFC) method is used for conducted emission measurements. IEC 61967-5 defines a method of measuring the conducted electromagnetic emissions at defined common-mode points in order to estimate the emissions radiated by connected cables in an application /16/. A schematic of the test setup are shown in Fig. 5.

The method will apply for those cases where the wires and cables connected to the sources and victims are much longer than the largest dimension of the integrated circuit.

IEC 61967-6 is defines a method of calculating the conducted electromagnetic emissions from an IC pin by using a magnetic field probe to measure the magnetic field as-

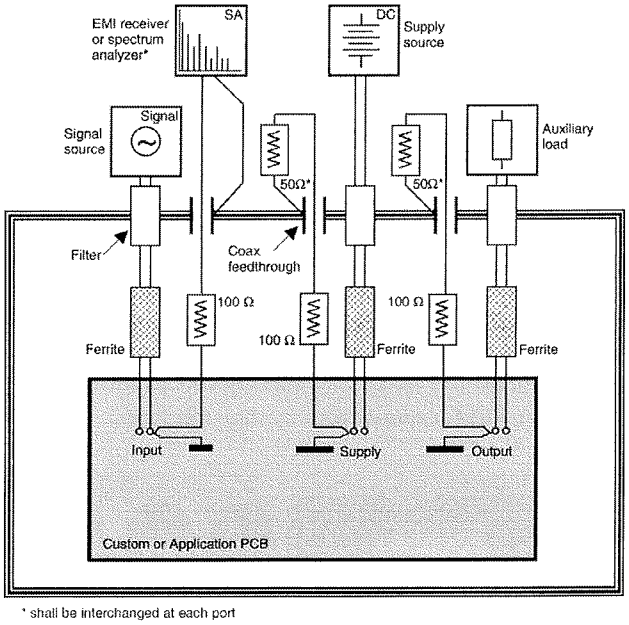


Fig. 5: IEC 61967-5 WBFC conducted emission test setup /13/

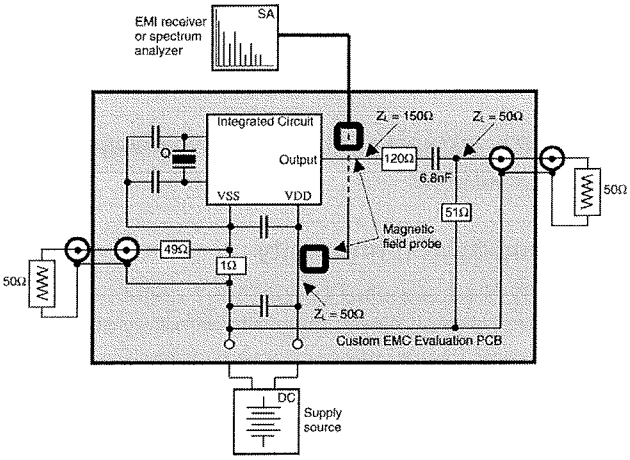


Fig. 6: IEC 61967-6 conducted emission test setup /13/

sociated with a connected PCB trace /16/. A diagram of the test setup is shown in Fig. 6. The preferred test configuration is with the DUT mounted on an IC EMC test board with standardized layout patterns to maximize repeatability and minimize probe coupling to other circuits. With care, this method can also be applied to application boards.

3.5. Measurement of IC immunity/ susceptibility

As for emissions, the assessment the conducted or radiated immunity performance of an IC, under controlled conditions, can yield useful information about the potential and severity of RF immunity in an application. The test methods in IEC 62132 standard utilize a continuous wave signal, either modulated or unmodulated. The required mod-

ulation, if any, is specified in the individual test method. This standard consists of five parts: a general guidance document and four immunity test methods /17/.

3.6. Radiated immunity

The TEM/GTEM cell can be used also as immunity tests setup. Test setup is similar to emissions setup only that EMI analyzer and pre-amplifier are exchanged with signal generator and power amplifier. DUT is tested in at least two orientations to ensure complete exposure to the generated electric field. Signal generator generates signal, which is amplified to achieve strength of electric field in the TEM/GTEM cell specified by the standard.

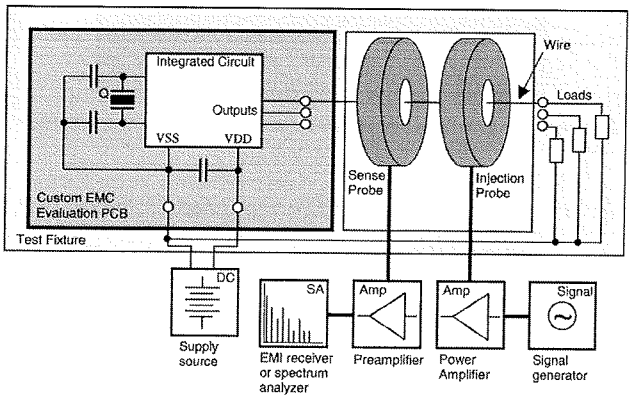


Fig. 7: IEC 62132-3 BCI conducted immunity test setup /13/

3.7. Conducted immunity

IEC 62132-3 is a standard that defines a method of evaluating the conducted electromagnetic immunity of an IC to noise currents injected on connected cables (Fig. 7). This method is based on bulk current injection (BCI) methods used for equipments and systems such as ISO 11452-4. The intent of this method is to evaluate the immunity of IC pins connected to cables. The injection coil induces by magnetic coupling a current (usually up to 100 mA) that may produce faults on the DUT at particular frequencies. The faulty state is recognized by a real-time monitoring of the device activity.

Direct power injection (DPI) method (Fig. 8) evaluates the immunity of IC pins connected to cables. The immunity signal is directly injected (capacitively coupled) onto the conducted conductor (cable, trace, etc.) instead of being inductively coupled to the cable or cable bundle. Signals issued from the DUT are analyzed in real-time by special oscilloscopes that alter the control software in case of abnormal waveform.

Other susceptibility measurement method for evaluating the conducted immunity of an IC to noise signals injected at defined common-mode points in order to simulate exposure to any connected cables to the radiated electromagnetic environment in an application is workbench faraday cage method (IEC 62132-5) (Fig. 9). The DUT is mounted on either IC EMC test board or an application board subject to the size limitations of the WBFC. With all input, output and power connections to the test board filtered and connected to common-mode chokes, the conducted noise is injected at PCB locations specified by the standard.

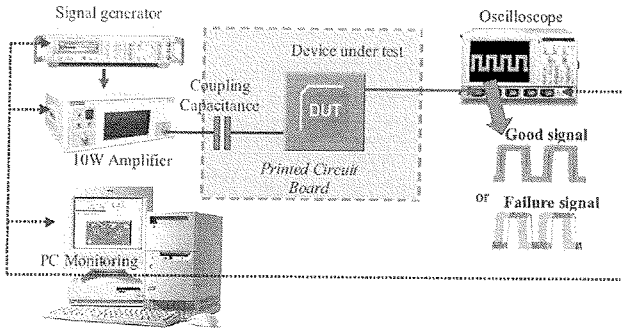


Fig. 8: IEC 62132-4 DPI conducted immunity test setup /18/

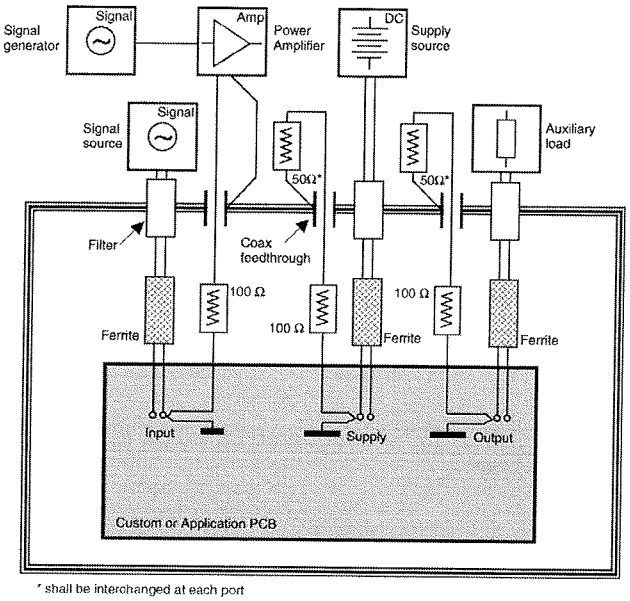


Fig. 9: IEC 62132-5 WBFC conducted immunity test setup /13/

3.8. Impulse immunity

A reason for investigating the IC immunity to impulse waveforms is that the impulse immunity tends to decrease with newer technology, as observed by researchers /19/. Reasons might be the decreased noise margins, the clock frequency increase or the increase of IC complexity /18/.

A new standard (IEC 62215) is under discussion to establish measurement techniques for measuring impulse immunity of ICs. Most of the documents are at the status of new proposals, as shown in Table 1 /20/.

4. Conclusions

This paper has presented some important aspects of EMC in ICs and its history. While the EMC evaluation of ICs is still in development, international standardized methods (IEC 61967, IEC 62132 and IEC 62215) are available for use. A comprehensive set of standardized IC EMC test methods with application has been presented. As it is shown in table 1, standards for radiated/conducted emissions and immunity has been developed and published by IEC in recently. In the future, focus will be on transient immunity evaluation, which is still in its infancy.

5. References

/1/ Cramming more components onto integrated circuits, G. E. Moore, Electronics, Vol. 38 (8), 1965

/2/ Automated Digital Computer Program for Determining Responses of Electronic Circuits to Transient Nuclear Radiation (SCEPTRE), S. R. Sidore, AFWL TR 66-126, Air Force Weapons Laboratory, 1967

/3/ A computer-aided evaluation of the 741 amplifier, B. A. Wooley, IEEE Journal of Solid-State Circuits, Vol. 6 (6), pp 357-366, 1971

/4/ The RF pulse susceptibility of UHF transistors, J. J. Whalen, IEEE trans. Electromagnetic compatibility, Vol. 17, pp. 220-225, 1975

/5/ Predicting RFI effects in semiconductor device at frequencies above 100 MHz, J. J. Whalen, IEEE Trans. Electromagnetic Compatibility, Vol. 21, pp. 283-290, 1979

/6/ The susceptibility of 1 K NMOS memory to conducted electromagnetic interference, J. N. Roach, Rec. 1981 Nat. Symp. Electromag. Compat. Symp., Boulder, CO, pp. 85-90, 1981

/7/ Predicting RFI upset of MOSFET digital IC's, J. G. Tront, IEEE Trans. Electromagnetic Compatibility, Vol. 27, pp. 64-69, 1985

/8/ Circuit, interconnections and packaging for VLSI, H. Bakoglu, Reading, MA: Addison-Wesley, ISBN 0-201-06008-6, 1990

/9/ RF upset susceptibilities of CMOS and low power Schottky 4-bit magnitude comparators, D. J. Kenneally et al., IEEE International Symposium on Electromagnetic Compatibility, pp. 671-677, 1990

/10/ Prediction of delays induced by in-band RFI in CMOS inverters, J. J. Laurin et al., IEEE Trans. Electromagn. Compat., Vol. 37 (2), pp. 167-174, 1995

/11/ Integrated Circuit Electromagnetic Immunity Handbook, J. G. Sketoe, NASA/CR-2000-210017, NASA Marshall Space Flight Center, AL 35812, pp. 64, 2000

/12/ A new nonlinear model of EMI-induced distortion phenomena in feedback CMOS operational amplifiers, F. Fiori, IEEE Transactions on electromagnetic compatibility, Vol. 44 (4), pp. 495-502, 2002

/13/ An overview of standards in electromagnetic compatibility for integrated circuits, R. M. Carlton, Microelectronics Journal, Vol. 35, pp. 487-495, 2004

/14/ Electromagnetic compatibility of Integrated circuits, S. Bendhia et al., Springer, 2005

/15/ Near-field measurements of VLSI devices, K. P. Slattery et al., IEEE Transaction on Electromagnetic Compatibility, Vol. 41 (4): 374-384, 1999

/16/ IEC 61967, Integrated Circuits, measurement of electromagnetic emissions, 150 kHz to 1 GHz, IEC standard; www.iec.ch

/17/ IEC 62132, Characterization of integrated circuits electromagnetic immunity, 150 kHz to 1 GHz, IEC standard; www.iec.ch

/18/ Issues in electromagnetic compatibility of integrated circuits: Emission and susceptibility, E. Sicard and J. M. Dienot, Microelectronics Reliability, Vol. 45, pp. 1277-1284, 2005

/19/ Predicting the breakdown behavior of microcontrollers under EMP/UWB impact using a statistical analysis, M. Camp et al., IEEE trans. on EMC, Vol 46 (3), pp. 368-379, 2004

/20/ IEC 62215, Integrated Circuits, measurement of impulse immunity, IEC standard; www.iec.ch

Jure Koselj
Kolektor Nanotesla Institut, 1521 Ljubljana

Drago Strle
University of Ljubljana, Faculty of Electrical Engineering

Prispelo (Arrived): 26.05.2008 *Sprejeto (Accepted): 19.03.2009*

OPTICAL ENCODER HEAD WITH IMPROVED LINEARITY

[1]Jernej Rozman, [2]Anton Pleteršek

[1]IDS d.o.o., Ljubljana, Slovenia
[2]University of Ljubljana, Ljubljana, Slovenia

Key words: encoder, optical scanner, Vernier scale, opto-sensors, interpolator.

Abstract: We present a high precision optical scanner, combining averaged optical sensors array with appropriate 10 um graduated scales on measurement - fix plate and Vernier sliced parallel scale on reading plate. The generated sine wave signals are at least 30dB less distorted by distributing and mismatching optical edges over a number of sine wave periods within a number of Vernier scaled periods. Reading plate which is positioned along optical array has a unit division smaller than those on a fixed scale - permit a far more precise positioned optically generated sine wave current. Position-like averaging of four generated signals was distributed over an opto-array and reading scale. Background of reduction as well as prototype is presented.

Good matching was found between mathematically analyzed optical scanner and measurement, where comparable measurements were performed on optical head having redistributed and fixed opto-edges.

Optični dekodirnik z izboljšano linearnostjo

Kjučne besede: dekodirnik, optični pretvornik, Vernierjeva skala, optični senzor, interpolator.

Izvleček: V članku predstavljamo optični dekodirnik z visoko točnostjo pretvorbe. Sestoji se iz polja optičnih senzorjev, namenjenih povprečenju signala ter iz fiksne in čitalne letve z 10 um rastrom in z Vernierjevo porazdelitvijo. Zgradba sistema omogoča periodično prerazporeditev optičnih robov znotraj periode sinusnega signala in znotraj Vernierjeve periode ter s tem zmanjšanje popačenja proizvedenih sinusnih signalov za 30dB. Čitalna letev ima raster zmanjšan v primerjavi z rastrom merilne letve kar posledično pomeni veliko boljše pozicioniranje in razpršen vpliv uklonskih pojavov. Princip delovanja in prototipni optični dekodirnik so opisani v članku.

Izdelana matematična analiza se dobro ujema z izsledki meritev. Poleg rezultatov meritev je podana primerjalna analiza sistemov z Vernierjevo porazdelitvijo - in sicer s konstantnim in s prerazporejenim rastrom.

1 Introduction

Accurate measurement of displacement is of prime importance in any computer controlled machine (CNC). The need to manufacture "something" requires the ability to move "something" with a very high level of accuracy. And for this accurate position sensors are needed, also called position encoders /4/.

Such encoders can be roughly split by the type of displacement they detect (rotary, linear) or by the quantity they detect (optical, magnetic) or by the type of output data they produce (incremental, absolute).

This work deals with linear optical incremental position encoders.

As this work deals with a very narrow field of research not much literature is available. Most information can be found in books about automatics and robotics /1, 2/. Our solution is also much differ from that in patents /10, 11/.

2 Optical encoder

An optical encoder (Figure 1) is typically composed of a light source (LED), the main scale with a built-in optical grating with period PM and the optical scanner that is composed of an optical sensor and a reading scale with a built-in optical grating with period PR /3/. The main and reading scales are glass plates on which a thin layer of metal (chromium Cr) is deposited and then a regular pattern is etched into this layer. If periods PM and PR are the same then as the scanner moves along the main scale the patterns on the main and reading scale overlap to a different degree, depending on the momentary position of the scanner. In short, as the scanner moves along the main scale the amount of light from the light source (LED) is modulated. Therefore the electrical signal produced by the photo sensor is also modulated. In our case the photo sensor is a reverse polarized photodiode and the signal is in the form of electrical current.

Real encoders use four such sensors placed apart ¼ of the grating period (PM, PR) so that they produce four signals that are 90° shifted between each other. Figure 2 shows the four ideal quadrature signals +A, -A, +B, -B that are produced by the four sensors. The two pairs of signals are amplified with a differential amplifier to remove the large DC components and suppress even harmonics to produce the signals A and B.

The signals in figure 2 are periodic; this corresponds to a movement of the scanner head with constant velocity along the main scale. One period of the signal corresponds to a movement of the scanner head equal to the grating period (PM, PR). The signals are not periodic in time but they are periodic in relation to the displacement along the main scale. The signals in figure 2 are also not pure sine/co-

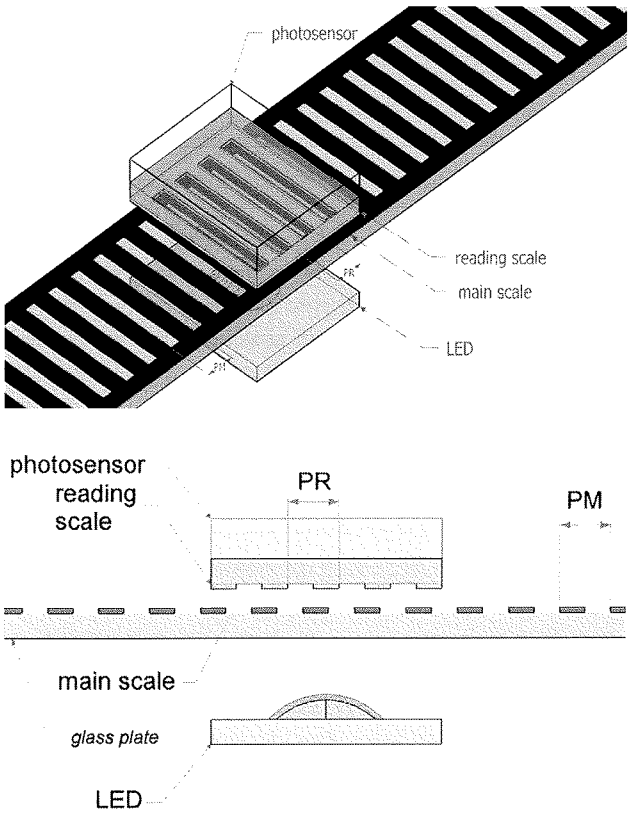


Fig. 1: Optical encoder.

sine in reality they are distorted and contain harmonic components.

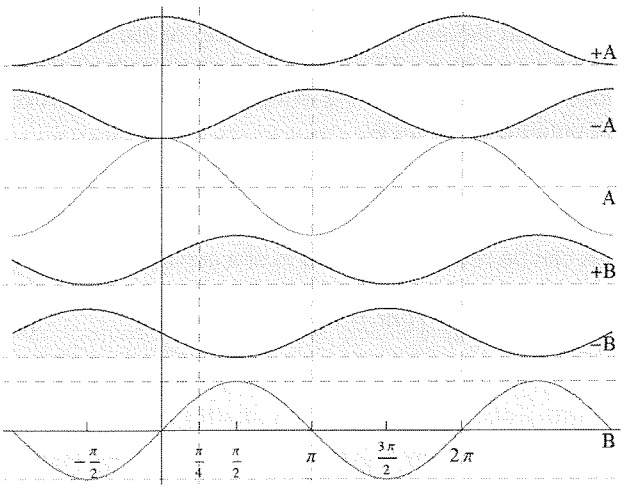


Fig. 2: Quadrature signals.

2.1 Interpolation

As we have seen the encoder produces two 90° shifted (quadrature) signals A and B as can be seen in figure 3. The two quadrature signals enable us to detect the position of the scanner head at all time just by measuring the values of signals.

The signals can be digitized directly so that the resolution of the encoder is determined by the grating period (PM, PR). Or interpolation can be used to increase resolution.

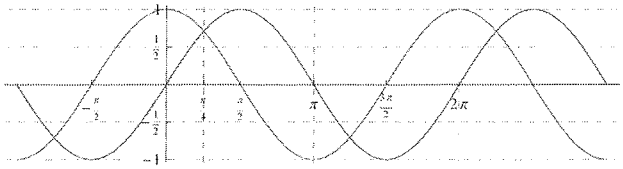


Fig. 3: Ideal signals.

If we have available two quadrature (sine/cosine) signals then we can by scaling and adding or subtracting of the two signals produce a new signal shifted by an arbitrary amount as is illustrated by the simple circuit in figure 4 and the two trigonometric formulas (1) and (2). We can produce a number of such shifted signals and make XOR operations on them (figure 5) to produce a quadrature signal with four times higher frequency which results in a four times higher encoder resolution.

$$\sin(x+\alpha) = \cos\alpha \sin x + \sin\alpha \cos x \quad (1)$$

$$\cos(x+\alpha) = -\sin\alpha \sin x + \cos\alpha \cos x \quad (2)$$

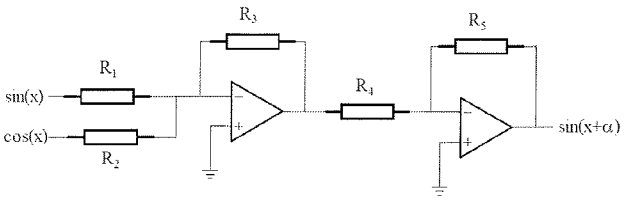


Fig. 4: Shifted signals.

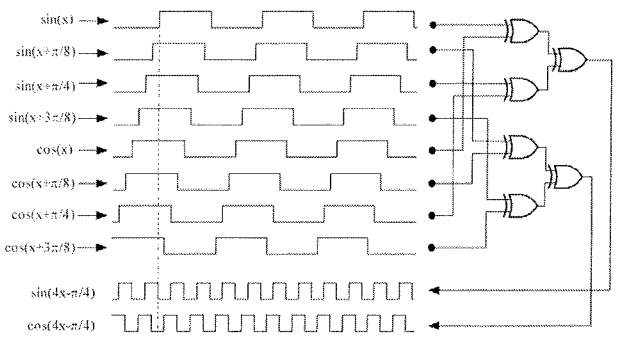


Fig. 5: Interpolation by factor 4.

One example of an integrated interpolation system is the IDS-EN400 /4/ which has a selectable interpolation factor 5, 10, 20, 25, 50, 100 and integrated signal conditioning circuitry.

The main emphasis of this work is the improvement of the quality of the quadrature signals to enable higher interpolation factors and higher encoder resolution. With the industry standard grating period (PM, PR) of 20µm, interpolation factors of 100 and more would enable resolutions in the nanometric range.

3 Improved encoder

The old scanning head (figure 6) was composed of four photodiodes (+A, -A, +B, -B) that produced the quardrature signals and an additional photodiode (RI) that generates the index signal which is used for absolute position encoding (for more information see /3/). Such a scanning head was basically described in Section 2. The spectral purity of such a scanning head is mainly determined by the diffraction and reflection of light by the two optical gratings in the main and reading scales. Without this the signals would have a triangular shape not sinusoidal.

We have used several techniques to improve the signal quality as described below.

3.1 Vernier (Nonius) pattern

In figure 7 we can see how two patterns with period P1=1 and P2=0.9 produce a third pattern which has a much larger period P=9 (9 is the least common multiple of P1 and P2). This new Vernier pattern corresponds with the movement of the scanning head along the main scale therefore we can place four photodiodes in one period of the Vernier pattern for generating the quadrature signals as depicted in figure 7.

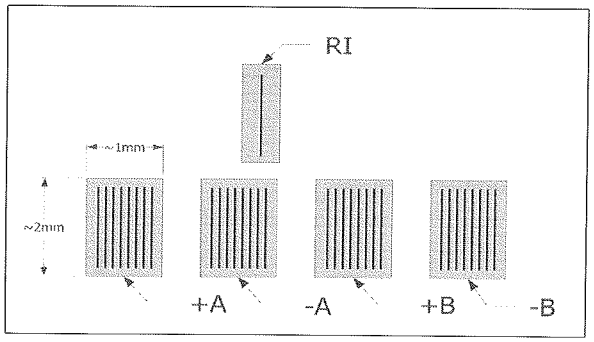
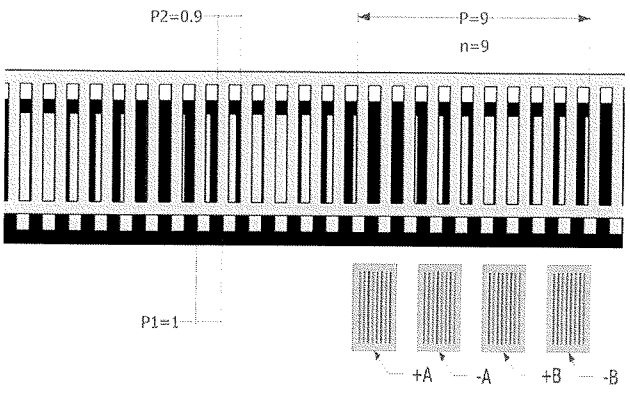


Fig. 6: Old scanning head.



Fig, 7: Vernier (Nonius) pattern.

The actual encoder uses a main scale with the grating period (*pitch*) of 20μm and we chose a reading scale period of 19μm to produce a Vernier pattern with a period of 380μm. Inside this 380μm we could place four photodiodes with maximum width of 95μm but the semiconductor technology chosen limits the width to 83μm. These dimensions are illustrated in figure 8. The two square shaped transmittances of the main and reading scales interact to create the Vernier pattern (380μm) which corresponds to the intensity of the light that falls on the photodiodes. As the scanning head moves 20μm the Vernier pattern moves 380μm. This movement of the pattern is detected by the four photodiodes placed 95μm apart to produce the 90° shifted signals. The "chopped" shaped of the Vernier pattern as seen in figure 8 is averaged (filtered) by the width of the photodiodes.

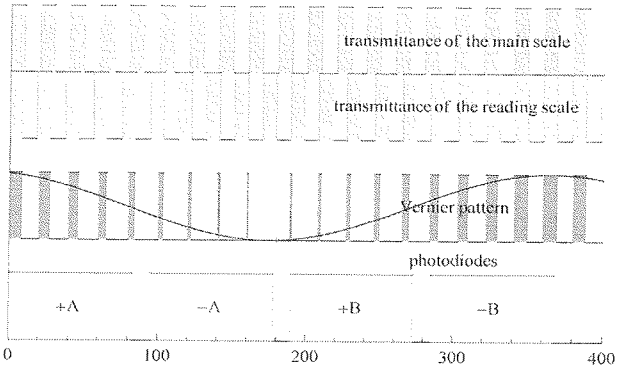


Fig. 8: Scanning cell.

This filtering can be simply described if we assume that each narrow strip of the diode produces the same small current i_n the only difference being that, as the pattern moves across the photodiodes, the currents i_n are shifted/delayed depending on the position of the strips that are producing them. Therefore we can find the total photodiode current i_D by summing all these small currents i_n over the photodiode width W,

$$i_D(x)=\frac{1}{W}\int_0^W i_n(x-\tau)d\tau \qquad (3)$$

This can be rewritten in the form of a convolution integral,

$$i_D(x)=\frac{1}{W}\int_{-\infty}^{\infty}(u(\tau)-u(\tau-W))i_n(x-\tau)d\tau \qquad (4)$$

where $u(\tau)$ is the unit step function. The impulse response of the linear system which is defined by (4) is $h(x)=\frac{1}{W}(u(\tau)-u(\tau-W))$. The frequency response is obtained by applying the Fourier transformation,

$$H(j\omega)=\int_{-\infty}^{\infty}h(x)e^{-j\omega x}dx=e^{-j\frac{W\omega}{2}}\left(\frac{\sin\left(\frac{W\omega}{2}\right)}{\frac{W\omega}{2}}\right)$$

and the amplitude response,

$$|H(j\omega)|=\frac{\sin\left(\frac{W\omega}{2}\right)}{\frac{W\omega}{2}} \qquad (5)$$

This is the typical $\frac{\sin x}{x}$ response which has periodic zeroes and these zeroes can be moved to a suitable place with the right selection of the photodiode width W. Please note that here x is displacement not time and $\omega = {}^{2\pi}/_{x}$.

Figure 9 shows an example where $W = {}^{380\mu m}/_{3} = 126.66\mu m$ therefore the first zero of the amplitude response is at the third harmonic. This arrangement would remove the third, sixth, ninth ... harmonic component. The disadvantage is that not all four photodiodes would fit in one Vernier period (380µm). Therefore we did not choose this option. Although the chosen $W = 83\mu m$ doesn't completely remove any higher harmonics it still suppresses them to some extent.

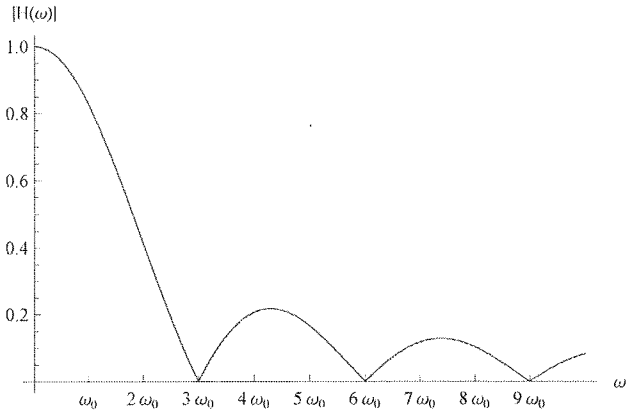


Fig. 9: Amplitude response ($W = \frac{1}{3}$ period).

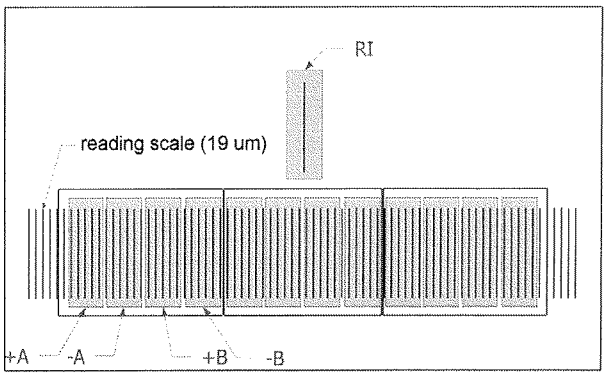


Fig. 10: Vernier distributed scanning head.

As shown in figure 8 four photodiodes were placed in the 380µm period so that a scanning cell was created. The whole scanning head contains 16 such cells (figure 10 shows only 3 cells). This was done to average out various unwanted effects like defects in the main and reading scale patterns or uneven illumination of the scanning head...

3.2 Shifted scanning cells

We could see in section 3.1 that by adding together shifted signals a favorable filter function was created therefore we tried this approach by shifting the scanning cells (figure 11) in the scanning head to remove the third harmonic component.

Let's write the combination of shifted signals as

$$g(x) = f(x) + f(x-a) + f(x-2a)..$$

Then by applying the Fourier transformation,

$$F[g(x)] = F[f(x)] + F[f(x-a)] + F[f(x-2a)]...$$

$$G(j\omega) = F(j\omega) + F(j\omega)e^{-ja\omega} + F(j\omega)e^{-j2a\omega}...$$

$$H(j\omega) = \frac{G(j\omega)}{F(j\omega)} = 1 + e^{-ja\omega} + e^{-j2a\omega}...$$

We get the frequency response $H(j\omega)$ of a system that adds signals that is shifted by x = a.

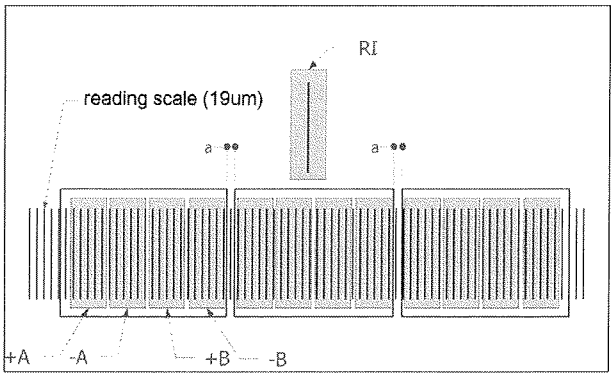


Fig. 11: Scanning head with shifted cells.

Our scanning head has 14 scanning cells and we wanted to remove the third component therefore $14a = {}^{380\mu m}/_{3}$, so we have to shift each cell by about 9 µm. The resulting filter function is shown in figure 12.

3.3 Rotation of the scanning head

We have also investigated the influence of scanning head rotation (figure 13) on the generated signals and discovered that this has a similar effect of shifting the signals as shown in equations (3)-(5) and a $\frac{\sin x}{x}$ filter function is created.

If for example the scanning head is rotated so that the upper edge of the scanning head is 6.66 µm (1/3 of the main scale period) left or right from the lower edge then the third harmonic component is removed similar as in figure 9.

4 Results

For comparison the harmonic components of the old scanning head (figure 6) were first measured (figure 14). As expected the largest is the third harmonic which is only 27dB smaller than the fundamental component.

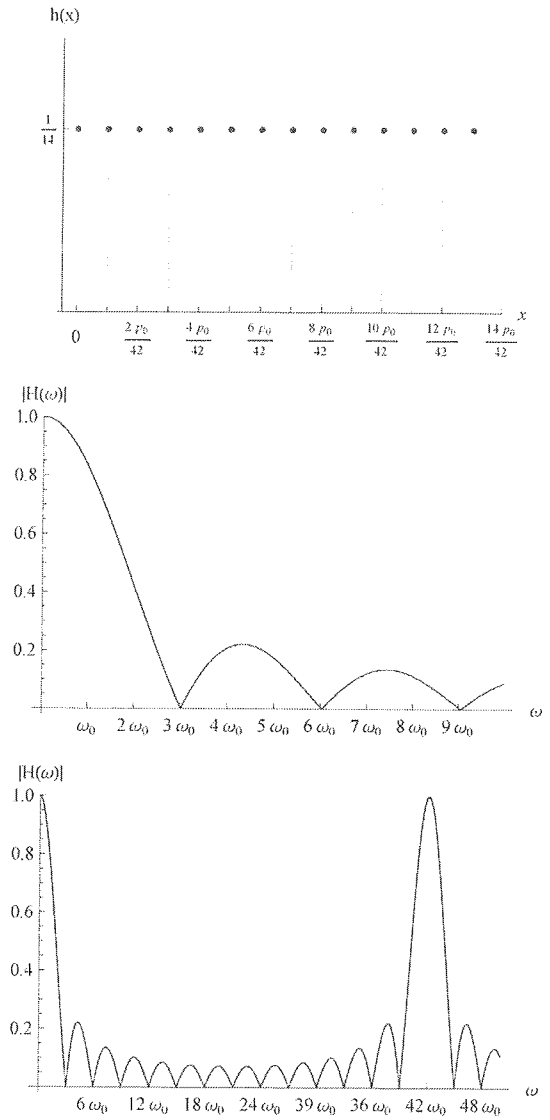


Fig. 12: Amplitude response (shifted cells).

Figures 15 and 16 show harmonic components of the distributed Vernier scanning head (figure 10) and the scanning head with shifted cells (figure 11). The Vernier distributed scanning head has the second and third harmonic 52dB below the fundamental component. For the scanning head with shifted cells only the second component can be seen 56db the third component is suppressed as predicted.

Figures 17 and 18 also show harmonic components of the two improved scanning heads. Here rotation was used to further suppress the second and third component. For the scanning head with shifted cells (figure 18) all harmonic components are below 60db.

In figures 15-18 the black line shows the FFT of the differential signal and the grey line shows the FFT of the single ended signal.

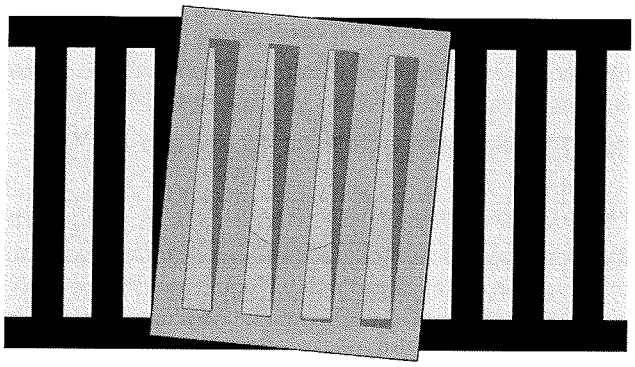


Fig. 13: Scanning head rotation.

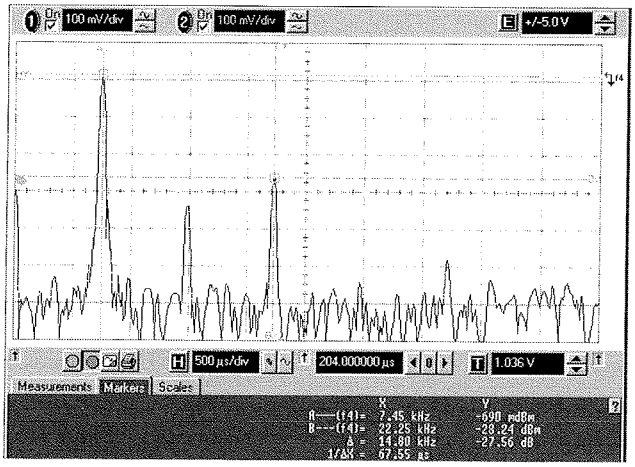


Fig. 14: Old scanning head.

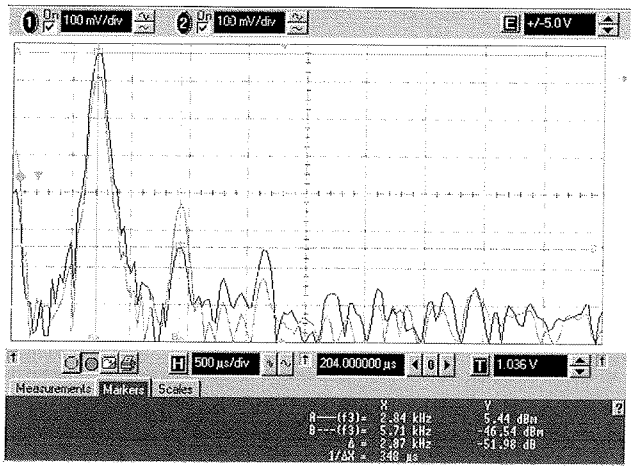


Fig. 15: Vernier distributed scanning head.

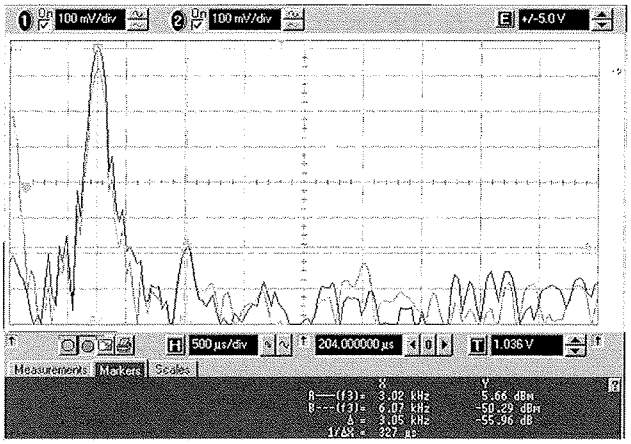


Fig. 16: Scanning head with shifted cells.

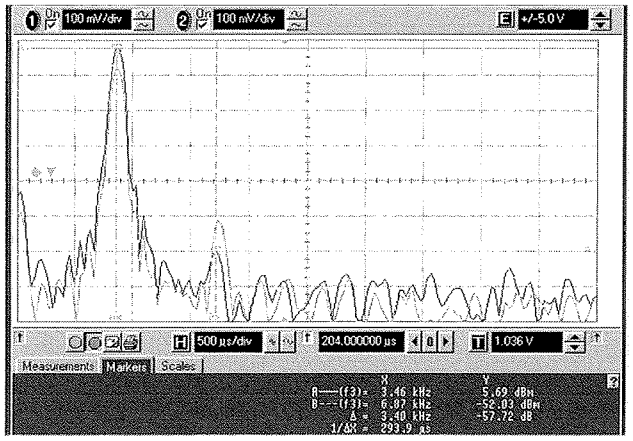


Fig. 17: Vernier distributed scanning head (rotated).

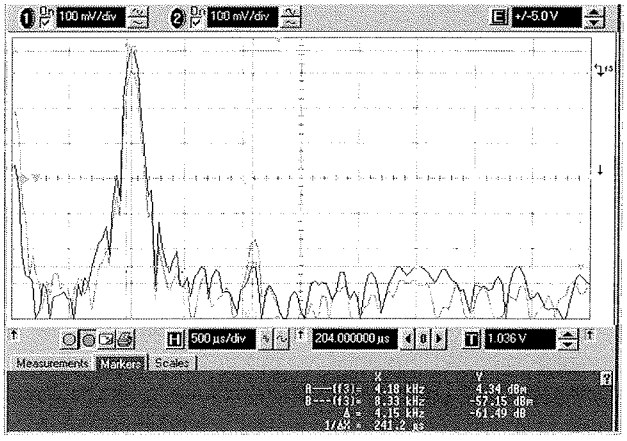


Fig. 18: Scanning head with shifted cells (rotated).

5 Conclusions

The aim of this work was to improve the scanning head of a linear encoder already used in the industry /5/. As shown in section 4 the improved scanning heads have the harmonic components suppressed by about 30db compared to the old scanning head. Furthermore with the introduction of multiple scanning cells into a single scanning head the influence of various unwanted effects were reduced like for example defects in the main and reading scale patterns or uneven illumination of the scanning head. Also by manufacturing all the photodiodes on a single chip diode matching was improved.

Acknowledgments

The authors would like to thank Miran Skrt (Iskra TELA) and Branko Cvetkovic (RLS) for their contributions to the project.

References

/1/ T. R. Kurfess: *Robotics and Automation Handbook*, CRC Press

/2/ Hyungsuck Cho: *Opto-Mechatronic Systems Handbook : Techniques and Applications*, CRC Press, 2003

/3/ J. Rozman: Optični merilni sistem dajalnika za merjenje linearnega pomika, Thessis, University of Ljubljana, Faculty of electrical engeneering, 2007

/4/ Sugiyama, Y. Matsui, Y. Toyoda, H. Mukozaka, N. Ihori, A. Abe, T. Takabe, M. Mizuno, S. ; "3.2 kHz, 14-Bit Optical Absolute Rotary Encoder With a CMOS Profile Sensor", Sensors Journal, IEEE, Aug. 2008, Volume: 8 , Issue: 8, Page(s): 1430 - 1436.

/5/ Cheung, N.C.; "An innovative method to increase the resolution of optical encoders in motion servo systems", Power Electronics and Drive Systems, 1999. PEDS '99. Proceedings of the IEEE 1999 International Conference on, Volume 2, 27-29 July 1999, Page(s):797 - 802.

/6/ Kok-Zuea Tang; Kok-Kiong Tan; Tong-Heng Lee; Chek-Sing Teo; "Neural network-based correction and interpolation of encoder signals for precision motion control", Advanced Motion Control, 2004, The 8th IEEE International Workshop on, 25-28 March 2004, Page(s):499 - 504.

/7/ Alhamadi, M.A.; Benammar, M.; Ben-brahim, L.; "Precise method for linearizing sine and cosine signals in resolvers and quadrature encoders applications", Industrial Electronics Society, 2004 - IECON 2004, 30th Annual Conference of IEEE, Volume 2, 2-6 Nov. 2004, Page(s):1935 - 1940.

/8/ www.ids.si: *IDS-EN400 Datasheet.*

/9/ www.iskra-tela.si: *Linerna scale TGM133 and TGM173 Datasheet.*

/10/ Thorburn William, G.; "Harmonic Suppressing Photodetector Array", Patent, International Publication Number WO 03/021197 A1.

/11/ Donald K. Mitchell, William G. Thorburn, "Optical encoder having slanted optical detector elements for harmonic suppression", US Patent Pub. Number US2007/0153292 A1.

/12/ William G. Thorburn, Bruce A. Horwitz, "Harmonic suppressing photodetector array", US Patent Pub. Number US2003/0047673 A1.

Jernej Rozman
IDS d.o.o. Sojerjeva 63, 1000 Ljubljana, Slovenia
E-mail: jernej.rozman@ids.si

Anton Pleteršek
University of Ljubljana, Tržaška 25,
1000 Ljubljana, Slovenia
E-mail: anton@kalvarija.fe.uni-lj.si,
anton.pletersek@ids.si

Prispelo (Arrived): 25.08.2008 Sprejeto (Accepted): 19.03.2009

IZKLOP TRIFAZNEGA KRATKOSTIČNEGA TOKA S TRIPOLNIM NIZKONAPETOSTNIM ODKLOPNIKOM

Martin Bizjak

Iskra MIS d. d., Kranj, Slovenija

Kjučne besede: nizkonapetostni odklopnik, enopolni odklopnik, tripolni odklopnik, kratkostični tok, izklop trifaznega kratkostičnega toka, tokovna omejitev izklopa, cepitev na delne obloke

Izvleček: Za izklop velikih tokov, ki nastanejo pri kratkih stikih v energetskih tokokrogih nizke izmenične napetosti, uporabljamo odklopnike. V njih izkoriščamo pojav tokovne omejitve, s katerim omejimo tako velikost toka, ki ga izklapljamo, kot tudi čas,v katerem doseže ničelno vrednost, kjer lahko tok prekinemo. V enopolnih odklopnikih s tokovno omejitvijo po nastanku kratkega stika napetost na priključkih odklopnika narašča najprej sorazmerno razmikanju kontaktnega para, potem pa hitreje, ko se oblok razširja v obliki zanke v deion komoro. Tam doseže napetost, ki je primerljiva z napetostjo vira, zato se tok hitro zmanjšuje proti ničli in se prekine. Časovni potek napetosti je kvantitativno odvisen od faznega kota nastanka kratkostičnega toka in pogojev v tokokrogu, vendar poteka kvalitativno po funkciji časa, ki je značilna za izklop izmeničnega toka z odklopnikom ali z varovalko na taljivi vložek.

Med izklopom trifaznega izmeničnega toka s tripolnim odklopnikom pa se potek napetosti v kakem od polov lahko bistveno razlikuje od tistega, ki je značilen za enopolni izklopni pojav. Vzrok je v medsebojnem vplivu tokov na sosednjih polov, zaradi česar gibanje obloka povzroča rezultanta sil, ki ga ne usmerja proti deion komori. Zato je kljub porazdelitvi tokov vseh treh faz enakomerno po celi periodi s faznim zamikom po 120° izklop kratkostičnega toka pri nekaterih faznih kotih nastanka kratkega stika težji, kot pri drugih.

Breaking of three-phase short-circuit current by three-pole molded-case circuit breaker

Key words: low-voltage high-current circuit breaker, single-pole MCB, three-pole three-phase MCCB, short-circuit current, breaking of three-phase short-circuit current, break with current limitation, arc splitting

Abstract: Circuit breakers are switching apparatus used to break short-circuit currents. Low voltage circuit breakers operate in short-circuit conditions by applying method of current limitation in order to reduce the magnitude and the duration of let-through current. The break is possible at current zero. The rapid increase of voltage between breaker's terminals, which forces the current rapidly toward zero, is provoked by the lengthening of arc column between opening contacts and forming of an arc loop, which finally enters the arc chutes, reaching voltage comparable to the source voltage. The waveform of voltage increase is typical for breaking events regardless of circuit conditions or even of the methods of current breaking, either by electromechanical circuit breaker or by melting fuse.

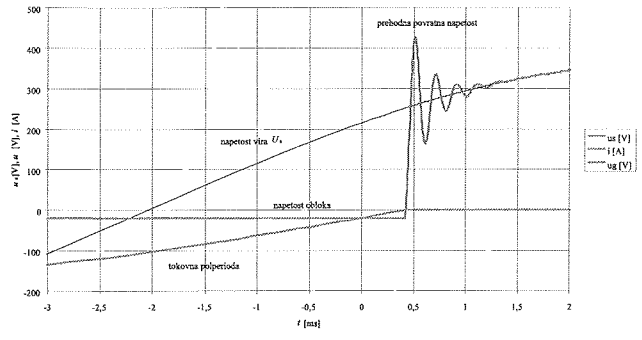
In contrary to breaking of single-phase short-circuit current, at three-phase short-circuit breaking event the voltage waveform of each particular phase does not always follow the typical single-phase one. The sum of influences of currents in the adjacent phases can result on the arc of particular phase in the sense to decelerate them on its way to arc chutes and the current limiting effect can be substantially delayed. Irregardless of uniform distribution of phase angle difference between phases of three-phase currents it is found out, that the conditions of break are harder for circuit-breaker at some specific phase angles of short-circuit event.

1. Uvod

V električnem omrežju imamo za varovanje pred učinki kratkega stika inštalirane odklopnike. Njihova naloga je hitro prekiniti tokovno pot, da ne pride do poškodb tokokroga in naprav, ki so nanj priključene. Odklopnik je stikalo, ki je sposobno izklopiti za več velikostnih redov večji tok, kot v normalnem stanju v tokokrogu. Vgrajeni nadtokovni sprožnik zazna prevelik tok in sprosti mehanizem, da se kontakti glavnega tokokroga stikala fizično razmaknejo. Pri tem pa nastane v reži med razmikajočimi se kontakti vroč plinski električni »vodnik«, oblok, ki omogoča, da tok še dalje teče po tokokrogu. Izklop je uspešno opravljen, ko med odprtimi kontakti nastali oblok »ugasne«, kar pomeni, da ni več električno prevoden. V nizkonapetostnih energetskih tokokrogih so kontakti stikalnih aparatov v atmosferskem zraku, torej oblok nastane v zraku kot izolacijskem mediju, ki zagotavlja električno ločitev odprtega kontaktnega para.

Večidel obloka izpolnjuje plazma obločnega »stolpca«, svetla električno prevodna sled delno ioniziranega plina, to je v tem primeru zrak, ki ima lastnosti električnega vodnika z neko od temperature odvisno ohmsko upornostjo. Ob kontaktnih površinah pa sta dve prielektrodni plasti, ki sta pri atmosferskem tlaku zelo tanki, reda velikosti 0,01 mm. V prielektrodnih plasteh potekajo procesi ionizacije plinskega medija in rekombinacija ionov v nevtralne osnovne delce plina. Potencialna razlika na anodni in katodni plasti je pri tokih reda 1 kA ali več je reda 10 V. Plazma v splošnem nima ostrih mej z okolico, a pri toku nad 0,5 A zaradi stanja z nižjo energijo kontrahira v omejen volumen »stolpca«, na katerem obstaja napetostna razlika, ki jo določa vsiljeni tok, ki teče skozi obločni stolpec. V prosto gorečem obloku, ki ni omejen s stenami, je potencialna razlika skoraj neodvisna od toka in je kakih 30 V. Le pri zmanjševanju toka pod 1 A potencialna razlika začne strmo naraščati. Stopnja ionizacije plina je eksponentno odvisna od njegove temperature in za oblok v zraku mora tok skozi obločni stolpec

vzdrževati temperaturo več kot 4000 K /1/. Pri dovolj majhnem toku oblok »ugasne«, ker energijski pogoji ne morejo vzdrževati zadostne stopnje ionizacije plina. Pri izklapljanju izmeničnega toka zato pričakujemo, da bo ta ugasnil pri prvem prehodu toka skozi ničlo, ker energijske razmere ne morejo vzdrževati stanja plazme. Prehod v breztokovno stanje tokokroga, ki vsebuje ohmsko, induktivno in kapacitivno komponento impedance prikazuje Slika 1.

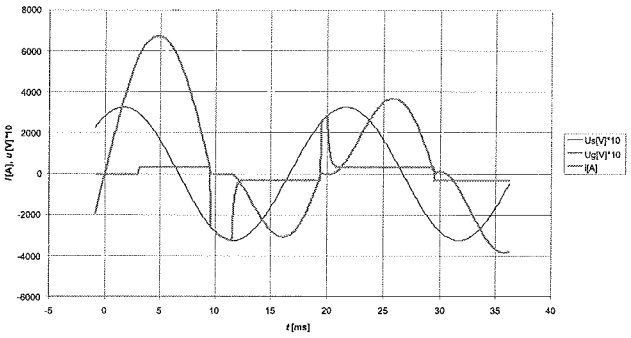


Slika 1: Prekinitev izmeničnega toka v tokovni ničli, simulacija z računskim modelom

Na grafu izklopa je prikazan časovni interval ob izteku ene polperiode izmeničnega toka 50 Hz in napetost na kontaktnem paru, med katerima je oblok s stalno napetostjo 30 V. V trenutku, ko tok *i* doseže ničlo, oblok ugasne in tok ostane na vrednosti nič, kar pomeni, da je izklop uspešno opravljen. Napetost med kontakti u_g se vrne v napetost vira s prehodnim pojavom dušenega nihanja s frekvenco v redu velikosti nekaj 10 kHz, kot reakcija tokokroga na prekinitev toka.

Tak primer uspešne prekinitve izmeničnega toka v »naravni« ničli tokovne polperiode z enostavnim razmikom (odprtjem) kontaktnega para je pri izklopu velikih tokov zelo redek. Plazma obločnega stolpca ima neko maso in toplotno kapaciteto, na mejni plasti odvaja v okoliški zrak toploto, večidel pa izseva svetlobo. Zato se oblok odziva na spremembo zunanjih pogojev z neko časovno konstanto, ki je reda velikosti ės /2/. V trenutku prehoda toka skozi ničlo je lahko temperatura plazme še tolikšna, da je v začetku naslednje tokovne polperiode oblok še dovolj električno prevoden, kar omogoči ponoven razvoj obloka. Ponovni vžig obloka po kratkotrajni ugasnitvi imenujemo »povratni vžig« obloka. Pojav je prikazan na Sliki 2, kjer po prehodu skozi ničlo tok za kratek čas preneha teči, a začne ponovno naraščati pri neki manjši napetosti med kontakti, ki se pri povratnem vžigu sesede na raven obločne napetosti. Vsak zaporedni povratni vžig nastane pri manjši napetosti, dokler tok oblok niti v tokovni ničli tudi za kratek čas ne ugasne. Povratni vžig je seveda uničujoč tako za odklopnik kot tudi za komponente tokokroga, ki naj bi jih varoval.

Povratek v električno prevodno stanje lahko povzročijo tudi velike napetostne amplitude oscilacij prehodnega pojava, ko se kontakti šele razmikajo in je napetostna trdnost med njimi manjša, kot v odprti legi. Če amplituda prehodne povratne napetosti preseže izolacijsko trdnost plina med odpirajočimi kontakti, se razelektritev »vžge« na dielektrični način. Tudi to privede v povratni vžig obloka.



Slika 2: Zaporedje povratnih vžigov v sledečih polperiodah izmeničnega toka, simulacija z računskim modelom

Pri izklopu kratkostičnega toka reda nekaj kA se v eni sami polperiodi izmeničnega toka (10 ms) lahko sprosti nekaj MJ toplote, kar termično zelo obremenjuje odklopnik in lahko preseže njegovo termično vzdržljivost. Izklopna zmogljivost odklopnika je podana z velikostjo razpoložljivega toka (prospective current), ki ga lahko daje vir pri dani napetosti v tokokrog brez odklopnika, ki pa ga je ta sposoben uspešno izklopiti in ostati pri tem sposoben opravljati osnovne stikalne funkcije. Pomemben podatek za izklop je joulski integral pri izklopu prepuščenega toka, to je integral i^2 po času v intervalu od nastanka kratkega stika do prekinitve kratkostičnega toka. Sodobni odklopniki za nizkonapetostne energetske tokokroge uspejo skrajšati tokovni interval na čas, ki je manj od ene polperiode izmeničnega toka, in amplitudo prepuščenega toka I_D za precej manj, kot jo omeji impedanca stikala. To dosežemo z metodo »tokovne omejitve« izklopnega toka, kjer med izklopom naraste napetost na priključkih odklopnika u_g na 1,5-kratnik do 2-kratnik napetosti vira. Učinek je načelno opisan z (1), kjer pri kateri koli impedanci *Z* v tokokrogu z zmanjšanjem razlike med napetostjo u_g in napetostjo vira u_s omejimo tok *i*, ki v tokokrogu doseže vrednost 0 takrat, ko sta obe napetosti izenačeni:

$$i\,Z(R,L,C) = u_s - u_g \qquad (1)$$

Obstaja nekaj načinov, kako povečati napetost med kontakti u_g s povečanjem potencialne razlike na obloku.

Če oblok stisnemo na majhen presek v kanal ali v režo s toplotno dobro prevodnimi stenami, mu zmanjšamo električno prevodnost, zato se mu pri danem toku poveča njegova napetost. Upornost obloka R_{oblok} v splošnem povečamo s podaljšanjem ali zoženjem obločnega »stolpca«, ter z zmanjšanjem električne prevodnosti medija, kot velja splošno za električni vodnik. V nizkonapetostnih odklopnikih je razmak med kontakti reda velikosti do 10 mm, kar predstavlja začetno dolžino obločnega stolpca, z oblikovanjem v zanke in zavoje ga lahko precej podaljšamo, a samo s tem obločne napetosti ne povečamo na želeno vrednost. Večji učinek nudi stisnjenje obloka v ozek kanal,

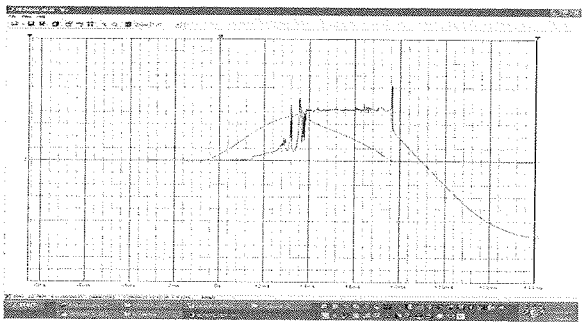
kar s pridom izkoriščamo v talilnem vložku varovalke /3/, stiskanje obloka med odpiranjem kontaktov v ozko špranjo z zaslonom, ki ga poganja mehanizem odklopnika, pa se zaradi počasnosti mehanizma ni izkazal kot upoštevanja vredna alternativa sedaj uporabljenim načinom /4/. Učinkovito pa je vodenje obloka po kovinskih letvah z Biot-Savartovo silo na tokovno zanko obločnega »stolpca«, vendar ga zaradi magnetohidrodinamskih zakonov ne moremo enostavno stlačili v poljubno ozko režo. Lahko pa ga speljemo med kovinske ploščice »deion« komore, kjer ga razdelimo na več zaporednih delnih oblokov. Napetost razsekanega obloka znatno naraste, čeprav se mu skupna dolžina ne poveča. Vsak delni oblok, ki je umeščen med dve plošči »deion« komore, ima svojo anodno in katodno plast z ustrezno anodno in katodno napetostjo. Vpliv dolžine obločnega stolpca delnega obloka je zanemarljiv, zato je razmak med ploščami deion komore majhen, od 1 mm do 2 mm. Obločna napetost U_{oblok} v zraku med jeklenima ploščama z razmikom za nekaj mm je podana z (2) in praktično neodvisna od toka /5/:

$$U_{oblok}\,[\mathrm{V}] = 18{,}5\,\mathrm{V} + 2{,}25\,\mathrm{V/mm}\cdot d\,[\mathrm{mm}] \qquad (2)$$

kjer je konstantni člen vsota obeh prielektrodnih napetosti, drugi člen pa vsebuje gradient napetosti v obločnem stolpcu. Čeprav ta zavzema skoraj vso dolžino med ploščama, k napetosti obloka največ doprinesejo tanke prielektrodne plasti. Jeklo je zaradi svojih feromagnetnih lastnostih pogosto uporabljen material za »deion« komore. Celotna potencialna razlika med koncema obloka, ki je razdeljen med N kovinskih plošč na N+1 zaporednih delnih oblokov, je U_{oblok} = (N+1)*25 V. Število plošč N načelno navzgor ni omejeno, vendar je praktična zgornja meja za N ta, da naj bo 1,5 < U_{oblok} < 2 U_s. V komoro z velikim številom plošč je oblok težko spraviti dovolj hitro in ga držati v njej dovolj dolgo.

2. Potek izklopa enofaznega toka v enopolnem odklopniku za izmenični tok nizke napetosti

Izklop kratkostičnega toka ene faze zadovoljivo opravi enopolni odklopnik. To pomeni, da izklapljamo tok v enem tokokrogu z enim sklopom sprožnika, kontaktov in deion komor. Današnji enopolni odklopniki večidel izkoriščajo način izklopa z učinkom tokovne omejitve in dosegajo izklopno zmogljivost do 10 kA pri napetosti vira 230 V. Hkrati je odločilno, da je odklopnik v primeru kratkega stika čim hitreje zaznati prevelik tok in razmakniti kontaktne dele kontaktnega para v varovanem tokokrogu. V ta namen izkoriščamo učinek odskoka kontaktov (blow-off) zaradi odpiralne sile toka skozi kontaktno mesto /6/ in sunek udarne kotve na odpiralni kontaktni del, ki jo pospešimo z magnetnim poljem izklopnega toka skozi tuljavo nadtokovnega sprožnika, katerega naloge je zaznati prevelik tok in sprožiti razmaknitev kontaktov. Zaželeno je doseči začetno odpiralno hitrost kontaktov vsaj 5 m/s. V trenutku fizičnega odmika kontaktov napetost obloka takoj naraste na minimalno vrednost U_{oblok}, ki znaša od 10 V do 12 V, in narašča sorazmerno z razmakom kontaktnega para. Vsak od kontaktnih delov ima podaljške oblikovane v »rogove«, ki olajšajo obloku oblikovanje v zanko in njeno širjenje zaradi Biot-Savartove sile, pa tudi približevanje ploščam deion komore. Ta razvoj obloka traja nekaj ms in U_{oblok} doseže vrednost do 100 V. Tok i, ki je naraščal skoraj po sinusoidi izmeničnega toka, doseže na koncu te faze izklopa svojo končno vrednost I_D. Oblok se nahaja pred vstopom v deion komoro oblok in U_{oblok} hitro narašča, ko se začne proces delitve na delne obloke. To ne uspe vedno v enem koraku, večkrat se vrne v prostor pred komor in ponovno vstopi med plošče. Ko je oblok dokončno razdeljen na delne obloke, je njegova napetost neodvisna od trenutnega toka in stalna. Na Sliki 3 je prikazan oscilogram napetosti med priključki enopolnega odklopnika, ki je približno enaka poteku obločne napetosti. Začetnemu počasnejšemu naraščanju napetosti, ko je oblok še med kontakti, sledi hitrejše naraščanje v fazi podaljševanja obloka v zanko in strm dvig, ko vstopa v deion komoro in se deli na delne obloke. Vstop uspe šele v treh korakih, kar se vidi iz treh napetostnih konic pred segmentom, ko napetost dokončno doseže kakih 400 V in drži to vrednost. Medtem pa se tok i strmo spušča od I_D proti vrednosti i = 0. Vrednost i je ostala v ničli in napetost je z U_{oblok} prešla na potek napetosti vira u_s, kar pomeni, da je oblok ugasnil. Prekinitev toka je bila uspešna in izklop je opravljen. Izklopne količine, kot je čas trajanja izklopa t_{izkl}, I_D in joulski integral (integral $i^2(t)\,dt$ od pojava kratkega stika do t_{izkl}), so odvisne od tega, v katerem tokovnem faznem kotu je bil izklop začet, vendar je časovni potek napetosti po teh fazah izklopa značilen za izklop izmeničnega toka z enopolnim odklopnikom. Potek izklopa je pri drugem faznem kotu tokovne sinusoide in pri drugih parametrih tokokroga kvalitativno identičen prikazanemu poteku izklopa, tudi če je izklop opravljen z drugačnim tipom odklopnika ali s taljivo varovalko.



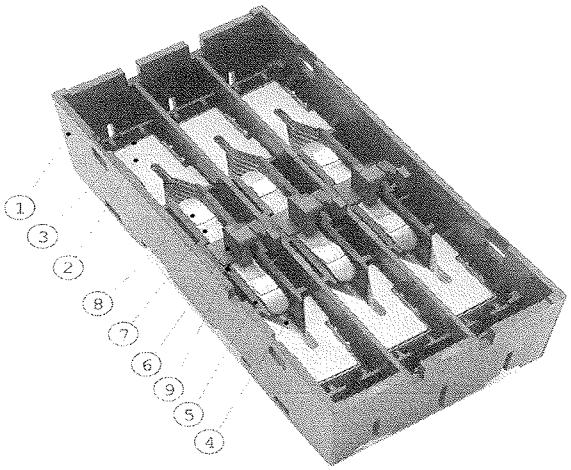
Slika 3: Oscilogram izklopa enopolnega odklopnika, preskusni enofazni tok 3 kA pri 230 V

3. Izklop trifaznega toka v tripolnem nizkonapetostnem odklopniku

Večpolni odklopnik ima priključke za izklop več neodvisnih tokokrogov. Vsak polu ima svoj lasten sistem zaznavanja prevelikega toka in stikalni sistem za izklop s tokovno ome-

jitvijo. Skupen za vse pole je le vklopno-izklopni mehanizem. Vsi poli, ki so namenjeni za priključitev posameznih faz, imajo identično zgradbo in način delovanja, le pol za nevtralni tokokrog (če ga odklopnik ima) se po zgradbi lahko razlikuje od polov za posamezne fazne tokokroge. Poli odklopnika so električno vezani vzporedno in tudi prostorsko razmeščeni vzporedno drug ob drugem. Medsebojno so izolirani s pregrado iz izolacijskega materiala.

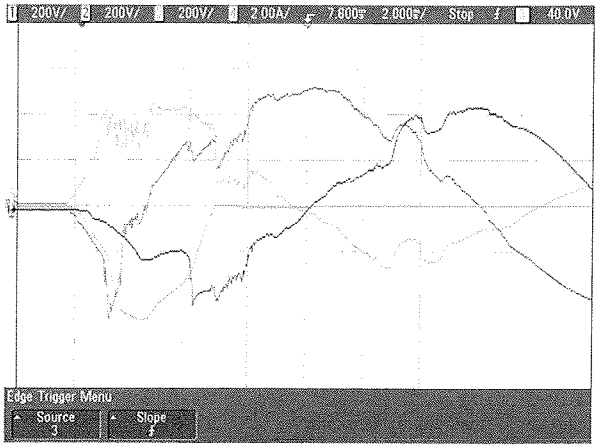
Primer razporeditve polov za posamezne fazne tokokroge v odklopnika za trifazni sistem prikazuje Slika 4. Prikazana je izvedba kontaktno-obločnega sistema z dvojnim parov kontaktov, ki dvakrat prekinjata tokokrog posameznega pola. Na sliki je videti le gibljivi kontaktni del z dvema kontaktnima mestoma in ob vsakem kontaktnem paru deion komoro, zaradi dvojne prekinitve ima odklopnik v vsakem polu po dve komori.



Slika 4: Razpored polov tripolnega odklopnika za izklop trifaznega tokokroga nizke napetosti

Izklop trifaznega toka ima v primerjavi z izklopom toka ene faze nekaj posebnosti. Kot faznega zaostanka med napetostmi in tokovi v posameznih fazah pri simetričnem bremenu je 120°. V primeru hkratnega nastanka kratkega stika v vseh treh fazah, kot pri preskusih kratkostične izklopne zmogljivosti odklopnika, sinhroniziramo trifazni kratki stik pri izbranem vklopnem faznem kotu na eni od faz, npr. v L1. V trenutku vklopa preskusnega vira kratkostičnega toka tok v vsaki fazi narašča zvezno od nič, potek pa je zaradi faznega zaostanka med fazami L1, L2 in L3 v vsaki fazi drugačen in odvisen od trenutka vklopa vira in impedance tokokroga. Zmogljivost preskusnega vira omogoča nastaviti razpoložljivi tok I_p na 50 kA pri 400 V medfazno. Z bremenskim uporom in dušilko nastavimo želeni tok in faktor moči (kosinus kota fazne diference med tokom in napetostjo) v vsaki fazi. Preskusi kratkostične izklopne zmogljivosti potekajo v specializiranih preskuševališčih po zahtevah standardov IEC 60947.

Ko s sinhronskim stikalom v izbranem trenutku v vseh treh fazah (L1, L2 in L3) vklopimo nastavljeni kratkostični tok I_p, odklopnik samodejno opravi izklopni manever. Pri tem osciloskopsko spremljamo potek napetosti in toka v vsaki od faz in ugotavljamo uspešnost izklopa (eventualen nastanek povratnega vžiga), učinek omejitve toka, če pa ima registrator hitrih pojavov dograjene matematične funkcije, pa še integral I^2 v času trajanja izklopa. Iz posnetih oscilogramov se da razbrati tudi obnašanje obloka med izklopom. Slika 5 prikazuje oscilogram izklopa trifaznega toka s tripolnim odklopnikom, ki ima v vsakem polu kontakte za dvojno prekinitev tokokroga.

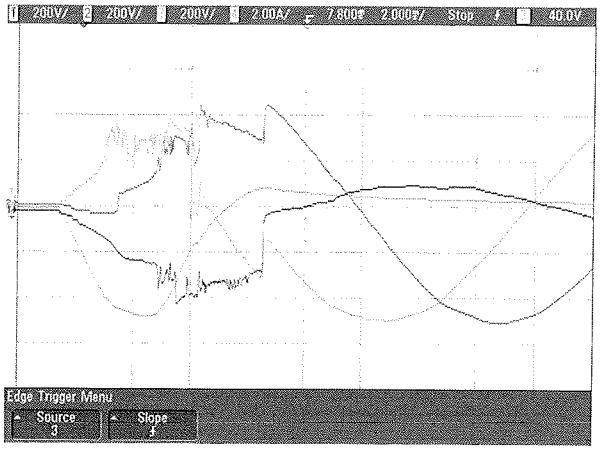


Slika 5: Oscilogram izklopa trifaznega toka 25 kA pri 400 V

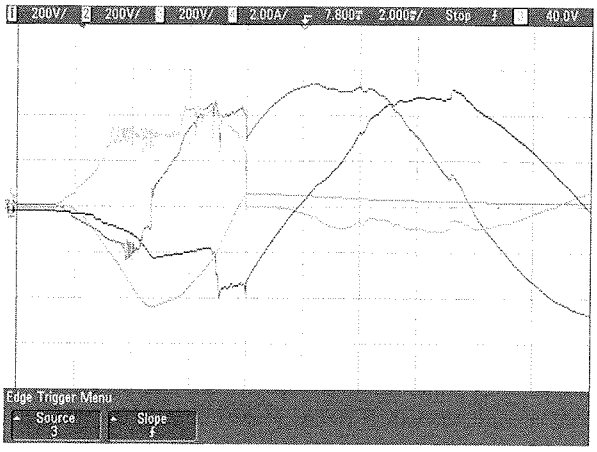
Napetost vira je bila 420 V medfazno, nastavljeni razpoložljivi tok I_p pa v vsaki fazi 25 kA. Faktor moči je bil nastavljen z dodatno zračno dušilko, da simulira induktivno breme. Za bremenom so bili fazni vodniki povezani v zvezdišče brez priključitve na nevtralni vodnik N. V takem primeru se tok vedno prekine najprej v eni fazi, v ostalih pa je tok po velikosti enak in nasprotne polaritete. Zato prekinitev toka v teh dveh fazah nastane hkrati.

Učinek tokovne omejitve v enem polu je določen s časovnim potekom napetosti na priključkih tega pola, tega pa večidel določa potek napetosti obloka. Na polu, ki prvi prekine tok, je potek napetosti u_g podoben temu pri izklopu enopolnega odklopnika (primerjaj z oscilogramom na Sliki 3). V enem od ostalih dveh polov tok v začetku narašča sorazmerno počasi, kar je pogojeno z vklopnim faznim kotom toka v tem polu, zato se tudi kontakti razmikajo počasneje. Obločna napetost narašča počasneje in trenutek vstopa obloka v deion komoro je kasnejši. Največja vrednost u_g je kakih 400 V, ko ta doseže stalno vrednost pred prekinitvijo toka. Čas trajanja izklopa je 5 ms do 6 ms od nastanka kratkega stika. Ta čas določa pol, na katerem u_g narašča najpočasneje. Potek napetosti na tem polu se razlikuje od tistega, ki je značilen za enopolni izklop: v začetku napetost narašča tako, kot narašča obločna napetost med razmikajočimi se kontakti in doseže približno 100 V na kontaktni par, potem pa njen dvig kaže zastoj, saj se ji vrednost za kake 2 ms ne spremeni. Šele potem se dvigne proti 400 V, kar kaže na vstop obloka v deion komoro in delitev na delne obloke. Primerjava med potekom napetosti u_g na obeh polih, v katerih je tok prekinjen nazadnje, kaže na medsebojni vpliv sosednjih faz. Biot-Savartova sila

na tokovno zanko ne deluje samo na večanje zanke obločnega stolpca v enem polu, ampak v sosednjem polu izriva oblok stran od deion komore zaradi istega učinka povečevanja tokovne zanke. Ker medsebojno vplivajo drug na drugega vsi trije poli, je potek u_g v L1, L2 in L3 odvisen od izbranega vklopnega faznega kota kratkostičnega toka. Izklopni pojav pri vklopnem faznem kotu 0° vira preskusnega toka v L1 je prikazan na oscilogramu Slike 6a, pri faznem kotu 30° v L1 pa na Sliki 6b.

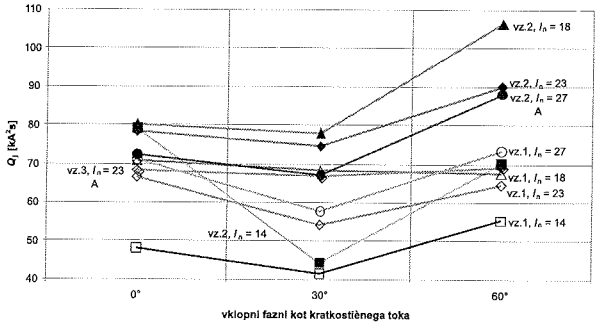


Slika 6a:Izklop trifaznega toka z začetkom pri faznem kotu 0°



Slika 6b:Izklop trifaznega toka z začetkom pri faznem kotu 30°

Posledica interference tokov med poli je, da pri nekem vklopnem faznem kotu kratkostičnega toka odklopnik teže izklopi. Merilo za oceno, koliko je izklop blizu zmogljivosti odklopnika, je izklopni integral I^2 po času t v trajanju izklopa. Za nekaj različic tripolnih odklopnikov istega tipa je bil med preskusom izklopa trifaznega toka kratkega stika I_p = 25 kA pri 400 V medfazno izmerjen izklopni integral in ugotovljena njegova največja vrednost, pri čemer so bili izbrani trije različni vklopni fazni koti: 0°, 30° in 60°. Rezultati so prikazani na grafu Slike 7: največje vrednosti izklopnega integrala najverjetneje nastopajo pri kotu 60°.



Slika 7: Joulski integral izklopa pri različnih faznih kotih začetka preskusnega toka

4. Sklep

V večpolnem odklopniku se med izklopom velikih tokov pojavlja interferenca med poli, zaradi česar je izklop toka v enem od zadnjih polov, v katerem se tok prekine, otežen. Interferenca se izraža kot oviranje gibanja obloka zaradi magnetnih učinkov toka sosednjega pola. Zato kasneje vstopi v deion komoro in kasneje doseže predvideno napetost, ki zagotavlja učinek tokovne omejitve. Ker nastopa interferenca med vsemi tremi poli odklopnika, odločujoča pa je v času trajanja ene polperiode izmeničnega toka, je njen učinek na izklop odvisen od razporeditve toka po vseh polih in njegovega časovnega poteka v času trajanja izklopa. Oboje pa je odvisno od faznega kota nastanka kratkega stika. Zato je težavnost izklopa za odklopnik odvisna tudi od tega kota.

5. Literatura

/1/ J. M. Yos: *Transport properties of nitrogen, hydrogen, oxigen and air to 30 000 K*, Techn. Memorandum RAD-TM-63-7, AVCO Corp., Mass., 1963

/2/ N. Reimann: *Das Stromnulldurchgang-Verhalten von Niederspannungs-Schaltlichtbögen unter Berücksichtigung von Kathoden-schicht und dynamischer Bogensäule*, disertacija, Technische Universiät Carolo-Wilhemina, Braunschweig, 1987

/3/ M. Bizjak, P. Žunko, S. Poberaj: *Characteristics of arc in melting fuse element filled with quartz sand*, Mediterranean Electritechn. Conf. MELECON 91, Ljubljana, 1991, pp. 1493-1496

/4/ E. M. Belbel, M. Lauraire: *A new breaking technology independent from the contact material*, 13th Internat. Conf. On Electric Contacts, Lausanne, 1986, pp. 150-155

/5/ H. Klepp: *Über den Einfluss der Löschkammerkonstruktion auf die Lichtbogenlöschung in Schützen gröser Nennstromstärken*, disertacija, Technische Universiät Carolo-Wilhemina, Braunschweig, 1982

/6/ M. Bizjak, S. N. Kharin, H. Nouri: *Influence of vapour pressure on the dynamic of repulsion by contact blow-off*, Proc. 21st Intern. Conf. On Electrical Contacts, Zurich, Switzerland, 2002, pp. 268-275

dr. Martin Bizjak, univ. dipl .ing fizike
R&R, Iskra MIS, d. d, Ljubljanska cesta 24a, 4000 Kranj,
e-mail: martin.bizjak@iskra-mis.si

Prispelo (Arrived): 01.09.2008 *Sprejeto (Accepted): 19.03.2009*

EFFICIENT SAMPLING FOR THE EVALUATION PROTOCOL FOR 2-D RIGID REGISTRATION

Andreja Jarc[1,2], Janez Perš[2], Peter Rogelj[2], Stanislav Kovačič[2]

[1]Sipronika d.o.o., Ljubljana, Slovenia

[2] University of Ljubljana, Faculty of Electrical Engineering, Machine Vision Laboratory, Ljubljana, Slovenia

Key words: image registration, evaluation protocol, pseudo-random vs. quasi-random sampling.

Abstract: In our research we aim to reduce the computation time spent on the evaluation protocol of a criterion function for rigid registration tasks. The basic evaluation protocol is performed on *N* uniformly distributed sampling lines in the *K*-dimensional transformation space. Similarity between two images is measured at each of the equidistantly placed points on each sampling line, which is a computationally intensive process. We hypothesize that the computational complexity can be reduced if attention is paid to the selection method of the sampling lines. In the research we compared four sampling methods which affect density and distribution of sampling lines: basic regular sampling, pseudo-random, Sobol quasi-random and Halton quasi-random sampling. We show that the use of Halton quasi-random generator yielded the most uniform distribution of sampling line. Thus, with proper sampling, the number of sampling lines could be systematically reduced in comparison to pseudo-randomly generated lines. This would result in shorter computation time spent on the protocol. The reduction of computation time is especially important when many criterion functions need to be evaluated (e.g. texture feature based image registration). The evaluation protocol was tested on a set of 11 2-D DRR (Digital Reconstructed Radiograph) and EPI (Electron Portal Image) image pairs. The tests have been conducted on intensity feature images as well as texture feature images extracted from the original intensity images. The paired Student's *t*-tests ($p<0.05$) indicated that results obtained from Halton quasi-random sampling were statistically significantly more consistent than the results based on regular or pseudo-random sampling.

Učinkovito vzorčenje za vrednotenje kriterijskih funkcij za 2-D togo poravnavo slik

Kjučne besede: poravnava slik, vrednotenje kriterijskih funkcij, pseudo-nakjučno vs. kvazi-naključno vzorčenje.

Izvleček: V svojih raziskavah poskušamo zmanjšati čas, ki je potreben za kvantitativno vrednotenje kriterijskih funkcij pri togi poravnavi slik. Osnovni protokol za vrednotenje kriterijskih funkcij temelji na *N* naključno porazdeljenih vzorčnih premicah v *K*-dimenzionalnem prostoru parametričnega transformacijskega modela. Podobnost med slikama izračunamo v ekvidistančnih točkah, ki ležijo na *N* vzorčnih premicah. Računanje podobnosti med slikama je računsko potraten postopek. Predpostavljamo, da bi računsko zahtevnost postopka zmanjšali, če uspemo najti optimalno metodo vzorčenja za generiranje vzorčnih premic. V ta namen med seboj primerjali štiri različne metode vzorčenja, ki vplivajo na gostoto in porazdelitev vzorčnih premic v prostoru. Metode vzorčenja, ki smo jih v testih med seboj primerjali, so bile naslednje: enakomerno vzorčenje po mreži, pseudo-naključno, Sobol kvazi-naključno in Haltonovo kvazi-naključno vzorčenje. Iz rezultatov testov se je izkazalo, da nam Haltonovo kvazi-naključno vzorčenje omogoča najbolj enakomerno in s tem »pravično« porazdelitev vzorčnih premic v prostoru. Enakomerna porazdelitev premic bi nam omogočila, da bi število vzorčnih premic lahko sistematično zmanjšali v primerjavi s številom pseudo-naključno generiranih premic. Zmanjšanje števila premic bi direktno pomenilo skrajšanje računskega časa, potrebnega za vrednotenje kriterijskih funkcij. Pohitritev postopka za vrednotenje kriterijskih funkcij še posebej pride do izraza v primerih, kjer je potrebno ovrednotiti večje število kriterijskih funkcij, za primer, poravnava s teksturnimi značilnicami. Teste različnih vzorčenj smo izvedli na setu 11-ih dvodimenzionalnih DRR (Digital Reconstructed Radiograph) in EPI (Electron Portal Imaging) slikovnih parih. Teste smo izvedli tako na originalnih svetlostnih slikah kot tudi na slikah teksturnih značilnic. Studentov t-test ($p<0.05$) je pokazal, da so rezultati, dobljeni na podlagi Haltonovega kvazi-naključnega vzorčenja statistično signifikantno bolj konsistentni od rezultatov, dobljenih z uporabo pseudo-naključnega vzorčenja.

1 Introduction

Clinical diagnosis, as well as therapy planning and evaluation rely increasingly on multiple images of different modalities. For example, in radiation therapy planning a CT (computer tomography) scan is needed for dose distribution calculations, while the contours of the target lesion are often best outlined on MRI (magnetic resonance image) /1/. Image registration is a procedure, where images of the same anatomical structures, acquired using the same or different imaging devices, are brought into the best possible spatial correspondence with respect to each other. Image registration is therefore a fundamental step of information integration. Detailed classifications of registration techniques applied to medical images have been reviewed in a number of surveys /2,3,4,5,6,7/.

In general, image registration is implemented as an optimization task of finding such transformation parameters that maximize or minimize a criterion function (CF), which measures a similarity between images as a function of registration. A criterion function can be considered as a function mapping from *K*-dimensional continuous space to a subset of a real line, where *K* is the number of parameters (degrees of freedom) of the parametrical spatial transformation model /8/. For example, for rigid registration of two-dimensional (2-D) or three-dimensional (3-D) images the value of *K* is 3 or 6, yielding in 3-D or 6-D optimization problem, respectively. The outcome of registration heavily depends on the criterion function profile.

The quality of CF in terms to registration, as proposed by Škerl et al, can be described by the following parameters:

accuracy (*ACC*), risk of non-convergence (*RON*), distinctiveness of the global extremum (*DO*) and capture range (*CR*) /8/. First, the accuracy of the CF is defined as a distance from an optimum of the CF to the 'gold standard' transformation, which corresponds to the aligned position of images. Next, the risk of non-convergence is a measure of robustness of CF. It includes the number, position and distinctiveness of local extrema. RON also describes how sensitive is a CF to interpolation, sampling, partial image overlap and noise. A *DO* measures how distinctive is a maximum (minimum) in respect to the decreasing (increasing) values of CF away from the optimum. Capture range is referred to a limited range of transformations around the optimum for which CF is a monotonic decreasing (increasing) function.

Exhaustive search of the parametrical transformation space would be an obvious and the most precise method to evaluate CF prior to registration at every transformation estimate in *K*-dimensional space. However, in terms of computational demands, this approach is prohibitory expensive.

Let us take for example a simple 2-D rigid registration problem with K=3 transformation parameters (two translations and one rotation), with a grid step size of 1 mm and a capture range of 50 mm. For these modest requirements we would get 50^3 (=125 000) transformation estimates at which CF should be evaluated. If the same requirements were to be met for a 3-D registration problem with K=6 degrees of freedom (3 translations, 3 rotations), $1.56 \cdot 10^{10}$ estimates of CF would follow.

The protocol for evaluation of similarity measures for rigid registration /8/ is an improvement of the exhaustive search method, as it applies random sampling to the parametrical space. The protocol has been tested for various multi-modal rigid registration tasks, therefore it is becoming a reference method for evaluation of similarity measures. It was devised by Škerl et al /8,9,10/. The continuous *K*-dimensional space is first normalized so that equal changes of each of the parameters in the normalized parametrical space produce the same mean voxel shift. Next, the normalized *K*-dimensional space is "pierced" by *N* randomly selected lines, where the intersection points with a hyper-sphere are uniformly distributed on the surface of the hyper-sphere with radius *R*. All sampling lines converge in the 'gold standard' (GS) transformation which corresponds to the aligned position of two images. Each sampling line is subsequently sampled by *M* equidistant points and the step size between points is defined as (2R/M). Let's denote X_0 as the origin or GS transformation and $X_{n,m}$ as one of the sampled points. Each of $X_{n,m}$ represents a *K*-dimensional vector of transformation parameters (see (Figure 1 [1]))

For the evaluation protocol as described above, the density and distribution of sampling lines are crucial to obtain representative estimates of the continuous *K*-dimensional transformation space. Following the recommendations in /8/, the number of sampling lines *N*, defined by a pseudo-random generator should be set to 50 for a 6-D optimization task. This way, sampling points should be uniformly distributed on the surface of a sphere.

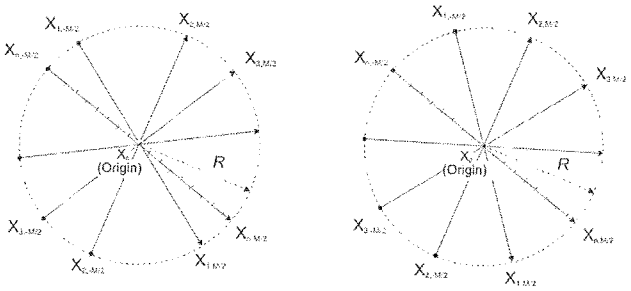


Fig. 1. 2-D parametrical space, sampled by N *lines and* M *points per line. The maximal displacement from the GS is denoted by* R, *which is a radius of the K-dimensional hyper-sphere. M and R define the step size between sampling points: 2·R/M. The left figure depicts sampling lines, which directions are generated by pseudo-random generator. The right figure depicts sampling lines, which are maximally avoiding each other.*

Nevertheless, examples in the literature show that pseudo-random generator is not the optimal choice to fill a *n*-dimensional space uniformly. Sequences of *n*-tuples that fill *n*-space more uniformly than uncorrelated random points are called quasi-random sequences /11/. The main property of sample points given by a quasi-random sequence is that the points are maximally avoiding each other. By this means the same number of sampling lines defined by a quasi-random sequence should cover *K*-dimensional space more evenly than if lines were defined by a pseudo-random sequence.

The following example (Figure 2) shows 2500 points on a sphere, generated by four different sampling methods. The first one is an example of a basic regular sampling of a sphere, the second one is an example of the pseudo-random generated points, the third one for the Sobol quasi-random /12/ and the last one for the Halton quasi-random sampling method /13/.

One can notice, that point density increases towards the poles when the basic regular sampling of a sphere is used. Furthermore, the pseudo-random generated points build small clusters and the vacant space between them is evidently large. The Sobol quasi-random generator delivers more uniform distribution of the points and less vacant space between them. One can see, that the most uniform distribution of the points is generated by Halton quasi-random generator.

The aim of this paper is to evaluate the performance of the evaluation protocol by comparing the consistency of its results for three random sampling methods: pseudo-random, Sobol quasi-random and Halton quasi-random. In

1 The figure is upgraded from /8/.

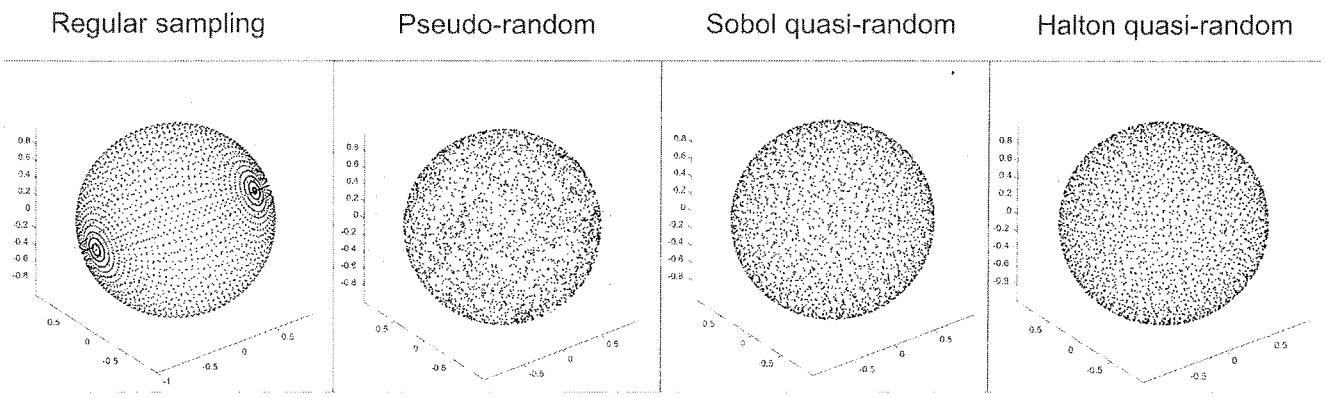


Fig. 2. 2500 points generated by four sampling methods.

addition, the comparison has been performed for the use of basic regular sampling method. Our goal is to find the sampling method that would exhibit similar consistency as the accepted pseudo-random sampling, but with significantly reduced number of sampling points. The comparisons were conducted on a set of 11 2-D DRR (Digital Reconstructed Radiograph) and EPI (Electron Portal image) images. The evaluation protocols have been applied not only to criterion functions based on intensity features but also to several texture feature images extracted from original intensity images/14/.

Generally, a large number of texture features may be extracted from intensity images. Therefore, the evaluation protocol should be conducted as many times as many features (or potentially their combinations) are available. In general, this may be more than 100 times. If the number of sampling lines was reduced in comparison to the pseudo-random sampling, the time spent on the protocol would be reduced as well and one could evaluate considerably larger number of criterion functions.

We anticipate that the modified evaluation protocol - as proposed in this paper - will be used for the evaluation of a larger number of texture feature based criterion functions prior to registration. The final goal is to select the most appropriate features for a specific registration task. Reduced computational time directly increases the number of texture features that we can afford to evaluate. This increases the chance the best features will be found.

This paper is organized as follows: first, the generation of sampling lines is described in detail. Then, the design of experiments is presented and the data set on which the tests have been conducted is introduced. Finally, some details about texture features used for registration are explained, and the comparisons of results among different sampling methods are shown. Discussion and conclusions complete the paper.

2 Methods and materials

2.1 Generation of sampling lines

In our modified evaluation protocol, the sampling lines in 3-D parametrical space (two translations and one rotation) are generated first by use of the Sobol quasi-random and second by Halton quasi-random generator. To compare with random sampling methods, the 3-D parametrical space is additionally sampled by a basic regular sampling.

The azimuthal angle φ and the polar angle θ of spherical coordinates are the outputs of the regular sampling or the quasi-random generators. r is a distance (radius) from a point to the origin (Figure 3).

The spherical coordinates (r, ϕ, θ) are related to the Cartesian coordinates (x, y, z) by the following equations /15/:

$$x = r\cos(\varphi)\sin(\theta) \quad (1)$$

$$y = r\sin(\varphi)\sin(\theta) \quad (2)$$

$$z = r\cos(\theta) \quad (3)$$

where $r \in [0,\infty)$, $\varphi \in [0,2\pi]$ and $\theta \in [0,\pi]$.

Each of the N sampling lines in 3-D parametrical space is defined by randomly selected starting position $X_{n,-M/2}$ on a sphere at the distance R from the origin and its mirror point $X_{n,M/2}$. The starting points are defined by a 3-D vector *[x,y,z]* (Eq. (1), (2) and (3)).

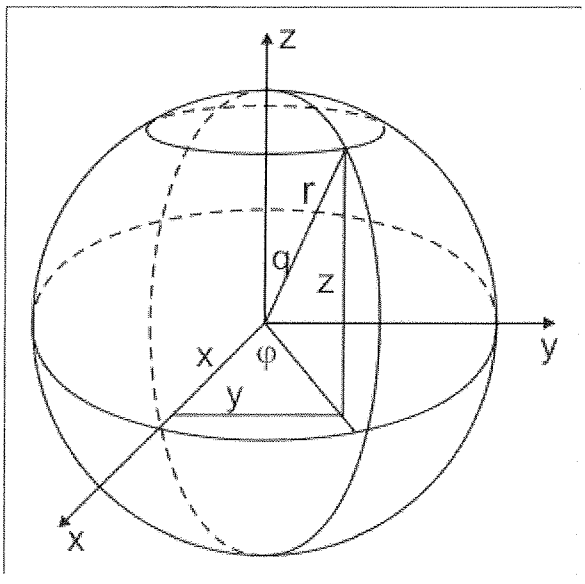


Fig. 3. Our notation of 3-D spherical coordinates.

The following two examples (Figure 4) show 2500 points generated by the method as described above. In the first example, the Cartesian coordinates of the points are defined by the Sobol quasi-random generator and in the second example by the Halton quasi-random generator.

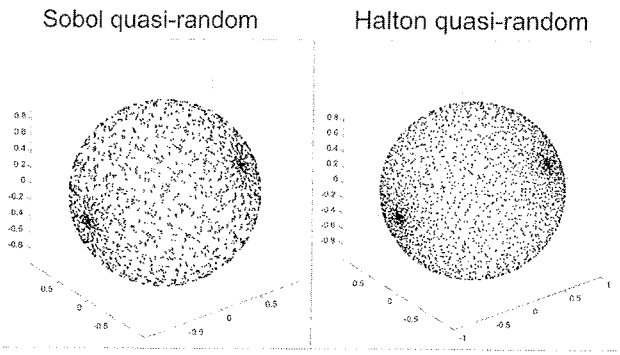


Fig. 4. Distribution of sampling points, where angles φ and θ have been selected by quasi-random generators.

It can be easily seen that points are not uniformly distributed over the sphere surface since they group in two clusters at the poles. The reason is, that the area element $d\Omega = sin(\theta)d\theta d\varphi$ depends on θ, and therefore points selected in this way are clustered near the poles /16/. To obtain points such that any small area on the sphere is expected to contain the same number of points, we choose u and v to be quasi-random variants on [0,1]. Then:

$$\varphi = 2\pi u$$
$$\theta = arccos(2v - 1) \quad (4)$$

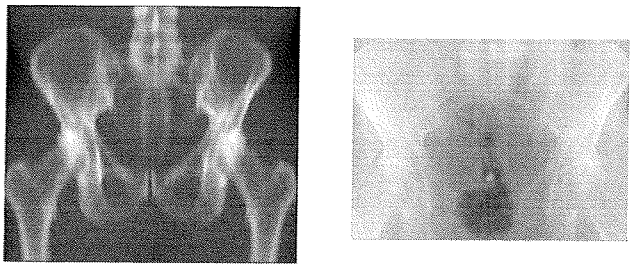
gives the spherical coordinates for a set of points, which are uniformly distributed over Ω.

Using this correction, the uniformity of sampling points improves as shown in Figure 2. The above correction was used throughout the paper for all but regular sampling methods.

2.2 Test image set

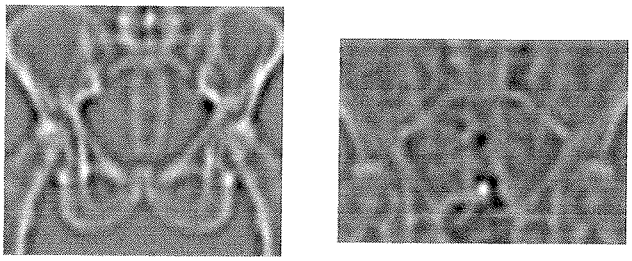
The tests have been conducted on 11 pairs of DRR (Digital Reconstructed Radiograph) and EPI (Electron Portal Imaging) images of the pelvis (Figure 5). By correctly matching the two modalities, it is possible to verify the positioning of the patient during radiation therapy and automatically adjust the positioning if necessary.

The registration of DRR/EPI images is not a trivial problem due to 2-D representation of 3-D data. Several papers have been published proposing and/or investigating various registration methods using these image modalities for patient positioning applications /17,18,19/. However, we found that intensity based registration is not reliable, since the intensity features do not comply with some global intensity relationship, expected by intensity-based registration approaches/20/. Therefore, an alternative registration approach based on texture features has been proposed to register DRR/EPI images/14/.



(a) DRR image of the pelvis (b) EPI image of the pelvis

Fig. 5. An example of one of the intensity image pair. (a)The reference image of resolution 582 x 517 pixels of size 0.56 x 0.56 mm. (b) The floating image of resolution 495 x 364 pixels of size 0.52 x 0.52 mm.



ı) Laws texture features from DRR (b) Laws texture features from EPI

Fig. 6. (a) Laws texture feature extracted from the intensity image of the reference DRR image. Dimensions of the texture images are smaller, since we had to crop the images to get rid of the filtering artifacts at the image borders. (b) Laws texture feature extracted from the intensity image of the floating EPI image.

The images used for the tests were initially aligned as they were used during the radiotherapy practice. The image alignment is achieved with employment of three lasers (sagittal, coronal and axial) for marking the patient reference coordinates/21/. The gold standard-GS registration in our tests was a transformation vector 0 with its tolerance of 3 mm. However, due to the design of our experiments imprecise GS registration did not influence our results.

2.3 Experiment Design

The comparison between the reference evaluation protocol, as implemented by the authors /22/ and three modified versions of the protocol has been performed. The modified versions used regular sampling, Halton and Sobol quasi-random sampling instead of pseudo-random sampling.

First, tests on the intensity images have been conducted. We assume, that images of each image pair are initially aligned. Mutual information (MI) /23,24,5/ has been used to measure similarity between the reference and the transformed floating image at each sampling estimate $X_{n,m}$. MI

was estimated from joint density, which was approximated by using a Parzen kernel. The Parzen kernel was applied to the 2-D joint histograms, quantized into 256 x 256 bins. Each joint histogram was created by plotting an *(i,j)* point for every pair of corresponding pixels in the overlap region, where *i* was the pixel intensity in the reference image, and *j* was the interpolated pixel intensity of the floating image.

The experiments were designed exactly as described in /22/ except for the number of sampling lines *N* and the way those lines were generated. The number of sampling lines has been systematically lowered from the recommended value of 50 in steps of 5: *N*=50, 45, 40, ..., 10, 5, 1. The normalization parameters used to generate sampling lines are listed in Table 1. For the parameter explanation see /8/.

Each sampling line provides a transformation profile of a criterion function. To observe the behavior of criterion functions we checked two parameters: accuracy (*ACC*) and risk of non-convergence (*RON*). Accuracy as it is defined in /8/ is a root mean-square of distances between the maximum value of a criterion function $X_{n,max}$ and origin X_0 on each of the *n* sampling lines; $n=1,2,...N$.

$$ACC = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left|X_{n,\max} - X_0\right|^2} \qquad (5)$$

Risk of non-convergence *RON(r)* is defined as the average of positive gradients $d_{n,m}$ within distance *r* from each of the *N* global maxima:

$$RON(r) = \frac{1}{2rN}\sum_{n=1}^{N}\sum_{m=\max-k}^{\max+k} d_{n,m} \qquad (6)$$

In our tests, the consistency of *ACC* and *RON* values between sampling lines has been compared for four different sampling methods. Furthermore, the consistency check of *ACC* and *RON* values have been performed for reduced numbers of sampling lines.

Moreover, the original and the modified evaluation protocols have been applied to the texture feature images (Figure 6) and the results have been compared. The conveyed texture information between texture images was measured again by MI.

The experiment details are identical as described above. Again, the effect of using the pseudo-random, regular or quasi-random generator was observed, while lowering the number of sampling lines from 50 in steps of 5.

2.4 Texture features used for registration

Apart from intensity images, the evaluation protocols have been tested on Laws texture features, which were extracted from both of the original intensity images. Laws /25/ developed a set of two-dimensional filter masks, which are composed of combinations of several one-dimensional filters /26/.

For the tests we chose Laws texture features extracted by combinations of level-L and spot-S filter masks. Both 1-D filters were of size 20 mm. L-S texture features were summed up with texture features obtained by S-L filter masks /25/, again of size 20 mm.

Each filtered image was subsequently converted to a texture energy image. Texture energy image was obtained by convolving the local texture feature image by a Gaussian averaging window. We used a Gaussian kernel of size 20 mm and cut off frequency at 3σ. Finally, the 2% extreme values of each texture energy image were saturated and the rest were scaled from 0 to 255 integer level yielding 8-bit quantization. See Figure 6.

2.5 Criterion functions

Our criterion functions are computed by measuring the conveyed intensity and texture feature information between images being registered. We choose mutual information (MI) to measure similarity between images. MI can be computed by using the following formula:

Tab.1. Floating image sizes, pixel sizes, translation and rotation units of normalized parametrical space, radius R, number of points along a line M and a step size between two sampling points.

Image set	Image size (mm)		Pixel size (mm)		Unit(mm)	Unit(rad)	R(mm)	M	δ(mm)
	X	Y	X	Y					
01	203	170	0.52	0.52	17.0	0.13	51.0	400	0.26
02	205	179	0.52	0.52	17.9	0.13	53.7	400	0.27
03	258	190	0.52	0.52	19.0	0.12	57.0	400	0.29
04	203	151	0.52	0.52	15.1	0.12	45.3	400	0.23
05	246	140	0.52	0.52	14.0	0.10	42.0	400	0.21
06	194	165	0.52	0.52	16.5	0.13	49.5	400	0.25
07	254	173	0.52	0.52	17.3	0.11	51.9	400	0.26
08	201	162	0.52	0.52	16.2	0.13	48.6	400	0.24
09	206	123	0.52	0.52	12.3	0.10	36.9	400	0.18
10	248	188	0.52	0.52	18.8	0.12	56.4	400	0.28
11	195	107	0.52	0.52	10.7	0.10	32.1	400	0.16

$$MI(A,B) = H(A) + H(B) - H(A,B) \quad (7)$$

with *H(A)* and *H(B)* are the Shannon entropies of image features for both of the images and *H(A,B)* is their joint entropy. Entropy $H(\cdot)$is computed as:

$$H(\cdot) = -\sum_i p(i) \cdot log_2 p(i) \quad (8)$$

where *p* is a probability distribution of features on an image.

3 Results and discussion

3.1 Test on intensity features

The tests are performed on 11 DDR/EPI image pairs (Figure 7, 8) following this protocol:

1. For each of the 11 DRR/EPI image pairs the sampling lines in 3-D transformation space are generated:
   - 400 sampling lines obtained by regular sampling of φ and θ,
   - 400 sampling lines for the reference evaluation protocol are obtained through the web-interface /22/,
   - 400 sampling lines generated by Sobol quasi-random generator, and
   - 400 sampling lines generated by Halton quasi-random generator.
2. The criterion function is evaluated in each of the 400 x 400 sampling points $X_{n,m}$.
3. The evaluation parameters of *ACC* and *RON* are - to consider statistics - calculated for the following consecutive sub-ranges of 400 sampling lines:
   - 8 sub-ranges of 50 sampling lines,
   - 8 sub-ranges of 45 sampling lines,
   - 10 sub-ranges of 40 sampling lines,
   - 11 sub-ranges of 35 sampling lines,
   - 13 sub-ranges of 30 sampling lines,
   - 16 sub-ranges of 25 sampling lines,
   - 20 sub-ranges of 20 sampling lines,
   - 26 sub-ranges of 15 sampling lines,
   - 40 sub-ranges of 10 sampling lines, and
   - 80 sub-ranges of 5 sampling lines.

The results are depicted as scatter of *ACC* and *RON* values. The scatter is computed as normalized standard deviation of the consecutive subranges of sampling lines. We expect that better random generator would yield lower scatter of the results. Lower scatter means more consistent results which is the aim of our tests. For purposes of clarity, results are shown only for *N*=50,40,30,20 and 10.

The paired Student's t-test (p<0.05) which compares the scatter of *ACC* values in Figure 7 of the four sampling methods indicated no significant difference between pseudo-random and Sobol-quasi random sampling at any of different numbers of sampling lines. On the other hand, there exists significant difference between results of pseudo-random and Halton-quasi random sampling when the analysis is performed on *N*=40,30,20 and 10 sampling lines. In

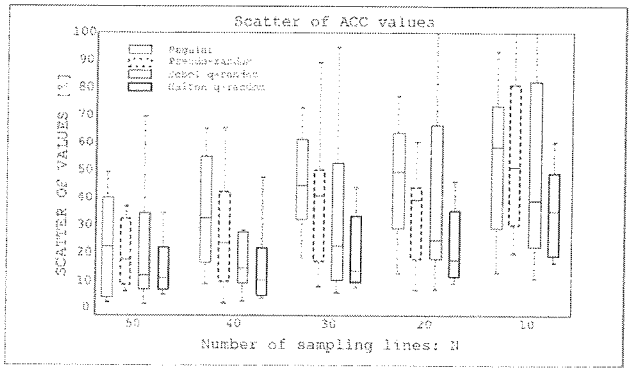


Fig.7. Results for intensity features. Box-and-whisker plots are showing scatter of values of 11 DRR/EPI image pairs for the parameter ACC *and four sampling methods: regular, pseudo-random, Sobol quasi-random and Halton quasi-random, respectively.*

these cases the pseudo-random generator based results show significant higher scatter of the values in comparison to Halton quasi-random sampling based results. However, there exists no significant difference between results of the two generators when the analysis is performed on *N*=50 and *N*=5 sampling lines. 50 sampling lines seems to be enough in 3-D transformation space to overcome a deficiency of pseudo-random generator, but 5 sampling lines are too few to achieve satisfactory consistence even with Halton quasi-random sampling.

The comparison between regular and pseudo-random sampling shows significantly higher scatter of *ACC* values for *N*=20 when the regular sampling method is employed.

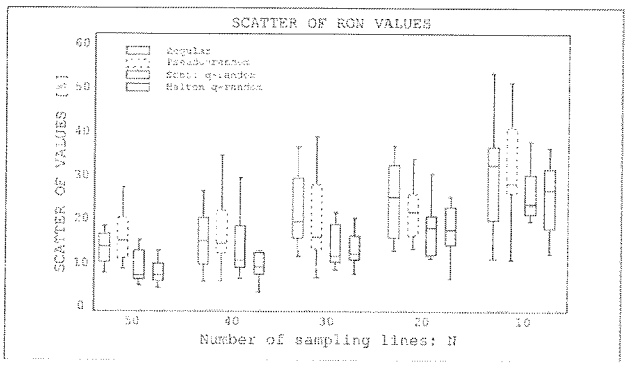


Fig.8. Results for intensity features. Box-and-whisker plots are showing scatter of values of 11 DRR/EPI image pairs for the parameter RON *and four sampling methods: regular, pseudo-random, Sobol quasi-random and Halton quasi-random, respectively.*

Similar to the scatter of *ACC* values, the paired Student's t-test (p<0.05) indicated no significant difference between pseudo-random and Sobol-quasi random sampling based scatter of *RON* values in Figure 8. However, Halton quasi-random sampling yielded significantly lower scatter of *RON*

values for all but N=5 sampling lines in comparison to pseudo-random sampling. Again, the reason is that 5 sampling lines are too few to cover 3-D transformation space dense enough with any of the generators.

The regular sampling shows no significant difference in the scatter of *RON* values compared to pseudo-random based results for any of N sampling lines.

3.2 Tests on texture energy images

The tests are performed on 11 texture image pairs derived from original intensity images (Figure 9, 10). The experiment protocol is the same as the one used for intensity features.

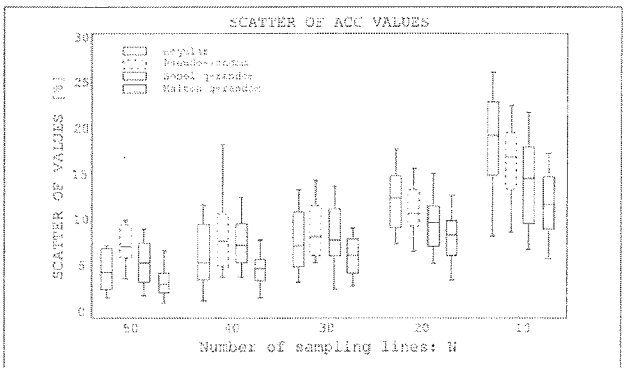


Fig. 9. Results for texture features. Box-and-whisker plots are showing scatter of values of 11 DRR/EPI image pairs for the parameter ACC *and four sampling methods: regular, pseudo-random, Sobol quasi-random and Halton quasi-random, respectively.*

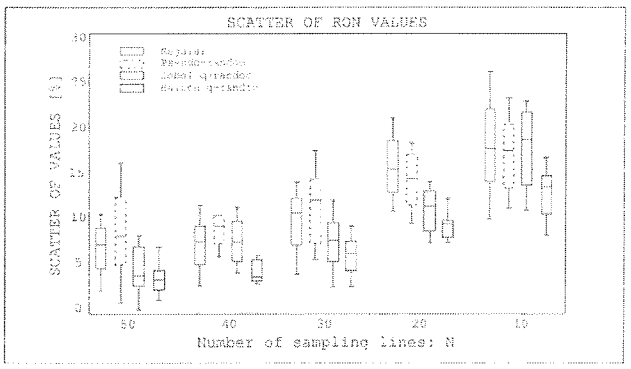


Fig. 10. Results for texture features. Box-and-whisker plots are showing scatter of values of 11 DRR/EPI image pairs for the parameter RON *and four sampling methods: regular, pseudo-random, Sobol quasi-random and Halton quasi-random, respectively.*

Note, that overall, the scatter values are much lower for texture features in comparison to intensity features. This is a strong argument to use texture features instead of intensities for this registration task. However, from the tests performed on the texture feature images similar conclusions may be drawn than from the results for intensity features. Again, the Halton quasi-random sampling outperformed both regular and pseudo-random sampling as it yielded more consistent results for both, *ACC* and *RON* values. Student t-test indicates significant difference between samplings for all systematically reduced number of sampling lines, except N=30 - *ACC* values and N=5 - *RON* values.

The recommended number of sampling lines N which would yield enough consistent results for our 3-D optimization task was initially not provided. Since our dimensionality was significantly lower than in the original work by Škerl et al, we started our tests with N=50 as reasonably safe margin. From the results of our tests we can hypothesize that by using Halton quasi-random generator the smallest number of sampling lines N, which would yield enough consistent results for 3-D optimization task, is as low as 10. This hypothesis is confirmed by paired Student t-test (p<0.05) which compared scatter results between pseudo-random N=50 sampling lines and systematically reduced N from 50 to 5 for Halton quasi-random generator. t-test indicated that for N=50 and N=40 Halton quasi-random delivered significantly lower scatter values in comparison to pseudo-random sampling with N=50. From N=35 to N=10 Halton quasi-random sampling indicated no significant difference to pseudo-random sampling with N=50. For N=5 Halton quasi-random sampling delivered significantly higher scatters in comparison to pseudo-random with N=50.

For our registration task the recommended pseudo-random sampling with 50 sampling lines yields comparable scatter of results to the Halton-random sampling with 10 sampling lines.

4 Conclusion

The evaluation protocol described in /8/ assesses the quality of similarity measure used in a specific registration problem prior to registration. This is done by evaluating the behaviour of a similarity measure for simulated transformations. The evaluation of a similarity measure includes the following parameters: accuracy, robustness and capture range.

We are using this evaluation protocol for assessment of criterion functions based on the intensity and a bank of texture features – such use requires the evaluation of many criterion functions. It is therefore important that the evaluation protocol is as efficient as possible, while retaining the truthfulness of the results.

The obtained results from our tests show that Halton quasi-random outperformed regular, pseudo-random and Sobol quasi-random sampling for our specific registration task. Additionally, the tests have shown, that if the computational time is of prime concern, the number of sampling

lines may be reduced, which yields a significant reduction of computation time spent on evaluation of one CF. Thus, a larger number of CFs may be evaluated to select those, that would provide the best registration.

Additionally, we can also see that values *ACC* and *RON* based on texture features deliver considerably lower scatters than values based on intensity features. It may be concluded that a lower scatter of results may be one of the additional parameters to assess the quality of a criterion function. The lower the scatter of the results the more representative are the criterion functions defined on each of the sampling lines. Moreover, it is less likely that a criterion function is ill-defined by containing a false global optimum and strong local maxima /27/.

5 Acknowledgement

The authors would like to thank the Institute of Oncology Ljubljana for the provision of the image data set and to the Slovenian Ministry of Higher Education, Science and Technology for the financial support, under grant 3211-05-000557.

6 References

/1/ P. A. Van den Elsen, J. B. A. Maintz, E. J. D. Pol, and M. A. Viergever. Automatic registration of ct and mr brain images using correlation of geometrical features. *IEEE Trans. Med. Imag*, 14(2), 1995.

/2/ A. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1-36, 1998.

/3/ H. Lester and S. R. Arridge. A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32:129-149, 1999.

/4/ D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Phys.Med.Biol.*, 46:1-45, 2001.

/5/ W. Pluim, J. P., A. Maintz, and M. A. Viergever. Mutual information based registration of medical images: a survery. *IEEE Trans. Med. Imag.*, 2003.

/6/ B. Zitova and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977-1000, 2003.

/7/ A. Gholipour, N. Kehtarnavaz, R. Briggs, M. Devous, and K. Gopinath. Brain functional localization: A survey of image registration techniques. *IEEE Trans. Med. Imag.*, 26(4):427-451, 2007.

/8/ D. Škerl, B. Likar, and F. Pernuš. A protocol for evaluation of similarity measures for rigid registration. *IEEE Trans. Med. Imag.*, 25(6), 2006.

/9/ D. Škerl, D. Tomaževič, B. Likar, and F. Pernuš. Evaluation of similarity measures for reconstruction-based registration in image-guided radiotherapy and surgery. *Int. J. Radiation Oncology Biol. Phys.*, 65(3):943-953, 2006.

/10/ D. Škerl, B. Likar, J. M. Fitzpatrick, and F. Pernuš. Comparative evaluation of similarity measures for the rigid registration of multi-modal head images. *Phys. Med. Biol.*, (52):5587-5601, 2007.

/11/ H. W. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University press, reprinted in 1996.

/12/ I. M. Sobol. The distribution of points in a cube and the approximate evaluation of integrals. 7(4):86-112, 1967.

/13/ J. H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2(1):84-90, 1960.

/14/ A. Jarc, P. Rogelj, and S. Kovačič. Texture feature based image registration. In *Symposium proceedings*, pages 17-20, Arlington Virginia, U.S.A., April 2007. IEEE ISBI.

/15/ E. W. Weisstein. Spherical coordinates, from mathworld-a wolfram web resource. *http://mathworld.wolfram.com/SphericalCoordinates.html*, accessed on 24.10.2007.

/16/ E. W. Weisstein. Sphere point picking, from mathworld-a wolfram web resource. *http://mathworld.wolfram.com/SpherePointPicking.html*, accessed on 24.10.2007.

/17/ G. Matsopoulos, P. Asvestas, K. Delibasis, V. Kouloulias, N. Uzunoglu, P. Karaiskos, and P. Sandilos. Registration of electronic portal images for patient set-up verification. *Phys. Med. Biol.*, 49:3279-3289, 2004.

/18/ L. Dong and A. Boyer. An image correlation procedure for digitally reconstructed radiographs and electronic portal images. *Int. J. Radiation Oncology Biol. Phys.*, 33(5):1053-1060, 1995.

/19/ A. Khamene, P. Bloch, W. Wein, M. Svatos, and F. Sauer. Automatic registration of portal images and volumetric ct for patient positioning in radiation therapy. *Med. Image Anal.*, 10:96-112, 2006.

/20/ A. Jarc, P. Rogelj, and S. Kovačič. Analysis of texture features for registration of *drr* and *epi* images. *International Journal of Computer Assisted Radiology and Surgery*, 2(Supplement 1):116-118, 2007.

/21/ O. Ureten and Erkal H. S. Measurement of patient setup errors using digitally reconstructed radiographs and electronic portal images. pages 88-90. 2nd International Biomedical Engineering Days, 1998.

/22/ D. Škerl, B. Likar, and F. Pernuš. An online protocol for evaluation of similarity measures. *www.lit.fe.uni-lj.si/Evaluation*, 2006.

/23/ F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imag.*, 16(2), 1997.

/24/ P. Viola and W. M. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137-154, 1997.

/25/ K. Laws. Rapid texture identification. pages 376-380. SPIE Image Processing for Missle Guidance, 1980.

/26/ M. Petrou and P. G. Sevilla. *Image Processing Dealing with Texture*. John Willey and Sons, 2006.

/27/ J. Pluim, Maintz J. B. A., and M. Viergever. Image registration by maximization of combined mutual information and gradient information. *IEEE Trans. Med. Imag.*, 19(8), 2000.

Andreja Jarc, M.Sc.
Sipronika d.o.o.
Tržaška 2, 1000 Ljubljana
andreja.jarc@sipronika.si

Dr. Janez Perš,
Dr. Peter Rogelj,
Prof. Dr. Stanislav Kovačič,
University of Ljubljana, Faculty of Electrical Engineering,
Tržaška 25, 1000 Ljubljana, Slovenia

Prispelo (Arrived): 16.06.2008 *Sprejeto (Accepted): 19.03.2009*

A NEW APPROACH TO THE MODELING OF NETWORK TRAFFIC IN SIMULATIONS

Matjaž Fras, Jože Mohorko, Žarko Čučej

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

Key words: self-similar, network traffic modeling, Pareto distribution, maximal transmission unit

Abstract: Simulations of telecommunications networks have become very important tools for their evaluation. A very important influence in simulations has network traffic. This paper introduces new concepts for the modeling of measured network traffic in simulation tools. With these new concepts, we can improve descriptions of the random packet-size process, especially for maximal-packets of network traffic, which have a very great impact on the bit or packet rates of network traffic. The suggested methods improve the contents of packets, especially maximal packets in modeled network traffic simulations, which leads to smaller differences in bit and packet-rates between measured and modeled network traffics.

Nov pristop k modeliranju samo-podobnega prometa v simulacijah

Kjučne besede: samo-podobnost, modeliranje omrežnega prometa, Pareto porazdelitev, maksimalna dolžina paketa

Izvleček: Simulacije telekomunikacijskih omrežij postajajo pomembno orodje za ovrednotenje le teh. Zelo pomemben in velik vpliv v simulacijah ima tudi omrežni promet. Ta članek predstavlja novi koncept modeliranja izmerjenega omrežnega prometa v simulacijskih orodjih. Z tem novim konceptom lahko izboljšamo opis naključnega procesa velikosti paketov omrežnega prometa, zlasti maksimalnih paketov, kateri imajo zelo velik vpliv na srednjo vrednost celotnega prometa v bitih in paketih na časovno enoto. Predlagane metode izboljšajo opis vsebnosti paketov v omrežnem prometu, še posebej maksimalnih paketov v modeliranem prometu,kar posledično vodi do manjših razlik v srednji vrednosti bitov in paketov na časovno enoto med izmerjenim in modeliranim prometom.

1. Introduction

Statistical analysis in Ethernet networks show that, in many cases, network traffic can be described by self-similarity /1/. This model appeared before fifteen years as an alternative, at that time, to the used models such as Poisson and Markov /2/. It was also shown, that heavy-tailed distributions, such as Pareto and Weibull, are more suitable for describing network processes, such as process packet-size and inter-arrival time /1, 3, 4, 5/.

One of the main goals of researchers was, and still is, the modeling of network traffic in simulations, such as OPNET /6, 7, 8/. In simulation we try to model the measured network traffic, which is the best possible approximation of the measured traffic in the sense of bit or packet-rates, bursts or variance. For evaluating discrepancies between measured and simulated network traffic, we chose different measures such as bit or packet rates, Hurst parameter, variance and also discrepancy between histograms of statistical network process for packet size and inter-arrival time.

During measuring and modeling we saw that discrepancies between measured and modeled traffic are derived from an inaccurate description of the packet-size process. We also saw that, especially for longer and maximal length packets (MTU- Maximal Transmission Unit), have a substantial influence on modeled network traffic. The captured histogram of the packet-size process had great discrepancy in regard to the measured histogram and chosen distribution, which is usually a consequence of maximal-packets. Maximal packets are a consequence of data fragmentation in TCP/IP stack. Usually with classical modeling, where a captured histogram of packet size process is described with distribution, we do not derive at a good enough description regarding the packet-size process of measured traffic, especially the content of maximal packets. This, consequently, leads to great discrepancy between measured and modeled network traffics, especially in bit and packet-rates, and also traffic bursts.

For this reason, we present three methods for describing a measured traffic histogram of packet-size which achieve more accurate descriptions of network traffic in simulations.

1. The first method is based on using "mixed distributions" for describing random processes, a similar concept is used in the area of image processing /9/, and already steps in the area of traffic modeling /10, 11, 12/.
2. The second method is based on estimating data files of a measured traffic histogram by defragmentation in a communications network /3, 4, 5/.
3. The third method combines the first and second methods.

This paper is organized as follows. The second section describes statistical modeling of network traffic by distribution and Hurst parameter. The next section describes

the packet-size process of network traffic. New approaches with suggested methods are in the forth section. The fifth section represents the simulation results. Finally, we finish this paper with the conclusion.

2 Statistical modeling of measured network traffic

Network traffic can be described as a combination of two random processes:

1. packet-size process $X(t)$
2. inter-arrival time $Y(t)$

Lets describe network traffic Z(t) as

$$Z(t) = \psi(X(t), Y(t)) \quad (1)$$

where ψ is the function of packet-size $X(t)$ and inter-arrival time process $Y(t)$. Both processes are described by probability distribution function (pdf). The choice of suitable distribution for a traffic process depend the measured network traffic's properties. For network traffic with a short-range dependence property, light-tailed distributions (exponential) are the more suitable for describing packet-size process, such as exponential. In the case of network traffic with long-range dependence, heavy tailed distributions are the more suitable distributions for describing such traffic, such as Pareto and Weibull. The probability density function (pdf) of Pareto distribution is

$$p(x) = \alpha k^{\alpha} \cdot x^{-\alpha-1}, \quad k \le x, \quad \alpha, k > 0 \quad (2)$$

where k is local parameter and α is shape parameter. Probability density function of Weibull distribution is:

$$p(x) = \frac{\alpha}{k} \cdot \left(\frac{x}{k}\right)^{\alpha-1} \cdot e^{-\left(\frac{x}{k}\right)}, \quad x \ge 0, \quad \alpha, k > 0 \quad (3)$$

where k is local parameter and α is shape parameter.

Definition of the self-similar random process /3, 13, 14, 15/ is based on autocorrelation function $r(k)$, which is described as

$$r(k) \approx k^{-\beta} L_1(k), \quad k \to \infty, \quad 0 < \beta < 1, \quad (4)$$

where $L_1(k)$ is slowly varying at infinity, that is for all $x > 0$ (i.e., $L_1(t)$ = constant, $L_1(t) = log(t)$). Hurst parameter H is used for described arrival process and it is defined by

$$H = 1 - \frac{\beta}{2}, \qquad 0 < \beta < 1 \quad (5)$$

and presents the measure of self-similarity. For describing arrival process, beside parameter H, are also needed parameters such are average arrival-rate, fractal onset time scale, source activity-ratio, and peak to mean ratio.

3 Problem of statistical packet size process

From measured traffic by sniffer /8/, we can obtain information about a packet-sizes, inter-arrival time, packet-rate... Based on histograms, we can evaluate both random traffic process $X(t)$ in $Y(t)$ and choose distributions, which are the best approximations of histograms. During research, where we estimate parameters of traffic processes we found that, in the case of estimating packet-size process parameters much larger discrepancies appear than in the case of inter-arrival time. Discrepancy between the histogram of measured traffic and distribution, which describe this process, can be evaluated by goodness of fit tests, such as Kolmogorov-Smirnov or Chi-square /16/. The greatest impact on these discrepancies is MTU, which as mentioned in the first section. MTU packets cause a strong discontinuity in the histogram and it is very difficult to describe such a histogram using the classical method. In our research, we paid attention to a statistical description the packet-size process of network traffic.

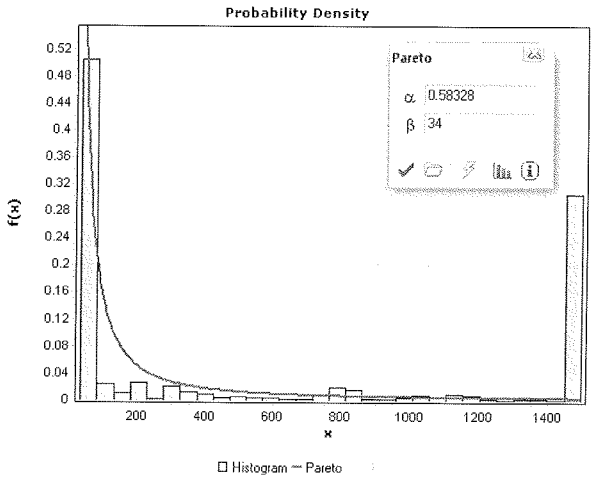


Fig. 1: Histogram of measured packet size process and distribution parameters estimation with classical method with EasyFit fitting tool.

Figure 1 shows an example of a packet-size histogram of measured network traffic and classical distribution parameters' estimation. From the captured histogram, we can see that minimal length size packets of around 54 B prevail. But there are also a lot of packets of maximal length, which also have a great influence on the bit-rate of the entire network traffic. The classical parameter estimation method (Figure 1) does not describe the process very well, especially those maximal packets, which usually lead to great discrepancies between measured and modeled traffic, in the sense of bit or packet rates. Such an estimation method also has very big difference between the contents of packets between measured and simulated traffic. We cannot solve this problem by using other methods for estimating distribution parameters for the packet-size process of network traffic, such as the CCDF method /3/.

The greatest discrepancies appear when describing network traffic with long-range dependence (LRD) property, where heavy tailed-distribution is used, such as Pareto. Smaller discrepancies also appear in the case of describing network traffic with short-range dependence (SRD), where exponential distribution was used, but these discrepancies are smaller than in the previous case.

3 Suggested methods for estimating distribution for packet-size process

All suggested methods are based on the transformation of captured-traffic. The first method is based on using mixed (multiple) distributions to the describe packet-size process, the second method is based on defragmentation of captured-packets and the third method is a combination of the first and second methods.

3.1 Mixed distributions

Using this method, we will describe network-traffic by multiple-distributions, which will be implemented using multiple traffic generators in the same simulation workstation. By using mixed distributions for describing the stochastic process of network traffic, we will achieve a smaller discrepancy between the measured histogram and the fitted distributions for packet-size process (Figure 2). Network traffic $Z(t)$ defined in (1) can be described as the sum or n-th data sources:

$$\begin{aligned} &Z(t) = Z_1(t) + Z_2(t), \ldots Z_n(t) \\ &Z(t) = \psi_1(X_1(t), Y_1(t)) + \ldots + \psi_n(X_n(t), Y_n(t)) = \\ &Z(t) = \sum_1^n Z_i(t) = \sum_1^n \psi_i(X_i(t), Y_i(t)) \end{aligned} \quad (6)$$

where $Z_i(t)$ is traffic for each traffic generator and ψ_i is a function of two random processes $X_i(t)$ and $Y_i(t)$, where $X_i(t)$ represent packet-size process and $Y_i(t)$ inter-arrival time. So, we can divide network traffic into separated segments modeled by different distributions. Points which separate the packet size process in multiple parties described by independent distribution, are threshold points. The simplest way to separate network-traffic for mixture distribution is to define the first traffic $Z_1(t)$, where are packets, which are longer than the threshold value, and another traffic $Z_2(t)$, with packets that are shorter than the threshold. In many cases, MTU size represents the threshold point.

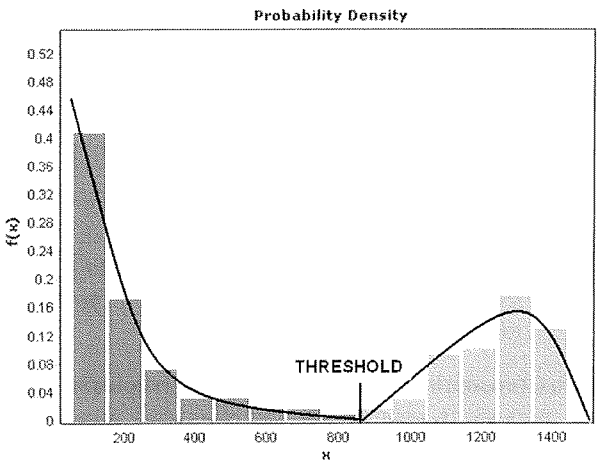


Fig. 2: Example of using two distributions for describing packet size process of captured network traffic.

$$\begin{aligned} &Z(t) = Z_1(t) + Z_2(t) = \\ &= \begin{cases} Z_1(t) = \psi_1(X_1(t), Y_1(t)); & packet\_size > threshold \\ Z_2(t) = \psi_2(X_2(t), Y_2(t)); & packet\_size \le threshold \end{cases} \end{aligned} \quad (7)$$

We must also estimate the belonging distributions for both inter-arrival time processes $Y_1(t)$ and $Y_2(t)$ and packet-size processes $X_1(t)$ and $X_2(t)$.

3.2 Defragmentation method

Whilst transmitting files across a network, IP packets are fragmented because of MTU limitations. The fragmentation process is executed in a model of IP encapsulation in TCP/IP stack. From the captured traffic in Figure 1, we can see that MTU packets impact on the discontinuity in the histogram, this causing the common distribution descriptions, with the help of the classical method. This new method is based on histogram estimation of the transmitted data file before fragmentation /4/. For a distribution estimation of the packet-size process we execute with the addition of maximal packets, which are fragmented in the fragmentation process during transmission. So, we combine all packets from a sequence of MTU packets, including the first packet shorter than the maximal size, from the same source in the new bigger packet. These newly derived at values, together with captured non-fragmented packets, are used designating the histogram of data, which will be described by new distribution.

$$\begin{aligned} &Z(t) = \psi(X(t), Y(t)) \rightarrow Z_T(t) = \psi(X_T(t), Y_T(t)) \\ &Z(t) \approx Z_f(t) \end{aligned} \quad (8)$$

$Z_T(t)$ represents the transformed traffic, which is a function of the transformed processes for packet-size X_T and inter-arrival time Y_T. The transformed histogram represents the originally transmitted files $Z_f(t)$. We spray the distribution of maximal packets in the captured histogram over a new range, using the defragmentation method, which represents the transmitted files. This method leads to more continuous histograms, such as in Figure 3, which can be described by the classical method more precisely using distribution, than the histogram in Figure 1. Estimation parameters of file sizes are used in traffic generators during simulations. Because of the limitation of MTU, which is a defined in model of a communication device, the files are fragmented into maximal packets during the simulation run. So estimate traffic is a good approximation of captured traffic.

3.3 Combination of distributions and defragmentation

The third method is based on a combination of mixed distributions and defragmentation methods. The basic idea is to describe captured traffic with two or more distributions, but for captured-traffic we can also execute the defragmentation process. For example, we can execute the defragmentation process of captured-traffic $Z(t)$, and then describe with one distribution $X_1(t)$ traffic of packets $Z_1(t)$,

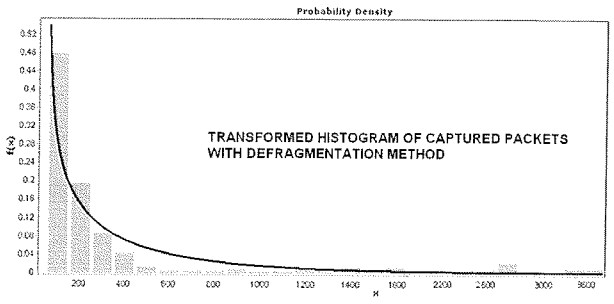


Fig. 3: Transformed histogram of captured histogram on Figure 1 with chosen distribution.

which are shorter than the maximal packets. With the second distribution $X_2(t)$, we can describe the traffic of the fragmentation packets $Z_2(t)$, which was equal to the maximal values before fragmentation.

$$Z(t) = Z_1(t) + Z_2(t) = \begin{cases} Z_1(t) = \psi_1(X_1(t), Y_1(t)); & packet\_size \neq threshold \\ Z_1(t) = \psi_2(X_2(t), Y_2(t)); & defragmentaion\_packets \end{cases} \quad (9)$$

For both processes $X_1(t)$ and $X_2(t)$ we also define and estimate distributions and their parameters for the belonging processes of inter-arrival time $Y_1(t)$, $Y_2(t)$, and also Hurst parameter, which can also be used in the modeling of arrival process.

4. Simulation results

We model the captured self-similar network traffic, which is shown on Figure 4, with short-range dependence by simulations with both classical and presented methods.

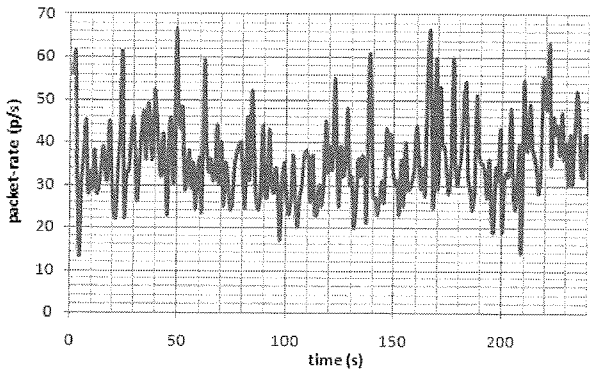


Fig. 4: Measured test network traffic captured by Wireshark sniffer.

In the case of classical estimation, we chose exponential distribution for describing the packet-size process, because the value of Hurst parameter is near 0.5 and also has short-range dependence. For the first and third methods we define threshold, which is equal to MTU size, because the bin with MTU packets is withdrawing from other neighboring local bins in the packet-size histogram (Figure 1). This bin is described by separated distribution, for the first and third methods. Table 1 shows parameters of measured network traffic and all estimated parameters for presented methods, which was used in OPNET simulations tool.

Table 1: Parameters of measured and simulated network traffic

	packet size process	inter-arrival time	p/s	kb/s	H	MSE
measured traffic	X	X	35.6	114.5	0.58	X
classical method	exponential $1/\lambda = 416{,}5$	Weibull $\alpha = 0.57326$ $\beta = 0.01895$	33.4	113.4	0.53	0.024
1. method	packets < MTU		34.1	124.9	0.54	0.016
	exponential $1/\lambda = 230.41$	Weibull $\alpha = 0.65792$ $\beta = 0.02587$				
	packets = MTU					
	constant 1482	Rayleigh $\sigma = 0.17435$				
2. method	exponential $1/\lambda = 452{,}48$	Weibull $\alpha = 0.6521$ $\beta = 0.0244$	31.0	114.5	0.52	0.026
3. method	packets < MTU		37.2	120.0	0.57	0.003
	exponential $1/\lambda = 106.7$	Weibull $\alpha = 0.677$ $\beta = 0.02932$				
	defragmentation data					
	Rayleigh $\sigma = 2181.7$	Rayleigh $\sigma = 0.1871$				

Table 1 shows the comparison between measured and modeled signals in bit and packet-rates, without method and suggested methods. There are also estimated parameters H, which are measure of a self-similarity. There are also mean square errors (MSE) between the measured and modeled histograms of the packet-size process, which also show the contents of the packets, shown in Figure 5. Using this test, we proved that presented methods impact the minimal discrepancy between measured and modeled signals and better describe measured traffic than classical estimation (without method).

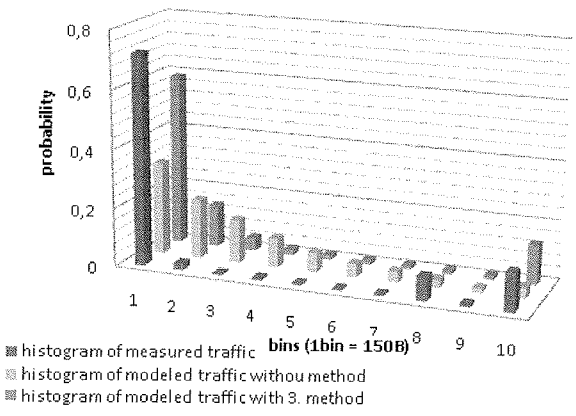


Fig. 5: Histograms of packets size process of measured traffic and modeled traffic with classical and 3.method

Figure 6 present the three simulated network traffics, which were modeled by estimated parameters from Table 1.

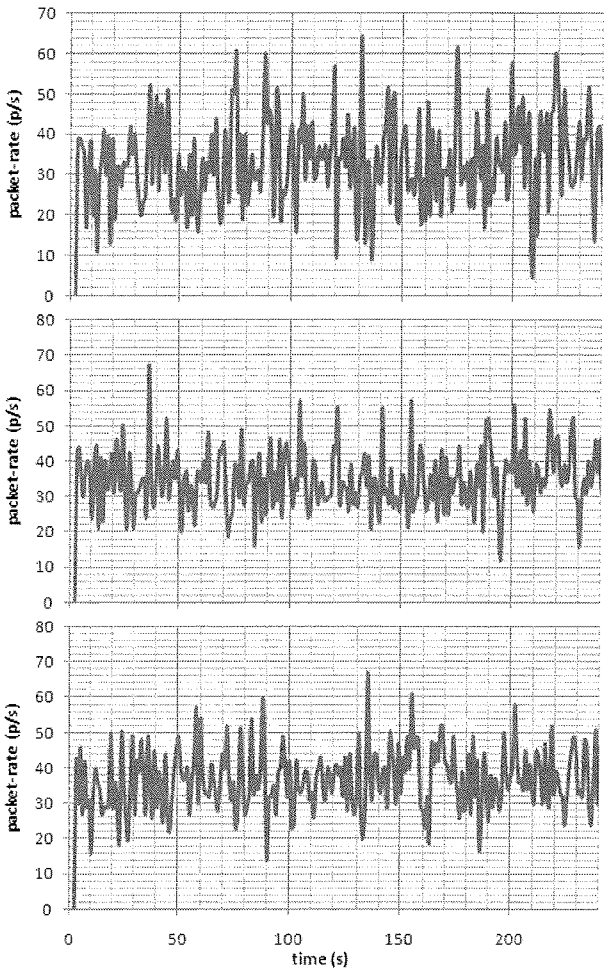


Fig. 6: Simulated network traffics in OPNET simulation tool. First graph presents simulated traffic, which was model by first method. Second graph presents simulated traffic, which was model by second method. Third graph presents simulated traffic, which was model by third method.

5. Conclusion

The presented methods show very good results in the case of modeling network traffic with short-range dependence, where we achieved better contents of packets, sometimes even better bit or packet-rates in the modeled traffic and a more accurate description of captured-traffic, then in the case of using classical manner of modeling the measured traffic. For future research we plan modeled network traffic with long-range dependence with purposeful methods, because in these cases classical estimation (without any methods) totally failed and lead to great discrepancy between measured and modeled traffic in the sense of bit and packet-rates, and also in bursts' intensities.

6. References

/1/ W. E.Leland, M. S. Taqqu, W. Willinger in D. V. Wilson, "On the self-similar nature of Ethernet traffic (Extended version)", IEEE/ACM Transactions on Networking, Vol.2, pp.1-15, 1994.

/2/ V. Paxon in S. Floyd, "Wide area traffic: the failure of Poisson modeling", IEEE/ACM Transactions on Networking, 3(3): 226-244, 1995.

/3/ K. Park in W. Willinger, "Self-Similar Network Traffic and Performance Evaluation", John Wiley & Sons, 2000.

/4/ K. Park, G. Kim in M. E. Crovella, "On the Relationship Between File Sizes Transport Protocols, and Self-Similar Network Traffic, International Conference on Network Protocols", 171-180, 1996.

/5/ W. Willinger, M. S. Taqqu, R. Sherman in D. V. Wilson, "Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level", IEEE/ACM Transactions on Networking, 5(1): 71-86, 1997.

/6/ F. Xue in S. J. Ben You, "The effect of aggregation on self-similar traffic", Department of Electrical and computing engineering, University of California.

/7/ J. Potemans, B. Van den Broeck, Y. Guan, J. Theunis, E. Van Lil in A. Van de Capelle, "Implementation of an Advanced Traffic Model in OPNET Modeler", OPNETWORK 2003, Washington D.C., USA, 2003.

/8/ M. Fras, J. Mohorko and Z.Cucej FRAS, "Estimating the parameters of measured self similar traffic for modeling in OPNET", IWSSIP Conference, 27.-30 June 2007, Maribor, Slovenia.

/9/ D. Persson, T. Eriksson, P. Hedelin, P," Packet Video Error Concealment With Gaussian Mixture Models", IEEE Transactions on Image Processing, Vol. 17, Issue 2, Feb. 2008 Page(s):145 - 154.

/10/ S. Luo, G. A. Marin, "Realistic internet traffic simulation through mixture modeling and a case study", Proceedings of the 37th conference on Winter simulation, 2005.

/11/ V. Karamcheti, D. Geiger, Z. Kedem and S. Muthukrishnan, "Detecting malicious network traffic using inverse distributions of packet contents", Sigcomm 2005, Philadelphia, USA.

/12/ A.Thümmler, P.r Buchholz and M. Telek, "A Novel Approach for Fitting Probability Distributions to Real Trace Data with the EM Algorithm", Dependable Systems and Networks, 2005.

/13/ H. Yölmaz, "IP over DVB: Management of self-similarity", Master of Science, Bođaziçi University, 2002.

/14/ M. Z. Jiang, "Analysis of wireless data network traffic", Master of Applied Science, Simon Fraser University, Vancouver, Canada, 2000.

/15/ O. Sheluhin, S. Smolskiy and A. Osin, "Self-Similar Processes in Telecommunications", John Wiley & Sons, 2007.

/16/ Chakravarti, Laha, and Roy, "Handbook of Methods of Applied Statistics", Volume I, John Wiley and Sons, pp. 392-394, 1967.

Matjaž Fras, Jože Mohorko, Žarko Čučej
Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko
Smetanova 17, 2000 Maribor, Slovenija
Epošta: matjaz.fras1@uni-mb.si

Prispelo (Arrived): 14.08.2008 *Sprejeto (Accepted): 19.03.2009*

EKSPERTNI SISTEM ZA ANALIZO REZULTATOV SIMULACIJ TAKTIČNIH RADIJSKIH OMREŽIJ

Saša Klampfer, Jože Mohorko, Žarko Čučej

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

Kjučne besede: ekspertna analiza, ekspertni sistem, simulacija, OPNET modeler, statistike, komunikacijske enote, uspešnost prenosa sporočil, prenosni kanal, izkoriščenost kanala, zakasnitve, radijska vidljivost, baza znanja, uporabniški vmesnik, pravila, sklepanje, mehke množice.

Izvleček: Članek opisuje metode analize in zasnovo sistema za avtomatizirano analizo simulacijskih rezultatov pridobljenih iz orodja za simulacije komunikacijskih omrežij - OPNET modelerja. Pri tem smo se omejili na temeljne gradnike, ki jih sistem mora vsebovati, da ga lahko uvrstimo v razred ekspertnih sistemov. V prvem delu članka smo za vsak gradnik podali temeljit opis. Dodali smo tudi opise mehanizmov mehkih množic, baze znanja in pravil, ki se uporabljajo v procedurah sklepanja in iskanja končnih rešitev. V drugem delu članka predstavljamo ekspertni sistem, ki smo ga zasnovali za analizo rezultatov simulacij taktičnih radijskih omrežij. V nadaljevanju je opisana funkcionalnost sistema, metode analize posameznih komunikacijskih parametrov in načini podajanja rezultatov.

Expert system for analysis of tactical radio networks simulations

Key words: expert analysis, expert system, simulation, OPNET, statistics, communication units, message completion rate, transfer channel, channel utilization, delay, radio visibility, knowledge base, user interface, rules, decision procedure, fuzzy sets.

Abstract: Now days the use of simulation tools rapidly increases because they are quite precise and gives satisfactory results, which are comparable with measured results on real communication networks. Most of available simulation tools have possibilities to present final results in graphical form. Such approach is suitable in cases, where expert user operate with small number of graphs and simulated scenarios. In cases where must user compare more than two individual graphs, analysis became complex and time voracious. We can decide that manual analysis of graphical results takes-up a lot of precious time, especially in cases of simultaneously analysis of multiple graphs. This time is economized by our system.

Expert system (ES in the following explanation) is defined as intelligent computer program which includes certain level of expert knowledge. Such knowledge is stored in knowledge data base. Knowledge data base quality is one of the most important factors for such systems. It is some kind of function concerning data base dimensions and knowledge quality. A wide-spread base with high expert knowledge leads to a high-performance expert system. Knowledge must be stored into the base in the right format, because an expert system has to understand this knowledge, to create right decisions based on that knowledge. Detailed description of knowledge base is given in section 2.1. Some of the most important parts of ES are user interface, reasoning mechanism and fuzzy set for sorting and arranging simulation values into proper classes.

Reasoning mechanism is one of the main parts of an expert system. It controls the operation of the whole ES. The mechanism must actively use the knowledge base for dealing with data, coming into the system, and for the derivation of suitable facts. The mechanism is composed of inquiring and reasoning processes, which helps in the solution search process. The most useful reasoning methods in the cases when we want to derive knowledge using production rules are: forward reasoning and backwards reasoning. Detailed description of reasoning mechanism is given in section 2.2.

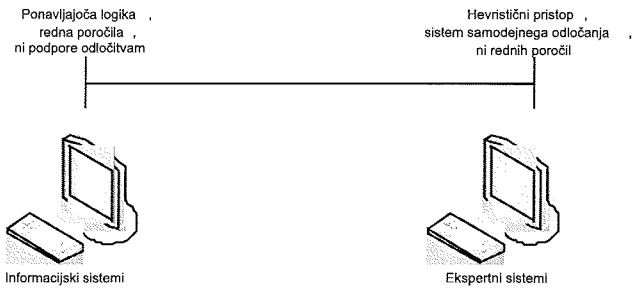
Fuzzy sets as main part of ES are generalization of regular crisp sets /1, 2/. Meanwhile, the appurtenance function of a crisp set has a stock value {0, 1} (specific element belongs or does not belong to this set) the appurtenance function of a fuzzy set (μ_A) has a stock value within the interval [0, 1]. We can reason, that a specific element in fuzzy set is contained by appurtenance, which is ε[0, 1].

For example we observe OPNET simulation graph for tactical radio received power data. For this data we define set A={x; data in x is acceptable}. Such a set contains all acceptable data. If we look at this set as an ordinary binary set, we can specify that data fully belongs to it or even does not fully belongs to it (two possibilities). A problem appears about 'acceptability' definition. In regular sets, passages between appurtenance and nonappurtenance are sharp (discrete). Passages between appurtenance and nonappurtenance in fuzzy sets are soft, slow and continuous (section 2.3). Fuzzy set is in our case used for analyzing results from transmitter utilization statistics and packet delay statistics, meanwhile radio visibility and message completion rate uses approach based on direct comparison between sent packets and received packets between communication units. Section 2.8 describes expert system user interface and manners to present results. Section four concludes the paper.

1. Uvod

Ekspertni sistem lahko opredelimo kot inteligentni računalniški program, ki uporablja znanje in procedure sklepanja za reševanje problemov. Vse definicije ekspertnega sistema v literaturi so si enotne, da vsebuje ozko specializirano znanje iz določenega strokovnega področja (problemske oz. raziskovalne domene) in da je le ta zmožen znotraj tega strokovnega področja oblikovati inteligentne odločitve. Posnema torej delovanje izvedenca ali eksperta za to strokovno področje in njegovo sposobnost analiziranja, reševanja in utemeljevanja odločitev znotraj problemske domene. Ekspertni sistemi tako niso splošni reševalci problemov širokega področja, temveč so namenjeni reševanju zaključenih, dobro definiranih problemov.

Literatura ekspertne sisteme uvršča med sisteme, ki temeljijo na znanju (Hart /11/, 1988, str. 7), oz. med sisteme za ravnanje z informacijami. Tako npr. Sauter /10/ uvršča ekspertne sisteme na desni konec daljice sistemov za ravnanje z informacijami, saj v nasprotju s sistemi, ki ležijo na levi strani daljice, ne delujejo s preprosto, temveč z visoko specializirano logiko in znajo sami oblikovati odločitev iz problemske domene.



Slika 1: Sistemi, ki ravnajo z informacijami
Fig. 1: Systems which handle with informations

Ekspertni sistem uporabniku omogoča spremljanje in spreminjanje procesa reševanja problema, zna pa tudi utemeljiti odločitev. Transparentnost ekspertnega sistema omogoča uporabniku razlago rezultatov in analizo, kako bi se rezultat spremenil, če bi vhodne podatke spremenili (kaj se zgodi če...).

Za delovanje ekspertnih sistemov so uporabljene metode umetne inteligence. Praviloma združujejo kvantitativne in kvalitativne informacije, teorijo verjetnosti, teorijo mehkih množic, aritmetiko števil in logična pravila, utemeljena na hevurističnih pričakovanjih. Pri tem so odločitve, ki jih poda ekspertni sistem, praviloma dobre, ne pa nujno tudi optimalne. Ekspertni sistemi se uporabljajo na praktično vseh področjih človekovega delovanja. Ukvarjajo se z različnimi vrstami problemov interpretacije, napovedovanja, diagnosticiranja, oblikovanja, načrtovanja, popravljanja, razhroščevanja, inštrukcij in nadzora. Ker naša aplikacija temelji na ekspertnem sistemu, si bomo najprej ogledali temeljne gradnike ekspertnega sistema, med katere spadajo baza znanja (poglavje 2.1), mehanizem sklepanja (poglavje 2.2), mehke množice (poglavje 2.3) in uporabniški vmesnik (poglavje 2.8). Poglavja 2.4, 2.5, 2.6 in 2.7 predstavljajo postopke analize opazovanih parametrov, ki podajo sliko o dogajanju v omrežju. Tretje poglavje predstavlja rezultate ekspertne analize v obliki izhodnega poročila, kjer so v slednjem zajete vrednosti povprečene na dolžino časovnega okna, ki ga določi uporabnik. S tem pridobi okvirno oceno o dogajanju v omrežju, brez potrebe po opazovanju celotnega intervala simulacije. V okviru že omenjenega poglavja predstavljamo še način analize rezultatov za določitev posameznega problema v omrežju ter funkcionalnosti ES sistema. Četrto poglavje s podanimi sklepi, ugotovitvami in povzetki rezultatov ekspertne analize zaključuje članek.

2 Gradniki ekspertnih sistemov

2.1 Baza znanja

V bazi znanja je shranjeno znanje iz strokovnega področja, ki ga podpira ekspertni sistem. Vsebuje znanje dveh vrst:

- deklarativno znanje, ki opisuje objekte (dejstva in pravila), ki jih obravnava ekspertni sistem in relacije med temi objekti,
- proceduralno znanje, ki vsebuje informacije, kako uporabljamo te objekte, da bi prišli do nekih sklepov in končne rešitve.

Baza znanja je najpomembnejši del ekspertnega sistema, saj velja, da je kakovost ekspertnega sistema v glavnem funkcija obsega in kakovosti baze znanja. Znanje je vanjo zapisano v obliki, ki jo ekspertni sistem razume in zna uporabiti za oblikovanje odločitve, za kar uporabljamo različne predstavitvene metode ali formalizme za predstavitev znanja. Znanje mora biti predstavljeno na način, ki omogoča prilagodljivo, hierarhično urejeno, heterogeno in aktivno strukturo zapisa. Prilagodljivost strukture zapisa znanja je potrebna zaradi naknadnega vključevanja novih spoznanj in omogočanja spreminjanja zapisov, hierarhičnost pa zaradi vertikalnih povezav med objekti nadrejenih in podrejenih tipov v bazi znanja. Heterogenost strukture pomeni možnost zapisa tako deklarativnega kot proceduralnega znanja, aktivnost strukture pa možnost povezovanja objektov v bazi znanja s pravili oz. metodami. Med formalizmi ali metodami za predstavitev znanja prevladujejo simbolične predstavitve, ki jih lahko razvrstimo v štiri glavne vrste: produkcijska pravila, logična predstavitev, semantične mreže in okviri.

Najpogosteje uporabljena metoda so *produkcijska pravila*. Logične relacije med objekti problemskega področja opišemo s pravili tipa *če-nato*, posplošeno torej, *če A, nato B*, ali, če je izpolnjen pogoj P, velja sklep S s faktorjem zaupanja G. Primer logične relacije je npr. *ČE* je količina padavin visoka *IN* še vedno dežuje, *POTEM* lahko z veliko verjetnostjo trdimo, da vode ne primanjkuje. Leva stran pravila (A) predstavlja pogoj ali situacijo, v kateri je pravilo uporabno, desna stran (B) pa določa posledico, sklep ali akcijo pravila. Vsaka stran lahko vsebuje več členov, ki so med seboj povezani z logičnimi operatorji *IN (and)*, *ALI (or)* in redkeje *NE (not)*. Produkcijsko pravilo si razlagamo v skladu z načelom, ki pravi, da če velja A in da iz A sledi B, lahko sklepamo, da velja tudi B. To načelo je osnova izpeljevanja dejstev oz. sklepanja iz aktiviranih produkcijskih pravil. Sklepanje je proces, sestavljen iz dveh korakov, selekcije in aktivacije. V prvem, selekciji, sistem ugotovi, katera pravila so primerna, ter eno od njih izbere za aktivacijo, kjer razlaga izbrano pravilo in iz njegovega jedra izpelje dejstvo, ki ga doda v bazo znanja. Vsa komunikacija poteka preko dejstev v bazi znanja, saj pravila ne morejo neposredno sprožiti drugih pravil.

Semantična mreža je struktura, ki predstavlja znanje na način, da ga organizira v vozlišča, med seboj povezana s povezavami /1/.

Okvir je opis objekta, v katerem je prostor za vsako informacijo o tem objektu. Kot okvir lahko štejemo zbirko med seboj povezanega znanja o določenem strokovnem področju ali njegovem delu, ki lahko vsebuje dejstva, relacije, pravila, dogodke ali privzete vrednosti. Teorija okvirov izhaja iz spoznanja, da človek, ko se nahaja v njemu do tedaj nepoznani situaciji, uporabi znanje, ki je plod izkušenj, povezanih z določenimi situacijami, objekti in podobno. Namesto da bi temeljito raziskal in analiziral novo situacijo, si v spomin prikliče podobne, že poznane situacije, elemente nove situacije pa skuša vključiti v že obstoječo strukturo vedenja o nečem, v tako imenovane okvire /3, 4/.

2.2 Mehanizmi sklepanja

So tisti del ekspertnega sistema, ki upravljajo in nadzorujejo delovanje celotnega ekspertnega sistema. Zadolženi so za aktivno uporabo znanja iz baze znanja, za ravnanje s podatki, ki vstopajo v sistem, in za izpeljevanje ustreznih sklepov. Mehanizme sklepanja v ekspertnem sistemu sestavljajo poizvedovalni postopki in postopki sklepanja, s katerimi ekspertni sistem izvaja iskanje ustreznih rešitev kot svoj rezultat. Njihova glavna lastnost je zmožnost vpogleda v svoje delovanje.

Najpogosteje uporabljeni metodi sklepanja v mehanizmih sklepanja pri zapisu znanja s produkcijskimi pravili sta sklepanje naprej in sklepanje nazaj. Pri sklepanju naprej sistem sklepa induktivno – iz množice znanih dejstev skuša priti do določenega sklepa oz. cilja. S sklepanjem išče končni cilj – neko zadovoljivo rešitev z znanim dejstvom, ki ga primerja z vzorci produkcijskih pravil na levi strani pravila. Če se leva stran ujema z dejstvom, se pravilo aktivira. Aktivirano pravilo v delovni pomnilnik doda novo dejstvo, izpeljano iz jedra oziroma desne strani pravila. Izpeljano dejstvo zdaj enakovredno nastopa v procesu sklepanja: pri ponovitvi postopka se lahko sproži pravilo, ki ima na levi strani prav to izpeljano dejstvo. Opisani postopek se ponavlja vse dotlej, dokler v množici produkcijskih pravil še obstajajo pravila, ki se lahko aktivirajo ali dokler mehanizem sklepanja ne izpelje dejstva, ki predstavlja zadovoljivo rešitev. Sklepanje naprej torej ugotavlja, do kakšnih zaključkov pridemo glede na dana dejstva, uporabljamo pa ga, ko ni mogoče vnaprej predvideti, kateri od možnih zaključkov je najprimernejši.

Sklepanje nazaj poteka deduktivno, njegov cilj se nanaša na potrjevanje ali zavračanje pravilnosti ciljne hipoteze. Mehanizem sklepanja najprej preveri, ali lahko ciljno hipotezo potrdi z dejstvom v delovnem pomnilniku, sicer išče pravilo, ki hipotezo lahko potrdi. Takšno pravilo ima na svoji desni strani vzorec, ki je enak ciljni hipotezi, levo stran pa potrjujejo dejstva iz delovnega pomnilnika. Če se leva stran izenači z dejstvi iz delovnega pomnilnika, je sklepanja konec in mehanizem sklepanja potrdi pravilnost ciljne hipoteze. V nasprotnem primeru levo stran razume kot podcilj, ki ga poskuša potrditi na enak način kot ciljno hipotezo. Postopek ponavlja, dokler mehanizem sklepanja ne potrdi vseh podciljev ali dokler ne zmanjka pravil, s katerimi bi podcilje potrdil. Pravilnost ciljne hipoteze je dokazana, ko so dokazani vsi podcilji, sicer ciljne hipoteze ni mogoče potrditi. Običajno so sistemi s sklepanjem nazaj učinkovitejši od sistemov s sklepanjem naprej, saj težijo k reduciranju iskalnega prostora in tako običajno hitreje pridejo do rešitve, uporabljamo pa ga, kadar obstaja manjše število ciljev oziroma zaključkov, ki jih je možno določiti vnaprej.

2.3 Mehke množice

Mehke množice (ang. *fuzzy sets*) so razširitev običajnih, ostrih množic (angleško *crisp sets*). Medtem ko ima lahko pripadnostna funkcija ostre množice zalogo vrednosti {0,1} (tj. določen element pripada ali ne pripada tej množici), ima pripadnostna funkcija mehke množice (μ_A) zalogo vrednosti znotraj intervala [0,1]. Torej je lahko določen element v mehki množici vsebovan s pripadnostjo ε[0, 1].

2.4 Uporabniški vmesnik sistema ekspertne analize simulacijskih rezultatov

Uporabniški vmesnik ekspertnega sistema skrbi za udobno sporazumevanje med sistemom in (neveščim) uporabnikom, ki mu omogoča vpogled v proces reševanja problema, ki ga izvajajo mehanizmi sklepanja. Uporabniški vmesnik prevaja podatke, ki jih poda uporabnik, v obliko, primerno za računalniško obdelavo, sklepe in pojasnila, ki jih poda sistem, pa predstavi uporabniku v razumljivi pisni ali grafični obliki. Navadno omogoča tudi interakcijo z okoljem in drugimi sistemi, npr. zunanjimi bazami podatkov. Najpogostejši vmesniki, ki jih uporabljajo ekspertni sistemi, so vprašanja in odgovori, meniji, hipertekst, naravni jezik, grafični vmesniki ipd.

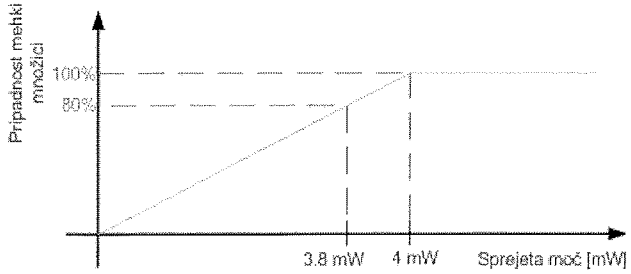
Uporabniški vmesnik je eden od najbolj kritičnih elementov ekspertnega sistema, saj slab uporabniški vmesnik lahko vodi v omejeno ali neučinkovito uporabo. Tudi oblikovanje uporabniškega vmesnika je na splošno zahtevnejše kot pri običajnih računalniških aplikacijah, saj so informacije, ki se izmenjujejo med uporabnikom in sistemom, navadno kompleksnejše, procesiranje, ki ga sistem izvaja, pa zahtevnejše.

2.5 Analiza izkoriščenosti pasovne širine oddajnika in sprejemnika s pomočjo mehke množice

V simulacijskem okolju OPNET Modeler sestavimo simulacijsko strukturo, ki je prikazana na sliki 4.

Promet med udeleženci v taktičnem radijskem omrežju je določen z tako imenovanimi pogodbami, ki definirajo, kdo s kom komunicira in kolikšna je intenzivnost podatkovnih izvorov. Iz primera na sliki 3 je moč razbrati da gre za P2P tip pogodb. V primeru analize izkoriščenosti prenosnih kanalov oddajnika in sprejemnika, je potrebno izbrati statistiko 'utilization' na vsaki enoti. To pomeni, da se bodo v izhodni vektor, ki določa statistiko za posamezno enoto shranile vrednosti obremenjevanja oddajnega in sprejemnega kanala. Za vsako enoto, ki ima izbrano statistiko

'utilization' se ustvari po en izhodni vektor s časovno kodo in amplitudno vrednostjo izkoriščenosti (ang. utilization). Če v drobnogled vzamemo enoto '1bataljon/brigada' le ta sprejema in oddaja promet v dve smeri. To je primer, ko se bodo vrednosti za obe smeri seštele in odražale na skupni izkoriščenosti, saj je enota omejena zgolj z enim sprejemnikom in oddajnikom. Z analizo vsake posamezne statistike za oddajnik in sprejemnik, posamezne enote lahko z uporabo mehke množice definiramo naslednje relacije: *if (izkoriščenost < 50%)*, potem postaja lahko potencialno sprejme še večjo količino prometa, *if (izkoriščenost > 50% && izkoriščenost < 80%)* potem je stanje relativno velika izkoriščenost...itd. Med 80 in 90% izkoriščenostjo začnejo enormno naraščati zakasnitve v Wi-Fi omrežju, kar pomeni, da je takšno stanje izkoriščenosti oddajnika oziroma sprejemnika lahko že alarmantno, podobno, kot je stanje 100%. Na ta način definiramo relacije, ki v celoti sestavljajo mehko množico in definirajo posamezne vrednosti, s kakšno pripadnostjo pripada vrednost mehki množici.



Slika 4: Primer mehke množice za sprejeto moč
Fig. 4: Example of Fuzzy Set for received power

Oglejmo si preprost primer vrednosti sprejete moči pri analizi rezultatov statistike sprejete moči na strani sprejemnika. Iz izmerjenih vrednosti sprejete moči brezžičnih omrežij lahko postavimo mejo, nad katero se brezžične postaje medsebojno slišijo, oziroma obratno, pod katero se komunicirajoče postaje ne slišijo. Predpostavimo, da je teoretična meja 4 *mW*, kar v praktični uporabi ostre množice pomeni, da se bodo postaje s sprejeto močjo npr. nad nivojem 4 *mW* med seboj slišale, v primeru 3.9999 *mW* pa ne, kar je nesmisel. Iz praktičnih eksperimentov lahko zagotovo potrdimo, da se bodo postaje medsebojno slišale celo pri 3.9 *mW* sprejete moči, vendar bo razmerje signal/šum nekoliko slabše. Iz tega razloga uporabimo mehko množico z definirano pripadnostno funkcijo, s čimer zagotovimo vključitev mejnih vrednosti, za katere še velja območje radijske slišnosti, vendar vrednost pripada funkciji z nekoliko manjšo pripadnostjo, kot podatek, ki se nahaja nad vrednostjo 4 *mW* (slika 4).

2.6 Analiza zakasnitev prometa na posamezni enoti

Analiza zakasnitev je identična analizi izkoriščenosti prenosnega kanala, le da v tem primeru uporabimo definirano mehko množico v bazi znanja za statistike tipa 'Delay'. Procedura analize je enaka predhodni metodi, le da so tukaj definirana območja z drugimi mejnimi vrednostmi. Navedimo primer območja zakasnitev, ki ustreza VoIP aplikaciji. *Če (zakasnitev < 150 ms)* potem označi vrednost kot primerno za izvedbo VoIP aplikacije, v primeru območja od 150 ms do 250 *ms* določi zakasnitev kot manj primerno za izvedbo VoIP aplikacije itd.

2.7 Analiza uspešnosti prenosa sporočil med posameznimi enotami z definiranimi komunikacijskimi pogodbami

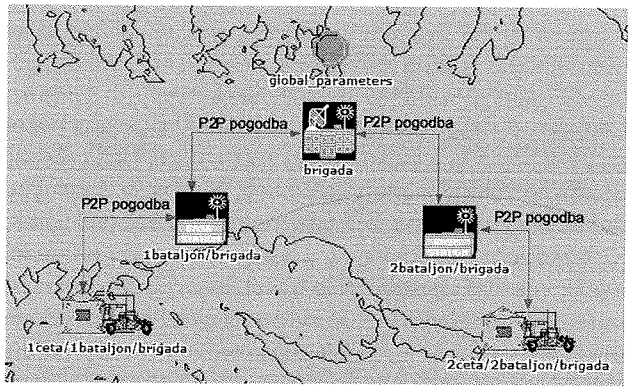
Kot smo že v predhodnem razdelku ugotovili imajo enote medsebojno lahko sklenjenih več komunikacijskih pogodb, izmed katerih lahko vsaka izmed njih zavzame tip 'broadcast' oziroma 'peer-to-peer'. V primeru 'broadcast' pogodbe lahko npr. 'brigada' pošilja eno ali več vrst sporočil po 'broadcast' pogodbi svojim podrejenim enotam, katere ob sprejemu sporočil ne odgovarjajo nadrejeni enoti. Za razliko od 'broadcast' pogodbe pa enota, ki ima sklenjeno 'peer-to-peer' pogodbo sprejema tudi povratni promet, ki se generira zaradi procedure potrjevanja. Iz tega govoreča postaja ugotovi, ali je bilo sporočilo sprejeto ali ne. V postopku analize uspešnosti prenosa sporočil, smo se osredotočili predvsem na P2P pogodbe. V kodni nivo (C jezik) strukture simulacijskega modela smo dodali funkciji izmed katerih prva ustvari datoteko oddanega prometa, druga pa datoteko sprejetega prometa za vsaki IP naslov sodelujoče postaje v komunikacijskem procesu. Vsaka datoteka vsebuje gledano po posamezni vrstici naslednje parametre; velikost generiranega paketa, čas nastanka paketa, izvorni IP naslov in ciljni IP naslov. Identična je struktura druge datoteke. Na tem mestu se pojavi vprašanje zakaj dve datoteki? Razlog je preprost, promet, ki je oddan, potuje po omrežju in prečka številne čakalne vrste, in ker lahko istočasno pošiljajo promet tudi druge komunikacijske enote ne pride do cilja v enakem zaporedju, da bi lahko gledali isto-ležne vrstice v obeh datotekah. Svoj vpliv doprinese še več-nitnost. Iz tega razloga je potrebno za oddan paket, ki je zapisan v prvi datoteki, poiskati isti paket v drugi datoteki, in s tem ugotoviti ali je paket prišel na cilj ali ne. Za posamezen poslan paket se torej postavi števec poslanega paketa na vrednost 1, in če le tega najdemo na cilju v drugi datoteki inkrementiramo še števec sprejetega prometa na vrednost 1, oziroma v nasprotnem primeru pustimo le tega na vrednosti 0. Iz primerjave števcev za vsak trenutek, kadar je kakšna enota oddajala pakete lahko ugotovimo, ali jih je ciljna enota tudi sprejela. Na osnovi primerjave lahko izračunamo, kolikšen je delež uspešnosti prenosa sporočil tekom celotne simulacije za posamično enoto. Tudi v primeru analize uspešnosti prenosa sporočil za vsako posamezno sporočilo k vrednosti dodamo še komentar, npr. '100% uspešnost prenosa' ipd. Na uspešnost prenosa, kakor tudi na ostale opazovane parametre lahko vplivajo številni dejavniki, kot na primer, razgibanost terena, vegetacija, moč oddajnika, ošumljenost kanala, razdalja med komunicirajočimi enotami itd.

2.8 Analiza radijske vidljivosti enot na terenu

Za razliko od predhodnih analiz uporabljamo tukaj poseben scenarij v simulaciji, kjer ima vsaka enota točno določeno časovno režo s periodo kjer lahko odda minimalni paket (npr. ping). V tem scenariju ne moreta oddajati dve enoti hkrati v istem časovnem okviru. Tako imajo komunicirajoče enote s pogodbami vsaka svojo časovno režo s takšno periodo, da se ne pokrije z nobenim drugim komunicirajočim parom. Pri analizi enostavno pogledamo za enote s sklenjenimi pogodbami, ali so v danem časovnem intervalu oddan paket sprejele ali ne. Obdobje, kjer paket ni bil sprejet se smatra kot področje časovnega območja simulacije, kjer radijske vidljivosti med posameznimi enotami ni bilo. Marsikomu se tukaj pojavi vprašanje, zakaj prenašati minimalne velikosti paketov? Razlog je preprost, s simulacijo želimo zgolj ugotoviti ali obstaja med enotama radijska vidljivost ali ne, ob tem pa ne želimo vplivati na zasičenost in zakasnitve v omrežju, zaradi katerih bi lahko pri analizi prišli do napačnih zaključkov.

3. Zasnova ekspertna sistema za vrednotenje simulacij taktičnih radijskih omrežij

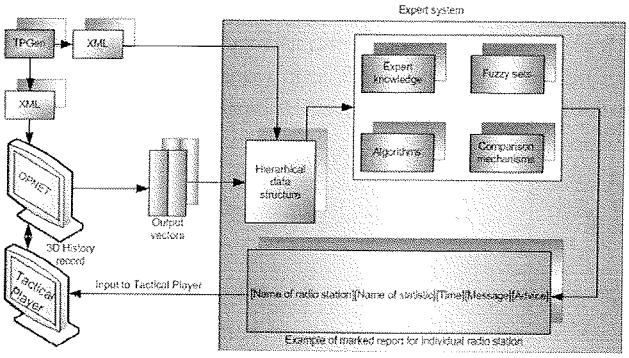
Ekspertni sistem analize rezultatov je bil razvit za potrebe analize simulacijskih rezultatov raziskovalnega projekta s tematsko vsebino modeliranja brezžičnih radijskih taktičnih omrežij za kontrolo in poveljevanje /5, 6, 7/.



Slika 3: Simulacijska struktura omrežja
Fig. 3: Network simulation structure

Sistem za simulacijo taktičnih omrežij smo oblikovali kot tri pomožne podprograme, pri čem TPGEN /5/ skrbi za pripravo parametrov simulacije, Expertni sistem za analizo rezultatov in Tactical Player za pregledovanje analiziranih podatkov. Ekspertni sistem je torej ključen sestavni del v sklopu gradnikov, ki so predstavljeni na sliki 2. Ker smo gradnike slednjega že opisali namenimo še par besed orodju OPNET in taktičnemu predvajalniku.

OPNET Modeler je vodilno simulacijsko okolje v komunikacijski industriji. Omogoča konstruiranje in študije telekomunikacijskih infrastruktur, posameznih naprav, protokolov, aplikacij ipd. Orodje stremi k objektno orientiranemu modeliranju. Ustvarjeni modeli predstavljajo zrcalo strukture dejanskih omrežij in omrežnih komponent. Prisotna je podpora za vse tipe komunikacijskih mrež z naprednimi tehnologijami kot so fast ethernet, WiFi, UMTS, GSM, itd. Simulacijski jezik bazira na seriji hierarhičnih urejevalnikov, ki vzporedno ponazorijo strukturo protokolov, opreme, mreže. Omogočena je tudi animacija dogajanja v omrežjih, kar še dodatno poenostavi razumevanje delovanja posameznega elementa. Nudi možnost ustvarjanja povsem novih enot oziroma preurejanja že obstoječih. Preureditev že obstoječe enote je možna celo na kodnem nivoju, ki je izveden s C/C++ programskim jezikom



Slika 2: Zgradba ekspertnega sistema in medsebojne relacije
Fig. 2: Block diagram of an expert system

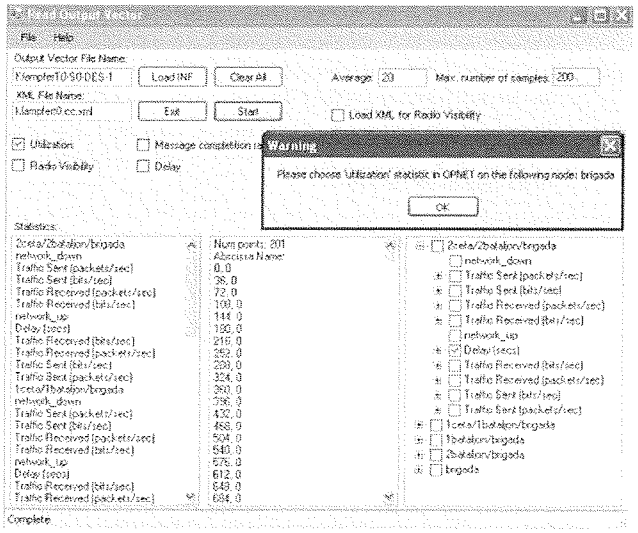
Taktični predvajalnik je programski paket namenjen prikazovanju rezultatov (slika 6), ki jih generira ekspertni sistem. Paket vključuje naslednje funkcionalnosti; branje XML datoteke, branje EHS datoteke, prikaz podatkov, pomikanje po podatkih hkrati pa ima tudi sposobnost krmiljenja OPNET History predvajalnika.

Komunikacija in izmenjava podatkov med orodjem OPNET, TPGen in Ekspertnim sistemom poteka na nivoju XML konfiguracijskih datotek, poročila, ki jih ES pošilja taktičnemu predvajalniku pa v oliki tekstovne datoteke (slika 6).

3.1 Rezultati ekspertnega sistema

Rezultati simulacijskih tekov se nahajajo v grafični obliki, le ti pa predstavljajo velik zalogaj pri analizi ugotavljanja, ali rezultati ustrezajo našim pričakovanjem ali ne. Najpreprostejša metoda je izgradnja aplikacije avtomatične računalniške analize, ki mora vsebovati specifične funkcionalnosti.

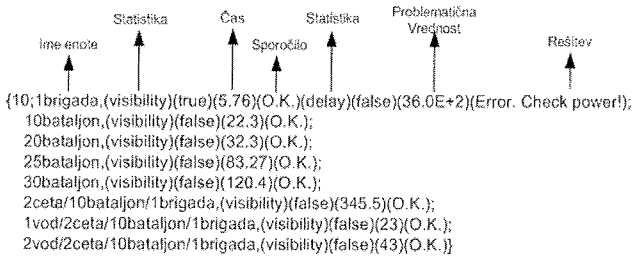
Če se na kratko povrnemo na sliko 2, na kratko predstavimo še orodje TPGen. Le to predstavlja programski paket namenjen vnosu parametrov v vsako posamezno radijsko postajo, ki sodeluje v procesu simulacije. Gre za most med OPNET simulacijskim okoljem in končnim uporabnikom. Vneseni parametri se shranjujejo v XML datoteke, le te pa se naknadno uvozijo v OPNET simulacijsko orodje.



Slika 5: Uporabniški vmesnik ekspertnega sistema
Fig. 5: User interface of an expert system

Po končani simulaciji prejmemo datoteko izhodnega vektorja (primer.ov), ki vsebuje podatkovne vektorje za vsako posamezno izbrano statistiko v simulaciji. Podatkovne vektorje sistem nato uskladi z vsebino XML datoteke, ki vsebuje informacije topologije omrežja. S tem poskrbimo za urejenost posameznih statistik, ki pripadajo posamezni enoti. Po tem koraku naredi ekspertni sistem analizo želenih podatkov namesto nas. Uporabniški vmesnik omogoča uporabniku izbiro statistik na posamezni komunikacijski enoti, ki jih želi opazovati in prikazovati v taktičnem predvajalniku. Ti rezultati so pokazatelji zmogljivosti omrežja. Sem spadajo prenesen promet, prenosne zakasnitve, izkoriščenost prenosnih kanalov, radijska vidljivost itd. Ko uporabnik izbere želeno statistiko v ES (slika 5) se lahko prične analiza s pomočjo baze podatkov (baze znanja). Kot primer; analiza uspešnosti prenosa med opazovanima postajama je izvedena s križno korelacijo med oddanim in sprejetim prometom na strani sprejemnika. Ekspertni sistem enostavno primerja isto-ležeče vrednosti poslanega in sprejetega prometa (paketi/s), kar je opisano v poglavju 2.6. Vrednosti lahko medsebojno v velikosti malenkostno odstopajo, kar je razlog za vpeljavo mehke množice, s čimer preprečimo napačno interpretacijo rezultatov (poglavje 2.3). Analiza je pokazala, da venomer ni mogoče rešiti problema z analizo posameznega grafa, temveč jih je za to potrebnih več. Sistem po opravljeni analizi pripravi poročilo (EHS), katerega izsek strukture prikazuje slika 6 in je namenjen taktičnemu predvajalniku.

Prva vrstica zapisa na sliki 6 prikazuje enega izmed številnih zapisov izbrane statistike. Gledano od leve strani proti desni najdemo najprej časovno kodo dogodka, nato ime enote '1brigada', ime statistike 'visibility', vrednost opazovanega parametra '36E+1' in sporočilo 'Check power!'. Takšna struktura velja za običajne vrednosti, v kolikor pa sistem zazna kritične vrednosti doda še dve dodatni vrednosti, kjer prva predstavlja postavljeno zastavico problema z vrednostjo 'true' (signalizira na problem). Druga vrednost



Slika 6: Struktura EHS datoteke z vključenimi sporočili

vključuje markirano kritično vrednost, ki se z rdečo barvo izpiše v taktičnem predvajalniku. Z drugačno barvo v taktičnem predvajalniku opozarjamo na območja s posameznimi problemi, ki potrebujejo posebno pozornost.

Sistem analize smo testirali na rezultatih simulacij brezžičnih radijskih omrežij, kjer se enote premikajo po vnaprej definiranih trajektorijah 3D prostora. V takšnem primeru imamo za analizirati na tisoče statistik. če še zraven omenjenega poudarimo, da ima lahko vsaka statistika po par tisoč vrednosti (odvisno od dolžine simulacije in števila točk na statistiko), lahko enostavno vidimo, da je ekspertni sistem najboljša rešitev. Končne rezultate uporabnik spremlja v taktičnem predvajalniku v obliki filma s podnapisi. Ob tem ima v predvajalniku na voljo izbiro, katero enoto želi opazovati (lahko eno, dve, deset ali pa tudi vse). Sistem še dodatno pripravi resolucijsko poročilo, ki vsebuje povprečene vrednosti na vnesen interval s strani uporabnika. Tretje poročilo je namenjeno uporabniku in vsebuje splošne statistične podatke od dogajanju v omrežju (koliko odstotkov celotnega časa ni bilo radijske vidljivost, kolikšen odstotek od vseh poslanih paketov je bil izgubljen, itd.).

4. Sklep

Kot smo že uvodoma omenili so ekspertni sistemi razvrščeni po Sauter-ju, glede na to, kako upravljajo informacije. Tako lahko najdemo na desni strani premice sisteme, kot ga prikazuje slika 1, na nasprotni strani pa sisteme s ponavljajočo logiko. Z razvrščanjem smo naš sistem umestili na sredino med informacijske sisteme in ekspertne sisteme. Razlog za to so specifične potrebe, ki so implementirane v naš sistem. Kot primer; iz področja ekspertnih sistemov smo uporabili uporabniški vmesnik, bazo znanja, mehanizme sklepanja, nismo pa uporabili vmesnika za zajem ekspertnega znanja. Obseg znanja je v našem primeru fiksen. Iz leve strani razvrščenih sistemov po Sauter-ju smo uporabili ponavljajočo logiko. Iz tega razloga predstavlja implementiran sistem za analizo grafičnih rezultatov OPNET-a hibridni ekspertni sistem. Ročna analiza grafičnih rezultatov zahteva precej dragocenega časa, še posebej v primerih analize več grafov hkrati. Tovrsten čas je iz ekonomskega vidika z implementiranim sistemom minimiziran na najnižjo možno vrednost. Sistem pripravi dve vrsti poročil, izmed katerih je prvo namenjeno operaterju (uporabniku) in zajema informacije o zmogljivosti omrežja tekom izvajanja simulacije, in hkrati vključuje še predloge za izboljšavo

lastnosti v omrežju. Drugo poročilo predstavlja EHS (Expert History Solution), ki vsebuje podatke za vsak simulacijski zapis posebej (slika 5). Z EHS datoteko zagotovimo še večjo uporabniško prijaznost, saj lahko uporabnik na ta način spremlja dogajanje simulacije v taktičnem predvajalniku, ki pripada višjemu programskemu nivoju. Celotna aplikacija je namenjena uporabniški prijaznosti in poenostavitvi analize rezultatov dogajanja v simuliranem omrežju.

5. Literatura

/1/ William Siler and James J. Buckley, Fuzzy Expert Systems and Fuzzy Reasoning, Willey & Sons, 2007

/2/ M. H. Van Emden and R. A. Kowalski , The Semantics of Predicate Logic as a Programming Language, University of Edinburgh, Edinburgh, Scotland

/3/ Jay Liebowitz, The Handbook of APPLIED EXPERT SYSTEMS

/4/ C.S. Krishnamoorthy, S. Rajeev, Artificial Intelligence and Expert Systems for Engineers, 1996

/5/ Mohorko Jože, Matjaž Fras, Žarko Čučej, Modeling of IRIS Replication Mechanism in Tactical Communication Network with OPNET

/6/ J. Mohorko, M. Fras, Ž. Čučej, Modeling methods in OPNET simulations of Tactical Command and Control Information Systems

/7/ Saša Klampfer, Jože Mohorko, Žarko Čučej, Simulation tools in telecommunications education process

/8/ D. Verber, M. Šprogar, M.Colnarič, Izvedba neintruzivne detekcije napak v vgrajenih krmilnih sistemih, Informacije MIDEM 2007

/9/ A. Assim, M. B. I. Reaz, M. I. Ibrahimy, A. F. Ismail, F. Choong, F. Mohd-Yasin, Pametni dom na osnovi umetne inteligence, Informacije MIDEM 2006

/10/ Sauter Vicki: Decision Support Systems. New York : John Wiley & Sons, 1997. 408 str.

/11/ Hart Anna: Expert Systems. London : Kogan Page, 1988. 208 str.

Saša Klampfer, Jože Mohorko, Žarko Čučej
Smetanova 17, Maribor, 2000, Slovenija
Epošta: sasa.klampfer@uni-mb.si

Prispelo (Arrived): 26.05.2008 *Sprejeto (Accepted): 19.03.2009*

A TOP DOWN APPROACH TO TEACHING EMBEDDED SYSTEMS PROGRAMMING*

I. Fajfar, T. Tuma, Á. Bűrmen, J. Puhan

Fakulteta za elektrotehniko, Ljubljana

Key words: embedded systems, higher education, computer programming, higher programming languages

Abstract: Over last decades we have been witnessing an ever faster technological development in the area of embedded systems. At the same time advances in IT and transition to mass higher education worldwide significantly changed the social and cultural environment in which we teach. We had to respond to these changes with updating the courses and making the teaching of embedded systems more efficient, attractive, and affordable.

The classical approach started with computer architecture combined with assembler. Higher level languages were taught independently. From a motivational point of view this has become questionable, since the gap between the assembler and students' pre-university computer experiences, which are mainly Windows-like applications and Internet, was getting too wide.

Our new curriculum starts with JavaScript, which resembles C, runs in environment familiar to any student, needs no additional software, and motivationally proved a success. Transition to embedded C is now much less painful. The assembler level is left for specialised courses in later semesters.

For the purpose of embedded C courses we developed special hardware platform. Each student can purchase their own professional system, at an affordable price. To achieve that, we needed a support from sponsors. Most courses related to embedded systems now use the same platform with only different add-on boards, which further reduces the price and getting-started overhead. That way students needn't spend too much time with unnecessary technicalies and can focus more on the subject. As a result, after only three years we have already received some quite positive feedback.

Pristop k učenju programiranja vgrajenih sistemov z vrha navzdol

Kjučne besede: vgrajeni sistemi, visoko izobraževanje, računalniško programiranje, višji programski jeziki

Izvleček: V zadnjih desetletjih smo priča vse hitrejšemu razvoju na področju vgrajenih sistemov. Obenem je tehnološki napredek in množično visoko izobraževanje v svetovnem merilu drastično spremenil socialno in kulturno okolje, v katerem učimo. Na te spremembe smo se morali odzvati s prenovo učnega pristopa in narediti študij vgrajenih sistemov učinkovitejši, privlačnejši in bolj dostopen.

Klasičen pristop začenja z opisom računalniške arhitekture v povezavi z zbirnikom. Višji programski jeziki, kot na primer C, se poučujejo neodvisno. Z motivacijskega stališča je postal tak pristop vprašljiv zaradi vse večjega razmika med zbirnikom in izkušnjami, ki jih imajo študentje z računalniki pred vpisom v visoko šolo. Te izkušnje zajemajo predvsem uporabo spleta in okenskih programov.

Naš nov pristop začenja z jezikom JavaScript. Ta je podoben jeziku C, deluje v okolju, ki je domače vsakemu študentu, ne zahteva dodatne programske opreme ter se je z motivacijskega stališča izkazal za zelo uspešnega. Naslednji korak je vgrajen C. Prehod na ta jezik je sedaj veliko enostavnejši. Zbirnik smo prepustili specializiranim predmetom v višjih letnikih.

V namen poučevanja vgrajenega Ceja smo razvili posebno strojno opremo. Vsak študent lahko kupi svoj lasten profesionalen razvojni sistem, čigar cena ne presega cene povprečnega učbenika. Cilj smo dosegli s sredstvi, ki so jih prispevali sponzorji. Večina predmetov sedaj uporablja enotno učno okolje, kar dodatno znižuje ceno in zmanjšuje količino uvajalnega dela. Vse to osvobaja študenta nepotrebnega poglabljanja v različne tehnične posebnosti in mu daje več časa za ukvarjanje z vsebino. Tri leta po vpeljavi novega pristopa smo že opazili nekaj zelo vzpodbudnih odzivov.

1. Introduction

An embedded system is a non-general computing system which comprises a microcontroller to implement some of its functionality. The area of embedded systems has undergone tremendous advances over the last few decades. Microcontrollers are getting cheaper and more powerful and are practically everywhere. It has been estimated that an average American came into contact with more than 100 microprocessors per day already a decade ago /1/, and this number is constantly growing. Apart from that, systems have become more sophisticated, running tens of thousands lines of code, and handcrafted approaches to developing small microcontroller applications of the past are no longer successful. More structured design methodology has become a must. All this rapid changes has made our embedded system curriculum dangerously outdated.

But before we plunge into designing a new curriculum we must not ignore the social and cultural changes, whose understanding is crucial to developing a successful educational programme. A higher education worldwide is facing a problem of transition from elitism to mass education

* Part of this research was presented at the 7th International Conference on Information Technology Based Higher Education and Training, held from 10th to 13th July, 2006, in Sydney, Australia

/2, 3, 4/. In Europe a significant growth of mass higher education has been observed over the past few decades, and Slovenia is no exception /5/. From the increased numbers stem many problems, like that of funding, organization, and of much changed conditions of teaching students with quite different motivation and academic talents.

Still another problem observed in all advanced societies is brought about by immense advances in IT development and extensive use of all kinds of multimedia technologies, especially TVs and computers, already from early childhood. These new video generations are less able to read, concentrate, and memorize /4/, which further influences the capacity of coping with abstract mathematical notations that university lecturers traditionally use to model real life systems. New means of communication has changed the very way information is absorbed and processed in students' heads, which is another challenge for traditional university teaching methods.

Paradoxically, already much increased number of students has at the same time intensified a struggle to attract more and more students to engineering sciences. Especially in western societies these sciences are amongst less popular, which too calls for an urgent change in teaching to attract women as well as some other underrepresented groups /6/.

All these changes are very hard to follow, in some part due to the inflexible European university systems. Most universities in the EU are state funded and consequently their curricula need to be government approved. This ensures a certain quality level and maintains compatibility. The down side, however, are relatively rigid curricula, since any major change needs to pass tedious bureaucratic procedures. During the past decade the contents of many courses were extended by microcontroller related topics. However, this was done by individual teachers without an overall concept, and only within the approved curriculum structure. The result was a patchwork of different microcontroller related courses on different levels. Accordingly, laboratories were based on all sorts of hardware platforms, so students had a lot of learning overhead to switch back and forth between different microcontrollers, compilers and languages. Furthermore, the lack of coordination necessarily led to overlapping course contents which further reduced the teaching efficiency. There was also a traditionally disjoined relation between teaching general purpose programming languages, which are considered hardware independent, and hardware specific embedded system topics.

Alongside a major restructure according to the Bologna process the Faculty of electrical engineering at the University of Ljubljana decided to carry out some needed reforms regarding the embedded system curriculum. Our ambition was to significantly increase the teaching efficiency in this field and remedy as much of the above mentioned shortcomings as possible. This may seem rather ambitious, but due to the rigid nature of our curriculum structure it was now or never. Once the enforced Bologna changes are clad in stone, there will be no more room for major restructures.

In the paper, we describe the design of new embedded systems curriculum, trying to cover the advances in the field itself as well as taking into account changed social and cultural environment. The course consists of three parts: introduction to general programming using JavaScript language, introduction to embedded systems using C language, and subject specific embedded system topics. We give some emphasis on developing embedded operating systems and the hardware development system we have developed for the educational purposes. At the end of the paper we comment some observations and experiences we have had over the last three years.

2. History, analysis of situation and preliminary steps

The first microprocessor course at the Faculty of Electrical Engineering in Ljubljana started already in the late 70s, short after an 8-bit Motorola 6800 was released. Because the 6800 was relatively simple yet complex enough to demonstrate all the basic principles of the microprocessor architecture and programming, and because of it's widespread usage, the 6800 (and some of it's improved versions like 6802 and HC11) was the core processor in almost all of our microprocessor oriented courses. The exception was an introductory course in the first year, where a so called hypothetic computer was used. It was an extremely simplified model of a real microcontroller using only 14 instructions and existed only as a simulated device running on a DOS. At that time that was quite a successful approach through which students got a basic idea of the happenings in an ALU, hardware registers, memory, and on a system bus when running a program. Later, this system was replaced by a real development board based on HC11, which was designed at our faculty. Before that, the same system was already used for sometime in higher level courses, but soon there emerged a need to replace old and increasingly uninteresting hypothetical computer. For that purpose we developed a special IO hardware and IDE with a graphical user interface, which enabled 1st year students to quickly write, compile, run and debug some simple programs. The programming language used was without exception assembly language.

There were also some lectures teaching higher languages, mostly C, but that was used for programming mainframes or PCs and many students saw little if any connection between both words. It even wasn't very rare that a student who was quite successful in mastering assembly language didn't come even close to passing the exam in C language and vice versa - that who had no difficulties with C language, was quite lost with assembler.

Today, the situation has dramatically changed. Modern microcontrollers are highly sophisticated in design and functionality. The development systems easily implement embedded C, or even C++ and compilers with user friendly debugging GUI environments. There is no need to start

off on the assembly level. Moreover, it has become extremely difficult to explain the complex machine level. Therefore it makes sense to skip the registers and start at a hardware independent level with a modern C based development system. The assembly language approach is by all means necessary, since this is the only way to show the students what exactly is happening on the machine level. Many times critical real time primitives are still coded in assembly language. However, this is not the approach for a novice any more. Instead, the assembler language approach has become a speciality and should be tackled in later courses. Having this concept in mind, it becomes also obvious that higher languages are not to be taught separately any longer.

At first we thought that starting with a high level language such as C would not be a problem since the accessibility of computers should have contributed to generally higher level of computer literacy in students entering the electrical engineering studies. But surprisingly, as many years' experience with teaching computer languages and architecture to the first-year students has shown, students show a considerable "fear" of computers. For many first-year students to be, the computer is simply a tool for accessing the Internet, and playing games. The rest is wrapped in mystery. It was easier to teach students programming a decade or two ago from scratch than now when they have certain (misleading) predispositions about what a computer is.

Understanding the needs and abilities of students is fundamental to designing an efficient learning environment. Many groups of students find computer science and even programming obscure and unattractive, and they cannot see any connection to the real life. It has been shown /5/ that women and other underrepresented groups are more likely to engage in discipline if they understand its connection to society. First steps for the student should therefore be concrete and simple, and they should produce (with as little effort as possible) quickly observable results the average student would appreciate. Embedded systems turned out to be too abstract for many beginners to be able to see their applicability in real life. We chose the area of computer engineering all students are familiar with from the user point of view, that is Internet and web based applications.

Apart from that, as we have learned, many students were so much frustrated by the failure at the first unsuccessful attempt to install an IDE that they gave up on programming altogether. Some of those that by some lucky coincidence have no problems installing an IDE were frustrated when, after first compiling a program, a single missing semicolon produced the list of error messages longer than a code itself. For lots of students that was enough to quit the further attempts altogether. To start out, we somehow wanted to avoid the need of compilers. Ideally, student should do with software already installed on his/her computer, e.g. a simple text editor and a browser.

According to those findings we decided to design a curriculum starting from the following prerequisites:

1. Computer architecture and assembly language courses should not be taught independently from the higher, general purpose languages.
2. Introductory course should start out with a higher language, supported with a lot of simple real-life problems that students will instantly recognise as familiar and therefore important.
3. Hardware architecture, programming model, assembly language, and lower software techniques such as bitwise logical operations or bit masking should be introduced gradually in later stages, again supported with solutions of realistic problems.

3. The curriculum - first year level

3.1. General purpose programming

The main objective of our curriculum is that after the first year students master basic algorithm development, programming and coding skills, learn the most basic elements of computer architecture, get some idea of system approach, and start to appreciate the role and importance of embedded systems in modern society. Apart from those subject-specific topics, we have made a significant effort to try and achieve some of the more general teaching goals. Above all we find it important that a student after a first year should to the certain extend be able to notice and solve problems, to think critically and to reflect, develop abstract and system-based thinking, to analyse and to draw syntheses, and gain some self-confidence and spirit of enterprise and activity.

Since our final goal after all is embedded system programming, we need to choose appropriate language to back that up. C language is definitely one of the most appropriate candidates, if nothing else for its widespread usage. C and its derivatives are by far the most widely used programming languages today. The most heard argument against C as a teaching language, that it is ugly - which by the way it is - can be avoided by following some simple rules like indenting code and strictly writing curly braces at every beginning end ending of a block, even when syntactically that is not required.

We have chosen a language according to the following criteria:

1. First steps should produce instant results with as little chance of failure as possible. We wanted to introduce a concept of "speaking" to a computer in a form of a simple text. This has proved a very important step for many students. Without the danger of getting tangled in different conditional and loop statements, compiler warning and error messages, students feel free to explore the familiar effect of an unfamiliar language. Even the most unenthusiastic students get some kick when they produce a first document which actually shows in a browser without much danger of something getting terribly wrong.

A mark-up language such as XHTML proved an appropriate candidate for that stage.

2. Very soon students have enough nerve to try and transfer to the computer some of the burdens of tedious typing of XHTML tags. Producing simple tables using JavaScript code is the next logical step. Unobligatory details such as declaring variables and putting semicolons at the end of statements are required by the lecturer whereas the computer is more forgiving. It is a matter of debate whether this is good or bad, but our experience has shown that the allowed sloppiness enabled a beginning student in general to focus more on the gist of coding, i.e. "explaining" the well formed idea to the computer.
3. Also very importantly, our language should resemble C as closely as possible, to avoid the problem of unnecessary learning overhead when switching between languages and make the transition to embedded C as seamless as possible. Again, JavaScript has proved the best candidate.

In summary, a combination of elementary XHTML and JavaScript formed an abstract (in view of embedded C programming) yet practical framework for introducing basic concepts of computer programming.

Having once eliminated as much of problems annoying for beginners as possible, we can focus on real problems right away. Motivation too is not an issue anymore, since practically every student has experience using browser and is keen to producing his/her own pages. Students quickly want to write more and more complex web pages and soon realise why programming is so important. That way, we can emphasize principles of programming like program development and stepwise refinement more efficiently rather than just present and explain sample programs. Students become very motivated to further develop programs that have been devised to certain stage in lectures. For example, we can present and demonstrate students a simple program that moves a window, and then tell them they can shake a window by multiple moves in different directions in a for loop. They are bound to try that at home.

As another example let us look at a simple code fragment producing a series of thumbnails in a browser window:

```
var i;
for (i = 0; i < num_thumbs; i++)
{
  document.write("<img src='");
  document.write(thumb[i]);
  document.write("' />");
}
```

Since students are familiar with pages displaying thumbnails, they in general appreciate and understand the benefits of using the for loop for producing such a page. This practical understanding motivates students directly for the deeper study of the logic of for loop itself.

Another benefit of using JavaScript is an early need to think modularly and write appropriate functions. Experiences have shown that students tend do avoid decomposing problems in modules and writing functions. Instead they put all code in a main() function. There can be up to several hundreds lines of code and they still don't see why this is wrong.

With JavaScript things are different. If one wants to respond to an event he/she must write a function to do it. It is crystal clear to anyone that trying to put more than a single line of code in a HTML tag verges on a suicide.

In order to support the XHTML/JavaScript part of the course we have written an on-line textbook with embedded live examples. A student is encouraged to experiment with them without a need to copy them elsewhere or even type them. If something gets wrong, there is a Reset button, which restores the original example. One such example is shown in Fig 1. The example demonstrates the use of a statement continue. Apart from the JavaScript code there are some instructions for the student as how to experiment with the code. A student is always encouraged first to try and forecast the effect of the changes to the code and only then to reload the page to see the actual output of the program.

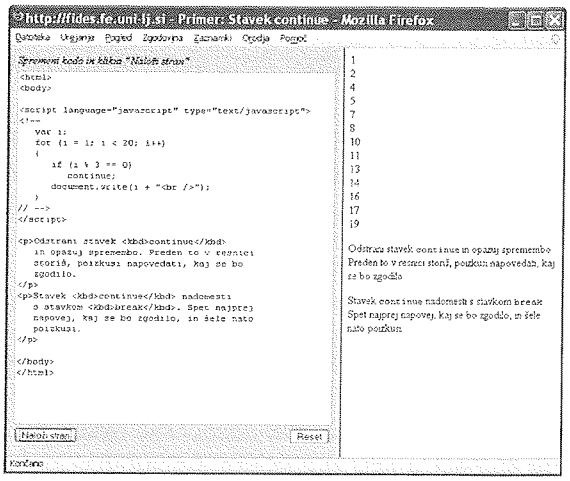


Fig. 1: An embedded live example in the textbook.

3.2. Transition to embedded C

In the second semester the students plunge into embedded C programming. They are already quite familiar with a basic syntax, program control structures, concept of calling and defining functions, and devising simple algorithms. The semester starts with pointing out most important differences between the languages JavaScript and C. Those aren't many since we have "hidden" most of the unnecessary parts of JavaScript that are too different from C language.

For the purpose of teaching embedded C we use a special training hardware platform with very rudimentary input and output, and without an operating system. We describe the platform in detail in section *Hardware Platform*. The

most important difference between C and JavaScript stems from the lack of the operating system: the hardware units used need to be initialized and our program must retain perpetual control over the system. Some other differences we notice at that stage are the consequence of strict typing rules of the C language; once we decided that a variable is of type double, we cannot change it in e.g. string. This is somewhat a relief to a minor population of students that already have some programming experiences. These differences are not very difficult to grasp for the students, and we quickly move on.

Connecting simple external hardware devices such as keys, sensors, and stepper motors is the next important step we take. Apart from some basic physical phenomena like bouncing, students learn that from the programmers point of view the problem of controlling such devices is in fact trivial.

The next example shows, how the problem of rotating a stepper motor for a certain degree is surprisingly similar to the problem of displaying an array of thumbnails we met in previous section:

```
int i;
for (i = curpos; i < stop; i++)
{
  outportb(switchseq[i % 4]);
  delay(20);
}
```

One just needs to replace the array of references to thumbnails with the array containing a four-step switching sequence. Due to the mechanical limitations a small delay between switches is also required.

Even more control over hardware is possible when we learn binary coding principles, bitwise logical operations, bit masking, and direct addressing of the hardware registers.

3.3. Early detection of multitasking and real-time problems.

A perceptive student will soon realise that the above example works good only as long as there are no other concurrent demands on the system apart from rotating a motor. Consider for example that the motor is driving a ramp. A closing ramp should stop immediately after a sensor has sensed an obstacle e.g. a car or person under it. It wouldn't be healthy if the system detected the obstacle only after the motor rotating has been completed.

Students learn that a computer can perform multiple tasks concurrently even if the microcontroller operation is serial in nature. The concurrency is achieved by simply distribute a processor time among different tasks that need to be done. The concept is not very difficult to understand, more interesting are the consequences. Without a proper guidance students quickly tangle into dead loops waiting for a key to be pressed, while a system has frozen.

In most cases concurrency alone is not enough to get the system working. Some tasks have to be executed in certain prescribed time span. Writing programs that meet certain performance deadline is quite different from ordinary, non-real-time programming that most students are used to. This concept again is easy to understand but to engineer a real time application requires a lot of system knowledge that is beyond the scope of a first-year student. So in first year we introduce somehow intuitive polling and assume that all partial tasks execute in a time span that is much smaller than the time available. To meet the timeliness of execution we constantly read the system clock and when the time is due, we simply execute what has to be executed. How really to engineer a real time program to meet performance demands, how the real time concepts affect the overall system performance, and how it complicates debugging is left for a later time, although students already get some idea that things are quite challenging and not so trivial.

4. Advanced Embedded Systems Programming

The curriculum at the Faculty of Electrical Engineering in Ljubljana, Slovenia, basically consists of four common semesters covering all fundamental EE topics followed by several specializing curricula branches. The latter can be roughly divided into four groups: Automatics, Electronics, Power Engineering and Telecommunications. All four groups include microcontroller based courses focusing on specific embedded applications. Typically, these would involve systems for control in robotics, power transmission, RF electronics, etc. So the notion of real-time multi-task programming is introduced at different levels. Either the courses discuss respective programming techniques or they build on embedded operating systems like µSmartX, which was developed by one of our post graduate students, and is freely available on the web /7/.

Of course all advanced courses engage students in practical project work. So far these projects have been based on arbitrary microcontrollers so there always was the typical getting-started-overhead. Also, the specific expensive equipment required the students to work in the laboratories on campus. With our new approach, the overhead is almost nil. Moreover, since the students are able to purchase their own development boards at affordable price, a considerable part of the project work can be done at home.

However, it is of utmost importance that the development board be powerful enough and flexible enough to allow the docking of any advanced hardware boards. This has been achieved by an inventive concept, as explained in the following section.

5. Hardware Platform

We are teaching embedded system knowledge in general but when it comes to giving students practical skills one

necessarily needs to resort to one specific microprocessor. This is just like getting a driving license. The goal is to acquire the skill of driving a car, any car. But you have to practice on one specific model. Although you will drive different cars in your life, we believe it is inefficient to switch back and forth between different car models while still in driving school. With teaching embedded systems it is no different, we need an affordable and robust workhorse to practice.

In the previous section we have already hinted at the idea of constructing a common hardware platform for second and third stage. The design specifications are tough. The development system obviously needs to be very flexible in order to accommodate simple user friendly sessions in the second semester as well as all semi-professional requirements of higher level courses.

Above all we wanted students to have an opportunity to buy their very own development system right from the beginning. Using the comparison to the driving school once more, it is clear that a student having his/her own car right from the beginning will be higher motivated, will be able to work after hours and will keep driving the same car after passing the license test.

5.1. Development Board

Looking for an all around workhorse between contemporary microcontrollers, we decided to take our chances with the ARM7 core by Philips. We're speculating that this technology will be around for at least one decade. In order to keep cost as low as average textbook and still meet professional standards we had to get sponsorship backup right from the beginning. However, in order to attract the attention of potential sponsors we had to present a faculty wide support for the project. This was a classic chicken-and-egg situation since the enthusiasm of participating teachers on the other hand very much depended on the price/performance of the development tool. After much negotiation on both sides a strong consortium of six companies was ready to develop and finance our new ARM7 development board.

According to the demands identified in previous sections we designed the basic development module as depicted in Fig. 2.

The highlight is the integrated on-board debugging hardware linking the ARM7 CPU to the well known professional development environment winIDEA™ by iSystem /8/ which is running on any standard personal computer. The PC is connected via USB and is providing the necessary power supply as well.

In this way we can offer full functionality of the entire development system to our students. The proprietary software on the PC is locked to the on-board debugging hardware in order to prevent unauthorized professional use of the system. This is an original concept protecting the copyright of winIDEA™ and giving the students full development power at the same time.

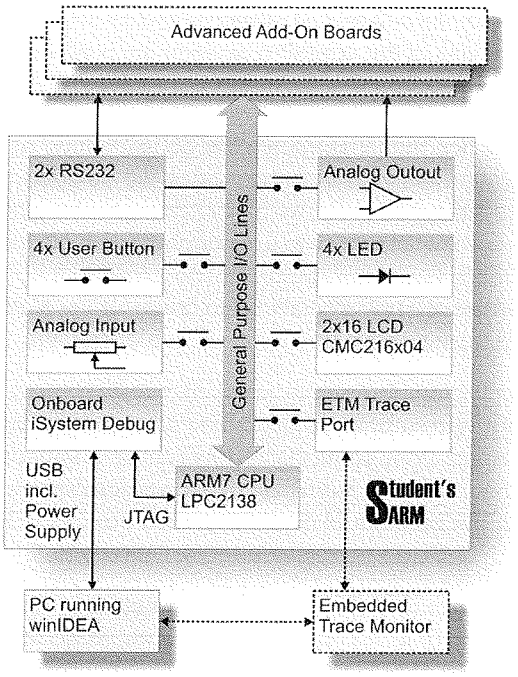


Fig. 2: ARM7 Development system overview.

As shown in Fig. 2, the development board has powerful debugging capabilities but very limited input/output devices. This is because we need to keep the initial costs as low as possible. Remember, the system is introduced in the second semester in support of teaching embedded C. It should provide just basic incentive for novice students. To this end we have included several very simple I/O devices supported with libraries of (semi)-standard C functions enabling students a quick start. There are four keys, four small LEDs, a potentiometer connected to one of the A/D inputs, a general purpose operational amplifier connected to one of the D/A outputs, a pair of RS232 serial ports, and the facilities to mount a standard LCD piggyback. That is more than enough for a beginner course. Advanced level course on the other hand requires more specific devices.

In order to accommodate these specific needs we have provided respective connectors to all CPU ports. Any number of sophisticated add-on boards can be attached to these connectors. For minimal interference with professional add-on equipment all on-board I/O devices except for the serial ports can be disconnected by jumper settings. Individual teachers are designing add-on boards for their specific needs in smaller quantities. Senior students are encouraged to experiment with add-ons in their project work. Many master theses are based on development and testing add-ons.

Optionally, an external embedded trace monitor can be connected to a special port, enabling students to trace their programs in real-time. This, however, requires relatively expensive additional hardware.

From a physical point of view the development board is manufactured in SMD technology, based on a four layer 10 by 10 cm PCB as seen in Fig. 3. In front, the four but-

tons and LEDs are visible. All ICs are covered by the blue plate, which serves for protection and for sporting the sponsor logos. The LCD piggyback is mounted over this area as well.

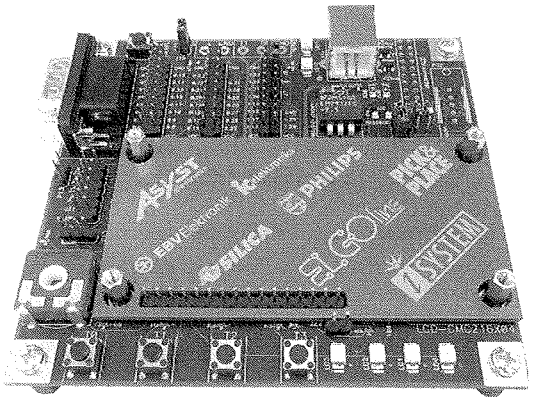


Fig. 3: Student's ARM7 Development board.

Thanks to our sponsors we are able to offer this board, including USB cable and winIDEA™ software to our students at a price less than 40 EUR. In fact the presented development board has become quite popular, so our sponsor iSystem is now offering it world wide as their LPC2138 evaluation board /8/, demonstrating the capabilities of winIDEA™.

5.2. Integrated Development Environment

As mentioned in previous section, we use winIDEA™ as an Integrated Development Environment. Since the software is locked to the on-board debugger, we are able to distribute a full version of the environment. It turned out that students quite appreciate the fact that they can work on a fully professional system at home. This is a strong motivational factor for them as well as for sponsors. They understandably expect that many electrical engineers will want to use exactly the same software in their professional life after graduation. This belief is secured by the saying that old habits die hard.

Fig. 4 shows a running winIDEA™ environment. We can see some basic elements of the environment such as source code and watch windows. The execution of the loaded program has stopped at a breakpoint and the user is able to observe the value of the variable *key*. The important fact is that the program is running on the target board. After two single steps through the code one is able to observe the third LED lighting as the consequence of the execution of the statement _setleds(0x4); This is extremely illustrative for an average first year student who still has difficulties grasping the sequential cause-and-effect concept of computer programming.

6. Observations

It has been almost three years since we introduced the approach described in the paper. Nevertheless it is extremely difficult to give any objective measure of the approach efficiency. We are aware that many parameters change each year upon which we have little or no control and are very difficult to measure. These are the quality of the students enrolled in the program, dates of examinations of other subjects, changes of teaching stuff, to mention just a few. Some may argue that any conclusions made can be of a post-hoc-ergo-propter-hoc type. There are however many indicators, mostly empirical, that make us believe that the approach presented has been successful in achieving the intended goal.

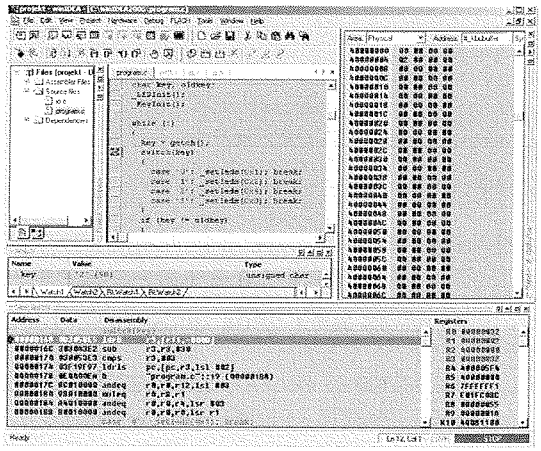


Fig. 4: The winIDEA™ integrated development environment.

The first thing we notice is drastic increase in students' interest in the subject already during the lectures. All of a sudden, out of their own initiative, many students are seeking further explanations and discussions on the subject even during the lecture breaks, and further through e-mail and after-hours. An increased number of students having problems with software and hardware is a sure indicator, that they actually try things out at home. Last month a group of students approached us with a wish that we organize additional summer classes covering the subject more in depth.

In exams, especially oral, where we test higher levels of abstraction according to Bloom's taxonomy, we noticed drastically raised levels of understanding of basic concepts that we have been teaching for more than two decades. This subjective observation was also partially confirmed in numbers. Fig. 5 shows percentage of students that passed the exam during the first examination period, i.e. during the first month after the end of the lectures, over the last seven years. When we introduced the new concept, a significant raise in success rate was observed (year 2005). The percentage is calculated against the population that took the exam, and not the total student population.

The results, however, are not surprising. Starting out on a too low level, which over the past years assembly language definitely has become, gives little motivation to the students. The gap between their experiences of everyday life and low level computing has simply become too wide. On the other hand, many students, already quite familiar with Internet, discover instant application of JavaScript in real life

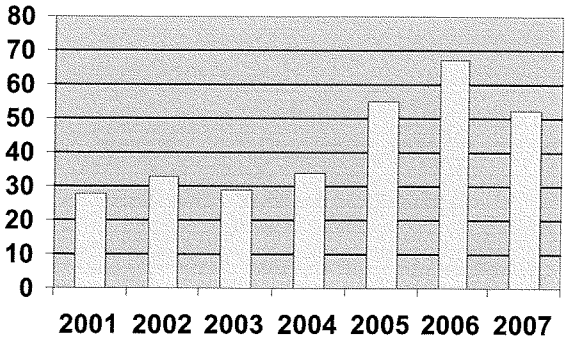


Fig. 5: Percentage of students passing the exam during the first examination period.

problems. This motivational factor is strong enough to lead many students effortlessly through the first, and for many the most difficult, part of learning computer programming. The next step, embedded C programming, turned out to be a natural sequel to the basic JavaScript programming.

7. Conclusion

In the paper we described a recent redesign of embedded systems curriculum at the University of Ljubljana. The old curriculum became dangerously outdated due to enormous technical advances accompanied by not neglect able changes in social and cultural environment through the last decade or two.

We carried out the reorganisation on a three point action plan. Firstly, we strove for a unified software/hardware platform, which would serve as much microcontroller and programming courses as possible. Secondly, we wanted each student to posses his/her own microcontroller development board right from the first year in order to get more involved and to be able to work more efficiently. Thirdly, we had to attract industrial partners in the project by using professional tools and getting respective sponsorships.

The three components mutually depend on each other: without uniting a critical mass of teachers, no sponsor could be attracted. Without a generous sponsorship we could not offer embedded boards for each student. Without students having their own board throughout their studies it was not possible to have a wide cooperation between teaches, which closes the loop. In fact, the uniqueness of our new curriculum lies in our ability to break this dead loop by implementing all three actions simultaneously.

The new curriculum is in effect for only three years so we don't have any long term feedback yet. The first experience however, is very encouraging. The only down side we can see so far is the fact, that our embedded system curriculum is strategically dependent on a single microprocessor architecture and a single development system.

8. Acknowledgment

The authors would like to thank the Ministry of Education Science and Sport (Ministrstvo za šolstvo, Znanost in šport) of the Republic of Slovenia for co-funding our research work through program P2-0246, Algorithms and Optimization Methods in Telecommunications.

9. References

/1/ W. Wolf, J. Madsen, *Embedded Systems Education for the future*, Proceedings of the IEEE vol. 88, no. 1, January 2000, pp. 23-30.

/2/ R. J. Wang, *From elitism to mass higher education in Taiwan: The problems faced*, Higher Education vol. 46, no. 3, 2003, pp. 261-287.

/3/ John Sharpham, *Managing the transition to mass higher education in Australia* Long Range Planning, vol. 26, no. 2, April 1993, pp. 51-58.

/4/ Martin A. Trow, From Mass Higher Education to Universal Access: The American Advantage (March 1, 2000). Center for Studies in Higher Education. Paper CSHE1-00. http://repositories.cdlib.org/cshe/CSHE1-00

/5/ OECD (2006): Education at a Glance OECD Indicators.

/6/ P. D. Stephenson, J. Peckham, Seeing is Believing: Using Computer Graphics to Enthuse Students, IEEE *Computer Graphics and Applications*, vol. 26, *no. 6*, Nov.-Dec. 2006, pp. 87-91.

/7/ *ěSmartX, the free real time operating system for the ARM7TDMI platform*, usmartx.sourceforge.net, 2006.

/8/ *iSystem AG* www.isystem.com, 2008.

/9/ T. Tuma, I. Fajfar, *A new curriculum for teaching embedded systems at the University of Ljubljana*, 7th International Conference on Information Technology Based Higher Education and Training : July, 2006, Sydney, Australia.

prof.dr. Iztok Fajfar, univ.dipl.ing.el.
Fakulteta za elektrotehniko Univerze v Ljubljani
Tržaška cesta 25, 1000 Ljubljana
tel.: (01) 4768 722, fax: (01) 4264 630
e-pošta: iztok.fajfar@fe.uni-lj.si

prof.dr. Tadej Tuma, univ.dipl.ing.el.
Fakulteta za elektrotehniko Univerze v Ljubljani
Tržaška cesta 25, 1000 Ljubljana
tel.: (01) 4768 329, fax: (01) 4264 630
e-pošta: tadej.tuma@fe.uni-lj.si

doc.dr. Arpad Burmen, univ.dipl.ing.el.
Fakulteta za elektrotehniko Univerze v Ljubljani
Tržaška cesta 25, 1000 Ljubljana
tel.: (01) 4768 322, fax: (01) 4264 630
e-pošta: arpad.buermen@fe.uni-lj.si

doc.dr. Janez Puhan, univ.dipl.ing.el.
Fakulteta za elektrotehniko Univerze v Ljubljani
Tržaška cesta 25, 1000 Ljubljana
tel.: (01) 4768 322, fax: (01) 4264 630
e-pošta: janez.puhan@fe.uni-lj.si

Prispelo (Arrived): 23.07.2008 *Sprejeto (Accepted): 19.03.2009*