



ISSN 1311-0829

ГОДИШНИК НА ТЕХНИЧЕСКИЯ УНИВЕРСИТЕТ – СОФИЯ

**Том 60, книга 3, 2010 г. – Юбилеен
„65 години ТУ – София” и „60 години
Годишник на ТУ – София”**



PROCEEDINGS OF THE TECHNICAL UNIVERSITY – SOFIA

**Volume 60, book 3, 2010 – Anniversary
“65 Years Technical University – Sofia” and “60 Years
Proceedings of the Technical University – Sofia”**

РЕДАКЦИОННА КОЛЕГИЯ

Главен редактор: проф. дтн Ганчо Божилов

Членове: проф. д-р Михаил Момчеджиков
проф. д-р Марин Христов
проф. д-р Стефан Овчаров
проф. д-р Георги Димчев
доц. д-р Вълко Кичев
доц. д-р Валери Стоилов
инж. Мая Петрунова

EDITOR'S BOARD

President: Prof. DSc Gantcho Bojilov

Members: Prof. PhD Michail Momchedjikov
Prof. PhD Marin Hristov
Prof. PhD Stefan Ovcharov
Prof. PhD Georgi Dimtchev
Assoc. Prof. PhD Valko Kitchev
Assoc. Prof. PhD Valeri Stoilow
Eng. Maya Petrunova

С Ъ Д Ъ Р Ж А Н И Е

1. Илиев, В., М. Момчеджиков. Методи за повишаване на шумоустойчивостта на широколентовите системи за комуникация	7
2. Илиев, В. Функционална устойчивост на алгоритъм за оптимална филтрация на сложни сигнали по критерий за точност	16
3. Кънчев, В., О. Бумбаров. Създаване на панорамни изображения тип “мозайка” с приложение при видеонаблюдения	23
4. Божилов, Г., А. Иванов. Аналитично определяне на ъгъла на изместване на четките от неутралната линия при колекторните електрически машини	32
5. Илиев, В., М. Момчеджиков. Алгоритъм за некохерентна оптимална филтрация на сложни сигнали по критерий за точност....	39
6. Илиев, В. Изследване на грешките в следене на фазата и закъснението на алгоритъм за оптимална филтрация на сложни сигнали	50
7. Захариев, З. Размити системи за извод чрез системи размити релационни уравнения	58
8. Велев, В., К. Филипов. Екологично чиста електроенергия от ядрени електроцентрали	67
9. Владов, Г. Метод за дългосрочно инвестиране с Марковска верига и динамично програмиране	74
10. Йончев, Е. Адаптивен по скорост и статорно съпротивление наблюдател на потокосцепление за системи за безсензорно векторно управление	84
11. Филипова, С. Софтуерна реализация на основните математически действия и функции посредством обобщени интервали	94
12. Илиев, В., М. Момчеджиков. Оптимално приемане в условията на априорна неопределеност на канала за връзка	107
13. Илиев, В. Метод за скрито предаване на информация с помощта на хаотични сигнали	113
14. Минковска, Д., Й. Марков. Проектиране на база от данни за обработка на резултатите от тестов контрол за оценяване на знанията	119
15. Георгиев, В. Взаимно ориентиран евристичен матод за описание на промените на лицето, използващ ефекта на соларизация	123

16. Димитрова, В. Подготовка на геометрична среда за целите на визуализация на данни от компютърна симулация и от експерименти	133
17. Ангелова, М. Определяне на пространственото положение на видимите стени от обект в 3D визуална сцена	142
18. Видеков, В., С. Цанова, Д. Нурков. Технологично изготвяне на компютърен микромодул	151
19. Врублевски, И., К. Чернякова, В. Видеков, Б. Цанева, А. Ячовски. Превключващ ефект в нанопорист оксид на алуминия при анодна поляризация	157
20. Ячовски, А. Контрол на процеса анодиране чрез микроконтактни структури	164

C O N T E N T S

1. Iliev, V. and M. Momchedjikov. Methods for noise immunity increase of broad band communication systems	7
2. Iliev, V. Functional stability of an algorithm for complex signals optimal filtration by precision criterion	16
3. Kanchev, V and O. Boumbarov, Creating panoramic image mosaic for video surveillance purposes	23
4. Bojilov, G. and A. Ivanov. Analytical evaluation of the angle of displacement of the brushes from the neutral line on the commutator electrical machines	32
5. Iliev, V. and M. Momchedjikov. Algorithm for non-coherent optimal filtration of complex signals by precision criterion	39
6. Iliev, V. Analysis of errors in the phase and delay tracking of algorithm for complex signals optimal filtration	50
7. Zahariev, Z. Fuzzy reasoning through fuzzy linear systems of equations.....	58
8. Veleev, V. and K. Filipov. Ecologically clean electricity from nuclear power plants	67
9. Vladov, G. A method for multiperiodic investment based on Markov chain and dynamic programming	74
10. Ionchev, E. Adaptive speed and stator resistance flux observer for the systems sensorless vector control	84
11. Filipova, S. Software realization on the basic mathematical operations and functions using generalized intervals	94
12. Iliev, V. and M. Momchedjikov. Optimal receiving in conditions of aprioric indeterminateness of the channel link	107
13. Iliev, V. Concealed transmitting of information based on chaotic signals	113
14. Minkovska, D. and J. Markov. Design of database for processing of test control results for assessment of the knowledge	119
15. Gueorguiev, V. A heuristic method for classification facial changes for visemes' selection based on solarization effect	123
16. Dimitrova, V. Geometric environment preparation for the visualization of data obtained through computer simulation on experiments	133
17. Angelova, M. Determination of spatial location of visible faces of objects in a 3D visual scene	142
18. Videkov, V., S. Tzanowa and D. Nourkov. Technological production of a wearable computer micromodule	151

19. Vrublevski, I., K. Chernyakova, V. Videkov, B. Tzaneva and A. Yachovski. Switching effect in nanoporous alumina at the anode polarization	157
20. Yachovski, A. Control of the process anodizing by micro structures.....	164

МЕТОДИ ЗА ПОВИШАВАНЕ НА ШУМОУСТОЙЧИВОСТТА НА ШИРОКОЛЕНТОВИТЕ СИСТЕМИ ЗА КОМУНИКАЦИЯ

Вася Илиев, Михаил Момчеджиков

Резюме. В работата е извършен сравнителен анализ на методите за повишаване на шумоустойчивостта на широколентовите системи за комуникация над осигуреният от базата на сигнала запас от шумоустойчивост. Анализиран са методите за повишаване на шумоустойчивостта чрез предварителна обработка и филтрация. В табличен вид е сравнена ефективността на разгледаните методи по различни видове критерии.

METHODS FOR NOISE IMMUNITY INCREASE OF BROAD BAND COMMUNICATION SYSTEMS

Vasya Iliev, Mikhail Momchedjиков

Abstract. In the work some methods for broad band communication systems noise immunity increase above the reserve provided by signal base are examined. Particular attention is paid to methods for noise immunity increase by preliminary processing and filtration. The examined methods effectiveness is compared by different criteria in table view.

1. Въведение

Системите за комуникации (СК) със сложни шумоподобни сигнали (СШПС) се прилагат за решаването на такива важни задачи като потискане на смущенията, кодово разделение при работа в обща честотна лента, потискане на ефектите от многолъчевото разпространение, измерване на параметрите на движение на обектите с висока точност и разрешаваща способност, повишаване на скритостта и криптоустойчивостта на предаваните съобщения и др.

От всички потенциални области на приложение на системите за комуникации със СШПС очевидно най-ценна е възможността да се противодейства както на умишлените, така и на случайните интерференционни смущения. Както е известно, шумоустойчивостта на този вид системи се определя от отношението на мощността на сигнала и мощността на смущението в изхода на съгласувания филтър (корелатор), което съгласно [1] е

$$(1) \quad q^2 = \frac{P_c}{P_{cm}} \cdot 2B_c,$$

където P_c и P_{cm} са съответно мощностите на сигнала и смущението на входа на приемното устройство (ПУ); $B_c = F_c T_c$ – база на използвания сложен сигнал.

Годишник на ТУ-София, т. 60, кн. 3, 2010

Proceedings of the Technical University - Sofia, v. 60, book 3, 2010

Съотношение (1.1) показва, че при фиксирани стойности на P_c и P_{cm} на входа на ПУ повишаването на отношението сигнал/смушение в изхода на ПУ е възможно само за сметка на увеличение на базата B_c на полезния сигнал. Използването на СШПС със свръх голяма база за борба със смушаващи сигнали, чието ниво на входа на ПУ надвишава нивото на полезния сигнал с няколко порядъка, е изключително гъвкаво и ефективно, но не винаги възможно и най-рационално.

В практиката за тази цел широко приложение са намерили редица методи за допълнителни подтискане на смушаващите сигнали, които позволяват получаването на необходимият запас от шумоустойчивост да се осигури и при работата със СШПС със нормална база. Тези методи включват редица допълнителни видове обработка на сигнала, чието описание е предмет на статията.

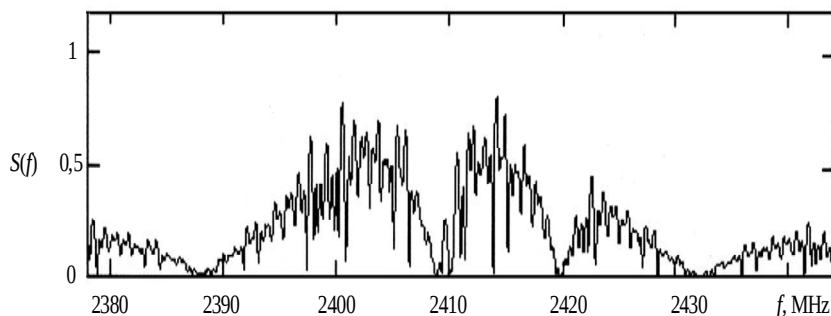
2. Методи за предварителна обработка и филтрация

Методи за повишаване на шумоустойчивостта при работа в условия на теснолентови смущения

За защита на ШСК от теснолентови смущения, чието ниво надвишава динамичния диапазон на ПУ и осигуреният от базата на сигнала допустим запас от шумоустойчивост, се прибегва към *режекция* на поразената част от спектъра на сигнала. При спектрална режекция на отделни участъци от спектъра на СШПС се предизвиква изкривяване на корелационната функция на сигнала в изхода на режекторния филтър, която съгласно [2] се определя чрез енергийния спектър на сигнала $G(\omega)$ и честотния коефициент на предаване на филтъра с израз

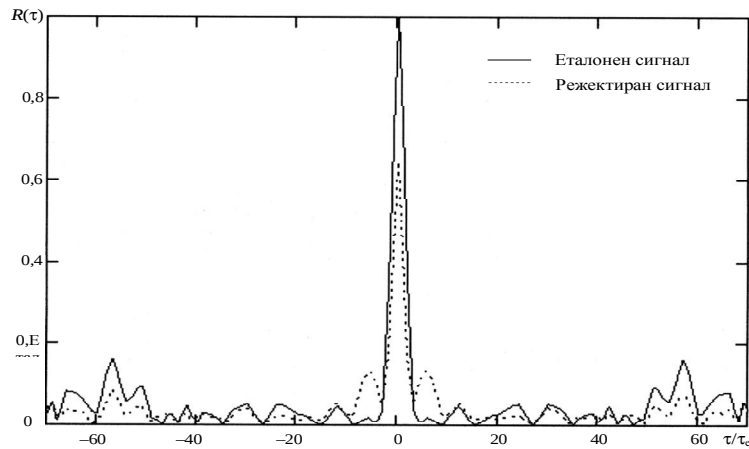
$$(2) \quad R(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) \cdot K(\omega) \cdot e^{j\omega t} d\omega$$

На фиг. 1 е показан спектърът на 63-елементен фазовоманипулиран СШПС в изхода на режекторен филтър с две произволни честоти на режекция – 2409MHz и 2419MHz.



Фиг. 1. Спектър на DSSS сигнал, преминал през два режекторни филтъра

На фиг. 2 с пунктирна линия е представена съответната му корелационна функция.



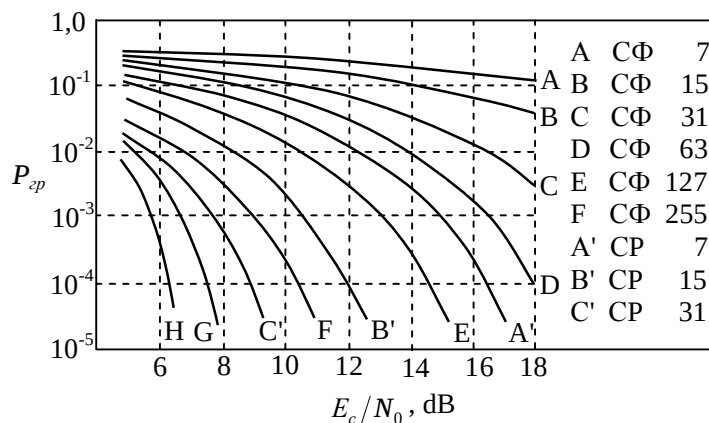
Фиг. 2. Корелационна функция на еталонен и режекторен DS-SSS

В [3] е изследвано, че режекцията на спектъра на сигнала до 25% предизвиква незначителна промяна във формата на КФ, но предизвиква понижаване на амплитудата на основния максимум на КФ. Коефициентът на нарастване на отношението сигнал/смущения в изхода на филтъра при липса на шум ($\sigma_N = 0$) е

$$(3) \quad G_p = 1 + \frac{P_{cm}}{2P_c} \left[2N - 1 + \frac{\sin(2N+1)\Omega T_e}{\sin \Omega T_e} \right],$$

където P_{cm} – мощност на смущението; N – брой на изводите на закъснителната линия на трансферзалния филтър; Ω – честотна разстройка.

На фиг. 3 в графичен вид е сравнена вероятността за грешно приет информационен символ при обработка на СШПС по традиционните методи и в случаите на допълнителна обработка по разглеждания метод [4]. На нея с цифри е означена необходимата стойност на базата $B_c = F_c T_c$ на използвания СШПС за постигане на конкретна стойност на вероятността за грешка.

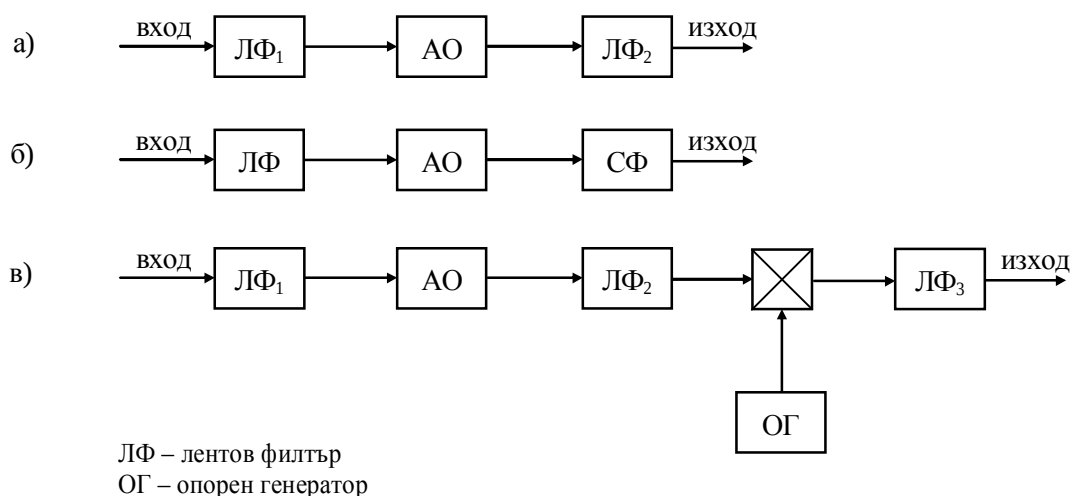


Фиг. 3. Сравнение на вероятностите за грешка при обработка със съгласуван филтър (СФ) и с допълнителна обработка със спектрална режекция (СР)

Показаните графики са получени за стойности на отношението сигнал/смущение $P_c/P_{cm} = -20$ dB. От тях се вижда, че при обикновените методи за обработка за постигане на $P_{гр} = 10^{-4}$ е необходима стойност $B_c = 255 \div 511$, докато в случаите на допълнителна обработка със спектрална режекция необходимата стойност на базата е $B_c = 31$.

Методи за повишаване на шумоустойчивостта при работа в условия на импулсни смущения

За борба с мощните импулсни смущения в практиката широко приложение е намерило амплитудното ограничение на сигналите. Използването на амплитудни ограничители (АО) в процеса на обработка на сигналите е целесъобразно както за потискане на смущаващите сигнали, така и за намаляване на динамичният им диапазон [5]. Най-често използваните АО са т.нар. *лентови амплитудни ограничители (ЛАО)*, показани на фиг.4а. От изхода на първия лентов филтър (ЛФ₁) сигналът и смущението постъпват на АО и след съответно ограничение се подават към ЛФ₂. Двата лентови филтри са с централна честота, равна на носещата честота на сигнала и честотна лента, равна на активната ширина на спектъра му. В случая ЛФ₁ осъществява предварителна селекция на полезния сигнал, а ЛФ₂ отстранява комбинационните съставлящи в спектъра на ограничения сумарен сигнал. При оптимална филтрация на мястото на ЛФ₂ се включва съгласуван филтър (фиг. 4б). При корелационна обработка (фиг. 4в) корелаторът се разполага след ЛФ₂.



Фиг. 4. Схеми на приемници с лентови амплитудни ограничители

Ако се приеме, че сигналът и смущението във времето се характеризират с нормални обвиващи – $a(t)$ и $\eta(t)$ съответно, то АО трябва да притежава импулсна характеристика $h_{АО}(t)$, определяща се от съотношението $h_{АО}(t) \cdot \eta(t) = 1$, [6]. Тогава за импулсната характеристика на АО може да се запише

$$(4) \quad h_{АО} = 1/\eta(t),$$

а обвиващата на сигнала и смущението в изхода на АО има следния аналитичен вид

$$(5) \quad \begin{aligned} a'(t) &= a(t)h_{АО} = a/\eta(t), \\ \eta'(t) &= 1. \end{aligned}$$

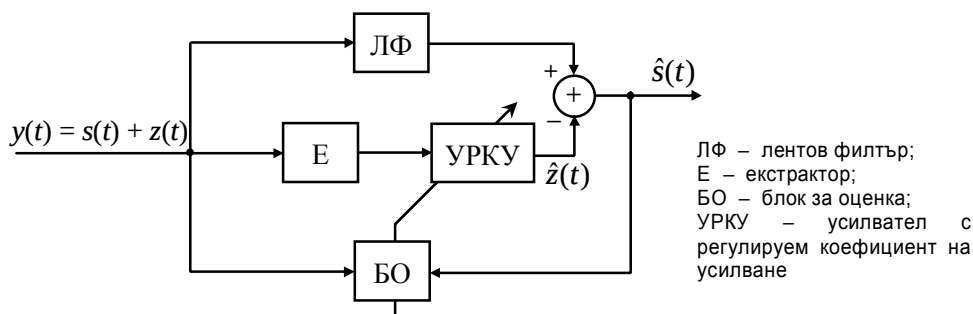
Максималната стойност на печалбата в отношението сигнал/смущение която се получава при съгласувана филтрация и допълнителна обработка с амплитуден ограничител с импулсна характеристика (4) е [5]

$$(6) \quad G_p = \left[\frac{\int_0^T \left[\frac{a(t)}{\eta(t)} \right]^2 dt}{\int_0^T \frac{a(t)}{\eta(t)} dt} \right] dt.$$

Методи за повишаване на шумоустойчивостта при работа в условия на структурни смущения

Структурните смущения представляват смущаващи сигнали със структура, аналогична на полезния сигнал и се различават от него единствено по формирания код. Поради тази причина те не могат да бъдат постискани непосредствено с методите за спектрална и амплитудна режекция. За борба с тях в ШСК най-често се прибегва към компенсационни *методи за режекция*. Същността на тези методи се състои в отделянето на смущаващия сигнал от входния сигнален процес и сумиране в противофаза с него, за да се осъществи компенсация.

На фиг. 5 е показана схема за компенсация на структурни смущения, основаваща се на статистическите различия между кодовете на полезния и смущаващия сигнал [7].



Фиг. 5. Схема за компенсиране на структурни смущения (предложена в [7])

В случая отделянето на смущаващия сигнал $z(t)$ се осъществява от екстрактора (Е) във вид на оценка $\hat{z}(t)$.

При правилно подбран коефициент на усиление на усилвателя с регулируем коефициент на усиление (УРКУ) може да бъде осъществена компенсация на смущаващия сигнал в изхода на сумиращото устройство. Регулирането на коефициента на усиление на УРКУ се осъществява от блока за оценка (БО), в съответствие с избрания алгоритъм на работа – минимизиране на средноквадратичната грешка в изхода на устройството. Чрез повдигане на сигнала $\hat{s}(t)$ на квадрат се получава

$$(7) \quad \hat{s}^2 = s^2 + (z - \hat{z})^2 + 2s(z - \hat{z}).$$

Чрез усредняване по време на двете части на уравнение (1.11) и отчитане на некорелираността на сигнала и смущението се получава

$$\begin{aligned}
 (8) \quad M[\hat{s}^2] &= M[s^2] + M[(z - \hat{z})^2] + 2M[s(z - \hat{z})] \\
 &= M[s^2] + M[(z - \hat{z})].
 \end{aligned}$$

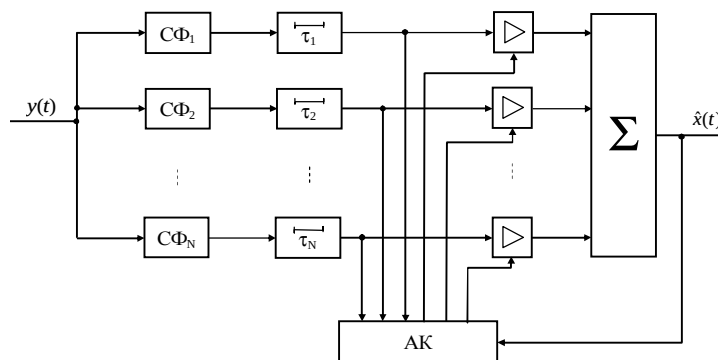
Тъй като в процеса на настройката мощността на сигнала $M[s^2]$ не се променя, то коефициентът на усилване на УРКУ се настройва по начин, по който се минимизира мощността на сигнала $M[\hat{s}^2]$ в изхода на устройството. При това положение и събираемост $M[(z - \hat{z})^2]$ също приема минимална стойност, т.е.

$$(9) \quad \min M[\hat{s}^2] = M[s^2] + \min M[(z - \hat{z})^2].$$

Когато $M[(z - \hat{z})]$ приема минимална стойност, съгласно уравнение (9), $M[(z - \hat{z})^2]$ също достига своя минимум. Следователно, сигналът $\hat{s}(t)$ в изхода на устройството се явява най-добрата оценка на сигнала $s(t)$ по критерия минимум на средноквадратичната грешка (МСКГ) и се получава при пълна компенсация на смущението $z(t)$.

Повишаване на шумоустойчивостта при работа в условия на априорна неопределеност

Разгледаните дотук методи за повишаване на шумоустойчивостта на ШСК изискват точна предварителна информация за смущаващите сигнали. Обикновено в реални условия на работа такава информация липсва. Повишаването на шумоустойчивостта на ШСК в тези случаи е възможно с помощта на *адаптивните методи за обработка и филтрация* [8]. В структурно отношение приемниците за адаптивна филтрация на СШПС са изградени като многоканален СФ, в който с помощта на анализатор се измерва съвместното разпределение на сигнала и смущенията в отделните канали. В съответствие с него се променят коефициентите на предаване на съответните канали за получаване на максимална стойност на отношението сигнал/шум в изхода на филтъра. На фиг. 6 е показана каноничната структура на адаптивен филтър за тази цел.



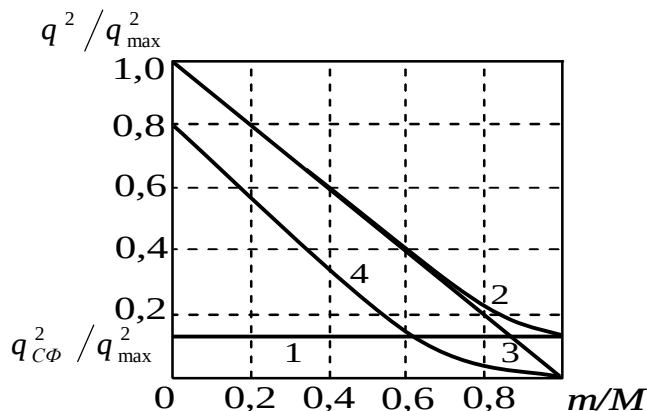
Фиг. 6. Схема на устройство за адаптивна филтрация на СШПС (предложена в [8])

От фиг. 6 се вижда, че адаптивният филтър се състои от N на брой канали, във всеки от които има съгласуван филтър (СФ), за точно определен елемент на

сигнала, закъснителна линия компенсираща закъснението им и усилватели с променлив коефициент на усилване. Анализаторът на каналите (АК) извършва анализ на напреженията в отделните канали и регулира коефициентите на усилване така, че отношението сигнал/смущение в изхода на суматора да е максимално. За тази цел е необходимо напреженията от отделните канали да се сумират синфазно с определени тегловни коефициенти.

На фиг. 7 е показана зависимостта на отношението $q^2/q_{\max}^2$ от отношението m/M (брой на поразените елементи на сигнала към общия брой елементи) за три вида филтри: съгласуван (графика 1), адаптивен линейен (графика 2) и адаптивен нелинеен (графика 4).

В случая $q_{\max}^2 = Mq_0^2 = 2E/N_0$ представлява максималното отношение сигнал/бял шум. При $m/M \ll 1$, графика 2 съвпада с графика 3, описана като $(1 - m/M)$. Графика 1 отразява зависимостта $q_{C\Phi}^2/q_{\max}^2$ от m/M . Тъй като $q_{C\Phi}^2$ не зависи от броя на поразените елементи m , то тя е успоредна на абсцисната ос, а отношението $q_{C\Phi}$ при въздействие на смущаващ сигнал е $q_{C\Phi}^2 = (P_c/P_{cm}).2B_c$.



Фиг. 7. Зависимост на отношението $q^2/q_{\max}^2$ от отношението m/M за три вида филтри

3. Заключение

От разгледаните в точка 2 методи за повишаване на шумоустойчивостта и Таблица 1 се вижда, че няма универсален метод, който да бъде еднакво ефективен за борба с различните видове смущения. Ефективността на разгледаните методи количествено може да се определи с коефициента на нарастване на отношението сигнал/шум G_p . За ефективността може да се съди и по редица други показатели, като изискване за априорна неопределеност, апаратна и изчислителна сложност, запазване или деструктуриране на полезния сигнал; възможност за многократна обработка; универсалност на метода и др. Теоретично е доказано, че ако смущенията са априорно известни, винаги може да бъде създаден такъв оптимален приемник, който да осигури максимална стойност на отношението сигнал/шум в изхода. На практика обаче смущаващите сигнали са предварително неизвестни, поради което пълното им отстраняване е практически невъз-

Таблица 1. Сравнение на ефективността на най-често използваните методи за борба с мощните смущаващи сигнали

Методи за отстраняване на смушчения	Коефициент на нарастване на отношението сигнал/шум G_p			Изискване за априорна определеност	Възможност за многократна обработка
	Теснолентови смушчения	Импулсни смушчения	Структурни смушчения		
Корелационни	$G_p = 2F_c T_c$	$G_p = (2F_c T_c)^2$	$G_p = 2F_c T_c [1 - R_{jk}^2]$	не	не
Спектрална режекция	До 20 dB на СВЧ До 40 dB на МЧ	–	$G_p = \left(1 + \frac{P_c}{P_{cm}}\right) \times \left(1 - \frac{\Delta F_{cm}}{F_c}\right)$	да	да
Амплитудна режекция	Енергийни загуби до 6 dB	$G_p = \left(1 + \frac{P_c}{P_{cm}}\right) \times \left(1 - \frac{t_u}{T_c}\right)$	Енергийни загуби от 1 до 6 dB	не	да
Компенсационни: - с екстрактор; - с амплитудно ограничение; - с опорен канал	–	–	$G_p \leq 20 \text{ dB}$	да	да
	$G_p = 1 + P_c / P_{cm}$	–	$G_p = 1 + P_c / P_{cm}$	да	да
	$G_p = 40 \text{ dB}$	–	$G_p \leq 40 \text{ dB}$	да	да
Адаптивна обработка	$G_p \approx 2F_c T_c \left(\frac{M}{m+1}\right)$	$G_p \approx (2F_c T_c)^2$	$G_p \approx 2F_c T_c [1 - R_{jk}^2] \times \left(\frac{M}{m+1}\right)$	не	не

можно. Независимо от факта, че в момента съществуват многобройни методи за потискане на едно или друго смущение, задачата за борба с тях като цяло остава неразрешена. В общия случай корелационните методи за обработка на сложни сигнали предлагат най-гъвкави начини за борба със смущенията, включително и работа в условия на априорна неопределеност. От Табл. 1 се вижда, че те потискат всички некорелирани смущения, без да е необходимо анализирането им и създаване на специални устройства за потискане и компенсиране на всеки отделен смущаващ сигнал.

ЛИТЕРАТУРА

- [1] Илиев, В., М. Момчеджиков. Метод за предаване на цифрова информация с помощта на интерфериращи шумови сигнали. Сборник с доклади и научни трудове на ИСТ-МВР, 2006.
- [2] Борисов, В. И., А. Е. Зинчук, Н. П. Лимарев, П. И. Мухин, Помехозащищенность систем радиосвязи с расширением спектра сигналов модуляцией несущей псевдослучайной последовательностью. – М., Радио и связь, 2003.
- [3] Боянов, Б., Цифрова обработка на сигналите. Бряг-принт-АД, Варна, 2003.
- [4] Илиев, В., М. Момчеджиков. Методи за инвариантна обработка на сложни сигнали, Годишник на ТУ-София, Том 55, 2005.
- [5] Ipatov, V. Spread Spectrum and CDMA. Principles and Applications. John Wiley & Sons, INC., 2005.
- [6] Bendat, J., A. Piersol. Engineering Applications of Correlation and spectral Analysis. Second Edition. John Wiley & Sons, New York., 1998.
- [7] Урядников, Ю., С.Аджемов. Сверхширокополосная связь. СОЛОН-Пресс, Москва, 2005.
- [8] Киррилов С. Н., А. А. Лоцманов. Многоканальный нелинейный адаптивный фильтр – компенсатор импульсных и узкополосных помех. Журнал Радиотехника № 6, 2006.
- [9] Iliev, I., S. Lishkov, Noise immunity of asynchronous direct spread spectrum systems with ternary sequences, PROC. OF THE TU – SOFIA, Vol. 53, 2004,pp. 159-168.

Автори: Вася Илиев, н. с. I ст. д-р от Институт за Специална Техника - София; Михаил Момчеджиков, проф. д-р – катедра „Радиокомуникации и видеотехнологии”

Постъпила на 14.01.2010

Рецензент доц. д-р Илия Илиев

ФУНКЦИОНАЛНА УСТОЙЧИВОСТ НА АЛГОРИТЪМ ЗА ОПТИМАЛНА ФИЛТРАЦИЯ НА СЛОЖНИ СИГНАЛИ ПО КРИТЕРИЙ ЗА ТОЧНОСТ

Вася Илиев

***Резюме.** В работата е изследвана функционалната устойчивост на метод за оптимална филтрация на сложни, при работа в условия на умишлени смущения и многолъчева интерференция. Първоначално е синтезиран математичен модел на метода, а в последствие е извършено числено изследване. Определени са характеристиките на устойчивост, като функция от параметрите на сигнала, системата и смущението.*

FUNCTIONAL STABILITY OF AN ALGORITHM FOR COMPLEX SIGNALS OPTIMAL FILTRATION BY PRECISION CRITERION

Vasya Iliev

***Abstract.** Functional stability of a method for signal optimal filtration in working conditions of complex signals and multi-ray interference is examined in the work. Initially, a mathematical model of the method is synthesized and then a numerical analysis has been carried. Stability characteristics are defined as a function of signal parameters, the system and the disturbances.*

1. Въведение

В [1] е предложен метод за корелационна обработка на сложни сигнали който запазва структурата им. В [2] е изведен математичен модел на предложениия метод, в резултат на решението на задачата за статистическия синтез. В [3] е извършено математическо моделиране, в резултат на което е доказана функционалната устойчивост на метода в отсъствие на шумове и смущаващи сигнали. Предложеният метод е предназначен основно да повишава шумоустойчивостта на широколентовите системи за комуникация. Във връзка с това в статията е изследвана функционалната му устойчивост при работа в условия на шум и корелирани смущения.

2. Математичен модел на метода при въздействие на отразени сигнали и умишлени смущения

Отразените сигнали, както и организираните смущения, представляват сигнали, еднотипни по структура с полезния сигнал, но отличаващи се от него по амплитуда, честота и монотонно изменящо се във времето закъснение [4]. При попадането им на входа на приемното устройство, в зависимост от момента на постъпване, те могат да предизвикат срив в работата на системата за синхронизация и блокиране на работата на устройството като цяло. Входният сигнален

процес в този случай представлява сума от полезния сигнал и смущението и се задава с израза

$$(1) \quad y(t) = A_s(t) \cdot g[t - \tau_{s_0}(t)] \cos \theta_{s_0}(t) + A_n(t) \cdot g[t - \tau_{n_0}(t)] \cos \theta_{n_0}(t).$$

В (1) означенията са както следва: $A_s, \tau_{s_0}, \theta_{s_0}, A_n, \tau_{n_0}, \theta_{n_0}$ - са съответно амплитудите, закъсненията във времето и фазовите ъгли на сигнала и смущението; $g[\cdot]$ – манипулираща функция, приемаща стойности 1 и –1 в съответствие с кодовата поредица на сигнала. В качеството на модел за промяна на времезакъснението на смущението е избран линеен закон [5]

$$(2) \quad t_{n_0}(t) = t_{s_0}(t) + t_0 + v_n t,$$

където t_0 е началната разлика във времезакъсненията между τ_{n_0} и τ_{s_0} ; v_n – скорост на изменение на времезакъснението на смущението τ_{n_0} . При направените предположения математическият модел на метода изведен в [2] във вид на система нелинейни диференциални уравнения за оценка на математическите очаквания на фазата и времезакъснението θ_s^* и τ_s^* на сигнала се преобразува във вида

$$(3) \quad \begin{cases} \theta^* = \frac{K_\theta}{p} [R^2(\tau_{s_0} - \tau_s^*) \sin(\theta_{s_0} - \theta_s^*) + \mu R^2(\tau_{n_0} - \tau_s^*) \sin(\theta_{n_0} - \theta_s^*)] \cdot F_\theta(p), \\ \tau^* = \frac{K_\tau}{p} [R(\tau_{s_0} - \tau_s^*) D(\tau_{s_0} - \tau_s^*) \cos(\theta_{s_0} - \theta_s^*) + \\ + \mu R(\tau_{s_0} - \tau_s^*) D(\tau_{n_0} - \tau_s^*) \cos(\theta_{n_0} - \theta_s^*)] \cdot F_\tau(p). \end{cases}$$

където $R(v)$ и $D(v)$ – корелационна функция и дискриминационна характеристика на системата за следене на закъснението (ССЗ). За опростяване на анализа се приема, че $F_\theta(p) = 1$ и $F_\tau(p) = 1/(1 + pT_1)$, където T_1 е времекопектанта на НЧФ във веригата на ССЗ; $\mu = A_n \cdot A_s^{-1}$, а $p = d/dt$. Чрез въвеждане на моментна грешка в оценката на фазата и относителна грешка в оценката на времезакъснението

$$(4) \quad \begin{aligned} \varphi &= \theta_{s_0} - \theta_s^* \\ x &= (\tau_{s_0} - \tau_s^*) \tau_u^{-1}, \end{aligned}$$

от система уравнения (3) се получава окончателният математичен модел на системата при работа в условия на умишлени смущения и отразени сигнали

$$(5) \quad \begin{cases} \frac{d\varphi}{dt'} = \gamma - R^2(x) \sin(\varphi) + \mu R^2(x + x_0 + \delta t') \sin(\varphi + \psi + qt'), \\ \frac{dx}{dt'} = b [\beta - x - \alpha R(x) D(x) \cos(\varphi) - \\ - \mu \alpha R(x + x_0 + \delta t') D(x + x_0 + \delta t') \cos(\varphi + \psi + qt')]. \end{cases}$$

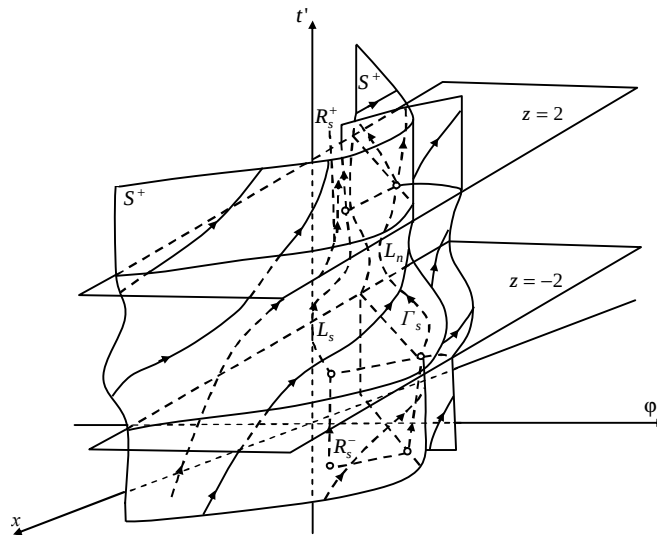
В случая $t' = K_\theta t$; $\gamma = p\theta_{s_0}/K_\theta$; $\beta = p(\tau_{s_0} - t_0)/K_\tau \tau_u$ са относителни начални разстройки на носещите и тактовите честоти; $x_0 = \tau_{s_0} \cdot \tau_u^{-1}$; $b = (K_\theta T_1)^{-1}$; $\delta = v_n K_\tau^{-1}$; $q_\Omega = \Omega K_\theta^{-1}$; Ω и ψ са съответно честотната разстройка и началната фазова разлика между полезния и смустващия сигнали; $\alpha = K_\tau \tau_u^{-1}$ е максималната стойност на относителното времезакъснение, което може да бъде компенсирано от ССЗ.

Когато параметъра $\mu < 1$ $A_n < A_s$, или в отсъствие на смущение $\mu = 0$, динамиката на системата може да бъде описана с помощта на изходния математичен модел съгласно [2].

В случаите, когато параметъра $m > 1$ в системата (5) могат да бъдат установени два вида синхронни стационарни режими: режим P_s , при който се осъществява синхронизация по полезния сигнал; режим P_n , при който се осъществява синхронизация по смущението. В областта $|x| \geq 2$ в системата (5) е еквивалентна на системата [6]

$$(6) \quad \begin{cases} \frac{d\theta}{dt_1'} = \gamma_1 - R^2(z) \sin \theta_\Delta, \\ \frac{dz}{dt_1'} = b_1 [\beta_1 + \delta_1 \tau_1 - z - KR(z)D(z) \cos \theta_\Delta], \end{cases}$$

където: $\theta_\Delta = \varphi + \psi + qt'$ и $z = x + x_0 + \delta t'$ са съответно разлики във фазите и времезакъсненията между смустващия и опорния сигнал; $\gamma_1 = (\gamma + q)\mu^{-1}$ и $\beta_1 = \beta + x_0 + \delta b^{-1}$ са началните разстройки на носещите и тактовите честоти на смустващия и опорния сигнал; $b_1 = b\mu^{-1}$ и $\delta_1 = \delta\mu^{-1}$ – безразмерни времеви константи; $K = \mu\alpha$ – максимална стойност на относителното закъснение което може да бъде компенсирано от ССЗ; $t_1' = \mu t'$.



Фиг. 1. Характер на поведението на траекториите на разглежданата система във фазовото пространство φ, x, t' за $\mu > 1$

Синхронните режими P_s и P_n се описват на фазовата плоскост φ, x, t' с помощта на устойчиви ограничени траектории L_s и L_n на системата (5), неизлизащи при $-\infty < t' < \infty$ извън границите на фазовите координати φ, x , както е показано на фиг.1.

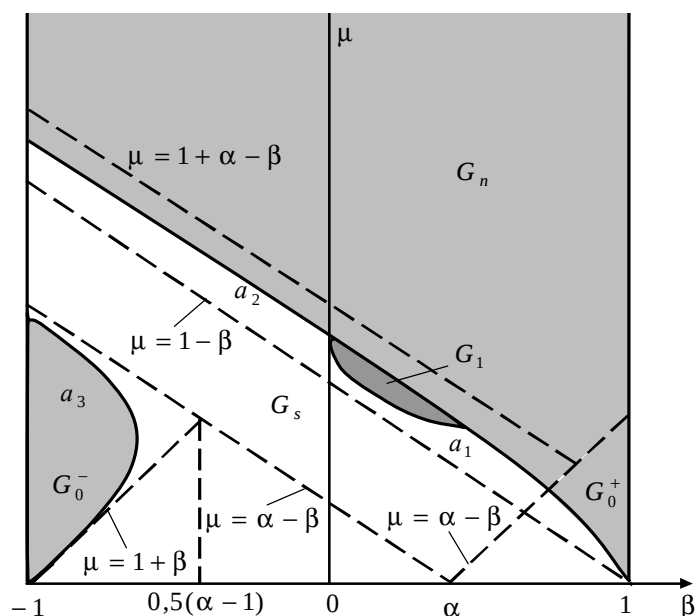
3. Изследване на шумоустойчивостта на метода при въздействие на смущението на захваната система

Въздействието на смущението на захваната система е налице в случаите, когато смущението е изостанало по време от полезния сигнал на стойност, при която грешката в следене на закъснението z е извън обхвата на дискриминационната характеристика на ССЗ. В зависимост от стойността на μ и параметрите на системата е възможно запазване на режима P_s с вероятност p_s или възникване на срыв в следенето и преминаване към режим следене на смущението P_n с вероятност $p_0 = 1 - p_s$.

В първия случай под въздействието на смущаващия сигнал, изобразяващата точка се придвижва във фазовото пространство φ, x, t' по траекторията L_s , фиг.1, отклонява се от стационарното състояние (φ_s, x_s) и с течение на времето излиза от областта $z = x + x_0 + \delta t' \leq -2$ на фазовото пространство и попада в областта на притегляне на полуправата R_s^+ фиг.1.

Преход към следене на смущението настъпва, ако изобразяващата точка в областта $x > 2$ при нарастване на t' не прескача областта $|z| < 2$ и попада в областта на притегляне

$$(7) \quad \Pi(R_n^+) \text{ полуправа } R_n^+ : z = (\beta - \alpha)\mu^{-1}, \quad x \geq 2.$$

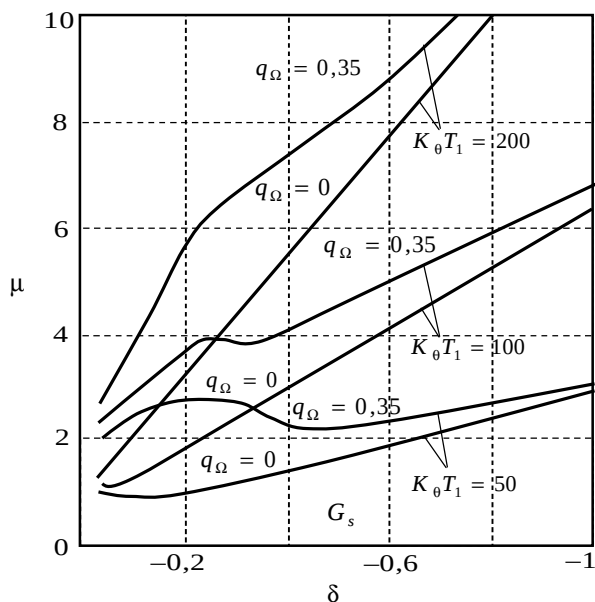


Фиг. 2. Области с параметри μ и β при $\alpha = \text{const} < 1, \varphi = 0$

На основа на числено изследване на установяващите се движения L_s и L_n на системата (5) е извършено разбиване на пространството на параметрите μ , β , α на области, съответстващи на различните режими на работа на системата. На фиг. 2 са показани тези области с параметри μ , β , при които $\alpha = \text{const} < 1$

Бифуркационните линии a_1 , a_2 , a_3 (фиг. 2) разделят пространството от параметри на системата (5) на области G_s, G_1, G_0^+, G_0^- и G_n . В областта G_s (областта на съхранение на следенето) отклоненията от стационарното състояние $x = \beta$, предизвикани от смущението, се компенсират от системата. За останалите стойности на параметрите смущението е ефективно и предизвиква срыв в следенето. При това за стойности на параметрите от областите G_0^+, G_0^- и G_1 е налице неограничено нарастване на абсолютните стойности на координатите $|x|$ и $|z|$. За стойности на параметрите от областта G_n се установява режим на синхронизация на смущаващия сигнал. Областта на съхранение на следенето G_s се явява **характеристика на устойчивост** на ССЗ към *корелирани смущения*. G_s е област от стойности на параметрите на системата и смущението, при които в системата (5) съществува устойчива траектория L_s , изцяло разположена в областта v на фазовото пространство φ, x, t' .

На фиг. 3 са показани графичните зависимости $\mu = \mu_s(\delta)$, определящи областта $G_s: \mu < \mu_s(\delta)$, съхраняваща режима на следене за полезния сигнал P_s с вероятност $p_s = 1$, определени числено за стойности на параметрите: $\gamma = 0,2$; $\beta = -0,5$; $\alpha = 5,0$; $K_\theta T_1 = 20, 50, 100$; $q_\Omega = 0, 0,35$ от скоростта на изменение на времезакъснението на смущаващия ШПС – δ , където възниква срыв в следенето за стойности на $\mu \approx 1..$



Фиг. 3. Графични зависимости $\mu = \mu_s(\delta)$, определящи областта G_s

При намаляване на началната честотна разстройка q_Ω , ефективността на смущението нараства, а големината на областта G_s намалява, така че срыв в следенето при едни и същи стойности на δ възниква при по-малки стойности на μ .

4. Изследване на шумоустойчивостта на метода при едновременно въздействие на сигнала и смущението

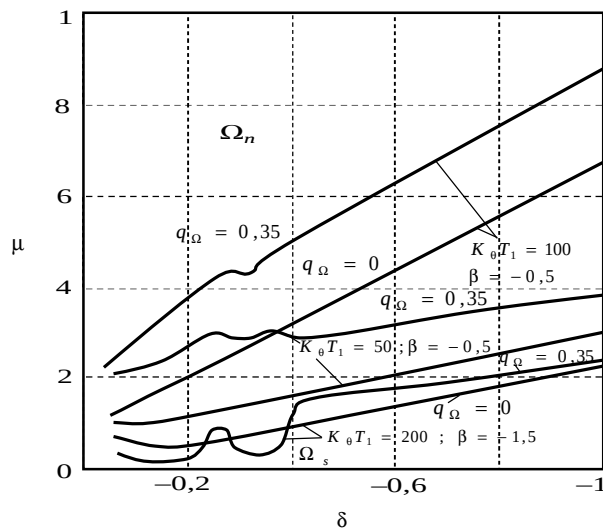
В случаите на едновременно въздействие на сигнала и смущението последното влияе на процеса на установяване на синхронизация.

Възможността за установяване на режим на работа P_s или P_n , под въздействието на смущението се определя от взаимното разположение във фазовото пространство на отрязъка L_0 областите на притягане на установяващите се движения на траекториите L_s или L_n .

Установяването на синхронен режим P_s настъпва, ако за всички стойности на параметрите x_0 и ψ от областта $V_n : \{0 \leq x_0 \leq 4, 0 \leq \psi \leq 2\pi\}$, всички траектории, започващи в момента $t' = 0$ на отрязъка L_0 , попадат при нарастване на t' в областта на притегляне на полуправата R_s^+ на състоянието на равновесие A_1 (фиг. 1) и не излизат от областта V_s . Областта от стойностите и параметрите на системата и смущаващия СШПС, при които това е налице, представлява *област на захващане на системата по полезния сигнал* - W_s .

Ако всички точки на отрязъка L_0 за всички стойности x_0 и ψ от областта V_n принадлежат на областта на притягане $\Pi(B_1)$ на стационарното състояние B_1 , то за тези стойности на параметрите на модела (5) принадлежи областта на захващане на системата по смущаващия сигнал Ω_n и съответно на установяване на режим на работа P_n .

Устойчивостта на модела (5) в разглеждания случай се оценява по големината на областта W_s , за която независимо от влиянието на смущението се установява режим P_s . Резултатът от численото изследване на процеса на захващане на полезния СШПС и определянето на областите Ω_s и Ω_n с вероятности $q_s=1$ и $q_n=1$ за конкретни стойности на параметрите γ , β , b и q_s са показани на фиг. 4. В случая параметрите на системата са: $\gamma = 0,2$; $\alpha = 5,0$; $K_0 T_1 = 20, 50$ и $\beta = -0,5, -1,5$ и принадлежат на областта V_s за стойности на параметъра $q_w = 0, 0,35$.



Фиг. 4. Графични зависимости $\mu = \mu_s^*(\delta)$, определящи областта на захващане Ω_s и областта на блокиране на полезния сигнал

5. Заключение

Получените резултати за областите на устойчивост G_s и W_s на модела (5) могат да бъдат използвани за оценка на устойчивостта на системата при въздействие на умишлени смущения, а така също за оценка на ефективността на смущението, изхождайки от условието за блокиране на системата.

По такъв начин, изследването на устойчивостта на двумерната корелационно екстремална следяща система при въздействие на корелирани смущения води към следните резултати. Синхронните режими на работа на системата се определят от ограничени по координатите x и z установяващи се движения на модела (5). Характеристиките на шумоустойчивост на системата се явяват областта на съхранение на режима на следене G_s с вероятност $p_s = 1$ и областта на захващане на системата по полезния сигнал W_s с вероятност $q_s = 1$. Нарушението на устойчивостта на системата е свързано с излизане на грешката в следене на закъснението x извън границите на дискриминационната характеристика на ССЗ или с преход към следене на смущението. Зависимостта на големината на областите G_s и W_s от параметрите на системата и смущението позволяват да бъдат решавани задачи свързани с оптимизиране на ССФ и ССЗ на двумерната система, изхождайки от изискванията за тяхната устойчивост към въздействие на корелирани смущения. Алгоритъма за определяне на областите на устойчивост осигурява решението на задачите свързани с оценка на влиянието на параметрите на ниско пропускащите филтри на характеристиките на шумоустойчивост на системата

ЛИТЕРАТУРА

- [1] Илиев, В., М. Момчеджиков. Метод за корелационна обработка на сложни шумоподобни сигнали. Списание "E+E", № 7-8, 2006г.
- [2] Илиев, В., М. Момчеджиков. Оптимална филтрация на DSSS сигнали. Годишник на Технически Университет – София. Том 58, кн. 1, 2008г.
- [3] Илиев, В., М. Момчеджиков. Изследване на преходните процеси и устойчивостта на алгоритъм за оптимална филтрация на DSSS сигнали. Годишник на Технически Университет – София. Том 58, кн. 1, 2008г.
- [4] Ipatov, V. Spread Spectrum and CDMA. Principles and Applications. John Wiley & Sons, INC., 2005.
- [5] Урядников, Ю., С. Аджемов. Сверхширокополосная связь, СОЛОН-Пресс, Москва, 2005.
- [6] Борисов В. И., А. Е. Зинчук, Н. П. Лимарев, П. И. Мухин Помехозащищенность систем радиосвязи с расширением спектра сигналов модуляцией несущей псевдослучайной последовательностью. М. Радио и связь, 2003.
- [7] Iliev, I., M. Nedelchev. Performance Analysis of a suboptimal multiuser detection algorithm, ICEST 2007, Macedonia, pp.531- 534.

Автор: Вася Илиев, н. с. I ст. д-р от Институт за Специална Техника – София.

Постъпила на 14.01.2010

Рецензент доц. д-р Илия Илиев

СЪЗДАВАНЕ НА ПАНОРАМНИ ИЗОБРАЖЕНИЕ ТИП „МОЗАЙКА” С ПРИЛОЖЕНИЕ ПРИ ВИДЕОНАБЛЮДЕНИЕ

Владимир Кънчев, Огнян Бумбаров

***Резюме.** Представен е модифициран метод за създаване на панорамно изображение на задния фон от тип ” мозайка ” в последователност от кадри, използващ геометрично хеширане. Прилагането на този тип изображение в системите за видеонаблюдение, улеснява отделянето и проследяването на движещи се обекти и процеса на Пан управление на ТВ камера.*

CREATING PANORAMIC IMAGE MOSAICS FOR VIDEO SURVEILLANCE PURPOSES

Vladimir Kanchev, Ognian Boumbarov

***Abstract.** The current paper presents a geometric hashing method for fast construction of background panoramic mosaics in video sequences. The created background panoramic mosaics can be used to facilitate extraction and tracking of moving objects, in order to have an accurate pan camera control.*

1. Introduction

Automatic construction of high resolution images has long been an interesting research topic in the area of video surveillance systems. Currently it is used frequently for stabilization and recognition of variations in scenes, video compression, automatic extending of Field of View (FOV) and camera resolution.

There are two basic approaches for the construction of mosaic images – hardware and software. In the first case, various technical devices are used during capturing of images, i.e. in catadioptric methods an optical system uses a set of special lenses and mirrors while in dioptric methods only common lenses are applied [1]. In the second case of mosaic image construction only software methods are used and there are no specific requirements for the hardware devices.

The software methods for creating image mosaics can be applied in two ways: during camera translation around X or/and Y axes plus rotation around Z axis on one hand, and rotation only around Y axis (Fig. 1) on the other hand. This paper is related to the latter case. The construction of panoramic image mosaics covers the following stages: detection of specific objects (lines, points and regions) from referent and test (current) images, and registration of their corresponding characteristics in relation to

referent image; calculation of the transformation function based on those characteristics, transformation of the test images, detection of overlapping areas followed by a “cut and paste” process and a final filtering of the constructed image mosaic.

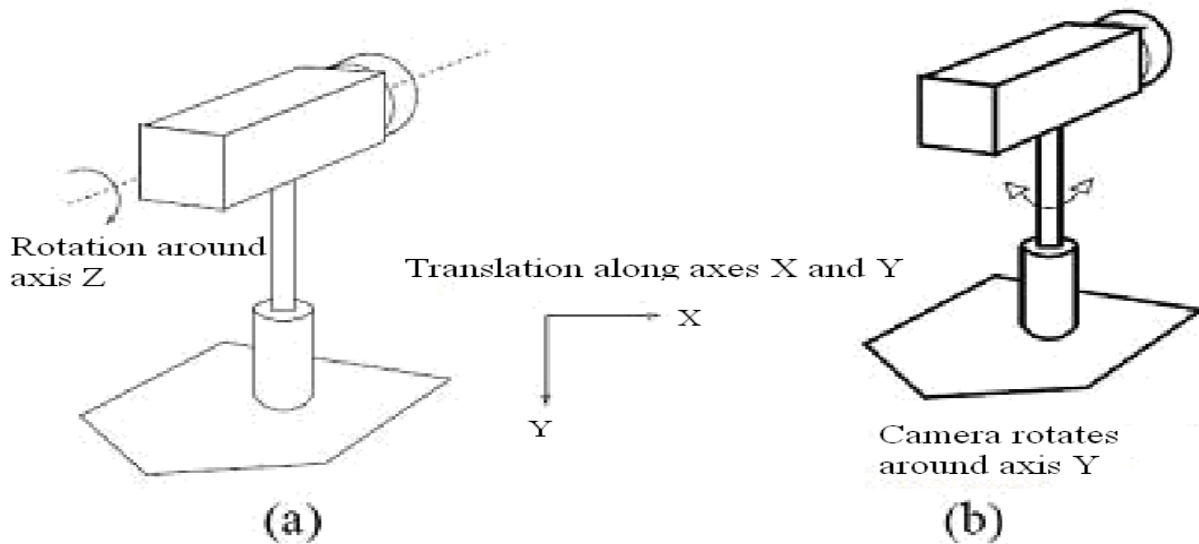


Fig. 1. (a) Planar mosaic (b) Panoramic mosaic

The registration of characteristics is carried out in following stages:

- Detection of characteristic objects – lines, points and regions. Therefore Canny, Laplacian and Gaussian detectors are applied.
- Registration or alignment of characteristics – finding correspondences between characteristic objects. For that purpose a degree of correspondences or cross-correlation between control points of two images is calculated [2] or an algorithm for detection of consecutive self-similarity is developed [3]. In order to increase the rate of registration, pyramidal and wavelet image decompositions are introduced [4].
- Calculation of the transformation function – finding the type and parameters of the transformation model. Goshtasby calculates the parameter values where the number of characteristic points should fall into a certain range [5], Stockman estimates parameters with highest probability values by clusterization [6]. Lamdan proposes fast 2D method for registration by similarity transformation where corresponding points are inserted into hash tables [7].
- Transformation of the current image pixels is done according to the referent image, based on the calculated function.

Overlapping areas of the current and referent images can be determined by intensity or color values either based on a single image or based on all images.

There are few types of modeling of overlapping areas: similarity and affinity transformation. It has been found out that local overlapping gives better results in the presence of local distortions [5]. At the same time methods with radial basis functions are global methods that correct such kind of distortions fairly well [8].

The overlapped areas are filtered to suppress differences into intensity and color values of different images. Their pixel values are substituted with one of their

estimated parameters: mean value, median value, last pixel value or bilinear merging of corresponding image areas.

In the current paper we present an algorithm for construction of panoramic image mosaics based on modified geometric hashing.

2. Description of the suggested method

Video surveillance systems operate in the following way: detection of intruder, his identification and subsequent tracking to make a description of his action [9]. A big number of such systems use only one stationary camera to construct a background model. Others use non-stationary Pan-Tilt cameras. Those cameras can dynamically change direction to extend FOV to be able to detect and track moving objects and finally create a background model.

The current system of video surveillance (Fig. 2) performs indoor tracking of moving objects using pan camera. The aim of this paper is to create an algorithm for background image mosaic, based on geometric hashing. That algorithm can be applied to detect, identify and extract a certain moving object (object of interest) from a set of frames. Afterwards the object of interest is tracked by Kalman filter and its predicted position can be used for controlling the pan movement of the camera.

The algorithm for constructing background image mosaics consists of the following stages:

- Extraction of characteristic features - angles of reference and test images with SUSAN operator [10].
- Correspondences between detected points are calculated by modified geometric hashing in order to achieve better rate of registration [11].
- Parameters of projective transformation between correspondent points are calculated.
- Then test images are transformed by calculated function and they are imposed on reference image through exact coincidence of overlapping areas.
- Filtering of the final image in order to remove redundant transitions. We use gradient method with fixed index table for overlapping areas.

The background model represents a mixture of Gaussian models that cover all background pixels. The model modifies its parameters when there is an alteration of external circumstances – illumination, for example.

In a sequence of frames the object is detected through its movement. The corresponding part of the background mosaic is subtracted from every single frame to facilitate tracking.

Tracking of an object from adjacent frames is performed by Kalman filtering, where the consecutive movement of the object is used for faster detection of its position.

The main purpose of camera pan movement control is to preserve the tracked objects within camera FOV.

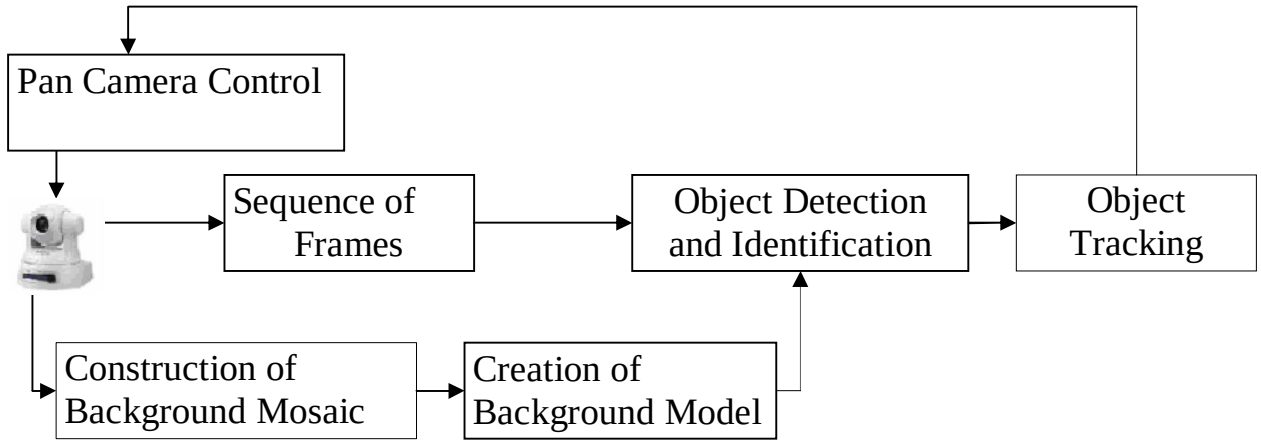


Fig. 2. A video surveillance system

The algorithm for geometric hashing has been proposed by Lamdan and its purpose is to find the characteristic objects through their local geometric features, which are invariant to various transformations and are less influenced by collision of objects in scene [7]. Its main advantage is the opportunity of parallel implementation.

It has been proved that the following number of characteristic points is sufficient for the introduction of the following transformation models in order to achieve invariance of geometrical features [12]:

- Translation – 1 point.
- Translation and rotation – 2 points.
- Translation, rotation and scaling – 2 points.
- Affine transformation -3 points.
- Projective transformation – 4 points.

The complexity of stage of a preliminary processing that includes construction and filling of hash table is $O(M.n^{c+1})$ and complexity of stage of recognition, composed of verification of model with detected points from test image, is $O(H.S^{c+1})$ where [12]:

- M - number of models with n features.
- S - number of features of scene.
- c - minimum number of features, necessary for forming a basis of model.
- H - complexity of operations on a row of hash table.

When we search coincidence between characteristic points between test and reference images, complexity of operation is $O(M^{c+1}.S^{c+1})$ [12].

In panoramic mosaics the following model for camera is used [13]:

$$l p = A[R/t] P, \quad (1)$$

where:

- $P = [X \ Y \ Z \ 1]^T$ is a vector with coordinates of 3-D point from reference frame.

- $p = [x \ y \ 1]^T$ is a point of image, corresponding to P .
- l is a projective constant.
- A represents a matrix of internal parameters.
- R is a rotation matrix.
- t is a vector of transformation.

When the camera is directed to a point P in a scene from two separate positions [13]:

$$P' = RP + t, \quad (2)$$

In the case of a panoramic image $t = 0$, thus $lA^{l-1}p' = lRA^{-1}p$. And it follows [11]:

$$mp' = Hp, \quad (3)$$

$$m \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad (4)$$

where H is a 3×3 nonsingular homographic matrix that represents 2D projective transformation. After short calculations we reach the following solution [11]:

$$x' = \frac{h_1x + h_2y + h_3}{h_7x + h_8y + 1} \quad y' = \frac{h_4x + h_5y + h_6}{h_7x + h_8y + 1} \quad (5)$$

Every correspondence gives two equations, hence for calculation of H are necessary four non-collinear correspondent points. Basis vector can get $\binom{m}{4}m!$ different values where m is a number of points in image.

3. Algorithm for registration of characteristic points

1. The model in reference image is introduced by a set of angles after application of SUSAN feature extraction operator.

2. For every quadruple of characteristic points (three of them non-collinear) are calculated two linear independent vectors l_1 and l_2 and the hash table is filled with $\binom{m}{4}$ groups for Q and l values (Fig. 3) [11].

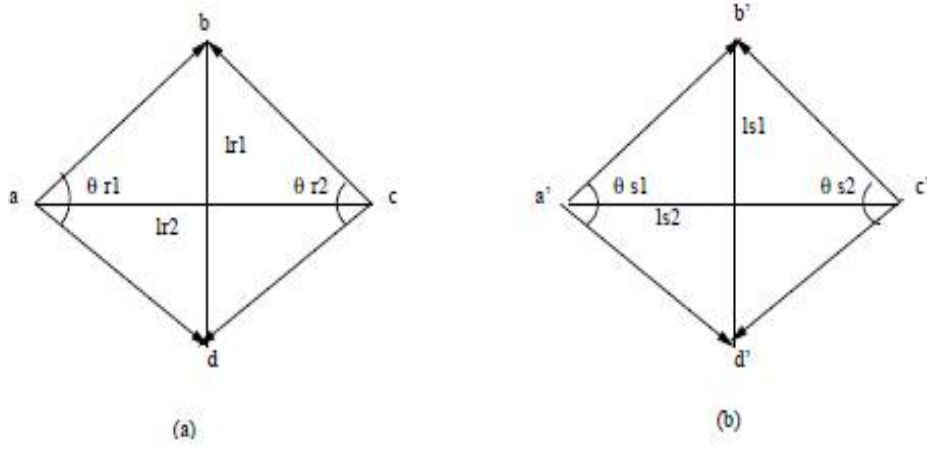


Fig. 3. (a, b, c, d) quadruple of reference image and (a', b', c', d') quadruple of test image.

3. For every quadruple $\left(\frac{n}{4}\right)$ from test image respective values for Q and l are determined where n is a number of characteristic points.

4. For every basis quadruple from test image differences between angles and $q_{s1(j)}, q_{r1(j)}, q_{s2(i)}, q_{r2(i)}$ are calculated:

$$d_{q1(i,j)} = \left| q_{s1(j)} - q_{r1(j)} \right| \quad (8)$$

$$d_{q2(i,j)} = \left| q_{s2(j)} - q_{s2(i)} \right| \quad (9)$$

Similarly differences between corresponding sides are calculated:

$$d_{l1(i,j)} = \left| l_{s1(j)} - l_{r1(i)} \right|, \quad (10)$$

$$d_{l2(i,j)} = \left| l_{s2(j)} - l_{r2(i)} \right|, \quad (11)$$

where $i = 1, 2, 3, \dots, \left(\frac{m}{4}\right)$ and $j = 1, 2, 3, \dots, \left(\frac{n}{4}\right)$.

Those $\left(\frac{m}{4}\right)\left(\frac{n}{4}\right)$ quadruples, whose difference values are higher than a certain threshold, are discarded. The rest quadruples are sorted into ascending way according to their $d_{l1(i,j)}$ and $d_{l2(i,j)}$ values, element with smallest values - q_1, q_2, l_1, l_2 from reference image is selected and we seek their correspondent points in test image. The parameters of transformation are determined by the method of least square and the whole test image is transformed by a calculated function.

The proposed algorithm consists of following stages:

1. Preliminary processing – selecting a model of characteristic points from reference image.
2. Selecting of characteristic points after applying SUSAN operator on a test image [10].
3. Construction of a hash table:
 - Pick out a basis of two points and calculate coordinates of all characteristic points from reference image.
 - After the quantization of coordinates we create a hash table and then make a registration with characteristic points from test image, obtained after applying SUSAN operator, as well.
4. Selection of optimal model and basis:
 - A histogram of coincidence between characteristic points from two images is constructed and we select a basis of points from a test image whose bin values are above certain threshold.
 - For the selected basis we calculate transformation T of points, whose square error between correspondent points is the smallest.
 - Difference between correspondent points is calculated and if it exceeds a certain threshold, the next quadruple from the list is loaded and we continue from point 2.
5. The test image is transformed by the selected function and is imposed on the reference image and finally we perform post filtering on mosaic image.

4. Results

The algorithm for creation of panoramic image mosaics has been developed on script language of Matlab2008a. We have worked with a sequence of test images, captured during pan movement of the camera at a certain small angle. The second image from a sequence serves as a reference image (Fig. 4) and the first and third ones - as a test set (Fig. 5). Four points are selected on each image. The points in the reference image are selected manually while the points on the test images are calculated after applying a SUSAN operator. The projected correspondence parameters are calculated on pairs of correspondent quadruples. Then all pixels of the two test images are transformed and imposed on top of the referent image. Finally the panoramic image mosaic is obtained (Fig. 6).



Fig. 4. A reference image

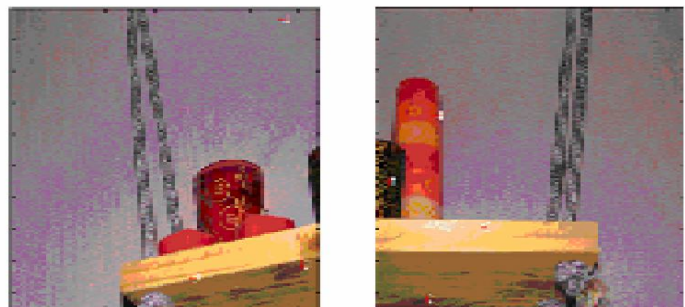


Fig. 5. Input test images

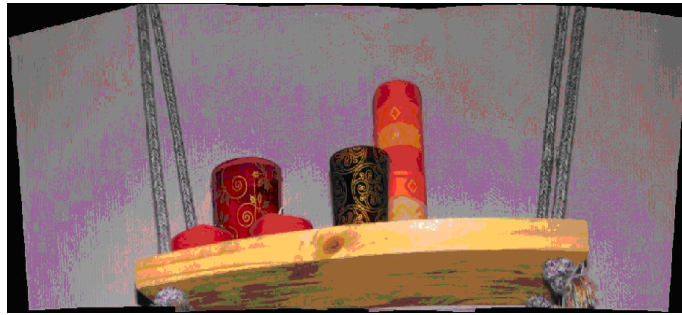


Fig. 6. An output panoramic mosaic image

5. Conclusions

In the current paper an algorithm for creation of panoramic image mosaic with geometric hashing from a sequence of images, captured after pan camera control, is proposed. The main purpose is to create a general background panoramic image mosaic of a scene. The general background panoramic image mosaic will facilitate detection, extraction and tracking of moving object. Information from these operations is used to control pan movement of camera during tracking.

ACKNOWLEDGEMENTS:

This work was supported by The Ministry of Education and Science, The National Science Fund by contract DO02-41/2008 “Biometrical Identification in Systems of Video Surveillance”.

REFERENCES

- [1] Parian, J. ,Panoramic Imaging: Techniques, Sensor Modeling and Applications
- [2] Roche, A., G. Malandain, X. Pennec, N. Ayache, “The correlation ratio as a new similarity measure for multimodal image registration,” in *Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention* (MICCAI’98).
- [3] Barnea, D., H. Silverman, F. Harvey, “A class of algorithms for fast digital image registration,” *IEEE Transactions*, Volume C-21 (2), pp. 179-186, 1972
- [4] Turcajova, R., J. Kautsky, “A hierarchical multiresolution technique for image registration,” *Proceedings of SPIE Mathematical Imaging: Wavelet Applications in Signal and Image Processing*, 1996.
- [5] Goshtasby, A., G. Stockman, C. Page,” A region-based approach to digital image registration with subpixel accuracy,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 24, no. pp. 390-399, 1986.
- [6] Stockman, G., S. Kopstein, S. Benett,” Matching images to models for registration and object detection via clustering,” *IEEE Transaction Pattern Analysis Machine Intelligence*, vol. PAMI-3, pp. 229-241, 1982.

- [7] Lamdan, Y., H. Wolfson, Geometric Hashing,” A General and Efficient Model Based Recognition Scheme, “ *Proc. Int'l Conf. Computer Vision, IEEE Press, Piscataway, N.J.*, pp. 238-249, 1988.
- [8] Arad, N., N. Dyn, D. Reissfeld, Y. Yeshurun, “Image warping by radial basis functions,” *Applications to facial expressions. CVGIP: Graphical Models and Image Processing*, 56, 161-172, 1994.
- [9] Biswas, A., A. Mukerjee, P. Guha, K. Venkatesh,” Detecting and Tracking Intruders using a Pan-Tilt Surveillance System,” *Proceedings of the Third International Conference on Visual Information Engineering (VIE)*, Bangalore, September 26-28, 2006, pp. 565-576.
- [10] Smith, S., J. Brandy, “SUSAN – A new approach to low level image processing,” *International Journal of Computer Vision*, 23(1), pp. 45-78, 1997.
- [11] Udhav, B., S. Chaudhuri, S. Roy, “A Fast Method for Image Mosaicing using Geometric Hashing,” *IETE Journal of Research: Special Issue On Visual Media Processing*, pp. 317-324, 2002.
- [12] Wolfson, H, I. Rigoutsos, “Geometric Hashing: An overview,” *IEEE Computational Science and Engineering*, 4(4), 10-21, 1997.
- [13] Hartley, R., A. Zisserman, Multiple View Geometry in Computer Vision, *Cambridge University Press*, 2000.

Authors: Vladimir Kanchev is Ing., PhD Student at Department Radio Communications and Video Technology, Faculty of Telecommunications; Ognian Boumbarov PhD, is with Department Radio Communications and Video Technology, Faculty of Telecommunications.

Received 14.01.2010

Reviewer Prof. Mikhail Momchedjikov

АНАЛИТИЧНО ОПРЕДЕЛЯНЕ НА ЪГЪЛА НА ИЗМЕСТВАНЕ НА ЧЕТКИТЕ ОТ НЕУТРАЛНАТА ЛИНИЯ ПРИ КОЛЕКТОРНИТЕ ЕЛЕКТРИЧЕСКИ МАШИНИ

Ганчо Божилов, Адриан Иванов

Резюме. В работата е предложен метод за аналитично определяне на ъгъла на изместване на четките от неутралната линия при колекторни електрически машини, необходим за постигане на благоприятна комутация. Предложени са формули за определяне на ъглите за компенсиране на реактивното електродвижещо напрежение в комутиращата секция и това от напречната реакция на котвата, както и оптималният ъгъл за получаване на линейна комутация. Дадени са и формули за определяне на електродвижещите напрежения при неизместени и изместени четки.

ANALYTICAL EVALUATION OF THE ANGLE OF DISPLACEMENT OF THE BRUSHES FROM THE NEUTRAL LINE ON THE COMMUTATOR ELECTRICAL MACHINES

Gantcho Bojilov, Adrian Ivanov

Abstract. In this paper a method for analytical evaluation of the angle of the displacement of the brushes from the neutral line of the commutator electrical machines is proposed. A formula of the evaluation of the angles for compensation of the electromotive forces in the commutator section and this of the armature reaction and the optimal angle for linear commutation is proposed. A formula for evaluation of the electromotive forces by non displaced and displaced brushes is given.

1. Увод

Както е известно, един от най-ефикасните начини за подобряване на комутацията при колекторните електрически машини без допълнителни полюси или компенсационна намотка, е изместването на четките от геометричната неутрална линия (г.н.л.). Конструктивно това се осъществява с помощта на подвижна траверса, като при генераторите изместването трябва да бъде в посоката на въртене на ротора, а при двигателите – в обратна посока. Стойността на ъгъла на това изместване се определя опитно – до получаване на удовлетворителна комутация. Обаче при машините с малка мощност по конструктивни съображения не

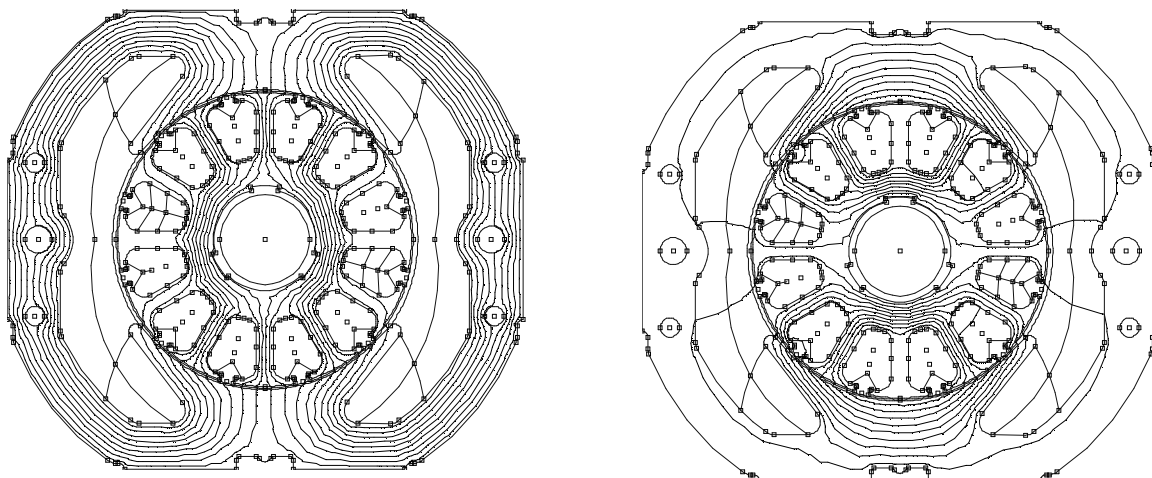
се поставя траверса, а необходимото изместване се осъществява чрез фиксирано изместване на свързването на секциите на котвената намотка към колектора – при генераторите в посока, обратна на посоката на въртене на ротора, а при двигателите – в посоката на въртене на ротора. Поради това опитното определяне на стойността на необходимия ъгъл е извънредно трудоемка и скъпа задача, свързана с изработване на не малък брой котви с различни ъгли на изместване. Ето защо е необходимо да се създаде подходящ аналитичен и приложен метод, който с достатъчна точност и прегледност да позволява предварителното определяне на този ъгъл още на стадия проектиране.

Съгласно класическата теория на електрическите машини [1,2,3,4], необходимият ъгъл α на изместване на четките от г.н.л. (или на свързването на секциите към колектора) се състои всъщност от два ъгъла – един основен ъгъл α_1 , необходим за компенсиране на изместването на физическата неутрална линия (ф.н.л.) от геометричната (г.н.л.), и още един допълнителен ъгъл $\Delta\alpha$, който заедно с първия да даде възможност за компенсиране на реактивното електродвижещо напрежение (е.д.н.) E_r в комутиращата секция, индукирано от самоиндуктивността и взаимната индуктивност на секциите. Едновременно с това изместването на ъгъл α_1 компенсира е.д.н. E_q , индукирано в комутиращата секция от потока на напречната реакция на котвата, което иначе се прибавя към първото.

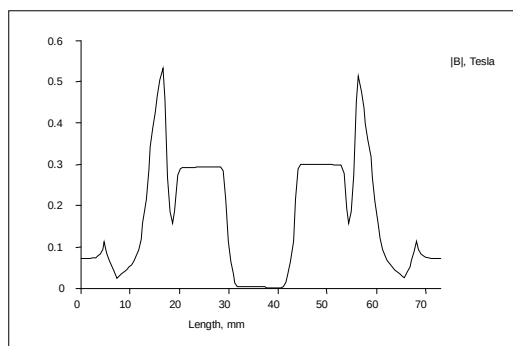
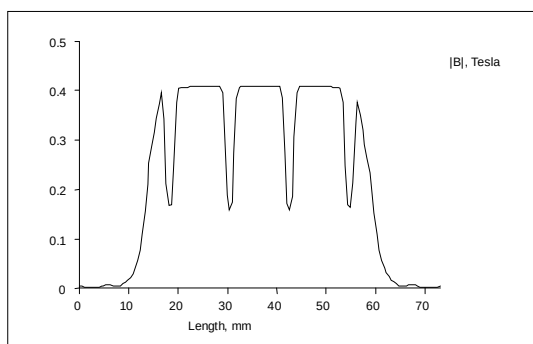
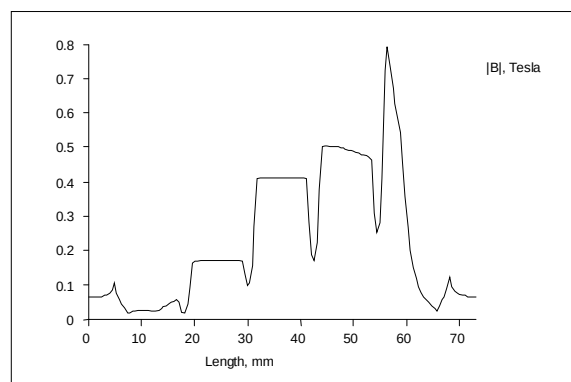
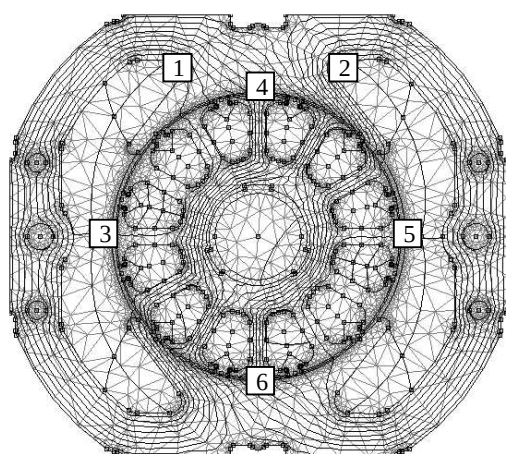
2. Описание на предлагания метод

Основният ъгъл α_1 , който е ъгълът между г.н.л. и ф.н.л., е всъщност мястото по обвода на котвата, където нормалната компонента на резултантната магнитна индукция във въздушната междина при натоварване е равна на нула. Това е мястото, където индукцията от възбудителното магнитно поле на полюсите е равна и противоположна на индукцията от размагнитващата реакция на котвата. Това място се намира в междуполусното пространство, където разпределенията на двете магнитни индукции по протежение на едно полюсно деление имат съвсем различен характер. Това е илюстрирано на фиг. 1, където разпределенията на магнитните индукции за един примерен двигател са получени с помощта на метода на крайните елементи чрез програмния продукт FEMM [5,6, 9,10].

Ако се абстрахираме от деформацията на магнитната индукция, дължаща се на назъбването на котвата, както това се прави обикновено в класическата теория на електрическите машини, разпределенията на двете магнитни индукции по периферната координата x в междуполусното пространство могат да се апроксимират със степенни функции с различни степенни показатели – фиг. 2.



Фиг. 1



Фиг. 2

За да се получи по-удобна формула за α_1 е подходящо да се приемат следните предпоставки.

Приемат се първоначално неизместени четки ($\alpha=0$), а оптималният ъгъл α_{opt} се намира впоследствие. Разпределенията на индукциите се апроксимират с параболи от четвърта и от втора степен [7,8]. Тогава се получава:

а) магнитна индукция от възбуждането

$$(1) \quad B_d(x) = B_{drez} \left(\frac{x}{a_m} \right)^4,$$

където B_{drez} е резултантната магнитна индукция на полето на полюсите с отчитане на размагнитващото действие на реакцията на котвата, $\alpha_m = 0,5 \left(\frac{180^\circ}{p} - \beta \right)$ е ъгъл, равен на широчината на междуполюсното пространство, β е централният ъгъл на полюса, изразен в геометрични градуси, p – броят на двойките полюси на индуктора.

б) магнитна индукция от напречната реакция на котвата

$$(2) \quad B_a(x) = B_{drez} \left(\frac{a_p F_a}{F_{rez}} \right) \left(\frac{x}{a_m} \right)^2,$$

където $\alpha_p = \frac{\beta p}{180}$ е коефициентът на полюсната дъга, $F_a = At k_y$ - максималната стойност на магнитодвижещото напрежение (м.д.н.) от реакцията на котвата, A – линейното токово натоварване (л.т.н.) на котвата, t - полюсното деление, k_y - коефициентът на скъсяване на котвената намотка, F_{rez} – резултантното м.д.н. за магнитната верига, което е сума от всички действащи м.д.н.

Ъгълът между ф.н.л. и г.н.л. е мястото, където разликата на индукциите $B_d(x) - B_a(x) = 0$, т.е.

$$(3) \quad \left(\frac{x}{a_m} \right)^4 = \left(\frac{a_p F_a}{F_{rez}} \right) \left(\frac{x}{a_m} \right)^2,$$

откъдето основният ъгъл на изместване на четките, необходим за компенсиране на изместването на ф.н.л. от г.н.л. и на е.д.н. E_q , ще бъде

$$(4) \quad a_1 = a_m \sqrt{\frac{a_p F_a}{F_{rez}}}.$$

Този резултат се потвърждава и чрез програмния продукт FEMM.

Ъгълът на изместване на четките, необходим за компенсиране на E_r , се получава от изравняването на реактивното е.д.н. E_r с комутационното E_k , индуктирано от разликата на индукциите $B_d(x) - B_a(x)$.

$$E_r = 2\omega_c k_y v_a l_d A k_y x$$

$$E_k = 2\omega_c k_y v_a l_d [B_d(x) - B_a(x)],$$

където w_c е броят на навивките в секция, v_a - периферната скорост на котвата, l_d - изчислителната дължина на котвата, x - коефициент на специфичната магнитна проводимост на самоиндукция и взаимна индукция на комутиращата секция. Тогава

$$\frac{Ak_y x}{B_{drez}} = \left(\frac{x}{a_m}\right)^4 - \left(\frac{a_p F_a}{F_{rez}}\right) \left(\frac{x}{a_m}\right)^2,$$

Като се има пред вид, че според ур. (4) $\frac{a_p F_a}{F_{rez}} = \left(\frac{a_1}{a_m}\right)^2$, се получава уравнението

$$(5) \quad x^4 - a_1^2 x^2 - c a_m^4 = 0,$$

решението на което е

$$(6) \quad a_2 = 0,707 a_1 \sqrt{1 + \sqrt{1 + 4c \left(\frac{a_m}{a_1}\right)^4}},$$

$$\text{където } c = \frac{Ak_y x}{B_{drez}}.$$

Получава се, че $\alpha_2 > \alpha_1$, т.е. допълнителният ъгъл на изместване за компенсиране на E_r е $\Delta\alpha = \alpha_2 - \alpha_1$.

Е.д.н. в комутиращата секция при изместени четки (или свързване към колектора) ще се получат от зависимостите

$$(7) \quad E_q' = E_q \left[1 - \text{sign}(a) \left(\frac{a}{a_1}\right)^2\right]$$

$$(8) \quad E_r' = E_r \left[1 - \text{sign}(a) \left(\frac{a}{a_2}\right)^2\right],$$

$$\text{където } E_q = 2w_c k_y v_a l_d Ak_y y; \quad y = \frac{m_0}{1 - a_p}.$$

Резултантно е.д.н. в комутиращата секция при изместени четки

$$(9) \quad E_{rez}' = E_q' + E_r'$$

Определяне на оптималния за благоприятна комутация ъгъл α_{opt} .

Оптималният ъгъл α_{opt} е този, при който се анулират не E_q или E_r поотделно, а този при който сумата от двете е.д.н. е равна на нула, което теоретично би осигурило линейна комутация. От $E_q' + E_r' = 0$ и от ур. (7) и (8) следва, че

$$(10) \quad a_{opt} = a_1 a_2 \sqrt{\frac{x + y}{x a_1^2 + y a_2^2}}.$$

Поради това, че $\xi > \psi$, получава се, че $\alpha_1 < \alpha_{opt} < \alpha_2$, като α_{opt} е поблизък до α_2 , отколкото до α_1 .

Заб. При еднофазните колекторни двигатели за променлив ток трябва да се има пред вид, че различните е.д.н. във формулите, както и л.т.н. A , трябва да бъдат ефективни стойности, а магнитните индукции и

м.д.н. - амплитудни стойности, различаващи се от ефективните с $\sqrt{2}$. Освен това към E_q и E_r трябва да се прибави и трансформаторното е.д.н. E_t , дефазирано от сумата на първите две приблизително на 90° . Тъй като тези двигатели са с последователно възбуждане и токът във възбудителната и в котвената намотка е един и същ, отношенията на м.д.н. F_a и F_{rez} за повечето двигатели са доста близки, откъдето следва, че ъглите α_1 , α_2 и α_{opt} се движат в едни същи граници (между 20° и 30° el.).

3. Числен пример

За илюстрация в табл.1 са дадени получените е.д.н. и ъглите на изместване на четките (респ. свързването) за няколко еднофазни променливотокови колекторни двигатели за ръчни електроинструменти, изчислени по описаната методика.

Таблица 1

$P_N[W]/U_N[V]$	α / α_m [°]	$\alpha_1 / \alpha_{opt} / \alpha_2$ [°]	E_r / E_r' [V]	E_q / E_q' [V]
650/230	0/30	19,6/23,8/26,3	11,9/11,9	4,4/4,4
700/230	7,5/30	17,1/21,3/24,1	9,3/8,4	4,3/2,8
1300/230	15/32,5	20/24,9/27,6	11,9/8,4	5,2/1,7
1000/110	20/30	19,6/24/26,7	3,7/1,6	1,8/-0,1
1600/230	22,5/32,5	20/24,9/27,7	13,2/4,5	5,7/-1,2
2100/230	25/30	18,7/23,1/26	8,1/0,6	3,8/-2,4
2300/110	35/30	22/26,1/28,4	10,3/-5,3	4,8/-5,8

Както се вижда от таблицата, при малки конструктивни ъгли на изместване ($0 < \alpha < 15^\circ$), намалението на E_q и E_r е малко, т.е. машината е недокомпенсирана и комутацията ще бъде забавена. При по-големите ъгли обаче ($20 < \alpha < 35^\circ$) E_q и E_r се намаляват значително и даже стават отрицателни, т.е. машината е прекомпенсирана и комутацията става ускорена, което е по-благоприятният случай.

ЛИТЕРАТУРА

- [1] Ермолин, Н. Расчет коллекторных машин малой мощности. Л., Энергия, 1973.
- [2] Попадиин, Ст. Электрически микромашины ч. II. С., Техника, 1970.
- [3] Ангелов, А., Д. Димитров. Электрически машины ч. I, II. С., Техника, 1988.
- [4] AEG. Eine Methode fur Berechnung der einphasen Kollektormotoren, 1979.
- [5] Meeker, D. Finite Element Method Magnetics, Ver.3.3, User's Manual, 2002.

- [6] Брандиски, К., И. Ячева. CAD системи в електромагнетизма. С., Сиела, 2002.
- [7] MATLAB – High performance numeric computation and visualization software – reference guide, The Mats works, 1992.
- [8] Sigma Plot 8.02 Demo, User's Guide, SPSS Inc. 2003.
- [9] Божилов, Г., А. Иванов. Нелинейна заместваща схема на еднофазен колекторен двигател, съчетана с метода на крайните елементи. Е+Е, No 7-8, 2004.
- [10] Bojilov, G., A. Ivanov, N. Ivanova. Experimental determination of non-linear parameters of a single-phase commutator motor, 11th Int. conf. of el. mach. ELMA 2005, Sofia, vol. 2.

Автори: Проф. д-р Ганчо Божилов, гл. ас. Адриан Иванов - кат. “Електрически машини”

Постъпила на 20.01.2010

Рецензент доц. д-р Михо Михов

АЛГОРИТЪМ ЗА НЕКОХЕРЕНТНА ОПТИМАЛНА ФИЛТРАЦИЯ НА СЛОЖНИ СИГНАЛИ ПО КРИТЕРИЙ ЗА ТОЧНОСТ

Вася Илиев, Михаил Момчеджиков

***Резюме.** Чрез комплексното използване на модифицираната нелинейна теория в приближение на Гаус и линейната теория на Фредхолм е синтезиран алгоритъм и приемник за некохерентна оптимална филтрация на сложни сигнали по критерий за точност. Изведени са изрази за ковариационните матрици за грешките във филтрацията на неизвестните параметри на сигнала. Изведени са също така формули за количествено определяне на шумоустойчивостта. Работоспособността на полученият алгоритъм се потвърждава от резултатите, получени вследствие на проведени натурални експерименти.*

ALGORITHM FOR NON-COHERENT OPTIMAL FILTRATION OF COMPLEX SIGNALS BY PRECISION CRITERION

Vasya Iliev, Mikhail Momchedjиков

***Abstract.** An algorithm and a receiver for non-coherent optimal filtration of complex signals by precision criterion are synthesized via complex usage of modified non-linear theory with Gaus approximation and linear theory of Fredholm. Equations are worked out of co-variation matrixes taking in account the errors in filtration of signal unknown parameters. Equations are worked out also for noise immunity quantitative determination. The obtained algorithm functionality is confirmed by the results from the carried natural experiments.*

1. Въведение

Кохерентните методи за обработка в най-пълна степен реализират преимуществата на оптималните методи за филтрация на сложни сигнали [1]. Наред с това използването им за обработка на DSSS сигнали с модифицирана псевдослучайна кодова поредица (ПСКП), както и при обработка на сложни сигнали с дискретна промяна на носещата честота (FHSS), е съпроводени с редица затруднения и невъзможност за постигане на желаните резултати. При свързка между подвижни обекти с флукутираща скорост, при разпространение на сигналите в среда, внасяща фазови изкривявания, използването на кохерентна обработка на основата на система за следене на фазата (ССФ) се оказва неефективно, вследствие на тясната лента на захващане и наличие на фазови прескачания при рязка промяна на честотата и фазата [2]. На практика в тези случаи, вместо кохерентна обработка на видео честота, се прилага

некохерентна обработка на сигнала на междинна честота, на основата на некохерентна система за следене на закъснението (ССЗ) и система за следене на честотата (ССЧ) [3].

На въпросите свързани със синтез на устройства за оптимална оценка на честотата са посветени редица статии и книги [1,4,7,11], но за съжаление в болшинството от тях се разглеждат въпросите свързани с некохерентна оптимална филтрация по критерий максимум на отношението сигнал/шум. В случаите, в които се разглеждат въпросите за некохерентна оптимална филтрация по критерий за точност, обикновено при решение на уравнението за оптимална нелинейна филтрация липсва честотен дискриминатор в явен вид [4].

В работата е синтезиран алгоритъм за оптимална некохерентна филтрация на сложни сигнали с неизвестна честота и закъснение по критерий за точност, в който е налице явно решение на уравнението за оптимална нелинейна филтрация относно ССЧ и ССЗ. Синтезът на алгоритъма е извършен съгласно метода за статистически синтез на системи за оптимална нелинейна филтрация предложен в [5].

2. Статистически синтез на алгоритъма

При положение, че входният сигнален процес $y(t)$ представлява адитивна смес от нормален шум и полезен сигнал,

$$(1) \quad y(t) = s[t, \bar{\lambda}(t)] + n(t),$$

където: $s[t, \bar{\lambda}(t)]$ - полезният сигнал, зависещ от времето и неизвестните параметрите $\lambda_{\alpha}(t)$, $n(t)$ - нормален гаусов шум с нулева средна стойност и делта корелационна функция, след заместването му в уравнението за оптимална нелинейна филтрация [7], то придобива следният аналитичен вид

$$(2) \quad \dot{W}_{ps}(t, \Delta\omega, \tau_0, \theta_0) = [F(t, \Delta\omega, \tau_0, \theta_0) - \bar{F}(t, \Delta\omega, \tau_0, \theta_0)] W_{ps}(t, \Delta\omega, \tau_0, \theta_0).$$

В (2) $F(t, \Delta\omega, \tau_0, \theta_0) = \frac{2A_c}{N_0} L_{s*}[y] \cdot g(t - \tau_0) \cos(\omega_0 t + \theta_0)$ - функционал на правдоподобие [3], $L_{s*}[y]$ - линеен интегрален оператор на Фредхолм, в случаите на филтрация по неенергийни параметри [6].

Апостериорната вероятностна плътност W_{ps} представлява функционал [8]

$$(3) \quad W_{ps}(t, \Delta\omega, \tau_0, \theta_0) = C \exp \left\{ -\frac{A_c^2 t}{N_0} \int_0^t g^2(t_1 - \tau_0) \cos^2(\omega_0 t_1 + \theta_0) dt_1 + \right. \\ \left. + \frac{2A_c}{N_0} \int_0^t g(t_1 - \tau_0) L_{s*}[y] \cos(\omega_0 t_1 + \theta_0) dt_1 \right\},$$

където $C = \exp \left\{ -\frac{1}{N_0} \int_0^T y^2(t) dt \right\}.$

В (3) ще бъдат извършени следните полагания:

$$(4) \quad \begin{cases} X(t, \Delta\omega, \tau_0) = \int_0^t L_{s^*}[y] \cdot g(t_1 - \tau_0) \cos \omega_0 t_1 dt_1, \\ Y(t, \Delta\omega, \tau_0) = \int_0^t L_{s^*}[y] \cdot g(t_1 - \tau_0) \sin \omega_0 t_1 dt_1, \end{cases}$$

$$(5) \quad f(t, \Delta\omega, \tau_0, \theta_0) = X(t, \Delta\omega, \tau_0) \cos \theta_0 - Y(t, \Delta\omega, \tau_0) \sin \theta_0.$$

За решение на нелинейното диференциално уравнение (2), двете му страни ще бъдат усреднени по θ_0 . За случаите на филтрация по неенергийни параметри това се свежда към усредняването на израза

$$(6) \quad \begin{aligned} F(t, \Delta\omega, \tau_0, \theta_0) W_{ps}(t, \Delta\omega, \tau_0, \theta_0) &= \frac{2A_c C}{N_0} L_{s^*}[y] \cdot g(t - \tau_0) \cos(\omega_0 t + \theta_0) \times \\ &\times \exp \left\{ -\frac{A_c^2 t}{2N_0} + \frac{2A_c}{N_0} \int_0^t L_{s^*}[y] g(t_1 - \tau_0) \cos(\omega_0 t_1 + \theta_0) dt_1 \right\}, \end{aligned}$$

който в резултат на усредняването се привежда във вида

$$(7) \quad \begin{aligned} F(t, \Delta\omega, \tau_0, \theta_0) W_{ps}(t, \Delta\omega, \tau_0, \theta_0) &= \\ &= \frac{2A_c C}{N_0} \exp \left\{ -\frac{A_c^2 t}{2N_0} + \frac{2A_c}{N_0} f(t, \Delta\omega, \tau_0, \theta_0) \right\}. \end{aligned}$$

Чрез усредняването на (6) по θ_0 , при условие на равномерно разпределение на фазата на интервала $[0, 2\pi]$, т.е. $W(\varphi) = 0,5/\pi$, се получава

$$(8) \quad \begin{aligned} \frac{1}{2\pi} \int_0^{2\pi} F(t, \Delta\omega, \tau_0, \theta_0) W_{ps}(t, \Delta\omega, \tau_0, \theta_0) d\theta &= \\ &= \frac{C}{2\pi} \exp \left\{ -\frac{A_c^2 t}{2N_0} \right\} \frac{\partial}{\partial t} \left[\int_0^{2\pi} \exp \left\{ \frac{2A_c}{N_0} f(t, \Delta\omega, \tau_0, \theta_0) \right\} d\theta \right] = \\ &= C \exp \left\{ -\frac{A_c^2 t}{2N_0} \right\} \frac{\partial}{\partial t} \left[I_0 \left\{ \frac{2A_c}{N_0} Z \right\} \right], \end{aligned}$$

където $Z^2 = X^2 + Y^2$; I_0 – функция на Бесел.

Тъй като $W_{ps}(t, \Delta\omega, \tau_0) = C \exp\left\{-\frac{A_c^2 t}{2N_0}\right\} I_0\left\{\frac{2A_c}{N_0} Z\right\}$, а отношението

$$\frac{\partial}{\partial t} \left[I_0\left\{\frac{2A_c}{N_0} Z\right\} \right] / I_0\left\{\frac{2A_c}{N_0} Z\right\} = \frac{\partial}{\partial t} \ln I_0\left\{\frac{2A_c}{N_0} Z\right\}, \text{ то}$$

$$(9) \quad \dot{W}_{ps}(t, \Delta\omega, \tau_0) = C \exp\left\{-\frac{A_c^2 t}{2N_0}\right\} \frac{\partial}{\partial t} I_0\left\{\frac{2A_c}{N_0} Z\right\} - \overline{F(t, \Delta\omega, \tau_0, \theta_0)} \cdot W_{ps}(t, \Delta\omega, \tau_0).$$

Като се изнесе $W_{ps}(t, \Delta\omega, \tau_0)$ пред скобите се получава окончателното уравнение за съвместната апостериорна вероятностна плътност за честотната разстройка $\Delta\omega$ и закъснението τ_0

$$(10) \quad \dot{W}_{ps}(t, \Delta\omega, \tau_0) = \left\{ \frac{\partial}{\partial t} \left[\ln I_0\left\{\frac{2A_c}{N_0} Z(t, \Delta\omega, \tau_0)\right\} \right] - \overline{F(t, \Delta\omega, \tau_0, \theta_0)} \right\} \cdot W_{ps}(t, \Delta\omega, \tau_0).$$

След прилагане на Гаусово приближение [9] в (2) и заместване с (10) се получава система нелинейни диференциални уравнения за оценка на честотата, закъснението и за вторите централни моменти на тези величини:

$$(11) \quad \begin{aligned} \dot{\omega}^* &= K_{\omega\omega} \frac{\partial}{\partial \omega} \frac{\partial}{\partial t} \left\{ \ln I_0\left[\frac{2A_c}{N_0} Z(t, \omega^*, \tau^*)\right] \right\} + \\ &+ K_{\omega\tau} \frac{\partial}{\partial \tau} \frac{\partial}{\partial t} \left\{ \ln I_0\left[\frac{2A_c}{N_0} Z(t, \omega^*, \tau^*)\right] \right\}, \end{aligned}$$

$$(12) \quad \begin{aligned} \dot{\tau}^* &= K_{\tau\tau} \frac{\partial}{\partial \tau} \frac{\partial}{\partial t} \left\{ \ln I_0\left[\frac{2A_c}{N_0} Z(t, \omega^*, \tau^*)\right] \right\} + \\ &+ K_{\omega\tau} \frac{\partial}{\partial \omega} \frac{\partial}{\partial t} \left\{ \ln I_0\left[\frac{2A_c}{N_0} Z(t, \omega^*, \tau^*)\right] \right\}, \end{aligned}$$

$$(13) \quad \frac{dK}{dt} = -KDK.$$

За удобство ще бъдат въведени означенията:

$$\begin{aligned}
F_1(t, \Delta w, t_0) &= \frac{\partial}{\partial t} \ln I_0 \left[\frac{2A_c}{N_0} Z(t, w^*, t^*) \right]; \\
(14) \quad D &= \begin{vmatrix} F_{2ww} & F_{2wt} \\ F_{2tw} & F_{2tt} \end{vmatrix}; \quad F_{2ij} = \left\| \frac{\partial^2 F_1}{\partial x_i \partial x_j} \right\|; \quad K = \begin{vmatrix} K_{ww} & K_{wt} \\ K_{tw} & K_{tt} \end{vmatrix}; \\
x_1 &= \Delta w; \quad x_2 = t_0; \quad \mathbf{x}^T = |x_1, x_2| = |\Delta w, t_0|.
\end{aligned}$$

В скаларен вид уравненията за централните корелационни моменти имат следния вид

$$\begin{aligned}
K_{ww}^2 F_{ww} + 2K_{ww}K_{wt} + K_{wt}^2 F_{tt} &= -\frac{dK_{ww}}{dt}, \\
(15) \quad K_{ww}K_{wt}F_{ww} + K_{wt}^2 F_{wt} + K_{ww}K_{tt}F_{wt} + K_{wt}K_{tt}F_{tt} &= -\frac{dK_{wt}}{dt}, \\
K_{wt}^2 F_{ww} + 2K_{wt}K_{tt}F_{wt} + K_{tt}^2 F_{tt} &= -\frac{dK_{tt}}{dt},
\end{aligned}$$

където :

$$\begin{aligned}
F_{\tau\tau} &= \frac{\partial^2 F_1}{\partial \tau^2} = \frac{\partial}{\partial t} \left[\frac{\partial B(Z)}{\partial Z} \left(\frac{\partial Z}{\partial \tau} \right)^2 + B(Z) \frac{\partial^2 Z}{\partial \tau^2} \right], \\
F_{\omega\omega} &= \frac{\partial^2 F_1}{\partial \omega^2} = \frac{\partial}{\partial t} \left[\frac{\partial B(Z)}{\partial Z} \left(\frac{\partial Z}{\partial \omega} \right)^2 + B(Z) \frac{\partial^2 Z}{\partial \omega^2} \right], \\
F_{\omega\tau} &= \frac{\partial^2 F_1}{\partial \omega \partial \tau} = \frac{\partial}{\partial t} \left[\frac{\partial B(Z)}{\partial Z} \frac{\partial Z}{\partial \omega} \frac{\partial Z}{\partial \tau} + B(Z) \frac{\partial^2 Z}{\partial \omega \partial \tau} \right], \\
B(Z) &= \frac{\partial}{\partial Z} \left\{ \ln I_0 \left[\frac{2A_c}{N_0} Z(t, \omega, \tau) \right] \right\}.
\end{aligned}$$

Решението на уравнение (13) за централните корелационни моменти има следния вид

$$(16) \quad K^{-1} = C''t + C'.$$

Решението на тези уравнения за стационарен случай дава нулеви стойности на коефициентите на усилване $K_{ij} = 0$ за $i \neq j$. Това означава, че отпада необходимостта от кръстосана връзка $K_{\omega\tau}$ и $K_{\tau\omega}$ между системите за следене по честота и закъснението.

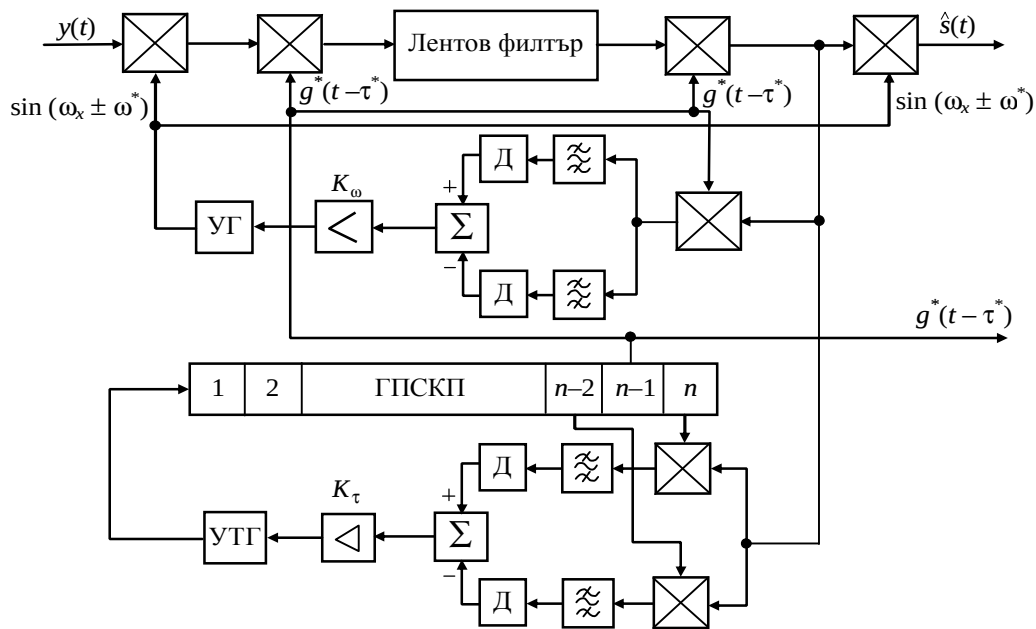
Средните стойности на дисперсията на грешките в следенето на честотата и времезакъснението, съгласно [10], е

$$(17) \quad \sigma_{\omega}^2 = K_{\omega\omega}^* \approx \frac{12N_0}{A_c^2 r_s(\tau) t^2}, \quad \sigma_{\tau}^2 = K_{\tau\tau}^* \approx \frac{12N_0}{A_c^2 r_s(\tau) t^2}.$$

Чрез преминаване от уравнения с частни производни към уравнения с крайните разлики [4] може да се запише

$$(18) \quad \begin{aligned} \omega^* &= \frac{K_{\omega\omega}}{2a} \left\{ \ln I_0 \left[\frac{2A_c}{N_0} Z(t, \omega^* + \Delta\omega, \tau^*) \right] - \ln I_0 \left[\frac{2A_c}{N_0} Z(t, \omega^* - \Delta\omega, \tau^*) \right] \right\}, \\ \tau^* &= \frac{K_{\tau\tau}}{2b} \left\{ \ln I_0 \left[\frac{2A_c}{N_0} Z(t, \omega^*, \tau^* - \tau_0) \right] - \ln I_0 \left[\frac{2A_c}{N_0} Z(t, \omega^*, \tau^* + \tau_0) \right] \right\}. \end{aligned}$$

Структурата на устройството, реализиращо алгоритъма (18), е показана на фиг. 1. Този корелатор (както и кохерентните) съдържа две подсистеми за следене, съответно за честотата и времезакъснението. В отличие от тях, оценката на честотата се осъществява от честотен дискриминатор за следене на честотата, а ССЗ съдържа некохерентен дискриминатор. Такова построяване на приемника осигурява устойчива работа при големи фазови и честотни флуктуации, а също така съществено се отразява на характера на нелинейната динамика на приемника, облекчавайки решаването на задачата за влизане в режим на синхронизация.



УГ – управляем генератор
Д – амплитуден детектор
ГПСКП – генератор на псевдослучайна кодова поредица
УТГ – управляем тактов генератор

Фиг. 1. Функционална схема на устройство, реализиращо алгоритъма за некохерентна оптимална филтрация на СШПС по критерий минимум на СКГ

3. Шумоустойчивост на алгоритъма

За определяне на шумоустойчивостта на алгоритъма е приложена методиката, използвана в [11]. При определено ниво на смущаващите сигнали на входа, шумоустойчивостта количествено може да се определи с коефициента q^2 , определящ отношението сигнал-шум в изхода на корелатора

$$(19) \quad q^2 = \frac{P_{c_{\hat{e}\hat{\zeta}\hat{\delta}}}}{P_{c_{\hat{i}\hat{e}\hat{\zeta}\hat{\delta}}}} = \frac{\hat{s}_r^2(t)}{\hat{s}_{ci}^2(t)}$$

където $P_{c_{\text{изх}}}$ и $P_{c_{\text{мизх}}}$ са мощностите на полезния и смущаващия сигнал в изхода на корелатора; $\hat{s}_{\tilde{n}i}(t) = \hat{n}(t) + \hat{x}(t)$ - сумарният смущаващ сигнал в изхода на корелатора; $n(t)$ - нормален бял шум; $x(t)$ - структурно смущение. На практика обикновено $x(t) \gg n(t)$, вследствие на което $\hat{s}_{\tilde{n}i}(t) \cong \hat{x}(t)$. Определянето на $\hat{s}_r^2(t)$ и $\hat{x}^2(t)$ в случая ще бъде извършено с помощта на енергийните им спектри и спектрите на опорните сигнали $s_{r_0}(t)$ и $s_{r_0}(t-t_3)$.

Мощност на сигнала в изхода на корелатора

Полезният сигнал на изхода на корелатора, определен чрез спектрите на съответните сигнали, е

$$(20) \quad \hat{s}_r(t) = \frac{1}{4p} \int_{-\infty}^{+\infty} [S_r^{N_1}(j\omega) S_{r_0}^*(j\omega) e^{-j\omega t_\zeta}] K_{w_2}(j\omega) e^{j\omega t} d\omega,$$

където $K_{w_2}(j\omega)$ - честотна характеристика на широколентовия филтър в изхода, $S_{r_0}^*(j\omega)$ - спектр, комплексно спрегнат с този на опорния сигнал, т.е. $S_{r_0}^*(j\omega) = S_{r_0}(-j\omega)$, $S_r^{(N_1)}(j\omega) = \frac{1}{2} \{S_{r_0}(j\omega) [K_{w_1}(j\omega) S_r^*(j\omega)]\} K_{N_1}(j\omega)$ - спектр на сигнала в изхода на лентовия филтър, $K_{N_1}(j\omega)$ - честотна характеристика на лентовия филтър.

След заместване на $S_r^{(N_1)}(j\omega)$ в (20), се получава

$$(21) \quad \hat{s}_r(t) = \frac{1}{8p} \int_{-\infty}^{+\infty} [S_{r_0}^*(j\omega) S_{r_0}(j\omega) e^{-j\omega t_\zeta}] K_{w_2}(j\omega) e^{j\omega t} d\omega,$$

където $S_{rr_0}(j\omega) = [S_r^*(j\omega) S_{r_0}(j\omega)] K_{N_1}(j\omega)$ е взаимният енергиен спектр на полезния и опорния сигнал в изхода на лентовия филтър, $S_r^*(j\omega) = S_r(-j\omega)$ - спектр, комплексно спрегнат със спектъра на полезния сигнал. При това положение

$$(22) \quad \hat{s}_r^2(t) = P_{r_0} \frac{1}{8p} \int_{-\infty}^{+\infty} G_r(w) G_{r_0}(w) g_{rr_0}^2(w) |K_{w_2}(jw)|^2 dw,$$

където $P_{r_0} = \frac{1}{2p} \int_{-\infty}^{+\infty} G_{r_0}(w) dw$ - мощност на опорния сигнал $s_{r_0}(t)$, $G_r(w), G_{r_0}(w)$ - енергийни спектри на полезния и опорния сигнали, $g_{rr_0}^2(w) = \frac{|S_{rr_0}(jw)|^2}{G_r(w) G_{r_0}(w)}$ - функция на взаимната честотна кохерентност на полезния и опорния сигнали.

Пресмятането на интеграла в (22) се извършва, съгласно [8], като решението му в конкретния случай е

$$\frac{1}{2p} \int_{-\infty}^{+\infty} G_r(w) G_{r_0}(w) g_{rr_0}^2(w) |K_{w_2}(jw)|^2 dw = P_{r_0} P_r R_{rr_0}(t),$$

където P_r - мощността на полезния сигнал на входа на корелатора, $R_{rr_0}(t)$ - коефициент на взаимна корелация между полезния и опорния сигнали. След заместване на решението на интеграла в (22) с неговата стойност се получава

$$(23) \quad \hat{s}_r^2(t) = P_{r_0}^2 P_r R_{rr_0}^2(t)$$

Мощност на смущението в изхода на корелатора

Определянето на мощността на смущаващия сигнал в изхода на корелатора се извършва по аналогичен начин

$$(24) \quad \hat{x}(t) = \frac{1}{4p} \int_{-\infty}^{+\infty} [S_x^{(N_1)}(jw) S_{r_0}^*(jw) e^{-jw t_\varphi}] K_{w_2}(jw) e^{jw t} dw, \quad \text{където}$$

$S_x^{(N_1)}(jw) = \frac{1}{2} \{S_{r_0}(jw) [K_{w_1}(jw) S_x^*(jw)]\} K_{N_1}(jw)$ - спектър на смущаващия сигнал в изхода на лентовия филтър, $S_x^*(jw) = S_x(-jw)$ - спектър, комплексно спрегнат със спектъра на смущаващия сигнал $x(t)$. След заместване на $S_x^{(N_1)}(jw)$ в (24) се получава

$$(25) \quad \hat{x}(t) = \frac{1}{8p} \int_{-\infty}^{+\infty} [S_{r_0}^*(jw) S_{xr_0}(jw) e^{-jw t_\varphi}] K_{w_2}(jw) e^{jw t} dw,$$

където $S_{xr_0}(jw) = [S_x^*(jw) S_{r_0}(jw) e^{-jw t_\varphi}] K_{w_1}(jw)$ - взаимен енергиен спектър на сигналите $x(t)$ и $s_{r_0}(t)$ в изхода на лентовия филтър, $S_x^*(jw) = S_x(-jw)$ - спектър, комплексно спрегнат със спектъра на смущаващия сигнал.

$$(26) \quad \hat{x}^2(t) = P_{r_0} \frac{1}{8p} \int_{-\infty}^{+\infty} G_x(w) G_{r_0}(w) g_{xr_0}^2(w) |K_{w_2}(jw)|^2 dw,$$

където $G_x(w)$ - енергиен спектър на смущаващия сигнал, $g_{xr_0}^2(w) = \frac{|S_{xr_0}(jw)|^2}{G_x(w)G_{r_0}(w)}$ -

функция на взаимната честотна кохерентност между опорния и смущаващия сигнал. Решението на интеграла в (26) е

$$\frac{1}{2p} \int_{-\infty}^{+\infty} G_x(w) G_{r_0}(w) g_{xr_0}^2(w) |K_{w_2}(jw)|^2 dw = P_{r_0} P_x R_{xr_0}(t),$$

където P_x - мощността на смущаващия сигнал на входа на корелатора, $R_{xr_0}(t)$ - коефициент на взаимната корелация между опорния и смущаващия сигнал. След заместването му в (26) се получава

$$(27) \quad \hat{x}^2(t) = P_{r_0}^2 P_x R_{xr_0}^2(t).$$

Отношение сигнал-шум в изхода на корелатора

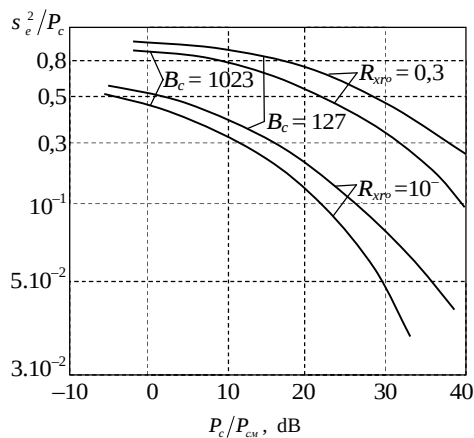
След заместване на (27) и (23) в (19), за отношението сигнал-шум q^2 в изхода на корелатора, в конкретния случай се получава

$$(28) \quad q^2 = \frac{\hat{s}_r^2(t)}{\hat{x}^2(t)} = \frac{P_{r_0}^2 P_r R_{rr_0}^2(t)}{P_{r_0}^2 P_x R_{xr_0}^2(t)} = r \cdot K_{\emptyset},$$

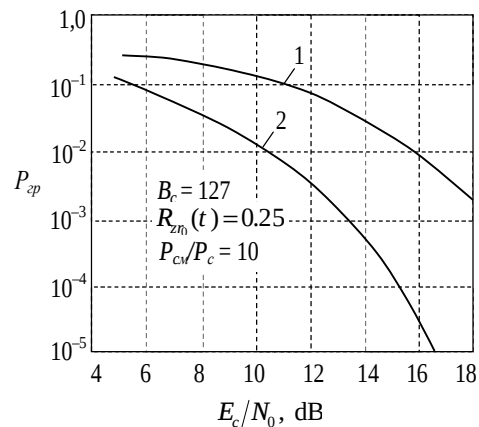
където $r = P_r / P_x$ - отношение сигнал-шум на входа на корелатора, $K_{\emptyset} = R_{rr_0}^2(t) / R_{xr_0}^2(t)$ - коефициент на шумоустойчивост.

На фиг. 2 е показана зависимостта на нормираната грешката във филтрацията ϵ от отношението сигнал/смущение на входа за стойности на $R_{xr_0}(t)$, равни на 0,3 и 0,1 при обработка с DSSS сигнали с бази $B_c = 127$, $B_c = 1024$ и отношение сигнал/шум на входа, равно на 16 dB, снети експериментално.

На фиг. 3 е показана графичната зависимост на вероятността за грешно приет информационен символ P_{zp} при обработка на DSSS сигнал с база $B_c = 127$ и коефициент на взаимна корелация $R_{xr_0}(t) = 0,25$. От нея се вижда, че в резултат на предварителната филтрация съгласно предложения метод вероятността за грешка силно намалява, което потвърждава шумопотискащите свойства на предлагания метод.



Фиг. 2. Зависимост на грешката във филтрацията е от отношението сигнал/смущение за различни стойности на ВКФ при обработка на DSSS сигнали за стойност на отношението сигнал/шум равно на 16 dB



Фиг. 3. Вероятност за грешка при обработка на DSSS сигнал във функция от отношението E_c/N_0 . Графика 1 – СФ, 2 – СФ плюс предварителна оптимална филтрация съгласно предложения метод

4. Заключение

Предлаганият метод за корелационна обработка позволява да бъде реализиран алгоритъм за многократна корелационна обработка, чрез последователно свързване на определен брой устройства от този вид, с обща система за синхронизация. Теоретично това би довело до получаването на безкрайно голяма стойност на коефициента на печалбата от обработката, т.е. $P_G = \sum_{i=1}^N P_{G_i}$,

където P_{G_i} - печалба от обработката в отделните устройства. На практика, поради действието на редица ограничаващи и дестабилизиращи фактори [12], по този начин е възможно реализирането на двукратна и трикратна корелационна обработка.

На основата на предложения метод е реализиран съгласуван филтър за предварителна филтрация на ШПС към съществуващ за тази цел приемник, с което е реализиран режим на двукратна корелационна обработка и е доказана практическата реализуемост на предлагания метод. Той успешно би могъл да бъде използван за шумоустойчиво предаване на информация в канал с ограничена пропускателна способност.

Качеството на филтрация и шумоустойчивостта на синтезираният алгоритъм в сравнение със случаите на кохерентна оптимална филтрация до голяма степен се определя от филтрите, включени в изходите на умножителите на ССЧ и ССЗ, осигуряващи филтрация на сместа от опорния сигнал и входния сигнален процес [11]. Тези филтри осигуряват нарастване на отношението сигнал/шум на входовете на амплитудните детектори. Лентата на пропускане на тези филтри, ако се изходи от условието за устойчивост на системата, може да

бъде значително по малка в сравнение със случаите на кохерентна оптимална филтрация с използване на (ССФ).

ЛИТЕРАТУРА

- [1] Ipatov, V. Spread Spectrum and CDMA: Principles and Applications. John Wiley & Sons, INC. New York, 2005.
- [2] Fazel, K., S. Kaiser. Multi-Carrier and Spread Spectrum Systems. John Wiley & Sons. INC. New York, 2008.
- [3] Torrieri, D. Principles of Spread-Spectrum Communication System. Springer, First Editions, New York, 2009.
- [4] Rodger, E., L. Peterson. Digital Communications and Spread Spectrum Systems. Macmillan Pub. Co, Second Editions, 2006.
- [5]. Илиев, В. Разработване на корелационен метод за оптимална филтрация на сложни сигнали по критерий за точност. Дисертационен труд за придобиване на образователна и научна степен ДОКТОР, ТУ-София, 2009г.
- [6] Berg, O., S. Haavik, R. Skaug. Spread Spectrum in Mobile Communication. IEE, 1998.
- [7] Dixon R. Spread Spectrum Systems with Commercial Applications. John Wiley & Sons, INC. Third Edition, New York, 1994.
- [8] Bendat, J., A. Piersol Engineering Applications of Correlation and Spectral Analysis. John Wiley & Sons, INC., Third Edition, New York, 1980.
- [9] Walker, J. Detectability of Spread-Spectrum Signals. Artech House, Sec. Editions, 2007.
- [10] Naguib, A. F., A. Paulraj. Capacity Improvement with Base-Station Antenna Arrays in Cellular CDMA, IEEE Transaction, 2008.
- [11] Борисов, В., А. Зинчук, Н. Лимарев, П. Мухин. Помехозащищенность систем радиосвязи с расширением спектра сигналов модуляцией несущей псевдослучайной последовательностью. М. Радио и связь, 2003.
- [12] Урядников, Ю., С. Аджемов, Сверхширокополосная связь, СОЛОН-Пресс, М, 2005.

Автори: Вася Илиев, н. с. I ст. д-р от Институт за Специална Техника - София; Михаил Момчеджиков, проф. д-р – катедра „Радиокомуникации и видеотехнологии”

Постъпила на 05.02.2010

Рецензент проф. д-р Добри Добрев

ИЗСЛЕДВАНЕ НА ГРЕШКИТЕ В СЛЕДЕНЕ НА ФАЗАТА И ЗАКЪСНЕНИЕТО НА АЛГОРИТЪМ ЗА ОПТИМАЛНА ФИЛТРАЦИЯ НА СЛОЖНИ СИГНАЛИ

Вася Илиев

Резюме. В работата е направено числено изследване на грешките в следене на фазата и закъснението на алгоритъм за оптимална филтрация на сложни DSSS сигнали по критерий за точност. Извършено е пробразуване на двумерната нелинейна следяща система в линейна, чрез структурна трансформация по метода на статистическото линеизиране. Изведен е математичен модел на трансформираната линейна система и е извършено математическо моделиране. Резултатите от изследването са представени в графичен вид.

ANALYSIS OF ERRORS IN THE PHASE AND DELAY TRACKING OF ALGORITHM FOR COMPLEX SIGNALS OPTIMAL FILTRATION

Vasya Iliev

Abstract. In the work a numerical analysis of errors in phase and delay tracking in algorithm for complex DSSS signals optimal filtration by precision criterion is made. A transformation of two-dimensional non-linear tracking system in linear is carried via structural transforming by method of statistic linearization. A mathematical model of transformed linear system and mathematical modeling is carried out. The analysis results are shown in graphical view.

1. Въведение

В [1] е синтезирана двумерна корелационно екстремална следяща система за оптимална филтрация на DSSS сигнали с неизвестна фаза и закъснение по критерий за точност. В [2] е извършено математическо моделиране на системата по метода на фазовото пространство, в резултат на което е доказана физическата реализуемост и функционалната и устойчивост. Трябва да се отбележи обаче, че синтезираната система е съществено нелинейна и направеният анализ на грешките в следене на фазата и закъснението в [2] е при съблюдаване на редица ограничителни условия (голямо отношение сигнал/шум, нормален закон на разпределение на грешката, проста конфигурация на ниско пропускащите филтри във веригите на обратните връзки и др.). Анализирани са грешките в следене на фазата θ_0 и закъснението τ_0 в голям обхват на изменение на отношението сигнал/шум, плътно до праговите стойности и при

произволни видове НЧФ във веригите на обратните връзки, е възможно с помощта на линейната теория чрез статистическото линеизиране [3].

При прилагането на този метод нелинейните звена на системите за следене на фазата и закъснението (ССФ, ССЗ) се заменят със статистически еквивалентни линейни звена със същия коефициент на усилване. Връзката между коефициентите на усилване на линейните и нелинейните звена в този случай се задава със съотношението

$$(1) \quad K_{E_j} = \int_{-\infty}^{\infty} K_j(x) W(x) dx,$$

където: $W(x)$ е вероятността плътност на сигнала на грешката; $K_j(x)$ – коефициент на усилване на j -тото нелинейно звено.

2. Преобразуване на оптимална нелинейна система в линейна

Преобразуването ще бъде осъществено чрез структурна трансформация на нелинейната система в двумерна линейна система за следене на фазата и закъснението в стационарен режим.

За целта, системата нелинейни диференциални уравнения (13) получени в [1] ще бъде записана чрез предварително отделяне от процеса $y(t)$ на съставящите на сигнала и шума във вида [4]

$$(2) \quad \begin{cases} \dot{\theta} = -\theta_0 - \frac{1}{p} \frac{K_{\theta} F_{\theta}(p)}{A_r} [A_s R_s^2(\tau) \sin \theta + n_{\theta}(t)] \\ \dot{\tau} = -\tau_0 - \frac{1}{p} \frac{K_{\tau} F_{\tau}(p)}{A_r} [A_s R_s(\tau) D(\tau) \cos \theta + \tau_u n_{\tau}(t)], \end{cases}$$

където: $R_s(\tau) = g(t - \tau^*) \cdot g(t - \tau_0)$ е автокорелационна функция (АКФ) на DSSS сигнал; $D(\tau) = g(t - \tau_0) [g(t - \tau^* + \tau_u) - g(t - \tau^* - \tau_u)]$ – дискриминационната характеристика на системата за следене на закъснението (ССЗ); $n_{\theta}(t)$ и $n_{\tau}(t)$ – съставящи на параметричния шум в системата за следене на фазата (ССФ) и ССЗ, получени в резултат на преобразуване на входния шум.

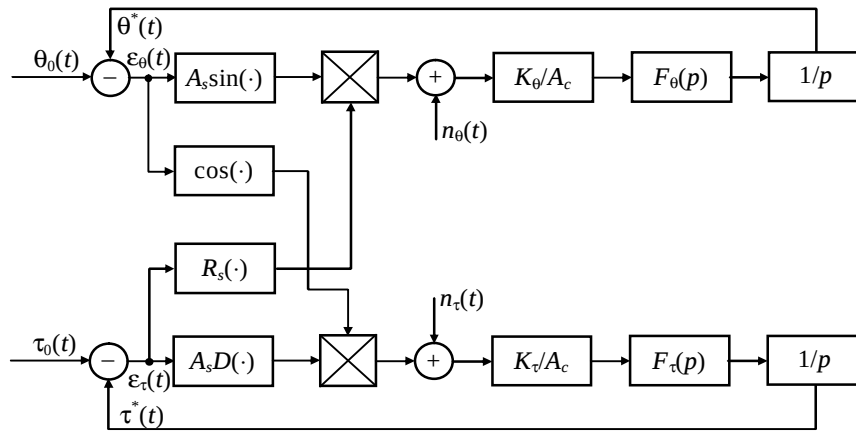
Ако се предположим, че входният шум има симетричен спектър около носещата честота на сигнала ω_0 , той може да бъде представен във вида [5]

$$(3) \quad n(t) = X_c(t) \cos(\omega_0 t + \theta_0) + X_s(t) \sin(\omega_0 t + \theta_0).$$

Чрез отделяне на високочестотните съставящи на шума, за ниско-честотните (параметричните) съставящи на шума се получават

$$(4) \quad n_{\theta}(t) = (1/2) [X_c(t) \cos \theta + X_s(t) \sin \theta],$$

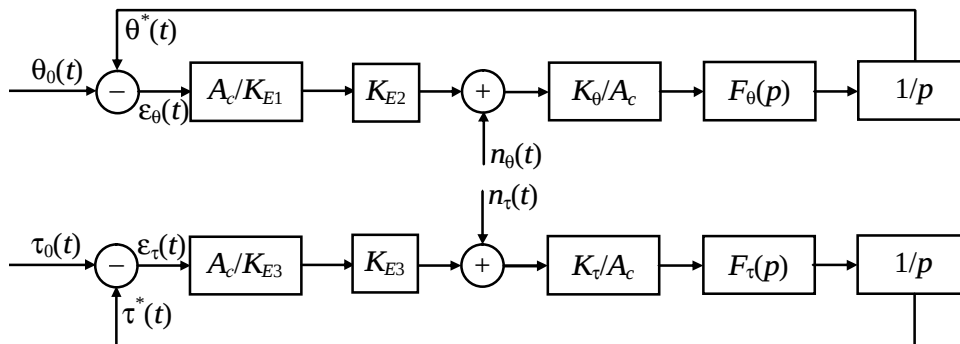
$$(5) \quad n_{\tau}(t) = \tau_u [X_c(t) \cos \theta + X_s(t) \sin \theta] \times [g(t - \tau^* + \tau_u) - g(t - \tau^* - \tau_u)].$$



Фиг. 1. Заместваща схема на нелинейна система

На основата на системата уравнения (2) изходната функционалната схема [1] относно фазовата разлика $\theta_0(t)$ и закъснението $\tau_0(t)$, се преобразува във вида, показан на фиг. 1.

Ако в заместващата схема, на основата на статистическата линеизация, бъдат заменени нелинейните елементи $\sin \theta, R_s(t), D(t), \cos \theta$ с линейни елементи, с еквивалентни коефициенти на усилване $K_{E1}, K_{E2}, K_{E3}, K_{E4}$, съгласно (1), се получава структурната схема на линеизираната система за следене на параметрите $\theta_0(t), \tau_0(t)$ – фиг. 2.



Фиг. 2. Структура на оптимална линеизирана система за следене на фазата и закъснението

Предавателните функции на веригата на ССФ и ССЗ в линейно приближение могат да бъдат получени от уравненията

$$(6) \quad \begin{cases} Y_\theta(p) = \frac{\theta^*(t)}{\theta_0(t)} = \frac{K_\theta F_\theta(p)}{p + K_\theta F_\theta(p)} \\ Y_\tau(p) = \frac{\tau^*(t)}{\tau_0(t)} = \frac{K_\tau F_\tau(p)}{p + K_\tau F_\tau(p)} \end{cases}$$

От (6) се вижда, че линеизираните ССФ и ССЗ се описват с линейни уравнения с еднаква структура. За определянето на грешката в следенето на ССФ и ССЗ е необходимо да бъдат определени коефициентите K_{Ej} .

3. Определяне на грешката в следене на фазата

$$(7) \quad K_{E_1} = \int_{-\infty}^{\infty} \left(\frac{d}{d\theta} \sin \theta \right) W(\theta) d\theta - \int_{-\infty}^{\infty} \cos \theta \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp \left[-\frac{(\theta - \bar{\theta})^2}{2\sigma_\theta^2} \right] d\theta,$$

където: $\bar{\theta}$ е средната съставляща на сигнала на грешката в следенето, равна на $\arcsin \gamma_p$; $\gamma_p = \Delta\omega/K_1$, $\Delta\omega$ – честотна разстройка; σ_θ^2 – средно квадратична грешка вследствие шума.

Чрез интегриране на (7) се получава

$$(8) \quad K_{E_1} = \cos \bar{\theta} - \exp \left(-\frac{\sigma_\theta^2}{2} \right) = \sqrt{1 - \left[\frac{\Delta\omega}{K_1 \exp(-\sigma_\theta^2)} \right]^2} \exp \left(-\frac{\sigma_\theta^2}{2} \right).$$

По такъв начин коефициентът K_{E_1} зависи от началната разстройка $\gamma_p = \Delta\omega/K_1$ и шумовата грешка σ_θ^2 на системата.

Еквивалентният коефициент на усилване K_{E_2} на фазовия дискриминатор на ССФ може да се изрази по следния начин

$$(9) \quad K_{E_2} = \int_{-\infty}^{\infty} R_s(x) W(x) dx = \int_{-1}^1 [1 - |x|] \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp \left[-\frac{(x - \bar{x})^2}{2\sigma_x^2} \right] dx,$$

където $x = \tau/\tau_u$; σ_x^2 – средно квадратична грешка; $\bar{x}$ – статистическа грешка.

В случая, коефициентът K_{E_2} определя амплитудата на сигнала в изхода на фазовия дискриминатор, в зависимост от грешката в следенето на ССФ. Чрез осъществяване на интегрирането в (9) и извършване на съответните преобразувания се получава

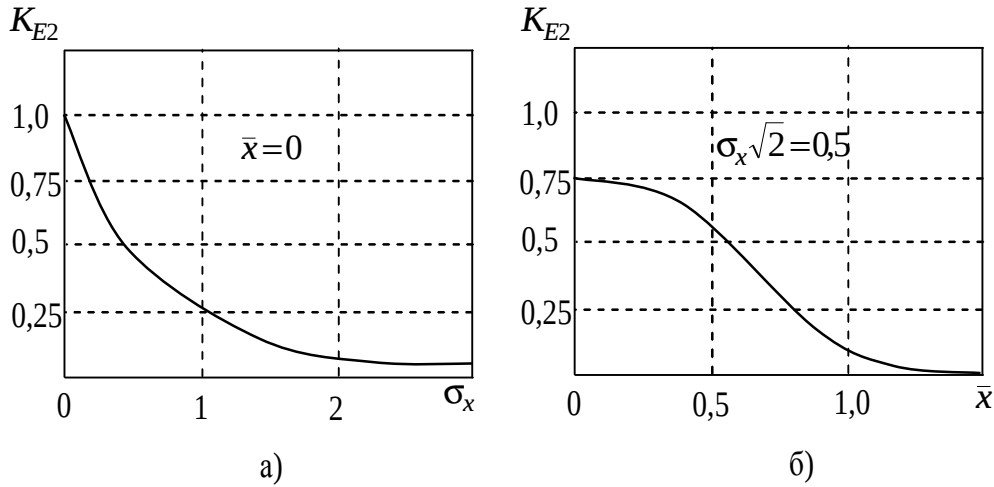
$$(10) \quad K_{E_2} = -\frac{\sigma_x}{(2\pi)^{1/2}} \left\{ 2 \exp \left[-\left(\frac{\bar{x}}{\sigma_x \sqrt{2}} \right)^2 \right] - \exp \left[-\left(\frac{1 + \bar{x}}{\sigma_x \sqrt{2}} \right)^2 \right] - \exp \left[-\left(\frac{1 - \bar{x}}{\sigma_x \sqrt{2}} \right)^2 \right] \right\} \\ - \frac{\bar{x}}{2} \left[\Phi \left(\frac{1 - \bar{x}}{\sigma_x \sqrt{2}} \right) - \Phi \left(\frac{1 + \bar{x}}{\sigma_x \sqrt{2}} \right) + 2\Phi \left(\frac{\bar{x}}{\sigma_x \sqrt{2}} \right) \right] + \frac{1}{2} \left[\Phi \left(\frac{1 - \bar{x}}{\sigma_x \sqrt{2}} \right) + \Phi \left(\frac{1 + \bar{x}}{\sigma_x \sqrt{2}} \right) \right],$$

където $\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp[-t^2] dt$ е интеграл на вероятностите.

В частен случай, когато $\bar{x} = 0$ за K_{E_2} се получава

$$(11) \quad K_{E_2} = \Phi \left(\frac{1}{\sigma_x \sqrt{2}} \right) - \frac{\sigma_x \sqrt{2}}{\sqrt{\pi}} \left[1 - \exp \left(-\frac{1}{2\sigma_x^2} \right) \right].$$

Зависимостта на K_{E_2} от средноквадратичната грешка σ_x и стойност на статистическата грешка $\bar{x}$ са показани на фиг. 3.



Фиг. 3. Зависимост на K_{E_2} от а) средноквадратичната грешка и б) статистическата грешка

Грешката от шума в изхода на веригата на ССФ ще бъде

$$(12) \quad \sigma_{\theta} = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_{\theta}'(\omega) |Y_{\phi}(j\omega)|^2 d\omega,$$

където $G_{\theta}'(\omega) = G_{\theta}(\omega) / (K_{E_1} K_{E_2})^2$ е спектралната плътност на мощността на флукуационните шумове с отчитане на влиянието на нелинейните елементи; $G_{\theta}(\omega) = A_c^2 / P_n \Delta f_{e\phi_n}$ – спектралната плътност на фазовите шумове; $Y_q(p) = K_{01} F_q(p) / [p + K_{01} F_q(p)]$ – предавателна функция на квазилинейната ССФ.

$$(13) \quad K_{01} = K_1 K_{E_1} K_{E_2}.$$

Ако се приеме, че спектралната плътност на фазовите шумове е равномерна, се получава

$$(14) \quad \sigma_{\theta}^2 = G_{\theta} \left| \frac{1}{K_{E_1} K_{E_2}} \right|^2 \Delta F'_{\text{шф}},$$

където

$$\Delta F'_{\text{шф}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} |Y_{\phi}(j\omega)|^2 d\omega,$$

е шумовата лента на ССФ с отчитане на лиенизирането. Чрез умножение и разделяне на формула (14) на шумовата лента $\Delta F'_{\text{шф}}$ на линейната ССФ, работеща с нулева начална разстройка и отчитайки, че произведението $\Delta F'_{\text{шф}} G_{\theta}$ участва в отношението сигнал/шум q в изхода на линейната система, се получава

$$(15) \quad \sigma_{\theta}^2 = \frac{1}{q} \left| \frac{1}{K_{E_1} K_{E_2}} \right|^2 \frac{\Delta F'_{\text{шф}}}{\Delta F_{\text{шф}}} = \frac{1}{q} \Pi_{\theta}.$$

В случая

$$(16) \quad \Pi_{\theta} = \left| \frac{1}{K_{E_1} K_{E_2}} \right|^2 \frac{\Delta F'_{\text{шф}}}{\Delta F_{\text{шф}}},$$

е коригиращ коефициент на грешката на линейната ССФ.

4. Определяне на грешката в следене на ССЗ

Еквивалентният коефициент на усилване на дискриминатора K_{E_3} може да бъде изразен по следния начин [6]

$$(17) \quad K_{E_3} = \int_{-2}^2 \Phi(x) \Big|_x \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left[-\frac{(x-\bar{x})^2}{2\sigma_x^2}\right] dx,$$

където

$$\Phi(x) \Big|_x = \begin{cases} 1 & \text{при } |x| \leq 1 \\ -1 & \text{при } 1 < |x| < 2. \end{cases}$$

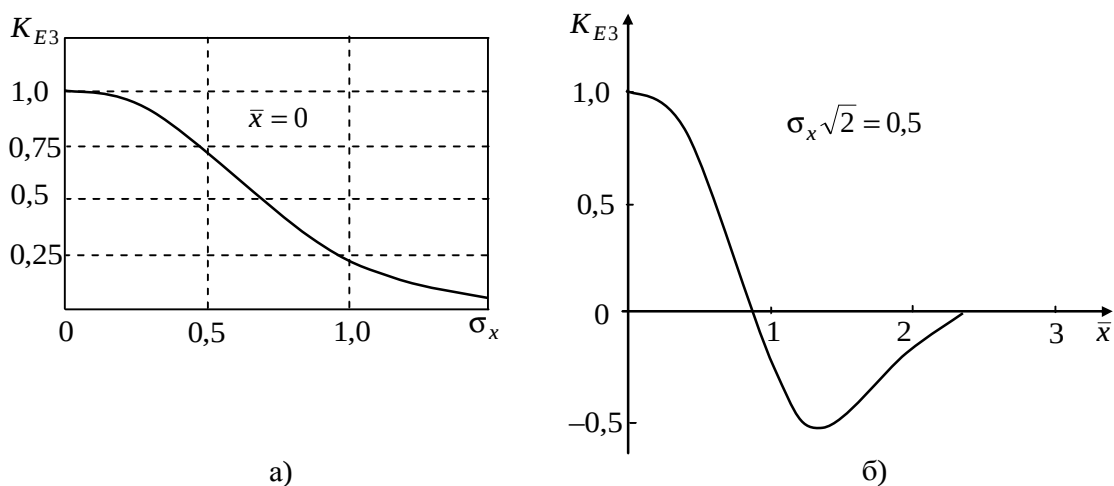
В резултат на интегрирането се получава

$$(18) \quad K_{E_3} = \frac{1}{2} \left[2\Phi\left(\frac{1+\bar{x}}{\sigma_x\sqrt{2}}\right) + \Phi\left(\frac{2+\bar{x}}{\sigma_x\sqrt{2}}\right) + 2\Phi\left(\frac{1-\bar{x}}{\sigma_x\sqrt{2}}\right) - \Phi\left(\frac{2-\bar{x}}{\sigma_x\sqrt{2}}\right) \right].$$

За частния случай при $\bar{x} = 0$ може да се запише

$$(19) \quad K_{E_3} = 2\Phi\left(\frac{1}{\sigma_x\sqrt{2}}\right) - \Phi\left(\frac{\sqrt{2}}{\sigma_x}\right).$$

На фиг. 4 е построена графичната зависимост на K_{E_3} във функция от средноквадратичната и статистическа грешки.



Фиг. 4. Зависимост на K_{E_3} от а) средноквадратичната грешка и б) статистическата грешка

От показаните графики следва, че K_{E_3} при $\sigma_x \leq 0,3$ и $\bar{x} = 0$ е практически равен на единица и значително намалява за $\sigma_x > 0,3$. За стойности на $\bar{x} > 1$, стойността на K_{E_3} може да бъде отрицателна, което свидетелства за неустойчивостта на системата при големи динамични грешки.

Грешка от шума на ССЗ по аналогия със ССФ може да бъде определено като

$$(20) \quad s_z^2 = \frac{1}{q_1} \left| \frac{1}{K_{E_3} K_{E_4}} \right|^2 \frac{\Delta F'_{\theta\zeta}}{\Delta F_{\theta\zeta}} = \frac{1}{q_1} \Pi_z,$$

където $\Delta F'_{uz}$ и ΔF_{uz} са съответно шумовата лента на ССЗ с отчитане лиенизацията и шумовата лента на линейната система; q_1 – отношението сигнал/шум на изхода на линейната ССЗ.

$$(21) \quad \Pi_z = \left| \frac{1}{K_{E_3} K_{E_4}} \right|^2 \frac{\Delta F'_{uz}}{\Delta F_{uz}}$$

е поправъчен коефициент в стойността на шумовата грешка на ССЗ, определена в линейно приближение.

Изрази (14) и (20) определят грешките в следенето на отделните подсистеми (ССФ и ССЗ). Записани в разгъната форма, те добиват следния вид

$$(22) \quad s_q^2 = \frac{1}{q} \frac{1}{|K_{E_1}(s_q, \bar{q}) K_{E_2}(s_x, \bar{x})|^2} \frac{\Delta F'_{\theta\delta}(s_q, s_x, \bar{q}, \bar{x})}{\Delta F_{\theta\delta}},$$

$$(23) \quad s_z^2 = \frac{1}{q_1} \frac{1}{|K_{E_3}(s_x, \bar{x}) K_{E_4}(s_q, \bar{q})|^2} \frac{\Delta F'_{\theta\zeta}(s_q, s_x, \bar{q}, \bar{x})}{\Delta F_{\theta\zeta}}.$$

5. Заключение

Въз основа на изрази (14), (20), (22) и (23), фиг. 3 и фиг. 4, грешките в следене на отделните подсистеми (ССЗ и ССФ) са взаимосвързани. Изменения в грешката на ССФ ще доведе до промяна в грешката на ССЗ и обратно – промяната в грешката на ССЗ ще доведе до промяна в работата на ССФ. Поради това, решението на системи уравнения (22) и (23) за количествено определяне на дисперсиите на грешките в следене е възможно с помощта на компютърно моделиране.

ЛИТЕРАТУРА

- [1] Илиев, В., М.Момчеджиков. Оптимална филтрация на DSSS сигнали. Годишника на Технически Университет – София. Том 58, кн. 1, 2008г.
- [2] Илиев, В., М. Момчеджиков, Изследване на преходните процеси и устойчивостта на алгоритъм за оптимална филтрация на DSSS сигнали. Годишник на Технически Университет – София. Том 58, кн. 1, 2008г.
- [3] Ipatov, V. Spread Spectrum and CDMA. Principles and Applications. John Wiley & Sons, INC., 2005.
- [4] Урядников, Ю., С. Аджемов. Сверхширокополосная связь, СОЛОН-Пресс, Москва, 2005.
- [5] Борисов, В., А. Зинчук, Н. Лимарев, П. Мухин. Помехозащищенность систем радиосвязи с расширением спектра сигналов модуляцией несущей псевдослучайной последовательностью. М. Радио и связь, 2003.

Автор: Вася Илиев, н. с. I ст. д-р от Институт за Специална Техника – София.

Постъпила на 05.02.2010

Рецензент проф. д-р Добри Добрев

РАЗМИТИ СИСТЕМИ ЗА ИЗВОД ЧРЕЗ СИСТЕМИ РАЗМИТИ РЕЛАЦИОНАЛНИ УРАВНЕНИЯ

Златко Захариев

***Резюме.** Представена е имплементация на размити системи за извод с използване на системи размити релационални уравнения. Разглеждат се правата и обратната задачи в размитите системи релационални уравнения. Даден е пример, както и няколко приложения на представената имплементация.*

FUZZY REASONING THROUGH FUZZY LINEAR SYSTEMS OF EQUATIONS

Zlatko Zahariev

***Abstract.** Application of fuzzy linear systems of equations (FLSE) in fuzzy inference module is presented here. It is a technique to represent fuzzy reasoning with FLSE. Direct and inverse problems are studied and an example is given. Applications of this approach are also described.*

1. Introduction

A technique to represent Mamdani type fuzzy inference systems (FIS) with fuzzy linear systems of equations (FLSE) is presented here. Some limitations of this technique are also described as well as possible methods to handle them.

The paper is divided in seven sections. In section 2 FIS are presented. Only information needed for this paper is given. In section 3 FLSE are introduced. Section 4 is the most important part of this paper where a technique to represent FIS with FLSE is described. Also, there is information about the usage, limitations and applications of this technique. An algorithm is also presented in this section. The example in section 5 is chosen to be simple and well known. Section 6 is about some methods to handle the limitations of the presented here technique. Conclusions are given in section 7.

2. Fuzzy inference systems

Fuzzy inference systems are quite classical topic. For this reason only a short description is given here. The purpose of this description is not to study such systems, but just to situate the current paper into a topic. A lot of literature on FIS and their usage can be found in [1], [2], [3], [5], [10], [11] and [12].

In general fuzzy inference process is separated in five steps.

1. Input data
2. Fuzzification
3. Fuzzy reasoning (FR)
4. Aggregation
5. Defuzzification

More about any of these steps can be found in [3]. In this paper a method to implement step 3 using FLSE is proposed and developed. This method gives to any FIS some intriguing improvements which are described in section 4.4.

Most kindly FR is permitted by a system of If-Then rules [3], [10]. Each rule actually defines a fuzzy relation between fuzzy variables or fuzzy sets and this relation can be represented by a fuzzy relational equation. Hence, a system of If-Then rules can be represented by a system of FLSE. Details are presented in section 4.

3. Fuzzy linear systems of equations (FLSE)

FLSE are not the main topic of this paper and because of this, only shallow view on them is presented here. More information on FLSE can be found in [4], [6], [7] and [9].

Max-min FLSE has the form:

$$\begin{cases} (a_{11} \wedge x_1) \vee (a_{12} \wedge x_2) \vee \dots \vee (a_{1n} \wedge x_n) = b_1 \\ (a_{21} \wedge x_1) \vee (a_{22} \wedge x_2) \vee \dots \vee (a_{2n} \wedge x_n) = b_2 \\ \dots \\ (a_{m1} \wedge x_1) \vee (a_{m2} \wedge x_2) \vee \dots \vee (a_{mn} \wedge x_n) = b_m \end{cases} \quad (1)$$

where $a_{ij}, b_i \in [0,1]$, $i = \overline{1, m}$, $j = \overline{1, n}$, are given and $x_j \in [0,1]$, $j = \overline{1, n}$ marks the unknown in the system. Operation $\vee$ between two numbers a and b is $a \vee b = \max(a, b)$. Operation $\wedge$ between two numbers a and b is $a \wedge b = \min(a, b)$.

The system (1) will be represented in matrix form:

$$A \bullet X = B \quad (2)$$

where $A = (a_{ij})_{m \times n}$ is a matrix of coefficients, $B = (b_i)_{m \times 1}$ holds the right hand side vector and $X = (x_j)_{1 \times n}$ is a vector of unknowns.

3.1. Solutions

The next notions are according to [7]:

A vector $X^0 = (x_j^0)_{1 \times n}$ with $x_j^0 \in [0,1]$, $j = \overline{1, n}$, is called solution of the system $A \bullet X = B$ if $A \bullet X^0 = B$ holds.

The set of all solutions of $A \bullet X = B$ is called complete solution set and is denoted by $\mathbf{X}^0$. If $\mathbf{X}^0 \neq \emptyset$ then the system is called consistent, otherwise it is called inconsistent.

A solution $X_{low}^0 \in \mathbf{X}^0$ is called lower solution of $A \bullet X = B$ if for any $X_{low} \in \mathbf{X}^0$ the relation $X^0 \leq X_{low}^0$ implies $X^0 = X_{low}^0$.

A solution $X_u^0 \in \mathbf{X}^0$ is called upper solution of $A \bullet X = B$ if for any $X^0 \in \mathbf{X}^0$ the relation $X_u^0 \leq X^0$ implies $X^0 = X_u^0$. If the upper solution is unique, it is called greatest (or maximum) solution.

$(X_1, \dots, X_n)$ with $X_j \subset [0, 1]$ is called an interval solution of the system $A \bullet X = B$ if any $X^0 = (x_j^0)_{n \times 1}$ with $x_j^0 \in X_j$ implies $X^0 = (x_j^0)_{n \times 1} \in \mathbf{X}^0$, for each j , $1 \leq j \leq n$. Any interval solution of $A \bullet X = B$ whose components (interval bounds) are determined by a lower solution from the left and by the greatest solution from the right, is called maximal interval solution of $A \bullet X = B$.

For the system (2) if the matrices A and X are given, computing their product is called direct problem resolution. If the matrices A and X are given, computing the unknown matrix X is called inverse problem resolution.

Direct and inverse problems resolution is very essential to the presented paper. Despite that direct problem resolution is a trivial task, solving the inverse one can be much more complicated. Methods and software packages are presented in [6], [7], [13] and [14].

4. Represent FR through FLSE

4.1. Limitations

There are two general limitations introduced here.

- Only one output value per rule is supported. Still, many input variables are supported.
- Only If-Then rules with OR connection between all their fuzzy values are supported.

In spite of that these limitations are very restrictive, some technique to handle them is proposed in section 6.

4.2. Representation

Representing FR through a FLSE can be proceeding by connecting all the parts from the system of If-Then rules with all the parts from the system (2). The connections are as follows:

- The If-Then rules are represented with the matrix A and the vectors X and B .
- The input is represented by the vector X .
- The output is represented by the vector B .

For the first step, there is need to select each column in the matrix A to represent one fuzzy value from the input variables. In this manner if in the FIS there

are two input variables and each has three values, the matrix A should have six columns. Each element in the vector X should map out one fuzzy value from all the input variables. Each element in the vector B should map corresponding fuzzy value from corresponding output variable.

Next, the matrix A should be filled in with all the rules in the system. Each row in A represents one If-Then rule.

Let for some FIS, r is the number of all If-Then rules and v is the number of all fuzzy values. Each element in the matrix A has the following value:

$$a_{ij} = \begin{cases} 1, & \text{if the } j^{th} \text{ fuzzy value is part of the } i^{th} \text{ rule} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where $i = \overline{1, r}$, $j = \overline{1, v}$

Also, the vector X should have v elements and vector B should have r elements.

The input is represented by the vector $X_{v \times 1}$ and the output by the vector $B_{r \times 1}$.

4.3. Direct problem resolution

Direct problem resolution for system (2) is equivalent to processing FR. After a FIS is represented with the system (2) the FR process can be obtained by composing the matrix A and the vector B with max-min composition. After defuzzification of the result vector X the output is obtained.

4.4. Inverse problem resolution

Solving the inverse problem for system (2) is much more interesting task. It is also much harder. In general it means to solve the system (2) when A and B are given and X is a vector of unknowns. In the terms of FR, this means that if the system and the output are given, feasible inputs can be determined. This can be really helpful for diagnostics of a system, adjusting the rules, reducing their numbers, etc. Other interesting issue is that in most cases, FLSE has more than one solution. This means that one output can be obtained through many inputs. Using presented below algorithms and software, all solutions can be obtained.

Because the inverse problem resolution is a complex task, complex theory, algorithms and software should be developed in order to solve the problem. Despite that nowadays the problem appears as almost classical, there are still only few relevant algorithms to solve it and even less software implementations.

More information can be found in [6], [7], [13] and [14].

Algorithm highly based on the algorithm described in [6] is used here. The algorithm is improved by the author in order to make it more efficient and a little bit simpler ([13] and [14]). Also a software package named FC²ore and presented in [14] is used here.

Just a short description of the problem, the algorithm and the software is given here. Detailed information can be found in [14].

4.4.1. Algorithm overview

Algorithm 1: Solving max-min FLSE.

1. Obtain input data for the matrices A and B i.e. the system of If-Then rules and the output.
2. Obtain greatest solution for the system and check for consistency. If the system (2) is consistent this means that the output values can be obtained from the current set of If-Then rules. Also the maximal value for all of the inputs is determined here.
3. If the system is inconsistent the outputs in B cannot be obtained with A , go to step 5.
4. Obtain all minimal solutions for (2). With this the whole solution set of the system is obtained i.e. all possible combinations of values for the inputs to obtain the given output.
5. Exit.

4.5. Applications

This approach gives a lot of new opportunities when working with FIS. Since this paper is not dedicated on applications, some of them are just mentioned in this section. Most general, described here approach makes each FIS (under the mentioned limitations) working both, in direct and inverse directions.

4.5.1. Diagnostics

It is very common approach to use FIS for determining some consequences. All the rules in such systems are with structure similar to: If (Conditions) Then (Consequences). As far as, the conditions are observable, the consequences can be obtained with direct problem resolution.

The opposite approach is when the consequence is already observed but there is need to diagnose the cause. Solving the inverse problem, all possible causes can be obtained.

4.5.2. Verification

Using presented approach a FIS can be verified with checking for consistency and solving it in inverse direction. If for some FIS there are also observed examples, the system can be checked against them for consistency and for correctness.

4.5.3. Adjustment

Adjustment to the FIS can be realized with adjusting the matrix A . As by default it contains only 0 and 1, it can contains any other numbers in $[0,1]$ interval, as both the algorithm and the software support this. This is similar to applying fuzzy modifiers and can be used for fine tuning a FIS or even to change its entire behavior.

4.5.4. Minimization

Any FLSE can be checked for linear independence [8]. Dependant rows in some FLSE means that they can be removed from the system and this will keep its behavior. Using this FIS can be minimized by removing unnecessary If-Then rules.

4.5.5. More

More applications can be found in [7].

5. Example

An example of representing FIS with FLSE is given in this section. Solving the direct and inverse problem is also demonstrated.

The chosen example is the famous fuzzy tipper, presented in [3] in the part of the MATLAB's fuzzy logic toolbox. It is selected because it is simple and thus very convenient for such an example. Also it perfectly covers all the limitations and finally it is easy available and thus well known.

5.1.Representation

First the model of FLSE should be obtained from the tipper.

Fuzzy tipper has two input variables: Service(S) and Food(F). Service variable has three values and Food variable has two. This means that the matrix A will have five columns. Their meanings are as follows: S:Poor, S:Good, S:Excellent, F:Rancid and F:Delicious.

There is one output variable called Tip(T) with three values: T:Cheap, T:Average and T:Generous. The tip is measured with a percentage of the bill. Also the “Then” part the rules (presented below) show the meaning for each component in B.

Three rules should be represented in the matrix A. They are:

1. If (S:Poor) or (F:Rancid) Then (T:Cheap)
2. If (S:Good) Then (T:Average)
3. If (S:Excellent) or (F:Delicious) Then (T:Generous)

From these rules matrix A is obtained to be:

$$A_{3 \times 5} = \begin{pmatrix} S : Poor & S : Good & S : Excellent & F : Rancid & F : Delicious \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

The vector B has three components. Their meaning is taken from the three rules. In this example b_1 stands for T:Cheap, b_2 stands for T:Average and b_3 stands for T:Generous.

With this the system is represented in the form

$$A_{3 \times 5} \bullet X_{5 \times 1} = B_{3 \times 1} \quad (4)$$

Next step is to solve the system (4). First the direct problem is applied then with concrete values for vectors X and B the inverse. FC^2 ore is used for this.

Let the input given values to be $S = 2$, $F = 2$. From here, after fuzzification there is: $S:Poor = 0.5$, $S:Good = 0.2$ and $F:Rancid = 0.2$. All other values are equal to 0. This means that $X_{5 \times 1} = (0.5 \ 0.2 \ 0 \ 0.5 \ 0)'$.

5.2. Direct problem

Direct problem resolution is equivalent to composing $A_{3 \times 5} \bullet X_{5 \times 1} = B_{3 \times 1}$ with max-min composition.

After this is done the answer is $B_{3 \times 1} = (0.5 \ 0.2 \ 0)'$. And after defuzzification based on centroid function, the final answer is $Tip = 7.52\%$.

5.3. Inverse problem

For solving the inverse problem the last obtained B will be used. Now $A_{3 \times 5}$ and $B_{3 \times 1}$ are given. Using FC^2 ore the FLSE is solved and all possible values of $X_{5 \times 1}$ are obtained. The FLSE has one greater solution and two lower solutions which are:

X_{gr}	X_{low_1}	X_{low_2}
0.5	0.5	0
0.2	0.2	0.2
0	0	0
0.5	0	0.5
0	0	0

That means that there are two maximal interval solution:

$$X_1 = \begin{pmatrix} 0.5 \\ 0.2 \\ 0 \\ [0, 0.5] \\ 0 \end{pmatrix} \quad X_2 = \begin{pmatrix} [0, 0.5] \\ 0.2 \\ 0 \\ 0.5 \\ 0 \end{pmatrix}$$

which are all possible values for $X_{5 \times 1}$ to obtain the same result $B_{3 \times 1}$. This basically means that either $S:Poor$ or $F:Rancid$ can be arbitrary number between 0 and 0.5 and this will not change the tip.

6. Handling the limitations

6.1. One output value per rule

Solving system with more than one output variables means that vector B should become matrix. This means that so called relational equations should be solved. Description for relational equations and efficient method for solving them can be found in [7].

Also, normally in FIS, for output only AND connection between different fuzzy values in different fuzzy variables are supported. So this limitation can be easily avoided if such a rule is split using that $a \Rightarrow (b \wedge c) \equiv (a \Rightarrow b) \wedge (a \Rightarrow c)$.

In other words one easy way to solve systems with many output variables is to split the system (2) into many systems. Every system should have for its right hand side one column of the matrix B.

6.2. Only OR connections

The limitation that only OR connection between terms are supported is much more complicated. It becomes because of the used composition in the system (2). With max-min composition only OR connection can be represented. It is possible to use min-max composition for AND connection and actually for any connection a convenient composition can be studied and used. Here the problem is that solving the inverse problem in different compositions can be a hard task. Also FC²ore can solve FLSE with six different compositions.

Using different composition for different connections unfortunately is not enough to deal with this limitation. Bigger problem here is how to manage the situation (which is very common) where in one FIS there are different rules with different connections and even different connections between the terms in same rule.

The first situation can be avoided again by splitting the system in many systems but this time all rules with same composition are separated in separated systems.

No solution for the second situation can be given by the author.

7. Conclusion

A new approach to present FIS is presented in this paper. It gives a lot of opportunities, but on the price of a lot of limitations.

REFERENCES

- [1] Baldwin, J.F. "Fuzzy logic and fuzzy reasoning," in Fuzzy Reasoning and Its Applications, E.H. Mamdani and B.R. Gaines (eds.), London: Academic Press, 1981.
- [2] Esragh, F. and E. H. Mamdani. A general approach to linguistic approximation, Fuzzy Reasoning and Its Applications, E.H. Mamdani and B.R. Gaines (eds.), London, Academic Press, 1981.
- [3] Jang, J. S. R. and N. Gulley. The Fuzzy Logic Toolbox for Use with MATLAB, Massachusetts, MathWorks Inc., 1995.

- [4] Klir, G., B. Yuan. Fuzzy Sets and Fuzzy logic: Theory and Applications, Prentice-Hall, PTR, Englewood Cliffs, NJ, 1995.
- [5] Mamdan,i E., S. Assilian. An experiment in linguist synthesis with fuzzy logic controller. *Int J Man-Mach Stud* 7(1), 1975, pp. 1-13.
- [6] Peeva, K. Universal algorithm for solving fuzzy relational equations, *Italian Journal of Pure and Applied Mathematics* 19, 2006, pp. 9-20.
- [7] Peeva, K. and Y. Kyosev. Fuzzy Relational Calculus-Theory, Applications and Software (with CD-ROM), In the series *Advances in Fuzzy Systems - Applications and Theory*, Vol. 22, World Scientific Publishing Company, 2004.
- [8] Peeva, K., Z. Zahariev. Software for testing linear dependence in Fuzzy Algebra, in *International Scientific Conference Computer Science*, September 2005, Chalkidiki, Greece, pp. 294-299.
- [9] Sanchez, E. Resolution of composite fuzzy relation equations, *Information and Control*, 30, 1976, pp. 38-48.
- [10] Siler, W, Buckley JJ (2004) Fuzzy expert systems and fuzzy reasoning. Wiley InterScience.
- [11] Yager, R.R., and L. A. Zadeh. *An Introduction to Fuzzy Logic Applications in Intelligent Systems*, Kluwer Academic Publishers, 1991.
- [12] Zadeh, L.A. Making computers think like people, *I.E.E.E. Spectrum*, 8/1984, pp. 26-32.
- [13] Zahariev, Z. Solving Max-min Relational Equations. Software and Applications, *Proceedings of 34th International Conference AME*, Sozopol, American Institute of Physics, 2008, pp. 516-523.
- [14] Zahariev, Z. Software package and API in MATLAB for working with fuzzy algebras, In *International Conference "Applications of Mathematics in Engineering and Economics (AMEE'09)"*, AIP Conference Proceedings, G. Venkov, R. Kovatcheva, V. Pasheva (eds.) American Institute of Physics, ISBN 978-0-7354-0750-9, 434-350.

Received 23.02.2010

Reviewer Prof. DSc Ketii Peeva

ЕКОЛОГИЧНО ЧИСТА ЕЛЕКТРОЕНЕРГИЯ ОТ ЯДРЕНИ ЕЛЕКТРОЦЕНТРАЛИ

Владимир Велев, Калин Филипов

***Резюме.** ЯЕЦ, в качеството си на електрогенерираща мощност, произвежда екологично чиста електроенергия. Това е свързано с типа на използваното гориво и с възприетата технология на енергопреобразуване. В статията е направено сравнение на предимствата и недостатъците от използването на ЯЕЦ и възобновяеми енергийни източници.*

ECOLOGICALLY CLEAN ELECTRICITY FROM NUCLEAR POWER PLANTS

Vladimir Velev, Kalin Filipov

***Abstract.** The nuclear power plants (NPP) in their capacity of electrical power producing facilities generate ecologically clean electricity. This is connected to the type of the fuel used and the perceived technology of energy conversion. In the article a comparison has been made to show the advantages and the disadvantages of the NPP utilization to the renewable energy resources.*

1. Увод

Ядрените електроцентрали заемат водещо положение в областта на неконвенционалните източници на енергия. По аналогия с електроцентралите, работещи с органични горива, в основата на процеса на получаване на електроенергия лежи изпарението на водата, използвана като работно тяло. Основната разлика между ЯЕЦ и конвенционалната ТЕЦ се състои в метода на нагряване на водата до състояние на кипене. Докато в парогенераторите на ТЕЦ се изгарят органични горива, то в ЯЕЦ за гориво се използва уран. Урановото гориво, обработено по подходящ начин, се разполага в реактора. В резултат на това възниква управляема верижна реакция на делене на ядрата на горивото, като отделената енергия се използва за подгръване на водата.

Под **ядрено гориво** се разбира материал, който съдържа нуклиди, делящи се при взаимодействие с неутрони, в резултат на което се освобождава енергия. Делящи се са изотопите на тежките елементи (с големи масови числа) уран и плутоний, а именно ^{233}U , ^{235}U , ^{239}Pu и ^{241}Pu . От тях в природата се среща единствено изотопът ^{235}U . Изотопите на плутония се получават при радиационен захват на неутрони от ^{238}U , а ^{233}U се получава в резултат на облъчване с неутрони на природния торий (^{232}Th). Изотопите на урана и

плутония могат да се делят при взаимодействие с неутрони с различни енергии (топлинни, бързи и с междинни енергии).

Ядреното гориво се използва в реакторите във вид на метали, сплави, металокерамика, оксиди, карбиди, нитриди и други горивни композиции, изпълнени в определена конструктивна форма.

2. Екологични аспекти при използването на ядрено гориво

Използването на ядреното гориво има редица важни особености, обусловени от физическите свойства и ядрения характер на протичащите процеси. Тези особености определят спецификата на ядрената енергетика, характера на нейната техника, особените условия на експлоатация, икономическите показатели и влиянието ѝ върху околната среда. Тези особености са следните [1], [4], [5]:

- *Висока топлотворна способност.*, т.е. голямо топлоотделяне, отнесено към единица маса. Тя е обусловена от значителната вътрешноядрена енергия, освобождавана при всеки акт на делене. При окисление на органично гориво (въглерод), което е екзотермична реакция, се освобождава енергия около 4 eV на всеки акт на взаимодействие, докато при делене на ядрото на урана се отделя енергия от порядъка на 200 MeV, от които 184 MeV се превръщат непосредствено в топлинна енергия [2]. Енергоотделянето в тези два процеса се различава $50 \cdot 10^6$ пъти за един акт на взаимодействие. Изхождайки от отношението на атомните маси на урана и въглерода (235:12), това съотношение, приведено за единица маса, се редуцира до $2.5 \cdot 10^6$ пъти. Такова висококонцентрирано отделяне на енергия от единица маса е съпроводено с мощно радиационно въздействие върху горивото и конструктивните материали от неутронни потоци с висока плътност. Това обстоятелство поставя високи изисквания към използваните материали в реакторостроенето, а също така и към реализирането на надеждно топлоотвеждане от всеки топлоотделящ елемент.
- *Необходимост от многократна циркулация на горивото в ядрения горивен цикъл.* Тази необходимост се дължи на факта, че горивото, заредено в активната зона на реактора, трябва винаги да притежава критична маса, поради което е невъзможно неговото пълно изгаряне.
- *Необходимост от многократна циркулация на горивото в ядрения горивен цикъл.* Невъзможността за пълно изгаряне на ядреното гориво се дължи на факта, че горивото, заредено в активната зона на реактора, трябва винаги да притежава критична маса, без която е невъзможно осъществяването на самоподдържаща се верижна реакция на делене. В действителност за получаване на енергия в

реактора е необходимо да има надкритична маса от делящи се нуклиди. Този излишък създава запаса от реактивност на реактора, необходим за осигуряване на продължителната му експлоатация между две презареждания с гориво (кампания на реактора). За реакторите с топлинни неутрони дълбочината на изгаряне на ядреното гориво е под 4%. В отработеното гориво обаче все още се съдържа голямо количество от делящи се и възпроизвеждащи нуклиди. То може да бъде върнато отново в горивния цикъл за повторно използване след радиохимична преработка и отстраняване на продуктите на делене (шлаки). Количеството на делящите се нуклиди в отработеното гориво може да представлява до 50% от първоначалното, което се съдържа в свежото ядрено гориво. От това следва, че ядреното гориво трябва многократно да циркулира през реактора и горивните предприятия на ядрената промишленост. Следователно, характерна особеност на пътя на движение на ядреното гориво в предприятията на ядрената енергетика е техническата възможност за рециклиране на неизползваните в отработеното гориво делящи се и възпроизвеждащи нуклиди. Върху възможностите за рециклиране на урана и плутония е основана и концепцията за развитие на ядрената енергетика на основата на реактори на бързи неутрони – реактори размножители.

- *Възпроизводство на ядреното гориво.* Това е процес на създаване на нови делящи се изотопи в горивото. Като сравнение, за реактор ВВЕР-1000 около 1/3 от топлинната енергия се получава в резултат на деленето на ^{239}Pu .
- *Специфика на „изгарянето“ на ядреното гориво.* Ядреното гориво не се нуждае от кислород за своето „изгаряне“. Също така процесът на „горене“ при ядрените горива не се съпровожда с отделяне на продукти на горенето – димни газове и твърди частици (пепел).
- *Радиоактивност, остатъчно топлоотделяне и гарантиране на безопасността при експлоатацията на ЯЕЦ.* Натрупването на радиоактивни продукти на делене в топлоотделящите елементи, тяхната много висока активност и свързаното с това дълговременно остатъчно топлоотделяне след спирането на реактора, както и високата придобита радиоактивност на конструктивните материали и на топлоносителя, съпътстват производството на електроенергия от ЯЕЦ [3]. Всичко това води до предявяването на особени изисквания към проектирането и експлоатацията на ядрените централи, тяхното основно оборудване, а също така на системите за контрол, управление и защита и на системите за гарантиране на ядрената и радиационна безопасност. Удовлетворяването на тези изисквания довежда до увеличаване на специфичните капиталовложения за изграждането на ЯЕЦ в сравнение с ТЕЦ.

Това увеличение се дължи най-вече на задължителното прилагане в ядрените електроцентрали на специализирани системи, устройства, прибори и материали. Особено внимание се обръща на топлотехническата надеждност и инженерното обезпечаване на ядрената и радиационна безопасност при аварийни ситуации, включително и при тежки аварии.

Спецификата на „изгарянето” на ядреното гориво обуславя екологично чистия характер на произвежданата от ЯЕЦ електроенергия. В действителност нещата не стоят точно така. Независимо от факта, че при нормална експлоатация на ЯЕЦ не се отделят вредни емисии в околната среда, отпадъчния продукт от електропроизводството изисква предприемането на специални мерки, свързани с неговото преработване и съхранение. Проблемът, свързан със съхранението и транспорта на отработеното ядрено гориво (ОЯГ), е основен въпрос, от рационалното решение на който зависи безопасната и икономична експлоатация на ядрените електроцентрали. Съществуват няколко начина за неговото решаване – погребване на отработеното ядрено гориво, дълговременно съхранение с последваща преработка и преработка без предварително съхранение. Повечето страни с развита ядрена енергетика обръщат сериозно внимание на радиохимичната преработка на отработеното ядрено гориво. Те гледат на това като необходимост, за да се намалят радиоактивните отпадъци, както и като главен способ за увеличаване на енергийните ресурси. Цената на операциите, свързани с транспорта и съхранението на отработеното ядрено гориво е включена в структурата на производствените разходи на произвежданата електроенергия. Независимо от придобитата радиоактивност на радиоактивните отпадъци, тяхното количество, отнесено към произведената електроенергия е много малко, което се дължи както на характера на третиране на РАО и на отработеното ядрено гориво, така и на високата „калоричност” на ядреното гориво.

Допълнително внимание заслужава фактът, че евентуална аварийна ситуация в ядрена електроцентрала, свързана с изхвърляне на радиоактивни продукти в околната среда, представлява рисков фактор, независимо от много ниската вероятност за нейното възникване. Тук трябва да се подчертаят две обстоятелства. Първото е свързано със завишените мерки за безопасност на съвременните ЯЕЦ, които значително увеличават специфичните капитални вложения за изграждането им. Второто обстоятелство представлява психологическият фактор, свързан с експлоатацията на ЯЕЦ. Повечето хора свързват подсъзнателно ядрената централа с атомната бомба. Този начин на мислене води до преекспониране на темата за ядрената и радиационна безопасност, независимо от много ниската честота на възникване на обществено опасни ядрени аварии.

Особен интерес представлява и въпросът, свързан с извеждането на ЯЕЦ от експлоатация. Независимо от значителната цена за провеждане на подобна операция (която зависи от възприетия подход за извеждане от експлоатация), тези разходи също са включени в структурата на ценообразуването на

произвежданата електроенергия. Също така е важно да се отбележи, че не съществува промишлено производство без съответните производствени отпадъци и необходими финансови средства за тяхното третиране. Това се отнася както за периода на експлоатация на съоръженията, така и за извеждането им от експлоатация.

Крайната цел на извеждането от експлоатация на ядрените съоръжения е пълното отстраняване на остатъчната радиоактивност и възвръщането на площадката за неограничено ползване. Процесът може да бъде извършен на последователни етапи, обусловени от физическото състояние на централата и оборудването и необходимият надзор от регулаторните органи.

Стадий 1: Безопасно съхранение. Състои се от минимална деконтаминация и физически и административен контрол за обезпечаване на ограничен достъп. Преди изпълнението отработеното гориво трябва да бъде изведено. Това ниво не променя лиценза на централата (зависимостта от надзорните органи).

Стадий 2: Ограничен демонтаж. Оборудването и сградите, които могат да бъдат лесно деконтаминирани се отстраняват или служат за други цели. Всички системи се дренират, биологичната защита на реактора се разширява и понякога вътрешнокорпусните елементи се отстраняват. Нивото на надзор може да бъде понижено.

Стадий 3: Неограничено използване. Сградите, оборудването и материалите, които не могат да бъдат деконтаминирани под установените нива на чистота се отстраняват, а отпадъците се преработват и складираат. Останалите части на централата се определят като зона за неограничено ползване. В някои случаи инсталациите се демонтират напълно и площадката се рекултивира до зелено поле.

Стратегията на извеждане от експлоатация е строго индивидуална за всяко ядрено съоръжение и се влияе както от неговите конкретни характеристики, така и от ядрената политика на експлоатиращата го държава, възможностите за преработване на радиоактивни отпадъци, бъдещото използване на площадката и други. Най-общо приетата стратегия обикновено представлява вариант или комбинация от следните три:

Метод 1: Деконтаминация. Всички компоненти и структури, които са радиоактивни, се почистват или демонтират, след което се пакетират и съхраняват в хранилище за радиоактивни отпадъци или временно на територията на площадката. След приключването на този етап и прекратяването на лиценза на ЯЕЦ, чистата част от площадката вече може да бъде използвана за други цели.

Метод 2: Безопасно съхранение. При този метод ядреното съоръжение се запазва цяло и се подлага на безопасно съхранение за срок от 20 до 150 години при осигуряване на поддържане и надеждна физическа охрана. Методът използва времето като деконтаминиращ ефект – след достатъчно понижаване на радиоактивността се преминава към дейности от обхвата на първия метод.

Метод 3: Погребване. Методът представлява поставянето на радиоактивните структури, компоненти и системи в дълготрайна субстанция

(например бетон). При този метод също трябва да се осигури подходящо поддържане, наблюдение, контрол и физическа охрана на погребваните елементи, дотогава, докато радиоактивността не се понижи до нива, които позволяват прекратяването на лиценза на съоръжението.

Трябва да се отбележи, че генерирането на промишлени отпадъци при експлоатация и извеждане от експлоатация важи косвено също така и за възобновяемите енергийни източници, тъй като за производството на отделните съоръжения (например вятърни турбини, фотоволтаични съоръжения и др.) също е необходима електроенергия, получена от конвенционални енергоизточници. Освен това, производството и бъдещото им извеждане от експлоатация са свързани и с генерирането на отпадъчни продукти.

От съществено значение е и съпоставимостта на производствените разходи и изкупните цени на електроенергията, получена от различни източници. При средна продажна цена на електроенергията от порядъка на 100 лв/MWh, цената на електроенергията, произвеждана от АЕЦ "Козлодуй", е около 40 лв/MWh цена, себестойността на произвежданата електроенергия от "Марица-изток 2" е около 80 лв/MWh, а от "Марица-изток 3" – 100 лв/MWh. Преференциалните цени на изкупуваната електроенергия от възобновяеми енергийни източници, според Решение на ДКЕВР №Ц-04/30.03.2009, са приложени в долната таблица.

Табл. 1. Преференциални изкупни цени на електроенергия от ВЕИ за 2009г.

ВЕЦ (до 10 MW)	105 лв/MWh
ВтЕЦ (в зависимост от типа и мощността)	145÷189 лв/MWh
ЕЦ с фотоволтаични модули (в зависимост от мощността)	755÷823 лв/MWh
ЕЦ работещи с биомаса (отпадна дървесина, отпадъци от земеделски култури или енергийни култури)	166÷217 лв/MWh

3. Заключение

В заключение може да се подчертае следното. Безспорно производството на електроенергия от възобновяеми енергийни източници, независимо от задължителния му характер към днешна дата, е необходимо, перспективно и екологично обосновано. Обаче по наше мнение, към него трябва да се подхожда по-предпазливо по следните причини. На първо място интензивността на първичните енергоизточници за производство на електроенергия от възобновяеми енергийни източници (ВЕИ) е в повечето случаи е на порядъци по-ниска спрямо тази при конвенционалните и ядрени електроцентрали. Това обуславя значителна заета площ от ВЕИ спрямо ТЕЦ и ЯЕЦ, както и ниските капацитивни възможности за енергопроизводство. Също

така, независимо от екологичния характер на електропроизводството от възобновяемите енергийни източници, в повечето случаи се нанасят значителни, а често и непоправими щети на флората и фауната в близост до съоръженията. Не на последно място високите преференциални изкупни цени на електроенергията от ВЕИ рефлексират върху крайната продажна цена на електроенергията. Това се отнася в известна степен и за мероприятията, свързани с повишаване на енергийната ефективност в промишлеността и бита. Независимо от съвременния характер на решенията за повишаване на енергийната ефективност, тяхното реализиране е свързано с разход на значителни финансови средства, които често са непостижими за крайните потребители на електрическа и топлинна енергия.

Литература

- [1] Велев, В., К., Филипов. Ядрени горива, Ифо Дизайн, София, 2008.
- [2] Глухов, Г. Ядрени енергийни реактори, ИФО Дизайн, С., 2004.
- [3] Ран, Ф., А. Адамантиадес, Дж. Кентон, Ч. Браун. Справочник по ядерной энерготехнологии, Энергоатомиздат, М., 1989.
- [4] Синев, Н., Б. Батуров. Экономика атомной энергетики, Атомиздат, М., 1980.
- [5] Синев, Н. Экономика ядерной энергетики, Энергоатомиздат, М., 1987.
- [6] Руководство по безопасности МАГАТЭ, №50-SG-D9

Автори: Проф. д-р. Владимир Велев, гл. ас. Калин Филипов, Технически Университет – София, катедра Топло и ядрена енергетика.

Постъпила на 24.02.2010

Рецензент доц. д-р Димитър Попов

МЕТОД ЗА ДЪЛГОСРОЧНО ИНВЕСТИРАНЕ С МАРКОВСКА ВЕРИГА И ДИНАМИЧНО ПРОГРАМИРАНЕ

Георги Владов

Резюме. Предлага се метод за дългосрочно инвестиране в портфейл съставен от предварително избрани класове активи. Въвеждат се краен брой състояния за описание случайните изменения на доходността на портфейла, а преходите между тях се моделират с Марковска верига. Дефинирани са две функции на печалба и е разработен марковски процес за взимане на решение с използване на динамично програмиране. Разгледан е илюстративен пример, в който резултатите от прилагането на метода са сравнени с два пазарни индекса.

A METHOD FOR MULTIPERIODIC INVESTMENT BASED ON MARKOV CHAIN AND DYNAMIC PROGRAMMING

Georgi Vladov

Abstract. A method for multiperiodic investment is proposed in portfolio of preliminarily selected asset classes. A finite number of states are introduced to describe the portfolio random return changes and Markov chain is applied to model the transitions between them. Two profit functions are defined and Markov decision process is developed using dynamic programming. An illustrative example is included, where the results of proposed method are compared with two market indexes.

1. Introduction

Many investment models employ a single-period framework, but people often have distant goals and are motivated to invest money for more than one year to achieve them, e.g. attaining a target standard of living many periods in the future or financing the higher education of their children. Different models are developed in recent years (see, e.g. [1]-[4]) that support investors to find optimal solution in this situation, which is generalized by economists as multiperiod investment problem. One of the most common used and effective approach of solving this problem, is to define it as a problem of optimization. A typical investment optimization model distributes investment capital among selected asset classes for a specified time horizon so as to maximize (minimize) the expected profit (risk) of the investments subjected to some constraints (limited amount of investment capital, investors return and risk preferences etc.).

In this paper we present a method for solving the multiperiod investment problem based on Markov chain and dynamic programming. The funds are allocated among portfolios from selected asset classes so as to maximize the portfolio profit at the end of the investment period. We introduce finite number of states to model the portfolio returns changes and apply Markov chain to describe the transitions between them. The optimal investment strategy is then determined period by period, using dynamic programming technique. We work out an illustrative example and provide comparative analysis of the obtained results with common used benchmarks.

2. The multiperiodic investment method

We consider n -year investment period. At the beginning of a period i (one year) the investor must choose among a set of defined investment alternatives $j, j = \overline{1, k}$. The investment alternative j represents a portfolio from M selected asset classes with fixed weight of each asset class $w_m^j, m = \overline{1, M}$. We assume that capital C is invested only once at the beginning of the n -year investment period. We suppose that the asset class annual returns r_m^i vary over the investment period in a random way. Given the annual returns of asset classes, the portfolio return R_j^i corresponding to alternative j in the period i can be calculated by the formula

$$R^j(i) = \sum_{m=1}^M w_m^j r_m^i, i = \overline{1, n} \quad (1)$$

where M is the number of asset classes, and w_m^j is the weight of asset class m in portfolio j , such that $\sum_{m=1}^M w_m^j = 1$.

Transition matrixes

Relying on the market situation the portfolio return in the period $(i + 1)$ goes up or down. It is assumed that these changes depend on whether or not the portfolio return increases or decreases in the periods (i) and $(i - 1)$. A single state s to each possible portfolio return changes, can be assigned, as represented in Table 1 [5]

Table 1

State	Description
0	The portfolio return increased in the period (i) and in the period (i-1)
1	The portfolio return increased in the period (i) but decreased in the period (i-1)
2	The portfolio return decreased in the period (i) but increased in the period (i-1)
3	The portfolio return decreased in the period (i) and in the period (i-1)

Denote by $p_{lq}^j(i)$ the one-step transition probability from state l to state q in the period i , evaluated for alternative j . Assuming that transition s can be described by a Markov chain, see [5] and [6], the probability $p_{lq}^j(i)$ can be calculated using

$$\begin{aligned} p_{lq}^j(i) &= p_{lq}^j(i-1) + \frac{1}{i} * [\aleph_{lq}^j(i) - p_{lq}^j(i-1)], i = \overline{2, n} \\ p_{lq}^j(1) &= p_{lq}^j(0) + \frac{1}{2} * [\aleph_{lq}^j(1) - p_{lq}^j(0)] \end{aligned} \quad (2)$$

where

$p_{lq}^j(i)$, $p_{lq}^j(i-1)$ the one-step transition probabilities from state l to state q in the periods (i) and $(i-1)$, evaluated for alternative j ;

$\aleph_{lq}^j(i)$ $(0,1)$ – variable that denotes, if the transition from state l to state q in the

period (i) is present, $\aleph_{lq}^j(i) = \begin{cases} 1, & \text{if } l \rightarrow q \\ 0, & \text{if } l \rightarrow s, s \neq q \end{cases}$

$p_{lq}^j(1)$ the initial one-step transition probability from state l to state q , evaluated for alternative j ; $p_{lq}^j(0)$ is a zero member that is determined experimentally [10].

If one step transition from state $s \neq l$ is present than we have $p_{lq}^j(i) = p_{lq}^j(i-1)$ (3).

A simple review on the set of states indicates that transition probabilities between some of the states in Table 1 do not exist, namely p_{01}^j , p_{03}^j , p_{11}^j , p_{13}^j , p_{20}^j , p_{22}^j , p_{30}^j and p_{32}^j .

The attractiveness of investment alternatives j is represented by their average profits $\bar{r}_{lq}^j(i)$, which are assumed to be function of the period i , the current state l and next state q , and are evaluated by the formula

$$\begin{aligned} \bar{r}_{lq}^j(i) &= 1 + \{\bar{r}_{lq}^j(i-1) + \frac{1}{i} * [r_{lq}^j(i) - \bar{r}_{lq}^j(i-1)]\} / 100, i = \overline{2, n} \\ \bar{r}_{lq}^j(1) &= 1 + \{\bar{r}_{lq}^j(0) + \frac{1}{2} * [r_{lq}^j(1) - \bar{r}_{lq}^j(0)]\} / 100 \end{aligned} \quad (4)$$

where

$\bar{r}_{lq}^j(i)$, $\bar{r}_{lq}^j(i-1)$ the one-step average profits by transition from state l to state q in the periods (i) and $(i-1)$, evaluated for alternative j ;

$r_{lq}^j(i)$ the one-step profits by transition from state l to state q in the period (i) , evaluated for alternative j ;

$\bar{r}_{lq}^j(1)$, $r_{lq}^j(1)$ the initial one-step average profit/ profit by transition from state l to state q , evaluated for alternative j ; $\bar{r}_{lq}^j(0)$ is a zero member that is determined experimentally [10].

If one step transition from state $s \neq l$ is present than we have $\bar{r}_{lq}^j(i) = \bar{r}_{lq}^j(i-1)$ (5).

The impossible transitions do not yield any profit, i.e. we set $r_{01}^j, r_{03}^j, r_{11}^j, r_{13}^j, r_{20}^j, r_{22}^j, r_{30}^j$ and r_{32}^j equal to zero. The one-step transition probability $p_{lq}^j(i)$ and average profit $\bar{r}_{lq}^j(i)$ in (2) - (5) are evaluated using statistics for annual returns of the selected asset classes, which are published in [7], and the definition of states given in Table 1.

Profit functions

Denote by $v_l^j(i)$ the expected conditional profit corresponding to alternative j , given one-step transition from state l in the period (i) . We have

$$v_l^j(i) = \sum_{s=1}^S p_{ls}^j(i) \bar{r}_{ls}^j(i), l = \overline{1, S} \quad (6)$$

This one-step conditional expected profit function is normally used in Markovian Decision Process, see e.g. [5], [6].

Another one-step conditional profit function $g_l^j(i)$ can be proposed, namely

$$g_l^j(i) = \min_s \bar{r}_{ls}^j(i), l = \overline{1, S} \quad (7)$$

It yields a conservative solution, which is based on maximizing the worst possible results from the transitions, i.e., a guarantied maximum profit is obtained.

Dynamic programming model

We determine the elements of the dynamic programming model as follows:

- 1) **Stage i** : corresponds to period i , $i = \overline{1, n}$;
- 2) **State x_i at stage i** : any state s at the beginning of period i , as defined in Table1;
- 3) **Variants j at stage i** , a_j^i : any portfolio of M asset classes with fixed weights w_m^j of each class, $m = \overline{1, M}$, $j = \overline{1, k}$.
- 4) **Objective function for stage i** : the profit from stages $i, i+1, \dots, n$, given the state x_i at stage i .

Denote by $f_i(x_i)$ the maximum profit from stages $i, i+1, \dots, n$, given the state x_i . Following the principle of optimality the dynamic programming recurrent equation is written in the form:

$$\begin{aligned} f_i(x_i) &= \max_{a_j^i} \left\{ \sum_{m=1}^S p_{x_i, m}^j(i) \bar{r}_{x_i, m}^j(i) f_{i+1}(m) \right\}, i = \overline{n-1, 1} \\ f_n(x_n) &= \max_{a_j^n} \{ v_{x_n}^j(n) \} \end{aligned} \quad (8)$$

When the profit function (7) is used $g_l^j(i)$ is replaced in (8) instead of $v_l^j(i)$.

3. Illustrative example

We consider a 10-year investment period, $n=10$, and we assume that portfolio consists of three asset classes ($M=3$): Large company stocks (LCS), Long-term government bonds (LTGB) and Treasury bills (TB). The investor puts in capital $C=10\,000\$$, which is invested into one of six possible alternatives ($k=6$): $a_1=(100;0;0)$, $a_2=(0;100;0)$, $a_3=(0;0;100)$, $a_4=(70;20;10)$, $a_5=(40;30;30)$ and $a_6=(10;20;70)$, i.e., a_1 represents the portfolio that consists of 100% LCS, 0% LTGB and 0% TB, etc. The annual returns of the selected asset classes are considered over 77 year period. A sample of their values for 22 year sub period is shown in the Table 2

Table 2

Period i	r_1^i	r_2^i	r_3^i	Period i	r_1^i	r_2^i	r_3^i
1980	32.42	-3.95	11.24	1992	7.67	8.05	3.51
1981	-4.91	1.86	14.71	1993	9.99	18.24	2.90
1982	21.41	40.36	10.54	1994	1.31	-7.77	3.90
1983	22.51	0.65	8.80	1995	37.43	31.67	5.60
1984	6.27	15.48	9.85	1996	23.07	-0.93	5.21
1985	32.16	30.97	7.72	1997	33.36	15.85	5.26
1986	18.47	24.53	6.16	1998	28.58	13.06	4.86
1987	5.23	-2.71	5.47	1999	21.04	-8.96	4.68
1988	16.81	9.67	6.35	2000	-9.11	21.48	5.89
1989	31.49	18.11	8.37	2001	-11.88	3.70	3.83
1990	-3.17	6.18	7.81	2002	-22.10	17.84	1.65
1991	30.55	19.30	5.60				

Using (1) and Table 2 we calculate the portfolio annual returns $R^j(i)$ and assign a state s , as defined in Table 1, to each type of portfolio return changes. The last 12 from all 77 results are summarized in Table 3, as an illustration

Table 3

Period i	R^1	S^1	R^2	S^2	R^3	S^3	R^4	S^4	R^5	S^5	R^6	S^6
1990	-3.17	2	6.18	2	7.81	2	-0.20	2	2.93	2	6.39	2
1991	30.55	1	19.30	1	5.60	3	25.81	1	19.69	1	10.84	1
1992	7.67	2	8.05	2	3.51	3	7.33	2	6.54	2	4.83	2
1993	9.99	1	18.24	1	2.90	3	10.93	1	10.34	1	6.68	1
1994	1.31	2	-7.77	2	3.90	1	-0.25	2	-0.64	2	1.31	2
1995	37.43	1	31.67	1	5.60	0	33.10	1	26.15	1	14.00	1
1996	23.07	2	-0.93	2	5.21	2	16.48	2	10.51	2	5.77	2
1997	33.36	1	15.85	1	5.26	1	27.05	1	19.68	1	10.19	1
1998	28.58	2	13.06	2	4.86	2	23.10	2	16.81	2	8.87	2
1999	21.04	3	-8.96	3	4.68	3	13.40	3	7.13	3	3.59	3
2000	-9.11	3	21.48	1	5.89	1	-1.49	3	4.57	3	7.51	1
2001	-11.88	3	3.70	2	3.83	2	-7.19	3	-2.49	3	2.23	2
2002	-22.10	3	17.84	1	1.65	3	-11.74	3	-2.99	3	2.51	1

For example, the annual returns of investment alternatives a_1 in the period 1990 and a_4 in the period 1994 are calculated as follows:

$$R^1(1990) = (-1.00 * 0.0317 + 0 * 0.0618 + 0 * 0.0781) * 100 = -3.17$$

$$R^4(1994) = (0.7 * 0.0131 - 0.2 * 0.0777 + 0.1 * 0.0390) * 100 = -0.25$$

Further the zero members $p_{lq}^j(0)$ and $\bar{r}_{lq}^j(0)$, as defined in (2) and (4), are obtained using (1), Table 1, Table 2 and formulas [10]

$$p_{lq}^j(0) = \frac{n^j(l|q)}{\sum_{s=1}^S n^j(l|s)} \text{ and } \bar{r}_{lq}^j(0) = 1 + \frac{\sum_{i=1}^{N_0} R_{lq}^j(i)}{n^j(l|q)} * \frac{1}{100} \quad (9)$$

where

$n^j(l|q)$ number of one-step transitions from state l to state q during a period 1926 – 1991 for investment alternative j ; s is a number of states ($S = 4$).

N_0 number of periods (e.g. $N_0 = 66$ for periods between 1926 – 1991).

Thus having the results from (9) we determine the one-step transition probability $p_{lq}^j(1)$ and average profit $\bar{r}_{lq}^j(1)$, using (2) - (5), for the initial period set to 1992. The one-step transition probabilities $p_{lq}^j(i)$ and average profits $\bar{r}_{lq}^j(i)$ of investment alternative j are calculated through (2) - (5) in the period 1993 – 2002 and presented in a matrix form, i.e. $P^j(i) = [p_{lq}^j(i)]$, $R^j(i) = [\bar{r}_{lq}^j(i)]$, where $j = \overline{1,6}$ and $i = \overline{1,10}$. For example the transition matrixes related to alternative 1 in the periods 1993 – 1994 are, as follows:

$$P^1(1) = \begin{bmatrix} 0.27 & - & 0.73 & - \\ 0.19 & - & 0.81 & - \\ - & 0.71 & - & 0.29 \\ - & 0.69 & - & 0.31 \end{bmatrix}, P^1(2) = \begin{bmatrix} 0.27 & - & 0.73 & - \\ 0.13 & - & 0.87 & - \\ - & 0.71 & - & 0.29 \\ - & 0.69 & - & 0.31 \end{bmatrix}$$

The profit matrixes $R^j(i) = [\bar{r}_{lq}^j(i)]$, $j = \overline{1,6}$ are

$$R^1(1) = \begin{bmatrix} 1.32 & - & 1.02 & - \\ 1.14 & - & 1.10 & - \\ - & 1.18 & - & 0.97 \\ - & 1.23 & - & 0.85 \end{bmatrix}, R^1(2) = \begin{bmatrix} 1.32 & - & 1.02 & - \\ 1.09 & - & 1.07 & - \\ - & 1.18 & - & 0.97 \\ - & 1.23 & - & 0.85 \end{bmatrix}$$

We find the optimal dynamic programming solution following the steps of the algorithm, described in section 2. We make calculations using (6), (7) and (8). For

$v_l^j(i) = \sum_{s=1}^S p_{ls}^j(i) \bar{r}_{ls}^j(i)$ the results are summarized in Tables 4 - 6 for stages 10 to 1.

Table 4

Stage 10

$f_{10}(x_n) = \max_{a_n^j} \{v_{x_n}^j(n)\}$							Optimal solution	
x_{10}	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$f_{10}(x_{10})$	j^*
0	11 022.27	10 411.67	10 377.40	11 007.58	10 841.09	10 565.46	11 022.27	1
1	11 389.70	10 246.86	10 295.54	11 102.02	10 798.09	10 418.57	11 389.70	1
2	11 567.47	11 146.66	10 240.74	11 312.38	11 001.72	10 535.98	11 567.47	1
3	8 889.56	11 127.00	10 236.54	9 452.56	9 910.92	10 493.99	11 127.00	2

Table 5

Stage 7

$f_7(x_i) = \max_{a_i^j} \{ \sum_{m=1}^4 p_{x_i, m}^j(i) \bar{r}_{x_i, m}^j(i) f_8(m) \}$							Optimal solution	
x_7	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$f_7(x_7)$	j^*
0	16 280.59	15 473.29	15 347.55	16 142.95	16 049.43	15 698.62	16 280.59	1
1	17 052.62	15 285.34	15 107.41	16 627.33	16 161.66	15 623.94	17 052.62	1
2	17 092.22	16 306.21	15 114.53	16 708.20	16 250.65	15 622.92	17 092.22	1
3	16 434.45	15 210.09	14 992.84	16 271.60	15 581.63	15 181.76	16 434.45	1

Table 6

Stage 1

$f_1(x_i) = \max_{a_i^j} \{ \sum_{m=1}^4 p_{x_i, m}^j(i) \bar{r}_{x_i, m}^j(i) f_2(m) \}$							Optimal solution	
x_1	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$f_1(x_1)$	j^*
0	36 086.11	33 939.61	34 381.84	36 218.68	35 436.13	34 445.97	36 218.68	4
1	36 105.97	34 133.34	34 611.50	35 581.91	35 043.11	34 032.87	36 105.97	1
2	36 820.05	36 463.39	33 632.53	36 211.00	35 585.42	34 682.16	36 820.05	1
3	36 474.92	33 758.00	33 065.26	36 112.62	34 541.27	33 673.04	36 474.92	1

It can be easily found that the optimal solutions j^* at stages 10 to 8 are the same as in stages 7 to 2. From Tables 4 - 6 we derive the optimal solution:

$$f_1(x_0) = 36\,218.68, j^* = 4; f_1(x_1) = 36\,105.97, j^* = 1; f_1(x_2) = 36\,820.05, j^* = 1; \\ f_1(x_3) = 36\,474.92, j^* = 1.$$

Hence, the best decision is to select the investment alternative $j^* = 1$ in states 1 and 2, $j^* = 4$ in state 0 for stage 1 or $j^* = 1$ in state 0 for stages 2 to 10, and $j^* = 2$ in states 3 for stages 8 to 10 or $j^* = 1$ in states 3 for stages 1 to 7.

In a similar way we derive the optimal solution for $g_l^j(i) = \min_s \bar{r}_{ls}^j(i)$

Table 7

Stage 10

$f_{10}(x_n) = \max_{a_n^j} \{g_{x_n}^j(n)\}$							Optimal solution	
x_{10}	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$f_{10}(x_{10})$	j^*
0	10 218.00	10 264.33	10 190.44	10 218.41	10 681.53	10 553.47	10 681.53	5
1	10 566.71	10 237.98	10 182.36	10 381.02	10 255.93	10 117.16	10 566.71	1
2	10 295.75	9 803.06	10 180.43	10 214.03	10 149.58	10 095.28	10 295.75	1
3	8 538.95	9 870.50	10 171.12	9 236.59	9 837.53	9 971.43	10 171.12	3

Table 8

Stage 1

$f_1(x_i) = \max_{a_j^i} \{g_{x_i, m}^j(i) * f_2(m)\}, m = \overline{1,4}$							Optimal solution	
x_1	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$	$f_1(x_1)$	j^*
0	12 185.63	12 240.88	12 379.87	12 186.12	12 738.41	12 585.70	12 738.41	5
1	13 117.94	12 331.45	12 338.28	13 028.65	12 688.67	12 344.81	13 117.94	1
2	11 573.42	11 776.28	12 180.74	11 785.48	11 967.65	12 074.09	12 180.74	3
3	10 112.83	11 637.49	12 153.67	10 738.54	11 280.28	11 878.64	12 153.67	3

It can be easily found that the optimal solutions j^* at stages 10 to 7 are the same as in stages 6 to 1. From Tables 7 - 8 we derive the optimal solution:

$$f_1(x_0) = 12\,738.41, j^* = 5; f_1(x_1) = 13\,117.94, j^* = 1; f_1(x_2) = 12\,180.74, j^* = 3; \\ f_1(x_3) = 12\,153.67, j^* = 3.$$

Hence, the best decision is to select the investment alternative $j^* = 5$ in state 0, $j^* = 1$ in state 1, $j^* = 3$ in state 3, $j^* = 1$ in state 2 for stages 10 to 7, or $j^* = 3$ in state 2 for stages 6 to 1.

An examination of the results in Tables 4 - 8 shows that the first case, when the expected profit function is used, is more risky and profit oriented than the second one. The investors may apply usually the first profit function (6) when they are years away from the target and second function (7) when they are near to the period end.

We compare now the results obtained for a 10-year investment period from both profit functions with the values of two selected market indexes, namely Russell 1000 for (6) and USTREAS Stat US T-Bill 90 for (7). Both indexes represent comprehensive and unbiased barometer for profit potential of high and low risky portfolios of security asset classes and are commonly used as a benchmark performance measures, see [8], [9]. Denote by $Index1_i$ the Russell 1000 and by $Index2_i$ the USTREAS Stat US T-Bill 90 values in the period (i). The details of comparison performed are depicted in fig. 1 and fig. 2

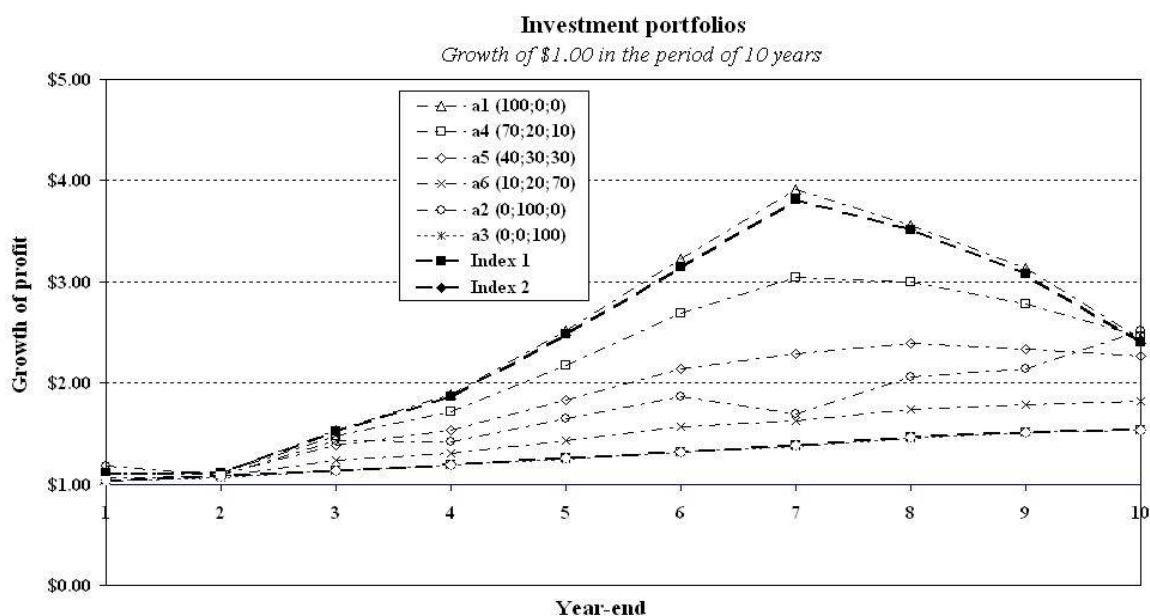


Fig. 1. Investment portfolios

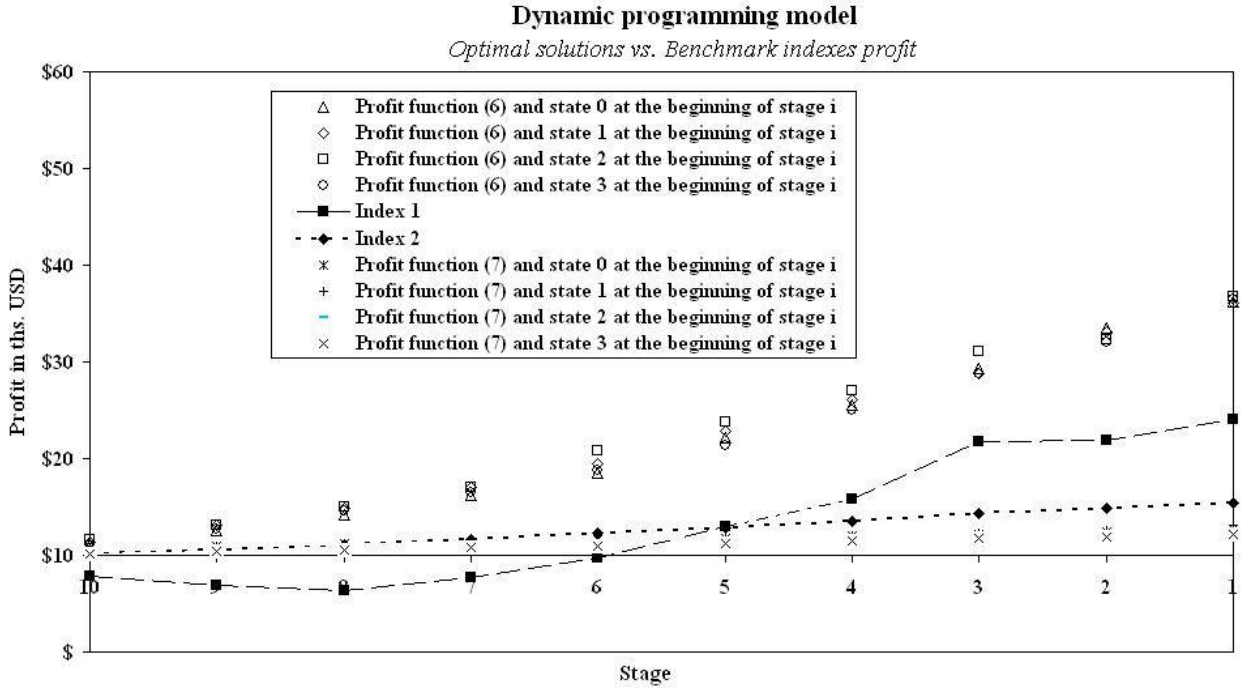


Fig. 2. Dynamic programming model

Fig. 1 indicated that the selected investment alternatives $a_j, j = \overline{1,6}$, produce results above zero profit level. The growth of \$1.00 invested in high performing alternative a_1 over the 10th year period is very close to Russell 1000 profit level, while the profit of second index is near to less risky alternative a_3 . The optimal dynamic programming solutions, represented in Fig. 2, for profit function (6) outperform both index profits (e.g. $\sum_{i=1}^{10} [f_i(x_i) - f_i^*(Index1_i)] = 81095.48 > 0$, where $f_i(x_i)$ is the minimum profit at stage i and $f_i^*(Index1_i)$ is the *Index1* profit at stage i). The results of second profit function in most of the stages are below the benchmark profit levels (e.g. $\sum_{i=1}^{10} [f_i(x_i) - f_i^*(Index2_i)] = -15536.01 < 0$, where $f_i(x_i)$ is the minimum profit at stage i and $f_i^*(Index2_i)$ is the *Index2* profit at stage i).

4. Conclusion

A method for multiperiod investment is proposed, which is based on Markovian Decision process and dynamic programming method. A finite number of states and Markov chain have been introduced to model the portfolio random return changes. An effective algorithm is proposed to evaluate the probabilities and average profits by one-step transitions between the states in Markov chain. Two one-step transition profit functions are defined and used, the second of which yields a conservative solution, i.e., a guaranteed maximum profit. The results of the method are analysed

and compared with two benchmark indexes. This method can be applied in the praxis for simulations of long term investment scenarios.

REFERENCES

- [1] Musumeci, Jim. Investing for a Distant Goal: Optimal Asset Allocation and Attitudes toward Risk., Research Dialogues. New York: TIAA-CREF, July 1998. Number 56.
- [2] Luenberger, D.G. Investment science., Oxford University Press, 1998.
- [3] Musumeci, Jim, and Joe Musumeci. A Dynamic-Programming Approach to Multiperiod Asset Allocation., Journal of Financial Services Research (February 1999): 5-21.
- [4] Ivanova, Z., K. Stoilova, T. Stoilov. Portfolio optimization – Information service in Internet, Academic Publishing House “Marin Drinov”, Sofia 2005.
- [5] Hillier, F.S., G. J. Lieberman. Operations Research, Fifth edit, Mc Graw – Hill, Inc., 1990.{4}
- [6] Taha, H. A. Operations Research. An Introduction, 5th Edition MacMillan Publ. Comp., N.Y., 1992.{5}
- [7] Ibbotson Associates, Stocks, Bonds, Bills and Inflation: 2003 Year Book., Chicago 2003.{6}
- [8] <http://www.russell.com/> {7}
- [9] <http://datalab.morningstar.com/> {8}
- [10] Gatev, G., G. Vladov. A method for solving the multiperiod investment problem based on Markov chain and dynamic programming, International conference Automatics and Informatics '05, Sofia 2005.

Autor Georgi Vladov – PhD Candidate, Dept. of Systems and Control Engineering

Received 30.04.2010

Reviewer Assoc. Prof. DSc Teofana Puleva

Адаптивен по скорост и статорно съпротивление наблюдател на потокосцепление за системи за безсензорно векторно управление

Евтим Йончев

Резюме. В работата е предложен дискретен адаптивен наблюдател на роторно потокосцепление за системи асинхронни електрозадвижвания с безсензорно векторно управление с повишена точност. Използван е модел на асинхронен мотор, описан в статорна координатна система, като в алгоритъма за оценка на неизвестните параметри – роторна скорост и статорно съпротивление, е използвана теоремата на Ляпунов. В работата се предлага нов подход - средноквадратичната грешката се минимизира чрез оптимален коефициент на обратната връзка на наблюдателя, определен по рекурсивен алгоритъм, базиран на метода на най-малките квадрати. Така е постигната повишена точност в сравнение с известните подходи – наблюдател с еталонен модел и наблюдател на Луенбъргър.

Adaptive speed and stator resistance flux observer for the systems sensorless vector control

Evtim Ionchev

Abstract. The work has proposed a discrete adaptive observer of rotor flux in the systems asynchronous electric drives with sensorless vector control with high accuracy. Used a model of induction motor described in the stator coordinate system as the key for assessing the unknown parameters - rotor speed and stator resistance is used Lyapunov theorem. The work proposes a new approach - the weighted error is minimized by an optimal feedback gains of the observer appointed by a recursive algorithm, based on the method of least squares. Thus, increased accuracy is achieved compared to known approaches - the standard model observer and Luenberger observer.

1. Въведение

Типичното векторно управление с ориентация по роторното потокосцепление изисква прецизен датчик на скорост, монтиран на вала на мотора. Допълнителният импулсен енкодер или прецизен тахогенератор, необходимият хардуер, електрическо опроводяване и прецизен монтаж компрометира предимствата на асинхронната машина с късосъединен ротор. Освен това за малките мощности (до 5 kW) цените на сензора и мотора са съизмерими. Даже при 50kW цената на сензора достига до 30% от стойността на мотора.

Поради горното голям интерес представлява възможността за изграждане на прецизни асинхронни електрозадвижвания, които не изискват допълнителен датчик на скорост или позиция на ротора, при които се изграждат специфични наблюдатели и online идентификатори.

2. Наблюдатели в системите за безсензорно векторно управление

Тъй като не всички променливи на състоянието на асинхронния двигател, необходими за ориентация на координатната система (директно или директно векторно управление с ориентация по полето) или за въвеждане на обратна връзка са измерими, се налага те да бъдат оценявани. Определянето на неизвестните променливи на състоянието при пренебрегване на смущенията и грешките при измерване се нарича наблюдение.

В оценителите по отворена схема, особено на ниски скорости, отклоненията в параметрите влияят значително върху показателите за качество на задвижването както в преходно така и в установено състояние. С цел повишаване на стабилността при вариране на параметрите и шумоустойчивостта се използват наблюдатели.

Най-често използваният наблюдател за детерминирани обекти е наблюдател на Луенбъргър[1].

Ако система може да бъде описана в дискретен момент на времето:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) \end{aligned} \quad (1)$$

и е наблюдаема, т.е. матрицата, M_0 , е от пълен ранг, състоянията могат да бъдат определени от (2) където:

$$M_0 = \begin{bmatrix} CA \\ CA^2 \\ \vdots \\ CA^n \end{bmatrix}$$

$$\begin{aligned} \hat{x}(k+1) &= A\hat{x}(k) + Bu(k) \\ \hat{x}(k+1) &= A\hat{x}(k) + Bu(k) + L(y(k+1) - C\hat{x}(k+1)) \end{aligned} \quad (2)$$

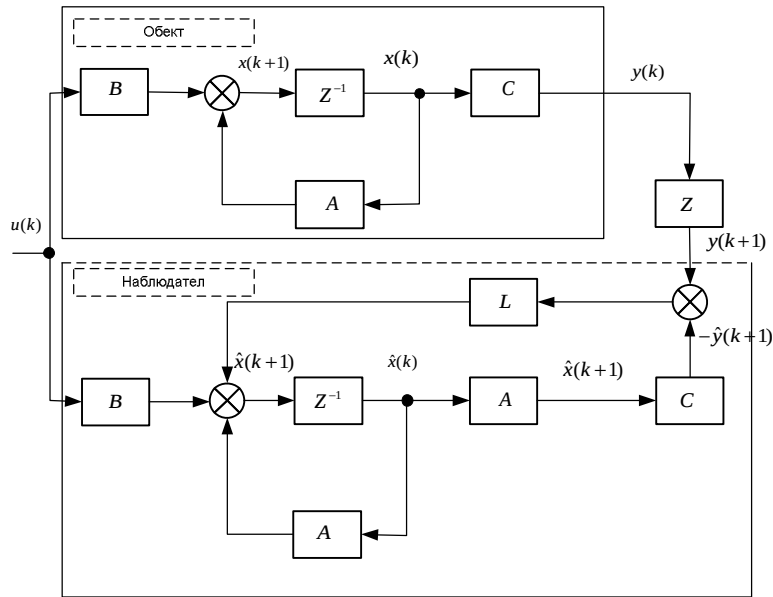
Фиг. 1. показва структурната схема на наблюдател, който е описан от (2). Изходния вектор y , се използва за изчисляване текущия вектор на състоянието x . Грешката на наблюдателя е:

$$\begin{aligned} e(k) &\triangleq x(k) - \hat{x}(k) \\ e(k+1) &\triangleq (A - LCA)e(k) \end{aligned} \quad (3)$$

където L е матрицата на обратната връзка (коефициент на усилване) на наблюдателя.

Основните стъпки при проектиране и изграждане на наблюдателя са както следва:

- избор на модела на двигателя, описан във временната област,
- дискретизация на модела,
- определяне на матрицата в обратната връзка на наблюдателя, изграждане на дискретизирания алгоритъм.



Фиг. 1. Структурна схема на наблюдател

3. Модел на двигателя

Възможно е използването на уравненията в роторна координатна система, ориентирана по полето на ротора или в статорна, ориентирана по полето на статора. За да се избегнат допълнителни преобразувания и нелинейни трансформации, стационарната координатна система е за предпочитане [2]. Основните предимства при използване на модел описан в статорна координатна система са:

- намалено време за изчисление;
- по-малко време на дискретизация;
- по-висока точност;
- по-стабилно поведение;

Моделът на асинхронния двигател описан в статорна координатна система, [1], [3] е даден по – долу.

Използван е стационарен модел на машината ($R_s = \text{const}$, $L_r = \text{const}$, $L_m = \text{const}$, $L_s = \text{const}$), без отчитане на токовото изместване в ротора ($R_r = \text{const}$) в статорна ab координатна система.

$$\frac{di_{sa}}{dt} = -\frac{(R_s + R_r \frac{L_m^2}{L_r^2})}{L_s s} i_{sa} + \frac{L_m R_r}{L_s s L_r} y_{ra} - \frac{L_m}{L_s s L_r} w_r y_{rb} + \frac{1}{s L_s} u_{sa} \quad (4)$$

$$\frac{di_{sb}}{dt} = -\frac{(R_s + R_r \frac{L_m^2}{L_r^2})}{L_s s} i_{sb} + \frac{L_m R_r}{L_s s L_r} y_{rb} + \frac{L_m}{L_s s L_r} w_r y_{ra} + \frac{1}{s L_s} u_{sb} \quad (5)$$

$$\frac{dy_{ra}}{dt} = \frac{L_m R_r}{s L_s} i_{sa} - \frac{R_r}{L_r} y_{ra} + w_r y_{rb} \quad (6)$$

$$\frac{dy_{rb}}{dt} = \frac{L_m R_r}{s L_s} i_{sb} - \frac{R_r}{L_r} y_{rb} - w_r y_{ra} \quad (7)$$

където $L_s = L_{ss} + L_m$, $L_r = L_{rs} + L_m$, $s = 1 - \frac{L_m^2}{L_s L_r}$

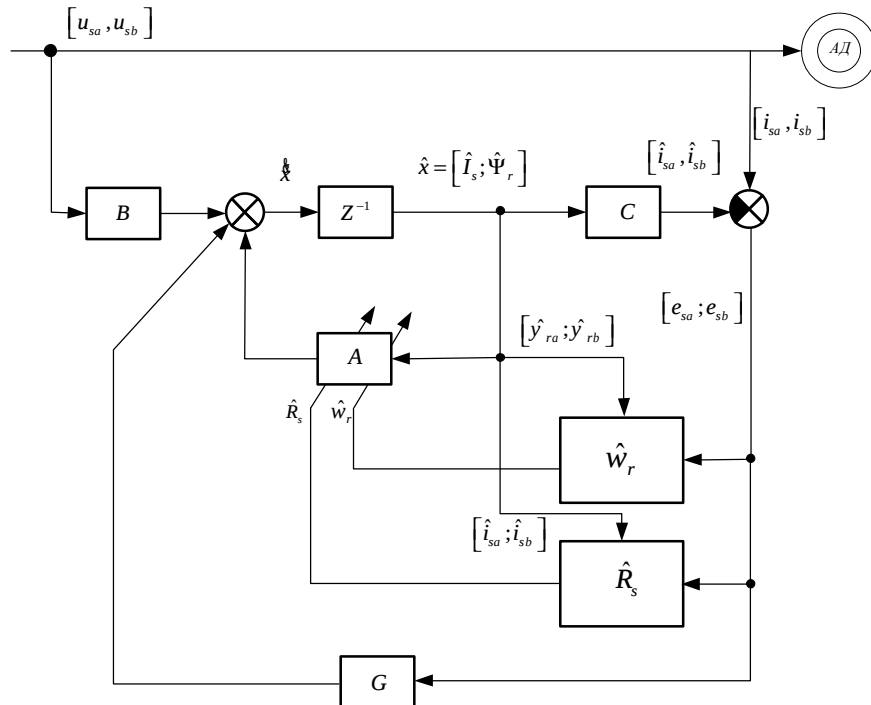
Уравнението за изходите е

$$\begin{bmatrix} i_{sa} \\ i_{sb} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} i_{sa} \\ i_{sb} \\ y_{ra} \\ y_{rb} \end{bmatrix} \quad (8)$$

Матрицата на състоянието на наблюдателя е функция на роторната скорост $\hat{A}(w_r)$ и поради това, освен потокосцеплението необходимо за ориентация на координатната система, трябва да се изчислява и роторната скорост. Изчислената роторна скорост е означена с $\hat{w}_r$.

$$\hat{A} = \begin{bmatrix} -\frac{(R_s + R_r \frac{L_m^2}{L_r^2})}{L_s s} & 0 & \frac{L_m R_r}{L_s s L_r} & -\frac{L_m}{L_s s L_r} \hat{w}_r \\ 0 & -\frac{(R_s + R_r \frac{L_m^2}{L_r^2})}{L_s s} & \frac{L_m}{L_s s L_r} \hat{w}_r & \frac{L_m R_r}{L_s s L_r} \\ \frac{L_m R_r}{s L_s} & 0 & -\frac{R_r}{L_r} & \hat{w}_r \\ 0 & \frac{L_m R_r}{s L_s} & -\hat{w}_r & -\frac{R_r}{L_r} \end{bmatrix} \quad (9)$$

Скоростния, адаптивен по поток наблюдател, позволява изчисляване на неизвестният скоростен параметър $\hat{w}_r$ в матрицата $\hat{A}$. За получаване на скоростния алгоритъм е използвана теоремата на Ляпунов.



Фиг. 2. Структурна схема на адаптивен по скорост и статорно съпротивление наблюдател на потокосцепление

По отношение на грешката уравнението на състоянието може да се изрази така:

$$\frac{de(t)}{dt} = \frac{d}{dt}(x - \hat{x}) = (A - GC)(x - \hat{x}) + (A - \hat{A})\hat{x} = (A - GC)e - \Delta A \hat{x} \quad (10)$$

където:

$$\Delta A = A - \hat{A} = \begin{bmatrix} 0 & 0 & 0 & -(\hat{w}_r - w_r)s \frac{L_m}{L_s L_r} \\ 0 & 0 & (\hat{w}_r - w_r)s \frac{L_m}{L_s L_r} & 0 \\ 0 & 0 & 0 & (\hat{w}_r - w_r) \\ 0 & 0 & -(\hat{w}_r - w_r) & 0 \end{bmatrix}$$

$e = x - \hat{x}$ е вектора на изчислителната грешка

Динамиката на уравнение (10) се определя от собствените стойности на матрицата $A - GC$. Ако тези собствени стойности имат отрицателни реални части, тогава изчисленото $\hat{x}$ се доближава до действителното x .

За постигане на устойчивост, адаптивния механизъм е получен чрез динамичните уравнения на състоянието и теоремата за устойчивост на Ляпунов.

Функцията:

$$V = e^T e + \frac{1}{2}(\hat{w}_r - w_r)^2 / c \quad (11)$$

където $c > 0 = \text{const}$

удовлетворява изискванията за непрекъснатост, диференцируемост и е положително определена.

Функцията е равна на 0 когато (e) е нула и когато изчислената скорост $(\hat{w}_r)$ е равна на действителната скорост (w_r) . Достатъчното условие за асимптотична устойчивост е производната dV/dt , да бъде отрицателно дефинирана.

$$\begin{aligned} \frac{dV}{dt} = e^T \left[\frac{d(e^T)}{dt} \right] + e^T \left[\frac{de}{dt} \right] + 2 \frac{d\hat{w}_r}{dt} \frac{(\hat{w}_r - w_r)}{c} = e^T [(A - GC)^T + (A - GC)] e - \\ - 2 \frac{L_m}{L_r} (\hat{w}_r - w_r) \frac{(e_{sa} y_{sb} - e_{sb} y_{sa})}{s L_s} + 2 \frac{d\hat{w}_r}{dt} \frac{(\hat{w}_r - w_r)}{c} \end{aligned} \quad (12)$$

където $e_{sa} = i_{sa} - \hat{i}_{sa}$; $e_{sb} = i_{sb} - \hat{i}_{sb}$

Условието за асимптотична устойчивост се изпълнява, ако $\frac{dV}{dt} < 0$ т.е. намалява ако $(\hat{w}_r - w_r) > 0$. Последното е вярно, ако сумата на последните два члена е нула.

$$\frac{d\hat{w}_r}{dt} = K(e_{sa} \hat{y}_{sb}^s - e_{sb} \hat{y}_{sa}^s), \quad K = c \frac{L_m}{s L_s L_r} \quad (13)$$

За подобряване динамиката на скоростния наблюдател е приложен PI закон:

$$\hat{w}_r = K_p(e_{sa} \hat{y}_{rb}^s - e_{sb} \hat{y}_{ra}^s) + K_i \int (e_{sa} \hat{y}_{rb}^s - e_{sb} \hat{y}_{ra}^s) dt \quad (14)$$

където K_p и K_i са пропорционалната и интегралната константи.

Влиянието на промяната на статорното съпротивление може да бъде премахнато, чрез използване на адаптивен алгоритъм за оценката му, получен чрез теоремата на Ляпунов, аналогично на роторната скорост. В този случай функцията на Ляпунов V_2 е избрана

$$V_2 = e^T e + \frac{1}{2} (\hat{R}_s - R_s)^2 \frac{1}{c_2}, \quad \text{където } c_2 > 0 = \text{const} \quad (15)$$

В матрицата на състоянието на грешката $\Delta \hat{A}_s$, $\hat{R}_s = R_s + \Delta R_s$ е оценяваното статорно съпротивление, а ΔR_s е грешката в статорното съпротивление.

$$\hat{A} - A = \Delta A_R = \begin{bmatrix} \frac{\hat{R}_s - R_s}{L_s s} & 0 & 0 & 0 \\ 0 & \frac{\hat{R}_s - R_s}{L_s s} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (16)$$

$$\begin{aligned} \frac{dV_2}{dx} &= e^T (A - GC)^T e^T + e^T (A - GC) e - e^T \Delta A_R \hat{x} - e \Delta A_R^T \hat{x}^T + \frac{d\hat{R}_s}{dx} (\hat{R}_s - R_s) = \\ &= e^T \left[(A - GC)^T + (A - GC) \right] e - 2 \frac{\hat{R}_s - R_s}{L_s s} (e_{ia} \hat{i}_{sa} + e_{ib} \hat{i}_{sb}) + 2 \frac{d\hat{R}_s}{dx} \frac{(\hat{R}_s - R_s)}{c_2} \end{aligned} \quad (17)$$

Условието за асимптотична устойчивост се изпълнява, ако $\frac{dV_2}{dt} < 0$ т.е. намалява ако $(\hat{R}_s - R_s) > 0$.

$$\frac{d\hat{R}_s}{dx} = \frac{c_2}{L_s s} (e_{ia} \hat{i}_{sa} + e_{ib} \hat{i}_{sb}) \quad (18)$$

Уравненията на двигателя (4) - (7) трябва да бъдат дискретизирани за цифрова реализация на алгоритъма:

$$x(t_{k+1}) = F(t_{k+1}, t_k) x(t_k) + \int_{t_k}^{t_{k+1}} F(t_{k+1}, t) B(t) u(t) dt \quad (19)$$

$$F(t_{k+1}, t_k) = e^{A(t_{k+1})(t_{k+1}-t_k)} = I + A(t_{k+1})(t_{k+1}-t_k) + A(t_{k+1})^2 \frac{(t_{k+1}-t_k)^2}{2!} + \dots \quad (20)$$

$$B(t_{k+1}) = B(t_{k+1}-t_k) + \frac{A(t_{k+1})B(t_{k+1}-t_k)^2}{2!} + \dots \quad (21)$$

Дискретизираният наблюдател се описва с:

$$\hat{x}(t_{k+1}) = F(t_{k+1}, t_k) \hat{x}(t_k) + B(t_{k+1}) u(t_{k+1}) + G(t_{k+1}) e(t_{k+1}) \quad (22)$$

където G е дискретизираната матрица в обратната връзка.

Тъй като матрицата A зависи от параметъра w_r , то получената дискретна преходна матрица $F(t_{k+1}, t_k)$ трябва да се изчислява на всяка стъпка. За изпълнение в реално време, обновяването на матрицата на обратната връзка, поради малката стъпка на дискретизация налага изисквания за сериозно бързодействие и изчислителни възможности на микроконтролера. Неблагоприятно влияние върху динамиката на наблюдателя, особено в

нискоскоростния диапазон оказват изчислителните грешки, поради близкото разположение на полюсите до границата на устойчивост.

За постигане на адекватна точност на дискретизирания модел е необходимо времето на дискретизация да бъде на порядък по-малък от характерните времеkonстанти на двигателя. Крайният избор се базира на полученото време за изпълнение на пълния алгоритъм, както и задоволителна точност и стабилност. Използване на два члена от реда очевидно увеличава времето за изчисление. Ако се пренебрегне вторият член в (19) и (21), тогава:

$$F(t_{k+1}, t_k) = e^{AT} \approx I + AT \quad (23)$$

$$B(t_{k+1}) = \int_0^T e^{Ax} B dx \approx BT \quad (24)$$

$$C(t_{k+1}) = C \quad (25)$$

където $T = t(k+1) - t(k)$ е стъпката на дискретизация.

$$F(t_{k+1}, t_k) = \begin{bmatrix} 1 - \frac{(R_s + R_r \frac{L_m^2}{L_r^2})}{L_s s} T & 0 & \frac{L_m R_r}{L_s s L_r} T & -\frac{L_m}{L_s s L_r} T \hat{w}_r \\ 0 & 1 - \frac{(R_s + R_r \frac{L_m^2}{L_r^2})}{L_s s} T & \frac{L_m}{L_s s L_r} T \hat{w}_r & \frac{L_m R_r}{L_s s L_r} T \\ \frac{L_m R_r}{s L_s} T & 0 & 1 - \frac{R_r}{L_r} T & -T \hat{w}_r \\ 0 & \frac{L_m R_r}{s L_s} T & \hat{w}_r T & 1 - \frac{R_r}{L_r} T \end{bmatrix} \quad (26)$$

$$B(t_{k+1}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{и} \quad C(t_{k+1}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (27)$$

$$x(k) = \begin{bmatrix} i_{sa}^s(k) & i_{sb}^s(k) & y_{ra}^s(k) & y_{rb}^s(k) \end{bmatrix}^T \quad (28)$$

$$u(k) = \begin{bmatrix} V_{sa}^s(k) & V_{sb}^s(k) \end{bmatrix}^T \quad (29)$$

За да се осигури устойчивост на наблюдателя в целия диапазон на изменение на скоростта, конвенционалният подход е да се изберат полюсите му пропорционални на полюсите на мотора. Константата на пропорционалност е $k \geq 1$, което прави наблюдателя динамично по-бърз от асинхронната машина. Но чувствителността към шума е свързано с константата на пропорционалност.

4. Синтез на рекурсивен алгоритъм по метода на най-малките квадрати

Измервателните и системни грешки е възможно да се окажат причина за несъвместност на решенията. В такъв случай се налага да се приеме оценката, която по най-добър начин удовлетворява системата уравнения. Един начин за решаване на тази задача е класическият метод на Гаус, известен като метод на

най-малките квадрати. Целта е минимизация на квадрата на евклидовата норма на вектора на грешката[4]

$$\|e\|_2^2 = e^T e = e_1^2 + e_2^2 + \dots + e_k^2 \quad (30)$$

Условието за достигане на минимум на тази сума е нулиране на всички частни производни или

$$\begin{bmatrix} F^T(k_0, k_1)C^T(k_0) \dots F^T(k_1, k_1)C^T(k_1) \end{bmatrix} \begin{bmatrix} y(k_0) \\ \vdots \\ y(k_1) \end{bmatrix} = \begin{bmatrix} F^T(k_0, k_1)C^T(k_0) \dots F^T(k_1, k_1)C^T(k_1) \end{bmatrix} \begin{bmatrix} C(k_0)F(k_0, k_1) \\ \vdots \\ C(k_1)F(k_1, k) \end{bmatrix} \hat{x}(k_1) \quad (31)$$

Дясната част на уравнението (31) може да бъде представена чрез дискретната матрица на наблюдаемост $M(k_1, k_0)$:

$$M(k_1, k_0) := \sum_{k=k_0}^{k_1} F^T(k, k_1)C^T(k)C(k)F(k, k_1) \quad (32)$$

Въвежда се допълнителната матрица P^* , чиито физически смисъл е ковариации на оценките.

$$P^*(k+1) := F(k)M^{-1}(k, k_0)F^T(k) \quad (33)$$

Окончателно, за матрицата в обратната връзка на наблюдателя се получава:

$$G(k+1) = P^*(k+1)C^T(k+1)[C(k+1)P^*(k+1)C^T(k+1) + I]^{-1} \quad (34)$$

а за M^{-1} изразът е:

$$M^{-1}(k+1, k_0) = P^*(k+1) - G(k+1)C(k+1)P^*(k+1) \quad (35)$$

Ако матрицата $M^{-1}(k, k_0)$ за определен момент от времето k е известна, то за следващия момент на измерване матрицата P^* се изчислява по уравнение (35) с което всички матрици в дясната част на (34) са определени и G за следващия дискретен момент също може да бъде определена. Уравнението за M^{-1} съответства на матричното диференциално уравнение на Рикати.

Начални условия.

Наблюденията започват в k_0 момент с осъществяване на първото измерване $y(k_0)$. Ако матрицата $M(k, k_0)$ е изродена, то изчисленията трябва да бъдат извършени по:

$$M(k+1, k_0) = F^{-1}(k)^T M(k, k_0)F^{-1}(k) + C^T(k+1)C(k+1)$$

$$M(k+1, k_0)\hat{x}(k+1) = F^{-1}(k)^T M(k, k_0)\hat{x}(k) + C^T(k+1)y(k+1)$$

с начални условия $M(k_0, k_0)\hat{x}(k+1) = C^T(k_0)y(k_0)$

В противен случай началните условия са $M(k_0, k_0) = C^T(k_0)C(k_0)$

Роторната скорост и статорното съпротивление са приети за неизвестни параметри и тъй-като дискретната преходна матрица зависи от тях е използван синтезирания по-горе механизъм за оценката им. На всяка стъпка дискретната преходна матрица се обновява и по уравнение (34) се изчислява актуалната

стойност на матрицата в обратната връзка на наблюдателя, осигуряваща минимум на квадрата на грешката на измеримите и оценени състояния.

Предсказване на вектора на състоянието:

$$\hat{x}(k/k-1) = F(k-1)\hat{x}(k-1/k-1) + B(k-1)u(k-1) \quad (36)$$

Ковариационна оценка на предсказването:

$$P^*(k) = F(k-1)M^{-1}(k-1)F^T(k-1) \quad (37)$$

Изчисляване матрицата в обратната връзка

$$G(k) = P^*(k)C^T(k)[C(k)P^*(k)C^T(k) + I]^{-1} \quad (38)$$

Обновяване на матрицата на наблюдаемост

$$M^{-1}(k/k+1) = [I - G(k)C(k)]P^*(k) \quad (39)$$

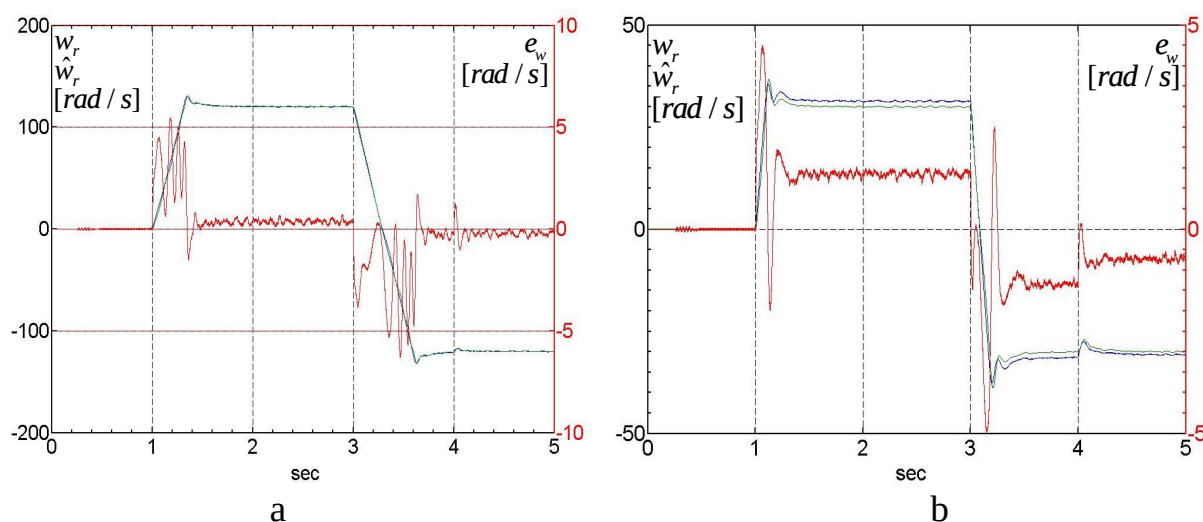
Оценка на вектора на състоянието:

$$\hat{x}(k/k) = \hat{x}(k/k-1) + L(k)[y(k) - C(k)\hat{x}(k/k-1)] \quad (40)$$

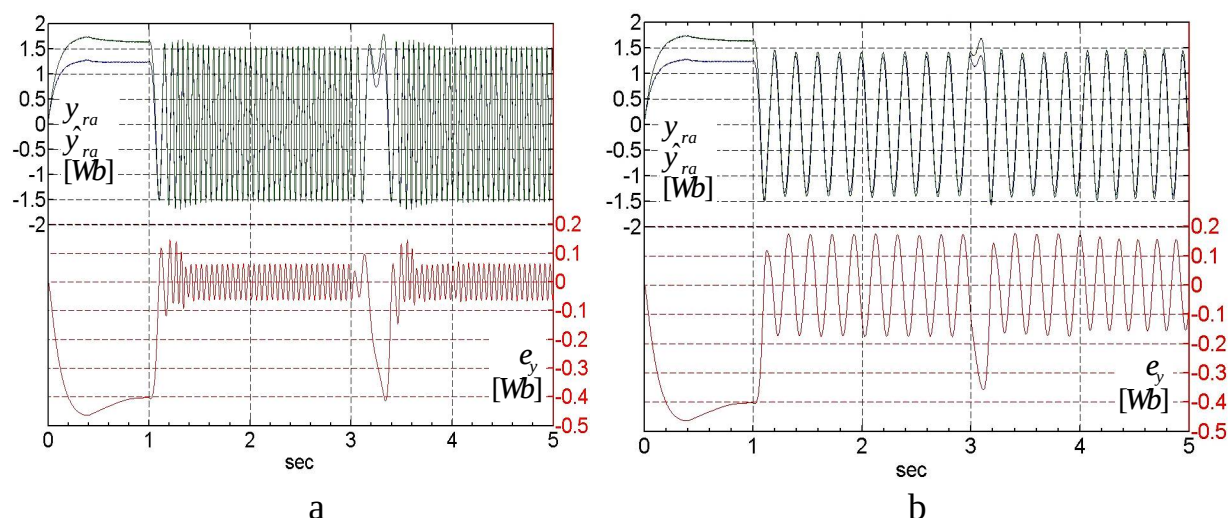
Табл. 1. Рекурсивен алгоритъм по метода на най-малките квадрати

5. Симуляционно изследване за оценка на показателите за качество на наблюдателя

Симулациите са проведени с Matlab/Simulink, при 20% увеличение на параметрите R_s и R_r на обекта спрямо модела, използван в наблюдателя съответстващо на реалните условия при нагряване в система за индиректно векторно управление с идеализиран инвертор на ток (без отчитане на комутационните процеси) с релейни хистерезисни регулатори и отрицателна обратна връзка по оценената роторна скорост.



Фиг. 3. Роторна скорост и оценката ѝ при задания: 200 rad/s (a) и 30 rad/s (b)



Фиг. 4. Роторно потоксцепление и оценката му при задания:
200 rad/s (a) и 30 rad/s (b)

6. Заключение

В сравнение с класическия Луенбъргъров наблюдател, наблюдателят синтезиран по метода на най-малките квадрати показва около 2 пъти по-малка динамична грешка, 7 пъти по-малка статична грешка при натоварване в оценките на скоростта и 2 пъти по-малка грешка в потокосцеплението в диапазона на високите скорости. При ниска скорост грешките са около 1.4 пъти по-малки при равни други условия. По високата точност във високоскоростния диапазон се обяснява с изчисляване на матрицата в обратната връзка на всяка стъпка, съгласно алгоритъма така, че да се минимизира квадрата на грешката, докато при наблюдателя на Луенбъргър същата е стационарна, определена при ниска скорост.

Благодарности. Настоящите изследвания са финансирани от ФНИ - МОМН по договор ДТК 02/1 от 2009 година.

ЛИТЕРАТУРА

- [1] Vas, P. "Sensorless Vector and Direct Torque Control" New York: Oxford University Press, (1998)
- [2] Jansen, P. and R. Lorenz "Observer based direct field orientation: Analysis and comparison of alternative methods", IEEE Tran. IA, vol. 30, no.4, pp.945-953, Jul/Aug, 1994
- [3] Schauder, C. "Adaptive Speed Identification for Vector Control of Induction Motor without Rotational Transducers" IEEE Tran. IA vol. 28, no. 5, pp. 1054-1061, Oct. 1992
- [4] Браммер, К., Г. Зиффлинг. "Филтър Калмана-Бьюси. Детерминированное наблюдение и стохастическая фильтрация", Москва, "Наука", 1982

Автор: гл. ас. Евтим Йончев, кат. "Автоматизация на електрозадвижванията".

Постъпила на 12.05.2010

Рецензент доц. д-р Борис Борисов

СОФТУЕРНА РЕАЛИЗАЦИЯ НА ОСНОВНИТЕ МАТЕМАТИЧЕСКИ ДЕЙСТВИЯ И ФУНКЦИИ ПОСРЕДСТВОМ ОБОБЩЕНИ ИНТЕРВАЛИ

Симона Филипова

Резюме. В настоящата статия е извършена софтуерна реализация на съществуващи в литературата методи за представяне на основните аритметични действия и елементарните математически функции чрез използване на обобщени интервали.

Методите са приложими за линеаризация на произволна нелинейна n -мерна функция $f(x)$ с интервален аргумент до афинна форма. Тази апроксимация е приложима за непрекъснато диференцируеми, и дори почасти непрекъснати многомерни функции с интервален аргумент.

В края на статията са предложени процедури за линеаризация на основните аритметични действия и елементарните математически функции, включени в програмно решение за линеаризация на система от 4 уравнения с 4 неизвестни.

SOFTWARE REALIZATION OF THE BASIC MATHEMATICAL OPERATIONS AND FUNCTIONS USING GENERALIZED INTERVALS

Simona Filipova

Abstract. In this paper, software realization of the existing methods about presentation the basic mathematical operations and functions using generalized intervals is discussed.

The methods consist of linearization of the arbitrary nonlinear n -dimensional function $f(x)$ with interval argument to affine form. This approximation can be applied for the continuous differential and piece-continuous differentiable multidimensional functions with interval argument.

Program solution for linearization of non-linear system consists of 4 equations with 4 variables is suggested in the end of the paper. It includes procedures for linearization of the basic mathematical operations and functions.

1. Introduction

Enclosing a real function $f(x)$ in a given interval argument x by an interval function $\mathbf{f}(x)$ is one of the fundamental problems in interval analysis.

Note: Here and henceforth, ordinary font letters will denote real quantities while bold face letters will stand for their interval counterparts.

Various enclosures based on the standard interval arithmetic have been proposed over the years depending on whether the argument x is interval or interval vector, on the one hand, and on the type of interval function $\mathbf{f}(x)$ used, on the other [1, 2].

Chebyshev affine approximation is valid for one dimensional function $f(x)$. But it is more precisely because of the fact that it guarantees closer interval enclosure than the above approximations [3]. It based on the linearization of the univariate functions with the following property:

Theorem 1: Let $f(x)$ be a bounded and continuous function in some real interval $x = [\underline{x}, \bar{x}]$, $x \in x$. Let h be the affine function that best approximate $f(x)$ in this interval under the minimax error criterion. Then, there exist three distinct points in interval $x = [\underline{x}, \bar{x}]$ where the error $d = f(x) - h(x)$ has maximum magnitude and the sign of the error alternates when the three points are considered in increasing order.

This theorem provides an algorithm for finding the optimal approximation in many cases, via the following corollary:

Corollary 1: Let $f(x)$ be a bounded and twice differentiable function defined on some interval $x = [\underline{x}, \bar{x}]$, whose second derivative $f''(x)$ does not change sign inside x . Let $f^a(x) = ax + b$, $x \in x$ be its minimax affine approximation in $x = [\underline{x}, \bar{x}]$. Then (Fig. 1):

- The coefficient $a = \frac{f(\bar{x}) - f(\underline{x})}{\bar{x} - \underline{x}}$ is the slope of the line $r(x)$ that interpolates the points $(\underline{x}, f(\underline{x}))$ and $(\bar{x}, f(\bar{x}))$.
- The maximum absolute error will occur twice (with the same sign) at the end points $\underline{x}$ and $\bar{x}$ of the range, and once (with the opposite sign) at the every interior point u in $x = [\underline{x}, \bar{x}]$ where $f'(u) = a$.
- The independent term b is such that $au + b = (f(u) + r(u))/2$ and the maximum absolute error is $d = |f(u) - r(u)|/2$.

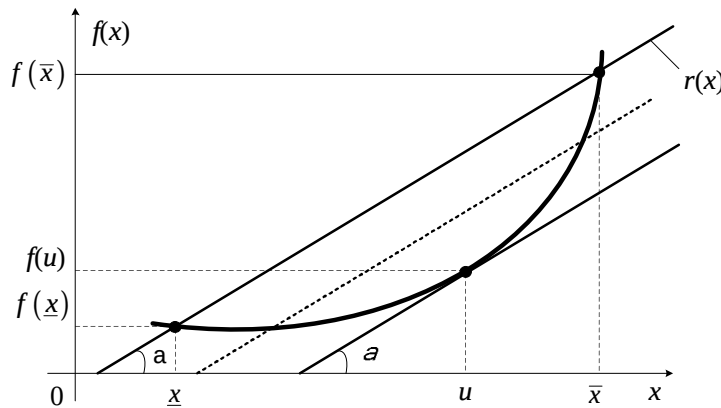


Fig. 1

2. Problem statement

Definition 1: A generalized (G) interval $\tilde{x}$ of length m is defined as follows

$$x = x_0 + \sum_{i=1}^m x_i e_i, \quad (1)$$

where x_i , $i = 0, 1, \dots, m$, are real numbers while e_i are unit symmetrical intervals, i.e.

$$e_i = [-1, 1]. \quad (1a)$$

The notation (1) is called affine form of interval presentation. Then the relation between this form and standard presentation of the interval

$$x = [\underline{x}, \bar{x}] = x_0 + [-r_x, r_x] \quad (2)$$

is following:

$$r_x = \sum_{i=1}^m |x_i| \Rightarrow \begin{cases} \underline{x} = x_0 - r_x \\ \bar{x} = x_0 + r_x \end{cases}. \quad (2a)$$

The affine form of presentation of the intervals will be used in the next sections of the paper.

3. Realization of the basic mathematical operations by affine form of the intervals

Let consider two intervals presented in the following affine form:

$$x = x_0 + \sum_{i=1}^{m_x} x_i e_i \quad (3a)$$

and

$$y = y_0 + \sum_{i=1}^{m_y} y_i e_i. \quad (3b)$$

Note: To simplify the presentation, we stand that $m_x = m_y = m$ (otherwise, we take $m = \max(m_x, m_y)$ and add zero components either in x or in y depending on whether m_x is smaller or larger than m_y).

Then we will present in brief the method for linear combination, multiplication and division of G-intervals in affine form (1). This method was suggested in [4].

3.1. Linear combination

Linear combination of the G-intervals x and y is another G-interval

$$z = z_0 + \sum_{i=1}^m z_i e_i = a x + b y, \quad (4)$$

where

$$\begin{aligned} z_0 &= a x_0 + b y_0 \\ z_i &= a x_i + b y_i, \quad i=1,2,\dots,m \end{aligned} \quad (4a)$$

Corollary 1: The addition of two G-intervals x and y gets from (4) in $a = b = 1$.

Corollary 2: The subtraction of two G-intervals x and y gets from (4) in $a = 1, b = -1$.

Note: After linear combination the length of initial G-intervals x and y and the resulting G-interval z is the same – m .

3.2. Multiplication

The multiplication product

$$z = z_0 + \sum_{i=1}^{m+1} z_i e_i = x \cdot y, \quad (5)$$

where

$$u = \sum_{i=1}^m |x_i|, \quad v = \sum_{i=1}^m |y_i|, \quad c = 0.5 \sum_{i=1}^m x_i y_i \quad (5a)$$

$$\begin{aligned} z_0 &= x_0 y_0 + c \\ z_i &= x_0 y_i + y_0 x_i, \quad i=1,\dots,m \\ z_{m+1} &= uv - |c| \end{aligned} \quad (5b)$$

Note: After multiplication the length of the resulting G-interval z increases with 1 with respect to the length of the initial G-intervals x and y .

3.3. Division

To define the operation division we have to introduce operation reciprocal $z = \frac{1}{x}$ in affine form as follows:

$$z = z_0 + \sum_{i=1}^{m+1} z_i e_i = \frac{1}{x}, \quad 0 \notin x, \quad (6)$$

where

$$s = -1 / (\underline{x} \cdot \bar{x}), \quad x_1 = -\sqrt{-1/s}, \quad x_2 = -x_1 \quad (6a)$$

$$x_s = \begin{cases} x_2, & \text{if } \underline{x} > 0 \\ x_1, & \text{if } \bar{x} < 0 \end{cases} \quad (6b)$$

$$\underline{f} = 1 / x_s - s x_s, \quad \bar{f} = 1 / \underline{x} - s \underline{x} \quad (6c)$$

$$f_0 = 0.5 (\underline{f} + \bar{f}), \quad r_f = \bar{f} - f_0 \quad (6d)$$

$$\begin{aligned}
z_0 &= s.x_0 + f_0 \\
z_i &= s.x_i, \quad i=1,...,m \\
z_{m+1} &= r_f
\end{aligned} \tag{6e}$$

Note: $\underline{x}$ and $\bar{x}$ are the endpoints of the interval x according to (2).

Note: After reciprocal operation the length of the resulting G-interval z increases with 1 with respect to the length of the initial G-intervals x and y .

The resulting G-interval $z = \frac{y}{x}$, $0 \notin x$ after division operation gets on two stages.

First, the reciprocal $z_1 = \frac{1}{x}$, $0 \notin x$ calculates. Next, the multiplication applies for the

product $y.z_1 = y.\frac{1}{x} = \frac{y}{x} = z$, $0 \notin x$. Hence, it is seen that the division operation increases the length of the resulting G-interval z by two because of the reciprocal (6e) and multiplication (5b).

4. Realization of the main mathematical functions affine form of the intervals

Let consider one dimensional nonlinear continuous or piecewise-continuous differentiable function $f(x)$, $x \in R^1$. The linear approximation of this function

$$L(x) = ax + b, \quad x \in x \subset R^1, \tag{7}$$

where: a – real number;

$b = [\underline{b}, \bar{b}]$ – interval vector.

was suggested for the first time in [5, 6] for the cases of continuous and piecewise-continuous differentiable nonlinear functions shown on Fig. 2a and Fig. 2b, respectively:

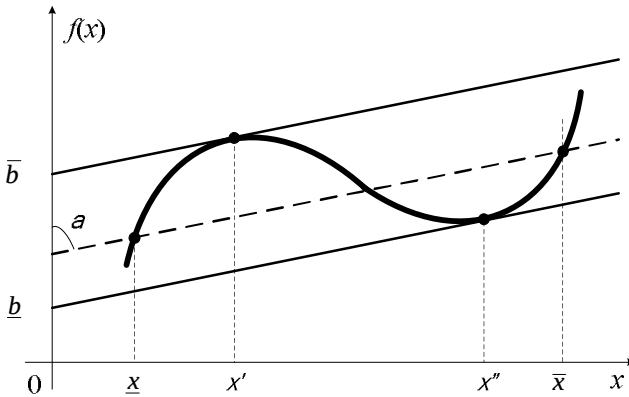


Fig. 2a

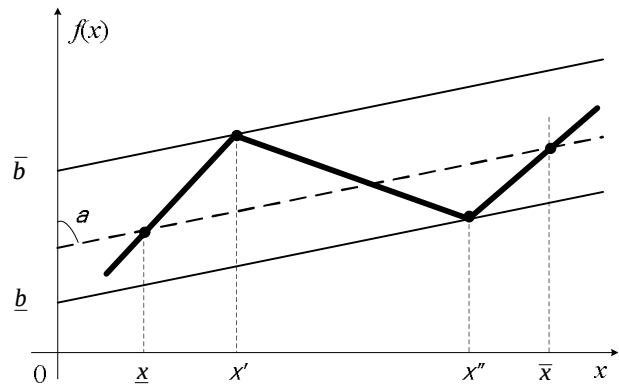


Fig. 2b

Note: Without loosing the generality of the problem analysis, the linearization method from [6] is applicable for family of n -dimensional non-linear functions $f_i(x) = \sum_j f_{ij}(x)$, $x \in X \subset R^n$.

The main points of the linearization (7) will be briefly presented here. Let consider the continuous function $f(x)$ shown on Fig. 3.

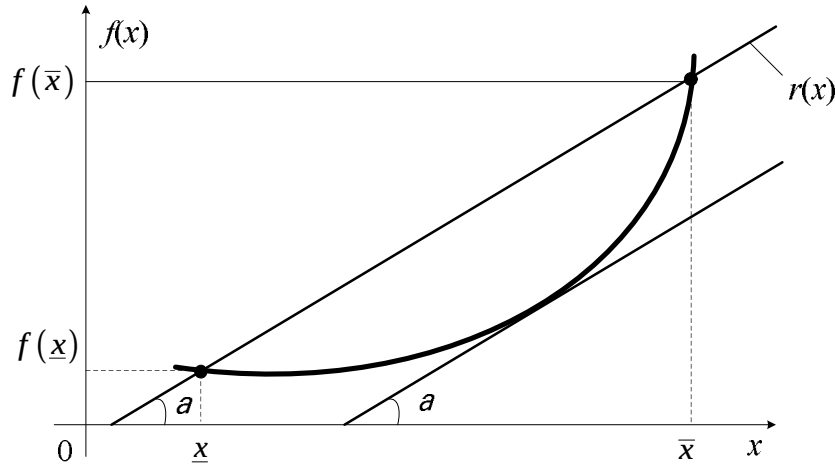


Fig. 3

Then, the procedure for determining the coefficients a and $b = [\underline{b}, \bar{b}]$ consists of the following steps.

1. Calculating the slope a of the linear line $r(x)$ as:

$$a = \frac{f(\bar{x}) - f(x)}{\bar{x} - x}. \quad (8)$$

2. Finding all real solutions x_k , $k = 1, 2, \dots, s$ of the equation

$$\frac{d(f(x))}{dx} - a = 0. \quad (9)$$

3. Using x_k and $x_0 = x$ and $x_{s+1} = \bar{x}$, the following quantities are computed

$$b_k = f(x_k) - a \cdot x_k, \quad k = 0, 1, \dots, (s+1) \quad (10)$$

4. Determining the low and the upper bounds of the interval $b = [\underline{b}, \bar{b}]$

$$\underline{b} = \min_k (b_k), \quad \bar{b} = \max_k (b_k), \quad k = 0, 1, \dots, (s+1). \quad (11)$$

The efficiency of this procedure depends principally on how much computational effort is needed to locate all real solutions of (9) in X . If the latter problem is easy to solve, then the procedure provides a simple and efficient way of determining the linear approximation (7). (Otherwise, it gets less accurate, i.e. broader but cheaper approximations are possible.

Software realization of linearization procedure uses affine form of presentation of the interval x according to formulas (1) and (2) from Section 2, i.e.

$$\mathbf{x} = x_0 + \sum_{i=1}^m x_i e_i, \quad e_i = [-1, 1]$$

$$\mathbf{x} = [\underline{x}, \bar{x}] = x_0 + [-r_x, r_x], \quad r_x = \sum_{i=1}^m |x_i| \Rightarrow \begin{cases} \underline{x} = x_0 - r_x \\ \bar{x} = x_0 + r_x \end{cases}.$$

5. Software realization

Programs according to the linearization's methods mentioned above are written in two aspects. First, procedures realizing a main mathematical operations concerning to theory from Section 3, are suggested. Second, a main program and the respectively subprograms transforming non-linear system consist of 4 non-linear equations to the linear one according to Section 4, are proposed.

5.1. Procedures for linearization the main mathematical operations

In this paper, four procedures are suggested. Each of them returns a result in affine form, i.e. G-interval:

$$\mathbf{z} = z_0 + \sum_{i=1}^m z_i e_i, \quad \text{where } e_i = [-1, 1], \quad (12)$$

Note: Transformation between G-interval and standard interval presentation is given in Section 2 of the paper.

Addition	Subtraction
<pre>function z = addl(x,y) nx = numel(x); ny = numel(y); if nx<m s = m-nx; x2 = zeros(1, s); x = [x x2]; end if ny<m s = m-ny; y2 = zeros(1, s); y = [y y2]; end z = x+y;</pre>	<pre>function z = subtrl(x,y) nx = numel(x); ny = numel(y); if nx<m s = m-nx; x2 = zeros(1, s); x = [x x2]; end if ny<m s = m-ny; y2 = zeros(1, s); y = [y y2]; end y(1) = -y(1); z = addl(x,y);</pre>
Multiplication	Reciprocal operation
<pre>function z = multl(x,y) nx = numel(x); ny = numel(y); if nx<m s = m-nx; x2 = zeros(1, s); x = [x x2]; end</pre>	<pre>function z = recl(x) nx = numel(x); if nx<m s = m-nx; x2 = zeros(1, s); x = [x x2]; end</pre>

<pre> if ny<m s = m-ny; y2 = zeros(1, s); y = [y y2]; end c1 = x(1)*y(1); xa = x(2:m); ya = y(2:m); c2 = 0.5*xa*ya'; z0 = c1 + c2; zi = x(1)*ya + y(1)*xa; u = sum (abs(xa)); v = sum (abs(ya)); znmax1 = u*v - abs(c2); z = [z0 zi znmax1]; m = m+1; </pre>	<pre> rx = sum(abs(x(2:nx))); xd = x(1)-rx; xu = x(1)+rx; eps1=1e-8; dx = xu - xd; if dx < eps1 ('Division by zero interval. Linearization is impossible'), z1 = 1/x(1); numx = nx-1; z2 = zeros(1, numx); z = [z1 z2]; return end p = 1/xd; a = - p/xu; t1 = 1/sqrt(-a); t2 = -1/sqrt(-a); if (xd > 0) s = t1; else s = t2; end bu = p - a*xd; bd = 1/s - a*s; b0 = 0.5*(bd + bu); rb = 0.5*(bu - bd); z = a*x; z(1)=z(1)+b0; z = [z rb]; m = m+1; </pre>
--	--

Each of above mentioned program modules gives an output $m = m_z$ -dimensional vector, which first component associates to z_0 in (12) and the other $(m_z - 1)$ elements are related to the coefficients z_i in the sum of (12).

If suppose that m_x, m_y and m_z are the lengths of the initial G-intervals x and y and the resulting G-interval z , respectively, then it is obvious the following relations after applying the main mathematical functions:

1. **Addition** ($z = x + y$)

$$m_z = \max(m_x, m_y) \quad (13a)$$

2. **Subtraction** ($z = x - y$)

$$m_z = \max(m_x, m_y) \quad (13b)$$

3. **Multiplication** ($z = x \cdot y$)

$$m_z = \max(m_x, m_y) + 1 \quad (13c)$$

4. **Reciprocal function** ($z = \frac{1}{x}, 0 \notin x$)

$$m_z = \max(m_x, m_y) + 1 \quad (13d)$$

5. **Division** ($z = \frac{x}{y}$, $0 \notin y$)

$$m_z = \max(m_x, m_y) + 2 \quad (13e)$$

5.2. Linearization of the 4-dimensional system of non-linear equations

Let consider the following 4-dimensional non-linear system:

$$\begin{cases} y_1 = \ln\left(\frac{x_1+1}{2}\right) + x_3 - 5 = 0 \\ y_2 = |x_1 - x_3| - x_2^2 + 5 = 0 \\ y_3 = \frac{x_3}{(x_1 + x_2 + 1)} + x_3 - 6 = 0 \\ y_4 = \tanh(x_3 - |x_4| + 2) - \arctg(x_4) - 7 = 0 \end{cases}, \quad (14)$$

where $\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{x}_3 = \mathbf{x}_4 = [0, 10^5]$. (14a)

Program realization of the linearization methods from Sections 3 and 4 includes main program and some functions which are called from the main program. The respective .m file associated main program is:

Main program
<pre>% n - number of interval variables Input data n=4; global m; xd=zeros(1,n); xg=100000.*ones(1,n); Xc=0.5*(xd+xg); R=0.5*(xg-xd); X = [Xc' diag(R)]; Determining the y₁ m=n+1; X100=X(1,:); X100(1,1)=X100(1,1)+1; X100=0.5*X100; X5=logl(X100); X5(2)=X5(2)/R(1); X5=X5; X200=X(3,:); X200(4)=X200(4)/R(3); X200=X200; X6=addl(X5, X200); X6(1)=X6(1)-5; y1=X6; Determining the y₂ X7=substrl(X(1,:), X(3,:)); X8=absl(X7); X8(2)=X8(2)/R(1); X8(4)=X8(4)/R(3); X8=X8; X9=step2l(X(2,:)); X9(3)=X9(3)/R(2); X9=X9; X10=substrl(X8, X9); X10(1)=X10(1)+5+t; y2=X10; Determining the y₃ X11=addl(X(1,:), X(2,:)); X11(1)=X11(1)+1; X12=divl(X(3,:), X11); X13=addl(X12, X(3,:)); X13(1)=X13(1)-6; X13(2)=X13(2)/R(1); X13(3)=X13(3)/R(2); X13(4)=X13(4)/R(3); y3=X13;</pre>

Determining the y_4

```

X14=absl(X(4,:));
X15=substrl(X(3,:), X14); X15(1)=X15(1)+2; X15=X15;
X16=tghl(X15); X16(4)=X16(4)/R(3); X16(5)=X16(5)/R(4); X16=X16;
X14(1)=X14(1)-7;
X17=arctgl(X14); X17(5)=X17(5)/R(4); X17=X17;
X18=substrl(X16, X17);
y4=X18;

h1=y1(2:(n+1)); h2=y2(2:(n+1)); h3=y3(2:(n+1)); h4=y4(2:(n+1)); H=[h1; h2; h3; h4];

b01=[y1(1) y2(1) y3(1) y4(1)]'; b02=H*Xc'; b0=b01-b02;

ny1=numel(y1); ny2=numel(y2); ny3=numel(y3); ny4=numel(y4);
yy=y1(n+2:ny1); rb1=sum(abs(yy));
yy=y2(n+2:ny2); rb2=sum(abs(yy));
yy=y3(n+2:ny3); rb3=sum(abs(yy));
yy=y4(n+2:ny4); rb4=sum(abs(yy));
rb=[rb1 rb2 rb3 rb4]';

```

Checking the correctives of linearization

Linearization system: $H \cdot X + B = 0$ ($H \cdot X + b_0 + [-rb, rb] = 0$)

```

pp=Xc;
f1=log(0.5*(pp(1)+1))+pp(3)-5;
f2=abs(pp(1)-pp(3))-pp(2)^2+5+t;
f3=pp(3)/(pp(1)+pp(2)+1)+pp(3)-6;
f4=tanh(pp(3)-abs(pp(4))+2)-atan(abs(pp(4))-7);
f=[f1 f2 f3 f4]';

```

```

bd1=b0-rb; bg1=b0+rb;
fd=H*pp'+bd1; fg=H*pp'+bg1;

```

```

eps=10e-12;
for i=1:n
    if fd(i)-f(i)>eps
        ('Wrong1 fd(i)'), i=i, fd(i)-f(i)
    elseif f(i)-fg(i)>eps
        ('Wrong2 fg(i)'), i=i, f(i)-fg(i)
    end
end

```

Final linear system: $H \cdot X = -b_0 + [-rb, rb]$

```

b0=-b0; bd=b0-rb; bg=b0+rb;

```

Procedures realizing the main mathematical operations (addition, subtraction, multiplication and division) are the same as the ones described in previous subsection. The additional necessary linearization procedures are following:

Function $\ln(x)$	Function $ x $
<pre> function z=logl(x) global m; nx=numel(x); if nx<m </pre>	<pre> function z=absl(x) global m; nx=numel(x); if nx<m </pre>

<pre> s = m-nx; x2 = zeros(1, s); x = [x x2]; end rx = sum(abs(x(2:nx))); xd = x(1)-rx; xu = x(1)+rx; eps2 = 1e-8; if xd<eps2 ('Function ln(x) is not defined'), return end p = xu/xd; dx = (xu - xd); eps1=1e-8; if dx<eps1 ('Division by zero interval. Linearization is impossible.') z1 = log(x(1)); numx = nx-1; z2 = zeros(1, numx); z = [z1 z2]; return end q = 1/dx; a = log(p)*q; t = 1/a; bd = log(xd)-a*xd; bu = log(t)-a*t; b0 = 0.5*(bd + bu); rb = 0.5*(bu - bd); z = a*x; z(1)=z(1)+b0; z = [z rb]; kk=log(la); kk1=1/kk; z=kk1*z; m = m+1; </pre>	<pre> s = m-nx; x2 = zeros(1, s); x = [x x2]; end rx = sum(abs(x(2:nx))); xd = x(1)-rx; xu=x(1)+rx; if xd>=0 a=1; bu = 0; bd = 0; elseif xu<=0 a = -1; bu = 0; bd = 0; else fd = abs(xd); fu = abs(xu); dx = xu - xd; eps1=1e-8; if dx<eps1 ('Division by zero interval. Linearization is impossible. '); z1 = abs(x(1)); numx = nx-1; z2 = zeros(1, numx); z = [z1 z2]; return end q = 1/dx; a = (fu -fd)*q; bu = fd-a*xd; bd = 0; end b0 = 0.5*(bd + bu); rb = 0.5*(bu - bd); z = a*x; z(1) = z(1)+b0; z = [z rb]; m = m+1; </pre>
Function $\tanh(x)$	Function $\arctg(x)$
<pre> function z = tghl(x) global m; nx = numel(x); if nx<m s = m-nx; x2 = zeros(1, s); x = [x x2]; end rx = sum(abs(x(2:nx))); xd = x(1)-rx; xu=x(1)+rx; fd = tanh(xd); fu = tanh(xu); </pre>	<pre> function z = arcctgl(x) global m; nx = numel(x); rx = sum(abs(x(2:nx))); xd = x(1)-rx; xu = x(1)+rx; if nx<m s = m-nx; x2 = zeros(1, s); x = [x x2]; end fu = acot(xu); fd = acot(xd); df = fu-fd; dx = xu-xd; </pre>

<pre> dx = xu - xd; eps1=1e-8; if dx<eps1 ('Division by zero interval. Linearization is impossible.'), z1 = tanh(x(1)); numx = nx-1; z2 = zeros(1, numx); z = [z1 z2]; return end q = 1/dx; a = (fu - fd)*q; a1 = sqrt(a); a2 = 1/a1; if a2<1 ('Linearization is impossible'), return end t1 = acosh(a2); t2 = -t1; bxd = tanh(xd)-a*xd; bt1 = tanh(t1)-a*t1; bt2 = tanh(t2)-a*t2; if t1>xd & t1<xu bd = min(bt1, bxd); bu = max(bt1, bxd); if t2>xd & t2<xu bd = min(bd, bt2); bu = max(bu, bt2); end elseif t2>xd & t2<xu bd = min(bxd, bt2); bu = max(bxd, bt2); end b0 = 0.5*(bd + bu); rb = 0.5*(bu - bd); z = a*x; z(1) = z(1)+b0; z = [z rb]; m = m+1; </pre>	<pre> eps1=1e-8; if dx<eps1 ('Division by zero interval. Linearization is impossible.'), z1 = acot(x(1)); numx = nx-1; z2 = zeros(1, numx); z = [z1 z2]; return end q = 1/dx; a = df*q; if a<(-1) ('Linearization is impossible'), return end a1 = (-1)/a; t1 = sqrt(a1 - 1); t2 = -t1; bxd = acot(xd)-a*xd; bt1 = acot(t1)-a*t1; bt2 = acot(t2)-a*t2; if t1>xd & t1<xu bd = min(bt1, bxd); bu = max(bt1, bxd); if t2>xd & t2<xu bd = min(bd, bt2); bu = max(bu, bt2); end elseif t2>xd & t2<xu bd = min(bxd, bt2); bu = max(bxd, bt2); end b0 = 0.5*(bd + bu); rb = 0.5*(bu - bd); z = a*x; z(1) = z(1)+b0; z = [z rb]; m = m+1; </pre>
--	--

The resulting linearized system is:

$$\begin{cases}
 y_1 = 230x_1 + 50x_3 - 26,2 \cdot 10^5 + [-1,1, 1,1] = 0 \\
 y_2 = -8,2x_1 - 5000x_2 + 16,3x_3 + 2495,9 \cdot 10^5 + [-1216,5, 1216,5] = 0 \\
 y_3 = -1,1x_1 - 2,2x_2 + 53,3x_3 - 25 \cdot 10^5 + [-55, 55] = 0 \\
 y_4 = 0,5x_3 - 2x_4 + 0,7 \cdot 10^5 + [-0,3, 0,3] = 0
 \end{cases} \quad (15)$$

6. Conclusion

Software realization of the existing method about linearization of the arbitrary n -dimensional nonlinear function $f(x)$ with interval argument x , presented in affine form, is suggested. After approximation the function $f(x)$ is presented in linear form

$L(x) = ax + b$, $x \in x \subset R^n$. The linearization method can be applied to the continuous and piecewise-continuous differentiable multidimensional functions.

In the paper, the procedures, realizing this linearization for all basic elementary mathematical functions (x^k , a^x (resp. e^x), $|x|$, $\sin(x)$, $\cos(x)$, $tg(x)$, $ctg(x)$, $\arcsin(x)$, $\arccos(x)$, $\arctgtg(x)$, $\arcctg(x)$, $\sinh(x)$, $\cosh(x)$, $tgh(x)$, $ctgh(x)$, $\arcsinh(x)$, $\arccosh(x)$, $\arctgh(x)$, $\arcctgh(x)$), are suggested. Additionally procedures, realizing the linearization of all main mathematical operations (addition, subtraction, multiplication and division), are proposed.

Based on the linearization procedures of the function x^k and the main mathematical elementary operations, the program for linearization of the polynomial function $p(x) = a_k x^k + a_{k-1} x^{k-1} + \dots + a_1 x + a_0$, is proposed.

All programs work in the Matlab environment, i.e. they are written as texts' files with extension .m.

REFERENCES

- [1] Moore, R. E. Methods and Applications of Interval Analysis, Philadelphia: SIAM, 1979.
- [2] Hansen, E. A Generalized Interval Arithmetic, in: Nickel, K. L. (ed.), Interval Mathematics, Springer Verlag, New York, 1975.
- [3] Figueiredo, L. H., J. Stolfi. Metodos Numericos Auto-Validados e Aplicacoes, 21 Coloquio Brasileiro de Matematica, IMPA 21-25 June, 1997, 116 p.
- [4] Kolev, L. "An Improved interval linearization for solving non-linear problems", Numerical Algorithms 37, 2004, pp. 213-224.
- [5] Kolev, L. A general interval method for global nonlinear dc analysis. Proc. of the 1997 European Conference on Circuit Theory and Design, ECCD'97, Technical University of Budapest, 30th August - 3rd Sept., 1997, Budapest, Hungary, volume 3, pp. 1460-1462.
- [6] Kolev, L. "A New Method for Global Solution of Systems of Non-Linear Equations", Reliable Computing 4, 1998, pp. 125-146.

Author: Simona Filipova-Petrakieva - Associate Professor in Technical University of Sofia, Department of Theoretical Electrical Engineering.

Received 12.10.2010

Reviewer Assoc. Prof. Valeri Mladenov

ОПТИМАЛНО ПРИЕМАНЕ В УСЛОВИЯТА НА АПРИОРНА НЕОПРЕДЕЛЕНОСТ НА КАНАЛА ЗА ВРЪЗКА

Вася Илиев, Михаил Момчеджиков

Резюме. В работата е предложен метод за адаптивна оптимална филтрация на DSSS-сигнали. Представен е обобщен модел на канала за връзка. Синтезирана е система за адаптивна филтрация на сигнала. Проведено е имитационно математическо моделиране. Показани са възможностите за реализация на системата с помощта на съвременна елементна база. В метода са залегнали алгоритмите за рекурентната процедура по преизчисляване на вектора на отчетите на импулсната характеристика на филтъра.

OPTIMAL RECEIVING IN CONDITIONS OF APRIORIC INDETERMINATENESS OF THE CHANNEL LINK

Vasya Iliev, Mikhail Momchedjиков

Abstract. An adaptive method for optimal DSS-signals filtration is suggested in this work. A general model of the channel link is introduced. A system for adaptive filtration of the signal is synthesized. Imitative mathematical modeling is realized. The options for the system realization with the contemporary elements are shown. The method is based on algorithms for recurrent procedure for re-estimating the vector of the readings of the pulse filter characteristic.

1. Въведение

Прилагането на методите за адаптивна обработка на сигналите и моделиране на параметрите на канала за свързка в задачата за съгласувано приемане позволява съществено да се повиши ефективността на предаваната информация, за сметка на допълнителни изчислителни операции и алгоритмична работа на системата. Съвременната техническа база на електронните компоненти позволява реализирането на такива системи с незначителни капиталовложения и значителни възможности за реализация. Разработването на методики за реализация на действащи системи и алгоритми за тяхната работа способства за построяването на системи работещи в условията на априорна неопределеност и позволява да бъде внесен принос в развитието на комуникационната техника.

2. Обобщен модел на канала за връзка

Съгласно съвременните теории като най съществени фактори, ограничаващи ефективността на комуникационните системи в реални условия, представляват маскиращите влияния на смущенията, действащи на приемната антена. Към източниците на електрически смущения се отнасят такива електромагнитни процеси, като шумовата температура на земята, атмосферни шумове, космически шумове, индустриални смущения, взаимни смущения, преднамерени смущения и др. На практика спектрално корелационните характеристики на полезния сигнал в една или друга степен се отличават от съответните характеристики на смущенията, което позволява селекцията им с помощта на съгласувани филтри.

Съгласуваният филтър от гледна точка на теорията за оптималното приемане [1] представлява пасивен четриполусник с предавателна характеристика от вида

$$M(j\omega) = (S(j\omega))^* / G(\omega)$$

където: $(S(j\omega))^*$ - спектър, комплексно спрегнат с полезния сигнал;

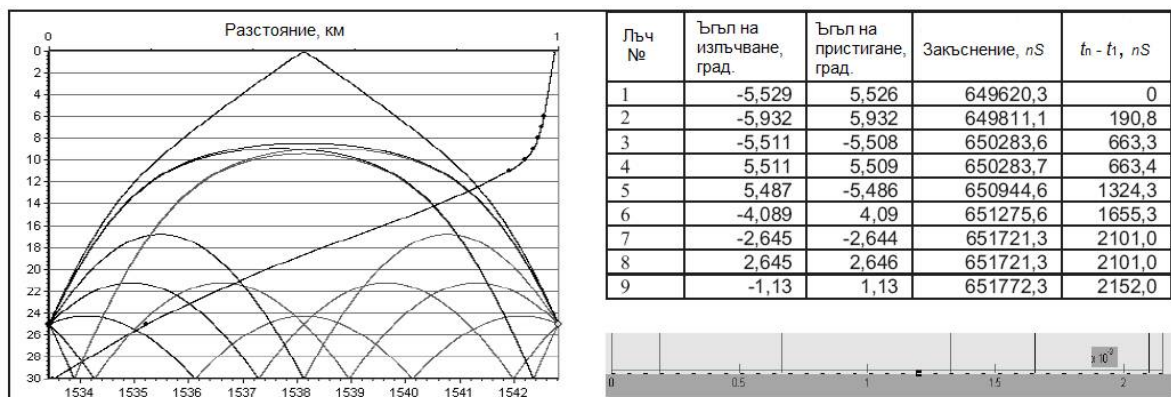
$G(\omega)$ - енергиен спектър на смущението в мястото на приемането

Прилагането на съгласувана филтрация позволява максимално да бъде повишено стойността на отношението сигнал/шум в изхода на системата, поради което при проектиране на комуникационни системи, като правило се поставя условието колкото е възможно по точно да бъде реализиран режим на съгласувана филтрация.

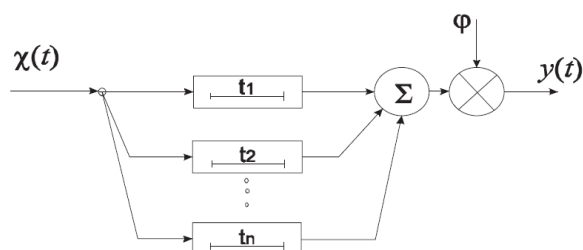
За да може приемният тракт на комуникационната система да притежава свойствата на съгласуван филтър е необходимо предавателната му функция да се описва с (1). Последното е възможно само ако в (1) и числителя и знаменателя са известни точно.

Параметрите на смущенията в болшинството от случаите като правило представляват бавно изменяща се във времето функция, което позволява тяхното регистриране и компенсацията им в реално време [2]. В реални условия на работа разпространението на полезния сигнал се явява многолъчево, поради което отделните лъчи на сигнала в мястото на приемане пристигат с различно закъснение (Фиг. 1), което влошава условията за съгласувано приемане.

Като се изходи от разгледаните по горе фактори, влияещи на преминаването на сигнала в средата за разпространение, каналът за връзка може да бъде представен с модела показан на фиг. 2, в който са отчетени влиянието както на адитивните и мултипликативни смущения, така и на ефектите от многолъчевото разпространение.



Фиг. 1. Закъснение на сигнала при многолъчево разпространение [3]



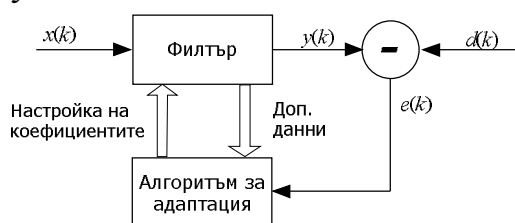
Фиг. 2. Обобщен модел на канал за връзка

3. Синтез на система за свързка с прилагане на адаптивна филтрация

Съгласно обобщеният модел от фиг. 2 задачата за съгласувана филтрация на сигнала в реални условия значително е усложнена и изисква специален подход за нейната реализация. Като един от компромисните варианти за нейното решение в този случай може да бъде посочен метода за широколентово предаване на информация по многолъчев канал с високо ниво на смущенията [1,4]. При този метод всеки информационен символ се предава с помощта на кодови поредици с подходящи корелационни свойства – кодови поредици на Баркер, Хафмен, Уолш, техните производни и др. На приемната страна се изчисляват корелационните функции на предадените кодови поредици с което те се декодират и се взема решение за вида на приетия информационен символ. По този начин приблизително се реализира условието за оптимално приемане изразено съгласно израз (1). На практика такъв вид системи се явяват особено ефективни при работа в условията на мощни адитивни смущения, но за сметка на съществено разширение на базата на сигнала. Ефектите от многолъчевото разпространение на сигнала при тях до голяма степен са потиснати, но при определени условия те могат да блокират работата им [1]. При съвместното използване на адаптивните и широколентови методи е възможно разделяне на лъчите на сигнала на приемната страна и практически пълно потискане на ефектите от многолъчевото разпространение.

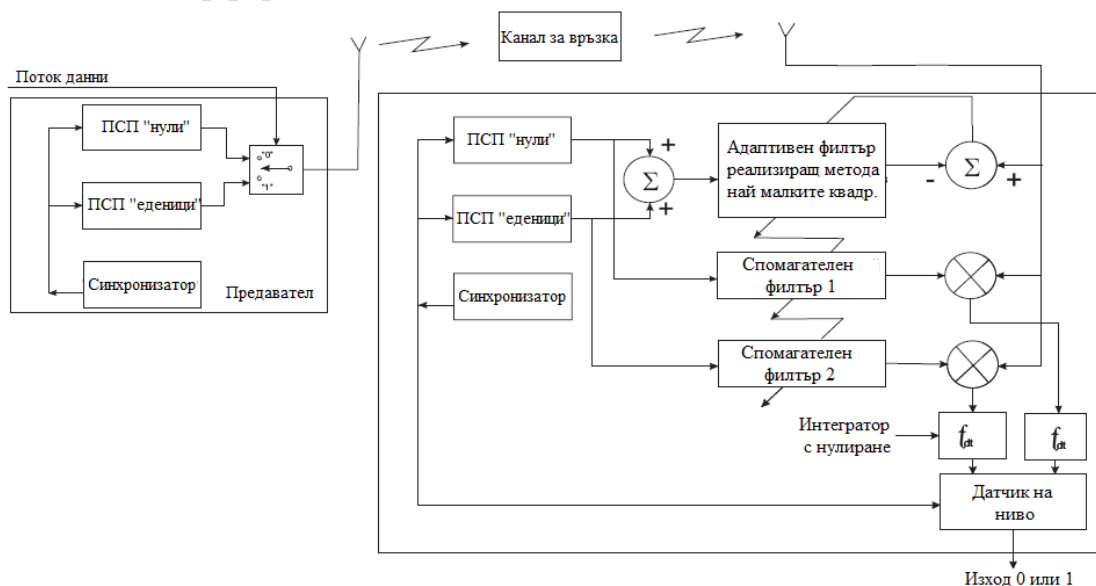
Същността на адаптивната цифрова филтрация на този вид сигнали [4] се състои в използването на обикновен КИХ – филтър с променлива импулсна

характеристика, както в цифровите филтри с честотна селекция, която не се променя по априорен закон, както е при каламановия тип цифрови филтри. В случая импулсната характеристика се коригира след постъпване на всеки нов отчет по такъв начин, че средноквадратичната грешка да бъде сведена до минимум на дадения етап от филтрацията. По такъв начин под адаптивен алгоритъм на работа се разбира рекурентната процедура по преизчисляване на вектора на отчетите на импулсната характеристика на филтъра на предната стъпка от обработката във вектор от „нови” отчети за следващата стъпка от обработката. Общата структура на адаптивния филтър за тази цел е показана на фиг. 3. Входния дискретен сигнал $x(k)$ се обработва с дискретен филтър, в резултат на което се получава изходния сигнал $y(k)$, който се сравнява с образцовия сигнал $d(k)$ в резултат на което се формира сигналът на грешката $e(k)$ като разлика между тях.



Фиг. 3. Обща структура на адаптивния филтър

На фиг. 4 е показана блокова схема на широколентова система за свързка с адаптивна филтрация, функционираща в условията на интензивни смущения и многолъчева интерференция.



Фиг. 4. Блокова схема на широколентова система за връзка с адаптивна филтрация

Принципът на действие на системата е следният: в предавателната страна в съответствие с информационните символи нула и единица се генерират две

псевдослучайни кодови последователности с помощта на които се формира носещият сигнал, подлежащ за предаване. Двете кодоводови поредици се генерират едновременно като се синхронизират с помощта на устройство за синхронизация и се повтарят във времето в зависимост от предаваната информация. Комутатора осигурява предаването на съответната последователност в зависимост от вида на предавания информационен символ от началото до края.

На приемната страна след определено закъснение сигнала постъпва на входа на приемното устройство смесен с шумовете от канала. Устройството за синхронизация на приемника формира отчети от сигнала синхронно и синфазно които постъпват на входа на корелаторите. В приемника се изчисляват взаимокорелационните функции между опорните кодови поредици и приетите зашумени кодови поредици и при правилна синхронизация в изхода на единия от корелаторите се получава максимум на сигнала. Тъй като стойността на закъснението внесено от канала за връзка е априорно неизвестно, фазата на системата за синхронизация постепенно се променя в съответствие с тези максимуми. Решението за приет сигнал се взема съобразно критериите за оптимално приемане [2].

За борба с ефектите от многолъчевото разпространение, дфилтър, работещ по метода на най малките квадрати. Този филтър непрекъснато моделира алгоритъма на работа на приемника в съответствие с параметрите на канала за връзка. Адаптирането на филтъра се осъществява по критерия за минимума на СКГ, представляваща разликата между сигнала в изхода на канала и адаптираният модел [5]. С други думи адаптивния филтър формира в изхода сигнал, максимално доближаващ се до постъпващия на входа на приемника.

Основно изискване към предлагания метод е периода на псевдослучайните кодови поредици да бъде по голям от времето за ефектите многолъчевото разпространение, а времеконстантата на адаптивния филтър да бъде не по малка от времето на интервала за многолъчевост. За канал с висока интензивност на смущенията процесът на адаптация трябва да бъде бавен. Предлагания метод се явява най ефективен за стационарен канал или нестационарен, но с бавно изменящи се параметри.

4. Перспективи за реализация на системата на основата на съвременна елементна база.

Практическата реализация на системата в пакета за визуално моделиране на динамични системи Simulink на програмната системата за математически изчисления Matlab на модела на адаптивната широколентова система за предаване на информация е показала нейната устойчива работа и добрата ефективност при високи нива на смущенията и многолъчево разпространение на сигнала.

Получените в хода на моделирането резултати и алгоритми за работа на системата позволяват нейната реализация на съвременна елементна база. За тази цел могат да бъдат използвани цифрови сигнални процесори от фамилията

ADSP-21XX на Analog Devices и TMS320C на Texas Instruments. Характеристиките на производителността на някои от тях са показани в таблица 1.

Таблица 1

Операция	ADSP 2185N – 80 MHz		TMS320C6416 – 600 MHz	
	При предаване на данни	Без предаване на данни	При предаване на данни	Без предаване на данни
Корелационна функция	27 mS	27 mS	1.7 mS	1.7 mS
БПФ 1024 компл. точки	160 μ S	160 μ S	11 μ S	11 μ S
FIR филтър с 35 извода	393 μ S	393 μ S	25,7 μ S	25,7 μ S

5. Изводи

Показано е, че прилагането на процеса на адаптация позволява проектирането на комуникационни системи работещи ефективно практически във всякакви условия на разпространение на сигнала. Чрез компютърно моделиране е показана работоспособността на системата при многолъчево разпространение на сигнала в канал с бавно изменящи се параметри при наличие на смущения.

Разгледаните в статията въпроси способстват за построяването на ефективно работещи системи с висока шумозащитеност работещи при отношения сигнл/шум по вход значително по малки от единица. Разгледани са и начините за техническата реализация на комуникационни системи за предаване на данни с прилагането на съвременна техническа и елементна база.

ЛИТЕРАТУРА

- [1] Ipatov, V. Spread Spectrum and CDMA. Principles and Applications. John Wiley & Sons, INC., 2005.
- [2] Bendat, J., A. Piersol. Engineering Applications of Correlation and Spectral Analysis. John Wiley & Sons, INC., Third Edition, New York, 2007.
- [3] Адаптивные фильтры: Пер. с англ. / Под ред. К.Ф.Н. Коуэна и П.М. Гранта. – М.: Мир, 1988. – 392 с.
- [4] Ibnkahla, M. Adaptive Signal Processing in Wireless Communications. Taylor and Francis Group, New York, 2009.
- [5] Киррилов, С. Н., А. А. Лоцманов. Многоканальный адаптивный фильтр – компенсатор импульсных и узкополосных помех. Журнал Радиотехника № 6, 2006.

Автори: Вася Илиев, н. с. I ст. д-р от Институт за Специална Техника - София; Михаил Момчеджиков, проф. д-р кат. „Радиокомуникации и видеотехнологии”

Постъпила на 29.11.2010

Рецензент доц. д-р Илия Илиев

МЕТОД ЗА СКРИТО ПРЕДАВАНЕ НА ИНФОРМАЦИЯ С ПОМОЩТА НА ХАОТИЧНИ СИГНАЛИ

Вася Илиев

Резюме.: *Разгледан е нов метод за скрито предаване на информация, основан на използване на закъснение за отделяне на полезния сигнал от шумоподобните трептения, които могат да бъдат както хаотични, така и стохастични. С числени методи е изследван математичния модел при използване на хаотични трептения. Разгледана е маскировката при предаване както на не кодиран сигнал, така и при цифрово предаване на информация.*

CONCEALED TRANSMITTING OF INFORMATION BASED ON CHAOTIC SIGNALS

Vasya Iliev

Abstract. *A new method for concealed transmitting is examined. It's based on the usage of delay for separating the useful signal from the noise-like oscillations, which can be either chaotic or stochastic. Estimating methods are used for examining the mathematical model with the chaotic oscillations usage. The mask is examined while transmitting a non-coded signal as well as digital transmitting of information.*

1. Въведение

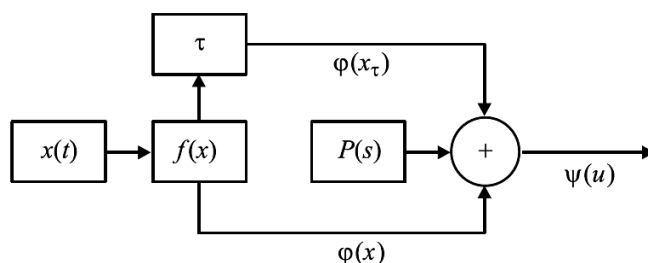
Хаотичните процеси се отнасят към едни от най интересните проблеми на физиката в началото на ХХІ в. [1], и на тяхното полезно приложение се отделя голямо влияние. Едно от важните приложения е използване на хаотичните сигнали за маскиране на предаваната информация. Много от предлаганите за тази цел методи се основават на използване на различни видове синхронизация на хаотичните осцилации [2,3]. Всички видове синхронизация на динамичния хаос се явяват много критични към параметрите на синхронизираните системи, и практическата реализация на скрито предаване на информация чрез използване на реални хаотични генератори е затруднена. Също така реалните генератори, притежаващи динамичен хаос, не се явяват чисто хаотични, тъй като вътрешните шумове оказват силно влияние на процеса на възбуждане на трептенията. В момента болшинството от известните експериментално реализирани методи за маскиране с помощта на динамичния хаос са създадени без използване на синхронизация [4,5].

В настоящата работа е моделиран нов метод за скрито предаване на информация чрез маскиране с шумоподобни трептения, основаващ се на използването на закъснението на сигнала с цел отделяне на полезната

информация в приемната страна. При това маскиращия сигнал може да бъде както хаотичен, така и стохастичен.

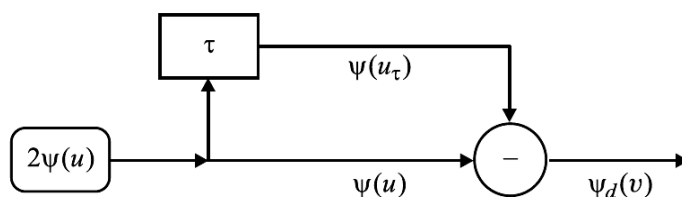
2. Същност на метода

Схемата, поясняваща описания метод за скрито предаване на информация е показана на фиг. 1 и 2. В предавателната страна (фиг.1) източникът на шумоподобни трептения генерира трептението $x = x(t)$ (хаотично или стохастично). От него се отделя радиоимпулса $f(x)$ с продължителност τ , който се разделя на два еднакви сигнала $\varphi(x)$ (при $f(x) = 2\varphi(x)$), постъпващи в различни канали. В един от каналите сигналът $\varphi(x)$ закъснява във времето на стойност τ и се преобразува в сигнала $\varphi(x_\tau)$ (при $x_\tau = x(t - \tau)$), след което постъпва на суматора. На другите входове на суматора постъпва незакъснелия радиоимпулс $\varphi(x)$ и предаваното съобщение $P(s)$ (при $s = s(t)$). Резултантния



Фиг. 1. Схема на предавателна част на система за скрито предаване на информация с използване на закъснение на маскиращото трептение

сигнал $\psi(u) = \varphi(x) + P(s) + \varphi(x_\tau)$ се усилва и се предава по каналът за връзка. В приемното устройство (фиг.2) сигнала се усилва до стойност, достатъчна за неговата обработка. За опростяване времето за разпространение на сигнала в каналът за свързка ще бъде пренебрегнато. Ще приемем, че сигналът се усилва до стойност $2\psi(u)$, след което се разделя в два канала. При това положение във всеки канал постъпва сигнал $\psi(u) = \varphi(x) + P(s) + \varphi(x_\tau)$.



Фиг. 2. Схема на примната част на ситема за скрито предаване на информация използваща закъснение на маскиращото трептение за отделяне на полезната информация

В единия от каналите сигналът $\psi(u)$ преминава без изменение, а в другия се подлага на закъснение τ , при което в изхода на този канал се получава сигнал във вида

$$(1) \quad \psi[u_\tau] = \varphi(x_\tau) + P[s(t - \tau)] + \varphi[x(t - 2\tau)]$$

Чрез изваждане от този сложен сигнал трептението $\psi(u)$, преминало без закъснение, се получава разликовият сигнал

$$(2) \quad \psi_d(v) = [\varphi(x) + P(s)] + P[s(t - \tau)] + \varphi[x(t - 2\tau)].$$

Полученият сигнал представлява сума от следващи един след друг три радиоимпулса. Първия радиоимпулс представлява обрнатата сума на незакъснелия маскиращ и информационния сигнал. Втория радиоимпулс представлява закъснелия информационен сигнал, а третия радиоимпулс – само маскиращото трептение, закъсняло на време 2τ . Доколкото полезен се явява само втория радиоимпулс, то първия и третия е целесъобразно да бъдат отстранени, което може да бъде осъществено чрез подходяща честотна или времева селекция.

3. Числена реализация на метода

Числената реализация на разгледания метод за скрито предаване на информация е осъществена чрез използване на многомодов източник на хаотични трептения със закъснение. Неговата динамика се описва със следната система диференциални уравнения [6,7]:

$$(3) \quad dx/dt = y - (\omega_0/Q)s; \quad dy/dt = F(z) - \omega_0^2 x; \quad dz/dt = [x_{\tau_0} - z]/\delta,$$

където $x = x(t)$, $y = y(t)$, $z = z(t)$, $x_{\tau_0} = x(t - \tau_0)$, τ_0 – време за закъснение на сигнала във веригата за обратна връзка, ω_0 и Q – резонансна честота и качествен фактор на филтъра във веригата за обратна връзка, δ – параметър за релаксация на инерционния елемент, $F(z)$ – нелинейна функция, определяща характеристиките на усилвателя. Изчислението за асиметричност на характеристиката на нелинейния елемент има следния вид [8]

$$(4) \quad F(z) = Bz/[1 + (z - \xi)^n],$$

където B, ξ, n – параметри на усилвателя, асиметрията и нелинейността. Полезният сигнал се формира с помощта на съотношението

$$(5) \quad s(t) = D[1 + m \sin(\omega_m t)] \sin(\omega_s t)$$

където D, m, ω_m, ω_s – константи.

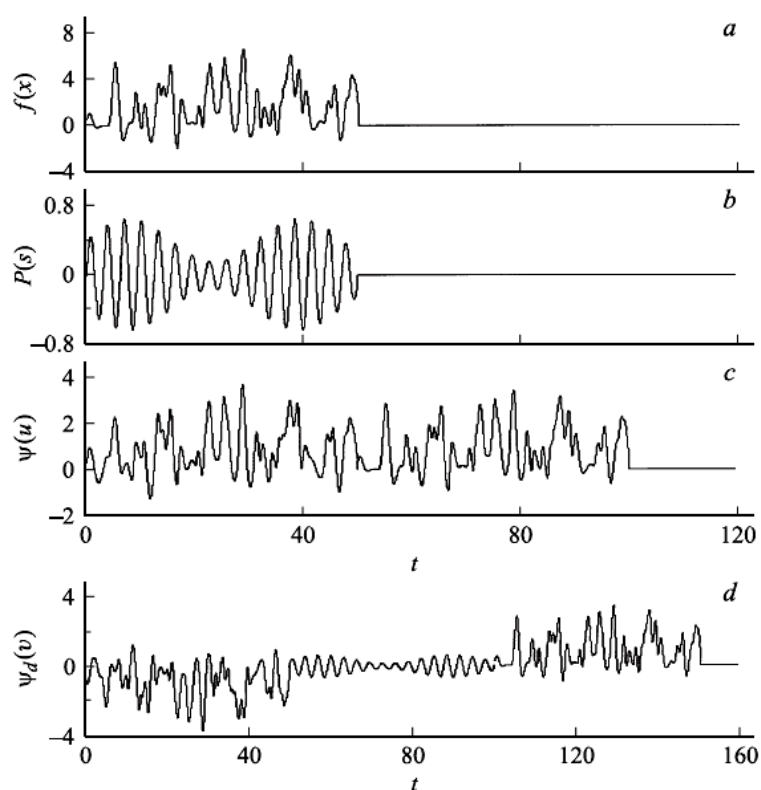
Изчисленията са проведени при следните стойности на управляващите параметри на генератора:

$\tau_0 = 4$, $\omega_0 = 2.8$, $Q = 1$, $\delta = 0.25$, $B = 6$, $\xi = 0.6$, $n = 4$, $D = 0.4$, $m = 0.6$, $\omega_s = 2$, $\omega_m = 0.6$.

При това продължителността на импулса, в границите на който се предава полезния сигнал и се реализира маскиране с помощта на хаотичния сигнал е равна на $\tau = 50$.

На фиг. 3,а е показан радиоимпулса на маскиращия сигнал $f(x)$, генериран в интервала от време $t \in [0; 50]$. Радиоимпулса е раздвоен в предавателната страна на два сигнала - $\varphi(x)$, като в единия от тях той е закъснял на стойност $\tau = 50$, и се е преобразувал в сигнала $\varphi(x_\tau)$. И двата маскиращи импулса постъпват на суматора, на който също се подава и подлежащия за маскиране сигнал $s(t)$, показан на фиг. 3,б. Резултантният сигнал в изхода на суматора $\psi(u)$ е показан на фиг. 3,с. Този сигнал на интервала от време $t \in [0; 50]$ изобразява сумата от маскиращия и полезния сигнал, а в интервала от време $t \in [50; 100]$ – маскиращия сигнал.

Ще бъде положено, че в канала за връзка сигнала $\psi_d(v)$, показан на фиг. 3,с, не закъснява, а в приемното устройство се усилва до ниво $2\psi(u)$.

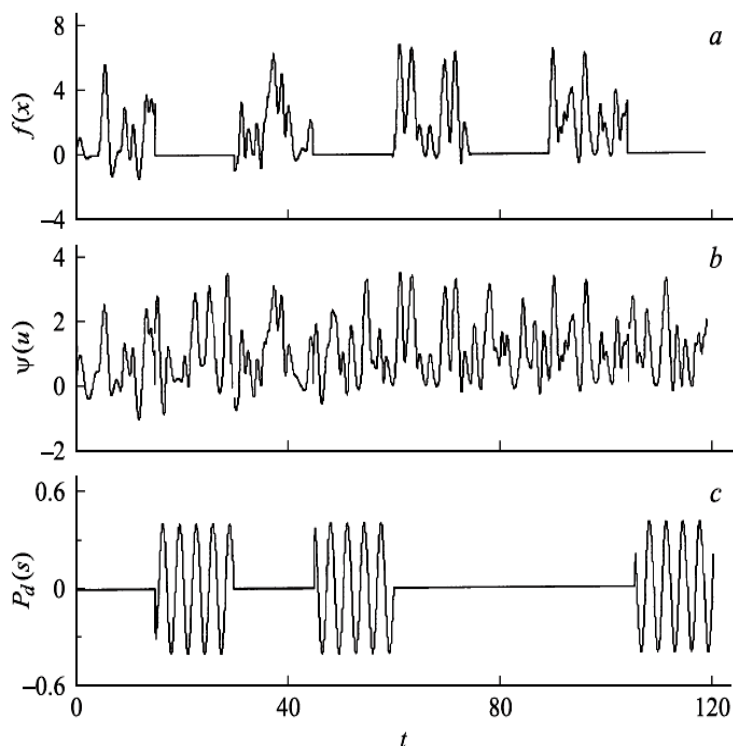


Фиг. 3. Времедиаграми на процесите по маскиране и демаскиране на аналогов сигнал

В този случай, след раздвояване на сигнала, в единия от каналите, не внасящ закъснение, той ще заема интервал от време $t \in [0; 100]$, а в другия (след закъснение на време τ) – интервал от време $t \in [50; 150]$, като видът му е подобен на реализацията от фиг. 3,с. Разликовият сигнал в изхода на изваждащото устройство $\psi_d(v)$ се изобразява с фрагмента от реализацията, показан на фиг. 3,д. На интервала от време $t \in [0; 150]$, този сигнал изобразява съвкупността от трите сигнала, следващи един след друг във вид на радиоимпулси. На интервала от време $t \in [0; 50]$ е реализирана обърнатата сума на маскиращия и маскируемия сигнал, на интервала $t \in [50; 100]$ - се

отделя полезния сигнал, а на интервала $t \in [100; 150]$ се реализира само закъснелия маскиращ сигнал. Ако изваждащото устройство се открива само на интервала от време $t \in [50; 100]$, то сигналят в изхода се възпроизвежда във вида, подобно на показания на фиг. 3.с.

Разглеждания метод за маскиране и отделяне на информация би представлявал голям интерес в случаите на използването му за битово маскиране и предаване на информация с помощта на информационни символи 0/1. Възможността за такова скрито предаване на информация е илюстрирано на фиг.4, която е получена за същите стойности на параметрите в изрази (3) и (4) както е при случая от фиг. 3, за изключение на дължината на импулса и структурата на информационния сигнал. В дадения случай той може да бъде всякакъв, в това число и хаотичен. За нагледност численото моделиране е осъществено при използване на хармонични бинарни битове.



Фиг. 4. Времедиаграми на процесите по маскиране и демаскиране на цифров сигнал

На фиг. 4,а е показан фрагмент от серия радиоимпулси на сигнала $f(x)$, с продължителност $\tau = 15$ и коефициент на запълване 1, формирани от непрекъснатото хаотично трептение $x(t)$. След разделяне на сигнала в два канала и закъснението му в един от тях на стойност равна на продължителността на импулсите $\tau = 15$, получените сигнали $\varphi(x)$ и $\varphi(x_\tau)$ се сумират в суматора заедно с полезния сигнал (5), за $m = 0$, $D = 0.4$, $\omega_s = 2$. Сигналят в изхода на предавателното устройство е показан на фиг.3, b. Ще бъде положено, че сигналят в канала за връзка не закъснява. За неговото отделяне в приемната страна се извършват същите тези операции, както в случаите на маскиране на аналогов сигнал, но при стойности на закъснението $\tau = 15$. В

резултат на това в изхода на изваждащото устройство се получава сложен сигнален процес с редуване на интервали от време, в които се наблюдават хаотични и детерминирани трептения. Ако това устройство бъде откривано с импулси, съответстващи на импулсите определящи трептящия процес в предавателната страна, то изходния сигнал при съответно закъснение на стробиращите импулси се преобразува в серия от радиоимпулси $P_d(s)$, показани на фиг. 3,с. Импулсите, получаващи се на интервала от време $t \in [15; 30]$, $t \in [45; 60]$ и $t \in [105; 120]$, изобразяват битове „1”. Пропуснатите импулси в интервала от време $t \in [75; 90]$ съответстват на битове „0”.

4. Заключение

Представените числени резултати свидетелствуват относно възможността за практическата реализация на предложения метод за скрито предаване на информация при използване както на хаотични, така и на стохастични маскиращи сигнали. Резултатите от численото моделиране показват, че този метод е приложим за маскиране както на аналогови, така и на дискретни цифрови радиосигнали.

ЛИТЕРАТУРА

- [1] Залогин, Н.Н., Кислов В.В. Широкополосные хаотические сигналы в радиотехнике и информационных системах. М.:Радиотехника, 2009.
- [2] Amann, A., Scholl E., Baba N., Just W. Time delayed feedback control of chaos by spatio-temporal filtering // 2003 Int. Conf. _Physics and Control_ , St.Petersburg, 2003 (<http://www.rusyscon.ru>).
- [3] Baptista, M.S., E.E. Macau, C. Grebogi. Conditions for efficient chaos-based communication // Chaos. 2007. 13 1. P. 145_150.
- [4] Dmitriev, A.S., A.I. Panas, S.O. Starkov. // Int. J. of Bifurcation and Chaos. 1996. V. 6. N 5. P. 851–865.
- [5] Guderian, A., A.F. Munster, M. Jinguji, M. Kraus, F.W. Schneider. Resonant chaos control by light in a chemiluminescent reaction // Chem. Phys. Lett. 1999. 312. P. 440_446.
- [6] IEEE Standard for Information technology. Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs); Amendment 1: Add Alternate PHYs, 2007.
- [7] Lukin, K. A. Noise radar Technology for short-range applications, Pros. of Intern. Conf. on Radar Systems, RADAR-1999, May 17-21, 1999, Brest, France.
- [8] Spec. Is. on applications of nonlinear dynamics to electronic and information engineering // Proc. IEEE. 2002. V. 90. N 5.

Автор: Вася Бойков Илиев, н. с. I ст. д-р от Институт за Специална Техника – София.

ПРОЕКТИРАНЕ НА БАЗА ОТ ДАННИ ЗА ОБРАБОТКА НА РЕЗУЛТАТИТЕ ОТ ТЕСТОВ КОНТРОЛ ЗА ОЦЕНЯВАНЕ НА ЗНАНИЯТА

Даниела Минковска, Йордан Марков

Резюме. Статията представя процеса на проектиране на реляционна база от данни, съдържаща информация от получените резултати от тестов контрол за оценка на знанията на студентите. Описана е методиката на провеждане на тестовия контрол като информационна основа за изработване на обектно-реляционния модел на базата от данни. Описани са правилата за създаване на модела и е изградена E-R диаграма, визуализираща асоциативните връзки между обектите. Правят се заключения за приложението и адаптацията на проектираната база от данни в учебния процес и за бъдещата ѝ реализация чрез Система за Управление на База от Данни.

DESIGN OF DATABASE FOR PROCESSING OF TEST CONTROL RESULTS FOR ASSESMENT OF THE KNOWLEDGE

Daniela Minkovska, Jordan Markov

Abstract. The paper presents the design process of a relation database that consist of information gathered from results of test control of assessment of the knowledge of the students. The methodology of the test control is described, as a foundation for the design of the entity-relationship model of the database. The rules for creating of the model are described and entity relationship diagram is build, visualizing the associative relations between the objects. The conclusions are drawn for the application and adaptation of the database in the study process and for its future implementation through Database Management System.

1. Introduction

Architecture for E-learning with the integrated structure is being designed to satisfy the need for adaptive education and integration of computer based learning in the education, and as a result from performed analysis with evaluation of the main characteristics of modern models, used in the learning procesis. Each subject is comprised of sections. Each structural section unit consists of: information part, exercises, problems to be solved, tests for self-training, and is finished with tests for check of the knowledge and the skills [1, 2]. To make the process of the examination of the knowledge of the students adaptive and effective it is necessity to save the results into a database, with an aim the teacher to be able to generate statistical information about the success of the students and to run analysis from which decisions to be done for the quality of the education.

2. Methodology and main characteristics of the process of implementation of the test control

According to the curriculum, students' knowledge is examined through two-stage continuous control. Student's knowledge constantly increases over time and examination of the level of acquired knowledge takes place in certain intervals of time [1, 2]. Using the model of periodical assessment that includes the solving of a test version by each student for the first and second continuous control, with the help of the developed computer system for test control the control and results assessment are received immediately.

At the two test controls the students' answers of the same number of questions are uniformly distributed at the learned material, respectively 35 questions for first control and 30 questions for second control. All questions are stored in one unified database, from which the questions at basic topics on the studied material separately, through random method [3] are generated. The numbers of the basic topics for the first control are 7, and for second control are 6. All generated tests variations have the same characteristics, for example: unified properties structure with the same number for choices, as one of them is true, while the others are the result of forecasted possible typical errors of students, difficulty, the same points for a correct answer, they have equal time for implementing of the control and one and the same predefined grade scale. In this way the students are stationed in equal conditions, because they have the possibility to solve equal by difficulty, but different by content tests.

After the computer test has been correctly finished, the environment generates the so called Log – file, of which variously information can be extracted. With this information the lecturer or the test observer can collect statistical data for the following indicators:

- Name of the student tested;
- Date of testing;
- Time: the time (in format hh:mm:ss) of the start and the end of the test is shown, as well as the amount a student spent thinking and answering the respective question and the total length for solving the test.

Information about the answer given can be obtained from three main parts:

- All answers are marked in the Score part – 1 is generated for a correct answer and 0 for a wrong one;
- If the student has answered a question in part Locked, a value of TRUE is generated, otherwise, if there is no marked answer – FALSE is generated;
- All answers that the student marked as right are shown in the part Last Response;

At the end of the first session, the Log file summarizes the number of right answers over all available in the test, the corresponding grade and the time for completing the test.

To make this technology more adaptive to the lecturer's needs the results of the testing need to be stored in database. This way through the use of an appropriate DBMS (database management system), the lecturer will be able to analyze the results

and drill down by year, semester, type of test, sections, received assessment or identification of the students.

3. Design of a relational database for processing of results from test control

According to the described methodology for implementing of test control, a database for storing, processing, analyzing and presenting of the results has been designed [4]. The data are structured with four groups:

- A list with students, that had taken the test;
- A list with types of tests – first and second test control;
- A list with implemented to this moment tests – by years, by semesters and by first and second test;
- Results of the tests – points for every student individually.

The business rules, described the logical relationships between the objects in the database are as follows:

- Each student is identified by his faculty number;
- One control can be made by many students, but one student can take only one control in a given moment of time;
- One control learning information can be include in many sections, as each section can contain many questions, but the result is only one from each question;
- The questions are multiple choice with only one correct answer;
- Each test has one and only one type (first or second continuous control);
- Each type of test has different number of sections and questions;
- Each student has only one result from every test type.

The conceptual scheme, describing the associations between the entities can be shown graphically by the following class-diagram (E-R diagram):

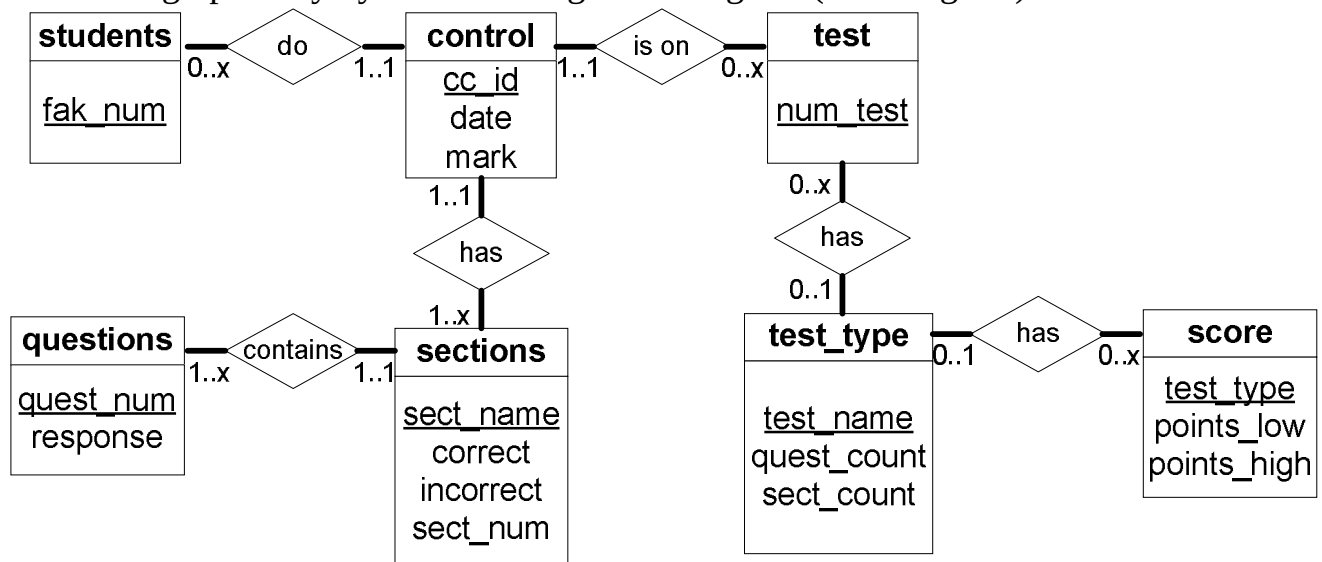


Fig. 1. E-R diagram of database for processing the test results

The students do control (first or the second) on a given date and they have a mark. In the entity set **students** the list with all students that have made control is

saved. The identification is done in the field **fak_num** for a given student. The entity set **control** contains as properties an identification on a control (first or the second control), a date of going, and the received mark of the each student. That permits filtering the students by made (or not made) controls and received marks. The studying material for every control is separated on the main topics (sections). The entity set **questions** saves the records with information for the results from received answers (**response**) of the each student by the number of questions (**quest_num**). Every question is a part from one main theme – the sections (the entity **section**) form studying material, identified by section name (**sect_name**) and reading on the numbers of the correct (**correct**) or wrong (**incorrect**) answers, as logic values.

On the other hand, the control is over distinct test (first or the second test) – entity set **test**, with property number of the test (**num_test**). Each test has a type – entity set **test_type**, identified by the property **test_name**, and containing information about number of question (**quest_count**), and number of sections (**sect_count**) in the every type of test. From the each test type we can receive the results (**score**), determining the minimum and the maximum number of the archived points from the test (**points_low** and **points_high**). Each unique combination of test and student is saved in one record.

4. Conclusion

The designed database has flexible character and permits permanently adding new information after each new run current control. It permits easy access and filtering by more different indications of the collected data. The relational model of the database allows it to be realized through any relational database management system.

As a future work, is coming the realization of the management of the designed database with a help of the database management system that will unify the research and analysis of the received results from the assessment of the students.

References

- [1] Stoyanova, L. An Structure of the WEB-based system for E-learning in Technological school “Electronic systems”, Proceedings of the Int. Conference Electronika’2004, Book 1, p.298-303, Sofia, 2004.
- [2] Minkovska, D., G. Jelev. Design, Architecture and Description of a Model for E-learning, Proceedings of the International Scientific Conference “Computer Science’2008”, Kavala, Greece, 2008.
- [3] Jelev, G., D. Minkovska. An Approach for Improving of Test Environment for the Knowledge Mastering, Proceedings of the International Scientific Conference “Computer Science’2005”, Chalkidiki, Greece, 2005.
- [4] C. Y. Date. An Introduction to Database Systems, Addison-Wesley, Philippines, 1987.

Authors: Assist. Prof. Daniela Minkovska, Department of Programming and Computer Technologies, Jordan Markov, Bachelor, Faculty of Computer Systems and Control.

ВИЗЕМНО ОРИЕНТИРАН ЕВРИСТИЧЕН МЕТОД ЗА ОПИСАНИЕ НА ПРОМЕНЕТЕ НА ЛИЦЕТО, ИЗПОЛЗВАЩ ЕФЕКТА НА СОЛАРИЗАЦИЯ

Веселин Георгиев

Резюме. Човешкото лице е система с голям брой степени на свобода, което обуславя и сложността на процеса на определяне на формата на лицето в моментите на говорене. Това е довело до разработка на различни методи за определяне на характеристиките на лицето при говорене като основа за формиране на виземно множество.

Представената публикация пописва нов евристичен метод за определяне на мястото и степеня на промяната върху лицето при произнасяне на дадена фонема. Това позволява дефиниране на критерий за различност/еднаквост като основа за формиране на виземното множество.

A HEURISTIC METHOD FOR CLASSIFICATION FACIAL CHANGES FOR VISEMES' SELECTION BASED ON SOLARIZATION EFFECT

Vesselin Gueorguiev

Abstract. Human face is a system with huge amount of degree of freedom. As a result, the facial expression analysis of talking humans is a complicate task. Existing methods for determining the locations, type and level of facial changes use different techniques for solving this task. This influence on the generated visemes' set used in character animation systems.

The paper presents a new heuristic method for classification facial changes for visemes' selection process.

1. Въведение

Човешкото лице е система с голям брой степени на свобода, което обуславя и сложността на процеса на определяне на формата на лицето в моментите на говорене – най-често, възприемането на говоренето много сериозно се влияе от изражението на лицето, а не от конкретното казаната фраза [6].

При определяне на виземите трябва да се има предвид, че няма абсолютна форма на лицето при произнасяне на дадена фонема – това зависи от конкретния човек, от съчетанието на фонемите в думи (т.нар. коартикулационни промени на фонемите), от логическите акценти (визуалната прозодия при говорене), от емоционалното и физическо състояние на говорещия и т.н. (фиг. 1). Затова, изборът на виземи изисква работа само с общите характеристики на промените при говоренето, докато индивидуалните

характеристики и емоцията се добавят като следващо ниво на реалистични детайли. Това определя основна задача при създаването на виземно множество за конкретен език - необходимо е да се направи първоначално определяне на зоните на промяна (първо ниво на клъстеризация), след което се извършва детайлизиране на промените. Резултатът от тези анализи е процедура за определяне на близост/различност между израженията на лицето при говорене, като част от процеса на клъстеризация/обединение, т.е. генерирането на виземно множеството за конкретния език.



Фиг. 1. Диаграма на факторите, влияещи върху формирането на изражението на лицето (по материали от [4])

В публикацията се описва нов евристичен метод, позволяващ определянето на мястото и степента на промяната при говорене. Описаният метод е устойчив спрямо промяна на момента и условията на генерация на обработвания видео материал. По този начин се осигурява възможност за сравнение на резултатите при по-продължителни проучвания.

2. Основна структура на движението на устата при говорене

Независимо от голямата степен на свобода, което има човешкото лице, в процеса на говорене могат да се определят няколко базови форми. Тези форми определят граничните позиции на лицето, разделящи движението при говорене на две базови движения (фиг. 2). Първото движение е резултат от преместването на челюстта нагоре-надолу и е прието да формира позициите 'затворена уста' (неутрално положение) и 'отворена уста'. Второто движение се определя като резултат от преместването на устните встрани, поради което формира позициите 'широка уста' и 'свита уста'. Всички другите форми устата при говорене представляват комбинация от тези базови форми (не е задължително на всички едновременно) [7].



Фиг. 2. Четирите основни положение на устата при говорене (отляво-надясно): неутрално положение, максимално отворена, максимално свита, максимално разтеглена встрани (източник:[7])

Наличието на базови форми теоретически позволява движението на устата да се анимират чрез метаморфоза между тях. Това изисква управление на тегловните коефициенти на участие на базовите форми при формиране на конкретната форма на лицето в процеса на говорене. Директното симулиране на реалистично говорещи хора на основата на само тези четири форми е изключително сложен процес. Резултатът от сложността на преходите между базовите форми е неправдоподобни и нереалистични форми на лицето при преходите, независимо че крайните форми са правдоподобни. Това е определило и този вид анимация да се използва само в случаите на елементарни аватарни генерации на говор.

Друг проблем е фактът, че тези форми не могат да опишат някои по-рядко срещани изражения, например силно стиснати устни, поради това че те са извън диапазона на изменение, описан с базовите форми. Решението на проблема е използването на преекспониране на участието на някоя от формите (коефициенти извън диапазона $[0,1]$) за постигане на този ефект, нещо което много често води до неправдоподобност не само на прехода, но и на крайната форма. Затова, при определяне на промените при говорене, участие вземат още промяната във вида, формата и ориентация на фокусните точки в областта около носа (фиг. 3), наличието на видим език и/или зъби, промените в зоната между устата и брадичката.

Независимо от тези недостатъци, подход за оценка на движенията при говорене, основан на четирите базови позиции, служи като основа за семантичното разделение на виземите на основни групи.



Фиг. 3. Основни видове фокусни точки около носа (източник: [7]).
Промяната им е фактор за оценка на наличието на емоция при говорене

В най-общия случай при говорене движенията на устата имат следните специфики:

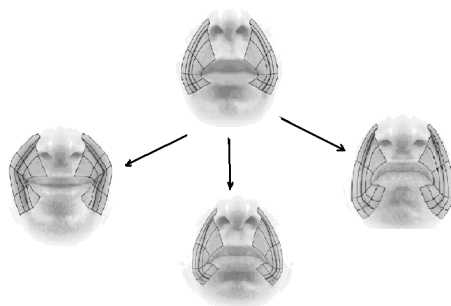
- движения много рядко се изпълняват от крайна позиция до крайна позиция – това е и причината, при анимации на основата на метаморфоза, виземите не винаги да не се избират по максимална позиция.

- движението нагоре/надолу съществува при всички фонемни – затова това е основната отправна точка при реализация на анимация на говорещи персонажи, независимо дали са реални, фантастични, антропоморфни и кукли от куклен театър.
- движението на устните встрани не се реализира при всички фонемни
- няма движение на устата при говорене, което да реализира две от основни движения едновременно до максималните стойности

3. Метод за определяне на промените в характеристикните места

За определяне на виземно множество за конкретно избран език е необходимо използването на метод, който позволява: а) да се определят областите на промяна, спрямо неутрална позиция; б) да позволява да се определи близост между промените, т.е. да се използва система за оценка на големината и вида на промяната. Съществуват различни методи, като най-общо те се делят на две основни групи: евристични методи, използващи специфични аспекти на биомеханиката на изследваните движения [7, 10, 11, 12, 13, 15, 18], и методи за директно извличане (data-driven techniques) [1, 2, 3, 5, 8, 9, 14, 16, 17], които използват входните данни за директно определяне на характеристикни параметри (най-често така са статистически методи).

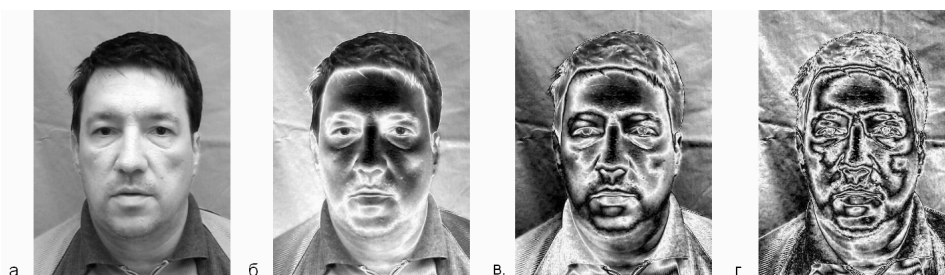
Един от най-анимационно ориентирани методи е методът на Osipa [7]. Той е представител на евристичните методи за ръчна класификация. При него, за определяне и анализ на промените в ярко осветените области и областите със сенки, се определя размера и дълбочината/изпъкналостта на тези зони от лицето. Това се постига на основата на интензитета (в тонове на сивото), като изследваните характеристикните области са: устни, областта около устата, област покрай устата по посока носа. Анализът се извършва на основата на деформирана двумерна мрежа (фиг. 4), като интензитета на пикселите определя съответната деформация по третата координата. Това е класически представител на евристичната група методи, т.е. класификационната процедура е семантично обвързана с важността на местата, в които са извършени промените.



Фиг. 4. Определяне на промените при метода на Osipa [7] – използва се деформация на двумерна мрежа, като интензитета на пикселите задава третата координата.

Предлаганият в тази статия метод представлява модификация на метода на Osira. Описаната модификация позволява използването на възможностите на цветните изображения и обработките върху тях, без необходимостта от помощни елементи. Класическият метод, по-същество, разглежда изображенията само като многотонови черно-бели (тонове на сивото), защото използва само интензитета на пикселите. Това намалява читаемостта на резултатите и затруднява използването им, което е породило необходимостта от помощната мрежа. В предлагания метод информацията се използва без необходимостта на помощни геометрични елементи, защото се използват възможностите, осигурявани от на цвета в изображенията. Това се постига чрез обработка, чиято задача е получаването на специфична цветна текстура в областите с промяна (сенки и осветени места). По този начин, чрез сравнението на цветните текстури, се определят характеристикните места, големината и вида на промяната.

За реализация на задача, за определяне на областите на промените и степента на промените, се използва обработка на двумерни цветни изображения. Идеята на метода е, че осветеността на обектите може да се използва като основа за определяне на цвят по избрана скала, т.е. според осветеността се получи и определен цвят. За постигане на тази обработка се използват възможностите на обработката на двумерни растерни изображения, наречена 'соларизацията'. При тази обработка оцветяването на обектите е според осветеността им, т.е. различният интензитет на осветяване на лицето води до различни цветове за съответните места. Последователното прилагане на тази обработка увеличава детайлността поради промяна на нивата на квантоване на интензитета. Резултатът от това е обхващането на все по-големи области и все по-голямо многообразие на цветове в обработеното изображение (фиг. 5).



Фиг. 5. Промени в изображение при приложение на ефекта на соларизация – задава различни цветове според осветеността. За примера: а) нормално изображение, еднократно (б), двукратно (в) и трикратно (г) приложение на соларизацията.

В предложения метод се използват еднократно и двукратно обработени изображения със соларизация – еднократната обработка се използва за определяне на разположението и приблизителна оценка на размерите на характеристикните области за промяната (фиг. 6), докато двукратното приложение позволява да се определят големината (височината/дълбочината) на промените по повърхността. За определяне на промените в изображението е

необходимо сравняване на обработеното изображение с неутралното положение. Това се реализира чрез използването на операцията ‘изваждане на изображения’, при която от ‘соларизираното’ форма на изследваното изражение се изважда ‘соларизираното’ форма на неутралната позиция, при което цветната текстура остава само в областите, които са променени. Оценката на големината на промените по повърхността (дълбочина или височина), е чрез количеството на цветовите зони в текстурата, докато броя и широчината им определя градиента на промяната (в случая това е наклона спрямо осветлението в засегнатите области).

За сравняване на промените, в процеса на определяне на близост между израженията на лицето, се използват двете нива на соларизация, според вида на обработката: определяне на групови характеристики и определяне на разликите в групата.

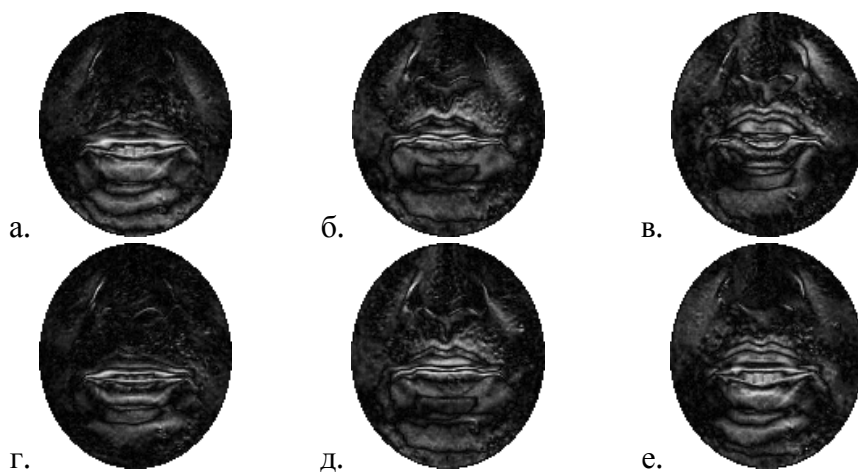


Фиг. 6. Използване на соларизация за определяне на местата на промяна и степента на промяна. а) неутрална позиция; б) обработвано изображение; в) резултат от изваждане на еднократно соларизирани изображения; г) резултат от изваждане на еднократно соларизирани изображения.

На етапа на определяне на групови характеристики се извършва определяне на големите различия – оценяват се формата и разположението на зоните на промяна, използвайки еднократно соларизирани изображения (фиг. 7). При тази оценка се следи за следните характерни промени:

- Оценява се наличието на оцветена област в бръчката от двете страни на долната устна. При говорене, това е свързано с движението на долната челюст, т.е. при отваряне на устата нагоре-надолу.
- Оценява се наличието на оцветена област между двете устни:
 - о Наличието на цвят в гамата на циана показва отваряне (основно е резултат от осветяването на зъбите при неиздадени напред устни).
 - о Появата на тонове на червеното и оранжевото определя видимостта на вътрешността на устата (обикновено е вътрешната страна на устните и/или венците)
- Оценява се оцветяването на горната устна:
 - о При движение само за отваряне/затваряне там не настъпват промени и това се определя чрез отсъствието на оцветяване.
 - о При движение свиване/разтягане на устата се получава оцветяване на горния ръб във виолетово.
 - о При разтягане встрани на устата се получава характерично оцветяване в синьо-цианово по долната страна на горната устна.

- Оцветяването на долната устна се оценява по два показателя:
 - о Оцветяването на предната страна на устната, около закръглението на горния ръб, като резултат на по-доброто осветяване на областта – получава се при движение на устата надолу (устната се обръща навън, което излага на светлината на нови части от нея).
 - о Оцветяването на основата на устната (от ъгълчетата на устната надолу и навътре), което е резултат от издаване на долната устна напред - степента на издаване се определя от броя на цветовете в тези участъци, а големината на участъците показва колко тясна е оформената тръбичка от устната (тесните тръбички имат плавно изменение, т.е. широки зона; при широка тръбичка на устните има голям градиент на промяна в тази част, т.е. тесни участъци).
- Оцветяването на зоната между устните и брадичката се получава по две причини:
 - о Движение на челюстта надолу – получава се преместване в същата посока, т.е. променя се осветяването на този участък.
 - о Разтягане на кожата в този участък – може да се получи от различни движения на долната устната.
- Оценява се наличието/отсъствието на оцветяване около носа, както и цвета на участъка (виолетовия цвят в тази област свидетелства за наличието на много слаба промяна на околоносната бръчка). При говорене това най-често е резултат от втория тип базово движение на устата – при разтягане встрани или при свиване към форма на устата, наподобяваща тръбичка. При фонемите, за които не участват мускули, осигуряващи разтягане/свиване встрани, тази промяна е показател за наличие на емоция в процеса на говорене.

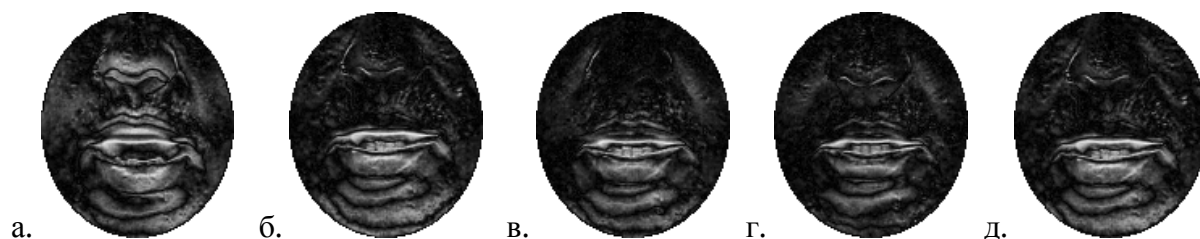


Фиг. 7. Използване на соларизация за определяне на местата на промяна и степента на промяна при гласните фонемите: а) [a]; б) [y]; в) [и]; г) [ъ]; д) [о]; е) [е];

4. Анализ на резултатите и приложимост на метода

Резултатите от анализите на изображенията на лицето при произнасяне на отделните фонемите в дума позволява да се определи принадлежност на съответните виземи към групи. Това може да се използва както основа за потенциалната възможност/невъзможност за обединение или клъстеризация. Например изображенията от фиг. 7 ясно показват потенциалните виземи за обединение (гласните [a]-[ъ] [o]-[y] и [e]-[и] в българския език), докато на фиг. 8 се вижда еднаквостта за гласни от една и съща група (фонемата [a] на различни позиции в думата спрямо ударението).

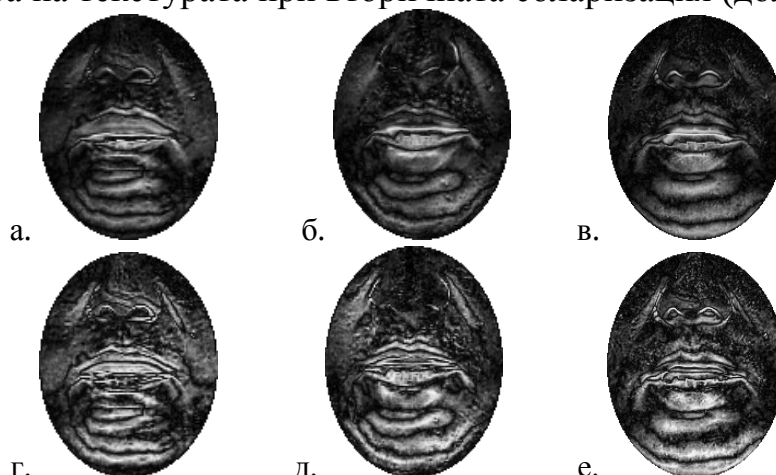
Използването на двойно соларизирани изображения, като второ ниво на сравнение, осигурява възможността за визуализация и анализ на по-малки промени. Получената цветна текстура се отличава с по-голямо количество цветове, по-тесни и контрастни зони. Като резултат от това класификацията позволява по-точното определяне както на промените в местата, изследвани на предната стъпка (зоните се увеличават по размер и се увеличава броя на цветовете), така и дава възможност за изследване на промени в нови зони: долния край на бузите, горния край на брадичката (под вдлъбнатината), долния край на брадичката (добър критерий за движението на челюстта), зоната над горната устна (промяната във вдлъбнатина под носа е критериите за самостоятелното движението на горната устна в посока нагоре). По този начин може да се определи наличието на влияние на емоционални промени на лицето и/или коартикулационно влияние върху дадената фонема.



Фиг. 8. Ефектът на соларизация при елементи от една и съща група. В примера: произнасяне на [a] в различни позиции спрямо ударението: преди ударението - (а) 2 срички и (б) 1 сричка; в) под ударение; след ударението - (г) 1 сричка; д) 2 срички.

Приложимостта на метода се определя и от възможността за използване за анализ на данни, записвани по различно време и от различни експериментатори. Единственото условие е необходимостта, за всяка група от записвания на произнасяни думи, да има и запис на неутралната позиция. По този начин се извършва специфична форма на нормализация на информацията, без да е необходимо използването на по-скъпа техника при реализацията на експериментите. На фиг. 9 е показано как изглеждат различни произнасяния на фонемата [a] – трите случая са от различни сеанси на запис, настройките на камерите са различни, разстоянието до камерата и ориентацията на главата, спрямо камерата, също са различни. Ясно се вижда запазването, както на

базовите характеристики на групата (при единична соларизация – горния ред), така и характера на текстурата при вторичната соларизация (долния ред).



Фиг. 9. Сравнение между различни изображения от една група, правени през различно време: запазва се разпределението и цветовата текстура като при еднократно обработените (горен ред), така и при двукратно обработените изображения (долен ред).

5. Заключение

Няма едно единствено множество от виземи за даден език. Виземните множества винаги се избират според необходимостта за постигане на определено ниво на достоверност на генерираната анимация. Това налага методът за определянето на различията между отделните виземи да позволява правилното определяне на спецификата на лицето на дадения човек като основа за определяне на общите характеристики на виземите за конкретен език.

Представения метод позволява определяне на мястото и степента на промените на лицето без необходимостта от използване на допълнителни елементи. Методът може да се прилага както за сравнение на експериментите между различни хора, в различни периоди от време и различни условия. Това позволява елиминиране на влиянието на физическо, физиологическо и емоционално състояние на хората при провеждане на експериментите.

ЛИТЕРАТУРА

- [1] Hyvarinen, A. and E. Oja. "Independent component analysis: Algorithms and applications". Neural Networks 13, pp. 411–430, 2000
- [2] Samaria, A. "HMM-Based Architecture for Face Identification", Image and Vision Computing, vol. 12, no. 8, pp. 537-543, 1994
- [3] Wang, A., M. Emmi, and P. Faloutsos. "Assembling an expressive facial animation system." In Sandbox '07: Proceedings of the 2007 ACM SIGGRAPH symposium on Video games, pages 21–26. ACM Press, 2007.
- [4] Fasel, B. and J. Luetttin. "Automatic facial expression analysis: A survey". Pattern Recognition, vol. 36, no. 2, 2003, pp. 259-275

- [5] E. Cosatto, J. Ostermann, H. P. Graf, and J. Schroeter, "Lifelike talking faces for interactive services", In Proceedings of the IEEE, vol. 91, no. 9, pp. 1406-1429, 2003
- [6] Parke, F. I., K. Waters, "Computer Facial Animation", 2nd Edition, AK Peters, Ltd., 978-1-56881-448-3, 2008
- [7] Osipa, J. "Stop Staring: Facial Modeling and Animation Done Right", 2nd Edition, Sybex, 2007, 0471789208
- [8] Choi, K. and J. N. Hwang. "Baum–Welch HMM Inversion for Audio-to-Visual Conversion", IEEE International Workshop on Multimedia Signal Processing, 1999, pp. 175–180.
- [9] Aleix, M. and K. Avinash, "PCA versus LDA", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 23, No 2, Feb. 2001, pp 228-233
- [10] Wallace, M., A. Raouzaoui, N. Tsapatsoulis, and S. Kollias. "Facial expression classification based on mpeg-4 faps: the use of evidence and prior knowledge for uncertainty removal." Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2004.
- [11] Magnenat-Thalmann N, Primeau E, Thalmann D, "Abstract Muscle Action Procedures for Human Face Animation", The Visual Computer, Vol. 3, No. 5, pp. 290-297, 1988
- [12] Ekman, P. and W. Friesen. "Facial Action Coding System: A Technique for the Measurement of Facial Movement." Palo Alto: Consulting Psychologists Press, 1978
- [13] Kalra, P., A. Mangili, N. Magnenat-Thalmann, D. Thalmann, "SMILE: A Multilayered Facial Animation System", Proc IFIP WG 5.10, Tokyo, Japan (Ed Kunii TL), 1991, pp. 189-198.
- [14] 182, P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", IEEE Transactions on pattern analysis and machine intelligence, vol. 19, no. 7, July 1997
- [15] Aleksic, P. S., J. J. Williams, Z. Wu, and A. K. Katsaggelos. "Audio-visual continuous speech recognition using mpeg-4 compliant visual features". ICIP 2002, vol.1, pages 960–963
- [16] Kshirsagar, S. and N. Magnenat-Thalmann, "Visyllable Based Speech Animation", EUROGRAPHICS 2003, Vol. 22, No. 3, 2003
- [17] Roweis, T. and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding", Science vol. 290, 22 December 2000, pp. 2323–2326.
- [18] T. Ezzat & T. Poggio, "Visual speech synthesis by morphing visemes". A.I Memo N°1658, C.B.C.L Paper N°173, M.I.T., 1999

Автор: гл.ас. Веселин Георгиев, катедра "Програмиране и Компютърни Технологии'

Постъпила на 01.12.2010

Рецензент доц. д-р Димитър Пенев

ПОДГОТОВКА НА ГЕОМЕТРИЧНА СРЕДА ЗА ЦЕЛИТЕ НА ВИЗУАЛИЗАЦИЯ НА ДАННИ ОТ КОМПЮТЪРНА СИМУЛАЦИЯ И ОТ ЕКСПЕРИМЕНТИ

Виргиния Димитрова

Резюме. Изчислението на поле и неговата визуализация е сложна задача, изискваща процесорно време и памет, зависещи до голяма степен от геометрията на средата, за която се изследва полето. Геометрията на средата играе определяща роля при генерирането на набора от точки, в равнината или пространството, за които се изпълняват изчисленията/измерванията, както и при визуализацията на получените данни. В статията се разглеждат различни случаи на геометрична среда – 2D и 3D, за изследване на различни пространствени конфигурации на източници на поле и приложените подходи за дискретизация.

GEOMETRIC ENVIRONMENT PREPARATION FOR THE VISUALIZATION OF DATA OBTAINED THROUGH COMPUTER SIMULATION OR EXPERIMENTS

Virginiya Dimitrova

Abstract. Field's calculation and visualization is a complex task, that requires a lot of processor time and memory, determined in a large scale by the geometric environment for which the field is investigated. The geometric environment is of great importance for the generation of the points, in the plane or in the space, for which the calculations/measurements are performed, as well as for the visualization of the accumulated data. In the article different cases of geometric environment, both 2D and 3D, used with different space configurations of field sources are explained together with the applied approaches for subdivision.

1. Увод

При симулация и визуализация на поле определящи (освен алгоритмите за изчисление и техниките за графично представяне) са два основни фактора. Единият е съставът и конфигурацията на източниците на полето, както и тяхното позициониране в пространството. Другият е дефинирането на равнинната или пространствена област, за която следва да се изпълнят симулацията и визуализацията на полето.

Най-общо източникът или източниците на поле ограждат област от равнината или пространството, която играе ролята на геометрична среда, върху която полето въздейства и за която се изследва неговото въздействие. В някои случаи геометричната среда е равнинна област с форма на стандартна

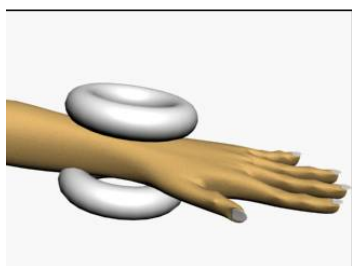
геометрична фигура (правоъгълник, окръжност и др.) или произволен затворен многоъгълник. В по-сложните случаи става въпрос за тримерен геометричен модел с произволна топология и различна сложност.

Независимо от вида на геометричната среда, за да се осъществят симулацията и визуализацията на полето е необходимо да се дискретизира равнинната или неравнинна повърхнина, т.е. да се генерира равнинна или пространствена мрежа. Тази мрежа може да бъде правилна или неправилна, а гъстотата ѝ обикновено се обозначава с термина „ниво на детайлизация” или „степен на подразделяне”. Гъстотата на мрежата е определяща за бързината и точността на изчисленията, както и за качеството на визуализацията. Независимо от вида на геометричната среда, подходът за дискретизация е тясно свързан с избора на координатна система (КС). Изборът на център и ориентация на КС, използвана при генерирането на мрежа, е определящ също за изчисленията и визуализацията на полето.

Трябва да се отбележи, че в цялостния процес на визуализация могат да участват и 3D модели или 2D изображения (векторни или растрерни), които не са свързани с процеса на симулация [1]. Тези допълнителни обекти или изображения обаче, са част от геометричната среда и допринасят за:

- Конфигурирането и позиционирането на източниците на полето;
- По-лесното възприемане и оценяване на крайните изображения.

Например, за по-лесно ориентиране на потребителя при позициониране на източниците на полето, могат да се предвидят 3D модели на източниците. За по-лесна визуална оценка на въздействието на полето може да се осигури 3D модел на геометричната среда [2], за която се изчислява и визуализира полето. При липса на разработен модел, със същата цел може да се използва 2D изображение (рентгенова снимка) или контур, както е показано на фиг. 3).



а) 3D модел

б) 2D изображение

в) Скица

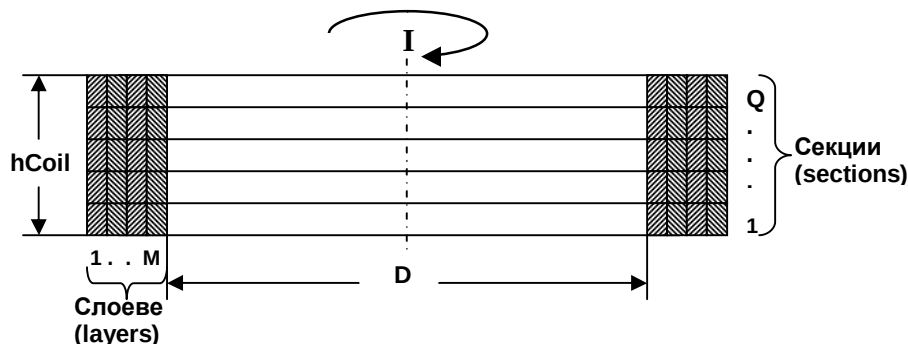
Фиг. 1. Какво обхваща геометричната среда, свързана с изчисление и визуализация на поле?

2. Дискретизация на геометрична среда за различни конфигурации на източници на поле

2.1. Един източник

Източникът на полето е многослойна цилиндрична бобина със сравнително голям радиус (фиг. 2), което позволява изследването на влиянието на полето да се ограничи в областта от пространството във вътрешността на

източника. Предвид формата на източника, генерираното поле се характеризира с осева симетрия, което позволява изчисленията да се изпълнят за правоъгълна равнинна област.

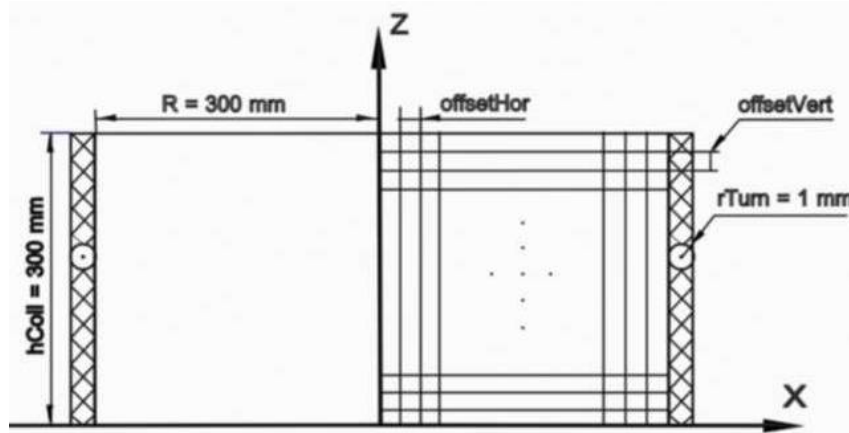


Фиг. 2. Напречно сечение на източник на електромагнитно поле

Избрана е декартова КС с център в средата на долната граница на правоъгълната равнинна област, за която се изпълнява симулацията и визуализацията. Ориентацията на КС е оста X – надясно, а оста Y – нагоре.

Предложен е подход за дискретизация на правоъгълната равнинна област, с оглед получаване на набор от краен брой точки, за които може да бъде изпълнена симулацията. Основно предимство на подхода (в комбинация с избора на КС) е, че позволява изчислението на координатите на точките да става *on-fly*, в процеса на изчисление на полето, без да е необходимо предварително дискретизиране на областта и запомняне на координатите на точките.

Правоъгълната област се разделя на равноотдалечени вертикални линии, разстоянието между които е параметър **offsetHor**, въвеждан от потребителя и определящ нивото на детайлност на мрежата (фиг. 3). Всяка от вертикалните линии се разделя на равноотдалечени точки, разстоянието между които е параметър **offsetVert**, въвеждан от потребителя. Стойностите на избраната величина, характеризираща полето, се изчисляват/визуализират за всяка точка от всяка от вертикалните линии.



Фиг. 3. Дискретизация на областта, за която се изчисляват/визуализират данните

Предлаганият подход на дискретизация позволява да не се съхраняват координати на точките, а те да се изчисляват в процеса на симулация. Всяко от изчисленията се извършва за k -тата точка от i -тата вертикална линия в дясната половина на разглежданото сечение (предвид наличието на симетрия); стойността се записва в двумерен масив $bList[i][k]$, където i е пореден номер на вертикална линия, а k е пореден номер на точка от вертикалната линия.

При визуализация, данните се извличат от двумерния масив, в който се съхраняват, а индексите на масива се преобразуват до (x,y) -координати, като се използва избраната координатна система.

2.2. Повече от един източник

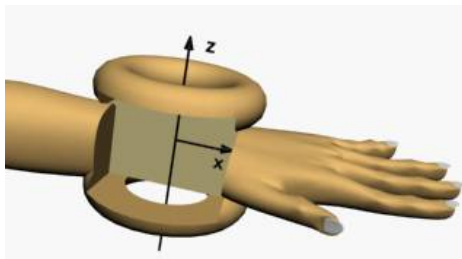
Източниците на полето са многослойни цилиндрични бобини с малък радиус, с идентични параметри и конструкция, които се позиционират (обикновено, но не задължително) по двойки в пространството. В случая, геометричната среда не е изцяло затворена във вътрешността на източника на полето и трябва да се дефинира равнинна или пространствена област, ограничена на базата на определен критерий. Този критерий може да се базира на целта на изследването или на разстоянията, отвъд които изчисленията не биха имали значение за изследването. Възможно е ограниченията да са наложени от конкретен 3D модел, единствено по повърхността на който се ограничават изчисленията и симулацията.

В случая на повече от един източник на поле съществуват няколко варианта на взаимното им разположение:

- източниците имат обща ос z ;
- източниците имат несъвпадащи z -оси, т.е. z -оси, сключващи ъгъл помежду си, но z -осите им лежат в една равнина;
- източниците имат несъвпадащи z -оси и z -осите им не лежат в една равнина.

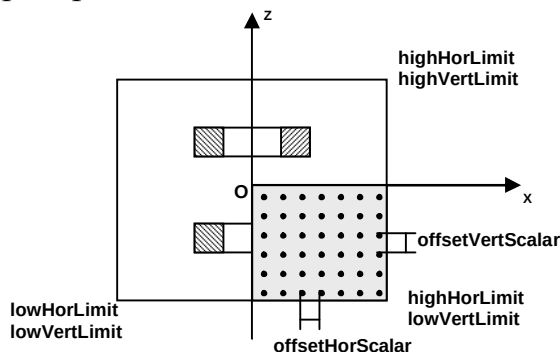
Съосни източници на поле

Избрана е КС, обща за двойката източници, чийто център лежи на върху общата ос Z , точно в средата на разстоянието между тях, както е показано на фиг. 4. Положителната посока на оста X е надясно, а положителната посока на оста Z – нагоре. Така дефинираната КС позволява да се оптимизира дискретизацията на областта. Подходът на дискретизация е аналогичен на изложения при един източник на поле.



Фиг. 4. Правоъгълна равнинна област и КС в случай на съосни източници на поле

В действителност, областта, за която се изпълняват изчисленията и визуализацията, излиза извън пространството между източниците и обхваща също самите източници и известен външен участък. Целта е получаване на по-пълна представа за разпространението на полето.



Фиг. 5. Сечение на пространствената област, за която се изследва полето

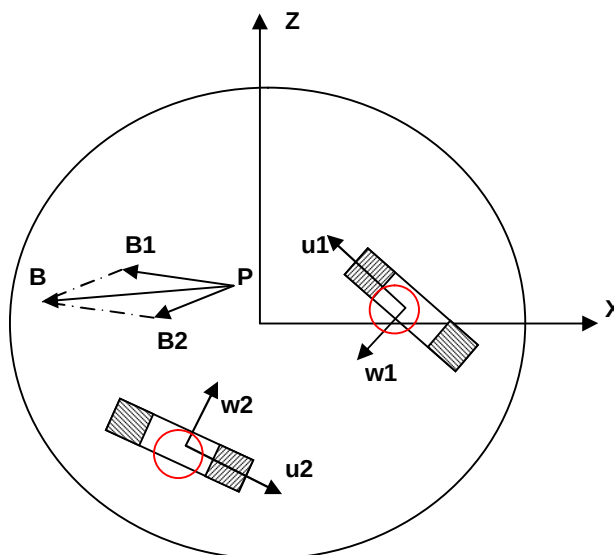
Размерите на областта могат да се задават интерактивно, като се въвеждат или променят стойностите на съответните параметри. По-големи стойности позволяват да се получи по-пълна картина на разпределение на полето, но увеличава значително броя на точките. Известно ускорение може да се постигне, ако се увеличат стъпките на дискретизация на областта по осите X и Z. Това обаче влошава качеството на изображението на разпределението на полето (назъбване на картината), по-специално за случая на визуализация на скаларно поле. За случая на визуализация на векторно поле по-голяма стъпка на дискретизация в известни граници може дори да даде по-ясна картина.

Несъосни, но ко-планарни източници на поле

Ако за двойката токови източници отпадне условието Z-осите им да лежат на една права, свободата за позиционирането им е по-голяма, както един спрямо друг, така и по отношение на геометричната среда, върху която въздействат. Тази свобода поражда усложнения в генерирането на набора от точки, за които трябва да бъде изчислено полето, както и в необходимостта от поддържане на различни КС: локални координатни системи за всеки от токовите източници (ЛКС) и глобална КС, в която се изчислява и визуализира сумарното поле. Центърът и ориентацията на ЛКС се определят от съображения за оптимизиране на изчисленията на локалното поле. Изборът на глобалната КС, в която се извършва симулацията и визуализацията на сумарното поле, се определя от два фактора – избраната геометрична среда, за която се изследва полето и пространственото разположение на конфигурацията токови източници. Едно конкретно решение е представено на фиг. 6.

За да се улесни и ускори дискретизацията, както и за да се ускорят и улеснят изчисленията, е подходящо областта от пространството да бъде апроксимирана с правилна геометрична фигура. Освен това, използваната геометрична фигура следва да покрива максимално изследваната област, като включва зоната на източниците и малък оградящ (външен) участък. При отчитане и на двата фактора и анализ на преобладаващите в практиката

приложения, предложението на автора е кръг, за двумерния случай и сфера – за тримерния.



Фиг. 6. Несъосни, но ко-планарни източници

По отношение на подхода за дискретизация, в случая на кръг е предложен алгоритъмът, представен на фиг. 7, който запълва кръга с точки, разположени по концентрични окръжности, започвайки от центъра. Този алгоритъм изисква два параметъра – разстояние между концентричните окръжности (параметър **offset**) и ъгъл (параметър **tita**). При всяка итерация се изчислява нова стойност на междинната променлива **dist**:

$$\mathbf{dist} += \mathbf{offset} \quad (1)$$

За всяка стойност на **dist** се изчислява нова стойност на междинната променлива **angle**:

$$\mathbf{angle} += \mathbf{tita} \quad (2)$$

За всяка точка от I квадрант (горната дясна $\frac{1}{4}$ на кръга) се намират нейните [x,z]-координати като:

$$\mathbf{x} = \mathbf{dist} * \sin(\mathbf{angle}) \quad (3)$$

$$\mathbf{z} = \mathbf{dist} * \cos(\mathbf{angle}) \quad (4)$$

Точките в останалите 3 квадранта се изчисляват чрез трансформация “огледален образ”.

Предимството на този алгоритъм е, че на практика позволява 100%-но запълване на областта. Недостатък е, че точките са с неравномерна гъстота, което силно снижава качеството на изображението на разпределение на полето.

Другият предлаган подход е да се генерират точки по линии, успоредни на осите X и Z, до пресичането им с ограждащата окръжност съгласно алгоритъма, представен на фиг. 8. Този алгоритъм допуска реализация и само с един параметър – еднаква стъпка по оста X и по оста Z, но може да се извлече допълнително предимство от въвеждането на два параметъра – различна стъпка по X и Z. X-координатите на точките се получават само с една операция сумиране при всяка итерация:

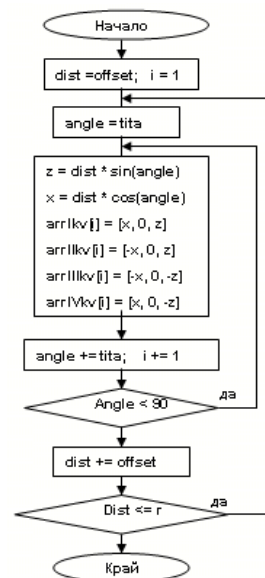
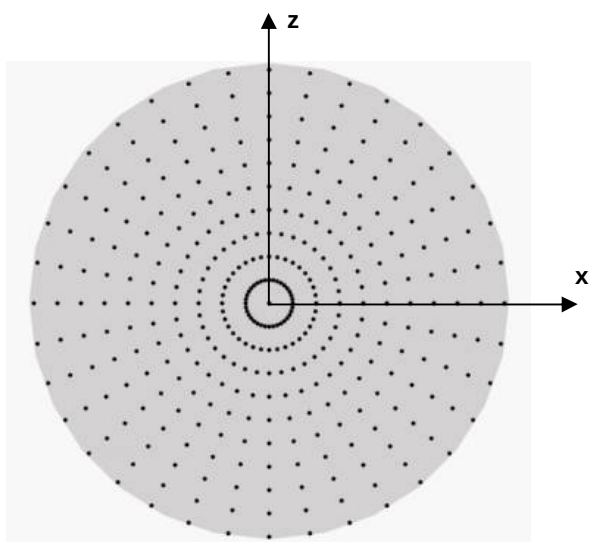
$$\mathbf{x} += \mathbf{offsetHor} \quad (5)$$

при начална стойност $x = \text{offsetHor}$ и крайна стойност $x = r$, където r е радиусът на кръговата област.

Z-координатите на точките също се изчисляват с една операция сумиране:
 $z += \text{offsetVert}$ (6)

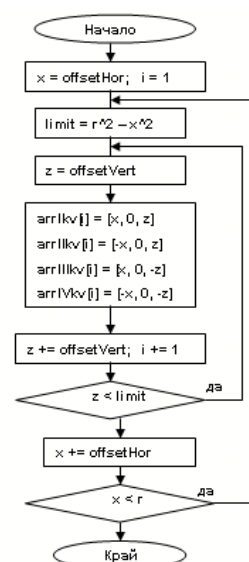
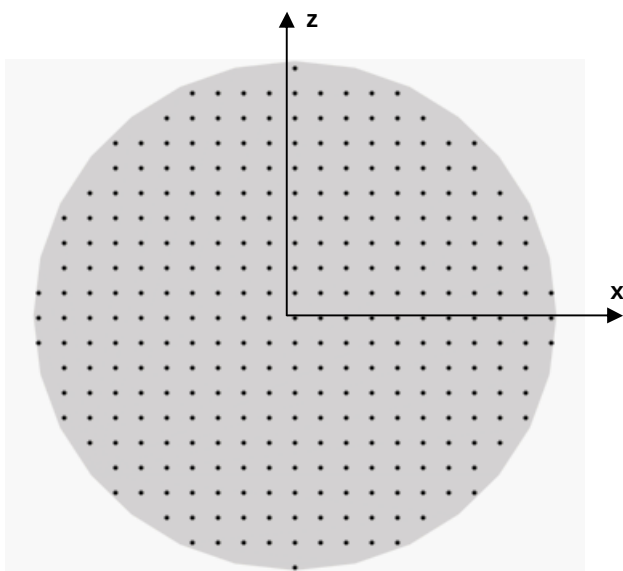
За всяка Z-координата обаче, се налага изпълнение на проверка спрямо междинната променлива **limit**, чиято стойност се преизчислява за всяка нова стойност на X-координатата по формула:

$$\text{limit} = \sqrt{r * r - x * x} \quad (7)$$



Фиг. 7. Дискретизация на кръг чрез запълването му с точки, разположени по концентрични окръжности

Недостатък на разглеждания алгоритъм е невъзможността да се постигне 100%-но запълване, което може да се компенсира с по-малка стойност на стъпката. Алгоритъмът обаче, има предимството, че генерираните точки са равномерно разпределени и в резултат, изображението на разпределението на полето е по-добро.



Фиг. 8. Дискретизация на кръг чрез запълването му с точки, разположени по линии, успоредни на осите X и Z

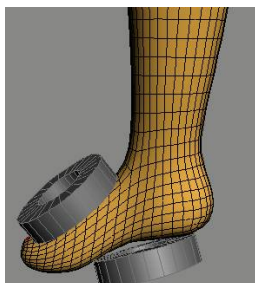
Несъосни източници с произволна пространствена ориентация

Отпадането на ограничението Z-осите на източниците да лежат в една равнина, налага промяна в дефинирането на областта, повлияна от полето и в начина на получаване на набора от точки, за които се изпълняват симулацията и визуализацията. Вместо равнинна кръгова област и алгоритмите, приведени на фиг. 7 и фиг. 8 за нейното дискретизиране, се използва 3D полигонален мрежов обект с желаната степен на детайлизация. Именно този 3D полигонален мрежов модел се използва за симулацията и визуализацията на сумарното поле.

Независимо от вида, начина на създаване, формата, размера и ориентацията в пространството на 3D полигоналният мрежов модел, от значение за качеството на симулацията и визуализацията е подразделянето на модела. За разлика от досега разглежданите случаи, не става въпрос за въвеждане на конкретна числова стойност за стъпка на дискретизация, а за осигуряване на възможност за интерактивно подразделяне на използваната 3D повърхнината до достигане на желаното ниво на детайлност.

Използването на мрежа от полигони позволява представяне на произволни 3D повърхнини и е приета за стандарт в повечето приложения за 3D симулация и визуализация, където се касае за сложни компютърни модели [5]. Качеството на мрежата от полигони се определя от нейните: интегритет, равномерност и ниво на детайлност. Интегритетът на мрежата от полигони се изразява в отсъствието на “дупки” в мрежата, а равномерността се определя от формата, приблизително еднаквите размери и взаимно разположение на елементите, на които е подразделена мрежата. И двете характеристики се определят от избрания алгоритъм на подразделяне. Възможните методи за подразделяне на повърхнина са много и голяма част от тях са реализирани на различни програмни езици и в различни среди [3, 4]. По правило програмните пакети за тримерно моделиране предоставят по няколко метода на подразделяне. Формата на елементите, на които се разбива мрежата от полигони варират от правилни многоъгълници (триъгълници, четириъгълници, шестоъгълници и осмоъгълници) до неправилни триъгълници и четириъгълници. На практика 3D полигоналните мрежи се представят като набор от топологично свързани равнинни триъгълни или четириъгълни полигони. Мрежите от триъгълни полигони са по-широко прилагани, тъй като са приети като *де-факто* стандарт в програмните пакети за компютърна графика. Те обаче водят до получаване на неравномерна мрежа и до не толкова гладко изображение на разпределението на полето [6]. Това определя избора на автора – четириъгълни полигони (фиг. 9).

Третата характеристика на мрежата – нейната степен на детайлност, се определя не от избрания алгоритъм на подразделяне, а от желаната прецизност на изчисление на полето и от “гладкостта” на получената картина на неговото разпределение.



Фиг. 9. 3D полигонална мрежа на геометрична среда,
за която се изчислява и визуализира полето

3. Заключение

В настоящата статия е разгледано значението на геометричната среда при изчисление и визуализация на поле. Тъй като геометричната среда не е нещо хомогенно, са разгледани различни нейни елементи, имащи отношение към цялостния процес на компютърна симулация и визуализация, в различните етапи на този процес – конфигуриране и позициониране на източниците на поле, дефиниране на повлияната от полето област, избор на координатни системи и дискретизация на тази област. Представени са конкретни практически решения, отговарящи на различна конфигурация на източници на поле. Извън обхвата на статията остават методите за повишаване на нивото на детайлизация на генерираната мрежа, по-специално в случая на тримерни повърхнини със сложна топология. В тази насока продължават усилията на автора, свързани с експорта и импорта на тримерни полигонални мрежи и с анализирането и имплементирането на алгоритми за тяхното подразделяне.

Литература

- [1] Foley, J.D., A. van Dam, S. Feiner, J. Hughe. *Computer Graphics: Principles and Practice*, Addison-Wesley, Reading, MA, 1990.
- [2] Hearn, D., M. P. Baker. *Computer Graphics*, Prentice-Hall, NJ, 1997.
- [3] Developing Microsoft .NET Applications for Windows (Visual C# .NET), MSDN Training, Microsoft Corporation, 2002.
- [4] Todorova, V., S. Maleshkov. "OpenGL programming in .NET environment: Problems and Solutions", Proceedings of the 18-th International Conference on Systems for Automation of Ingeneering and Research (SAER-2004), Varna, Bulgaria, 2004.
- [5] Todorova, V., S. Maleshkov. "Methods for 3d surface subdivision in calculation and visualization of static magnetic field distribution", *"Electrotechnic and Electronics"*, Bulgaria, 2006.
- [6] Тодорова, В. "3D Computer Simulation of the Static Magnetic Field Distribution over the Virtual Human Body", "Information Technologies and Control", ISSN 1312-2622, 2005

Автор: Виргиния Димитрова, д-р, главен асистент в катедра „Програмиране и компютърни технологии”, факултет „Компютърни системи и управление”.

ОПРЕДЕЛЯНЕ НА ПРОСТРАНСТВЕНОТО ПОЛОЖЕНИЕ НА ВИДИМИТЕ СТЕНИ ОТ ОБЕКТ В 3D ВИЗУАЛНА СЦЕНА

Мария Ангелова

Резюме. *Определянето на пространственото положение на видимите стени от обект е важен етап в процеса на разпознаването на образи и изграждането на модел на тримерни визуални сцени. За решаването на тази задача, авторът предлага използването на йерархична структурно-лингвистична граматика. Чрез нея постепенно се извършва пространствено структурно обобщение на използваните примитиви от едномерни към тримерни обобщени признаци. Анализът се извършва с помощта на евристични правила и каталози за пространствените отношения на стени от обект, конструирани от възли тип "ВИЛА" и "СТРЕЛКА". Възлите се анализират в следния ред: "ВИЛА" → "СТРЕЛКА" → "L". Случаите на вдлъбнати и самозакриващи се обекти са също разгледани.*

DETERMINATION OF SPATIAL LOCATION OF VISIBLE FACES OF OBJECTS IN A 3D VISUAL SCENE

Maria Angelova

Abstract. *Determination of spatial location of visible faces of objects is an important step in the process or pattern recognition and model creation for a 3D visual scene. The author proposes the use of a hierarchical structural-linguistic grammar. Here a gradual structural generalization of the grammar primitives from 1D to 3D structural parts is made. The analysis are made with the help of heuristical rules and catalogues for the spatial relations of visible faces constructed by nodes of types "Y" and "ARROW". The nodes are examined in the following order of types: "Y" → "ARROW" → "L". The cases of concave and self-covering faces are also considered.*

1. Въведение

Крайната цел при анализа на тримерни визуални сцени е да се изгради модел на сцената. Това означава да се разпознаят всички присъстващи в сцената обекти и да се определи тяхното взаимно пространствено разположение [1]. Неотменна част от процеса на разпознаване е определяне на пространственото положение на видимите стени от обект. За решаването на тази задача, авторът предлага използването на йерархична структурно-лингвистична граматика. Терминалният и нетерминалният речник и правилата за трансформация на тази граматика са подробно описани в [2].

При анализа на сцената се използва една двумерна проекция на визуалната сцена. Стремешът е анализ на високо ниво, максимално близко до

човешкото мислене и начин на възприемане на околната среда. Първоначално се извършва разпознаване на обектите и от получените описания се извлича информация за техните пространствени свойства.

Анализът на тримерна визуална сцена с цел изграждане на нейното структурно-лингвистично описание се извършва при приемане на следните ограничения:

1. Всеки връх, принадлежащ на обект от сцената, се образува от пресичането на три стени и обектите в сцената са както изпъкнали, така и вдлъбнати.

2. Обектите са в такава (основна) позиция, че нейното незначително изменение не води до изменение на типовете на върховете в изображението.

3. Точката на наблюдение е избрана така, че нейното незначително изменение не води до изменение на типовете на върховете в изображението.

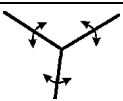
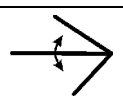
4. Не се отчита наличието на сенки, т.е. осветеността на сцената е избрана така, че не се образуват сенки.

5. Допуска се сливане на ръбове и/или съвпадение на възли и закриване между обектите, но не и цялостно закриване на сегмент от стена.

6. Предполага се, че разглежданите изображения са с извършена предварителна обработка и в тях няма наличие на шум и разкъсвания в контура.

Това са в основни линии ограниченията на Хъфман [3] и Клаус [4], но анализът допуска наличието на прости криволинейни повърхнини (цилиндрична, конична, сферична) и на закриване на обектите в сцената. Първото ограничение води до наличие само на четири типа възли в разглежданите визуални сцени: "L", "ВИЛА", "СТРЕЛКА" и "Т", което опростява разглеждането. Те са представени в таблица 1.

Таблица 1. Класификация на възлите

	"L" - възел, в който се свързват две линии
	"ВИЛА" - разклонение от три линии с ъгли $<180^\circ$
	"СТРЕЛКА" - възел от три линии, единият от ъглите е $>180^\circ$ и връзки между областите
	"Т" - възел с разклонение от три линии, две от които са колинеарни

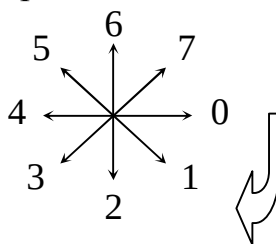
2. Каталогизация на отношенията на видимите стени от обект

В граматиката се използват три йерархични нива, на всяко от които се избират подходящи примитиви за изграждане на структурно-лингвистични описания на разглежданата тримерна визуална сцена. Извършваното постепенно пространствено структурно обобщение на използваните примитиви се движи от едномерни към тримерни обобщени признаци:

<възли и сегменти> → <стени> → <обекти>

На второ структурно йерархично ниво видимите стени от обект се използват за примитиви при изграждането на структурно-лингвистилните описания на сцената. На този етап от анализа с помощта на множество евристични правила и каталози за отношенията на стените за всяка видима стена се определя нейната геометрична форма и пространствено положение.

За да може да се направи оценка на пространственото положение на стените на обектите е необходимо да се изберат посоки в пространството. При анализа на всички визуални сцени се използват осем номерирани посоки в пространството (фиг. 1). Номерацията им е по часовниковата стрелка.



Фиг. 1. Посоки в пространството и тяхната номерация

Като входна информация за работата на системата се използва контурно бинарно изображение с извършена сегментация и отделени топологично-особени (характерни) признаци. Основните особености за всеки възел и сегмент в изображението като номер, форма, начален и краен възел, съседни възли или сегменти са част от базата от данни, изградена за всяка разглеждана сцена. Тук за всеки възел е изграден двоичен вектор, показващ наличието на съседни възли в разглежданите осем пространствени посоки.

В [1] е предложен каталог за възлите от тип “ВИЛА” при изпъкнали тела в сцената. В съответствие с избраната номерация на посоките в пространството той изглежда както е показано в таблица 2.

Резултатите от анализа на възлите тип “ВИЛА” не винаги са достатъчни. В този случай се преминава към анализ на възлите тип “СТРЕЛКА”. Такъв е случая например при тела с изпъкнали прости криволинейни повърхнини (конична, цилиндрична, сферична). При възли тип “СТРЕЛКА” възможните комбинации от пространствени разположения на стените са 24. Авторът предлага употребата в този случай на организирания от него каталога, даден в таблица 3.

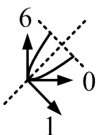
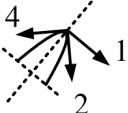
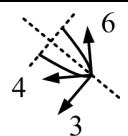
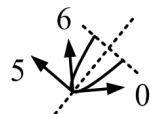
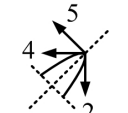
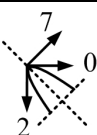
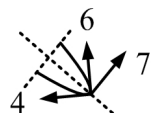
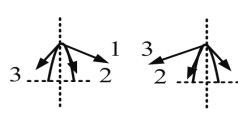
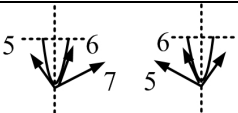
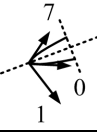
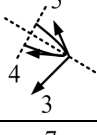
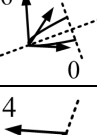
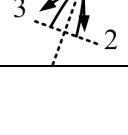
Таблица 2. Каталог за отношенията между стените, определени от признак тип “ВИЛА”, където ПС – предна стена; ГС – горна стена; ЛС – лява стена; ДС – дясна стена

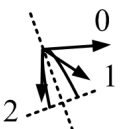
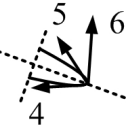
	Тип на признака	Двоичен вектор								Разположение на стените
		0	1	2	3	4	5	6	7	
1		1	0	1	0	0	1	0	0	0, 2–ПС; 2, 5–ЛС; 0, 5–ГС
2		1	0	0	1	0	1	0	0	0, 3–ДС; 3, 5– ЛС; 0, 5–ГС

3		1	0	0	1	0	0	1	0	0, 3-ДС; 3, 6-ЛС; 0, 6-ГС
4		0	1	0	1	0	0	1	0	1, 3-ПС; 3, 6-ЛС; 1, 6-ГС
5		0	1	0	0	1	0	1	0	1, 4-ЛС; 4, 6-ГС; 1, 6-ДС
6		0	1	0	0	1	0	0	1	1, 4-ЛС; 4, 7-ГС; 1, 7-ДС
7		0	0	1	0	1	0	0	1	2, 4-ПС; 4, 7-ГС; 2, 7-ДС
8		0	0	1	0	0	1	0	1	2, 5-ЛС; 5, 7-ГС; 2, 7-ДС

Таблица 3. Каталог за отношенията между стените, определени от признак тип “СТРЕЛКА”, където ПС – предна стена; ГС – горна стена; ЛС – лява стена; ДС – дясна стена

	Тип на признака	Двоичен вектор								Разположение на стените
		0	1	2	3	4	5	6	7	
1		0	1	1	0	0	0	0	1	1, 2-ПС 1, 7-ГС
2		0	0	1	1	0	1	0	0	2, 3-ПС 3, 5-ГС
3		1	1	0	1	0	0	0	0	0, 1-ПС 1, 3-ЛС
4		1	0	0	0	0	1	0	1	0, 7-ПС 5, 7-ЛС
5		0	1	0	1	1	0	0	0	1, 3-ДС 3, 4-ПС
6		0	0	0	0	1	1	0	1	4, 5-ПС 5, 7-ДС
7		0	1	0	0	0	0	1	1	1, 7-ПС 6, 7-ГС
8		0	0	0	1	0	1	1	0	3, 5-ПС 5, 6-ГС

9		1	1	0	0	0	0	1	0	0, 1-ПС 6, 0-ЛС
10		0	1	1	0	1	0	0	0	1, 2-ПС 2, 4-ЛС
11		1	0	1	1	0	0	0	0	0, 2-ДС 2, 3-ПС
12		0	0	0	1	1	0	1	0	3, 4-ПС 4, 6-ДС
13		1	0	0	0	0	1	1	0	0, 6-ПС 5, 6-ЛС
14		0	0	1	0	1	1	0	0	2, 4-ПС 4, 5-ЛС
15		1	0	1	0	0	0	0	1	0, 2-ПС 0, 7-ДС
16		0	0	0	0	1	0	1	1	4, 6-ПС 6, 7-ДС
17		0	1	1	1	0	0	0	0	1, 2-ДС 3, 2-ЛС
18		0	0	0	0	0	1	1	1	5, 6-ЛС 6, 7-ДС
19		1	1	0	0	0	0	0	1	1, 0-ПС 0, 7-ГС
20		0	0	0	1	1	1	0	0	3, 4-ПС 4, 5-ГС
21		1	0	0	0	0	0	1	1	0, 7-ПС 6, 7-ЛС
22		0	0	1	1	1	0	0	0	2, 3-ПС 3, 4-ЛС

23		1	1	1	0	0	0	0	0	0, 1–ДС 1, 2–ПС
24		0	0	0	0	1	1	1	0	4, 5–ПС 5, 6–ДС

Друг възможен случай се получава, ако в разглежданото тяло няма възли от тип “ВИЛА” или “СТРЕЛКА”, а само от тип “L”.

Правило 1: Ако в дадена видима стена от обект участват само възли от тип “L”, то тя се класифицира като предна (ПС) независимо от формата на сегментите, изграждащи възела.

Правило 2: За да се определи пространственото положение на видимите стени на всеки обект в разглежданата тримерна визуална сцена, трябва да се анализират възлите на тялото в зависимост от техния тип в следния ред:

“ВИЛА” → “СТРЕЛКА” → “L”

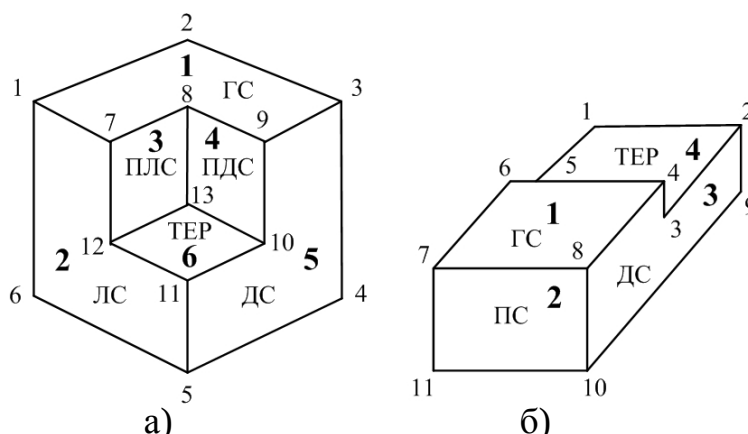
Приоритетът на получаваната информация е в същия ред.

Търсенето в каталозите за пространствените отношения се извършва чрез сравнение на двоичния вектор за посоките за образуващите възела сегменти, който е част от атрибутите на всеки характерен признак (възел) в сцената. Това е входна информация за анализа на всяка тримерна визуална сцена.

3. Обработка на вдлъбнати и самозакриващи се стени

Каталозите за пространствените отношения на стените, определени от наличието на възел “ВИЛА” или “СТРЕЛКА” дават добри резултати когато се анализират изпъкнали тела. Появата на вдлъбнати повърхнини води до необходимостта от извършване на допълнителен анализ, за да се изяснят конфликтните ситуации като например получаване на няколко стени с едно и също пространствено положение за едно и също тяло.

Тялото от фиг. 2 а) има четири възела от тип “ВИЛА”: 7, 9, 11 и 13. Възел 13 се образува в следствие на допиране на вдлъбнатите повърхнини на тялото и затова се изключва от разглеждането. Наличието на такъв възел може да се установи чрез анализ на съседните за него три възела.



Фиг. 2. Тримерни сцени с вдлъбнати тела

а) вдлъбнатото тяло ; б) вдлъбнатото тяло със самозакриване

Правило 3: Ако поне един от съседните за разглеждан възел тип “ВИЛА” възли не е част от фоновия контур на КСГТ (контурно свързана група тела), в изграждането на която участва разглеждания обект, то той е образуван от допира на вдлъбнати повърхнини и не участва в анализа за определяне на пространственото положение на стените.

Разглеждайки останалите три възела “ВИЛА” се получава следното разпределение според каталога от Таблица 3.

Таблица 4. Пространствено разпределение на стените спрямо възли “ВИЛА” за обект от фиг. 2 а)

Възел	Разположение
7	4 – ЛС; 2 – ГС; 5 – ДС
9	1 – ЛС; 6 – ГС; 5 – ДС
11	1 – ЛС; 2 – ГС; 3 - ДС

Видно е, че за някои от стените се получават еднакви разположения (две ЛС: 1 и 4; две ДС: 3 и 5; две ГС: 2 и 6), което е неприемливо. Анализът на координатите на изследваните възли може да помогне да се реши част от проблема.

Правило 4: При получаване на две различни ГС за един и същи обект тази с по-малката стойност на у координатата на съответния възел тип “ВИЛА” се определя като тераса (ТЕР), а другата е ГС.

Правило 5: При получаване на две различни ЛС за един и същи обект тази с по-малката стойност на х координатата на съответния възел тип “ВИЛА” е ЛС, а другата се определя като предна дясна стена (ПДС).

Правило 6: При получаване на две различни ДС за един и същи обект тази с по-голямата стойност на х координатата на съответния възел тип “ВИЛА” е ДС, а другата се определя като предна лява стена (ПЛС).

Отчитайки горните правила получаваме следното разпределение:

1 – ЛС; 2 – ГС; 3 – ПЛС; 4 – ПДС; 5 – ДС; 6 – ТЕР

За тялото от фиг.2 б) се анализират един възел 8, който е от тип “ВИЛА” . За обекта се получава следното разпределение:

1 – ПС; 2 – ГС; 3 - ДС

За стена 4 се използва възел 2, който е от тип “СТРЕЛКА”. Според каталога от таблица 3 полученото разпределение е отново ГС.

Правило 7: При получаване на две различни ГС за един и същи обект в резултат на анализ на възел “СТРЕЛКА” стената, разпределена като ГС спрямо възела “СТРЕЛКА”, става ТЕР, ако в тялото до момента няма получена стена с такова разпределение. Ако в резултата на анализа на възел “ВИЛА” в тялото е получена стена, разпределена като ТЕР, то изследваната стена става ПС.

По този начин стена 4 става ТЕР и анализът приключва.

За други възможни пространствени разположения се въвеждат следните правила:

Правило 8: При получаване на две ЛС за един и същи обект и равенство на х координатите на съответните възли тип “ВИЛА”, тази стена, за която у

координата на възела е по-голяма, се определя като горна лява стена (ГЛС), а другата е долна лява стена (ДЛС).

Правило 9: При получаване на две ДС за един и същи обект и равенство на x координатите на съответните възли тип “ВИЛА”, тази стена, за която y координата на възела е по-голяма, се определя като горна дясна стена (ГДС), а другата е долна дясна стена (ДДС).

Правило 10: При получаване на две ПС за един и същи обект и равенство на x координатите на съответните възли тип “ВИЛА”, тази стена, за която y координата на възела е по-голяма, се определя като предна горна стена (ПГС), а другата е предна долна стена (ПДЛС).

Правило 11: При получаване на две ПС за един и същи обект и различни x координати на съответните възли тип “ВИЛА”, тази стена, за която x координата на възела е с по-малка стойност, се определя като предна лява стена (ПЛС), а другата е предна дясна стена (ПДС).

4. Алгоритъм за определяне на пространственото положение на стените

За да се определи пространственото положение на всяка видима стена от обект се извършва следната последователност от действия:

1. Прочитане на информацията за откритите затворени контури в сцената (възли и свързващите ги сегменти).

2. Прочитане на информацията за откритите фоновы контури в сцената (възли и свързващите ги сегменти).

3. Търсене на възли тип “ВИЛА” за всеки затворен контур (стена от обект) и в зависимост от съответстващото им разположение чрез каталога от таблица 2 се определя пространственото положение на стената.

Този анализ се извършва за всеки обект в сцената поотделно. Записват се само не повтарящи се възли “ВИЛА” и то такива, които имат поне един съсед, който е част от фоновия контур за съответната КСГТ (Правило 3).

4. Преразпределение на получените пространствени релации.

За тази цел се използват правилата от Правило 4 до Правило 11 без Правило 7.

5. Проверка дали всички видими стени в сцената са получили пространствено разпределение и ако не е така, за не разпределените стени се прави анализ спрямо участващите в тях възли тип “СТРЕЛКА” с помощта на каталога от Таблица 3.

Отново се записват само не повтарящи се възли “СТРЕЛКА”.

6. Добавяне на ново разпределените стени от възлите тип “СТРЕЛКА” към вече анализирани стени от възли тип “ВИЛА”.

7. Преразпределение на стените, получили разпределение спрямо възел тип “СТРЕЛКА”, ако това е необходимо. Разпределението се съобразява с Правило 7.

8. Проверка дали все още има не разпределени стени в сцената и ако това е така, в тях се търсят възли тип “L”, за които се използва Правило 1.

9. Запис на получената при анализа информация в БД.

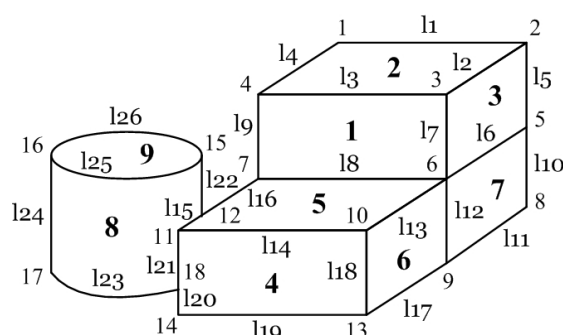
10. Край.

5. Получени резултати

На базата на предложеният подход и описаният алгоритъм е разработена програмна система за анализ на тримерната структура на сцени. Проведени са редица практически експерименти с разработената система. При тях са получени добри резултати за тримерни сцени с различна сложност, започвайки от сцена с един изпъкнал обект до такава, съдържаща вдлъбнат обект и самозакриващ се обект.

На фиг. 3 е представена сцена със съвпадение на възли и ръбове и закриване между обектите в сцената. Един от обектите е с криволинейни повърхнини. Резултатите за нея при прилагане на гореописания алгоритъм са дадени в Таблица 5.

Таблица 5 Получени резултати за сцената от фиг. 3



Фиг. 3. Тестова визуална сцена

обект	контур	Описание на контура	Позиция
1	8	17 16 l23,16 18 l22+l25,18 17 l24+l26:	ПС
1	9	16 15 l25, 15 16 l26:	ГС
2	2	3 4 l3,4 1 l4,1 2 l1,2 3 l2:	ГС
2	1	3 6 l7,6 7 l8,7 4 l9,4 3 l3:	ПС
2	3	3 2 l2,2 5 l5,5 6 l6,6 3 l7:	ДС
3	4	15 10 l14,10 11 l15,11 13 l16+l17,13 15 l19:	ПС
3	6	15 6 l18,6 9 l12,9 10 l13,10 15 l14:	ДС
3	5	15 13 l19,13 7 l20+l21,7 6 l8,6 15 l18:	ГС
4	7	8 9 l11,9 6 l12,6 5 l6,5 8 l10:	ПС

Вижда се, че алгоритъмът дава коректни резултати въпреки сложния характер на сцената.

В бъдеще усилията на автора ще бъдат насочени към разширяване на разнообразието на разглежданите визуални сцени след премахване на някои от налаганите в момента ограничения върху тяхната структура.

ЛИТЕРАТУРА

- [1] Гочев, Г. Компютърно зрение и невронни мрежи, ТУ - София, 1998.
- [2] Angelova M. Structural Hierarchical Grammar for Description of Visual Scenes, Proc. 17th International Conference of Systems for Automation of Engineering and Research, 245-250, Varna, 2003.
- [3] Huffman, D. Impossible Objects as Nonsense Sentences, Machine Intell., vol. 6, Edinburgh Univ. Press, 1971.
- [4] Клоувз, М. Как видеть вещи, Интегральные работы, т. 2, Мир, Москва, 1973.

Автор: Мария Ангелова, гл.ас. д-р в кат. "Програмиране и компютърни технологии" към Факултет по Компютърни системи и управление.

ТЕХНОЛОГИЧНО ИЗГОТВЯНЕ НА КОМПЮТЪРЕН МИКРОМОДУЛ

Валентин Видеков, Славка Цанова, Димитър Нурков

Резюме. В статията се представя конструктивно технологично оптимизиране и изготвяне на компютърен микромодул. Модулът представлява микросистема със специализирано приложение за обработка на данни и предаването им на разстояние. Разгледани са основните стъпки при топологичното проектиране и последващото технологично изпълнение чрез технологията на повърхностния монтаж.

TECHNOLOGICAL PRODUCTION OF A WEARABLE COMPUTER MICROMODULE

Valentin Videkov, Slavka Tzanova, Dimitar Nourkov

Abstract. *The paper presents the optimization and technological development of computer micro modules. The module represents a microsystem with specific application for sensing, data processing and transmission. The basic steps in topological design and fabrication on technology of surface mounting are discussed.*

1. Увод

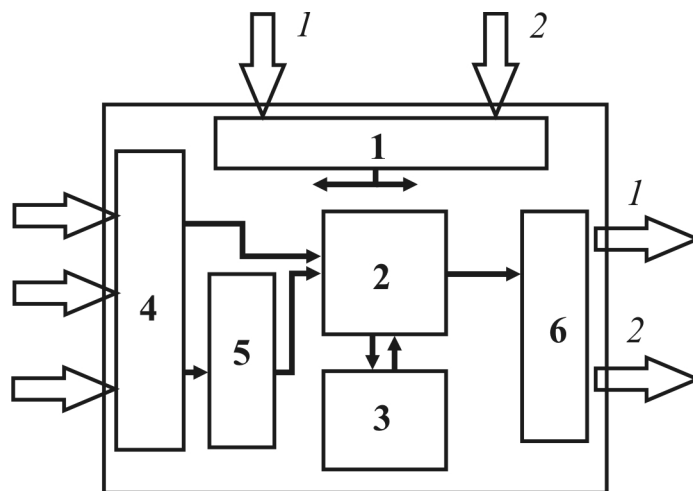
През последните години освен стандартната тенденция за намаляване размерите на електронните изделия [1] и функционалното усложняване на интегралните схеми [2], се появиха нови тенденции в приложението на микромодулите. Едно от тези приложения е създаването на устройства за събиране и обработка на информация по определени критерии реализирани като микросистеми за автономно преносимо приложение [3]. Един от инициаторите на това развитие бе медицинската електроника изискваща непрекъснат контрол от разстояние на различни параметри. Наличието на достатъчно миниятюрни сензори, памети, процесори и други елементи, и развитието на техниката за монтаж на системи позволи реализиране на микросистеми с автономно функциониране и преносимо приложение. Една такава система е [4] представляваща микромодул за обработка на звукови, светлинни, инерционни въздействия и предаване на разстояние на получените резултати.

2. Микрокомпютър за преносимо предназначение

Основните блокове съставлящи микрокомпютъра са блок за събиране на данни от външни въздействия, микропроцесор, захранване, памет и блок за извеждане на данните. В зависимост от приложението блокът за събиране на

данни може да бъде съставен от различни сензори: например микрофони, фотодетектори в различен спектрален обхват, измерители на ускорение и други. Извеждането на данните може да се извършва непрекъснато или по заявка. В зависимост от предназначението самото извеждане може да става контактно (по проводна схема) или безконтактно (по радиоканал).

На фиг. 1 е показана блоковата схема на такъв микрокомпютър.



Фиг. 1. Микрокомпютър за преносимо приложение

Захранването **1** може да се осъществи чрез акумулатор, батерия или проводно **1**, а също така и дистанционно **2**, чрез използване на електромагнитно поле или фотоелементи. Управлението се извършва от микропроцесор **2**, свързан с останалите блокове. Паметта **3** може да се използва за съхранение на данни или предварително зададени функционални приложения. Сензорните елементи **4** могат да бъдат с цифров изход и тогава се свързват директно с процесора, или ако са аналогови преминават през аналого-цифров преобразувател **5**. Външния интерфейс **6** може да бъде контактен **1** или по радиоканал **2**.

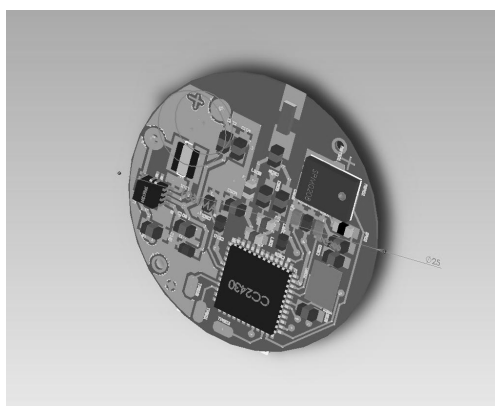
На фиг. 2 е показано конструктивното решение на такава микросистема, реализирана като свързване на два микромодула.



Фиг. 2. Конструктивно решение на базовия вариант

3. Конструкция на подобрен вариант

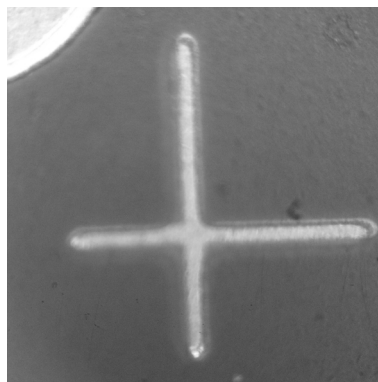
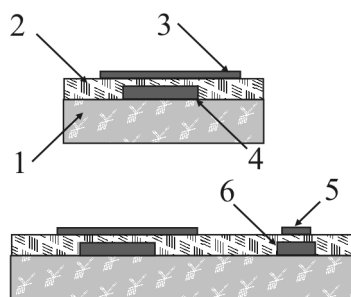
Основната цел на подобрената конструкция е да се намалят размерите на устройството и броя на технологичните операции при изготвянето му. В базовия образец имаме два технологични цикъла за изготвяне на двата микромодула и операции за монтаж на двата модула в един. Първата стъпка в оптимизирането е чрез преглед на елементната база да се намерят елементи изпълняващи същите функции, но реализирани в по-малък корпус. Втората стъпка е да се намерят елементи изпълняващи функции на няколко елемента от базовия образец. Трета стъпка е да се направи подобрена схемно решение. След горните три стъпки се преминава към конструктивно проектиране на новия модул. На фигура 3 е показан примерен изглед на новия модул като резултат от проектирането чрез програмния продукт „Altium Designer”.



Фиг. 3. Конструктивно решение на оптимизирания вариант

При оптимизирането на конструкцията бяха постигнати и двете цели. Същата функционалност се изпълнява от един микромодул изпълнен чрез двустранен повърхностен монтаж при което се намаляват технологичните процеси от 5 на 2. През целия етап на конструктивно проектиране е проследявана технологичността на модула, така че тя да позволява реализация в катедра „Микроелектроника”. Освен изискванията за функциониране на елементите (например антената) бяха предвидени специални маркери позволяващи прецизно съвместяване на фотошаблоните при отчитане на технологичния слой – фигура 4.

При проектирането му бяха използвани изисквания от технологията микромодулите и приложени за технологична схема с повърхностен монтаж



Фиг. 4. Специален маркер за точно съвместяване (30 μm)

Подложката от FR4 (1) се покрива с фоточувствителна епоксидна маска (2), като тя покрива и топологията на контактната площадка 4. Отвора около площадката трябва да е по-голям от нея, което изисква топологичен елемент от фотошаблона (3), който да е съответно по-голям по размери. Като се отчете, че маската е негативна, то шаблона трябва да е непрозрачен, което означава невъзможност за наблюдаване на контактната площадка. Следователно е необходим специален елемент за съвместяване формиран върху подложката (6) и елемент от шаблона за маската (5). Размерите на последния са съобразени с технологичното подецване на елемента (6) и визуалната различимост през нанесената маска, която е цветна. В дясно е показан резултата от прилагането на проектирания елемент.

4. Технологично изготвяне

Преди изготвянето на шаблоните за микромодула бяха проведени серия от тестове за обределяне параметрите на фотомаската. Същата се изготвяше



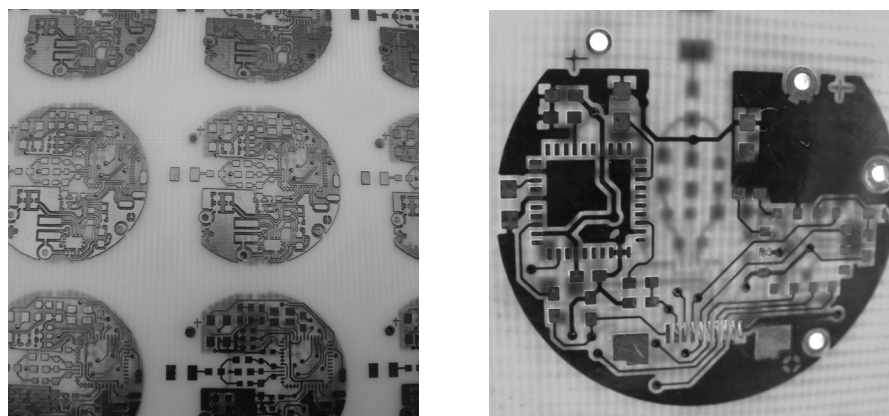
Фиг. 5. Тест за определяне качеството на фотомаските.

върху гъвкава основа и експониране чрез лазерен лъч. Бяха определени плътността на експонирания слой, назъбеността на ръба и концентрацията на дефекти. На фиг. 5 е показан един от тестовите елементи наблюдаван през микроскоп и увеличение $\times 100$. Назъбеността на ръба бе от порядъка на $5\text{ }\mu\text{m}$, което при използвани фотолакове с дебелина $30\text{ }\mu\text{m}$ и топологични елементи с минимален размер $100\text{ }\mu\text{m}$ се отчете, че няма да доведе до технологични проблеми. При подходящо експониране и проявяване на фотолака и епоксидната маска се получава

интегриране (заглаждане) на ръба.

Технологично микромодулите се изготвяха по класическа субтрактивна технология върху FR4 с двустранна медна метализация. Свързващите отвори са с диаметър $350\text{ }\mu\text{m}$ след метализация. Топологията се формира чрез фотолитографска обработка. Фотолака е сух с дебелина $30\text{ }\mu\text{m}$ и се нанася чрез двустранно ламиниране. Следва експониране с живачна лампа (100 W) през двустранна маска. Същата предварително е съвместена и оформена като джоб. Експонирането на двете страни е последователно едно след друго и след това е общото проявяване. Проявяването е при $30\text{ }^{\circ}\text{C}$ в 1% разтвор на натриев карбонат. Проявяването се контролира визуално, като осветлението е филтрувано (жълто-зелен филтър). Следва промиване в дестилирана вода. На фиг. 6 са показани фотографии от различни етапи на обработка.

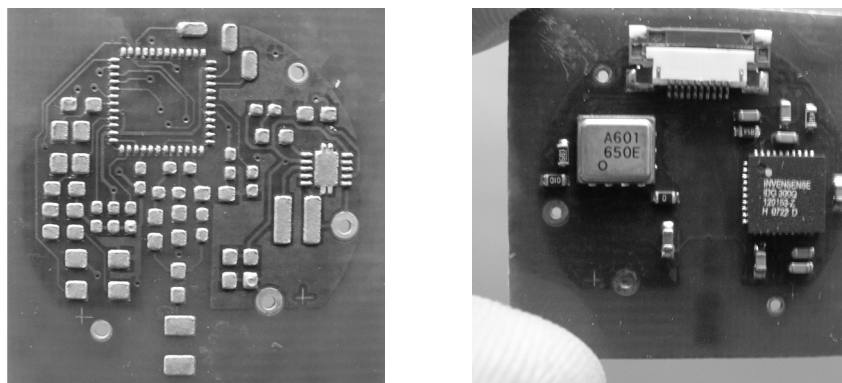
Основата на модула след проявяване се вижда в лявата част на фиг. 6, а след сваляне на фотолака е показана в дясно. Подложката е $0,5\text{ mm}$ и се виждат двата слоя мед.



Фиг. 6. Микромодули след ецване и модул след свален фотолак

Следващите операции са свързани с нанасянето на защитната маска. Ецва се и след експонация се проявява. След това се извършва аналогична обработка за втората страна. След оптичен контрол се извършва термично стабилизиране на епоксидната маска.

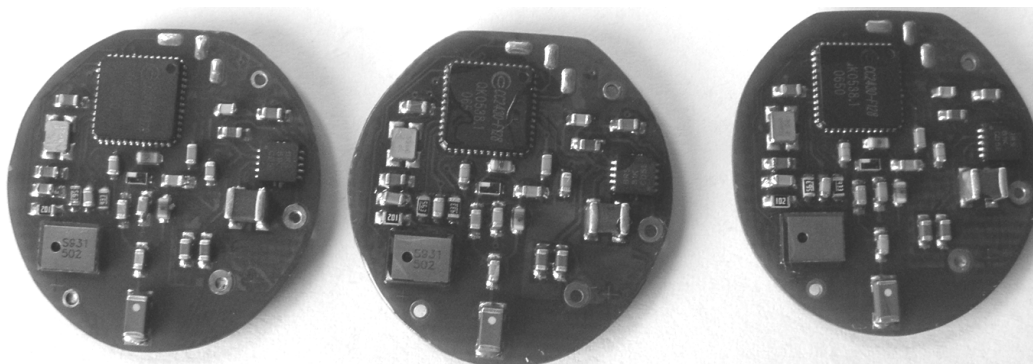
Монтирането на елементите се извършва след нанасяне на спойваща паста през метална маска. Нанасянето се извършва на ръчен принтер и с метален ракел. Обработват се едновременно 9 образеца. Първоначално се нанася паста от едната страна и се разполагат елементите. Разполагането се извършва ръчно под микроскоп. Следва конвекционно спояване на първата страна. На фигура 7 е показана нанесена спойваща паста и монтираните елементи от едната страна.



Фиг. 7. Модул с нанесена спойваща паста и модул със споени елементи

При втората страна нанасянето на пастата е аналогично, но под подложките се поставят опори за да повдигнат нивото (от долу има елементи). При спояването на втората страна се отчита и вертикалният градиент на температурата с цел неразпояване на елементите от долната страна.

На фигура 8 е показана втората страна на микромодула. В левия ъгъл долу се вижда микрофона, а в долната част е чипът на антената.



Фиг. 8. Монтирани елементи от втората страна на микромодули

Получените микромодули са защитени допълнително с лаково покритие за по-висока устойчивост. Специално внимание е отделено на защитата на микрофона и антената.

5. Изводи

Разглеждането на представеното електронно изделие, като единна конструктивно-технологична задача позволи да бъдат намалени технологичните операции при изготвянето му и същевременно да се намали обема му. От съществено значение при решаването на такава задача е използването на системния подход и то при отчитане на конкретните технологични възможности.

В дадения случай е показана възможността да се създаде микросистемен модул при използване на съществуващото оборудване и технологични възможности на катедра „Микроелектроника“.

ЛИТЕРАТУРА

- [1] Sundaram, V. Tummala, R.R. Fuhan Liu Kohl, P.A. Jun Li Bidstrup-Allen, S.A. Fukuoka, Y. , Next-generation microvia and global wiring technologies for SOP , Advanced Packaging, IEEE Transactions on, May 2004, V. 27 I. 2 pp. 315 - 325
- [2] Rajendran, R. Devarajan, N., FPGA Implementation of Space Vector PWM Technique for Voltage Source Inverter Fed Induction Motor Drive , Computer and Electrical Engineering, 2009. ICCEE '09. Second International Conference on, 28-30 Dec. 2009, Dubai, ISBN: 978-1-4244-5365-8, pp. 422 - 426
- [3] Jongtae Yu, Chengqi Guo, Wearable Computing as next Interaction Technology and its Value for Virtual Team, <http://www.decisionsciences.org/Proceedings/DSI2008/docs/529-2371.pdf> 2010,11
- [4] N.B. Bharatula and S. Ossevoort and M. Staeger and G. Troester, “Towards Wearable Autonomous Microsystems,” Proceedings of 2 nd International Conference on Pervasive Computing, Vienna,Austria, April 21-23,2004

Автори: Валентин Видеков, д-р инж. доц., Славка Цанова д-р инж. доц., катедра „Микроелектроника, Димитър Нурков, инж., докторант във ФФОЕ

Постъпила на 17.01.2011

Рецензент доц. д-р Радосвет Арнаудов

ПРЕВКЛЮЧАЩ ЕФЕКТ В НАНОПОРИСТ ОКСИД НА АЛУМИНИЯ ПРИ АНОДНА ПОЛЯРИЗАЦИЯ

Игор Врублевски, Катерина Чернякова, Валентин Видеков, Боряна
Цанева, Александър Ячовски

***Резюме.** Разгледан е ефект на превключване при анодна поляризация на нанопорист аноден оксид на алуминий. Оксида е получен в потенциостатичен режим във воден разтвор на фосфорна киселина. Измерванията показват възпроизводимост на резултата. Коментирано е изменението съпротивлението на пората за приложение като енергонезависими памети.*

SWITCHING EFFECT IN NANOPOROUS ALUMINA AT THE ANODE POLARIZATION

Igor Vrublevski, Katerina Chernyakova, Valentin Videkov, Boryana Tzaneva,
Alexander Yachovski

***Abstract.** Switching effect of porous alumina has been studied by the anodic polarization of structures electrolyte-anodic alumina-aluminum. The anodic alumina films were formed in the aqueous solutions of phosphoric acid in the potentiostatic mode. In accordance with our measurements switching characteristics of anodic alumina were of the high reproducibility. Made a comment about the change of resistance of the pores for applications in nonvolatile memory*

1. Увод

Пористия оксид на алуминия, като наноструктуриран материал, намира широко приложение в нанотехнологиите, например за изготвяне на различни нанокомпозити, пористи мембрани и маски за фотонни, електронни и оптоелектронни системи [1]. В тези системи се използва и такова свойство на анодния оксид, като високорегулярната пориста структура. От друга страна особената двуслойна структура на пористия оксид води до необикновени свойства. Външният слой оксид прилежащ към пората е „легиран” с аниони от електролита, а вътрешният е относително чист.

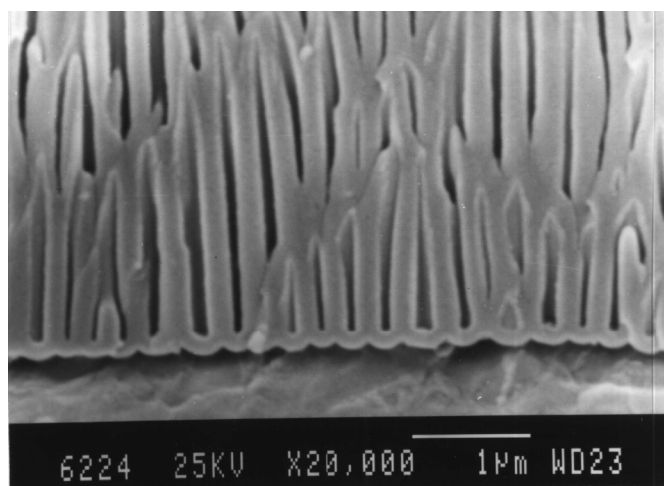
Неотдавна бе показано, че при реанодиране на слоеве от аноден оксид на алуминий е възможна появата на отрицателно електрическо съпротивление [2]. Авторите в [3] показват открити бистабилни превключващи свойства и ефект на памет в анодния оксид на алуминия, които могат да се използват за създаване на енергонезависими памети. Предполага се, че ефекта на памет е свързан с изменението съпротивлението на анодния оксид, предизвикан от натрупването на електрически заряд при протичането на електрически ток.

Такава памет представлява интерес поради възможността за разработка на ново поколение универсална високоплътна енергонезависима памет. В работата се разглеждат резултатите от изследването на превключващите свойства на анодния оксид на алуминия при анодна поляризация на структурата електролит – оксид – метал.

2. Експериментална част

Ленти от алуминий, получени от алуминиево фолио с чистота 99,99% се обезмасляваха в горещ разтвор на натриев хидрооксид, след това щателно се промиваха в дестилирана вода и се изсушаваха на въздух. Образците се анодираха при постоянно напрежение в 0,41M воден разтвор на фосфорна киселина в продължение на 6 минути. Температурата на електролита се поддържаше постоянна на $(18 \pm 0,5) ^\circ\text{C}$. Напрежението на клетката се подаваше с използване на потенциостат-гальваностат Р582М. Слоеве се израстваха при потенциал 43 V. След анодирането слой порист оксид на алуминия отново щателно се промиваше с дестилирана вода. Една част от образците се прогряваха на въздух при 200°C в продължение на 1 час. В същата електрохимична клетка в режим на развивка на анодния потенциал със скорост 2,6 V/min се провеждаха поляризационни измервания на отгряти и прясно анодирани слоеве.

Напречното сечение на пористия оксид на алуминия, формиран в разтвор на фосфорна киселина, се изследваше с помощта на сканиращ електронен микроскоп JOEL 840А. На фигура 1 е показано напречното сечение на порист оксид на алуминия върху алуминиево фолио.

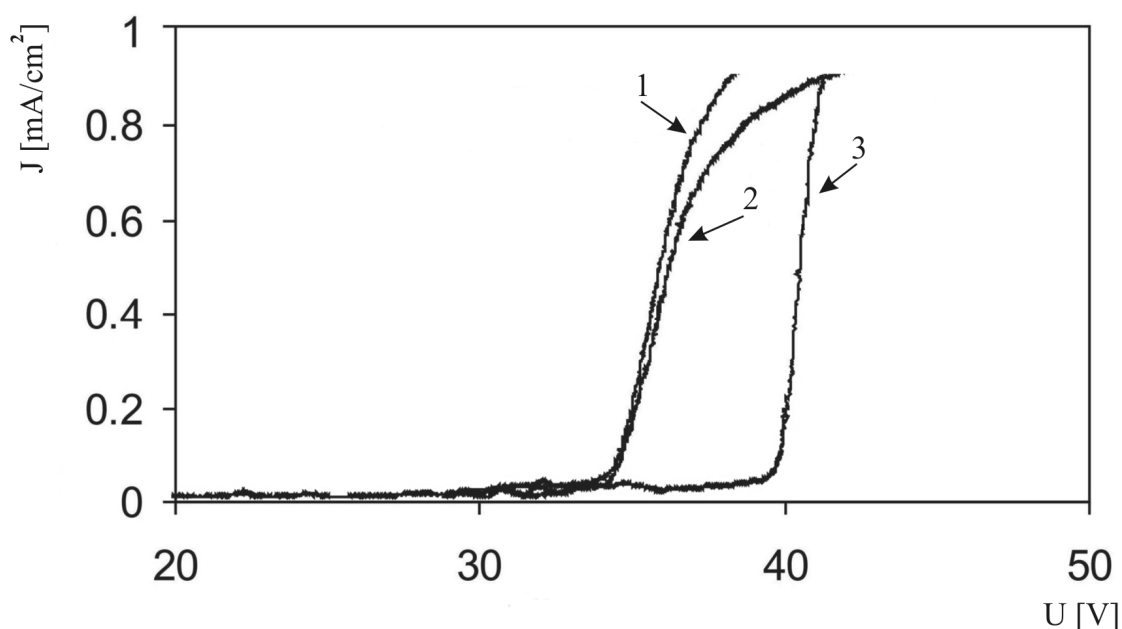


Фиг. 1. Напречно сечение на слой от порист аноден оксид на алуминий получен във воден разтвор на фосфорна киселина

Както се вижда от фигурата пористия слой има достатъчна регулярна клетъчна структура.

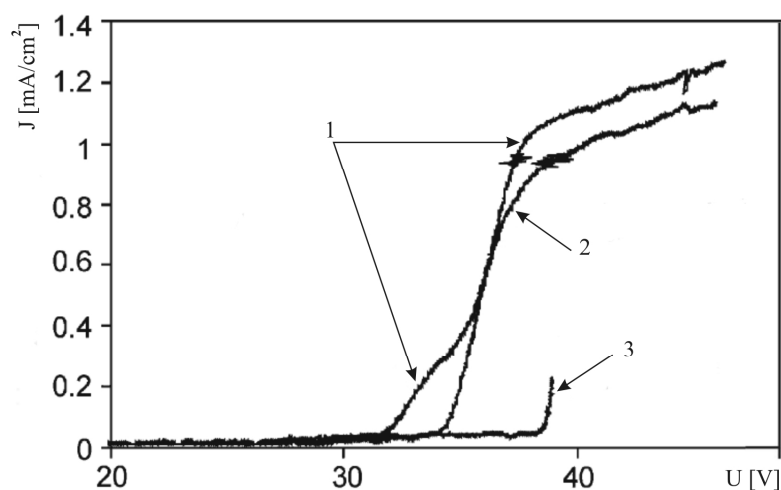
3. Резултати

Ефектът на превключване в пористия оксид се изследваше с използване на анодна поляризация както на отгряти, така и на прясно анодирани слоеве от аноден оксид. За целта се използваше двустепенна анодна поляризация на отгряти слоеве, резултатите на които се сравняваха с поведението на прясно анодирани слоеве. За първия стадий на анодна поляризация се използваше режим на развивка на напрежението до момента на спиране рязкото нарастване на анодния ток. След това поляризационните измервания се повтаряха отново при развивка на напрежението до 0V. На фигура 2 е показан ефекта на превключване в слоеве от порист алуминиев оксид получени във воден разтвор на фосфорна киселина, при анодна поляризация.

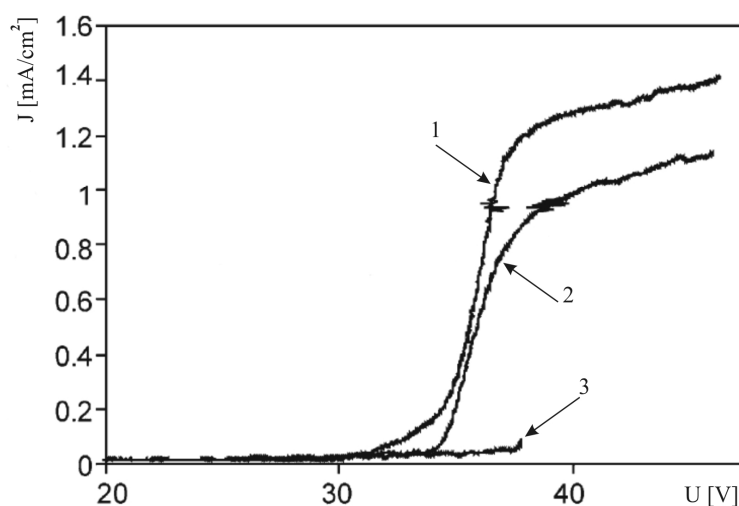


Фиг. 2. Ефект на превключване в слоеве нанопорист алуминиев оксид при анодна поляризация. 1 – прясно анодиран оксид; 2 – отгрят оксид (втори етап); 3 – отгрят оксид (първи етап)

Може да се забележи следната обща закономерност при поляризационното поведение на анодния алуминиев оксид – участък на стръмно нарастване на анодния ток с последващ преход към стадий на плавно нарастване при приближаване на анодния потенциал до напрежението на формиране на слоя. За отгритите слоеве началото на участъка с плавно нарастване на анодния ток съвпада с напрежението на формиране на анодния слой при 43 V. За получаване на непълна поляризация на анодните слоеве при експериментите се използваше поляризационен процес, който се прекъсваше в началото и в средата на участъка на бързо нарастване на анодния ток. На фигури 3 и 4 са показани поляризационните криви на слоеве от порист алуминиев оксид, прекъсвани на различни етапи от първата поляризация.



Фиг. 3. Поляризационно поведение на слоеве от нанопорест алуминиев оксид при прекъсване на тока на първи етап на поляризацията при $0,23 \text{ mA/cm}^2$. 1 – отгрят оксид на втори етап; 2 – прясно анодиран оксид; 3 – отгрят оксид на първи етап



Фиг. 4. Поляризационно поведение на слоеве от нанопорест алуминиев оксид при прекъсване на тока на първи етап на поляризацията при $0,09 \text{ mA/cm}^2$. 1 – отгрят оксид на втори етап; 2 – прясно анодиран оксид; 3 – отгрят оксид на първи етап

Непълната поляризация на отгрити слоеве след първи етап довеждаше до изменение на поляризационното поведение на тези слоеве на началния стадий при втория етап на поляризация. Анодния ток възникваше при по-ниско напрежение, отколкото при прясно анодираните слоеве и се характеризираше със слабо повишаване на анодния потенциал. При това обаче участъка на бързо

нарастване на анодния ток съвпаднаше с аналогичния участък за прясно анодирани слоеве. Може да се предположи, че такова различие е резултат от преходни процеси възникващи в непълно поляризирани отгтрети слоеве от алуминиев оксид.

Ефектът на превключване в порестия алуминиев оксид при анодна поляризация свидетелства за наличие на постоянна поляризация в прясно анодиран оксид. В случая на прясно анодирани слоеве токовия пик се появява при по-ниско напрежение отколкото при отгтрети слоеве. Може да се предположи, че причината за тази разлика в напреженията е свързана с вътрешно поле, което присъства в прясно анодирания слой. Вътрешното поле се сумира с външното електрическо поле, което води до увеличаване на резултатната напрегнатост на полето. При отгтрети слоеве напрежението на токовия пик при втората поляризация съвпада с напрежението на прясно анодиран слой, ако поляризационния процес на първия етап се прекъсва веднага след прекратяване на бързото токово нарастване. В този случай видът на поляризационната крива на отгтретия слой напълно повтаря кривата на прясно анодиран слой. Следователно можем да предположим, че заряда вграден в прясно анодирания слой изчезващ след отгряването, отново възниква след зъвършване на анодната поляризация.

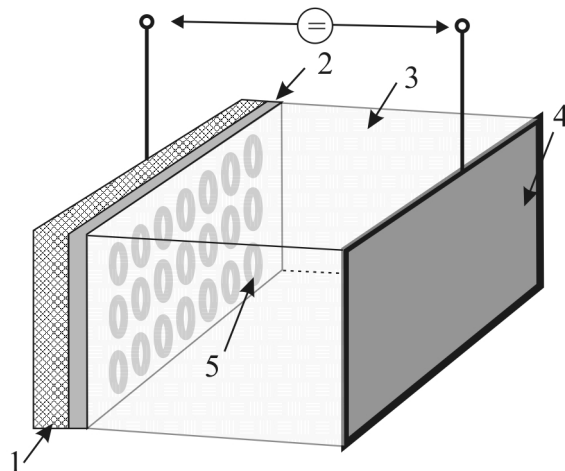
Както е показано в [4, 5], при прясно анодирани слоеве имаме запълнени електронни центрове в повърхностния слой. Отгряването на такива слоеве при 200°C води до освобождаването на електроните от центровете и като резултат отгтретия оксид има празни електронни центрове в повърхностния слой.

Пропускането на електрически заряд през отгтрят слой довежда до изменение на електрическото съпротивление. При анодна поляризация под действие на електрическото поле с висока напрегнатост и инжекция на електрони от електролита се променят електронинните свойства на алуминиевия оксид – запълване на електронните центрове образувани в повърхностния слой. В резултат токовата характеристика на такъв аноден слой се възстановява до характеристиката на прясно анодиран слой. Изменението на електронните свойства се придружава от намаляване на електрическото съпротивление в няколко порядъка. При изключване на източника на ток анодният слой запазва своето ново състояние възникнало след преминаване на електрическия заряд.

Изследвания за приложение на алуминиевия оксид като памети се извършват в различни насоки, както чрез запълване на порите с различни материали така и използвайки характеристиките на оксида [6, 7]. Предимство на наноструктурирания порест алуминиев оксид е това, че всяка елементарна клетка може да бъде използвана като превключващ елемент. Това открива перспективи за използването му като енергонезависими памети с висока плътност на записа.

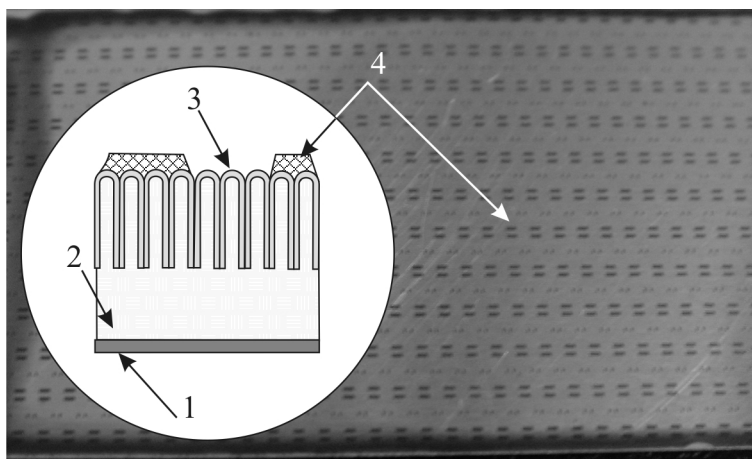
4. Конструкция

В разгледаните експерименти характеристиките са резултат от интегрално измерване по схемата от фигура 5.



Фиг. 5. Интегрално измерване на характеристиките на оксида. 1 – алуминий; 2 – оксид; 3 – електролит; 4 – метален електрод; 5 – пора

Получаването на единичен контакт между алуминий, пора от оксид, електролит и метал е затруднено. Това изисква прецизна фотолитография на нанониво и съответен монтаж на изовод. Междинен вариант може да бъде създаване на микроконтакт обединяващ няколко пори. На фигура 6 е показана конструкция на този вариант. Фотолитографията е извършена по стандартна схема с използване на негативен фоторезист.



Фиг. 6. Микро-интегрално измерване на характеристиките на оксида. 1 – метален електрод; 2 – електролит; 3 – оксид; 4 – алуминий

След получаване на осидния слой върху тилната страна на алуминиевото фолио се извършва фотолитография и селективно се ецва алуминия за получаване на микроостровчета. Същите обхващат няколко десетки пори. Контактването към контакта 4 от тилната страна се извършва чрез залепване на микроизвод с проводящо лепило. При измерванията характеристиките запазват вида си с леко променяне на наклона.

5. Изводи

За изследване превключващите свойства на порестия алуминиев оксид може да бъде използван процеса на анодна поляризация на структура електролит-окис-метал. Предварителните изследвания на електрическите свойства върху специално подготвени образци на порест алуминиев оксид показаха ефект на превключване с изменение на съпротивлението от няколко порядъка. Тези свойства са особено перспективни за създаване на енергонезависими памети.

Тези памети могат да бъдат реализирани както на нанониво с размер на единична пора, така и като микро структура от няколко пори.

ЛИТЕРАТУРА

- [1] Sulka, G. D. Highly ordered anodic porous alumina formation by self-organized anodizing. Nanostructured Materials in Electrochemistry. Wiley–VCH GmbH & Co. KGaA, Weinheim. 2008. P. 1–116.
- [2] Jagminas, A., A. Čečūnienė, I. Vrublevsky, V. Jasulaitienė, R. Ragalevičius. Behavior of alumina barrier layer in the supporting electrolytes for deposition of nanowired materials. Electrochim. Acta. 55. 2008. P. 3361–3367.
- [3] Lee, J., S. Nigo, Y. Nakano, S. Kato, H. Kitazawa, G. Kido. Structural analysis of anodic porous alumina used for resistive random access memory. Sci. Technol. Adv. Mater. 11. 2010. P. 025002–025006.
- [4] Vrublevsky, I., A. Jagminas, J. Schreckenbach, W. A. Goedel. Electronic properties of electrolyte/anodic alumina junction during porous anodizing. J. Appl. Surf. Sci. 253. 2007. P. 4680–4687.
- [5] Vrublevsky, I., V. Parkoun, J. Schreckenbach, W. A. Goedel. Dissolution behaviour of the barrier layer of porous oxide films on aluminum formed in phosphoric acid studied by a re-anodizing technique. J. Appl. Surf. Sci. 252. 2006. P. 5100–5108.
- [6] Nakata Shunjiq, Non-volatile Al₂O₃ Memory using Nanoscale Al-rich Al₂O₃ Thin Film as a Charge Storage Layer, Japanese Journal of Applied Physics. Part 2. Letters vol: 45 issue: 4B page: 3176-3178 year: 2006
- [7] Lin Chih-Yang ; Wu Chen-Yu ; Wu Chung-Yi ; Chenming Hu ; Tseng Tseung-Yuen, Bistable resistive switching in Al₂O₃ memory thin films, Journal of the Electrochemical Society ISSN 0013-4651 2007, vol. 154, N9, pp. 189-192

Автори: Игорь Врублевски д-р инж. ст.н.с. кат. „Микро и наноелектроника” БДУИР, Екатерина Чернякова д-р инж. кат. „Химия” БДУИР, Валентин Видеков д-р инж. доцент , кат. „Микроелектроника”, Боряна Цанева д-р инж. гл. ас. кат. „Химия”, Александър Ячовски инж. маг. докторант катедра „Микроелектроника”

Постъпила на 17.01.2011

Рецензент доц. д-р Ива Бетова

КОНТРОЛ НА ПРОЦЕСА АНОДИРАНЕ ЧРЕЗ МИКРОКОНТАКТНИ СТРУКТУРИ

Александър Ячовски

Резюме. Разгледани са основни приложения на анодния оксид и контролирането на технологичния процес чрез измерване на микроконтактни структури. Разгледан е метод за контролиране равномерността на анодирането при получаване на наноструктурирани мембрани от аноден оксид на алуминия като е измервано контактното съпротивление. Запълването е с електролитно отложена мед.

CONTROL OF THE PROCESS ANODIZING BY MICRO STRUCTURES

Alexander Yachovski

Abstract. The basic applications of the anode oxide and control of the technological process by measuring the micro contact structures are described. A method for controlling uniformity of anodizing in obtaining nanostructured membranes of anode aluminum oxide with following measurement of the contact resistance is also described. The filling is done using electrolytically deposited copper.

1. Увод

Приложението на анодния оксид на алуминия в електрониката е известно от десетилетия. Известни са приложения за електролитни кондензатори [1], като подложка за микромодули с приложение в различни честотни области [2, 3], а също така и за големи мощности [4]. Особен интерес представлява използването на анодния оксид на алуминия като материал в областта на нанотехнологиите. В [5] са разгледани широк кръг въпроси от прилагането му като матрица за създаване на нанокomпозитни материали, а също така и като наноматрица с различни приложения.

Същевременно в редица устройства на микроелектрониката се използват контактни системи с размери в микронната и под микронна област [6], включително и изводи от матричен тип [7]. Някой от тези контакти се изготвят чрез LIGA процес [8], съчетаващи фотолитографски техники за получаване на отвори и запълването им с метал. Съществува вариант за микроконтактен извод изготвян чрез електрохимичен процес, обединяващ групово създаване на наноизводи формиращи микроконтакт [9]. Наноструктурираната матрица позволява създаване на анизотропни материали.

Във връзка с горното интерес представлява възможността да се контролира равномерността на получения оксид след технологичното му изготвяне чрез използване на микроконтакти.

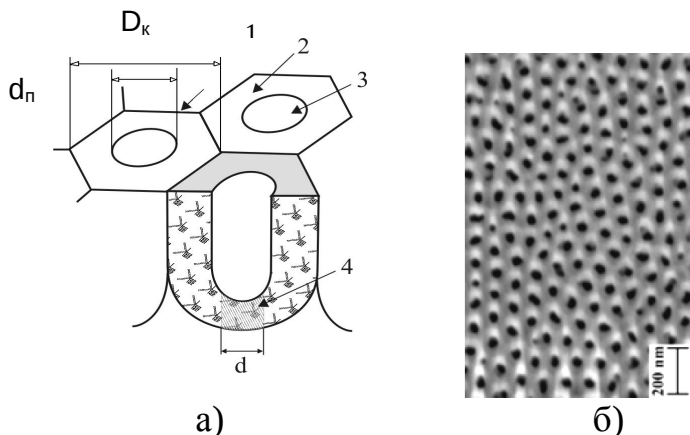
2. Аноден оксид на алуминия

Процесът анодиране на алуминия в различни електролити води до изграждане на собствен окис израстващ от подложката. Този процес е известен отдавна и широко се прилага в техниката. Уникалното в случая е получаването на самоструктурирана матрица от елементарни клетки окис, като и до днес се откриват нови свойства. При получаването на окиса могат да се използват различни състави на електролити и различни режими – таблица 1.

Таблица 1

Електролит	J=30A/м2			α	J=50A/м2			α
	pH	Уст. В	та мин.		pH	Уст. В	та мин	
H ₂ SO ₄ (5%)	0,5	12,3	19,7	0,326	0,5	17	11,75	0,318
H ₂ SO ₄ (10%)	0,3	9,4	19,8	0,333	0,5	13	11,74	0,317
H ₂ SO ₄ (10%)	1,0	15,4	19,6	0,319	1,0	22	11,76	0,319
(COOH) ₂ (2%)	1,0	20	19,5	0,313	1,0	30	11,83	0,328
(COOH) ₂ (2%)	1,8	33	19,7	0,326	1,8	47	11,66	0,307
(COOH) ₂ (4%)	0,9	18	19,6	0,319	0,9	27	11,7	0,312
(COOH) ₂ (4%)	1,6	29	19,5	0,313	1,6	39	11,8	0,324
H ₃ PO ₄ (4%)	1,3	48	19,6	0,319	1,3	85	11,67	0,308
H ₃ PO ₄ (4%)	2,0	68	19,7	0,326	2,0	100	11,72	0,315

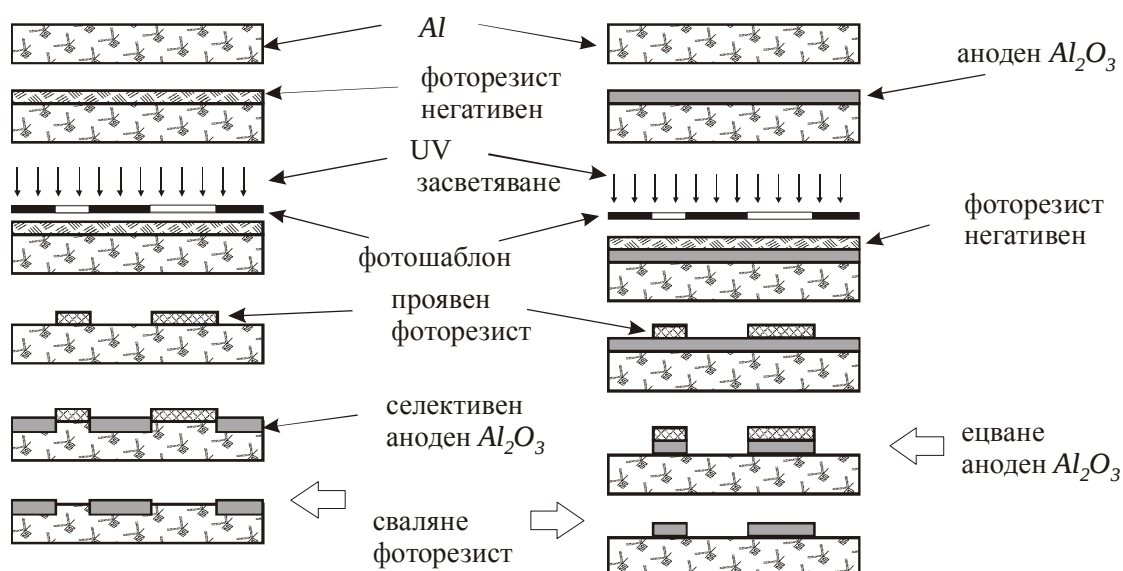
При анодизирането могат да се променят различни параметри на оксида и неговата структура. Параметърът α е отношението на диаметрите на пората и клетката – фиг. 1. Структурата му се състои от елементарни клетки (фиг. 1) които от своя страна също са съставни. Основните параметри са диаметър на клетката D_k и диаметър на пората d_n . Клетката се формира от два типа оксид, като вътрешният 2 е по-чист от замърсявания и по-дебел. В центъра на пората се намира отвор 3. Съществуват технологични методи при които дъното 4 на пората може да бъде отворено.



Фиг. 1. Структура на аноден Al₂O₃ а) и поглед отгоре б)

Променяйки режимите на израстване можем да променяме диаметъра на порите, дебелината на оксида и свойствата му. На практика анодния оксид се явява наноструктуриран материал. Дебелината (височината на порите) също може да се променя в широки граници и варира от части на микрона до няколко стотин микрона.

Съществено предимство на процеса за анодиране на алуминия е, че същият се съчетава отлично с планарната технология в микроелектрониката. Възможни са и двата процеса адитивен и субтрактивен. Принципът на селективното анодиране можем да разглеждаме като вариант от LIGA технология. На фиг. 2 са показани двата варианта за структуриране на анодния оксид.



Фиг. 2. Основни схеми за топологично структуриране на аноден Al_2O_3

Използвайки горните схеми и съчетавайки ги с технологии за запълване на порите могат да бъдат реализирани различни елементи в микро и наноелектрониката.

3. Приложения на оксида

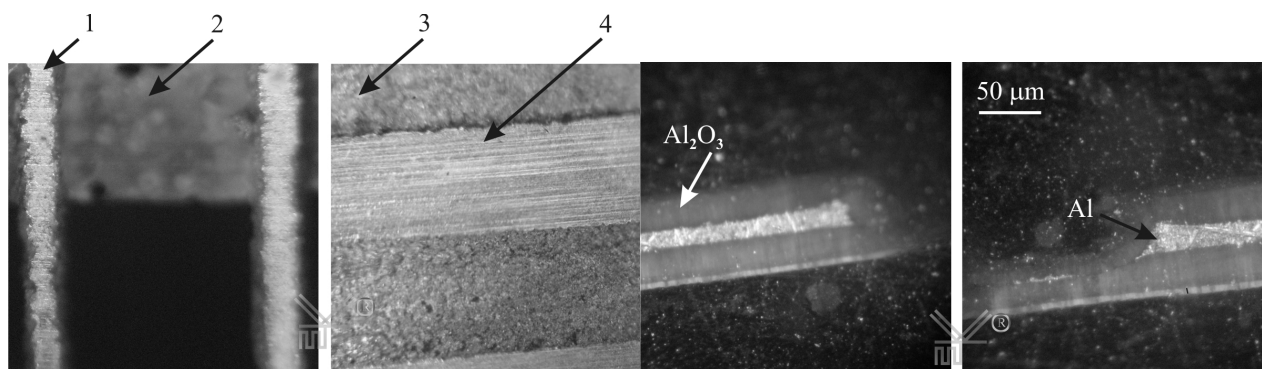
Едно от първите приложения на анодния оксид на алуминия е като диелектрик в хибридните интегрални схеми и модули. Използват се както изолационните му свойства при реализиране на едно или многослойна метализация, така и при реализиране на вградени пасивни елементи (резистори и кондензатори). Тук основната задача е повишаване плътността на окиса и запълването на порите с цел получаване на по-добро пробивно напрежение и управление на диелектричната проникваемост. Влияние оказват както изходния материал така и режимите на израстване. Контролират се дебелината на покритието, загуби, изолационно съпротивление, концентрация на примеси и други [XX].

Анодният оксид може да бъде използван и като конструктивен елемент в различни микроелектромеханични устройства[10] във формата на различни

мембрани и греди. Мембраните могат да бъдат реализирани изцяло от оксид или чрез вградени метални ядра [11].

За проверка на предложените технологични схеми са проведени експерименти с алуминиево фолио и чистота 99%. От него се изготвят основи с размери 2x5 cm. Същите са използвани за изготвяне на мембрани. Работи се с електролит 5% разтвор на оксалова киселина и потенциал 40 волта. След израстване на оксид от порядъка на 5 до 15 μm се провежда фотолитография с фоторезист AZ1350J нанесен на центрофуга и стандартен процес за обработка. Ецването се извършва в селективен етчер на основата на фосфорна киселина и хромен триоксид при температура 80°C. Контролирането на ецването се извършва визуално чрез междинно промиване в дестилирана вода.

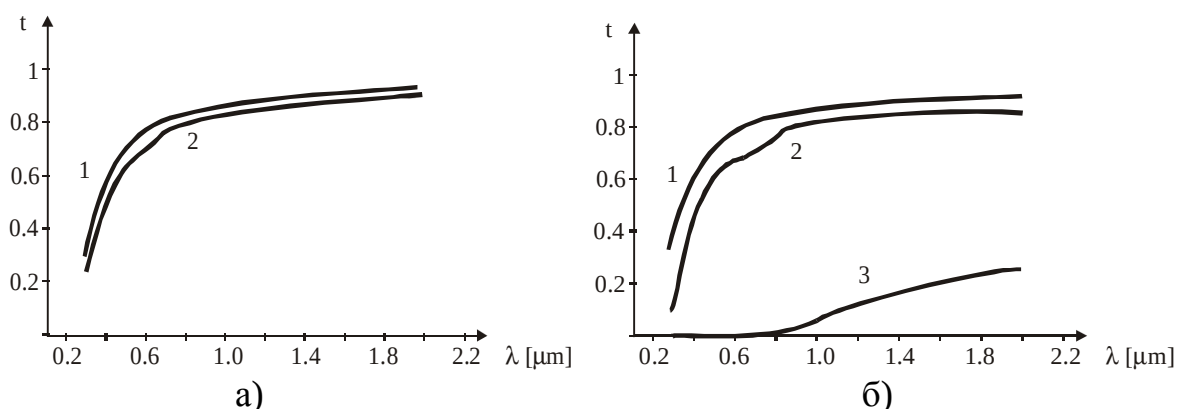
Съществуват два варианта за формиране на мембрани с метално ядро. При вариантите трябва да се отчита необходимостта за електрически контакт на ядрото необходим за анодизирането му. На долната фигура 3 е показан вариант с ецване на алуминия 1 до нивото на оксида 2 който е топологично структуриран. Алуминият има връзка с общ електрод позволяващ последващото му анодизиране. При вторият образец имаме частично ецнат алуминий 3 с дебелина около 60 μm и област 4 която не е ецвана.



Фиг. 3. Елементи на мембрани: 1 – Al; 2 - Al_2O_3 ; 3 – ецван Al; 4 – Al

Отварянето на дъното на порите превръща мембраната в нанофилтър или основа за наномрежа с огромна контактна повърхност. В редица случаи горните приложения са свързани с определяне на параметри на оксида. Същите се определят от състава на алуминия, електролита и режимите на анодиране [10]. В зависимост от приложението се търсят и различни параметри. Например при използване за високочестотни приложения е важна диелектричната проницаемост, която зависи от остатъчното съдържание на електролит в порите и т.н. За такива цели анодирането се провежда при ниски температури, с минимална концентрация на електролит и големи плътности на тока. При други приложения са важни оптичните характеристики, а при трети механичните параметри.

На фигура 4 са дадени основни оптични характеристики за слоеве от аноден оксид.



Фиг. 4. Оптични характеристики на аноден Al_2O_3 : а) 1- пропускане от страна на порите, 2 - пропускане от страна на дъното; б) 1 – аморфен оксид, 2 – γ модификация, 3 - α модификация

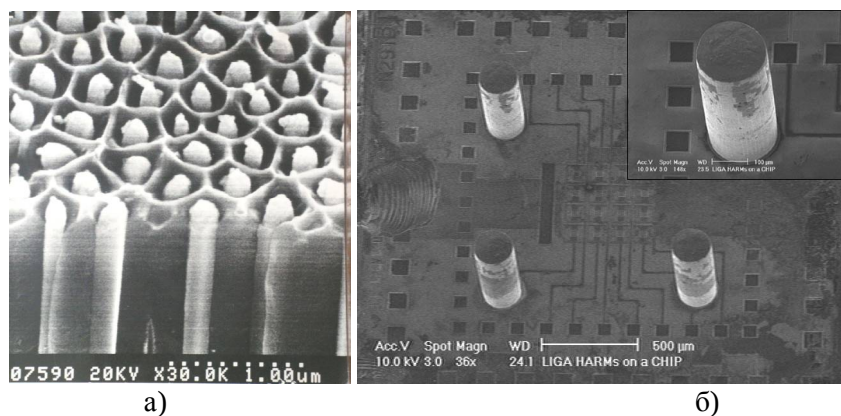
За микромеханични структури от значение са електромеханичните параметри на оксида. Част от тях са дадени в таблица 2.

Таблица 2

Материал	Еластичност σ_e [MPa]	Якост σ_h [MPa]	Модул на еластич. E [GPa]	Микро твърдост H_v [MPa]	Обемно съпрот. ρ_v Ω/m	Повърх. съпрот. ρ_s $\Omega/\square$
аморфен	400	400	140	4700	$5 \cdot 10^{10}$	$\approx 1 \cdot 10^{15}$
γ Al_2O_3	370	370	-	5600	$6 \cdot 10^6$	$\approx 1 \cdot 10^{12}$

Контролирането на тези параметри може да се извършва по различни методи, но един от основните въпроси е доколко ефективно можем да определим еднаквостта на параметрите за различните области от оксида.

Едно от приложенията на матрицата от аноден оксид на алуминия е използването и като маска в модифициран „LIGA” процес за израстване на наноелектроди от матричен тип. Матрични електроди се израстват електрохимично в диелектрична маска с отвори. Единият вид електроди могат да бъдат електроди за студена емисия на електрони от вида показан на фиг. 5а израстани в аноден оксид, или чрез стандартен „LIGA” процес – фиг. 5б [12].



Фиг. 5. Електрохимически израстнати наноигли в наноматрица от анодиран алуминий

Този процес може да бъде модифициран за създаване на контактни елементи представляващи четка от наноигли [13], или за елементи с анизотропна проводимост.

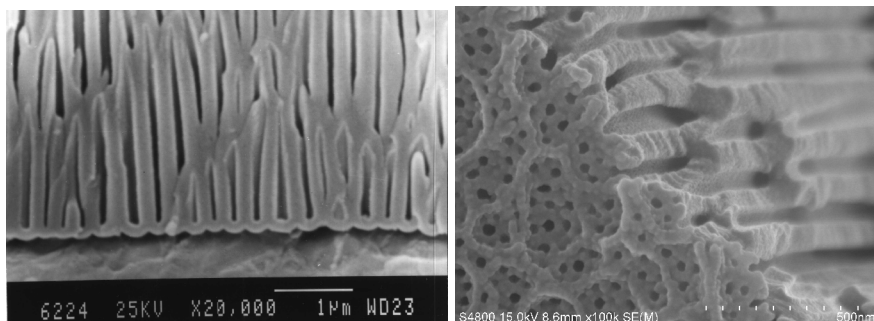
4. Контрол на технологичния процес

Получаването на микроелектронните елементи е свързано с групови процеси, които трябва да бъдат прецизно контролирани. Това е свързано с изискването за висока еднаквост на параметрите по цялата площ на пластината/подложката. Практически е невъзможно да се контролира процесът на изготвяне на отделния елемент, а се контролират параметрите на процеса. Такива могат да бъдат температура, налягане, скорост на поток, концентрация и други. Въпреки всички взети мерки съществуват технологични отклонения които трябва да бъдат определени. Те се определят на база после технологични измервания.

Контрол на технологичния процес по резултати от измерване на готови елементи след технологичен процес е основен за определяне разброса на параметрите по пластината. Този контрол може да се извършва чрез измерване на самите структури или чрез специални тестови структури. Обикновено контролът се извършва чрез сондови установки [14]. Когато имаме елементи с много малки размери можем да извършваме измерване на всеки елемент, изборно измерване на елементи представляващ извадка, или групово измерване с усредняване.

При провеждане на технологични процеси свързани с наноструктурни елементи е характерно голямо количество елементи с много малки размери. Непосредственото измерване на отделната нанообласт изисква прецизна апаратура с висока разделителна способност и голям брой измервания. В този случай имаме редица предизвикателства при измерване на параметрите след технологичния процес. Като пример можем да разгледаме анодния оксид на алуминия.

Ако искаме да определим разпределението на диаметъра на порите можем да използваме сканиращ електронен микроскоп и обработка на изображението. В случая обаче не можем да дадем отговор за състоянието на порите в дълбочина. Изследването на лом или специални шлифове – фиг. 6 е трудоемко.



Фиг. 6. Вертикален строеж на нанопори в алуминиев оксид

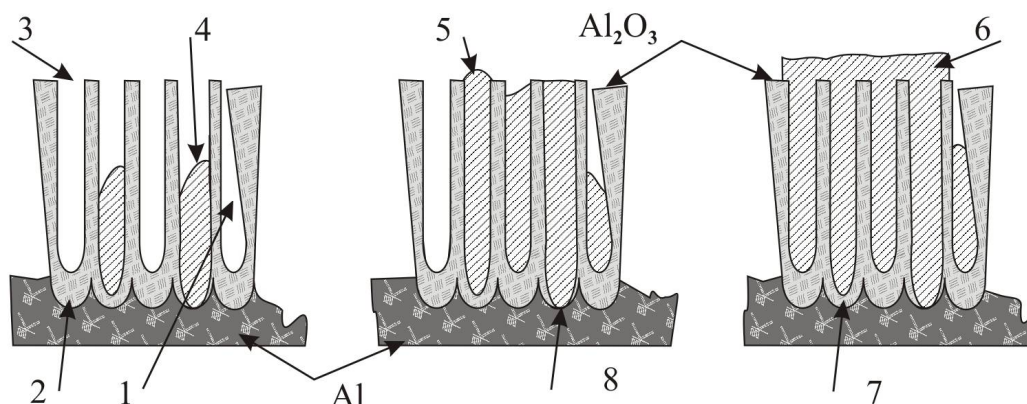
Можем да разгледаме следната теза – възможно ли е да се окачествява технологичния процес за наноструктуриране на анодния оксид на алуминия чрез използване на микроконтактни структури в зависимост от приложението му. Микроконтактните структури да се изготвят и измерват по стандартните микроелектронни технологии.

5. Експериментална част

Оценката на анодния оксид на алуминия се извършва чрез измерване на микроконтакти реализирани чрез фотолитографски процес и измерване на техните параметри.

Видове микроконтактни структури:

- 1- Капацитивни структури реализирани върху аноден оксид и алуминиев електрод представляващ основата. Втори електрод върху анодния оксид получен чрез механично контактуване или чрез метализация. Дава информация за диелектрична проникваемост, диелектрични загуби и пробивно напрежение.
- 2- Структури използващи запълване на порите с метал. В зависимост от запълването могат да бъдат с частично запълване, пълно запълване и препълване (фиг. 7)

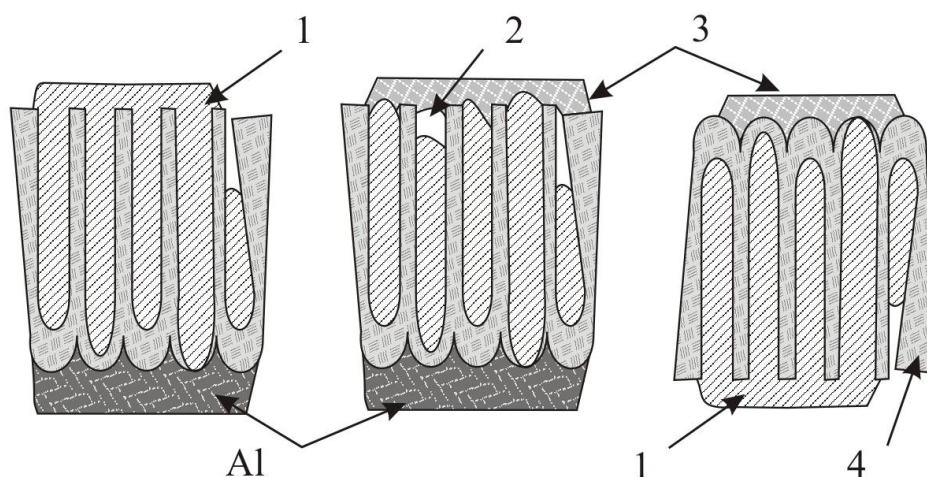


Фиг. 7. Запълване на порите в аноден оксид на алуминий. 1 – частично затворена пора, 2 – дъно на пора, 3 – отвор на пора, 4 – частично запълнена пора, 5 – запълнена пора, 6 – препълнени пори (съединена метализация), 7 изтънена пора/дъно, 8 – пробито дъно

При настоящите експерименти е оценявана възможността за използване на микроконтактни структури реализирани чрез запълване на порите. Запълването на порите може да даде информация за количеството метал отложен в пората, което е пропорционално на протичащия ток. Това от своя страна е показател за процеса в дъното на пората и самата пора. При равномерен процес върху цялата повърхност на подложката трябва да имаме равномерно запълване на порите.

Задачата която трябва да бъде решена е реализирането и измерването на микроконтактната структура. Единият от вариантите е създаване на микрочетков контакт предложен в [13] и измерване на неговото контактно съпротивление. Проведени са изследвания в тази насока, но там възникват редица неопределености свързани с усилието на контактуване, дълбочината на контакта, статистическото разпределение на контактната повърхност по вертикала и други.

Вторият вариант е създаване на микроконтакт от неподвижен тип, с използване на електрохимичен процес и/или вакуумно нанасяне на метал (конструкция - фиг. 8).

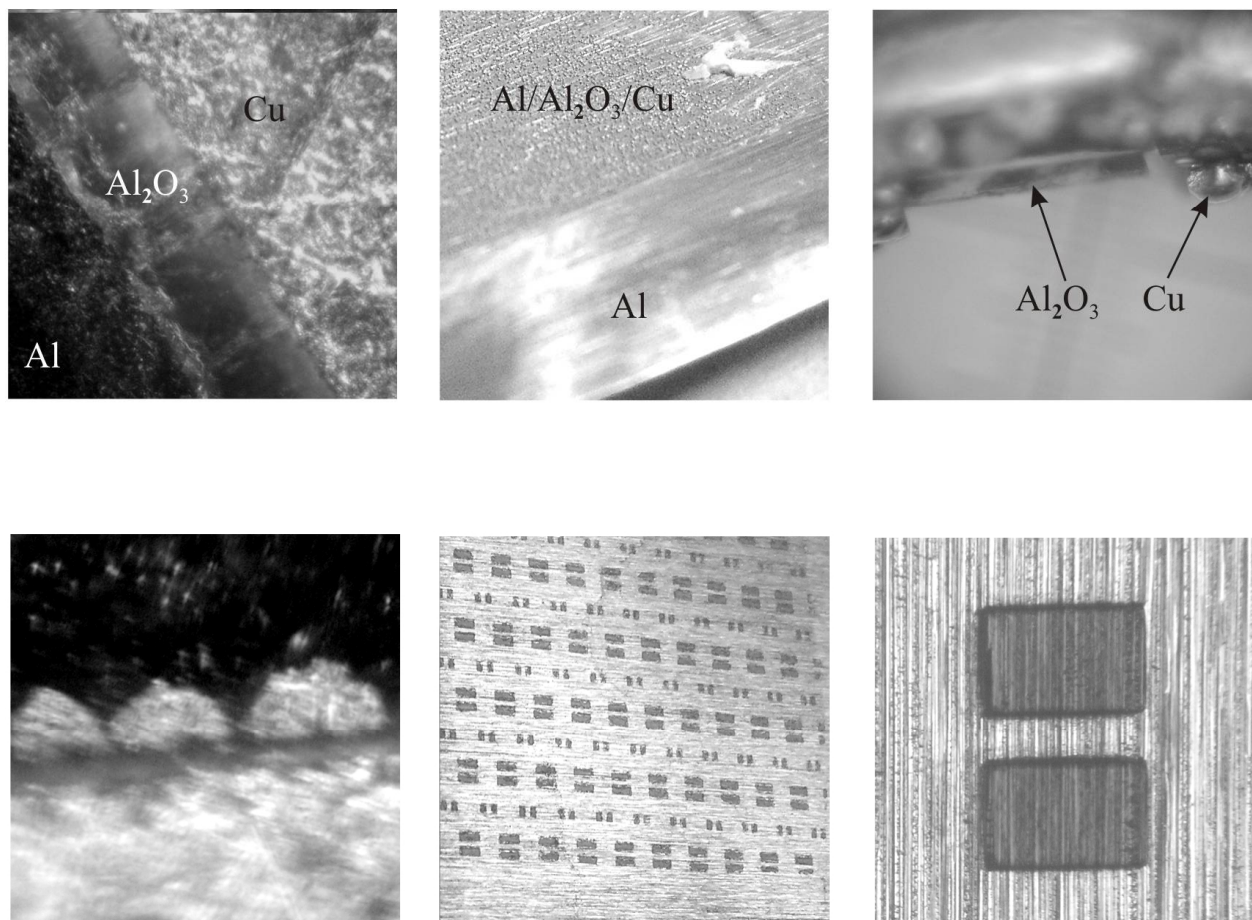


Фиг. 8. Микроконтакт към пори на Al_2O_3 . 1 – препълнени пори, 2 – незапълнена пора, 3 – вакуумно нанесен слой (контакт), 4 – оксид

При препълнени пори микроконтактът се формира в слоя от съединени с метал пори. Електрическата връзка се осъществява между алуминиевата основа и плътния метал 1. Съществува разновидност при която алуминиевата основа се разтваря (ецва) селективно и контактът се осъществява чрез вакуумна метализация 3 на тилната страна (откъм дъното на порите). При запълнени пори (или частично незапълнени 2) контактът се реализира чрез вакуумна метализация и съответно топологично формиране.

Изготвянето на микроконтактите може да се извърши по две технологични схеми. Чрез селективно анодиране на алуминия и последващо запълване на порите в селективните участъци или чрез тотално анодиране и последващо формиране на селективни участъци за запълване. Вакуумното метализиране се извършва с метал съответстващ на запълнените пори. При запълване с мед се изпарява мед, а при запълване с никел, се нанася никел. Целта е да се избегнат контактни явления при различие на метализациите.

На фиг. 9 са показани елементи от микроконтактни структури в анодния оксид и вертикална структура на презапълнени пори.

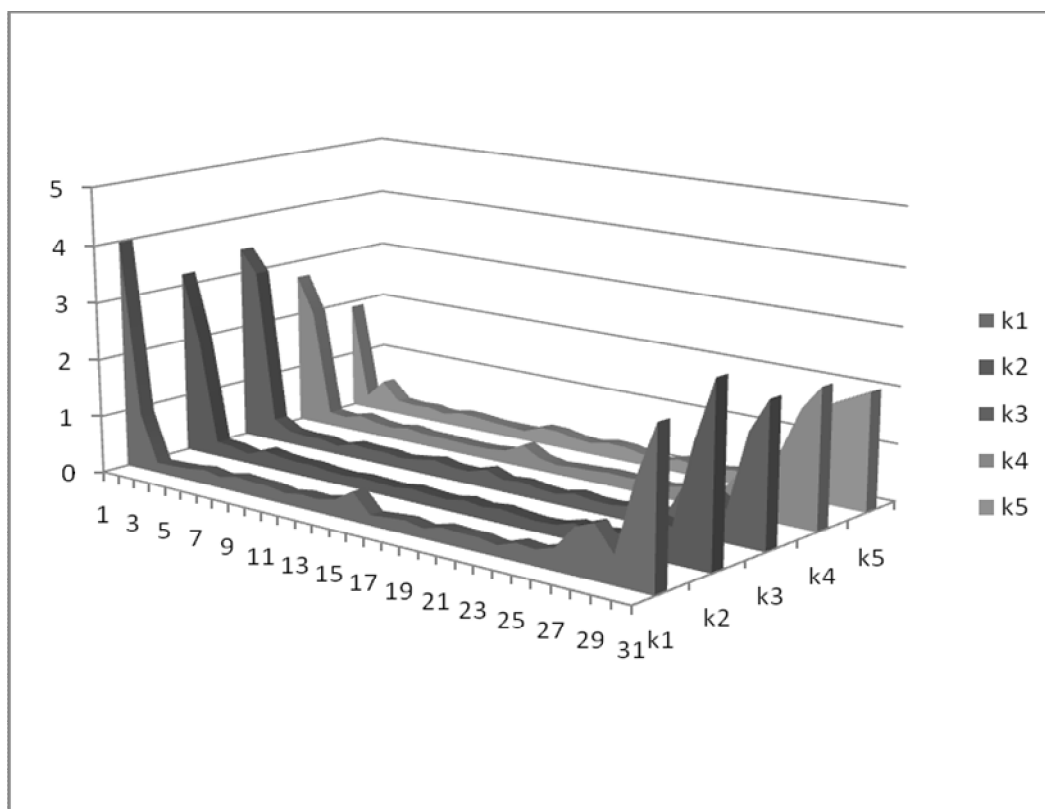


Фиг. 9. Етапи от изготвени микроконтактни структури с презапълнени пори.
Алуминиево фолио/оксид/електрохимична мед

Показаните образци са изготвени върху алуминий с чистота 99% . Дебелината на оксида е 12 μm . На предпоследната снимка е показана матрица от микроконтакти, а след нея двойка контакти с единични размери 120x180 μm .

Измерването на контактното съпротивление се извършва по два метода, три и четири сондов. При три сондовия метод единия електрод контактува с основата, а другите два с микроконтакта. При четири сондовия се използват двойките микроконтакти. Към всяка двойка се подават две сонди, като едната е токова, а другата напрежителна. Измерва се относителното изменение на съпротивлението на двойките микроконтакти.

Токът се задава от стабилизирани източник като се контролира чрез PHILIPS PM2525, а потенциала се измерва с KEITHLEY 195A. Измерителният ток е 1 и 10 mA. Усилието на контактуване се контролира чрез електронна везна с точност до 1 g. Средното усилие при измерването е 14 ± 2 g. На фиг.10 са дадени резултатите от измерване на област от тестови микроконтакти включваща 5 колони по 31 елемента. Полученото съпротивление е в оме.



Фиг. 10. Разпределение на съпротивлението на микроконтакти в аноден оксид на алуминий запълнени с мед

Периферните високи контактни съпротивления вероятно се дължат на дефекти в оксида и фотолитографската обработка.

6. Изводи

В доклада е показана възможността за прилагане на микроконтактни структури за окачествяване равномерността на технологичен процес. Като пример е представено запълването на порите в нанопорьозен аноден оксид на алуминия. Методът прилагащ микроконтакти е компромисно решение за окачествяване на процеси протичащи на нано ниво чрез измерване на микро области. Всяка област дава интегрална оценка, а отделните области са диференциална оценка за процеса. Предимство е и възможността да се използват по достъпни средства за измерване.

Методът икономисва време при оценяването и при подходящо подбиране на структурите може да дава оценка и за конструктивни параметри. Недостатък е възможността за изкривяване на резултатите при интегрирането на отделните нано структури.

ЛИТЕРАТУРА

- [1] Chen, S. B. at all, High-Density MIM Capacitors Using Al₂O₃ and AlTiO_x Dielectrics, IEEE Electron Device Letters, VOL. 23, NO. 4, APRIL 2002 pp. 185 – 187.
- [2] Seong-Ho Shin Young-Se Kwon, Selectively anodized aluminum substrates for microwave power module package, Electronic Components and Technology Conference, 2005. Proceedings. 55th, ISSN: 0569-5503, Vol. 2, pp1904-1908
- [3] <http://americas.kyocera.com/> 10.2010.
- [4] Brakspear, S. Anodised Aluminium as an Alternative Substrate for High Power Hybrid Microelectronics, Microelectronics International, ISSN 1356-5362, V7 / 2 1990, pp. 8 – 11.
- [5] Позняк, А.А. Модифицированный анодный оксид алюминия и композитные материалы на его основе, Минск, „Издат. Центр БТУ” 2007, ISBN 978-985-476-561-7.
- [6] Micro Bump Switch Contact Points, <http://e-fab.com/products/micro-contacts-micro-bumps/> отворен 11.2010.
- [7] Kosuke MIURA, at all, Development of High-density Grid-array Contact Technology, Sei technical review · number 60 · june 2005, <http://global-sei.com/tr/pdf/industrial/60-5.pdf>.
- [8] Tsuyoshi HAGA, at all, Development of Micro Contact Probe, <http://global-sei.com/tr/pdf/industrial/55-13.pdf>.
- [9] Видеков, В. Микроконтакти тип „четка”[®] с наноиглена структура Сп. Електротехника и електроника ISSN 0861-4717 бр. 9-10 2006 стр. 44 -48
- [10] Михуров, Н. И. Алюмооксидные микро-наноструктуры для микроэлектромеханических систем Минск „Беспринт” 2004 ISBN 985-6722-84-5
- [11] Видеков, В., Б. Цанева, Ал. Ячовски, Г. Фърков, Диелектрични наноструктурирани мембрани с метално ядро 9 МНТК по Авангардни материали и обработки АМО`2009 Сборник доклади ISSN 1313-4264 стр. 337-342
- [12] Daniel Sang-Won Park at all, Polydimethylsiloxane-based pattern transfer process for the post-IC integration of MEMS onto CMOS chips, Journal Of Micromechanics And Microengineering, © 2004 IOP Publishing Ltd V14 pp. 335–340
- [13] Videkov, V., V. Porkun, I. Vrublevsky, S. Tzanova. Brush Type Bonding with Nano-Needle Structure for Carrier Wafer Bonding in Microelectronics, Nanotech Northern Europe (NTNE 2007), 28-30 March, Helsinki, Finland, pp.70-74.
- [14] Farid Matta. Advances in IC testing: the membrane probe card – technical, Hewlett-Packard Journal, June, 1990, http://findarticles.com/p/articles/mi_m0HPJ/is_n3_v41/ai_9409219/

Автор: Александър Ячовски, инж. маг. докторант катедра „Микроелектроника”

Постъпила на 17.01.2011

Рецензент доц. д-р Никола Йорданов