UDK 621.3:(53+54+621+66)(05)(497.1)=00

ISSN 0352-9045

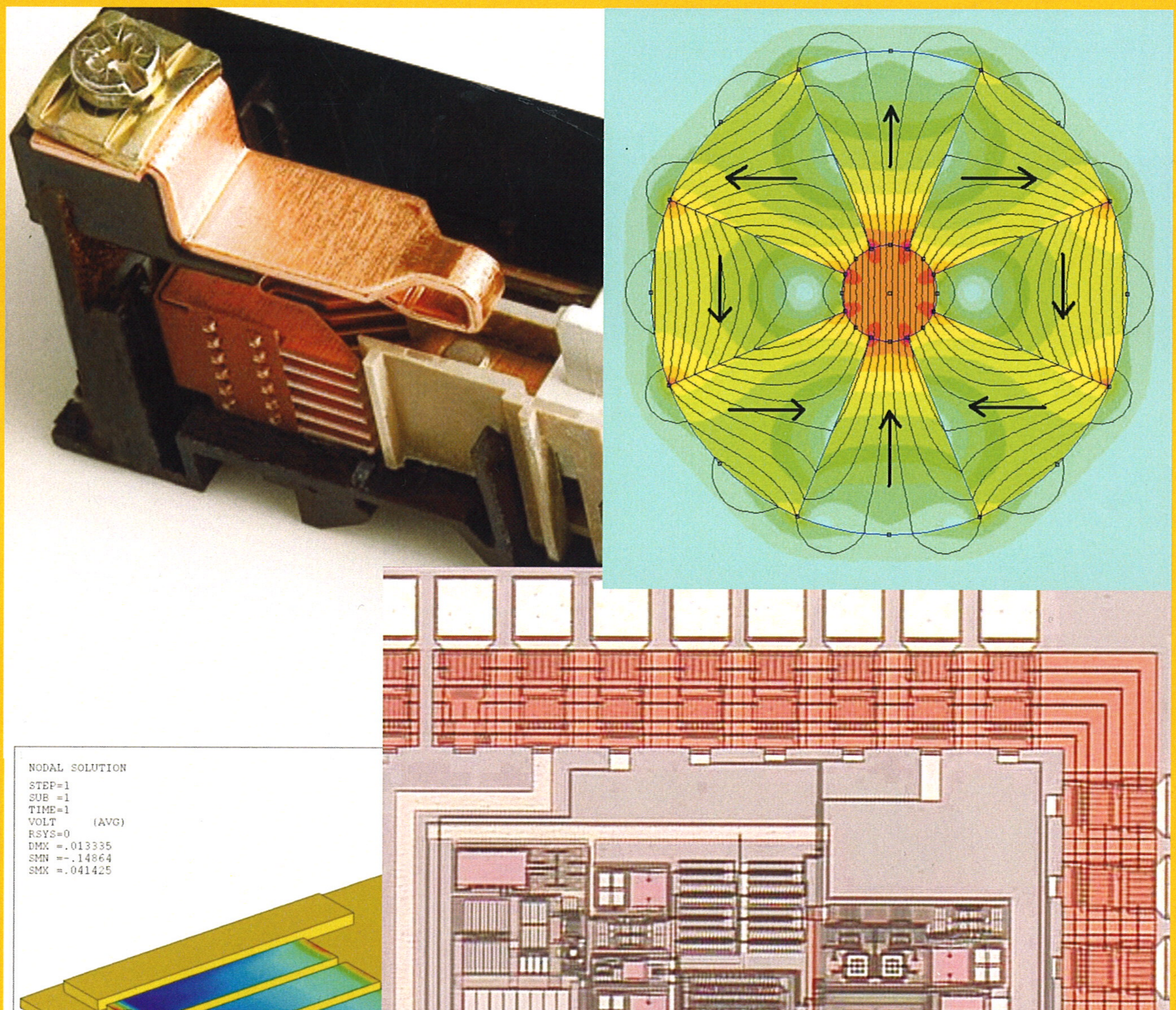
INFORMACIJE MIDEM

2○2009

Strokovno društvo za mikroelektroniko
elektronske sestavne dele in materiale

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

INFORMACIJE MIDEM, LETNIK 39, ŠT. 2(130), LJUBLJANA, junij 2009



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

INFORMACIJE **MIDEM** **2** ○ 2009

INFORMACIJE MIDEM LETNIK 39, ŠT. 2(130), LJUBLJANA, JUNIJ 2009

INFORMACIJE MIDEM VOLUME 39, NO. 2(130), LJUBLJANA, JUNE 2009

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

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

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

Znanstveni svet za tehnične vede je podal pozitivno mnenje o reviji kot znanstveno-strokovni reviji za mikroelektroniko, elektronske sestavne dele in materiale. Izdajo revije sofinancirajo ARRS in sponzorji društva.
Scientific Council for Technical Sciences of Slovene Research Agency has recognized Informacije MIDEM as scientific Journal for microelectronics, electronic components and materials.
Publishing of the Journal is financed by Slovene Research Agency and by Society sponsors.

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

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

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

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

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

Grafična priprava in tisk
Printed by
BIRO M, Ljubljana

Naklada
Circulation
1000 izvodov
1000 issues

Poštnina plačana pri pošti 1102 Ljubljana
Slovenia Taxe Percue

UDK621.3:(53+54+621+66), ISSN0352-9045 Informacije MIDEM 39(2009)2, Ljubljana

Slika na naslovnici je skupek fotografij in slik iz posameznih prispevkov v tej številki.	Front page is constructed of photos and pictures taken from contributions published in this issue.

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 2009, brisati iz seznama članstva.

Prijavnice pošljite na naslov:

MIDEM pri MIKROIKS

Stegne 11

1521 Ljubljana

Ljubljana, junij 2009

Izvršilni odbor društva

UDK621.3:(53+54+621+66), ISSN0352-9045 Informacije MIDEM 39(2009)2, Ljubljana

VPLIV STEN OB KONTAKTIH Z OBLOKOM NA IZKLOPNO ZMOGLJIVOST NIZKONAPETOSTNEGA ODKLOPNIKA

Martin Bizjak

Iskra MIS, Kranj, Slovenija

Kjučne besede: nizkonapetostni odklopnik, napetost obloka, tokovna omejitev, kratki oblok, delitev obloka, deion komora, prepustni tok, joulski integral izklopa (integral I^2t), vpliv plastične mase na izklopno zmogljivost.

Izvleček: Izklopna zmogljivost nizkonapetostnega odklopnika je odvisna od razvoja električnega obloka in njegovih značilnosti. Odklopnik naj pri izklopu kratkostičnega toka zagotovi čim manjšo amplitudo prepuščenega toka in čim manjši joulski integral (integral I^2t). To je omogoča učinek tokovne omejitve kratkostičnega toka. Tokovno omejitev določa razvoj obločne napetosti, ki naj dovolj hitro naraste preko neke kritične vrednosti. Hitrost naraščanja obločne napetosti in njena maksimalna vrednost sta odvisni od hitrosti odpiranja kontaktov, ki prekinjajo varovani tokokrog, formiranja zanke obloka med kontakti in delitve obloka na ustrezno število delnih oblokov.

Računska simulacija poteka toka in napetosti obloka med kontakti z ad hoc računskim modelom razkrije, katere značilne količine obloka odločujoče vplivajo na tokovno omejitev pri izklopu. S simulacijo ponazorimo časovni potek obločne napetosti in prepuščenega toka, kot ga registriramo eksperimentalno na oscilogramu pri preizkusu izklopne zmogljivosti. Nekatere vrednosti, ki jih potrebujemo za računsko simulacijo najdemo v literaturi, druge pa vzamemo kot variabilne parametre. Iz primerjave simulacijskih in eksperimentalnih rezultatov sklepamo na vplive, ki so posledica hitrosti odpiranja kontaktov in značilnosti v odklopniku vgrajenega materiala. Tako je bil kvalitativno ugotovljen vpliv plastične mase iz obloge kontaktno-obločnega prostora (s primerjavo dveh vrst materiala) na izklopno zmogljivost obravnavanega tipa odklopnika.

The influence of lateral walls near arcing contacts on breaking capacity of low-voltage circuit-breaker

Key words: low-voltage circuit breaker, arc voltage, current limiting, short arc, arc splitting, arc chute, let-through current, joule integral (I^2t integral), influence of plastic material on breaking capacity.

Abstract: The breaking capacity of low-voltage circuit breaker is closely related to the development of arc voltage and its magnitude during breaking of short-circuits current. The magnitude of let-through current shall be reduced and the joule integral minimized as well in order to achieve significant current-limiting effect of circuit breaker. Current limiting requires the magnitude of arc voltage to increase in the sufficiently short time interval to the relatively stationary value above the certain critical voltage. Both of them are dependent on the movement of contact members during opening and on the elongation speed of arc column in the initial phase as well as on the subdividing of arc in the arc chute.

Numerical simulation of the variation of arc voltage and let-through current during the breaking time reveals the influencing quantities of arc crucially affecting the current limiting action. The numerical procedure simulates the current and voltage waveforms normally recorded in oscillograms of breaking test shots. Some values necessary for the calculation were found in literature, others are used for simulation as variable parameters. The mutual combination of experimental and simulation results led us to the conclusion which physical effects, either movement of contacts or characteristics of materials (metallic and non-metallic) are dominant at break. By this way a significant influence of plastic material forming lateral walls around arcing contacts on the breaking capacity of circuit breaker under discussion was indicated qualitatively by comparing effects of two types of plastics.

1. Uvod

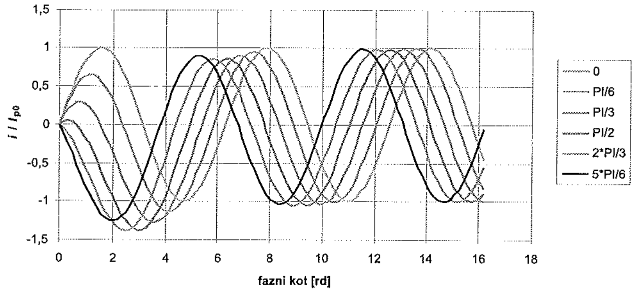
Pri nastanku nenormalnih stanj v nizkonapetostnem energetskem tokokrogu, kot je pojav prenapetosti ali tokovne preobremenitve v kratkem stiku, je treba tako stanje čim hitreje odpraviti ali tokokrog zaščititi pred njim. Prevelik tok skušamo hitro prekiniti z odklopnikom, kar pa v nizkonapetostnem energetskem tokokrogu traja vsaj toliko časa, da trenutna vrednost toka doseže prvo ničelno vrednost. To preide izmenični tok na »naravni« način v vsaki polperiodi izmeničnega toka in odklopnik bi lahko uspešno opravil izklop toka, če bi po prehodu toka skozi »naravno ničlo« obdržal tok na ničelni vrednosti še naprej. V tokokrogih z veliko razpoložljivo močjo (in z velikim razpoložljivim tokom I_p v kratkem stiku) pa bi izklop trajal že predolgo in bi bil vpliv preobremenitve z joulskim integralom (integral I^2t) in z končno vrednostjo I_D prepuščenega toka že kritičen. Zato odklopnik pri operaciji izklopa omeji tok po amplitudi (ki je bistveno manjša od amplitude izmeničnega toka) in po času do prve tokovne ničle (bistveno krajši od trajanja polperiode izmeničnega toka), po kateri mora biti tok prekinjen, to je: ostati na vrednosti nič. Izklop je opravljen z učinkom »tokovne omejitve« /1/.

Časovni potek prepustnega toka od trenutka, ko se kontakti odklopnika začnejo odpirati, je odvisen predvsem od naraščanja napetosti med kontakti. Režo med razmikajočima se kontaktoma kontaktnega para takoj premosti električni oblok. Kot aktivno komponento tokokroga ga izkoristimo za omejitev prepustnega toka izklopa /2/.

Delovanje odklopnika v nenormalnih stanjih v tokokrogu, npr. pri kratkem stiku, ne moremo poljubno preskušati »in situ«, ampak v simuliranih razmerah, kar lahko opravljamo v specializiranih preskuševališčih. Tak laboratorij razpolaga z viri izmeničnega enofaznega ali trifaznega toka velike kratkostične zmogljivosti z možnostjo nastavitve razpoložljivega toka na nekaj 10 kA ali celo več kot 100 kA in z nastavitvijo preskusne napetosti do 1000 V. Preskusni tokokrog, v katerem preskušamo delovanje odklopnika, vsebuje ohmske in induktivne komponente, s katerimi simuliramo razmere v realnem tokokrogu v kratkem stiku. Pogoje preskusa glede na namen odklopnika podajajo ustrezni elektrotehniški standardi /3/, /4/. S posebnim stikalom lahko vklopimo preskusni tok pri vnaprej izbranem faznem kotu sinusne napetosti vira, s čimer simuliramo nastanek kratkega stika pri poljubnem tokovnem faznem kotu.

2. Potek izklopa s tokovno omejitvijo

Sposobnost odklopnika, koliko velik kratkostični tok še uspešno prekine in ostane v stanju delovanja, je podana z njegovo kratkostično izklopno zmogljivostjo. Zmogljivi odklopniki opravijo izklop v prvi polperiodi izmeničnega toka. Tokokrogi, ki dajo v kratkem stiku večji razpoložljivi tok (prospective current) I_p, imajo zaradi večje induktivne komponente impedance manjši faktor moči, toda amplituda toka v prvi polperiodi je lahko večja od amplitud v naslednjih periodah. Pri tokih nad 10 000 A pri nekaterih faznih kotih začetka kratkega stika je povečanje amplitude znatno. Na grafu Slike 1 je računsko simuliran vklop toka v pretežno induktivnem tokokrogu pri faktorju moči 0,3 in pri različnih vklopnih faznih kotih. Prva amplituda pri neugodnem vklopnem faznem kotu naraste glede na ostale za faktor 1,3.



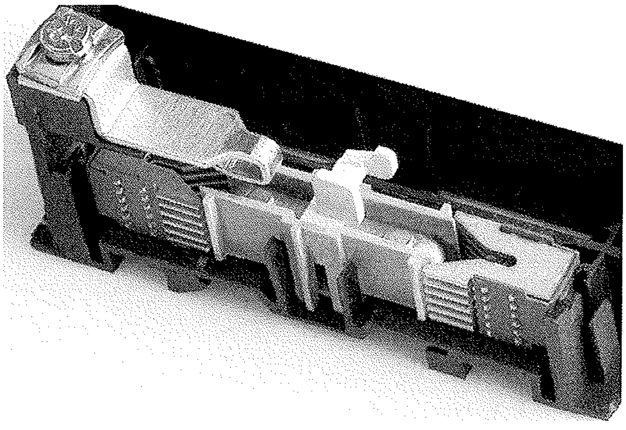
Slika 1: Vklop toka v pretežno induktivnem tokokrogu (faktor moči 0,3) pri različnih vklopnih faznih kotih, računska simulacija

Pri nenadnem povečanju toka, ki ga povzroči nenormalno stanje v tokokrogu, odklopnik reagira samodejno: pri nastavljenem faktorju nadtoka, ki je običajno med 10- in 15-kratnikom toka normalne obremenitve, sproži mehanizem za odpiranje kontaktov. Poleg tega ne glede na nastavljeno prožilno vrednost se na kontaktnem paru v stiku poveča odrivna sila zaradi toka, ki teče skozi majhno stično ploskev med kontaktnima površinama in ta prevlada nad silo kontaktnega stiska, tako da razmakne (odpre) kontaktaktni par (contact blow-off) /5/. V odklopniku za tok delovanja v normalnem stanju pri 25 A se odskok kontaktov zgodi pri trenutni vrednosti toka kakih 1000 A. Vgrajeni magnetni sprožnik deluje z nekaj več zakasnitve od trenutka, ko trenutna vrednost toka preseže nastavljeno vrednost: udarna igla kotve nadtokovnega magnetnega sprožnika, ki jo požene magnetno polje toka v njegovi tuljavi, pospeši odpiranje gibljivega kontakta do hitrosti nekaj m/s. Hkrati sprosti še napeto vzmet mehanizma za izklop, ki požene vklopno-izklopno ročičje, da dokončno razmakne kontakte in jih obdrži odprte na ustrezni medkontaktni razdalji izklopljenega stanja. Ta zadnji proces je zaradi vztrajnosti relativno velikih mas mehanizma najpočasnejši, vendar poteka v časovnem merilu milisekunde. Skupni učinek zaporednega delovanja vseh navedenih izklopnih dejavnosti je velika hitrost pri odpiranju gibljivega kontakta, ki doseže ob ločitvi kontaktnega para od 2 m/s do 5 m/s, v izklopljenem stanju pa je med njima razmik od 9 mm do 10 mm.

V trenutku ločitve kontaktov napetost na kontaktnem paru takoj naraste na 10 V do 20 V in se na oscilogramu poteka napetosti na kontaktnem paru jasno opazi kot skokovita sprememba vrednosti. Velikost »stopnice« je odvisna le od vrste kontaktnega materiala, ker v reži velikosti 0,1 mm oblok gori v kovinskih parah s površine kontaktov. Med kontakti z materialom na osnovi srebra z dodatki, npr. Ag/Ni, Ag/C, in na bakru (Cu) znaša (16 ± 1) V /6/. Pri nekoliko večji reži velikosti kak milimeter napetost naraste na 30 V do 35 V; ta vrednost je odvisna tudi od vrste plina, ki obdaja kontakte, ker pri tej medkontaktni reži oblok gori pretežno v okoliškem mediju /7/. Dolžina obločnega stolpca reda velikosti 1 mm pri napetostnem gradientu približno 2 V/mm vzdolž reže /8/ prispeva k celotni napetosti zanemarljivo malo. Dokler dolžina obločnega stolpca ne preseže kritične vrednosti, je oblok zasidran med kontakte in se ga ne da speljati drugam.

Ko med odpiranjem kontaktov reža med kontaktnim parom naraste preko kritične dolžine (pribl. 2 mm), lastno magnetno polje obloka že lahko obločni stolpec ukrivi v lok ali del zanke in ta se pod vplivom magnetnega polja giblje v smeri Biot-Savartove sile po podaljških kontaktnih delov, kontaktnih »rogov« in obločnih letev. Potencialno razliko na kontaktnem paru z oblokom med njima podajamo kot »napetost obloka«. V začetku odpiranja kontaktov v stanju negibljivega obloka (1 ms do 2 ms od trenutka ločitve kontaktnega para) je približno sorazmerna naraščanju medkontaktne reže (po oceni 2 V/mm × 5 m/s), v stanju gibljivega obloka, pa začne naraščati veliko hitreje. Značilni časovni potek napetosti obloka v stanju odpiranja kontaktov je razpoznaven na oscilogramih izklopa kot »stopnica« napetosti na kontaktnem paru pri prehodu napetostne razlike na sklenjenem kontaktnem paru zaradi toka skozi prehodno upornost na kontaktnem mestu, oz. pri zaprtih kontaktih, na napetost »kratkega obloka« pri ločitvi kontaktnega para. Pogosto namesto enega kontaktnega para v odklopniku uporabljamo dvojni kontaktni par, ki prekine tokokrog na dveh zaporednih mestih. Gibljivi kontaktni del, lahek in za mehansko izvedbo enostavnejši »kontaktni mostiček«, premosti dva mirujoča kontaktna dela, tako da nastaneta pri izklopu dva zaporedna obloka. Na oscilogramu obločne napetosti se pri ločitvi kontaktnega mostička od mirujočih kontaktov opazi dve zaporedni napetostni »stopnici«, visoki po kakih 15 V .

Med naraščanjem trenutne vrednosti toka se povečuje tudi presek električno prevodnega obločnega stolpca. Prosto goreči oblok ima pri toku reda 1 kA do 10 kA premer med 10 mm in 20 mm /9/. Kontaktne dele nizkonapetostnega odklopnika obdajata bočno dve stranski steni, ki sta razmaknjeni za kakih 5 mm. Slika 2 prikazuje razporeditev sistema para mirujočih kontaktov (na sliki je zaradi preglednosti prikazan le eden od para) in kontaktnega »mostička«, vloženih med stranski steni, ki se na obeh koncih podaljšujeta v deion komori s paketom kovinskih lamel, med katere se oblok razdeli na delne obloke. Dvojni kontaktni par poskrbi, da pri razmaknitvi kontaktnega mostička nastaneta dva izklopna obloka, med vsakim kontaktnim parom eden. Vsak od njiju nekaj časa gori med razmikajočim se kontaktnim mostičkom in mirujočim kontaktnim delom enega in drugega para, potem pa se začneta bolj ali manj sočasno premikati proti koncema kontaktnega mostička in proti obema deion komorama. Pred vstopom v komoro oba obloka gorita v prostoru med stranskima stenama. Skozi vsakega od obeh obločnih stolpcev tedaj teče tok velikosti nekaj kA in glede na toku ustrezni premer prosto gorečega obloka, povzet iz /9/, lahko ugotovimo, da ga stranski steni omejujeta s precej manjšim razmakom in stiskata njegov presek. Oblok se ob stiku s steno učinkovito hladi, pri tem pa se material stene na površini segreje na nekaj 100°C. Taka temperatura presega mejo termične obstojnosti večine plastičnih mas. Površina stranskih sten začne emitirati snovne komponente z izparevanjem in ablacijo. Napetostni gradient obločnega stolpca se zaradi vpliva sten lahko precej poveča in s tem tudi celotna napetost obloka /10/.



Slika 2: Gibljivi kontaktni mostiček med stenama kontaktno-obločnega prostora z parom deion komor

Dokončno pa obločno napetost povečamo na želeno vrednost v deion komori, kjer oblok razdelimo na več zaporednih delnih oblokov. Napetost obloka U_{oblok} lahko uravnamo s številom lamel deion komore, ki določa število delnih oblokov n_{do}. Za oblok med jeklenimi lamelami z razmakom 2,5 mm je bila izmerjena napetost med 26,5 V in 27,5 V, pri ocenjenem gradientu 2,25 V/mm dolžine obloka /8/. Za delni oblok v deion komori nizkonapetostnega odklopnika, ki ima lamele običajno razmaknjene za 1 mm, lahko za oceno napetosti celotnega obloka U_{oblok} vzamemo napetost delnega obloka 28 V, U_{oblok} pa je v približku (1) sorazmerna številu delnih oblokov n_{do}:

$$U_{oblok} = 28 \cdot n_{do} \text{ [V]} \tag{1}$$

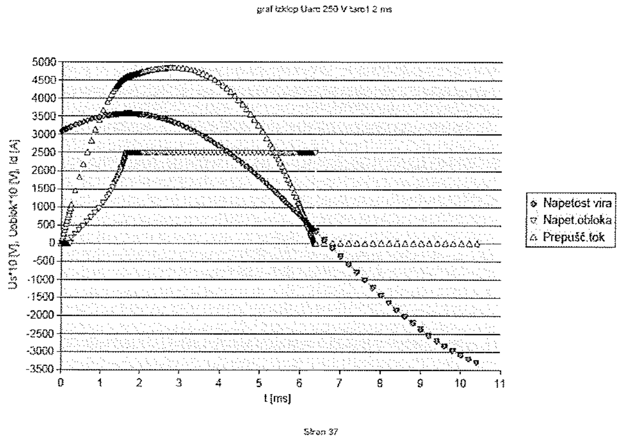
Velikost napetosti obloka v deion komori in čas, v katerem to vrednost doseže, pogojujeta učinek tokovne omejitve pri izklopu. Z njim želimo doseči omejitev konične vrednosti prepuščenega toka I_D in čim manjši joulski integral I^2t, ki je določen z (2):

$$I^2t = \int_0^{t_0} i^2 \mathrm{d}t \tag{2}$$

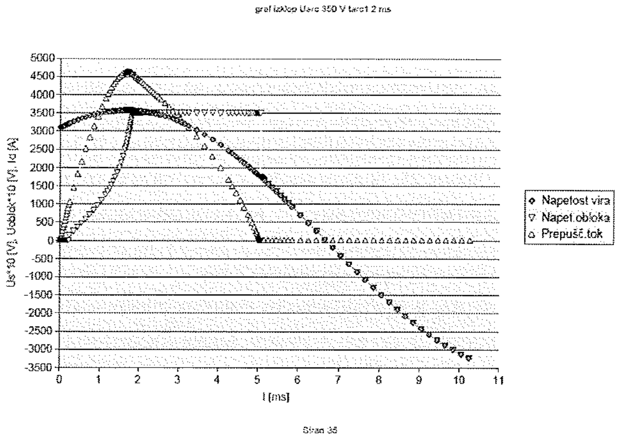
kjer je $i = i(t)$ tok med skozi odklopnik od trenutka nastanka v času $t = 0$ do trenutka t_0, ko $i(t)$ doseže ničelno vrednost, $i(t_0) = 0$, in jo obdrži poljubno dolgo, $i(t \geq t_0) = 0$. Če je čas trajanja izklopa t_0 krajši, je tudi vrednost I^2t običajno manjša.

3. Ocena vplivov obločne napetosti na tokovno omejitev

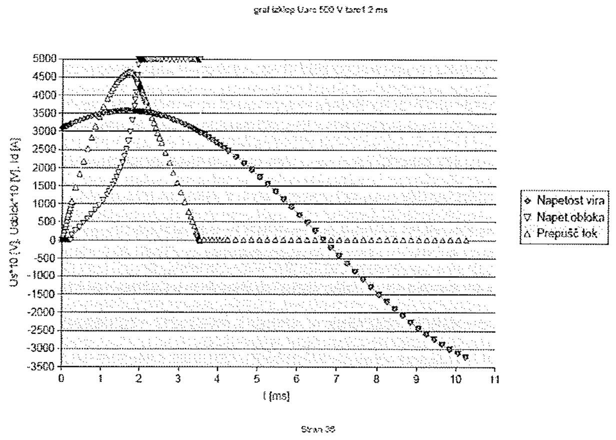
Vpliv velikosti napetosti obloka U_{oblok} in časa naraščanja $u_{oblok}(t)$ do U_{oblok} na potek prepuščenega toka $i(t)$ pokaže računska simulacija izklopa, kot je prikazana na Slikah 3a, 3b, 3c, 3d, 3e in 3f. S simulacijo so raziskane razmere v pretežno induktivnem tokokrogu, ki ga napaja vir izmenične napetosti 50 Hz z vrednostjo U_S = 230 V in pričakovano variacijo + 10%, pri razpoložljivem toku I_p = 10 kA, kjer je faktor moči 0,45 z pričakovanim odstopanjem - 0,05. V neugodnem primeru pri U_S = 253 V, faktor moči 0,40 kratek stik nastopi pri napetostnem faznem kotu 60°. Stikalna pot skozi odklopnik ima neko impedanco, ki je pretežno ohmska in je določena z zgornjo mejo dopustne disipacije $P \leq$ 2,0 W pri nazivnem toku odklopnika I_n. Ko trenutna vrednost $i(t)$ doseže 1000 A (izmerjeno v primeru obravnavanega tipa odklopnika), se zaradi elektrodinamičnega (blow-off) učinka kontaktni par loči, napetost med kontakti $u_{oblok}(t)$ takoj naraste od 0 v $t = 0$ na 15 V in narašča po t najprej linearno v skladu s predpostavko, da je v začetku odpiranje kontaktnega para linearno po času t, potem pa hitreje proti asimptotični vrednosti. Za presojo učinka hitrosti naraščanja $u_{oblok}(t)$ proti U_{oblok}, je opravljena simulacija izklopa za asimptoto $u_{oblok}(t) \rightarrow \infty$ pri 1 ms in pri 2 ms. Simulacija izklopa za obe predpostavljeni vrednosti je prikazana na Slikah 3a, b in c za asimptoto $u_{oblok}(t)$ pri 2 ms, na Slikah 3d, e in f pa pri 1 ms od nastanka kratkega stika. Po tem času se oblok nahaja v deion komori razdeljen med lamelami komore na delne obloke, zato je napetost obloka po času konstantna, $u_{oblok}(t) = U_{oblok}$. Vpliv vrednosti U_{oblok} je obdelana s simulacijo pri vrednostih 250 V, 350 V in 500 V. Te vrednosti so izbrane glede na izbrano napetost vira U_S = 253 V: prva je približno enaka U_S, druga je blizu temenski vrednosti U_S (358 V), zadnja pa je približno enaka 2 U_S. Rezultati simulacije za U_{oblok} = 250 V so prikazani na grafih Slike 3a in Slike 3d, za U_{oblok} = 350 V na grafih Slike 3b in Slike 3e, za U_{oblok} = 500 V pa na Sliki 3c in Sliki 3f.



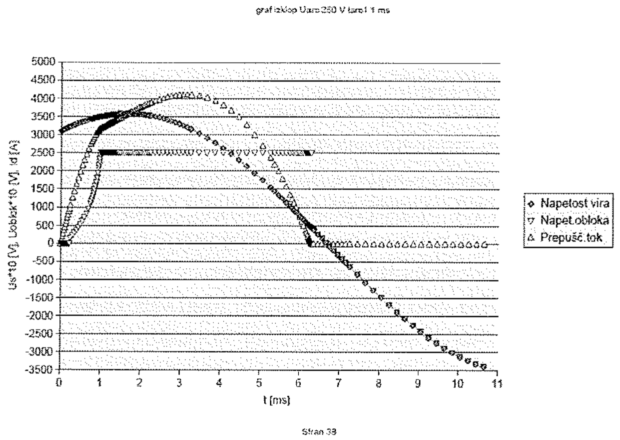
Slika 3a: *Simulacija izklopa I_p = 10 kA pri U_{oblok} = 250 V, čas dviga u_{oblok} 2 ms.*



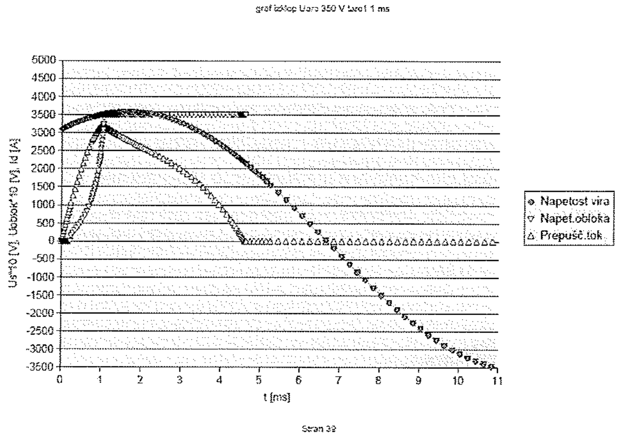
Slika 3b: *Simulacija izklopa I_p = 10 kA pri U_{oblok} = 350 V, čas dviga u_{oblok} 2 ms.*



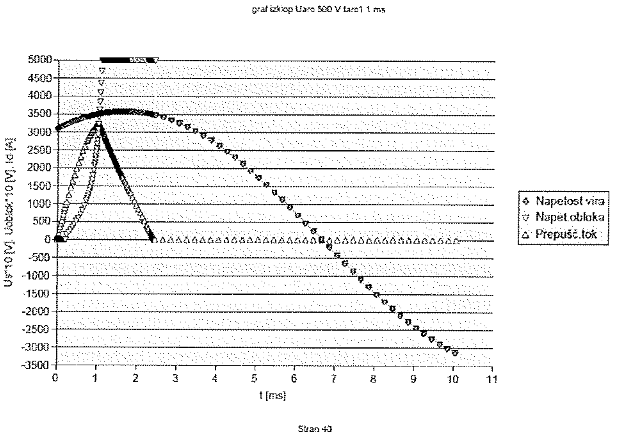
Slika 3c: *Simulacija izklopa I_p = 10 kA pri U_{oblok} = 500 V, čas dviga u_{oblok} 2 ms.*



Slika 3d: *Simulacija izklopa I_p = 10 kA pri U_{oblok} = 250 V, čas dviga u_{oblok} 1 ms.*



Slika 3e: *Simulacija izklopa I_p = 10 kA pri U_{oblok} = 350 V, čas dviga u_{oblok} 1 ms.*



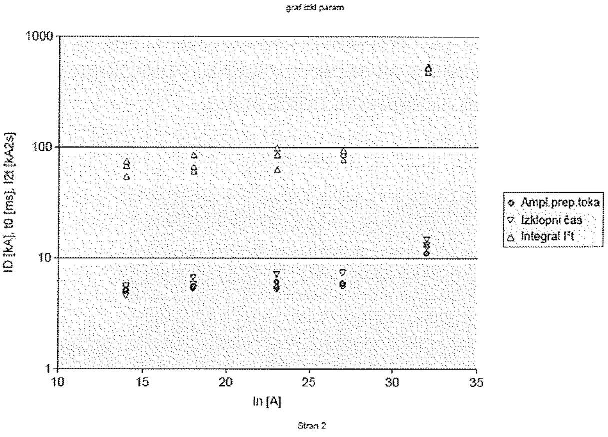
Slika 3f: *Simulacija izklopa I_p = 10 kA pri U_{oblok} = 500 V, čas dviga u_{oblok} 1 ms.*

Pri presoji vplivov hitrosti razvoja obloka iz »kratkega« obloka do popolnoma na delne obloke v deion komori »razdeljnega« obloka iz rezultatov simulacije razberemo, da hitrost prehoda iz začetne faze obloka v končno vpliva predvsem na konično vrednost prepuščenega toka I_D, če je $U_{\text{oblok}} \geq U_S \cdot \sqrt{2}$. Če torej U_{oblok} ne dosega vsaj temenske vrednosti vira U_S, je učinek tokovne omejitve pri izklopu majhen in izklop bo mogoč šele v trenutku blizu prve naravne ničle sinusnega toka. Pri večanju $U_{\text{oblok}} \rightarrow 2\ U_S$ je učinek tokovne omejitve že zelo velik, saj tok $i(t)$ upada od $I_D \rightarrow 0$ že s hitrostjo začetnega naraščanja proti I_D. Večji učinek tokovne omejitve v praksi ni sorazmeren vložku, s katerim bi ga dosegli. Da omejimo joulski integral izklopa I^2t, je treba predvsem zadostiti kriteriju, da je U_{oblok} večja ali vsaj približno enaka temenski vrednosti napetosti vira U_S, velikost I_D je v šibki korelaciji z vrednostjo I^2t. Zato pri izklopu izmeničnega toka 50 Hz (s polperiodo 10 ms) čas naraščanja obločne napetosti pod 1 ms nima odločilnega vpliva na vrednost I^2t, če vrednost U_{oblok} ne dosega kritične vrednosti, dokler ga omejujemo na vrednost, ki je za red velikosti manjša od trajanja ene polperiode izmeničnega toka. Pri ekstremno dolgih časih naraščanja $u_{\text{oblok}}(t) \rightarrow U_{\text{oblok}}$ v trajanju 1/2 polperiode ali dalj, seveda učinek tokovne omejitve na velikost I^2t tudi pri preseganju kriterija za vrednost U_{oblok} nima uporabnega vpliva.

4. Rezultati preskusov izklopa velikih tokov

Nizkonapetostni odklopniki za splošne namene zaščite inštalacij ali motorjev se izdelujejo za različne vrednosti nazivnih tokov. Za določeno območje nazivnih tokov, npr. po stopnjah za In od 2 A, 2,5 A, 4 A, 6,3 A, 10 A ... do zgornje vrednosti 25 A so istega tipa, to je iste velikosti in konstrukcije, za I_n od 63 A navzgor so drugega tipa, ki je večji in drugačne zgradbe. Zaradi funkcionalnih zahtev so včasih med različicami istega tipa, ki pripadajo različnim vrednostim I_n, nekatere razlike v materialu in izvedbi posameznih vgrajenih delov. Pri večjih vrednostih nazivnih tokov so potrebni večji preseki vodnikov ali termično bolj odporen material in podobno. Zaradi neizbežnega kompromisa pri spremembi materiala se včasih na preskušanju prototipov ali izdelkov v redni proizvodnji pojavijo neželeni stranski učinki.

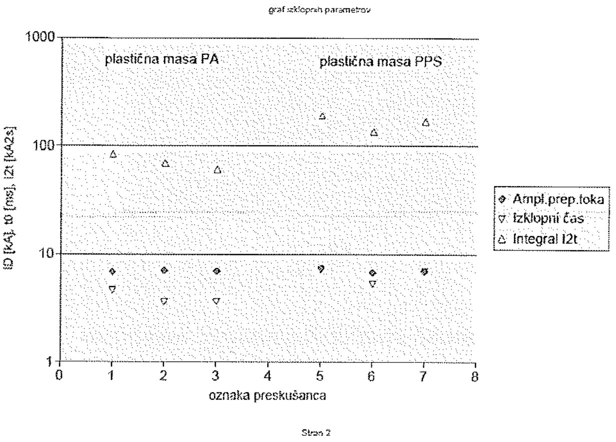
Na osnovi rezultatov simulacije izklopa, kot so prikazani na Slikah 3(a ... f) so bile določene smernice za konstrukcijsko zasnovo kontaktno-obločnega sistema odklopnika za zaporedje nazivnih tokov I_n do 32 A. Odklopniki za nekaj največjih vrednosti I_n so bili preskušeni na kratkostično izklopno zmogljivost v laboratorijskih pogojih pri nastavljeni napetosti vira 253 V z razpoložljivim tokom I_p = 25 kA. Preskusi izklopa so bili opravljeni pri vklopnih faznih kotih laboratorijskega »kratkostičnega« toka α, $\alpha + 30°$ in $\alpha + 60°$. Kot α je bil nastavljen na sinusoidi napetosti vira, vklopni fazni kot preskusnega toka pa je odvisen od fazne diference med napetostjo in tokom v preskusnem tokokrogu. Pri vsakem preskusu izklopa (operacija »O«) je bil posnet oscilogram prepuščenega toka $i(t)$. Potek $u_{\text{oblok}}(t)$ na oscilogramu ni bil registriran zaradi nevarnosti za merilni sistem, ker bi bilo treba za to meriti napetostno razliko med deloma kontaktnega para v prisotnosti obloka. Iz registriranih oscilogramov je bila določena vrednost I_D, izklopni čas t_0, ki je časovni interval med vklopom preskusnega toka in dokončno prekinitvijo toka, ko tok doseže ničelno vednost in ostane na tej poljubno dolgo, s pomočjo vgrajenega integratorja izračunan joulski integral I^2t. Preskus je bil opravljen na vsaj treh odklopnikih z istim nazivnim tokom. Za preskušance z I_n 14 A, 18 A, 23 A, 27 A in 32 A so merski rezultati urejeni kot zveza med izklopnimi količinami in nazivnim tokom I_n: $\{t_0, I_D, I^2t\} = f(I_n)$. Grafično jih podaja Slika 4: za odklopnike nazivnih tokov I_n z vrednostjo 14 A, 18 A, 23 A, in 27 A korelacijo med izmerjenimi količinami t_0, I_D, I^2t in I_n lahko prikažemo kot funkcijsko zvezo. Za I_n = 32 A pa rezultati odstopajo od te za več, kot je interval deviacije izmerjenih vrednosti.



Slika 4: Vrednosti značilnih izklopnih količin za odklopnike z I_n = 14A, 18A, 23A, 27A in 32A

Razlog za opisano razlika je v zgradbi tipske različice odklopnika za I_n = 32 A glede na različice za I_n < 32 A, kar se odraža tudi v zveznem spreminjanju izklopnih količin t_0, I_D in I^2t z I_n do 32 A in z nezveznostjo pri različici za 32 A. V tej je bilo treba zagotoviti večji kontaktni stisk in večjo odpornost kontaktov na zavaritev pri velikih tokih, zaradi velike termične obremenitve odklopnika pri nazivnih pogojih obremenitve pa je bilo treba izbrati za stene kontaktno-obločnega prostora in nosilec gibljivega kontakta termično odpornejši plastični material. Ta običajno v stiku z vročim obločnim medijem manj degradira in zato manj emitira v plazmo obloka. Eksperimentalni rezultati (Slika 4) kažejo, da so se pri odklopniku za 32 A vse vrednosti prikazanih izklopnih količin prekomerno povečale glede na pričakovane, ki bi jih dobili z ekstrapolacije rezultatov za 14 A, 18 A, 23 A in 27 A. S pomočjo rezultatov simulacije izklopa, prikazanih na Slikah 3 (a - f), presojamo, da je pri odklopniku za 32 A povečanje vrednosti I^2t verjetno posledica manjše U_{oblok}, kar hkrati podaljša tudi izklopni čas t_0, povečanje I_D pa posledica počasnejšega naraščanja $u_{\text{oblok}}(t)$ do končne U_{oblok}, ki ga pogojuje počasnejše odpiranje kontaktov

Hipoteza, da ima v odklopniku za 32 A vgrajen plastični material za stene kontaktnega prostora manjši učinek gašenja obloka, je bila preverjena eksperimentalno na odklopniku za I_n = 18 A. Ta v standardni izvedbi izkazuje dobro izklopno zmogljivost, ki dobro korelira z rezultati preskusov na odklopnikih ostalih nazivnih vrednosti. Zato sta bili za preskus vpliva plastičnega materiala na izklopno zmogljivost odklopnika izdelani dve različica odklopnika za 18 A: prva, v kateri so bile stene kontaktno-obločnega prostora iz poliamida (PA) in druga, v kateri je bil uporabljen polifenilensulfid (PPS). V obeh različicah odklopnika za 18 A so bili izdelani po trije preskušanci ene in druge verzije in preskušeni pri enakih pogojih. Izmerjene vrednosti značilnih izklopnih količin t_0, I_D in I^2t so prikazane na grafu Slike 5. Pri obeh verzijah je konična vrednost prepuščenega toka I_D v mejah merske nenatančnosti enaka. Vrednost I^2t pa je za preskušance z vgrajenim PPS materialom skoraj dvakrat večja, kot pri različici z vgrajenim PA materialom. Pri tem se je tudi izklopni čas t_0 povečal iz povprečno 4 ms na povprečno 7 ms. Rezultati preskusa očitno kažejo kvalitativni vpliv materiala na stene kontaktno-obločnega prostora na izklopno zmogljivost odklopnika.



Slika 5: Odvisnost značilnih izklopnih količin t_0, I_D in I^2t od vrste plastičnega materiala ob kontaktih, izklop I_p = 25 kA

5. Sklep

Značilnosti izklopnega pojav pri preskusu kratkostične izklopne zmogljivosti se odražajo v oscilogramih časovnega poteka napetosti med kontakti in prepuščenega toka. Merjenje medkontaktne napetosti in napetosti obloka zahteva poseg v merjeni vzorec, ki ni zaželen, nevarno pa je tudi za merilni inštrumentarij. Zato smo za razlago pojavov, ki so registrirani na oscilogramih, opravili še simulacijo izklopa z ad hoc računskim modelom. S hkratno obravnavo eksperimentalnih in simulacijskih rezultatov smo kvalitativno razpoznali vpliv plastičnega materiala za stene kontaktno-obločnega prostora na oblok pri izklopu kratkostičnega toka in izbrali za praktično uporabo ustreznejši material.

6. Literatura

/1/ M. Lindmayer: Schaltgeräte, Springer, 1987

/2/ G. Burkhard: Schaltgeräte der Enenergietechnik, 1. Auflage, VEB Verlag Technik, Berlin, 1985

/3/ Standard SIST EN 60947-2: 2006, Nizkonapetostni odklopniki

/4/ Standard SIST EN 60898-1: 2004, Inštalacijski odklopniki

/5/ E. Vinaricky: Elektrische Kontakte, Werkstoffe und Anwendungen, 2. Auflage, Springer, 2002

/6/ A. Erk, H. Finke: Über das Verhalten unterschiedlicher Kontaktwerkstoffe beim Einschalten prellender Starkstrom-Schaltglieder, ETZ - A 86 (1965), pp.297-302

/7/ E. Vinaricky: Das Abbrand- und Schweissverhalten verschiedener Silber-Grafit-Kontaktwerkstoffe in unterschiedlichen Atmosphären, disertacija, TH Wien, 1994

/8/ H. Klepp: Über den Einfluss der Löschkammerkonstruktion auf die Lichtbogelöschung in Schützen grosser Nennstromstärken, disertacija, TU Carolo-Wilhemina, Braunschweig, 1982

/9/ P. G. Slade: Electrical Contacts, Marcel Dekker, 1999

/10/ M. Bizjak: Model ablacijsko stabiliziranega obloka v odklopniku, disertacija, Univerza v Ljubljani, Fakulteta za elektrotehniko in računalništvo, 1992

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

Prispelo (Arrived): 02.09.2008 Sprejeto (Accepted): 09.06.2009

SIMULACIJA KOMUNIKACIJSKIH SISTEMOV V REALNEM ČASU Z REALNO KOMUNIKACIJSKO OPREMO V SIMULACIJSKI ZANKI

Matjaž Fras, Jože Mohorko

Univerza v Mariboru, Maribor, Slovenija

Kjučne besede: simulacija, taktična omrežja, realni čas, radijska vidljivost

Izvleček: V članku smo predstavili nov pristop k simulacijam komunikacijskih omrežij, ki omogoča simulacije v realnem času z realno komunikacijsko opremo v simulacijski zanki. Pri tem pristopu je mogoče v simulacijsko okolje priključiti realne komunikacijske naprave, ki lahko komunicirajo s simuliranimi napravami ali pa z drugimi realnimi napravami preko simuliranih povezav ali omrežij. Tovrstne simulacije omogoča simulacijsko okolje OPNET Modeler z dodatnim novim modulom SITL (System-in-the-Loop), ki omogoča povezavo med realnimi in simuliranimi napravami. Uporabnost tovrstnih simulacij smo predstavili na primeru »real-sim-real« tipa simulacije, kjer realne komunikacijske naprave komunicirajo preko simuliranega omrežja v realnem času. V tem primeru smo, preko simuliranega brezžičnega omrežja na virtualnem terenu, prenašali realen promet digitalne telefonije (VoIP).

Real time communication systems' simulation with real communication devices in simulation loop

Key words: simulation, tactical network, real time, radio visibility

Abstract: This paper presents the new concept of the communication networks' simulations, which enable real time simulations, with the real communication equipment in the simulation loop. In the context of project »Modeling of Command and Control information systems« we develop the simulation system, which can be used in the tactical network's planning and evaluation process. This simulation system is based on the OPNET Modeler simulation tool, which is combined by our utilities (TPGen, ModCom), that enables traffic modeling of Command and Control information system (C2IS) used in Slovenian army /1, 2/. We developed also the expert system for automatic simulation results analysis to evaluate tactical networks' performances /3, 4/ . In the context of this project we also research the possibilities for real time tactical network simulations, where real tactical units' computers with installed C2IS communication are connected to the simulated wireless communication infrastructure on virtual terrain. The first such simulator is *Battlespace communication network planner and simulator*, which are described in /5, 6/. This simulator is based on special hardware communication equipment controlled by OPNET simulation tool, where real traffic is influenced by simulation results in real time. In our research we used, instead to the special hardware, the new OPNET Modeler's software module *System-in-the-Loop* (SITL) /7, 8/. This module allows interconnections and packets exchange between real and simulated communication devices or networks. This simulation concept is one of the most important novelties, in the last years and it was also presented in the OPNET conference OPNETWORK 2007 in Washington /10/.

Figure 1 shows all possible types of the SITL simulations. Figure 2 shows the example of the real-sim-real SITL simulation type, where real computers communicate through the simulated network. The mechanisms of real time SITL simulations with required software and hardware equipment are shown on the block scheme from Figure 3. In our SITL simulation test case of transmitting voice over IP (VoIP), we connect two external laptop computers across the simulated network as shown in Figure 4. The simulated network, which was modeled in OPNET simulation tool, is more detailed shown in Figure 5. The connectivity between external laptops was tested using ping utility as shown in Figure 6. As VoIP application /15/ we used freeware D-Voicer /12/ installed on real laptop computers. Figure 7 shows the user interface of active D-Voicer application on both, client and server, side. The results of SITL simulation, where real digital voice stream is transmitted through simulated network are shown in Figures 8 and 9.

In the third part we present the proposed concept of real time SITL tactical wireless networks' suitable for Slovenian army. The simulation concept is similar as in VoIP example case, described in previous section. The main difference is in the used communication applications, which are for this case Command and Control Information System (C2IS). In Slovenian case this C2IS consists from Iris Replication Mechanism (IRM) in conjunction with Sitaware graphical interface /1, 2, 16/ as shown in Figure 10. Tactical radios of military units we modeled, in OPNET Modeler simulation tool, using standard wireless routers connected over SITL gateway to the real tactical computers. Communication parameters of wireless routers are defined according to communication parameters of real tactical radios and modems. In OPNET Modeler simulation tools are considered influences modeled virtual terrain on wireless links using different radio wave propagation models /13, 19-22/. Simulated tactical network can be very realistically visualized using 3DNV visualization tool as shown in Figure 11.

1. Uvod

Visoka kompleksnost sodobnih komunikacijskih sistemov in kratek čas, ki je na voljo za iskanje kakovostnih rešitev, narekujeta potrebo po simulacijskih orodjih, ki so namenjeni simulacijam telekomunikacijskih omrežij, naprav, protokolov. Takšno potrebo je začutila tudi slovenska vojska ko je razpisala projekt[1] »Modeliranje taktičnih informacijskih sistemov poveljevanja in kontrole (TISPINK)« v okvirju katerega smo razvili simulacijski sistem, ki omogoča kakovostno načrtovanje, ovrednotenje in optimizacijo taktičnih komunikacijskih omrežij. Simulacijski sistem /1, 2/ temelji na simulatorju OPNET Modeler, ki smo ga nadgradili z aplikacijami TPGen, za avtomatsko modeliranje prometa med vozlišči omrežja v TISPINK, ModCom za avtomatsko modeliranje taktičnih radijskih naprav za prenos podatkov, Ek-

---

[1] *Rezultati, predstavljeni v tem članku so rezultat raziskav narejenih v okviru CRP projekta "Modeliranje informacijskih sistemov poveljevanja in kontrole", financiranega s strani Slovenskega ministrstva za obrambo.*

spertnim sistemom /3, 4/ za avtomatsko analizo in vrednotenje simulacijskih rezultatov, ter Taktičnim predvajalnikom /23/ za nazorno vizualizacijo rezultatov ekspertnega sistema. V takšnem simulacijskem sistemu je najtežavnejši problem natančno modeliranje kompleksnih aplikacij, kakršne so uporabljene tudi v TISPNK. V opisanem simulacijskem sistemu smo ta problem poenostavili s statističnim modeliranje podatkovnih izvorov.

V tem članku bomo pokazali drugačen pristop, kjer se izognemo modeliranju na aplikacijskem nivoju tako, da v simulacijsko zanko, v realnem času, povežemo realno komunikacijsko opremo. Eden prvih takšnih simulatorjev je predstavljen v člankih /5, 6/, kjer so avtorji predstavili simulator taktičnih omrežij v realnem času z imenom »Battlespace communication network planner and simulator«. Izdelan simulator temelji na simulacijskem okolju OPNET v povezavi s specialno v ta namen razvito komunikacijsko opremo, ki omogoča emulacijo vplivov simuliranih radijskih naprav na podatkovne prenose med realnimi komunikacijskimi napravami in aplikacijami.

V naši raziskavi smo za tovrstne simulacije uporabili namesto specialne emulacijske strojne opreme programsko rešitev, ki jo ponuja najnovejši OPNET modul System-in-the-Loop (SITL) /11/, katerega smo že na kratko opisali v /9/. Modul SITL omogoča izmenjavo paketov med realnimi in simuliranimi napravami ter predstavlja zadnjo večjo novost na področju simulacij komunikacijskih omrežij, ki je bila predstavljena na konferenci OPNETWORK 2007 /10/. V članku bomo predstavili nekaj rezultatov, ki smo jih pridobili v fazi testiranja tega koncepta, pri načrtovanju simulatorja TISPINK.

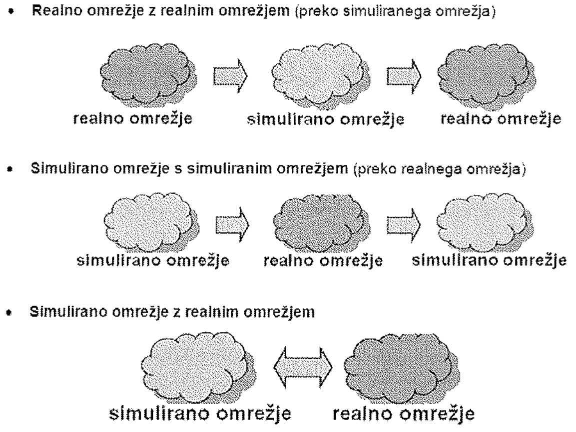
V fazi testiranj smo izvedli »real-sim-real« tip simulacij, kjer realne komunikacijske naprave komunicirajo preko simuliranega omrežja. Preko simulatorja (računalnik z nameščenim simulacijskim okoljem OPNET Modeler in dodatnim SITL modulom) smo povezali dva realna prenosna računalnika. Komunikacijo med priključenima prenosnikoma smo najprej testirali s pomočjo ICMP ping aplikacije, nato pa smo izvedli še prenos digitalnega govora (VoIP). Kot VoIP aplikacijo smo uporabili D-Voicer /12/, ki omogoča prenos digitalnega govora z uporabo širokopasovne internetne povezave.

V drugem delu tega članka smo predstavili zasnovo sistema za simulacijo taktičnih omrežij slovenske vojske. V tem primeru so preko simulatorja povezani reali taktični računalniki, z nameščeno specialno programsko opremo za sisteme poveljevanja in kontrole, kot je npr. replikacijski mehanizem IRM /1, 2/. V simulatorju je modelirano brezžično radijsko omrežje na virtualnem terenu, kjer je upoštevan vpliv terena na širjenje radijskih valov /13/. Takšen simulacijski sistem omogoča zelo realistično urjenje poveljniškega kadra.

Članek je sestavljen iz naslednjih poglavij. Drugo poglavje opisuje modul SITL. Tretje poglavje opisuje »real-sim-real« tip simulacije, kjer realne komunikacijske naprave komunicirajo preko simuliranega omrežja. Uporabnost tega koncepta smo prikazali na primeru prenosa digitalnega govora (VoIP) preko preprostega simuliranega omrežja. V četrtem poglavju je predstavljena zasnova trenažnega sistema poveljevanja in kontrole Slovenske vojske. V zaključku so podani predlogi za možnosti uporabe tovrstnih simulacij.

2. Modul SITL

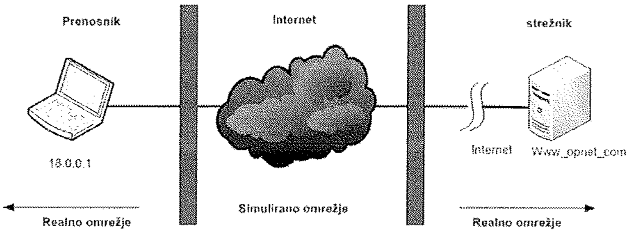
Simulacijo v realnem času z realnimi napravami v simulacijski zanki smo izvedli s pomočjo OPNET modula System-in-the-Loop (SITL) /9, 11/. Modul SITL omogoča povezavo in izmenjavo paketov med realnimi omrežnimi napravami in simuliranimi omrežnimi napravami v realnem času. Obstajajo tri tipične konfiguracije povezovanja zunanjimi omrežnimi napravami, ki jih prikazuje slika 1.



Slika 1: Možnosti povezav realnih in simuliranih naprav ali omrežij s pomočjo modula System-in-the-Loop /10/.

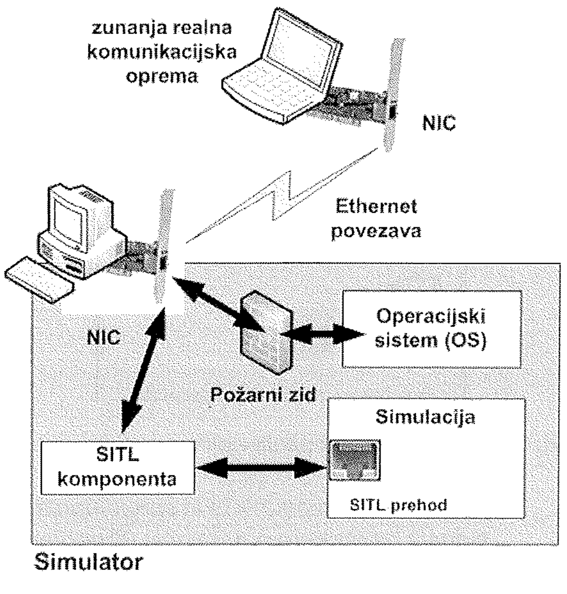
Najzanimivejši primer povezav predstavlja prvi tip simulacije (»real-sim-real«), kjer povežemo realne komunikacijske naprave preko simuliranega omrežja. Simulirano omrežje vpliva na realne pakete v obliki zakasnitev, izgub, podvojitev, itd. SITL omogoča, s pomočjo prevajalnih funkcij, povezavo med simulacijskim okoljem z OPNET Modelerejem ter Ethernet fizičnimi vmesniki gostujočega računalnika. S pomočjo prevajalnih funkcij se realni paketi pretvorijo v simulacijske pakete (in obratno). V primerih uporabe protokolov, ki s SITL niso podprti, pa lahko tudi sami napišemo svojo prevajalno funkcijo/11/. V verziji OPNET Modeler 14.5 so, v kombinaciji s SITL, podprti naslednji protokoli: Ethernet, IPv4 in IPv6, ICMP, ICMPv6, OSPFv2, RIPv1, RIPv2, TCP, UDP, okrnjen FTP.

V primeru tipa simulacij, kjer realna omrežja komunicirajo preko simuliranega omrežja, SITL praktično podpira vse potrebne protokole, saj simulirano omrežje uporabimo le kot transportno omrežje. Primer takega simulacijskega scenarija je prikazan na sliki 2, kjer realni klient dostopa do realnega strežnika preko simuliranega omrežja, predstavljenim z internetnim oblakom. Na SITL vmesniku, kjer je priključen na simulacijo realni prenosni računalnik, se paketi preoblikujejo v simulacijske, ti paketi potujejo preko simuliranega interneta, nato se pri drugem SITL vmesniki, ponovno pretvorijo v realne pakete, ki potujejo do realnega strežnika.



Slika 2: Primer uporabe »real-sim-real« tip simulacij, kjer realni gostitelj in strežnik komunicirata preko simuliranega omrežja v simulacijskem okolju OPNET.

Slika 3 prikazuje obravnavanje paketov na simulacijskem računalniku v primeru uporabe v SITL modula. Realne pakete, ki prispejo na Ethernet vmesnik (mrežne kartica NIC) in so namenjeni za simulacijo, blokira na poti do lokalnega operacijskega sistema požarni zid. Hkrati se ti paketi, s pomočjo gonilnikov WinPCap /14/ zajamejo, filtrirajo in preusmerijo preko SITL prehoda v simulirano omrežje.



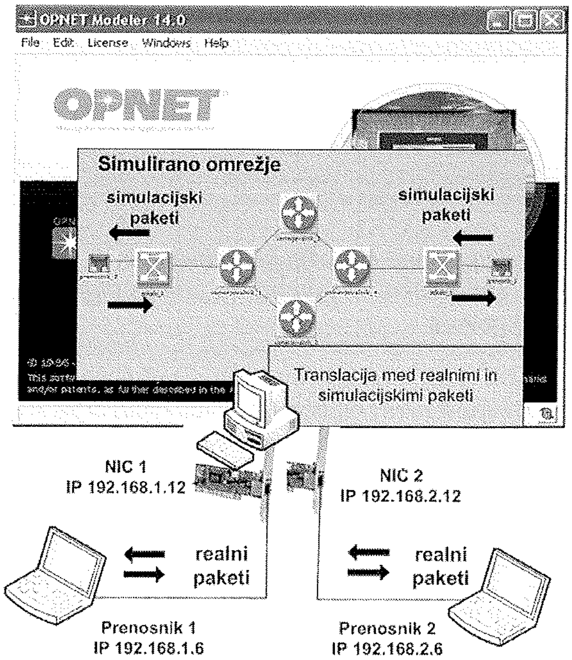
Slika 3: Blokovna shema simulacijskega sistema z nameščeno programsko opremo OPNET in SITL, kjer se izvaja simulacija v realnem času.

Modul SITL je primeren za uporabo v naslednjih primerih /10/:

- opazovanje vplivov simuliranega omrežja na realno omrežje in obratno.
- simuliranje vpliva obremenjevanja realnega omrežja s pomočjo simuliranega generatorja prometa.
- razvoj in preizkušanje prototipov novih protokolov in naprav z povezavo realnih naprav ali protokolov z simuliranimi.
- zunanje krmiljenje simuliranih naprav v realnem času
- testiranje zmogljivosti novih komunikacijskih protokolov.
- testiranje skalabilnosti s povečanjem števila realnih naprav v omrežju s pomočjo dodatnih simuliranih naprav
- itd.

3. Prenos realnega digitalnega govora preko simuliranega omrežja

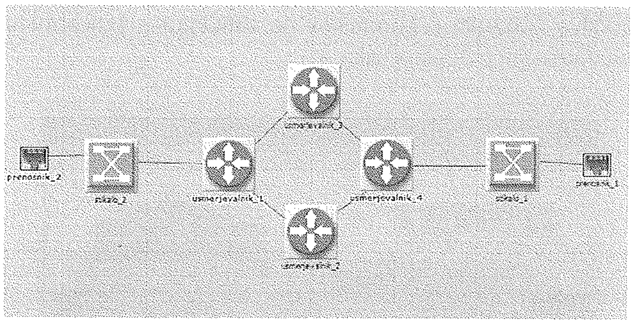
Metodologijo simulacij tipa »real-sim-real« bomo pokazali najprej na primeru prenosa digitalnega govora preko simulirane omrežne infrastrukture. Na simulacijski računalnik smo priključili dva prenosnika (Prenosnik 1 in Prenosnik 2) preko dveh mrežnih vmesnikov (NIC1 in NIC2). Princip priključitve zunanjih prenosnikov na simulator s potrebnimi nastavitvami prikazuje slika 4.



Slika 4: Povezava zunanjih realnih prenosnikov (Prenosnik 1 in Prenosnik 2) na simulator, kjer se izvaja simulacija komunikacijske infrastrukture.

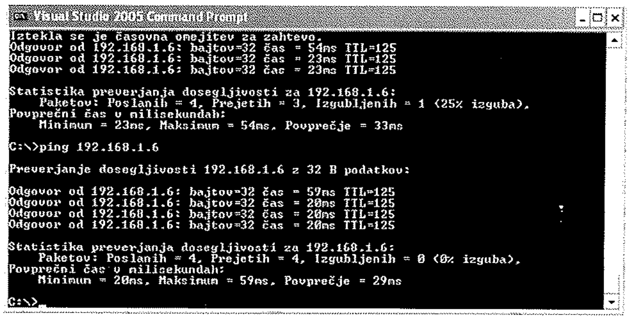
Konfiguracija z dvema mrežnima karticama na simulacijskem računalniku je nujna v primeru, ko so internetni naslovi obeh računalnikov iz istega podomrežja. V nasprotnem primeru bi zadostoval, na simulacijskem računalniku, en sam vmesnik, na katerega bi priključili oba prenosnika preko zvezdišča. S temi ukrepi zagotovimo, da med zunanjimi napravami ni direktne komunikacije.

Potem, ko smo povezali zunanje računalnike na simulator in nastavili komunikacijske parametre (IP naslovi, nastavitev požarnih zidov, itd.), smo se lotili še modeliranja simuliranega omrežja, preko katerega bo potekala komunikacija z realnimi zunanjimi računalniki. V simulacijskem okolju OPNET imamo na voljo posebno knjižnico s SITL gradniki kot sta SITL prehod (SITL_Gateway) in SITL povezavo (SITL_link). SITL prehod omogoča prehod paketov med simuliranimi in realnimi komunikacijskimi napravami. Na SITL prehodu lahko izbiramo prevajalne funkcijo za pretvorbo paketov ter definiramo filter za zajem paketov na mrežni kartici. S tem lahko določimo katere pakete bomo usmerili v simulacijo in katere blokirali. Pomembna je tudi definicija izvora IP paketov, kjer določimo, s katere mrežne kartice bodo zajeti realni paketi. SITL povezava se uporablja za povezavo med SITL prehodom in prvo simulacijsko napravo. Slika 5 prikazuje simulirano omrežje, ki smo ga uporabili v našem primeru za prenos digitalnega govora med realnima prenosnima računalnikoma (Laptop 1 in Laptop 2). Simulirano omrežje je sestavljeno iz dveh stikal ter štirih usmerjevalnikov.



Slika 5: Simulirano omrežje uporabljeno v simulacijskem primeru za prenos digitalnega govora.

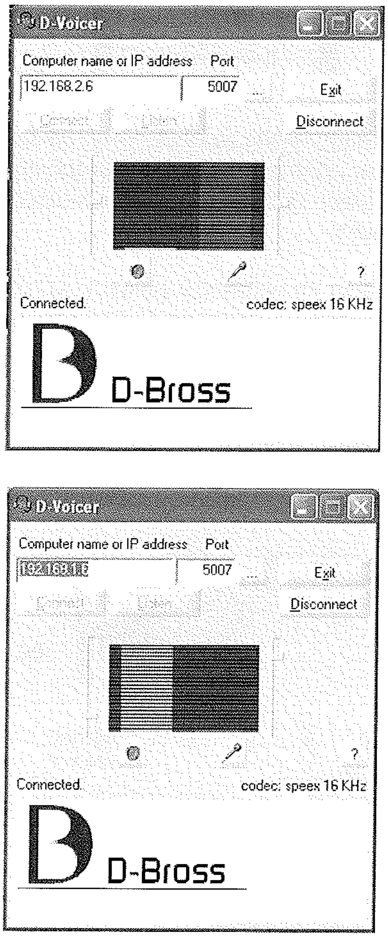
Povezljivost med prenosnikoma preko simuliranega omrežja smo najprej testirali s pomočjo ICMP aplikacije ping, ki smo jo izvedli na enem izmed obeh prenosnikov. Komunikacija med realnima računalnika je možna le v primeru, da je simulacija aktivna in so vsi komunikacijski parametri pravilno nastavljeni. Na sliki 6 je prikazan primer uspešnega testa povezljivosti.



```
Iztekla se je časovna omejitev za zahtevo.
Odgovor od 192.168.1.6: bajtov=32 čas = 54ms TTL=125
Odgovor od 192.168.1.6: bajtov=32 čas = 23ms TTL=125
Odgovor od 192.168.1.6: bajtov=32 čas = 23ms TTL=125

Statistika preverjanja dosegljivosti za 192.168.1.6:
    Paketov: Poslanih = 4, Prejetih = 3, Izgubljenih = 1 (25% izguba),
Povprečni čas v milisekundah:
    Minimum = 23ms, Maksimum = 54ms, Povprečje = 33ms

C:\>ping 192.168.1.6

Preverjanje dosegljivosti 192.168.1.6 z 32 B podatkov:

Odgovor od 192.168.1.6: bajtov=32 čas = 59ms TTL=125
Odgovor od 192.168.1.6: bajtov=32 čas = 20ms TTL=125
Odgovor od 192.168.1.6: bajtov=32 čas = 20ms TTL=125
Odgovor od 192.168.1.6: bajtov=32 čas = 20ms TTL=125

Statistika preverjanja dosegljivosti za 192.168.1.6:
    Paketov: Poslanih = 4, Prejetih = 4, Izgubljenih = 0 (0% izguba),
Povprečni čas v milisekundah:
    Minimum = 20ms, Maksimum = 59ms, Povprečje = 29ms

C:\>_
```

Slika 6: Preverjanje povezljivosti med realnima računalnikoma (Prenosnik 1 in prenosnik 2) povezanima preko simuliranega omrežja.

Ko smo vzpostavili povezavo med obema računalnikoma, smo izvedli še preizkus prenosa digitalnega govora preko simuliranega omrežja. Na prenosna računalnika (Prenosnik 1 in Prenosnik 2) smo namestili Voice over IP (VoIP) aplikacijo D-Voicer /12, 15/. D-Voicer je preprosta aplikacija, ki omogoča prenos govora z uporabo širokopasovne internetne povezave. Uporabniški vmesnik te aplikacije je prikazan na sliki 7. VoIP storitve pretvorijo analogni govor iz mikrofona v digitalni signal, katerega nato prenašamo po omrežju. Na drugi strani omrežja se digitalni signal ponovno pretvori v govor in reproducira na slušalkah.



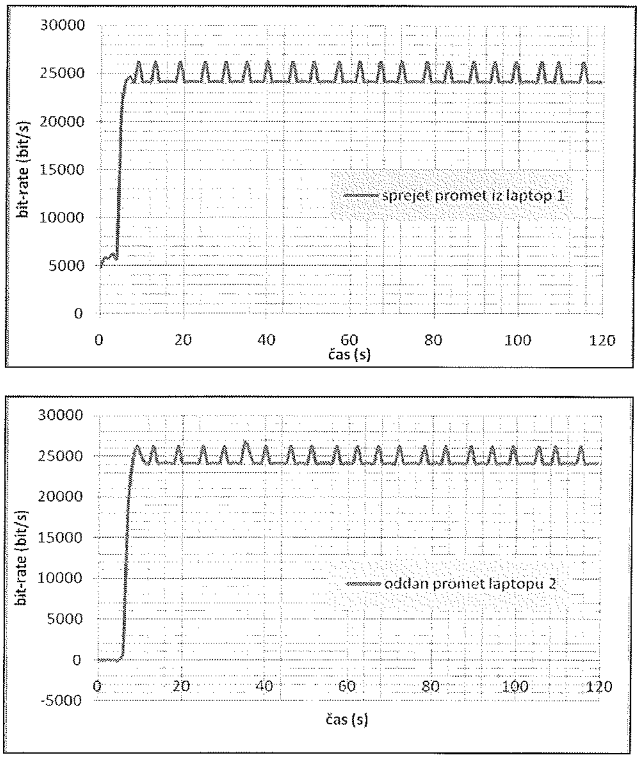
Computer name or IP address: 192.168.2.6
Port: 5007
Connected.
codec: speex 16 KHz

Computer name or IP address: [illegible]
Port: 5007
Connected.
codec: speex 16 KHz

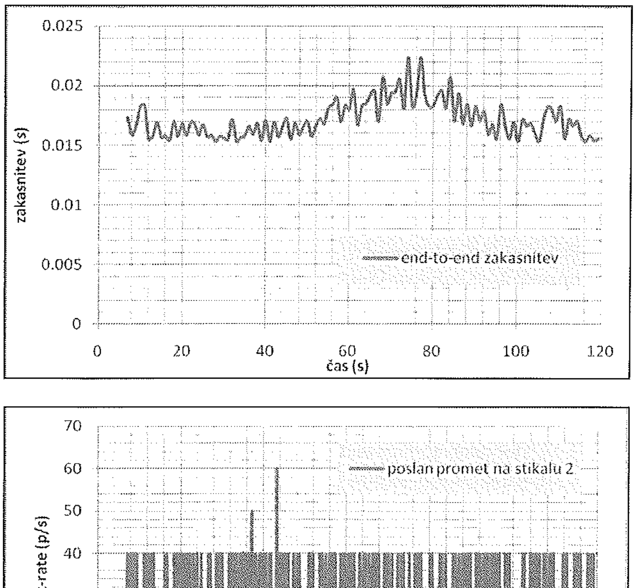
Slika 7: Uporabniški vmesnik aplikacije D-Voicer za primer oddajanje digitalnega govora na prenosniku (Prenosnik1) in sprejemanje tega signala na drugem prenosniku (Prenosnik 2) preko simuliranega omrežja.

Po izvedeni simulaciji lahko na podlagi zajetih statistik, v OPNET Modelerju, opravimo podrobno analizo rezultatov. Slika 8 prikazuje realni promet digitalnega govora izmerjenega na SITL prehodih, ki predstavljajo točko vstopa podatkov digitalnega govora v simulirano omrežje (zgornji graf) ter točko izstopa (spodnji graf).

Iz rezultatov simulacij vidimo, da je promet, ki ga oddaja prenosnika 1 enak tistemu, ki ga sprejema prenosnik 2, kar pomeni, da ni bilo izgubljenih podatkov. Slika 9 prikazuje izmerjene zakasnitev (zgornji graf), ki nastajajo v simuliranem omrežju pri prenosu digitalnega govora, ter promet izražen v paketih/s, ki smo ga izmerili na simuliranem stikalu 2.



Slika 8: *Promet digitalnega govora v bit/s izmerjen na SITL prehodih pri vhodu in izhodu iz simuliranega omrežja.*



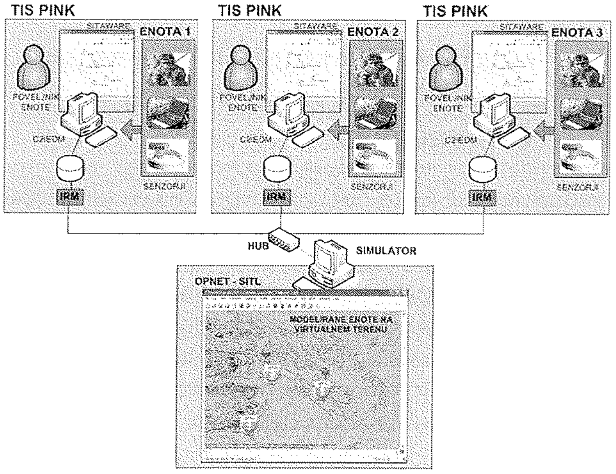
Slika 9: *Prvi graf prikazuje zakasnitve digitalnega govora pri prenosu preko simuliranega omrežja. Drugi graf prikazuje poslan promet v paketih/s preko drugega stikala v simuliranem omrežju.*

Iz rezultatov na sliki 9 vidimo, da so zakasnitve v simuliranem omrežju pri prenosu digitalnega govora minimalne. To je tudi za pričakovati, saj so bile uporabljene simulirane povezave s kapaciteto 100Mb/s minimalno obremenjene.

4. Načrtovanje lastnosti sistema za simulacije taktičnih omrežij v realnem času

V zadnjih letih se je vse bolj pojavila potreba po simulacijah v realnem času z realno opremo v simulacijski zanki. Ta trend je še posebej izražen na vojaškem področju. Eden pravih eksperimentov na tem področju je »Battlespace communication network planner and simulator« /5, 6/, kjer je bila razvita namenska programska in strojna komunikacijska oprema, ki je omogočila povezavo med realnim in simuliranim komunikacijskimi napravami.

Koncept, ki ga bomo predstavili tukaj, temelji na programski rešitvi, kjer smo namesto posebej razvite strojne opreme uporabili v prejšnjih poglavjih predstavljen OPNET modul SITL. Na osnovi tega koncepta želimo zasnovati trenažer taktičnih informacijskih sistemov poveljevanja in kontrole (TISPINK) (C2IS - Command and Control Information Systems), ki bo omogočal zelo realistično urjenje poveljniškega kadra. V takšnem sistemu bi povezali realne taktične računalnike, na katerih je nameščena TISPINK programska oprema (IRM in Sitaware), s simuliranim taktičnim radijskim omrežjem na virtualnem terenu. Na sliki 10 je predstavljena zasnova takšnega simulacijskega sistema. Vsaka taktična enota ima vsaj en taktični računalnik, ki se povezuje z ostalimi enotami preko TISPINK sistema. V trenažnem sistemu smo vse taktične računalnike povezali v Ethernet lokalno omrežje preko zvezdišča. Osnovo TISPINK programske opreme slovenske vojske predstavljata grafični vmesnik Sitaware in replikacijski mehanizem IRM /1, 2/, ki skrbi za izmenjavo podatkov med podatkovnimi bazami enot. Obe programski opremi sta produkta Danskega proizvajalca Systematic /16/. V isto lokalno omrežje povežemo tudi računalnik, na katerem je nameščena simulacijska programska oprema OPNET Modeler z dodatnim modulom SITL. V simulacijskem okolju OPNET na virtualen teren postavimo komunikacijske modele taktičnih enot (za vsak taktični računalnik po en model). Taktične radijske postaje so modelirane z brezžičnimi usmerjevalniki, katerim lahko nastavljamo parametre, kot so oddajna moč, tip antene, modulacija, frekvenca kanala, kapaciteta kanal, itd. Trenažni sistem temelji na že predstavljenem tipu »real-sim-real« SITL simulacij. V takšnem trenažnem sistemu se bodo realni TISPINK podatki replicirali preko simuliranega brezžičnega omrežja v simulatorju. Simulator bo zelo realistično vplival na realne pakete v obliki zakasnitev, izgube podatkov zaradi radijskih motilnikov (jamerjev), izgube radijske vidljivosti, itd. TISPINK operaterji bodo komunikacijske probleme zaznali na enak način, kot da bi bili na realnem terenu z realno komunikacijsko opremo. Takšen sistem omogoča, zraven realističnega treninga poveljnikov in TISPINK operaterjev, tudi kakovostno načrtovanje radijskih omrežij in razvoj taktičnih strategij, saj je mogoče že v laboratoriju predvideti velik del komunikacijskih problemov, ki bi se drugače pokazali šele na terenu.



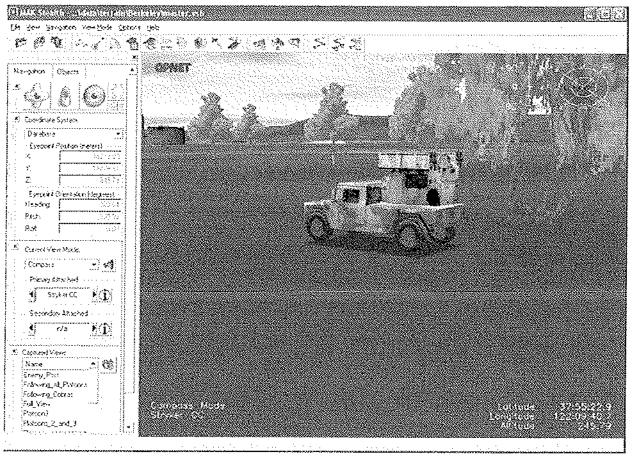
Slika 10: Trenažer taktičnih informacijskih sistemov poveljevanja in kontrole.

Za zagotovitev takšnih simulacij so potrebni zraven OPNET Modelerja še naslednji moduli:

- TMM (Terrain Modelling Module) je modul, ki omogoča upoštevanje vpliva terena na širjenje radijskih valov. Virtualni teren je modeliran v obliki višinskih zemljevidov formata DTED (Digital Terrain Elevation Data) in USGS (U.S. Geological Survey).
- Paket za simulacijo brezžičnih omrežij (Wireless Suite) omogoča simulacijo brezžičnih omrežij različnih tipov in tehnologij kot so MANET, 802.11, 3G/4G, Ultra Wide Band, WiMAX, Bluetooth, Satelitski linki, itd. /11, 17, 18/. Imamo na voljo tudi več različnih modelov širjenja radijskih valov kot so Free-space/19/, Longley-Rice/20, 21/ ali pa model TIREM /22/, ki ga uporablja tudi ameriška vojska.
- 3D vizualizacija omrežij (3DNV) /9, 11/ omogoča prikazovanje 3D animacij simuliranih omrežij na virtualnem terenu. Takšna vizualizacija je uporabna predvsem za ponazoritev vpliva terena na brezžično komunikacijo ali vpliv premikanja mobilnih enot na brezžično komunikacijo. 3DNV omogoča ponazoritev komunikacijskih enot z 3D modeli vojaških vozil, ljudi, letal, vojaških baz, ... na v 3D modeliranem terenu. Slika 11 prikazuje primer ponazoritve taktične radijske postaje z vojaškim vozilom.

5. Zaključek

V članku smo predstavili nov princip simulacije komunikacijskih omrežij in sicer simulacije v realnem času z realno komunikacijsko opremo v simulacijski zanki. V našem primeru smo za tovrstne simulacije uporabili modul SITL programskega paketa OPNET Modeler. V prvem delu članka smo tovrstne simulacije prikazali na primeru prenosa realnega digitalnega govora (VoIP) preko preprostega simuliranega omrežja. Priloženi simulacijski rezultati, za oddani in sprejet promet, zakasnitve itd., nakazujejo nekaj možnosti, ki ji takšne simulacije nudijo. V zadnjem delu članka smo predstavili zasnovo trenažerja taktičnih informacijskih sistemov poveljevanja in kontrole (TISPINK), ki je zasnovan na enakem konceptu, ki smo ga pokazali na primeru prenosa govora. V tem primeru bodo realni taktični računalniki s TISPINK programsko opremo, komunicirali preko simuliranega radijskega omrežja na virtualnem terenu. Takšen sistem bo omogočal realističen trening poveljnikom enot, po drugi strani pa bo omogočal tudi kakovostno načrtovanje taktičnih komunikacijskih omrežij in taktičnih vojaških aktivnosti.



Slika 11: Realistična 3D vizualizacija simulirane komunikacijske opreme na virtualnem terenu.

Uporabnost tovrstnih simulacij je mogoče najti tudi na civilnih področjih komunikacij, kot na primer pri razvoju in testiranju novih komunikacijskih protokolov in naprav (na primer za področje VoIP).

Reference

/1/ J. Mohorko, M. Fras and Ž. Čučej, Modeling of IRIS Replication Mechanism in a Tactical Communication network, using OPNET, Communication Network (accepted but not published yet).

/2/ J. Mohorko, M. Fras and Ž. Čučej: "Modeling of IRIS replication mechanism in tactical communication network with OPNET", OPNETWORK 2007 - the eleventh annual OPNET technology Conference, August 27th-31st, Washington, D.C., 2007.

/3/ Jože Mohorko, Saša Klampfer, Žarko Čučej, Ekspertni sistem za analizo rezultatov simulacij taktičnih radijskih omrežij, Inf. MIDEM, 2008, ????

/4/ J. Mohorko, S. Klampfer, "A system for expert analysis of tactical networks' performances", SoftCOM 2008, Sep-tember 25-27, 2008, Split - Dubrovnik, Croatia.

/5/ L. B. H. Ernest and L. C. W. Clement, BattleSpace Communications Network Planner and Simulator (BCNPS), OPNETWORK 2004 - Aug 30th - Sept 3rd - Washington, D.C.

/6/ L. J. Phang, A. Y. P. Chua, T. H. Beck and E. B. H. Lee, Battle Space Communications Network Planner and Simulator (BCNPS), Proceedings to 10th International Command and Control

Research and Technology Symposium: The Future of Command and Control, June 13-16, McLean, VA. Command and Control Research Program (CCRP), Washington, D.C, 2005.

/7/ M. Fras, J. Mohorko, Ž. Čučej, Analiza, modeliranje in simulacija vpliva prometa aplikacij za izmenjavo datotek P2P na zmogljivost omrežij, Inf. MIDEM, 2008, ????.

/8/ M. Fras, J. Mohorko, Ž. Čucej, "Estimating the parameters of measured self similar traffic for modeling in OPNET", in: Proceedings of IWSSIP2007 Conference, Maribor, Slovenia, 2007.

/9/ M. Fras, G. Globačnik in J. Mohorko, Advanced method of network simulations with OPNET Modeler, Zbornik radova, Novi Sad, Serbia, 2008, str. 215-218.

/10/ OPNET Technologies, Inc, Session 1933: Developmental and Interoperability Testing with OPNET System-in-the-Loop, Proceedings, OPNETWORK 2007 - the eleventh annual OPNET technology Conference, August 27th-31st, Washington, D.C., 2007.

/11/ OPNET 14.5 PL5 dokumentacija.

/12/ http://www.d-voicer.com/.

/13/ M. Fras, J. Mohorko, Ž. Čučej, Simuliranje taktičnih omrežij z upoštevanjem modela širjenja radijskih valov. ERK 2007, 24. - 26. September 2007, Portorož, Slovenija.

/14/ http://www.winpcap.org/.

/15/ T. Wallingford, Switching to VoIP , O'Reilly Media, Inc., CA, USA, 2005.

/16/ http://www.systematic.dk/.

/17/ B. Vujičić, Modeling and Characterization of Traffic in Public Safety Wireless Networks, Master of Applied science, Simon Fraser University, Vancouver, Canada, 2006.

/18/ M. Z. Jiang, Analysis of wireless data network traffic, Master of Applied Science, Simon Fraser University, Vancouver, Canada, 2000.

/19/ S. R. Robinson and P. S. Idell, Free-space propagation model for coherence-separable broadband optical fields, Optical Society of America, Journal, vol. 70, Apr. 1980, p. 432-437.

/20/ K. A. Chamberlin and R. J. Luebbers, An Evaluation of Longley-Rice and GTD Propagation Models, IEEE Transactions on Antennas and Propagation, vol. AP-30, Nov. 1982, p. 1093-1098.

/21/ M. M. Weiner, Use of the Longley-Rice and Johnson-Gierhart Tropospheric Radio Propagation Programs: 0.02-20 GHz, IEEE Journal on Selected Areas in Communications, VOL. SAC-4, No. 2, March 1986.

/22/ D. Eppink and W. Kuebler, TIREM/SEM Handbook, Department of defense, Electromagnetic compatibility analysis center, Annapolis, Maryland, 1995.

/23/ G. Globačnik, J. Mohorko, Ž. Čučej, Result visualization in tactical network simulations, International Conference on Software, Telecommunications and Computer Networks -SoftCOM 2008, September 25-27, 2008 Split - Dubrovnik, Croatia.

Matjaž Fras, Jože Mohorko
Fakulteta za elektrotehniko,
računalništvo in informatiko
Univerza v Mariboru, Smetanova 17,
2000 Maribor, Slovenija
Tel: (+386 2) 220-7120 Email: matjaz.fras1@uni-mb.si

Prispelo (Arrived): 29.10.2008 *Sprejeto (Accepted): 09.06.2009*

UDK621.3:(53+54+621+66), ISSN0352-9045

Informacije MIDEM 39(2009)2, Ljubljana

IP PACKET QUEUING DISCIPLINES AS BASIC PART OF QOS ASSURANCE WITHIN THE NETWORK

Saša Klampfer, Jože Mohorko, Žarko Čučej

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

Key words: waiting queues, queuing mechanisms, delay, dropped traffic, queuing, quality of service

Abstract: This paper represents results of study how different queuing methods influences on service quality as basic part of QoS system in wired networks. We briefly introduces six queuing methods, where two of them are high sophisticated (WFQ - Weighted Fair Queuing and CBWFQ - Class Based Weighted Fair Queuing) used in Cisco's routers, the third one (CQ - Custom Queuing) is used as basic method for comparison with all others sophisticated methods. All mentioned queuing principles are tested in OPNET Modeler simulation tool. During many simulation runs we get the best results with combined PQ-CBWFQ method which is also presenter of a low latency queuing. The satisfying results are also obtained with CBWFQ method where high priority traffic undisturbed travel through the network, but majority low-priority traffic must wait. This is also the main reason why we obtain highest delays within network in case of CBWFQ method. When are present only two different traffic flows, WFQ is the best choice concerning delays and dropped traffic. All other queuing disciplines don't fulfill our expectations.

Mehanizmi paketnega uvrščanja kot osnovni pogoj za zagotavljanje kvalitete storitev v omrežju

Kjučne besede: čakalne vrste, mehanizmi uvrščanja, zakasnitve, zavržen promet, uvrščanje, kvaliteta storitev

Izvleček: V članku predstavljamo rezultate študije, ki se nanaša na opazovanje vpliva posameznih metod uvrščanja paketov v čakalne vrste na kakovost storitev v omrežjih. Mehanizmi uvrščanja paketov v čakalne vrste ob enem tudi predstavljajo ključen del sistema QoS. Na kratko predstavljamo šest različnih metod uvrščanja, med katerimi sta dve visoko sofisticirani (WFQ - utežnostno pravično uvršanje in CBWFQ - utežnostno pravično uvrščanje na osnovi razredov) uporabljeni tudi v Cisco-vih usmerjevalnikih. CQ (ang. Custom Queuing) je uporabljena kot osnovna metoda uvrščanja in nam v zaključni fazi predstavlja referenco pri primerjavi rezultatov naprednejših mehanizmov uvrščanja z osnovnimi načini uvrščanja, kamor spada CQ. Vse čakalne mehanizme smo preizkusili v simulacijskem orodju OPNET Modeler. Tekom številnih simulacij smo najboljše rezultate prejeli s kombinirano metodo PQ-CBWFQ, ki je ob enem predstavnik čakalnih vrst z najnižjimi zakasnitvami. Zadovoljive rezultate smo prav tako prejeli z CBWFQ metodo, kjer visoko-prioritetni podatkovni tokovi nemotene tečejo skozi omrežje, vendar pri tem morajo nižje-prioritetni tokovi, ki predstavljajo glavnino vsega prometa, čakati. To je ob enem tudi glavni vzrok za dobljene visoke zakasnitve znotraj omrežja v primeru uporabe CBWFQ načina uvrščanja. Študija je pokazala, da v primeru prisotnosti samo dveh različnih podatkovnih tokov, zadovolji naše potrebe že WFQ metoda uvrščanja, predvsem v smislu zavrženega prometa. Vse ostale metode uvrščanja naših pričakovanj niso izpolnile.

1. Introduction

Demand for internet quality of service (QoS) in wired networks has been growing rapidly over recent years, especially in networks of large companies and organizations, with limited wired link bandwidth, where are in use many different applications, such as VoIP, video on demand, video conferencing, FTP sessions, web browsing, E-mail etc. Such organizations often have internet link which also serves hundreds or even thousands of internal network computers, named nodes. In those cases we must be able to provide the sufficient bandwidth and low latency for high priority applications, for example VoIP. Without distinguishing between the different priority applications we cannot be able to provide service quality on the desired level. Recent, strong growth of bandwidth hungry peer-to-peer applications has made this task especially difficult and complex. One of the important parts of QoS service are also queuing methods which manages traffic flows in the normal operation conditions and also in the situations when the network becomes overloaded. Our goal, in this research, is to explore how each of mentioned queuing methods influences on dropped packets and delays in network when we use high priority applications in cohabitation with all other low level priority applications. The effectiveness of each queuing method will be introduced in continuation. Queuing plays an important part within whole QoS system. This area presents issue for many research processes. For example, /1/ describes influence of different queuing methods on the common utilization network characteristic, meanwhile /7/ represents providing QoS system for multimodal systems in wired networks. As wired networks, are in last few years also very important wireless networks where is also need upon quality of service. Such solutions are described in /6/. Congestion control with use of queuing mechanisms for Byte-Stream networks is described in /10/. Regarding well spread VoIP traffic within wired and also wireless networks, are many research procedures regarding QoS concentrated into already mentioned area.

Mostly used queuing methods are detail described in second and third section, meanwhile fourth section represents OPNET modeler simulation tool, which is used for proving our assumptions, and decisions. Fifth section gives a description of wired network simulation structure, in sixth section are presented simulation results meanwhile seventh section gives a short conclusion with decisions and ascertainments, afterwards follow directions and points of departure for further work and researches in this area.

2. Queue

Wide-area networks (WAN) consist from wide segments where each of them is again combination of smaller segments - local-area networks (LAN). Similar topology is at local area networks level, where each segment is linked to network stations using basic network components such as routers, switches, bridges. Each of mentioned elements contain queuing mechanisms which carries each individual arrived packet according to priority number, entered into the type of service (ToS) field of IP packet header. Most sophisticated solutions on the market contain more than one queues and each of them is intend to provide place for different traffic priority flows, respectively each of them belongs to specific priority class.

Firstly, the queuing mechanisms will be explained on the simplest example of single queue, which is known as FIFO (first in, first out). FIFO queue is data structure which acts as settled list. On one side data comes into such list, and on the other side they come out in the same sequence as they came in. Operations which are executed upon waiting queue are: element insertion, element removal, which element is placed at the start place of queue, occupies places and free places in the queue /1/.

3. Queuing mechanisms

Traffic link congestions are today one of the most bothering things because they affect on quality of service of other applications presented in network. One way for handling congestions, represents queuing algorithms, which sorts traffic according with queuing methods and predefined priority levels applied by ToS field. Cisco is one of the most renowned companies for network equipment. Their networking algorithms /8/ contain the following queuing mechanisms which are capable to handle with congestions in network:

- FIFO queuing
- Priority queuing (PQ)
- Custom queuing (CQ)
- Weighted Fair Queuing (WFQ)
- Class Based Weighted Fair Queuing (CBWFQ)
- Multiclass weighted round robin (MWRR)
- Deficit weighted round robin (DWRR)

Each algorithm is designed to solve the specific problem concerned with the network traffic. Each solution can affect on network performances in sense of link throughput, delays, dropped traffic /1, 2/ etc.

Custom queuing (CQ) task is to share link bandwidth between active applications in the network respectively between the organizations. In such circumstances the bandwidth must be divided proportionally between applications and end users in sense that congestions cannot appear. CQ mechanism provides satisfactory bandwidth at congestion point and at the same time protects specific traffic with constant bandwidth amount, meanwhile remains bandwidth CQ leaves to other active data traffic in the network. Traffic is managed with assigning weight amount respectively to place in queue for each class of packets. In that case we have 17 waiting queues - classes or more which are mutual equivalent. In each waiting queue class has annotation which tells us, how much bandwidth should be provided for transferring data into the output link. Afterwards, algorithm arrange messages into one of seventeen available queues, where queue with index 0 is intended for storing system messages, such as keepalive messages and many other signal messages. Queue empty procedure executes upon principle weight largeness, obtained from priorities, what means, that message with high priority amount has a largest weight comparison with message with lover priority amount. Router manage waiting queues from index one to index 16 in round robin manner. Figure 1 show custom queuing mechanism with up to seventeen waiting queues served in round robin manner.

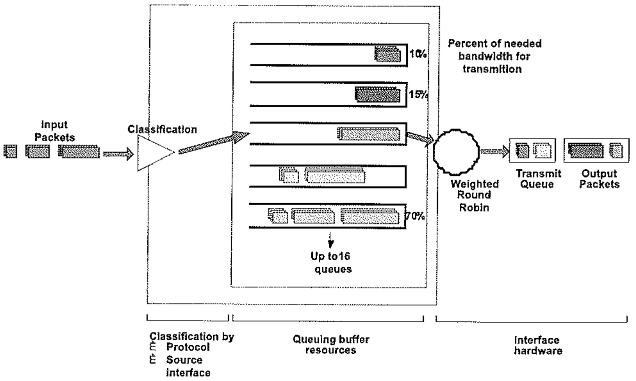


Fig. 1. Custom cueuing (CQ) discipline

CQ is statically configured /11/ as many other mechanisms, such as PQ etc. Statically means that queuing mechanism is not capable automatically change router settings regarding to dynamically changes within the network. This is the reason why CQ is not capable to automatically accommodate to dynamical changes in network /4/.

Weighted fair queuing (WFQ). In situations where we want to provide constant response time and keep delays in required range without assigning excessive bandwidth, is optimal solution weighted fair queuing. WFQ is algorithm/mechanism which introduces bit-wise fairness, and it allows, that is each queue fair served. Fairness is provided using mechanism which counts bytes. For the simplest example we observe two queues of the same length. First of them contains 100 packets and second one contains only 50 packets. In that case will WFQ work on the following manner; firstly it will take two packets from first queue then it takes one packet from second queue and then again two from first and one from second queue etc.

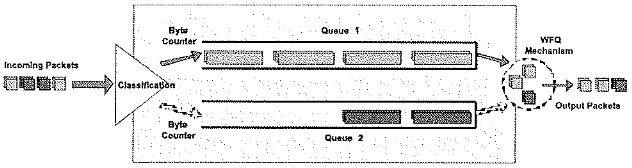


Fig. 2. Weighted fair queuing (WFQ) discipline

Such algorithm makes possible service fairness for each participating queue. Low level priority traffic flows undisturbed travel through network, what is a good compromise for each participating traffic flow in the network. WFQ also have other benefits such as minimizing configuring costs, because it is capable automatically accommodate to the dynamic network changes. Because of its good properties is in use on the majority of the serial E1/T1 interfaces. Weight is, also in this case, defined with IP priority amount (ToS field of IP packet). For IP priorities are in use settings in range from 0 (best effort) to 5 (IP quality speech), meanwhile settings 6 and 7 are reserved. Algorithm then uses these data to figure out, how many additional services it must provide/predict for each individual queue. It is capable to use each of available bandwidth for low priority traffic flows even if is presented high-priority traffic flow which don't use all bandwidth. In basis this principle is opposite comparison with time-division multiplexing (TDM), which reserves all available bandwidth and lives it unused if traffic flow is not presented. WFQ operates with IP priority settings as well with resource reservation protocol (RSVP) and also is capable to manage round trip delay problem. Such queuing clearly improves algorithms such as SNA, logical link control (LCC) and transmission control protocol (TCP), and at the same time it is capable accelerating flows and removing congestions in the network. Results became more predictable during whole routing path, meanwhile delays can be multiple reduced/decreased comparison with other queuing disciplines (CQ, PQ, FIFO) /8/.

Multiclass weighted round robin (MWRR) bases on a best effort connection scheduling discipline. It represents some kind of simplest emulation of generalized processor sharing discipline. Comparison, while generalized processor sharing (GPS) /12/ serves infinitesimal amount of data from each non empty connection, MWRR serves a number of packets for each connection which is not empty, where is this number (denoted with *x*) defined with the following equation

$$x = normalised\left(\frac{Weight}{Mean\_Packet\_Size}\right) \qquad (1)$$

That the MWRR emulation of generalized processor sharing discipline can be correctly in cases, when packets can take different sizes, WRR server must know a source's mean packet size in advance. In practice, source's mean packet size can be unpredictable, and this is the reason why MWRR server doesn't predict/allocate bandwidth fairly. Additionally there is also the problem that algorithm is fair only over time scales longer than a round time. If the time scale is shorter than round time then some connections may get more services than others. If a connection has a small weight or respectively if number of connections is large, then these scenario leads to long periods of unfairness /3, 5/.

Deficit weighted round robin discipline (DWRR) is some kind of a modified ordinary weighted round robin scheduling. It is capable handling variable size packets without knowing their main size in advance. A DWRR scheduler associates each connection with a deficit counter initialized to 0. Such scheduler visits each connection in turn and tries to serve one quantum of bits from each visited connection. For example, the packet at the head of the queue is served only, if it isn't larger than the quantum size. Opposite, if it is larger, the quantum is added to the connections deficit counter. If the scheduler visits a connection such that the sum of the connection's deficit counter and the quantum is larger than or equal to the size of the packet at the head of the queue, then the packet at the head of the queue is served and the deficit counter is reduced by a packet size.

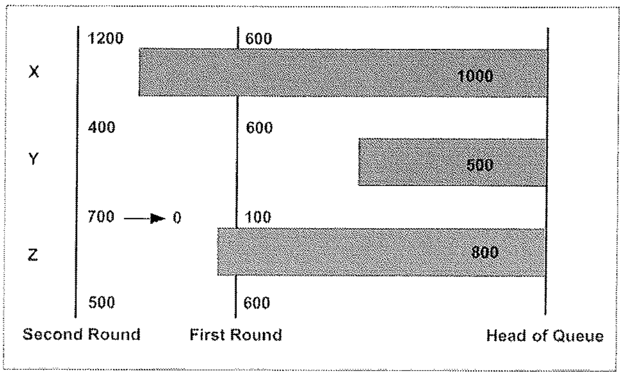


Fig. 3. Deficit weighted round robin queuing discipline

In the above figure one packet from each of connections X, Y and Z is queued at the DWRR scheduler. The quantum size is 600 units. In the first round Y's packet is served, and deficit counters for X and Z increase to 600. In the second round, X's packet is served and its deficit counter is set to 400. Z's packet is also served and its deficit counter is set to 500. Y's counter is cleared, because it has no packets to send /3, 6, 7/.

Class based weighted fair queuing (CBWFQ) represents the newest scheduling mechanism intended for handling with congestions' and at the same time provides better flexibility. We can use it when want to provide minimal required amount of bandwidth to specific application (in our case VoIP). For that case, network administrator must provide classes with predefined bandwidth amounts, where one class is for example intended for video conferencing application, another to VoIP application and so on. Instead of waiting queue assurance for each individual traffic flow, CBWFQ determines different traffic flows. For each such class is assured minimal bandwidth. We take a look at example, where could majority lower priority traffic flow take off an important part of bandwidth form highest priority traffic flow in case of WFQ use. Typical example is video transmission, which needs half of T1 connection bandwidth /8/. Sufficient link bandwidth can be assured using WFQ mechanism, but only when present are only two traffic data flows. In cases when more than two traffic flows appear, video session will suffers for bandwidth because WFQ mechanism works on fairness principle. For example, if nine additional traffic flows make demands upon same bandwidth, video session gets only 1/10 of whole bandwidth, and that is not enough. Even, if we put IP priority level 5 into ToS field of IP packet header, circumstances will not change. In that case, video conference gets only 6/15 of bandwidth, and that is still not enough, because mechanism must provide half of all available bandwidth on T1 link. With use of CBWFQ mechanism we can assure required bandwidth. Network administrator just needs to define for example video conferencing class and install video session into that class, the same principal can be used for all other applications which need specific amount of bandwidth. Such classes are served with flow based WFQ algorithm, which allocates remain bandwidth to other active applications in the network.

4. Opnet modeler simulation tool

OPNET Modeler is a leading simulation tool used for modeling, simulations and analysis of communication networks, devices, protocols and applications. Modeler is a graphically oriented simulation tool, which uses project, node and process editors for building communication models. The project editor offers graphic topological description, whilst in the meantime the node editor is used for describing protocols. The process editor is an upgrade of C language, which uses a finite-state machine for algorithm and protocol descriptions. OPNET includes many standard communication models for constructing wired, radio, optical and satellite communication structures. Tool offers also 2D and 3D animation to accompany changes in the network structure /9/.

5. Simulation scenario for queuing disciplines evaluation on VoIP application

Test simulation network architecture is a copy of the real private network which belongs to private company. Our main goal is oriented to improve network performances.

The highest level on Figure 4 represents network server architecture offered from internet service provider (ISP). Servers' subnet consists of five Intel servers where each of them offers own profile, such as; web profile (web server), VoIP, E-mail, FTP and video profile. These servers are connected over 16 port switch through wired link onto private company's router. Company network consists of four LAN segments including different kind of users. On the west wing of the company are VoIP users which represent technical support to their customers. At south wing of company is placed conferencing room where employees make meetings with other outland partners. Here are two places intended for 2 simultaneous sessions. At north wing is a smallest office with only 10 employees which presents the development part of company, and they uses different applications needed for their work. For example, they searching information on the web; call their suppliers, mutual exchange files over FTP and so on. Remain west wing includes fifty disloyal employees which surfs on the net (web) during work time, downloading files etc (heavy browsing).

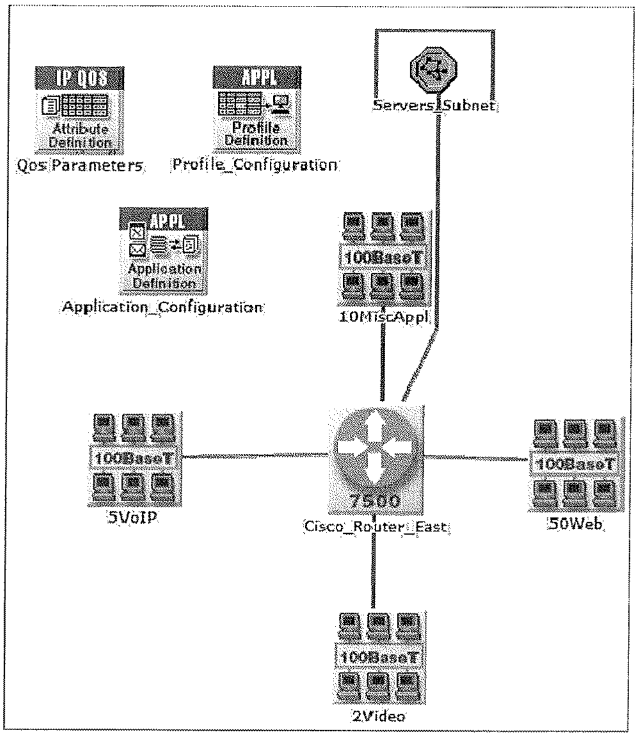


Fig. 4. *Tested wired network arhitecture which is a copy of a real network*

The highest level on Figure 4 represents network server architecture offered from internet service provider (ISP). Servers' subnet consists of five Intel servers where each of them offers own profile, such as; web profile (web server), VoIP, E-mail, FTP and video profile. These servers are connected over 16 port switch through wired link onto private company's router. Company network consists of four LAN segments including different kind of users. On the west wing of the company are VoIP users which represent technical support to their customers. At south wing of company is placed conferencing room where employees make meetings with other outland partners. Here are two places intended for 2 simultaneous sessions. At north wing is a

smallest office with only 10 employees which presents the development part of company, and they uses different applications needed for their work. For example, they searching information on the web; call their suppliers, mutual exchange files over FTP and so on. Remain west wing includes fifty disloyal employees which surfs on the net (web) during work time, downloading files etc (heavy browsing).

Each wing of company is connected over 100BaseT link to Cisco 7500 router. Further, this router is connected with ISP servers switch over a wired (VDSL2) 10Mbit/s ISPs' link. Connections between servers and switch are also 100BaseT. Wired link in this case represents bottleneck, so for that case we have to involve QoS system and their different queuing disciplines.

Table 1. Used applications and number of their clients

Application	Number of users
Heavy web browsing	50 clients
FTP	4 clients
Video conferencing	7 clients
VoIP	5 clients
E-mail	1 client

We create six scenarios; in the first scenario we tested custom queuing discipline which represents a basic for comparison with WFQ (second), MWRR (third), DWRR (fourth), CBWFQ (fifth) and with combined hybrid PQ-CBWFQ (sixth scenario) queuing disciplines. Network topology in all scenarios remain the same, differences are only in used queuing disciplines. By comparison of simulations' results for different scenarios we tried to prove, how each of queuing discipline serves used network applications.

6. Simulation results

As we mentioned before, we collected delay statistics from six different queuing disciplines scenarios (CQ, WFQ, MWRR, DWRR, CBWFQ and PQ-CBWFQ) for two different active applications (VoIP and HTTP) in the network and with different applied priorities by ToS field of IP packet header. For that case we defined VoIP traffic flows between clients where such flows represents high priority traffic, meanwhile HTTP traffic represents low priority flow, based on best effort type of service. In our scenarios we have 82.09% users which use lower priority HTTP traffic and only 17.91% users which use high priority VoIP application. Considering on Figure 5, means, that only 17.91% of users takes a whole majority part of bandwidth, so lowest priority HTTP traffic which represents majority of all traffic must wait. Because such majority traffic must wait, delays rapidly increase what is clearly seen on a Figure 5. Evident is, that VoIP traffic has lowest delays in comparison with HTTP traffic. Best results we obtain with custom queuing method which ensures required bandwidth on possible congestion points and serves all traffic fairly. After CQ queuing scheme follows WFQ, DWRR, MWRR, CBWFQ and PQ-CBWFQ. WFQ, DWRR, MWRR and CBWFQ reaches worse delay results, because of fairness queuing discipline. Similar results are obtained also in HTTP case. IF CBWFQ scheme is in use then high-priority traffic will be ensured with fix amount of bandwidth defined from administrator side. For example, administrators, using CBWFQ, define 9Mbit/s for VoIP, so in that case remains only 1Mbit/s for all other majority applications, so majority low level traffic will affect on rapidly delay increasing, as is shown on figure 6.

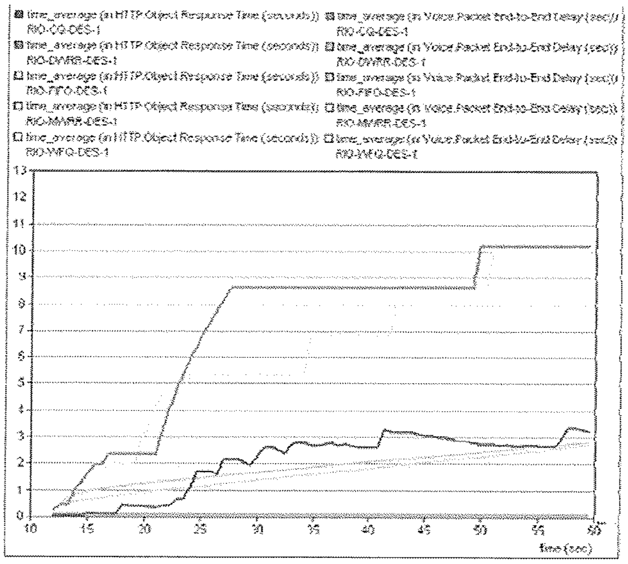


Fig. 5. Used different queuing disciplines upon VoIP and HTTP traffic

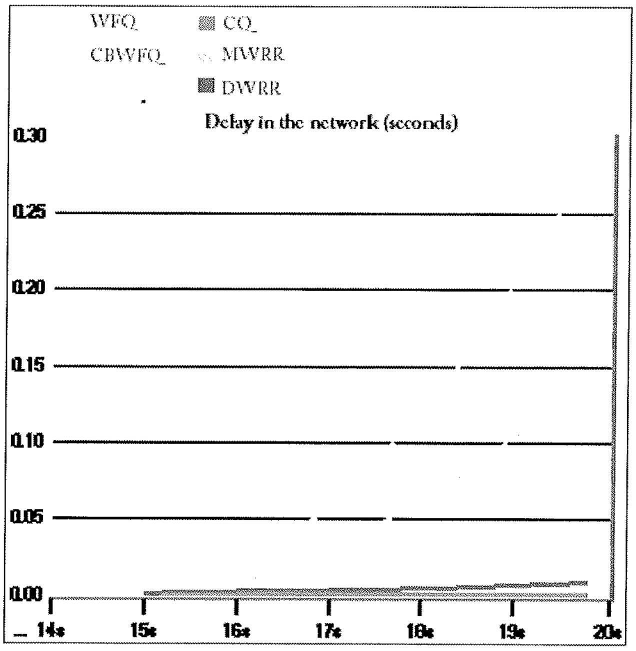


Fig. 6. Time avarage global delay in the network

Figure 7 shows the amount of VoIP dropped packets, when using different queuing schemes. As we mentioned above, best results in that case we obtain with CBWFQ method which have fixed guaranteed amount of bandwidth. Then follow WFQ, DWRR, MWRR and CQ queuing scheme. The opposite situation is when we take in to the consideration delays. There selection of CBWFQ introduces biggest delay, because majority low level traffic must wait.

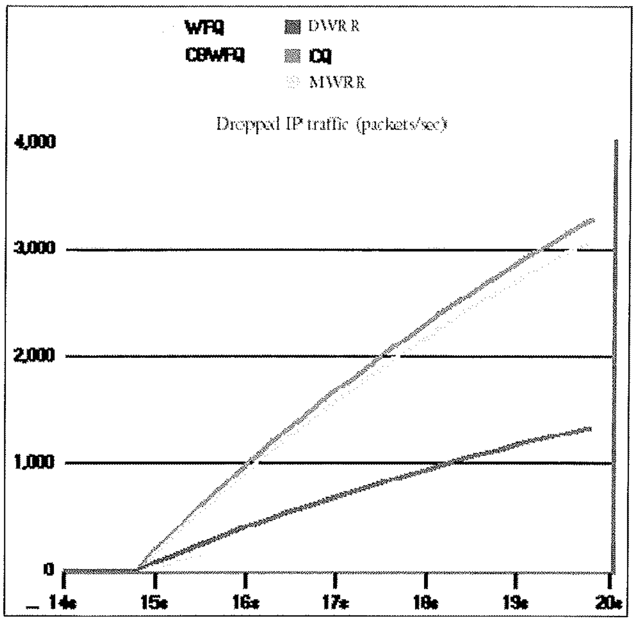


Fig. 7. Tested wired network arhitecture which is a copy of real network

On the Figure 8 we will show, how combined queuing method PQCBWFQ improves the delays comparison with before presented delays, shown on figure 6.

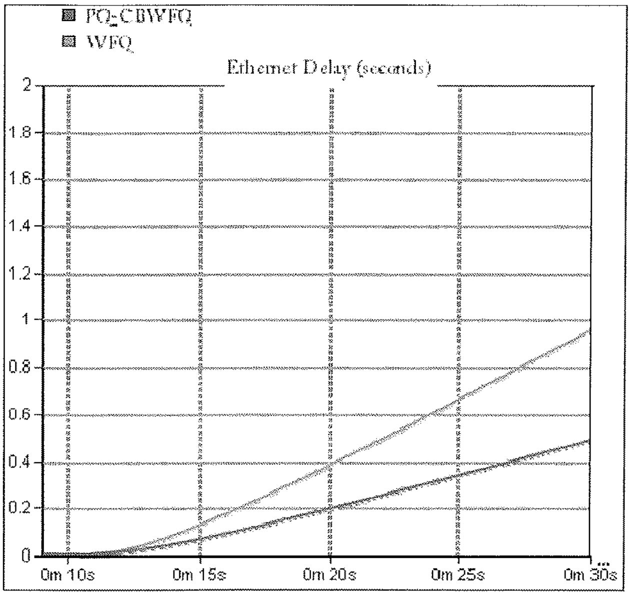


Fig. 8. Ethernet delay (seconds) for combined PQCBWFQ method comparison with WFQ

As we can see on Figure 8, PQ-CBWFQ delay is smallest than WFQ delay, but on figure 6 has ordinary CBWFQ method bigger delay than WFQ, observed in whole ethernet segment. Such combinations can perceivable improves network performances. Similar effect that is shown as on Figure 8 can be seen also in Figure 9 for VoIP delay.

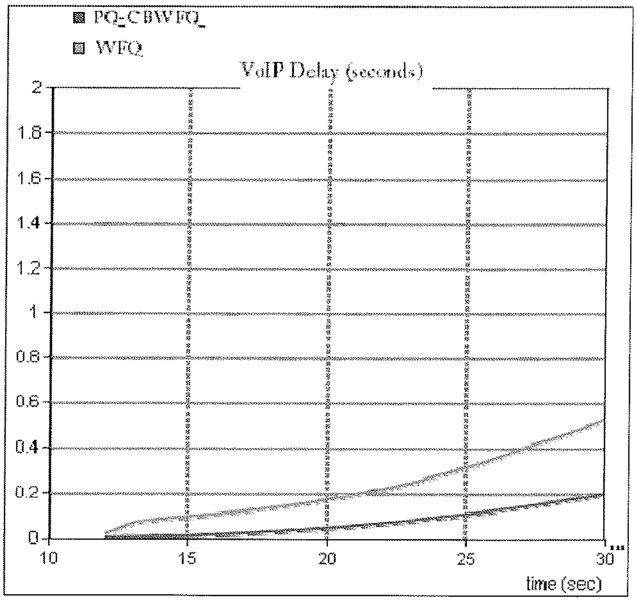


Fig. 9. VoIP delay (seconds) for combined PQCBWFQ method comparison with WFQ

Using combined queuing method the delay is also reduced in comparison with ordinary WFQ queuing for VoIP traffic. Delay in VoIP application plays an important role on quality of perception. The smaller it is, the better voice quality can be offered.

7. Conclusion

During many simulation runs and graph analysis we can say that queuing policy discipline influences on quality of service for the network applications. In many cases CQ queuing discipline was the best choice, in the case, when we have only two traffic flows WFQ has been the best choice, but looking from other perspective, when we must handle multiple traffic flows CBWFQ was the best solution. Such method (CBWFQ) has also disadvantages; in our case we defined only one class with bandwidth amount 9Mbit/s reserved for VoIP, rest of bandwidth belongs to majority low priority HTTP traffic. In that point, majority traffic hasn't enough bandwidth and must wait, what influences on common delay. This is the main reason why CBWFQ has highest average delay in network. No matter to that delay, VoIP delay is constant during simulation because of ensured bandwidth by defined class. Looking from other side, if we want fairness queuing discipline which serves fair all applications then we use WFQ or CQ mechanism, but if we want only that highest priority traffic flows pass through the network, we should use priority queuing PQ.

Delays in CBWFQ case can be reduced using PQ-CBWFQ queuing scheme. This is a combination of a priority PQ and a CBWFQ queuing mechanisms. PQ-CBWFQ is a queuing method from family that offers smallest delays. Such queuing allows that specific flow class defined with IP priorities is served as strict priority queue. Highest priority class is served before all other priority classes. Such combination and their improvement from view of Ethernet delays are shown on Figure 9. Our simulations show that we must look for solutions also in using combined queuing methods. All other available combinations represent a challenge for further researches on that area.

References

/1/ S. Klampfer, J. Mohorko, and Č. Žarko, "Vpliv različnih načinov uvrščanja na karakteristiko prepustnosti omrežja," ERK 2007

/2/ M. Fras, S. Klampfer, Ž. Čučej, "Impact of P2P traffic to the IP communication network's performance," IWSSIP 2008

/3/ T. Subash and S. IndiraGandhi, "Performance Analysis of Scheduling Disciplines in Optical Networks," MADRAS Institute od Technology, Anna University

/4/ Jernej Kozak: Podatkovne strukture in Algoritmi. Društvo MFA SRS, Ljubljana.

/5/ Lary L. Peterson and Bruce S. Davie, Computer Networks, Edition 3, San Francisco 2003

/6/ Stefan Bucheli, "Compensation Modeling for QoS Support on a Wireless Network," Master degree thesis

/7/ Kian Meng Yap, Alan Marshall, Wai Yu, "Providing QoS for Multimodal System Traffic Flows in Distributed Haptic Virtual Environments," Queen's University Belfast

/8/ Internetworking Technology Handbook - Quality of Service (QoS), Cisco Systems

/9/ OPNET Modeler Techical Documentat

/10/ Samuel P. Morgan, "Queuing Disciplines and Passive Congestion Control in Byte-Stream Networks," IEEE Transactions On Communications

/11/ http://www.cisco.com/en/US/docs/ios/12_0/qos/configuration/guide/qccq.html

/12/ Maria Johanna Gerarda van Uitert, "Generalized Processor Sharing Queues," Masterdam 2003

Saša Klampfer, Jože Mohorko, Žarko Čučej
University of Maribor, Faculty of Electrical Engineering and Computer Science, Maribor, Slovenia
sasa.klampfer@uni-mb.si

Prispelo (Arrived): 28.10.2008 Sprejeto (Accepted): 09.06.2009

UDK621.3:(53+54+621+66), ISSN0352-9045 Informacije MIDEM 39(2009)2, Ljubljana

CANTILEVER METHOD FOR DETERMINATION OF D_{31} COEFFICIENT IN THIN PIEZOELECTRIC FILMS

Samo Penič, Uroš Aljančič, Drago Resnik, Danilo Vrtačnik, Matej Možek, Slavko Amon

Laboratory of Microsensor Structures and Electronics (LMSE), Faculty of Electrical Engineering, University of Ljubljana, Ljubljana, Slovenia

Key words: piezoelectric, characterization, d_{31} coefficient, PZT, simulation, FEM, ANSYS

Abstract: A cantilever method for characetrization of thin piezoelectric films is proposed. Using the proposed cantilever method, piezoelectric coefficient d_{31} of thin film piezoelectric material on various samples was determined. Cantilever based characterization method provides a fast comparison of different piezoelectric material samples, since multiple samples can be mounted simultaneously on the testing structure. It is shown how, when combined with numerical simulation, piezoelectric coefficient d_{31} can be determined from fitting measured voltage response with simulated response.

Exact knowledge of geometry and material properties of cantilever and samples proved to be important in order to determine piezoelectric coefficients with sufficient accuracy. Stainless steel cantilever was adequately characterized by measuring its Young's modulus. Silicon properties are adequately determined by published data. Mechanical properties of PZT layers are on the other hand more difficult to acquire, since they are rather dependent on the actual PZT preparation procedure and composition. Nevertheless, we expect that error here introduced is small due to very thin PZT layer compared to stainless steel cantilever and silicon substrate. To improve the proposed method, based on numerical simulation results, guard chips were mounted at the side of the cantilever to reduce stress variation over samples.

Determined values of piezoelectric coefficients d_{31} for PZT layers under test were in reasonable agreement with results available in the literature.

Metoda za določanje koeficienta d_{31} tankih piezoelektričnih filmov

Kjučne besede: piezoelektrik, karakterizacija, d_{31} koeficient, PZT, simulacija, silicij, FEM, ANSYS

Izvleček: V članku je predstavljena metoda za karakterizacijo tankih piezoelektričnih plasti. Z uporabo ročice smo določili piezoelektrični koeficient d_{31} tankih piezoelektričnih filmov. Metoda omogoča hitro primerjavo lastnosti različnih materialov, ter s pomočjo numerične simulacije hkratno karakterizacijo parametra d_{31} večih vzorcev.

Poznavanje geometrije in materialnih lastnosti ročice in vzorjev je ključno za natančno določitev piezoelektričnih koeficientov. Mehanske lastnosti jeklene ročice smo določili z meritvijo Youngovega modula, za mehanske lastnosti silicijevega substrata pa smo uporabili podatke v literaturi. Mehanske lastnosti tankih PZT plasti so težje dostopne, saj se razlikujejo zaradi same zgradbe PZT keramike ter njene priprave. Zaradi tanke plasti PZT materiala, ocenjujemo, da je napaka pri uporabi vrednosti za debele materiale zanemarljiva. Na osnovi simulacij smo predstavljeno metodo izboljšali z dodatnimi stranskimi čipi, ki izboljšajo homogenost stresa na vzorcih.

Vrednosti za piezoelektrični koeficient d_{31}, ki smo jih dolovili s predlagano metodo, se ujemajo s podatki iz literature.

1. Introduction

When designing a new product or device, proper material selection is of basic importance. Material properties are also used in numerical analysis, when predicting device behavior. In case of piezoelectric microstructures, the properties of thin film piezoelectrics are influenced by chemical composition and other parameters of piezoelectric manufacturing process. It is thus important to have means for analyzing specific samples of piezoelectric thin films.

Due to unique properties of piezoelectric effect, piezoelectrics are important materials in micro-electromechanical system (MEMS) technology, used for actuation or sensing, energy harvesting etc. Characteristics of piezoelectrics, especially piezoelectric coefficients d, play important role in device design, simulation and behavior prediction. In general, thin film materials used in microengineering behave differently than bulk, thus requiring an adequate characterization of their core properties. Several methods are in use /1/ and new ones are being developed. Selection of the appropriate material for a certain application requires comparison of different materials using datasheet specifying core information about these materials.

Properties of piezoelectric materials vary with chemical composition, preparation technique e.g. sintering temperature and other effects. These influences present difficulties for comparison of different materials prepared by different methods, of different thicknesses and possibly from different producers, usually taking plain datasheet information from catalogue as a starting point.

To overcome this obstacle a comparative method for characterization of different thin film piezoelectric samples bonded to a stainless steel cantilever is proposed. The relative response of different piezoelectric samples to the same mechanical stress gives immediate comparison of their basic properties such as sensitivity and linearity. Furthermore, coupling the measured results with numerical simulation based on finite element method (FEM) enables determination of absolute value for piezoelectric coefficient d_{31}.

The paper presents in detail the proposed technique for thin film piezoelectrics characterization and introduces a comparative method for simultaneous evaluation of multiple piezoelectric samples based on numerical simulation in combination with measured results. The result of this characterization is the absolute value of d_{31} coefficient for multiple samples and comparison of piezoelectric response to mechanical stimulus. The method is practically tested on different thin film Lead Zirconate Titanate (PZT) chip samples prepared on silicon substrates. Measured results are matched with numerical simulation and piezoelectric coefficients are determined using ANSYS finite element analysis software.

2. Basic properties of piezoelectrics

Piezoelectrics are materials that respond to the applied mechanical stress with electric voltage on the electrodes. This is called the direct piezoelectric effect, which serves as a basis for sensors and generators. The effect can be reversed and it is then called converse or inverse piezoelectric effect. Here mechanical strain is induced when voltage is applied. The response is dependent on the polarity of applied voltage and can therefore vary between elongation and contraction.

Equations that describe electromechanical relations in a piezoelectric material are given in Voight notation with relations /2/

$$\begin{aligned} \{T\} &= [c]\{S\} - [e]\{E\} \\ \{D\} &= [e]^{\mathrm{T}}\{S\} - [\check{n}]\{E\} \end{aligned} \tag{1}$$

where $\{T\}$ is stress tensor, $\{S\}$ strain tensor, $\{E\}$ electric field vector and $\{D\}$ electric displacement vector. Material properties are described with stiffness matrix $[c]$ which includes information about Young's modulus Y and Poisson ratio σ of the material, with piezoelectric stress matrix $[e]$ (superscript T denotes matrix transpose) related to piezoelectric strain matrix $[d]$ and with permittivity matrix $[\check{n}]$. Piezoelectric strain coefficients d_{ij} and piezoelectric stress coefficients e_{ij} are related with stiffness coefficients c_{ij} by matrix equation $[e]= [c]\,[d]$.

Piezoelectrics can be used for sensing or actuation, depending on whether the applied input load is mechanical or electrical, respectively. The two modes of operation can also be used interchangeably which makes piezoelectrics extremely versatile electromechanical materials since the same structure can act as a sensor or an actuator. Though the effect is reversible, certain considerations must be taken into account during the design of the structure /3/.

3. Piezoelectrics characterization

3.1 Bulk piezoelectrics characterization

A complete characterization process of bulk piezoelectric material includes determination of stiffness coefficients c_{ij} (including Young's modulus Y and Poisson ratio σ), permittivity ($\check{n}_{ii}$) and piezoelectric coefficients (d_{ij}).

Most widely used method adopted as IEEE standard for piezoelectric characterization is the resonance method /4/. For such characterization, piezoelectric material is prepared as a flat rectangular plate between two electrodes, forming a capacitor. The capacitor impedance Z is measured at different frequencies. From $Z(f)$ diagram, the resonant (f_r) and anti-resonant (f_a) frequencies are found. Then, the elastic compliance (inverse stiffness matrix) and piezoelectric coefficients for practical purposes usually d_{31} and d_{33} can be derived /1/.

Direct methods for determining piezoelectric coefficients d_{ij} include deformation measurements when voltage is applied to the electrodes. These methods are used to quantify the direct and converse piezoelectric effect. Direct methods are also used to investigate the behavior of the piezoelectric material in terms of hysteresis and nonlinearity, thermal behavior and aging. Mechanical deformation measurement of piezoelectric sample vs. applied voltage is used to determine piezoelectric coefficients d_{ij}, calculated from relation in Voight notation $S_j= d_{ij}\, E_i$/1/.

A different method for measuring piezoelectric coefficients d_{ij} is based on direct piezoelectric effect. Here, sample is mechanically loaded, therefore the bounded electric charge becomes free, ready to flow out from the electrodes /5/. Electrodes are short circuited and electric displacement D is measured. Piezoelectric coefficient d_{ij} is here calculated from equation in Voight notation $D_i = d_{ij}\, T_j$ /1/.

In order to determine the relative permittivity $\check{n}_r$, capacitance measurements are carried out at low frequency, usually 1 kHz and for low AC voltage excitation levels, ranging few mV /1/. The relative dielectric constant is then calculated as

$$\check{n}_r = \frac{Ct}{\check{n}_0 A} \tag{2}$$

where t is thickness of piezoelectric layer, A electrode area, C measured capacitance and $\check{n}_0$ permittivity of free space.

3.2 Thin film piezoelectrics characterization

In general, the properties of thin film materials can differ significantly from its bulk counterparts. Therefore, adequate characterization of piezoelectric thin film properties is essential. Thin film characterization methods are usually based on similar principles as for bulk. The prevailing methods use converse piezoelectric effect where electrically excited thin piezoelectric film results in mechanical displacement, which is typically in the order of a few angstroms /6/. Sometimes the direct piezoelectric effect is used. Thin film piezoelectric together with electrodes are deposited on a substrate wafer and fixed in a rigid frame above pneumatic pressure cavity /7/. Pressure in the cavity is varied thus applying different mechanical stress to the piezoelectric layer. The charge integrator is used to measure the induced charge which is used in combination with excitation pressure to determine piezoelectric coefficients d_{ij}.

For determining Young's modulus of thin piezoelectric films, several approaches exist. One of the possibilities to characterize mechanical thin film properties is presented in /8/. The experiment consists of loading a membrane with a line load applied to the middle of the span using nanoindenter. A Mireau microscope interferometer is used to observe fringes that are formed on the loaded sample. Using a CCD camera these fringes are recorded and strains determined. From known stresses and strains in the material, Young's modulus can be determined.

3.3 Cantilever method for characterization of thin piezoelectric films

In this case, characterization method is focused on piezoelectric coefficient d_{31} using direct piezoelectric effect.

In the proposed characterization method we introduce a cantilever with mounted piezoelectric samples on silicon substrate, with exact control of deflection. Mounting several samples simultaneously to the same cantilever provides us a comparison of piezoelectric responses of various piezoelectric materials to the same stimulus. This provides fast and accurate comparison of different piezoelectric materials appropriate for R&D work. When comparing responses of different materials, relative comparative method is usually sufficient and sometimes preferred to comparing absolute values due to its simplicity. However, determination of absolute values of piezoelectric coefficients is also possible, upgrading the proposed method with analysis of mechanical setup using appropriate numerical simulation as shown later. For this purpose, finite element analysis (FEA) software ANSYS was used.

Mechanical properties of piezoelectric and silicon were taken from literature /9,10/. Permittivity was determined from capacitance measurements.

4. Experimental setup

Experimental setup consisting of rectangular cross-section cantilever with mounted samples is shown in Fig. 1. Due to the simplicity of cantilever with rectangular cross-section, also analytical expressions for stress distribution exist, enabling comparison with numerical results. Proposed characterization method uses samples with thin film piezoelectric capacitor structure on silicon substrate, mounted on stainless steel cantilever. The selection of optimal samples placement is essential, usually selected for high sensitivity as the region of maximum stress distribution in the beam still having sufficient uniformity. Stress decreases in cantilever longitudinal direction towards the cantilever free end where it reaches zero. Therefore, the samples are mounted in the region of maximum stress being at the root of the cantilever. Following our simulation results, care must be taken not to induce an excessive error in the placement of samples.

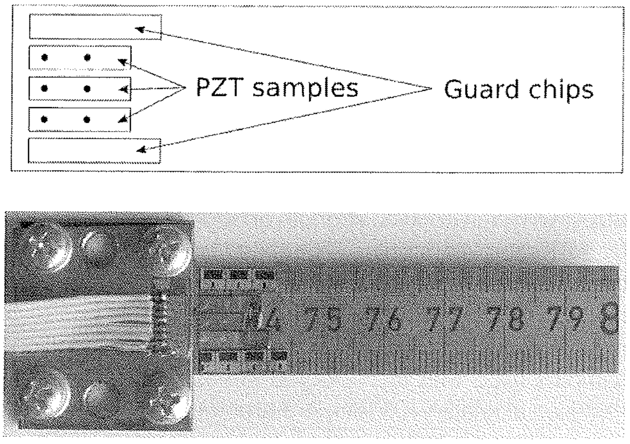


Fig. 1: Top view of the cantilever with mounted samples and side guards: (a) schematic, (b) photograph

For adequate characterization of piezoelectric thin film samples, high repeatability of sample loading is essential. The testing cantilever setup, together with bonded samples represents such a test structure. Stainless steel was selected as the material for cantilever, providing possibility of high repeatable deflections. Furthermore, stainless steel cantilever is mechanically resistant and can be reused after replacing samples.

During characterization, samples are often exposed to higher mechanical stresses as during the normal sensor or actuator operation. To achieve such a wide measurement range, cold rolled austenitic stainless steel (1.4310) was selected for the cantilever realization. This material has an extended elastic range due to a special treatment during the fabrication. In this case, the cantilever returns to its initial position even after extremely large deflections.

To achieve large measured range of stresses for samples under test, the mechanical part of testing system has to provide adaptability. Therefore, 10 cm long and 18 mm wide stainless steel strips (cantilevers) of thickness 0.5 mm were cut by milling and then pressed between two rigid stainless steel plates acting as a fixed support. In this approach, the cantilever length is adjustable, resulting in increased measured range with high repeatability and accuracy.

To illustrate the characterization of piezoelectric samples with described experimental setup, various thin PZT layers were deposited by sol-gel method on silicon chips covered by Pt/Ti as reported elsewhere /11/. Gold electrodes were placed on top of PZT layer by sputtering and shaped by shadow mask method. Thin Ti and Pt layers with thicknesses of 10 and 100 nm respectively are not significant for the overall mechanical properties of the relatively thick samples and were thus neglected in numerical simulations.

As an example of the proposed characterization procedure, three samples with two different thicknesses of PZT layer were introduced, marked as samples PZT1a, PZT1b and PZT2. Due to our numerical simulations, two dummy guard chips were added at cantilever sides to achieve better stress uniformity over the samples (Fig. 1).

To assure a reliable transfer of induced mechanical stress from the cantilever to the PZT samples, a strong and stable bond between the cantilever and the samples has to be achieved. Therefore, an epoxy adhesive (UHU endfest 3000) with high bonding strength of 3000 N/cm^2 was used for PZT samples bonding. The extended elastic range of the selected stainless steel, in the combination with the mentioned adhesive enable highly reliable loads on testing samples, up to the silicon tensile strength. In addition, samples fixed with the selected adhesive can be easily removed at relativly low temperatures what makes the testing cantilever reusable /11/.

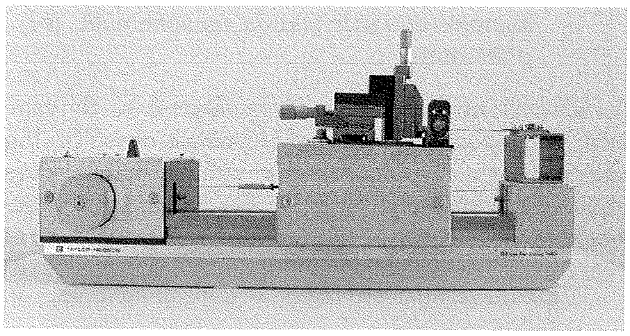


Fig. 2: Experimental setup: Taylor-Hobson traversing table and micromanipulator are used to achieve high deflection repeatability.

To achieve highly repeatable stresses, testing cantilever with bonded samples is mounted on the fixed part of modified Taylor-Hobson 150mm Traversing Table, as shown in Fig. 2. The computer controlled worktable is motor driven in both directions, but can also be moved manually. Straightness accuracy of the worktable is within ±1 μm over the full 150mm range. In order to assure deflection repeatibility, a micromanipulator with 8 mm tall pointed pin is mounted at the top of the worktable, as described in detail elsewhere /12/.

Voltage response of PZT samples is measured by Semiconductor Parameter Analyzer HP4155A, including SMU and PMU Generator Expander HP41501A.

For determination of piezoelectrics permittivity, capacitance on test capacitors is measured with HP4284A Precision LCR Meter at various frequencies, at excitation amplitude 1 V and DC bias 0 V.

5. Numerical modeling

For the purpose of simulation, commercial FEM modeling and simulation software ANSYS was used. Simulator input for cantilever test structure with samples is built using ANSYS proprietary scripting language APDL. Meshing is done using built-in automatic mesh generator. The resulting hexahedral mesh of simulated test structure is shown in Fig. 3. Local improvement of the mesh was done manually to refine mesh in structure critical regions such as thin PZT layer and to avoid badly shaped elements.

The test structure basically consists of several different layers - stainless steel (SS) cantilever, silicon (Si) substrate chip, metal and PZT layer. Electrodes and interface layers were neglected at mechanical simulation due to their small thicknesses. For modeling SS and Si materials, three-dimensional SOLID95 elements were used. PZT layer was modeled with SOLID226 elements with capability to couple mechanical and electrical quantities using piezoelectric effect.

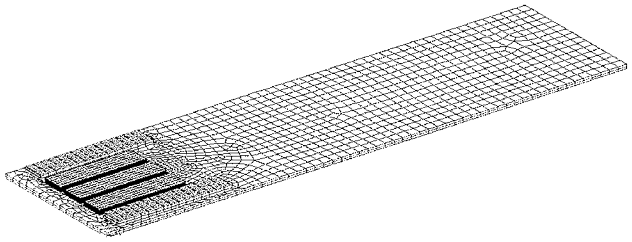


Fig. 3: Generated mesh of cantilever with 3 bonded samples and two side guards.

When we take into account material symmetry, general form of stiffness matrix [c] for ceramics, permittivity matrix [ň] and piezoelectric coefficients matrix [d] can be simplified /1/.

$$[c]=\begin{bmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{13} & 0 & 0 & 0 \\ c_{13} & c_{13} & c_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & (c_{11}-c_{12})/2 \end{bmatrix} \qquad (3)$$

$$[\check{n}]=\begin{bmatrix} \check{n}_1 & 00 & \\ 0 & \check{n}_1 & 0 \\ 0 & 0 & \check{n}_3 \end{bmatrix} \qquad (4)$$

$$[d]=\begin{bmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{15} & 0 & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{bmatrix} \qquad (5)$$

Due to the lack of exact information in the literature, mechanical properties of thin PZT layer were approximated by bulk values. Therefore, values $c_{11}=13.9\cdot10^{10}$ Pa, $c_{33}=11.5\cdot10^{10}$ Pa, $c_{44}=2.56\cdot10^{10}$ Pa, $c_{13}=7.43\cdot10^{10}$ Pa,

$c_{12} = 7.78\cdot10^{10}$ Pa were taken from literature /10/. Due to the small thickness of PZT compared to the cantilever and Si substrate, the error introduced is negligible.

SS material is usually considered isotropic. The Young's modulus of SS material was measured using nanoindentation method /13/. The measured value of SS Young's modulus is Y = 167.56 GPa.

Silicon is very well known material. Due to Si crystal symmetry, it is described by 3 stiffness coefficients c_{11}, c_{12} and c_{44}. In our case Si was modeled using anisotropic symmetric matrix with coefficients $c_{11} = 0.1657\cdot10^{6}$ Pa, $c_{12} = 0.0639\cdot10^{6}$ Pa, $c_{44} = 0.0796\cdot10^{6}$ Pa /14/.

Due to the longitudinal stress dominating in our case as confirmed by our numerical simulation, only piezoelectric coefficient d_{31} was taken into account.

Boundary conditions for cantilever at FEM simulation were fixed support on the cantilever left side (deflection and its derivative equal to 0) and free deflection on the right side. To allow simple load variation, the deflection was described in the program as a parameter. Electrical ground boundary condition was set on the bottom electrode.

Standard sparse direct linear solver was used for solving the model having 85000 elements with 4 basic variables (degrees of freedom) of the problem: electric potential and displacements in x, y, z direction. Sparse direct solver is a robust and fast solver for linear and nonlinear analysis, appropriate when poorly shaped elements are present in the model, such as the high aspect ratio (thickness vs. width) elements in the model of PZT layer. The sparse direct solver is based on a direct solution of equations by elimination, as opposed to iterative solvers where the solution is obtained through an iterative process that successively refines an initial guess to the final solution that is within a prescribed tolerance of the final solution. Direct elimination requires the factorization of an initial very sparse linear system of equations into a lower triangular matrix followed by forward and backward substitution. Drawback of this solver is that it requires a significant amount of memory, thus it is not suitable for larger scale models with more than a half million variables. Because sparse direct solver is based on direct elimination, poorly conditioned matrices do not pose difficulty in producing the solution /15/. Direct solver was chosen for our simulated approach since it does not exceed the recommended number of equations and there was enough computer memory available to perform computation.

Simulations were performed on Intel Core Duo 6600 64-bit processor architecture with 4GB RAM memory, running at 2.4GHz. A single simulation run with chosen solver required typically 6 minutes.

6. Procedure for determination of piezoelectric coefficient d$_{31}$

The described experimental setup was used to deflect cantilever. Corresponding voltage response of the mechanically loaded PZT samples was measured with parametric analyzer as described previously. Numerical simulator was configured as discussed in the previous section, to translate the test structure into numerical model. The characterization of piezoelectric effect and related d_{31} parameter was performed by fitting the measured voltage response with simulated response: d_{31} parameter value was varied in the simulator until a good match between measured and calculated voltage response was found. The value of d_{31} that provided best fit throughout all deflections between calculated and measured voltage response was selected as the final result for the piezoelectric coefficient d_{31}.

7. Results and Discussion

The PZT samples capacitance was measured using LCR meter at frequencies ranging from 20 Hz to 10 kHz, at excitation voltage of 10 mV. A relatively small dependence of capacitance vs. frequency was detected (Fig. 4). Measured capacitance value at 1 kHz was taken, as stated in /1/. Top electrode area was measured under the microscope. PZT layer thickness was measured after the fabrication of the layer. From data given in Table 1 the relative permittivity $\check{n}_r$ for samples was calculated. Samples PZT1a and PZT1b are built on the same PZT layer differing only in their electrode position, regarding to the cantilever support (Fig. 1). Electrode of PZT1a was located 2.7 mm from the support, while the electrode of PZT1b was located 5.4 mm from the cantilever support. Sample PZT2 was prepared with modified processing for double thickness of PZT1, resulting in changed value of PZT permittivity. The electrode location for PZT2 was the same as for PZT1a, 2.7 mm from the support.

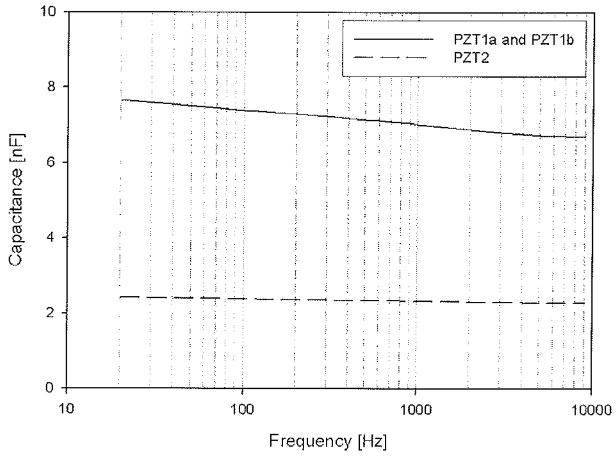


Fig. 4: Measured samples capacitance vs. frequency.

Table 1: Measured sample parameters and calculated relative permittivity of PZT layers

Sample	PZT Thickness [nm]	Electrodes [mm^2]	Capacitance [nF]	Rel. permittivity ň$_r$
PZT1a	740	0.87	7.67	737
PZT1b	740	0.87	7.67	737
PZT2	1554	0.87	2.40	484

Simulated stress profile in PZT layer is shown in Fig. 5 (simulation path is shown in the inset). From Fig. 5 can be concluded that electrode exact position is important when performing characterization of multiple samples. Following our numerical simulations results, to minimize the difference of stress profile in neighbor samples, two longer guard chips are added at the sides, as shown in Fig. 1. Calculated stress distribution in the cantilever and samples is shown in Fig. 6a.

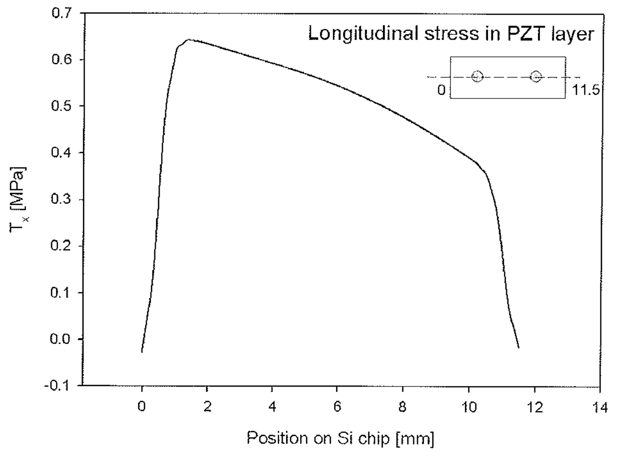


Fig. 5: Simulated longitudinal stress profile in PZT layer vs. position on Si chip.

The effect of guard chips is quantified in Fig. 6 and Table 2. As shown, the absolute stress in samples is decreased when guard chips are present. However, stress uniformity over the samples improves significantly. The relative difference in stress in both cases, without and with guard chips, was calculated between central and side samples. Guard chips thus provide more homogenous stress conditions on all samples.

According to the piezoelectric effect, voltage response is proportional to the stress, what is described by piezoelectric coefficients. Calculated voltage response of PZT samples due to calculated stress is given in Fig. 6b.

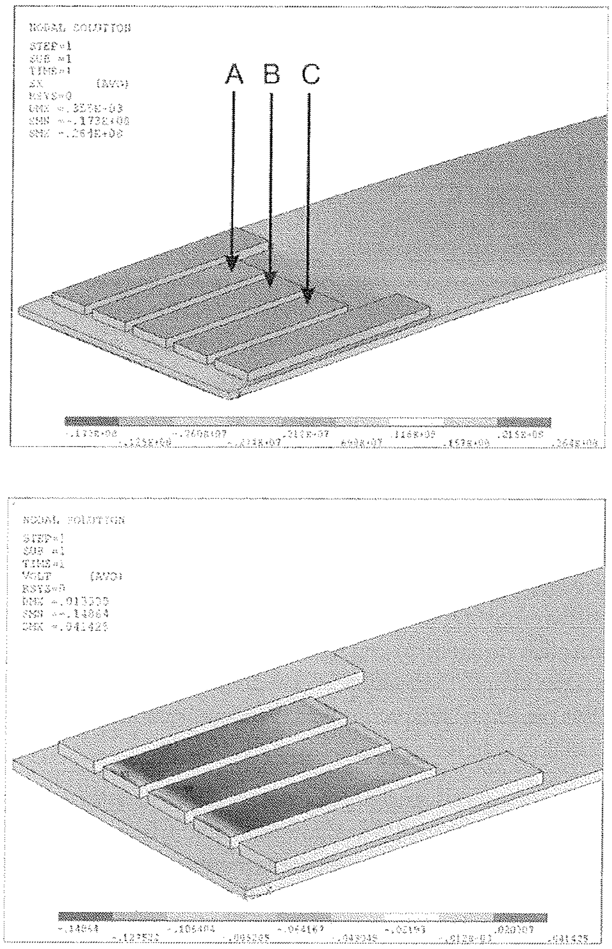


Fig. 6: Simulated longitudinal stress distribution in stainless steel cantilever and silicon chips (a) and corresponding voltage on top of PZT layer due to accumulated charge (b). Positions on the chips A, B and C show where stresses were compared.

Measured time dependent voltage response of PZT samples during testing is shown in Fig. 7. Here, the cantilever was deflected to predefined values using the micromanipulator as previously described. At start, the cantilever was first manually deflected over the desired deflection value, and then after this it was released to rest in final position determined by micromanipulator. Similar procedure was applied also during the end of testing. Consequently, voltage spikes always occurred at the start and at the end of loading.

As also seen in Fig. 7, the response for constantly deflected cantilever slowly decreases with time, probably due to piezoelectric internal effects such as leakage and recombinations, and due to external effects such as input im-

Table 2: Improvement of the stress uniformity over samples when guard chips are used.

Stress in samples **without guard chips** [MPa]			Stress in samples **with guard chips** [MPa]		
Position	Positions A, C	Rel. difference	Position	Positions A, C	Rel. difference
13.19	14.74	10.5 %	10.04	10.08	3.7 %

pedance of HP4155A connected to the sample. Therefore, measurement of the response was done after the spike settled down, typically after 10 seconds.

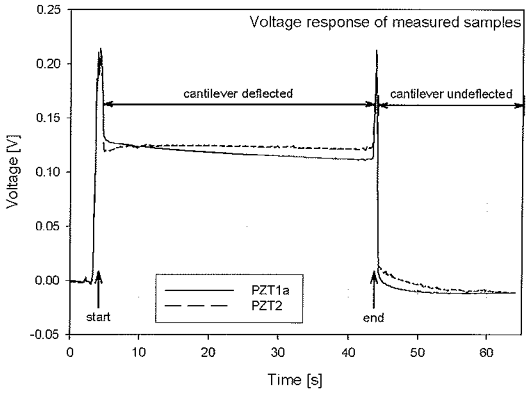


Fig. 7: Measured voltage response vs. time during testing

Measured voltage response results are graphically displayed in Fig. 8. The response amplitude is dependent on electrode distance from the cantilever support and is in correlation with simulated stress profile in PZT layer shown in Fig. 5. The voltage on PZT1a is thus considerably higher than voltage on PZT1b. PZT2 that differs in thickness and material properties produces response that is slightly higher than with PZT1a.

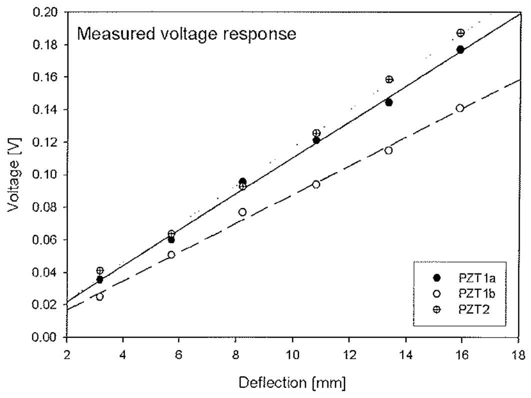


Fig. 8: Measured voltage response of PZT samples vs. deflection.

Determination of piezoelectric coefficient d_{31} was done by using numerical simulation as described previously. Successive simulations were performed for various values of coefficient d_{31} until close agreement between simulated and measured voltage response was obtained. Measured and simulated responses at various deflections for best values of piezoelectric coefficient d_{31} are given in Table 3. The summary of measured values for relative permittivity ň$_r$ and piezoelectric coefficient d_{31} for PZT materials under test is given in Table 4. Results obtained are in reasonable agreement with available values from the literature /7/.

Table 4: Measured properties of PZT layer.

Sample	Relative permittivity - ň	Piezoelectric coef. - d_{31} [pC/N]
PZT1a	737	66.1
PZT1b	737	-66.1
PZT2	484	-20.7

Graphical representation of measured and simulated voltage responses vs. deflection for all three PZT samples are shown in Fig. 9. In the range of measured deflections, the simulated response displays linearity while it is slightly distorted for measured values, probably due to measurement error.

8. Conclusion

Using the proposed cantilever method, piezoelectric coefficients d_{31} for various thin film piezoelectrics were determined. Cantilever based characterization method provides a fast comparison of different piezoelectric material samples, since multiple samples can be mounted simultaneously on the testing structure. Furthermore, when combining experimental data with numerical simulation, piezoelectric coefficient d_{31} can be determined by matching simulated results with voltage response measurements.

Exact knowledge of geometry and material properties of cantilever and samples proved to be important in order to measure piezoelectric coefficients with sufficient accuracy. Stainless steel cantilever was adequately characterized by measuring its Young's modulus. Silicon properties are adequately determined by published data in the literature. Mechanical properties of PZT layers are on the other hand more difficult to acquire, since they are rather dependent on the actual PZT preparation procedure and composition. Nevertheless, we expect that error here introduced is small due to very thin PZT layer compared to stainless steel cantilever and silicon substrate. To improve the presented method, based on numerical simulation results guard chips were mounted at the side of the cantilever to

Table 3: Measured voltage response and simulated values for different deflections of cantilever with best fit value for d_{31} parameter.

	PZT1a		PZT1b		PZT2	
Deflection	Meas[mV]	Sim [mV]	Meas[mV]	Sim [mV]	Meas[mV]	Sim [mV]
3.175 mm	35.5	35.24	24.9	27.80	41	37.38
5.715 mm	60.0	63.40	50.8	50.10	63.7	67.30
8.255 mm	95.4	91.62	76.9	72.43	92.5	97.19
10.795 mm	121	119.8	93.9	94.70	125	127.1
13.335 mm	144	148.0	114	117.0	158	157.1
15.875 mm	173	176.2	141	139.1	187	186.9

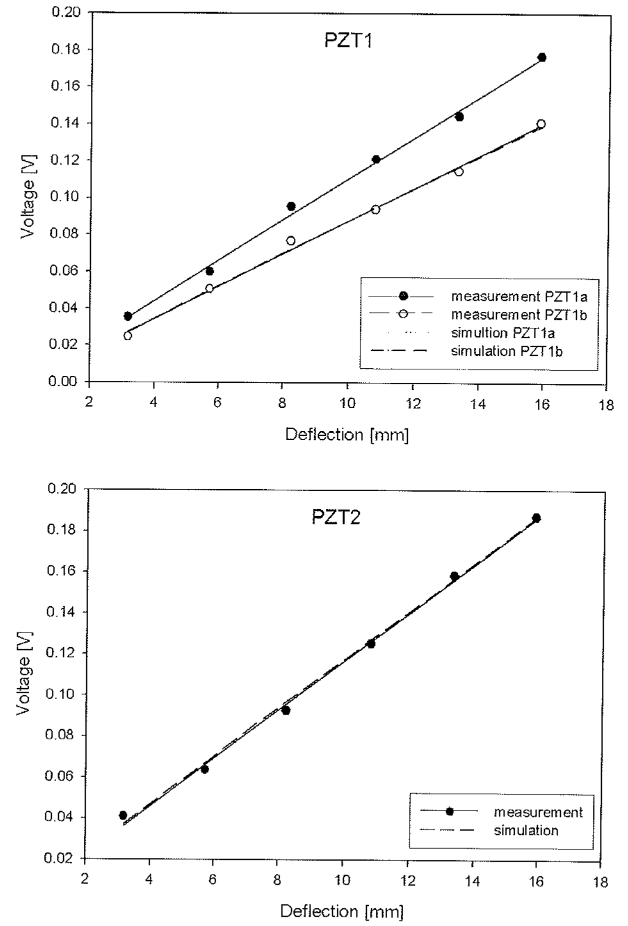


Fig. 9: Graphical representation of measured and simulated voltage response of PZT samples vs. deflection.

reduce stress variation over the samples. Determined values of piezoelectric coefficients d_{31} for PZT layers under test were in reasonable agreement with results available in the literature.

Acknowledgment

Authors would like to acknowledge Electronic Ceramics Department - K5, Jožef Stefan Institute, Slovenia for PZT samples preparation.

This work was supported by Ministry of Higher Education, Science and Technology and Slovenian Research Agency.

References

/1/ T.L. Jordan, Z. Ounaies, "Piezoelectric Ceramics Characterization". NASA/ CR-2001-211225 ICASE report to NASA Langley Research Center. Report No. 2001-28, September 2001.

/2/ Ansys Inc. "Ansys Inc. Theory reference". Ansys Inc., 2005.

/3/ S. Penič, U. Aljančič, D. Vrtačnik, D. Resnik, M. Možek and S. Amon "Numerical modeling of PZT/SiO_2 microcantilever with interdigitated electrodes", Proc. 43rd International Conference on Microelectronics, Devices and Materials and the Workshop on Electronic Testing, Bled, Slovenia, September 2007, pp. 63-68.

/4/ "IEEE Standard on Piezoelectricity", (IEEE Standard 176-1987), Institute of Electrical and Electronic Engineers, 345 East 47th St, New York, NY 10017.

/5/ K. C. Kao, "Dielectric phenomena in solids", Elsevier Academic Press, San Diego, California, 2004.

/6/ J.T. Dawley, G. Teowee, B.J.J. Zelinski and D.R. Uhlmann "Piezoelectric Characterization of Bulk and Thin Film Ferroelectric Materials using Fiber Optics". MTI Instruments application note, http://www.mtiinstruments.com/

/7/ J.F. Shepard Jr., P.J. Moses and S. Trolier-McKinstry "The wafer flexure technique for the determination of the transverse piezoelectric coefficient (d_{31}) of PZT thin films". Sensors and Actuators A, vol. 71, 1998, pp. 133-138.

/8/ H.D. Espinosa , B.C. Prorok, M. Fischer "A methodology for determining mechanical properties of freestanding thin films and MEMS materials". Journal of the Mechanics and Physics of Solids, vol. 51, 2003, pp. 47-67.

/9/ Efunda, http://www.efunda.com/

/10/ X.J. Zheng, Y. C. Zhou and J.Y. Li "Nano-indentation fracture test of $Pb(Zr_{0.52}Ti_{0.48})O_3$ ferroelectric thin films". Acta materialia, vol. 51, 2003, pp. 3985-3997.

/11/ U. Aljančič, B. Malič, M. Mandeljc, M. Vukadinovič, D. Vrtačnik, D. Resnik, M. Možek, M.Kosec, S. Amon "Cantilever as Testing Structure for Characterization of PZT Thin Films on Pt/Si Substrates". Proc. 42nd International Conference on Microelectronics, Devices and Materials and the Workshop on MEMS and NEMS, Strunjan, Slovenia, September 2006, pp. 271-276.

/12/ U. Aljančič, M. Vukadinovič, D. Resnik, D. Vrtačnik, M. Možek, S. Peniv, S. Amon "Cantilever Characterization Method for Static Behavior of PZT Thin Films". Proc. 43rd International Conference on Microelectronics, Devices and Materials and the Workshop on Electronic Testing, Bled, Slovenia, September 2007, pp. 115-120.

/13/ S. Penič, U. Aljančič, D. Vrtačnik, D. Resnik, M. Možek, M. Makovec, R. Bošnjak and S. Amon "FEM modeling of piezoresistive force sensor for medical retractor and design verification". Proc. 6th EUROSIM Congress on Modelling and Simulation, Ljubljana, Slovenia, September 2007, p. 158.

/14/ A. M. Fitzgerald "Practical Issues in Finite Element Analysis of MEMS". Ansys Workshop, March 2006.

/15/ Ansys Inc. "Ansys Inc. Basic Analysis Guide". Ansys Inc., 2005.

Samo Penič, univ. dipl. inž. el.
mag. Uroš Aljančič
doc.dr. Drago Resnik
doc.dr. Danilo Vrtačnik
mag. Matej Možek
prof.dr. Slavko Amon

University of Ljubljana, Faculty
of Electrical Engineering,
Laboratory of Microsensor Structures and Electronics
Trzaska 25, Ljubljana 1000, SLOVENIA
e-mail: matej.mozek@fe.uni-lj.si
Telefon: 01 4768 303, Telefax: 01 4264 630

Prispelo (Arrived): 12.08.2008 Sprejeto (Accepted): 09.06.2009

UDK621.3:(53+54+621+66), ISSN0352-9045 Informacije MIDEM 39(2009)2, Ljubljana

MODEL DETEKTORJA KAOTIČNOSTI

Matej Šalamon, Tomaž Dogša

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

Kjučne besede: kaotična vezja, bifurkacijski diagram, modeliranje, simulatorji električnih vezij.

Izvleček: Kaotično vedenje elektronskih vezij je mogoče analizirati tudi s pomočjo simulatorjev analognih vezij. Začetek kaotičnega režima delovanja vezja lahko ocenimo s pomočjo univerzalne Feigenbaumove konstante in bifurkacijskih točk, ki jih odčitamo iz bifurkacijskega diagrama.

V prispevku je predstavljen model detektorja kaotičnosti, ki omogoča direktno tvorjenje bifurkacijskega diagrama s samim simulatorjem SPICE /6/. Brez tega bi sicer potrebovali dodatni program, ki bi omogočal avtomatsko izvajanje simulacij vezja pri različnih vrednostih bifurkacijskega parametra ter shranjevanje ekstremnih vrednosti signalov.

The Model of Chaoticness Detector

Key words: chaotic circuits, bifurcation diagram, modelling, circuit simulators.

Abstract: Almost all circuits under specific circumstances (choice of parameters, initial conditions, input signals etc.) can become chaotic /1/. Chaotic circuits can be used in the cryptography or as a test that shows simulator's non-repeatability /8/. In general, chaotic behaviour of the electronic circuits is not tolerated. Simulation experiments allow us to determine chaotic responses for various sets of parameters, calculation of Lyapunov exponents, entropy, construction of Poincaré planes, construction of bifurcation diagrams etc.

The determination of the chaotic regions requires two steps. First one is the manual construction of a bifurcation diagram which is a tedious work. While carefully changing the bifurcation parameter and running the simulation we must find peak values in the selected output signal. The bifurcation points are values of the bifurcation parameter where a qualitative change in circuit dynamics is noticed. In the bifurcation diagram we can observe bifurcation sequences or period-doublings that construct a tree (*Fig. 1*). If a period-doubling bifurcation from period-1 to period-2 occurs at a bifurcation parameter value a_1 and from period-2 to period-4 at a parameter value a_2, then we can make a prediction of the parameter value a_∞ where chaos should appear (*eq. (1)*). However, we can not expect that this prediction is totally accurate. Firstly, because it is based on numbers a_1 and a_2 determined by the simulation and secondly, we have assumed that all measured bifurcation ratios are described by the Feigenbaum's constant. Nevertheless, after the bifurcation diagram is constructed it is easy to calculate the parameter value a_∞ which is a reasonable estimation of the region where chaos begins. This simple calculation is the second step toward the determination of the chaotic region.

Our goal was to simplify the construction of the bifurcation diagram by using the circuit simulator. The process of generating the bifurcation diagram was modelled by the special circuit that was called the chaoticness detector. With this bifurcation points based model we can easy detect the presence of circuit's chaotic behaviour.

This paper is structured as follows. In the second chapter the description of the bifurcation diagram and its use is presented. In the third chapter we describe in the detail the model of the chaoticness detector which comprises analogue memory where the peak values (*Fig. 3*) are stored and the model of time varying bifurcation parameter. The basic components of the analogue memory model are two capacitors and switches. The switches ensure that within each signal's period the capacitors store the maximum and minimum values of the signal.

In the fourth chapter we present the chaoticness detection of the Chua's circuit /9/. While the simulation runs, bifurcation parameter alters in time from the selected start to stop values. Signals peak values are kept in the peak values memory and with appropriate time step they are also stored in the output file. Based on these stored data we can easy plot a bifurcation diagram (*Fig. 6*) that enables a parameter a_∞ calculation.

The chaoticness detector can be used as an effective tool for analysis of the chaotic behaviour on the level of circuit simulators. The concept of the detector can also be used as a starting point for the practical implementation.

1 Uvod

Odkritje determinističnega kaosa v sredini šestdesetih let je pritegnilo pozornost številnih raziskav različnih znanstvenih področij: biologije, meteorologije, fizike, astrologije, kemije, medicine, elektronike, kriptologije. Analize različnih nelinearnih sistemov potrjujejo obstoj fenomena kaosa z značilnimi univerzalnimi zakonitostmi, lastnostmi in pojavi /2/, /3/. Eden takšnih pojavov je pojav bifurkacij, katerih število postane v primeru kaotičnega obnašanja neskončno veliko.

Kaotičnost je mogoče zaznati /1/, /2/, /4/, /5/, /6/, /7/ npr. z opazovanjem trajektorij v prostoru stanj, z različnimi analizami časovnih potekov signalov v vezju, s Poincaréjevimi prerezi atraktorjev, z izračuni Lyapunovih eksponentov, z entropijo, z analizo bifurkacijskih diagramov itd.

V prispevku smo se osredotočili na detekcijo kaotičnosti elektronskih vezij s pomočjo bifurkacijskega diagrama. Ker obnašanje analognih vezij najpogosteje analiziramo s pomočjo simulatorjev analognih vezij /6/, smo izdelali model detektorja kaotičnosti, ki je primeren za uporabo v tovrstnih simulatorjih. Namenjen je izračunu ocene mejne vrednosti bifurkacijskega parametra, pri kateri se običajno obnašanje vezja spremeni v kaotično.

Delovanje modela temelji na načinu detektiranja kaotičnosti s pomočjo bifurkacijskih točk. Ta način je opisan v drugem poglavju. Bifurkacijske točke odčitamo iz bifurkacijskega diagrama, ki ga dobimo na osnovi analize, v vezju izbranega signala. Iz njega izločimo vse lokalne ekstremne vrednosti (minimume ali maksimume) in jih, v odvisnosti od spremenljivega bifurkacijskega parametra, grafično prikažemo v bifurkacijskem diagramu. Ker simulatorji ne omogočajo iskanja lokalnih ekstremnih vrednosti, je za to potrebno uporabiti še dodatna matematična orodja /8/, /7/, ki omogočijo naknadno analizo signalov, dobljenih z več predhodno izvedenimi simulacijami vezja, pri različnih vrednostih bifurkacijskega parametra. Dodatnim programom in naknadnim analizam se lahko izognemo z uporabo modela detektorja kaotičnosti, s katerim je mogoče bifurkacijski diagram tvoriti s samim simulatorjem.

Model detektorja kaotičnosti, ki je opisan v tretjem poglavju, sestavljata model časovno spremenljivega bifurkacijskega parametra in model analognega pomnilnika ekstremnih vrednosti.

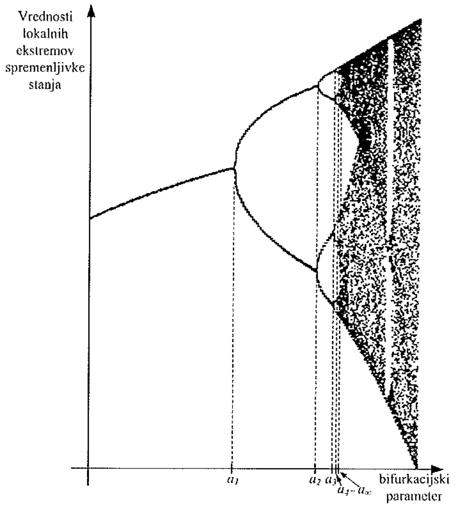
V četrtem poglavju je predstavljen konkretni primer detekcije kaotičnosti Chujevega oscilatorja /9/.

2 Detektiranje kaotičnosti

Kvalitativne spremembe v obnašanju nekega vezja, ki nastopijo zaradi spremembe njegovih parametrov, imenujemo *bifurkacije*. Parameter, s katerim je na te spremembe mogoče vplivati, imenujemo *bifurkacijski parameter*. Vrednosti bifurkacijskega parametra, pri katerih te spremembe nastopijo, imenujemo *bifurkacijske točke*.

Bifurkacij je več vrst /1/. Pri vezjih višjega reda so bifurkacije lahko zelo kompleksne. Različne vrste bifurkacij lahko vodijo v kaotično obnašanje vezja, če se le-te zaradi spremembe bifurkacijskega parametra pričnejo ponavljati.

Prehod bifurkacij v kaotični režim delovanja je mogoče zelo nazorno prikazati s pomočjo bifurkacijskega drevesa oziroma *bifurkacijskega diagrama*. Zgled takšnega diagrama prikazuje slika 1.



Slika 1: Bifurkacijski diagram in bifurkacijske točke $a_1, a_2 \ldots a''$.

Fig. 1: The bifurcation diagram and the bifurcation points $a_1, a_2 \ldots a''$.

Bifurkacijski diagram predstavlja odvisnost lokalnih ekstremnih vrednosti izbrane spremenljivke stanj od vrednosti bifurkacijskega parametra. Vrednosti bifurkacijskega parametra, pri katerih nastopijo kvalitativne spremembe obnašanja vezja, so označene kot bifurkacijske točke: $a_1, a_2 \ldots a''$. Pri vrednostih bifurkacijskega parametra manjših od a_1, se vezje obnaša kot običajen oscilator, kar kaže zvezno naraščajoč potek ekstremnih vrednosti. Spremenljivka stanja zavzame v tem primeru le eno ekstremno vrednost, kar pomeni, da vezje oscilira s konstantno amplitudo.

Pri bifurkacijskem parametru, večjem od a_1, nastopi prva cepitev - bifurkacija. Vezje se obnaša kot oscilator, katerega izhodni signal je sicer periodičen, znotraj periode pa sta prisotni dve različni amplitudi oziroma lokalna ekstrema. Naslednje cepitve, ki se pojavijo v točkah a_2, a_3, a_4 itd., so vse pogostejše, kar v praksi otežuje njihovo natančno določitev. Neperiodično, kaotično obnašanje vezja nastopi šele v točki a_∞, ko postane število različnih lokalnih ekstremov neskončno. Ker je to točko z meritvijo ali simulacijo nemogoče natančno določiti, jo lahko na osnovi univerzalne konstante *ä* ter poznavanja točk a_1 in a_2, le ocenimo /10/:

$$a_\infty = (a_2 - a_1) \cdot \left(\frac{1}{\delta - 1} \right) + a_2 \tag{1}$$

δ je Feigenbaumova konstanta oziroma limita vrste, ki jo tvorijo razmerja bifurkacijskih točk:

$$\delta \equiv \lim_{k \to \infty} \frac{a_k - a_{k-1}}{a_{k+1} - a_k} = 4{,}6692016\ldots \tag{2}$$

Gre za univerzalno in eksperimentalno določeno konstanto, ki se pojavlja v vseh kaotičnih sistemih.

Ker so bifurkacijske točke zmeraj povezane s to konstanto, lahko na osnovi poznavanja vsaj dveh ocenimo vrednost bifurkacijskega parametra a_∞ in tako ocenimo začetek kaotičnega vedenja vezja.

Detekcija kaotičnega režima delovanja vezja je torej povezana z iskanjem in določitvijo lokalnih ekstremnih vrednosti izbranega signala vezja, določitvijo bifurkacijskega diagrama ter z oceno bifurkacijske točke a_∞.

3 Model detektorja kaotičnosti analognih elektronskih vezij

Kaotično obnašanje vezij lahko analiziramo s pomočjo simulatorjev analognih vezij, v povezavi z dodatnimi matematičnimi orodji, ki omogočajo dodatno, naknadno obdelavo rezultatov simulacij. Tej se lahko izognemo z uporabo dodatnih modelov za simulator, ki omogočajo neposredno obdelavo vmesnih rezultatov simulacije. Tovrstni modeli predstavljajo vezja, ki se lahko priključijo ali dodajo k vezju, katerega obnašanje želimo analizirati. Običajno vsebujejo poleg osnovnih gradnikov (uporov, kondenzatorjev, tuljav, napetostnih in tokovnih virov...) še elemente, krmiljene na osnovi različnih matematičnih izrazov ali ukazov posebnega skriptnega jezika[1] /11/.

---

[1] *Npr. ICL - (Interactive Command Language) skriptni jezik simulatorja IsSpice4.*

Prispevek predstavlja model detektorja kaotičnosti, ki omogoča direktno tvorjenje bifurkacijskega diagrama s samim simulatorjem. Brez tega bi sicer potrebovali dodatni program, ki bi omogočal avtomatsko izvajanje simulacij vezja pri različnih vrednostih bifurkacijskega parametra ter detekcijo in določitev ekstremnih vrednosti v simuliranih časovnih potekih.

Model detektorja sestavljata:

- model časovno spremenljivega bifurkacijskega parametra in
- model analognega pomnilnika lokalnih ekstremnih vrednosti.

Bifurkacijski parameter je lahko v splošnem poljuben parameter vezja, npr.: upornost, kapacitivnost, ojačenje, temperatura, napetost. Njegovo spreminjanje v izbranem intervalu lahko zagotovimo s pomočjo spremenljive napetosti ali toka.

Izvedba modela časovno spremenljivega bifurkacijskega parametra je odvisna predvsem od vrste samega parametra in možnosti deklaracije njegovih vrednosti. Zaradi tega je potrebno model spremenljivega bifurkacijskega parametra prilagoditi obravnavanemu kaotičnem vezju in ga ni mogoče posplošiti.

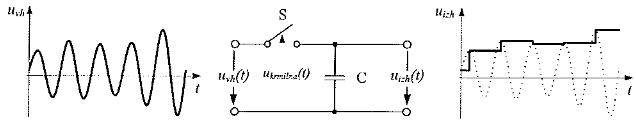
V četrtem poglavju je opisan primer modela časovno spremenljive upornosti, uporabljene kot bifurkacijski parameter Chujevega oscilatorja.

Ker simulatorji analognih vezij ne vsebujejo modelov, ki bi omogočali iskanje in shranjevanje posameznih ekstremnih vrednosti signalov, smo za model detektorja kaotičnosti potrebovali model analognega pomnilnika lokalnih ekstremnih vrednosti.

3.1 Model analognega pomnilnika ekstremnih vrednosti signala

Analogni pomnilnik ekstremnih vrednosti vhodnega signala $u_{vh}(t)$ omogoča začasno shranjevanje njegovih ekstremnih vrednosti. Ekstremna vrednost signala se najprej detektira in nato zadrži na izhodu vezja $u_{izh}(t)$ vse do naslednje, nove, ekstremne vrednosti. Način pomnjenja ekstremnih vrednosti temelji na polnjenju in praznjenju kondenzatorja C, z napetostjo vhodnega signala $u_{vh}(t)$, preko stikala S, kar prikazuje slika 2.

Stikalo S, ki je krmiljeno s krmilno napetostjo $u_{krmilna}(t)$, se mora vklopiti tik pred nastopom ekstremne vrednosti, izklopiti pa v trenutku, ko ekstremna vrednost nastopi. Trenutek vklopa stikala in časovna konstanta polnjenja kondenzatorja, določena z vklopno upornostjo stikala R_{ON} in kapacitivnostjo kondenzatorja C, morata zagotoviti, da se kondenzator v času vklopa napolni na napetost, čim bližjo ekstremni.



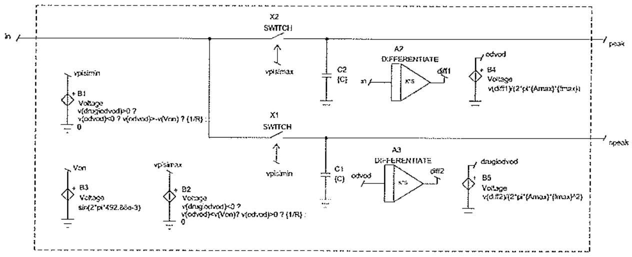
Slika 2: Način delovanja pomnilnika ekstremnih vrednosti. $u_{vh}(t)$ in $u_{izh}(t)$ - časovna poteka vhodnega in izhodnega signala.

Fig. 2: A signal peak values storage method. $u_{vh}(t)$ and $u_{izh}(t)$ are an input and an output signals.

V trenutku prisotnosti ekstremne vrednosti se stikalo izklopi. S tem se povezava med vhodom in kondenzatorjem prekine, kar povzroči praznjenje kondenzatorja preko izklopne upornosti stikala R_{OFF}. Če je ta dovolj velika, se vrednost napetosti na kondenzatorju, do naslednjega ponovnega vklopa stikala, bistveno ne spremeni. V takšnem primeru predpostavljamo, da se v času izklopljenega stikala na kondenzatorju ohranja konstantna, ekstremna vrednost vhodnega signala.

Kadar želimo shranjevati tako minimalne kot maksimalne ekstremne vrednosti, moramo uporabiti dva kondenzatorja in dve stikali, ki ju je potrebno ločeno krmiliti.

Model analognega pomnilnika lokalnih ekstremnih vrednosti vhodnega signala, primeren za simulator[2] SPICE, prikazuje slika 3. Na vhod (*in*) modela je priključen vhodni signal, iz katerega se izločijo njegove lokalne ekstremne vrednosti. Minimalne se ohranijo na izhodu *npeak*, maksimalne pa na izhodu *peak*.



Slika 3: Model analognega pomnilnika ekstremnih vrednosti signala, primeren za simulator IsSpice4.

Fig. 3: The IsSpice4 model of the peak values analog memory.

Kondenzator C1 služi začasnemu shranjevanju minimalnih vrednosti, kondenzator C2 pa maksimalnih. Z ustreznim krmiljenjem stikal X1 in X2 zagotovimo, da se oba kondenzatorja pravočasno napolnita in zadržita ekstremni vrednosti. Krmiljenje stikal se izvaja na osnovi izračunanih trenutnih vrednosti prvega in drugega odvoda vhodnega signala. Oglejmo si podrobnosti.

Ekstremne vrednosti vhodne napetosti nastopijo zmeraj v trenutkih, ko je vrednost prvega odvoda vhodnega signala

---

[2] IsSpice4 verzija 8.11.

enak nič, detekcija lokalnih minimumov oziroma maksimumov pa zahteva še poznavanje vrednosti drugega odvoda vhodnega signala. Obstoj lokalnega minimuma je pogojen s pozitivno vrednostjo drugega odvoda, lokalnega maksimuma pa z negativno vrednostjo. Izračun trenutnih vrednosti prvega in drugega odvoda vhodnega signala omogočata diferenciatorja A2 (prvi odvod) in A3 (drugi odvod).

Zaradi lažjega razumevanja krmiljenja stikal v nadaljevanju predpostavimo, da je vhodni signal *u_{vh}(t)* kosinusne oblike, s časovno spremenljivo amplitudo *A(t)* in frekvenco *f(t)*:

$$u_{vh}(t) = A(t) \cdot \cos(2 \cdot \pi \cdot f(t) \cdot t) \tag{3}$$

Njegov prvi odvod opisuje izraz:

$$\begin{aligned} \frac{du_{vh}(t)}{dt} &= \frac{dA(t)}{dt} \cdot \cos(2 \cdot \pi \cdot f(t) \cdot t) - \\ &-2 \cdot \pi \cdot A(t) \cdot \left(\frac{df(t)}{dt} \cdot t + f(t) \right) \cdot \sin(2 \cdot \pi \cdot f(t) \cdot t) \end{aligned} \tag{4}$$

Če predpostavimo, da so časovne spremembe amplitud in frekvence dovolj majhne, se izraz (4) lahko poenostavi:

$$\frac{du_{vh}(t)}{dt} \approx -2 \cdot \pi \cdot A(t) \cdot f(t) \cdot \sin(2 \cdot \pi \cdot f(t) \cdot t) \tag{5}$$

Vidimo, da je amplituda časovnega poteka odvoda v takšnem primeru kar sorazmerna produktu trenutnih vrednosti frekvence in amplitude vhodnega signala.

Če želimo stikali X1 in X2 krmiliti z vrednostmi odvodov, je potrebno te predhodno normirati. Normiramo jih s predvideno maksimalno vrednostjo frekvence f_{max} in amplitude A_{max} vhodnega signala: $(2 \cdot \pi \cdot A_{\max} \cdot f_{\max})$. Na podoben način izvedemo še normiranje drugega odvoda. Normiranje prvega in drugega odvoda izvedeta napetostno krmiljena vira B4 oziroma B5.

Stikali X1 in X2 sta neposredno krmiljeni s krmilnima napetostma v vozliščih *v(vpisimax)* in *v(vpisimin)*, ki jih dajeta krmiljena vira B1 in B2. Vklopljeni sta, kadar sta krmilni napetosti pozitivni, izklopljeni kadar sta enaki nič.

Upornost stikal ob vklopu R_{ON} je obratno sorazmerna vrednosti krmilne napetosti. Če izberemo, da je vrednost krmilne napetosti ob vklopu $1/R_{ON}$ in, da je $R_{ON}=1\Omega$, je časovna konstanta polnjenja kondenzatorja:

$$\tau_{ON} = R_{ON} \cdot C \tag{6}$$

določena samo z vrednostjo kondenzatorja C.

Napetost krmiljenega vira B1 skrbi za vklop in izklop stikala X1 oziroma določa interval polnjenja kondenzatorja C1 ter interval zadrževanja minimalne vrednosti na izhodu *npeak*. Podobno funkcijo ima krmiljen vir B2, s katerim je krmiljen vklop in izklop stikala X2 oziroma interval polnjenja kondenzatorja C2 ter interval zadrževanja maksimalne vrednosti na izhodu *peak*.

Stikali se vklopita, če so izpolnjeni določeni pogoji. X1 je vklopljeno kadar je vrednost drugega odvoda pozitivna in hkrati vrednost prvega odvoda negativna in večja od negativne referenčne napetosti V_{ON}. Referenčno napetost zagotavlja napetostni vir B3 in je prisotna v vozlišču *v(VON)*.

Podobno velja za krmiljenje stikala X2 preko katerega se polni kondenzator C2. Vklopljeno je samo, če je vrednost drugega odvoda negativna in hkrati vrednost prvega odvoda pozitivna ter manjša od referenčne napetosti V_{ON}.

Referenčna napetost predstavlja tisto vrednost prvega odvoda, pri kateri se mora stikalo vklopiti, da se bo kondenzator v času "*t* napolnil na ekstremno vrednost napetosti vhodnega signala z dopustnim odstopanjem $\pm\Delta u$.

Če zahtevamo, da se pri maksimalni frekvenci vhodnega signala in pri spremembi ekstremne vrednosti iz 0V na 1V zadržana napetost na kondenzatorju ne sme razlikovati od prave ekstremne vrednosti za več kot ±0,001%, lahko izračunamo potreben čas polnjenja kondenzatorja C:

$$\Delta t = \frac{a\cos(1-\Delta u)}{360 \cdot f_{\max}} = \frac{a\cos(0{,}999)}{360 \cdot f_{\max}} = \frac{2{,}56}{360 \cdot f_{\max}} \tag{7}$$

Sedaj lahko izračunamo še referenčno napetost V_{ON} pri kateri se mora stikalo vklopiti, da bo čas polnjenja kondenzatorja enak Δt. Ker se kondenzator prične polniti zmeraj za čas Δt pred nastopom ekstremne vrednosti, ko je vrednost odvoda enaka nič, lahko ob upoštevanju izraza (7) in normiranega prvega odvoda, izračunamo vrednost referenčne napetosti:

$$\begin{aligned} V_{ON} &= \sin(2 \cdot \pi \cdot f_{\max} \cdot t_{ON}) = \sin\left(2 \cdot \pi \cdot f_{\max} \cdot \left(\frac{1}{2 \cdot f_{\max}} - \Delta t \right) \right) = \\ &= \sin\left(\pi \cdot \left(1 - \frac{2 \cdot a\cos(1-\Delta u)}{360} \right) \right) = 54{,}02 mV \end{aligned} \tag{8}$$

Vidimo, da je ob zgornjih zahtevah referenčna napetost konstantna oziroma neodvisna od frekvence vhodnega signala.

Zahtevajmo še, da se kondenzator napolni na 99,9% končne vrednosti v času $6{,}9 \cdot \tau_{ON}$. Če upoštevamo, da je ta čas enak času Δt in, da je $R_{ON}=1\Omega$, lahko izračunamo vrednost kondenzatorja *C=C1=C2* pri kateri bo, pri maksimalni frekvenci vhodnega signala, odstopanje zadržanih ekstremnih vrednosti od dejanskih, manjša od Δu=0,999V:

$$C = \frac{\Delta t}{6{,}9 \cdot R_{ON}} = \frac{a\cos(1-\Delta u)}{6{,}9 \cdot 360 \cdot f_{\max} \cdot R_{ON}} = \frac{1}{969{,}3 \cdot f_{\max}} \tag{9}$$

Pri frekvencah nižjih od f_{max} bo omenjeno odstopanje kvečjemu manjše od ±0,001%.

Kakor hitro pogoja za vklop stikal nista izpolnjena, stikali izklopita in s tem prekineta povezavo med vhodom in kondenzatorjema. V času izklopljenih stikal se na kondenzatorjih ohranita trenutni ekstremni vrednosti vhodnega signala. Ti se bistveno ne spremenita, če izberemo dovolj veliko izklopno upornost (R_{OFF}) stikal oziroma dovolj veliko izklopno časovno konstanto:

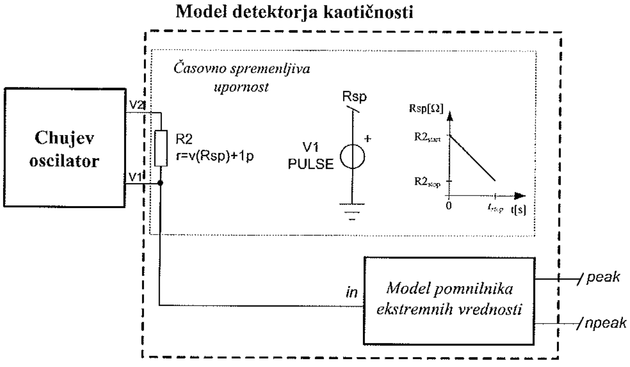
$$\tau_{OFF} = R_{OFF} \cdot C \qquad (10)$$

Natančnost opisanega modela pomnilnika je odvisna od dopustnega odstopanja zadržanih ekstremnih vrednosti, izbranih vrednosti časovnih konstant ob vklopu in izklopu stikal, dopustnega maksimalnega časa integriranja[3] in izbranega časa trajanja simulacije. Večja natančnost je pogojena z dlje trajajočo simulacijo.

4 Rezultati detekcije kaotičnosti Chujevega oscilatorja

V številnih znanstvenih prispevkih zasledimo različna vezja, ki se lahko obnašajo kaotično. Gre za preprosta RLC-vezja, raznovrstne oscilatorje, kapacitivno " preklopna vezja, digitalne filtre, flip-flope, adaptivna sita, napajalnike in pretvornike, močnostna vezja /1/. Najpogosteje obravnavano kaotično vezje je prav gotovo *Chujev oscilator* /9/ (slika 4), ki smo ga v prispevku obravnavali kot vzorčno vezje v katerem smo, s pomočjo modela detektorja kaotičnosti, detektirali mejo med periodičnim in kaotičnim osciliranjem.

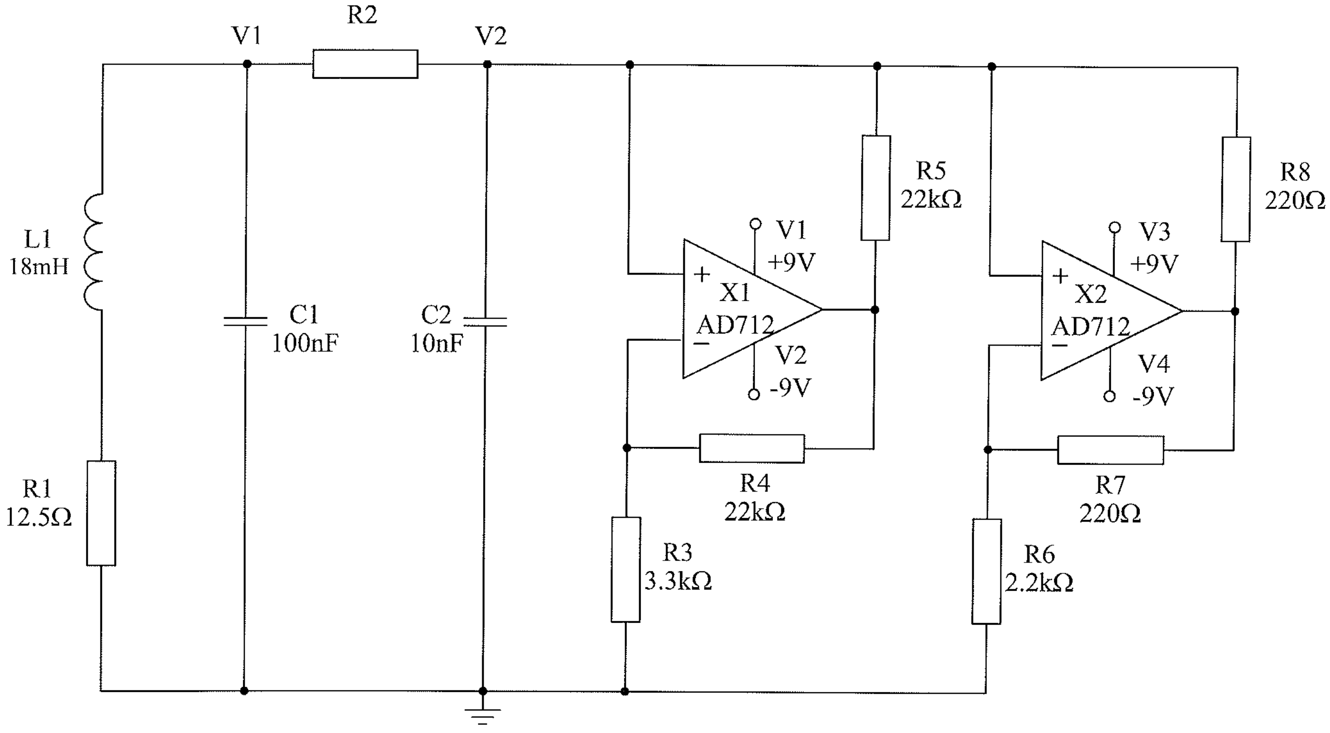
Detekcijo kaotičnosti smo izvedli na osnovi analize časovnega poteka napetosti[4] na kondenzatorju C1. Pri tem smo uporabili simulator IsSpice4, verzijo 8.11 /11/. Za bifurkacijski parameter smo izbrali[5] upornost R2, ki smo jo spreminjali v intervalu med 1800Ω≤R2≤1860Ω, s čimer smo dosegli vse pomembne kvalitativne spremembe v obnašanju vezja: običajno harmonično osciliranje vezja se je preko bifurkacij spremenilo v kaotično.



Slika 5: Model detektorja kaotičnosti v povezavi s Chujevim oscilatorjem.

Fig. 5: The Chua's oscillator wired to the model of chaoticness detector.

Mejno vrednost bifurkacijskega parametra, pri kateri se nekaotični režim delovanja vezja spremeni v kaotičnega, smo detektirali glede na enačbo (1), na osnovi prvih dveh bifurkacijskih točk v bifurkacijskem diagramu. Slednjega smo določili s simulacijo vezja, prikazanega na sliki 5. Ta



Slika 4: Chujev oscilator /9/.

Fig. 4: The Chua's oscillator /9/.

---

[3] *Parameter t_{max} pri časovni analizi.*

[4] *Enake rezultate bi dobili z analizo napetosti na kondenzatorju C2 ali toka skozi tuljavo L1.*

[5] *Bifurkacijski parameter je lahko tudi kapacitivnost kondenzatorja C2.*

slika prikazuje povezavo Chujevega oscilatorja z modelom detektorja kaotičnosti, ki ga sestavljata časovno spremenljiva upornost R2 in model analognega pomnilnika lokalnih ekstremnih vrednosti.

Časovno spremenljivo upornost smo modelirali s pomočjo časovno odvisnega napetostnega vira V1 in upornosti R2. Napetost vira V1 oziroma napetost v vozlišču v(Rsp) smo spreminjali linearno od vrednosti 1860 do 1800. Padajočo napetost smo povezali z upornostjo R2 tako, da smo vrednost upora R2 opisali z izrazom: r=v(Rsp)+1p.

Glede na osnovno frekvenco oscilatorja (f=3751Hz), določeno z vrednostjo tuljave L1 in kondenzatorja C1 ter pričakovano maksimalno amplitudo signala v opazovanem vozlišču (A_{max}=1V), smo izračunali potrebne parametre modela pomnilnika ekstremnih vrednosti.

Pri izbrani maksimalni frekvenci opazovanega signala f_{max}=3,8kHz, smo po enačbi (9) izračunali še vrednosti kondenzatorjev C1 in C2: C1=C2=C=271.5nF. Za vrednosti vklopne in izklopne upornosti stikal smo izbrali: R_{ON}=1Ω, R_{OFF}=1TΩ.

Izvedli smo časovno analizo (TRAN), z naslednjimi vrednostmi parametrov: *tstep*=1/3751, *tstart*=50·*tstep*, *tmax*=6,9·R_{ON}·C/9, *tstop*=4s. Z izbrano vrednostjo parametra *tstep* smo dosegli, da so se v izhodno datoteko shranjevale ekstremne vrednosti le enkrat na periodo opazovanega signala. S tem smo onemogočili nepotrebno shranjevanje velikega števila enakih ekstremnih vrednosti znotraj periode.

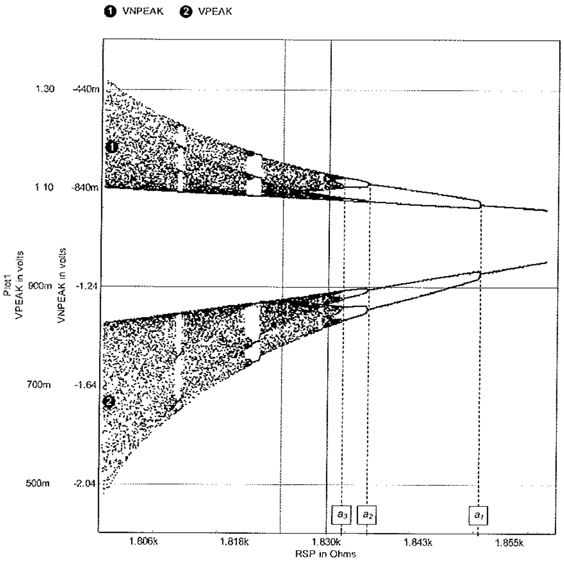
S parametrom *tstart* smo določili čas pričetka shranjevanja ekstremnih vrednosti v izhodno datoteko in s tem izločili nepotrebno informacijo o začetnem prehodnem pojavu v vezju.

Parameter *tmax*, ki določa maksimalni dopustni korak integriranja, smo izbrali glede na najmanjšo časovno konstanto v vezju. V našem primeru je to časovna konstanta polnjenja kondenzatorjev v modelu pomnilnika ekstremnih vrednosti, ki znaša 1,87μs.

Končni čas simulacije *tstop* smo izbrali tako, da se je v času simulacije, pri spremenljivi vrednosti upora R2, v izhodno datoteko shranilo dovolj ekstremnih vrednosti, potrebnih za določitev bifurkacijskega diagrama. Z izbranim časom *tstop*=4s smo tako dosegli, da se je upornost R2 vsako periodo zmanjšala za 4mΩ, kar predstavlja tudi izbran korak spreminjanja bifurkacijskega parametra.

Rezultat izvedene simulacije sta bifurkacijska diagrama, prikazana na sliki 6.

Pri bifurkacijskem diagramu, označenim s številko 1, predstavljajo vrednosti na ordinatni osi vrednosti minimalnih napetosti, pri diagramu označenem s številko 2 pa vrednosti maksimalnih napetosti na kondenzatorju C1. Vrednosti na abscisni osi predstavljajo vrednosti bifurkacijskega parametra oziroma upornost upora R2.



Slika 6: Bifurkacijska diagrama Chujevega oscilatorja, dobljena s simulatorjem IsSpice4.

Fig. 6: The bifurcation diagrams of the Chua's oscillator, simulated with the IsSpice4 simulator.

Bifurkacijske točke, ki jih lahko odčitamo iz enega ali drugega bifurkacijskega diagrama, so zapisane v tabeli 1.

Table 1: The bifurcation points of the Chua's oscillator.

Bifurkacijska točka	Vrednost upora R2	Vrsta trajektorije
a_1	1851 25 Ω	Pojav limitnega cikla s periodo 2
a_2	1835,79 Ω	Pojav limitnega ci kla s periodo 4
a_3	1832 50 Ω	Pojav limitnega cikla s periodo 8

Natančnost položaja posameznih točk bifurkacijskega diagrama je odvisna predvsem od izbrane natančnosti merjenja lokalnih ekstremnih vrednosti, izbranega koraka bifurkacijskega parametra in maksimalnega časa integriranja. Z ozirom na prej izbrane zahteve pričakujemo, da se bifurkacijska točka pojavi kakor hitro se sosednji ekstremni vrednosti amplitud razlikujeta za več kot ±0,001%.

S pomočjo dobljenih bifurkacijskih točk in enačbe (1) smo ocenili, da bo osciliranje Chujevega oscilatorja kaotično, kakor hitro bo vrednost upornosti R2 manjša od 1831,57 Ω.

5 Sklep

Pri približevanju bifurkacijskega parametra k vrednosti, ki ločuje običajno in deterministično naključno - kaotično vedenje, postane vezje izredno občutljivo. Zaradi hiperobčutljivosti te mejne vrednosti natančno ni mogoče predvideti. S pomočjo bifurkacijskega diagrama jo lahko le ocenimo.

V prispevku smo predstavili model detektorja kaotičnosti, ki omogoča direktno tvorjenje bifurkacijskega diagrama s pomočjo samega simulatorja SPICE. Tvorjenje bifurkacijskega diagrama bi sicer zahtevalo uporabo dodatnega orodja, ki bi omogočalo avtomatsko izvajanje simulacij vezja pri različnih vrednostih bifurkacijskega parametra ter shranjevanje ekstremnih vrednosti signalov.

Pokazali smo, da lahko z ustreznimi parametri modela detektorja in časovne analize dosežemo 0,001% natančnost izračuna bifurkacijskih točk oziroma mejne vrednosti bifurkacijskega parametra, ki ločuje običajni režim delovanja vezja od kaotičnega.

Uporabnost modela detektorja kaotičnosti smo predstavili s primerom detekcije kaotičnosti Chujevega oscilatorja. Sicer je model detektorja mogoče uporabiti tudi pri ocenjevanju neponovljivosti rezultatov simulacij /8/ in kot izhodišče za praktično implementacijo.

6 Literatura

/1/ M. J. Ogorzałek: *Chaos and complexity in nonlinear electronic circuits*, World Scientific, Series A, letnik 22, 1997.

/2/ P.Faure, H. Korn: *Is there chaos in the brain? I. Concepts of nonlinear dynamics and methods of investigation*, letnik 324, št. 9, september 2001 , str. 773-793.

/3/ T. S. Parker, L.O. Chua: *Chaos: A tutorial for engineers*, Proceedings of the IEEE, letnik 75, št. 8, avgust 1987, str. 982-1008.

/4/ C.W. Wu, N. F. Rul'kov: *Studying chaos via 1-D maps-A Tutorial*, IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, letnik 40, št. 10, oktober 1993, str. 707 -721.

/5/ M. Suneel: *Electronic Circuit Realization of the Logistic Map*, Sadhana - Academy Proceedings in Engineering Sciences, Indian Academy of Sciences, letnik 31, št. 1, februar 2006, str. 69-78.

/6/ D. C. Hamill: *Learning about chaotic circuits with SPICE*, IEEE Transactions on Education, letnik 36, št. 1, februar 1993, str. 28 - 35.

/7/ M. Stork, J. Hrusak, D. Mayer: *Chaos in Simple Nonlinear Systems and Chaotic Systems Simulation and Implementation*, Applied Electronics, 2006, International Conference on 6-7 Sept. 2006 str. 193-196.

/8/ M. Šalamon, T. Dogša: *Problem neponovljivosti simulacij električnih vezij*, Informacije MIDEM, 2004, letnik 34, št. 1, str. 11-17.

/9/ M. P. Kennedy: *Three steps to chaos. II. A Chua's circuit primer*, IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, letnik 40, št. 10, oktober 1993, str. 657-674.

/10/ R. C. Hilborn: Chaos *and Nonlinear Dynamics, an Introduction for Scientists and Engineers*, Oxford University Press, 2000, Second Edition.

/11/ Intusoft: *ICAP/4 IsSpice4 user's guide*, Intusoft 1988-1996.

Doc. dr. Matej Šalamon
Izr. prof. dr. Tomaž Dogša
oba UNIVERZA V MARIBORU
FAKULTETA ZA ELEKTROTEHNIKO,
RAČUNALNIŠTVO IN INFORMATIKO
2000 Maribor, Smetanova 17, Slovenija
E-mail: matej.salamon@uni-mb.si, tdogsa@uni-mb.si

Prispelo (Arrived): 01.09.2008 Sprejeto (Accepted): 09.06.2009

UDK621.3:(53+54+621+66), ISSN0352-9045

A LOW-COST SINGLE-CHIP READOUT CIRCUIT FOR PH SENSING

Hasmayadi Abdul-Majid, Yuzman Yusoff, Rohana Musa, Tan Kong Yew,
Mohd-Shahiman Sulaiman

MIMOS Berhad, Technology Park Malaysia, Kuala Lumpur, Malaysia

Key words: ROIC, ISFET sensor, sensor interfacing circuit, single-chip sensor solution

Abstract: A readout interface circuit for ISFET-based pH sensor with temperature compensation is described. The ROIC, together with our in-house developed enhancement-type ISFET could lead to single-chip pH sensor solution, leading to low cost sensor for precision agri- and aqua-culture. Temperature coefficient (TCF) cancellation technique has resulted in ISFET temperature dependency of being a negligible 0.2mV/^{o}C over a temperature range of 0^{o}C to 65^{o}C - close to ideal case of 0 mV/^{o}C. Preliminary results of sub-blocks that were fabricated in earlier runs indicate close resemblance of test results to the simulated results hence the conclusion that the silicon is functioning properly.

Nizkocenovno vezje na čipu za odčitavanje pH vrednosti

Kjučne besede: ROIC, ISFET senzor, senzorsko vmesno vezje, senzor na enem čipu

Izvleček: V prispevku opišemo vmesno čitalno vezje (ROIC) za pH senzor na osnovi tranzistorja ISFET. Vezje ROIC skupaj z ISFET tranzistorjem lahko pripelje do cenene izvedbe pH senzorja na enem čipu, primernega za široko uporabo. Temperaturna odvisnost tranzistorja je zanemarljivih 0,2 mV/°C v temperaturnem območju 0°C do 65°C, zahvaljujoč tehniki ničenja temperaturnega koeficienta, TCF. Preliminarne meritve na izdelanih blokih kažejo na dobro ujemanje s simuliranimi rezultati.

1. Introduction

Advancement in the microelectronics and semiconductor technology has led to various research and development activities in the field of sensor and electronics interfacing. One such area is FET-based pH sensor development, or also known as Ion-Sensitive Field Effect Transistor (ISFET) development. Slight modification of a field effect transistor by replacing certain layers with pH sensitive materials such as silicon nitride, aluminum and tantalum has resulted in a miniaturized FET-based pH sensors (ion-sensitive field-effect transistor, ISFET)/1/. This also promises the possibility of lab-on-a-chip solution, i.e. integration of sensor and the electronics on the same silicon. The end product is targeted to be low cost and suitable for field deployment.

An ISFET sensor is cheap and has good sensitivity characteristics. These features, coupled with the fact that it is silicon-based, have made these sensors as one of the main candidate for single-chip implementation. On top of the positive sides of ISFET pH sensor, among the problems faced in ISFET pH sensor development is reproducibility, stability, drift and temperature dependency /2/. Improvements made in each of the problems will, thus, make the use of ISFET sensor more appealing.

Analog readout interface circuit (ROIC) plays an important role in acquiring and processing the signal produced by ISFET sensor and passing it to a data converter for further signal processing. Many research works in the area of ISFET ROIC design have been reported in open literature /3/-/5/. Some of the approaches used as part of the sensing mechanisms are bridge type technique /3/, complementary ISFET-MOSFET pair technique /4/ and Constant-Voltage-Constant-Current (CVCC) technique /5/. CVCC technique has the advantage in terms of robustness and suitability for precision agriculture application.

In view of precision agriculture application where the sensors are placed in a field for a long period, with periodic data collection time at a center, temperature effects and drifting are some of the issues that need to be handled by the ROIC circuit properly. This means, a need for temperature sensing and cancelling the effects either at the ROIC stage or at the DSP stage for a true pH reading. The latter could be expensive in terms of processing power, leading to the solution not being suitable for field deployment where power is a concern. To do correction due to temperature effects at the ROIC stage, work in /5/, although using CVCC technique but it uses resistor temperature sensor. Resistor temperature sensor is resistive-based, hence producing thermal noise. Due to very low signal being read from the sensor, the thermal noise generated might hinder a single-chip solution hence could potentially increase the cost of the sensor solution. Our work solves this problem through the use of parasitics bipolar transistor to sense

---

This work was supported by the Malaysian Government through the Ministry of Science, Technology and Innovation (MOSTI) and Ministry of Finance (MoF) under the 9th Malaysia Plan (9MP). All of the authors are with MIMOS Berhad, a national R&D institute located at Technology Park Malaysia, Kuala Lumpur, Malaysia

the temperature. /4/ cancels out the temperature effect by using complementary ISFET-MOSFET pair technique.

In this work, CVCC technique is implemented for sensing a signal from sensor due to its robustness. Parasitic bipolar is used as a temperature sensing element to correct the actual pH reading (a temperature-independent pH measurement). This also solves the problem faced by /5/, hence making it suitable for single-chip implementation.

This paper is divided into four sections. Section 2 briefly describes the ISFET pH sensor. The proposed readout interface circuit with temperature coefficient cancellation method is presented in section 3. The SPICE simulation results of readout interface circuit with and without temperature compensation are presented in section 4. Finally, the conclusions are summarized in section 5.

2. Isfet Device

A basic cross-section of an ISFET structure is shown in Fig. 1. It is similar to the conventional MOSFET device, except for the standard metal gate which is replaced by a reference electrode and an electrolyte /1/. In base or acidity (pH) sensor, a sensing membrane sensitive to hydrogen ion concentration such as Si_3Ni_4, Al_2O_3 and Ta_2O_5 is used as gate insulator /3,4,5/. A change of pH or hydrogen ion concentration in an electrolyte induces a change in the electric field in the insulator-semiconductor interface and hence changing ISFET threshold voltage (V_T). Thus, different levels of pH can be represented by the drain currents (I_{DS}) of an ISFET device.

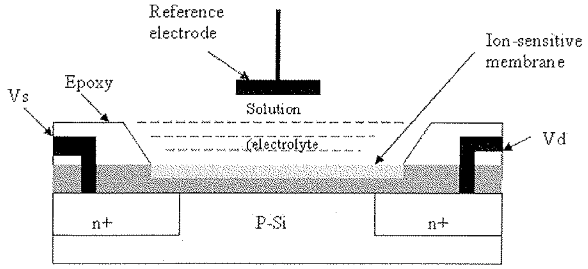


Fig. 1. Basic cross-section of Ion Sensitive Field Effect Transistor (ISFET)

To investigate the characteristic of ISFET, an established behavioral macro-model from previous publication is used in the design simulation. The model was firstly introduced by Sergio Martinoia and Guiseppe Massobrio in 1999 /6/ . It is based on a modified SPICE MOS transistor model in which the threshold voltage is influenced by the gate surface potential induced by hydrogen ions present in electrolyte flowing over the gate structure. It is suitable for the simulations of the device operating in a relatively wide range of temperature and hydrogen ion concentration.

3. pH Sensor interfacing circuit

The proposed ROIC for ISFET pH sensor with temperature compensation consists of three componets: a parasitic bipolar vertical pnp temperature sensor, constant-voltage-constant-current (CVCC) circuit and a summing circuit (Fig. 2). The principle of this readout circuit is to cancel the positive temperature coefficient (+ve TCF) of ISFET with the negative temperature coefficient (-veTCF) of parasitic bipolar vertical pnp device. Temperature coefficient of both devices can be combined through a summation circuit. As a result, a temperature-dependent characteristic of an ISFET is improved compared to the solution proposed by /5/. A non-temperature independent solution such as the one by /5/ is not suitable for deployment in an uncontrolled environment such as agriculture.

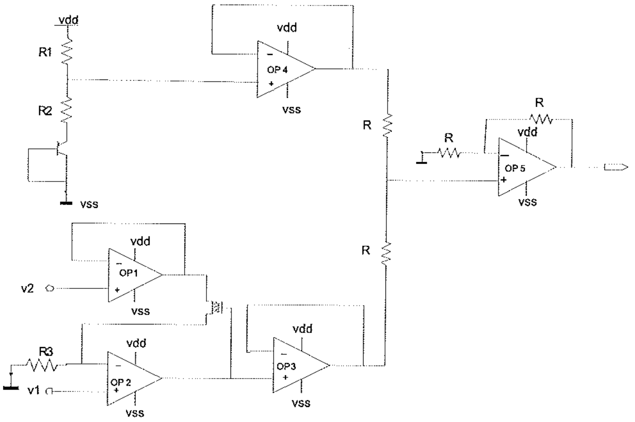


Fig. 2. Readout circuitry with temperature compensation

3.1 Constant Voltage Constant Current (CVCC) Circuit

In order to ensure the ISFET device operates in linear region at fixed VDS, a CVCC circuit is implemented. The schematic of this circuit is shown in Fig. 3.

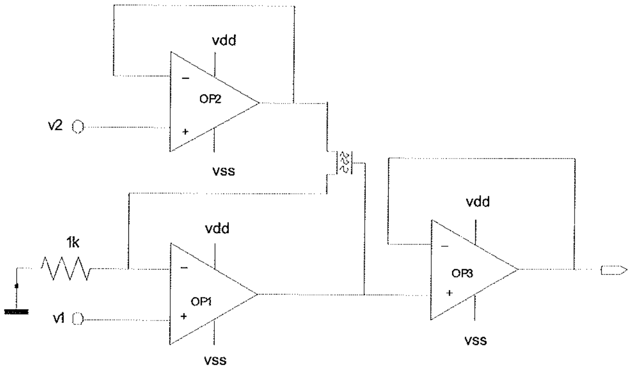


Fig. 3. Constant voltage constant current (CVCC) circuit

It consists of three operational amplifiers OP1, OP2 and OP3. OP1 is connected to the source of ISFET while OP2 is connected to the drain and provides a feedback to the ISFET gate. The gate voltage serves as an output signal in response to electrolyte solutions. In addition the feedback gate voltage, which is connected to a reference electrode will establish a stable electrolyte potential with respect to the surface potential of ISFET sensing gate. The varying of the gate voltage will restore the drain current as electrolyte parameters are changed. Another amplifier, OP3, serves as an output buffer.

3.2 Vertical PNP Temperature Sensor

In order to develop temperature sensor using CMOS compatible process with no required additional mask, a parasitic bipolar pnp device is employed. The circuit configuration of a bipolar pnp temperature sensor is shown in Fig. 4(a). Fig 4(b) shows the layout structure of bipolar vertical pnp. Area used for the emitter is 100µm2. Generally, the base-emitter voltage of parasitic bipolar pnp provides a negative temperature coefficient. By adding and controlling the value of the two resistors (R1 and R2) that are connected to the parasitic bipolar pnp, the temperature coefficient of the temperature sensor can be tuned closely to the negative of the ISFET temperature coefficient.

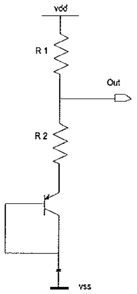


Fig. 4(a) Schematic of bipolar temperature sensor.

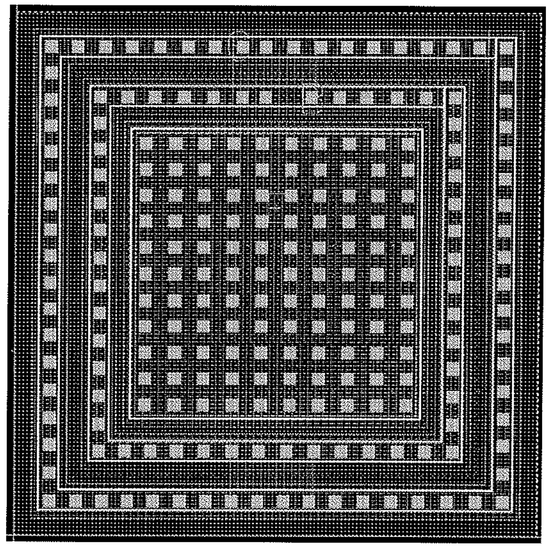


Fig. 4(b) Layout of bipolar vertical pnp device. © 2008 MIMOS Berhad

3.3 Summation Circuit

A summation circuit consists of a simple amplifier with several resistors. Fig. 5 illustrates the circuit configuration of the summation circuit employed in this work. This summation circuit is used to combine the output of ISFET and bipolar pnp temperature sensor. The positive temperature coefficient (+TCF) of ISFET is cancelled by adding to negative temperature coefficient (-TCF) of pnp temperature sensor.

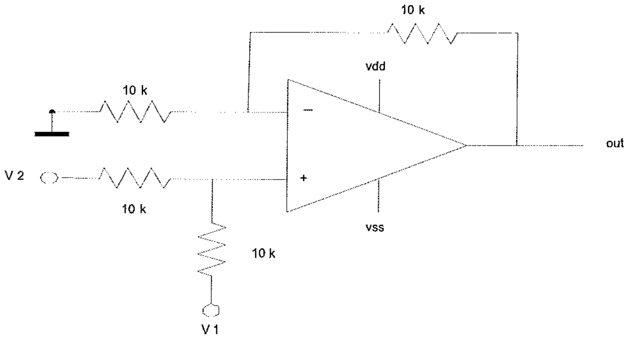


Fig. 5. Circuit schematic of summation circuit.

4. Results

In this work, the ROIC is simulated and fabricated using MIMOS 0.35µm process technology. Fig. 6 shows measured data for the ROIC. The average sensitivity is 50.16 mV/pH, showing adherence to the specification which is from 50 to 55 mV/pH. Fig. 7 shows measured output voltage for 3 different pH values, confirming the functionality of the ROICs. Fig. 8 shows test setup for the ROIC. Values for R1 and R2 are set to bias the ISFET's Vds to 0.5V. Fig. 9 and Fig. 10 show the simulation result of the output voltage signal from the ISFET and temperature sensor, respectively. This result is obtained for the temperature range of 0^{O}C to 65^{O}C. The simulated temperature coefficient (TCF) for ISFET is +1.420mV/^{O}C. Meanwhile the designed temperature sensor gives temperature coefficient (TCF) of -1.418mV/^{O}C.

The readout circuitry is completed by feeding both the temperature sensor and ISFET output voltages with different gains into a summing circuit that mutually offset their temperature coefficient and produce a temperature independent output signal. To validate the improvements of the proposed compensation method, the output voltage signal with and without compensation is shown in Fig. 11 and Fig. 12 respectively. For pH 4, 7 and 10, the corresponding temperature coefficient (TCF) before compensation in 0^{O}C to 65^{O}C ranges is 1.42mV/^{O}C. On the contrary, temperature coefficient (TCF) with compensation has a near zero slope (temperature independent) curves. Thus, the temperature sensitivity compensation method presents a more accurate pH measurement.

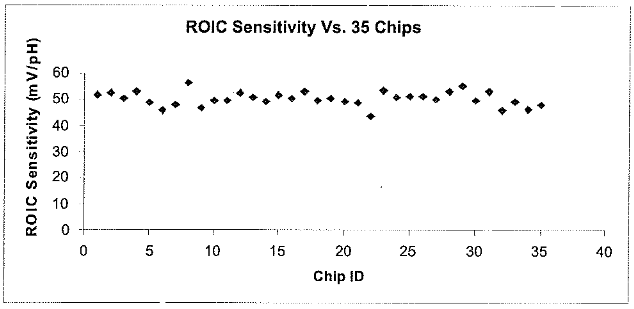


Fig. 6. Measured ROIC's sensitivity. © 2008 MIMOS Berhad

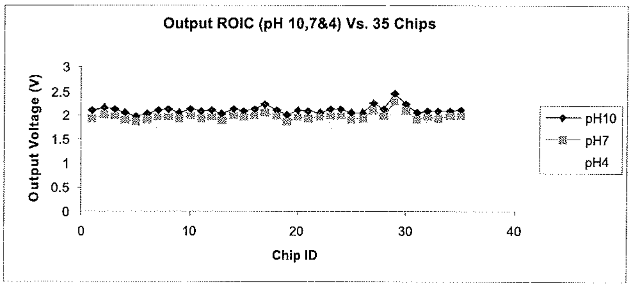


Fig. 7. Measured of ROIC's output voltage for three pH values. © 2008 MIMOS Berhad

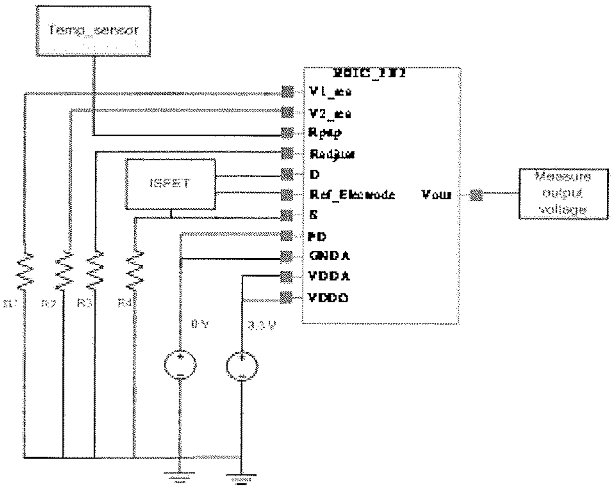


Fig. 8. Test setup for the ROIC. © 2008 MIMOS Berhad

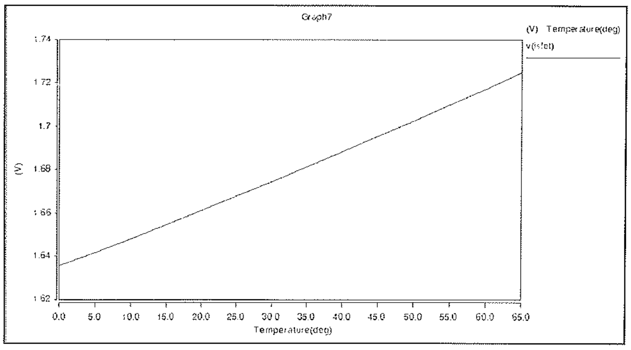


Fig. 9. Simulated temperature coefficient (TCF) for ISFET using CVCC configuration. © 2008 MIMOS Berhad

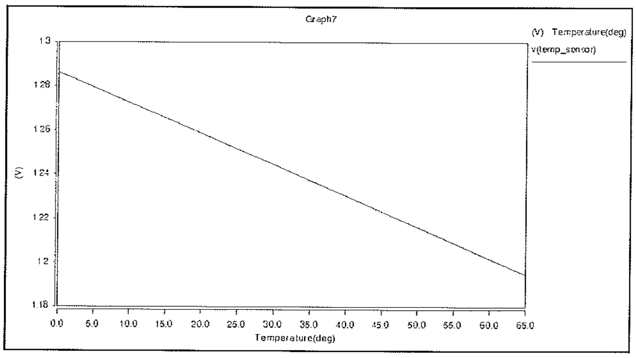


Fig. 10. Simulated temperature coefficient (TCF) for temperature sensor. © 2008 MIMOS Berhad

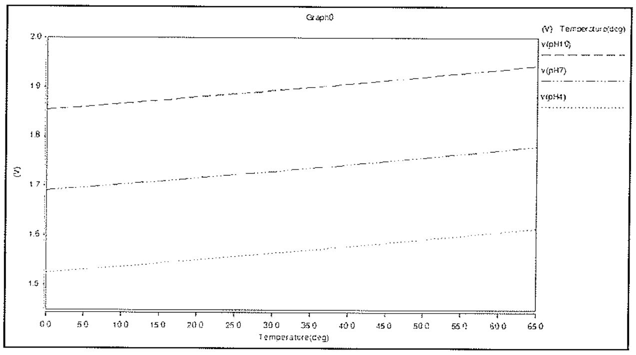


Fig. 11. Output signal voltage of readout circuit without temperature compensation. © 2008 MIMOS Berhad

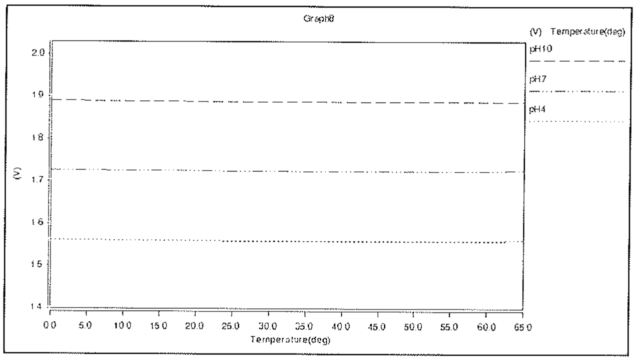


Fig. 12. Output signal voltage of ROIC with temperature compensation. © 2008 MIMOS Berhad

5. Conclusion

A readout circuitry for ISFET based sensor with temperature compensation has been developed and presented. Using temperature coefficient (TCF) cancellation technique, the ISFET temperature dependence is improved from 1.42mV/^{O}C to less than 0.2mV/^{O}C over a temperature range of 0^{O}C to 65^{O}C. The design has been targeted to MIMOS 0.35µm CMOS technology and has been fabricated as shown in Fig 13. Preliminary results of indicated close resemblance of test results to the simulated results hence we can safely conclude that the silicon is functioning properly and meeting all the specifications. The circuit, single-chip sensor solutions and the precision agriculture sensor solutions have been filed for patent under patent file: PI 20072136.

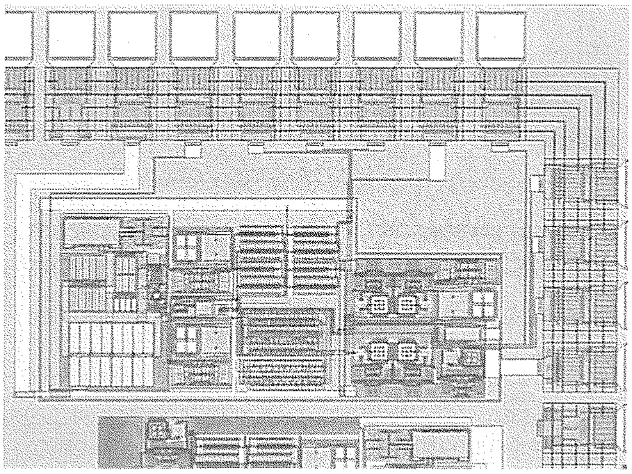


Fig. 13. ROIC on wafer. © 2008 MIMOS Berhad

6. References

/1/ P. Bergveld, "Development of an ion-sensitive solid-state device for neurophysiological measurement," IEEE Trans. On Biomed. Eng., vol. BME-17, pp. 70-71, 1970.

/2/ H. V. D. Vlekkert, L.Bousse and N.D. Rooij," The temperature dependence of the surface potential at the Al_2O_3/electrolyte interface, "J. of Colloid and interface Science, 122, pp.336-345, 1988.

/3/ Chung et al, "Electronic Circuit for Ion Sensor," U.S. Patent 6906524, Jun. 14, 2005.

/4/ Chodavarapu et al,"PH-Change Sensor and Method," U.S. Patent 2007/0138028, Jun. 21, 2007.

/5/ J. G. Connery et al, "Temperature Compensation for Potentiometrically Operated ISFETs," U.S. Patent 4879517, Nov. 7, 1989.

/6/ B. D. Liu, Y. K. Su and S. C. Chen, "Ion sensitive field effect transistor with silicon nitride gate for pH sensing," *Int. J. Electronics*, 67, pp. 59-63, 1989.

/7/ L. Bousse, H. V. D. Vlekkert and N. D. Rooij, " Hysterisis in Al_2O_3 gate ISFETs," *Sensors and Actuators B*, 2, pp.103-110, 1990.

/8/ D. H. Kwon, B. W. Cho, C. S. Kim and B. K. Sohn, "Effect of heat treatment on Ta_2O_5 sensing membrane for low drift and high sensitivity pH-ISFET," *Sensors and Actuators B*, 34, pp.441-445, 1996.

/9/ S. Martinoia and G. Massobrio, "A behavioral macromodel of the ISFET in SPICE," *Sensors and Actuators B*, 62, pp.182-189, 1999.

/10/ W. Y. Chung, C. H. Yang, M. C. Wang, D. G. Pijanowska and W. Torbicz, "Temperature compensation electronics for ISFET readout application," *IEEE International Workshop on Biomedical Circuit and system*, 2004.

/11/ W. Y. Chung, Y. T. Lin, D. G. Pjanowska, C. H. Yang, M. C. Wang, A. Krzyskow and W. Torbicz, "New ISFET interface circuit design with temperature compensation," *Elsevier, Microelectronic Journal 37*, pp.1105-1114, 2006.

/12/ MIMOS Malaysian Patent Pending PI 20072136.

Hasmayadi Abdul-Majid, Yuzman Yusoff, Rohana Musa, Tan Kong Yew Member IEEE, *Mohd-Shahiman Sulaiman* Senior Member IEEE
MIMOS Berhad, Technology Park Malaysia,
57000 Kuala Lumpur, Malaysia

Prispelo (Arrived): 04.12.2008 Sprejeto (Accepted): 09.06.2009

UDK621.3:(53+54+621+66), ISSN0352-9045 Informacije MIDEM 39(2009)2, Ljubljana

OPTIMIZATION OF THE MAGNETIC FIELD IN A MAGNETIC REFRIGERATOR

Jaka Tušek[1], Alen Šarlah[1], Alojz Poredoš[1], Dušan Fefer[2]
[1]Faculty of Mechanical Engineering, University of Ljubljana, Ljubljana, Slovenia
[2]Faculty of Electrical Engineering, University of Ljubljana, Ljubljana, Slovenia

Key words: magnetic refrigeration, magnetic field, permanent magnets, FEMM program, Hall sensor

Abstract: In this article we describe the development and the optimization of the structure for generation the magnetic field in the magnetic refrigerator. This refrigerator is located in the Laboratory for Refrigeration (LHT) at the Faculty of Mechanical Engineering, University of Ljubljana. Initially, we carried out a numerical simulation of the magnetic field that is generated by the structure, using the FEMM program, and then optimized its geometry. In the second part, we analyzed measurements of the magnetic field density in the device, using a magnetometer with a three-axis Hall sensor, with the aim of confirming the numerical results. The values of the magnetic field density were 0.97 T (measured) and 0.98 T (numerical). At the same time we were able to confirm the suitability of the FEMM program for estimating the static magnetic field, which is generated with magnetic circuits.

Optimiranje magnetnega polja v magnetnem hladilniku

Kjučne besede: magnetno hlajenje, magnetno polje, permanentni magneti, program FEMM, Hallova sonda

Izvleček: Članek prikazuje razvoj in optimiranje strukture za ustvarjanje magnetnega polja v magnetnem hladilniku, ki smo ga izdelali v Laboratoriju za hladilno tehniko (LHT) na Fakulteti za strojništvo, Univerza v Ljubljani. Sprva smo izvedli numerično simulacijo magnetnega polja, ki ga takšna struktura ustvarja (program FEMM) ter na osnovi tega tudi optimizacijo same geometrije strukture. V drugem delu smo, v želji po potrditvi numeričnih rezultatov, prikazali še meritve gostote magnetnega pretoka v magnetnem hladilniku (magnetometer s tri-osno Hallovo sondo). S tem smo v celoti ovrednotili obravnavano strukturo ter prišli do zaključka, da le-ta ustvarja 0,97 T (izmerjeno) oziroma 0,98 T (numerično) gostote magnetnega pretoka. Hkrati smo potrdili primernost programa FEMM za oceno vrednosti statičnega magnetnega polja, ki ga ustvarjajo strukture na osnovi magnetnih krogov.

1. Introduction

Magnetic refrigeration at room temperature is a new, developing technology for refrigeration that could, in the next few years, represent a viable alternative to the vapor-compression refrigeration technology in widespread use today. It is important to realize that, despite continuous development, classical, vapor-compression refrigeration technology is still very energy inefficient and its operation makes use of ozone-depleting refrigerants.

For these reasons researchers and engineers working in refrigeration have started to investigate new technologies for refrigeration, among which the most promising is magnetic refrigeration. Magnetic refrigerators can be 15-20% more efficient than classical refrigerators, using a magnetocaloric material as the refrigerant and water, or even air, as the heat-transfer fluid.

Magnetic cooling has been widely used for refrigeration at very low temperatures, i.e., cryogenic refrigeration, for a long time, but in recent years magnetic refrigeration technology has become interesting for refrigeration at room temperature, too. This is due to the development of new magnetocaloric materials, including gadolinium (Gd) and its alloys and other magnetocaloric materials that exhibit the so-called giant magnetocaloric effect.

Magnetic refrigeration is based on the magnetocaloric effect, which is defined as the change in temperature of a material as it is subjected to an external magnetic field. So in addition to the magnetocaloric material a basic magnetic refrigerator requires the generation of a magnetic field in an air gap, where the magnetocaloric material can be situated during the operation of the magnetic refrigerator.

In general, we can generate a magnetic field in two ways. For the generation of a large magnetic field density, e.g., a few Tesla, we need to use electromagnets or even superconducting magnets; however, they consume a large amount of additional energy for their operation and, in general, they are expensive and have a large volume and mass. Another way to generate a magnetic field is to use permanent magnets, which can generate lower magnetic field densities than electromagnetic devices, but they do not consume any additional energy during their operation, which is a great advantage over electromagnetic devices. In the past few years, investigations have concentrated on developing stronger permanent magnets and constructing ferromagnetic cores for the generation of a magnetic circuit through an air gap.

2. The basis of magnetic refrigeration

The development of magnetic refrigeration started with the discovery of the magnetocaloric effect by Warburg in 1881 /1/. The magnetocaloric effect is now very well investigated and is applied in a variety of fields in cryogenics, i.e., refrigeration at temperatures below -70°C. However, in the past few years the use of the magnetocaloric effect and magnetic refrigeration at room temperature has become increasingly interesting. The main reason for this was the discovery and development of new materials and technologies that offered the possibility of many applications in magnetic refrigeration for industrial and domestic situations.

A magnetic refrigerator is composed of a porous magnetocaloric material, the so-called active magnetic regenerator (AMR), which is the "heart" of the magnetic refrigeration device. The basic elements of a magnetic refrigerator are the device for generating the magnetic field, two external heat exchangers and a heat-transfer fluid, which is used to transfer the heat from the magnetocaloric material over the external heat exchangers to the surroundings.

A magnetic refrigerator performs four basic processes during its operation. In general, a magnetic refrigeration cycle consist of the magnetization and demagnetization processes, which are achieved by changing the external magnetic field, and two other, middle, processes of fluid flow from a cold heat exchanger over the magnetocaloric material to a hot heat exchanger and vice versa.

2.1. The magnetocaloric effect

We can view the magnetocaloric effect as the heating or cooling of a magnetocaloric material as it is subjected to an external magnetic field.

For a better understanding of the magnetocaloric effect and its application in magnetic refrigeration, let us consider an adiabatic system or an isolated magnetocaloric material (Fig. 1). When the magnetocaloric material is not exposed to a magnetic field the magnetic moments in the material are disordered or randomly orientated. However, when a magnetic field is applied, the magnetic moments become oriented in the direction of the applied magnetic field. From the magnetic point of view the system has reduced magnetic entropy, which means that in an adiabatic system the temperature of the material must increase in compliance with the law of the conservation of energy. We can see the reverse process taking place when we remove the magnetocaloric material from the magnetic field, and the magnetic moments revert to random orientations, which causes an increase in the magnetic entropy and a corresponding decrease in the temperature. As a result of this the system cools down.

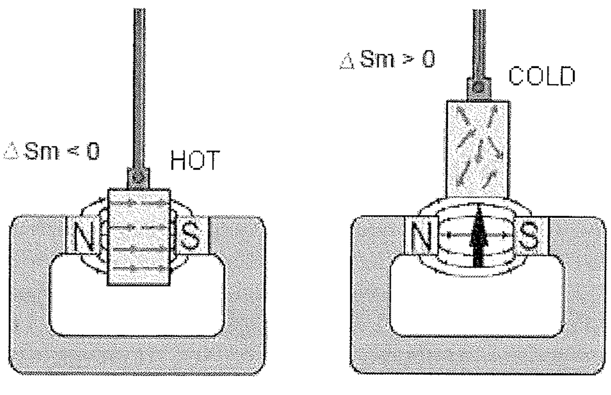


Fig. 1: Schematic illustration of the magnetocaloric effect

3. The form of permanent magnets for the generation of a magnetic field in a magnetic refrigerator

So far there have been 17 prototype magnetic refrigerators built and tested with the magnetic field being generated with permanent magnets. Of these 13 magnetic refrigerators had the magnetic field sources designed on the basis of a classical magnetic circuit and 4 had magnetic refrigerators designed on the basis of a Halbach array. /2/

3.1. Magnetic circuit

In general, the magnetic circuit is designed with a magnetic field source, which in our case was a high-quality permanent magnet, a magnetic flux conductor, usually a soft ferromagnetic material that conducts and directs the magnetic flux, and an air gap, which provides access to the generated magnetic field. A typical example of a magnetic circuit is shown in Fig. 2. The basic purpose of the magnetic circuit is to generate as large a magnetic field density in the air gap as possible. This is achieved by directing the magnetic flux in such a way that it forms a closed path, with the losses being as small as possible. In order to generate the largest possible magnetic field density in the air gap, this gap should be as small as possible to reduce the losses of the magnetic flux to the surroundings.

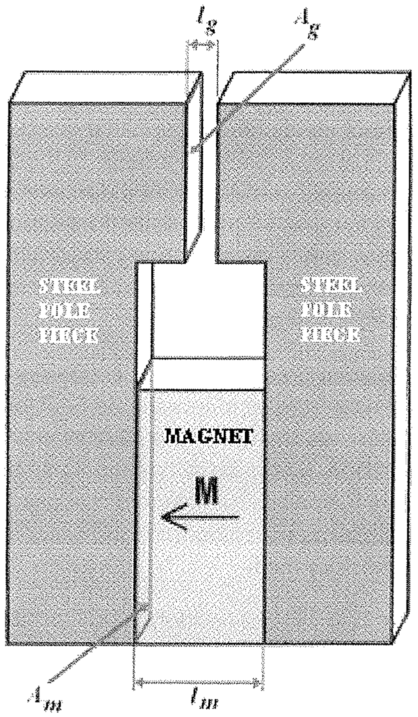


Fig. 2: Example of a simple magnetic circuit

3.2. Halbach array

Conventional structures of permanent magnets, like the magnetic circuit in Fig. 2, can generate a magnetic field density of up to 1 T in the air gap. Until the discovery of the so-called Halbach array in the 1970s, to generate larger magnetic field densities (of about 2 T and more) required the use of electromagnetic devices. However, permanent magnets arranged in a Halbach array can achieve magnetic field densities that were previously only possible with an electromagnet.

The basic concept of a Halbach array is in the structure of several permanent magnets that are magnetized in such a way as to cooperate in generating a stronger, more coherent magnetic field density, which can be even larger than the remanence of the permanent magnets being used. A circular Halbach array, which is the basis of some structures for the generation of a magnetic field in a magnetic refrigerator, is shown in Fig. 3. /3/

With the discovery of the Halbach effect a lot of researchers from all over the world started to investigate the effect and its applications. Since then there have been many modified Halbach arrays, with which even larger magnetic field density than are possible with the classical circular Halbach array have been reported. In the early 1990s Leupold et al. /4/ presented the possibility of augmenting the magnetic field density in the air gap by using a soft ferromagnetic material in the air gap of a circular Halbach array.

The development of magnetic refrigerators has indicated that modified Halbach arrays are very convenient for the needs of magnetic refrigeration. Among the modified Halbach arrays that are the most promising for magnetic refrigeration, two in particular stand out. Both were developed by Lee et al. /5, 6/ and could generate about 2.5 T of magnetic field density in a 10-mm air gap.

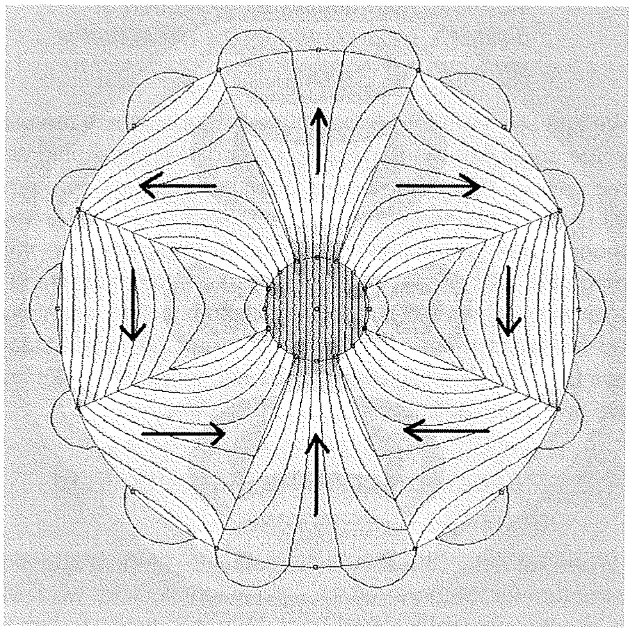


Fig. 3: Magnetic field density generated with a circular Halbach array of 8 permanent magnets

4. Development and analysis of the magnets' structure for the generation of a magnetic field in the magnetic refrigerator in the LHT

The generation of a magnetic field in the device for magnetic refrigeration in the Laboratory for Refrigeration (LHT) in the Faculty of Mechanical Engineering is designed on the basis of two parallel magnetic circuits (Fig. 5). This geometry was chosen based on price and the future trends in the development of magnetic refrigeration, which is focused on the development of rotating magnetic refrigerators with several regions of large magnetic fields /2/. In this way we can ensure a higher operating frequency of the magnetic refrigerator.

First, we carried out a numerical simulation of the magnetic field that is generated by the magnet structure and then we measured the magnetic field, with the intention being to confirm the numerical results. In this way we obtain the exact values of the magnetic field density in the magnet structure.

4.1. Numerical simulation

The finite-element method is the most useful way to numerically simulate a magnetic field because there are several commercial programs that can solve very complex problems without the need to develop specific algorithms. We decided to use the FEMM (Finite Element Method) program for the numerical simulation of the magnetic field. FEMM is able to solve two-dimensional, low-frequency electromagnetic problems /7/. In its solving phase the program solves the appropriate Maxwell equations for each node of the numerical mesh. The Maxwell equations are a system of six equations that describe the electromagnetic field. We only work with a static magnetic field, so only three of the Maxwell equations are required in our case. If we consider the magnetic vector potential (A_{mag}), we could write down only one equation, which is the basis for the numerical simulation of a static magnetic field with the FEMM program /7/:

$$\nabla \times \left(\frac{1}{\mu_0} \nabla \times \vec{A}_{mag} \right) = \vec{j} \tag{1}$$

Where μ_0 is the permeability of a vacuum, which is $4\pi \cdot 10^{-7}$ H/m, and j is the electric flux density.

We simulated the magnetic field that is generated by the structure of permanent magnets and soft ferromagnetic material with the FEMM program. First, we had to optimize the geometry, because we want to use the minimum amount of material, and at the same time we want to have the strongest magnetic field possible in the air gaps, which means in the areas where the magnetocaloric material would be magnetized, and the smallest magnetic field possible in areas where the magnetocaloric material would be demagnetized. We also optimized the height of the air gaps, because from the magnetic point of view we want to have the smallest air gap, but from the refrigeration point of view we want to use the largest amount of magnetocaloric material possible, and so for this reason we need to have the largest air gap possible. The optimization was performed by plotting charts of average magnetic field density in the air gap as a function of the different dimensions of the basic parts of the structure (the width and the height of the magnet, the thickness of the external ring and the height of the air gap) /8/. The scheme of the structure with optimized dimensions is shown in Fig. 4 in two-dimensional form. The depth of the structure in axial direction is 170 mm (external ring) and 90 mm (magnets and inner yoke), while in area of air gaps it is further focused on 55 mm.

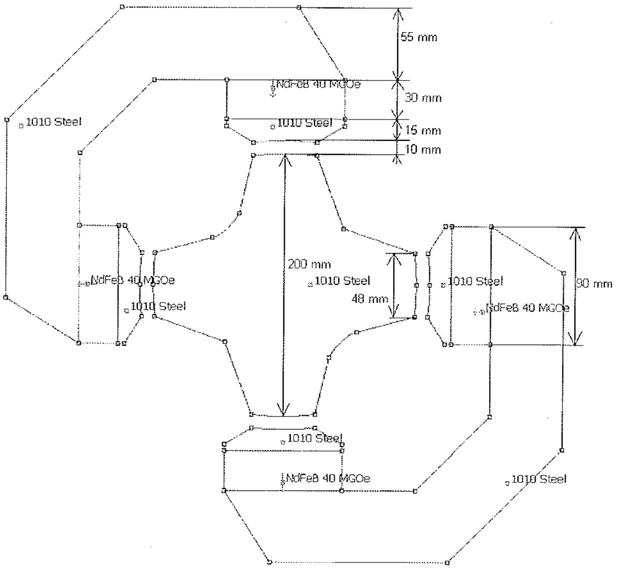


Fig. 4: The scheme of the structure for generating the magnetic field with marked optimized dimensions and used materials

The magnet structure is designed on the basis of four neodymium-iron-boron permanent magnets (Nd-Fe-B with 40 MGOe maximum energy product). These magnets are currently some of the strongest permanent magnets available, based on their maximum energy product, which is the most important factor when selecting a permanent magnet. As a soft ferromagnetic material for conducting and focusing the magnetic flux we used low-carbon 1010 steel, which is magnetically not ideal, but we chose it because of its low price and good forming properties. The structure has four air gaps with a strong magnetic field and four areas of low magnetic field where the magnetocaloric material is circulating during the operation of the magnetic refrigerator.

After defining the final geometry of the magnetic structure we simulated the magnetic field that is generated by it. When the program completed the calculations the results were outputted in graphical form with the distribution of the magnetic field density (Fig. 5) or with a graph of the magnetic field density (B) as a function of distance (Fig. 6); this represents the circle where the magnetocaloric material would be placed during the operation of the magnetic refrigerator and is shown in Fig. 5.

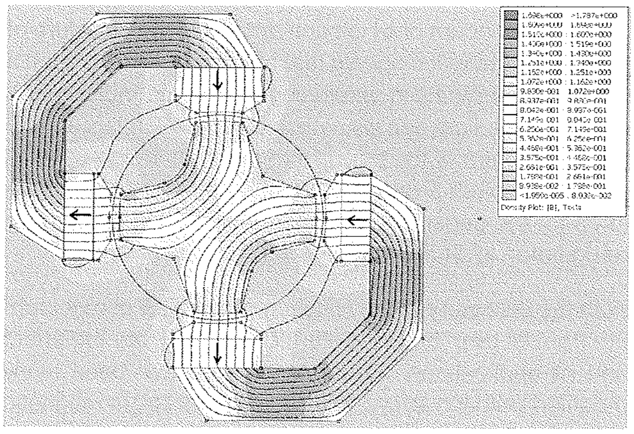


Fig. 5: Distribution of the magnetic field density

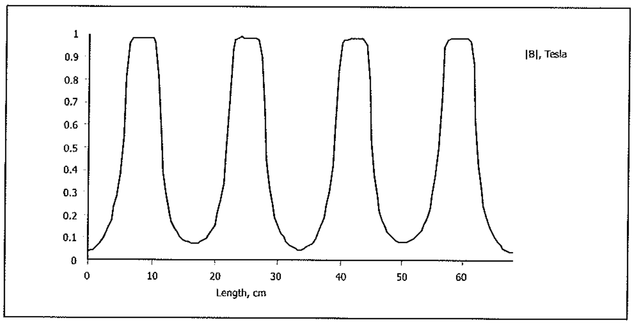


Fig. 6: Magnetic field density as a function of distance, represented by the circle that is shown in Fig. 5

We can see from Fig. 6 that in the air gaps, which means in the areas where the magnetocaloric material should be magnetized during the operation of the magnetic refrigerator, the magnetic field density is 0.98 T and suitably homogeneous for efficient operation. At the same time the magnetic field density in the areas where the magnetocaloric material should be demagnetized is a little less than 0.05 T, which means that the magnets and the carbon steel are far enough away from the demagnetization areas so we have a suitably low magnetic field density.

4.2. Measurement of the magnetic field density

When we established the final geometry of the magnetic structure for the generation of the magnetic field, we built it into the magnetic refrigerator. Out of a desire to know accurately the magnetic field we also measured the magnetic field density in the magnet structure.

The measurements were made using a three-axis magnetometer with an integrated three-axis Hall probe (SENIS transducer x-H3x-xx_E3D-2.5kHz-0.1-2T /9/), which is the most appropriate for this kind of measurement because of its accuracy and small dimensions.

The structure and the measurement points are shown in Fig. 7 (the front supporting plate covers the view to the basic elements of the structure). The magnetic field density was measured at 40 measuring points, which are marked in Fig. 7. The measurement points are arranged in such a way that three measurement points are in the middle of each air gap, two in the internal edge and two in the external edge of each air gap, while three measurement points are in each demagnetizing area. In this way we cover the whole of the circle in which the magnetocaloric material is situated.

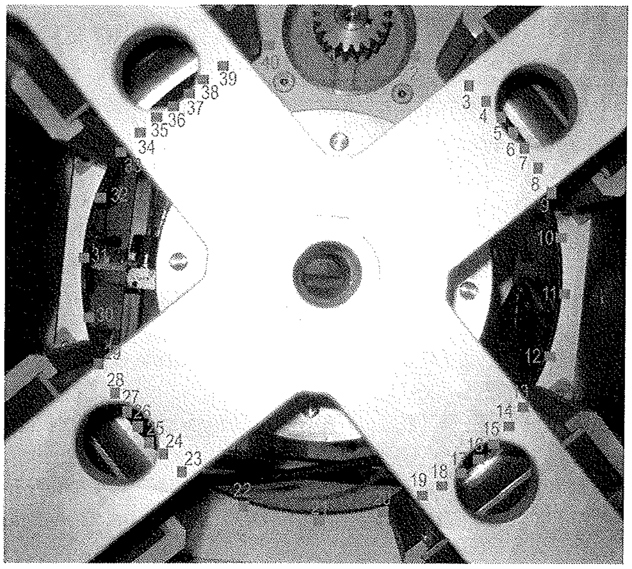


Fig. 7: Structure for generating the magnetic field in the magnetic refrigerator and the points where the measurements were made

At each of the marked measurement points we measured the magnetic field density and the results are shown in Fig. 8. For comparison the results obtained with the FEMM program are also presented in Fig. 8.

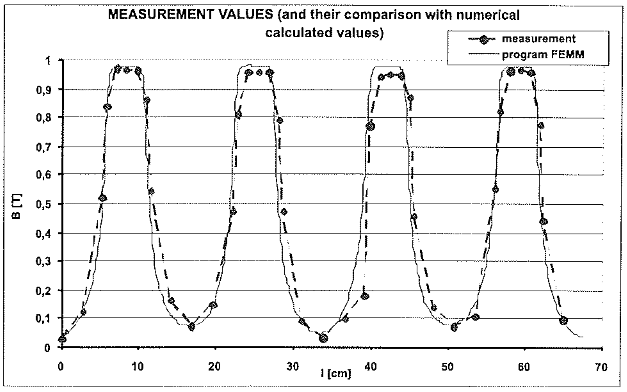


Fig. 8: Magnetic field density in the structure

It is clear that the magnetic field density in the air gaps is 0.97 T and in the demagnetizing areas it is around 0.05 T.

In addition, to estimate the homogeneity of the magnetic field density in the air gaps, which is very important for the efficient operation of the magnetic refrigerator, we also measured the magnetic field density at different heights (radial direction) and depths (axial direction) of the air gaps. We concluded that the magnetic field density varies a lot with the height in the air gap. In the middle of the air gap the magnetic field density was almost perfectly homogenous (Fig. 8), but close to the magnets and far from the magnets, which means on the upper and lower edges of the air gaps, the homogeneity of the magnetic field density is much worse and varies in the air gaps by as much as 0.2 T. At the same time the homogeneity of the magnetic field density is much better for different depths of the air gaps, because at the front and back edges of the air gaps the magnetic field density is 0.95 T.

The uncertainty in the measurement results is a combination of two factors. First, is the uncertainty due to the accuracy of the magnetometer, which is 0.1 % of the linear measurement range (0-2 T). On this basis we can calculate the relative measurement uncertainty due to the accuracy of the magnetometer, which is between ±0.1 % and ±2.3 %. The accuracy is the poorest in the air gaps and the best in the demagnetizing areas, where the values of the magnetic field densities are the smallest.

Second, is the uncertainty that is caused by the positioning of the magnetometer's probe during the measurement. We were not able to use mechanical positioning because of the compactness of the structure, which is why the measurement was made manually. The inaccuracy due to the positioning of the probe is the main contribution to the uncertainty of the measured values in the intermediate areas, where the inhomogeneity of the magnetic field density is at its greatest, whereas in the air gaps and the demagnetizing areas, because of the good homogeneity, the error in the positioning was negligible. We were able to estimate the absolute accuracy due to the positioning of the probe as ±2.5 mm. On this basis and with the distribution of the magnetic field density in the intermediate areas we calculated the relative measurement uncertainty in the intermediate areas due to the positioning of the probe to be between ±5 % (near the air gaps) and ±46 % (near the demagnetizing areas). This latter value is large and so in those areas the measured results are clearly not very accurate.

5. Conclusion

If we compare the results obtained with the FEMM program and the measured values of the magnetic field density (Fig. 8) we can conclude, on the basis of the numerical results, that our structure provides a 0.93 T change in the magnetic field density. On the basis of the measurement results the structure provides a 0.92 T change. The small difference between the values means that the agreement is very good, and so we can confirm the suitability of the FEMM program for estimating the magnetic field that is generated by the symmetrical magnetic circuits. The difference between the measured and the calculated values is in the range of the measurement uncertainty of the magnetometer we used and the uncertainty due to the positioning of the Hall probe. Furthermore, the reason for the deviation of the results can be attributed to three sources of error. First, is that the FEMM program allows only two-dimensional simulations, which can cause some error. Second, is that some changes were made to the structure, i.e., the inhomogeneity of the structure (e.g., the screws and the holders for the magnets), but these were not considered in the simulation. Third, we did not know accurately the properties of the permanent magnets and the carbon steel that were used in the simulation. This is why we used these assumed materials in the simulation.

With the development and analysis of the structure we were able to obtain sufficiently accurate values for the change in the magnetic field density that is possible with this magnet structure. This represents the basic information for further analyses and research on our magnetic refrigerator.

References:

/1/ A. M. Tishin, Y.I. Spichkin: The Magnetocaloric Effect and its Applications. Institute of Physics Publishing, London, 2003

/2/ K.A. Gschneidner Jr., V.K. Pecharsky: Thirty years of near room temperature magnetic cooling: Where we are today and future prospect. International Journal of Refrigeration, vol 31, 2008, str. 945-961

/3/ K. Halbach: Design of permanent multipole magnets with oriented rare earth cobalt material. Nuclear Instruments and Methods, vol. 169, 1980, str. 1 - 10

/4/ H.A. Leupold, A.S. Tilak, E. Potenziani,II: Adjustable Multi-Tesla Permanent Field Sources. IEEE Trans. Magn., 1993, str. 2902 - 2904

/5/ S.-J. Lee, J.M. Kenkle, D. Jiles: Design of Permanent - Magnet Field Source for Rotary - Magnetic Refrigeration Systems. IEEE Transactions on Magnetics, vol. 38, 2002, str. 2991 - 2993

/6/ S.-J. Lee, D. Jiles, K.A. Gschneidner, Jr., V. Pecharsky: Permanent magnet structure for generation of magnetic fields. United States Patent, Patent No.: US 6,680,663 B1; 2004

/7/ D. Meeker: Finite Element Method Magnetics, version 4.0., User's manual, 2006

/8/ Jaka Tušek: Developing of a regenerator for the magnetic refrigerator. University of Ljubljana, Faculty of Mechanical Engineering, Ljubljana, 2007

/9/ D.R. Popovic, S. Dimitrijevic, M. Blagojevic, P. Kejik, E. Schurig, R.S. Popovic: Three-Axis Teslameter With Integrated Hall Probe. IEEE Transactions on Instrumentation and Measurement, vol. 56, 2007, str. 1396 - 1402

Jaka Tušek, univ. dipl. inž. str.
Dr. Alen Šarlah, univ. dipl. inž. str.
Prof. Dr. Alojz Poredoš, univ. dipl. inž. str.

University of Ljubljana
Faculty of Mechanical Engineering
Laboratory for Refrigeration
Aškerčeva 6, SI-1000 Ljubljana, Slovenia
E-mail: jaka.tusek@fs.uni-lj.si

Prof. Dr. Dušan Fefer, univ. dipl. inž. el.

University of Ljubljana
Faculty of Electrical Engineering
Laboratory for Magnetic Measurements
Tržaška cesta 25, SI-1000 Ljubljana, Slovenia

Prispelo (Arrived): 12.03.2009 *Sprejeto (Accepted): 09.06.2009*

UDK621.3:(53+54+621+66), ISSN0352-9045 Informacije MIDEM 39(2009)2, Ljubljana

MODELING OF SURFACE ACOUSTIC WAVE CHEMICAL VAPOR SENSORS

Zdravko Živković[1], Marija Hribšek[1] and Dejan Tošić[2]

[1]Institute Goša, Belgrade, Serbia

[2]School of Electrical Engineering, University of Belgrade, Belgrade, Serbia

Key words: SAW, surface acoustic wave, vapor sensor, polymer

Abstract: New approach to modeling and analysis of transversal surface acoustic wave (SAW) chemical vapor sensors is presented. The sensor is modeled as a two-port device with parts represented by equivalent circuits. Change of output voltage, or frequency, as a function of vapor concentration is calculated. The model is general and includes propagation losses which are usually neglected in analysis methods. Closed form expressions for vapor concentration estimations are obtained. Simulation results are compared to experimental data. The approach enables better insight in the sensor operation and therefore the optimal design of vapor sensors.

Modeliranje SAW senzorja kemičnih hlapov

Kjučne besede: SAW, površinski akustični valovi, senzor hlapov, polimer

Izvleček: V prispevku predstavimo nov pristop k modeliranju in analizi transverzalnega SAW senzorja kemičnih hlapov. Senzor z dvema priključkoma modeliramo z nadomestnimi vezji. Izračunavamo spremembo izhodne napetosti ali frekvence v odvisnosti od koncentracije hlapov. Model je splošen in upošteva izgube, ki jih druge analize zanemarijo. Simulacije primerjamo z merjenimi rezultati. Naš pristop omogoča boljše razumevanje delovanja senzorja zatorej tudi možnost optimalnega načrtovanja kemičnih senzorjev hlapov.

1. Introduction

In the last two decades surface acoustic wave (SAW) chemical vapor sensors have found numerous applications due to their compact structure, high sensitivity, small size, outstanding stability, low cost, fast real-time response, passivity, and above all their ability to be incorporated in complex data processing systems. They can be used for *in situ* monitoring and sensing systems. /1,2,3/

The basic principle of SAW sensors is the reversible sorption of chemical vapors by a solvent coating which is sensitive to the vapor to be detected. It is interesting that a SAW-based sensor system is used as a volatile organic contamination monitoring system for the satellite and space vehicle assembly rooms in NASA. SAW sensors have been able to distinguish organophosphates, chlorinated hydrocarbons, ketones, alcohol, aromatic hydrocarbons, saturated hydrocarbons, and water /1/. SAW sensors are particularly useful for wireless monitoring in harsh environment /4/.

Surface acoustic waves were discovered in 1885 by Lord Rayleigh and are often named after him as Rayleigh waves /5/. A surface acoustic wave is a type of mechanical wave motion which travels along the surface of a solid material, referred to as substrate. The amplitude of the wave decays exponentially with distance from the surface into the substrate, so that the most of the wave energy is confined to within one wavelength of the surface /6,7/. The velocity of acoustic waves is typically 3000 m/s , which is much lower than the velocity of the electromagnetic waves.

A basic SAW device was originally developed in 1965 /8/. It consists of two interdigital transducers (IDTs) on a piezoelectric substrate such as quartz. Each IDT is a reversible element made of interleaved metal electrodes, which are used to launch and receive the waves: an electrical signal is converted to an acoustic wave and then back to an electrical signal. An IDT is a bidirectional transducer that radiates energy equally on both sides of the electrodes. Consequently, theoretical insertion loss introduced by an IDT is at least 6 dB.

Starting around 1970, different kinds of SAW devices were developed for applications in pulse compression radars, satellite communications and signal processing systems, mobile radio, and cellular telephones /9,10/. There are very broad ranges of commercial and military system applications that include components for radars, front-end and IF filters, CATV and VCR components, cellular radio and pagers, synthesizers and analyzers, navigation, computer clocks, tags, and many, many others /11,12/. SAW devices have found numerous applications outside their conventional fields, communications and signal processing. In the last two decades considerable work has been done in the development of SAW sensors of different types. High quality SAW filters are used as temperature, pressure and stress sensors /4,13/ as well as chemical and biosensors /14,15/. Generally, two types of chemical SAW sensors are used: transversal (or delay) and resonant.

Analysis of SAW devices can be approached to in three ways: (1) exact analysis by solving the wave equation, (2) approximate analysis by means of equivalent electro-mechanical circuits, and (3) approximate analysis via the delta function model /9,pp.55-122/. It is well known that the exact analysis of SAW devices using surface wave theory is very complex (even in the case of a free surface)

/6,7,10,14/. It starts from the second Newton's law and a set of partial differential equations. The equations are solved for the appropriate boundary conditions and relations between mechanical and electrical quantities of a piezoelectric substrate. In the most general case, e.g. in the presence of electrodes on the surface, the Maxwell's equations for the electromagnetic field should be taken into account, as well. Consequently, the exact analysis can be effectively applied only for IDTs with a small number of electrodes.

The simplest approximate method of analysis is based on the delta function model. It gives the results relatively fast, but its use is limited to small loads and substrates with lower coupling constants.

Better approximate methods use equivalent circuit models for IDTs, where the analysis tools known in electrical engineering can be applied. In these methods the accuracy depends on the complexity of the model. The closed-form solutions are derived for simple IDTs on quartz and lithiumniobate /9, pp.55-122,15/. Recent development of MEMS-based SAW chemical sensors, on new piezoelectric materials, also utilize equivalent circuits but only for modeling frequency characteristics of uniform IDTs /16,17/. In addition, closed-form solutions for more complex IDT structures have been developed by means of advanced electrical engineering analysis methods /18,19,20,21/.

One of the main objectives in a chemical sensor analysis is derivation of formulas which connect the change of electrical signals (e.g., voltages and frequency shifts) and chemical quantities (e.g., vapor concentration). The existing analysis approaches are usually: (a) the exact analysis via the wave equation /2,3,7/ and (b) the analysis based on published formulas derived from the wave equation /22,23/. The most complete treatment of the exact analysis has been reported in /7/. Generally, chemical SAW sensors have been analyzed mainly from the chemical point of view with less attention given to the relations between the electrical and chemical quantities and matching conditions at the electrical ports.

Typically, the analysis that is based on published formulas (which connect electrical signals and chemical compounds) neglects many properties of a real SAW delay line, such as propagation losses, technological constraints, and production tolerances. This is the reason why some researchers perform more experiments than needed, or have difficulties in explaining discrepancies between the expected and measured values /24/.

In this paper, a new modeling algorithm for the analysis of transversal chemical SAW sensors, based on the electrical equivalent circuit method, is presented. The algorithm develops in a straight forward manner explicit general relations between electrical signals, voltages or frequencies, and vapor detection estimations taking into account properties of real SAW devices, which are usually neglected. The whole sensor is modeled as a two-port network consisting of three parts: (1) the input interdigital transducer, (2) the delay line that is the sensing part, and (3) the output interdigital transducer. The transducers are modeled as three-port networks and the delay line as a two-port network.

This paper focuses on the essential problem of modeling the delay line with acoustically thin films and the influence of the gas concentration on its behavior and, consequently, on the sensor's output voltage or frequency. The proposed algorithm is used for the vapor concentration estimation and the results are compared with the experimental data. It has been shown that the concentration prediction is better if the properties of practical SAW devices are properly taken into account, especially at higher center frequencies.

2. Principles of SAW Sensor Operation

A transversal SAW chemical sensor can be schematically presented as in Figure 1. It consists of two interdigital transducers on a piezoelectric substrate, such as quartz. Piezoelectric materials are anisotropic, which will yield different material properties versus the cut of the material and the direction of propagation. Commonly used substrates are ST-cut quartz and lithiumniobate. ST-cut quartz crystal is cut at a specified angle (0°, 132.75°, 0°) to the crystallographic axes so that it has a small or vanishing dependence of wave velocity upon temperature at room temperature /7/. The SAW propagation is in the x-direction with velocity v = 3158 m/s. Usually, Y-cut lithiumniobate crystal is used. The propagation is in the z-direction with velocity v = 3488 m/s, but the temperature dependence is not negligible.

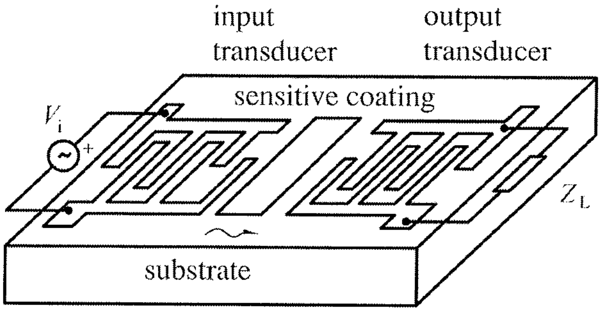


Fig. 1. The basic configuration of a chemical SAW sensor. (Acoustic absorbers and matching networks are not shown.)

A chemically sensitive thin layer is placed between the interdigital transducers on the top surface of the piezoelectric substrate. The surface wave is induced by an electrical signal applied to the input IDT.

The output signal (voltage) is taken from the output IDT. The velocity and attenuation of the wave are sensitive to mass and viscosity of the thin layer. The purpose of the thin layer – a polymer film – is to absorb chemicals of interest. When the chemical is absorbed, the mass of the polymer increases causing a change in velocity and phase of the acoustic signal, which causes a change in amplitude and frequency of the output voltage at the load impedance Z_L. Acoustic absorbers (not shown in Figure 1) should be appropriately placed on the substrate edges to damp unwanted SAW energy and eliminate spurious reflections that could cause signal distortions.

The IDTs are identical with uniformly spaced electrodes of equal lengths and equal ratio of electrodes width and spacing. The number of electrodes defines the frequency bandwidth of a SAW device. The electrode's length and number, and matching networks at the electrical ports, should be chosen to match the IDT input resistance, at the center frequency f_0, to the load resistance R_L and the generator resistance R_g. In that case, the overall minimal loss due to IDTs is 12 dB. The wavelength corresponding to the center frequency equals the distance between the electrodes of the same polarity. The center frequency and the bandwidth are determined by the IDT's geometry and the substrate type.

The middle part of a SAW sensor is a delay line, generally treated as lossless. However, its losses can be neglected only for lower frequencies and small delays (small distances between the transducers). The transfer function of the delay line is normally assumed unity, although this may not be true for high frequencies ($f > 0.5$ GHz) or if there are films in the propagation path /12,p.I.6-10/. In communications, in electrical filtering applications, the distance between the IDTs is small. Quite opposite, in chemical sensors this part is essential and must have a certain length, usually 100-200 wavelengths /7/, which should be taken into account.

3. New Model of SAW Chemical Sensors

The configuration presented in Figure 1 can be modeled by a general equivalent electro-mechanical circuit given in Figure 2. The IDTs are three-port networks and the sensing part is a two-port network designated by DL in Figure 2. The characteristic SAW acoustic impedance of the unloaded substrate is designated by Z_0 and the acoustic impedance due to the mass loading of the thin film is Z_m:

$$Z_0 = A\rho_s v \tag{1}$$

$$Z_m = A_m \rho_m v \tag{2}$$

where A is the substrate cross-section area through which the waves propagate, r_s is the mass density of the piezoelectric substrate, v is the SAW velocity in the piezoelectric substrate, A_m is the cross-section area of the thin film, and r_m is the mass density of the film. $Z_g = R_g$ and $Z_L = R_L$ are purely resistive electrical impedances of the generator and the electrical load, respectively.

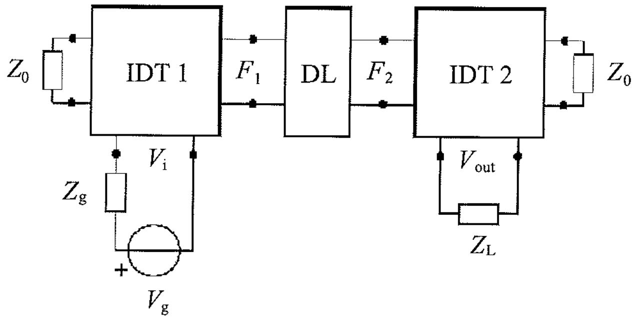


Fig. 2. The equivalent circuit of a SAW sensor.

Since the analysis of IDTs using equivalent electrical circuit models is well-known, the focus of this paper is to model the sensing part properties. The key observations relevant to the chemical sensor analysis are the following: (a) the sensor operates near the center frequency and (b) the IDTs are uniform with equal length electrodes. In that case, the IDT driving-point admittance at the electrical port, $Y_{IDT} = G_a(f) + jB_a(f) + j2\pi f C_0$, $j = \sqrt{-1}$, where C_0 is the static capacitance, can be calculated using well-known formulas /12,p.I.6-6/.

$$G_a(f_0) = 8k^2 f_0 C_s W_a N_p^2, \quad B_a(f_0) = 0 \tag{3}$$

where k is the piezoelectric coupling coefficient, f_0 is the center frequency, C_s is the capacitance per unit electrode length, W_a is the electrode length (that is, the width of the wave front), N_p is the number of electrode pairs. Equation (3) is used for designing proper matching of IDTs at the electrical ports. It should be noted that the characteristics of IDTs outside the narrow band around the center frequency are of no interest.

The output voltage across the load V_{out} is proportional to the mass loading of the sensing part. First, the output voltage in the presence of sensing material (polymer without vapor) is calculated and it serves as a reference voltage V_b, also referred to as the baseline voltage. The difference of the output voltage in the presence of vapor and the reference voltage is proportional to the vapor concentration. In some applications the output voltage is directly measured, but usually the sensor is a part of a more sophisticated system. In that case two equal SAW sensors are used: one is vapor-free and serves as a reference, the other one is exposed to vapor and actually performs the sensing function. The two SAW sensors are embedded into electrical oscillator circuits and the frequency shift between the oscillators is proportional to the gas concentration. Using an electronic circuit called the mixer the voltage proportional to the vapor concentration is obtained from the frequency shift. In any case the voltage across the load has to be found, which implies that the electrical transfer function of the sensor has to be determined.

According to Figure 2 the electrical transfer function can be expressed as follows:

$$T(f) = \frac{V_{out}}{V_g} = \frac{V_{out}}{F_2}\frac{F_2}{F_1}\frac{F_1}{V_g} \tag{4}$$

where F_1 and F_2 are mechanical forces.

Since the transducers are identical, V_{out}/F_2 is the conjugate complex value of F_1/V_g, and these terms are only functions of frequency. Therefore,

$$|T(f)|=|T_{13}(f)|^2\left|\frac{F_2}{F_1}\right| \tag{5}$$

where $T_{13}(f) = V_{out}/F_2$ represents the transfer function of the transducer. Fortunately, in this case, since the sensors work close to resonance in matched conditions at the input and output, the elaborate work of computing $T_{13}(f)$ can be avoided. Instead of Eq. (5) a much simpler expression can be used:

$$|T(f_0)|=|T_{13}(f_0)|^2\left|\frac{F_2}{F_1}\right| \tag{6}$$

At resonance, for a delay line without the polymer film, and negligible losses, $F_2/F_1 = 1$ and $|T_{13}(f_0)|^2 = 1/4 \approx -12\text{dB}$. Therefore, the relative variation of the output voltage V_{out} due to the mass loading is equal to the relative variation of F_2. The delay line of Figure 2 can be schematically represented as shown in Figure 3.

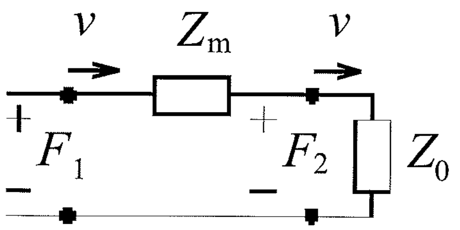


Fig. 3. The equivalent circuit of a mass loaded delay line.

By analogy between electrical and mechanical quantities, the relative variation of F_2 and the relative variation of velocity, for Z_m much smaller than Z_0, are determined by

$$\frac{\Delta F_2}{F_{20}}=\frac{\Delta v}{v}=\frac{F_2-F_{20}}{F_{20}}=\frac{-Z_m}{Z_0+Z_m}\approx\left.\frac{-Z_m}{Z_0}\right|_{Z_m<<Z_0} \tag{7}$$

where F_{20} denotes F_2 without mass loading, and v is the corresponding wave velocity. Using Eqs. (6) and (7) the variation of the output voltage due to the mass loading, can be calculated as follows:

$$\frac{\Delta V_{out}}{V_0}=-\frac{Z_m}{Z_0} \tag{8}$$

Where V_0 is the output voltage without the mass loading.

Using Eqs. (1),(2),(7),(8) and the well-known relationship between wavelength, velocity and frequency, the frequency shift due to the mass loading can be calculated as:

$$\frac{\Delta f}{f_0}=\frac{\Delta v}{v}=\frac{-Z_m}{Z_0}=\frac{\Delta V_{out}}{V_0}=-\frac{\rho_m h_m}{\rho_s \lambda_0}K_w \tag{9}$$

where ρ_m and h_m are the density and thickness of the thin layer, ρ_s is the density of the piezoelectric substrate, and K_w is a coefficient that depends on the technological process and implementation of the sensor. K_w is defined as the ratio of the polymer film width W_m and the width of the SAW front W_a : $K_w = W_m/W_a$. The components of the wave decay exponentially inside the substrate and the penetration is of the order of one wavelength. Therefore, in Eq.(9), instead of the substrate thickness, one wavelength λ_0 is used. From the last equation Δf can be determined as:

$$\Delta f=-\frac{\rho_m h_m}{\rho_s v}f_0^2 K_w \tag{10}$$

The last equation shows that the higher sensitivity will be obtained if the center frequency is higher, thickness and density of the film larger, and the substrate density and velocity smaller. This means that quartz ($\rho_s = 2.62$ g/cm^3) is a better choice for the substrate than lithiumniobate ($\rho_s = 4.7$ g/cm^3). Furthermore, if ST-cut quartz is used temperature dependence can be neglected /7/. Using the last equation the frequency shift, or the output voltage change, due to the polymer sensing film (without vapor) can be determined:

$$-\frac{\Delta V_{out}}{V_0}=-\frac{\Delta f}{f_0}=\frac{\rho_p h_p}{\rho_s v}f_0 K_w \tag{11}$$

where ρ_p and h_p are the density and thickness of the polymer, respectively. The reference voltage is

$$V_b=V_0-|\Delta V_{out}|=V_0\left(1-\frac{|\Delta V_{out}|}{V_0}\right) \tag{12}$$

Since ΔV_{out} is very small, V_b is very close to V_0.

Mass sensitivity S_m is an important characteristic of SAW sensors and is defined as $S_m = \Delta f/\Delta(\rho_p h_p)$ /22/. According to Eq.(11), assuming $K_w = 1$, it follows:

$$S_m=\frac{\Delta f}{\Delta(\rho_p h_p)}=\frac{1}{\rho_s v}f_0^2 \tag{13}$$

The mass sensitivity is determined only by the substrate (ρ_s and v) and the geometry of IDTs (f_0).

When vapor is absorbed, an additional voltage change occurs. Using the same reasoning and the fact that h_p is much smaller than λ_0, the voltage change due to vapor in sorbent phase, can be calculated as

$$-\frac{\Delta V_{vap}}{V_b}=\frac{\rho_{vap} h_p}{\rho_s v}f_0 K_w \tag{14}$$

where ρ_{vap} is the density of the vapor in sorbent phase. The sorbent phase of a volatile chemical compound is that part of the compound which is absorbed by polymer. Since the reference voltage shift without vapor ΔV_{out} is known, Eq.(11), the last equation can be expressed as follows:

$$\frac{\Delta V_{vap}}{V_b}=\frac{\rho_{vap}}{\rho_p}\frac{\Delta V_{out}}{V_0} \tag{15}$$

From Eq.(15) concentration of the chemical compound in vapor phase can be predicted using the known relation between the concentrations in sorbent and vapor phases /22/:

$$K = C_s / C_v \qquad (16)$$

where K is the partition coefficient, C_s is the concentration of the chemical compound in sorbent phase (in the sorbent coating /14,p.291/), and C_v is the concentration of the chemical compound in vapor phase (concentration in the ambient /14,p.291/). Vapor sensitivity depends on the choice of the sorbent coating material, polymer, and its strength of sorption, which is given by the partition coefficient K. The voltage shift ΔV_{vap} as a function of concentration C_v can be obtained as

$$\frac{\Delta V_{vap}}{V_b} = KC_v \frac{1}{\rho_p} \frac{\Delta V_{out}}{V_0} \qquad (17)$$

The concentration in Eq.(17) is in g/cm^3. It should be noted that different concentration units can be used in literature, but the units used here are as in /14,p.291/. Consequently, the value of K depends on the concentration units used. Equivalent relationship holds for Δf and the frequency shift due to vapor Δf_{vap}:

$$\Delta f_{vap} = KC_v \Delta f / \rho_p \qquad (18)$$

Equations (14)-(17) are derived for a delay line with negligible propagation losses. These equations contain normalized voltages (dimensionless quantities) so that they remain valid even for lossy delay lines. However, the propagation losses affect voltage shifts and voltages (quantities in Volts) and should be taken into account. The propagation loss is a nonlinear function of frequency, substrate and delay. According to /6,10/ the propagation loss for quartz can be calculated using

$$a_{1\mu s} = (2.15 f_{GHz}^2 + 0.45 f_{GHz}) \qquad (19)$$

where $a_{1\mu s}$ is the attenuation coefficient in dB/µs and f_{GHz} is the frequency in GHz. The propagation loss a_{dB} (in dB) is the product of the attenuation coefficient $a_{1\mu s}$ and the delay τ (in µs): $a_{dB} = a_{1\mu s}\tau$. Any voltage V or a voltage shift ΔV obtained from a lossless model should be divided by the factor $a = 10^{a_{dB}/20}$. If the desired accuracy in the voltage shift prediction is better than 1‰ (that is, better than 0.001), the propagation loss should be neglected only in the case when its value is less that 0.01 dB.

4. Simulation Results and Discussion

Prediction of mass sensitivity as a function of frequency is demonstrated first.

For the quartz substrate (ρ_s = 2.62 g/cm^3, v = 3158 m/s), according to Eq.(13), the frequency dependence of the mass sensitivity is calculated and presented in Figure 4.

As can be seen, the simulation results, based on the proposed model, are in a good agreement with the experimental results. The points in Figure 4 are obtained from Figure 6 reported in /22/ for the 158 MHz device of the transversal type and the (200, 300, 400) MHz devices of the resonator type.

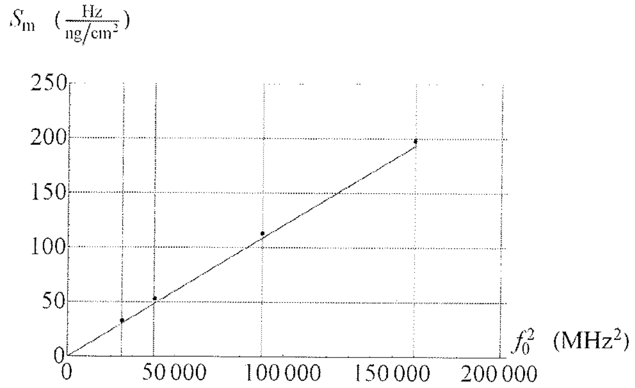


Fig. 4. Mass sensitivity versus f_0^2. Simulated results are represented by solid line and reported measured values /22,Figure 6/ are designated by dots.

Prediction of the frequency shift due to the polymer film on quartz is illustrated by the next example. Polymer is polyvinyl tetradecanal (PVTD) applied by Langmuir-Blodgett (LB) method /22/. Using Eq.(13), and the reported values for h_p = (79.5, 49.6, 22, 12.4)nm and ρ_p = 1g/cm^3, the frequency shifts for four different center frequencies are calculated and presented in Table 1 along with the experimental results /22/.

Table 1. Frequency shifts due to polymer loading.

Frequency	158 MHz	200 MHz	300 MHz	400 MHz
Predicted Δf	240 kHz	240 kHz	239 kHz	240 kHz
Measured Δf	247 kHz	243 kHz*	250 kHz*	205 kHz*

The relative error in Δf is (probably) caused by the fact that the experimental values of h_p are higher than the reported calculated ones /22/, which were used in the prediction. The "*" in Table 1 marks measured reported values for resonator type devices. However, the predictions are made for transversal devices, which explains the discrepancy between the experimental and predicted values at higher frequencies.

Finally, predictions of the frequency shift due to vapor are calculated. The polymer is PVTD and the vapor 1,2-dichloethane (DCE). From Eq.(18) one can conclude that for the same polymer, vapor concentration, and Δf the same value of Δf_{vap} should be obtained regardless the center frequency.

For the vapor concentration of C_v = 650.8 g/m^3 and the center frequency of f_0 = 158 MHz, the predicted frequency shift is Δf_{vap} = 16.6 kHz and the measured shift is about 13 kHz /22/. The corresponding partition coefficient $K = 10^{2.0279}$ is taken from the literature /1/.

The proposed algorithm is also applied to the calculation of vapor concentrations given in /1/. Vapor is trichloroethylene (TCE), and the polymer is PVTD. In this case, since the frequency is high, attenuation due to the propagation losses should be taken into account, but it does not effect the ratio $\Delta V_{vap}/V_b$. Using data from /1/ (f_0 = 500 MHz, $\log(K) = 2.65$, $\rho_p = 0.96$ g/cm^3, $\Delta f = 1.5$ MHz, $MW = 131.4$ g/mol, $V_m = 24.46$ l/mol) and Eq.(17), the vapor concentration is calculated as a function of $\Delta V_{vap}/V_b$. The results are presented in Figure 5, where dots represent the measured data /1/. According to /1, Figure 3/ the correction factor is $K_w = 0.76$. The prediction is in a good agreement with the experimental results.

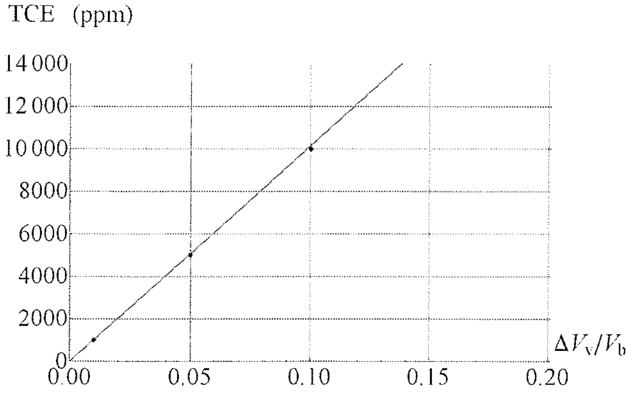


Fig. 5. TCE concentration versus the normalized voltage shift.

The proposed method is verified using experimental data from /24/. The experiments were carried out without any predictions, for devices on quartz at the center frequencies (39.6, 99, 132, 198, 264) MHz, different polymers, and three gases which simulate warfare chemical agents. Polymers were polyisobutylene (PIB), polyepichlorohydrin (PECH), and polydimethylsiloxane (PDMS) deposited by the spin coating technique. Space between the transducers was 1500μm with the aperture 1800μm. The characteristics were measured directly using E-5061A network analyzer.

Predictions are made at f_0 = 99 MHz for PECH and dichloromethane (CH_2Cl_2, DCM) with $\rho_p = 1.36$ g/cm^3, $h_p = 0.24$mm, $K = 10^{2.0743}$, $MW = 85$ g/mol, $V_m = 24.46$ l/mol. The delay τ is calculated from the distance between transducers d_{IDT} = 1500μm and the wave velocity v = 3158 m/s: $\tau = d_{IDT}/\mathrm{v} = 0.475\mu s \approx 0.5\mu s$. The correction factor 1/a due to attenuation loss is calculated according to Eq.(19), $f_{GHz} = 0.099$, $a_{1\mu s} = (2.15 f_{GHz}^2 + 0.45 f_{GHz})$, $\tau = 0.5$, $a_{dB} = a_{1\mu s}\tau$, $a = 10^{a_{dB}/20}$, $1/a = 0.996$, as expected since the frequency is lower than 500 MHz and the delay is only about 0.5μs. For 5 ppm concentration of DCM, the predicted frequency shift is 584 Hz and the measured value is 574 Hz /24/.

Difficulties were encountered in measuring the concentration of the same gas on PIB /24/. That can be explained by the proposed model: the constant K for PIB is about three times smaller than for PECH and therefore the detected voltage, which represents the gas concentration, will be also three times smaller, which gives insufficient voltage for the analyzer to be detected.

5. Conclusions

A new model for acoustically thin SAW transversal chemical vapor sensors has been developed. The model is based on electrical equivalent circuits of SAW devices, which connect electrical signals and chemical compounds, and takes into account important properties of real SAW devices, such as propagation losses, technological constraints, and production tolerances. The unique feature of the model is a set of closed form analytic expressions for vapor concentration estimations. The expressions explicitly relate the vapor concentration, substrate parameters, and center frequency. They enable insight into the influence of the sensor design parameters on the sensor performance.

The presented method predicts very efficiently and correctly the frequency and voltage shifts due to the vapor concentrations in chemical sensors. The simulation results, based on the proposed model, are in a good agreement with the experimental results. The results presented can be used in future for the design of optimal sensors for a given vapor.

Acknowledgements

The authors would like to thank the Ministry of Science and Technological Development of Serbia for financial support under the project number TR 11026.

References

/1/ Ho, C.K.; Lindgren, E.R.; Rawlinson, K.S.; McGrath, L.K.; Wright, J.L. Development of a Surface Acoustic Wave Sensor for In-Situ Monitoring of Volatile Organic Compounds. Sensors 2003, 3, 236-247.

/2/ Wohltjen, H.; Dessy, R. Surface acoustic wave probe for chemical analysis. Analytical Chemistry 1979, 51, 1458-1464.

/3/ Wohltjen, H. Mechanism of operation and design considerations for surface acoustic wave device vapour sensors. Sensors and Actuators 1984, 5, 307-325.

/4 / Pohl A. A Review of Wireless SAW Sensors. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 2000, 47, 317-332.

/5 / Rayleigh L. On waves propagated along the plane surface of an elastic solid. Proc. London Math. Soc. 1885, 17, 4-11.

/6 / Farnell, G.W. Elastic Surface Waves. In Surface Wave Filters; Matthews, H., Ed.; John Wiley: New York, USA, 1977; pp. 1-55.

/7 / Martin, S.J.; Frye, G.C.; Senturia, S.D. Dynamics and Response of Polymer-Coated Surface Acoustic Wave Devices: Effect of Viscoelastic Properties and Film Resonance. Analytical Chemistry 1994, 66, 2201-2219.

/8 / White, R.M.; Voltmer, F.W. Direct Piezoelectric Coupling to Surface Electric Waves. Appl. Phys. Lett. 1965, 7, 314-316.

/9 / Matthews, H. Surface Wave Filters; John Wiley: New York, USA. 1977; pp. 443-476.

/10 / Morgan, D.P. Surface Wave Devices for Signal Processing; Elsevier: London, UK. 1985; pp. 15-57.

/11 / Campbell, C. Surface Acoustic Wave Devices and their Signal Processing Applications; Academic Press: San Diego, USA. 1989, pp. 238-315.

/12 / Golio, M. The RF and Microwave Handbook, Second Edition; CRC Press LLC: Boca Raton, USA. 2008; pp. I.6.1-I.6.15.

/13 / Seifert, F.; Bulst, W.E.; Ruppel, C. Mechanical sensors based on surface acoustic waves. Sensors and Actuators 1994, A44, 231-239.

/14 / Ballantine, D.S.; White, R.M.; Martin, S.J.; Ricco, A.J.; Zellers, E.T.; Frye, G.C.; Wohltjen, H. Acoustic Wave Sensors: Theory, Design, Physico-Chemical Applications; Academic Press: San Diego, USA. 1997, pp. 1-7.

/15 / Smith, W.R., et al. Analysis of interdigital surface wave transducer by use of equivalent circuit model. IEEE Transaction on Microwave Theory and Techniques 1969, 17, 856-864.

/16 / Rufer, L.; Torres, A.; Mir, S.; Alam, M. O.; Lalinsky, T.; Chan, Y. C. SAW chemical sensors based on AlGaN/GaN piezoelectric material system: acoustic design and packaging considerations, in Proceedings of the 7th International Conference on Electronics Materials and Packaging, EMAP 2005, 2005, Tokyo, Japan, pp. 204-208.

/17 / Rufer, L.; Lalinský, T.; Grobelný, D.; Mir, S.; Vanko, G.; Őszi, Zs.; Mozolová, Ž.; Gregus, J. GaAs and GaN based SAW chemical sensors: acoustic part design and technology, in Proceedings of the 6th International Conference on Advanced Semiconductor Devices and Microsystems, ASDAM 2006, 2006, Smolenice, Slovakia, pp. 165-168.

/18 / Debnath, N.; Ajmera, J.C.; Hribšek, M.F.; Newcomb, R.W. Scattering and Admittance Matrices of SAW Transducers. Circuits, Systems and Signal Processing 1983, 2, 161-178.

/19 / Hribšek, M.; Tošić, D. An Improved Algorithm for Analysis of Uniform SAW Devices, in Proceedings of the 26th Midwest Symposium on Circuits and Systems, 1983, Puebla, Mexico, pp. 243-246.

/20 / Hribšek, M.; Tošić, D. An Improved algorithm for Analysis of SAW Pulse Compression Filters, in Proceedings of the 8th Microcoll Conference, 1986, Budapest, Hungary, pp. 373-374.

/21 / Hribšek, M. Transfer Function of a SAW Device with Apodized Transducers, in Proceedings of the International Symposium on Circuits and Systems ISCAS'82, 1982, Rome, Italy, pp. 636-638.

/22 / Grate, J.W.; Klusty, M. Surface Acoustic Wave Vapor Sensor Based on Resonator Devices, NRL Memorandum report 6829, May 23, 1991, pp. 1-38.

/23 / Grate, J.W.; Zellers, E.T. The Fractional Free Volume of the Sorbed Vapor in Modeling the Viscoelastic Contribution to Polymer-Coated Surface Acoustic Wave Vapor Sensor Responses. Analytical Chemistry 2000, 72, 2861-2868.

/24 / Joo, B.-S.; Lee, J.-H.; Lee, E.-W.; Song, K.-D.; Lee, D.-D. Polymer Film SAW Sensors for Chemical Agent Detection, in Proceedings of the 1st International Conference on Sensing Technology, Nov. 21-23, 2005, Palmerston North, New Zealand, pp. 307-310.

*Zdravko Živković, Marija Hribšek **
Institute Goša, Milana Rakića 35, 11000 Belgrade, Serbia. E-Mail: zdravko.zivkovic@ymail.com

Dejan Tošić
School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, PO Box 35-54, 11120 Belgrade, Serbia. E-Mail: tosic@etf.rs

** Corresponding author: marija.hribsek@yahoo.com; tel.: +381-11-2413332; fax: +381-11-2410977*

Prispelo (Arrived): 20.02.2009 Sprejeto (Accepted): 09.06.2009

UDK621.3:(53+54+621+66), ISSN0352-9045 Informacije MIDEM 39(2009)2, Ljubljana

DVB-ASI DISTRIBUTION AND SELECTION IN DVB-T/H REDUNDANCY SYSTEMS

Andrej Kosi, Mitja Solar

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

Key words: Telecommunications networks and systems, DVB-T/H Digital Video Broadcasting - Terrestrial/Handheld, Redundancy system, DVB - Digital Video Broadcasting, ASI - Asynchronous Serial Interface, DVB-ASI distributor, DVB-ASI selector

Abstract: DVB-ASI signal is used as a standard interconnection interface in digital broadcasting equipment. In article main focus is on DVB-T/H transmitter and redundancy systems, where specific demands regarding quality, size, price and flexibility are present. On more important transmission sites usually different types of redundancy systems are used for higher reliability. The most known types of redundancy systems are dual drive (DD), N+1 and N+M.

When redundancy system is used the need for additional DVB-ASI signals and control over signal is demanded. For this purpose DVB-ASI distributor and DVB-ASI selector units have been developed. Basic DVB-ASI distributor module has one DVB-ASI input and four DVB-ASI outputs. For achieving distribution to more outputs chaining of modules is used. Basic DVB-ASI selector module allows selection from four inputs to one main output and one spare (monitoring) output. With combination of selector modules more complex selections are possible. Decision for four outputs for DVB-ASI distributor and four inputs for DVB-ASI selector is a tradeoff that was made on basis of known demands in dual drive and usual size of N+1 system. In this way costs are reduced and with modular approach extension of system is still possible.

Delitev in izbiranje DVB-ASI signala v redundantnih DVB-T/H oddajnih sistemih

Kjučne besede: telekomunikacijski omrežja in sistemi, digitalna video radiodifuzija (DVB) DVB-ASI signal, digitalna prizemeljska televizija (DVB-T), mobilna prizemeljska televizija (DVB-H), redundantni sistemi, DVB-ASI delilnik, DVB-ASI selektor

Izvleček: V digitalni radiodifuzije je DVB-ASI signal standardiziran in najbolj razširjen način distribucije do oddajnih točk oziroma oddajnikov DVB-T/H signala. V članku je poudarek na DVB-T/H oddajnikih in redundantnih oddajniških sistemih, kjer so specifične zahteve za kvaliteto, velikost, ceno in prilagodljivost oddajniškega sistema. Na večjih in pomembnejših oddajnih točkah se pogosto uporabljajo redundantni sistemi oddajanja, ki zagotavljajo višjo zanesljivost oddajniškega sistema. Najbolj razširjeni sistemi so oddajniki z dvema gonilnikoma (tako imenovani Dual Drive), N+1 sistem (sistem kjer je na N oddajnikov na voljo en rezervni oddajnik) in N+M sistem (na N oddajnikov je M rezervnih oddajnikov). Kljub temu, da N+M sistem zagotavlja visoko stopnjo zanesljivosti se v praksi najpogosteje uporabljata sistem oddajnika z dvema gonilnikoma in N+1 sistem.

Ne glede na uporabljen sistem je osnovna naloga zagotoviti nemoteno oddajanje signala v primeru napake na oddajnem sistemu. Na oddajni točki je običajno na voljo en izvor DVB-ASI signala na posamezen oddajnik. Signal je potrebno zaradi podvajanja sklopov razdeliti na več enakih signalov in imeti možnost izbiranja vhodnega signala. Na tržišču sicer obstaja nekaj produktov, ki pa so običajno namenjeni za uporabo v večjih distribucijskih mrežah in so zato neekonomični za uporabo v zgoraj omenjenih redundantnih sistemih. Zaradi večje prilagodljivosti pri izvedbi različnih tipov in različnih velikosti redundantnih sistemih smo rešitev zasnovali modularno. Osnovni modul DVB-ASI delilnika deli DVB-ASI signal na štiri izhodne DVB-ASI signale. V primeru, da potrebujemo več izhodnih signalov zaporedno povežemo posamezne module. Pri DVB-ASI selektorju osnovni modul omogoča izbiro DVB-ASI signala med štirimi DVB-ASI vhodi. S pomočjo kombiniranja več osnovnih modulov DVB-ASI selektorja je možna razširitev izbire števila vhodnih signalov.

Pri sami izvedbi rešitve je veliko pozornosti potrebno posveti pravilni zasnovi in izbiri komponent, saj imamo opravka z visokofrekvenčnimi signali. Popačenje, ki nastanejo pri prehodu visokofrekvenčnega pravokotnega signala skozi koaksialni kabel zmanjšamo s pomočjo kabelskega izravnalnika. V nadaljevanju zmanjšamo še odstopanje (trepetanje) urinega takta. Pri DVB-ASI selektorji je za razliko od DVB-ASI delilnika pred časovnim sklopom dana še možnost izbira vhodnega signala. Amplitudno in časovno izboljšan signal je potem razdeljen in amplitudno ojačen na več izhodnih DVB-ASI signalov.

1 Introduction

For distribution of digital television and mobile television content from television studio to transmitter different distribution approaches (optic cables, wireless, satellite, coaxial cable and other) are used. Input signal is transformed (if needed) to standard input for DVB-T/H (Digital Video Broadcasting - Terrestrial/Handheld) transmitter regardless which distribution path is used. As standard input for transmitter DVB-ASI (Digital Video Broadcasting - Asynchronous Serial Interface) signal is used. Formal document describing professional interfaces for Digital Video Broadcasting is ETSI TR 101 891 (European Telecommunications Standards Institute) /1/. Physical characteristic of the DVB-ASI are similar to digital video signal SDI (Serial Digital Interface) defined by organization SMPTE (Society of Motion Picture and Television Engineers) under designation 259M. DVB-ASI defines interface and way of transmission of digital data. Signal is transmitted serial over 75Ω coaxial cable with nominal amplitude of 800mVpp at source. DVB-ASI is using 8B10 encoding and 270 Mbps bit rate /2/. Equipment used for transmission of SDI signal is not always compatible with DVB-ASI signal and using such equipment in transmission chain can cause problems. One of possible causes could be wrong clock detection, polarity of signal (DVB-ASI is polarity sensitive and SDI is not) and others.

Depending on size and importance of transmitting site different redundancy systems are used for achieving higher reliability of transmitting system. In most cases one of the following solutions is used for redundancy systems: dual drive (DD), N+1 and N+M. In case of dual drive system dividing DVB-ASI signal for two PA (Power Amplifier) drivers is needed. Basic concept of dual drive redundancy system is shown in picture 1.1.

In case of N+1 system dividing and selection unit is needed. For this purpose we need N dividing modules, one for each transmitter. One DVB-ASI output signal is used for transmitter and other one as input for selection unit. With selection unit appropriate DVB-ASI signal is selected for reserve transmitter. Basic concept of N+1 redundancy system is shown in picture 1.2. In case of failure of one transmitter, ASI selector will receive information from control unit about which input to select.

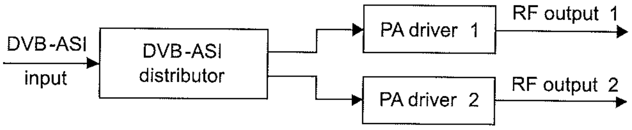


Fig. 1.1: Example of dividing DVB-ASI signal for dual drive system.

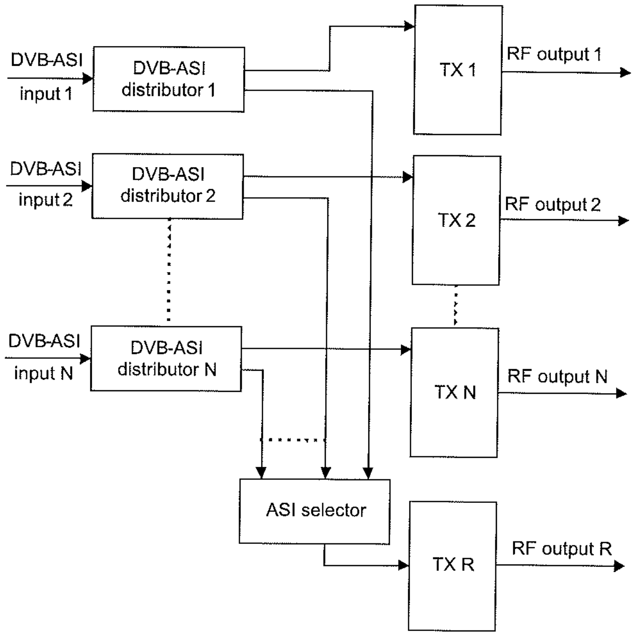


Fig. 1.2: Example of dividing and selection of DVB-ASI signal for N+1 system.

N+M redundancy system is extension of N+1 system. On each dividing module more outputs are used and M selection units are needed. Although N+M system ensures higher reliability it is rarely used because of higher price and complex architecture. If on transmitting site additional reserve DVB-ASI signal is present and one of described redundancy system is used, then demands for dividing and selecting units are doubled for every system.

2 Proposed solution

On basis of presented demands for redundancy systems DVB-ASI distribution (DVB-ASI divider) and DVB-ASI selector units have been developed. To have same basic construction block for all redundancy systems optimum solution is DVB-ASI distributor with four outputs per module and DVB-ASI selector with four inputs per module. Besides main output DVB-ASI selector has one spare monitoring output. To extend distribution to more outputs, chaining of modules is used. To extend number of inputs for DVB-ASI selector combining selector modules is also possible.

2.1 DVB-ASI distributor

Basic function of DVB-ASI distributor is dividing input signal on more equivalent outputs. Simply passive dividing of signal is not appropriate solution because in equipment that is following passive dividing system wrong detection could happen. For this reason active dividing is used. Example of active dividing system is shown on picture 2.1, where separated blocks with their functions are presented.

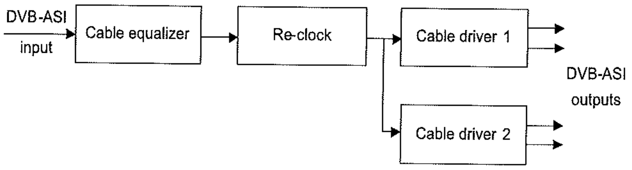


Fig. 2.1: DVB-ASI distributor module.

When high frequency digital signal is traveling through long coaxial cable signal will loose amplitude and shape because higher frequencies are more attenuated as lower ones /3/. On picture 2.2 /3/ example of dependence between frequency and attenuation for 100m coaxial cable is shown. Attenuation depends on frequency and cable length. Because of this property input signal is first recovered against cable length. For this purpose cable equalizer with inverse cable characteristic for compensating cable influence on signal is used.

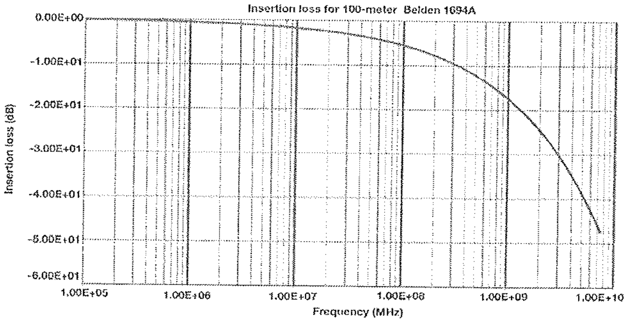


Fig. 2.2: Attenuation of signal against frequency for 100m long coaxial cable.

Long cables and different equipment in transmission chain have also influence on embedded clock of DVB-ASI signal. Such influence is manifested as jitter in signal. With reclocking embedded clock is recovered again. Depending on usage of ASI distributor reclocking function can be disabled. Such improved signal is then divided to four output signals.

At the end of circuit cable driver is amplifying signal to nominal amplitude of 800 mVpp. On picture 2.3 DVB-ASI signal is shown after passing long coaxial cable. If same signal is then used as input for DVB-ASI distributor the signal on output will be recovered to signal shown in picture 2.4. For such measurements special oscilloscopes with infinitive persistence and color display are used. One of such measuring instrument is Agilent Infiniium MSO8104A that was used for measuring DVB-ASI signal before and after usage of developed DVB-ASI distributor and DVB-ASI selector module /7/.

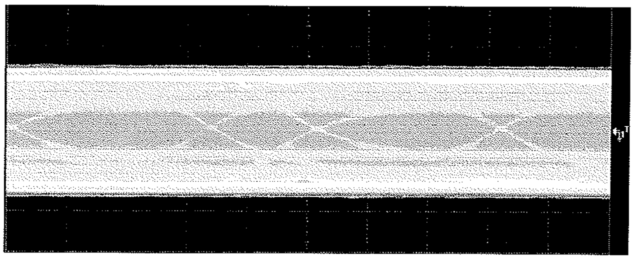


Fig. 2.3: Example of DVB-ASI signal after traveling through long coaxial cable.

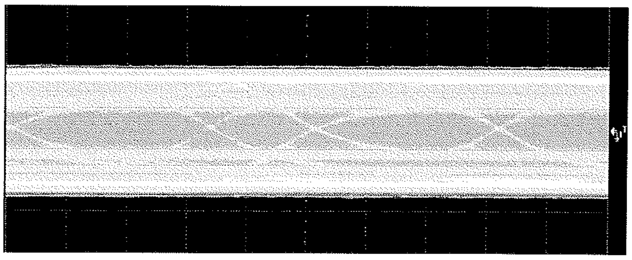


Fig. 2.4: Example of DVB-ASI signal on picture 2.3 after using DVB-ASI distributor.

2.2 DVB-ASI selector

For DVB-ASI selector module similar components as for DVB-ASI distributor are used. On picture 2.5 basic block of DVB-ASI selector module are shown. Input DVB-ASI signals are first reconstructed against coaxial cable deformation with cable equalizer. Selection block is main part of DVB-ASI selector. In this part selection of input signal is done. After selecting signal and reclocking cable driver is amplifying signal to nominal amplitude 800 mVpp.

3 Measurements and basic technical data

In table 3.1 demands for professional equipment that is using DVB-ASI according to European standard EN 50083-9 are presented /4/.

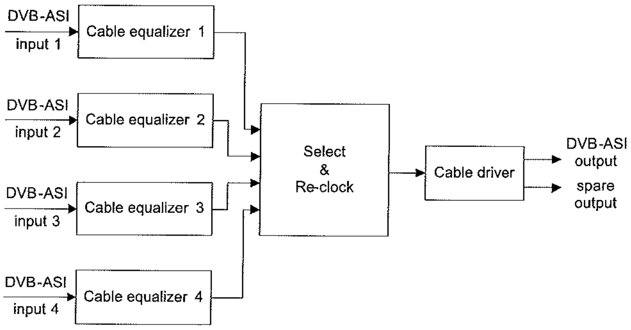


Fig. 2.5: DVB-ASI selector module.

Table 3.1: Demands for DVB-ASI equipment.

Transmitter output characteristics	Units	
Output voltage (p -p)	mV	800 ± 10%
Deterministic Jitter (DJ) (p -p)	%	10
Random Jitter (RJ) (p-p)	%	8
Return loss	dB	under consideration
Max. rise/fall time (20 -80%)	ns	1,2

Receiver input characteristics	Units	
Min. sensitivity (D21.5 idle pattern)	mV	200
Max. input voltage (p -p)	mV	880
S_{11} (range: 0,1 to 1,0 x bit rate)	dB	-17
Min. discrete connector return loss (0,3 MHz – 1 GHz)	dB	-15

Because DVB-ASI physical characteristic are similar to SDI signal descriptions and measurements proposed from EBU (European Broadcasting Union) document Tech 3283 (Measurements in digital component television studios) can be used /5/. On picture 3.1 basic DVB-ASI signal characteristic and measurement definitions are shown.

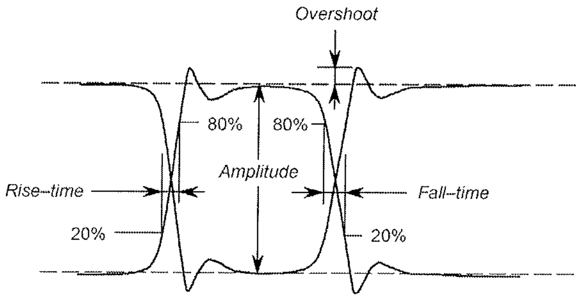


Fig. 3.1: Basic measurements for DVB-ASI signal.

Besides basic measurements for DVB-ASI signal important information is jitter. Jitter is short term variation of digital signal from their ideal position in Time /6, 8/. There are several ways how measuring of jitter is done. On basis of type of measure we know periodic jitter, cycle - cycle jitter, TIE - time interval error. Another way to visually represent jitter is eye diagram. Quick estimation of signal characteristics is possible from eye diagram. On picture 3.2 /reference 6 Figure 4.2.1b, 4.2.3b, 4.2.4c and 4.2.5a/ typical eye diagrams with their distributions for different types of jitter are shown. In the upper left part of picture random jitter is shown. In the upper right periodic jitter is shown. In the lower left part of picture data depended jitter and lower right duty-cycle depended jitter is shown.

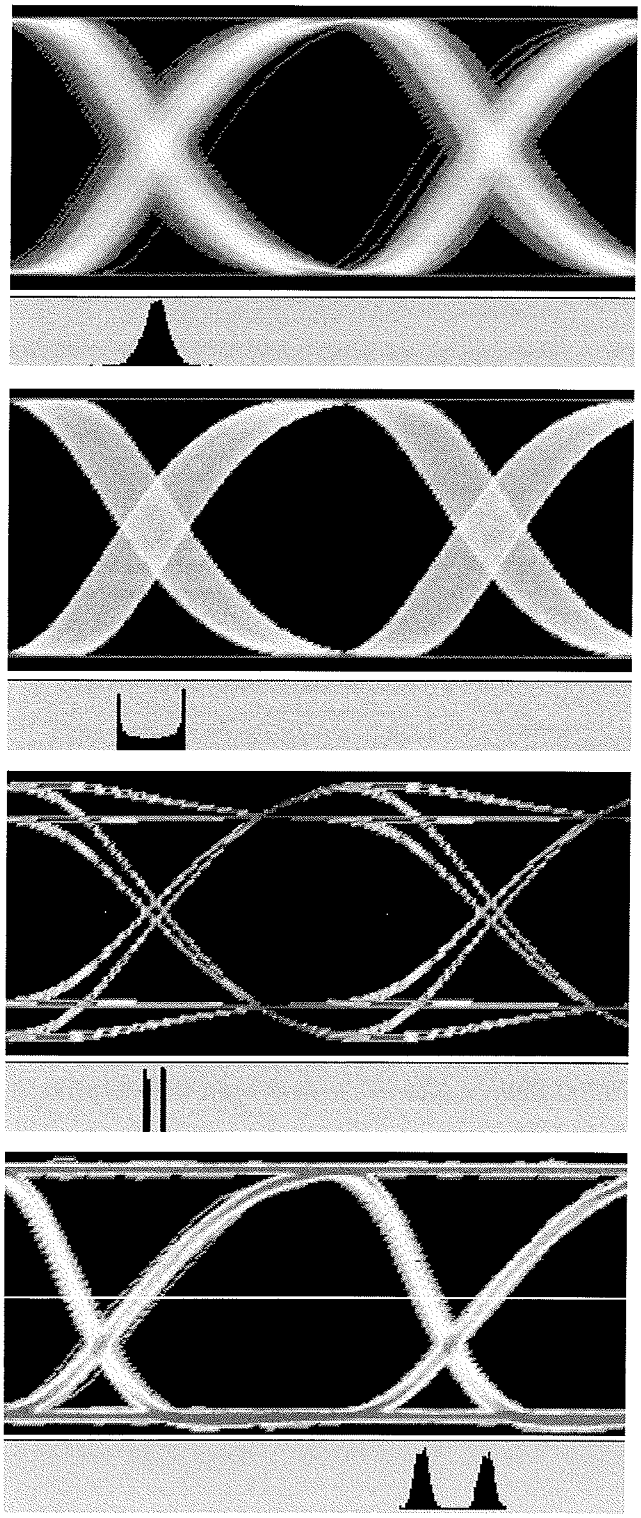
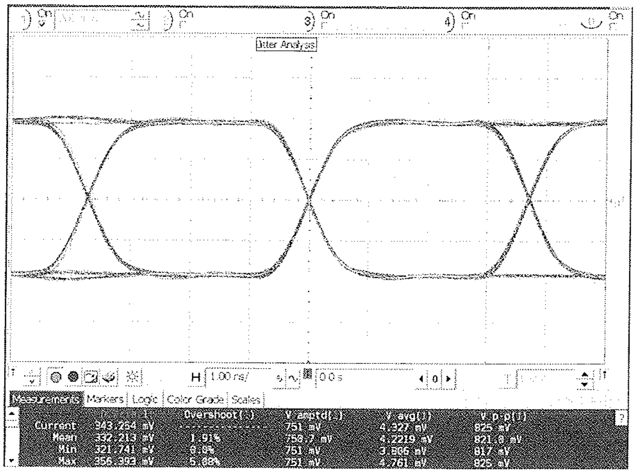


Fig. 3.2: Typical eye diagrams for different types of jitter.

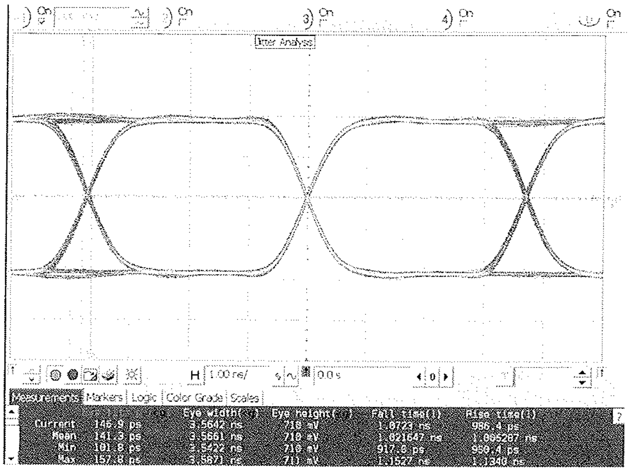
On picture 3.3 and 3.4 results of measurements are shown. For measuring instrument Agilent Infiniium MSO8104A was used /7/. Both pictures are presenting typical values measured on DVB-ASI distributor and DVB-ASI selector modules. On picture 3.3 amplitude characteristic are shown with their current, mean, minimum and maximum values. Measured values for peak to peak amplitude (Vp-p), overshot and averaged amplitude (DC offset) are visible.



	[illegible]	Overshoot(1)	V amptd(1)	V avg(1)	V p-p(1)
Current	343.254 mV	----------------	751 mV	4.327 mV	825 mV
Mean	332.213 mV	1.91%	758.7 mV	4.2219 mV	821.8 mV
Min	321.741 mV	0.0%	751 mV	3.806 mV	817 mV
Max	356.393 mV	5.08%	751 mV	4.761 mV	825 mV

Fig. 3.3: Example of measure results.

On picture 3.4 time and eye diagram characteristic are shown with their current, mean, minimum and maximum values. Measured values for jitter, eye width, eye height, fall time and rise time are visible.



	[illegible]	Eye width(1)	Eye height(1)	Fall time(1)	Rise time(1)
Current	146.9 ps	3.5642 ns	718 mV	1.0723 ns	986.4 ps
Mean	141.3 ps	3.5661 ns	718 mV	1.021647 ns	1.005287 ns
Min	101.8 ps	3.5422 ns	718 mV	917.8 ps	950.4 ps
Max	157.8 ps	3.5671 ns	718 mV	1.1527 ns	1.1340 ns

Fig. 3.4: Example of measure results.

In table 3.2 basic technical data for DVB-ASI distributor are shown.

4 Conclusions

In the article a concept for distribution and selection of DVB-ASI signal in DVB-T/H transmitting redundancy systems was shown. Solution with costs effective DVB-ASI distributor and DVB-ASI selector was presented. DVB-ASI distributor module has one DVB-ASI input and four DVB-ASI outputs per module. DVB-ASI selector module has four DVB-ASI inputs, one main and one monitoring DVB-ASI output per module. Because of modular concept different

Table 3.2: Technical data for DVB-ASI distributor.

Data	Value
Number of inputs	1 for each module
Signal type	DVB-ASI 270 Mb/s
Connector	BNC (IEC169-8)
Impedance	75 ohms
Return loss	18 dB to 540 MHz
Number of outputs	4 for each module (up to 3 modules in one rack)
Output amplitude	800 mV ± 10%
DC offset	0,0 V ± 0,5V
Overshoot	<10% of amplitude
Rise time	<1,2 ns
Jitter	< 0,2 UI peak-to-peak
External dimensions	19" rack, 1U, 180 mm depth
Mains AC voltage	85-264 V AC, 47-60Hz
Power consumption	<10 VA

In table 3.3 basic technical data for DVB-ASI selector are shown.

Table 3.3: Technical data for DVB-ASI selector.

Data	Value
Number of inputs	4 for each module
Signal type	DVB-ASI 270 Mb/s
Connector	BNC (IEC169-8)
Impedance	75 ohms
Return loss	18 dB to 540 MHz
Number of outputs	2 for each module (up to 2 modules in one rack)
Output amplitude	800 mV ± 10%
DC offset	0,0 V ± 0,5V
Overshoot	<10% of amplitude
Rise time	<1,2 ns
Jitter	< 0,2 UI peak-to-peak
Communication	RS232/RS422/RS485
Communication protocol	ASCII
External dimensions	19" rack, 1U, 180 mm depth
Mains AC voltage	85-264 V AC, 47-60Hz
Power consumption	<10 VA

types and sizes of redundancy systems are possible. Additional flexibility in planning or changing structure of redundancy systems is achieved. Usage of developed products is possible in any other system where distribution or selection of DVB-ASI signal is needed. Measuring results confirmed accordance to different standards for professional equipment and DVB-ASI signal.

5 References

/1/ ETSI TR 101 891 V1.1.1, Digital Video Broadcasting (DVB); Professional Interfaces: Guidelines for the implementation and usage of the DVB Asynchronous Serial Interface (ASI), 2001-02.

/2/ Asynchronous Interfaces For Video Servers by Karl Paulsen, November 2003, http://www.tv-technology.com/pages/s.0069/t.1519.html

/3/ Use equalization to drive digital video through long cables, Mark Sauerwald, National Semiconductor Corp, June 2006, http://www.planetanalog.com/article/printableArticle.jhtml?articleID=188702457

/4/ European Standard EN 50083 Part 9: Interfaces for CATV/SMATV headends and similar professional equipment for DVB/MPEG-2 transport streams, CENELEC European Committee for Electrotechnical Standardization, March 1997.

/5/ EBU document Tech 3283, Measurements in digital component television studios 625-line systems at the 4:2:2 and 4:4:4 levels using parallel and serial interfaces (SDI), December 1996.

/6/ Understanding and Characterizing Timing Jitter, Tektronix, 2003 http://www.tek.com/Measurement/scopes/jitter/55W_16146_1.pdf

/7/ Infiniium 8000 Series Oscilloscopes Superior Signal Viewing and Analysis, Agilent Technologies, 2007, http://cp.literature.agilent.com/litweb/pdf/5989-4271EN.pdf

/8/ Eye Patterns in Scopes Peter J. Pupalaikis, Eric Yudin, LeCroy Corporation, 2005 http://www.lecroy.com/tm/Library/WhitePapers/PDF/Eye_Patterns_in_Scopes-designcon_2005.pdf

Manuscript submitted on August 26, 2008. This work was supported by company ELTI d.o.o Slovenia.

A. Kosi is with the ELTI d.o.o, Panonska 23, Gornja Radgona, Slovenia (e-mail: andrej.kosi@uni-mb.si).

M. Solar is with the Faculty of Electrical Engineering and Computer Science in Maribor, Smetanova 17, 2000 Maribor, Slovenia (e-mail: mitja.solar@uni-mb.si).

Prispelo (Arrived): 04.12.2008 Sprejeto (Accepted): 09.06.2009