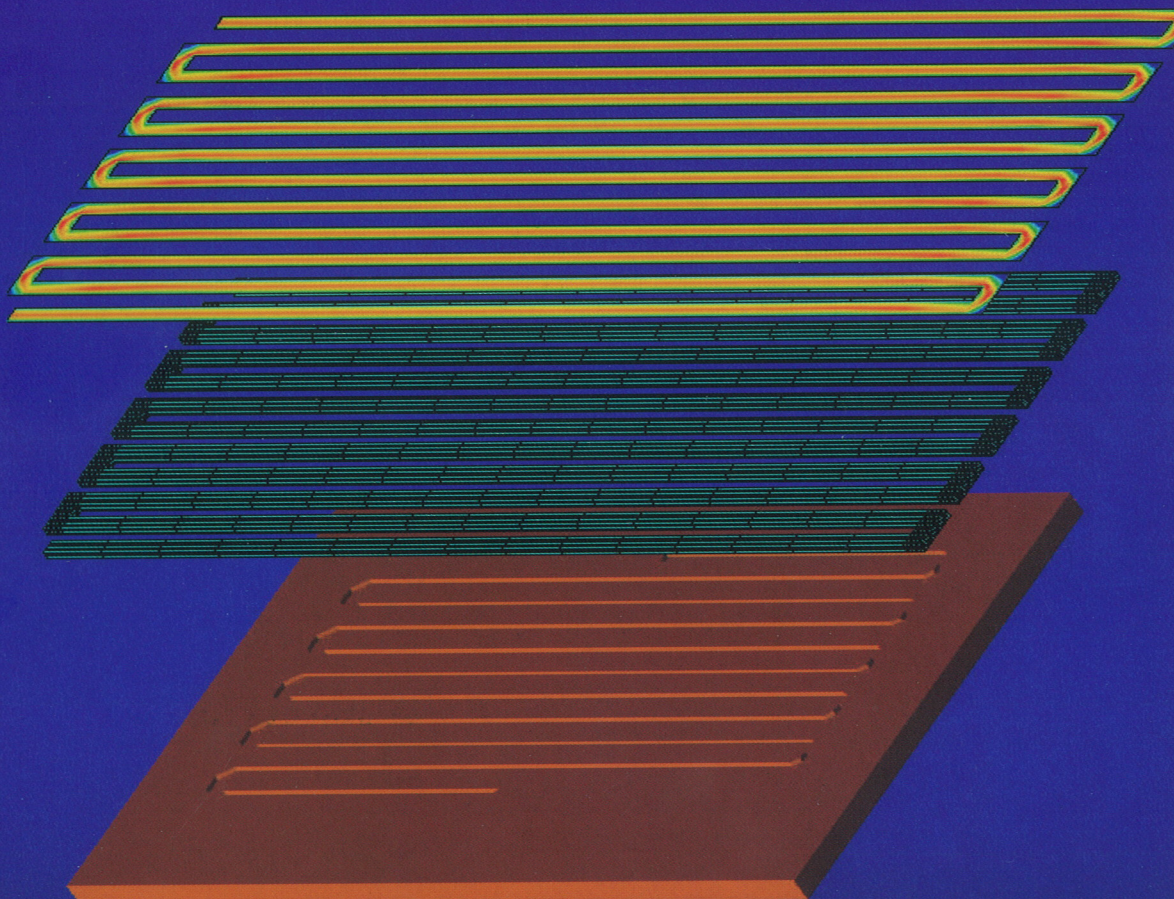# INFORMACIJE MIDEM 3·2007

Strokovno društvo za mikroelektroniko
elektronske sestavne dele in materiale

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

Slika na naslovnici: Front page:
RR aktivnosti na področju mikrofluidnih struktur v   R&D activities in the field of microfluidic structures in
LMSE, UL FE   LMSE, UL FE
LMSE - Labaratorij za mikrosenzorske strukture in   LMSE - Laboratory of Microsensor Structures and
elektroniko, Univerza v Ljubljani,   Electronics, University of Ljubljana,
Fakulteta za elektrotehniko   Faculty of Electrical Engineering

VSEBINA   CONTENT

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

Spoštovani,

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

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

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

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

Upamo, da vas delovanje društva še vedno zanima in da boste članstvo obnovili. Žal pa bomo morali dosedanje člane, ki članstva ne boste obnovili do konca leta 2007, brisati iz seznama članstva.

Prijavnice pošljite na naslov:

MIDEM pri MIKROIKS

Stegne 11

1521 Ljubljana

Ljubljana, september 2007

Izvršilni odbor društva

CHARACTERISTICS OF ARGON AND HELIUM PLASMAS CREATED BY MICROWAVE DISCHARGE AT ATMOSPHERIC PRESSURE

T. Belmonte[*], R.P. Cardoso, C. Noël, G. Henrion, F. Kosior

Laboratoire de Science et Génie des Surfaces, Nancy-Université, CNRS, Nancy Cedex, France

Key words: Microwave plasma, atmospheric pressure.

Abstract: A review of recent developments in the field of microwave plasmas created in noble gases at atmospheric pressure is presented. Several possible designs of discharge configuration are presented and the evolution of plasma jet is explained. The contraction and filamentation of plasma is explained and illustrated. Such plasma is characterized by a high temperature of both neutral and ionized gaseous atoms, which can easily reach several 1000K. The electron temperature is often between 15000 and 30000K, and the electron density usually exceeds 10^{20} m^{-3}. At such conditions, several reactions untypical for low pressure plasmas occur. Among them, formation of dimmers is of particular interest. In some cases, the density of dimmers such as He$_2^+$ may exceed the density of common He$^+$ ions.

Značilnosti argonske in helijeve plazme ustvarjene z mikrovalovno razelektritvijo pri atmosferskem tlaku

Kjučne besede: mikrovalovna plazma, atmosferski tlak

Izvleček: V prispevku opisujemo zadnja dognanja na področju mikrovalovne plazme, ki jo ustvarimo v žlahtnih plinih pri atmosferskem tlaku. Opisujemo nekatere možne konfiguracije razelektritve in razložimo razvoj plazemskega curka znotraj razelektritve. Curek plazme se pri teh razmerah običajno zoži znotraj razelektritvene cevi, pojavijo pa se tudi plazemske niti, kar ilustriramo in razložimo v tem prispevku. Takšno plazmo odlikuje visoka temperatura nevtralnega plina in ioniziranih atomov, ki zlahka doseže več 1000K. Temperatura elektronov je pogosto med 15000 in 30000K, medtem ko njihova gostota običajno preseže vrednost 10^{20} m^{-3}. Pri takšnih razmerah opazimo nekatere reakcije, ki niso značilne za nizkotlačne plazme. Med njimi je posebej zanimiva tvorba dvoatomnih molekul žlahtnih plinov. V nekaterih primerih, kot npr. He$_2^+$, lahko gostota takšnih molekul celo preseže gostoto običajnih helijevih ionov He$^+$.

1. Introduction

High frequency plasmas are widely spread sources of active species which can be used in various applications /1-42/. When these sources are set up on processes working at atmospheric pressure, lower cost treatments can be achieved if efficiency is unchanged with respect to processes under low pressure. Therefore, there is a need to better understand and characterize high pressure sources to optimize the treatments /43-59/.

Microwave plasmas in CW mode operate at high treatment temperatures, namely a few thousands Kelvin. Several sources are widely studied like surface wave excited plasmas /60-64/, waveguide-based microwave torches, like the *"Torche à Injection Axiale"* /65-72/ and resonant cavities /73-76/. In Fig. 1, three sources are presented showing possible designs for atmospheric pressure microwave plasmas. The plasma is either created in open air, confined without wall contact or guided by a fused silica tube where the electric field propagates. Some of these sources cannot operate at high power since fused silica which is commonly employed as confinement vessel melts at ~2000 K. Discharges in rare gases are mostly studied, especially in argon and helium. Beyond design and arrange-ment of the sources, strong efforts, both theoretical and experimental, were performed recently to better understand the physical processes that govern these plasmas. However, complementary studies are still necessary to improve our knowledge on these small-scale plasmas where huge gradients exist. Indeed, diagnostics with high spatial resolution are needed. Some basic data for kinetic processes and even for some species should be determined experimentally at high temperature. This would help develop predictive self-consistent collisional radiative models in optically thick media.

In the present paper, we review recent advances in microwave sources at atmospheric pressure. We describe the difficulties to overcome in order to progress in the understanding of these plasmas. The outline of the paper is then as follows: In section 2, a special attention is paid to recent theories available to describe contraction of some rare gas plasmas. A brief comment will also be given on filamentation. In section 3, we present the most relevant question on basic data needed in the field of atmospheric microwave sources. Finally, concluding remarks are provided in section 4.

Informacije MIDEM 37(2007)3, str. 117-122

T. Belmonte, R.P. Cardoso, C. Noël, G. Henrion, F. Kosior:
Characteristics of Argon and Helium Plasmas Created by ...
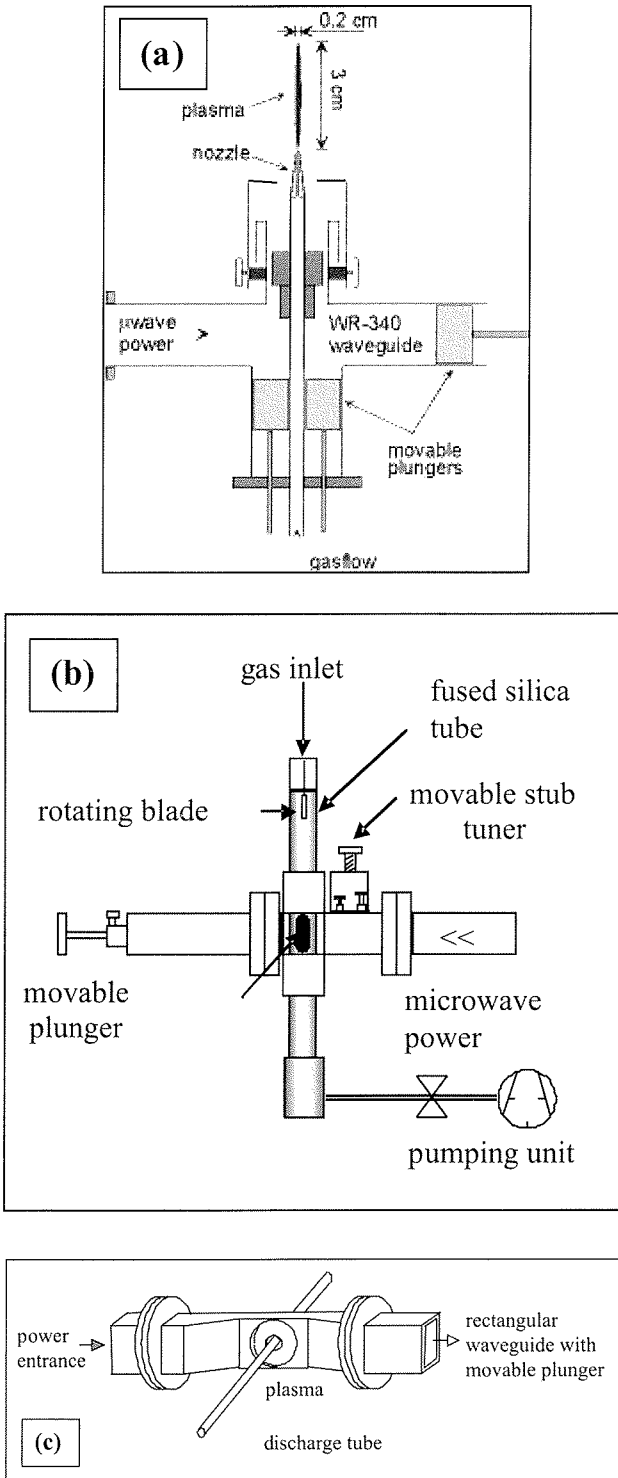
Fig. 1: Some possible designs of microwave sources operating at atmospheric pressure. (a) The "Torche à Injection Axiale" /65/, (b) resonant cavity /73/ and (c) surfaguide wave launcher /60/. The plasma can be sustained in open air, confined without wall contact or guided along a dielectric, respectively.

2. Contraction and filamentation in rare gas plasmas

Contraction can be defined as the compression of plasma into a filament located at the discharge axis. It can be observed in various gases and excitation sources, not only microwave (see Fig. 2). It is characterized by a local increase in both the electron density and the gas temperature and by a decrease in the electron temperature. In the other hand, filamentation is observed only in high frequency discharges only. One single plasma filament splits into two or more filaments of smaller diameters /77, 78/ once the electron density is sufficiently increased radially.
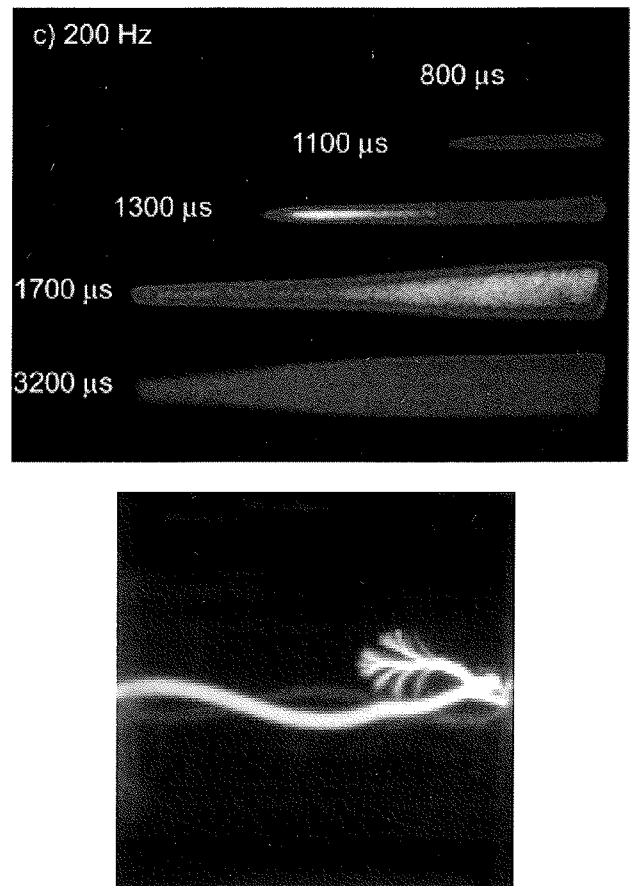
Fig. 2: Illustration of contraction (left) and filamentation (right) phenomena. Left figure is the time evolution of a neon plasma column sustained at atmospheric pressure at 200 Hz (after /78/). Right figure is an argon plasma sustained in a microwave resonant cavity at atmospheric pressure pulsed at 2kHz (duty cycle is 41%).

Recently, study of contraction has gained interest because of the potentialities of these plasmas at atmospheric pressure. Two main mechanisms relying on the same idea, the radial decrease in the ionization rate, can be evoked to explain the contraction process. The first one deals with electron-electron collisions. Energetic electrons which are depleted by inelastic collision can be heated by electron-electron collisions that tend to give a Maxwellian distribu-

T. Belmonte, R.P. Cardoso, C. Noël, G. Henrion, F. Kosior:
Characteristics of Argon and Helium Plasmas Created by ...

Informacije MIDEM 37(2007)3, str. 117-122

tion. However, this mechanism is only possible if the electron density is high enough, usually close to the discharge axis. Then, at the edge of the plasma, the tail of electron energy distribution function is likely depleted and the ionization rate decreases radially /79-80/. The second possible explanation is nonuniform heating of the gas along the discharge radius. It creates contraction through its influence on the kinetics of dimer ions, which controls the charged-particle balance. These descriptions rely on basic data that now need to be examined in detail for rare gases. Helium and argon are considered next.

3.   Basic data

A common difficulty present in available works deals with the temperature dependence of rate constant of chemical reactions over wide ranges. Indeed, depending on the power delivered to the plasma, the temperature of the gas can vary over large range of values. Moreover, due to high temperature gradients, it varies spatially. The main consequence of this change in the gas temperature is the control of the kinetics of excimers, whether they are neutral or charged. There is no measurement on the excimer density nor on their vibrational distribution above 1500 K. In Fig. 3, we give an overview of the gas temperature dependence of the rate constant for the three possible recombination processes of He_2^+ ions /81-84/. Available data are spread over a wide range of values and experimental data are all given below 600 K.

Let's consider for example the case of He^+ and He_2^+ ions in helium atmospheric microwave discharges where several difficulties arise. Dimer ions are mainly produced from He^+ by a three-body collision:

$$He^+ + He + He \rightarrow He_2^+ + He \qquad (1)$$

with a rate constant of $1.2 \pm 0.2 \times 10^{-31}$ cm^6 s^{-1} at 300 K /85, 86/.

First, the temperature dependence of this process, extrapolated from data by Böhringer /85/ measured between 50 K and 350 K ($T_g^{-0.60\pm0.10}$) is rather different from that determined by Russel /86/ between 80 K and 300 K ($T_g^{-0.38\pm0.06}$). However, the two sets of experimental data differ by less than 30%. The accuracy of the measurements within this range of temperature is probably not high enough to determine correctly the exponent of the power law. The former dependence was used for example by /67/ to determine to role of dimer ions between 500 K and 3000 K.
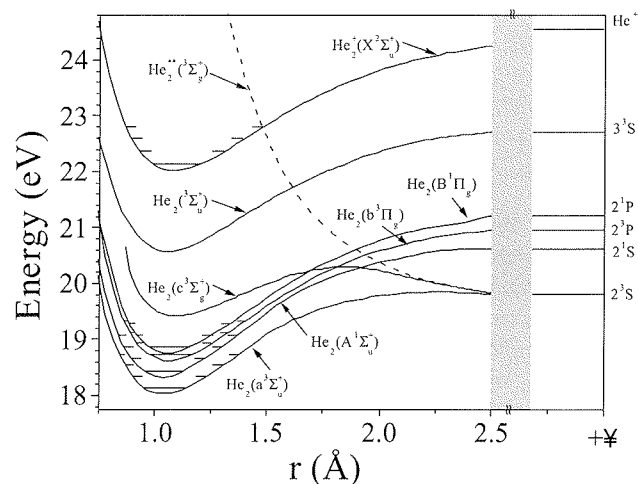
Fig. 4:   Energy diagram of helium dimers. The doubly excited state He_2^{**} is an antibonding triplet state intersecting the dimer ion close to the v=3 vibrational level.

Second, to estimate the ratio $[He_2^+]/[He^+]$, three main processes are considered by /67/ to account for the reaction pathways possibly followed by He_2^+: direct electron recombination, He_2^+ creation by reaction (1) and loss by the reverse process, dissociation of dimer ions by electron impact being negligible. Assuming the Saha equilibrium, the rate constant of the reverse process of reaction (1) is estimated. This choice gives a ratio $[He_2^+]/[He^+]\sim0.55$ at 2500 K for $n_e=10^{14}$ cm^{-3}, i.e. similar number densities for the two species. The rate constant used for direct electron recombination is 5.0×10^{-9} $/T_e^{0.5}$ cm^3 s^{-1} where T_e is the electron temperature in eV. This rate constant is given with no gas temperature dependence. However, it does depend on it as shown in recent works based on ring storage experiments. Indeed, due to the unfavourable location of the doubly excited potential energy curves (see Fig. 4), the recombination route and rate (Eq. (2)) of ions depend

Fig. 3:   Equivalent two-body rate constants for recombination processes of He_2^+ ions as a function of the gas temperature. An electron temperature of 2 eV and an electron density of 10^{14} cm^{-3} are chosen. Data deduced from experiments are by /81/-/84/. Data used by /67/ is added for comparison. Extrapolations of available temperature dependences over the range [300-3200 K] are also given (solid lines). When available, the accuracy of the equivalent rate constant is reported.

Informacije MIDEM 37(2007)3, str. 117-122

T. Belmonte, R.P. Cardoso, C. Noël, G. Henrion, F. Kosior:
Characteristics of Argon and Helium Plasmas Created by ...

both on the electron energy and the initial rovibrational distribution (/87-89/).

$$He_2^+ (J, v) + e \rightarrow \left\{ He_2^*; He_2^{**} \right\} \rightarrow He^* + He \quad (2)$$
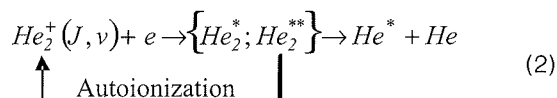
Autoionization

A competing process has to be taken into account, namely the autoionization of the intermediate Rydberg He_2^* or doubly excited He_2^{**} neutral molecule, which can fragment back into an electron and a molecular ion in the (J',v') rovibrational level.

The unusually low values of the direct recombination rate found in experimental works are at least partially explained (see /82, 90/ for details). The dependence of the rate of the direct recombination process with the electron temperature was determined to be $\propto T_e^{-0.9}$ /90/. The gas temperature dependence is not available on a large range of temperatures. However, such dependence is available for neon dimer ions /91/. Results obtained confirm a fall-off in efficiency of the direct recombination for vibrationally excited rare-gas dimer ions with increasing temperature, the contribution of autoionisation being not negligible /92/. Direct recombination should only be considered at low temperature and high electron energy.

Therefore, the two-body recombination of electrons with dimer ions should not be considered at high temperature in the estimation made by /67/. In fact, this consideration does not change significantly the predicted result. Despite ambipolar diffusion prevails at 2500 K and is much lower than direct recombination as used by /67/, the density ratio He_2^+/He^+ is nearly the same than that these authors found. One of the two processes, the atom assisted association or its reverse process, has rate always much higher than the loss of He_2^+ by recombination or ambipolar diffusion which are known to be slow recombination processes. If the absolute value of the density of He_2^+ depends on direct and reverse rates of atom assisted association, the density ratio He_2^+/He^+ does not (see Eqs. 10 and 11 p 466 in /67/). In fact, to modify the temperature dependence of the density ratio He_2^+/He^+, one should either increase the rate of the recombination or decrease the rate of dissociation by atom impact. A last remark deals with the role of associative ionization. Rate constants for these processes at high temperature are unknown and the reaction processes should be included in the calculation of the density ratio He_2^+/He^+ if they are not negligible.

Considering now the temperature dependence by Russel /86/, if the balance of association and dissociation via atom impact is in equilibrium, one should find a ratio $He_2^+/He^+ \sim 9.2$ at 2500 K for $n_e = 10^{14}$ cm^{-3}, showing that at high temperature, He_2^+ ions may possibly still prevail.

In argon, these same authors find a ratio $Ar_2^+/Ar^+ \sim 3.0 \times 10^{-3}$ at 2500 K for $n_e = 10^{15}$ cm^{-3}, a result used by /63/ more recently. This difference between helium and argon is attributed to the dissociation energy of the dimers (2.4 eV

versus 1.3 eV), which is nearly twice higher for the former and only partially compensated by the larger partition function of argon ion dimers. Here again, no experimental data is available to determine how the Ar_2^+/Ar^+ ratio evolves versus gas temperature.

The same kind of problem arises for excimer creation or associative ionization despite theoretical data are available in the case of this latter processes for some triplet states of helium /87/. However, only experimental measurements would confirm the validity of available estimations.

4. Conclusion

Atmospheric microwave plasmas are high-temperature non-equilibrium media where complex phenomena that are still under investigation occur. Better understanding these media requires specific measurements to determine rate constants at high temperature. Diagnostics with high spatial and temporal resolutions are needed to provide basic data that lack today to provide satisfying description of these sources of active species.

5. Acknowledgement

References

/1/   U. Cvelbar, M. Mozetic and M. Klanjsek-Gunde, IEEE Trans. Plasma Sci. **33**, 236 (2005).

/2/   M. Mozetic, Informacije-MIDEM **33**, 222 (2003).

/3/   Ricard, V. Monna and M. Mozetic, Surf. Coat. Technol. **174-175**, 905 (2003).

/4/   K. Norikazu, H. Yamada, T. Yajima and K. Sugiyama, Thin Solid Films **515**, 4192 (2007).

/5/   R. Peelamedu, D. Kumar and S. Kumar, Surf. Coat. Technol. **201**, 4008 (2006).

/6/   A. Anders, Surf. Coat. Technol. **200**, 1893 (2005).

/7/   S. Guruvenket, G. Mohan Rao, Manoj Komath and Ashok M. Raichur, Appl. Surf. Sci. **236**, 278 (2004).

/8/   P. Ganachev and H. Sugai, Plasma Sources Sci. Technol. **11**, A178 (2002).

/9/   Ricard, M. Gaillad, V. Monna, A. Vesel and M. Mozetič, Surf. Coat. Technol. **142-144**, 333 (2001).

/10/   Vesel, M. Mozetič and A. Zalar, Appl. Surf. Sci. **200**, 94 (2002).

/11/   Vesel, E. Vamvakopoulos, M. Mozetič and G. A. Evangelakis, Phys. B Condens. Matter. **324**, 261 (2002).

/12/   Drenik, U. Cvelbar, A. Vesel and M. Mozetič. Inf. MIDEM **35**, 85 (2005).

/13/   Vesel, M. Mozetič, A. Drenik, S. Milosevic, N. Krstulovic, M. Balat-Pichelin, I. Poberaj and D. Babič, Plasma Chem. Plasma Process **26**, 577 (2006).

/14/   M. Mozetič, A. Vesel, A. Drenik, I. Poberaj and D. Babič, J. Nucl. Mater. **363-365**, 1457 (2007).

T. Belmonte, R.P. Cardoso, C. Noël, G. Henrion, F. Kosior:
Characteristics of Argon and Helium Plasmas Created by ...

Informacije MIDEM 37(2007)3, str. 117-122

/15/ T. Vrlinič, A. Vesel, U. Cvelbar, M. Krajnc and M. Mozetič, Surf. Interface Anal. **39**, 476 (2007).

/16/ Vesel, M. Mozetic, A. Hladnik, J. Dolenc, J. Zule, S. Milosevic, N. Krstulovic, M. Klanjšek-Gunde and N. Hauptmann, J. Phys. D: Appl. Phys. **40**, (2007) in press.

/17/ U. Cvelbar, S. Pejovnik, M. Mozetič and A. Zalar, Appl. Surf. Sci. **210**, 255 (2003).

/18/ M. Mozetič, A. Zalar, U. Cvelbar and I. Poberaj, Appl. Surf. Sci. **211**, 96 (2003).

/19/ U. Cvelbar, M. Mozetič and A. Zalar, Vacuum **71**, 207 (2003).

/20/ M. Mozetič, A. Zalar, U. Cvelbar and D. Babič, Surf. Interface Anal. **36**, 986 (2004).

/21/ M. Mozetič and U. Cvelbar, Adv. Mater. **17**, 2138 (2005).

/22/ U. Cvelbar, M. Mozetič and A. Ricard, IEEE Trans. Plasma Sci. **33**, 834 (2005).

/23/ M. Klanjšek-Gunde, M. Kunaver, A. Hrovat and U. Cvelbar, Prog. Org. Coat. **54**, 113 (2005).

/24/ U. Cvelbar, M. Mozetič, I. Poberaj, D. Babič and A. Ricard, Thin Solid Films **475**, 12 (2005).

/25/ M. Klanjšek-Gunde, M. Kunaver, U. Cvelbar, N. Barle, Vacuum **80**, 189 (2005).

/26/ U. Cvelbar, B. Markoli, I. Poberaj, A. Zalar, L. Kosec and S. Spaić, Appl. Surf. Sci. **253**, 1861 (2006).

/27/ U. Cvelbar, D. Vujošević, Z. Vratnica and M. Mozetič, J. Phys. D: Appl. Phys. **39**, 3487 (2006).

/28/ U. Cvelbar and M. Mozetič, J. Phys. D: Appl. Phys. **40**, 2300 (2007).

/29/ M. Mozetič, A. Zalar and M. Drobnič, Appl. Surf. Sci. **144-145**, 399 (1999).

/30/ M. Mozetič, A. Zalar and M. Drobnič, Thin solid films **343-344**, 101 (1999).

/31/ M. Mozetič and A. Zalar, Appl. Surf. Sci. **158**, 263 (2000).

/32/ M. Mozetič, A. Zalar, P. Panjan, M. Bele, S. Pejovnik and R. Grmek, Thin Solid Films **376**, 5 (2000).

/33/ M. Mozetič, Vacuum **61**, 367 (2001).

/34/ M. Kunaver, M. Klanjšek-Gunde, M. Mozetič and A. Hrovat, Dyes Pigm. **57**, 235 (2003).

/35/ M. Mozetič and A. Zalar, Mater. Sci. Forum **437-438**, 81 (2003).

/36/ M. Kunaver, M. Klanjšek-Gunde, M. Mozetič, M. Kunaver and A. Hrovat, Surf. Coat. Int. B **86**, 175 (2003).

/37/ M. Mozetič and A. Zalar, Vacuum **71**, 233 (2003).

/38/ M. Mozetič, Vacuum **71**, 237 (2003).

/39/ M. Klanjšek-Gunde, M. Kunaver, M. Mozetič and A. Hrovat, Powder Technol. **148**, 64 (2004).

/40/ M. Kunaver, M. Mozetič and M. Klanjšek-Gunde, Thin Solid Films **459**, 115 (2004).

/41/ Hassanien, M. Tokumoto, P. Umek, D. Vrbanić, M. Mozetič, D. Mihailović, P. Venturini and S. Pejovnik, Nanotechnology **16**, 278 (2005).

/42/ Arčon, M. Mozetič and A. Kodre, Vacuum **80**, 178 (2005).

/43/ U. Cvelbar, M. Mozetič, D. Babič, I. Poberaj and A. Ricard, Vacuum **80**, 904 (2006).

/44/ M. Mozetič, Surf. Coat. Technol. **201**, 4837 (2007).

/45/ M. Balat-Pichelin and A. Vesel, Chem. Physic **327**, 112 (2006).

/46/ Vesel and M. Mozetič, Vacuum **61**, 373 (2001).

/47/ M. Mozetič, A. Vesel, V. Monna and A. Ricard, Vacuum **71**, 201 (2003).

/48/ M. Mozetič, U. Cvelbar, A. Vesel, A. Ricard, D. Babič and I. Poberaj, J. Appl. Phys. **97**, 103308-1 (2005).

/49/ N. Krstulovic, I. Labazan, S. Milosevic, U. Cvelbar, A. Vesel and M. Mozetič, J. Phys. D: Appl. Phys. **39**, 3799 (2006).

/50/ M. Mozetič, A. Vesel, U. Cvelbar and A. Ricard, Plasma Chem. Plasma Process **26**, 103 (2006).

/51/ M. Mozetič, A. Ricard, D. Babič, I. Poberaj, J. Levaton, V. Monna and U. Cvelbar, J. Vac. Sci. Technol. A **21**, 369 (2003).

/52/ D. Babič, I. Poberaj and M. Mozetič, Rev. Sci. Instrum. **72**, 4110 (2001).

/53/ Ricard, V. Monna and M. Mozetič, Surf. Coat. Technol. **174-175**, 905 (2003).

/54/ Poberaj, M. Mozetič and D. Babič, J. Vac. Sci. Technol. A **20**, 189 (2002).

/55/ G. Cicconi, J. Phys. D: Appl. Phys. **15**, 1403 (1982).

/56/ C. Hollenstein, J.-L. Dorier, J. Dutta, L. Sansonnens and A.A. Howling, Plasma Sources Sci. Technol. **3**, 278 (1994).

/57/ V.J. Law, A.J. Kenyon, N.F. Thornhill, A.J. Seeds and I. Batty, J. Phys. D: Appl. Phys. **34**, 2726 (2001).

/58/ M. Mozetič, A. Vesel, A. Drenik, I. Poberaj and D. Babič, J. Nucl. Mat. **363-365**, 1457 (2007).

/59/ Vesel, M. Mozetic and M. Balat-Pichelin, Vacuum **81**, (2007) 1088.

/60/ M. Moisan and Z. Zakrzewski, J. Phys. D: Appl. Phys. **24**, 1025 (1991).

/61/ U. Kortshagen, H. Schlüter and A. V. Maximov, Phys. Scr. **46**, 450 (1992).

/62/ Yu. M. Aliev, H Schlüter and A Shivarova, Plasma Sources Sci. Technol. **5**, 514 (1996).

/63/ E. Castaños-Martinez, Y. Kabouzi, K. Makasheva and M. Moisan, Phys. Rev. E **70**, 066405 (2004).

/64/ Y. Kabouzi, D. B. Graves, E. Castaños-Martinez and M Moisan, Phys. Rev. E **75**, 016402 (2007).

/65/ M. Moisan, G. Sauvé, Z. Zakrzewski and J. Hubert, Plasma Sources Sci. Technol. **3**, 584 (1994).

/66/ Jonkers, M. van de Sande, A. Sola, A. Gamero and J. van der Mullen, Plasma Sources Sci. Technol. **12**, 30 (2003).

/67/ Jonkers, M. van de Sande, A. Sola, A. Gamero, A. Rodero and J. van der Mullen, Plasma Sources Sci. Technol. **12**, 464 (2003).

/68/ R. Alvarez, M. C. Quintero and A. Rodero, J. Phys. D: Appl. Phys. **38**, 3768 (2005).

/69/ S. Y. Moon and W. Choe, Spectrochim. Act. B **58**, 249 (2003).

/70/ C. Tendero, C. Tixier, P. Tristant, J. Desmaison and P. Leprince, Spectrochim. Act. B **61**, 2 (2006).

/71/ R. Stonies, S. Schermer, E. Voges and J. A. C. Broekaert, Plasma Sources Sci. Technol. **13**, 604 (2004).

/72/ I. Al-Shamma'a, S. R. Wylie, J. Lucas and C. F. Pau, J. Phys. D: Appl. Phys. **34**, 2734 (2001).

/73/ R. P. Cardoso, T. Belmonte, G. Henrion and N. Sadeghi, J. Phys. D: Appl. Phys. **39** 4178 (2006).

/74/ R. P. Cardoso, T. Belmonte, P. Keravec, F. Kosior and G. Henrion, J. Phys. D: Appl. Phys. **40** 1394 (2007).

/75/ A. Skovoroda and A. V. Zvonkov, J. Exp. Theoret. Phys. **92**, 78 (2001).

/76/ H. Potts and J. Hugill, Plasma Sources Sci. Technol. **9**, 18 (2000).

/77/ N. Djermanova, D. Grozev, K. Kirov, K. Makasheva, A. Shivarova and Ts. Tsvetkov, J. Appl. Phys. **86**, 738 (1999).

/78/ Y. Kabouzi and M. Moisan, IEEE Trans. Plasma Sci. **33**, 292 (2005).

/79/ G. M. Petrov and C. M. Ferreira, Phys. Rev. E **59**, 3571 (1999).

/80/ Yu. B. Golubovskii, H. Lange, V. A. Maiorov, I. A. Porokhova and V. P. Sushkov, J. Phys. D: Appl. Phys. **36**, 694 (2003).

/81/ R. Deloche, P. Monchicourt, M. Cheret and F. Lambert, Phys. Rev. A **13**, 1140 (1976).

/82/ Carata, A. E. Orel and A. Suzor-Weiner, Phys. Rev. A **59**, 2804 (1999).

/83/ P. C. Hill and P. R. Herman, Phys. Rev. A **47**, 4837 (1993).

Informacije MIDEM 37(2007)3, str. 117-122

T. Belmonte, R.P. Cardoso, C. Noël, G. Henrion, F. Kosior:
Characteristics of Argon and Helium Plasmas Created by ...

/84/ R. J. van Sonsbeek, R. Cooper and R. N. Bhave, J. Chem. Phys. **97**, 1800 (1992).

/85/ H. Bohringer, W. Glebe and F. Arnold J. Phys. B: At. Mol. Phys. **16**, 2619 (1983).

/86/ J. E. Russel, J. Chem. Phys. **84**, 4394 (1986).

/87/ J. S. Cohen, Phys. Rev. A **13**, 86 (1976).

/88/ J. B. Birks, Rep. Prog. Phys. **38**, 903 (1975).

/89/ C. Focsa, P. F. Bernath and R. Colin, J. Molec. Spectr. **191**, 209 (1998).

/90/ X. Urbain, N. Djuric, C. P. Safvan, M. J. Jensen, H. B. Pedersen, L. Vejby Sogaard and L. H. Andersen, J. Phys. B: At. Mol. Opt. Phys. **38**, 43 (2005).

/91/ Jiang, J. A. Gutherie, R. C. Chaney and A. J. Cunningham, J. Phys. B: At. Mol. Opt. Phys. **22**, 3047 (1989).

/92/ J. Cunningham, T. F. O'Malley and R. M. Hobson, J. Phys. B: At. Mol. Phys. **14**, 773 (1981).

*T. Belmonte[*], R.P. Cardoso, C. Noël, G. Henrion, F. Kosior.
Laboratoire de Science et Génie des Surfaces,
Nancy-Université, CNRS, Parc de Saurupt – CS 14234
– 54042 Nancy Cedex – France
[*]Corresponding author.
Fax: +33.3.83.53.47.64
E-mail: Thierry.Belmonte@mines.inpl-nancy.fr*

INTERACTION OF HIGHLY DISSOCIATED OXYGEN PLASMA WITH POLYMERS AND THEIR COMPOSITES

P. Eiselt

Plasmabull Engineering GmbH, Lebring, Austria

Key words: Oxygen plasma, Composite, Polymer, Coating, Particle dispersion, Image analysis, Surface characterization

Abstract: Recent application of fully dissociated oxygen plasmas with a low density of charged particles for treatment of polymers and polymer-matrix composite materials is described. Plasma is often created in a high frequency inductively coupled discharge to avoid ion acceleration. At the pressure of few mbar the plasma density is often of the order of 10^{16} m^{-3}, and the density of neutral oxygen atoms of the order of 10^{21} m^{-3}. The dissociation fraction of oxygen molecules may approach 100%. Plasma with such characteristics causes modification of solid materials. It is extensively used for surface cleaning and activation, selective etching and sterilization. The first effect of plasma treatment is surface activation. The wettability of materials is increased dramatically enabling their painting, printing and metallization. The treatment time depends on the type of materials and plasma parameters. Optimal wettability is often obtained in less than 1s of plasma treatment thus making the technology suitable for industrial use. Prolonged plasma treatment of polymer matrix composites causes selective etching. Different components are etched at different rates. The highest is the etching rate of the polymer matrix, while inorganic fillings are not etched at all. Oxygen plasma treatment of the composites thus represents a unique method for studying the distribution as well as the orientation of different fillings in composites. The application of this technology is illustrated with several examples.

Interakcija visoko disociirane kisikove plazme s polimeri in njihovimi kompoziti

Kjučne besede: Kisikova plazma, Kompozit, Polimer, Prevleka, Porazdelitev delcev, Analiza slik, Preiskava površin

Izvleček: V prispevku je opisana uporabnost nizkotlačne popolnoma disociirane kisikove plazme za obdelavo polimerov in kompozitnih materialov s polimerno matriko. Takšno plazmo pogosto generiramo v visokofrekvenčni induktivno sklopljeni razelektritvi, s čimer se izognemo pospeševanju ionov v električnem polju. Pri tlaku nekaj mbar je gostota plazme reda velikosti 10^{16} m^{-3}, gostota nevtralnih kisikovih atomov pa reda 10^{21} m^{-3}. Stopnja disociiranosti kisikovih molekul se lahko približa 100%. Plazma s tovrstnimi značilnostmi povzroča spremembo trdnih materialov in se široko uporablja za čiščenje, aktivacijo, selektivno jedkanje in sterilizacijo. Prvi pojav, ki ga opazimo pri izpostavi trdnih materialov kisikovi plazmi, je površinska aktivacija. Omočljivost tako obdelanih materialov se dramatično poveča, kar omogoča dober oprijem materiala pri barvanju, tiskanju in metalizaciji. Značilni čas obdelave je odvisen od vrste materiala in plazemskih parametrov. Optimalno omočljivost pogosto dosežemo že v času, manjšem od 1s, kar omogoča industrijsko uporabo. Podaljšana plazemska obdelava kompozitov s polimerno matriko vodi k selektivnemu jedkanju. Različne komponente v kompozitu se jedkajo z različno hitrostjo. Največja je hitrost jedkanja polimerne matrike, medtem ko se anorganska polnila sploh ne jedkajo. Plazemska obdelava kompozitov tako predstavlja edinstveno metodo za preiskavo porazdelitve in celo orientacije polnil v kompozitih. Aplikacija te tehnologije je ilustrirana z različnimi primeri.

1 Introduction

In the past decade, oxygen plasma has been successfully applied to novel technologies such as plasma ashing, plasma cleaning, selective plasma etching and plasma sterilization /1-18/. All the technologies are based on controlled oxidation of organic compounds. In contrary to standard oxidation that is carried on close to the thermodynamic equilibrium, the oxidation in plasma is a well non-equilibrium process. The main advantage of the non-equilibrium oxidation is the capability of controlling the oxidation rate independently from the sample temperature. This is possible due to a low potential barrier for oxidation with reactive particles from oxygen plasma. Oxygen plasma is a source of different reactive particles including excited molecules and atoms, positive and negative molecular and atomic ions, ozone and neutral oxygen atoms. The concentration of these particles depends largely on discharge parameters (i.e. type of discharge, discharge power and frequency, magnetic field, size and shape of discharge vessel, type of material facing plasma, pressure and gas mixture, etc.) Interaction of plasma radicals with solid materials is both

physical and chemical. Physical interaction is usually performed with charged radicals that can be accelerated by biasing samples, while neutral radicals usually do not have substantial kinetic energy so the interaction is purely potential and therefore very selective. Optimal oxidation selectivity is therefore obtained with plasma of low ion density and high neutral radical density.

2 Inductively coupled RF oxygen plasma

Plasma is usually created in a gaseous discharge. Electrons are accelerated in electric field and thermalized at elastic collisions. Their energy distribution function is therefore rather Maxwellian with the temperature of several eV. Electrons in the high-energy tail of the distribution function have enough energy for direct ionisation of gaseous molecules, while those in the low energy part of the distribution function are only capable to excite rotational and vibrational excited states of molecules. In the case of oxygen molecules, the excitation energy for ro-vibrational states is well

below 1 eV, the excitation energy for metastable molecules $O_2{}^1\Delta$ and $O_2{}^1\Sigma$ is about 1 and 2 eV, respectively, the dissociation energy is 5.2 eV, while the ionisation energy is 12 eV. Electrons with the average energy of few eV are therefore likely to excite molecules into metastable states and ro-vibrational states, while dissociation and ionisation in less probable.

Excited particles tend to de-excite. There are many channels for de-excitation, some of them occur in the gas phase, while the others take place on the surface of the discharge chamber. In any case, the conservation of energy and momentum as well as rules of quantum mechanics should be obeyed. Vibrationally excited states are de-excited in the gas phase primarily by vibration interchanges (V-V transitions) and super-elastic collisions with atoms (V-T transitions) /8/. Neutral oxygen atoms in the ground state can recombine to molecules only at three-body collisions that are unlikely to occur at low pressure (say below few mbar) so they are rather stable in the gas phase.

Surface de-excitations often play a dominant role in low-pressure plasmas. The probability for ion recombination as well as metastables relaxation is close to unity /19/, while the recombination of neutral oxygen atoms depends largely on the type of material facing plasma as well as its temperature and morphology. The probability for recombination at heterogeneous surface recombination (O + O → O_2) for many glasses and some ceramics is often low (the typical order of magnitude in 10^{-4}), while for many metals and some porous ceramics it is of the order of 10^{-1}. The recombination probability does not depend only on the type of material, but also on other parameters.

At low flux of oxygen atoms onto the surface, the recombination probability depends on the surface coverage with O atoms. This phenomenon was often observed at high vacuum, when the flux of O atoms onto the surface is below say 10^{22} $m^{-2}s^{-1}$. When the flux is increased, the surface becomes saturated with O atoms and the recombination probability approaches a constant value, that has been known as the recombination coefficient (γ). The recombination coefficients depend on the material, the surface morphology, and often also on the temperature. There seems to be no general rule, but the recombination coefficient tends to increase with increasing surface roughness as well as increasing surface temperature.

The density of different excited species in plasma depends on excitation and de-excitation probabilities. The excitation probabilities depend mostly on electron density and temperature, while the de-excitation probabilities depend particularly on surface properties. By choosing smooth materials with low recombination coefficients for recombination of O atoms it is often possible to achieve plasma with a low density of ions but a high density of neutral atoms. It is often possible to obtain oxygen plasma with the ion density below 10^{16} m^{-3} and the neutral atom density above 10^{21} m^{-3} /20-23/.

High oxidation selectivity can only be obtained with cold plasmas. Cold plasma is a state of gas with a low kinetic energy of heavy particles, i.e. all particles except electrons. There are some channels for heating heavy particles in plasmas. Happily enough, elastic collisions between fast electrons and heavy particles do not lead to substantial kinetic energy exchange due to a small mass of the electrons. The major channel for heating heavy particles at collisions with energetic electrons is a dissociation event. At direct electron impact dissociation, the excessive kinetic energy of the electron can be observed as the kinetic energy of newly formed atoms, which can move apart with a substantial kinetic energy. Such energetic atoms get effectively thermallized at elastic collisions with other heavy particles. A good way to avoid this sort of plasma heating is application of plasma with a rather low electron temperature.

Another mechanism of heating heavy particles is super-elastic collisions between vibrationally excited molecules and oxygen atoms. The reaction cross-section is large / 19/. The only way to avoid such collisions is application of fully dissociated plasma where such collisions are unlikely to occur.

An important channel for heating heavy particles may be acceleration of ions in electric field. The ions are accelerated in electric field as are the electrons. As long as the electric field frequency is low, the ions can follow any change of the local electric field. The energy an oxygen ion can pick from the electric field is

$$W = (e^2E^2)/(2m\omega^2). \tag{1}$$

Here, e is the ion charge, E peak electric field, m ion mass and ω electric field frequency. The kinetic energy an oxygen ion in a high vacuum can gain in electric field versus the frequency is plotted in Figure 1. As long as the frequency is low (say below 100 kHz), the ions are well accelerated in the field. But as the frequency is increased,
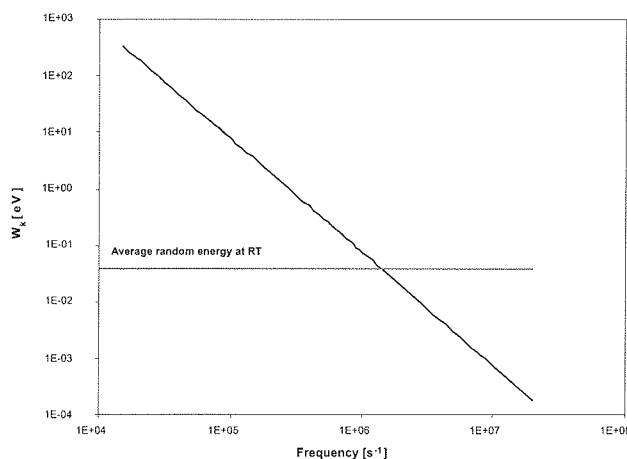
Fig. 1. *Maximal kinetic energy of oxygen ions oscillating in a high frequency electric field of 100V/cm.*

the ions are not able to follow the field. At the frequency of about few MHz, the kinetic energy an ion can pick from the field is less that average thermal energy of an ion at room temperature. This means that ions in the electric field with a frequency above 10 MHz cannot pick energy worth mentioning. As long as the electric field frequency is higher than 10 MHz, the ions thus cannot be accelerated in the field and cannot contribute to neutral gas heating.

Upper considerations lead to the conclusion that best oxygen plasma for selective oxidation of different materials is created in a radio-frequency discharge. A high degree of dissociation of oxygen molecules is obtained in a discharge chamber made from glass, which has a smooth surface and a low coefficient for recombination of oxygen atoms. There are two extreme modes for RF generator coupling: i) capacitive, and ii) inductive. In practice, the coupling is often a mixture of both extremes. Inductive coupling is often obtained using a coil wounded around a glass tube. In this case, electrons are accelerated in induced electric field sustained due to alternating magnetic field in the coil. The electric field in the axis is rather low, and is increasing towards the edge of the glass tube.

Another extreme is capacitive coupling. In this case, the charged particles are accelerated in the alternating electric field between two parallel electrodes. A sheath with a substantial potential is established next to the powered electrode. As long as the sheath is almost collisionless (i.e. the mean free path is larger than the sheath thickness) the ions entering the sheath from the gas phase are accelerated towards the electrode and do not transfer kinetic energy to other particles. They bombard the electrode and some are reflected as neutral fast atoms or molecules. These fast particles do heat the neutral gas. The heat exchange between positive ions and other heavy particles is increased in the case the sheath is not collisionless. In such cases there are more channels for kinetic energy exchange in the gas phase. This often occurs at elevated pressure, say above 0.1 mbar, where the density of atoms increases with increasing pressure.

As the electric field frequency increases towards the microwave range, the ions can gain practically no energy from the field, and also electrons cannot pick as much energy as in the case of radio-frequency discharges. As a general rule, the electron temperature in simple microwave discharges is always lower than in radio-frequency discharges with comparable power. More energy is transferred to neutral gas heating so the microwave plasmas are never as cold as the RF plasmas.

3   Plasma parameters

Parameters of low pressure plasmas created by inductively coupled RF discharge in a glass tube depend on discharge power and pressure. Typical values are as follows: neutral gas kinetic temperature is often equal to ion kinetic temperature and is a bit more than room temperature - values between 300 and 500 K are common. The electron temperature is often about 50000 K or more. At the pressure of few 10 Pa, the density of electrons and ions is often between 10^{15} and 10^{16} m^{-3}, the density of neutral oxygen atoms of the order of 10^{21} m^{-3} while the density of metastable oxygen molecules of the order of 10^{19} m^{-3}. The plasma potential is often of the order of 10 V and the Deby length about 10^{-4} m. The density of neutral oxygen atoms is certainly the most important parameter.

Several methods have appeared to measure the O density in highly dissociated oxygen. The methods include optical spectroscopy /24-27/, mass spectrometry /28-29/, gas titration /22, 30/ and catalytic probes /31-42/. The latter was found to have some advantages over other techniques, as catalytic probes enable real/time measurements and do not disturb the original concentration of O atoms. The disadvantages include a poor understanding of surface recombination phenomena and sensitivity to high-frequency interferences. From the latter point of view, fibre optics catalytic probes (FOCP) have a definite advantage: as any connection is made optical, they are completely immune to stray effects caused by high frequency electromagnetic field /35-38/. On the other hand, the FOCPs cannot measure low densities of O atoms.

Figure 2 represents measured values of O density in inductively coupled RF plasma created in a discharge tube made from borosilicate glass. Measurements were performed with a FOCP. One can observe that the O density does not depend much on RF power as long as the pressure is low. As the pressure is increased, the O density increases monotonously until it reaches a broad maximum. At high pressure, the O density decreases with increasing power. The appearance of maxima on the curves presented in Figure 2 is explained by different mechanisms of oxygen atom production and loss. At low pressure the O density is limited by surface effects rather than the discharge power. At high pressure the limiting factor is the poor density of electrons as well as their temperature. At even high-
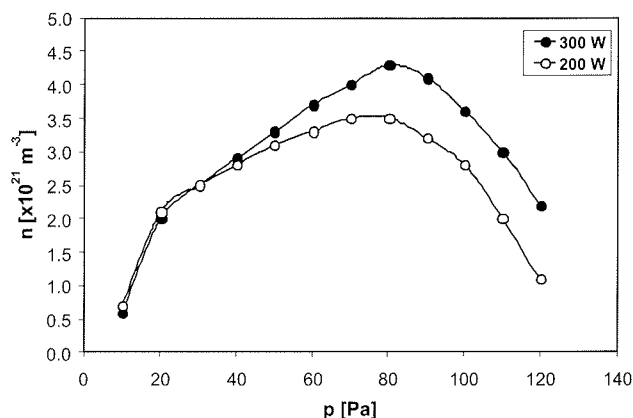
Fig. 2.   Density of neutral oxygen atoms in a glass discharge tube with the inner diameter of 3.6mm versus pressure. The parameter is the discharge power.

er pressure the gas-phase atom loss by three-body would become important if the power were further increased. The optimal conditions for a large O density are met at pressure between about 50 and 100 Pa. In this range, the O density depends largely on the discharge power: a higher power causes a higher density. The pressure at which the maximum appears depends on power: at higher power the maximum is shifted to a higher pressure. In any case, the theoretical limit of the O density is full dissociation.

4 Interaction of plasma radicals with sample surfaces

The interaction of inductively coupled oxygen plasma with solid materials is almost entirely potential. As shown in upper text, the ion density is usually below 10^{16} m^{-3} and their kinetic energy at the sample surface about 10 eV. On the other hand we have neutral atoms with the density often exceeding 10^{21} m^{-3}. The dissociation degree therefore exceeds the ionization fraction by 5 orders of magnitude. A large flux of O atoms onto a sample surface assures rich surface chemistry if the samples are organic materials. The first effect of interaction is surface function-

a

b
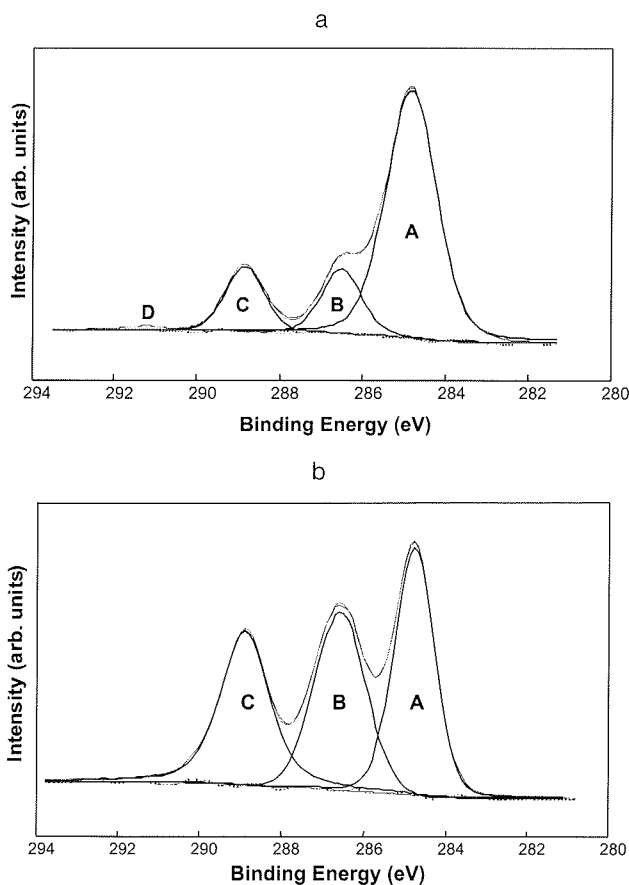
Fig. 3. High resolution XPS C1s peak of polyethyleneterephalate. a) before and b) after plasma treatment for 1s. The O atom density is $6x10^{21}$ m^{-3}. The peak A corresponds to C-C bond, B corresponds to C-O bond, C corresponds to O=C-C bond. D is the π^*- π sh up.

a

b
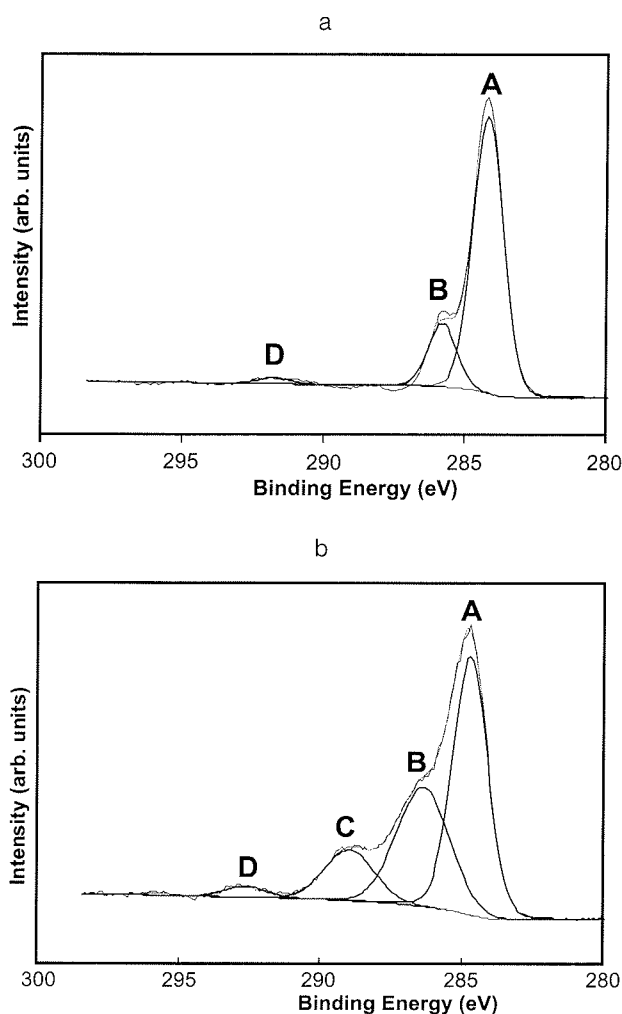
Fig. 4. High resolution XPS C1s peak of polyethersulphone. a) before and b) after plasma treatment for 1s. The O atom density is $6x10^{21}$ m^{-3}.

alization with oxygen-rich functional groups. Next effect is slow etching of the organic material. Since the interaction is almost purely potential, the etching largely depends on the nature and structure of organic materials.

The appearance of the O-rich functional groups is best monitored by X-ray Photoelectron Spectroscopy (XPS) / 43-45/. The resulting surface activation (change of surface energy) is often measured by a contact angle of a water drop, while the distribution of different particles (fillings) in a polymer matrix composite is best monitored by a Scanning Electron Microscopy /46-50/. The appearance of the reaction products is often detected by Optical Emission Spectroscopy (OES) [/51/].

4.1 Surface activation

The first effect of oxygen plasma treatment of organic materials is surface activation. Figure 3 represents the effect of oxygen plasma treatment of polyethyleneterephalate. Figure 3a is the carbon C1s peak obtained by a high-resolution XPS. The carbon is almost entirely bonded to other
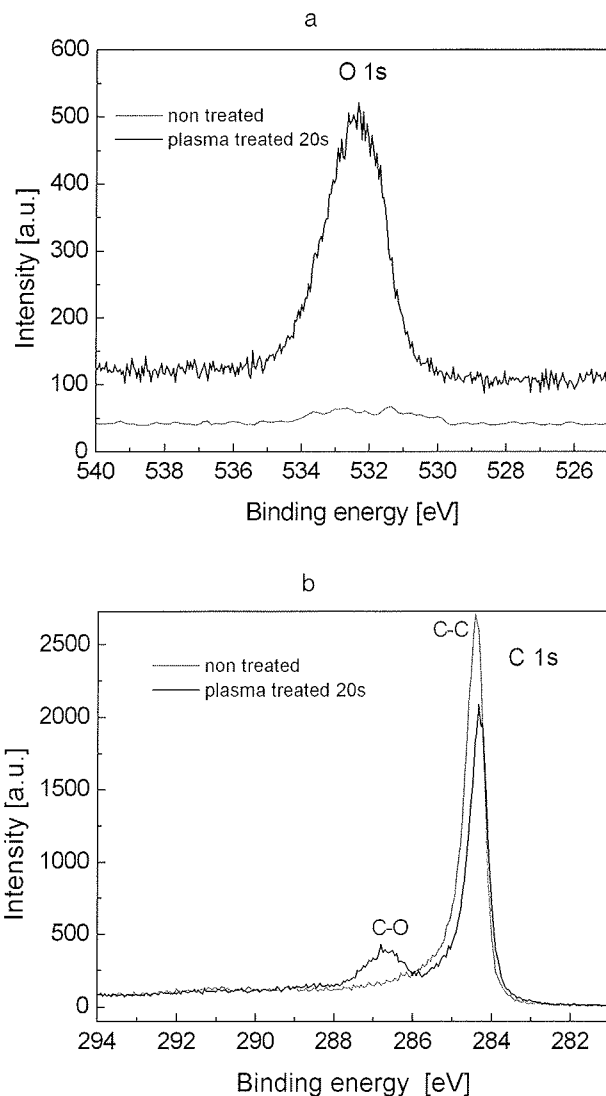
Fig. 5. High resolution XPS analysis of electrolytic graphite. a) O1s peak before and after plasma treatment for 20s, b) C1s peak before and after plasma treatment for 20s.

carbon atoms and some oxygen is presented in the form of ester group. The sample is exposed to inductively coupled oxygen plasma for 1 second. The carbon peak after this short plasma treatment is shown in Figure 3b. The peak is now enriched with oxygen functional groups. The surface of this material is saturated with functional groups in a second – further treatment does not influence the concentration of the groups on the surface.

Next example of quick plasma activation is presented in Figure 4. In this case, the organic material to be activated is polyethersulphone – a polymer containing sulphur. The C1s peak before plasma treatment is presented in Figure 4a, while after the treatment it is shown in Figure 4b. As in the case of polyethyleneterephalate, the treatment time is only 1 second. The surface is quickly saturated with oxygen rich functional groups such as C-O (peak B in the spectrum) and C=O (peak C). The most energetic peak (D) is a satellite peak ($\pi \rightarrow \pi^*$).
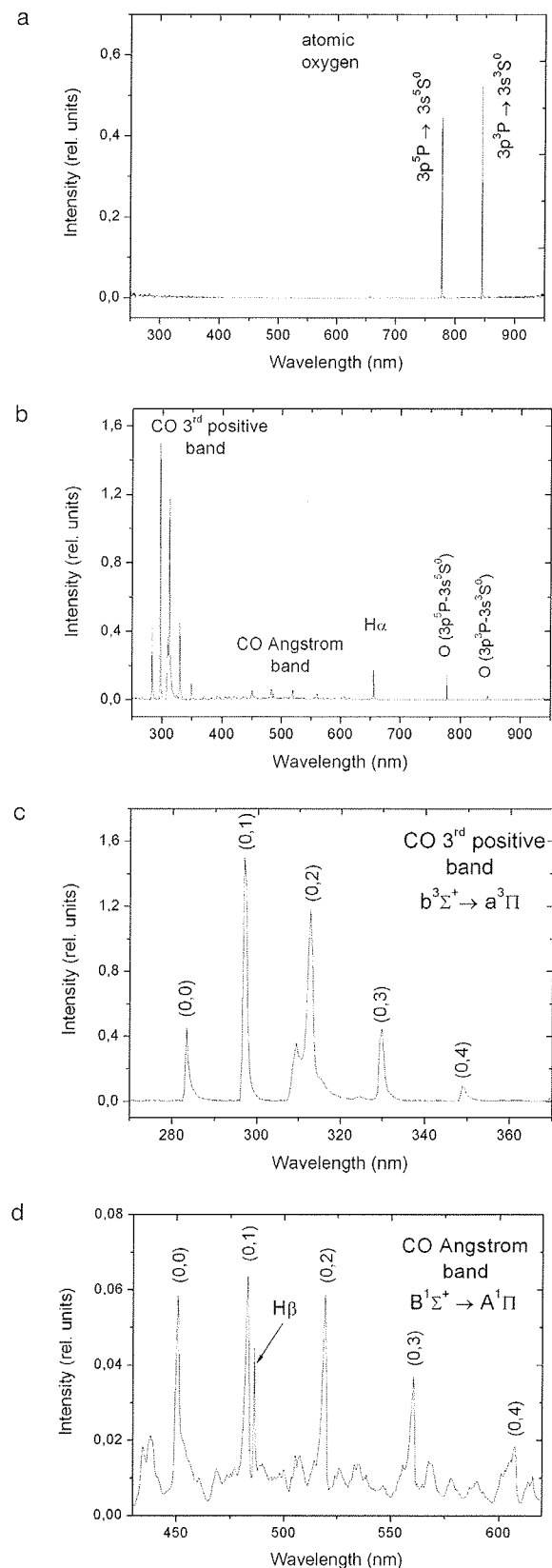
Fig. 6 Optical emission spectroscopy of oxygen plasma. a) before etching of organic material, b) during extensive etching, c) detail spectrum of the CO band corresponding to transitions within the 3rd positive system, d) detailed spectrum of the Angstom band.

Apart from organic materials, other forms of carbon are also activated by oxygen plasma treatment, but the required treatment time may be longer. Figure 5 demonstrates activation of pure electrolytic graphite. In this case, the required treatment time is about 20 s. Figure 5a represents the high resolution XPS O1s peak before and after the plasma treatment, while Figure 5b represents the C1s peak. As expected, there is no oxygen bonded to untreated graphite, but the 20 s treatment cause an appearance of oxygen bonded in the form of C - O bond.

4.2 Polymer etching

As the surface is saturated with chemically bonded oxygen, the next step is slow chemical etching of carbon material. The etching rate depends on the type of material, its temperature, and the flux of oxygen atoms on the surface of the sample. The evidence of etching can be obtained by Optical Emission Spectroscopy. A simple optical spectrometer is good enough to monitor the appearance of the oxidation products. Figure 6a is a typical OES spectrum

a

b
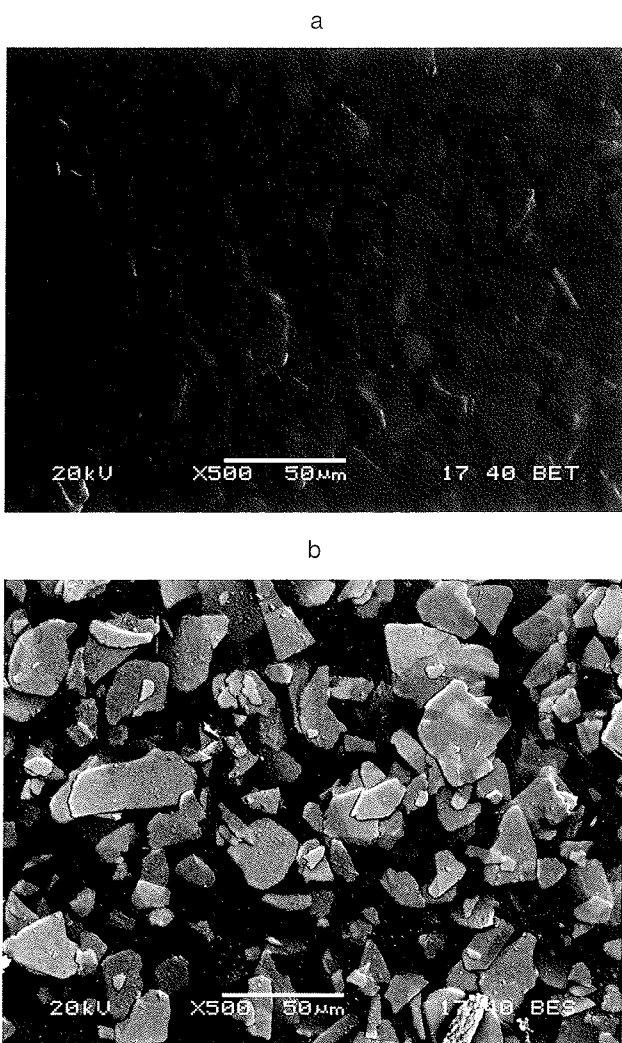
Fig. 7.   SEM image of a metal paint with a pearl effect. a) untreated sample, b) sample treated in oxygen plasma for 40s. The O atom density is $8x10^{21}$ m^{-3}.

a

b
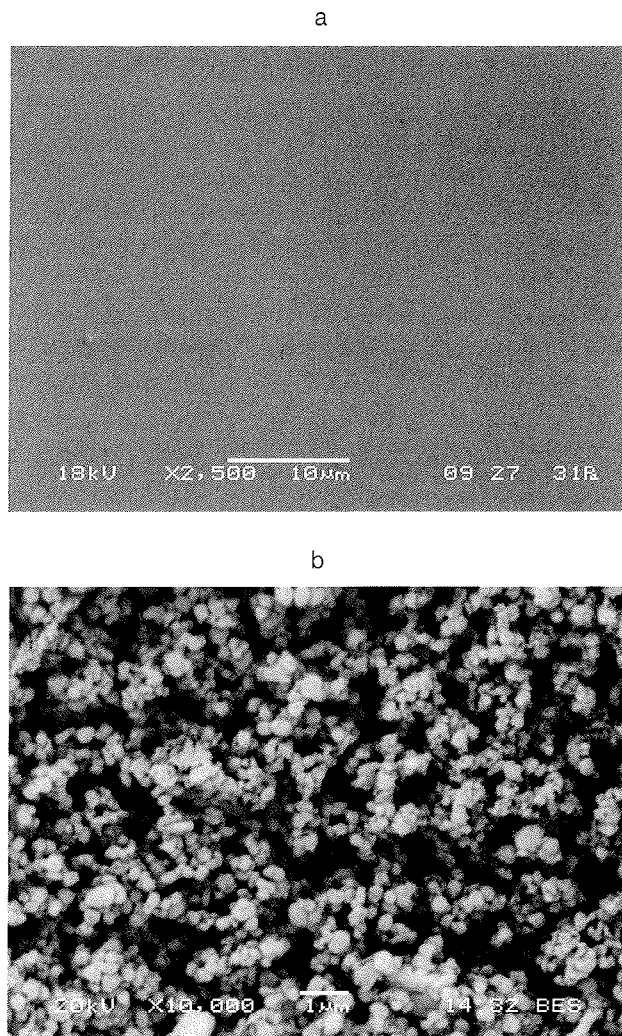
Fig. 8.   SEM image of a powder coating. a) untreated sample, b) sample treated in oxygen plasma for 300s. The O atom density is $6x10^{20}$ m^{-3}.

from oxygen plasma before etching of organic materials. The only features worth mentioning are the oxygen atoms peaks at 777 nm and 845 nm. Other peaks are so small that cannot be visible in Figure 6a. As etching of organic materials starts, the OES spectrum becomes richer as shown in Figure 6b. The peaks that correspond to CO bands appear. Figure 6c is a detail of the CO band corresponding to the transition in the 3rd positive system, while Figure 6d represents the emission due to CO transitions in the Angstrom band. Optical emission spectroscopy therefore represents a powerful tool for detection etching of carbon-containing compounds. Since CO is normally not presented in oxygen plasma in reasonable concentration, the appearance of the CO bands indicate etching of carbon containing materials.

4.3 Selective etching of composites

Composite materials with a polymer matrix are nowadays widely used as bulk materials and films. They combine the properties of polymers (easy instillation, low price) and the

a

b
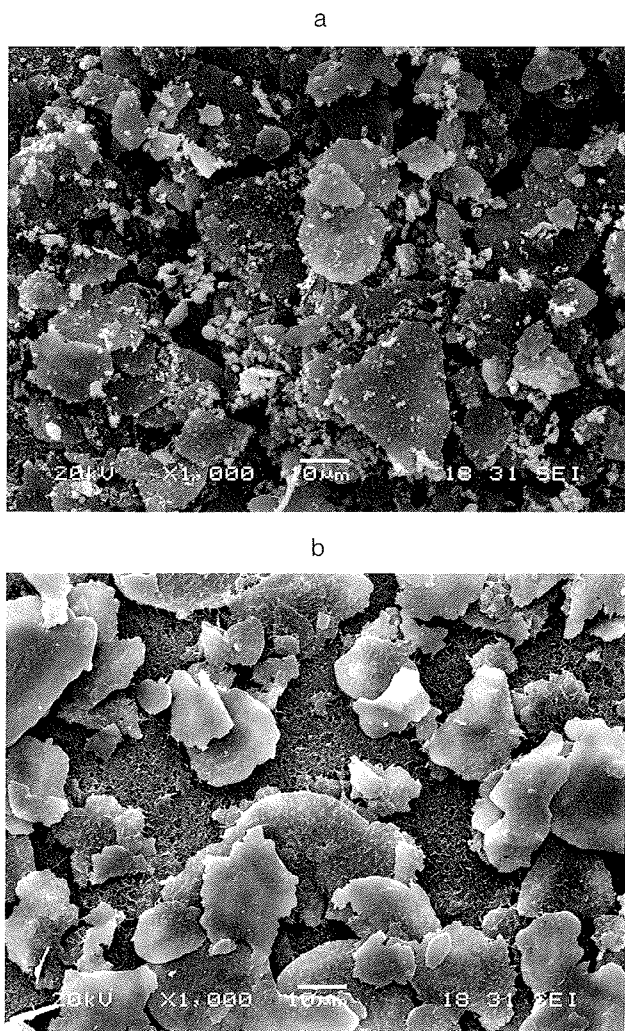
Fig. 9.   SEM image of similar samples produced by two
different producers. The plasma parameters
were as follows: treatment time 120s, O atom
density $2x10^{21}$ m^{-3}.

fillers (good mechanical, electrical and optical properties). Many materials referred as plastics are actually composites. The most usual fillers are pigments – they are used to give the material its colour. Different fibres are used to increase the material strength and toughness. Graphite is

often used to increase electrical conductivity, while different coatings and paints are actually composites with a variety of fillers. A typical paint for metal, for instance is a composite of et least 5 different fillers distributed in the polymer matrix. The characteristics of the composites depend on the type of polymer, the type and concentration of the fillers, and the production procedure. They often depend also on distribution and in some cases even orientation of the fillers in the polymer matrix. The term "pigment dispersion" in paint coatings describes the relative amount of pigment aggregates and agglomerates in solid media. It affects gloss and haze of the final coating and may also change viscosity of the coating. Dispersion of pigments is influenced by the properties of coating components and by the production process.

It is difficult to detect the exact distribution of fillers in the polymer matrix. Namely, the uppermost material is polymer that hides the fillers. The best way of making fillers visible is a gentle removal of the polymer. The treatment should leave the fillers intact and should be performed at low temperature to avoid any deformation of the polymer matrix. The best way of doing so is to perform low temperature etching by inductively coupled oxygen plasma. As mentioned before, the interaction of such oxygen plasma with the solid material is almost entirely potential: oxygen atoms react with polymer matrix while leaving the fillers fairly intact. The first example of such treatment is a coating with mica flakes. Figure 7a is the SEM image of the untreated sample, while Figure 7b represents the image of the same sample exposed to oxygen plasma for 30 s. Comparison of Figures 7a and 7b clearly shows a high etching selectivity obtained by inductively coupled oxygen plasma: the polymer is effectively etched, while the mica flakes are left intact.

Mild oxygen plasma treatment is even selective enough to have different etching rates for two different organic materials. Figure 8 represents results of the etching of paint for metal. Figure 8a is the SEM image of untreated sample, while Figure 8b represents the same paint after plasma treatment. While the original surface is perfectly flat (covered by the polymer), the surface of plasma treated sample is covered with small perfectly spherical features. These spheres are actually organic pigments. The structure of
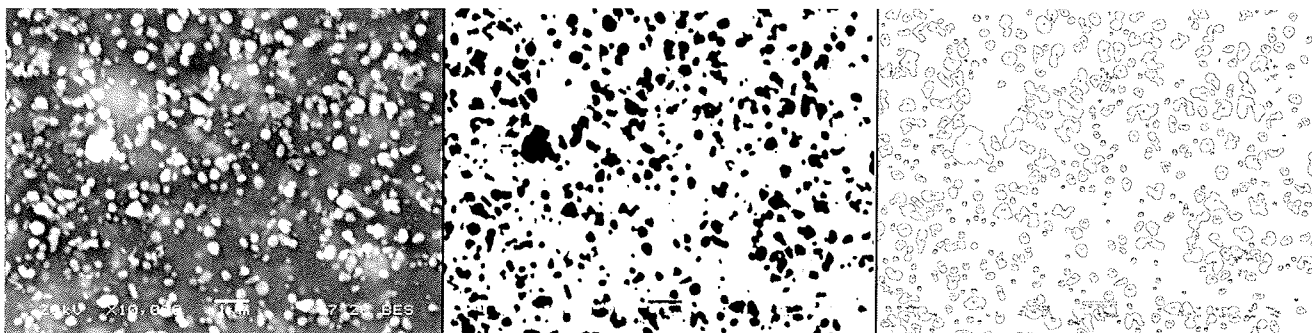
Fig. 10.   SEM image of a sample before and after the image modification and outlined particles detected with
appropriate program.

the organic pigments is different from the polymer matrix. The oxygen plasma is obviously gentle enough to distinguish between the polymer and the organic pigment. While polymer is etched at a high rate, the pigments are practically not etched at all, as observed by comparison of Figure 8 a and b.

The oxygen plasma treatment is used for study of fillers type, size and distribution in composites in order to detect small variations of similar products produced by different procedures. Figure 9 represents SEM images of similar products by two different producers. The similarity of the products is so high that classical testing of the two materials does not show any appreciable difference between them. The oxygen plasma etching, however, reveals small but important differences between the products in Figure 9 a and b. The product presented in Figure 9a is rich in small particles with the dimension of about 1êm, while these particles are absent in the product shown in Figure 9b. Plasma etching of such products actually enables the reverse engineering. If the type and concentration of fillers in a product are unknown, one can learn about them by etching the material with oxygen plasma and image the surface by a SEM.

Advanced software allows for quantification of the filler distribution in the polymer matrix composites. Figure 10 shows the appropriated steps done to study the size distribution of filler grains. The left figure is the SEM image of the plasma etched material. The middle figure is the negative contrast black/white micrograph. The right figure 3 shows the results of particle detection using the appropriate software program. The results of the image analysis of these three micrographs are presented in the Figure 11, where the size distribution of pigment particles is shown. Large particles, of diameter greater than 2 mm, were excluded from the calculation because of their small population in the micrographs. The results in Figure 11 are shown for three mate-

rials with same ingredients but different production procedure.

5   Conclusions

Highly reactive oxygen plasma is obtained in inductively coupled RF discharges. The degree of ionization is often of the order of 10^{-6} while the dissociation fraction easily exceeds 10%. The neutral gas temperature is kept close to room temperature since there is practically no mechanism heating the neutral gas. The high dissociation rate, low kinetic temperature, low ion density and low plasma potential allow for practically pure potential interaction of plasma radicals with solid materials exposed to plasma. The etching of samples is extremely selective: while organic materials are etched, the inorganic materials are not etched at all. The etching rate depends on the type of organic materials. Simple polymers are etched at relatively high rate while graphite is practically not etched at all. Even different types of organic materials are etched at different rates. While epoxy resign is etched at a relatively high rate, the organic pigments are virtually untouched. This extremely high etching selectivity allows for development a method for determination of the distribution and orientation of fillers in polymer matrix composites. Several examples of the practical application of this technology are presented. The choice of plasma parameters depend on characteristics of particular samples. High etching rate is obtained using plasma with a high atom density, but extremely high etching selectivity is obtained using plasma with a moderate O density.

References

/1/   M. R. Wertheimer, L. Martinu, E. M. Liston, Plasma sources for polymer surface treatment, Handbook of thin film Process Technology, ed. by D.A.Glocker and S.I. Shah. Bristol, Inst. of Physics Publishing, Bristol, 1998.

/2/   K. G. Pruden, G. B. Raupp, S. P. Beaudoin, J. Vac. Sci. Technol. B, vol. 21, pp. 1496, 2003.

/3/   J. S. Kim, F. Cacialli, R. Friend, Thin Solid Films, vol. 445, pp. 358, 2003.

/4/   S. Gomez, P. G. Steen, W. G. Graham, Appl. Phys. Lett., vol. 81, pp. 19, 2002.

/5/   H. Singh, J. W. Coburn, D. B. Graves, J. Appl. Phys., vol. 88, pp. 3748, 2000.

/6/   D. J. Wilson, N. P. Rhodes, R. L. Williams, Biomaterials, vol. 24, pp. 5069, 2003.

/7/   Z. Y. Wu, N. Xanthopoulos, F. Reymond, J. S. Rossier, H. H. Girault, Electrophoresis, vol. 23, pp. 782, 2002.

/8/   M. Mozetic, A. Zalar, Mater. Sci. Forum, vol. 437, pp. 81, 2003.

/9/   A. G. Whittaker, E. M. Graham, R. L. Baxter, A. C. Jones, P. R. Richardson, G. Meek, G. A. Campbell, A. Aitken, H. C. Baxter, J. Hosp. Infect., vol. 56, pp. 37, 2004.

/10/   K. Gorna, S. Gogolewski, Polym. Degrad. Stabil., vol. 79, pp. 475, 2003.

/11/   M. Nagatsu, F. Terashita, Y. Koide, Jap. J. Appl. Phys, vol. 42, pp. L856, 2003.

/12/   M. K. Gunde, M. Kunaver, Appl. Spect., vol. 57, pp. 1266, 2003.

Fig. 11. Particle diameter distribution in samples produced by three different procedures marked as Y1, Y2 and Y3.

/13/   M. Kunaver, M. K. Gunde, M. Mozetic, M. Kunaver, A. Hrovat, Surf. Coat. Int. B, vol. 86, pp. 175, 2003.

/14/   M. Mozetic, U. Cvelbar, M. K. Sunkara, S. Vaddiraju, Adv. Mater., vol. 17, pp. 2138, 2005.

/15/   A. Vesel, M. Mozetic, A. Drenik, S. Milosevic, N. Krstulovic, M. Balat-Pichelin, I. Poberaj, D. Babic, Plasma Chem. Plasma Proc., vol. 26, pp. 577, 2006.

/16/   A. Vesel, M. Mozetic, A. Zalar, Appl. Surf. Sci., vol. 200, pp.94, 2002.

/17/   M. Mozetic, Vacuum, vol. 71, pp. 237, 2003.

/18/   U. Cvelbar, D. Vujosevic, Z. Vratnica, M. Mozetic. J. Phys. D Appl. Phys., vol. 39, pp. 3487, 2006.

/19/   A. Ricard, Reactive plasmas, SFV, 1996.

/20/   A. Ricard, V. Monna, Plasma Sourc. Sci. Technol., vol. 11, pp. A150, 2002.

/21/   A. Vesel, M. Mozetic, Vacuum, vol. 61, pp. 373, 2001.

/22/   S. Villeger, S. Cousty, A. Ricard, M. Sixou, J. Phys. D Appl. Phys., vol. 36, pp. L60, 2003.

/23/   A. Drenik, U. Cvelbar, A. Vesel, M. Mozetic, Inf. MIDEM, vol. 35, pp. 85, 2005.

/24/   J. J. Robbins, R. T. Alexander, W. Xiao, T. L. Vincent, C. A. Wolden, Thin Solid Films, vol. 406, pp. 145, 2002.

/25/   M. C. Kim, S. H. Yang, J. H. Boo, J. G. Han, Surf. Coat. Technol., vol. 174, pp. 839, 2003.

/26/   A. Granier, M. Vervloet, K. Aumaille, C. Vallee, Plasma Sourc. Sci. Technol., vol. 12, pp. 89, 2003.

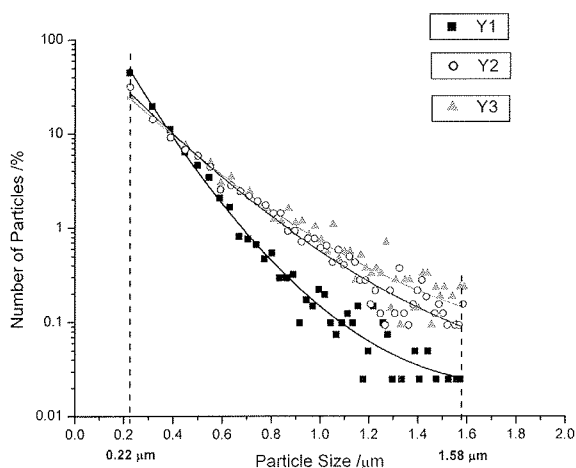/27/   H. Biederman, V. Stelmashuk, I. Kholodkov, A. Choukourov, D. Slavinska, Surf. Coat. Technol., vol. 174, pp. 27, 2003.

/28/   G. F. Leu, A. Brockhaus, J. Engemann, Surf. Coat. Technol., vol. 174, pp. 928, 2003.

/29/   Y. C. Hong, H. S. Uhm, Phys. Plasmas, vol. 10, pp. 3410, 2003.

/30/   G. B. I. Scott, D. A. Fairley, D. B. Milligan, C. G. Freeman, M. J. McEwan, J. Phys. Chem. A, vol. 103, pp. 7470, 1999.

/31/   L. Elias, E. A. Ogrizlo, H. I. Schiff, Can. J. Chem., vol. 37, pp. 1680, 1959.

/32/   M. R. Carruth, R. F. DeHaye, J. K. Norwood, A. F. Whitaker, Rev. Sci. Instr., vol. 61, pp. 1211, 1990.

/33/   I. Sorli, R. J. Rocak, Vac. Sci. Technol. A, vol. 18, pp. 338, 2000.

/34/   M. Mozetic, A. Vesel, M. Gaillard, A. Ricard, Plasmas Polym., vol. 6, pp. 41, 2001.

/35/   D. Babič, I. Poberaj, M. Mozetič, Rev. Sci. Instr., vol. 72, pp. 4110, 2001.

/36/   I. Poberaj, D. Babič, M. Mozetič, J. Vac. Sci. Technol. A, vol. 20, pp. 189, 2002.

/37/   M. Mozetic, A. Ricard, D. Babic, I. Poberaj, J. Levaton, V. Monna, U. Cvelbar, J. Vac. Sci. Technol. A, vol. 21, pp. 369, 2003.

/38/   M. Balat-Pichelin, A. Vesel, Chem. Phys., vol. 327, pp. 112, 2006.

/39/   A. Vesel, M. Mozetic, M. Balat-Pichelin, Vacuum, in press, (doi:10.1016/j.vacuum.2007.02.003), 2007.

/40/   M. Mozetic, A. Vesel, U. Cvelbar, A. Ricard, Plasma Chem. Plasma Process., vol. 26, pp. 103, 2006.

/41/   M. Mozetic, A. Vesel, V. Monna, A. Ricard, Vacuum, vol. 71, pp. 201, 2003.

/42/   M. Mozetic, U. Cvelbar, A. Vesel, A. Ricard, D. Babic, I. Poberaj. J. Appl. Physic., vol. 97, pp. 103308-1, 2005.

/43/   T. Vrlinic, A. Vesel, U. Cvelbar, M. Krajnc, M. Mozetic, Surf. Interface Anal., in press, (DOI: 10.1002/sia.2548), 2007.

/44/   A. Vesel, M. Mozetic, A. Zalar, Vacuum, in press, 2007.

/45/   A. Vesel, M. Mozetic, J. Kovac, A. Zalar, Appl. Surf. Sci., vol. 253, pp. 2941, 2006.

/46/   M. Mozetic, A. Zalar, P. Panjan, M. Bele, S. Pejovnik, R. Grmek, Thin solid films, vol. 376, pp. 5, 2000.

/47/   M K. Gunde, M. Kunaver, M. Mozetic, P. Pelicon, J. Simcic, M. Budnar, M. Bele, Surf. Coat. Int., Part B, Coat. Trans., vol. 85, pp. 115, 2002.

/48/   M. Kunaver, M. K. Gunde, M. Mozetic, A. Hrovat, Dyes Pigm., vol. 57, pp. 235, 2003.

/49/   M. K. Gunde, M. Kunaver, M. Mozetic, A. Hrovat, Powder Technol., vol. 148, pp. 64, 2004.

/50/   M. Kunaver, M. Mozetic, M. K. Gunde, Thin solid films, vol. 459, pp. 115, 2004.

/51/   N. Krstulovic, I. Labazan, S. Milosevic, U. Cvelbar, A. Vesel, M. Mozetic, J. Phys. D Appl. Phys., vol. 39, pp. 3799, 2006.

P. Eiselt
Plasmabull Engineering GmbH, Parkring 6,
A – 8403 Lebring, Austria

GRAPHICAL FRAMEWORK FOR SYSTEM LEVEL DESIGN SPACE EXPLORATION

Klemen Perko[1], Andrej Trost[2]

[1]Sipronika d.o.o., Ljubljana, Slovenia
[2]Faculty of Electrical Engineering, Laboratory for Integrated Circuits Design, University of Ljubljana, Ljubljana, Slovenia

Key words: abstraction, design-space exploration, graphical modeling, high-level design, system-level simulation.

Abstract: As technology advances, options for realization of heterogeneous systems increase. Designers use a variety of hardware (HW) and software (SW) co-design methodologies in order to meet application constraints as fast as possible. The paper presents a graphical modeling framework used for high-level modeling and design-space exploration of heterogeneous systems. The framework provides designer graphical elements for using modeling concepts from system modeling libraries. Graphical modeling relieves the designer of the manual-typing source code and thus hides many details of system-level design languages that normally need to be taken care of. The graphical framework also provides different constraint checks during modeling and automatically generates an executable model for evaluation of a heterogeneous system. Our case study exemplifies the use of the framework and shows what information is obtained from an executable model built on a high-level of abstraction. Evaluation of results serves as a basis for further design decisions. Graphical modeling enables rapid changes in the model and thus speeds-up design-space exploration.

Grafično okolje za raziskovanje načrtovalskega prostora na nivoju sistemov

Kjučne besede: abstrakcija, modeliranje, raziskovanje načrtovalskega prostora, visokonivojsko načrtovanje.

Izvleček: S tehnološkim napredkom se povečuje nabor možnih realizacij heterogenih sistemov. Načrtovalci za čimprejšnje izpolnjevanje načrtovalskih zahtev uporabljajo širok spekter metodologij za sočasno načrtovanje strojne in programske opreme. Članek predstavlja grafično modelirno okolje za modeliranje in raziskovanje načrtovalskega prostora heterogenih sistemov. To okolje omogoča načrtovalcu uporabo grafičnih elementov pri modeliranju konceptov iz knjižnic za modeliranje na sistemskem nivoju. Grafično okolje načrtovalca razbremenjuje ročnega pisanja programske kode, tako da mu ni več potrebno poznati točne sintakse ukazov programskih jezikov za modeliranje na sistemskem nivoju. Okolje med izdelavo modela preverja njegovo skladnost z različnimi omejitvami. Po končanem modeliranju heterogenega sistema, za njegovo ovrednotenje okolje avtomatsko ustvari izvršljiv model. Uporaba okolja je prikazana na praktičnem primeru. Prikazano je, katere informacije dobimo iz izvršljivega modela zgrajenega na visokem nivoju abstrakcije. Ovrednoteni rezultati predstavljajo podlago nadaljnjim načrtovalskim odločitvam. Grafično modelirno okolje omogoča hitre spremembe modela in tako pospeši raziskovanje načrtovalskega prostora.

1 Introduction

Advances in technology provide various options for realization of embedded systems. Designers are encouraged to use a variety of HW and SW implementation technologies in order to meet application constraints and provide quick time-to-market solutions. The increasing complexity of modern embedded systems requires new design methodologies and system-level design tools /1/.

Many research studies are concentrated on the issues of HW/SW co-design, co-simulation and various optimization techniques. The research activity is slowly drifting away from modeling heterogeneous aspects of the system towards system description on a higher abstraction level /2/.

This paper will present a design framework for system-level design space exploration. The presented framework is used for a quick evaluation of design decisions in the first stages of the design process. The evaluation is based on the results obtained from a high-level model of the system composed in a graphical framework.

2 Design space exploration

HW and SW components of digital systems are designed by using specialized languages. The HW description language VHDL /3/ or Verilog is used for design and implementation of HW components and the C or C++ is used for SW description. These languages are mature and provide automatic implementation and various optimization possibilities.

On the system level, we need tools and languages for modeling systems composed of HW and SW components. The result of research in this area is several system level design languages (SLDL) and HW/SW co-design methodologies /4/.

A typical design flow starts with a high-level system model containing architecture description, functionality description and mapping information /5/. During design-space exploration, the model is repeatedly evaluated and changed until the application constraints are met. If the design methodology supports different levels of abstraction, we have

to repeat design-space exploration on each level. In each level we add new information thus lowering the level of abstraction. Finally, a description of HW and SW components prepared for automatic implementation tools is obtained.

2.1 SystemC Design Flow

The SystemC is a system-level description language based on the C++ language. The programming language C++ can be used also as an extensible object-oriented modeling language. The SystemC extends the capabilities of the C++ by enabling modeling of hardware descriptions /6/, /7/.

The SystemC language is implemented as a C++ class library. It adds important concepts to the C++ such as concurrent processes execution, modeling timed events and hardware data types. The SystemC enables designer to describe the whole system model in one language, verify it by using the same language, and further refine it all the way to the implementation level (typically the register transfer level). A system can be modeled at the behavioral or architectural level and then iteratively refined to the register transfer level.

Building a detailed model of an embedded system in the SystemC can be a very time-consuming task. In order to speed-up the design space exploration process, we need to identify important modeling concepts for the model evaluation at the current abstraction level. When a satisfactory model is obtained, more details can be described (for example timing and communication) and the design-space exploration is repeated on a lower level of abstraction /5/. Model refinement continues until all the details necessary for implementation are obtained.

In this paper we will focus on design exploration on the highest abstraction level. In the first stage of the design process we can identify some concepts repeatedly needed by designers for any new model. A model of an embedded system is composed of HW (architecture) and SW (functionality) units. SW units are running on the model architecture using its resources. SW can be further modeled as a composition of some tasks. A high-level architectural model contains execution (processing), communication and data storage units. To relieve the designer of the burden of repeatedly implementing models of these basic concepts in the SystemC, system modeling libraries supporting them were developed in our Laboratory /8/, /9/. They provide wrappers for modeling functionality and architecture on a high-level of abstraction.

System functionality is intuitively described as a network of tasks. The tasks are modeled in terms of architectural resource usage and no actual algorithm is specified. Each task is assigned an execution unit responsible for characterizing the cost of executing services (e.g. time, energy and size by means of logical blocks or transistors count). Execution and communication units are parts of the architecture description.

A library with functionality wrappers provides mechanisms for modeling parallel task execution. Event modeling is used for triggering task execution. From the functionality point of view, the task execution is limited only by their data dependency. The maximum level of parallel operations that a specific algorithm permits is first examined and later decreased by applying restrictions of execution units. When more than one task specifies the same execution unit, it is up to the execution-unit scheduling policy to determine the outcome of such request.

An architecture-wrappers library provides support for high-level modeling of architectural resources. Using these wrappers, designers can instantiate and connect any number of hardware units in their model and build an architectural model. Concepts of execution and communication units present HW resources and give the designer only information about architecture resource utilization in interaction with algorithm functionality.

One of the integral parts of our libraries is also a built-in support for logging relevant information about the system during execution of simulation. The system modeling libraries enable a component-based construction of the system model at a higher abstraction level. The concept of components promotes reuse of the already developed models which can leverage design productivity.
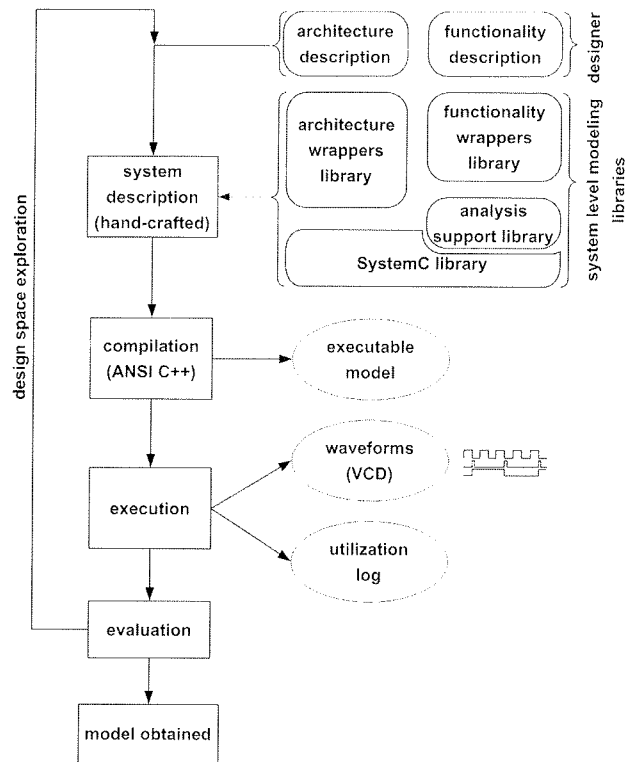
Fig. 1.   Design flow using a system modeling library

The design-space exploration flow starts by composing a system model containing descriptions of architecture and functionality by using prepared wrappers from the system modeling library (see Figure 1). The model composition
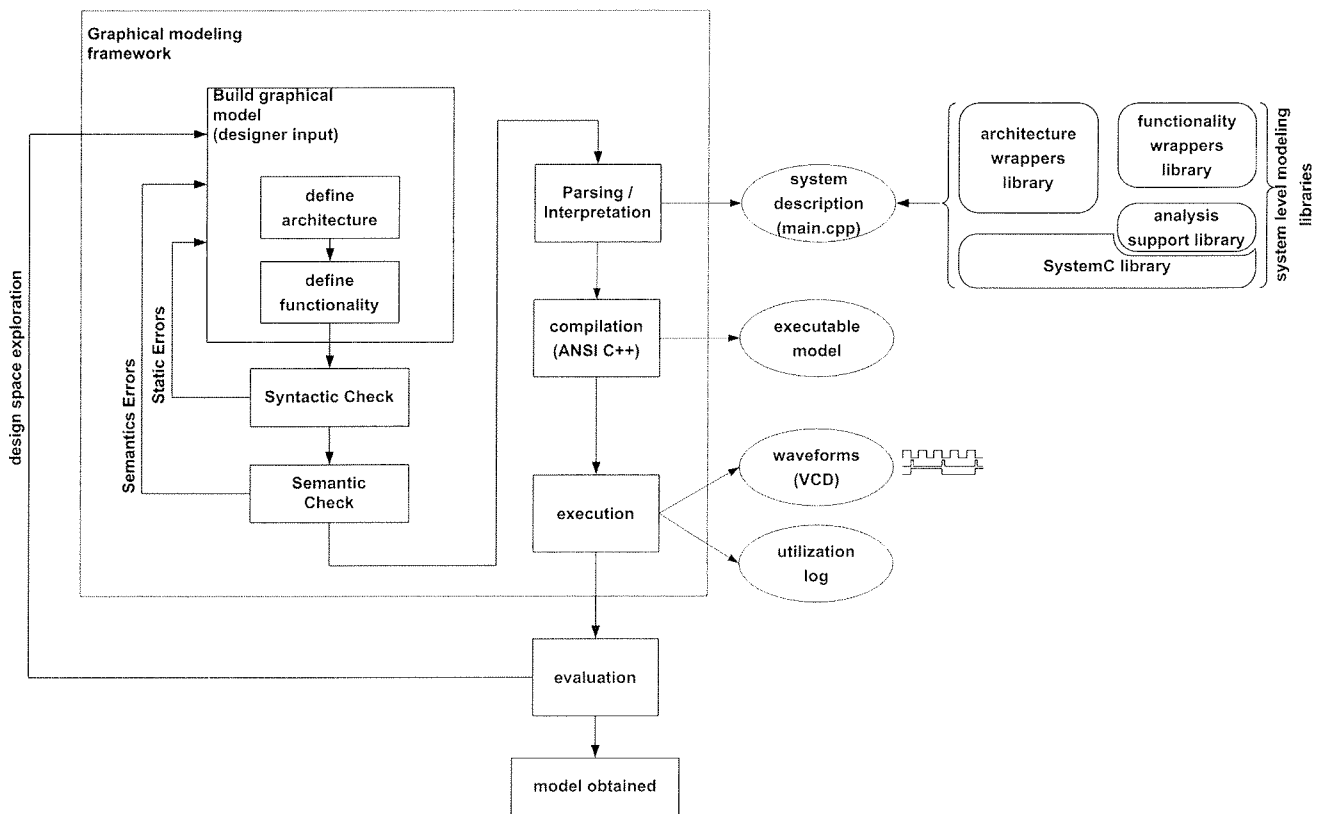
Fig. 2.   Graphical framework for design space exploration

process can be divided into four steps:

- In the first step, the designer builds functionality. He/she defines descriptions of each task (time estimations for HW resource usage) and its time dependency (task-start triggers).

- In the second step, the designer determines the system architecture.

- In the third step, the designer specifies mapping of all tasks to appropriate architecture resources.

- In the final step, the designer defines the necessary simulation settings (e.g. simulation step size, max. simulation time, which variables to log into waveforms, and settings for reports of resource utilization).

The design flow continues with compilation of the system model together with wrappers and the SystemC library. An executable model is obtained which produces upon execution waveforms and resource utilization files.

In the next stage, data relevant for further design decisions (e.g. resource utilization, task being idle because of resource contention) are evaluated. Evaluation results are compared with the system specification constraints. If they are not satisfactory, the designer repeats the design cycle with a different system implementation.

During design-space exploration various system implementations can be relatively easy to build and feedback about their evaluations results can be used for finding a path towards a solution best meeting system constraints /8/, /9/.

2.2   Graphical Design Flow

While the system modeling libraries provide a great support for simulation and evaluation, the designer still needs to manually describe the system by means of coding. Consequently, this means that coding has to be changed in each repetition of the design cycle, which is an error-prone process. For designers this still represents a heavy burden for quick and efficient design space exploration.

To simplify and speed-up the process of building a system model and enable its faster exploration we developed a domain-specific graphical modeling framework (GF). The framework enables a graphical creation of a high-level system model and interpretation of the model into the SystemC source code. The system architecture is described by inserting and interconnecting reusable library components in a graphical framework. The system functionality is defined in tasks written in the SystemC and presented as blocks in the graphical framework. Connections are used for a graphical presentation of the tasks time dependency and mapping to architectural resources.

The modeling framework also supports simulation settings, selection of reports and variables being logged during simulation execution and definition of stimulators for simulating external signals that this model is dependent of. The graphical framework performs different syntactic checks during model building and interpretation phase thus minimizing designer errors. In this way it greatly helps designers to build an appropriate model more quickly. Automation of compilation and execution stages is also supported.

3 Graphical modeling framework

Graphical modeling environments are used extensively in different domain-specific areas (e.g. Matlab/Simulink for signal processing).

When a system-model developer decides to switch from the design language to a graphical framework, he/she can take one of the two different approaches; either starts developing a new domain-specific GF from the ground up or using one of the already developed generic graphical frameworks that can be configured for particular domain-specific needs. Each approach has some advantages and disadvantages over the opponent.

The first approach allows the developer to fully control his/her design. As developing such framework is quite expensive, this approach is limited to applications with large potential market. The cost of the second approach is lower and the developer's control over the framework is limited. Only toolsets offered by a selected generic framework can be used. This approach is much more appropriate for applications needing only a small amount of installations.

Open source graphical modeling environments found suitable for us are Eclipse Graphical Editing Framework /10/ and Generic Modeling Environment (GME) /11/. We decided to use GME since it is more mature, offers very good user support through online forum and provides tools for easy integration of the interpreter for translating the graphical model.

3.1 Generic Modeling Environment - GME

The Generic Modeling Environment (GME) /12/ is a configurable toolkit used for creating domain-specific modeling, model analysis, model transformation and program synthesis environments. The configuration is accomplished through meta-models specifying the modeling paradigm (modeling language) of the application domain. The modeling paradigm contains all the syntactic, semantic and presentation information regarding the application domain. It defines concepts used to construct models, their relationship, organization and graphical presentation, and rules governing model construction.

The modeling paradigm is created by configuring a meta-model using the GME meta-modeling language. Meta-models are used to automatically generate target domain-specific environment. An interesting aspect of this approach is that the environment itself is used to build meta-models. This top-level environment is called a Meta-metamodel.

The generated domain-specific environment is then used to build domain models that are stored in the model database. They are used to automatically generate applications or to synthesize input to different Commercial Off-The-Shelf (COTS) analysis tools. This process is called model interpretation.

Figure 3 depicts how GME is configured to suit domain-specific modeling environment needs. The role of the meta-
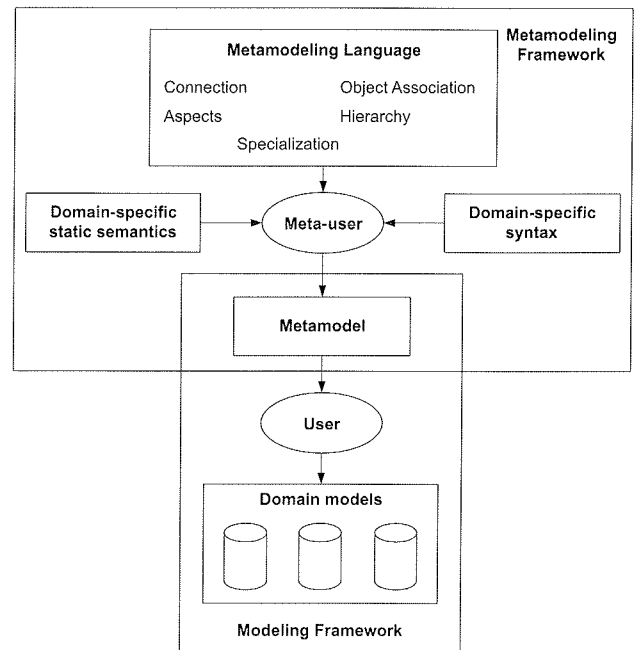
Fig. 3. Configuration of GME needed to obtain a domain-specific modeling framework

user is to construct a domain-specific meta-model with all syntactic, static semantic and presentation information regarding this specific domain. The meta-modeling paradigm is based on the Unified Modeling Language (UML) /14/. The syntactic definitions are modeled using pure UML class diagrams and the static semantics are specified with constraints using the Object Constraint Language (OCL). This process needs to be done just once and the developer of a domain-specific modeling framework takes over the role of a meta-user. Users of this domain-specific framework can build their specific models according to rules defined in the meta-model.
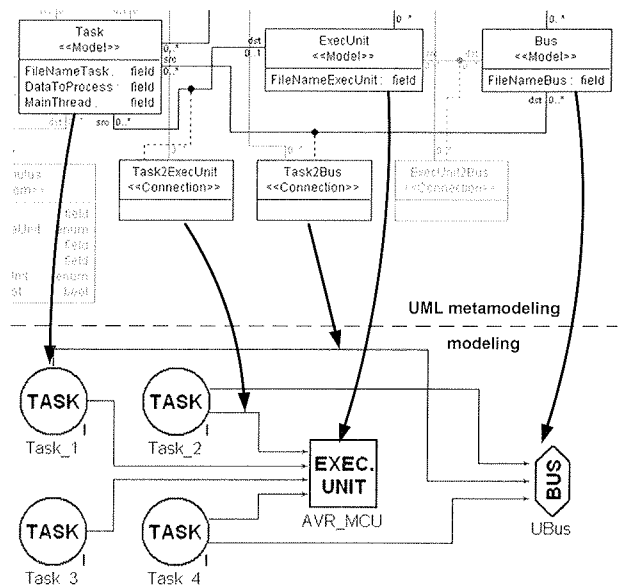
Fig. 4. Creating a domain-specific modeling framework

Figure 4 illustrates a snippet of the UML meta-modeling paradigm and its actual corresponding presentation in GME. The curvy arrows show how individual modeling elements and their relations are defined by different parts of the meta-model.

GME has a built-in set of generic concepts: folders, models, atoms, connections, roles, constraints and aspects. These concepts are the main elements used by the meta-model developer. We will not make a detailed presentation of all of them as this would exceed the scope of this paper. The reader can find it in /12/, /13/. We will just focus on the concept of aspects. Aspects provide visibility control. They are used to allow models to be constructed and observed from different viewpoints. Existence of parts of the domain in a particular aspect is determined by the meta-model. Each part can be either visible or hidden. The concept of aspects allows the user to employ just the parts suited for a selected viewpoint and hide all the others irrelevant for it.

GME also provides high-level C++ and Java interfaces for writing plug-in components to traverse, manipulate and interpret graphical models into an appropriate text description suiting as input to COTS analysis tools. The interpreter needs to be written by the meta-user because interpret-

er must be able to translate graphical models built according to the meta-model.

3.2   Building paradigm

To configure GME for specific needs of our high-level system modeling, we built a meta-model containing information of all the concepts supported in our system modeling libraries. As mentioned above, the libraries provide wrappers for creating abstract HW resource units (execution and communication units) and wrappers for abstract task creation. Tasks serve for creating a description of algorithm functionality.

Figure 5 shows a part of our meta-model designed by using generic concepts of the GME environment and static UML diagram. For clarity of presentation the only most important concepts of our system-level modeling methodology are presented. The meta-model enables a model of a typical embedded system on a high abstraction level to be made-up as a composition of:

-    at least one execution unit (*ExecUnit*),
-    any number of communication units (*CommUnit*), and
-    at least one task (*Task*).
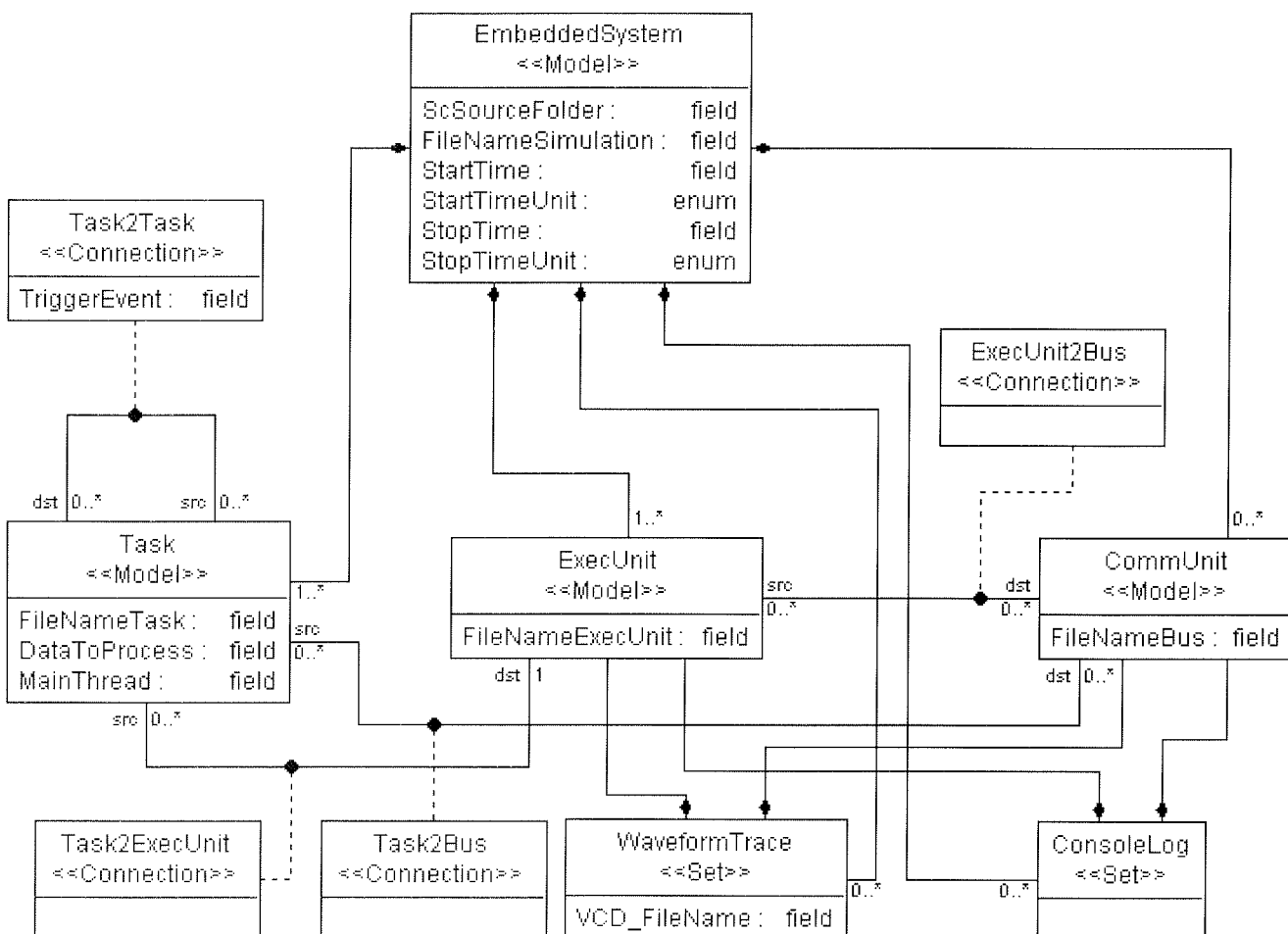
Fig. 5.   Snippet of our meta-model

The restrictions for the numbers of instances in the actual model are set by multiplicity constraints (e.g. constraint for the *ExecUnit* is set to: "1..*"). These three units are directly compatible with wrappers in our system modeling libraries. The meta-model also defines possible connections between these elements. The designer can make just the connections permitted in the meta-model. The connections shown in Figure 5 are:

- *ExecUnit2CommUnit*: with these connections the designer defines the communication units available for a selected execution unit. Generally an execution unit can have more than one communication unit and many different execution units can share the same communication units. Instances of execution and communication units connected together compose system architectural resources.

- *Task2Task*: with these connections the designer defines the order of task execution. The order is governed by the tasks' data dependency and the direction from the source to destination has to be followed. Instances of the tasks connected together with the *Task2Task* connections compose system functional description.

- *Task2ExecUnit*: with these connections the designer assigns execution units responsible for execution of a selected task. Each task can be assigned to only one execution unit.

- *Task2CommUnit*: with these connections the designer defines the communication units available for data-manipulation operations of tasks. Generally, a task can use more than one communication unit, but only those available to the assigned execution unit can be used. This means that the designer can select only between those communication units that have been previously attached with *ExecUnit2CommUnit* to the execution unit. Actual constraints for creation of these connections are implemented in a syntactic check performed before starting the model interpretation.

Two blocks representing the GME concepts of *sets* are also shown in Figure 5. The sets are used for selecting and grouping object instances in the system model. With the set *WaveformTrace* the designer defines which HW resources will be traced during simulation. Multiple *WaveformTrace* sets with different members can be used. Each of them represents a different VCD (Value Change Dump) waveform file. A *ConsoleLog* set defines the HW resources used for printing the resource utilization log. This information is gathered and printed after the actual simulation ends.

Besides the presented blocks, the meta-model contains also some other elements required for model construction and simulation setup. All of them are listed in Table 1. The event splitter and event joiner are used for defining the order of task execution. The event joiner performs an addition of multiple input events when starting a specific task depends on execution ending of multiple tasks. Event split-

ter triggers multiple tasks in a certain order and can be used for modeling a SW scheduler. Start and stop events are used for control of the simulation process. External event-generator elements serve for imitating input signals coming from the surroundings where our system will be operating.

The concept of aspects in GME provides visibility control. The aspects allow models to be constructed and viewed from different viewpoints. They show only elements relevant in a particular aspect. In our meta-model we implemented four different aspects in which a model of an embedded system can be viewed.

- In the task triggering aspect, the designer enters functionality of the system by placing and connecting task instances. The simulation setup elements (start and stop events) and external-event generators are also defined in this aspect.

- In the architecture aspect, instances of hardware resources (execution and communication units) are placed and connections *ExecUnit2CommUnit* are defined.

- The mapping aspect serves for mapping tasks to appropriate hardware resources. Only connections among the already defined instances can be made.

- In the simulation setting aspect, *WaveformTrace* and *ConsoleLog* set elements are instantiated and their appropriate members defined.

Table 1 lists all of the implemented elements of our meta-model in conjunction with the visibility aspects. Even if a specific element is visible in more than one aspect, it can be instantiated or modified only in its primary aspect. The primary aspect is denoted with a shadowed cell.

Aspect / Visibility	Task Triggering	Architecture	Mapping	Simulation Settings
Task	•		•	
Event Splitter	•			
Event Joiner	•			
Execution Unit		•	•	•
Bus		•	•	•
Start Event	•			
Stop Event	•			
External Event Gen.	•			•
Waveform Trace				•
Console Log (usage)				•

Table 1. Visibility of elements depends on the aspect

For connecting all the elements together, we defined proper connections in the meta-model. As mentioned above, we did not describe all of them since this is not crucial for understanding the idea of our approach. At this point it needs just to be noted that the possibility of making connections also depends on the aspect. All the possible connections implemented in our meta-model and the ability of making them dependent on a particular aspect, are presented in Table 2.

Connection \ Aspect	Task Triggering	Architecture	Mapping
Task2Task	•		
Task2Event Splitter	•		
Task2Event Joiner	•		
EventJoiner2EventSplitter	•		
ExternalEventGenerator2Task	•		
StartEvent2Task	•		
Task2StopEvent	•		
Task2ExecUnit			•
Task2CommUnit			•
ExecUnit2CommUnit		•	

Table 2. Possibility of making connections depends on the aspect

3.3 Model interpretation

Very important part of our graphical modeling framework is the model interpreter. The GME provides high-level C++ and Java interfaces for writing plug-in components to traverse, manipulate and interpret models. The purpose of the interpreter is to translate all information captured in the graphical model into a textual description.

We designed an interpreter which produces a source code description of the system components compatible with the SystemC and our system modeling libraries. The interpreter is capable of handling all the concepts defined in our meta-model. It is written in the C++ and based on the MFC library.

Before an actual interpretation begins, different syntactic and semantic checks are performed to verify the graphical model. Errors are reported and the designer is guided to repair the model. The interpreter generates the SystemC source code together with appropriate project files for automatic compilation and linking. Finally, an executable description of the system model is obtained.

4 Case study

To see how our graphical modeling framework operates in practice, system-level modeling of an existing real-time embedded system will be presented. Since the embedded system is actually already built, its performance can be extracted from the implementation model or measured in the system. Performance estimation before actual implementation was not possible since no modeling framework suitable for heterogeneous system simulation was available at the time. We will show that using our graphical framework for modeling the observed system on a high abstraction level enables performance estimation before the implementation is made. The framework allows very easy exploration of different system implementations.

The case study presents an Illumination and Camera Controller (ICC) /15/ used in computer vision applications for high-speed control of illumination units and triggering line cameras. This is a typical control-oriented embedded sys-

tem where two position encoders are used for triggering events and computing outputs in real time. A USB communication port is available for setting the operating parameters.

Figure 6 shows a hardware platform for implementation of the ICC. The available hardware resources include an AVR microprocessor, CPLD, USB transceiver, RAM data memory and some other peripheral devices. As this system operates in a time-critical environment, it is crucial to assure that it operates as a hard real-time system. For all these reasons it is reasonable to develop it at the system-level.
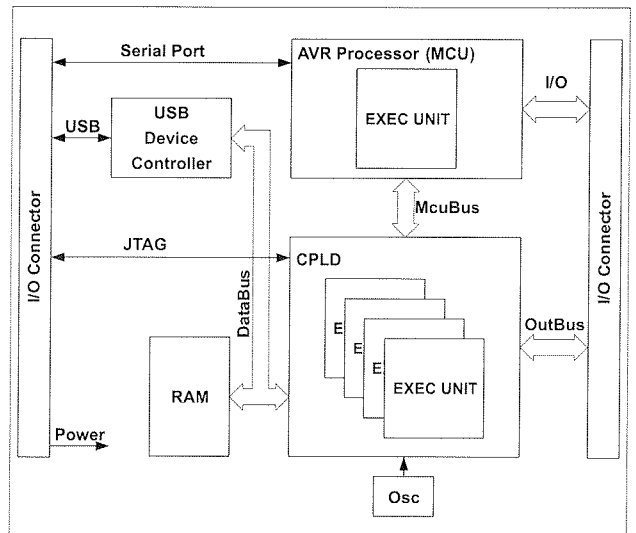
Fig. 6. Platform of illumination and camera controller

4.1 Model construction

Construction of the ICC model on the high abstraction level was performed in four aspects using graphical elements from the meta-model.

In the first aspect, functionality of the ICC is defined. Functionality of the system can be divided into eight different tasks presented in Figure 7. Tasks *T0H* and *T0S* serve for setting the operating parameters after system start-up and for communication with the operator. Task *T0H* performs communication with the USB transceiver and passing of the parameters to task *T0S* which configures the microprocessor. For simulating the parameter setting right after power-up, a start event element *Event_T0* is used.

Operation of the ICC is triggered by two external encoders: absolute position encoder (APE) and incremental position encoder (IPE). We used two external event generator elements *AbsEncIRQ* and *IncEncIRQ* for modeling APE and IPE, respectively. Task *T1* reads IPE events and triggers tasks *T2* and *T3A*. Task *T2* reads new illumination and camera control data from RAM and sends them to the output bus. Since the IPE does not give an absolute position, task *T3A* performs actual re-calculation of the inspected object position based on the data obtained from both encoders. Task *T3B* reads the data from the APE and performs transformation from the Gray to binary code. The

new value of the object position is calculated in task *T3C*. Tasks *T3A* and *T3C* generate output events when a specified position boundary is reached. The events are combined into the *EventJoiner* connected to task *T3D*. This task generates page trigger signals for cameras and resets the illumination control. The output event is in our model connected to the *Event_Stop* element for simulation purposes.
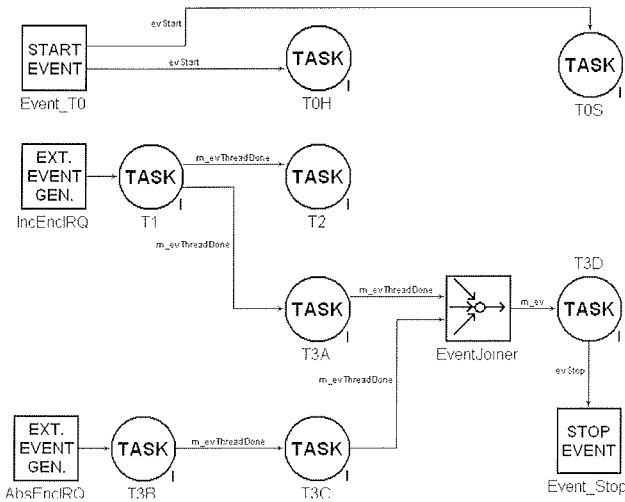
Fig. 7. Functionality description in the task triggering aspect

The possible parallel execution of the tasks is limited by their data dependency and available HW resources. The hardware resources are defined in the architecture aspect by placing instances of execution and communication units. The microprocessor contains only one execution unit capable of executing many different software tasks. On the other hand, the CPLD device can implement more special purpose execution units operating in parallel, but is limited with its size.

In the proposed model, tasks *TOH, T1, T2* and *T3B* are implemented in CPLD. The architecture description contains four CPLD and one AVR execution unit instances, as presented in Figure 8. The ICC platform communication buses *McuBus, DataBus* and *OutBus* are also instantiated and connected to appropriate execution units.
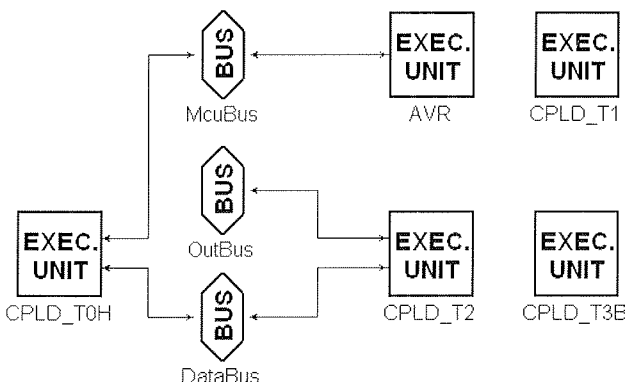
Fig. 8. Architecture of ICC on a high abstraction level

Mapping of the tasks to execution and communication units is actually defined in the mapping aspect, as shown in Figure 9. Each task is mapped to one execution unit and zero or more communication units. The actual use of the communication units is defined in the task description. Tasks *TOS, T3A, T3C* and *T3D* are assigned to execution unit *AVR* representing the microprocessor in the actual ICC system. All other tasks have their own execution units. It should also be mentioned that task *T2* can use two communication units (*DataBus* and *OutBus*).
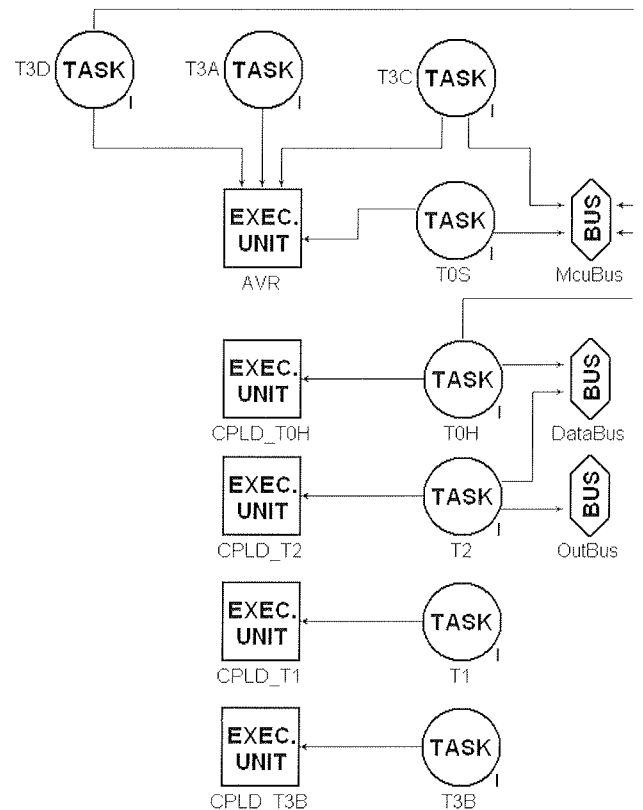
Fig. 9. Mapping tasks on architecture resources

At this point let us briefly explain how the tasks are actually described in SystemC. The task model is defined in terms of architectural resource usage and no actual algorithm is specified. Figure 10 lists a high-level description of task *T2*. This task reads data from the data bus and sends it to the output bus. In order to transfer a block of 8 bytes, the process is repeated eight times.

```
for(int i=0; i<8; i++)
{               // 8x - for each D/A channel
    m_pExecUnit ->GetData(this,4,false, s
        (ICommUnit*) &DataBus,  -1);
            // gets data from data storage
            // uses DataBus
            // need four cycles
    m_pExecUnit ->WriteData(this,1,false,
        (ICommUnit*) &OutBus,  -1);
            // write data to D/A converter
            // uses OutBus
            // needs one cycle
}
```

Fig. 10. High-level description of task T2 in SystemC

The predefined methods *GetData* and *WriteData* from our system modeling libraries are used for modeling data transfer. The designer needs to supply the pointer to an appropriate communication unit and estimation of the cycle duration. Both data transfer requests are performed through execution unit interface (*m_pExecUnit*). The libraries also provide methods for modeling and estimation of computational tasks (e.g. *Add, Mult, Wait*) which can be used for high-level task descriptions.

An image from the simulation setting aspect is shown in Figure 11. A set named *WaveVCD_MCU* is selected and its members (both external-event generators, execution unit *AVR* and communication units *McuBus* and *OutBus)* are shown. All the other architectural resources are shadowed. In this way, the designer defines resources used for producing VCD traces and console log files during model execution.
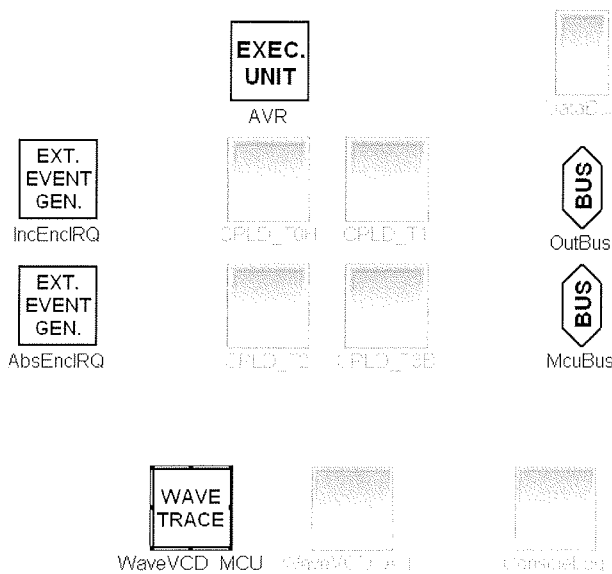
Fig. 11. Selecting hardware resources for waveform traces

4.2 Results

System model evaluation results are obtained after interpretation, compilation and execution of the designed graphical model. Their analysis provides basis for further design decisions.

Table 3 summarizes a part of the utilization log regarding the microprocessor (*AVR*) and the bus *McuBus* that transfers data to the CPLD device. The table is split into the architectural part and functionality part (Tasks). The architectural part shows the resource execution time (RET), i.e. summation of the time the services are required from a specific resource. Complementary, the functionality part presents the task active and wait timings (in % of RET). The numbers stated in the *active* column represent the percentage of the time a specific task is being actively executed on a specific resource. Similarly, the numbers stated in the *wait* column represent the time a specific task has to wait for a specific resource to become available – this is the time interval during which a task may be executed regarding data dependency, but its execution is not started because of the unavailability of HW resources.

Architecture	AVR		McuBus	
RET[ns]	358 144		134 500	
Tasks	active	wait	Active	wait
T0S [% RET]	26.8	0	59.5	0
T3A [% RET]	2.1	5.1	/	/
T3C [% RET]	1.3	0	3.3	0
T3D [% RET]	69.8	0	37.2	0
total [ns]	552 644			

Table 3. Utilization log for AVR and McuBus

Analysis of the results from Table 3 shows that the *AVR* microprocessor is active for about 65% of the simulation time and *McuBus* is active for about 24% of it. Task *T3D* requires most of the processors active time (69.8%) and produces 37.2% of the *McuBus* activity. Waiting can be observed for task T3A (5.1%) caused by the resource contention.

If utilization of a particular resource is not found appropriate, the mapping on architectural resources can be revised. If the real-time constraints are still not met, the functionality description can be revised (e.g. revise algorithm).

The timing diagram, as presented in Figure 12, provides detailed information for the model evaluation. The actual waiting time during resource contention and task execution times can be observed and used for verification of real-time constraints. Under resource contention, task *T3A,* for example, waits 2030ns for an execution unit to become available, but it takes only 84ns to actually execute it.

5 Conclusion and future work

We present a graphical modeling framework used for high-level modeling of heterogeneous systems. It provides graphical design elements for using modeling wrappers from system modeling libraries. Graphical modeling relieves the designer of manual typing the source code and thus hides many details of the SystemC code that normally need to be taken care of. Thus the designer can put more effort on actual modeling. Our graphical framework also provides different constraint checks during modeling and integrates support for simulation settings. When modeling is completed, an executable model is automatically generated to simulate the system behavior on a high abstraction level. Our case study exemplifies the use of our framework and shows information obtained from the executable model built on a high-abstraction level. Evaluation of this information serves as a basis for model evaluation and further design decisions. Graphical modeling enables rapid changes in the model (e.g. changes in the mapping aspect can give
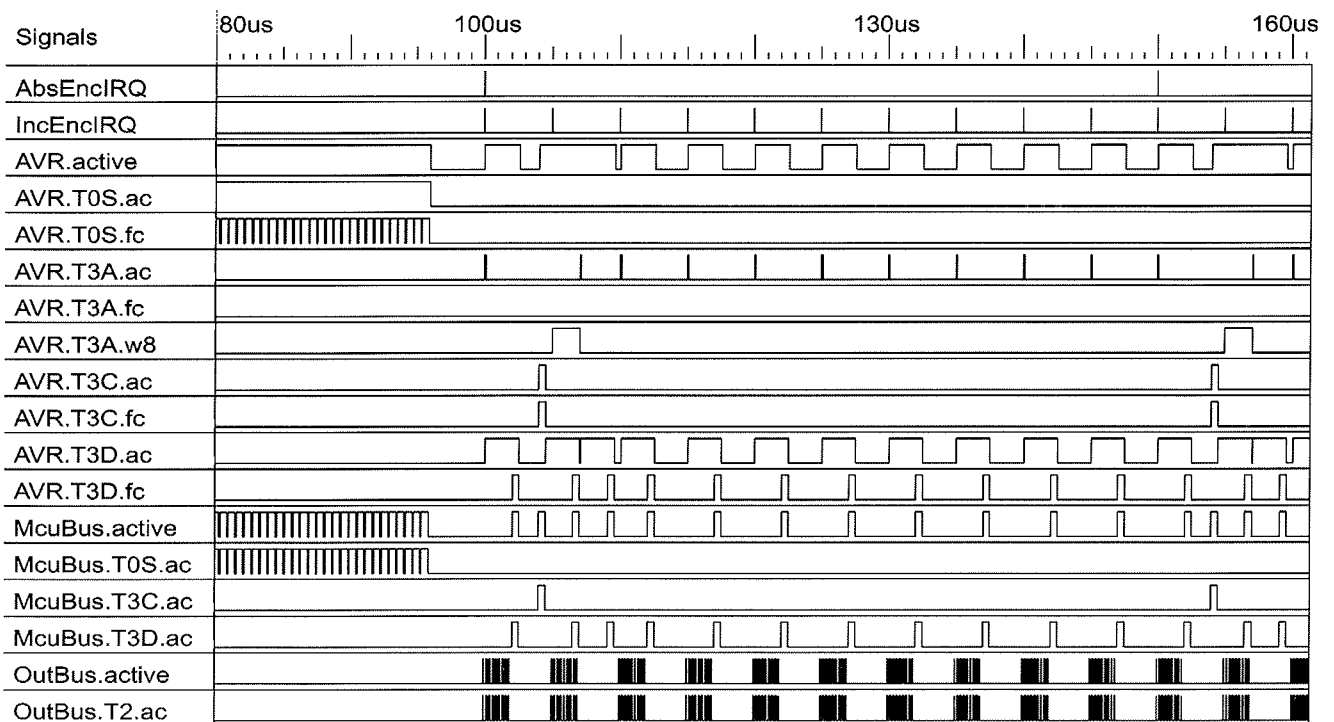
Fig. 12. Timing diagram of selected tasks execution

better results) without time-consuming manual SystemC code rewriting.

In our future work we intend to include support for modeling task interruption and to handle priorities of tasks execution in order to achieve efficient modeling of operating systems. We plan to provide a hierarchical approach to model building as it will significantly improve handling complexity of large models. Our intention is to enable a graphical composition of the task description by providing a library of commonly used methods. Predefined methods for construction of a task high-level description (e.g. *Add, Multiply, GetData, WriteData*) will be thus made available to the designer. We also wish to implement some constraint checks using the OCL language rather than performing them before the interpretation phase. Using OCL, these constraint checks will be performed during the model construction phase.

6   References

/1/   A. A. Jerraya, Long Term Trends for Embedded System Design, CEPA 2 Workshop – Digital Platforms for Defence, Brussels, Belgium, March 15-16, 2005.

/2/   A. A. Jerraya, W. Wolf, Hardware/Software Interface Codesign for Embedded Systems, IEEE Computer Society, vol. 38, no. 2, pp. 63-69, February 2005

/3/   VHDL homepage, http://www.vhdl.org/.

/4/   A. Habibi, S. Tahar: A Survey on System-On-a-Chip Design Languages, Proc. IEEE 3rd International Workshop on System-on-Chip (IWSOC'03), IEEE Computer Society Press, pp. 212-215, June-July 2003

/5/   L. Cai, D. Gajski. Transaction Level Modeling: An Overview. CODES+ISSS'03, October 2003, Newport Beach, California, USA, Pp. 19-24, ISBN:1-58113-742-7

/6/   SystemC OSCI homepage: http://www.systemc.org/

/7/   D. C. Black & J. Donovan. SystemC from the ground up. Springer, ISBN: 1402079885, June 2004

/8/   J. Dedič: Enovito razvojno okolje za sočasno načrtovanje strojne in programske opreme, doktorska disertacija, Univerza v Ljubljani, Fakulteta za elektrotehniko, 2006

/9/   J. Dedič, M. Finc, A Trost: A Framework For High-Level System Design Exploration, Informacije MIDEM, vol. 36, no. 3(119) , pp. 151-160, 2006

/10/   Eclipse GEF project homepage: http://www.eclipse.org/gef/

/11/   GME project homepage: http://www.isis.vanderbilt.edu/Projects/gme

/12/   A. Ledeczi, et al. The Generic Modeling Environment. Workshop on Intelligent Signal Processing, Budapest, Hungary, May 17, 2001.

/13/   Institute for Software Integrated Systems, Vanderbilt University, A Generic Modeling Environment: GME 6 User's Manual, Version 6.0

/14/   OMG UML homepage: http://www.uml.org/

/15/   A. Trost, B Likar, Embedded Development Platform and Applications, Proc. 39th MIDEM Conference, pp. 255-260, October 2003

Klemen Perko, B.Sc.
Sipronika d.o.o., Tržaška 2, 1000 Ljubljana
klemen.perko@sipronika.si

Assistant Prof., Dr. Andrej Trost,
University of Ljubljana, Faculty of Electrical Engineering,
Tržaška 25, 1000 Ljubljana, Slovenia

DESIGN CONSIDERATION FOR POWER MODULES OF ELECTRO-MOTOR DRIVES

Jurij Podržaj, Janez Trontelj

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

Key words: electro-motor drive, power modules, PWM current switch, optimized AC drive

Abstract: he major component of electro-motor drive is the power module producing a PWM signal to drive the motor from the Battery DC source. The power consumption or the efficiency of the electric drive is dominantly the function of the high side and low side semiconductor switch which performs the PWM current to the AC motorcoil. The design of such switch is therefore crucial for the drive performance.

In the paper the design considerations are discussed for the optimization of electrical mechanical and thermal performances of the three phase motor drive power module.

The measurement results on the optimized module are presented together with the price/performance analysis of the power module.

Načrtovalski vidiki za močnostne krmilnike elektromotorjev

Kjučne besede: trifazni krmilnik, tokovna stikala, PWM krmiljenje, optimizacija krmilnika

Izvleček: Glavni del krmilnika elektromotorja je močnostni modul, ki generira PWM pulzno-širinsko moduliran signal, ki napaja motor iz baterije. Poraba moči oziroma učinkovitost močnostnega krmilnika je v večini funkcija zgornjih in spodnjih polprevodniških stikal, ki priklapljajo PWM tok na navitje motorja. Pravilno načrtovanje takega stikala je zato bistvena za delovanje krmilnika. V članku so predstavljene možnosti optimizacije električnih, mehanskih in termičnih vel#in in njihov učinek na delovanje trifaznih krmilnikov za elektromotorje. Predstavljeni so rezultati meritev z optimiziranim krmilnikom. Prav tako je predstavljena cenovna analiza za izboljšave glede na osnovni sistem.

1. Introduction

The PWM signal is used to provide power AC supply to the three phase coil of the AC motor /1/ as shown in fig.1.
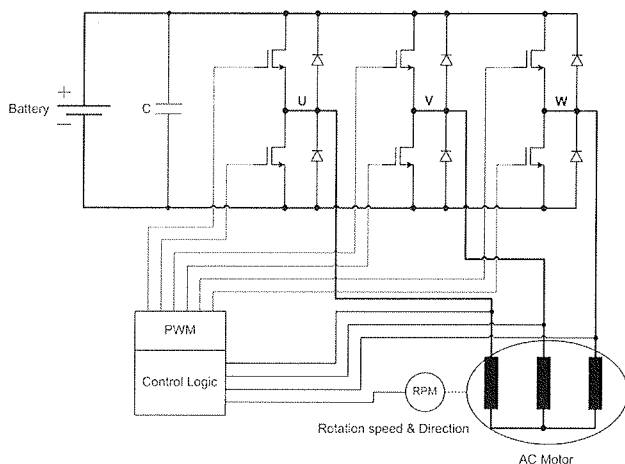
Fig. 1: AC motor system

2. System partitioning

For low power motors the optimal solution is to combine both the low power PWM control signals generation with the power switches in a single ASIC.

This could work as long as the motor power is low enough to be able to design the three low side and the three high side switches on the same ASIC. In this case we need to consider the ASIC power dissipation to be able to make a decision if this approach can be a cost effective solution. To analyze the cost-performance of a single ASIC the simplified models can be used considering the maximal allowed power dissipation of the ASIC, the cost of the mixed signal silicon area, the cost of a single device silicon area, the power dissipation of the bonding wires, the packaging cost and the assembly cost.

The power dissipation of the ASIC is a combination of the power dissipation due to the losses on the semiconductor switch when conducting and the power dissipation due to the so-called switching losses which are caused during the switch conductance transition, and the losses due to charging and discharging the battery buffer capacitor during switching.

The switching losses can be as high as one third of the total power dissipation so a conservative figure for the maximal power dissipation expressed with maximal allowed motor current is approximately

$$r_{on} I_{max}^2 = 2 P_{dm} / 3 \qquad 2.1$$

where the r_{on} is the resistance of the switch in ON state including resistance of connections (bonding and bus bars),

I_{max} is the maximal motor current and P_{dm} is a maximal allowed power dissipation of the drive.

The silicon area is in linear proportion to the r_{on} but is different for mixed signal silicon (Am) and single device silicon (Ad).

A simple rule of thumb gives us the following approximate relations.

One square mm is required for a 50V drain-source break down voltage to achieve 100mOhms on resistance. On the other hand the same performance can be achieved with only 0,7 mm^2 using single device silicon Ad.

Taking into account that the cost of Am silicon area is about 1.5 times the cost of Ad silicon area, this simply means that the cost efficiency of the single silicon area is more than two times higher.

Unfortunately the simple linear model for the silicon area cost is not applicable for larger silicon area due to final defect density of the silicon processing which further reduces the benefits of larger integration. It has been estimated that using some simplification the break point between a single ASIC approach which combines the low power drivers and the power switches is close to the area of about 30 mm^2.

In this case the required area for low power control electronic is about 9mm^2, so the remaining 21mm^2 is used for six switches with γ_{on} of about 30mOhms.

3. Analysis of maximal power capability of single ASIC motor drive

Using equation 2.1. the maximal current can be calculated:

$$I^2{}_{max} = 2P_{dm} / 3 \cdot 0.03 \qquad 3.1$$

P_{dm} can be calculated from the best possible thermal resistance from the packaged ASIC to the ambient.

Fig. 2 shows the thermal profile of the ASIC packaged in a power package with thermal resistance of 0.5 degree C/Watt mounted to the heat sink with 230mm x 200mm x 40mm.

The maximal possible power P_{dm} can be calculated from the temperature difference between maximal ambient temperature and maximal junction temperature, taking into the account the total thermal resistance of the set-up. In our case the maximal ambient temperature is 120 degrees C and the maximal junction on temperature 150 degrees C. The worst case thermal resistance is 1 degree C/Watt, which gives us the next power P_{dm} is 30W.

30W maximal power dissipation of the ASIC for three phase AC motor can provide the maximal motor phase current
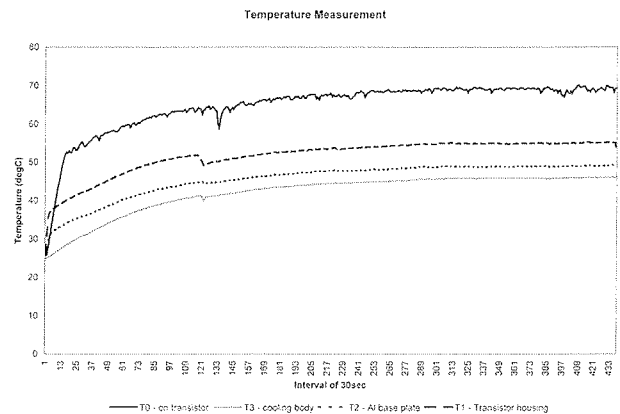
Fig. 2: Thermal profile of the dedicated ASIC model
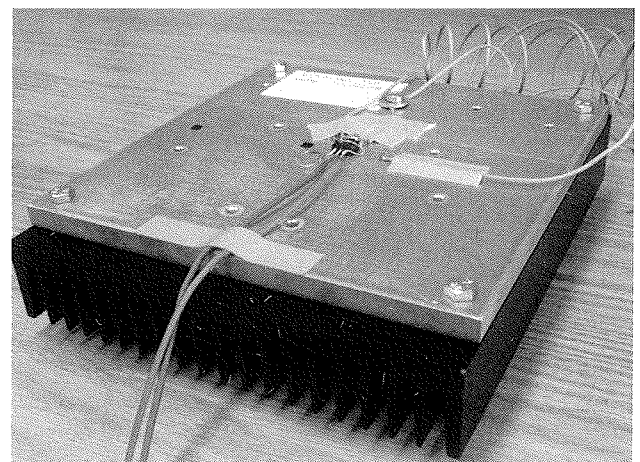
Fig. 3: Single ASIC model set-up

according to equation 2.1, where the total power dissipation is the sum of the power dissipation of six switches, i.e. two per phase. This leaves us with 30W/6 per switch, so only 5W can be entered into equation 2.1.

The result is:

$$I_{max} = \sqrt{2 \cdot P_{dm} / 3 \cdot 0,03} \cong 10A$$

This is therefore the limit for single ASIC power drive. Depending on the maximal battery voltage and transistor break-down voltage the typical maximal electro-motor power would range from 300W to max 1kW.

4. Design consideration for power motor using external power switches

From the above analysis it is clear that the most critical figure which needs to be improved is the maximal power dissipation. Here we have two measures which could be improved.

The first one is the reduction of r_{on} resistance which can be lowered based on large silicon area using low cost sin-

gle device silicon, and which can be further minimized by the use of parallel devices. However it is required that the device is connected using low resistance connections. An example of such low resistance is shown in figure 4, where the transistor drain is soldered to 3mm thick copper bus-bar and the source is bonded with thick multiple Al wires. This arrangement using 5 transistors with the area of 30 mm^2 provides r_{on} resistance of 500µOhms.
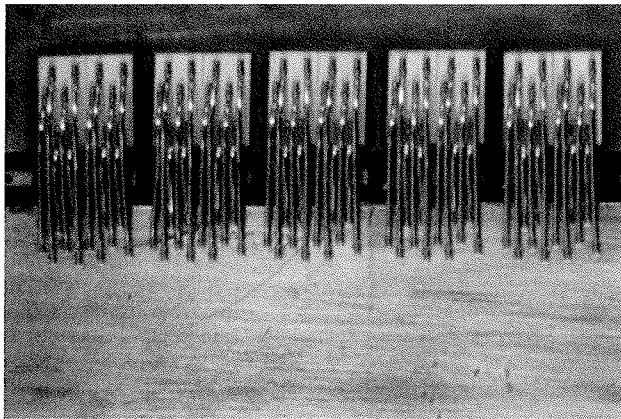
Fig. 4: Transistor connections

The second improvement is to reduce the thermal resistance of the transistor junction to the ambient. Direct soldering to the copper bus-bar reduces this thermal resistance to less than 0,05 degrees C/Watt, so the only remaining critical thermal path is the heat-sink to ambient air for passive. The photograph of the power drive three + 1 phase switch is shown in figure 5.
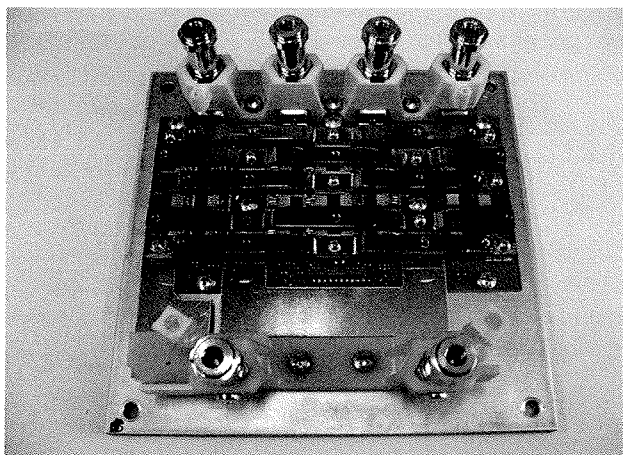
Fig. 5: The photograph of the power drive

For this arrangement a thermal resistance for transistor junction to the ambient has been reduced to 0,15 degree C/Watt. Taking into the account the r_{on} resistance improvement from 30mOhms to 400µOhms, i.e. 75 times and the thermal resistance improved which is approximately 7 times the total maximal current increase is $\sqrt{75 \cdot 7} = 17$ times, which leads us to maximal current of about 220A.

The power drive which has been designed based on described principles is shown in figure 6.
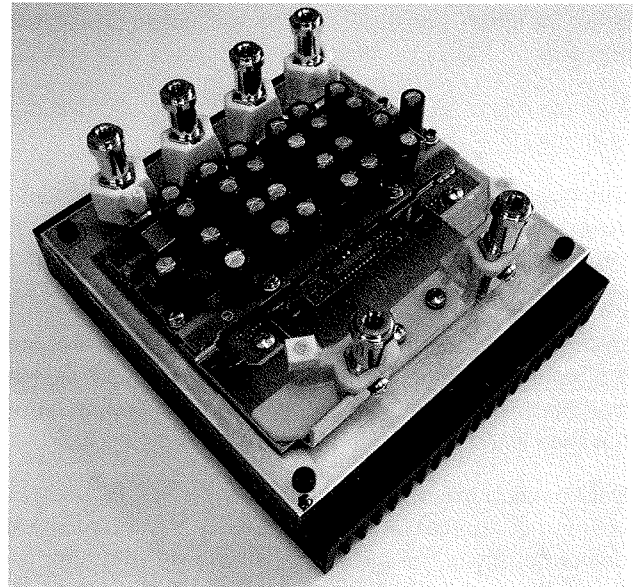
Fig. 6: Power drive

The measurement results using the peak required current for 60 minutes are shown in figure 7.
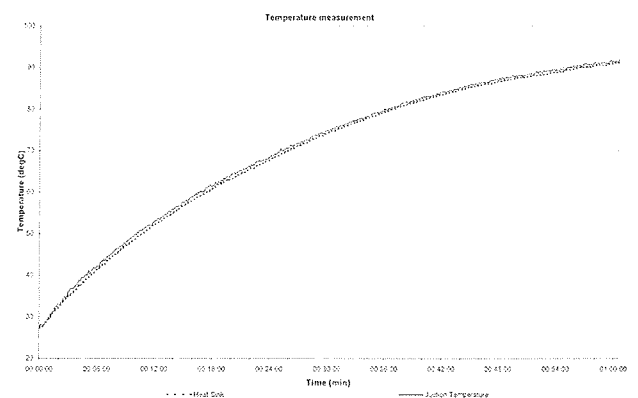
Fig. 7: Measurement results of power drive: The upper curve shows temperature behaviour of transistor junction; the lower curve shows temperature behaviour of the heat sink

5. Conclusions

The analysis of single ASIC power drive with integrated power switch has been presented. The results of this analysis have been implemented on the design of power drive with external power switches. The proposed design guide lines have improved the cost-effectiveness by the factor of around 1,6 times compared to leading manufacturers of such devices /2/. The designed power drive has been measured to show superior performance compared to the existing power drives in the market /3/.

6. Acknowledges

7. References

/1/ V. Ambrožič, Sodobne regulacije pogonov z izmeničnimi stroji, Fakulteta za elektrotehniko, 1996

/2/ U. Nicolai, Application Manual Power Modules, ISLE 2000, 2000

/3/ M. Fukada, D. Nakajima, K. Takanashi, Power module, Patent Nr.: US 6,501,172 B1, 31. Dec 2002

Jurij Podržaj
Univerza v Ljubljani, Fakulteta za elektrotehniko
Tržaška c. 25, 1000 Ljubljana, Slovenija
Tel: 01/4768 – 340; Email: jure.podrzaj@gmail.com

Dr. Janez Trontelj
Univerza v Ljubljani, Fakulteta za elektrotehniko
Tržaška c. 25, 1000 Ljubljana, Slovenija
Tel: 01/4768 – 333; Email: janez.trontelj1@guest.arnes.si

ANALYSIS OF POTENTIAL ATTACK SCENARIOS FOR SYSTEMS WITH IEEE STD 1149.1 SECURITY EXTENSION

Anton Biasizzo

Jozef Stefan Institute, Ljubljana, Slovenia

Abstract: The paper addresses the security problems of boundary-scan design. Recently proposed security extension for IEEE Std. 1149.1 providing a locking mechanism is discussed. Possible attack scenarios are analysed. Complete attack time is calculated for different lengths of Key/Lock registers. For a large length of the Key/Lock registers it is practically impossible to perform a complete attack. Assuming that the attacker has some limited time interval to perform the attack, the probability of compromising the system is explored and the probability of successful attack within a given time interval is calculated. Test Access Port control logic with locking mechanism was implemented in Xilinx Spartan3 FPGA. The mechanism requires small hardware overhead and can be easily included in the IEEE Std 1149.1 test infrastructure.

Analiza možnih scenarijev vdora v sisteme z vgrajeno varnostno razširitvijo standarda IEEE 1149.1

Izvleček: V zadnjem času postajajo aktualni problemi zagotavljanja varnosti v sistemih z vgrajeno preizkusno linijo (boundary-scan). Za vsako integrirano vezje z vgrajeno preizkusno linijo in vsak sistem zgrajen iz takšnih vezij namreč obstaja nevarnost vdora. Na panelni diskusiji konference ITC 2004, ki jo je moderiral E.J. Marinissen, so obravnavali možnosti vdora v sistem in kraje intelektualne lastnine preko preizkusne infrastrukture [1-5]. Predlagana je bila rešitev osnovana na kriptiranju preizkusnih podatkov, tako da se v integrirano vezje na vhodu preizkusne linije doda vezje za dekodiranje podatkov, na izhodu linije pa vezje za kodiranje podatkov. Pomanjkljivost takšne rešitve je v tem, da je že samo vezje za dekodiranje in kodiranje dokaj kompleksno sekvenčno vezje in je zanj potrebno zasnovati vgrajen samodejni preizkus. Kraja intelektualne lastnine pa ni edini problem povezan z varnostjo sistemov z vgrajeno preizkusno linijo. Standardi IEEE 1149.1, IEEE 1149.4 in IEEE 1500 predvidevajo izvajanje preizkusa sistema preko namenskega preizkusnega vodila. V nekaterih izvedbah je to preizkusno vodilo priključeno na računalnik, ki je daljinsko dostopen preko interneta. V takšnih primerih je omogočeno daljinsko preizkušanje in vzdrževanje sistema, hkrati pa obstaja možnost vdora v sistem s strani neavtoriziranih oseb. Ob vdoru v sistem lahko napadalec sproži izvajanje preizkusnega ukaza, ki zmoti normalno delovanje sistema in ima zato lahko katastrofalne posledice. Za preprečevanje dostopa neavtoriziranim osebam je bila predlagana [6] varnostna razširitev standarda IEEE 1149.1. V tem prispevku analiziramo možne scenarije vdora v sisteme z vgrajeno varnostno razširitvijo standarda. Obravnavana sta dva osnovna načina vdora. Za različne velikosti varnostnega mehanizma so izračunane verjetnosti vdora v danih časovnih okvirih. Na kratko je predstavljena tudi praktična izvedba mehanizma v vezju FPGA.

1. Introduction

In recent years, discussion about the security problems of systems incorporating scan design has emerged. Any chip that uses scan design and any system built around it provides access to the system's internal logic and may be vulnerable to hackers. Possible theft of intellectual property via scan test infrastructure was discussed at a panel discussion at ITC 2004 moderated by E.J. Marinissen /1/. R. Kapur proposed a solution based on data encryption to protect the data in scan chains /2/, /3/. The application of cryptographic algorithms in scan design chain is, however, not trivial. The logic implementing a cryptographic algorithm is itself a complex sequential circuit which requires some design-for-test (DFT) solution: a conventional way to solve the problem is by organizing the flip-flops in a scan chain. On the same conference B. Yang, K. Wu, and R. Karri presented a paper in which they demonstrated the vulnerability of the implementation of DES algorithm with inserted scan chain using Synopsys Test Compiler /4/. Some further research in this topics has been reported recently by D.Hely et al. /5/.

Beside possible theft of intellectual property, scan design can be potentially misused for breaking in the system which may lead to a serious damage. Scan design is often combined with the test infrastructure of DFT standards IEEE Std. 1149.1, IEEE Std. 1149.4 and IEEE Std. 1500. In some implementations of remote system maintenance, test access port (TAP) is interfaced to a computer connected to internet. An attacker may crack the computer system and get access to the test port. Executing EXTEST or some other pin-permission instruction during normal system operation could have catastrophic consequences in safety critical applications. In order to prevent unauthorised users to access the system via IEEE Std. 1149.1 TAP, a security extension for IEEE Std 1149.1 has been proposed /6/. It provides a locking mechanism that can be activated manually or automatically after a predefined time-out. The security extension requires small hardware overhead and allows full conformance with IEEE Std. 1149.1. The proposed solution is also applicable to IEEE Std. 1149.4. Similar to the approach reported in /7/, it can be included as an extension in full conformance with IEEE Std. 1149.4. In this paper we analyze potential attack scenarios for the

systems with implemented IEEE Std 1149.1 security extension. The results may be of interest to the designers and managers when making decisions about the level of security of the prospective products.

2. IEEE Std 1149.1 security extension

A chip with implemented IEEE Std 1149.1 infrastructure and security extension is shown in Figure 1. The locking mechanism is shown in more details in Figure 2. The security extension of IEEE Std 1149.1 includes two additional instructions: LOCK and UNLOCK. When LOCK instruction is active the TAP control logic maps all instructions (except UNLOCK) to a harmless BYPASS instruction until the UNLOCK instruction with valid key code is applied.
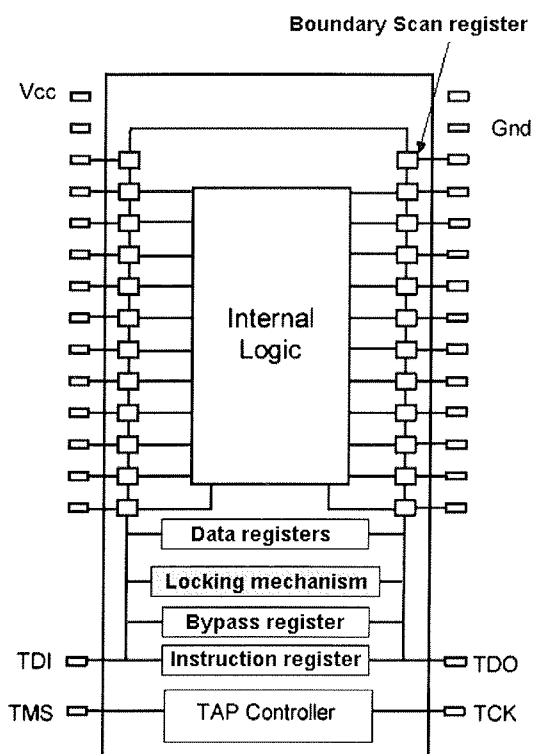
Fig. 1: IEEE Std 1149.1 infrastructure and security extension

The process of locking the TAP controller consists of the following steps:

1. LOCK instruction is entered into Instruction Register via TDI and decoded.
2. Lock Register and Key/Lock Shift Register are enabled.
3. The contents of the Key/Lock Shift Register is cleared at active Capture DR.
4. Lock code is entered into the Key/Lock Shift Register via TDI.
5. Lock code is transferred from the Key/Lock Shift Register to the Lock Register and Key Register is cleared at active Update DR.
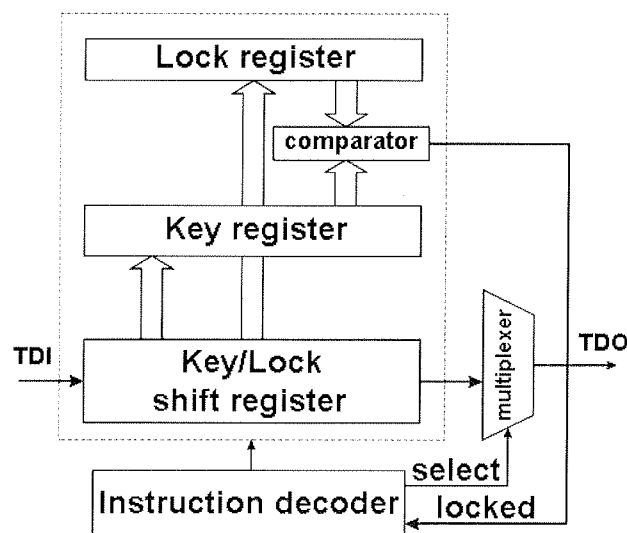
Fig. 2: IEEE Std 1149.1 locking mechanism

Comparator compares the contents of Lock Register and Key Register. If the contents are different, the Locked signal fed to the Instruction Decoder is activated. Consequently, the instruction decode logic maps all instructions except UNLOCK to the BYPASS instruction. This mapping is active until the Locked signal is released by executing UNLOCK instruction with the current key code.

Notice that the contents of Lock Register and Key Register are the same if the lock code is equal to zero. In this case, the TAP control logic remains unlocked.

The process of unlocking the TAP controller consists of the following steps:

1. UNLOCK instruction is entered into Instruction Register via TDI and decoded.
2. Key Register and Key/Lock Shift Register are enabled.
3. The contents of the Key/Lock Shift Register is cleared at active Capture DR.
4. Key code is entered into the Key/Lock Shift Register via TDI.
5. Key code is transferred from the Key/Lock Shift Register to the Key Register at active Update DR.

Comparator compares the contents of Lock Register and Key Register. If the contents are equal, it deactivates Locked signal and the next instruction entered via TDI can be executed. If the contents are different (i.e., wrong key code) the Lock signal remains activated and the TAP control logic remains locked.

3. Possible attack scenarios

We assume that the attacker has access to the Boundary-Scan infrastructure and is familiar with the locking mechanism described in the paper. However the attacker neither

knows the length of the Key/Lock registers nor the key code.

We can distinguish between two kinds of attacks to the locking mechanism:

1. Invasive attack, where the attacker's intention is to interfere with the operation of the circuit regardless if he/she leaves any traces of the attack (i.e., after the attack it is possible to deduce that someone interfered with the device).

2. Non-invasive attack, where the attacker first determines the lock code. The knowledge of the lock code enables the attacker to cover the tracks of the attack by locking the device with original lock at the end of the attack.

3.1. Invasive attack

3.1.1. Invasive attack scenario

We assume that the boundary-scan logic is in the Run-Test-Idle state. The attacker can easily change from any state to this state by maximum 6 additional clocks: 5 clocks to change to Test-Logic-Reset state (TMS=1) and one clock to change to Run-Test-Idle state (TMS=1) and this additional transitions are negligible part in comparison with the whole the attack.

For the invasive attack the attacker does not care about the value of the lock code. He/she merely tries to overwrite it by executing the LOCK instruction after each UN-LOCK attempt.

The attack consists of:

1. Determining the length of the Key/Lock registers.

    The fact that the Key/Lock Shift Register is cleared at active Capture-DR can be also used for determining the length of the Key/Lock registers. The attacker executes the UNLOCK instruction and feeds values 1 to the input of the boundary-scan chain (TDI). By counting zeros at the output, the length of Key/Lock registers can be determined.

2. Repeating the following steps:
    -   UNLOCK instruction, where the attacker generates a guess value of the lock code and tries to unlock the lock mechanism.
    -   LOCK instruction, where the attacker overwrites the value of the Lock Register with 0. (Since the Key/Lock Shift Register is cleared at active Capture-DR the value 0 is the best choice for overwriting the Lock Register. In this way no Shift-DR cycles are required.) If previous attempt was unsuccessful the circuit remains locked and the LOCK instruction is ignored.

    It is worth noting that the attacker does not care in which step the Lock Register was successfully overwritten. The only goal is to have the value 0 placed in the Lock Register after the successive number of UNLOCK/LOCK executions.

3. Unlocking the circuit with 0 as the key and performing the malicious actions.

3.1.2. Invasive attack time analysis

Let L be the length of the Instruction Register and M the length of the Key/Lock registers, respectively. Let us denote the frequency of the boundary-scan clock (TCK) by f.

Determination of the length of Key/Lock registers starts in Run-Test-Idle state and stops in Select-IR Scan state of the boundary-scan test logic. It consists of:

-   transition to Capture-IR state (3 cycles - TMS="110"),
-   loading UNLOCK instruction (L cycles - TMS="0..0"),
-   transition to Capture-DR state (4 cycles - TMS="1110"),
-   loading values "1" to TDI and monitoring output TDO ($M+1$ cycles - TMS="0..0"),
-   transition to Select-IR Scan state (4 cycles - TMS="1111").

Complete determination of the length of Key/Lock registers consists of

$L+M+12$ ($3+L+4+M+1+4$) cycles of boundary-scan clock.

Unlock attempt starts in Select-IR Scan state and consists of:

-   transition to Capture-IR state (1 cycle - TMS="0"),
-   loading UNLOCK instruction (L cycles - TMS="0..0"),
-   transition to Capture-DR state (4 cycles - TMS="1110"),
-   loading unlock code (M cycles - TMS="0..0"), (first unlock code has one additional clock)
-   transition to Select-IR Scan state (4 cycles - TMS="1111").

Complete unlock attempt consists of $L+M+9$ ($1+L+4+M+4$) cycles of boundary-scan clock.

Lock attempt also starts in Select-IR Scan state and consists of:

-   transition to Capture-IR state (1 cycle - TMS="0"),
-   loading LOCK instruction (L cycles - TMS="0..0"),
-   transition to Capture-DR state (4 cycles - TMS="1110"),
-   transition to Select-IR Scan state (4 cycles - TMS="1111").

Complete lock command consists of $L+9$ ($1+L+4+4$) cycles of boundary-scan clock.

In the final UNLOCK command the unlock code is 0 and since Key/Lock Shift Register is cleared at Capture-DR state there is no need to load unlock code. It consists of:

-   transition to Capture-IR state (1 cycle - TMS="0"),
-   loading UNLOCK instruction (L cycles - TMS="0..0"),
-   transition to Capture-DR state (4 cycles - TMS="1110"),

- transition to Run-Test-Idle state (3 cycles - TMS="110").

The length of the final unlock code is therefore L+8 cycles of boundary-scan clock.

The length of the complete attack must cover all possible lock codes. Since the length of Key/Lock registers is M there are 2^M possible codes and complete length of the attack is

Length = 2^M(L+M+9+L+9)+L+M+12+L+8 = 2^M(2L+M+18)+2L+M+20

The total time of the attack is given by

$$t = \frac{2^M (M + 2L + 18) + 2L + M + 20}{f}$$

Let us assume that the length of the Instruction Register is 8 bit and that the boundary-scan clock frequency is 100MHz. Some results for the time required for the complete attack for different lengths of Key/Lock registers are given in the table below.

M	time
8	108 μs
16	33 ms
32	2835 s
48	7.3 years
56	2056 years
64	573000 years

3.1.3. Incomplete attack and the probability of the successful attack

For a large length of the Key/Lock registers it is practically impossible to perform a complete attack. Suppose that the attacker has some time interval to perform the attack. Let us estimate the probability that the system will be compromised.

First we determine the number of unlock codes that the attacker can exploit in a time interval of length t:

$$N = \frac{t \cdot f - 2L - M - 20}{M + 2L + 18}$$

The probability that the system will be compromised is

$$p = \frac{N}{2^M} = \frac{t \cdot f - 2L - M - 20}{2^M (M + 2L + 18)}$$

Let us, determine the probability that the system will be compromised if the time interval is one hour for previous example

M	probability
32	100 %
36	7.5 %
40	0.44 %
48	0.0016 %
56	5.6 10^{-8}
64	2 10^{-10}

From the above equations we can estimate the lower bound of the length of the Key/Lock registers that would assure the required system security for a given time interval.

$$M = \log_2\left(\frac{t \cdot f - 2L - M - 20}{p\,(2L + M + 18)}\right) =$$
$$= \log_2(t \cdot f - 2L - M - 20) -$$
$$- \log_2 p - \log_2(2L + M + 18)$$

The derived equation cannot be solved analytically. Yet, the effect of the 2L+M+21 cycles is negligible and can be omitted. Since we are determining the lower bound of the length of the Key/Lock registers M we can replace the term (M + 2 L + 20) by s smaller value (2 L + 20), which gives

$$M_{(est)} = \log_2(t \cdot f) - \log_2 p - \log_2(2L + 18)$$

Minimal length of the Key/Lock registers for several probabilities that the system will be compromised in one day are given in the table below.

probability	$M_{(est)}$
100 %	37.9
10 %	41.2
1%	44.5
0.1%	47.9
10^{-4}	51.2
10^{-5}	54.5
10^{-6}	57.8

3.2. Non-Invasive attack

3.2.1. Non- Invasive attack scenario

Assumptions for non-invasive attack are the same as in the case of invasive attack: the Boundary-Scan logic is in the Run-Test-Idle state. If this is not the case the attacker can easily change from any test logic state to Test-Logic-Reset in 5 test clocks and to Run-Test-Idle in one additional clock.

The goal of the attacker is not just to get access to the Boundary-Scan test infrastructure but also to cover his/her tracks. This can be achieved only by determining the lock code. In order to determine the lock code the attacker has to check if the boundary-scan test logic is unlocked after every unlock command. This can be accomplished by determining the length of the data path for any restricted instruction, which uses data register longer than bypass register. When the boundary-scan test logic is locked bypass register is placed in the data path instead of the protected data register. The LOCK instruction is an obvious choice since it can be used also to erase the lock code. As was the case with the evaluation of the lock register length the attacker can take advantage of the fact that the value of the lock shift register is cleared prior the shifting.

The Non-Invasive attack consists of:

1. Evaluation of the key/lock register length as described in the case of Invasive attack.

2. Repeating of the following steps:
   - UNLOCK instruction with the guessed value of the lock code,
   - LOCK instruction and check if the length of the data path:
     - if the length of the data path is 1 then the boundary-scan test logic remains locked and step 2 is repeated with new guess value,
     - if the length of the data path is longer than 1 then the guess value is correct lock code. Use 0 as the new lock code (unlock the boundary-scan test logic) and stop the attack.

After the circuit exploitation with unlocked boundary-scan test logic the test logic can be locked with the original lock code in order to cover the track of the intrusion.

3.2.2. Non-Invasive attack time analysis

Let L be the length of the Instruction Register and M the length of the Key/Lock registers, respectively. Let us denote the frequency of the boundary-scan clock (TCK) by f.

Non-Invasive attack is performed in the following steps:

1. Evaluation of the length of the key/lock register starts in Run-Test-Idle state and stops in Select-IR Scan state of the boundary-scan test logic. It consists of:
   - transition to Capture-IR state (3 cycles – TMS="110"),
   - loading UNLOCK instruction (L cycles – TMS="0...0"),
   - transition to Capture-DR state (4 cycles – TMS="1110"),
   - loading values "1" to TDI and monitoring output TDO (M+1 cycles – TMS="0...0"),
   - transition to Select-IR Scan state (4 cycles – TMS="1111").

   This step is done in L+M+12 test cycles.

2. Unlock attempt starts at Select-IR Scan state of the boundary-scan test logic and consists of:
   - transition to Capture-IR state (1 cycle – TMS="0"),
   - loading UNLOCK instruction (L cycles – TMS="0...0"),
   - transition to Capture-DR state (4 cycles – TMS="1110"),
   - loading unlock code (M cycles – TMS="0...0"),
   - transition to Select-IR Scan state (4 cycles – TMS="1111").

   This step is done in L+M+9 test cycles.

3. It is followed by data path length check, which consists of:
   - transition to Capture-IR state (1 cycle – TMS="0"),
   - loading LOCK instruction (L cycles – TMS="0...0"),
   - transition to Capture-DR state (4 cycles – TMS="1110"),

- length check (1 cycle – TMS="0", TDI="1"):
  a. if the length of data path is 1 ("1" shifted out of TDO)
     - make transition to Select-IR Scan state (4 cycles – TMS="1111"),
     - repeat step 2 with new guessed unlock code,
  b. if the length of data path is greater then 1 ("0" shifted out of TDO):
     - load lock code 0 (M cycles – TMS="0...0", TDI="0...0"),
     - transition to Run-Test-Idle state (3 cycles – TMS="110").

This step is done in L+10 test cycles, when the guessed lock value is wrong and in L+M+9, when the guessed lock value is correct.

After successful attack the Boundary-Scan test logic is in Run-Test-Idle state and current lock code is 0 (i.e. test logic is unlocked).

In the worst case for the attacker all 2^M codes are tried where 2^{M-1} were wrong. The length of such attack is:

$$Length = 2^{M-1} \cdot (L + M + 9 + L + 10) +$$
$$+ (L + M + 9 + L + M + 9) + 2 =$$
$$= 2^M \cdot (2L + M + 19) + M + 1$$

The total time of the Non-Invasive attack is given by

$$t = \frac{2^M (M + 2L + 19) + M + 1}{f}$$

With the same assumptions of the length of the Instruction register and boundary-scan clock frequency as in the case of the Invasive attack the required time for the complete attack with respect to different lengths of Key/Lock registers are given in the table below.

M	time
8	110 µs
16	33.4 ms
32	2880 s
48	7.4 years
56	2080 years
64	580000 years

3.2.3. Incomplete Non-Invasive attack and the probability of successful attack

As was the case for Invasive attack it is practically impossible to perform complete Non-Invasive attack of boundary-scan test logic with large Key/Lock registers. It neither makes sense, since the attack is stopped as soon as the valid key is found. Therefore more useful measure of the security strength of the circuit is given by the probability that the system will be compromised in the given time span.

First we determine the number of unlock codes that the attacker can exploit in given time interval t :

$$N = \frac{t \cdot f - M - 1}{M + 2L + 19}$$

The probability that the system will be compromised is

$$p = \frac{N}{2^M} = \frac{t \cdot f - M - 1}{2^M (M + 2L + 19)}$$

The probabilities that the system will be compromised in the time interval of one hour for the previous example is given in the following table

M	probability
32	100 %
36	7.4 %
40	0.44 %
48	0.0015 %
56	$5.5 \ 10^{-8}$
64	$2 \ 10^{-10}$

From the above equations we can estimate the lower bound of the length of the Key/Lock register that would assure the required system security for given time interval:

$$M = \log_2 \left(\frac{t \cdot f - M - 1}{p (2L + M + 19)} \right) =$$
$$= \log_2 (t \cdot f - M - 1) - \log_2 p - \log_2 (2L + M + 19)$$

This equation cannot be solved analytically yet the impact of M+1 is negligible and can be omitted. The term can be replaced with smaller value since we are estimating lower bound of the Key/Lock register length, which gives

$$M_{(est)} = \log_2 (t \cdot f) - \log_2 p - \log_2 (2L + 19)$$

In the following table the estimations of the lower bound of the Key/Lock register length are given.

probability	$M_{(est)}$
100 %	37.8
10 %	41.2
1%	44.5
0.1%	47.8
10^{-4}	51.1
10^{-5}	54.5
10^{-6}	57.8

4. Implementation

Test Access Port control logic with locking mechanism was implemented in Xilinx Spartan3 FPGA. The mechanism requires small hardware overhead and does not slow down the conventional boundary-scan tests. The implementation details for configurations with three different lengths of Lock Register / Key Register of are summarised in the table below. For comparison, the configuration data of the test logic without security extension is also shown in the

table. In all cases, the length of the Boundary Data Register is 2 bits and the length of the Instruction Register is 4 bits.

	without security extension	included Lock Register / Key Register		
		8 bit	16 bit	32 bit
number of slices	34	48	61	92
number of slice Flip Flops	45	69	93	141
number of 4 input LUTs	61	86	107	149

5. Conclusions

Security extension of IEEE Std 1149.1 based on a locking mechanism was investigated: typical attack scenarios were considered and analysed. Test Access Port control logic with the locking mechanism has been implemented in Xilinx Spartan3 FPGA. The mechanism requires small hardware overhead and can be easily included in the IEEE Std 1149.1 test infrastructure.

References

/1./ E. J. Marinissen, moderator, "Security vs. Test Quality: Can We Really Only Have One at a Time?" Proceedings of the International Test Conference, 2004. pp. 1411.

/2./ R. Kapur, "Security vs. Test Quality: Are they mutually exclusive?" Proceedings of the International Test Conference, 2004. pp. 1414.

/3./ B. Yang, K. Wu, R. Karri, "Scan based side channel attack on dedicated hardware implementations of data encryption standard", Proceedings of the International Test Conference, 2004, pp. 339-344.

/4./ EE Times On Line, Latest News, Scan design called portal for hackers http://www.us.design-reuse.com/news/news8974.html

/5./ D. Hely, F. Bancel, M-L. Flottes, B. Rouzeyre, " Securing Scan Control in Crypto Chips", Journal of Electronic Testing, Theory and Practice, Vol. 23, No. 5, 2007, pp. 457 - 464.

/6./ F. Novak, A. Biasizzo, "Security extension for IEEE Std 1149.1", Journal of Electronic Testing, Theory and Practice, Vol. 22, No. 3, 2006, pp. 301-303.

/7./ U. Kac, F. Novak, F. Azais, P. Nouet, M. Renovell, "Extending IEEE Std. 1149.4 analog boundary modules to enhance mixed-signal test", IEEE Design & Test of Computers, Vol. 20, No. 2, 2003, pp. 32-39.

Anton Biasizzo
Jozef Stefan Institute, Jamova 39, 1000 Ljubljana
anton.biasizzo@ijs.si

ANALYSES OF SHAFT CURRENTS IN LOW-VOLTAGE INDUCTION MOTOR FOR FORKLIFT DRIVE WITH ELECTRONIC EQUIPMENT

Stjepan Štefanko[1,2], Željko Hederić[2], Miralem Hadžiselimović[3], Ivan Zagradišnik[3]

[1] KONČAR - Institut of electrotechnics, Zagreb, Croatia
[2] University J. J. Strossmayer, Faculty of electrical engineering, Osijek, Croatia
[3] University of Maribor, Faculty of electrical engineering and computer science, Maribor, Slovenia

Key words: low-voltage induction motor, shaft currents, measurements, electronic equipment

Abstract: At the power supply of the low-voltage asynchronous motors (with homogeneous yokes) from the network, shaft currents resulting in damage of bearings, can emerge. The main cause of these currents is the eccentric position of rotor in the stator bore (static and dynamic eccentricity). The required condition for the emergence of these currents is nonlinearity of magnetizing curve of electrical steel of the motor stack lamination. The article describes performed measurements on the four-pole motor. First, the measuring equipment: Rogowski coil and his mounting, AD converter NI-DAQPad-6015 and power analyzers NORMA-D6000 have been described. Then, ways of measurements and processing of measurements have been explained. Data obtained by measurements have been presented at the nominal load, for the star- and delta-connected stator winding, measured with AD converter NI-DAQPad-6015 and power analyzer NORMA-D6000. Differences in measured results of shaft currents measured by the power analyzer NORMA-D6000 related to measurements with the AD converter NI-DAQPad-6015 are explained.

Analiza meritev tokov gredi nizkonapetostnega asinhronskega motorja za pogon vilicarja z elektronsko opremo

Kjučne besede: nizkonapetostni asinhronski motor, tokovi gredi, meritve, elektronska oprema

Izvleček: Pri napajanju nizkonapetostnih asinhronskih motorjev (s homogenimi jarmi) iz omrežja lahko nastanejo tokovi v gredi, ki imajo za posledico okvaro ležajev. Glavni vzrok teh tokov je ekscentrični položaj rotorja in statorja (statična in dinamična ekscentričnost). Potrebni pogoj za nastanek teh tokov je nelinearnost magnetilne krivulje materiala - lamel paketa, tj. dinamo pločevine motorja.

V članku so opisane izvedene meritve na štiripolnem motorju za pogon vilicarja. Najprej je opisana merilna oprema: tuljava Rogowskega in njena vgradnja, AD kartica NI-DAQPad-6015 in digitalni analizator NORMA-D6000. Nato so pojasnjeni načini meritev in predvsem obdelava meritev.

Za izvedene meritve so prikazani rezultati meritev v nazivni točki, za vezavo zvezda in trikot statorskega navitja z AD kartico NI-DAQPad-6015 in inštrumentom NORMA-D6000. Pojasnjene so razlike merilnih rezultatov za vrednosti tokov gredi pri merjenju z AD kartico NI-DAQPad-6015 glede na meritve z analizatorjem moči NORMA-D6000.

1. Introduction

The phenomena of shaft currents at big synchronous and induction machines was discovered and explained in the last century. Air gaps in the stator yoke that exist in the stator stacks put together from parts (segments) or completed stacks composed of steel segments can cause shaft currents. They are closed in the circuit: shaft - bearing - bearing shield - housing - bearing shield - bearing - shaft and they damage the bearings.

The mechanisms of the shaft currents beginning at low voltage induction motors that have as a rule homogeneous yoke are very complicated /1/. In this article it is presented that the nonlinearity of magnetization curve of iron is a necessarily condition of the shaft currents origin.

All manners of shaft currents origins at low voltage induction motors are not explained to the end, therefore it is necessary to determine, from the view of the shaft current measurements, with frequency analyses the harmonic content of the shaft currents and then to determine the ground as well as the mechanism of the harmonic components origin.

In this article the measurements of shaft currents on one four pole low voltage induction motor with the help of Rogowski coil /2/ are analyzed. The method of shaft currents measurements with Rogowski coil is recommended in /3/, where Rogowski coil was mounted around the shaft of the rotor.

At shaft current measurements described in /4/ the Rogowski coil was mounted (around the shaft) on the sta-

S. Štefanko, Ž. Hederić, M. Hadžiselimović, I. Zagradišnik:
Analyses of Shaft Currents in Low-voltage Induction Motor for ...

Informacije MIDEM 37(2007)3, str. 152-157

tor and at measurements described in this article the Rogowski coil was also placed on the stator.

At the end it should be reminded that we may not mistake the notion of the shaft currents with the notion of the bearing currents that are closed thru the bearings without regard to mechanism of the origin and of the currents frequency. The shaft currents begin by supply of the motor from the network and should be named "inductive", but important harmonic components of these currents are in the range form 0 to 1000 Hz. At the supply from the frequency converter except low and high frequency shaft currents (Circulating Bearings Currents), therefore "inductive" currents also the "capacitive" bearing currents, begin. Characteristic high frequency of "inductive and "capacitive" harmonics currents components in dependence of the control and realization of the converter and motor may exceed 100 kHz.

The consequences of the difference of the shaft currents are necessarily the different ways in the calculation (different machine models) and measurements of these currents. The subject of this article are the measurements of the shaft currents by supply of the low voltage induction motor with squirrel cage from the network.

2. Measuring equipment for shaft currents

By measurements of the shaft currents, the standard performance Rogowski coil AmpFLEX A100 20/200 A, is used and for data assembling from the Rogowski coil two parallel devices are used: AD card NI-DAQPad-6015 and power analyzer NORMA-D6000. Parallel with the shaft currents measurements on two different ways the stator currents of the motor were measured with the help of: Rogowski coil and (AmpFLEX A100 20/200 A) and by using current shunts (Triaxial shunt 6 A...300 A).

AmpFLEX A100 20/200A is a coil of Rogowski with flexible air core, of the extent of 45 cm, produced by Chauvin Arnoux which enables the measurements of alternating currents to 200 A. The coil has two measurement ranges: to 20 A with output voltage ratio 100 mV/A and to 200 A with output voltage ratio 10 mV/A. The accuracy of the measurements is 1 % from 10 Hz to 20 kHz frequency range and angle error 1°; the output impedance is 1 kΩ.

At measurements of the shaft currents the Rogowski coil is placed in the inner space of the bearing shield of the motor in such a manner that it symmetrically embrace the bearing and the shaft is perpendicular on the area by which the Rogowski coil is closed (Figure 1.), that is essentially for the accuracy of the measurements /2, 5/.

At low voltage induction motors the distance between winding overhang and the bearing edge is only some centimeters. The vicinity of the stator winding and the Rogowski coil is the reason why the additional shaft current is meas-
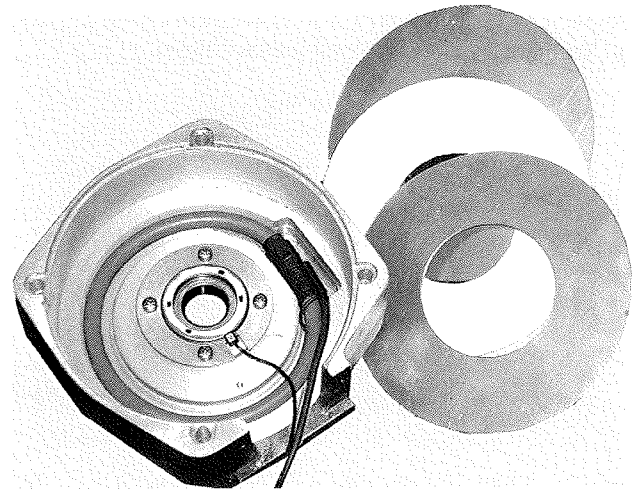
Fig. 1. Bearing shield with Rogovski coil and magnetic screen from two iron sheets and isolation

ured by the coil owing to leakage flux of the winding overhang, which size by the observed motor is approximately equal to the shaft current. To assure the correct shaft current measurements by Rogowski coil it is necessary to avoid the influence of the winding overhang leakage flux on the measured current size.
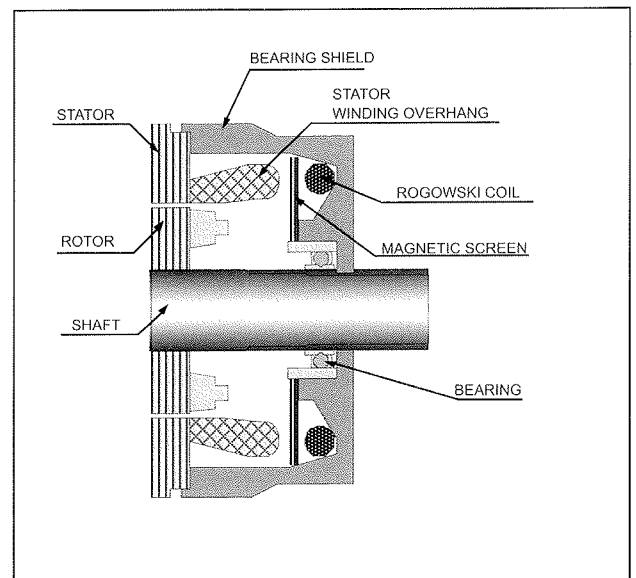
Fig. 2. Position of Rogowski coil and magnetic screen inside of the motor

With the intention to prevent the leakage flux breakthrough into the Rogowski coil two iron sheets 1 mm thickness are mounted, they are galvanic separated between each other (Figure 2.). With insertion of such magnetic screen the leakage flux from winding overhang is chocked to the level that is smaller then quantum noise (Figure 3.). Then the voltage signal from the Rogowski coil, which represents the contribution of the leakage flux of the stator winding overhang, has the smaller size than the reading resolutions (equation (7) in chapter 3. The method of measurements) and after the AD conversion the signal size of the distur-

Informacije MIDEM 37(2007)3, str. 152-157

S. Štefanko, Ž. Hederić, M. Hadžiselimović, I. Zagradišnik:
Analyses of Shaft Currents in Low-voltage Induction Motor for ...

bance could not differentiate from the error owing to round up the sample on the nearest quantum level.
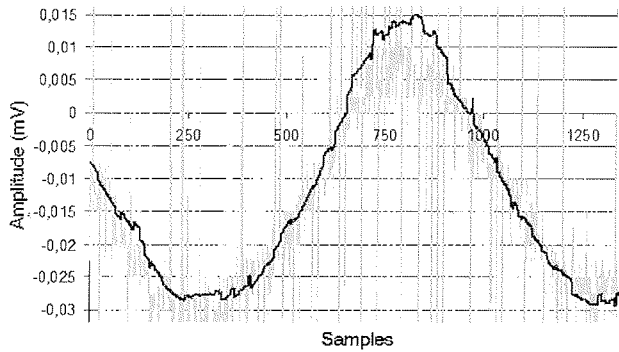
Fig. 3.   Voltage output form card NI-DAQPad-6015 by determination test of winding overhang leakage influence on the disturbances from Rogowski coil (at rated current and without rotor)

NI-DAQPad-6015 is an AD card of the producer National Instruments, which enables simple connection (plug and play) with PC via USB connector with maximum sampling frequency of 200 kS/s. The card has 16 analog inputs, 8 digital I/O, 2 analog outputs and all connectors are BNC type. The characteristics of the devices is 16 bits accuracy and the voltage range of the card is ±0,05 V (resolution 1,52 µV) to ±10 V (resolution 0,305 µV). The AD card assembled data from four Rogowski coils. One measured shaft current but three measured stator currents of the motor. The voltage output from the Rogowski coil integrator was of the span-width ±3 V and so is this selected voltage range of the measured device. The AD card doesn't have his own memory, so the data in the real time measurement via USB are transmitted in the PC where they are adjusted and stored by the help of the program package LabVIEW.

NORMA-D6000 is the power analyzer suitable for all measurements of the motors and generators. Beside of the current and voltage measurements, simultaneously the torque, speed, mechanical power and slip (together 12 input channel) could be measured. The high accuracy makes possi-

ble the precisely determination of the losses. The device has its own memory and it is possible to make FFT analysis of the measured signals on the site. With the analyzer the data about shaft currents from the Rogowski coil, stator current of the motor which is measured with the help of the current shunts, further about the motor speed, torque on the shaft and the stator voltage of the motor are collected. All data are simultaneously collected and stored in the device memory and from there the measurements are than transmitted into the PC. The current and voltage channels have the accuracy: ± (0,05 % from the measuring range, +0,02 % from measuring magnitude), therefore this instrument may be used in the lab test. Owing to the relative small voltage signal from the Rogowski coil (which measured the shaft currents) into the ratio to the smallest voltage range of ±25 V in the measured signal a 20 dB bigger noise as at AD card was present, which had the voltage range of ±3 V.

3.   Method of measurements

Shaft currents measurements demand fulfilling of some conditions so that later quality frequency processing of the measured signal would be realized. Before the beginning of the measurement it is essentially to determine the frequency range, common measured time of the signal (T_s), number of samples in this time (N), as well as the resolution of the measured signals.

By shaft current measurement the frequencies of some important harmonic component should be expected in the range from 0 to 1000 Hz. From the Nyquist theorem of sampling of signals

$$f_N \geq 2 \cdot f_{max} \qquad (1)$$

it follows that the sampling frequency (f_N) must be at least twice as big as the maximum frequency (f_{max}) that should be expected, therefore it is necessary to do the measurement with more than 2000 samples in one second. Further, that in the FFT analyses it should be possible to differentiated frequencies in the range of the 0,1 Hz (frequency resolution Δf), it follows that by the equation
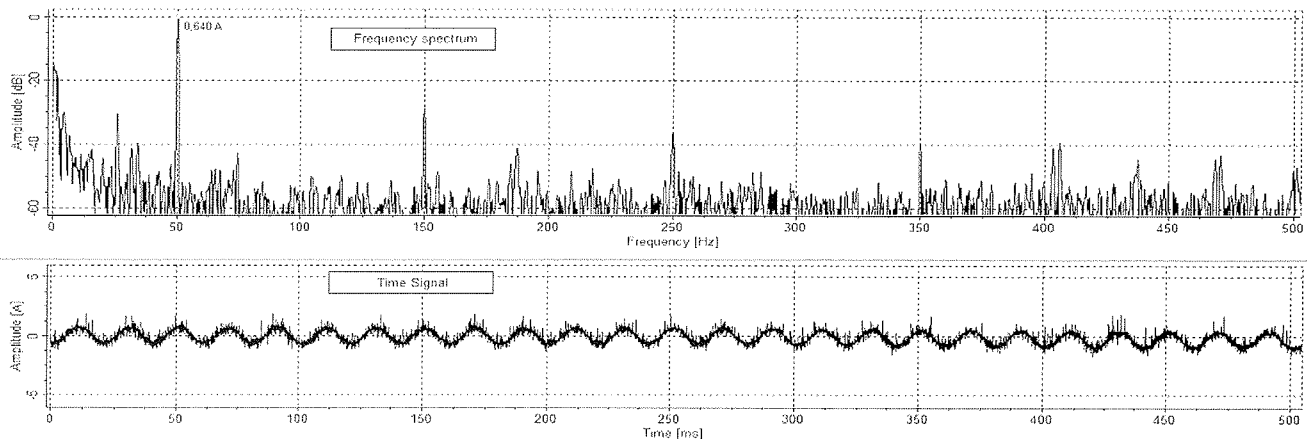
Fig. 4.   Shaft current in Y connection (NI-DAQPad-6015) – frequency specter and time signal
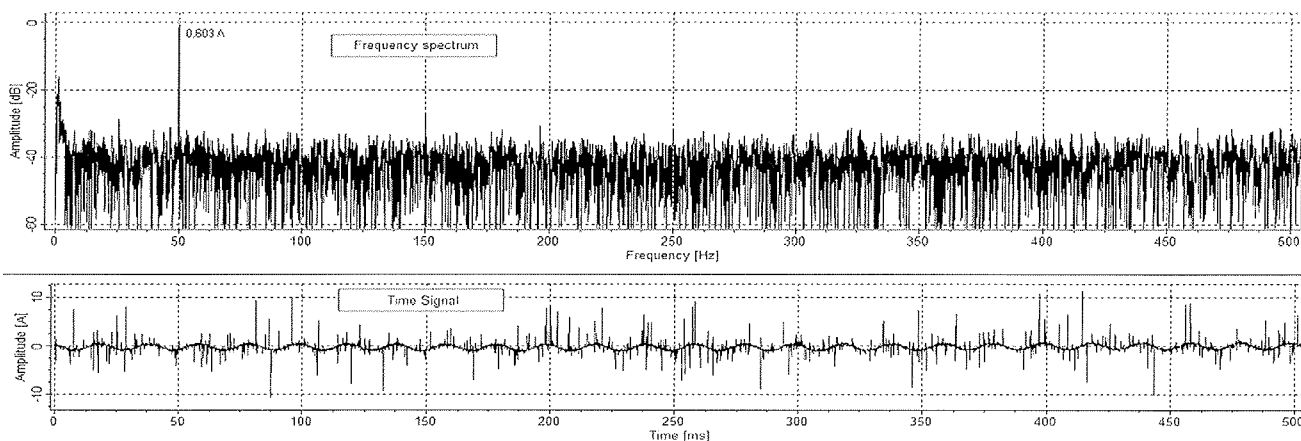
Fig. 5.   Shaft current in Y connection (NORMA-D6000) – frequency specter and time signal

$$T_s = \frac{1}{\Delta f} \qquad (2)$$

sufficient common measured time of the signal (sampling time) is 10 second.

The ratio between the sampling time and the number of the samples in this time is called the time resolution (Δt) and it represents time displacement between two samples

$$\Delta t = \frac{T_s}{N} \qquad (3)$$

The signal, which is processed by the DFT/FFT analysis, has limited time interval but it has to represent the signal of the infinite duration i.e. the supposition is that the observed interval has the periodical repetition and so the infinite duration signal is formed. This leads in the FFT analysis to the error and to the phenomena of the artificial frequencies which result in the spectrum - leakage. The leakage is manifested so that beside existent harmonic components in the original signal, after the frequency treatment, harmonic components appear which are not in the original signal. In the specter this leads to: appearance of increased noise, by which the less expressed harmonics are covered, or decrease of the amplitude of the salient harmonic components.

With the aim to decrease the leakage in the DFT/FFT analysis, filters called windows /6/ are used to be applied as the standard. Practically many types of the windows exist, but two types represent the extremes: Hanning's because of the simplicity of the algorithm is mostly used and Blackman's, which gives smaller leakage, bigger accuracy of the amplitude defining but it needs more memory space as Hanning's windows. By all types of windows their influence has improved with reduced resolution of the signal and this can be achieved in the given time of the sampling signal (T_s) by increasing in relation to (3).

The number of the samples (N_s) in the time of sampling (T_s) is called a sampling rate and represents the reciprocal value of the time resolution

$$f_s = \frac{1}{\Delta t} \qquad (4)$$

The sampling rate is limited by the possibility of measuring equipment. For data sampling at measured equipment two parallel devices are used: NI-DAQPad-6015 in NORMA-D6000.

The Card NI-DAQPad-6015 has upper boundary sampling rate f_{smax} = 200 kS/s (number of the samples in the sec) and this value is reduced reciprocal proportionally with the
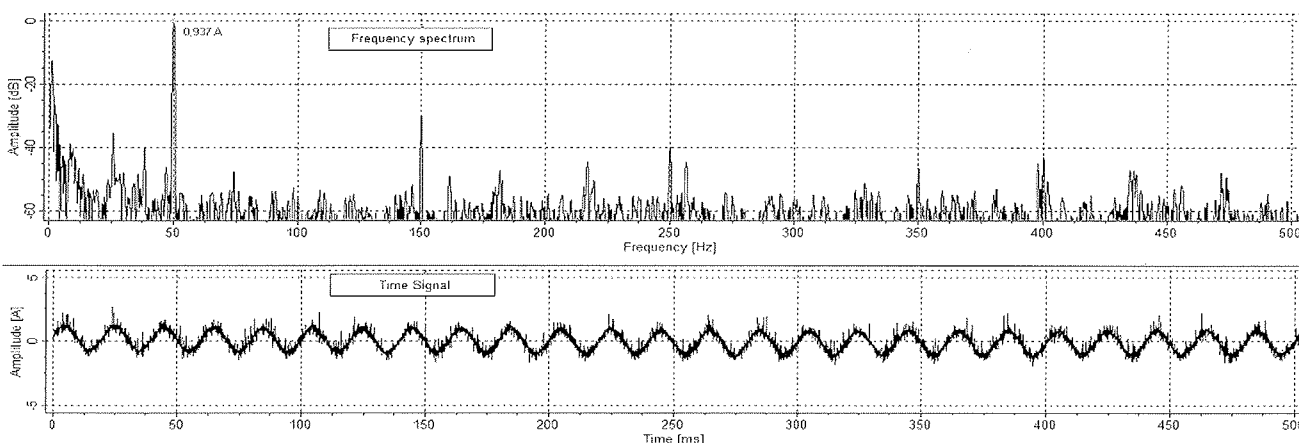
Fig. 6.   Shaft current in Δ connection (NI-DAQPad-6015) – frequency specter and time signal

Informacije MIDEM 37(2007)3, str. 152-157

S. Štefanko, Ž. Hederić, M. Hadžiselimović, I. Zagradišnik:
Analyses of Shaft Currents in Low-voltage Induction Motor for ...

number of the input signals, owing to the transmission of data via USB in the PC. Regarding to $n_i = 4$ input signals (three stator currents and shaft current) sampled by this card the maximum number of the samples is

$$N_{max} = \frac{f_{smax}T_s}{n_i} = 500.000 \text{ samples} \qquad (5)$$

Due to the operation in the real time and possible choking by the transmission and depositing of the data $N = 400.000$ samples is picked out in the total signal measurements time which gives sampling rate

$$f_s = \frac{N}{T_s} = 40 \text{ kS/s} \qquad (6)$$

for every single input signal.

The power analyzer NORMA-D6000 has the upper boundary sampling rate $f_{smax} = 50 \text{ kS/s}$ independently of the number of the input signals but limited memory storage. The power analyzer has, except of current measurements, registered also the input voltage, speed, power and torque of the motor so that on the occupation of 80 % of the memory capacity the chosen sampling rate is $f_s = 12,5 \text{ kS/s}$, which is 125.000 samples in the 10 sec sampling time.

The reading resolutions of both devices have been satisfying ($n = 16$ bits) which gives for the voltage inputs into the devices the resolution

$$\Delta u = \frac{u_{max} - u_{min}}{2^n - 1} \qquad (7)$$

By the equation (7) for AD card NI-DAQPad-6015 is

$\Delta u = \frac{3 - (-3)}{2^{16} - 1} = 91,55 \text{ } \mu V$ and for power analyzer NORMA-

D6000 $\Delta u = \frac{25 - (-25)}{2^{16} - 1} = 763 \text{ } \mu V$.

In order that by the FFT treatment of the measured signals the possibility of the appearance of the aliasing effect would be neglected the low-pass filter is used. Considering that the lowest sampling rate of the signals was by the power analyzer NORMA-D6000, then definition of Nyquist frequency by equation (1) for the filter frequency 6 kHz would be taken. Blackman window is chosen owing to bigger accuracy, but the problem of the limited memory space by FFT treatment of the signal didn't exist. The chosen sampling rates of both devices have assured 800 samples (NI-DAQPad-6015) and 250 samples (NORMA-D6000) in the period of fundamental harmonic component (50 Hz) that was for the chosen frequency resolution $\Delta f = 0,1$ Hz a satisfactory number of samples for the re-construction of signals at the inverse FFT analyze.

4. Analyses of the measurements results

All measurements are executed on four pole squirrel cage induction motor for forklift trucks /8/ with rated data in Δ connection of stator winding:

— voltage	22,5 V
— current	190 A
— power factor	0,76
— frequency	50 Hz
— rotation	1455 min^{-1}
— torque	31,5 Nm.

Shaft currents are measured in delta and star connection at lo-load, at half load and full load. Owing to better differentiation of the frequencies in the fig. 4, 5 and 6 the measurement results are given at the rated point. As to result from item 2 and 3 the time dependence of the shaft currents and frequency specter on fig. 4 and 5 are differentiated only by the convertibility analog into the digital signals - in the AD card NI-DAQPad-6015 and power analyzer NORMA-D6000. For all that the level of the noise at the power analyzer NORMA-D6000 is substantial bigger (approximately 18 dB).

In the table 1 the comparison of the r.m.s. shaft current values for star and delta connection and for AD card NI-DAQ-Pad-6015 and power analyzer NORMA-D6000 is given.

	NI-DAQPad-6015	NORMA-D6000
Y connection	0,495 A	0,920 A
Δ connection	0,705 A	1,073 A

Table 1. R.M.S. values of shat currents in Y and Δ connection of assembled data by AD card NI-DAQPad-6015 and power analyzer NORMA-D6000

With the measurements by the power analyzer NORMA-D6000 86 % bigger current is obtained in comparison with measurements by the card NI-DAQPad-6015 for the star connection and 52 % for delta connection. Mainly (but not completely) these differences can be explained with different measurement errors by equation (7). The measurement results are given on Figure 7 by card NI-DAQPad-6015 and power analyzer NORMA-D6000 in the star con-
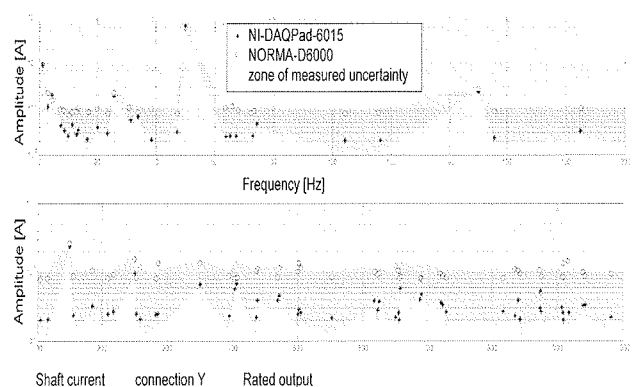
Fig.. 7. Measurements error of power analyzer NORMA-D6000 in comparison with card NI-DAQPad-6015 of important harmonics of shaft current in star connection

S. Štefanko, Ž. Hederić, M. Hadžiselimović, I. Zagradišnik:
Analyses of Shaft Currents in Low-voltage Induction Motor for ...

Informacije MIDEM 37(2007)3, str. 152-157

nection for important harmonics in the range from 0 to 1000 Hz. The shaded zone is calculated so that with regard to card NI-DAQPad-6015 measured heights (amplitudes) of the current single harmonics are for the plus added errors (Δi_{card} + Δi_{NORMA}), which come out from equation (7) and for the minus errors (-Δi_{card}). Figure 7 shows that mainly points measured by instrument NORMA-D6000 are within the zone owing to measured uncertainty.
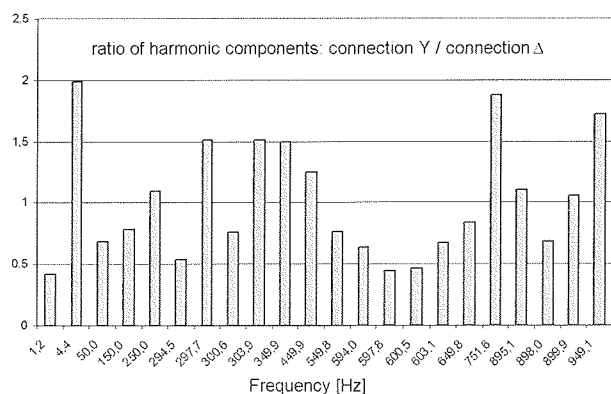
Fig. 8.   Ratio of amplitudes important harmonic
            components of shaft current spectrum:
            connection Y/connection Δ

Figure 8 shows the ratio of the amplitudes of the important harmonics of the shaft current for star and delta connection. To clear up those results, the comprehensive theoretical analyze is required, which is not subject of this article. Let us remind that theoretically determined mechanism of the beginning single harmonic current component and (theoretically) calculated frequency which differ from the measured frequencies for less then 1 % are stated. Likewise it is determined that in delta connection more different frequencies appeared as in star connection that is confirmed by the measurements.

5.   Conclusion

In the article the measurements of shaft currents on (one) low voltage induction motor with electronic equipment are represented. Special care is consecrated to the choice and preparing of the measured equipment and the method of the measurements. Analyses of some results of the measurements and the error of the measurements are done and from this it can be concluded:

-    that by relative small size of shaft currents (like in the observed motor) the special attention must be paid to the influence of the leakage field from winding overhang on the size of the measured shaft current (magnetic screen),

-    that the equipment for the measurements has to be chosen so, that the quantum noise after equation (7) is as small as possible but the ratio of the voltage with respect to the Rogowski coil current is as big as possible.

6.   References

/1/   Seinsch, H. O.; "Lagerstrome bei Drehstrom Induktionsmaschinen, Ursachen und Methoden ihrer Unterdrückung", Conti Elektro-Berichte 15, Jan./Juni 1969., pp. 43-51

/2/   Vujević, Dušan; "Rogowskijev svitak kao strujni tran-sformator", Automatika, Zagreb, 39 (1998), 3-4; str. 125-133

/3/   Ong, R.; Dymond, J.H.; Findlay, R.D.; "Comparison of techniques for measurement of shaft currents in rotating machines", IEEE Transactions on Energy Conversion, Volume 12, Issue 4, Dec. 1997, pp. 363-367

/4/   Štefanko S.; Kurtović I.; Bogut M.; Kovačević M.; Momić M.; "Broken Rotor Bar Detection in Induction Machines Using Measurement of Shaft Currents", The 2001 IEEE International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives, Grado, Italy, September 13., 2001., pp. 623-626

/5/   Erdman, J. M.; Kerkman, R. J.; Schlegel , D. W.; Skibinski, G. L.; "Effect of PWM Inverters on AC Motor Bearing Currents and Shaft Voltages", IEEE Transactions on Industry Applications, Volume 32., No. 2., March/April 1996, pp. 250-259

/6/   Harris, F.J.; "On the use of windows for harmonic analysis with the discrete Fourier transform", Proceedings of the IEEE, Volume 66, Issue 1, Jan. 1978, pp. 51-83

/7/   Bičanić, K.; Kuzle, I.; Tomiša, T.; "Nekonven-cionalni mjerni pretvarači", Energija, Zagreb, god. 55 (2006), br. 3., str. 328-351

/8/   Baša, K.; Žemva, A.; "Lead-acid Battery state-of-charge estimation for induction motor forklift trucks", Informacije MIDEM 36(2006)1, Ljubljana, str. 37-43

Stjepan Štefanko,
KONČAR - Institut za elektrotehniku, d.d.,
Fallerovo šetalište 22, 10000 Zagreb, Croatia

Željko Hederić,
Sveučilište J. J. Strossmayera, Elektrotehnički
fakultet, Kneza Trpimira 2b, 31000 Osijek, Hrvatska

Miralem Hadžiselimović,
Univerza v Mariboru, Fakulteta za elektrotehniko,
računalništvo in informatiko, Smetanova 17,
2000 Maribor, Slovenija

Ivan Zagradišnik,
Univerza v Mariboru, Fakulteta za elektrotehniko,
računalništvo in informatiko, Smetanova 17,
2000 Maribor, Slovenija

AN EFFICIENT UNIT-SELECTION METHOD FOR EMBEDDED CONCATENATIVE SPEECH SYNTHESIS

Jerneja Žganec Gros, Mario Žganec

Alpineon, Ljubljana, Slovenia

Key words: text-to-speech synthesis, embedded speech synthesis, unit-selection methods

Abstract: This paper presents a method for selecting speech units for polyphone concatenative speech synthesis, in which the simplification of procedures for search paths in a graph accelerated the speed of the unit-selection procedure with minimum effects on the speech quality. The speech units selected are still optimal; only the costs of merging the units on which the selection is based are less accurately determined. Due to its low processing power and memory footprint requirements, the method is suitable for use in embedded speech synthesizers.

Postopek za izbiro govornih segmentov pri vgrajeni polifonski združevalni sintezi govora

Kjučne besede: sinteza govora, vgrajeni sintetizator govora, izbira govornih segmentov

Izvleček: V prispevku predstavljamo postopek za izbiro govornih segmentov pri polifonski združevalni sintezi govora, pri katerem smo s poenostavitvami postopkov iskanja poti po grafu vplivali na hitrost postopka za izbiro govornih segmentov, vendar tako, da se to čim manj odraža na kvaliteti govora. Izbrani segmenti so še vedno optimalni, le cene lepljenja segmentov, na katerih temelji izbira, so manj natančne. Zaradi računske preprostosti ter majhnih pomnilniških zahtev je postopek primeren za uporabo v vgrajenih sintetizatorjih govora.

1. Introduction

Polyphone or corpus concatenative speech synthesis systems usually use extensive speech corpora containing tens of hours of recorded, segmented, and labeled speech, and use memory of several gigabytes. In such a corpus, each basic speech unit or each speech segment constituting a specific series of basic speech units or polyphones occurs repeatedly in various contexts and with different prosodic characteristics /1/.

Limitations in computational processing power and memory footprint used in embedded systems affect the planning of the unit-selection process /2/. The selection of speech units is the part of concatenative or corpus-based speech synthesis that can exert the most influence on the speed of the entire speech synthesis process.

It is necessary to find a favorable compromise between the size of the speech corpus and the computational complexity of the unit-selection procedure /1/. If the unit-selection procedure is very simplified and thus also very fast, a selection of units in a larger speech corpus can be performed in the same amount of time. Oversimplification of the procedure can, however, result in the selection of inappropriate speech units and therefore reduce the speech quality despite using a larger corpus. In contrast, choosing a complex unit-selection procedure can ensure an optimal unit selection, but because of time restrictions this can only be performed on a small speech corpus.

The paper is structured in the following way. In section 2, unit-selection in polyphone concatenative speech synthe-

sis is introduced as a graph-search problem. An overview of unit-selection methods is presented.

The unit-selection procedure with which we succeeded in accelerating the speed of the procedure without significantly affecting the speech quality is presented in Section 3. This is achieved by simplifying the calculation of the concatenation cost and thus creating conditions enabling a specific structure of the algorithm for finding the optimal path in the graph.

The evaluation of the speed and the speech quality of the proposed unit-selection procedures is presented in Section 4.

2. Unit-selection in polyphone concatenative speech synthesis

The task of unit-selection procedures is to find the most appropriate speech units in the corpus such that they produce a maximum-quality signal when merged.

Input data that the unit-selection procedure receives from language processing modules in the speech synthesizer are sequences of phonemes to be pronounced, whereby prosodic parameters for the pronunciation of each phoneme are provided. These parameters contain data on the fundamental frequency and duration of the phoneme pronunciation.

Output data that the unit-selection procedure must convey to the module for concatenating speech segments into a

speech signal are sequences of specific fragments from the speech corpus called polyphones, or the speech units that the concatenation module will have to merge. These sequences can also be equipped with prosodic parameters for each fragment, which enables the concatenating module to convert the original prosodic parameters from the corpus such that they resemble the desired prosodic parameters to the greatest extent possible.

2.1. Search graph for finding the optimal sequence of speech units

The problem of finding the optimal sequence of recorded units for quality speech signal synthesis can be presented as finding an optimal path in a graph. This kind of presentation clearly demonstrates the problem of selecting speech units and, at the same time, enables the use of recognized procedures for solving this problem. Each vertex of the graph represents a basic speech unit from the speech corpus. The basic speech segments may be allophones, diphones, triphones, or any other basic speech unit. The graph is divided into individual levels. The first level contains the initial vertices; that is, all basic speech units in the speech corpus that correspond to the first basic speech unit in the input character sequence that needs to be synthesized.

The edge between the vertices determines the possibility of merging the basic speech units represented by the connected vertices. In merging speech units, the unit of a higher-level vertex chronologically follows the unit of the lower-level vertex. This is why the edges between the vertices are directed. The vertices are interconnected such that each n-level vertex is connected to all $n+1$ level vertices.

In this kind of graph, finding the optimal speech unit sequence can be defined as finding the optimal path between any initial vertex in the graph (first level of the graph) and any final vertex in the graph (last level of the graph), whereby the edges between the graph's vertices determine the possible paths.

To start searching for the best path in the graph, criteria expressing the final goal of the speech unit selection must be defined as numeric relations between the data represented in the graph. The final goal of the speech unit selection is the maximum possible intelligibility and naturalness of the synthetic speech. In general, the following criteria have been implemented to make the speech as intelligible and natural as possible:

- The smallest possible number of speech unit concatenations,

- The smallest possible discontinuity of concatenated units at the point of concatenation,

- The best fit between the concatenated units' prosodic features and the desired speech prosody.
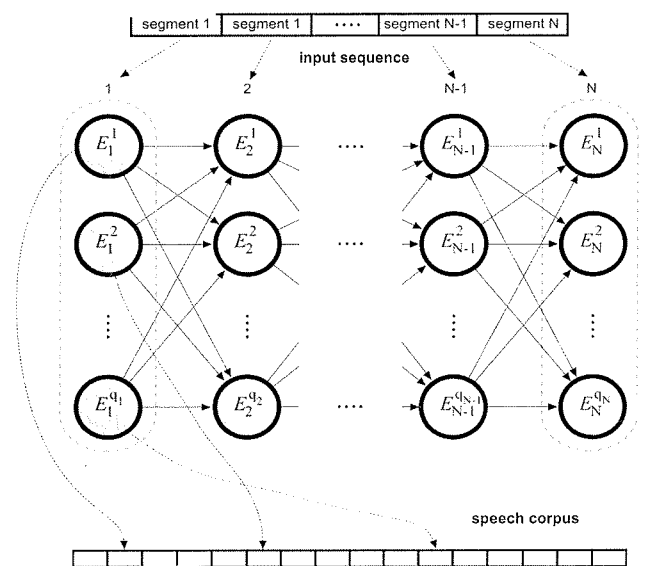
Fig. 1. Structure of the graph for finding the optimal speech unit sequence; E^i_1 are the graph initial-level vertices, E^i_N are the graph final-level vertices.

The first two criteria are evaluated by defining the concatenation cost for every edge between the vertices in the graph, whereas the last criterion is evaluated by defining the cost of fit of prosodic features for every vertex. The cost of an individual path in the graph equals the sum of the costs of vertices through which the path runs plus the sum of costs of all the edges the path contains. The optimal path in the graph is the path with the lowest cost.

2.2. The cost of fit of prosodic features

The cost of fit of prosodic features expresses the similarity or difference between the prosodic features of a specific speech unit from the speech corpus and the desired prosodic features of the part of the speech signal that the speech unit is to form. The required prosodic features can be determined as in /3/ and /4/. The cost of fit of prosodic features usually consists of the weighted result of comparing the speech unit duration and its desired duration, and of the weighted result of comparing the profile of the speech unit basic frequency and the desired fundamental frequency profile. In most cases, the ratio in which the unit's duration and the fundamental frequency profile influence the cost is determined experimentally.

In order to find the optimal speech unit sequence in the graph, the cost of fit of prosodic features is determined for each vertex. It is necessary to calculate this cost for each vertex. Although speech corpora can be very extensive, the calculation of the cost does not constitute a numeric obstacle in finding the optimal path in the graph.

2.3. Concatenation cost

A speech signal is formed by merging or concatenating speech units from the pre-recorded speech corpus. Dur-

ing the process of merging, audible speech signal discontinuities can occur. We try to evaluate the influence of signal discontinuity on the speech quality through the cost of concatenation.

There are several possible approaches to evaluating the influence of concatenation on the speech quality. The simplest method is to define the cost as "0" for concatenating speech units that directly follow one another in the speech corpus, and to define the cost as "1" for all other speech unit combinations. The use of the cost "0" in units that directly follow one another in the speech corpus is logical because they are already linked together and therefore merging is not necessary. With the use of the cost "1" in units that do not follow one another in the speech corpus, all the concatenations were equally evaluated, regardless of the characteristics of the units being merged. With this kind of concatenation cost, the procedure for finding the optimal speech unit sequence would select the sequence with the smallest number of mergers, regardless of the type of speech units.

A better evaluation of the influence of concatenation on speech quality is achieved if the cost of speech unit merging depends on the allophones that are concatenated. Similar to the previous approach, the cost "0" is defined for the merging of speech units that directly follow one another in the speech corpus.

The most accurate evaluation of the influence of concatenation on speech quality is achieved by taking into account the phonological features of both units merged when calculating the concatenation cost. In this, the differences in the fundamental frequency, formant frequencies, the amplitude, noise factor, noise spectral features, and so on can be taken into account. However, it should be noted that the use of a large number of parameters requires the determination of a large number of weights evaluating the influence of the difference in every parameter on the cost of merging. Determining these weights can be very time-consuming and often includes long-term experiments, empirical solutions, and suppositions. A great deficiency of this method of determining the cost of merging is its numeric complexity. With regard to the fact that concatenation costs are determined individually for every pair of basic speech units from the speech corpus, they are impossible to calculate in advance.

To solve this problem, we propose a compromise solution that is considerably faster, and nonetheless partly takes into account the phonological features of concatenated speech units, is determining the concatenation cost in advance for the individual groups of basic speech units from the speech corpus. In this approach, all the basic speech units in a speech corpus are classified into groups on the basis of their phonological features such that the speech units within an individual group phonologically resemble one another to the best extent possible. This is achieved by using clustering techniques. The concatena-

tion costs are calculated in advance for all group combinations and saved.

2.4. Related work

The optimal path in the graph can be reliably determined by graph traversal whereby all the possible paths in the graph are examined and the best one among them can be selected. The number of possible paths between any initial and final vertex of the graph depends on the number of graph levels and the number of occurrences of the basic speech units in the speech corpus.

Considering that a recording of a speech unit in the speech corpus can occur several thousand times and that input sequences can consist of dozens of basic speech units, it becomes clear that the number of possible paths in the graph is very large. Therefore not all of the possible paths in the graph are investigated, but various procedures are used to simplify and accelerate the search. Some procedures preserve the optimality of the solution, whereas other sacrifice optimality for the sake of faster operation.

The optimal sequence of speech units is determined by minimizing the cost that reflects a decrease in the quality of the synthesized speech due to spectral differences, differences in the phonetic environment, and mutual merging of speech units. The system that was among the first to use the selection of speech units of variable length was the ATR v-Talk /5/. In addition to all the parameters used up until then, Hirokawa also suggested the use of prosodic differences in selecting the optimal sequence of speech units /6/. In this approach, synthesized speech is created by concatenating the selected speech units and changing their prosodic features if necessary. The use of information on prosody in the speech unit selection was proposed by Campbell /7/, /8/.

The procedure for minimizing the sums of both costs employs a search based on dynamic programming or one of its derivatives such as A*. Basic speech units or phonemes are usually used as the basic search units. The existing systems that synthesize speech by concatenating speech segments from an extensive speech corpus use this procedure most frequently. The CHATR speech synthesis system was developed on the basis of these methods /9/.

By increasing the number of parameters used in finding or selecting speech segments, the size of the speech corpus has to be large enough. With a sufficiently extensive speech corpus, speech segments that resemble the required input prosodic parameters of the segments can be selected from it. In this case it is not necessary to change the prosodic features before merging the selected speech segments /10/.

Many recent studies that deal with improving the procedures for searching and defining the parameters were taken into account when calculating the cost of segments /11/, /12/. Modeling functions for calculating costs is a complex issue.

In the selection of speech segments, search procedures can use additional labeling of segments of various lengths that mark the critical parts where concatenation could result in the potential distortion of the final speech signal /13/.

Another approach to speech unit selection is the use of static modeling: FSM /14/, DCD /15/, GRM /16/, /17/, and Bulyko /18/.

3. The speech unit selection method with a simplified cost of merging

This section proposes a new and simplified speech unit selection method that is very fast and thus appropriate for implementing the concatenative speech synthesizer in embedded systems.

The basic simplification in this method is that the cost of merging two speech segments depends only on the phonemes that are being joined by merging. If merging is carried out at the center of the phonemes, such as in diphonic synthesis, the cost of merging for each phoneme is defined in the center of the phoneme.

If merging is carried out at the phoneme boundaries, the cost of merging must be defined for all the sequences of two phonemes that can occur in speech. These costs of merging can be defined in advance and are not calculated during synthesis. In addition to these costs of merging, it is presumed that the cost of merging equals "0" if the segments that are being merged directly follow one another in the speech corpus, regardless of the phonemes joined at the concatenation point.
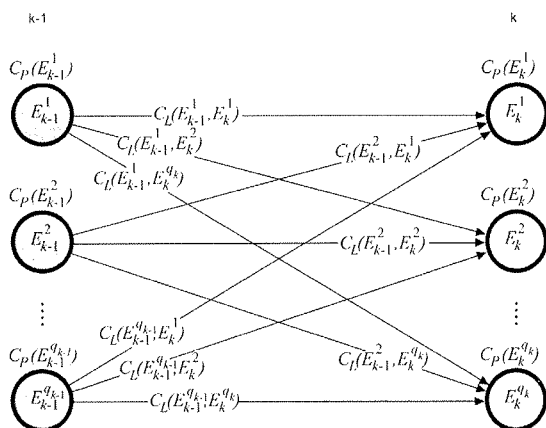
Fig. 2: The costs of merging speech segments are defined for the connections between the graph's vertices; k represents the level of the graph. The costs of fit of the prosodic features are defined for the graph's vertices.

The graph used in speech unit selection is created as described in the previous paragraph. It comprises N levels, whereby each level corresponds to exactly one basic speech segment in the input sequence that is to be synthesized. At level k of the graph, which corresponds to the speech segment S_k, q_k vertices are located; at level $k+1$, which corresponds to the speech segment S_{k+1}, q_{k+1} vertices are located; and so forth.

Every vertex E^i_k ($1 \leq i \leq q_k$) at level k of the graph represents a specific recording of the speech segment S_k in the speech corpus. For every vertex E^i_k, the cost of fit of the prosodic features of the corpus speech segment represented by the vertex is also calculated, as well as the required prosodies for the speech segment S_k in the input sequence. This cost is labeled $C_P(E^i_k)$. The vertices are connected by linking every vertex E^i_k ($1 \leq i \leq q_k$) at level k with all the vertices E^j_{k-1} ($1 \leq j \leq q_{k-1}$) at level $k-1$. The cost of the connection between vertices E^i_k and E^j_{k-1} equals the cost of merging the speech corpus segments represented by the vertices. This is labeled $C_L(E^j_{k-1}, E^i_k)$.

In finding the optimal path in the graph it must be established which path between any initial vertex E^i_1 ($1 \leq i \leq q_1$) and any final vertex E^i_N ($1 \leq i \leq q_N$) of the graph has the lowest cost.

The cost of the entire path is calculated by adding the costs of merging or the costs of edges between the vertices traversed (C_L), and the costs of fit of the prosodic features or the costs of the vertices visited (C_P). Thus, at every level k ($1 \leq k \leq N$) of the graph only one of the vertices E^i_k ($1 \leq i \leq q_k$) must be selected, or only one of the speech segments in the speech corpus that will be used in speech synthesis. This vertex is labeled $E^{x(k)}_k$. The cost of the optimal path in the graph can be expressed as:

$$C = \min_{x(1),x(2)...x(N)} \left(C_P(E^{x(1)}_1) + \sum_{k=2}^{N} \left(C_P(E^{x(k)}_k) + C_L(E^{x(k)}_k, E^{x(k-1)}_{k-1}) \right) \right).$$

The cost of the optimal path as the function of selecting a vertex $x(k)$ at the individual level of the graph is a decomposable function. If the cost of the optimal path between the graph's initial vertices and the vertex E^i_k at level k of the graph is labeled $C_O(E^i_k)$, and if the cost of the optimal path between the graph's initial vertices and any k-level vertex is labeled C_k, the following applies:

$$C_k = \min_{x(k)} \left(C_o(E^{x(k)}_k) \right)$$

and

$$C_O(E^i_k) = C_P(E^i_k) + \min_{x(k-1)} \left(C_L(E^i_k, E^{x(k-1)}_{k-1}) + C_O(E^{x(k-1)}_{k-1}) \right).$$

It can be seen that the function of the cost can be defined recursively or that the cost of the path to vertex E^i_k at level k of the graph depends only on the cost of the prosodic fit for vertex E^i_k and the costs of optimal paths to the vertices of the previous level ($C_O(E^i_{k-1})$), to which the costs of merging are added.

In optimizing such a function, dynamic programming can be used to find the optimal path in the graph. This method simplifies the search for the optimal path by dividing it into searches for partial optimal paths for every level of the graph.

In practice, the procedure is designed such that four parameters are defined for every vertex of the graph. The first, parameter $I(E^i_k)$, is an index of the basic speech unit in the speech corpus represented by the vertex. This parameter is already defined for the vertex at the start of the procedure, when the graph is being created. The second parameter equals the cost of fit of prosodic features $C_P(E^i_k)$, which is also calculated when creating the graph. The third parameter equals the lowest cumulative cost or the lowest cost of the path between any initial vertex and the current $C_O(E^i_k)$ vertex. This cost is calculated during the optimal path calculation procedure. The fourth parameter is an index of the vertex $P(E^i_k)$ from the previous level of the graph located on the optimal path between the initial vertices and the current vertex. This parameter is also calculated during the graph search procedure.

The procedure begins by defining the cost of fit of the prosodic features of the same vertices for the lowest cumulative cost of initial vertices:

$$C_O(E^i_1) = C_P(E^i_1), \ (1 \leq i \leq q_1).$$

In the initial vertices, the indicator of the vertex from the previous level of the graph is set to "0" because initial vertices have no precursor. Then the lowest cost of the path to individual vertices at the second level of the graph is defined:

$$C_O(E^i_2) = C_P(E^i_2) + \min_{j=1}^{q_1}\left(C_L(E^i_2, E^j_1) + C_O(E^j_1)\right), \ (1 \leq i \leq q_k).$$

In addition, the *(j)* index of the vertex at the pervious level of the graph located on this path with the lowest cost is recorded. This procedure is repeated sequentially for all remaining levels of the graph:

$$C_O(E^i_k) = C_P(E^i_k) + \min_{j=1}^{q_{k-1}}\left(C_L(E^i_k, E^j_{k-1}) + C_O(E^j_{k-1})\right), \ (1 \leq i \leq q; 2 \leq k \leq N) \quad (1)$$
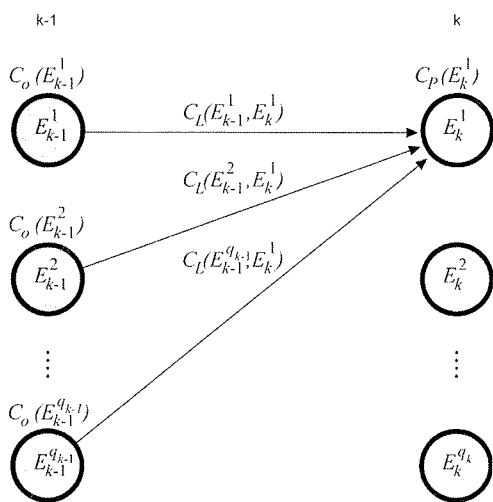
Fig. 3: *The cost of the optimal path to vertex E^i_k depends on the costs of optimal paths to the vertices at the graph's previous level $C_O(E^i_{k-1})$, the costs of merging $C_L(E^i_{k-1}, E^i_k)$, and the cost of fit of the prosodic features $C_P(E^i_k)$; k represents the level of the graph.*

The cost of the optimal path is the lowest among the costs of optimal paths to individual final vertices of the graph:

$$C = \min_{j=1}^{q_N}\left(C_o(E^j_N)\right).$$

The optimal final vertex is the final vertex with the lowest cumulative cost.

After the procedure is concluded, the sequence of vertices located on the optimal path is compiled by tracing in reverse the indices of vertices at the previous levels of the graph $P(E^i_k)$ that were saved during the procedure.

With the simplification of the cost of merging introduced in this procedure, the concatenation costs can be determined in advance, so that the cost of merging $C_L(E^i_{k-1}, E^i_k)$ depends only on the type (phonological group) of speech segments S_k and S_{k-1}. This also means that all the costs of the edges between the vertices of the graphs E^j_{k-1} and E^i_k are the same for any *j* and *i*. This does not apply only if the speech segments represented by vertices E^j_{k-1} and E^i_k, directly follow one another in the speech corpus. In this case, the cost of merging equals 0:

$$C_L(E^i_{k-1}, E^j_k) = \begin{cases} C_L(S_{k-1}, S_k) & ; I(E^j_k) - I(E^i_{k-1}) \neq 1 \\ 0 & ; I(E^j_k) - I(E^i_{k-1}) = 1 \end{cases}$$

$I(E^i_k)$ is the index or the consecutive site of the speech segment represented by vertex E^i_k in the speech corpus.

This means that the calculation of the lowest cost of the path can be further simplified. The recursive equation for calculating the lowest cost of the path to vertex E^i_k is shown in equation (1).

Taking into account the simplifications above, equation (1) can also be expressed as:

$$C_O(E^i_k) = C_P(E^i_k) + \min_{j=1}^{q_{k-1}}\begin{cases} C_L(S_{k-1}, S_k) + C_O(E^j_{k-1}) & ; I(E^j_k) - I(E^i_{k-1}) \neq 1 \\ C_O(E^j_{k-1}) & ; I(E^j_k) - I(E^i_{k-1}) = 1 \end{cases} \quad (2)$$

$C_L(S_{k-1}, S_k)$ is always a positive number. Therefore, equation (2) can also be expressed as:

$$C_O(E^i_k) = C_P(E^i_k) + \begin{cases} \min\left(C_O(E^j_{k-1}), \min_{j=1}^{q_{k-1}} C_L(S_{k-1}, S_k) + C_O(E^j_{k-1})\right), \\ \quad if \ \exists J; \ I(E^i_k) - I(E^j_{k-1}) = 1 \\ \min_{j=1}^{q_{k-1}}\left(C_L(S_{k-1}, S_k) + C_O(E^j_{k-1})\right), \quad otherwise \end{cases} \quad (3)$$

Because the calculation of the minimum in the equation above does not depend on *i*, this calculation can be performed only once for all the vertices incident to the same level S_k of the graph:

$$C'_O(S_k) = \min_{j=1}^{q_{k-1}}\left(C_L(S_{k-1}, S_k) + C_O(E^j_{k-1})\right) = C_L(S_{k-1}, S_k) + \min_{j=1}^{q_{k-1}}\left(C_O(E^j_{k-1})\right)$$

Equation (3) can now be expressed as:

$$C_O(E^i_k) = C_P(E^i_k) + \begin{cases} \min\left(C_O(E^j_{k-1}), C'_O(S_k)\right) & if \ \exists J; \ I(E^i_k) - I(E^j_{k-1}) = 1 \\ C'_O(S_k) & otherwise \end{cases}$$
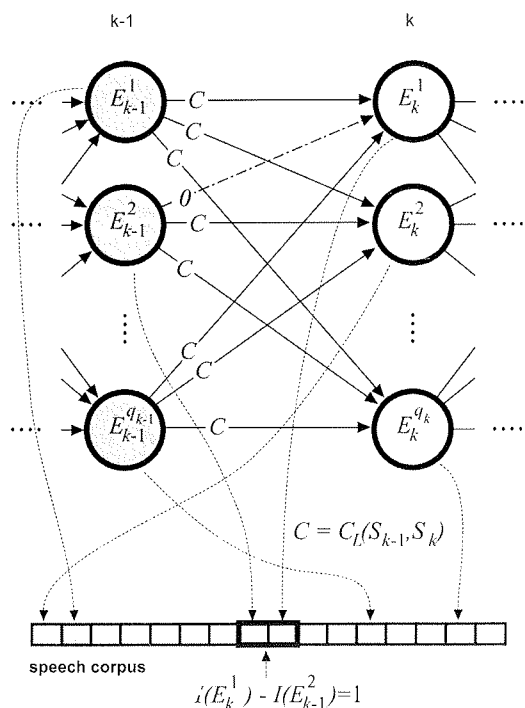
Fig. 4: The cost of merging two speech segments directly following one another in the speech corpus equals 0; k represents the level of the graph. The concatenation costs of all other segments depend only on the type (group) of speech segments that are being merged, and are therefore the same for all the connections between two levels of the graph.

It can be established that, by using the unit-selection procedure with the simplified concatenation cost described above, only one calculation of the minimum is required for every level of the graph, and only one sum and one comparison for every vertex of the graph. The time required to calculate the optimal path increases almost linearly with the increase in the size of the speech corpus.

4. Evaluation

4.1. Objective evaluation

Two versions of embedded concatenative speech synthesis using two different methods for selecting speech units were compared according to the quality of synthesized speech and the computational speed.

The first version used a unit-selection method with a simplified concatenation to select speech units described in Section 3, while the second used a simplified method for speech unit selection /19/. Using the second method, the quality of synthesized speech was slightly lower than the quality of the synthesized speech in the first procedure.

The search time in both speech unit selection methods increases linearly with the length of the utterance that is to
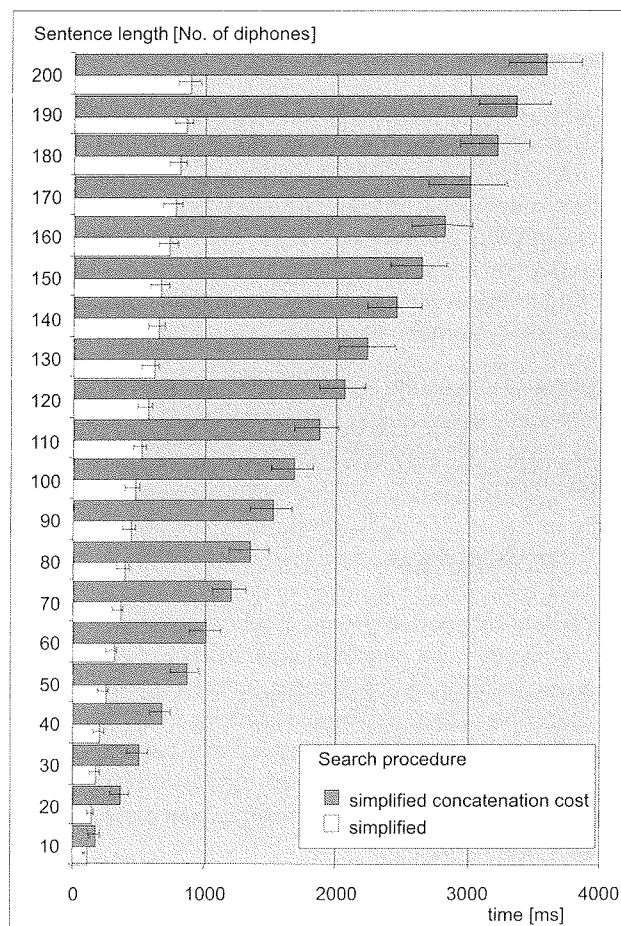
Fig. 5. Comparison of computational speed for two unit-selection search methods. As anticipated, the simplified method /19/ is faster because it does not search through all the possible paths in the graph, but limits itself only to the most promising ones. The search speed increases with the length of the sentence for which the procedure must find a suitable speech unit sequence in the speech corpus.

be synthesized, and also linearly with the size of the speech corpus, which is an improvement compared to traditional procedures for finding paths in the graph used in speech unit selection of concatenative or corpus speech synthesis. As anticipated, both methods operated increasingly more slowly when increasing the sentence size for which they were seeking the segments required for synthesis.

The simplified search method is faster than the search method with a simplified concatenation cost because it is less complex. As shown in Figure 5, it can find segments for synthesizing shorter utterances twice as fast, and segments for longer sentences four times as fast.

4.2. Subjective evaluation

The proposed method for polyphone concatenative speech synthesis was tested on an embedded device developed for this purpose /20/. A synthesizer for Slovenian speech

was embedded into an automatic system for providing information on honey yields at apicultural observation points.

The intelligibility and naturalness of the synthesized speech was evaluated using an extensive experiment prepared in line with ITU-T recommendations for testing the quality of synthesized speech. The general impression of the speech synthesizer was evaluated as 3.2, or "fair," and corresponds to the general impression of evaluations of state-of-the-art embedded speech synthesizers for other languages that usually receive grades of approx. 3.5 on the MOS scale.

Listeners evaluated the synthesized speech as intelligible, appropriately dynamic and fast, and suitable for use in automatic systems for providing oral information via telephone or the Internet.

5.   Conclusion

This article presents a new method for selecting speech units in polyphone concatenative speech synthesis, in which simplifications of procedures for finding the path in the graph increase the speed of the speech unit-selection procedure with minimum effects on the speech quality. The units selected are still optimal; only the costs of merging the units on which the selection is based are less accurately determined.

Due to its low computational speed and memory footprint requirements, the method is suitable for use in embedded speech synthesizers.

6.   Acknowledgements

7.   References

/1/   Beutnagel, M., Conkie, A., Schroeter, J. and Stylianou, Y., "The AT&T Next-Gen TTS System", in Proceedings of the 137th Meeting of the Acoustic Society of America, 2000.

/2/   Lévy, C., Linares, G., Nocera, P., Bonastre, J. F., (2004). Reducing Computational and Memory Cost for Cellular Phone Embedded Speech Recognition System, Proceedings of the ICASSP '04, Montreal, Canada, Vol. IV, pp. 489-492.

/3/   Vesnicer, B., Mihelič, F., (2004). Sinteza slovenskega govora z uporabo prikritih Markovovih modelov. Elektrotehniški vestnik, Vol. 71, No. 4, pp. 223-228. /9/ Black, A.W., Taylor, P., (1994). CHATR: a generic speech synthesis system, Proceedings of the COLING, Kyoto, Japan, pp. 983-986.

/4/   Mihelič, F., Vesnicer, B., Žibert, J., Noeth, E., (2007). Prosody Evaluation for Embedded Slovene Speech-Synthesis Systems, Inf. MIDEM, Sep. 2007, Vol. 37, No. 3.pp 176 - 181.

/5/   Sagisaka, Y., Kaiki, N., Iwahashi, N., Mimura, K, (1992). ATR ff-talk speech synthesis system, Proceedings of the ICSLP'92, Banff, Canada, pp. 483-486.

/6/   Hirokawa, T., Hakoda, K, (1990). Segment selection and pitch modification for high quality speech synthesis using waveform segments, Proceedings of the ICSLP'90, Kobe, Japan, pp. 337-340.

/7/   Campbell, W.N., Wightman, C.W., (1992). Prosodic encoding of syntactic structure for speech synthesis. Proceedings of the ICSLP. Banff, Canada. pp. 369-372.

/8/   Campbell, W.N., (1994). Prosody and the selection of units for concatenation synthesis, Proceedings of the 2nd ESCA/IEEE Workshop on Speech Synthesis. New York, USA. pp. 61-64.

/10/   Campbell, W.N., (1997). Processing a speech corpus for CHATR synthesis, Proceedings of the ICSP, Seul, Korea, pp. 183-186.

/11/   Toda T., Kawa, H. , Tsuzak, M., (2004). Optimizing Sub-Cost Functions For Segment Selection Based On Perceptual Evaluations In Concatenative Speech Synthesis, Proceedings of the ICASSP'04, pp. 657-660.

/12/   Vepa, J., King, S., (2004). Subjective Evaluation Of Join Cost Functions Used In Unit Selection Speech Synthesis, Proceedings of the In INTERSPEECH '04, pp. 1181-1184.

/13/   Breuer, S., Abresch, J., (2004). Phoxsy: Multi-phone Segments for Unit Selection Speech Synthesis, Institute for Communication Research and Phonetics (IKP) University of Bonn, Prceedings of the Interspeech'04.

/14/ Mohri, M., Pereira, F. C. N., Riley, M., (2000). The Design Principles of a Weighted Finite-State Transducer Library, Theoretical Computer Science, Vol. 231, No.1, pp.17–32.

/15/   Allauzen, C., Mohri, M., Riley, M., (2003). DCD Library - Decoder Library, software collection for decoding and related functions, In AT&T Labs - Research.

/16/   Allauzen, C., Mohri, M., Roark, B., (2004). A General Weighted Grammar Library, Proceedings of the Ninth International Conference on Automata (CIAA 2004), Kingston, Canada.

/17/   Yi, J. R. W., (2003). Corpus-Based Unit Selection for Natural-Sounding Speech Synthesis, PhD Thesis, Massachusetts Institute of Technology.

/18/   Bulyko, I., Ostendorf, M., (2001). Unit Selection for Speech Synthesis Using Splicing Costs with Weighted Finite State Transducers, Proceedings of the EUROSPEECH '01, Aalborg, Danmark. Vol. 2, pp. 987–990.

/19/   Mihelič, A., (2006). Sistem za umetno tvorjenje slovenskega govora, ki temelji na izbiri in združevanju nizov osnovnih govornih enot, PhD Thesis, Faculty of Electrical Enginering, University of Ljubljana.

/20/   Mihelič, A., Žganec Gros, J., Pavešič, N., Žganec, M. (2006). Efficient Subset Selection from Phonetically Transcribed Text Corpora for Concatenation-based Embedded Text-to-speech Synthesis. Inf. MIDEM, Mar. 2006, Vol. 36, No. 1, pp. 19-24.

dr. Jerneja Žganec Gros, dr. Mario Žganec
Alpineon , Ulica Iga Grudna 15, SI-1000 Ljubljana,
Slovenia
info@alpineon.com
tel +386 1 423 9440
tel +386 1 423 9445

SINGLE CORE HARDWARE MODULE TO IMPLEMENT ENCRYPTION IN TECB MODE

M. B. I. Reaz[1], M. I. Ibrahimy[1], F. Mohd-Yasin[2], C. S. Wei[2], M. Kamada[3]

[1]Department of Electrical and Computer Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia
[2]Faculty of Engineering, Multimedia University, Selangor, Malaysia
[3]Department of Computer and Information Sciences, Ibaraki University, Hitachi, Japan

Key words: Encryption, DES, 3DES, FPGA, Synthesis, Hardware

Abstract: The growth of the Internet as a vehicle for secure communication has resulted in Data Encryption Standard (DES) no longer capable of providing high-level security for data protection. Triple Data Encryption Standard (3DES) is a symmetric block cipher with 192 bits key proposed to further enhance DES. Many applications crave for the speed of a hardware encryption implementation while trying to preserve the flexibility and low cost of a software implementation. This project used single core module to implement encryption in Triple DES Electronic Code Book (TECB) mode, which was modeled using hardware description language VHDL. The architecture was mapped in Altera EPF10K100EFC484-1 and EP20K200EFC672-1X for performance investigations and resulted in achieving encryption rate of 102.56 Mbps, area utilization of 2111 logic cells (25%) and a higher maximum operating frequency of 78.59 MHz by implementing on the larger FPGA device EP20K200EFC672-1X. It also suggested that 3DES hardware was 2.4 times faster than its software counterpart.

Elektronski modul za izvedbo šifriranja v TECB načinu

Kjučne besede: šifriranje, DES, 3DES, FPGA, sinteza, strojna oprema

Izvleček: Porast zahtev po uporabi varnih internetnih storitev je privedel do spoznanja, da DES standard (Data Encryption Standard) ne omogoča več zelo visoke zaščite podatkov. Predlagani trojni DES (3DES), ki je simetrična šifra s 192-bitnim ključem, naj bi dodatno izboljšal DES. Izvedba 3DES standarda v strojni opremi omogoča visoke hitrosti šifriranja in poskuša obdržati fleksibilnost in nizko ceno programskih rešitev. V delu opišemo uporabo elektronskega modula za izvedbo 3DES TECB (3DES Electronic Code Block), ki smo ga modelirali z uporabo VHDL jezika. Arhitekturo smo preslikali v Alterini FPGA vezji in dosegli šifrirne hitrosti 102.56 Mbps, izkoristek površine 2111 logičnih celic (25%), in višjo delovno frekvenco 78.59MHz pri uporabi večjega vezja EP20K200EFC672-1X. Ocenili smo, da je elektronska izvedba 3DES do 2.4-krat hitrejša od programske rešitve.

1. Introduction

In the wake of advancement in computer technology and increasingly volatile information flow, we are faced with challenges of safeguarding information that is not meant for public knowledge /1/. It is common to see all sorts of electronic inventions such as the cellular phone, various devices in the military system and smart cards today /2/.

The growth of the Internet has contributed to the increase in the amount of data transferred daily across regions. These data transmissions may contain funds amounting to millions of dollars or government records. However, these applications require high data security. The ease in obtaining and duplicating these data through resourceful parties /e.g. hackers/ has resulted in a decline in confidence amongst Internet users towards online transaction. As such, it is essential to ensure the privacy and authenticity of these data. One of the existing methods that can be used to guard the security and authenticity of data through the Internet is through cryptography.

Data Encryption Standard, DES has been the world wide standard for more than 20 years /3/. DES is used in IPSec

protocols, secure socket layer (SSL) protocol and ATM cell encryption. During those years, bundles of software and hardware had been developed to implement this algorithm. However due to the need of higher security, 3DES had been chosen based on its close relationship to DES /4/. Triple DES is an improved version of DES and provides better security compared to DES. This is due to its longer key length and more rounds of DES encryptions. DES only has an effective key length of 56 bits, which is insufficient to resist any brute force attack today /3/. Research has shown that a key-breaking machine that costs less than $1 million can find a key in an average of 3.5 hours and the cost is estimated to drop by a factor of 5 every 10 years /3/. Even though 3DES is three times slower than DES, if used properly, it can be as strong as the 2304-bit public key algorithm because it has longer key length. With an increase in its security standards and compatibility to the DES software and hardware, 3DES is clearly a better choice compared to other algorithms such as RSA and ECC /5/.

Due to its symmetric nature, 3DES is a better choice in encrypting bulk data and is therefore less expensive /1, 6/.

Informacije MIDEM 37(2007)3, str. 165-171

M. B. I. Reaz, M. I. Ibrahimy, F. Mohd-Yasin, C. S. Wei, M. Kamada:
Single Core Hardware Module to Implement Encryption In TECB Mode

3DES uses only 128 or 196 bits symmetric keys and has simpler algorithm. It is less complicated, less computationally intensive and does not introduce much overhead. Thus, it requires relatively inexpensive hardware /1, 3/. 3DES is faster than RSA. Due to its much longer key length, RSA causes high-level resource utilization and is not suitable to be used in mobile or wireless devices as these devices have underpowered processors /6/.

In comparison to AES, 3DES is faster /7/. The limitation of AES exists because the cipher and its inverse use different codes and/or tables. As such, it does not have bidirectional architecture for encryption and decryption as that of the 3DES. The inverse cipher can only partially reuse the circuitry that implements the cipher, resulting in a larger hardware presumably.

Hardware realization of cryptography has the advantages of being more secured and faster in speed /8/. It gives a higher performance as desired /9/. Even though software implementation of cryptography uses general-purpose processors that offer enough power to satisfy the needs of individuals, hardware realization is the only way to achieve speeds that is more significant than the general-purpose microprocessor /10/. This feature is important for commercial and communication purposes as this is shown in /9/ that security related processing can consume up to 95% of a server's processing capacity. By using a dedicated hardware to run encryption application, more computing can be done within a stipulated period due to parallel processing. Field Programmable Gate Array (FPGA) offers a potential alternative to speed up the hardware realization. From the perspective of computer-aided design, FPGA comes with the merits of lower cost, higher density, and shorter design cycle. The programmability and simplicity of FPGA made it favorable for prototyping digital system.

In this paper, the framework of FPGA-based hardware realization of cryptography using 3DES is proposed. With this approach, both the speed and performance are preserved without the need to trade-off between these two important criteria in encryption and decryption. In this method, VHDL (Very High Speed Integrated Circuit Hardware Description Language) is selected as the hardware description language to realize the system.

2. Triple DES Algorithm

Triple DES encrypts a block of 64-bit data using two or three unrelated 64 bits keys /5/. The internal operation done on these data is similar to that of DES where the only difference is that DES consists of 16 iterations whereas 3DES consists of 48 iterations. In other word, 3DES contains three successive DES operations. Out of the 64-bit key used in DES, the effective key size is only 56 bits. The eighth bit in each byte is used for odd parity checking and is thus ignored. As such the total effective key size for 3DES is 168 bits.

A DES encryption operation is divided into two stages involving the key and the data. In the first stage, 16 subkeys are created from the key whereas the encryption of data message is occurred in the second stage.

During the first stage, permutation is initially done on the 64-bit key and resulting in a 56-bit permutated key. After key permutation, this 56-bit permutated key is divided into left and right halves C_0 and D_0, where each half has 28 bits. With C_0 and D_0 defined, sixteen blocks C_n and D_n, where n = 1, 2, 3,..., 16 are formed by left shifting C_{n-1} and D_{n-1} (once or twice). C_n and D_n are then concatenated to form a 56-bit data, C_nD_n. This 56-bit data is then permutated and resulted in 48-bit subkeys formed. After 16 iterations, 16 sets of subkeys are created. These subkeys are used for data encryption during the second stage.

During the second stage, permutation is done on the 64-bit message data block, M. As this is the first permutation process being done on the data, it is called Initial Permutation. The permutated data are then divided into left half L_0 and right half R_0, each having 32 bits. It is followed by 16 iterations of operations, using function f, which operates on two blocks: data block of 32 bits and subkey block of 48 bits to produce an output block of 32 bits.

$$L_n = R_{n-1} \qquad (1)$$

$$R_n = L_{n-1} + f(R_{n-1}, K_n) \text{ where } n = 1, 2, 3,..., 16 \qquad (2)$$

As shown in (1) and (2), during each of the 16 iterations, the right 32 bits of the previous iteration, R_{n-1} is used as the left 32 bits of the current iteration, L_n. The right 32 bits in the current iteration, R_n is obtained by implementing XOR to the left 32 bits of the previous step with f function.

To calculate f function, each block R_{n-1} is expanded from 32 bits to 48 bits and the expanded R_{n-1}, $E(R_{n-1})$ is then XORed with the block of subkey K_n, i.e.,

$$K_n + E(R_{n-1}) = B_1B_2B_3B_4B_5B_6B_7B_8 \qquad (3)$$

where n = 1, 2, 3,..., 16 and B_i is a group of 6 bits. This results in a 48-bit block, which is then divided into $B_1B_2B_3B_4B_5B_6B_7B_8$. Each B_i gives an address in a different S box, S_i. The 4-bit blocks for the entire eight S boxes are combined to form a 32 -bit block.

Function f is obtained by implementing permutation on the group output such as,

$$f = P(S_1(B_1)S_2(B_2)... S_8(B_8)) \qquad (4)$$

At the end of the sixteenth iteration, the order of the two blocks $L_{16}R_{16}$ is reversed to $R_{16}L_{16}$ before applying the permutation on the reversed block. This is the last permutation to be done on the data, thus being called the Final Permutation.

Decryption in DES uses the same process as the encryption operation. The only difference lies in the order in which

the subkeys are used. In the decryption process, the subkeys are used in reverse order, meaning that K_{16} is applied first with K_1 being applied last.

Triple DES shows a high level of similarity in operation to that of DES. Encryption and decryption in 3DES are done by compounding the operation of DES encryption $E_k(l)$ and decryption $D_k(l)$ operations. Encryption operation in 3DES is defined by,

$$Encryption = E_{K3}(D_{K2}(E_{K1}(l))) \qquad (5)$$

whereas the decryption operation is defined by,

$$Decryption = D_{K1}(E_{K2}(D_{K3}(l))) \qquad (6)$$

From equation (5), it shows that the plaintext is first encrypted by $K1$ using DES. The encrypted data is then decrypted by $K2$ before being encrypted by $K3$. In contrast to that, equation (6) indicates that the 3DES cipher text is initially decrypted by $K3$ using DES, whereby the result is then being encrypted by $K2$. The plaintext is recovered by decrypting the output from second DES operation by $K1$.

Final permutation is actually the inverse operation of initial permutation. As such in a 3DES operation, the initial permutation of the second DES round cancels the final permutation of the first DES round. This is the same in the third DES round where its initial permutation cancels off the final permutation of the second DES round, leaving only an initial permutation and a final permutation during the whole 3DES operation.

3. Design Flow of 3DES Single Core Module

The specification of the 3DES core is set prior to the start of the design process. Different 3DES operation mode could result in different design complexity and different level of security. As such, a trade off between these two conditions must be taken into consideration during the design stage. As to avoid complicated design, 3DES Electronic Code Book (TECB) had been chosen as the mode of operation in this project. This resulted in reduced area utilization and compromised security level in the core design.

Due to varying number of bits being shifted during the different iteration rounds, normal shift register could not be used. A counter had to be added in the design so as to determine the current iteration round. The input signals to the shifting module were shifted appropriately depending on the output of the counter. The output of counter must be passed correctly to the shifting module. Error in connections such as MSB of the counter output being connected to the LSB of the shifting module input could result in error in bit shifting.

During DES encryption operation, the subkeys were transmitted in the sequence of 1 to 16 whereas during DES decryption operation, the subkeys were transmitted in the sequence of 16 to 1. The initial design in mind was to have

a multiplexer and a demultiplexer. The 16 subkeys were multiplexed. This was then followed by demultiplexing these multiplexed subkeys either in the sequence of 1 to 16 or from 16 to 1. The sequence in which the subkeys were sent out was determined by the select signal of the demultiplexer. However, this design was difficult to implement as complicated control signals were needed to obtain the 16 subkeys in the correct sequence. These subkeys had to be stored in registers before being demultiplexed.

Based on this, the design of the full implementation of 3DES encryption engine was produced as shown in Figure 1. The multiplexers and demultiplexers in Figure 1 played the role of realising the 16 iterations in a DES operation and the 48 iterations in the 3DES operation.
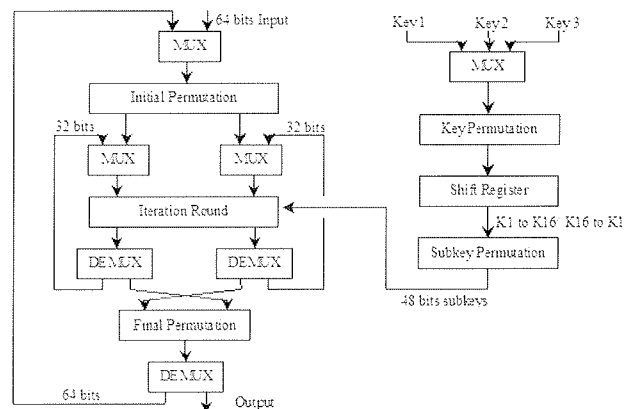
Fig. 1:   Full implementation of 3DES

4. Simulation and Synthesis

4.1 Functional Simulation

Functional simulation was done to verify that the design behaved as expected on the VHDL coding using MAX PLUS II software. Since the delay of the combinational logic and wires were not known, the signal suffered only a constant signal change delay of 0.1ns. This delay must be taken into consideration. As the design operated on positive clock edge, this delay could cause the response of the circuit to be delayed by 1 clock cycle. As such, the processes of the other modules were also delayed 1 clock cycle so as to synchronize the operation of the whole design.

The validation of the 3DES operation was done by referring to /11/, where the data message was encrypted based on DES. Validation of the above design after 16 iterations showed the same result as that of DES in /11/. To further verify the design, a simulation done for the whole 3DES operation showed that the encrypted input data could be decrypted to recover the original data message. The two aforementioned verification methods indicated that this design implemented 3DES correctly. The timing diagrams obtained from the simulation that was done during the verification process are shown in Figure 2 and Figure 3 for encryption and decryption respectively.

Informacije MIDEM 37(2007)3, str. 165-171

M. B. I. Reaz, M. I. Ibrahimy, F. Mohd-Yasin, C. S. Wei, M. Kamada:
Single Core Hardware Module to Implement Encryption In TECB Mode
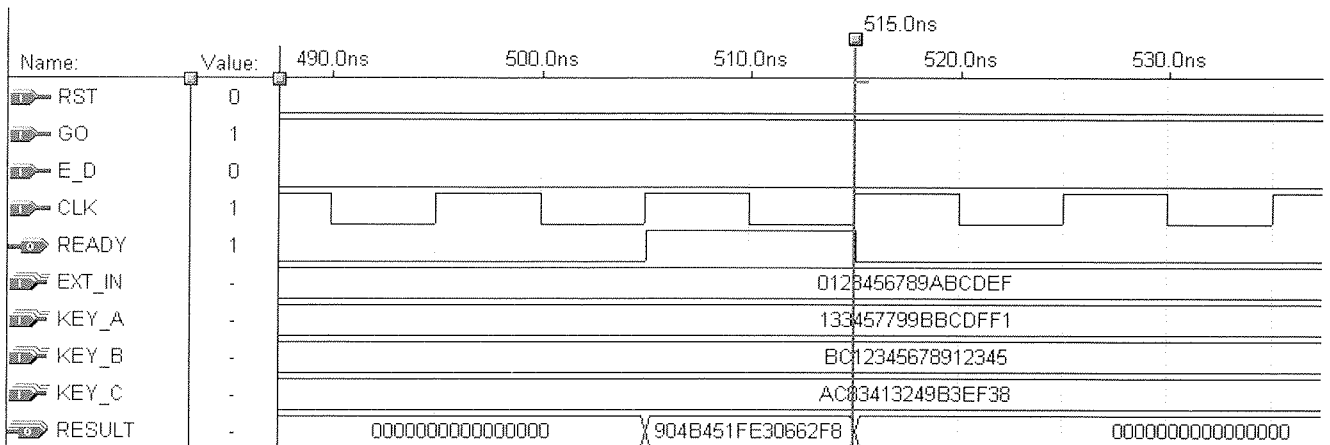
Fig. 2: Functional simulation of encryption operation
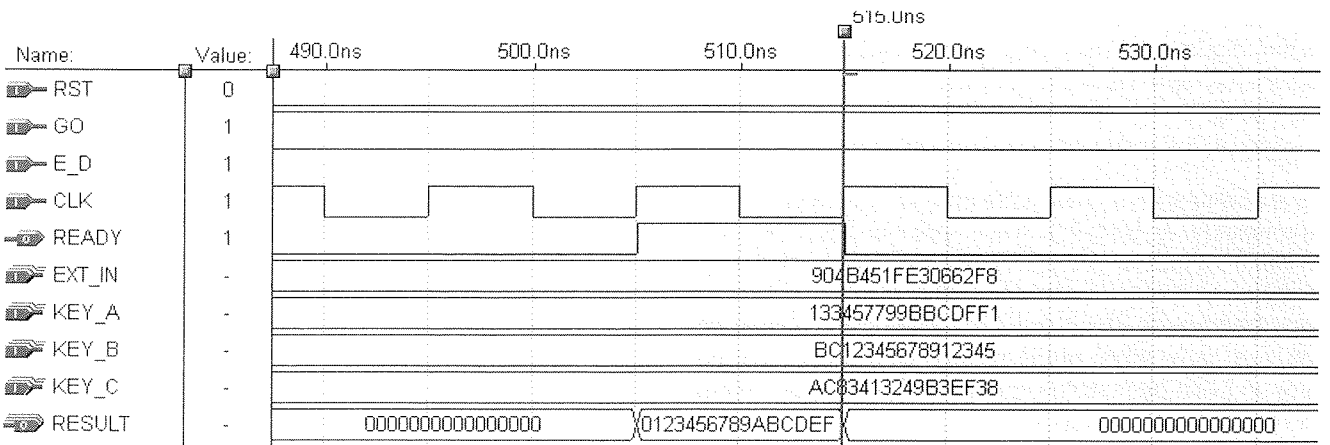
Fig. 3: Function simulation of decryption operation

From Figure 2, the keys used in the encryption of data message 0123456789ABCDEF were 133457799BB-CDFF1, BC12345678912345 and AC83413249B3EF38. The cipher resulted text was 904B451FE30662FB. During the decryption operation, the cipher text was decrypted to obtain the original data message of 0123456789ABCDEF as shown in Figure 3.

4.2 Synthesis and Optimization

With the functional simulation showing the correct behavioural result, synthesis was done using Altera Quartus II 4.0 software on the core design implemented into FPGA. Device family that could fit the design into it was chosen and the timing requirements were set.

Different FPGA could result in different maximum frequency obtained. A larger FPGA device family such as APEX20KE gave a higher maximum clock frequency than FLEX10KE. This was an important criterion to be considered while deciding on the FPGA to be used even though the design could be fitted into both. The smaller device family had a higher resource utilization percentage. By deciding to use a larger device family, speed optimization had been given priority in view of excessive amount of resource in the FPGA selected.

There was a trade off between area and speed. A higher number of logic elements used that resulted in higher maximum operating frequency. Initial synthesis of the design on APEX20KE family gave a maximum frequency of 72.77 MHz. However, after switching off the 'Remove Duplicate Registers' and 'Remove Duplicate Logic' setting, the maximum operating frequency achieved approximately 77 MHz. The logic cells used that summed up to be 25%, which was 2% more than the previous setting. By setting the maximum frequency requirement to 80MHz, a higher value of 78.59 MHz was achieved. This was the highest maximum clock frequency value with APEX20KE family that could be obtained from the optimization process.

4.3 Timing Simulation

Timing simulation was performed to verify that the module functioned correctly and there were no timing violations in the implemented design. The functional simulation was done using MAX PLUS II software but the timing simulation was done using Quartus II 4.0 software.

During timing simulation, the total delay of the wires and combinational logic was taken into account. Initial testing using the clock signal having frequency that was higher than that of maximum operating frequency resulted in er-

roneous output. The result obtained was not the encrypted data message. This was because the encrypted data cannot be decrypted to recover the initial data message. The total delay had exceeded one clock cycle period.

The clock signal period was then set to 13ns. This clocking period was larger than the total wire and combinational logic delay. Different sets of keys and input data blocks were used during the simulation. It was found that the encrypted data could be decrypted to recover the original data. Besides that, the reset pin had also been tested. Reset signal was set to 'high' to reset the design.

4.4   Synthesis Results

Table 1: Synthesis results

Family	APEX20KE
Device	EP20K200EFC672-1X
Name	Core
Total logic elements	2111 / 8320 (25%)
Total I/O pins	325 / 376 (86%)
Total memory bits	2048 / 106496 (1%)
Total PLLs	0 / 2 (0%)
Total combinational functions	2110
Total registers	408
Performance, f_{max}	78.59 MHz
Clock period	12.724 ns

Table 1 shows the synthesis results of the 3DES encryption engine. The FPGA family that had been selected for the realization of 3DES encryption engine was APEX20KE (more precisely, EP20K200EFC672-1X).

Out of the 8320 logic elements contained in the device, a total of 2111 logic cells were used. A total of 325 I/O pins were utilized, which is equivalent to 86 percent of the total pins in the device. Out of these 325 pins, 65 pins were output pins while the remaining 260 pins were input pins. Out of a total of 106496 memory bits in the device, 2048 of them were utilized. This is equivalent to 1 percent of the total memory bits resource. Besides that, the total number of registers used in the EP20K200EFC672-1X device

summed up to be 408. A maximum clock frequency of 78.59 MHz was obtained. The clock signal that was used in the device must have a period of at least 12.724 ns. Any period below this value gave a faulty result.

4.5   Timing and Area Analysis

The results for timing and area analysis of the main modules are presented in terms of maximum operating frequency and logic cell (LC). The analysis was done using Quartus II software. The devices chosen for the implementation were EP20K200EFC672-1X of APEX20KE family and EPF10K100EFC484-1 of FLEX10KE family. Comparison was done between the two devices.

Tables 2 and 3 show the effect of registers and logic cells duplication in EP20K200EFC672-1X and EPF10K100EFC484-1 respectively when the full 3DES architecture was mapped into them. To implement these features, the 'Remove Duplicate Registers' and 'Remove Duplicate Logic' settings were selected or deselected.

When the 'Remove Duplicate Registers' and 'Remove Duplicate Logic' settings were selected during the hardware implementation of the encryption module in EP20K200EFC672-1X, this resulted in lower area utilization of 1984 logic cells and lower maximum operating frequency of 72.77 MHz. When these settings were deselected, higher area utilization of 2111 logic cells and higher maximum operating frequency of 78.59 MHz was obtained.

However, that is not the case when EPF10K100EFC484-1 was used. Selecting the 'Remove Duplicate Registers' and 'Remove Duplicate Logic' setting resulted in lower area utilization but higher maximum operating frequency.

Table 4 shows the synthesis results for the final design of the project. Two devices were used, namely EP20K200EFC672-1X and EPF10K100EFC484-1. EP20K200EFC672-1X is a larger device compared to EPF10K100EFC484-1.

Table 2: Effect of registers and logic cells duplication in EP20K200EFC672-1X

'Remove Duplicate Registers' and 'Remove Duplicate Logic'	Area (LC)	Clock Period (ns)	Maximum Operating Frequency (MHz)
On	1984 / 8320 (23%)	13.742	72.77
Off	2111 / 8320 (25%)	12.724	78.59

Table 3: Effect of registers and logic cells duplication in EPF10K100EFC484-1

'Remove Duplicate Registers' and 'Remove Duplicate Logic'	Area (LC)	Clock Period (ns)	Maximum Operating Frequency (MHz)
On	1924 / 4992 (38%)	17.5	57.14
Off	2080 / 4992 (42%)	17.6	56.82

Informacije MIDEM 37(2007)3, str. 165-171

M. B. I. Reaz, M. I. Ibrahimy, F. Mohd-Yasin, C. S. Wei, M. Kamada:
Single Core Hardware Module to Implement Encryption In TECB Mode

Table 4: Synthesis results

Device	Area (LC)	Clock Period (ns)	Maximum Operating Frequency (MHz)
EP20K200EFC672-1X	2111/ 8320 (25%)	12.724	78.59
EPF10K100EFC484-1	1924 / 4992 (38%)	17.5	57.14

When the larger device was used, it was found that the final design had a higher maximum operating frequency of 78.59 MHz. It utilized more logic cells. However, when the smaller device from FLEX10KE family was used, it only had a maximum operating frequency of 57.14 MHz. Besides that, the design used only 1924 logic cells of the resource, which was lesser than the 2111 logic cells used in EP20K200EFC672-1X.

With this, it can be concluded that the mapping of the design architecture on different devices can result in different maximum operating frequency and area utilization. A larger device results in higher maximum operating frequency and larger area utilization. As such, considerable decision must be taken on whether a faster operation is needed or a smaller device is required.

Figure 4 demonstrates the RTL view of the core entity. It is shown that core was formed by three smaller entities, namely s_mac, key_block and inp_block. Each of these entities had its own unique function. S_mac controlled and synchronized the operations of the other two entities while key_block processed the three keys, producing the subkeys needed before sending them to inp_block. Inp_block was the entity where the actual encryption and decryption of the plaintext occurred.

Table 5: Comparisons between hardware and software implementation

	3DES (FPGA)	3DES (software)
Key size (bits)	192	192
Data rate (Mbps)	102.56	42.9

Table 5 shows the comparisons done on the performances of the hardware and software implementation of 3DES.

Triple DES was implemented into FPGA and as well as MATLAB using an Intel Pentium III 866 MHz machine. It shows that 3DES hardware was significantly (2.4 times) faster than its software counterpart. The 3DES software could only manage a data rate of 42.9 Mbps compared to 102.56 Mbps of 3DES hardware.

5. Conclusions

The hardware implementation of 3DES encryption engine on FPGA chip was realized. The chip selected was EP20K200EFC672-1X of APEX20KE family. It could encrypt data at a rate of 102.56 Mbps, with a maximum operating frequency of 78.59 MHz and area utilization of 2111 logic cells.

The throughput of 102.56 Mbps in the current full implementation of 3DES core can be considered as low by industry standard. As such, to improve the throughput of the design, pipelining of the iterations process can be implemented. Registers can be added to store data during the pipelining process. This will invariably reduce the maximum clock frequency; however the number of clock cycles being used for one complete 3DES operation can be greatly reduced, thus reducing the latency.

To allow more secured encryption process, additional 3DES operation modes can be added to the core module. Currently, the encryption hardware only operates under TECB mode. By including more modes of operation, users can choose to operate under certain mode, depending on their preference.
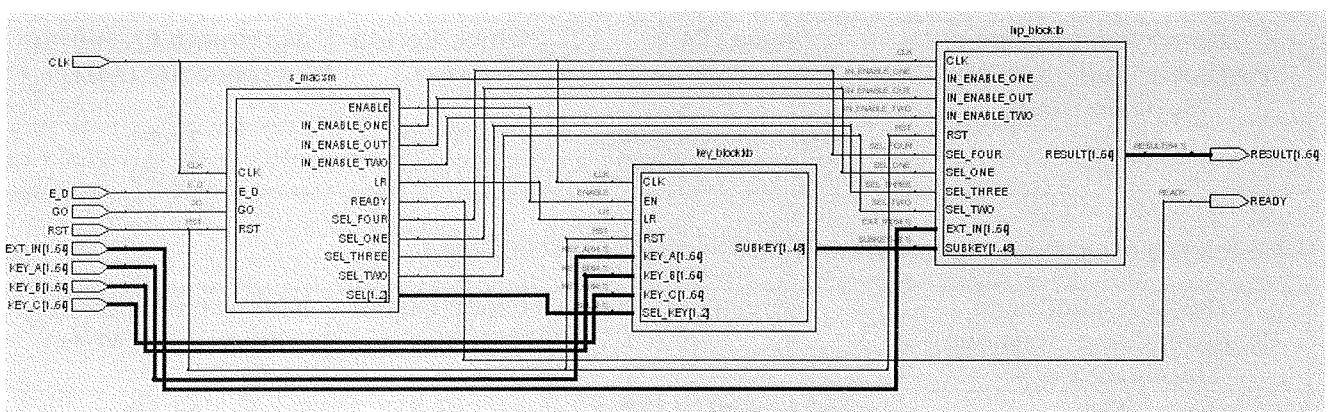

Fig. 4:   RTL view of core entity

References

/1/   Aladdin Knowledge System, "The enduring Value of Symmetric Encryption", White Paper, pp: 5-8, August 2000.

/2/   Harper, S. and Athanas P., "A Security Policy Based Upon Hardware Encryption", System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference, pp: 190 – 197, Virginia, 5-8 Jan. 2004.

/3/   Davor Runje, Mario Kovac, "Universal Strong Encryption FPGA Core Implementation", Design, Automation and Test in Europe, 1998, Proc., pp: 923-924, France, 23-26 Feb 1998.

/4/   O.Y.H.Cheung, P.H.W.Leong, "Implementation of an FPGA Based Accelerator for Virtual Private Networks", Proceedings of IEEE International Conference on Field-Programmable Technology (ICFPT), pp: 34-41, Hong Kong, 2002.

/5/   "Data Encryption Standard", Federal Information Processing Standards (FIPS) Publication 46-7, National Institute of Standards and Technology (NIST), USA, 1999.

/6/   Young Sae Kim, Woo Seok Kang and Jun Rim Choi, "Implementation of 1024-bit Modular Processor for RSA Cryptosystem", AP-ASIC 2000, Proceedings of the Second IEEE Asia Pacific Conference, pp: 187-190, Korea, 28-30 Aug. 2000.

/7/   C. Sanchez-Avilla & R. Sanchez-Reillo, "The Rijndael Block Cipher (AES Proposal): A Comparison with DES", Security Technology, 2001 IEEE 35th International Carnahan Conference, pp: 229-234, London, 16-19 Oct 2001.

/8/   Raghuram, S.S. and Chakrabarti, C, "A Programmable Processor for Cryptography", Proceedings. The 2000 IEEE International Symposium on Circuits and Systems, Volume: 5, pp: 985-688, Geneva Switzerland, 28-31 May 2000.

/9/   Lisa Wu, Chris Weaver, and Todd Austin, "Crypto Maniac: A Fast Flexible Architecture for Secure Communication", Computer Architecture, 2001. Proceeding. 28th Annual International Symposium, pp: 110-119, Goteborg, Sweden, 30 June-4 July 2001.

/10/   Pawel R. Chodowiec, "Comparison of the Hardware Performance of the AES Candidates Using Reconfigurable Hardware", Master Thesis, 150 pages, 2002, George Mason University.

/11/   J. Orlin Grabbe, www.aci.net/kalliste/des.htm, 5 Jan 2004.

M. B. I. Reaz[1], M. I. Ibrahimy[1],
F. Mohd-Yasin[2], C. S. Wei[2], M. Kamada[3]
[1]Department of Electrical and Computer Engineering,
International Islamic University Malaysia,
53100 Kuala Lumpur, Malaysia
[2]Faculty of Engineering, Multimedia University,
63100 Cyberjaya, Selangor, Malaysia
[3]Department of Computer and Information Sciences,
Ibaraki University, Hitachi, Ibaraki 316-8511, Japan
ibrahimy@iiu.edu.my

GAIN CONTROL LOOP FOR A 2.4 GHZ VARIABLE GAIN LOW NOISE AMPLIFIER (VGLNA)

Lini Lee, Roslina Mohd Sidek, Sudhanshu Shekhar Jamuar and *Sabira Khatun

Department of Electrical and Electronic Engineering,
*Department of Computer System and Communication Engineering, Faculty of
Engineering, University Putra Malaysia (UPM), Selangor, Malaysia.

Key words: continuous gain variation, high gain, linearity, low NF, radio frequency, variable-gain LNA

Abstract: The most critical point of a integrated receiver is the radio frequency (RF) input and the first stage of the receiver is a low noise amplifier (LNA). Thus, LNAs with low power consumption and excellent properties in terms of gain, noise and linearity are highly in demand. Moreover, a variable-gain LNA which can prevent saturation in the receiver when the input signal becomes relatively large has the advantage of no additional attenuator or variable gain amplifier is required. Thus, power consumption, chip size and cost can be minimized at the same time. Hence, this work proposes a two-stage variable gain LNA with large and continuous gain variation. By introducing a simple gain-control loop at the second stage of a cascade LNA, a continuous gain tuning range of approximately 14 dB is achieved. The VGLNA is designed using 0.18 μm CMOS technology and targeted for applications at 2.4 GHz. The maximum gain and NF of this VGLNA are 23 dB and 1.12 dB respectively. The DC power consumption is reported to be 12.4 mW from a 1 V power supply. Comparison is made with available circuits to show that this VGLNA has the advantage of a continuous tuning range and low NF, and this is accomplished without degrading its performance.

Kontrolna zanka ojačanja za 2.4GHz nizkošumni ojačevalnik s spremenljivim ojačanjem (VGLNA)

Kjučne besede: spreminjanje ojačanja, visoko ojačanje, linearnost, nizko šumno število, radijske frekvence, ojačevalniki s spremenljivim ojačanjem

Izvleček: Najbolj občutljiva točka integranega sprejemnika je RF vhod z nizkošumnim ojačevalnikom (LNA). Zatorej je povpraševanje po LNA ojačeval-nikih z nizko porabo ter drugimi odličnimi lastnostmi, kot so ojačanje, nizek šum in linearnost, veliko. Prednost imajo predvsem LNA ojačevalniki s spre-menljivim ojačanjem, ki ne gredo v zasičenje, ko zunanji signal postane relativno močan. Zatorej lahko porabo, velikost čipa in ceno hkrati zmanjšamo.V prispevku opišemo dvostopenjski LNA ojačevalnik z velikim in nepretrganim ojačanjem. Z uvedbo kontrolne zanke ojačanja v drugo stopnjo stopničastega LNA dosežemo nastavljivo nepretrgano ojačanje v območju 14dB. Tak VGLNA je načrtan v 0.18μm CMOS tehnologiji s ciljno uporabo na frekvenci 2.4GHz. Največje ojačenje in šumno število sta 23dB in 1.12dB. DC poraba je 12.4mW pri napajalni napetosti 1V. Primerjava s podobnimi vezji pokaže prednosti tako načrtanega VGLNA.

1.   Introduction

The ceaseless advance of Complementary Metal-Oxide Semiconductor (CMOS) technology such as the shrink of the gate length and the improvement of the cutoff frequency makes CMOS more and more attractive for many radio frequency (RF) circuits in the several GHz range /1-3/. There is no doubt that a complicated RF system can be realized in CMOS and some chips have been fabricated and reported /4-7/.

The first block of wireless receiver following the antenna is a low noise amplifier (LNA) which plays a significant role as its noise figure (NF) sets a lower bound on the NF of the entire system. Moreover it should be able to accommodate large signals without distortion and must match the input impedance. A good input impedance match is more critical if a pre-select filter precedes the LNA since the transfer characteristics of many filters are sensitive to the quality of the termination. The additional requirement is the low

power consumption which plays an important role in the applications of portable communications systems /8/.

It is not a simple task to design a LNA with high gain and low NF by considering both linearity and input impedance match, with low power consumption. Inductive degeneration is adopted as the LNA architecture popularly because it does not introduce extra noise source and it uses a source degenerative inductor to realize input impedance match. The theoretical analysis and calculation of NF of this topology has been done using an extended MOS noise model /8/. If the gain of LNA block is too low, noise level of this block might dominates the overall NF and if it is too high, the input signal would create nonlinearity. Inversely, if the gain is too low, the mixer noise dominates the overall NF and if it is too high, the input signal to the mixer creates large intermodulation products. For these reasons, the design of the LNA with a variable gain stage or termed as variable gain low noise amplifier (VGLNA) is adopted not only to achieve high gain and low noise but also high linearity. The function of the controllable gain is to prevent sat-

uration in the receiver when the input signal becomes relatively large.

Conventional design techniques for VGLNA can be seen in using bypass switch /9/, which achieves gain variation by on-off control of the bypass switch. However, continuous gain control is difficult with this method. Thus unnecessary power would be consumed at low gain mode. Other gain variation methods reported is the current variation according to target gain /10-12/. These methods reduce its power consumption at low gain mode by controlling its biased current of a LNA. However, this method suffers from small gain variation or discontinuous gain control.

This paper focuses on the CMOS VGLNA designed for 2.4 GHz with a gain-control loop. The VGLNA can operate from 2.1 GHz to 2.5 GHz and can be used for Wideband Code Division Multiple Access (WCDMA) and IEEE 802.11b/g Wireless Local Area Network (WLAN) applications. By introducing a simple gain control loop composed of a gain control transistor and a capacitor, a continuous gain tuning range of approximately 14 dB is achieved.

2.   Variable gain block of a LNA

Fig. 1 shows the fundamental architecture of proposed gain control LNA. The gain control loop is composed of a gain control transistor (M_{gc}) and a capacitor C_f. If the M_{gc} is turned off, the gain control loop is like an open circuit and the VGLNA operates with minimum gain. On the other hand, if M_{gc} is turned on with sufficiently high voltage (like V_{DD}), then M_{gc} operates as a bypass switch and it operates at maximum gain. Therefore, a large gain variation range similar to that of the bypass switch method is achieved.
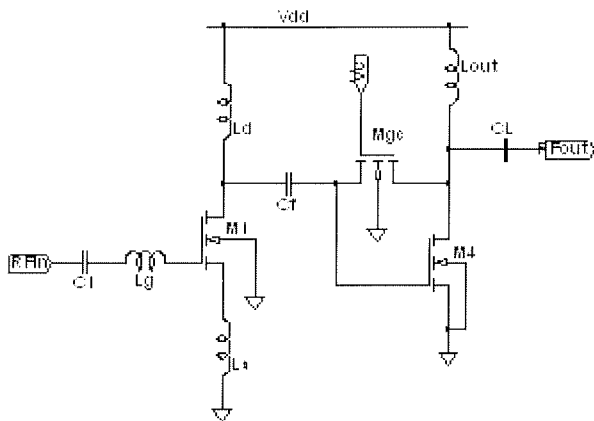

Fig. 1.   Fundamental architecture of proposed variable gain LNA.

The control voltage of the M_{gc} is between threshold voltage and V_{DD} with M_{gc} operating as a variable resistor. This resistance value is equal to the channel resistance R_{gc} which can be expressed as

$$R_{gc} = \frac{1}{\mu_n C_{ox} \frac{W}{L} \left(V_c - V_{gs4} - V_{th-gc} \right)}$$ (1)

where V_c and V_{th-gc} are the control or gate voltage and threshold voltage of M_{gc} respectively. V_{gs4} is the gate-to-source voltage of M4. From (1), R_{gc} varies inversely proportional to the variation of V_{c-}. As V_c increases, the impedance of the gain control loop becomes smaller, thus the inter-stage matching changes. Therefore, the gain of the second stage increases continuously as V_c increases. This is verified in the simulation results in Section 4.

The capacitor C_f is added to support the operation of M_{gc} as a variable resistor. It operates as a DC blocking capacitor, making sure that the V_{DS} of M_{gc} stays small. With the V_{DS} stays small, the M_{gc} operates as a variable resistor.

3.   Design of a Variable Gain LNA

Based on the variable gain block described in Section 2, a 2.4 GHz VGLNA is designed. The VGLNA is composed of a cascode stage and variable gain block. Fig. 2 shows the cascode stage of the designed VGLNA. The cascode stage is made of transistors M1 and M2. On-chip inductors of L_g and L_s are used respectively for matching the input stage of the cascode stage and as a degeneration inductor to suppress the NF of the VGLNA.

Fig. 3 shows the schematic of the variable gain block in a VGLNA. The gain control characteristics of designed variable gain block depends on the size of M_{gc}. Large size of M_{gc} gives large gain variation. However, too large of gain variation would changes too rapidly making continuous control difficult. Moreover, large size of M_{gc} increases parasitic capacitance and may reduce the maximum gain. On the other hand, if the size of the M_{gc} is very small, the gain variation would become small. This result might not satisfy the linearity requirement of the following stage in a receiver such as a mixer. Therefore, the size of the M_{gc} must be selected within the proper gain variation range.
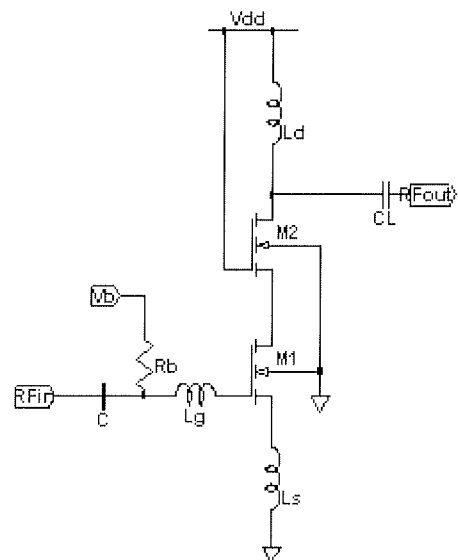

Fig. 2.   The cascode stage of the designed VGLNA.

Informacije MIDEM 37(2007)3, str. 172-175

L. Lee, R. Mohd Sidek, S. Shekhar Jamuar, S. Khatun: Gain Control
Loop for a 2.4 GHz Variable Gain Low Noise Amplifier (VGLNA)
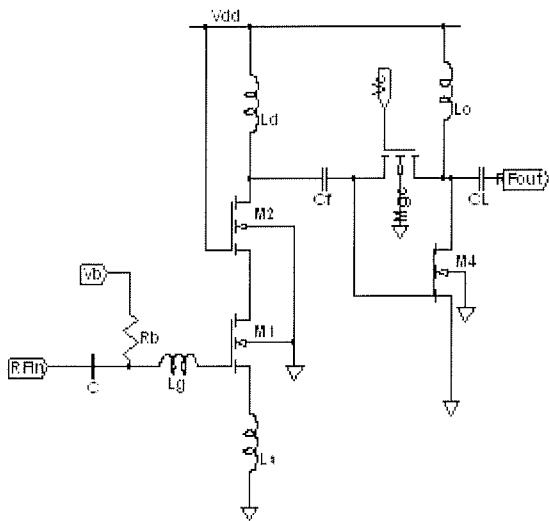
Fig. 3. The variable gain block in the proposed VGLNA.

4. Simulation results

The VGLNA has been designed for WCDMA and 802.11b/
g WLAN applications in 0.18 μm CMOS technology. The
design has been simulated using Agilent's Advanced De-
sign System (ADS).

Fig. 4 and 5 show the S21, S11 and NF of the designed
VGLNA at high gain mode. With the power consumption of
12.4 mW from a supply voltage of 1.0 V, the VGLNA
achieves power gain of 23.29 dB at 2.4 GHz. The gain is
maximized at 2.4 GHz because input matching is optimum
at 2.4 GHz. The input return loss is -20 dB. The NF of the
designed VGLNA is 1.116 dB at 2.4 GHz for the maximum
gain mode.

As for the minimum gain mode, S21 and S11 are reported
to be 9.54 dB and -21 dB as shown in Fig. 6. Fig. 7 shows
the NF of the VGLNA at low gain mode which is slightly
higher with 1.123 dB. This result is acceptable in compar-
ison with other VGLNAs /13-15/. The designed VGLNA
can achieve low NF because the gain of cascode stage
mitigates the influence of NF at the gain control block.
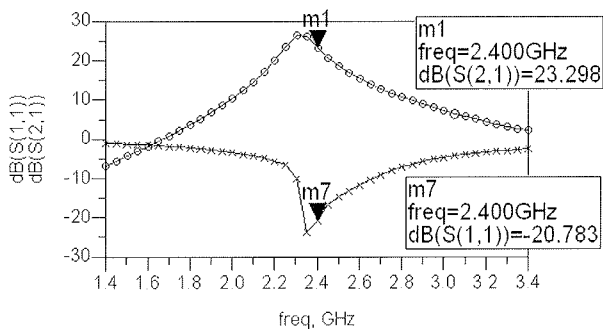
*Fig. 4. S21 and S11 of the designed VGLNA at high
gain mode.*
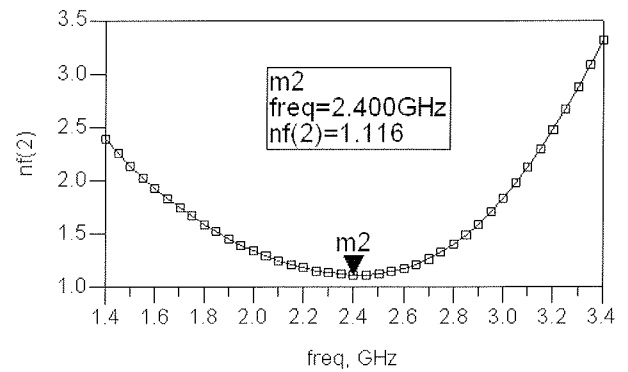
Fig. 5. NF of the proposed VGLNA at high gain mode.
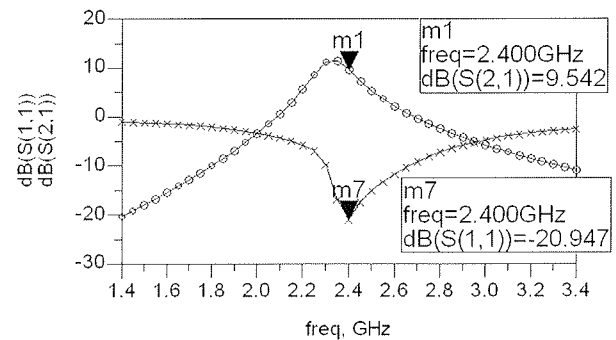
*Fig. 6. S21 and S11 of the designed VGLNA at low gain
mode.*
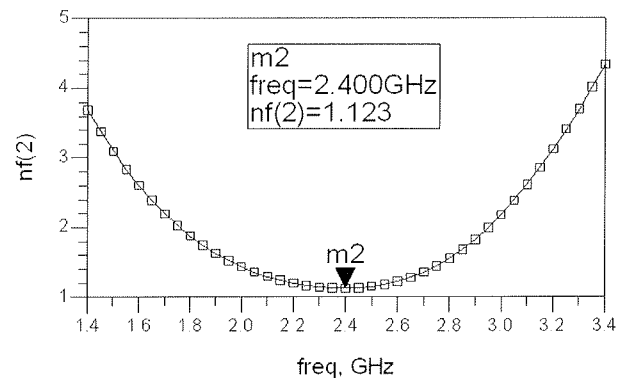
Fig. 7. NF of the proposed VGLNA at low gain mode.

As shown in Fig. 8, the VGLNA achieves continuous gain
tuning range of 13.75dB from 23.29 dB to 9.54 dB at 2.4
GHz. The gain variation is similar to that of the LNA with the
bypass switch method /9/. In addition, the power con-
sumption of the following stages in a receiver such as mix-
er and automatic gain control (AGC) can be further reduced
by achieving the continuous gain control. The NF varies
from 1.116 dB to 1.123 dB at minimum gain mode. NF var-
ies very little in comparison with the gain changes because
the increase of NF is minimized by the gain of the first stage.

As for the linearity, the third order input intercept point (IIP3)
is simulated at low gain mode and it is -2.18 dBm at 2.4
GHz. Thus, the LNA can receive maximum input power of
802.11b WLAN without much distortion.

L. Lee, R. Mohd Sidek, S. Shekhar Jamuar, S. Khatun: Gain Control
Loop for a 2.4 GHz Variable Gain Low Noise Amplifier (VGLNA)

Informacije MIDEM 37(2007)3, str. 172-175

Fig. 8.   Gain and NF variation of the VGLNA.

Table 1 shows the comparison of performances for the designed VGLNA with other variable gain LNAs. Though gain variation range of proposed VGLNA is smaller than other VGLNAs /9-11/ but it has a much better NF and continuous gain control is achieved. Therefore, overall performance of the proposed VGLNA is better in terms of continuous tuning range and NF.

Table 1 Comparison of Variable Gain LNA Performances

	This work (CMOS)	[9]	[10]	[11]	[12]	[13]
Technology (um)	0.18	0.25	-	0.25	0.18	0.18
Operating frequency (GHz)	2.4	2.4	0.9	5.6	5.7	5.2
Gain (dB)	23.29	14.7	26	19.5	16.5	20
Gain variation (dB)	13.75	29	29	19.5	8	20
Continuous gain control	Yes	No	No	No	Yes	Yes
Minimum NF (dB)	1.12	2.88	2.1	3.1	3.5	3.5
Biased current (mA)	12.4	11.7	15	10	3.2	17

5.   Conclusion

In this paper, a VGLNA for WCDMA and 802.11b/g is presented. The designed VGLNA composed of a cascode stage and a variable gain block. The VGLNA achieves low NF of 1.12 dB, gain of 23 dB and power consumption of 12.4 mW from 1.0 V power supply. By adding a simple gain control loop composed of a transistor M_{gc} and a capacitor, the VGLNA achieves a gain tuning range of 13.75 dB with continuous gain control. This is accomplished with a low NF which is reported to be only 1.12 dB. This VGLNA is able to mitigate the dynamic range and linearity requirement of the stages that followed such as mixer and AGC.

Acknowledgment

References

/1/   T. H. Lee, *The Design of CMOS Radio Frequency Integrated Circuit*, Cambridge University Press, Cambridge (1998).
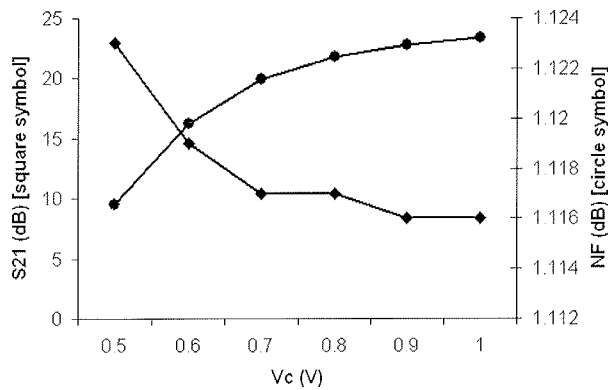
/2/   L. E. Larson, "Integrated circuit technology options for RFIC's – present status and future directions," *IEEE Journal of Solid-State Circuits*, vol. 33 no. 3, March 1998, pp. 387-399.

/3/   E. Abou-Allam, T. Manku, M. Ting and M. S. Obrecht, "Impact of Technology scaling on CMOS RF devices and circuits," *IEEE Custom Integrated Circuits Conference (CICC 2000)*, 2000, pp. 361-364.

/4/   T. Kadoyama, N. Suzuki, N. Sasho, H. Iizuka, I. Nagase, H. Usukubo and M. Katakura, "A complete single-chip GPS receiver with 1.6-V 24-mW Radio in 0.18-um CMOS," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 4, April 2004, pp. 562-568.

/5/   G. Gramegna, M. Franciotta, V. Mandara, N. G. Bellantone, M. Vaiana, M. Paparo, M. Losi, S. Das and P. Mattos, "23mm² single-chip 0.18um CMOS GPS receiver with 28mW-4.1mm² radio and CPU/DSP/Ram/ROM," *IEEE Custom Integrated Circuits Conference (CICC 2004)*, 2004, pp. 81-84.

/6/   R. Ahola, et. al, "A single-chip CMOS transceiver for 802.11a/b/g wireless LANs," *IEEE Journals of Solid-State Circuits*, vol. 39, no. 12, Dec. 2004, pp. 2250-2258.

/7/   L. Leung and et. al, "A 1V Low-power Single-Chip CMOS WLAN IEEE 802.11a Transceiver," *Proc. Solid-State Circuits Conf. 2006 (ESSCIRC 2006)*, Sept. 2006, pp. 283-286.

/8/   D. K. Shaeffer and T. H. Lee, " A 1.5 V, 1.5 GHz CMOS low noise amplifier," *IEEE Journal of Solid-State Circuits*,vol. 35, no. 2, 1999, pp. 745-759.

/9/   R. Point, M. Mendes and W. Foley, "A differential 2.4 GHz switched-gain CMOS LNA for 802.11b and Bluetooth," *Radio and Wireless Conferences 2002 (RAWCON)*, 2002, pp. 221-224.

/10/   S. Pennisi, S. Scaccianose and G. Palmisano, "A new design approach for variable-gain low noise amplifier," *Radio Frequency Integrated Circuits (RFIC) Symposium 2000*, June 2000, pp. 11-13.

/11/   M. Rajashekharaiah, P. Upadhyaya and Heo Deukhyoun, "A compact 5.6 GHz low noise amplifier with new on-chip gain controllable active balun," *IEEE Workshop on Microelectronics and Electron Devices 2004*, 2004, pp. 131-132.

/12/   Y. S. Wang and L. -H. Lu, "5.7 GHz low-power variable-gain LNA in 0.18 um CMOS," *Electronics Letters*, vol. 41, no. 2, Jan. 2005, pp. 66-68.

/13/   M. -D. Tsai, R.-C. Liu, C. -S. Lin and H. Wang, "A low-voltage fully-integrated 4.5-6-GHz CMOS variable gain low noise amplifier," *33ʳᵈ European Microwave Conference 2003*, vol. 1, Oct. 2003, pp. 13-16.

*Lini Lee, Roslina Mohd Sidek, Sudhanshu Shekhar Jamuar and *Sabira Khatun
Department of Electrical and Electronic Engineering,
*Department of Computer System and Communication Engineering, Faculty of Engineering,
University Putra Malaysia (UPM),
43400 UPM Serdang, Selangor, Malaysia.
Corresponding e-mail: linilee@gmail.com*

PROSODY EVALUATION FOR EMBEDDED SLOVENE SPEECH-SYNTHESIS SYSTEMS

France Mihelič*, Boštjan Vesnicer*, Janez Žibert*, and Elmar Nöth[†]

*University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia
[†]Universität Erlangen-Nürnberg, Erlangen, Germany

Abstract: This paper describes an evaluation of the prosody modeling in an HMM-based Slovene speech-synthesis system that is suitable for embedded systems due to its relatively small memory footprint. The objective-evaluation procedure is based on the results of the automatic recognition of syntactic-prosodic boundary positions and accented words in the synthetic speech. We have shown that the recognition results represent a close match with the prosodic notations, labeled by the human expert on the natural-speech counterpart produced by the speaker whose speech was used to train the speech-synthesis system. Therefore, the recognition rate of the prosodic events is proposed as an objective evaluation measure for the quality of the prosodic modeling in the speech-synthesis system. The results of the proposed evaluation method are also in accordance with previous subjective-listening evaluation tests, where high scores for the naturalness for such a type of speech synthesis were observed.

Ocenjevenje prozodije za vgrajene sisteme za sintezo slovenskega govora

Kjučne besede: vgrajeni sistemi, sinteza govora, akustično modeliranje s PMM, modeliranje prozodije, evaluacija sinteze govora, razpoznavanje prozodičnih oznak, metoda podpornih vektorjev, jedro RGB

Izvleček: Članek opisuje vrednotenje modeliranja prozodije v sistemu za sintezo slovenskega govora, ki deluje na osnovi prikritih Markovovih modelov (PMM). Zaradi relativno skromne zasedbe pomnilnika programa za sintezo, je tak način zelo uporaben za realizacijo v vgrajenih sistemih. Objektivna metoda vrednotenja je zasnovana na osnovi samodejnega razpoznavanja mest sintaktično-prozodičnih mej in poudarjenih besed v sintetiziranem govoru. Rezultati razpoznavanja kažejo veliko ujemanje med položaji sintaktično-prozodičnih mej in poudarjenimi besedami, ki jih je ekspert označil na pripadajočem naravnem govoru govorca, katerega govor smo uporabili za učenje sistema za sintezo govora. V tem smislu delež pravilno razpoznanih prozodičnih dogodkov predlagamo za merilo uspešnosti prozodičnega oblikovanja v sistemu za sintezo govora. Rezultati predstavljene evaluacije so tudi skladni s predhodnimi subjektivnimi slušnimi ocenjevanji, ki so dala relativno visoke ocene v smislu naravnosti takega načina sinteze govora.

1. Introduction

Modern speech synthesizers are able to achieve high intelligibility; however, they still suffer from rather unnatural speech. Recently, to increase the naturalness of speech-synthesis systems, there has been a noticeable shift from diphone-based toward corpus-based unit-selection speech synthesis /3/. This has been made possible by the rapid increase in the speed and capacity of computer resources. The main idea in unit-selection speech-synthesis systems is to dynamically select appropriate speech units (e.g., diphones) from a large speech database, and in this way reduce the need for signal-manipulation algorithms (e.g., PSOLA) which significantly degrade the quality of the speech.

In such systems the emphasis is more on engineering techniques (searching, optimization, statistical modeling) than on linguistic-rule development /8/. Many of these algorithms are borrowed from the automatic-speech-recognition (ASR) community. For example, hidden Markov models (HMMs) are widely used for the automatic segmentation and labeling of speech databases (e.g. /7/).

However, to achieve a reasonable performance with such corpus-based systems large computational resources are required, and these are often not available when dealing with embedded systems. Since our objective was to develop a speech-synthesis system that should operate with embedded-systems applications, we focused our research on systems with a relatively small memory footprint.

In accordance with current trends an HMM-based approach to speech synthesis for Slovene was implemented. In contrast to other corpus-based speech-synthesis systems, a system using acoustic models trained on a large speech database using HMM techniques provides a relatively small memory footprint, comparable to – or even smaller than – the footprint of embedded diphone-based systems /5, 17/, and demands no special computational load.

Subjective listening-evaluation tests of the HMM-based system showed a surprisingly high level of prosodic quality for such synthesized speech, although no special proso-

dy modeling was used in the system /14/. In our present research we tried to apply the established methods already used in ASR for prosodic-events recognition to evaluate the prosodic content in the synthesized speech.

The rest of the paper is organized as follows. Section 2 overviews the main idea behind the HMM-based speech-synthesis system. In Section 3 we present the prosodic labeling and databases used for the training and the evaluations. Section 4 gives a short description of the prosodic features and classification procedure for the detection of prosodic events in a speech-synthesis system. The results of the experiments are discussed in Section 5. Finally, concluding remarks and plans for future work are presented in the last section.

2   HMM-based speech synthesis

The selected HMM-based approach for speech synthesis differs from other approaches because it uses the statistical framework of HMMs not only for the segmentation and labeling of the database but also as a model for speech production. The method was originally proposed in /11/ and later extended by Yoshimura et al. /15/.

A schematic representation of building and using an HMM-based speech-synthesis system is shown in Figure 1. The top pane shows a *training step* of such a system, where a statistical model of the speech is estimated (middle part). In the bottom pane a *synthesis step* is shown, where the speech signal is generated.
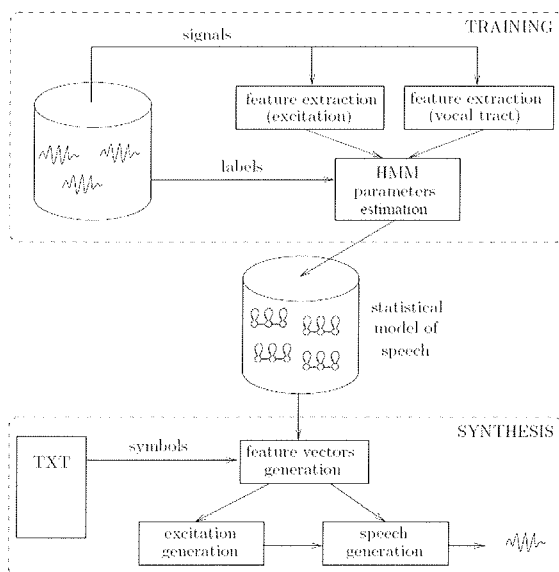

Fig. 1:   Schematic diagram of HMM-based training and synthesis procedures.

For a reliable estimation of the parameters of the statistical model a speech parametrization is required. Since we want to be able to synthesize high-quality speech, the parameters should contain enough information for the reconstruction of speech, which should be perceptually similar to the original. For this purpose, the *source-filter* theory of speech production /9/ is applicable. In order to follow this theory, the parameters of the *vocal tract* transfer function and the *excitation* (f_0) need to be estimated. We used MFCC and log f_0 parameters along with their Δ and $\Delta\Delta$ counterparts.

The procedures for estimating the parameters of HMMs are well known from the field of ASR, and can also be applied in our case. A difference here is that additional parameters of the excitation should be modeled. For this purpose, new types of HMMs are introduced. They are called multi-space-distribution (MSD) HMMs, and are presented in /13/.

A detailed description of the speech parametrization and the HMM structure that we used is given in /14/.

For duration modeling a possible solution is to incorporate *explicit* state duration densities (non-parametric or parametric). However, this increases the storage and computational time significantly, and also increases the need for more training material. To avoid those problems we make a simplification and estimate the duration densities only when the training process is already finished /15/. Evaluation tests using this kind of duration prediction give comparable results /14/ to the previously developed two-stage duration model for speech synthesis using the diphone-concatenation technique /4/.

The memory size of the statistical acoustical and duration models, the pronunciation dictionaries and the object code of the current version of the speech-synthesis system was approximately 1.5 MB.

3   Prosody labeling

We have decided to investigate the possibility of generalizing the described speech-synthesis approach to enable some simple prosody modeling. Therefore, the same speech corpora that were used for the training of the HMM acoustic models used in the speech synthesis were manually annotated with syntactic-prosodic boundaries and word accents in the utterances. These corpora consist of 578 sentences (37 minutes of speech) uttered by the speaker 02m from the Slovenian Weather Forecast Speech Database (VNTV) /7/. Recently, when SiBN speech corpora were recorded and annotated /17/ we acquired additional records of weather forecasts from the same speaker consisting of an additional 253 sentences (17 min of speech). These data were also prosodically annotated in the same manner. The syntactic-prosodic boundaries along the lines of /1/ were annotated for the transliterations of speech utterances, and word accents were labeled via acoustic perceptual sessions. We used the same prosodic annotation as in our previously reported work /6/. Similar annotations for Slovene speech were also used by /10/. Three-class annotation was used in each utterance for the prosody boundaries and the word accents. They are listed in Table 1 and Table 2.

Table 1: Syntactic-prosodic boundary labels

M3 :	Clause boundaries
M2 :	Constituent boundaries likely to be marked prosodically
M0 :	Every other word boundary

Table 2: Word accent labels

PA :	The most prominent (primary) accent within a prosodic clause
SA :	All other accented words are marked as carrying a secondary accent
UA	Unaccented words

An example of the labeled text utterance is given below[1]:

V prihodnjih SA *dneh bo vroče* PA M3. *Ob koncu* SA *ted-na* PA M2 *pa se bo vročini* SA *pridružila še* M2 *sopar-nost* SA M3.

English translation: *In the following days it will be hot. During the weekend the hot weather will be accompanied by sultry conditions.*

4 Prosodic features selection and classification

Classification was performed using prosodic feature sets derived from duration segmental characteristics on the word level, speaking rate, energy and pitch. The duration and energy features were additionally normalized by applying normalization procedures based on statistical parameters that were estimated from the training data /2/ (pp. 38–59).

The energy and pitch features were based on the short-term energy and f_0 contour, respectively. Some of the features that were used to describe a pitch contour are shown in Figure 2. Additionally, we used the mean and the median as features /2/ (pp. 59–62).

We derived the same features set (95 features) that were proposed in the experiments on German speech /2/ (page 103). All 95 features were computed on the Slovene speech databases that were used in our evaluations.

The prosody events were detected by support vector machines (SVMs) /19/. In our previous studies a classification with neural networks was also performed /6/, but we gained a significant improvement in terms of recognition scores when SVMs were applied. The LIBSVM software /18/ was used for the training and recognition tests. We used the RBF kernel with $C = 2^3$ and $\gamma = 2^{-5}$. Note that all the data were linearly scaled into the [-1, +1] range during the pre-processing stages.

---

1   *M0 and UA labels are not indicated in the example.*
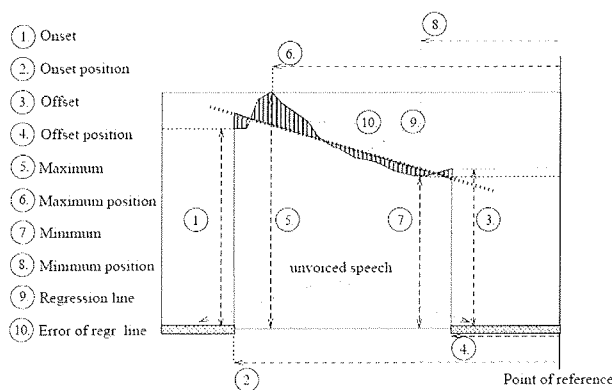
Fig. 2: Example of features used to describe a pitch contour.

5 Experimental results

Our first experiments on the recognition of prosody events were made on the same data set (02m VNTV part) that was used for the training of the HMM-based Slovene speech-synthesis system. We used the data text transcription and the Viterbi forced alignment method to label the speech on the phoneme- and word-duration level. After-wards, 95 dimensional feature vectors were computed for each word of the uttered speech. In total, 6363 vectors with an accompanying 6363 syntactic-prosodic boundary markers and word-accent markers were determined.

5.1 Cross evaluation on natural speech

The first evaluation was performed on natural speech. The cross evaluation was accomplished by dividing the data into 5 training and test subsets. The cross-evaluation results gave an 81% average overall recognition rate and a 75% class-wise recognition rate for the detection of syntactic-prosodic boundaries, and a 69% overall recognition rate and a 66% class-wise recognition rate for word-accent detection. The results proved the consistency of the speech-data labeling procedure and the appropriateness of using SVM for classification process. The cross-evaluation results from this step also served to determine the applicable pair of RGB-kernel parameters reported in Section 4.

5.2 Recognition of the prosodic events in the synthesized speech counterpart

In our next experiment a new database consisting of the synthesized speech generated from text transcriptions of speech-data recordings used for the training of the speech-synthesis system was acquired. In this way we obtained the synthesized-speech database counterpart of the original natural-speech database (02m VNTV part). We were also able to use the same prosodic markers as in the previous experiment. To determine the prosodic feature vectors for synthesized speech this database was also labeled on the phoneme- and word-duration levels. Surprisingly,

the cross-evaluation check on this database gave entirely comparable results to those obtained from the natural-speech counterpart. And even when we were using the natural-speech database for training and its synthesized counterpart for testing, we still got relatively high recognition scores, with an 83.8% overall recognition rate, a 70.7% class-wise recognition rate for prosodic boundaries, and a 77.5% overall recognition rate and a 69.9% class-wise recognition rate for word accents. The confusion matrices for each class are given in Table 3[2] and Table 4.

Table 3: Confusion matrices for syntactic-prosodic boundaries recognition in absolute and relative figures for synthesized speech. The recognition system was trained on the natural-speech counterpart.

actual/predicted	MO	M2	M3	all actual
MO	3662	429	23	4114
M2	266	1034	43	1343
M3	52	125	151	328
all predicted	3980	1588	217	

actual/predicted	MO	M2	M3
MO	89.0%	10.4%	0.6%
M2	19.8%	77.0%	3.2%
M3	5.9%	38.1%	46.0%

Table 4: Confusion matrices for word-accent recognition in absolute and relative figures for synthesized speech. The recognition system was trained on the natural-speech counterpart.

actual/predicted	UA	SA	PA	all actual
UA	3456	417	151	4024
SA	341	955	137	1433
PA	200	188	518	906
all predicted	3997	1560	806	

actual/predicted	UA	SA	PA
UA	85.9%	10.4%	3.7%
SA	23.8%	66.6%	9.6%
PA	22.1%	20.7%	57.2%

Although, as already mentioned in the Introduction section, subjective listening-evaluation tests of the system showed a high level of prosodic quality /14/, we were surprised with these figures. The relatively high recognition results could be due to the method used to acquire the HMM acoustical models for sub-word units of the speech-synthesis system. In our case, triphone units that represented an acoustic model of each phoneme of Slovene with a specific left and right phoneme context were modeled /14/. Since our training database was very domain specific (nearly consecutive season weather forecasts) consisting of only 770 different words, we could expect that the training material for the specific triphone unit was obtained in a large part from uttered word(s) in a specific prosodic context. Based on this assumption, synthesized words in the same context – we have used text transcriptions of training speech material – possess similar basic prosodic attributes in terms of duration, pitch and energy, and could therefore enable an appropriate prosodic impression.

Since in this case we got a very close match of the prosodic features between natural and synthetic speech, the question arises as to what degradation in terms of automatic prosodic-events recognition could we expect if some different text in the same domain of weather forecasts were to be synthesized?

5.3 Recognition of the prosodic events in synthesized speech from the test set

As mentioned in Section 3, we have recently acquired new speech material from the same speaker that was used for the training of our speech-synthesis system. This speech material includes 267 different words (35%) that were not uttered in the training database, and almost all the uttered sentences – except some short sentences expressing opening and closing remarks – were different from those in the training set. This new database was prosodically annotated. On the basis of its text transcription we were again able to produce its synthesized counterpart, which was additionally labeled on the phoneme and word-duration level. Afterwards, prosodic feature vectors for this new synthetic speech data were computed.

In our next recognition experiments the database that consisted of the synthesized speech from the VNTV database was used to train the SVM classifier, and this new data was used for the recognition tests. In this case we still got a relatively close match between the annotation of the syntactic-prosodic boundaries and the recognition results, which is shown in Table 5[3]. The overall recognition rate was 75% and the class-wise recognition rate was 66%.

---

2   *In syntactic-prosodic boundaries recognition we did not count the recognition of the M3 marker at the end of each utterance, since the recognition of this marker is trivial.*

3   *As in Table 3, we did not count the recognition of the M3 marker at the end of each utterance.*

Table 5: Confusion matrices for syntactic-prosodic boundaries recognition in absolute and relative figures for the test part of the synthesized speech. The recognition system was trained on the VNTV synthesized speech.

actual/predicted	M0	M2	M3	all actual
M0	1606	303	87	1996
M2	160	328	50	538
M3	31	36	86	153
all predicted	1797	667	223	

actual/predicted	M0	M2	M3
M0	80.5%	15.2%	4.3%
M2	29.7%	61.0%	9.3%
M3	20.3%	23.5%	56.2%

With the recognition of word accents we encountered strong confusion[4] between the primary accent (PA) and other accented words (SA) in comparison with the human labeler's annotations, and therefore a stronger degradation in the recognition scores, where we got a 62% overall recognition rate and 53% class-wise recognition rate. Nevertheless, the automatic distinction between the accented words and those without accent still remains relatively high (Table 6), showing an adequate prosodic content of the test part of the synthesized speech.

Table 6: Confusion matrices for word-accent recognition in absolute and relative figures for the test part of the synthesized speech. The recognition system was trained on the VNTV synthesized speech.

actual/predicted	UA	PA and SA	all actual
UA	1250	417	1783
PA and SA	221	936	1157
all predicted	1471	1353	

actual/predicted	UA	PA and SA
UA	70. 1%	29.9%
PA and SA	19.1%	80.9%

We also separately explored the recognition results for three subsets of test utterances. The first set of utterances (Test 1) consisted of sentences where the speaker used words that were already used in the training process. The second subset was composed of utterances where at least one word was different from the words used in the training stage (Test 2), and in the third subset there were utterances that differed by two or more words (Test 3).

As expected, the recognition results depicted in Tables 7 and 8 depend on the selection of utterances, showing a decreasing recognition rate when more unseen events (words) are included in the test samples. This is also an indication that such a type of evaluation method is reasonably sensitive to small changes in synthesized speech.

Table 7: Recognition results for prosodic-syntactic boundaries in synthesized speech for three different test subsets.

Test sets	Number of utterances	Recognition overall	Recognition class - wise
Test 1	103	76.0%	66.8%
Test 2	150	74.8%	65.4%
Test 3	86	73.8%	62.7%

Table 8: Recognition results of accented words (PA and SA) in synthesized speech for three different test subsets.

Test sets	Number of utterances	Recognition overall	Recognition class - wise
Test 1	103	77.4%	78.5%
Test 2	150	72.9%	74.0%
Test 3	86	72.3%	73.3%

6   Discussion

On the basis of the presented results, the prosodic quality of domain-constrained synthesized speech produced by an HMM-based speech-synthesis system is relatively high, even though no additional prosodic modeling was included in the system. This could be an important issue when dealing with embedded systems where an expansion of the system model usually leads to an additional demand for a larger memory footprint and more computational power.

A further improvement in the prosody modeling could be achieved if the prosodic markers in the training set were to be used to build different prosody-specific acoustic models of sub-word units. Note, however, that in this case the text-to-speech system should be able to extract prosodic markers from the text that is to be synthesized. Some previous studies indicate /1/ that automatic prosodic annotation based on appropriate statistical n-gram modeling gives reasonably good results and could be used for this task. Our evaluations also showed that better results can be expected for limited-domain speech synthesis, which is also the most common synthesis approach when using embedded devices.

---

4   *This confusion could also partly depend on the in consistency of the human labeler, since there is a 4-year gap between the labeling of the training and test sets.*

After analyzing the recognition results, the automatic recognition of prosodic events was also shown to be an effective tool for the objective evaluation of the performance of speech-synthesis systems on the prosody-modeling level and could therefore also be used for side-by-side comparisons of the different systems. The application of such a type of evaluation in comparison with subjective evaluation tests – where a set of representative listeners should be gathered for each evaluation test, questionnaires prepared, filled in and analyzed – is much less time consuming and could be easily reproduced.

7 Conclusion

We have described experimental results from an automatic recognition of prosodic events in the synthesized speech produced by an HMM-based speech-synthesis system. The results indicate that such kinds of tests could be used as an objective measure for the evaluation of prosody modeling in speech-synthesis systems. They also confirm a relatively good impression of naturalness with HMM-based speech synthesis, which was also noticed during previously performed subjective listening tests. The results of this study also suggest that it could be worthwhile to make use of the presented (or similar) prosodic annotations of training speech corpora for constructing prosody-specific sub-word acoustic models. However, the effectiveness of such an approach is still to be confirmed empirically, which we plan to do in our future experiments.

8 Acknowledgements

9 References

/1/ Batliner A., Kompe R., Kiessling A., Mast M., Niemann H., Nöth E., "M = Syntax + Prosody: A syntactic-prosodic labelling scheme for large spontaneous speech databases", Speech Communication, 25, 1998, pp. 193-222.

/2/ Buckow J., "Multilingual Prosody in Automatic Speech Understanding", Logos Verlag Berlin, 2004.

/3/ Campbell N., Black A., "Prosody and the Selection of Source Units for Concatenative Synthesis", J. van Santen, R. Sproat, J. Olive and J. Hirschberg (Eds.), in Progress in Speech Synthesis, pp. 279-282, Springer Verlag, 1996.

/4/ Gros J., "A two-level duration model for the Slovenian speech", Electrotechnical Review, vol. 66, no. 2, 1999, pp. 92-97.

/5/ Mihelič A., Žganec Gros J., Pavešić N., Žganec M., "Efficient subset selection from phonetically transcribed text corpora for concatenation-based embedded text-to-speech synthesis", Informacije MIDEM 36, nr. 1, 2006, pp. 19-24.

/6/ Mihelič F., Gros J., Nöth E., Warnke V., "Recognition and Labeling of Prosodic Events in Slovenian Speech", Lecture Notes in Artificial Intelligence 1902, Springer, 2000, pp. 165-170.

/7/ Mihelič F., Gros J., Dobrišek S., Žibert J., Pavešić N., "Spoken Language Resources at LUKS of the University of Ljubljana",

International Journal of Speech Technology, vol. 6, 2003, pp. 221-232.

/8/ Ostendorf M., Bulyko I., "The Impact of Speech Recognition on Speech Synthesis", Proc. of the IEEE Workshop on Speech Synthesis, 2002.

/9/ Rabiner L., Huang B.-H., "Fundamentals of Speech Recognition", Prentice Hall, Englewood Cliffs, NJ, USA, 1993.

/10/ Stergar J., Horvat B., "Prediction of Symbolic Prosody Breaks with Neural Nets", Informacije MIDEM 32, nr. 3, 2002, pp. 213-218.

/11/ Tokuda K., Kobayashi T., Imai S., "Speech parameter generation from HMM using dynamic features", Proc. of ICASSP, vol. 1, 1995, pp. 660-663.

/12/ Tokuda K., Yoshimura T., Masuko T., Kobayashi T., Kitamura T., "Speech Parameter Generation Algorithms for HMM-based Speech Synthesis", Proc. ICASSP, vol. 3, 2000, pp. 1315-1318.

/13/ Tokuda K., Masuko T., Miyazaki N., Kobayashi T., "Multi-Space Probability Distribution HMM", IEICE Transactions on Information and Systems, vol. E85-D, no. 3, 2002, pp. 455-464.

/14/ Vesnicer B., Mihelič F., "Evaluation of Slovenian HMM-Based Speech Synthesis System", Lecture Notes in Artificial Intelignece 3206, Springer 2004, pp. 513-520.

/15/ Yoshimura T., Tokuda K., Masuko T., Kobayashi T., Kitamura T., "Duration Modeling for HMM-based Speech Synthesis", Proc. ICSLP, vol. 2, 1998, pp. 29-32.

/16/ Zemljak M., Kačič Z., Dobrišek S., Gros J., Weiss P., "Computer-based Symbols for Slovene Speech", Journal for Linguistics and Literary Studies, vol. 2, 2002, pp. 159-294.

/17/ Žganec Gros J., Žganec M., "An Efficient Unit-Selection Method for Embedded Concatenative Speech Synthesis", Informacije MIDEM, vol. 37, 2007, no. 3, pp 158 - 164.

/18/ Žibert J., Mihelič F., "Development of Slovenian broadcast news speech database", Proceedings of Fourth International Conference on Language Resources and Evaluation, Lisbon, Portugal, 2004, pp. 2095-2098.

10 Web References

/19/ Chang C.-C., Lin C.-J., "LIBSVM: a library for support vector machines", http://www.csie.ntu.edu.tw/$\sim$cjlin/libsvm .

/20/ Hsu C.-W., Chang C.-C., Lin C.-J., "A Practical Guide to Support Vector Classification", http://www.csie.ntu.edu.tw/$\sim$cjlin/papers/guide .

prof. dr. France Mihelič,
mag. Boštjan Vesnicer, dr. Janez Žibert
Faculty of Electrical Engineering,
University of Ljubljana
Tržaška 25, 1000 Ljubljana, Slovenija
france.mihelic@fe.uni-lj.si
tel +386 1 4768 841; fax +386 1 4768 316

prof. dr. Elmar Nöth
IMMD5, Universität Erlangen-Nürnberg
Martensstrsse 3, 91058 Erlangen, Germany
noeth@informatik.uni-erlangen.de
tel +49 9131 85 27888; fax +49 9131 303811

OPTIMIZACIJA NAVORNE KARAKTERISTIKE ELEKTRONSKO KOMUTIRANEGA MOTORJA V HIBRIDNEM POGONU

Boštjan Pevec[1], Primož Bajec[2], Janez Nastran[1], Danijel Vončina[1]

[1]Univerza v Ljubljani, Fakulteta za elektrotehniko, Ljubljana, Slovenija
[2]Hidria Rotomatika d.o.o., Spodnja Idrija, Slovenija

Kjučne besede: hibridno vozilo, elektronsko komutirani stroj, integrirani zaganjalnik-generator, navorna karakteristika, močnostni stikalni pretvornik, krmiljenje EK motorja

Izvleček: V članku je predstavljena problematika zagona hibridnega pogonskega agregata, sestavljenega iz enosmernega elektronsko komutiranega (EK) stroja s trajnimi magneti (NdFeB) in štiritaktnega motorja z notranjim zgorevanjem (NZ). Pogonski agregat je namenjen pogonu cestnega motocikla. Zahteve po visokih zagonskih navorih motorjev z NZ, omejitve fizičnih dimenzij EK stroja in širok razpon vrtilne hitrosti narekujejo posebno pazljivost pri načrtovanju električnega stroja in v nadaljevanju tudi analizo krmilnih algoritmov EK stroja, predvsem v območju nizkih vrtilnih hitrosti. Obravnavana sta dva pristopa k optimizaciji navorne karakteristike EK stroja in sicer slabljenje glavnega magnetnega pretoka v stroju in nastavljanje kota prevajanja tranzistorjev močnostnega stikalnega pretvornika. Predlagan je nov princip krmiljenja obravnavanega tipa motorja, ki vključuje oba navedena pristopa. Predstavljene so primerjalne meritve časovnih potekov vrtilnih hitrosti zagona motorja z NZ, med predlaganim in običajnim načinom krmiljenja EK motorja v hibridnem pogonskem agregatu. Eksperimentalni rezultati potrjujejo izboljšave, ki jih doprinese predlagano krmiljenje EK stroja v motorskem režimu obratovanja.

Torque Characteristic Optimization of Brushless DC Motor in the Hybrid Vehicle

Key words: hybrid vehicle, brushless DC motor, integrated starter-generator, torque characteristic, switched mode power converter, BLDC motor control

Abstract: The paper outlines a case study on a brushless direct current (BLDC) motor for a hybrid propulsion system, with a particular focus on the motor torque characteristic. The discussed hybrid propulsion system is in its origin intended to drive the motorcycle. The subject of our research is the Integrated Starter–Generator (ISG) coupled to the internal combustion (IC) engine. Requirements for a high power density and an acceptable construction of the electrical machine have led to the permanent–magnet BLDC machine with surface mounted rare–earth magnets on the external rotor. Very wide rotational speed range of modern IC engines (up to 15 000 rpm), dictates particular properties and capabilities of the switch–mode power converter (SPC) (Fig. 1) in order to support all operating modes of the ISG and to enable a bi–directional energy flow. High efficiency is of course an imperative. During the design of the converter, our attention was paid also to efficient operation of the ISG at the operation range limits such as providing an efficient and fast starting maneuver of the IC engine and extending the low–speed generator operation range /7/.

Stringent starting torque demands, electrical machine geometry limitations and a wide rotational speed range of the BLDC are reasons for control algorithm analysis in the low speed operation mode to improve electric motor torque characteristic. Two approaches for the optimization of torque characteristic are discussed in the paper: magnetic flux weakening method and a modification of transistor conduction angle.

The flux-weakening method is a well known principle to extend the rotational speed range of the BLDC described by many authors /1-3/, which is achieved with a phase current leading technique. The decreasing of the main magnetic flux is influencing on decreasing of the back EMF in the stator winding which consequently increases the maximal possible stator current. On the other hand with this approach the main magnetic flux is decreased which directly decreases the electromagnetic torque (Fig. 4). That is why the influence of the phase current leading is studied in detail to determine the optimal angle of the phase current at rotational speed range directed by the starting manoeuvre of the IC engine.

To achieve optimal motor regime of the BLDC, modification of the phase current conduction is studied in detail. Two most common principles of BLDC control are a $2\pi/3$ angle switch-on mode (Fig. 5) and a π angle switch-on mode /1, 2/ (Fig. 6). The first control principle has a higher efficiency, whereas with the second one we can produce a higher maximal electromagnetic torque. Conduction angles of switches between both control principles are studied in detail to determine the optimal angle for switches of the AC/DC converter. Consequently that means that during one electrical turn the transistor conduction angle is between $2\pi/3$ and π (Fig. 7).

A novel control principle is proposed which includes both of the above discussed approaches and is based on laboratory measurements of the BLDC motor (Fig. 7-10). A torque characteristic comparison between a commonly used control method and the proposed control method is presented (Fig. 11). Experimental results verify improvements on the starting procedure of the hybrid vehicle made by the proposed control algorithm of the BLDC (Fig. 12).

1. Uvod

Prve ideje o vozilih na električni pogon segajo daleč v leto 1830. Kot rezultat nenehnega razvoja motorjev z notranjim zgorevanjem (NZ), električnih strojev, akumulatorjev, sklopov močnostne elektronike, mikroprocesorske tehnike in nenazadnje čedalje večje okoljske zavesti ter s tem pove-

zanimi ukrepi za omejevanje onesnaževanja okolja se je kot najprimernejša izkazala združitev več različnih tipov pogonskih strojev v skupni pogonski agregat. Možnih je več kombinacij, med katerimi je tudi kombinacija motorja z NZ in električnega stroja, napajanega iz obnovljivega vira električne energije. Ena izmed možnih izvedb hibridnega pogona je tudi integrirani zaganjalnik-generator (IZG) na

skupni mehanski gredi z motorjem z NZ. Pri izbiri električnega stroja za IZG so bile ključne zahteve po obratovalnih karakteristikah stroja v obeh možnih režimih delovanja, in sicer v motorskem in v generatorskem režimu /7/. Po rezultatih študije različnih tipov električnih strojev se je kot najprimernejši izkazal enosmerni elektronsko komutirani (EK) stroj (*angl.* BLDC) s trajnimi magneti iz redkih zemelj na zunanjem rotorju, ki ob pravilni konstrukciji in izbiri materialov nudi sprejemljivo specifično mehansko oz. električno moč. Zaradi velikega razpona vrtilne hitrosti sodobnih motorjev z NZ za cestne motocikle je za načrtovalce velik izziv konstrukcija električnega stroja, ki bi dosegala optimalne obratovalne karakteristike tako v generatorskem kot tudi v motorskem režimu delovanja. Omenjena zahteva in tudi sam princip delovanja EK stroja narekujeta uporabo namenskega močnostnega pretvornika za napajanje električnega stroja, ki mora optimalno delovati v vseh obratovalnih režimih IZG-ja in omogočati dvosmerni pretok električne energije (Slika 1).
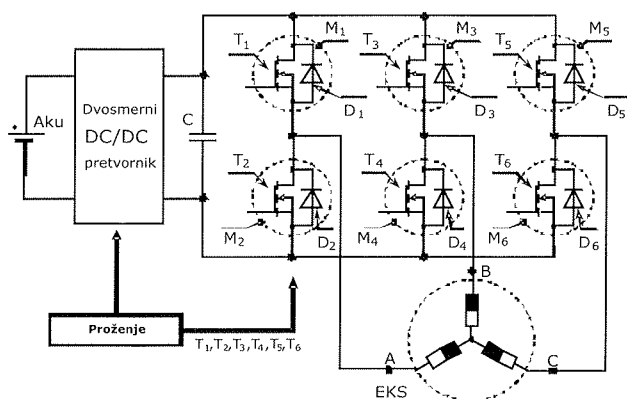
Slika 1:   Močnostni stikalni pretvornik za vodenje IZG.
Fig. 1:   The switch-mode power converter topology.

Močnostni stikalni pretvornik je sestavljen iz dvosmernega DC/DC pretvornika, ki skrbi za nastavljanje ustreznih napetostnih nivojev v odvisnosti od režima delovanja EK stroja in trifaznega pretvornika, ki omogoča delovanje stroja v motorskem in v generatorskem režimu. Pri zagonu motorja z NZ je zaradi zmanjšanja parazitnega padca napetosti dvosmerni DC/DC pretvornik premoščen.

Med načrtovanjem celotnega hibridnega pogona, predvsem pa močnostnega pretvornika /8/, je bila posvečena posebna pozornost obratovalnim karakteristikam električnega stroja pri mejnih vrednostih vrtilne hitrosti, kot na primer optimizacija procesa zagona motorja z NZ in obratovanje IZG v generatorskem režimu vse do najvišjih vrtilnih hitrosti pogonskega agregata /7/.

Zahteve po visokih zagonskih navorih motorjev z NZ (zagon pri -10°C s polovično napolnjenostjo akumulatorja), omejitve pri geometriji in fizičnih dimenzijah EK stroja ter širok razpon vrtilne hitrosti narekujejo skrbno analizo navorne karakteristike EK stroja. Analiza motorskega obratovanja IZG je potekala predvsem v smeri dviga navorne karakteristike v obsegu vrtilnih hitrosti, ki jih narekuje za-

gon motorja z NZ, kar pripomore k hitrejšemu, zanesljivejšemu in energijsko manj potratnemu zagonu. V članku sta analizirana dva pristopa k doseganju predhodno zastavljenega cilja in sicer slabljenje glavnega magnetnega pretoka in nastavljanje kota prevajanja posameznih tranzistorjev trifaznega pretvornika.

V literaturi je pogosto navedena metoda slabljenja polja za razširitev območja delovanja EK motorja /1-3/, ki je dosežena s tehniko faznega prehitevanja statorskega magnetnega pretoka. Zmanjšanje glavnega magnetnega pretoka vpliva na znižanje inducirane protinapetosti v statorskem navitju, kar posledično povečuje maksimalno možno vrednost statorskega toka. Po drugi strani pa s tem zmanjšujemo delovno komponento magnetnega pretoka, ki vpliva na generiranje elektromagnetnega navora. Zato je v nadaljevanju podrobno raziskan vpliv faznega prehitevanja statorskega magnetnega pretoka na elektromagnetni navor enosmernega EK motorja med zagonom motorja z NZ.

Drugi ukrep za doseganje optimalne navorne karakteristike EK motorja je nastavljanje kota prevajanja tranzistorjev trifaznega pretvornika. V praksi sta najpogosteje uporabljena dva principa proženja tranzistorjev in s tem dva načina krmiljenja enosmernega EK motorja /1, 2/: tako imenovano 120° proženje in 180° proženje tranzistorjev. Prednost 120° proženja tranzistorjev je nekoliko višji izkoristek, prednost 180° proženja pa večji maksimalni navor, ki ga motor lahko razvije. Podrobno je raziskano področje, ki zajema kote proženja tranzistorjev med obema načinoma proženja z namenom, ugotoviti optimalen kot prevajanja posameznih tranzistorjev. To pomeni, da bo dejanski kot prevajanja posameznega tranzistorja znotraj električnega obrata pri predlaganem načinu krmiljenja med 120° in 180°, odvisno od trenutne vrtilne hitrosti EK motorja. Povečanje kota prevajanja posameznega tranzistorja ima za posledico povečanje elektromagnetnega navora in zmanjšanje izkoristka stroja.

2. Teoretično ozadje predlaganega pristopa

A) Slabljenje glavnega magnetnega pretoka

Razširitev področja obratovalnih vrtilnih hitrosti električnih strojev z metodo slabljenja glavnega magnetnega pretoka v stroju je pogosto uporabljen ukrep v elektromotorskih pogonih /1-3/. Iz fizikalnih zakonitostih EK motorja izhaja, da je ključnega pomena odvisnost med amplitudo inducirane protinapetosti v statorskem navitju motorja in vrtilno hitrostjo rotorja. Pozitivna napetostna razlika med enosmerno napajalno napetostjo in vsoto induciranih napetosti v dveh faznih navitjih (v primeru, da so navitja zvezana v zvezdo) pogojuje velikost električnega toka v EK motorju in s tem velikost razvitega navora. Z višanjem vrtilne hitrosti se napetostna razlika zmanjšuje, s čimer posledično upada električni tok in tudi navor motorja.

Informacije MIDEM 37(2007)3, str. 182-188

B. Pevec, P. Bajec, J. Nastran, D. Vončina: Optimizacija navorne
karakteristike elektronsko komutiranega motorja v hibridnem pogonu

Princip delovanja enosmernega EK motorja v področju slabljenja polja oziroma v področju konstantne moči je zelo podoben vektorsko reguliranemu pogonu z asinhronskim motorjem ali pogonu s tuje vzbujanim enosmernim motorjem. Pri vektorski regulaciji asinhronskega motorja je z *d*-komponento statorskega toka (v smeri *d*-osi pri modelu motorja v *dq* koordinatnem sistemu) neposredno določeno celotno magnetenje (vzbujanje) stroja, medtem ko pri enosmernem tuje vzbujanem motorju in tudi EK motorju z dodatno komponento statorskega toka (vzdolžna komponenta i_d) le slabimo izhodiščni magnetni pretok v motorju, ki je lahko rezultat tujega vzbujanja ali magnetnega pretoka zaradi vgrajenih trajnih magnetov.

Za natančno analizo delovanja EK motorja v področju slabljenja polja je treba zapisati model EK motorja v *dq* koordinatnem sistemu, vezanem na rotor motorja in sicer tako, da je *d* os usmerjena skladno z rezultančnim vektorjem magnetnega pretoka rotorja $\vec{\phi}_{rot}$. Pri EK motorju s sinusno obliko induciranih napetosti in pri sinusnem vzbujanju sta i_d in i_q konstantni vrednosti v rotorskem *dq* koordinatnem sistemu. Pri EK motorju s trapezno obliko induciranih faznih napetosti je modeliranje motorja in zapis v rotorskem *dq* koordinatnem sistemu sorazmerno zahtevno. Trapezno obliko induciranih napetosti v statorskem navitju lahko zapišemo z vsoto prispevkov osnovne in dodatnih višjeharmonskih komponent statorske napetosti /6/. Pri simetrični trapezni obliki inducirane napetosti s srednjo vrednostjo nič nastopajo v Fourierovi vrsti le lihi višji harmoniki. V končni obliki je model EK motorja zapisan v mešanem *dq* in $\alpha\beta$ koordinatnem sistemu /6/.

Delovanje EK motorja s trapezno obliko inducirane napetosti v področju slabljenja polja lahko opišemo na osnovi modela motorja s sinusno obliko inducirane napetosti v rotorskem *dq* koordinatnem sistemu. *dq* koordinatni sistem je definiran tako, da vzdolžna komponenta toka i_d neposredno vpliva na vzbujanje motorja in sovpada z rezultančno komponento rotorskega magnetnega pretoka $\vec{\phi}_{rot}$, prečna komponenta i_q pa vpliva na proizveden navor (Slika 2).
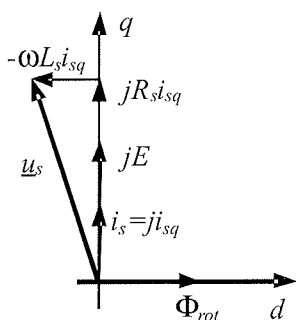
Slika 2: Kazalčni diagram EK motorja.
Fig. 2: Brushless AC motor phasor diagram.

V optimalnem režimu delovanja (maksimalen navor) EK motorja je vzdolžna komponenta toka $i_d = 0$. Režim delovanja je možen do vrtilne hitrosti, pri kateri je pri dani napajalni napetosti regulacija želenega statorskega toka še izve-

dljiva. Obratovanje EK motorja v področju slabljenja glavnega magnetnega polja je razvidno iz slike 3.
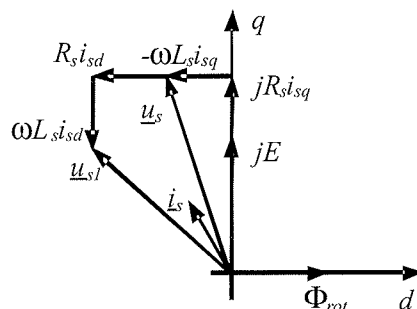
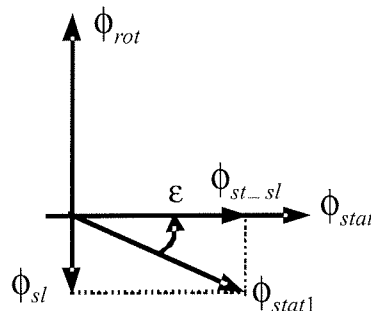Slika 3: Kazalčni diagram EK motorja v področju slabljenja polja.
Fig. 3: Phasor diagram of the Brushless AC motor in the flux-weakening area.

Statorski tok EK motorja lahko razstavimo na komponenti v smeri *d* in *q* osi, pri čemer komponenta toka i_q vpliva na proizveden elektromagnetni navor, negativna komponenta i_d pa vpliva na rezultančni magnetni pretok rotorja. Pri tuje vzbujanem enosmernem motorju ima podobno vlogo reakcija indukta. Posledica dodatne komponente toka i_d in dopustne maksimalne vrednosti statorskega toka, določene z

$$I_s = \sqrt{i^2_{sq} + i^2_{sd}},$$ (1)

je zmanjšanje komponente toka i_q, kar se odraža v zmanjšanju navora EK motorja.

Poenostavljen diagram EK motorja s trapezno obliko induciranih napetosti, ki zajema razmere s slik 2 in 3, kaže slika 4. Pri obratovanju električnega motorja brez slabljenja polja je kot med vektorjema magnetnih pretokov rotorja $\vec{\phi}_{rot}$ in statorja $\vec{\phi}_{stat}$ stroja enak 90° (največji elektromagnetni navor). Pri obratovanju s slabljenjem polja vektor magnetnega pretoka $\vec{\phi}_{stat1}$ prehiteva vektor rotorskega pretoka za več kot 90°, kar je na sliki označeno s kotom ε.

Slika 4: Kazalčni diagram EK motorja s trapezno obliko induciranih napetosti.
Fig. 4: Phasor diagram of the BLDC motor in the flux-weakening area.

Vektor $\vec{\phi}_{stat1}$ razstavimo na komponento $\vec{\phi}_{sl}$, ki slabi magnetni pretok rotorja in na komponento $\vec{\phi}_{st\_sl}$, ki ustvarja ele-

ktromagnetni navor. Vidimo, da je slednji nekoliko manjši od $\bar{\phi}_{stat}$, kar bi pri nižjih vrtilnih hitrostih pomenilo tudi nižji navor. Pri višjih vrtilnih hitrostih pa zaradi posledično nižje inducirane napetosti teče v motor večji tok.

B) Nastavljanje kota prevajanja tranzistorjev

Osnovna dva principa krmiljenja enosmernega EK motorja sta 120° prevajanje tranzistorjev in 180° prevajanje tranzistorjev v enem električnem obratu. Pri prvem načinu krmiljenja sočasno prevajata po dva tranzistorja in električni tok teče skozi dve fazni navitji. V prikazanem primeru (Slika 5) teče tok v navitje faze A in iz navitja faze B. V naslednjem koraku (vrtenje v smeri urinega kazalca) bi tekel tok v navitje faze A in iz navitja faze C. Komutacija toka iz enega faznega navitja v drugega se vrši vsakih 60° (električnih) in sicer 30° pred največjo vrednostjo maksimalnega navora. Na tak način dosežemo največji povprečni navor in najmanjšo valovitost navora (angl. Torque Ripple).
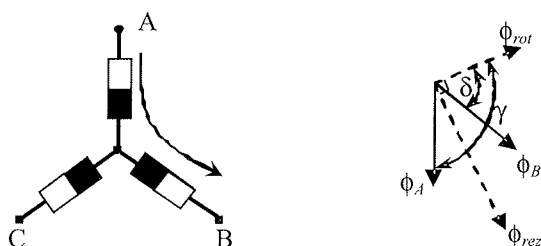
Slika 5: 120° prevajanje tranzistorjev.
Fig. 5: 2π/3 angle switch-on mode.

EK motor razvija maksimalen navor pri zavrtem rotorju (ni inducirane protinapetosti). Takrat lahko zapišemo skupni tok iz vira kot:

$$I_{vh} = \frac{U}{R_F \cdot 2},$$ (2)

kjer je U napajalna napetost, R_F pa je upornost faznega navitja.

Če označimo količnik $U/R_F = I$, lahko zapišemo:

$$I_{vh} = \frac{I}{2},$$ (3)

Elektromagnetni navor motorja je zapisan kot:

$$M = k \cdot \vec{\phi}_{rot} \times \vec{I}_{mot},$$ (4)

kjer je ϕ_{rot} magnetni pretok rotorja, v k pa združimo vse konstrukcijske konstante motorja. Če želimo izračunati elektromagnetni navor, moramo upoštevati vse komponente statorskega toka (slika 5):

$$M = k \cdot \phi_{rot} \cdot (I_A \sin\gamma + I_B \sin\delta + I_C \sin\lambda),$$ (5)

kjer je γ kot med vektorjema magnetnega pretoka rotorja in magnetnega pretoka faze A (120°), δ kot med vektorjema magnetnega pretoka rotorja ter magnetnega pretoka faze B (60°) in λ kot med vektorjema magnetnega pretoka rotorja ter magnetnega pretoka faze C.

Tako dobimo največji navor motorja pri 120° prevajanju tranzistorjev:

$$M_{max} = k \cdot \phi_{rot} \cdot \left(\frac{I}{2}\sin 120° + \frac{I}{2}\sin 60°\right) = k \cdot \phi_{rot} \cdot I \cdot \frac{\sqrt{3}}{2}.$$ (6)

Če zavrtimo rotor za 30° dobimo najnižjo vrednost navora, ki znaša:

$$M_{min} = k \cdot \phi_{rot} \cdot \left(\frac{I}{2}\sin 90° + \frac{I}{2}\sin 30°\right) = k \cdot \phi_{rot} \cdot I \cdot \frac{3}{4}.$$ (7)

Razmerje med navorom, ki ga razvija motor in tokom, ki teče iz vira, je pokazatelj energijskega izkoristka motorja; večje kot je razmerje višji je energijski izkoristek. Tako znaša navedeno razmerje pri zavrtem rotorju in pri 120° prevajanju tranzistorjev:

$$\frac{M_{max}}{k \cdot \phi_{rot} \cdot I_{vh}} = \sqrt{3} \quad ; \quad \frac{M_{min}}{k \cdot \phi_{rot} \cdot I_{vh}} = \frac{3}{2}.$$ (8)

Pri 180° prevajanju tranzistorjev so hkrati odprti trije tranzistorji, tok pa vedno teče v vseh treh faznih navitjih EK motorja. Na sliki 6 je prikazan primer, ko tok teče v navitje faze A in iz navitij faz B in C. Na desnem delu slike je kazalčni diagram magnetnih pretokov v trenutku, ko motor razvija največji navor.
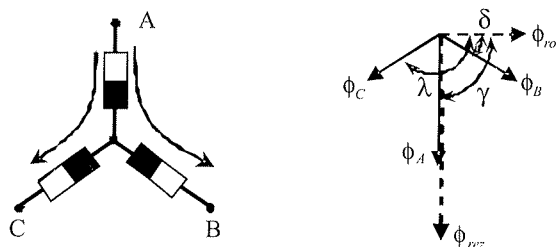
Slika 6: 180° prevajanje tranzistorjev.
Fig. 6: π angle switch-on mode.

Podobno kot pri 120° prevajanju stikal moramo pri izračunu največjega in najmanjšega navora pri zavrtem rotorju upoštevati vse tri komponente faznih tokov. Skupni tok iz vira znaša:

$$I_{vh} = \frac{U}{R_F \frac{3}{2}} = \frac{2}{3} \cdot I.$$ (9)

Če upoštevamo razmere s slike 6, lahko s pomočjo enačbe 5 zapišemo največji navor motorja pri 180° prevajanju tranzistorjev:

$$M_{max} = k \cdot \phi_{rot} \cdot \left(I\frac{2}{3}\sin 90° + I\frac{1}{3}\sin 30° + I\frac{1}{3}\sin 150°\right) = k \cdot \phi_{rot} \cdot I.$$ (10)

Najnižjo vrednost navora dobimo, po istem postopku kot pri 120° prevajanju stikal, če zavrtimo rotor za 30°:

$$M_{min} = k \cdot \phi_{rot} \cdot \left(I\frac{2}{3}\sin 60° + I\frac{1}{3}\sin 0° + I\frac{1}{3}\sin 120°\right) = k \cdot \phi_{rot} \cdot I \cdot \frac{\sqrt{3}}{2}.$$ (11)

Razmerje med največjim oziroma najmanjšim navorom in tokom iz vira pri zavrtem rotorju in pri 180° kotu prevajanja stikal znaša:

In nacije MIDEM 37(2007)3, str. 182-188

B. Pevec, P. Bajec, J. Nastran, D. Vončina: Optimizacija navorne
karakteristike elektronsko komutiranega motorja v hibridnem pogonu

$$\frac{M_{max}}{k \cdot \phi_{rot} \cdot I_{vh}} = \frac{3}{2} \quad ; \quad \frac{M_{min}}{k \cdot \phi_{rot} \cdot I_{vh}} = \frac{3 \cdot \sqrt{3}}{4} = 1,3. \quad (12)$$

3. Predlagan pristop optimizacije navorne karakteristike

V prejšnjem poglavju je podrobno obdelano teoretično ozadje predlaganih pristopov za optimizacijo navorne karakteristike EK motorja, iz katerih je moč predvideti njihovo odvisnost od vrtilne hitrosti motorja in od kota prevajanja tranzistorjev trifaznega pretvornika. Slabljenje glavnega magnetnega pretoka enosmernega EK motorja znižuje inducirano protinapetost in s tem povečuje največji možni tok faznega navitja. Po drugi strani pa se z nižanjem vrednosti glavnega magnetnega pretoka manjša tudi elektromagnetni navor. Zato je treba za optimalno delovanje EK motorja poiskati ustrezen nivo slabljenja glavnega magnetnega pretoka v celotnem poteku zagona motorja z NZ. Slabljenje glavnega magnetnega pretoka pri enosmernem EK motorju dosežemo s tehniko faznega prehitevanja statorskega magnetnega pretoka (Slika 4). To pomeni, da tranzistorje prožimo nekoliko prej od preklopa dajalnika pozicije (običajno trije Hall-ovi senzorji), ki je nameščen tako, da motor razvije maksimalen navor 30° po spremembi njegovega logičnega stanja. Iz enačb 6 in 10 je vidna razlika v največjem navoru, ki ga enosmerni EK motor lahko razvija v dveh različnih režimih proženja tranzistorjev močnostnega stikalnega pretvornika. Iz enačb 8 in 12 pa je razvidno znižanje izkoristka motorja pri 180° prevajanju tranzistorjev. Iz navedenih primerjav je moč sklepati, da optimalno delovanje EK motorja dosežemo s kotom prevajanja tranzistorjev trifaznega pretvornika med 120° in 180°. Zato je treba poiskati optimalne delovne točke v celotnem obravnavanem področju vrtilnih hitrosti, kjer je razmerje navor/izkoristek največje.

Optimalno delovanje EK motorja dosežemo z upoštevanjem obeh obravnavanih principov krmiljenja, kar pomeni, da se preklopi tranzistorjev ne vršijo več ob spremembah logičnega stanja Hall-ovih signalov $H_{1,2,3}$ (Slika 7), ampak so pogojeni s kotom predproženja tranzistorjev trifaznega pretvornika (β) (slabljenje glavnega magnetnega pretoka) in kotom podaljšanja prevajanja tranzistorjev (α) pri določeni vrtilni hitrosti enosmernega EK motorja glede na 120° način krmiljenja.
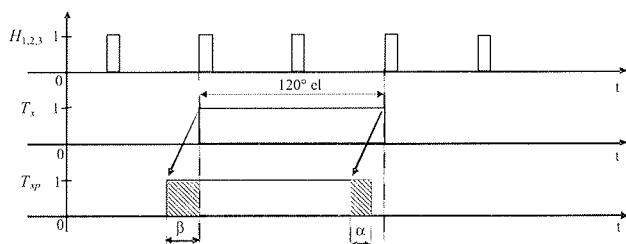

Slika 7: Proženje tranzistorjev v režimu slabljenja polja in podaljšanega prevajanja tranzistorjev.
Fig. 7: The proposed transistor switching.

Zaradi nesinusne oblike inducirane protinapetosti in nelinearnosti enosmernega EK motorja je odvisnost navora enosmernega EK motorja od kota predproženja β, od kota podaljšanja prevajanja tranzistorjev α in od vrtilne hitrosti motorja analitično zelo zahtevno opisati. Določitev optimalnega poteka navorne karakteristike oziroma največjega navora v celotnem obsegu vrtilnih hitrosti zagona motorja z NZ je obsegala laboratorijske meritve navornih karakteristik. Na osnovi izmerjenih 3D karakteristik sta bila določena kota α in β za posamezne vrtilne hitrosti EK motorja.

Izmerjeno odvisnost elektromagnetnega navora od kota predproženja β in od vrtilne hitrosti pri kotu α = 22,5° kaže slika 8.
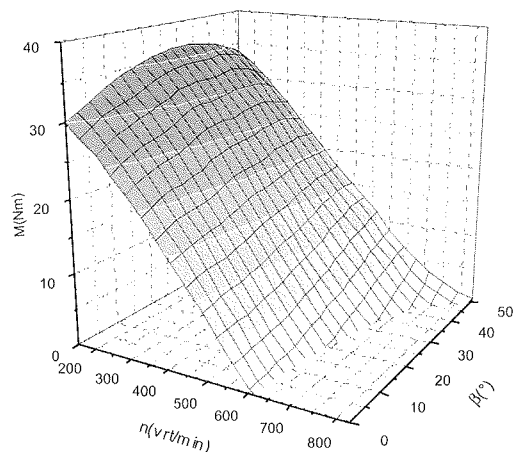
Slika 8: Odvisnost navora od kota β in od vrtilne hitrosti pri α = 22,5°.
Fig. 8: Torque dependence on β and rotational speed at α = 22,5°.

3D karakteristika na sliki 8 je le ena iz niza karakteristik, ki so bile izmerjene pri različnih kotih α. Iz merilnih rezultatov je vidna pričakovana odvisnost navora EK motorja od vrtilne hitrosti in od kotov α in β. Slika 9 kaže primer te odvisnosti pri vrtilni hitrosti motorja 500 vrt/min.
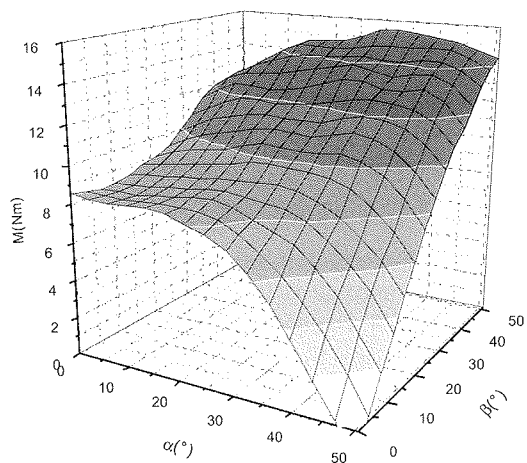
Slika 9: Odvisnost navora EK motorja od kota α in kota β pri vrtilni hitrosti 500 vrt/min.
Fig. 9: Torque dependence on α and β at rotational speed 500 rpm.

Slika 10 kaže odvisnost energijskega izkoristka motorja od kotov α in β pri vrtilni hitrosti 500 vrt/min. Iz slike je razvidno zniževanje energijskega izkoristka pri povečevanju kota α.
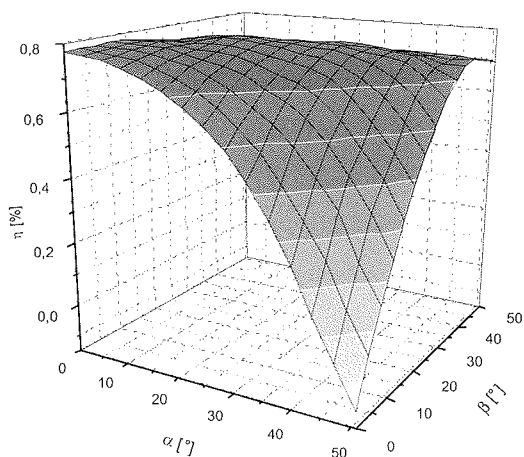
Slika 10:Odvisnost izkoristka motorja od kotov α in β
pri vrtilni hitrosti 500 vrt/min.
Fig. 10: Efficiency dependence on α and β at
rotational speed 500 rpm.

Pri strategiji zagona motorja z NZ je treba upoštevati trenutne razmere, ki vplivajo na zagonsko proceduro (temperatura bloka motorja, stanje napolnjenosti baterije,...). Da dosežemo optimalno delovanje enosmernega EK motorja kot tudi celotnega hibridnega sklopa, je treba upoštevati tako vrednosti največjega navora kot tudi izkoristek pri določeni vrtilni hitrosti. Tako sta predlagani dve zagonski strategiji, ki upoštevata zgoraj navedene zahteve. Pri hladnem zagonu je smiselno upoštevati kote pri katerih enosmerni EK motor ustvarja maksimalen navor, kar pripomore k zanesljivemu zagonu motorja z NZ. V drugem primeru, ko je temperatura bloka motorja z NZ dovolj visoka, pa krmilimo EK motor tako, da dosežemo najvišji energijski izkoristek in s tem pripomoremo k učinkovitejši rabi električne energije baterije hibridnega vozila (Tabela 1).

Tabela 1: Strategija zagona EK motorja pri dveh
*obratovalnih pogojih motorja z NZ.*
Table 1: Starting manoeuvre strategy at two operating
modes of IC engine.

Optimalen navor					Optimalen energijski izkoristek				
n [vrt/min]	α	β	M [Nm]	η[%]	n [vrt/min]	α	β	M [Nm]	η[%]
800	30	50,6	1,32	6,40	800	30	50,6	1,32	6,40
750	30	50,6	2,65	24,21	750	30	50,6	2,65	24,21
700	30	50,6	4,28	42,03	700	30	50,6	4,28	42,03
650	30	50,6	6,31	56,33	650	30	37,5	3,18	56,33
600	30	50,6	9,30	67,93	600	7,5	22,5	3,09	75,03
550	30	50,6	12,47	72,72	550	0	7,5	4,31	83,69
500	30	50,6	15,48	71,94	500	7,5	15	10,87	84,27
450	30	50,6	18,52	68,62	450	0	15	12,36	79,32
400	45	50,6	21,95	63,94	400	0	15	16,35	73,80
350	45	45	25,54	58,30	350	0	15	20,32	67,23
300	45	45	28,88	53,15	300	7,5	22,5	25,25	59,37
250	30	37,5	32,22	45,51	250	7,5	22,5	28,87	49,62
200	30	37,5	34,99	40,11	200	7,5	22,5	31,90	41,31
150	30	37,5	37,38	31,82	150	7,5	30	34,70	32,78
100	22,5	30	40,24	23,61	100	22,5	30	40,24	24,42
50	10	15	45,57	16,18	50	10	15	45,57	16,18
0	0	0	62,99	11,28	0	0	0	62,99	11,28

4. Eksperimentalni rezultati

Izsledki teoretičnih raziskav so bili preverjeni na prototipnem hibridnem motociklu v katerem sta na skupni mehanski gredi 1100 cm^3 štiritaktni motor z NZ in enosmerni EK stroj v funkciji IZG z nazivnimi parametri P_n = 900 W, $R_{A,B,C}$ = 8 mΩ, $L_{A,B,C}$ = 50 μH. Vse merilne, regulacijske in krmilne funkcije so bile izvedene z mikrokrmilnikom PIC 18F452.

Slika 11 kaže dvig navorne karakteristike in razširitev obsega vrtilnih hitrosti s predlaganim načinom krmiljenja enosmernega EK motorja ter primerjavo izkoristkov v obeh režimih delovanja glede na običajno 120° krmiljenje EK motorja.
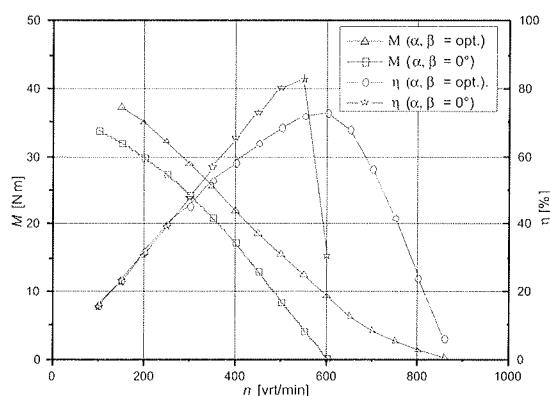
Slika 11:Primerjava navora in izkoristka enosmernega
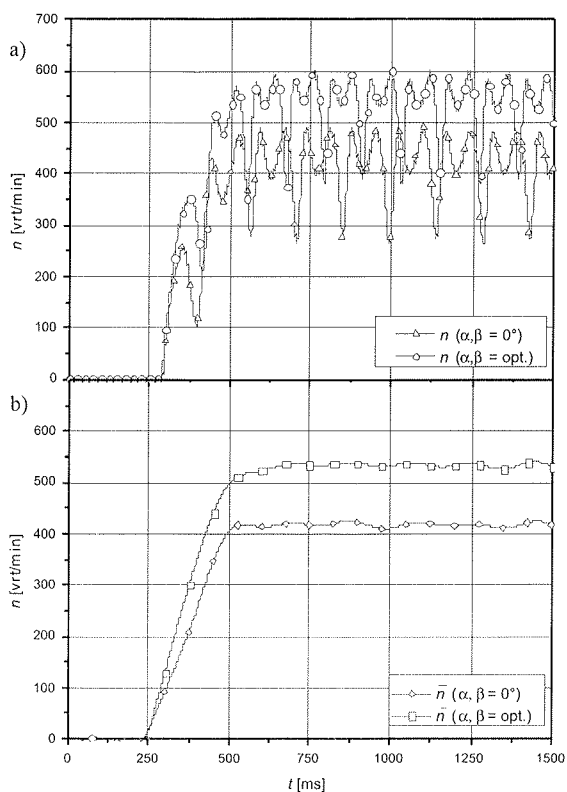*EK motorja pri α in β = 0 ter optimalnem α in β.*
Fig.11: Comparison of torque and efficiency at α, β =
0° and at optimal α and β.

Iz slike je razvidno, da EK motor s predlaganim načinom krmiljenja doseže v povprečju za 15% višji navor in dvig zgornje meje motorskega območja delovanja za 30%, kar je v skladu s teoretičnimi izsledki v drugem poglavju. Pri nižjih vrtilnih hitrostih rotorja je izkoristek motorja pri predlaganem načinu krmiljenja primerljiv z izkoristkom pri 120° krmiljenju, tako zaradi krajšega kota podaljšanja prevajanja tranzistorjev kot tudi zaradi manjšega vpliva slabljenja magnetnega polja. Z višanjem vrtilne hitrosti povečujemo komponento toka, ki slabi glavni magnetni pretok v motorju in hkrati prilagajamo kot prevajanja tranzistorjev, kar posledično povečuje elektromagnetni navor obenem pa niža izkoristek EK motorja.

Ustreznost predlaganega pristopa krmiljenja enosmernega EK motorja je bila ovrednotena na podlagi primerjave časovnega poteka vrtilne hitrosti zagona motorja z NZ z običajnim (120°) krmiljenjem in krmiljenjem po predlagani strategiji (Slika 12).

Valovitost vrtilne hitrosti (Slika 12a) je posledica navorne karakteristike motorja z NZ (štiritaktni motor). Doprinos predlaganega pristopa je bolje razviden iz primerjav povprečnih vrednosti vrtilne hitrosti (Slika 12b), kjer je očitna razlika v hitrosti naraščanja vrtilne hitrosti (pospeška) in tudi razlika v doseženi končni vrtilni hitrosti hibridnega agregata.

Informacije MIDEM 37(2007)3, str. 182-188

B. Pevec, P. Bajec, J. Nastran, D. Vončina: Optimizacija navorne
karakteristike elektronsko komutiranega motorja v hibridnem pogonu

*Slika 12: Primerjava časovnih potekov vrtilnih hitrosti
zagona motorja z NZ.*

Fig.12:   IC engine staring maneuver comparison.

Med izvajanjem optimalnega zagonskega postopka se na
podlagi izmerjene vrtilne hitrosti rotorja določata kota α in
β. Do vrtilne hitrosti 50 vrt/min sta oba kota enaka nič, saj
je zaradi visoke dinamike spreminjanja vrtilne hitrosti gnane
gredi hibridnega agregata in zaradi enostavnega načina
detektiranja pozicije rotorja EK motorja težko pravilno aplici-
rati strategijo kotno premaknjenih prožilnih pulzov za
posamezne tranzistorje močnostnega pretvornika. Visoka
dinamika vrtilne hitrosti izvira iz visokih trenutnih vrednosti
navora motorja z NZ, kot posledica kompresijskih in ek-
spanzijskih ciklov. Nad 50 vrt/min sta kota nastavljena v
skladu s predlagano strategijo.

5.   Zaključek

V članku je opisana problematika motorskega obratovanja
enosmernega EK stroja kot izvršilnega člena integriranega
zaganjalnik-generatorja v hibridnem motociklu. Obravna-
vana sta dva ukrepa za izboljšanje navorne karakteristike
in sicer slabljenje glavnega magnetnega pretoka ter nas-
tavljanje kota prevajanja tranzistorjev, ki sta se v praksi iz-
kazala kot izjemno učinkovita. Največji navor, ki ga obravna-
vani stroj lahko razvije, se je v povprečju zvišal za 15%,
povečal se je tudi razpon vrtilnih hitrosti delovanja stroja v
motorskem režimu s 600 vrt/min na 860 vrt/min. Posle-
dično se je izboljšala dinamika zagona motorja z notranjim
zgorevanjem, kar je verificirano z meritvami na hibridnem
vozilu. Slabost predlaganega pristopa je nižji izkoristek, ki
je posledica slabljenja glavnega magnetnega pretoka in

daljšega kota prevajanja toka skozi posamezna fazna navitja.
Predlagan pristop krmiljenja enosmernega EK motorja je
moč uporabiti v aplikacijah, ki zahtevajo obratovanje z mak-
simalnim navorom, izkoristkom ali pa s kombinacijo obeh v
celotnem območju vrtilnih hitrosti.

Literatura

/1/   B. K. Bose, Modern Power Electronics and AC Drives, Prentice –
      Hall, New Jersey, 2002.

/2/   T. M. Jahns, Torque Production in Permanent-Magnet Synchronous
      Motor Drives with Rectangular Current Excitation, *IEEE Trans-
      action on Industry Applications*, Vol. 1A-20, No. 4, str. 803-
      813, julij/avgust, 1984.

/3/   S. D. Sudhoff, K. A. Corzine, H. J. Hegner, A Flux-Weakening Strat-
      egy for Current-Regulated Surface-Mounted Permanent-Magnet
      Machine Drives, *IEEE Transaction on Energy Conversion*, Vol.
      10, No. 3, str. 431-437, september 1995.

/4/   P. Bajec, Analiza in sinteza integriranega zaganjalnika, genera-
      torja in ojačevalnika navora v hibridnem pogonu, Doktorska dis-
      ertacija, Fakulteta za elektrotehniko, Ljubljana, 2005.

/5/   P. Uršič, Močnostni pretvornik za vodenje integriranega zagan-
      jalnika in generatorja v hibridnem pogonu, Magistrsko delo,
      Fakulteta za elektrotehniko, Ljubljana, 2003.

/6/   P. L. Chapman, S. D. Sudhoff, C. A. Whitcomb, Multiple Reference
      Frame Analysis of Non-Sinusoidal Brushless DC Drive, *IEEE
      Transactions on Energy conversion*, Vol. 14, No. 3, str. 440-
      446, september 1999.

/7/   P. Bajec, B. Pevec, D. Vončina, D. Miljavec, J. Nastran, Extend-
      ing the Low-Speed Operation Range of PM Generator in Auto-
      motive Applications Using Novel AC-DC Converter Control, *IEEE
      Transactions on Industrial Electronic*, Vol. 52, No. 2, str. 436-
      443, april 2005.

/8/   H. Lavrič, D. Vončina, P. Zajec, F. Pavlovčič, J. Nastran, A Preci-
      sion Hybrid Amplifier for Voltage Calibration Systems, *Inf. MI-
      DEM*, vol. 34, No. 1, 2004, pp. 37 – 42

*Boštjan Pevec, univ. dipl. inž. el.
prof. dr. Janez Nastran, univ. dipl. inž. el.
izr. prof. dr. Danijel Vončina, univ. dipl. inž. el.*

*Laboratorij za močnostno
elektroniko in regulacijsko tehniko
Univerza v Ljubljani, Fakulteta za elektrotehnko
Tržaška cesta 25, 1000 Ljubljana, Slovenija
e-mail: bostjan.pevec@fe.uni-lj.si, janez.nastran@fe.uni-
lj.si, voncina@fe.uni-lj.si
tel: +386 1 47 68 466, fax: +386 1 47 68 487*

*dr. Primož Bajec, univ. dipl. inž. el.
Hidria Rotomatika d.o.o.
Spodnja Kanomlja 23, 5281 Spodnja Idrija, Slovenija
e-mail: primoz.bajec@aet.si
tel: +386 1 47 68 466*

DEFINICIJA KRITERIJSKE FUNKCIJE ZA RUBUSTNO OPTIMIZACIJO LASTNOSTI OPERACIJSKEGA OJAČEVALNIKA

Janez Puhan, Árpád Burmen, Sašo Tomažič, Tadej Tuma

Fakulteta za elektrotehniko, Ljubljana

Kjučne besede: ASIC, nastavljanje parametrov vezja, optimizacija vezij, kriterijska funkcija, robustnost

Izvleček: Načrtovanje analognega dela namenskega integriranega vezja (ASIC) je v grobem sestavljeno iz dveh korakov. Najprej načrtovalec izbere topologijo vezja, ki bo zastavljeni nalogi kos. Sledi nastavljanje parametrov vezja (npr. širin in dolžin kanalov MOS tranzistorjev), dokler postavljene zahteve niso izpolnjene. Drugi korak predstavlja optimizacijski proces, v katerem se načrtovalec odloča glede na svoje znanje in izkušnje. Odločitve načrtovalca v matematičnem smislu predstavljajo kriterijsko funkcijo. V članku je predstavljen matematični zapis kriterijske funkcije za MOSFET ojačevalnik, ki poskuša posnemati načrtovalske odločitve. Kriterijska funkcija je preizkušena na realnih primerih p- in n-kanalnega ojačevalnika. Rezultati racunalniške optimizacije pokažejo, da je mogoče z ustrezno izbiro kriterijske funkcije karakteristike klasično načrtovanega vezja celo izboljšati.

Cost Function Definition for Robust Optimisation of Operational Amplifier

Key words: ASIC, circuit sizing, circuit optimisation, cost function, robustness

Abstract: Designing of analogue part of an arbitrary application specific integrated circuit (ASIC) is a process which requires senior designer skills. It is mainly composed of analysing designer ideas by computer simulator.

Initially an adequate circuit configuration has to be set, afterwards he/she just varies circuit parameters (e.g. MOSFET channel widths and lengths) in order to achieve required attributes (circuit sizing). So the second stage of designer's work is a kind of optimisation process. Key to the final solution is an evaluation of the circuits tested during the process, which is done subjectively with regard to all his/her knowledge and experience. Translated to optimisation process this represents cost function. Mathematical definition of cost function for MOSFET operational amplifier is presented. It follows the designer's knowledge to avoid degenerated solutions. Robustness across a range of operating conditions is also taken into account. The definition is tested on a real operating amplifier case with p- and n-channel input differential pair. The results obtained with direct optimisation method show that an improvement over the human design can be achieved. Therefore it could represent a helpful tool for solving circuit sizing problem.

1 Uvod

Pri načrtovanju analognega dela integriranega vezja poskušamo zadovoljiti postavljene zahteve ali specifikacije vezja. To so lastnosti vezja, ki so podane z realnimi števili, kot npr. ojačenje, fazna varnost, pasovna širina ...

Ko vezje izpolnjuje vse zahteve v zadovoljivi meri, je postopek načrtovanja končan. Razdelimo ga lahko v dve stopnji. Najprej moramo določiti konfiguracijo vezja, za katero predpostavljamo, da bo z njo mogoče izpolniti postavljene zahteve. Pri tem se moramo opreti predvsem na izkušnje in teoretično znanje iz analize znanih tipov vezij. V literaturi zasledimo poskuse avtomatske sinteze topologije oziroma konfiguracije analognega vezja /1, 2/, ki pa se v praksi niso prijeli. V drugi stopnji se poskušamo s spreminjanjem parametrov vezja (npr. s spreminjanjem dolžin in širin kanalov MOSFET) postavljenim zahtevam čimbolj približati. V praksi to navadno pomeni, da vezja z različnimi nabori parametrov preizkušamo s pomočjo računalniških simulacij. Torej izvajamo optimizacijski postopek, ki je bil do sedaj obravnavan s stališč nominalnega vezja /3, 4/ in toleranc /5, 6, 7/. Na tej osnovi so bila izdelana optimizacijska orodja, ki temeljijo na analitično dobljenih enačbah /8, 9/, ali

na simulacijah /10, 11/ ipd. Nadalje številni članki s področja parametrske optimizacije /12, 13, 14, 15, 16/ nakazujejo njeno ključno vlogo pri avtomatizaciji načrtovanja integriranih vezij.

Vendar načrtovalci le redko posežejo po optimizacijskih orodjih. Odgovor na vprašanje, zakaj je temu tako, se skriva v vrednotenju vezij s posameznimi nabori parametrov, oziroma v odločanju katero vezje je boljše. Človek se o tem odloča subjektivno na podlagi izkušenj, znanja in preteklih poizkusov.

Tako se že v osnovi izogiba izrojenim rešitvam, ki imajo nekatere zahteve izpolnjene več kot zadovoljivo, ali pa niso robustne. Do sedaj so se s to težavo spoprijeli s postavljanjem različnih eksplicitnih ali implicitnih omejitev /12, 13, 15, 17/, ter tako močno omejili parametrski prostor. Takšne rešitve za načrtovalca niso priročne. Potrebovali bi matematično definicijo kriterijske funkcije, ki bo optimizacijski postopek vodila v pravo smer, ter se ne bo izgubljala v posameznih izrojenih rešitvah. Ob tem je potrebno dodati, da je kriterijska funkcija že po svoji naravi vezana na določen tip vezja, če ne kar na določeno konfiguracijo. Tako je potrebno za vsak tip (ali celo konfiguracijo) vezja (npr.

Informacije MIDEM 37(2007)3, str. 189-194

J. Puhan, Á. Burmen, S. Tomažič, T. Tuma: Definicija kriterijske
funkcije za rubustno optimizacijo lastnosti operacijskega ojačevalnika

ojačevalnik, sledilnik ...) definirati novo kriterijsko funkcijo. V članku je podana definicija kriterijske funkcije za klasično konfiguracijo CMOS operacijskega ojačevalnika.

2 Splošna matematična definicija kriterijske funkcije

Kot je bilo povedano že v uvodu mora kriterijska funkcija ovrednotiti vse relevantne lastnosti vezja. V nasprotnem primeru nas lahko optimizacijski postopek privede do nesmiselnih rešitev. (Primer: ojačevalnik s čim širšim frekvenčnim pasom, rezultat je vezje z zadovoljivo pasovno širino, vendar zelo visoko tokovno porabo) Tako ima kriterijska funkcija obliko vsote, v kateri se seštevajo prispevki posameznih lastnosti.

Naj vektor p označuje n spremenljivih parametrov vezja. Za vsak nabor parametrov dobimo m lastnosti vezja, ki jih združimo v vektor l.

$$l=l(p) \qquad l= [l_1, l_2 \ldots l_m]^T \quad p = [p_1, p_2 \ldots p_n]^T \qquad (1)$$

Lastnosti vezja se morajo v končnem rezultatu čimbolj približati postavljenim zahtevam, oziroma jih v celoti izpolniti. Zahteve popišemo z vektorjema s = [s1, s2 . . . sm]^T in z = [z1, z2 . . . zm]^T spodnjih in zgornjih mej. Pri tem ima lahko spodnja meja vrednost $-\infty$, zgornja pa ∞.

Naj bo g(x) poljubna zvezna monotono naraščajoča funkcija za x > 0, ki gre skozi koordinatno izhodišče (g(0) = 0). Definiramo funkcijo f(x), ki naj bo enaka:

$$f(x) = \begin{cases} 0 & x < 0 \\ g(x) & x \geq 0 \end{cases} \qquad (2)$$

Kriterijska funkcija E(p) naj bo vsota prispevkov posameznih lastnosti, in sicer bolj kot je lastnost oddaljena od zahtevanega intervala, večji je njen prispevek. Prav tako naj bodo lastnosti med seboj enako utežene, kar pomeni, da oddaljenost merimo relativno.

$$E(p) = \sum_{i=1}^{m} \left(f\left(\frac{s_i - l_i(p)}{u(s_i)}\right) + f\left(\frac{l_i(p) - z_i}{u(z_i)}\right) \right) \qquad (3)$$

Pri tem je funkcija u(x) = |x| za vsak x ≠ 0 in u(0) = konstanta večja od 0. Navadno vzamemo kar konstanta = 1. Da bi bila kriterijska funkcija dokončno določena si moramo izbrati še g(x). Najenostavnejša izbira je g(x) = x.

Vendar z optimizacijskim postopkom še ne moremo pričeti. Definirati je potrebno spremenljive parametre vezja p, vse relevantne lastnosti l, ter seveda želene vrednosti s in z. Vse našteto se navezuje na izbran tip vezja, oziroma konkretno konfiguracijo vezja. V razdelku 4 bomo podrobneje razčlenili CMOS operacijski ojačevalnik.

3 Robustnost

Optimizacijski postopek bo našel minimum kriterijske funkcije (3), kjer bodo lastnosti vezja l izpolnjene v najboljši

možni meri. Vendar bo to res le ob nominalnih pogojih izdelave in delovanja, kar pa ne zadošča. Ojačevalnik z odličnimi lastnostmi pri 25 °C, ki se pri 75 °C prelevi v oscilator, je neuporaben. Zato morajo postavljene specifikacije veljati preko širšega območja, znotraj katerega se lahko v vnaprej predpisanih mejah spreminjajo tako pogoji delovanja (npr. temperatura), kot pogoji izdelave (npr. spremembe procesa izdelave integriranih vezij). Oziroma vezje mora biti načrtovano robustno.

Spremenljive pogoje izdelave nam poda proizvajalec z r nabori parametrov modelov polprevodniških elementov. Poleg tega imamo še t spremenljivih navzgor in navzdol omejenih pogojev delovanja, ki definirajo množico $D \subset R^t$. Lastnosti l naj bi veljale v vsaki točki množice D za vsakega izmed r naborov, kar je zaradi neskončnega števila točk v D nemogoče preveriti. Zato množico D diskretiziramo in za vsakega izmed t pogojev izberemo v značilnih in ekstremnih vrednosti (npr. tri diskretne temperature). Množica D sedaj vsebuje končno število $\prod_{i=1}^{t} v_i$ točk ali kotov. Z upoštevanjem r naborov lahko robustnost vezja preverimo z analizo oziroma določitvijo lastnosti l v $r \prod_{i=1}^{t} v_i$ kotih. V tem kontekstu lahko nominalne pogoje obravnavamo kot en kot.

Da bi optimizacijski postopek zanesljivo privedel do robustnega vezja, bi morali določiti lastnosti l v vsakem izmed $r \prod_{i=1}^{t} v_i$ kotov v vsaki iteraciji. To v praksi ni izvedljivo, kakor tudi pri ročnem načrtovanju ne preverjamo vseh kotov po vsaki spremembi. Izkaže se, da je mogoče z optimizacijskim postopkom vseeno doseči robustnost z naslednjim hevrističnim postopkom. Sprva optimizacijski postopek izvedemo v nominalnih pogojih in na koncu preverimo vse kote. Za vsako lastnost, ki vsaj v enem kotu krši specifikacije, si zapomnimo tisti kot, v katerem je odstopanje največje. Optimizacijski postopek ponovimo, vendar tokrat v vsaki iteraciji poleg nominalnih lastnosti določamo še lastnosti v najbolj težavnih kotih. V najslabšem primeru moramo tako v vsaki iteraciji vezje analizirati v $m+1 \ll r \prod_{i=1}^{t} v_i$ kotih, kar je izvedljivo.

4 Lastnosti operacijskega ojačevalnika

Da bi optimizacijska zanka dala smiseln rezultat, moramo v kriterijski funkciji upoštevati vse relevantne lastnosti operacijskega ojačevalnika. Kot relevantne lastnosti lahko razumemo pomembnejše, oziroma ključne lastnosti vezja. Če katero izmed njih izpustimo, bo po optimizacijskem postopku ta lastnost vezja postavljena bodisi povsem naključno, ali pa bo imela celo zelo slabo vrednost. V prvem primeru je njena naključna vrednost posledica tega, da lastnost ni bila vključena v kriterijsko funkcijo. V drugem, usodnejšem primeru, pa so se na račun naše lastnosti poskušale izboljšati druge vključene lastnosti. Kriterijska funkcija se navezuje na določen tip vezja. Slika 1 prikazuje testno simulacijsko vezje za preizkušanje operacijskih ojačevalnikov.
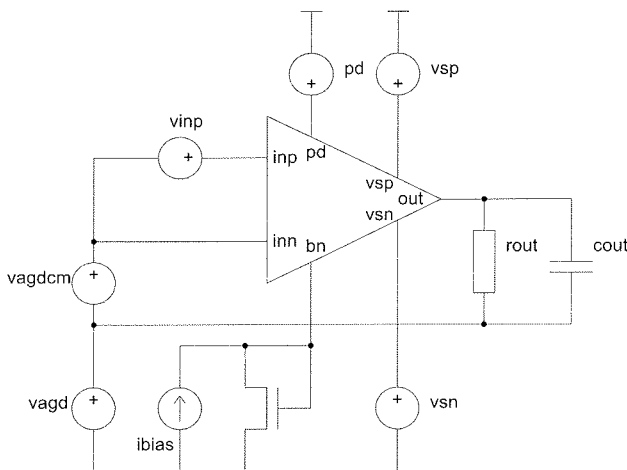
Slika 1: Testno simulacijsko vezje

Pri načrtovanju operacijskega ojačevalnika so pomembne naslednje lastnosti: površina silicija, ki ga ojačevalnik zavzema, enosmerni odziv vezja (dc analiza), odziv vezja na majhne izmenične signale v delovni točki (ac analiza), šum, ki se generira v vezju (noise analiza), ter hitrost vezja (tran analiza). Površino silicija A določimo enostavno z vsoto površin posameznih elementov. Prva pomembnejša odločitev nastopi po dc analizi, ki jo izvršimo s spreminjanjem vira vinp iz negativnega do pozitivnega nasičenja. Analogno maso zagotavlja vir vagd = (vsp − vsn)/2. Ojačevalnik pojmujemo za neizrojenega, če izhodna napetost vout objame analogno maso, in hkrati je njegovo aktivno področje med skrajnima vrednostima napetosti vinp. Oziroma veljati morata pogoja:

$$\min(vout) < vagd < \max(vout)$$

$$\frac{dvout}{dvinp}\bigg|_{\min(vinp),\max(vinp)} < \frac{\max\left(\frac{dvout}{dvinp}\right)}{2} \qquad (4)$$

Če pogoja (4) nista izpolnjena, potem vsem še nedoločenim lastnostim vezja (za zdaj še vse lastnosti razen površine) pripišemo najslabše možne vrednosti. To vodi v visoko vrednost kriterijske funkcije, kar izrojenemu vezju tudi pritiče. Nadaljnjih analiz v tem primeru ne izvajamo. V nasprotnem določimo dc lastnosti (slika 2), ki so: razpon v_{pp}, ojačenje v_{pp}/vin_{pp}, v_{offset}, simetrija $v_{outoffset}$, tokovna poraba ip in razlike napetosti v_{gs}−v_{th} tranzistorjev v tokovnih zrcalih. Slednje podajajo področje delovanja. Tokovna zrcala določajo delovno točko, zaradi česar moramo zagotoviti zanesljive preslikave tokov, oziroma ti tranzistorji morajo delovati v nasičenju.

Naslednja pride na vrsto izmenična analiza v delovni točki. Tokrat zahtevamo monotono padajočo fazo prevajalne funkcije. Nemonotono padajoči fazni poteki lahko v optimizacijskem postopku privedejo do le navidezno velike fazne varnosti, zato jih prepovemo.

$$\frac{d\arctan \frac{\text{im}\frac{vout}{vinp}}{\text{re}\frac{vout}{vinp}}}{df} < 0 \qquad (5)$$
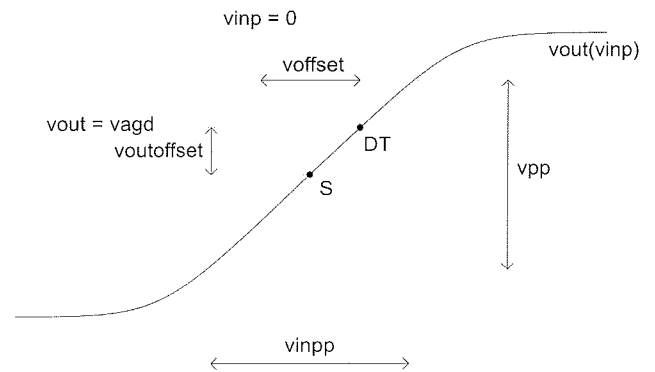
Slika 2: dc lastnosti (DT . . . delovna točka, S . . . simetrična točka)

Če je pogoj (5) izpolnjen določimo relevantne lastnosti v frekvenčnem prostoru, ki so: pasovna širina f_{0dB}, fazna varnost pm, amplitudna varnost am, ter maksimalne vrednosti za CMRR, $PSRR_p$ in $PSRR_n$ za frekvence nižje od f_{rr}. Poleg naštetih lastnosti vključimo še ojačenje takoj za vhodnim diferencialnim parom G_{int}. Načrtovalske izkušnje nam kažejo, da je modro ojačenje v vozlišču t_1 pri frekvenci, ko ojačenje v t_2 pade na polovico (sliki 4 in 5), držati na določenem nivoju.
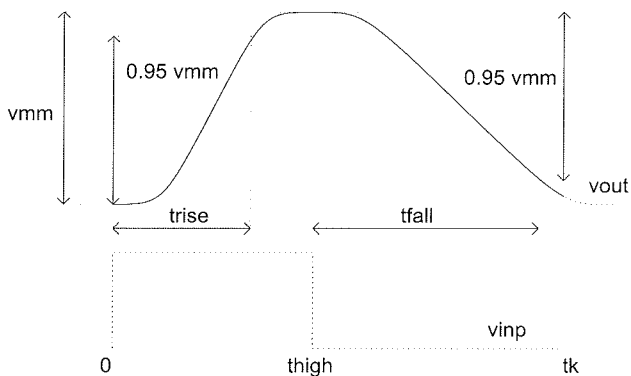
Nadaljujemo s šumno analizo, kjer nimamo posebnih zahtev. Relevantni šumni lastnosti sta spektralni gostoti šuma ojačevalnika pri frekvencah $f_{1/f}$ in f_{term}. Frekvenci sta izbrani tako, da podajata informacijo o 1/f šumu pri nizkih frekvencah in o termičnem šumu pri nekaj višjih frekvencah. Poleg tega se izkaže, da je prav, če nadzorujemo tudi razmerje med integriranim šumom diferencialnega para in pripadajočega tokovnega zrcala $\frac{noise_{diff}}{noise_{mir}}$. S tem poskušamo relativno znižati šum tokovnega zrcala, ki ga vidimo navzven.

Zadnja pride na vrsto časovna analiza. Vir vinp ima tokrat obliko pravokotnega impulza, ki ojačevalnik potisne iz negativnega v pozitivno nasičenje in spet nazaj (slika 3). Merimo hitrost odziva ojačevalnika, ki je izražena s časom naraščanja t_{rise} in padanja t_{fall}. Časovna analiza je najzahtevnejša in kot taka računsko najbolj obremenjujoča. V optimizacijskem postopku se zanesljivo pojavijo kombinacije parametrov, ki dajo zelo počasna vezja. Da bi v teh primerih lahko izmerili t_{rise} in t_{fall}, bi moral biti vhodni impulz zelo širok, časovno analizo pa bi bilo potrebno izvesti za mnogo daljši čas. Ker vnaprej ne vemo, katero vezje je počasno, bi bilo potrebno to narediti za vsako kombinacijo parametrov, tudi za hitra vezja. To je nesprejemljivo, zato počasna vezja enostavno razglasimo za izrojena.

$$v_{mm} > 0.95(\max(vout_{dc}) - \min(vout_{dc}))$$

$$vout(t_k) < vout(0) + 0.05v_{mm} \qquad (6)$$

Razpon dosežen v časovni analizi mora obsegati vsaj 95% razpona iz dc analize. Hkrati zahtevamo, da izhodna napetost v končnem času t_k pade na manj kot 5% napetosti v pozitivnem nasičenju.

Informacije MIDEM 37(2007)3, str. 189-194

J. Puhan, Á. Burmen, S. Tomažič, T. Tuma: Definicija kriterijske
funkcije za rubustno optimizacijo lastnosti operacijskega ojačevalnika

Slika 3: tran lastnosti

Na tem mestu se postavi vprašanje, ali je vključitev časovne analize v kriterijsko funkcijo sploh potrebna? V kriterijski funkciji se že nahajajo izmenične lastnosti vezja (ac analiza). Z višanjem pasovne širine vezja avtomatično dvigujemo tudi hitrost vezja. Vendar ne smemo pozabiti, da je izmenična analiza narejena na vezju lineariziranem v delovni točki. Tako nam izmenične lastnosti vezja posredno nekaj povedo o hitrosti spreminjanja izhodne napetosti preko aktivnega področja. Za merjenje časov naraščanja in padanja pa je ključnega pomena tudi podatek, kako globoko v nasičenju se vezje nahaja in koliko časa potrebuje, da se iz tega stanja zbudi. Časovna analiza torej v kriterijsko funkcijo vnaša velikosignalne lastnosti in je kot taka potrebna.

5   Rezultati na realnih primerih

Opisano kriterijsko funkcijo smo preizkusili na dveh realnih primerih CMOS operacijskih ojačevalnikov (sliki 4 in 5). Ojačevalnika sta bila načrtovana v 0.8 μm tehnologiji firme Austria Micro Systems. Spremenljivi parametri so bile vse širine in dolžine kanalov tranzistorjev v vezju, ter vrednost upornosti in kapacitivnosti v povratni vezavi izhodnega tranzistorja.
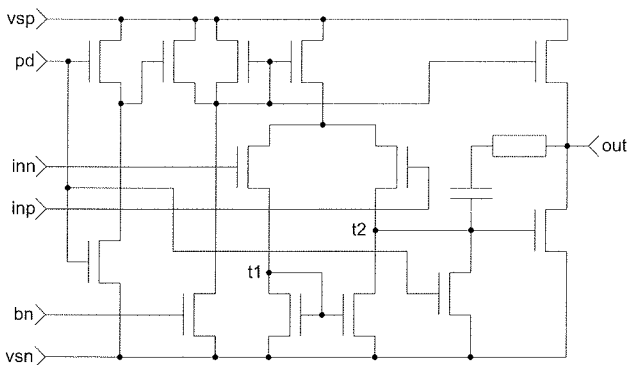
Slika 4: Operacijski ojačevalnik s p-kanalnim diferencialnim parom

Privzeti so konstantni pogoji izdelave (r = 1), kar ne okrni nazornosti obeh primerov. Robustnost vezja smo preizkusili z naslednjimi spremenljivimi pogoji delovanja: napajalna napetost vsp − vsn, enosmerni vhodni potencial vagd-cm, odmični tok ibias, rezistivni in kapacitivni del bremena

rout in cout, ter temperatura. Za vsako izmed veličin smo izbrali tri značilne vrednosti (t_i = 3 za i = 1, 2 . . . 6), in sicer minimalno, nominalno in maksimalno.
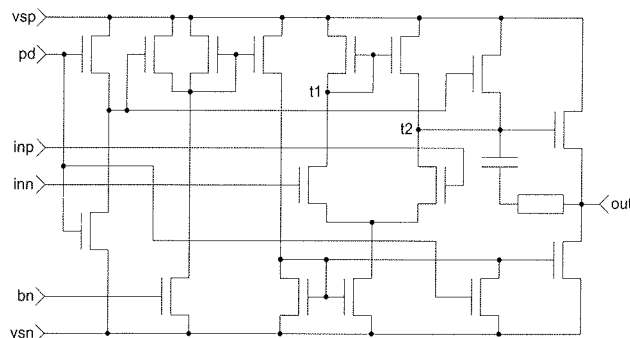
Slika 5: Operacijski ojačevalnik z n-kanalnim diferencialnim parom

Torej imamo 3^6 = 729 kotov. Da bi dosegli zadovoljivo robustnost, smo morali poleg lastnosti v nominalnih pogojih, v kriterijsko vključiti še fazno varnost pri maksimalnem kapacitivnem delu cout bremena. S tem postanejo zadovoljivo robustne tudi vse ostale lastnosti. V optimizacijskem postopku smo tako v vsaki iteraciji izračunavali le dva kota. Rezultati za oba primera so zbrani v tabelah 1 in 2.

Tabela 1: Rezultati za ojačevalnik s p-kanalnim diferencialnim parom

p-kanalni dif. par		zaht.	ref.	opt.
A	μm^2	10000↓	11669	10000
v_{pp}	V	4↑	3.6	3.6
v_{pp}/v_{inpp}		2000↑	1692	2052
v_{offset}	μV	100↓	182	35
$v_{outoffset}$	mV	200↓	239	200
i_p	μA	500↓	570	*618
$v_{gs} - v_{th}$	mV	600↑	506	755
f_{0dB}	MHz	20↑	17	20
G_{int}	dB	-0.7↑	-0.71	0.25
pm	°	70↑	69	70
am	dB	-30↓	-25	-30
$CMRR$	dB	-100↓	-96	*-92
$PSRRp$	dB	-100↓	-95	-100
$PSRRn$	dB	-60↓	-59	-82
$noise_{1/f}$	$nV/\sqrt{Hz}$	100↓	119	81
$noise_{term}$	$nV/\sqrt{Hz}$	9↓	9.6	9.3
$\frac{noise_{diff}}{noise_{mrr}}$		0.4↑	0.35	0.57
t_{rise}	ns	400↓	478	289
t_{fall}	ns	200↓	257	171

V tabelah 1 in 2 so zbrane lastnosti, ki so bile podrobneje opisane v odstavku 4. Vse načrtovalske zahteve oziroma specifikacije vezja so omejene le na eno stran. Znaka ↑ in ↓ povesta, da je to spodnja, oziroma zgornja meja. Zadnja

dva stolpca podajata lastnosti referenčnega in opti-
miziranega vezja. Referenčno vezje je bilo načrtovano klasi-
čno ročno. Optimizirano vezje je rezultat optimizacijskega
postopka, zagnanega iz skrajne točke parametrskega pros-
tora, ko imajo vsi parametri najmanjše možne vrednosti.

*Tabela 2: Rezultati za ojačevalnik z n-kanalnim
differencialnim parom*

n-kanalni dif. par		zaht.	ref.	opt.
A	μm^2	10000↓	14688	13315
v_{pp}	V	4↑	3.9	*3.8
v_{pp}/v_{inpp}		4000↑	3145	5069
v_{offset}	μV	80↓	89	67
$v_{outoffset}$	mV	100↓	122	100
i_p	μA	600↓	634	*736
$v_{gs} - v_{th}$	mV	700↑	311	697
f_{0dB}	MHz	20↑	15	20
G_{int}	dB	1↓	1.26	1
pm	°	80↑	76	86
am	dB	-40↓	-32	*-25
$CMRR$	dB	-110↓	-108	-110
$PSRRp$	dB	-50↓	-45	-47
$PSRRn$	dB	-50↓	-46	-46
$noise_{1/f}$	$nV/\sqrt{Hz}$	100↓	133	123
$noise_{term}$	$nV/\sqrt{Hz}$	9↓	9.4	*9.9
$\dfrac{noise_{diff}}{noise_{mirr}}$		5↑	4.95	5
t_{rise}	ns	400↓	440	260
t_{fall}	ns	600↓	671	368

Že hiter prelet rezultatov pokaže, da je optimizirano vezje v
večini lastnosti boljše od referenčnega. Slabše, vendar še
vedno zadovoljive lastnosti*, so se pojavile pravzaprav le
pri tokovni porabi in amplitudni varnosti (n-kanalni par).

6   Zaključek

Predlagana splošna formulacija kriterijske funkcije je lahko
uporabljena na poljubnem tipu, oziroma konfiguraciji vez-
ja. Robustnost vezja nadalje zagotovimo z razširitvijo kriter-
ijske funkcije s tako imenovanimi težavnimi lastnostmi v
skrajnih pogojih izdelave in delovanja. Pri tem so koti iz-
brani selektivno, kar pripomore k učinkovitejšemu optimiza-
cijskemu postopku.

Kriterijska funkcija opisane oblike je bila preizkušena v op-
timizacijskem postopku na dveh industrijskih primerih
CMOS operacijskega ojačevalnika. Pred tem so bile seve-
da določene vse relevantne lastnosti tega tipa vezja, ki vod-
ijo v smiseln rezultat. Iz rezultatov je razvidno, da lahko pred-
stavlja splošno optimizacijsko orodje s tako definirano kri-
terijsko funkcijo uporaben pripomoček pri načrtovanju in-
tegriranih vezij. Kljub temu pa načrtovanja na tej stopnji še
ni mogoče povsem avtomatizirati. Končna odločitev je še

vedno v načrtovalčevih rokah, ki z rezultati optimizacijske-
ga teka ni nujno zadovoljen. Postopek lahko v tem primeru
ponovi s spremenjenimi zahtevami, ali pa na roke nadalju-
je iz optimiziranega vezja.

V kriterijski funkciji ostajajo nedoločene relevantne in
težavne lastnosti za posamezne tipe, oziroma konfiguraci-
je vezij, ki bi zagotovile smiseln in robusten rezultat. S šir-
šim naborom lastnosti za več tipov vezij bi bilo mogoče
zgraditi splošnejše orodje za načrtovanje analognih in-
tegriranih vezij.

Literatura

/1/ H. Y. Koh, C. H. Sequin, P. R. Gray. OPASYN: a compiler for CMOS
     operational amplifiers. IEEE Transactions on Computer Aided
     Design of Integrated Circuits and Systems, vol. 9, no. 2, Feb.
     1990, pp. 113-25

/2/ J. P. Harvey, M. I. Elmasry, B. Leung. STAIC: an interactive frame-
     work for synthesizing CMOS and BiCMOS analog circuits. IEEE
     Transactions on Computer Aided Design of Integrated Circuits
     and Systems, vol. 11, no. 11, Nov. 1992, pp. 1402-17

/3/ R. K. Brayton, G. D. Hachtel, A. L. Sangiovanni-Vincentelli. A survey
     of optimization techniques for integrated-circuit design. Proceed-
     ings of the IEEE, vol. 69, no. 10, Oct. 1981, pp. 1334-62

/4/ W. Nye, D. C. Riley, A. Sangiovanni-Vincentelli, A. L. Tits. DELIGHT.
     SPICE: an optimization-based system for the design of integrat-
     ed circuits. IEEE Transactions on Computer Aided Design of In-
     tegrated Circuits and Systems, vol. 7, no. 4, April 1988, pp.
     501-19

/5/ S. W. Director. The simplicial approximation approach to design center-
     ing. IEEE-Transactions-on-Circuits-and-Systems, vol. CAS-24,
     no. 7, July 1977, pp. 363-72

/6/ K. J. Antreich, R. K. Koblitz. Design centering by yield prediction.
     IEEE Transactions on Circuits and Systems, vol. CAS-29, no. 2,
     Feb. 1982, pp. 88-95

/7/ P. Feldmann, S. W. Director. Integrated circuit quality optimization
     using surface integrals. IEEE Transactions on Computer Aided
     Design of Integrated Circuits and Systems, vol. 12, no. 12, Dec.
     1993; pp. 1868-79

/8/ M. del Mar Hershenson, S. P. Boyd, T. H. Lee. GPCAD: a tool for
     CMOS op-amp synthesis. 1998 IEEE/ACM International Con-
     ference on Computer-Aided Design, Digest of Technical Papers
     (IEEE Cat. No. 98CB36287), ACM, New York, NY, USA, 1998,
     pp. 296-303

/9/ G. Gielen, G. Debyser, K. Lampaert, F. Leyn, K. Swings, G. Van der
     Plas, W. Sansen, D. Leenaerts, P. Veselinovic, W. van Bokhov-
     en. An analogue module generator for mixed analogue/digital
     ASIC design. International Journal of Circuit Theory and Appli-
     cations, vol. 23, no. 4, July-Aug. 1995, pp. 269-83

/10/ E. S. Ochotta, R. A. Rutenbar, L. R. Carley. Synthesis of highper-
     formance analog circuits in ASTRX/OBLX. IEEE Transactions
     on Computer Aided Design of Integrated Circuits and Systems,
     vol. 15, no. 3, March 1996, pp. 273-94

/11/ A. Dharchoudhury, S. M. Kang. Worst-case analysis and optimiza-
     tion of VLSI circuit performances. IEEE Transactions on Compu-

Informacije MIDEM 37(2007)3, str. 189-194

J. Puhan, Á. Burmen, S. Tomažič, T. Tuma: Definicija kriterijske
funkcije za rubustno optimizacijo lastnosti operacijskega ojačevalnika

ter Aided Design of Integrated Circuits and Systems, vol. 14, no. 4, April 1995, pp. 481-92

/12/ M. del Mar Hershenson, S. P. Boyd, T. H. Lee. Optimal design of a CMOS op-amp via geometric programming. IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems, vol. 20, no. 1, Jan. 2001, pp. 1-21

/13/ G. Debyser, G. Gielen. Efficient analog circuit synthesis with simultaneous yield and robustness optimization. 1998 IEEE/ACM International Conference on Computer-Aided Design, Digest of Technical Papers (IEEE Cat. No. 98CB36287), ACM, New York, NY, USA, 1998, pp. 308-11

/14/ T. Mukherjee, L. R. Carley, R. A. Rutenbar. Efficient handling of operating range and manufacturing line variations in analog cell synthesis. IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems, vol. 19, no. 8, Aug. 2000, pp. 825-39

/15/ R. Phelps, M. Krasnicki, R. A. Rutenbar, L. R. Carley, J. R. Hellums. Anaconda: simulation-based synthesis of analog circuits via stochastic pattern search. IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems, vol. 19, no. 6, June 2000, pp. 703-17

/16/ K. Antreich, J. Eckmueller, H. Graeb, M. Pronath, F. Schenkel, R. Schwencker, S. Zizala. WiCkeD: analog circuit synthesis incorporating mismatch. Proceedings of the IEEE 2000 Custom Integrated Circuits Conference, (Cat. No. 00CH37044), IEEE, Piscataway, NJ, USA, 2000, pp. 511-14

/17/ P. Mandal, V. Visvanathan. CMOS op-amp sizing using a geometric programming formulation. IEEE Transactions on Computer Aided Design of Integrated Circuits and Systems, vol. 20, no. 1, Jan. 2001, pp. 22-38

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