

Запорізький національний технічний університет



Радіоелектроніка Інформатика Управління

2(16)'2006

Науковий журнал

Видається двічі на рік

Зареєстрований **29 січня 2003 року**
Державним комітетом інформаційної політики,
телебачення та радіомовлення України.

Свідоцтво – серія **КВ № 6904**.

Засновник і видавник – *Запорізький національний технічний університет*

Запоріжжя, ЗНТУ
2006

Передплатний індекс – 22914

Журнал «Радіоелектроніка. Інформатика. Управління» видається з 1999 року.

Журнал входить до переліку наукових фахових видань України, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора і кандидата технічних наук та фізико-математичних наук за спеціальністю «Радіофізика». Статті, що публікуються в журналі, реферуються в базах даних та РЖ ВІНТІ (Росія) і «Джерело» (Україна)

РЕДАКЦІЙНА КОЛЕГІЯ

Головний редактор – д.т.н. Потапенко Є. М.

Заст. головного редактора – к.т.н. Дубровін В. І.

Члени редколегії :

д.ф.-м.н. Ахметшин А. М.

д.т.н. Волков О. В.

д.ф.-м.н. Горбань О. М.

д.ф.-м.н. Горр Г. В.

д.т.н. Гостєв В. І.

д.ф.-м.н. Дробахин О. О.

д.ф.-м.н. Ковальов О. М.

д.ф.-м.н. Корніч Г. В.

д.ф.-м.н. Онуфрієнко В. М.

д.т.н. Павлов О. А.

д.т.н. Піза Д. М.

д.ф.-м.н. Погосов В. В.

д.т.н. Труфанов І. Д.

д.ф.-м.н. Чумаченко В. П.

Редакційно-видавнича рада:

Беліков С. Б., Внуков Ю. М., Піза Д. М., Яримбаш С. Т., Потапенко Є. М., Дубровін В. І., Петруша Ю. П., Флора В. Д., Соболь Ю. О., Дедков М. В., Рибіна Ю. А., Табунщик Г. В., М'ясников М. Л., Богданова Л. Ф., Кобзар Н. М., Федоренко Е. С.

Рекомендовано до видання вченою радою Запорізького національного технічного університету, протокол № 2 від 30.10.2006 р.

Рукописи проходять незалежне рецензування з залученням провідних фахівців, за результатами якого редакційна колегія приймає рішення про опублікування

Журнал зверстаний редакційно-видавничим відділом
Запорізького національного технічного університету

Адреса редакції: 69063, м. Запоріжжя, вул. Жуковського, 64, ЗНТУ,
редакція журналу «РІУ»

Тел: (061)7-698-2-96 – редакційно-видавничий відділ

Факс: (0612)64-21-41

E-mail: rvv@zntu.edu.ua

TQM@zntu.edu.ua

Web-сайт журналу розташований в Internet за адресою: <http://www.zntu.edu.ua/base/news2005/ric.htm>

© 2006 Запорізький національний технічний університет

ЗМІСТ

РАДІОФІЗИКА	5	
<i>А. В. Крапивной, В. П. Чумаченко</i> ОБ ОЦЕНИВАНИИ ВОЛНОВОГО ПОЛЯ, РАССЕЯННОГО МЯГКОЙ СФЕРОЙ		5
РАДІОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ	9	
<i>А. А. Баркалов, Р. В. Мальцева, Баракат Халед</i> ПРЕОБРАЗОВАНИЕ АДРЕСОВ В КОМПОЗИЦИОННОМ МИКРОПРОГРАММНОМ УСТРОЙСТВЕ УПРАВЛЕНИЯ С РАЗДЕЛЕНИЕМ КОДОВ		9
<i>В. Г. Лихограй, А. Е. Стрельницкий, А. А. Стрельницкий, А. И. Цопа, В. М. Шокало</i> ЭФФЕКТИВНОСТЬ РАБОТЫ СИСТЕМ АБОНЕНТСКОГО РАДИОДОСТУПА САРД ПРИ НАЛИЧИИ ПОМЕХИ. МОДЕЛЬ И АНАЛИЗ		31
<i>В. В. Брайлловський, П. П. Ватаманюк, М. Г. Рождественська</i> ІМІТАТОР СИГНАЛІВ ЯДЕРНОЇ ІНДУКЦІЇ ДЛЯ ІМПУЛЬСНОГО ЯКР-СПЕКТРОМЕТРА		14
<i>В. І. Грядун</i> ВЛИЯНИЕ АКТИВНЫХ АТОМОВ НА ЭНЕРГИЮ ОБРАЗОВАНИЯ ТОЧЕЧНЫХ ДЕФЕКТОВ В КРЕМНИИ		16
<i>В. Н. Крищук, Л. М. Карпуков, М. В. Заскоцкая, А. Ю. Фарафонов</i> ИССЛЕДОВАНИЕ ДОПУСКОВЫХ ОГРАНИЧЕНИЙ В КОНСТРУКЦИЯХ МИКРОПОЛОСКОВЫХ ФИЛЬТРОВ С УЧЕТОМ ВИДА АППРОКСИМАЦИИ АЧХ		20
<i>В. М. Крищук, Г. М. Шило, Б. А. Артюшенко</i> ГЕНЕТИЧНИЙ АЛГОРИТМ ПРИЗНАЧЕННЯ ДОПУСКІВ НА РАДІОЕЛЕМЕНТИ З ІНТЕРВАЛЬНИМ ОЦІНЮВАННЯМ		26
МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ	50	
<i>С. Н. Гребенюк, С. И. Гоменюк, Е. Л. Мизерная, В. В. Киричевский</i> МОДЕЛИРОВАНИЕ НАПРЯЖЕННО- ДЕФОРМИРОВАННОГО СОСТОЯНИЯ КОМПОЗИТОВ И КОНСТРУКЦИЙ НА ИХ ОСНОВЕ		50
<i>М. А. Новотарський</i> МЕРЕЖІ ДЛЯ МОДЕЛЮВАННЯ СКЛАДНИХ СИСТЕМ		60
<i>Р. В. Киричевский, Е. С. Решевская, В. М. Тархова, Е. В. Прокopenко, В. В. Киричевский</i> ИССЛЕДОВАНИЕ ЗАДАЧ МЕХАНИКИ ДЕФОРМИРУЕМОГО ТВЕРДОГО ТЕЛА В ВЫЧИСЛИТЕЛЬНОМ КОМПЛЕКСЕ «MIRELA+»		54
<i>С. Н. Сердюк, К. Х. Ибрагим, Ж. К. Каминская</i> УСОВЕРШЕНСТВОВАНИЕ МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЭКСПЕРТНЫХ ПРОБЛЕМНЫХ ЗНАНИЙ		66
<i>А. Я. Скляр, И. А. Макрушан</i> МОДЕЛИРОВАНИЕ ТРАНСПОРТНЫХ КОММУНИКАЦИЙ КОРПОРАТИВНЫХ КОМПЬЮТЕРНЫХ СЕТЕЙ: РАСПРЕДЕЛЕНИЕ ТРАНЗАКЦИЙ ПО МАРШРУТАМ		70

<i>Е. Е. Сыревич</i> ВЕРИФИКАЦИЯ ВЫСОКОУРОВНЕВЫХ МОДЕЛЕЙ ЦИФРОВЫХ УСТРОЙСТВ74	<i>В. И. Хаханов, А. Н. Парфентий, Хассан Ктейман, Ваде Гриби</i> HEDEFS – ПРОГРАММНО-АППАРАТНАЯ СИСТЕМА МОДЕЛИРОВАНИЯ НЕИСПРАВНОСТЕЙ 77
---	---

НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ 85

<i>Е. В. Бодянский, Е. А. Винокурова, Н. С. Ламонова</i> ДВОЙНОЙ ВЭЙВЛЕТ-НЕЙРОН И АЛГОРИТМ ЕГО ОБУЧЕНИЯ85	<i>А. И. Липчанский, Механна Сами, Хабис А. А. Зидат</i> ИСПОЛЬЗОВАНИЕ ПРОДУКЦИОННЫХ СИСТЕМ И БИНАРНЫХ ОТНОШЕНИЙ В ЭКСПЕРТНЫХ СИСТЕМАХ ДИАГНОСТИКИ КОМПЬЮТЕРНОЙ ТЕХНИКИ 93
<i>А. О. Дранкова, Е. В. Ткаченко</i> ИССЛЕДОВАНИЕ СЕТИ ХОПФИЛДА ДЛЯ РАСПОЗНАВАНИЯ ЦИФРО-ПОДОБНЫХ ОБРАЗОВ91	

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ 102

<i>Н. В. Алипов, Н. М. Кораблев, М. И. Хиль, М. В. Гусятин</i> СТРУКТУРА ЦИФРОВОГО АВТОМАТА С ПСЕВДОСЛУЧАЙНЫМИ ПЕРЕХОДАМИ ИЗ НАЧАЛЬНОГО СОСТОЯНИЯ В ОДНО И ТО ЖЕ КОНЕЧНОЕ СОСТОЯНИЕ 102	<i>В. И. Дубровин, Е. Н. Федорченко</i> ДИАГНОСТИКА НА ОСНОВЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ 115
<i>Г. Г. Грездов</i> НОВЫЙ СПОСОБ ОЦЕНКИ ВЕЛИЧИНЫ ОСТАТОЧНОГО РИСКА ПРИ ФОРМИРОВАНИИ ЭКОНОМИЧЕСКИ ЭФФЕКТИВНОЙ КОМПЛЕКСНОЙ СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ В АВТОМАТИЗИРОВАННЫХ СИСТЕМАХ 109	<i>А. С. Кулик, А. Ю. Завгородний</i> ДИАГНОСТИРОВАНИЕ ДАННЫХ КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ НА ОСНОВЕ ЕСТЕСТВЕННОЙ ИЗБЫТОЧНОСТИ 120
	<i>А. И. Рязанцев, В. С. Кардашук</i> СИСТЕМА ЭКОЛОГИЧЕСКОГО МОНИТОРИНГА ОКРУЖАЮЩЕЙ СРЕДЫ 128

ТЕОРІЯ І МЕТОДИ АВТОМАТИЧНОГО УПРАВЛІННЯ 133

А. К. Новиков СИНТЕЗ ОПТИМАЛЬНЫХ УПРАВЛЕНИЙ НА ОСНОВЕ АНАЛИТИЧЕСКОГО РЕШЕНИЯ ЛИНЕЙНОЙ СТАЦИОНАРНОЙ ДВУХТОЧЕЧНОЙ КРАЕВОЙ ЗАДАЧИ Часть 1. Постановка и решение задачи. Базовый алгоритм 133	А. К. Новиков СИНТЕЗ ОПТИМАЛЬНЫХ УПРАВЛЕНИЙ НА ОСНОВЕ АНАЛИТИЧЕСКОГО РЕШЕНИЯ ЛИНЕЙНОЙ СТАЦИОНАРНОЙ ДВУХТОЧЕЧНОЙ КРАЕВОЙ ЗАДАЧИ Часть 2. Примеры применения базового алгоритма 143
--	--

КЕРУВАННЯ У ТЕХНІЧНИХ СИСТЕМАХ 157

<i>Н. В. Ефименко, Н. В. Луценко</i> ПОСТРОЕНИЕ РЕГУЛЯРИЗОВАННЫХ ЭЛЛИПСОИДАЛЬНЫХ ОЦЕНОК В ЗАДАЧЕ ОПРЕДЕЛЕНИЯ ВЕКТОРА СОСТОЯНИЯ ЛИНЕЙНОЙ ДИСКРЕТНОЙ СИСТЕМЫ С МНОГОМЕРНЫМ ВЫХОДОМ 157

АВТОРЫ НОМЕРА..... 162

РАДІОФІЗИКА

РАДИОФИЗИКА

RADIOPHYSICS

УДК 537.874.6

А. В. Крапивной, В. П. Чумаченко

ОБ ОЦЕНИВАНИИ ВОЛНОВОГО ПОЛЯ, РАССЕЯННОГО МЯГКОЙ СФЕРОЙ

На примере коротковолнового рассеяния плоской волны мягкой сферой установлена возможность использования для решения трехмерных задач нового подхода, апробированного ранее на двумерных объектах в случае их возбуждения электромагнитной волной. Теория основывается на свойстве локальности высокочастотного рассеяния, принимает во внимание кривизну граничной поверхности и не предполагает обращения в нуль полей на затененной ее части. Метод более точен, чем приближение физической оптики, однако сохраняет сравнимую с ним простоту.

ВВЕДЕНИЕ

В процессе решения задач рассеяния коротких электромагнитных или акустических волн простым и удобным инструментом является метод физической оптики (приближение Киргофа) [1]. Для оценки значений поля на освещенной части рассеивателя он использует замену его граничной поверхности бесконечной плоскостью касательной к поверхности в рассматриваемой точке. На затененной части поверхности поле предполагается равным нулю. Поле вне рассеивателя определяется затем с помощью определенного вида поверхностного интеграла. Такой подход является весьма эффективным, так как не требует матричного решения. Однако применим он только для больших гладких объектов. К тому же оценка рассеянного поля ухудшается с увеличением отклонения рассматриваемого направления от зеркального. Более точными (и значительно более сложными) высокочастотными методами, которые могут быть использованы для нахождения рассеянных полей, являются геометрическая [2] и физическая [3] теории дифракции.

В недавних работах [4, 5] был предложен новый подход к оценке поля рассеянного выпуклыми двумерными объектами, который, сохраняя простоту метода физической оптики, позволяет заметно улучшить оценки для незеркальных направлений и дает приемлемую точность и в среднечастотном диапазоне. Подобно методу физической оптики он основывается на свойстве локальности коротковолнового рассеяния, однако принимает во внимание кривизну поверхности и не предполагает обращения в нуль поля на теневой части рассеивателя. В настоящей работе исследуется возможность использования развитого подхода при решении трехмерных задач. Рассмотрено рассеяние плоской волны простейшим выпуклым объектом – мягкой сферой. В последующем развиваемый подход предполагается распространить на выпуклые препятствия общего вида, включая случай электромагнитного рассеяния.

ПОСТАНОВКА И РЕШЕНИЕ ЗАДАЧИ

Рассмотрим стационарное звуковое поле, которое устанавливается в среде, характеризуемой плотностью ρ и скоростью звука c , при наличии в ней мягкой сферы, показанной на рис. 1. Звуковое поле будем характеризовать давлением p и круговой частотой колебаний ω . Пусть $p(x, y, z, t) = p(x, y, z)e^{i\omega t}$ и ось z направлена навстречу падающей плоской волне

$$p_i = e^{ikz}. \quad (1)$$

Тогда давление

$$p = p_i + p_s \quad (2)$$

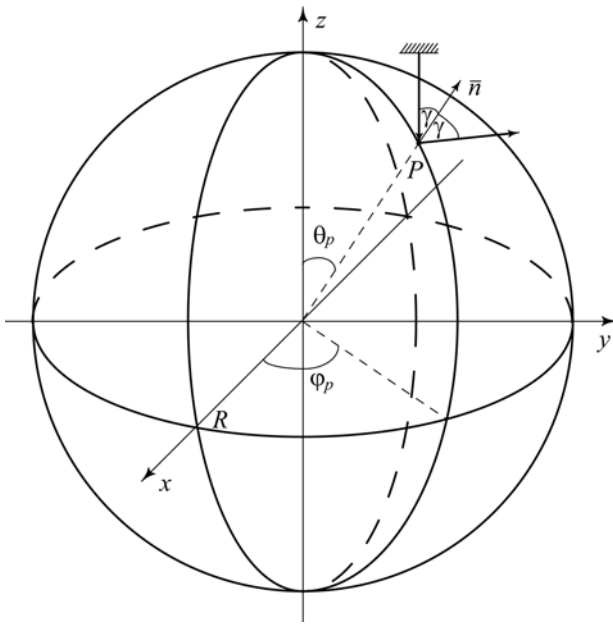


Рисунок 1 – Геометрия задачи

удовлетворяет уравнению Гельмгольца

$$\Delta p + k^2 p = 0, \quad k = \frac{\omega}{c} = \frac{2\pi}{\lambda} \quad (3)$$

и граничному условию

$$p(P) = 0, \quad P \in S, \quad (4)$$

где S – поверхность рассеивателя.

Рассеянное поле p_s на бесконечности удовлетворяет условию излучения. Вне S оно может быть представлено интегралом

$$\begin{aligned} p_s(M) &= -\frac{1}{4\pi} \iint_S \frac{\partial p}{\partial n} \frac{e^{-ikr_{PM}}}{r_{PM}} dS = \\ &= \frac{ikZ_0}{4\pi} \iint_S v_n \frac{e^{-ikr_{PM}}}{r_{PM}} dS, \end{aligned} \quad (5)$$

где

$$v_n = -\frac{1}{ikZ_0} \left. \frac{\partial p}{\partial n} \right|_P = \frac{i}{kZ_0} \left[\left. \frac{\partial p_i}{\partial n} + \frac{\partial p_s}{\partial n} \right] \right|_P \quad (6)$$

– скорость колебания частиц среды в направлении нормали к поверхности (направление \vec{n} показано на рис. 1), а $Z_0 = c\rho$ – волновое сопротивление среды.

Пусть (r, θ, ϕ) – сферическая система координат с началом в центре шара и полярной осью вдоль Oz . Испробуем два способа определения приближенных значений p_s у поверхности шара.

Приближим сначала рассеивающую поверхность в окрестности точки P поверхностью кругового цилиндра, который перпендикулярен плоскости падения волны и ось которого проходит через центр сферы. Кривизна поверхности в направлении перпендикулярном плоскости падения при этом очевидно не учитывается. Повторив выкладки, сделанные в [4] в случае цилиндрического рассеивателя, мы убедимся, что величина $\left. \frac{\partial p}{\partial n} \right|_P$ в (6) может быть приближенно представлена выражением

$$\left. \frac{\partial p}{\partial n} \right|_P = p_i(P) [ik \cos \gamma - X'(R)], \quad (7)$$

где γ – угол падения в точке P , а $X'(R)$ имеет вид

$$X'(R) = k \left[\frac{\mu}{kR} - \frac{H_{\mu+1}^{(2)}(kR)}{H_{\mu}^{(2)}(kR)} \right]. \quad (8)$$

Здесь $H_{\mu}^{(2)}(kR)$ – функция Ганкеля и

$$\mu = kR \sin \gamma. \quad (9)$$

Попытаемся теперь учесть кривизну поверхности и в направлении перпендикулярном плоскости падения. С этой целью в окрестности точки P введем локальные сферические координаты (r, ϑ, ϕ) с началом отсчета в центре сферы и полярной осью вдоль оси упомянутого выше цилиндра. При этом точке P отвечает значение ϑ , равное $\frac{\pi}{2}$ и $\phi = \theta - \theta_P$ при $\vartheta = \frac{\pi}{2}$. Разделяя переменные в уравнении Гельмгольца [6], мы получим

$$p_s = X(r)T(\vartheta)\Phi(\phi), \quad (10)$$

$$\frac{d}{dr} \left(r^2 \frac{dX}{dr} \right) + (k^2 r^2 - \chi) X = 0, \quad (11)$$

$$\frac{1}{\sin \vartheta} \frac{d}{d\vartheta} \left(\sin \vartheta \frac{dT}{d\vartheta} \right) + \left(\chi - \frac{m^2}{\sin^2 \vartheta} \right) T = 0, \quad (12)$$

$$\frac{d^2 \Phi}{d\phi^2} + m^2 \Phi = 0. \quad (13)$$

Будем считать далее, что в малой окрестности точки P , как и в случае плоской поверхности, функциональные зависимости падающей и отраженной волн совпадают вдоль поверхности с точностью до постоянного множителя. Это дает $T = C = \text{const}$ и $m^2 = \chi \sin^2 \vartheta$. Отсюда $m = \sqrt{\chi}$ и

$$\Phi = e^{-i\sqrt{\chi}\phi} \quad (14)$$

при $\vartheta = \frac{\pi}{2}$. Чтобы сравнить зависимости от ϕ , используем в выражении $p_i = e^{ikR \cos \theta} = e^{ikR \cos(\phi + \theta_p)}$ разложение косинуса в ряд Маклорена по степеням ϕ :

$$p_i = e^{ikR(\cos \theta_p - \sin \theta_p \phi + \dots)} \approx e^{ikR(\cos \theta_p - \sin \theta_p \phi)}. \quad (15)$$

Сравнивая (14) и (15), мы получим, что эти зависимости совпадут до постоянного множителя, если

$$\chi = k^2 R^2 \sin^2 \gamma. \quad (16)$$

Удовлетворяя условиям на бесконечности, выберем решение уравнения (11) в виде $X(r) = h_{\mu}^{(2)}(kr)/h_{\mu}^{(2)}(kR)$, где $h_{\mu}^{(2)}(kr)$ – сферическая функция Бесселя, а

$$\mu = \sqrt{\chi + 0,25} - 0,5. \quad (17)$$

Подставив найденное значение p_s в (4), мы определим, что $C = -p_i(P)$. Таким образом отраженное поле в окрестности точки P может быть приближенно представлено как

$$p_s = -p_i(P) \frac{h_{\mu}^{(2)}(kr)}{h_{\mu}^{(2)}(kR)} e^{-i\sqrt{\chi}\phi}. \quad (18)$$

Нахождение $\left. \frac{\partial p}{\partial n} \right|_P$ с учетом выражения (18) снова приводит нас к формуле (7), где $X'(R)$ имеет вид

$$X'(R) = k \left[\frac{\mu}{kR} - \frac{h_{\mu+1}^{(2)}(kR)}{h_{\mu}^{(2)}(kR)} \right]. \quad (19)$$

Знание нормальной производной давления на поверхности сферы позволяет с помощью формул (5) и (6) определить полностью все характеристики рассеянного поля.

ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

На рис. 2 значения v_n , рассчитанные описанными выше способами, сравниваются с соответствующими значениями, полученными методом физической оптики, и точными значениями. Под точным решением мы подразумеваем решение, полученное в виде ряда методом разделения переменных [7]. Как видно, приближенные результаты, полученные с помощью развитого подхода, являются значительно более точными, чем приближение физической оптики. При этом расчетные данные для $\arg(v_n)$ в теневой области являются правильными вплоть до углов, где $|v_n|$ становится весьма малым.

Определим бистатистический поперечник рассеяния как

$$\sigma = \lim_{r \rightarrow \infty} 4\pi r^2 \frac{Q_s}{Q_i}, \quad (20)$$

где $Q_i = \frac{1}{2} \text{Re}(p_i^* v_i)$ – интенсивность падающей волны, $Q_s = \frac{1}{2} \text{Re}(p_s^* v_s)$ – интенсивность рассеянной волны, а v_i

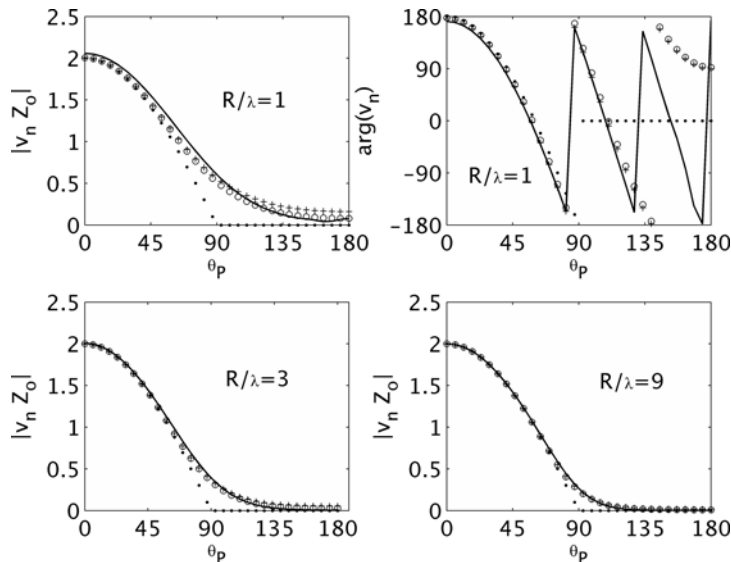


Рисунок 2 – Скорость колебания частиц среды в радиальном направлении на поверхности сферы: сплошная линия – точно, пунктир – метод физической оптики, кружки – формула (8), крестики – формула (19)

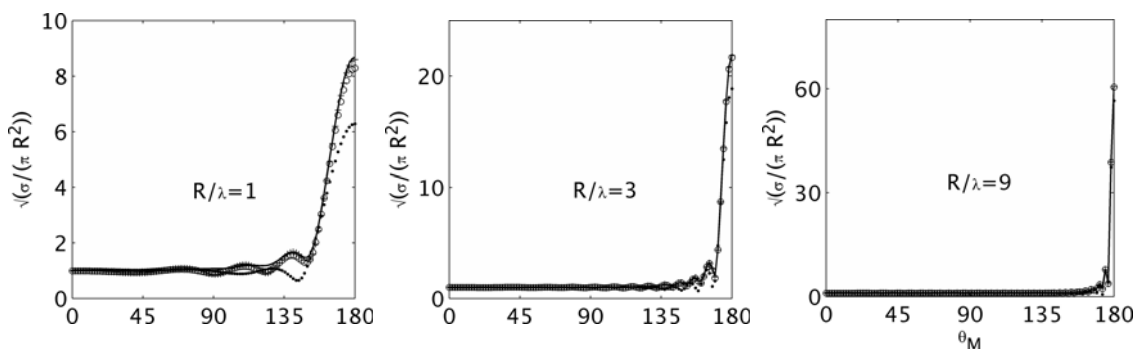


Рисунок 3 – Угловая зависимость рассеянного поля в дальней зоне:

сплошная линия – точно, пунктир – метод физической оптики, кружки – формула (8), крестики – формула (19)

и v_s – соответствующие скорости. Нахождение предела (20) и последующее интегрирование в (5) по φ приводят к следующему выражению для поперечника

$$\sigma(\theta_M) = \pi R^4 \left| \int_0^\pi \frac{\partial p}{\partial n} J_0(kR \sin \theta_M \sin \theta) e^{ikR \cos \theta_M \cos \theta} \sin \theta d\theta \right|^2 \quad (21)$$

На рис. 3 представлены результаты расчета значений поля в дальней зоне. Видно, что для направлений, отличных от зеркального, данные, полученные с помощью нашего метода, являются значительно более точными, чем в случае, когда $\frac{\partial p}{\partial n}$ в формуле (21) находится методом физической оптики. Из рисунков 2 и 3 также следует, что два рассмотренных варианта, ведущие, соответственно, к формулам (8) и (19), являются для рассматриваемого рассеивателя приблизительно равноценными. Использование формулы (19) вместо (8) лишь незначительно улучшает точность расчета дальнего поля в направлении распространения падающей плоской волны.

ЗАКЛЮЧЕНИЕ

На примере мягкой сферы установлена возможность использования в случае трехмерных выпуклых препятствий высокочастотного метода, развитого ранее для оценки полей, рассеянных проводящими или импедансными цилиндрами при возбуждении их плоской электромагнитной волной. Установлено, что, как и в случае двумерных объектов, предлагаемый подход является более точным, чем метод физической оптики, и обладает приблизительно той же эффективностью. Он позволяет улучшить оценки характеристик поля на поверхности препятствия и заметно повысить точность

расчета поля, рассеянного в незеркальных направлениях.

ПЕРЕЧЕНЬ ССЫЛОК

1. Knott E. F., Shaeffer J. F., Tuley M. T. Radar cross section. – Norwood, MA: Artech House, 1993. – 611 p.
2. Keller J. B. Diffraction by a convex cylinder // IEEE Trans. Antennas Propagat. – Vol. AP-4. – July 1956. – P. 312–321.
3. Уфимцев П. Я. Метод краевых волн в физической теории дифракции. – М.: Советское радио, 1962. – 244 с.
4. Chumachenko V. P. On the estimation of scattering from convex conducting cylinders // Microwave and Optical Technology Letters. – Vol. 45. – May 2005. – P. 191–194.
5. Krapivny A. V., Chumachenko V. P. On the estimation of scattering from convex impedance cylinders // Proceedings of 2006 Int. Conf. on Mathematical Methods in Electromagnetic Theory, Kharkiv, Ukraine. – June 26–29, 2006. – P. 248–250.
6. Кошляков Н. С., Глинер Э. Б., Смирнов М. М. Уравнения в частных производных математической физики. – М.: Высшая школа, 1970. – 712 с.
7. Лепендин Л. Ф. Акустика. – М.: Высшая школа, 1978. – 448 с.

Надійшла 5.06.06

На прикладі короткохвильового розсіювання плоскої хвилі м'якою сферою встановлена можливість використання при розв'язуванні тривимірних задач нового підходу апробованого раніше на двовимірних об'єктах у випадку їх збудження електромагнітною хвилею. Теорія ґрунтується на властивості локальності високочастотного розсіювання, приймає до уваги кривину граничної поверхні і не передбачає перетворення в нуль полів на тінювій її частині. Метод є більш точним, чим наближення фізичної оптики, але зберігає порівнянню з ним простоту.

The problem of plane-wave scattering off a soft sphere is studied by using a new method, which was recently introduced to approximately solve two-dimensional problems for conducting and impedance cylinders. The investigation has shown applicability of the technique to three-dimensional scatterers. The theory is based on the locality property of short-wave scattering and considers the curvature of the target and the shadow-side fields. The approach is more accurate than the physical optics approximation and keeps the simplicity comparable with that.

РАДІОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ

РАДИОЕЛЕКТРОНИКА И ТЕЛЕКОММУНИКАЦИИ

RADIO ELECTRONICS AND TELECOMMUNICATIONS

УДК 681. 324

А. А. Баркалов, Р. В. Мальчева, Баракат Халед

ПРЕОБРАЗОВАНИЕ АДРЕСОВ В КОМПОЗИЦИОННОМ МИКРОПРОГРАММНОМ УСТРОЙСТВЕ УПРАВЛЕНИЯ С РАЗДЕЛЕНИЕМ КОДОВ

Предложен метод оптимизации композиционного микропрограммного устройства управления с разделением кодов. Метод основан на введении преобразователей адреса микрокоманды и кодов операторных линейных цепей в схему устройства управления. Метод позволяет уменьшить емкость управляющей памяти. Рассмотрен пример применения предложенного метода.

ВВЕДЕНИЕ

Одним из методов реализации устройства управления (УУ) любой цифровой системы является использование модели композиционного микропрограммного устройства (КМУУ) [1, 2]. Подобные устройства идеально подходят при реализации цифровых систем в виде «систем на кристалле» [3, 4], где произвольная логика реализуется на FPGA или CPLD, а регулярная логика – на встроенных блоках памяти. Стремление к увеличению количества реализуемых системой функций приводит к необходимости уменьшения аппаратных затрат в отдельных блоках системы, включая и УУ. Одним из методов решения этой задачи является уменьшение числа термов в системе функций возбуждения [5, 6], что возможно, например, при исполь-

зовании в КМУУ метода разделения кодов [2]. Этот метод позволяет использовать наличие псевдоэквивалентных операторных линейных цепей (ОЛЦ) в исходной граф-схеме алгоритма (ГСА). Однако разделение кодов целесообразно только при сохранении минимальной разрядности адреса микрокоманды. В противном случае требуемая емкость управляющей памяти КМУУ резко увеличивается по сравнению с известными методами синтеза [2, 4]. В настоящей работе предлагается метод синтеза КМУУ с разделением кодов, позволяющий сохранить минимальную емкость управляющей памяти при произвольных характеристиках исходной ГСА.

1 ПОСТАНОВКА ЗАДАЧИ

Пусть исходная ГСА Γ имеет начальную вершину b_0 , конечную вершину b_E , операторные вершины из множества B_1 и условные вершины из множества B_2 . В вершинах $b_q \in B_1$ записываются микрокоманды $Y(b_q) \subseteq Y$, где $Y = \{y_1, \dots, y_N\}$ – множество микроопераций. В вершинах $b_q \in B_2$ записаны элементы множества логичес-

ких условий $X = \{x_1, \dots, x_L\}$. Вершины ГСА Γ связаны дугами $\langle b_q, b_t \rangle$, образующими множество дуг E . Введем ряд определений [1, 2], необходимых для дальнейшего изложения материала.

Определение 1. Операторной линейной цепью ГСА Γ называется конечная последовательность операторных вершин $\alpha_g = \langle b_{g1}, \dots, b_{gF_g} \rangle$, такая, что для любой пары соседних компонент вектора α_g существует дуга $\langle b_{gi}, b_{gi+1} \rangle \in E$ ($i = 1, \dots, F_g - 1$).

Определение 2. Входом ОЛЦ α_g называется вершина $b_q \in D^g$, где $D^g \subseteq B_1$ – множество операторных вершин, входящих в ОЛЦ α_g , если существует дуга $\langle b_t, b_q \rangle \in E$, такая, что $b_t = b_0$ или $b_t \in B_2$ или $b_t \notin D^g$.

Определение 3. Выходом ОЛЦ α_g называется вершина $b_q \in D^g$, такая, что существует дуга $\langle b_q, b_t \rangle \in E$, где $b_t = b_E$ или $b_t \in B_2$ или $b_t \notin D^g$.

Определение 4. Операторные линейные цепи α_g и α_j называются псевдоэквивалентными ОЛЦ, если их выходы связаны с входом одной и той же вершины ГСА.

Произвольная ОЛЦ α_g может иметь более одного входа. Обозначим j -й вход ОЛЦ α_g как I_g^j . Каждая ОЛЦ имеет только один выход O_g , входящий в множество $O(\Gamma)$ выходов ОЛЦ ГСА Γ .

Пусть для ГСА Γ сформировано множество ОЛЦ $C = \{\alpha_1, \dots, \alpha_G\}$, удовлетворяющее условию

$$D^j \cap D^i = \emptyset; \quad (j \neq i; j, i \in \{1, \dots, G\});$$

$$B_1 = D^1 \cup D^2 \cup \dots \cup D^G; \quad G \rightarrow \min. \quad (1)$$

При выполнении (1) каждая вершина $b_q \in B_1$ входит точно в одну ОЛЦ $\alpha_g \in C$, число которых – минимально. Пусть F_{\max} – максимальное число компонент в ОЛЦ $\alpha_g \in C$, то есть $F_{\max} = \max\{F_1, \dots, F_G\}$. Закодируем ОЛЦ $\alpha_g \in C$ R_1 -разрядными кодами $K(\alpha_g)$, где $R_1 = \lceil \log_2 G \rceil$. Закодируем компоненты ОЛЦ $\alpha_g \in C$ R_2 -разрядными кодами $K(b_t)$, где $R_2 = \lceil \log_2 F_{\max} \rceil$, причем кодирование выполним таким образом, чтобы для соседних компонент $b_{g_i}, b_{g_{i+1}}$ ($i = 1, \dots, F_g - 1$) выполнялось условие:

$$K(b_{g_{i+1}}) = K(b_{g_i}) + 1 \quad (g = 1, \dots, G). \quad (2)$$

В этом случае адрес $A(b_q)$ микрокоманды $Y(b_q)$ такой, что $b_q \in D^g$, представляется в виде конкатенации

$$A(b_q) = K(\alpha_g) * K(b_q), \quad (3)$$

где * – знак конкатенации. Такое представление адреса называется принципом разделения кодов [2] и ему соответствует КМУУ U_1 (рис. 1).

Здесь комбинационная схема СС реализует систему функций возбуждения счетчика СТ

$$\Phi = \Phi(\tau, X) \quad (4)$$

и систему функций возбуждения регистра RG

$$\psi = \psi(\tau, X), \quad (5)$$

где τ – внутренние переменные, кодирующие ОЛЦ $\alpha_g \in C$ и $|\tau| = R_1$. Компоненты ОЛЦ $\alpha_g \in C$ кодируются внутренними переменными T и $|T| = R_2$. Управляющая память СМ хранит микрооперации Y , сигнал управления синхронизацией y_0 и сигнал управления выборкой из управляющей памяти y_E .

Устройство управления U_1 функционирует следующим образом. По сигналу Start = 1 в счетчик СТ и регистр RG заносятся нулевые коды, что соответствует адресу $A(b_q)$, где $\langle b_0, b_q \rangle \in E$, одновременно триггер TF устанавливается в единичное состояние и сигнал Fetch разрешает выборку микрокоманд из СМ. Очередная микрокоманда $Y(b_q)$ считывается из СМ. Если $b_q \neq O_g$ ($g = \overline{1, G}$), то одновременно с микрооперациями Y формируется сигнал y_0 . При $y_0 = 1$ по сигналу Clock к содержимому СТ прибавляется единица, что соответствует адресации микрокоманд в пределах одной ОЛЦ, содержимое RG при этом не меняется. Если $b_q = O_g$ ($g = \overline{1, G}$), то сигнал y_0 не формируется, и по сигналу Clock в RG заносится код очередной ОЛЦ, определяемый функциями Ψ , а в СТ заносится код компоненты, определяемый функциями Φ . При достижении адреса $A(b_t)$, где $\langle b_t, b_E \rangle \in E$, формируется сигнал y_E , триггер TF обнуляется и выборка микрокоманд из СМ прекращается.

Такая организация КМУУ имеет ряд положительных качеств [2]:

- код ОЛЦ не зависит от адреса микрокоманды, что может быть использовано для оптимального кодирования ОЛЦ и оптимизации числа термов в схеме СС;
- число сигналов обратной связи меньше, чем длина адреса микрокоманды, что позволяет уменьшить число входов схемы СС по сравнению с КМУУ с общей памятью [1].

Однако при выполнении условия

$$R_1 + R_2 > R, \quad (6)$$

где $R = \lceil \log_2 M \rceil$, $M = |B_1|$, емкость управляющей памяти КМУУ U_1 резко увеличивается по сравнению с ее минимальным размером

$$V_{\min} = 2^R(N + 2). \quad (7)$$

Кроме того, оптимальное кодирование псевдоэквивалентных ОЛЦ, при котором коды ОЛЦ одного класса эквивалентности входят в один обобщенный интервал R_1 -мерного булевого пространства, не всегда возможно и зависит от характеристик исходной ГСА.

В настоящей работе предлагается метод организации КМУУ, позволяющий сохранить положительные каче-

ства КМУУ U_1 при выполнении (6) и невозможности оптимального кодирования псевдоэквивалентных ОЛЦ. Предлагаемый метод позволяет при выполнении определенных условий уменьшить емкость управляющей памяти по сравнению с параметром V_{\min} .

2 ОСНОВНАЯ ИДЕЯ МЕТОДА

Предлагаемый метод основан на введении:

– преобразователя АТ адреса микрокоманды, выраженного как (3), в адрес $A'(b_q)$ минимальной разрядности $R_{\min} \leq R$;

– преобразователя СОТ кодов ОЛЦ $\alpha_g \in C$ в коды классов псевдоэквивалентных ОЛЦ по аналогии с преобразованием кодов состояний автомата Мура [7].

Применение этих методов приводит к КМУУ U_2 (рис. 2).

Здесь преобразователь АТ реализует систему функций

$$\Lambda = \Lambda(\tau, T), \quad (8)$$

используемых для адресации микрокоманд, а преобразователь СОТ реализует систему функций

$$Z = Z(\tau), \quad (9)$$

используемых для формирования функций возбуждения регистром и счетчиком

$$\Phi = \Phi(Z, X), \quad (10)$$

$$\Psi = \Psi(Z, X). \quad (11)$$

Пусть в операторных вершинах $b_q \in B_1$ содержится Q попарно различных наборов микроопераций, тогда для кодирования этих наборов достаточно

$$R_0 = \lceil \log_2 Q \rceil \quad (12)$$

разрядов. Если

$$R_0 < R, \quad (13)$$

то емкость СМ КМУУ U_2 уменьшается по сравнению с параметром V_{\min} .

В настоящей работе предлагается метод синтеза КМУУ U_2 , включающий следующие этапы.

1. Формирование преобразованной ГСА Г.
2. Формирование множества ОЛЦ преобразованной ГСА.
3. Кодирование операторных линейных цепей.
4. Кодирование операторных вершин.
5. Адресация микрокоманд.
6. Формирование содержимого управляющей памяти.
7. Кодирование классов псевдоэквивалентных ОЛЦ.

8. Формирование таблицы переходов КМУУ.

9. Формирование таблицы преобразователя адресов.

10. Формирование таблицы преобразователя кодов.

11. Формирование систем функций Φ, Ψ, Λ, Z .

12. Реализация схем КМУУ в заданном элементе базисе.

3 ПРИМЕР ПРИМЕНЕНИЯ ПРЕДЛОЖЕННОГО МЕТОДА

Рассмотрим применение предложенного метода на примере ГСА Γ_1 (рис. 3).

Преобразование исходной ГСА. Преобразование сводится к введению дополнительных операторных вершин, а также к введению дополнительных сигналов в вершины исходной ГСА. Если существует дуга $\langle b_0, b_q \rangle \in E$, где $b_q \in B_2$, то в ГСА Γ вводится пустая вершина $b_t \in B_1$, и дуга $\langle b_0, b_q \rangle$ заменяется дугами $\langle b_0, b_t \rangle$ и $\langle b_t, b_q \rangle$. Если существует дуга $\langle b_q, b_E \rangle$, где $b_q \in B_1$, то в вершину b_q вводится сигнал y_E (в нашем примере – в b_{17}). Если существует дуга $\langle b_q, b_E \rangle$, где $b_q \in B_2$, то в ГСА Γ вводится вершина $b_t \in B_1$ с сигналом y_E , а дуга $\langle b_q, b_E \rangle$ заменяется дугами $\langle b_q, b_t \rangle$ и $\langle b_t, b_E \rangle$.

Для рассматриваемого примера структура преобразованной ГСА совпадает со структурой исходной ГСА.

Формирование множества ОЛЦ. Применение процедуры из работы [2] приводит к формированию следующего множества операторных линейных цепей $C = \{\alpha_1, \dots, \alpha_8\}$, где $\alpha_1 = \langle b_1 \rangle$, $I_1^1 = O_1 = b_1$; $\alpha_2 = \langle b_2, b_3 \rangle$, $I_2^1 = b_2$, $O_2 = b_3$; $\alpha_3 = \langle b_4, b_5 \rangle$, $I_3^1 = b_4$, $O_3 = b_5$; $\alpha_4 = \langle b_6, \dots, b_{10} \rangle$, $I_4^1 = b_6$, $I_4^2 = b_9$, $O_4 = b_{10}$; $\alpha_5 = \langle b_{11}, b_{12} \rangle$, $I_5^1 = b_{11}$, $O_5 = b_{12}$; $\alpha_6 = \langle b_{13}, b_{14} \rangle$, $I_6^1 = b_{13}$, $O_6 = b_{14}$; $\alpha_7 = \langle b_{15}, b_{16} \rangle$, $I_7^1 = b_{15}$, $O_7 = b_{16}$; $\alpha_8 = \langle b_{17} \rangle$, $I_8^1 = O_8 = b_{17}$. Итак, $G = 8$, $F_{\max} = 5$, $O(\Gamma_1) = \{b_1, b_3, b_5, b_{10}, b_{12}, b_{14}, b_{16}, b_{17}\}$. Отметим, что во все вершины $b_q \in O(\Gamma_1)$ вводится сигнал y_0 .

Кодирование операторных линейных цепей. В рассматриваемом примере разбиение P_C множества ОЛЦ C на классы псевдоэквивалентных ОЛЦ имеет $I = 4$ класса, где $B_1 = \{\alpha_1\}$, $B_2 = \{\alpha_2, \alpha_3, \alpha_4\}$, $B_3 = \{\alpha_5, \alpha_6, \alpha_7\}$, $B_4 = \{\alpha_8\}$. Отметим, что в данном случае оптимальное кодирование классов [7] невозможно. В этой связи кодируем ОЛЦ $\alpha_g \in C$ кодами $K(\alpha_g)$ разрядности $R_1 = \lceil \log_2 8 \rceil = 3$ произвольным образом: $K(\alpha_1) = 000$, ..., $K(\alpha_8) = 111$.

Кодирование операторных вершин. В рассматриваемом примере $F_{\max} = 5$, следовательно, $R_2 = 3$. При своем первом компонентом ОЛЦ $\alpha_g \in C$ нулевой код: $K(b_1) = K(b_2) = K(b_4) = K(b_6) = K(b_{11}) = K(b_{13}) = K(b_{15}) = K(b_{17}) = 000$. Тогда, согласно условию (2), имеем $K(b_3) = K(b_5) = K(b_{10}) = K(b_{12}) = K(b_{14}) = K(b_{16}) = 001$, $K(b_8) = 010$, $K(b_9) = 011$, $K(b_{10}) = 100$.

Адресация микрокоманд. Пусть $Y_q \subseteq Y$ – набор микроопераций, включающий элементы множества Y , y_0 и y_E . С учетом всех преобразований в ГСА Γ_1 можно выделить $Q = 11$ наборов микроопераций:

- $Y_1 = Y(b_1) = Y(b_5) = \{y_1, y_2\};$
- $Y_2 = Y(b_2) = Y(b_7) = \{y_0, y_3\};$
- $Y_3 = Y(b_3) = Y(b_{12}) = \{y_2, y_4\};$
- $Y_4 = Y(b_4) = \{y_0, y_2, y_5\};$
- $Y_5 = Y(b_6) = Y(b_{13}) = \{y_0, y_1, y_2\};$
- $Y_6 = Y(b_8) = Y(b_{15}) = \{y_2, y_4\};$
- $Y_7 = Y(b_9) = \{y_0, y_3, y_4, y_5\};$
- $Y_8 = Y(b_{10}) = Y(b_{14}) = \{y_3\};$
- $Y_9 = Y(b_{11}) = \{y_0, y_3, y_6\};$
- $Y_{10} = Y(b_{16}) = \{y_3, y_6\};$
- $Y_{11} = Y(b_{17}) = \{y_3, y_E\}.$

Итак, в рассматриваемом примере общее число операторных вершин $M = 17$ и для их адресации необходимо $R = 5$ разрядов, а число наборов микрокоманд $Q = 11$ и для их кодирования достаточно $R_0 = 4$ разряда. Следовательно, условие (13) выполняется и емкость управляющей памяти в КМУУ U_2 будет меньше параметра V_{\min} в два раза.

Условимся рассматривать наборы $Y_q \subseteq Y \cup \{y_0, y_E\}$ как микрокоманды и назовем их двоичные коды $A(Y_q)$ адресами микрокоманд. В рассматриваемом примере $\Lambda = \{\lambda_1, \dots, \lambda_4\}$, пусть $A(Y_1) = 0000, \dots, A(Y_{10}) = 1010$. Отметим, что в данном случае адресация микрокоманд проводится производным образом.

Формирование содержимого управляющей памяти.

Управляющая память СМ задается таблицей, фрагмент которой для нашего случая представлен в табл. 1.

Таблица 1 – Фрагмент содержимого управляющей памяти КМУУ $U_2(\Gamma_1)$

Адрес	Y_q	y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_E	Примечание
0000	Y_1	0	1	1	0	0	0	0	0	$Y(b_1); Y(b_5)$
0001	Y_2	1	0	0	1	0	0	0	0	$Y(b_2); Y(b_7)$
0010	Y_3	0	0	1	0	1	0	0	0	$Y(b_3); Y(b_{12})$

Здесь $U_2(\Gamma_1)$ означает, что КМУУ U_2 синтезируется по ГСА Γ_1 , в столбце «Примечание» записаны микрокоманды исходной ГСА, соответствующие микрокоманде Y_q ($q = 1, \dots, Q$).

Кодирование классов псевдоэквивалентных ОЛЦ.

В рассматриваемом примере $\Pi_C = \{B_1, \dots, B_4\}$, $I = 4$, следовательно, для кодирования классов $B_i \in \Pi_C$ достаточно $R_3 = \lceil \log_2 I \rceil = 2$ переменных и $Z = \{z_1, z_2\}$. Пусть $K(B_1) = 00, \dots, K(B_4) = 11$.

Формирование таблицы переходов КМУУ.

Эта таблица отображает закон формирования функций (10)–(11). Для формирования таблицы переходов построим систему обобщенных формул перехода для классов $B_i \in \Pi'_C$, где класс $B_i \notin \Pi'_C$, если выходы ОЛЦ

$\alpha_g \in B_i$ содержат сигнал y_E . Обобщенная формула имеет вид:

$$B \rightarrow \vee X_i(I_g^j)I_g^j, \tag{14}$$

где $X_i(I_g^j)$ – конъюнкция логических условий, определяющая переход из выхода любой ОЛЦ $\alpha_t \in B_i$ во вход I_g^j ($B_i \in \Pi'_C$).

В рассматриваемом примере $\Pi'_C = \{B_1, B_2, B_3\}$ и система (14) имеет вид:

$$\begin{aligned} B_1 &\rightarrow x_1 I_2^1 \vee \bar{x}_1 x_2 I_3^1 \vee \bar{x}_1 \bar{x}_2 I_4^1; \\ B_2 &\rightarrow x_2 I_5^1 \vee \bar{x}_2 x_3 I_6^1 \vee \bar{x}_2 \bar{x}_3 I_7^1; \\ B_3 &\rightarrow x_4 I_8^1 \vee \bar{x}_4 I_4^2. \end{aligned}$$

Таблица переходов КМУУ U_2 имеет столбцы: $B_i, K(B_i), I_g^j, A(I_g^j), X_h, \Phi_h, \Psi_h, h$, где адрес входа I_g^j представляется в виде (3), Ψ_h – набор функций возбуждения, формирующий в RG код $K(\alpha_g)$, Φ_h – набор функций возбуждения, формирующий в СТ код компоненты $K(b_t)$, соответствующий входу I_g^j ; $h = \bar{1}, \bar{N}$ – номер перехода. Очевидно, параметр N равен числу дизъюнктивных членов в системе (14).

Фрагмент таблицы переходов КМУУ $U_2(\Gamma_1)$ представлен в табл. 2.

Таблица 2 – Фрагмент таблицы переходов КМУУ $U_2(\Gamma_1)$

B_i	$K(B_i)$	I_g^j	$A(I_g^j)$	$X_h(I_g^j)$	Ψ_h	Φ_h	h
B_1	00	I_2^1	001000	x_1	D_3	–	1
		I_3^1	010000	$\bar{x}_1 x_2$	D_2	–	2
		I_4^1	011000	$\bar{x}_1 \bar{x}_2$	$D_2 D_3$	–	3

Формирование таблицы преобразователя адреса.

Эта таблица задает закон формирования системы (8) и включает столбцы: $b_m, A(b_m), Y_q, A(Y_q), \lambda_m, m$, где λ_m – набор переменных $\lambda_r \in \Lambda$, принимающих единичное значение в адресе микрокоманды Y_q , соответствующей операторной вершине $b_m \in B_1$; m – номер строки таблицы ($m = 1, \dots, M$).

Для рассматриваемого примера число строк $M = 17$, фрагмент таблицы преобразователя адресов КМУУ $U_2(\Gamma_1)$ приведен в табл. 3.

Таблица 3 – Фрагмент таблицы преобразователя адресов КМУУ $U_2(\Gamma_1)$

b_m	$A(b_m)$	Y_q	$A(Y_q)$	λ_m	m
b_1	000000	Y_1	0000	–	1
b_2	001000	Y_2	0001	λ_4	2
b_3	001001	Y_3	0010	λ_3	3
b_4	010000	Y_4	0011	$\lambda_3 \lambda_4$	4

Формирование таблицы преобразователя кодов.

Эта таблица задает систему (9) и включает столбцы: $\alpha_g, K(\alpha_g), B_i, K(B_i), Z_g, g$, где Z_g – набор переменных $z_r \in Z$, принимающих единичное значение в коде $K(B_i)$, где $B_i \in \Pi'_C$. В случае КМУУ $U_2(\Gamma_1)$ эта таблица включает $G_0 = 7$ (табл. 4).

Таблица 4 – Таблица преобразователя кодов КМУУ $U_2(\Gamma_1)$

α_g	$K(\alpha_g)$	B_i	$K(B_i)$	Z_g	G
α_1	000	B_1	00	–	1
α_2	001	B_2	01	z_2	2
α_3	010	B_2	01	z_2	3
α_4	011	B_2	01	z_2	4
α_5	100	B_3	10	z_1	5
α_6	101	B_3	10	z_1	6
α_7	110	B_3	10	z_1	7

Формирование систем функций КМУУ.

Функции Φ и Ψ формируются по таблице переходов КМУУ, дизъюнктивные нормальные формы функций (10)–(11) включают термы, соответствующие строкам таблицы переходов. Например, из табл. 2 имеем $D_3 = \bar{z}_1 \bar{z}_2 x_1 \vee (\bar{z}_1 \bar{z}_2 \bar{x}_1 \bar{x}_2)$. Функции Λ формируются по таблице преобразователя адресов. Например, из табл. 3 имеем $\lambda_4 = \bar{\tau}_1 \bar{\tau}_2 \tau_3 \bar{T}_1 \bar{T}_2 \bar{T}_3 \vee \bar{\tau}_1 \tau_2 \bar{\tau}_3 \bar{T}_1 \bar{T}_2 \bar{T}_3$. Функции (9) реализуются по таблице преобразователя кодов. Например, из табл. 4 имеем $z_1 = \tau_1 \bar{\tau}_2 \bar{\tau}_3 \vee \tau_1 \bar{\tau}_2 \tau_3 \vee \tau_1 \tau_2 \bar{\tau}_3 = \tau_1 \bar{\tau}_2 \vee \tau_1 \bar{\tau}_3$. Функции $Y \cup \{y_0, y_E\}$ формируются по таблице содержимого управляющей памяти. Отметим, что наиболее естественным способом реализации схем СМ и СОТ является использование ПЗУ или других блоков памяти.

Реализация схемы КМУУ в заданном элементном базисе. Методы реализации схем в базисе CPLD, FPGA и PROM достаточно полно рассмотрены в литературе [4, 5, 6, 8]. Таким образом, для реализации схем $Y, \Lambda, \Phi, \Psi, Z$ могут быть использованы стандартные методы, что выходит за рамки данной работы.

ЗАКЛЮЧЕНИЕ

Применение метода преобразования адресов позволяет использовать метод разделения кодов при синтезе КМУУ вне зависимости от характеристик исходной ГСА, определяющих разрядности кодов ОЛЦ, компонент ОЛЦ и минимальную разрядность адреса микрокоманды. При этом сохраняется возможность оптимального кодирования псевдоэквивалентных ОЛЦ, что

позволяет уменьшить аппаратные затраты в схеме СС по сравнению с произвольным кодированием ОЛЦ. При выполнении условия (6) уменьшение аппаратных затрат возможно за счет применения преобразователя кодов ОЛЦ в коды классов псевдоэквивалентных ОЛЦ. Отметим, что применение преобразователя адреса приводит к увеличению времени цикла КМУУ U_2 по сравнению с соответствующим параметром КМУУ U_1 . Однако, такой подход дает потенциальную возможность уменьшения емкости управляющей памяти КМУУ U_2 по сравнению со всеми известными методами реализации КМУУ [2]. Исследования авторов показали, что при выполнении условия (6), невозможности оптимального кодирования ОЛЦ и уменьшении емкости управляющей памяти предлагаемый метод позволяет на 14–17 % уменьшить аппаратные затраты по сравнению с известными методами реализации КМУУ.

ПЕРЕЧЕНЬ ССЫЛОК

1. Баркалов А. А., Палагин А. В. Синтез микропрограммных устройств управления – Киев: ИК НАН Украины, 1997. – 136 с.
2. Баркалов А. А. Синтез устройств управления на программируемых логических устройствах. – Донецк: ДонНТУ, 2002. – 262 с.
3. Грушвицкий Р. И., Мурсаев А. Х., Угрюмов Е. П. Проектирование систем на микросхемах программируемой логики. – Петербург: БХВ. – 2002. – 636 с.
4. Luba T. Synteza ukladow logicznych. – Warszawa: WKL, 2001. – 238 s.
5. Salcic Z. VHDL and FPLDs in Digital systems Design, Prototyping and Customization. – Kluwer Academic Publishers, 1998. – 412 p.
6. Sasao T. Switching theory for logic synthesis. – Kluwer Academic Publishers, 1999. – 355 p.
7. Баркалов А. А. Принципы оптимизации логической схемы микропрограммного автомата Мура. // Кибернетика и системный анализ. – 1998. – № 1. – С. 65–72.
8. Соловьев В. В. Проектирование цифровых схем на основе программируемых логических интегральных схем. – М.: Горячая линия-ТЕЛЕКОМ, 2001. – 636 с.

Надійшла 7.07.06
Після доробки 7.08.06

Запропоновано метод оптимізації композиційного мікропрограмного пристрою керування з розділенням кодів. Метод заснований на введенні перетворювачів адреси мікрокоманди та кодів операторних лінійних ланцюгів до схеми пристрою керування. Метод дозволяє зменшити ємність керуючої пам'яті. Розглянуто приклад застосування запропонованого методу.

The method of compositional microprogram control unit with codes sharing optimization is proposed. Method is based on adding of the converters of microinstruction address and operator linear chains' codes in the circuit of the control unit. The method allows reducing the capacity of control memory. The example of application of the offered method is given.

УДК 539.1.075

В. В. Брайловський, П. П. Ватаманюк, М. Г. Рождественська

ІМІТАТОР СИГНАЛІВ ЯДЕРНОЇ ІНДУКЦІЇ ДЛЯ ІМПУЛЬСНОГО ЯКР-СПЕКТРОМЕТРА

Запропоновано схему імітатора сигналів для налагодження імпульсних спектрометрів ядерного квадрупольного резонансу (ЯКР). Коливання, подібні до сигналів ядерної індукції, генеруються кварцовим резонатором. Ударне збудження кварцового резонатора забезпечується радіоімпульсами накачування спінової системи досліджуваного зразка речовини.

ВСТУП

На сьогодні актуальною залишається проблема розробки і вдосконалення радіотехнічних засобів виявлення, ідентифікації та дослідження азотомістких речовин. Зокрема, триває вдосконалення схемотехніки радіоспектрометрів, робота яких базується на явищі ядерного квадрупольного резонансу, та розвиваються методики їх застосування [1, 2].

Існують два основні типи ЯКР-спектрометрів – імпульсні та стаціонарні (неперервні). Імпульсний метод дозволяє спостерігати сигнали ЯКР від легких ядер, які мають квадрупольні властивості. Це стосується і ядер ^{14}N , що входять до складу таких речовин, як вибухівки (гексоген, ТНТ, октоген) та наркотики [3].

Квадрупольні моменти легких ядер малі, у зв'язку з чим електромагнітні відгуки атомної спінової системи на збуджуючий радіоімпульс лежать в низькочастотній ділянці спектру (орієнтовно 0,6–6,0 МГц).

Чутливість прямих імпульсних методів детектування сигналів ЯКР погіршується з пониженням частоти. Величина сигналів ЯКР-відгуку знаходиться на рівні теплового шуму робочого контура спектрометра, і вони, здебільшого, не можуть бути виділені без спеціальної обробки.

Тому перший етап налагодження ЯКР-установки після її побудови чи модернізації з використанням безпосередньо досліджуваної речовини досить складний. Робота з вибуховими речовинами на етапі оптимізації установки може бути небезпечною. Деякі тестові речовини, наприклад уротропін ($(\text{CH}_2)_6\text{NO}_4$), дають відносно сильний сигнал ядерної індукції, проте його очікувана частота точно невідома, оскільки вона залежить від температури і хімічної чистоти речовини. Так, для хімічно чистого уротропіну частота ЯКР-відгуку в діапазоні температур 273–300 К змінюється від 3319,9 кГц до 3306,2 кГц.

© Брайловський В. В., Ватаманюк П. П., Рождественська М. Г., 2006

ПОСТАНОВКА ЗАВДАННЯ

Для налагодження спектрометра бажано застосувати імітатор ЯКР-відгуку з точно відомою частотою та регулюванням сили сигналу в широких межах: від сотень мілівольт до рівня теплового шуму на робочому контурі з досліджуваною речовиною. Форма згасаючих радіоімпульсів імітатора та їх часові характеристики повинні наближатись до відповідних показників сигналів ЯКР-відгуку реальних речовин. Наприклад, при ідентифікації гексогену збудженням квадрупольних моментів ядер його атомів ^{14}N радіоімпульсами з частотою заповнення $f_{36} = 5192$ кГц, час спин-спінової релаксації при температурі 300 К складає близько 1,5 мс, а спин-граткової релаксації – 8 мс.

МЕТОД ТА АЛГОРИТМ ВИРІШЕННЯ

Сформульованим вище вимогам у значній мірі відповідає простий імітатор на кварцовому резонаторі, схема якого наведена на рис. 1.

Для грубої оцінки працездатності спектрометра кварцовий резонатор спочатку встановлюється в рознім X1. Частота заповнення збуджуючих радіоімпульсів, що підводяться до входу робочої камери, та частота настройки

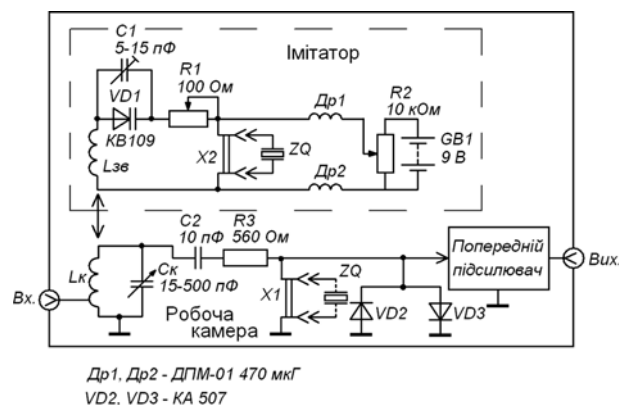


Рисунок 1 – Схема імітатора та робочої камери імпульсного ЯКР-спектрометра

робочого контура L_K , C_K встановлюються близькими до частоти послідовного резонансу кварца.

Під час дії на робочому контурі потужного радіоімпульса збудження та відразу після його закінчення напруга на резонаторі ZQ , обмежена $p-i-n$ діодами $VD2$, $VD3$, не перевищує $U_{\max} = 1$ В, що запобігає руйнуванню кварцової пластини. Наступне згасання амплітуди коливань до деякого рівня U_{\min} відбувається за час

$$\tau = \frac{Q_{\text{кв}}}{\pi f_0} \ln \frac{U_{\max}}{U_{\min}}, \quad (1)$$

де $Q_{\text{кв}}$ – добротність кварцового резонатора, f_0 – його резонансна частота.

Якщо за U_{\min} прийняти напругу теплового шуму на робочому контурі L_K , C_K , то вираз (1) дозволяє розрахувати тривалість згасання радіоімпульса, що імітує потужний сигнал ЯКР-відгуку. Величину шумової напруги на контурі можна знайти за формулою Найквіста:

$$U_{\text{ш}} = \sqrt{4kT_{\text{ш}}\Delta f_{\text{пр}}Z_K}. \quad (2)$$

Тут $T_{\text{ш}}$ – шумова температура контура, $\Delta f_{\text{пр}}$ – смуга пропускання приймача спектрометра, Z_K – резонансний опір контура. Вираз (2) дозволяє оцінити шумову напругу неохолодженого контура, настроєного на частоту $f_0 = 5$ МГц, на рівні $U_{\text{ш}} \approx 2$ мкВ. Приймавши добротність кварцового резонатора $Q_{\text{кв}} = 10^4$, з (1) отримуємо: $\tau \approx 8$ мс.

На наступному етапі налагодження резонатор перевстановлюється в рознім $X2$. Імітатор зв'язаний індуктивно з котушкою робочого контура. Котушка зв'язку діаметром 1 см складається з двох витків мідного проводу.

Плата імітатора може переміщуватись відносно котушки робочого контура. Наведений в робочому контурі згасаючими коливаннями кварцового генератора сигнал залежить від відстані між котушками $L_{\text{зв}}$ та L_K і їх взаємної орієнтації. Пристрій дозволяє змінювати рівень імітаційних сигналів у широкому діапазоні: від сотень мілівольт до одиниць мікрівольт, тобто до рівня напруги теплового шуму на робочому контурі.

Частота відгуку в межах декількох кілогерц може змінюватись за допомогою варикапу $VD1$. Імітація зміни ширини лінії в спектрі ЯКР досягається варіацією добротності кола резонатора ZQ за допомогою регулювання резистором $R1$ величини внесеного активного опору втрат.

Імітатор може бути застосований також для налаштування радіоспектрометра, що працює в режимі стимулювання квадрупольних переходів імпульсом збудження з частотою заповнення, розстроєною відносно частоти лінії в спектрі досліджуваної речовини. Величина можливого розстроювання при роботі імітатора

на частоті $f_0 = 5$ МГц та регулюванні резистором $R1$ добротності кварцового резонатора в межах $Q_{\text{кв}} = 10^3-10^4$ може бути оцінена як $\Delta f = f_0/Q_{\text{кв}} = 0,5-5,0$ кГц.

РЕЗУЛЬТАТИ

Робота описаного пристрою ілюструється осцилограмою сигналу імітації ЯКР-відгуку (рис. 2). Осцилограма одержана за допомогою цифрового запам'ятовуючого осцилографа типу Tektronix TDS 1012, під'єднаного до виходу імпульсного спектрометра.

В даному випадку застосовано кварцовий резонатор на частоту $f_0 = 5000$ кГц. Частота заповнення радіоімпульсів збудження становила $f_{\text{зб}} = 5001,5$ кГц. В процесі оптимізації режимів всіх вузлів спектрометра зв'язок імітатора з робочим контуром поступово зменшувався. В результаті зафіксовано сигнал імітації, співрозмірний з рівнем шумової напруги на контурі. Виділення корисного сигналу здійснено програмно-апа-

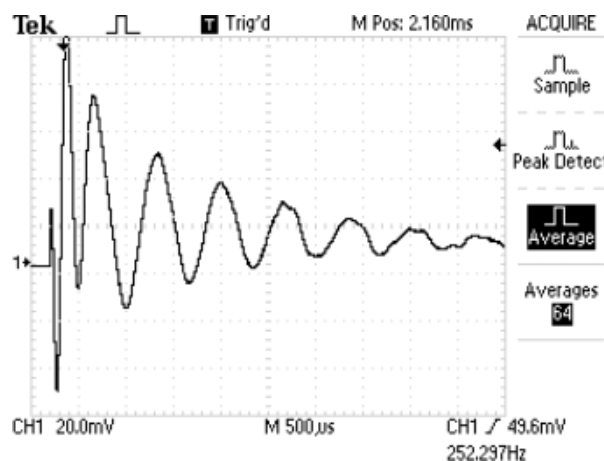


Рисунок 2 – Вигляд сигналів імітації на виході спектрометра

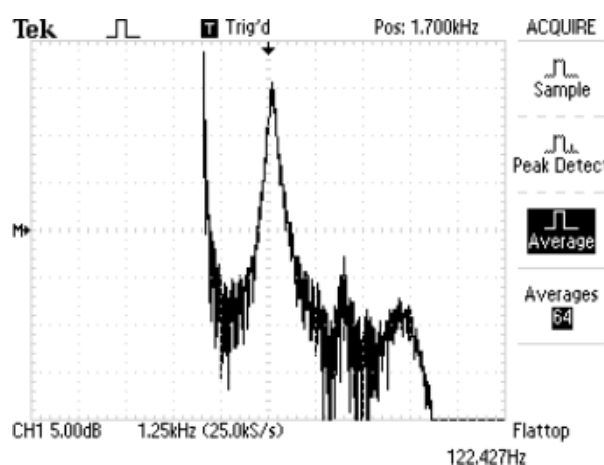


Рисунок 3 – Сигнали ЯКР-відгуку ядер атомів ^{14}N зразка тестової речовини NaN_2O

ратним методом багаторазового (64 реалізації) усереднення зашумленого сигналу.

Для порівняння на рис. 3 приведено реальні сигнали ЯКР-відгуку, одержані збудженням квадрупольних моментів ядер атомів ^{14}N зразка тестової речовини NaNO_2 .

Тривалість збуджуючих радіоімпульсів τ становила 40 мкс. Для виконання умови $\gamma\tau H = \pi/2$, що необхідно для ефективного збудження спінової системи ядер азоту в NaNO_2 , імпульсна потужність передавача спектрометра була встановлена на рівні близько 200 Вт. Тут γ – гіромагнітне відношення ядра, H – напруженість магнітної компоненти поля. Робота велась в режимі розстройки частоти заповнення збуджуючих радіоімпульсів ($f_{36} = 4645$ кГц) відносно частоти ЯКР-відгуку. Об'єм робочого контура спектрометра складав 0,1 л, а маса зразка тестової речовини – 0,5 г. Зразок знаходився при кімнатній температурі. На осцилограмі видно биття сигналу ЯКР-відгуку із збуджуючими коливаннями. При цьому огинаюча згасаючих осциляцій співпадає з кривою спін-граткової релаксації збуджених ядер атомів ^{14}N в NaNO_2 , характерною для наведених умов експерименту.

ВИСНОВКИ

Як видно, форма згасаючих радіоімпульсів імітатора та їх амплітудно-часові характеристики набли-

жаються до відповідних показників сигналу ЯКР-відгуку ядер реальної речовини.

Дана робота проведена в рамках проектної угоди 2369 між ЧНУ та Науково-технологічним центром України.

ПЕРЕЛІК ПОСИЛАНЬ

1. Rudakov T. N., Shevchenko S. N. A modified method for measuring the spin relaxation in nuclear quadrupole resonance // Instruments and Experimental techniques. – 2003. – Vol. 46, No. 3. – P. 373–375.
2. Белый Ю. И., Гарцев Н. А, Семейкин Н. П., Шаршин Ю. А. Разработка обнаружителей взрывчатки на основе ядерного квадрупольного резонанса (ЯКР) // Научные технологии. – 2005. – Т. 6, № 2. – С. 38–40.
3. Гречишкин В. С., Синявский Н. Я. Новые физические технологии: обнаружение взрывчатых и наркотических веществ методом ядерного квадрупольного резонанса // УФН. – 1997. – Т. 167, № 4. – С. 413–427.

Надійшла 12.04.06

Предложена схема имитатора сигналов для настройки импульсных спектрометров ядерного квадрупольного резонанса (ЯКР). Колебания, схожие с сигналами ядерной индукции, генерируются кварцевым резонатором. Ударное возбуждение кварцевого резонатора обеспечивается радиоимпульсами накачки спиновой системы исследуемого образца вещества.

The circuit of signal imitator for adjusting of nuclear quadrupole resonance (NQR) pulsed spectrometer is proposed. Oscillations, which are similar to nuclear induction signals, are generated by a quartz resonator. Impact excitation of the quartz resonator is realized by pump pulses of spin system of the investigated matter sample.

УДК 539.2

В. И. Грядун

ВЛИЯНИЕ АКТИВНЫХ АТОМОВ НА ЭНЕРГИЮ ОБРАЗОВАНИЯ ТОЧЕЧНЫХ ДЕФЕКТОВ В КРЕМНИИ

Методом механико-молекулярного моделирования AMBER рассматриваются точечные дефекты и комплексы атомов с вакансиями и междоузельными атомами кремния. Рассчитаны энергии образования точечных дефектов, а также комплексов V-X и IA-X. Показано, что только атомы водорода в комплексе с вакансией снижают энергию ее образования, а для комплексов Si-X она значительно превышает энергию образования собственного междоузельного атома.

ВВЕДЕНИЕ

Новые технологии СБИС заостряют проблему микродефектов полупроводниковых структур, так как отношение размера микродефекта к минимальному тех-

нологическому размеру постоянно возрастает. Это влечет за собой необходимость более глубокого изучения физических явлений в полупроводниках, в частности, процессов генерации точечных дефектов в кристаллах кремния. Пока что нет точных теоретических расчетов энергии образования дефектов Френкеля и Шоттки в полупроводниках, поэтому является целесообразным сделать квантово-механические оценки в этом направлении, как для идеальных дефектов так и для дефектов, возмущенных полем дополнительных атомов, с помощью современных методов молекулярного моделирования. Целью настоящей работы является выяснение влияния активных атомов на энергию образования точечных дефектов в кремнии.

© Грядун В. И., 2006

ОСНОВНАЯ ЧАСТЬ

Рассматривается наноструктура из восьми, ($2 \times 2 \times 2$), ячеек кремния, в которых находится 95 атомов. При механико-молекулярном моделировании [1] связанные атомы рассматриваются как молекулы: нулевая молекула включает 91 атом; первая, вторая, третья и четвертая молекулы представляют собой отдельные, не имеющие связей, атомы. Координаты атомов и межатомные связи представлены в следующем файле:

ChemSite system file Si22200.sync

PRMS 10.000 0 0 3.100 1 1 20.000 20.000 20.000
NMOL 0

ATOM0 14 -5.401 -5.401 -5.401 1 1 SUP 0 Si 0
ATOM1 14 -4.050 -4.050 -4.050 4 4 SUP 0 Si 0
ATOM2 14 -5.402 -2.700 -2.700 4 2 SUP 0 Si 0
ATOM3 14 -4.051 -1.350 -1.350 4 4 SUP 0 Si 0
ATOM4 14 -2.702 -2.702 0.001 4 4 SUP 0 Si 0
ATOM5 14 -2.702 0.001 -2.702 4 4 SUP 0 Si 0
ATOM6 14 -5.403 -0.000 -0.000 4 2 SUP 0 Si 0
ATOM7 14 -1.351 -1.351 1.351 4 4 SUP 0 Si 0
ATOM8 14 -4.053 -4.053 1.353 4 4 SUP 0 Si 0
ATOM9 14 -1.351 1.351 -1.351 4 4 SUP 0 Si 0
ATOM10 14 -4.053 1.353 -4.053 4 4 SUP 0 Si 0
ATOM11 14 -4.051 1.350 1.350 4 4 SUP 0 Si 0
ATOM12 14 -2.702 -0.001 2.702 4 4 SUP 0 Si 0
ATOM13 14 -0.001 -2.702 2.702 4 4 SUP 0 Si 0
ATOM14 14 -0.000 -0.000 -0.000 4 4 SUP 0 Si 0
ATOM15 14 -2.702 2.702 -0.001 4 4 SUP 0 Si 0
ATOM16 14 -0.001 2.702 -2.702 4 4 SUP 0 Si 0
ATOM17 14 -1.350 1.350 4.051 4 4 SUP 0 Si 0
ATOM18 14 1.350 -4.051 1.350 4 4 SUP 0 Si 0
ATOM19 14 1.350 -1.350 4.051 4 4 SUP 0 Si 0
ATOM20 14 1.351 1.351 1.351 4 4 SUP 0 Si 0
ATOM21 14 -1.350 4.051 1.350 4 4 SUP 0 Si 0
ATOM22 14 1.350 1.350 -4.051 4 4 SUP 0 Si 0
ATOM23 14 1.350 4.051 -1.350 4 4 SUP 0 Si 0
ATOM24 14 -2.700 2.700 5.402 4 2 SUP 0 Si 0
ATOM25 14 0.001 2.702 2.702 4 4 SUP 0 Si 0
ATOM26 14 0.000 -0.000 5.403 4 2 SUP 0 Si 0
ATOM27 14 2.700 -2.700 5.402 4 2 SUP 0 Si 0
ATOM28 14 2.702 0.001 2.702 4 4 SUP 0 Si 0
ATOM29 14 -2.700 5.402 2.700 4 2 SUP 0 Si 0
ATOM30 14 0.000 5.403 -0.000 4 2 SUP 0 Si 0
ATOM31 14 2.702 2.702 0.001 4 4 SUP 0 Si 0
ATOM32 14 2.700 5.402 -2.700 4 2 SUP 0 Si 0
ATOM33 14 1.353 4.053 4.053 4 4 SUP 0 Si 0
ATOM34 14 4.053 1.353 4.053 4 4 SUP 0 Si 0
ATOM35 14 4.053 4.053 1.353 4 4 SUP 0 Si 0
ATOM36 14 2.706 2.706 5.405 4 2 SUP 0 Si 0
ATOM37 14 2.706 5.405 2.706 4 2 SUP 0 Si 0
ATOM38 14 0.025 5.416 5.416 1 1 SUP 0 Si 0
ATOM39 14 5.405 2.706 2.706 4 2 SUP 0 Si 0
ATOM40 14 5.416 0.025 5.416 1 1 SUP 0 Si 0
ATOM41 14 5.416 5.416 0.025 1 1 SUP 0 Si 0
ATOM42 14 -2.700 -5.402 -2.700 4 2 SUP 0 Si 0
ATOM43 14 -1.350 -4.051 -1.350 4 4 SUP 0 Si 0
ATOM44 14 0.001 -2.702 -2.702 4 4 SUP 0 Si 0
ATOM45 14 -0.000 -5.403 -0.000 4 2 SUP 0 Si 0
ATOM46 14 1.353 -4.053 -4.053 4 4 SUP 0 Si 0
ATOM47 14 1.351 -1.351 -1.351 4 4 SUP 0 Si 0
ATOM48 14 2.702 -2.702 -0.001 4 4 SUP 0 Si 0
ATOM49 14 2.702 -0.001 -2.702 4 4 SUP 0 Si 0

ATOM50 14 4.051 -1.350 1.350 4 4 SUP 0 Si 0
ATOM51 14 4.051 1.350 -1.350 4 4 SUP 0 Si 0
ATOM52 14 5.402 -2.700 2.700 4 2 SUP 0 Si 0
ATOM53 14 5.403 -0.000 -0.000 4 2 SUP 0 Si 0
ATOM54 14 5.402 2.700 -2.700 4 2 SUP 0 Si 0
ATOM55 14 -2.700 -2.700 -5.402 4 2 SUP 0 Si 0
ATOM56 14 -1.350 -1.350 -4.051 4 4 SUP 0 Si 0
ATOM57 14 -0.000 -0.000 -5.403 4 2 SUP 0 Si 0
ATOM58 14 -5.416 -5.416 0.025 1 1 SUP 0 Si 0
ATOM59 14 -5.405 -2.706 2.706 4 2 SUP 0 Si 0
ATOM60 14 -4.053 -1.353 4.053 4 4 SUP 0 Si 0
ATOM61 14 -2.706 -2.706 5.405 4 2 SUP 0 Si 0
ATOM62 14 -5.416 -0.025 5.416 1 1 SUP 0 Si 0
ATOM63 14 -2.706 -5.405 2.706 4 2 SUP 0 Si 0
ATOM64 14 -1.353 -4.053 4.053 4 4 SUP 0 Si 0
ATOM65 14 -0.025 -5.416 5.416 1 1 SUP 0 Si 0
ATOM66 14 -5.416 0.025 -5.416 1 1 SUP 0 Si 0
ATOM67 14 -5.405 2.706 -2.706 4 2 SUP 0 Si 0
ATOM68 14 -4.053 4.053 -1.353 4 4 SUP 0 Si 0
ATOM69 14 -2.706 5.405 -2.706 4 2 SUP 0 Si 0
ATOM70 14 -5.416 5.416 -0.025 1 1 SUP 0 Si 0
ATOM71 14 -2.706 2.706 -5.405 4 2 SUP 0 Si 0
ATOM72 14 -1.353 4.053 -4.053 4 4 SUP 0 Si 0
ATOM73 14 -0.025 5.416 -5.416 1 1 SUP 0 Si 0
ATOM74 14 -5.402 2.700 2.700 4 2 SUP 0 Si 0
ATOM75 14 -4.050 4.050 4.050 4 4 SUP 0 Si 0
ATOM76 14 -5.401 5.401 5.401 1 1 SUP 0 Si 0
ATOM77 14 0.025 -5.416 -5.416 1 1 SUP 0 Si 0
ATOM78 14 2.706 -5.405 -2.706 4 2 SUP 0 Si 0
ATOM79 14 4.053 -4.053 -1.353 4 4 SUP 0 Si 0
ATOM80 14 5.405 -2.706 -2.706 4 2 SUP 0 Si 0
ATOM81 14 5.416 -5.416 -0.025 1 1 SUP 0 Si 0
ATOM82 14 2.706 -2.706 -5.405 4 2 SUP 0 Si 0
ATOM83 14 4.053 -1.353 -4.053 4 4 SUP 0 Si 0
ATOM84 14 5.416 -0.025 -5.416 1 1 SUP 0 Si 0
ATOM85 14 2.700 -5.402 2.700 4 2 SUP 0 Si 0
ATOM86 14 4.050 -4.050 4.050 4 4 SUP 0 Si 0
ATOM87 14 5.401 -5.401 5.401 1 1 SUP 0 Si 0
ATOM88 14 2.700 2.700 -5.402 4 2 SUP 0 Si 0
ATOM89 14 4.050 4.050 -4.050 4 4 SUP 0 Si 0
ATOM90 14 5.401 5.401 -5.401 1 1 SUP 0 Si 0
BOND0 1 0
BOND1 2 0
BOND1 42 0
BOND1 55 0
BOND2 3 0
BOND3 4 0
BOND3 5 0
BOND3 6 0
BOND4 7 0
BOND4 8 0
BOND4 43 0
BOND5 9 0
BOND5 10 0
BOND5 56 0
BOND6 11 0
BOND7 12 0
BOND7 13 0
BOND7 14 0
BOND8 58 0
BOND8 59 0
BOND8 63 0
BOND9 14 1
BOND9 15 0
BOND9 16 0
BOND10 66 0

BOND10 67 0	BOND47 49 1
BOND10 71 0	BOND48 50 1
BOND11 12 1	BOND48 79 0
BOND11 15 1	BOND49 51 1
BOND11 74 0	BOND49 83 0
BOND12 17 0	BOND50 52 0
BOND12 60 0	BOND50 53 0
BOND13 18 0	BOND51 53 1
BOND13 19 0	BOND51 54 0
BOND13 64 0	BOND52 86 1
BOND14 20 0	BOND54 89 1
BOND14 47 0	BOND55 56 1
BOND15 21 0	BOND56 57 1
BOND15 68 0	BOND59 60 1
BOND16 22 0	BOND60 61 0
BOND16 23 0	BOND60 62 0
BOND16 72 0	BOND61 64 1
BOND17 24 0	BOND63 64 1
BOND17 25 0	BOND64 65 0
BOND17 26 0	BOND67 68 1
BOND18 45 0	BOND68 69 0
BOND18 48 0	BOND68 70 0
BOND18 85 0	BOND69 72 1
BOND19 26 1	BOND71 72 1
BOND19 27 0	BOND72 73 0
BOND19 28 0	BOND74 75 1
BOND20 25 1	BOND75 76 0
BOND20 28 1	BOND78 79 1
BOND20 31 0	BOND79 80 0
BOND21 25 1	BOND79 81 0
BOND21 29 0	BOND80 83 1
BOND21 30 0	BOND82 83 1
BOND22 49 0	BOND83 84 0
BOND22 57 0	BOND85 86 1
BOND22 88 0	BOND86 87 0
BOND23 30 1	BOND88 89 1
BOND23 31 1	BOND89 90 0
BOND23 32 0	NMOL 16
BOND24 75 0	ATOMO 14 -5.038 -5.038 5.038 1 0 SUP 0 Si 0
BOND25 33 0	NMOL 36
BOND27 86 0	ATOMO 14 -5.038 5.038 -5.038 1 0 SUP 0 Si 0
BOND28 34 0	NMOL 82
BOND28 50 0	ATOMO 14 5.038 -5.038 -5.038 1 0 SUP 0 Si 0
BOND29 75 1	NMOL 94
BOND31 35 0	ATOMO 14 5.038 5.038 5.038 1 0 SUP 0 Si 0
BOND31 51 0	
BOND32 89 0	
BOND33 36 0	
BOND33 37 0	
BOND33 38 0	
BOND34 36 1	
BOND34 39 0	
BOND34 40 0	
BOND35 37 1	
BOND35 39 1	
BOND35 41 0	
BOND42 43 1	
BOND43 44 0	
BOND43 45 1	
BOND44 46 0	
BOND44 47 1	
BOND44 56 1	
BOND46 77 0	
BOND46 78 0	
BOND46 82 0	
BOND47 48 1	

На рис. 1 показана базовая молекула, в которой мы будем моделировать микродефекты. Энергия рассматриваемой модели рассчитывается методом AMBER и составляет – 36,92 ккал/моль.

Собственный междоузельный атом моделируется программно путем удаления произвольного атома в объеме решетки и размещения дополнительного атома кремния в междоузельном пространстве кристалла (рис. 2). Тотальная минимальная энергия кристалла после геометрической оптимизации модели составила 25,76 ккал/моль и, следовательно, энергия образования междоузельного атома кремния составляет величину 2,72 эВ. В случае, когда собственный междоузельный атом кремния локализуется в непосредственной близости от вакансии, энергия его формирования несколько уменьшается и составляет величину 2,34 эВ. Занятое атомом кремния междоузелье оказалось окта-

эдрическим с расстояниями до ближайших атомов кремния: 2,424; 2, 2,439; 2,64; 2,657; 2,66; 2,706; 2,728; 2,730 А.

Рассматривались также модели междоузельных комплексов Si-H, Si-N, Si-O в кремнии, энергии которых составили, соответственно: 16,33; 20,16 и 18, 63 эВ. На рис. 3 показана геометрически оптимизированная наноструктура с междоузельным комплексом Si-N.

Вакансия в кристалле кремния моделировалась путем программного переноса объемного атома на поверхность (рис. 4). Минимизация энергии полученной модели кристалла с вакансией позволили рассчитать энергию образования вакансии, которая составила величину 1,44 эВ.

Вакансионные комплексы V-X моделировались программно путем добавления соответствующих атомов к одной их четырех свободных связей вакансии. Вакансионный комплекс V-N изображен на рис. 5.

Геометрическая оптимизация комплекса V-N в наноструктуре позволяет рассчитать его энергию образования, которая составила 1,25 эВ, и установить координаты атома водорода. Длина связи Si-H составляет 1,47 А, а ближайшие расстояния до трех других атомов кремния составляют 2,759; 2,759 и 2,733 А.

Подобные расчеты проведены также для комплексов V-N, V-O, V-F, V-S и V-Cl, энергии образования которых, соответственно, составили: 2,24; 2,42; 2,18; 3,6 и 3,38 эВ.

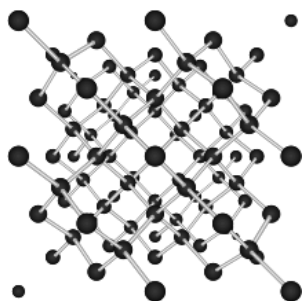


Рисунок 1 – Исходная модель кристалла кремния

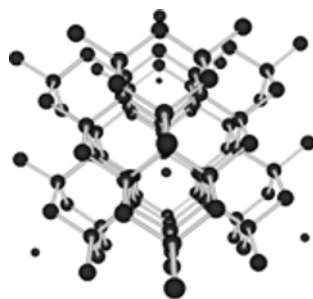


Рисунок 2 – Собственный междоузельный атом в кремнии

Длины связей вакансионных комплексов, соответственно, составили: 1,802; 1,766; 1,823; 2,105; 2,06 А.

Учитывая, что только атомы водорода уменьшают энергию формирования вакансий, были проведены также расчеты энергии формирования вакансий в присутствии двух, трех и четырех атомов водорода на произвольных связях вакансии, которые составили, соответственно, 1,31; 1,58 и 2,02 эВ.

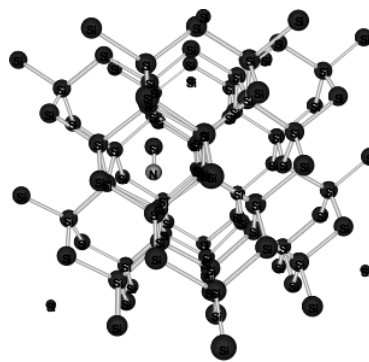


Рисунок 3 – Междоузельный комплекс Si-N в кремнии

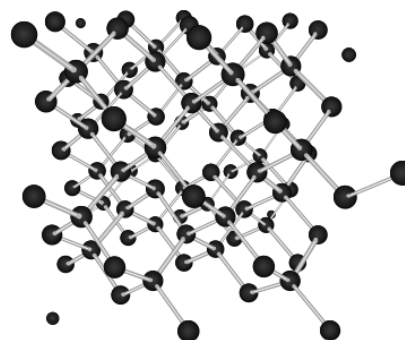


Рисунок 4 – Вакансия в кристалле кремния

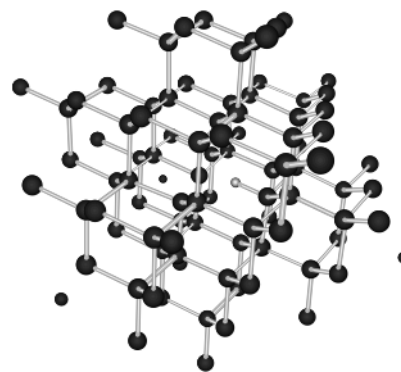


Рисунок 5 – Вакансия с атомом водорода на одной из четырех образовавшихся незаполненных связей в кристалле кремния

ЗАКЛЮЧЕНИЕ

Расчет энергии формирования собственного междоузельного атома кремния в рассматриваемой наноструктуре показывает, что ее величина примерно на одну восьмую часть меньше, если междоузельный атом формируется в непосредственной близости от вакансии, то есть

$$\varepsilon_{ia}^v \approx \frac{7}{8} \varepsilon_{ia}.$$

Удивительной оказалась локализация собственного междоузельного атома кремния в октаэдрическом междоузлии рассматриваемой наноструктуры, вопреки ожидаемым тетраэдрическому или гексаэдрическому междоузлиям, описанным в книге Ланно [2].

Для образования междоузельных комплексов Si-X необходимы значительно большие энергии чем энергия образования собственного междоузельного атома.

Влияние активных атомов на вакансии более разнообразное – атомы и молекулы водорода снижают энергию образования вакансионных комплексов V-H и V-H₂ на 0,19 и 0,13 эВ, соответственно, по сравнению с энергией образования простой вакансии. Другие атомы увеличивают соответствующую энергию: атом азота – на 0,22; кислорода – на 0,40; серы – на 1,58; фтора – на 0,16; хлора – на 1,36 эВ.

Следует заметить, что в работе [3] представлены результаты расчета энергии формирования вакансий при переходе молекулы водорода на две свободные связи вакансии, которая составила величину 1,2 эВ и меньше полученной нами на 0,11 эВ.

Таким образом, наличие в кристалле таких активных атомов как H, F, N, O, S, Cl и других может существенно перераспределять концентрации вакансий

и междоузельных атомов, оказывая тем самым влияние на технологические процессы, в частности на диффузионные операции.

Результаты работы были частично представлены в тезисах докладов конференции по физике лазерных кристаллов ICPLC 2005 [4].

ПЕРЕЧЕНЬ ССЫЛОК

1. Грядун В. І. Вплив атомів водню на енергетичні квантові властивості кремнію // Новые технологии. – 2004. – № 3 (6). – С. 75–77.
2. Ланно М., Бургуэн Ж. Точечные дефекты в полупроводниках. Теория. – Москва: Мир, 1984. – 263 с.
3. Завадинский В. П., Гнеденко А. А., Мисюк А., Бак-Мисюк Я. Влияние давления и водорода на образование вакансий в кремнии // Физика и техника полупроводников. – 2004. – Т. 38, вып. 11. – С. 1281–1284.
4. Griadun V. I. Atoms Effect on Energy of Formation of Point Defects in Silicon // Book of Abstracts and Program 2nd International Conference on Physics of Laser Crystals. – Yalta, Crimea, 25–30 September. – Kharkiv, Lyon, Yalta, 2005. – P. NT16.

Надійшла 10.05.06
Після доробки 26.06.06

Методом механіко-молекулярного моделювання AMBER розглядаються точкові дефекти і комплекси атомів з вакансіями та міжвузловими атомами кремнію. Розраховані енергії утворення точкових дефектів, а також комплексів V-X та IA-X. Показано, що тільки атоми водню у комплексі з вакансією знижують енергію її утворення, а для комплексів Si-X вона значно перевищує енергію утворення власного між вузлового атома.

Dot defects and complexes of atoms with vacancies and own interstitial atoms are considered in the cubic silicon nanostructure by the method of mechanics-molecular modeling AMBER. Are designed energy of generation of dot defects, and also complexes V-X and IA-X. It is shown, that only atoms of hydrogen in a complex with vacancy reduce energy of its making up, and for complexes Si-X it considerably exceeds energy of generation own interstitial atom.

УДК 621.372.852.001.11

В. Н. Кришук, Л. М. Карпуков, М. В. Заскоцкая, А. Ю. Фарафонов

ИССЛЕДОВАНИЕ ДОПУСКОВЫХ ОГРАНИЧЕНИЙ В КОНСТРУКЦИЯХ МИКРОПОЛОСКОВЫХ ФИЛЬТРОВ С УЧЕТОМ ВИДА АППРОКСИМАЦИИ АЧХ

Проведено дослідження допускових обмежень на геометричні розміри конструкцій мікрополоскових фільтрів з урахуванням виду апроксимації амплітудно-частотної характеристики, що забезпечують задані характеристики фільтра. Синтез фільтрів виконується на основі фільтрів-прототипів нижніх частот. Розрахунок допусков проводиться з використанням інтервальних моделей. Урахування впливу технології виготовлення топології

фільтрів здійснюється шляхом введення вагових коефіцієнтів.

В настоящее время все острее становится вопрос сокращения сроков проектирования и производства при условии повышения качества выпускаемой продукции. Неизменный интерес к системам радиолокации, авто-

© Кришук В. Н., Карпуков Л. М., Заскоцкая М. В., Фарафонов А. Ю., 2006

материзованным системам связи, системам телекоммуникации и радионавигации приводит к необходимости создания СВЧ устройств в максимально короткий срок с повышением требований к их характеристикам. Высокая сложность расчета и жесткие требования к точности изготовления наиболее характерны для микрополосковых фильтров. Таким образом, актуальной и важной проблемой является определение допусков при проектировании микрополосковых фильтров и выбор оптимальных конструкций фильтров.

Проблематике проектирования микрополосковых устройств посвящено значительное количество работ. Наибольшую актуальность приобретают задачи точного определения геометрических размеров синтезируемой конструкции и определения допусков на параметры конструкции в зависимости от требований к выходным характеристикам. Методики определения геометрических размеров микрополосковых фильтров представлены в работах [1, 2, 3, 4], имеют разную степень сложности и точности. Сравнение этих методик выполнено в работе [5]. Использование для определения гарантированных допусков интервальных моделей, приводит к определенным ограничениям на вид выходной функции, однако такие модели просты в использовании и обладают довольно высокой точностью [6, 7].

Целью настоящей работы является исследование допусковых ограничений в конструкциях полосно-пропускающих фильтров (ППФ) и фильтрах нижних частот (ФНЧ) с учетом вида аппроксимации АЧХ и сравнение допусков, полученных при введении весовых коэффициентов на каждой итерации и в конце вычислений.

1 МЕТОДИКА СИНТЕЗА ФИЛЬТРОВ НА СВЯЗАННЫХ ЛИНИЯХ

При проектировании ППФ на связанных линиях с использованием метода синтеза на основе фильтров-прототипов нижних частот (НЧ) необходимо определить количество звеньев фильтра n и параметры элементов схемы замещения фильтра g_i . Далее определяются волновые сопротивления четной Z_{oe} и нечетной Z_{oo} волн для каждого звена фильтра. На основе синтезированных волновых сопротивлений рассчитываются геометрические размеры топологии фильтра.

Количество элементов фильтра прототипа НЧ определяется видом и параметрами частотной характеристики проектируемого фильтра. При этом необходимо учитывать, что число каскадно-включенных звеньев фильтра на связанных линиях на 1 больше числа элементов фильтра-прототипа [8].

Параметры g_i для средних элементов фильтра-прототипа определяются по методике, изложенной в [9], а для крайних элементов определяются при помощи соотношений [8]:

$$g_0 = \frac{\pi}{\omega'_n} \cdot \left(\frac{f_n - f_{-n}}{f_n + f_{-n}} \right), \quad (1)$$

$$g_{n+1} = \frac{\pi}{r \cdot \omega'_n} \cdot \left(\frac{f_n - f_{-n}}{f_n + f_{-n}} \right), \quad (2)$$

где ω'_n – приведенная граничная частота фильтра нижних частот; f_n, f_{-n} – граничные частоты, определенные по уровню рабочего затухания, для верхней и нижней границы полосы пропускания фильтра; n – число элементов фильтра-прототипа нижних частот; r – величина нагрузочного сопротивления [9].

При синтезе с помощью фильтров-прототипов НЧ осуществляется переход от одиночного звена фильтра на связанных линиях к четырехполюснику, содержащему два отрезка линии передачи и инвертор импеданса. Далее осуществляется переход к фильтру-прототипу НЧ, основанный на системе уравнений, выражающих зависимость волновых сопротивлений звеньев фильтра на связанных линиях для четного и нечетного видов возбуждения от параметров g_i фильтра-прототипа НЧ [8].

Волновые сопротивления звеньев фильтра на связанных линиях для четного и нечетного видов возбуждения находятся по формулам [9]:

$$Z_{\frac{oe}{oo_i}} = Z_0 \left[1 \pm \frac{Z}{k_{i-1,i}} + \left(\frac{Z}{k_{i-1,i}} \right)^2 \right], \quad (3)$$

где $k_{i-1,i} = \frac{Z}{\frac{\pi}{\omega'_n} \cdot \left(\frac{f_n - f_{-n}}{f_n + f_{-n}} \right) \cdot \sqrt{g_{i-1} \cdot g_i}}$, Z_0 – волновое сопротивление подводящих линий передачи, $i = 1 \dots n + 1$ – число звеньев фильтра на связанных линиях.

Длины звеньев фильтра l_i вычисляются из соотношения [8].

2 ИССЛЕДОВАНИЕ ДОПУСКОВ ЭЛЕМЕНТОВ КОНСТРУКЦИИ

При анализе допусков элементов конструкции фильтра на связанных линиях необходимо определить функцию вносимого затухания фильтра в зависимости от частоты. Звенья фильтра на связанных линиях, разомкнутых на концах, могут быть представлены в виде четырехполюсников, а частотная характеристика фильтра может быть получена перемножением классических матриц передачи отдельных звеньев и описана выражением:

$$a(f) = 20 \cdot \log \frac{1}{2} \cdot \left(A_{11} + \frac{A_{12}}{Z} + A_{21} \cdot Z + A_{22} \right), \quad (4)$$

где $A_{11}, A_{12}, A_{21}, A_{22}$ – коэффициенты результирующей матрицы передачи фильтра.

Непосредственное использование выражения (4) для синтеза допусков связано с большим объемом вычислений. Для синтеза допусков на параметры конструкции фильтра используются интервальные модели, которые учитывают нелинейную зависимость затухания фильтра от параметров конструкции и взаимную компенсацию отклонений, а также обладают достаточной точностью [6].

Для создания упрощенных линейных интервальных моделей использована внутренняя и внешняя интерполяция [10]. Синтез допусков произведен на верхней границе полосы пропускания фильтра.

Исследование допусков выполнено на примере микрополоскового ППФ с частотной характеристикой Баттервортовского и Чебышевского типа и следующими параметрами:

– граничные частоты полосы пропускания фильтра $f_{-п} = 960$ МГц, $f_{п} = 1040$ МГц по уровню вносимого затухания $a_{п} = 3$ дБ;

– граничные частоты полос заграждения $f_{-з} = 925$ МГц, $f_{з} = 1075$ МГц по уровню затухания $a_{з} = 30$ дБ;

– волновое сопротивление подводящих линий передачи $Z_0 = 50$ Ом.

Заданным исходным данным, при аппроксимации Баттерворта, соответствует фильтр из семи звеньев связанных микрополосковых линий ($n=7$) и из пяти звеньев при аппроксимации Чебышева. Фильтры реализованы на подложке с диэлектрической проницаемостью $\epsilon_r = 10$ и толщиной $h = 1$ мм. Толщина металлизации в расчетах не учитывалась. Номинальные размеры ППФ (W – ширина полоска; S – зазор между связанными линиями) приведены в таблице 1.

На рисунке 1 представлены АЧХ фильтров при различных видах аппроксимации. Для построения АЧХ используется пакет программ Microwave Office. Полученные данные электромагнитного анализа свидетель-

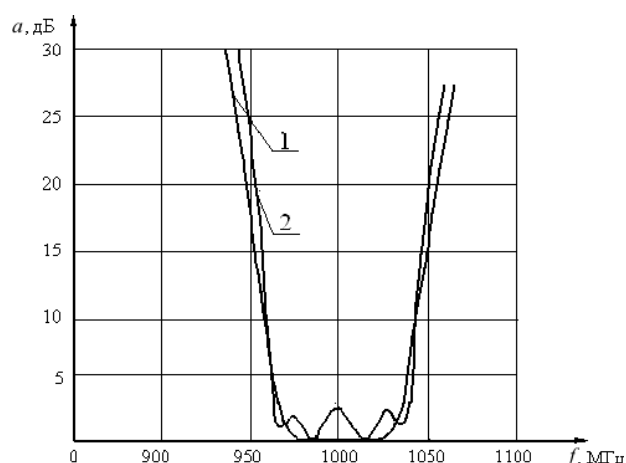


Рисунок 1 – Частотные характеристики фильтра при разных видах аппроксимации АЧХ:

1 – аппроксимация Баттерворта; 2 – аппроксимация Чебышева

ствуют о высокой точности используемой методики проектирования.

Интервальная модель для расчета допусков, в связи с возникающими искажениями вида выходной функции фильтра при назначении отклонений на параметры фильтра с использованием внешней интерполяции построена на основе данных внутренней интерполяции. Начальные отклонения размеров конструкции фильтра и диэлектрической проницаемости материала подложки приняты равными $\delta_i = \pm 1,5$ %. Отклонения функции затухания определялись на частоте 1042 МГц (баттервортовская аппроксимация) и 1041 МГц (чебышевская аппроксимация), которые соответствуют уровню затухания 5 дБ. Для определения коэффициентов интервальной модели и допустимых отклонений параметров конструкции фильтра отклонение функции вносимого затухания было выбрано $\delta_a = \pm 5$ %.

Таблица 1 – Номинальные размеры участков фильтра

Аппроксимация Баттерворта			Аппроксимация Чебышева		
№ звена	Волновые сопротивления, Ом	Размеры, мм	№ звена	Волновые сопротивления, Ом	Размеры, мм
Подвод. линия	$Z_0 = 50$	$W = 0,8817$	Подвод. линия	$Z_0 = 50$	$W = 0,8817$
1,7	$Z_{oe} = 86,7833$	$W = 0,45742$ $S = 0,24456$	1,5	$Z_{oe} = 61,384652$	$W = 0,8041$ $S = 0,7057$
	$Z_{oo} = 37,5025$			$Z_{oo} = 42,269311$	
2,6	$Z_{oe} = 58,4259$	$W = 0,84274$ $S = 0,92016$	2,4	$Z_{oe} = 54,223478$	$W = 0,891$ $S = 1,6227$
	$Z_{oo} = 43,7329$			$Z_{oo} = 46,390135$	
3,5	$Z_{oe} = 54,09206$	$W = 0,89343$ $S = 1,66583$	3	$Z_{oe} = 53,726339$	$W = 0,897$ $S = 1,80226$
	$Z_{oo} = 46,4863$			$Z_{oo} = 46,759087$	
4	$Z_{oe} = 53,4654$	$W = 0,9011$ $S = 1,91914$	–	–	–
	$Z_{oo} = 46,95802$				

Таблица 2 – Отклонения параметров конструкции фильтров

Параметр	Аппроксимация Баттерворта		Аппроксимация Чебышева	
	$\omega_{\text{нижнее}}, \text{ мкм}$	$\omega_{\text{верхнее}}, \text{ мкм}$	$\omega_{\text{нижнее}}, \text{ мкм}$	$\omega_{\text{верхнее}}, \text{ мкм}$
W1	1,706	1,727	2,199	2,174
W2	5,380	4,791	0,470	0,486
W3	0,632	0,503	0,378	0,394
W4	0,445	0,500	0,470	0,486
W5	0,632	0,503	2,199	2,174
W6	5,380	4,791	–	–
W7	1,706	1,727	–	–
S1	3,338	3,527	5,183	5,06
S2	2,688	2,872	1,003	1,039
S3	1,220	1,040	0,911	0,76
S4	0,888	1,130	1,003	1,039
S5	1,220	1,040	5,183	5,06
S6	2,688	2,872	–	–
S7	3,338	3,527	–	–
l1	10,99	11,00	2,277	2,777
l2	3,770	3,752	0,897	0,897
l3	1,337	1,335	0,665	0,666
l4	0,778	0,778	0,897	0,897
l5	1,337	1,335	2,277	2,777
l6	3,770	3,752	–	–
l7	10,99	11,00	–	–
h	0,092	0,135	0,081	0,071
ε , отн. ед.	$2,645 \cdot 10^{-3}$	$2,103 \cdot 10^{-3}$	$1,635 \cdot 10^{-3}$	$2,813 \cdot 10^{-3}$

Таблица 3 – Допусковые ограничения на параметры ППФ ($n = 5$) для чебышевской аппроксимации

Параметр	На каждой итерации		По завершению вычислений	
	$\omega_{\text{нижнее}}, \text{ мкм}$	$\omega_{\text{верхнее}}, \text{ мкм}$	$\omega_{\text{нижнее}}, \text{ мкм}$	$\omega_{\text{верхнее}}, \text{ мкм}$
W1	0,944	0,938	0,950	0,938
W2	0,938	1,015	0,938	0,971
W3	0,938	1,082	0,938	0,98
W4	0,938	1,015	0,938	0,971
W5	0,944	0,938	0,949	0,938
S1	0,938	0,943	0,962	0,938
S2	0,938	0,972	0,938	0,972
S3	0,938	1,054	1,125	0,938
S4	0,938	0,972	0,938	0,972
S5	0,938	0,943	0,962	0,938
l1	0,938	0,938	0,938	0,939
l2	0,938	0,938	0,938	0,939
l3	0,938	0,939	0,938	0,939
l4	0,938	0,938	0,938	0,939
l5	0,938	0,938	0,938	0,939
h	0,008	0,007	0,008	0,007
ε , отн. ед.	$1,635 \cdot 10^{-3}$	$2,813 \cdot 10^{-3}$	$1,635 \cdot 10^{-3}$	$2,813 \cdot 10^{-3}$

Абсолютные величины отклонений параметров конструкции фильтров от размеров были определены с использованием критерия максимального объема (табл. 2).

Анализируя данные таблицы 2, видно, что наиболее критичными к допускам являются центральные звенья фильтров.

Снизить требования к точности изготовления фильтров при сохранении заданных характеристик возможно путем перераспределения поля допуска с целью увеличения допусков на наиболее критичные элементы конструкций.

Перераспределение поля допуска может быть выполнено с учетом этапов технологического процесса изготовления микрополосковых фильтров. Примером может служить операция химического травления фольгированного диэлектрика. На данном этапе абсолютная погрешность изготовления элементов конструкции определяется величиной подтравливания, а не габаритными размерами. Поэтому необходимо распределить поле допуска выходной функции по элементам топологии с целью получения одинаковых допусковых ограничений путем введения весовых коэффициентов.

Алгоритм определения одинаковых допусковых отклонений можно представить так: на основе данных внутренней интерполяции из передаточных коэффициентов модели составляется массив наибольших коэффициентов по каждому параметру, в котором определяют наибольший коэффициент a_{\max} . Далее вычисляется соотношение $\vartheta_i = a_i/a_{\max}$ и рассчитывается сумма передаточных коэффициентов $\sum a_i$, по которым определяются весовые коэффициенты $P_i = \vartheta_i/\sum a_i$.

Алгоритм может быть применен на каждой итерации и по завершению итерационного процесса.

В таблице 3 представлены допусковые ограничения на параметры ППФ с количеством звеньев $n = 5$ при введении весовых коэффициентов на каждой итерации и по завершению расчетов.

Таблица 4 – Исходные данные для синтеза ППФ и ФНЧ

Параметр	ППФ	ФНЧ
Z – волновое сопротивление линии, Ом	50	50
ϵ – диэлектрическая проницаемость подложки	10	9,6
h – толщина диэлектрической подложки, мм	0,635	1
$a_{\text{п}}$ – затухание в полосе пропускания, дБ	0,1	3
$a_{\text{з}}$ – затухание в полосе заграждения, дБ	10	30
$f_{\text{п}}$ – граничная частота полосы пропускания, МГц	1700	–
$f_{\text{п}}$ – граничная частота полосы пропускания, МГц	2300	–
$f_{\text{з}}$ – граничная частота полосы заграждения, МГц	800	–
$f_{\text{з}}$ – граничная частота полосы заграждения, МГц	3844,7	1150
$f_{\text{с}}$ – частота среза, МГц	–	700
n – количество звеньев	3	5

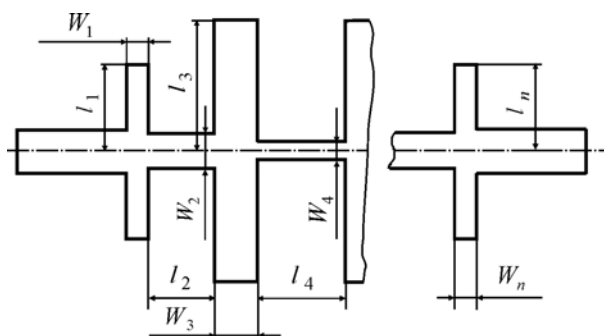


Рисунок 2 – Токонесущая часть микрополоскового ФНЧ

Из таблицы 3 видно, что введение весовых коэффициентов на каждой итерации дает более жесткие допуски на параметры фильтра, чем при введении весовых коэффициентов по завершению расчетов. Результаты отличаются не более, чем на 9,47 %.

Аналогично проведено сравнение допусковых ограничений на параметры ППФ и ФНЧ.

В таблице 4 представлены исходные данные для синтеза ППФ и ФНЧ при чебышевской аппроксимации.

В таблице 5 представлены допусковые ограничения на параметры ППФ при введении весовых коэффициентов на каждой итерации и по завершению расчетов.

Как видно из таблицы 5, результаты отличаются в пределах 3,74 %, а при аппроксимации Баттерворта различие результатов при введении весовых коэффициентов на каждой итерации и по завершению вычислений будет составлять около 3 %.

Топология и размеры токонесущей части микрополоскового ФНЧ рассчитаны с использованием методик [11, 12] (рис. 2).

В таблице 6 представлены допусковые ограничения на параметры ФНЧ при введении весовых коэффициентов на каждой итерации и по завершению расчетов.

Таблица 5 – Допусковые ограничения на параметры ППФ для чебышевской аппроксимации

Параметр	На каждой итерации		По завершению вычислений	
	$\omega_{\text{нижнее}}, \text{ МКМ}$	$\omega_{\text{верхнее}}, \text{ МКМ}$	$\omega_{\text{нижнее}}, \text{ МКМ}$	$\omega_{\text{верхнее}}, \text{ МКМ}$
W1	0,817	0,777	0,847	0,773
W2	0,777	0,785	0,773	0,780
W3	0,817	0,777	0,847	0,773
S1	0,802	0,777	0,799	0,773
S2	0,778	0,777	0,773	0,787
S3	0,802	0,777	0,799	0,773
l1	0,777	0,777	0,779	0,773
l2	0,776	0,777	0,775	0,773
l3	0,777	0,777	0,779	0,773
h	0,605	0,589	0,605	0,589
ϵ , отн. ед.	$2,5197 \cdot 10^{-3}$	$2,5845 \cdot 10^{-3}$	$2,5197 \cdot 10^{-3}$	$2,5845 \cdot 10^{-3}$

Таблица 6 – Допусковые ограничения на параметры ФНЧ для чебышевской аппроксимации

Параметр	На каждой итерации		По завершению вычислений	
	$\omega_{\text{нижнее}}, \text{ МКМ}$	$\omega_{\text{верхнее}}, \text{ МКМ}$	$\omega_{\text{нижнее}}, \text{ МКМ}$	$\omega_{\text{верхнее}}, \text{ МКМ}$
W1	7,293	7,309	7,453	7,490
W2	7,293	7,802	7,453	7,543
W3	7,293	7,303	7,453	7,464
W4	7,293	7,802	7,453	7,543
W5	7,293	7,309	7,453	7,490
l1	7,299	7,293	7,498	7,453
l2	7,293	7,293	7,453	7,453
l3	7,301	7,293	7,484	7,453
l4	7,293	7,293	7,453	7,453
l5	7,299	7,293	7,498	7,453
h	3,427	3,456	3,431	3,452
ϵ , отн. ед.	0,011	0,011	0,011	0,011

В данном случае различие результатов при аппроксимации Чебышева составляет 3,33 %, а при аппроксимации Баттерворта – 6,13 %. При этом тенденция получения более жестких допусков при введении весовых коэффициентов на каждой итерации сохраняется для разных видов фильтров.

ВЫВОДЫ

Проведено исследование допусковых ограничений в конструкциях микрополосковых фильтров при различных видах аппроксимации АЧХ. Показано, что использование аппроксимации Баттерворта позволяет расширить допуски на параметры конструкции фильтра, хотя и ведет к увеличению габаритных размеров топологии. Анализ особенностей возникновения погрешностей на этапах технологии производства позволяет расширить гарантированные допуски на параметры фильтров. Выявлена необходимость введения весо-

вых коэффициентов на каждой итерации синтеза допусков в связи с получением более жестких гарантированных допусков.

Направление дальнейших исследований связано с исследованием допусков СВЧ устройств с учетом неоднородностей микрополосковых линий.

ПЕРЕЧЕНЬ ССЫЛОК

1. Аналоговые полупроводниковые интегральные схемы СВЧ / В. Н. Данилин, А. И. Кушниренко, Г. В. Петров. – М.: Радио и связь, 1985. – 192 с.
2. Полосковые платы и узлы. Проектирование и изготовление / Е. П. Котов, В. Д. Капун, А. А. Тер-Маркарян, В. П. Лисицын, Ю. И. Фаянс / Под ред. Е. П. Котова и В. Д. Капуна. – М.: «Сов. радио», 1979. – 248 с.
3. Проектирование интегральных устройств СВЧ: Справочник / Ю. Г. Ефремов, В. В. Конин, Б. Д. Солганик и др. – К.: Техника, 1990. – 159 с.
4. Akhtarzad S., Rowbotham T., Johns P. The design of coupled microstrip lines // IEEE Transaction on microwave theory and techniques – 1975. – Vol. MTT-23, No. 6, June. – P. 486–492.

5. *Krishuk V., Farafonov A., Romanenko S.* Analysis of projection methods of microstrip coupled lines filters // Proceedings of the International Conference "Modern problems of radio engineering, telecommunications and computer science". – Lviv-Slavsko (Ukraine). – 2004. – P. 583–586.
6. *Шило Г. М.* Формування інтервальних моделей для обчислення допусків // *Радіоелектроніка. Інформатика. Управління.* – 2002. – № 1. – С. 90–95.
7. *Krischuk V., Shilo G., Gaponenko N.* Optimization of ISLAE solutions in the problems of assigning tolerances for parameters of electronic devices // Proceedings of the International Conference "Modern problems of radio engineering, telecommunications and computer science". – Lviv (Ukraine). – 2002. – P.114–115.
8. *Малорацкий Л. Г., Явич Л. Р.* Проектирование и расчет СВЧ элементов на полосковых линиях. – М.: Сов. радио, 1972. – 232 с.
9. *Фельдштейн А. Л., Явич Л. Р., Смирнов В. П.* Справочник по элементам волноводной техники. – М.: Сов. радио, 1967. – 651 с.
10. *Krischuk V., Shilo G., Gaponenko N.* Optimization of ISLAE solutions in the problems of assigning tolerances for parameters of electronic devices // Proceedings of the International Conference "Modern problems of radio engineering, telecommunications and computer science". – Lviv (Ukraine). – 2002. – P.114–115.
11. *Конструирование и расчет полосковых устройств /* Под ред. И. С. Ковалева. – М.: Сов. радио, 1974. – 295 с.
12. *Крищук В., Фарафонов О., Шило Г., Гапоненко М.* Оптимізація допусків мікросмужкових фільтрів // *Вісник*

Національного університету «Львівська політехніка» «Комп'ютерні системи проектування. Теорія і практика». – 2003. – № 470. – С. 173–178.

Надійшла 10.07.06
Після доробки 21.08.06

Проведено дослідження допускових обмежень на геометричні розміри конструкції мікросмужкових фільтрів з урахуванням виду апроксимації амплітудно-частотної характеристики, які забезпечують задані характеристики фільтра. Синтез фільтрів виконується на основі фільтрів-прототипів нижніх частот. Розрахунок допусків виконується з використанням інтервальних моделей. Врахування впливу технології виготовлення топології фільтрів здійснюється шляхом введення вагових коефіцієнтів.

Probing tolerance limitations on the geometrical sizes constructions of the microstrip filters is carried out. Also, view of approximation amplitude-frequency characteristics, providing the necessary characteristics of the filter take into account. The synthesis of filters is fulfilled on the basis filters-prototypes of the lower frequencies. Calculation of tolerances is made with usage of interval models. The discount agency of manufacture technology of the filters topology is carried out by introduction of weight coefficient.

УДК 621.396.6

В. М. Крищук, Г. М. Шило, Б. А. Артюшенко

ГЕНЕТИЧНИЙ АЛГОРИТМ ПРИЗНАЧЕННЯ ДОПУСКІВ НА РАДІОЕЛЕМЕНТИ З ІНТЕРВАЛЬНИМ ОЦІНЮВАННЯМ

Запропоновано метод підвищення ефективності генетичного алгоритму для призначення інтервальних допусків на радіоелементи. Окрім відомих способів оцінки працездатності рішень використовується інтервальне розширення вихідних функцій та пошук межових вершин методом впливів. Алгоритм адаптовано до роботи з обмеженнями, що виникають при проектуванні радіоапаратури.

ВСТУП

Синтез та аналіз допусків є невід'ємною частиною проектування якісного і водночас доступного за ціною виробу. В умовах скорочення часу на проектування, тестування та налагодження масового виробництва радіоапаратури постає задача автоматизації цих процедур. Здебільшого синтез та аналіз допусків базується на використанні детермінованих (переважно інтервальних [1]) ітераційних методів з лінійними моделями. Основний недолік таких методів – це неможливість роботи з вихідними функціями, які містять екстремуми в інтервалах їх допуску, хоча останнім часом і з'явилися роботи присвячені розв'язку цієї проблеми [2].

Використання інтервальних методів ускладнено також при існуванні великої кількості обмежень.

На противагу детермінованим методам набувають поширення стохастичні методи: Монте-Карло, моделюючий віджиг, генетичний алгоритм (ГА). Ці методи не потребують від вихідної функції таких властивостей як неперервність, диференційованість чи монотонність. Але слід зазначити, що метод Монте-Карло вимагає великої кількості обчислень, а використання моделюючого віджигу не завжди можливо [3]. Використання ГА допускового аналізу базуються в тому числі і на інтервальних методах [4, 5]. Існують також приклади застосування ГА для синтезу допусків [4, 6, 7]. Синтез допусків за допомогою ГА пов'язаний з визначенням працездатності знайденої області з точки зору обмежень на вихідні функції. Для цього в існуючих ГА синтезу допусків використовуються: метод Монте-Карло, аналіз вершин [7], ANOVA (аналіз середніх значень і варіацій) [4] та нечітка логіка [6]. Але всі ці методи або не надають гарантованих відповідей, або потребують великої кількості обчислень.

© Крищук В. М., Шило Г. М., Артюшенко Б. А., 2006

Метою роботи є розробка засобів щодо підвищення ефективності ГА синтезу інтервальних допусків на параметри радіоапаратури з допомогою нових підходів до оцінювання працездатності сформованих областей. Зокрема, може бути використано інтервальне розширення вихідних функцій та пошук межових вершин методом впливів. Для вирішення поставленої задачі необхідно:

- розробити нові методи оцінювання працездатності області, що базується на використанні інтервального розширення вихідних функцій;
- адаптувати ГА для роботи з виникаючими при проектуванні радіоелектронної апаратури обмеженнями;
- провести аналіз впливу параметрів ГА на ефективність його застосування.

ОЦІНЮВАННЯ ПРАЦЕЗДАТНОСТІ ОБЛАСТІ

Для адаптації ГА в задачах призначення допусків на параметри радіоелементів слід визначити найсуттєвіші обмеження. При проектуванні радіоелектронних засобів найбільш поширеними є наступні обмеження:

- допускова область повинна вмщувати початкові рішення:

$$x_0 \in \mathbf{X}, \quad (1)$$

де x_0 – номінальне рішення, $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ – інтервальна допускова область (брус), $x_i = [x_i, \bar{x}_i]$ – інтервал значень параметру i , n – кількість параметрів;

- обмеження на параметри елементів:

$$x_i \in sp_i = [sp_i, \bar{sp}_i], \quad (2)$$

де sp_i – інтервал, що обмежує можливі значення параметру i ;

- обмеження на допуск (виробничі обмеження):

$$wid(x_i)/2 \in [mint_i, maxt_i]; \quad (3)$$

- обмеження на вихідні функції (обмеження працездатності):

$$\underline{F}_j \leq f_j(x_1 \dots x_n) \leq \bar{F}_j, \quad \forall x_i \in x_i, \quad j = \overline{1, m}, \quad (4)$$

де $f_j(x_1 \dots x_n)$ – вихідні функції, F_j та \bar{F}_j – нижні та верхні обмеження на вихідні функції, m – кількість вихідних функцій.

Вираз (4) еквівалентно співвідношенню:

$$\mathbf{X} \subset \mathbf{W}, \quad (5)$$

де \mathbf{W} – область працездатності.

Якщо виконуються всі обмеження, то брус вважається припустимим:

$$\mathbf{X} \subset \mathbf{ROA}, \quad (6)$$

де \mathbf{ROA} – область припустимих значень параметрів.

Крім цього, більшість радіоелементів випускається з регламентованими допусками ($\pm 1\%$, $\pm 2\%$, $\pm 5\%$, $\pm 10\%$, $\pm 20\%$) та номінальним рядом параметрів.

Приклад таких обмежень дається на рис. 1.

Для перевірки виразів (1)–(3) знадобиться до $3 \cdot n$ обчислень. Найбільші труднощі виникають з перевіркою виразу (4). Кількість обчислень при перевірці працездатності залежить від монотонності вихідних функцій, опуклості та неперервності області \mathbf{W} та способу оцінювання. Для методу Монте-Карло необхідна кількість обчислень пропорційна об'єму області, при аналізі вершин проводиться до 2^n обчислень, а для методу ANOVA – 3^n . Скоротити кількість обчислень можливо при використанні інтервальних розширень ви-

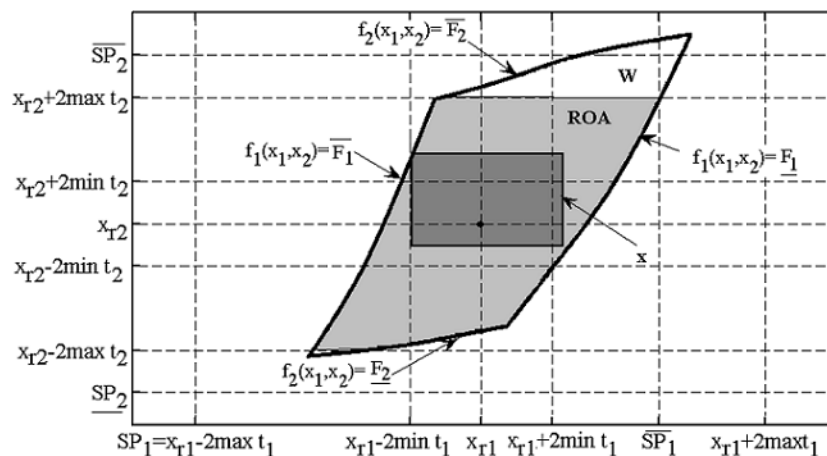


Рисунок 1 – Приклад обмежень та допускової області

хідних функцій та пошуку межових значень методом впливів.

Знаходження адекватного інтервального розширення може бути досить складною задачею, особливо у випадку немонотонних функцій. Процедура обчислень значно скоротиться при використанні природного інтервального розширення функцій, які утворюються заміною усіх змінних та арифметичних операцій на інтервальні. В цьому випадку перевірка працездатності може проводитись при разовому обчисленні інтервального розширення вихідної функції. Працездатності бруса відповідає умова:

$$f(\mathbf{X}) \subset [E, \bar{F}], (\mathbf{X} \subset \mathbf{W}), \quad (7)$$

де $f(\mathbf{X})$ – природне інтервальне розширення функції $f(x_1 \dots x_n)$, $x_i \in \mathbf{x}_i$.

Область є гарантовано непрацездатною, якщо виконується:

$$f(\mathbf{X}) \cap [E, \bar{F}] = 0, (\mathbf{X} \not\subset \mathbf{W}). \quad (8)$$

В інших випадках, за умови монотонності вихідних функцій, необхідно обчислити найменше та найбільше значення межової функції на вершинах бруса. Для цього може бути використано пошук межових значень методом впливів. Цей метод базується на зміні кожного з параметрів відносно номінальної точки і з'ясування, які зміни сприяють зменшенню та збільшенню вихідної функції. Координати вершин з мінімальним (максимальним) значенням вихідної функції утворюються використанням усіх змін, що сприяють зменшенню (збільшенню) вихідної функції. Таким чином формується адекватне інтервальне розширення вихідної функції за $2(n+1)$ обчислень. Ця процедура також дозволяє визначити монотонність вихідної функції на брус \mathbf{X} та необхідність переходу до перевірки працездатності бруса іншими методами.

ГЕНЕТИЧНИЙ АЛГОРИТМ

В основу роботи ГА покладено ідею запозичену з еволюційного процесу в біології [8]: найприспособаніші істоти виживають та, схрещуючись між собою, дають нове, краще, покоління. Мутація, схрещення та відбір дозволяють видам пристосовуватись до різноманітних умов існування. Ці ж процеси дають можливість знаходити оптимальні рішення і для різноманітних прикладних задач.

Елементи популяції (істоти або генотипи) в ГА є бітовими рядками G довжиною l . Отже, постає задача їх відображення в інтервальну область. Для цього найбільшого поширення набули співвідношення:

$$\begin{aligned} \bar{x}_i &= \min\{\text{mid}x_i + t_i, \overline{sp}_i\}, \\ \underline{x}_i &= \max\{\text{mid}x_i - t_i, \underline{sp}_i\}, \end{aligned} \quad (9)$$

де

$$\text{mid}x_i = \sum_{j=1}^l 2^{j-1} G_{2id+j} \cdot \frac{(\overline{sp}_i - \underline{sp}_i)}{2} + \underline{sp}_i, \quad (10)$$

$$t_i = \sum_{j=1}^l 2^{j-1} G_{(2i+1)d+j} \cdot \max t_i. \quad (11)$$

При використанні подібних співвідношень інтервальна область задовольняє обмеженню (2) та частково (3). Враховуючи порівняну простоту виробничих обмежень, їх можна задовольнити ще при відображенні генотипу, для цього необхідно (11) замінити на співвідношення:

$$\begin{aligned} t_i &= \max \left\{ \min t_i, |\text{mid}x_i - x_{0i}| \right\} + \sum_{j=1}^l 2^{j-1} \times \\ &\times G_{(2i+1)d+j} \cdot \min \left\{ \overline{sp}_i - \text{mid}x_i, \text{mid}x_i - \underline{sp}_i \right\}. \end{aligned} \quad (12)$$

Таким чином будуть автоматично задовільнені всі існуючі обмеження крім обмежень працездатності. Тим самим скорочується час роботи ГА. Але при цьому слід враховувати, що подібне відображення прийнятне лише коли виконується:

$$\begin{aligned} \text{wid}(\mathbf{sp}_i) &\gg 2 \min t_i, \\ \max \left\{ \overline{roa}_i - x_{0i}, x_{0i} - \underline{roa}_i \right\} &\gg \min t_i, \end{aligned} \quad (13)$$

де $\mathbf{roa}_i = [\underline{roa}_i, \overline{roa}_i]$ – інтервал припустимих значень параметру i .

Мірою пристосованості генотипу в ГА є значення функції фітнесу цього генотипу. Для задач знаходження інтервальної допускової області, зазвичай використовують:

$$\text{fitness}(G) = \begin{cases} \text{performance}(\mathbf{X}_G) & \text{if } (\mathbf{X}_G \subset \mathbf{ROA}); \\ 0 & \text{otherwise,} \end{cases} \quad (14)$$

де $\text{performance}(\mathbf{X}_G)$ – функція, що її потрібно максимізувати (здебільшого об'єм допускової області, показник ціни виробу, тощо), \mathbf{X}_G – допускова область, що є відображенням (9)–(12) генотипу G .

Але при використанні подібної функції фітнесу всі неприйнятні рішення вважаються однаково пристосованими. Враховуючи існування номінальних рішень, подібна функція фітнесу має право на існування, якщо не виконується:

$$SP \gg W. \quad (15)$$

В інших випадках, при відсутності на будь-якому поколінні припустимих рішень, ГА фактично зводиться до методу Монте-Карло. Кількість обчислень зменшується при використанні ступеню неприйнятності. Для врахування ступеню неприйнятності достатньо використати інтервальні розширення вихідних функцій:

$$\text{fitness}(G) = \begin{cases} \text{performance}(\mathbf{X}_G) & \text{if } (\mathbf{X}_G \subset \mathbf{ROA}); \\ -\text{dist}(\mathbf{f}(\mathbf{X}_G), [E, \bar{F}]) & \text{if } (\mathbf{X}_G \notin \mathbf{W}); \\ 0 & \text{otherwise,} \end{cases} \quad (16)$$

де $\text{dist}(\mathbf{f}(\mathbf{X}_G), [E, \bar{F}])$ – метрика Хаусдорфа:

$$\text{dist}(\mathbf{f}(\mathbf{X}_G), [E, \bar{F}]) = \max \left\{ \left| \bar{F} - \overline{\mathbf{f}(\mathbf{X}_G)} \right|, \left| E - \underline{\mathbf{f}(\mathbf{X}_G)} \right| \right\}. \quad (17)$$

Тоді кожен крок ГА набуває вигляду:

```

10 for i:=1 to popnum
20 begin
30 select g1, g2 from PrevG by PrevFitnessG;
40 G[i]:=cross (g1, g2);
50 mutate (G[i]);
60 XG:=map(G);
70 FitnessG [i]:=fitness(G)
80 end;
90 PrevG=G;
100 PrevFitnessG=FitnessG,
    
```

де popnum – величина популяції; select – селекція, випадковий вибір генотипів $g1, g2$ відповідно до їх пристосованості; PrevG, G – масиви генотипів попереднього покоління та покоління, що формується; $\text{FitnessG}, \text{PrevfitnessG}$ – масиви значень функції фітнесу генотипів попереднього покоління та покоління, що формується; cross – кросингвер (перехрещення хромосом генотипів, утворення нового генотипу); mutate – мутація (внесення випадкових змін в генотип); map – відображення генотипу в брус за виразами (9), (10), (12), $\text{fitness}(G)$ – функція фітнесу (пристосованості) за виразом (14).

ВИКОРИСТАННЯ ГЕНЕТИЧНОГО АЛГОРИТМУ

Особливості використання ГА розглянемо для задачі призначення допусків на параметри коливального

контуру [9]. Його коефіцієнт передачі визначається за допомогою виразу:

$$H = \left((1 + R_1/R_2)^2 + R_1^2(wC - 1/(wL))^2 \right)^{-\frac{1}{2}}, \quad (18)$$

де $x_1 = R_1, x_2 = R_2, x_3 = C, x_4 = L$, при частоті $w = 1,05$.

Необхідно задовольнити наступні обмеження: ширина допуску на всі параметри елементів не нижче $\text{mint}_i = \pm 0,5\%$; $i = \overline{1, 4}$, нормовані номінальні значення $x_0 = (1,05; 1,05; 1,05; 1,05)$, $sp_i = x_{0i} \pm 20\%$, $H = [0,454; 0,561]$. Коефіцієнти вартості елементів TC_i задано:

$$TC_i = \begin{cases} 8 & \text{if } 0,5\% < \text{wid}x_i/2 < 2\%; \\ 4 & \text{if } 2\% < \text{wid}x_i/2 < 5\%; \\ 2 & \text{if } 5\% < \text{wid}x_i/2 < 10\%; \\ 1 & \text{if } 10\% < \text{wid}x_i/2 < 20\%. \end{cases} \quad (19)$$

Провести параметричний синтез та синтез регламентованих допусків. При виборі цільової функції слід виходити з того, що метою допускового синтезу є зменшення вартості виробу та підвищення його надійності. Отже, слід використовувати таку цільову функцію, що комбінує цінові показники виробу та об'єм допускової області:

$$\text{performance}(\mathbf{X}) = \prod_{i=1}^4 \frac{\text{wid}x_i}{\text{wid}sp_i} + 0,25 \cdot 10^{-3} \times \sum_{i=1}^4 \begin{cases} 8/TC_i & \text{if } (\text{wid}x_i/2 > 2\%) \\ 0 & \text{otherwise} \end{cases}. \quad (20)$$

На рис. 2 приведено хід роботи ГА з вірогідністю мутації 100%, щільністю мутації 5%, величині популяції 100 (ці значення параметрів будуть надалі використовуватись, якщо прямо не сказано зворотне), фітнесом за виразом (15), по осі абсцис відкладено номер покоління, по осі ординат відкладено: максимальне значення фітнесу maxfitness (рис. 2, а), середнє значення фітнесу midfitness (рис. 2, б), та відносна кількість припустимих рішень $N_{ROA}(\%)$ (рис. 2, в). Як видно з рисунків, для визначення кращого рішення знадобилось приблизно 100 поколінь.

Процес визначення рішення завершується на 100-му поколінні, результати обчислень приведено в табл. 1. Рис. 2 і табл. 1 відповідають оптимальному режиму роботи ГА, який було визначено після дослідження впливу функції фітнесу, відображення генотипу та параметрів ГА.

В табл. 2 порівнюється використання функцій фітнесу за виразами (14) та (16), відображення генотипу за допомогою виразів (9), (10), (12), та (9)–(11) на 100-му поколінні. З таблиці видно, що доцільно ви-

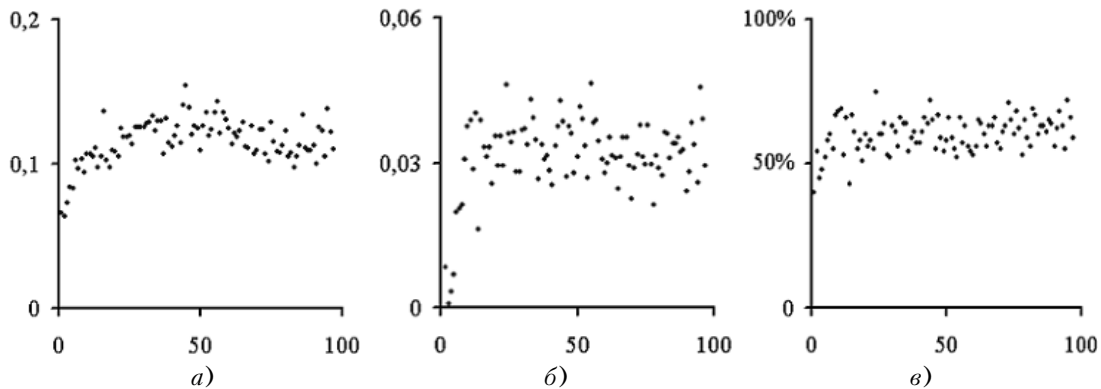


Рисунок 2 – Хід роботи алгоритму:

a – *maxfitness*; *б* – *midfitness*; *в* – N_{ROA}

користувати функцію фітнесу, яка враховує ступінь неприйнятності брусу, а виробничі обмеження слід враховувати при відображенні генотипу.

Таблиця 1 – Найкраще рішення

Параметр	R_1	R_2	C	L
$midx_i$	1,0628	1,1550	1,0429	1,0399
$wid x_{i/2}$	$\pm 7,68\%$	$\pm 9,09\%$	$\pm 19,46\%$	$\pm 18,89\%$

Таблиця 2 – Порівняння використання різних функцій фітнесу та відображення генотипу

fitness(G)	map(G)	N_{ROA}	<i>midfitness</i>	<i>max fitness</i>
Вираз (16)	Вирази (9), (10), (11)	55 %...60 %	0,02...0,03	0,14
Вираз (16)	Вирази (9), (10), (12)	55 %...70 %	0,03...0,04	0,18
Вираз (14)	Вирази (9), (10), (12)	30 %...40 %	-0,40...-0,22	0,16

В табл. 3 надається залежність ефективності роботи ГА (N_{ROA} , *midfitness*, *maxfitness*) від щільності мутації на 100-му поколінні. Як видно з таблиці, при щільності мутації 0,005–0,025 % ГА дає найкращі показники.

Таблиця 3 – Ефективність роботи ГА при різних щільностях мутації

Щільність мутації	N_{ROA}	<i>midfitness</i>	<i>maxfitness</i>
0,001	85 %...90 %	0,10...0,12	0,18
0,0025	55 %...75 %	0,04...0,05	0,16
0,05	55 %...70 %	0,03...0,04	0,18
0,1	30 %...50 %	0,01...0,02	0,15
0,2	20 %...35 %	-0,02...-0,01	0,12

Залежність ефективності роботи ГА (N_{ROA} , *midfitness*, *maxfitness*) від обсягу популяції на 100-му поколінні при обсязі популяції 100 (при інших – відповідно до їх обсягу) надано в табл. 4. Як видно з таблиці, при величині популяції ~100 ГА дає найкращі показники. Ці значення параметрів приблизно співпадають з обраними авторами робіт [5, 6, 7, 8].

Таблиця 4 – Ефективність роботи ГА при різних об'ємах популяції

popnum	N_{ROA}	<i>midfitness</i>	<i>maxfitness</i>
50	40 %...55 %	0,02...0,04	0,14
100	55 %...70 %	0,03...0,04	0,18
150	60 %...70 %	0,03...0,04	0,17
200	60 %...70 %	0,04...0,04	0,15
300	65 %...75 %	0,02...0,03	0,11

Проведені обчислення показали, що збереження з покоління в покоління перших найкращих елементів без змін не надає суттєвих переваг. Використання інтервального розширення вихідної функції дозволило скоротити час на аналіз працездатності приблизно на 30 % порівняно з аналізом вершин.

ВИСНОВКИ

В запропонованому генетичному алгоритмі використовується оцінювання працездатності за допомогою інтервального розширення вихідних функцій. Найменшу кількість обчислень забезпечує природне інтервальне розширення. При недотриманні умов працездатності використовується адекватне інтервальне розширення, що формується за допомогою методу впливів. Проведене тестування алгоритму показало, що воно забезпечує скорочення часу обчислень не менш як на 30 % при порівнянні з аналізом вершин. Збільшення кількості еле-

ментів приводить до подальшого відносного зменшення часу обчислень.

Для підвищення ефективності роботи генетичного алгоритму доцільно використовувати такі функції фітнесу, що враховують ступінь відхилення вихідних функцій від припустимих значень. Кількість обчислень значно скорочується при використанні таких відображень генотипу, що задовольняють виробничим обмеженням.

ПЕРЕЛІК ПОСИЛАНЬ

1. Шило Г. М., Воропай О. Ю., Гапоненко М. П. Интервальные методы назначения эксплуатационных допусков // Радиоэлектроника. Информатика. Управление. – 2003. – № 2. – С. 78–82.
2. Гапоненко Н. П., Воропай А. Ю. Анализ допусков для немонотонных выходных функций // Радиоэлектроника и информатика. – 2005. – № 2(31). – С. 29–32.
3. Минаков И. А. Сравнительный анализ некоторых методов случайного поиска и оптимизации // Известия Самарского центра Российской академии наук. – 1999. – № 2. – С. 286–293.
4. Foroueaghi B. Worst Case Tolerance Design and Quality Assurance via Genetic Algorithm // Journal of Optimization Theory and Applications – 2002. – Vol. 113, No. 2. – Pp. 251–268.
5. Femia, N., and Spagnuolo, G., Genetic Optimization of Interval Arithmetic-Based Worst-Case Circuit Tolerance Analysis // IEEE Transactions on Circuits and Systems 1 Fundamental Theory and Application. – 1999. – Vol. 113, Issue 12. – Pp. 1441–1456.

6. Ji S., Li X., Ma Y., Cai H. Optimal Tolerance Assignment Based on Fuzzy Comprehensive Evaluation and Genetic Algorithm // Int J Adv Manuf Technol. – 2000. – No. 16. – Pp. 461–468.
7. Spagnuolo G., Vitelli M. Worst-case tolerance design by genetic algorithm // Proceedings of the 2002 IEEE International Symposium on Industrial Electronics. – 2002. – Pp. 1178–1183.
8. Holland J. H. Adaptation in natural and artificial systems: an introductory analysis with application to biology, control and artificial intelligence. – University of Michigan Press. – 1975. – 175 p.
9. Воропай О. Ю. Призначення інтервальних допусків для немонотонних вхідних впливів // Радіоелектроніка. Інформатика. Управління. – 2006. – № 1. – С. 19–23.

Надійшла 12.06.06
Після доробки 12.08.06

Предложено метод повышения эффективности генетического алгоритма для назначения интервальных допусков на радиоэлементы. Помимо известных методов оценки работопригодности полученных решений применяется интервальное расширение выходных функций и поиск граничных вершин методом влияний. Алгоритм адаптирован для работы с возникающими при проектировании радиоаппаратуры ограничениями.

A new method to increase effectivity of genetic algorithm for electronics' interval tolerance assignment is presented. Target function interval natural inclusion and method of influence are used to test found domains. The algorithm is adopted to work out constraints of electronics' design.

УДК 681.324.50

В. Г. Лихограй, А. Е. Стрельницкий, А. А. Стрельницкий, А. И. Цопа,
В. М. Шокало

ЭФФЕКТИВНОСТЬ РАБОТЫ СИСТЕМ АБОНЕНТСКОГО РАДИОДОСТУПА САРД ПРИ НАЛИЧИИ ПОМЕХИ. МОДЕЛЬ И АНАЛИЗ

В статье предложена модель для расчета эффективности работы САРД при наличии непрерывной и импульсной помехи, приведен ее математический анализ, разработана методика построения кривых постоянной вероятности битовой ошибки, которые ограничивают области гарантированной эффективности работы САРД.

АКТУАЛЬНОСТЬ И ЦЕЛЬ РАБОТЫ

Сети с радиодоступом стали одним из перспективных направлений развития телекоммуникаций. Разнообразием таких сетей являются локальные сети абонентского радиодоступа (САРД) с протоколом IEEE 802.11b. Они нашли широкое применение для связи на расстояниях до 90 м и возможности их описаны во многих работах. Однако до сих пор остается открытым вопрос о реальной эффективности локальных САРД

с учетом влияния непрерывных и импульсных помех, а также неравномерности характеристик направленности клиентских адаптеров (КА) и точек доступа (ТД).

Целью работы является получение аналитических выражений вероятности битовой ошибки с учетом влияния непрерывных и импульсных помех, а также разработка методики построения кривых постоянной вероятности битовой ошибки.

ОСНОВНАЯ ЧАСТЬ

Проведем анализ влияния непрерывных и импульсных помех на широкополосную САРД с прямым расширением спектра на основе M -арной фазовой манипуляции, что широко применяется в стандартах семейства IEEE 802.11.

Пусть на широкополосную САРД воздействует непрерывный мешающий сигнал со средней мощностью P_J , который полностью покрывает ее рабочую полосу B и аналогично стационарному гауссовскому шуму имеет нулевое среднее и равномерную спектральную плотность мощности $J_0 = P_J/B$. Тогда соотношение сигнал/(шум+помеха) на входе приемника САРД определим следующим образом:

$$\frac{E_b}{N_0 + J_0} = \frac{E_b}{N_0} \frac{1}{1 + \frac{E_b P_J}{N_0 P_s G_p}}, \quad (1)$$

где $E_b = P_s T_b = P_s / C_b$ – энергия бита, P_s – мощность бита, $C_b = 1/T_b$ – скорость передачи бита, N_0 и J_0 – спектральные плотности мощности теплового шума и помехи соответственно, $G_p = BT_b$ – коэффициент расширения спектра.

Вероятность битовой ошибки P_b при передаче равновероятных сигналов M -арной фазовой модуляции с когерентной демодуляцией равна [1]:

$$P_B = 2Q \left[\sqrt{\frac{2E_b}{(N_0 + J_0)}} k \cdot \sin \frac{\pi}{2^k} \right] = 2Q \left[\sqrt{2 \cdot k \cdot \frac{\frac{E_b}{N_0}}{1 + \left(\frac{E_b}{N_0}\right) \cdot \left(\frac{P_J}{P_s}\right) \cdot \left(\frac{1}{G_p}\right)}} \cdot \sin \frac{\pi}{2^k} \right], \quad (2)$$

где $Q(x) \approx \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left[-\frac{u^2}{2}\right] du$ – функция ошибок; $k = \log_2 M$ – разрядность множества передаваемых символов размерности M .

Пусть распространение радиоволн от требуемой САРД и постановщика помех осуществляется по разным направлениям и описывается моделью, изложенной в [2, 3]. Тогда отношение мощностей помехи и полезного сигнала на входе приемника САРД можно представить в следующем виде:

$$\frac{P_J}{P_s} = \frac{P_{TJ} G_{TJ} G_{RJ} (r_s)^n}{P_{Ts} G_{Ts} G_{Rs} (r_J)^n} = \frac{P_{TJ} G_J (r_s)^n}{P_{Ts} G_s (r_J)^n}, \quad (3)$$

где P_{Ts} – мощность бита переданного сигнала; P_{TJ} – средняя мощность передатчика мешающего сигнала; G_{Ts} , G_{TJ} – максимальный коэффициент усиления (КУ) передающих антенн соответственно САРД и постановщика помех; G_{Rs} , G_{RJ} – КУ приемной антенны САРД соответственно по направлению прихода полезного сигнала и помехи; $G_J = G_{TJ} G_{RJ}$ – коэффициент усиления канала связи для помехи; $G_s = G_{Ts} G_{Rs}$ – коэффициент

усиления канала связи для сигнала; r_s , r_J – расстояние (разнесение) между передающими и приемными антеннами соответственно данной САРД и постановщика помех; n – показатель степени затухания. Показатель степени затухания n указывает, насколько быстро растут потери при распространении с увеличением расстояния.

Окончательный вид выражения (2) с учетом (3) будет следующим:

$$P_B = 2Q \left[\sqrt{2 \cdot k \cdot \frac{\frac{E_b}{N_0}}{1 + \left(\frac{E_b}{N_0}\right) \cdot \frac{P_{TJ} G_J (r_s)^n}{P_{Ts} G_s (r_J)^n} \cdot \left(\frac{1}{G_p}\right)}} \cdot \sin \frac{\pi}{2^k} \right]. \quad (4)$$

Предположим теперь, что вместо непрерывной широкополосной помехи действует импульсная помеха повышенной мощности P_J/c , что достигается за счет уменьшения времени ее воздействия на САРД с ($0 < c < 1$) относительно общего времени воздействия, которое считается $c = 1$. Вероятность возникновения импульсной преднамеренной помехи в данный момент времени можно считать равной c . Из-за воздействия преднамеренных помех в течение времени передачи с спектральная плотность мощности преднамеренной помехи с учетом теплового шума возрастает до $(N_0 + J_0/c)$. В оставшийся промежуток времени с вероятностью $(1 - c)$ источник помех не выдает и вероятность битовой ошибки (ВБО) принятого сигнала определяется только наличием теплового шума со спектральной плотностью мощности N_0 (выражение (2) при $J_0 = 0$).

Таким образом, выражение для ВБО при воздействии импульсной помехи представляет собой сумму ВБО с учетом тепловых шумов и ВБО с учетом преднамеренных импульсных помех:

$$P_b = 2 \left\{ (1 - \rho) \cdot Q \left[\sqrt{2k \frac{E_b}{N_0}} \cdot \sin \frac{\pi}{2^k} \right] + \rho \cdot Q \left[\sqrt{2k \frac{E_b}{N_0 + J_0/\rho}} \cdot \sin \frac{\pi}{2^k} \right] \right\} = 2 \left\{ (1 - \rho) \cdot Q \left[\sqrt{2k \frac{E_b}{N_0}} \cdot \sin \frac{\pi}{2^k} \right] + \rho \cdot Q \left[\sqrt{2k \frac{\frac{E_b}{N_0}}{1 + \left(\frac{1}{\rho}\right) \cdot \left(\frac{E_b}{N_0}\right) \cdot \left(\frac{P_J}{P_s}\right) \cdot \left(\frac{1}{G_p}\right)}} \cdot \sin \frac{\pi}{2^k} \right] \right\}, \quad (5)$$

где отношение P_J/P_s также определяется выражением (3).

В [2, 3] показано, что на открытом пространстве САРД работает как в дальней ($n = 4$), так и в промежуточной ($n = 2$) и ближней ($n = 1$) зонах. Применение выражений (5) для расчетов в ближней зоне является весьма приближенным. Согласно [4] понятие коэффициента усиления в ближней зоне может быть использовано в предположении, что вся мощность излучения антенны площадью апертуры S содержится в пределах «прожекторного» пучка. При таком нестрогом приближении [4]

$$G = \frac{4\pi \cdot r^2}{S} \quad (6)$$

и зависит от расстояния.

Проведем анализ полученных результатов расчетов на основании формул (3), (5), которые представлены в виде графиков на рис. 1–3. При расчетах в качестве критерия эффективности работы САРД была исполь-

зована ВБО, которая определяется как вероятность ошибочно принятого символа P_b , причем при $P_b = 10^{-6}$ достоверность принимаемой информации считается высокой, при $P_b = 10^{-4}$ – средней и при $P_b = 10^{-2}$ – низкой [1]. В дальнейших расчетах принята $P_b = -60$ дБ, что соответствует высокому качеству обслуживания (Quality of Service – QoS), в том числе и для передачи мультимедийного трафика [1].

Поверхности ВБО для случая непрерывной помехи приведены на рис. 1–3, в базе $E_b/N_0 - G_j/G_s$. Из представленных данных следует, что при $n = 1$ (ближняя зона для открытого пространства) САРД является

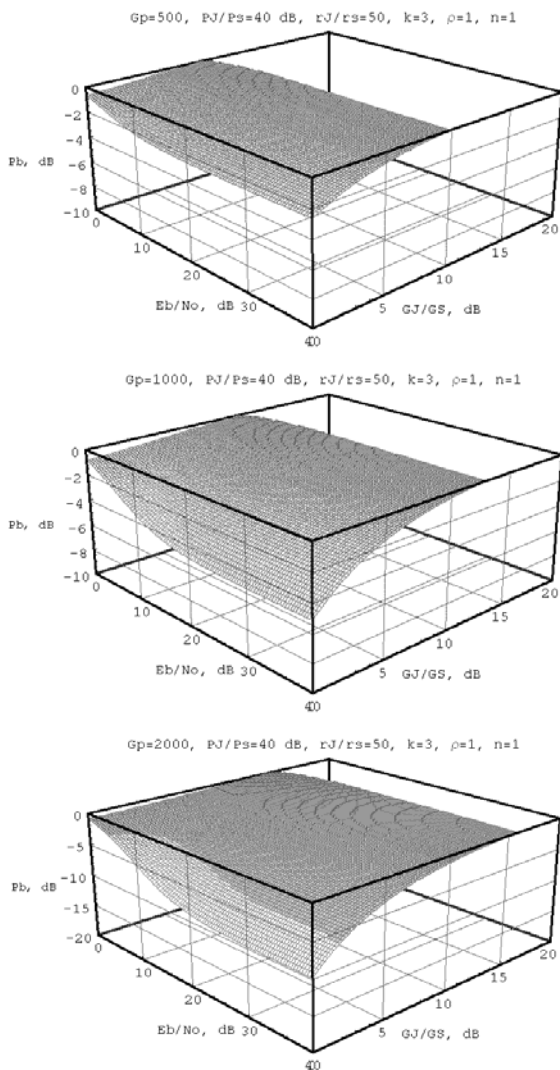


Рисунок 1 – ВБО САРД при $n = 1$

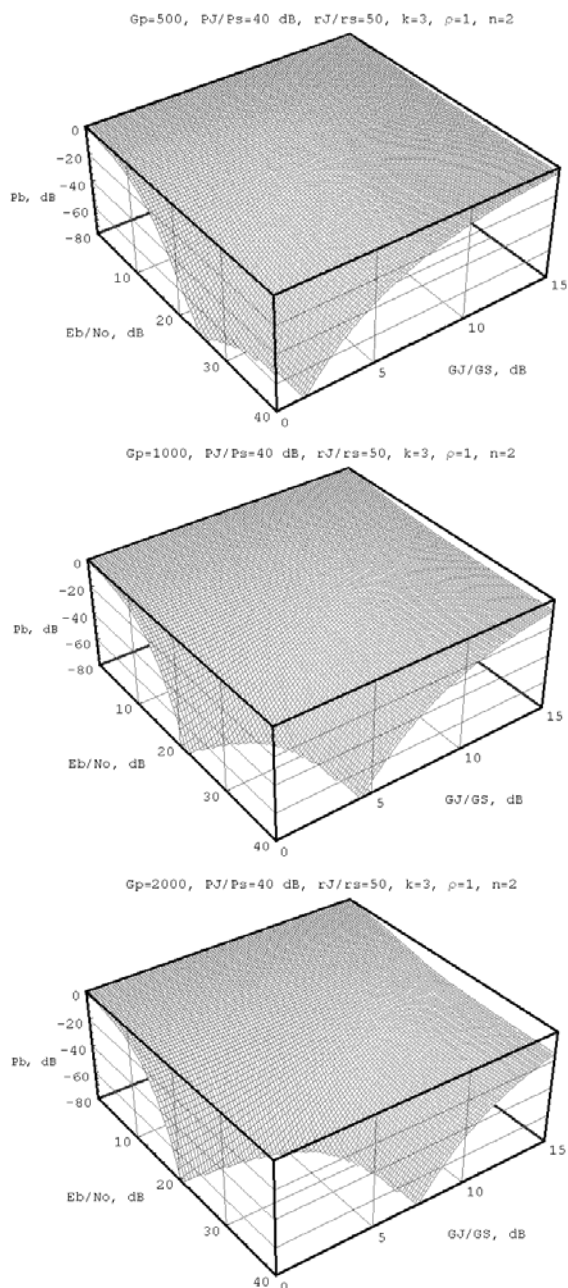


Рисунок 2 – ВБО САРД при $n = 2$

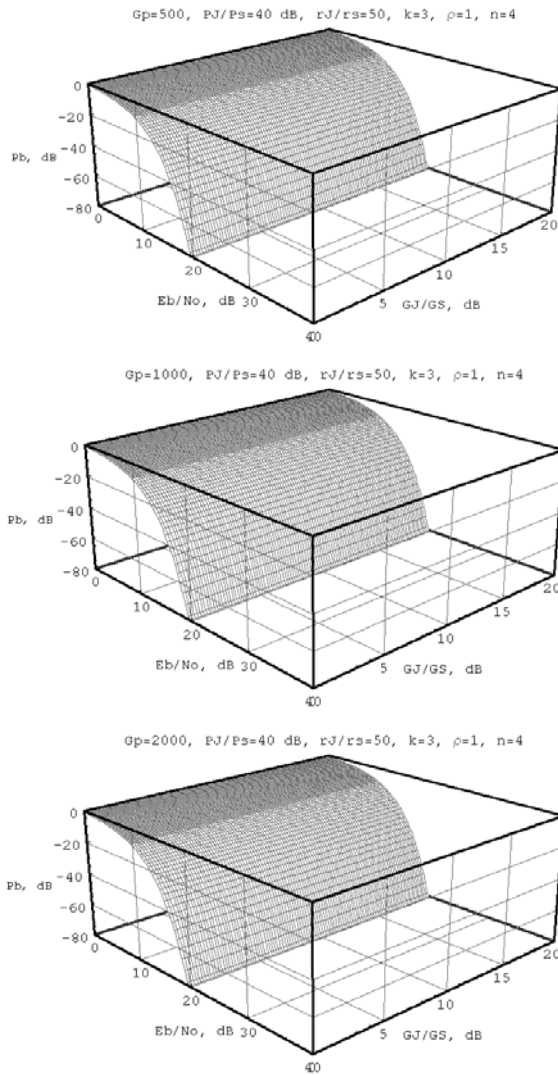


Рисунок 3 – ВБО САРД при $n = 4$

системой с весьма низкой достоверностью передачи информации. Критерий $P_b = -20$ дБ – и тот трудно достигается только при $G_p = 2000$.

По мере роста затухания на трассе распространения (от $n = 2$ до $n = 4$) формируется рабочая область САРД с требуемой ВБО $P_b = -60$ дБ (рис. 2, 3) и условия обеспечения высокой эффективности системы связи существенно облегчаются.

Сравнительный анализ воздействия непрерывных и импульсных помех можно провести из сопоставления данных, представленных на рис. 4, 5. При расчетах рассматривался известный из практики случай, когда на САРД действует мощная помеха $P_{Tj}/P_{Ts} = 40$ дБ, а относительная направленность канала помеха/сигнал G_j/G_s изменяется от 0 до 12 дБ. Этот пример отвечает такой ситуации: САРД расположена возле РЛС с частотной или импульсной модуляцией, в которых несущая частота близка к рабочей частоте САРД,

а направленность антенн выше, чем направленность антенн САРД ($G_j/G_s > 0$ дБ).

Из анализа данных, представленных на рис. 4, 5, выявлен новый эффект. Он заключается в том, что при увеличении относительной направленности канала связи помеха/сигнал кривые $P_b = f(E_0/N_0)$ сглаживаются и при определенном значении G_j/G_s вероятность ошибочного отражения бита информации практически не зависит от увеличения отношения E_0/N_0 . Другими словами – увеличение энергии бита не приводит к росту эффективности работы САРД.

Этот вывод справедлив как для случая непрерывной, так и для случая импульсной помехи. Следует заметить, что с ростом G_j/G_s ВБО ухудшается от 10^{-6} при $G_j/G_s = 0$ дБ до $P_b < 10^{-2}$ при $G_j/G_s = 6$ дБ (см., например, рис. 4, а при $E_0/N_0 = 15$ дБ), если $G_j/G_s = 12$ дБ, то $P_b \approx 10^{-2}$, и фактически не изменяется при $E_b \rightarrow \infty$.

Таким образом, эффективным постановщиком помех, который нарушает работу САРД, может быть мощный генератор с хорошо направленной антенной. И наоборот, САРД сохраняет нужную величину ВБО, если при действии помехи $G_{Tj} \ll G_{Ts}$, а $G_{Rj} \ll G_{Rs}$. Первое условие выполняется за счет использования в САРД передающей антенны с большим значением коэффициента усиления, а второе условие реализуется, если приемная антенна САРД в направлении на помеху имеет значение функции направленности, близкое к нулю.

Перейдем к рассмотрению методики построения кривых ВБО, имеющих постоянное значение на плоскости r, θ . Такие кривые позволяют упростить процесс взаимного размещения КА для достижения заданной достоверности передачи информации при воздействии на САРД помехи.

Свяжем величину P_b с производительностью системы C/B . Из [1] известно, что

$$\frac{E_b}{N_0} = \frac{S}{N} \left(\frac{B}{C} \right), \text{ а } \frac{S}{N} = 2^{(C/B)} - 1.$$

Тогда

$$\frac{E_b}{N_0} = 2^{(C/B)} - 1 \cdot \frac{B}{C}. \tag{7}$$

Подставив (4.7) в (4.4), окончательно получим

$$P_b = 2 \cdot Q \left\{ \sqrt{\frac{(2^{C/B} - 1) \cdot \frac{B}{C}}{1 + \left[(2^{C/B} - 1) \cdot \frac{B}{C} \cdot \frac{P_j}{P_s} \cdot \frac{1}{G_p} \right]}} \cdot \sin \frac{\pi}{2^k} \right\}. \tag{8}$$

Из (4.8) следует, что при $C/B = \text{const}$ и $P_b = \text{const}$.

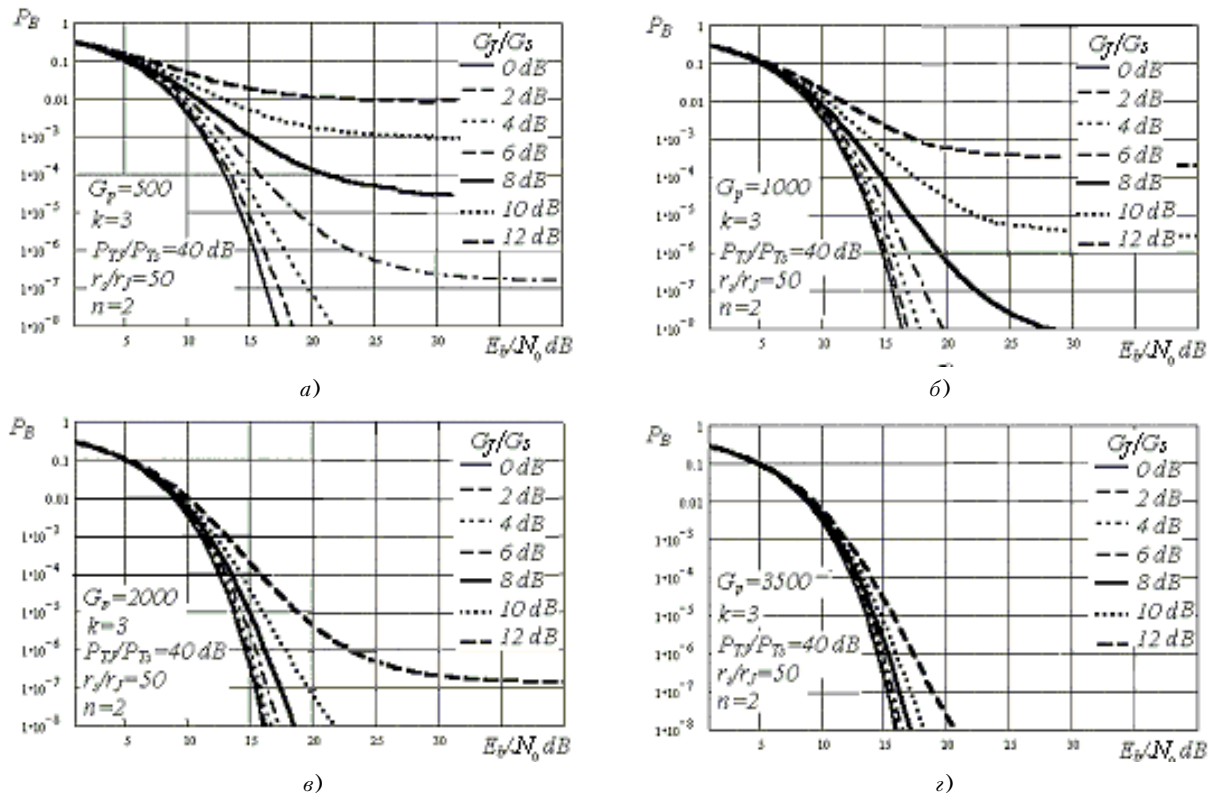


Рисунок 4 – Влияние непрерывной помехи на работу САРД

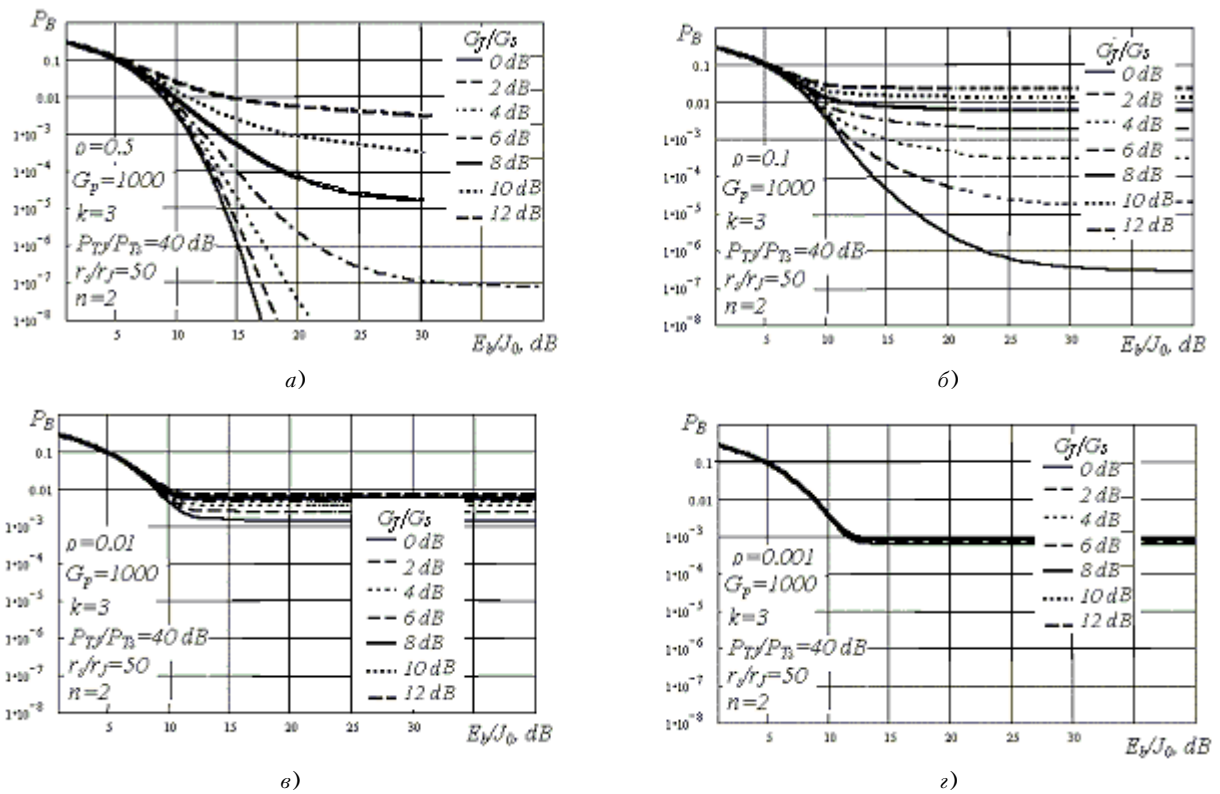


Рисунок 5 – Влияние импульсной помехи на работу САРД

Таким образом, конфигурация кривых постоянной производительности (см. [5]) полностью совпадает с кривыми постоянной вероятности битовой ошибки, которые можно вычислить, записав выражение (8) в виде

$$P_b(r, \Theta) = \text{const.} \quad (9)$$

Кривые постоянной вероятности битовой ошибки строятся по такой же методике, как и кривые постоянной производительности (см. [5]). Они ограничивают области допустимой вероятности битовой ошибки САРД. Пример построения приведен для САРД с протоколом IEEE 802.11b на рис. 6–7.

Здесь изображены кривые $P_b(r, \Theta) = \text{const}$ для случаев открытого пространства и помещения. Эти кривые получены на основании данных о производительности САРД, приведенных в п. 3.3. На каждом из рисунков кривым 1, 2, 3 соответствует различные значения P_b ,

полученные при разных уровнях помехи $\frac{P_j}{P_s}$. Все данные расчетов сведены в табл. 1.

Таблица 1 – Значения P_b для различных уровней $\frac{P_j}{P_s}$

Тип трассы	Зона распространения	$\frac{P_j}{P_s}$	Значение P_b для кривых на рис. 6–8		
			1	2	3
Открытое пространство	Промежуточная	1	$6 \cdot 10^{-18}$	$3 \cdot 10^{-7}$	$3 \cdot 10^{-3}$
		10	$6 \cdot 10^{-18}$	$3 \cdot 10^{-3}$	$5 \cdot 10^{-3}$
		10^2	$4 \cdot 10^{-2}$	$5 \cdot 10^{-2}$	$8 \cdot 10^{-2}$
	Ближняя	1	$7 \cdot 10^{-69}$	$6 \cdot 10^{-18}$	$3 \cdot 10^{-7}$
		10	$9 \cdot 10^{-11}$	$6 \cdot 10^{-8}$	$4 \cdot 10^{-5}$
		10^2	$3 \cdot 10^{-2}$	$4 \cdot 10^{-2}$	$5 \cdot 10^{-2}$
Помещение	Ближняя	1	$3 \cdot 10^{-7}$	$2 \cdot 10^{-3}$	$7 \cdot 10^{-2}$
		10^2	$5 \cdot 10^{-2}$	$8 \cdot 10^{-2}$	10^{-1}

Рассчитанные по предложенной методике кривые $P_b(r, \Theta) = \text{const}$ предоставляют разработчику полезные данные о гарантированной эффективности работы радиоканала САРД при том или ином уровне помехи и при расположении КА внутри и на границе областей допустимых ВБО.

ВЫВОДЫ

1. В рамках известной из [1] модели цифровой системы связи для случаев действия непрерывной и импульсной специальной помехи получены новые выражения для расчета вероятности битовой ошибки, которые учитывают предложенную в разделе 2 модель

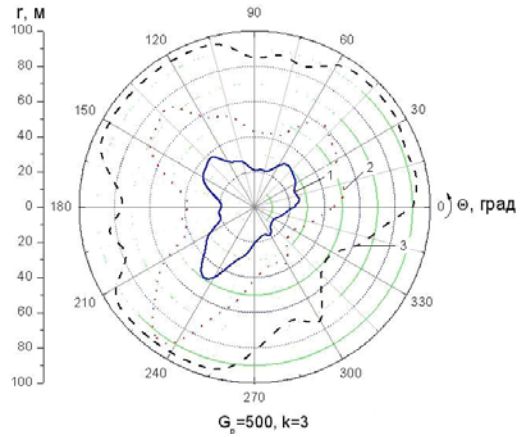


Рисунок 6 – Семейство кривых постоянной вероятности битовой ошибки для открытого пространства в промежуточной зоне

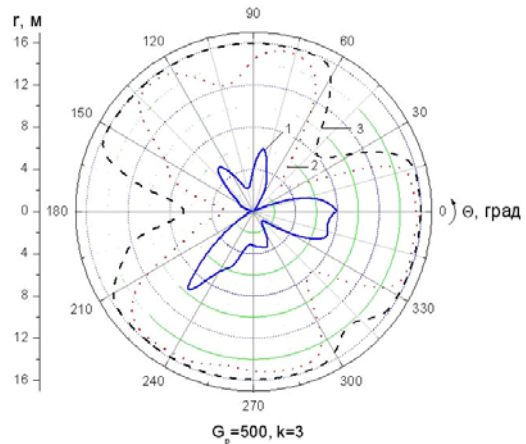


Рисунок 7 – Семейство кривых постоянной вероятности битовой ошибки для открытого пространства в ближней зоне

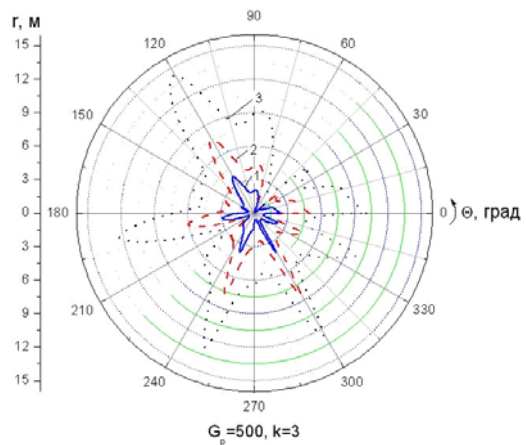


Рисунок 8 – Семейство кривых постоянной вероятности битовой ошибки для помещения в ближней зоне

распространения радиоволн, характеристики направленности антенн радиоканала и постановщика помех, а также производительность САРД.

2. По полученным формулам проведены расчеты и обнаружено, что цифровым системам связи присущ эффект насыщения от деконструктивного действия помех. Этот эффект наступает при определенных значениях отношения мощностей помеха/сигнал и направленностей канала связи в направлении приема помехи и сигнала и состоит в полной независимости вероятности битовой ошибки от энергии бита информации. На основе выявленного эффекта можно сделать рекомендации как по проектированию эффективных постановщиков помех, так и по разработке САРД, слабочувствительных к деконструктивным действиям помех определенной мощности и структуры.

3. Также выяснено, что в ближней зоне САРД трудно реализовать величину $P_b < 10^{-2}$. Это можно сделать только лишь при низком уровне помехи ($P_f/P_S \leq 1$). В частности показано, что при необходимости обеспечения связи в любом направлении, в силу большой неравномерности ДН антенн КА соотношение сигнал/шум в радиоканале САРД должно быть не менее 40 дБ, а при наличии движущейся техники на трассе радиоканала еще на 25 дБ больше.

4. Проведены экспериментальные исследования сигнально-помехового окружения САРД, работающей по протоколу IEEE 802.11b на частоте 2.43 ГГц. Получена совокупность ранее не известных результатов, которые дают возможность сделать рекомендации по планированию радиосетей в помещении и в полевых условиях при деконструктивном действии прицельной помехи и подвижных объектов.

5. На основании проведенных ранее исследований производительности САРД [5] предложена методика

расчета радиоканала с гарантированной вероятностью битовой ошибки (с гарантированной эффективностью).

ПЕРЕЧЕНЬ ССЫЛОК

1. *Бернард Скляр*. Цифровая связь. – Москва–Санкт-Петербург–Киев: Издательский дом «Вильямс», 2003. – 1104 с.
2. *Шокало В. М., Лихограй В. Г., Стрельницкий А. Е., Стрельницкий А. А., Роздоловский Ю. М., Бабанская Е. В.* Моделирование и измерение ослабления сигнала в радиоканале локальных САРД. Часть 1 // Всеукраинский межведомственный научно-технический сборник «Радиотехника». – 2005. – 143. – С. 143–150.
3. *Шокало В. М., Лихограй В. Г., Стрельницкий А. Е., Стрельницкий А. А.* Моделирование и измерение ослабления сигнала в радиоканале локальных САРД. Часть 2 // Всеукраинский межведомственный научно-технический сборник «Радиотехника». – 2006. – № 144. – С. 166–171.
4. ЭМС радиоэлектронных средств и непреднамеренные помехи / Составитель Т. Р. Ж. Уайт. – Джерман Таун, Мериленд, 1971–1973. – Выпуск 1. Общие вопросы ЭМС. Межсистемные помехи: Сокращ. пер. с англ. / Под ред. А. И. Саггира. – М.: Сов. радио, 1977. – 352 с.
5. *Шокало В. М., Стрельницкий А. Е., Стрельницкий А. А., Бабанская Е. В.* Прогнозирование производительности радиоканала локальной системы абонентского радиодоступа без учета влияния помех // Всеукраинский межведомственный научно-технический сборник «Радиотехника». – 2006. – Вып. 145. – С. 60–72.

Надійшла 17.05.06

У статті запропонована модель для розрахунку ефективності роботи САРД при наявності безперервної й імпульсної завади, проведений її математичний аналіз, розроблена методика побудови кривих постійної ймовірності бітрової помилки, що обмежують області гарантованої ефективності роботи САРД.

In the article a model for computation of efficiency of work of the system of subscriber radio of access at presence of continuous and impulsive hindrance is offered, its mathematical analysis is resulted, the method of construction of curves of permanent probability of bit error, which limit the regions of the assured efficiency of the SSRA work, is developed.

УДК 621.372.8.049.75

В. І. Оборжицький

РОЗРАХУНОК ЕЛЕКТРИЧНИХ ПАРАМЕТРІВ НАПРЯМЛЕНИХ ВІДГАЛУЖУВАЧІВ НА ЗВ'ЯЗАНИХ ЛІНІЯХ ПЕРЕДАЧІ З КОМПЕНСУЮЧИМИ РЕАКТИВНОСТЯМИ

Запропоновано метод розрахунку електричних параметрів напрямленого відгалужувача на двох однакових зв'язаних лініях передачі з додатковим і реактивними елементами, який забезпечує можливість компенсації впливу різниці фазових швидкостей парної і непарної мод і впливу неоднорідностей з'єднань вхідних ліній зі зв'язаними лініями.

© Оборжицький В. І., 2006

ПОСТАНОВКА ЗАДАЧІ

Напрявлені відгалужувачі на зв'язаних лініях передачі відносяться до мікрохвильових вузлів, які знаходяться широке застосування в різноманітних радіотехнічних пристроях НВЧ діапазону в інтегрованому

чи монолітно-інтегрованому виконанні. Це зумовлено простою конструкції таких відгалужувачів, допустимими значеннями їх електричних параметрів та хорошими частотними властивостями. Проте при розробці смужкового варіанту конструкції відгалужувача виникають проблеми, які необхідно розв'язувати в процесі проектування. Одна з них зумовлена наявністю неоднорідного діелектричного заповнення зв'язаних ліній передачі. У цьому випадку постійні поширення мод парного і непарного збудження приймають різні значення. В результаті зі збільшенням рівня зв'язку, а також з ростом діелектричної проникності підкладки зменшується напрямленість відгалужувача, погіршуються його частотні характеристики, що особливо суттєво при проектуванні монолітних конструкцій.

Вирішення вказаної проблеми здійснюється як технологічним, так і схемотехнічним шляхом. У першому випадку фазові швидкості парної і непарної мод можна вирівняти покриттям конструкції відгалужувача в області зв'язку додатковим шаром діелектрика [1, 2], введенням додаткового провідника з вільним потенціалом [1, 3], використанням багат шарової діелектричної підкладки [1, 4], наданням області зв'язку періодичного профілю (пилоподібного, ступінчастого) [1, 5], що видовжує шлях непарної моди, так як її струми витісняються до внутрішніх країв провідників, а струми парної моди – до зовнішніх.

Достатньо ефективною і технологічно простою є схемотехнічна компенсація впливу різниці фазових швидкостей, яка ґрунтується на введенні в область зв'язку додаткових реактивних елементів. При цьому найчастіше використовують дві ємності, які включаються між зв'язаними лініями по обидві сторони відрізка [1, 6, 7]. Можна також досягти компенсації за допомогою однієї реактивності [8]: ємності, введеної посередині відрізка зв'язаних ліній; ємності чи індуктивності, під'єднаної до відрізка на одному з його країв. Другий, несиметричний випадок застосовується при необхідності забезпечення лише високого рівня розв'язки чи кращого узгодження на вході. Описані у публікаціях методи розрахунку електричних параметрів напрямлених відгалужувачів містять ряд суттєвих обмежень. Так, метод розрахунку [6, 7] справедливий тільки для випадку сильного зв'язку. Це обмеження знято в [9], але розв'язок отримано для розміщення ємностей на краях області зв'язку. Крім того, всі методи розглядають випадок з нульовою фазовою затримкою сигналу при його переході з основної лінії у відгалужене плече.

Ще одне суттєве обмеження існуючих методів проектування напрямлених відгалужувачів на зв'язаних лініях стосується можливості врахування впливу неоднорідностей, які виникають в місцях з'єднання (стиках) одиночних входних ліній передачі зі зв'язаними лініями. Наявність таких неоднорідностей призводить до розузгодження і, як результат, до зміни характе-

стик реальної конструкції відгалужувача. Необхідні при цьому корективи, як правило, вносяться в процесі проектування на етапі комп'ютерної оптимізації. З метою компенсації впливу неоднорідності ще на етапі визначення електричних параметрів мікрохвильового пристрою запропоновано [10] доповнювати схему пристрою еквівалентною схемою неоднорідності і далі розробляти метод розрахунку параметрів доповненої схеми з урахуванням параметрів цієї неоднорідності.

Оскільки від названих факторів суттєво залежать якісні показники проектування, то за мету даної роботи ставилась розробка методу визначення електричних параметрів напрямлених відгалужувачів з відрізка двох однакових зв'язаних ліній передачі, який би дозволяв компенсувати вплив як різниці фазових сталей парної і непарної мод, так і вплив неоднорідностей стиків шляхом застосування одного чи двох реактивних елементів без обмежень на місце їх включення та на значення фазової затримки сигналу при відгалуженні.

1 ЕКВІВАЛЕНТНА СХЕМА ТА ОСНОВНІ РОЗРАХУНКОВІ СПІВВІДНОШЕННЯ

Розробка методу розрахунку електричних параметрів напрямленого відгалужувача базується на припущенні, згідно з яким у відрізку зв'язаних ліній поширюються квазі-Т хвилі парного (синфазного) і непарного (протифазного) збудження, основними параметрами яких є хвильові опори і фазові сталі парної Z_{oe} , β_e та непарної Z_{oo} , β_o мод (втрати у лініях можна знехтувати завдяки малій довжині відрізків). При такому підході узагальнена еквівалентна схема відгалужувача матиме вигляд, наведений на рис. 1. На ній до кінців відрізка двох зв'язаних ліній передачі довжиною $2l$ під'єднуються одиночні входні лінії з хвильовим опором Z_c . Ці з'єднання (стики) показані на схемі у вигляді чотириполюсників. Всі стики вважаються однотипними. На відстані l_1 від середини відгалужувача ввімкнено додаткові реактивні опори jX . При такій структурі зберігається вертикальна симетрія відгалужувача, що дає змогу застосувати до всієї схеми

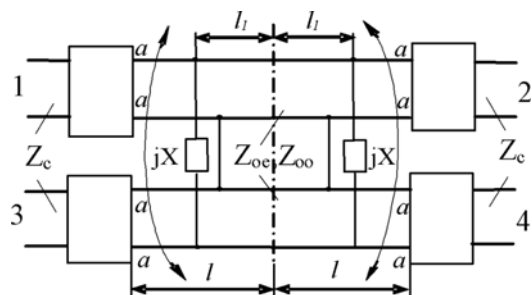


Рисунок 1 – Еквівалентна схема напрямленого відгалужувача

метод синфазно-протифазного збудження. В результаті восьмиполосник з рис. 1 розпадається на два парціальних чотириполосники, еквівалентні схеми яких утворені половиною загальної схеми відгалужувача. Відрізняються вони лише кінцевими навантаженнями відрізків зв'язаних ліній довжиною l : при синфазному збудженні кінці ліній будуть розімкненими (режим холостого ходу – 1); при протифазному збудженні кінці ліній будуть закороченими (режим короткого замикання – 2).

В свою чергу парціальні чотириполосники мають горизонтальну симетрію, що дозволяє розкласти їх на відповідні парціальні двополосники. Вхідні імпеданси цих двополосників можуть бути визначені через параметри розсіяння восьмиполосника:

$$\begin{aligned} \dot{z}_{ie1} &= \frac{1 + \dot{S}_{21} + \dot{S}_{31}}{1 - (\dot{S}_{21} + \dot{S}_{31})}; \quad \dot{z}_{io1} = \frac{1 + \dot{S}_{21} - \dot{S}_{31}}{1 - (\dot{S}_{21} - \dot{S}_{31})}; \\ \dot{z}_{ie2} &= \frac{1}{\dot{z}_{io1}}; \quad \dot{z}_{io2} = \frac{1}{\dot{z}_{ie1}}, \end{aligned} \quad (1)$$

де $\dot{z}_{ie,o1}$ – нормовані до Z_c вхідні імпеданси двополосника синфазного (e) і протифазного (o) збудження у випадку режиму 1; $\dot{z}_{ie,o2}$ – вхідні імпеданси у випадку режиму 2. Співвідношення (1) були отримані з врахуванням умов взаємності, ідеального узгодження ($S_{11} = S_{22} = S_{33} = S_{44} = 0$), а також ідеальної розв'язки ($S_{41} = S_{32} = 0$) напрямленого відгалужувача.

Оскільки еквівалентна схема на рис. 1 утворена тільки реактивними елементами (стики вважаються безвтратними), то вхідні імпеданси також повинні бути реактивними. Вирази (1) забезпечують $z_{ie,o1,2} = jx_{ie,o1,2}$ при виконанні умов квадратурності та реактивності напрямленого відгалужувача, тобто:

$$\varphi_{21} - \varphi_{31} = (2 \cdot k + 1) \cdot \pi / 2; \quad (2)$$

$$|\dot{S}_{21}|^2 + |\dot{S}_{31}|^2 = 1, \quad (3)$$

де φ_{n1} – початкові фази ($k = 0, 1, 2, \dots$), $|\dot{S}_{n1}|$ – модулі параметрів розсіяння $\dot{S}_{n1} = |\dot{S}_{n1}| \cdot \exp(-j\varphi_{n1})$ восьмиполосника ($n = 2, 3$). Тоді для вхідних опорів парціальних двополосників з врахуванням умови (2) можна записати:

$$\begin{aligned} x_{ie,o1} &= \frac{|\dot{S}_{21}| \cdot \sin \varphi_{21} \pm |\dot{S}_{31}| \cdot \sin \varphi_{31}}{|\dot{S}_{21}| \cdot \cos \varphi_{21} \pm |\dot{S}_{31}| \cdot \cos \varphi_{31} - 1}; \\ x_{ie2} &= -\frac{1}{x_{io1}}; \quad x_{io2} = -\frac{1}{x_{ie1}}, \end{aligned} \quad (4)$$

де знак «мінус» відповідає їх синфазному, а знак «плюс» протифазному збудженню.

Оскільки неоднорідність стику входить до складу парціальних двополосників, то розраховані за (4) зна-

чення вхідних реактивних опорів залежать від її параметрів. Щоб в подальших розрахунках позбутися цієї залежності, необхідно, застосовуючи відповідну реактивну еквівалентну схему неоднорідності, наприклад [11], виконати перерахунок вхідних опорів $x_{ie,o1,2}$ у опори $x_{ae,o1,2}$ в перерізі $a-a$ (рис. 1) на виході утвореного неоднорідністю чотириполосника. Трансформація цих опорів через еквівалентну схему неоднорідності повинна давати вихідні значення $x_{ie,o1,2}$. При використанні для неоднорідності, наприклад, поширеної еквівалентної Т-схеми нормовані до Z_c значення вхідних опорів в перерізі $a-a$ розраховуються за співвідношенням

$$\begin{aligned} x_{ae,o1,2} &= \left[\left(1 + \frac{x_3}{x_2} \right) \cdot (x_1 - x_{ie,o1,2}) + x_3 \right] \div \\ &\div \left[\frac{x_{ie,o1,2}}{x_2} - \left(1 + \frac{x_1}{x_2} \right) \right], \end{aligned} \quad (5)$$

де $x_{1,2,3}$ – нормовані до Z_c значення реактивностей Т-схеми (x_2 – паралельна реактивність).

Використовуючи систему з двох рівнянь для вхідних опорів парціальних двополосників, утворених відрізками ліній з додатковою реактивністю, при їх синфазному збудженні в режимах 1 і 2 можна отримати співвідношення, за якими розраховуються електричні параметри відрізка зв'язаних ліній:

$$z_e = \sqrt{-x_{ae1} \cdot x_{ae2}}; \quad (6)$$

$$\operatorname{tg} \Theta_e = -z_e / x_{ae1} = x_{ae2} / z_e, \quad (7)$$

де $z_e = Z_{oe} / Z_c$ – нормоване значення хвильового опору, а $\Theta_e = \beta_e l$ – електрична довжина відрізка для парної моди. Задавши значення хвильового опору непарної моди Z_{oo} , далі можна визначити фазові сталі β_e, β_o , а за ними, використовуючи (7), розрахувати значення електричної довжини відрізка $\Theta_o = \beta_o l$ для непарної моди.

Прирівнюючи вирази для додаткового реактивного опору, отримані з двох рівнянь для вхідних опорів парціальних двополосників при їх протифазному збудженні в режимах 1 і 2, можна записати квадратне рівняння відносно змінної t_1 , яка залежить від відстані l_1 до місця включення додаткової реактивності:

$$a \cdot t_1^2 + 2 \cdot b \cdot t_1 + c = 0, \quad (8)$$

де

$$\begin{aligned} a &= -(z_o + x_{ao1} \cdot t_o) \cdot (z_o + x_{ao2} \cdot t_o); \\ b &= (z_o + x_{ao1} \cdot t_o) \cdot (z_o \cdot t_o); \\ c &= -(z_o \cdot t_o - x_{ao1}) \cdot (z_o \cdot t_o - x_{ao2}); \\ t_1 &= \operatorname{tg}(\beta_o \cdot l_1); \quad t_o = \operatorname{tg} \Theta_o; \end{aligned}$$

$z_o = Z_{oo}/Z_c$ – нормоване значення хвильового опору непарної моди. Саме ж значення нормованого до Z_c додаткового реактивного опору X розраховується за співвідношенням:

$$x = \frac{2 \cdot z_o \cdot (z_o \cdot t_2 - x_{ao1})}{z_o \cdot (t_1 \cdot t_2 - 1) - (t_1 + t_2) \cdot x_{ao1}}, \quad (9)$$

де $x = (X/Z_c)$; $t_2 = \text{tg}(\Theta_o - \beta_o \cdot l_1)$. Ці реактивності під'єднуються до країв області зв'язку відгалужувача при $l_1 = l$, що матиме місце, коли нормований хвильовий опір Z_{oo} становитиме:

$$z_o = \frac{x_{ao1} \cdot x_{ao2} \cdot (t_o^2 + 1)}{(x_{ao1} - x_{ao2}) \cdot t_o}. \quad (10)$$

Як видно з рис. 1, при $l_1 = 0$ в схему відгалужувача вводиться тільки одна додаткова реактивність. При цьому значення хвильового опору непарної моди дорівнюватиме

$$z_o = x_{ao2}/t_o, \quad (11)$$

а нормований додатковий опір згідно з (9) визначатиметься за співвідношенням

$$x = z_o \cdot (x_{ao1} - z_o \cdot t_o) / (z_o + x_{ao1} \cdot t_o). \quad (12)$$

При фазовій затримці $\phi_{31} = 0$ та без врахування впливу стиків формули (6), (7), (11), (12) співпадають з виразами [8], отриманими для відгалужувача з однією реактивністю.

2 ПОСЛІДОВНІСТЬ РОЗРАХУНКІВ ТА ПРИКЛАДИ ЇХ ЗАСТОСУВАННЯ

Вихідними даними для розрахунків є значення перехідного загасання S , фазовий зсув сигналу ϕ_{31} при відгалуженні а також значення хвильового опору Z_c вхідних ліній.

Використовуючи S та ϕ_{31} з врахуванням (2), (3) розраховуються значення модулів хвильових параметрів S_{21} , S_{31} і далі за (4) – значення вхідних опорів $x_{ie,o1,2}$ парціальних двополосників, при яких будуть забезпечуватись ці параметри відгалужувача.

Далі необхідно виконати перерахунок $x_{ie,o1,2}$ у вхідні опори $x_{ae,o1,2}$ в перерізі $a-a$. Оскільки еквівалентні параметри неоднорідності стику ще не відомі, то процес обчислень відбуватиметься в ітераційному циклі, в якому за перше наближення приймаються значення $x_{ae,o1,2} = x_{ie,o1,2}$ і за (6), (7) розраховуються параметри парної моди Z_{oe} , Θ_e . Потім вибирається чи при вимозі, щоб $l_1 = l$ чи $l_1 = 0$, розраховується значення хвильового опору непарної моди Z_{oo} . Якщо значення Z_{oo} вибирається, то шукаються корені рівняння (8), які в значній мірі залежать від цього вибору, і далі за ними визначається відстань l_1 та за (9) розраховується

реактивність X . Якщо ж хвильовий опір Z_{oo} обчислюється, то знову виникає потреба в ітераційній процедурі, оскільки електрична довжина Θ_o ще не відома, за перше наближення якої приймається значення $\Theta_o = \Theta_e$. Далі за (10) чи (11) розраховується Z_{oo} , визначаються фазові сталі β_e , β_o , приймається наступне наближення $\Theta_o = \beta_o l$ і виконується повторний розрахунок Z_{oo} . По закінченню цього ітераційного циклу за (9) чи (12) розраховується значення додаткового опору X . Слід відмітити, що для зв'язаних ліній передачі, у яких $\beta_e > \beta_o$, додаткова реактивність буде від'ємною, тобто для компенсації впливу різниці фазових швидкостей мод використовуються тільки ємності.

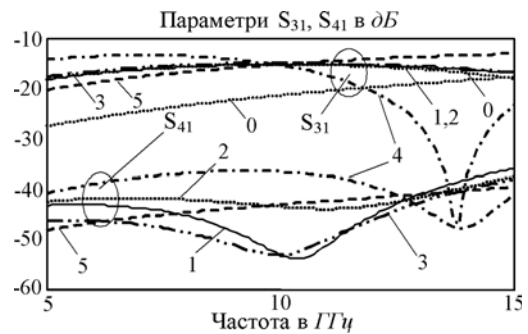


Рисунок 2 – Залежність модулів коефіцієнтів передачі S_{31} , S_{41} від частоти

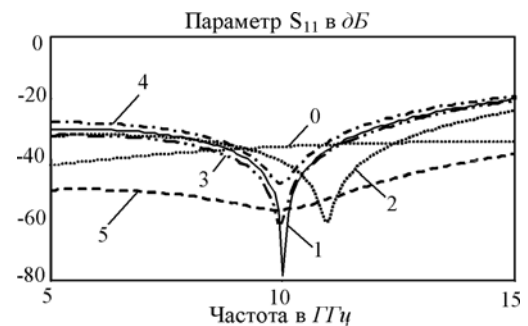


Рисунок 3 – Залежність модуля коефіцієнта відбиття S_{11} від частоти

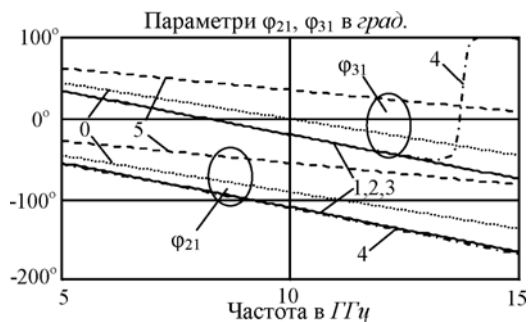


Рисунок 4 – Залежність фазових зсувів ϕ_{21} , ϕ_{31} від частоти

Наступним кроком в ітераційному циклі компенсації впливу неоднорідностей стиків є розрахунок еквівалентних параметрів цих неоднорідностей на основі значень Z_c , Z_{oe} , Z_{oo} . Далі за (5) розраховуються вхідні опори $x_{ae,01,2}$, які в повторних обчисленнях використовуються як наступне наближення. Швидка збіжність ітерацій зумовлена достатньо малою зміною параметрів зв'язаних ліній під впливом неоднорідностей.

Прикладом застосування розробленого методу розрахунку можуть служити результати комп'ютерного моделювання напрямлених відгалужувачів з перехідним загасанням $C = 15$ дБ на робочій частоті 10 ГГц у мікросмужковому виконанні на підкладці з GaAs з діелектричною проникністю 12,9 товщиною 100 мкм. Хвильовий опір вхідних ліній становить 50 Ом. На рис. 2 приведено частотні залежності модулів коефіцієнтів передачі, які відповідають перехідному загасанню та розв'язці напрямлених відгалужувачів, на рис. 3 – частотні залежності модуля коефіцієнта відбиття на вході 1, а на рис. 4 – частотні характеристики фазових затримок φ_{21} , φ_{31} сигналів на виходах 2 та 3. На цих рисунках криві 0 відносяться до відгалужувача з параметрами $Z_{oe} = 59,8$ Ом, $Z_{oo} = 41,8$ Ом, $\Theta_e = 45^\circ$, $2l = 2,64$ мм, розрахованими класичним методом [1] без компенсації різниці фазових швидкостей та неоднорідності, що є причиною досить низького рівня розв'язки (рис. 2). У відгалужувача, який на робочій частоті при $\varphi_{31} = 20^\circ$ без врахування впливу стиків має розраховані запропонованим методом параметри $Z_{oe} = 60,6$ Ом, $Z_{oo} = 49,6$ Ом, $\Theta_e = 55^\circ$, $2l = 3,10$ мм, $l_1 = 0$, додаткова ємність $C_d = 0,056$ пФ, вказаний недолік усунуто (криві 1). Під впливом неоднорідностей стиків частотні характеристики такого відгалужувача змінюються (криві 2). Компенсація цього впливу досягається зміною параметрів до значень $Z_{oe} = 60,7$ Ом, $Z_{oo} = 49,5$ Ом, $\Theta_e = 58^\circ$, $2l = 3,28$ мм, $l_1 = 0$, $C_d = 0,062$ пФ, яким відповідають криві 3.

Характеристики відгалужувачів залежать від місця включення ємностей. Наприклад, при $l_1 = l$ у відгалужувача з параметрами, аналогічними попередньому випадку, і з $C_d = 0,034$ пФ погіршується його напрямленість та звужується робоча смуга частот (криві 4).

Шляхом вибору значення фазового зсуву φ_{31} можна змінювати довжину відрізка зв'язаних ліній. Криві 5 відповідають відгалужувачу з електричними параметрами $Z_{oe} = 62,3$ Ом, $Z_{oo} = 44,8$ Ом, $\Theta_e = 27^\circ$, $2l = 1,43$ мм, $l_1 = 0$, $C_d = 0,03$ пФ при $\varphi_{31} = -35^\circ$. Порівняно з попереднім випадком довжина області зв'язку скоротилася вдвічі при деякому зниженні напрямленості.

ВИСНОВКИ

Запропонований в роботі метод розрахунку напрямлених відгалужувачів на зв'язаних лініях передачі дозволяє визначати його електричні параметри з одночас-

ною компенсацією за допомогою додаткових реактивних опорів впливу різниці фазових швидкостей мод парного і непарного збудження, а також впливу неоднорідностей з'єднань одиночних і зв'язаних ліній. При цьому допускається використання як одного, так і двох опорів, місце під'єднання яких залежить від значення хвильового опору непарної моди. Крім того, шляхом вибору значень фазового зсуву сигналу при відгалуженні можна зменшити розміри відгалужувача. Результати моделювання напрямлених відгалужувачів з електричними параметрами, розрахованими запропонованим методом, підтверджують його ефективність та доцільність застосування в інженерній практиці і, в першу чергу, в структурі систем автоматизованого проектування мікрохвильових пристроїв.

ПЕРЕЛІК ПОСИЛАНЬ

1. Микроелектронные устройства СВЧ / Г. И. Веселов, Е. Н. Егоров, Ю. Н. Алехин и др. – М.: Высш. шк., 1988. – 280 с.
2. *Buntschuh C.* High directivity microstrip couplers, using dielectric overlays / IEEE MTT-S Symp. Dig., 1975. – Pp. 125–128.
3. *Banba S., Ogawa H.* Multilayer MMIC directional couplers using thin dielectric layers / IEEE Trans. Microwave Theory Tech. – June 1995. – Vol. 43, No. 6. – Pp. 1270–1275.
4. *Mar S. L.* Phase velocity compensation in parallel-coupled microstrip / IEEE MTT-S Symp. Dig., June 1982. – Pp. 410–412.
5. *Podell A.* A high directivity microstrip coupler technique / IEEE MTT-S Symp. Dig., May 1970. – Pp. 33–36.
6. *Schaller G.* Optimization of microstrip directional coupler with lumped capacitors / Arch. Elek. Uebertrag. Tech.. – 1977. – Vol. 31, July-Aug.. – Pp. 371–375.
7. *Kajfez C.* Raise coupler directivity with lumped compensation / Microwaves. – 1978. – Vol. 27, Mar.. – Pp. 64–70.
8. *Dydyk M.* Microstrip directional couplers with ideal performance via single-element compensation / IEEE Trans. Microwave Theory Tech.. – June 1999. – Vol. 47, No. 6. – Pp. 956–964.
9. *Dydyk M.* Accurate design of microstrip directional couplers with capacitive compensation / IEEE MTT-S Symp. Dig., May 1990. – Pp. 581–584.
10. *Оборжицкий В. И.* Моделирование линейных СВЧ устройств с синтезом их электрических параметров, учитывающим влияние неоднородностей / Труды 6-й Междунар. НПК «Современные информационные и электронные технологии», Одесса, 24–26 мая 2005 г. – С. 207.
11. *Sabban A., Gupta K. C.* A planar-lumped model for coupled microstrip line discontinuities / IEEE Trans. Microwave Theory Tech. – Feb. 1992. – Vol. 40, No. 2. – Pp. 245–252.

Надійшла 20.06.06
Після доробки 3.08.06

Предложен метод расчета электрических параметров направленного ответвителя на двух одинаковых связанных линиях передачи с дополнительными реактивными элементами, обеспечивающий возможность компенсации влияния разности фазовых скоростей парной и непарной мод и влияния неоднородностей соединений входных линий со связанными линиями.

The method of electrical parameters calculation of two-coupled line directional coupler with additional lumped reactances is proposed. The possibility to compensate the effect of difference of even and odd mode phase velocities and also the effect of input-line-to-coupled-line junction discontinuities is provided at that.

В. В. Топалов

СИНТЕЗ СИГНАЛЬНЫХ КОНСТРУКЦИЙ НА ОСНОВЕ МНОГОПОЗИЦИОННЫХ ВРЕМЕННЫХ СИГНАЛОВ И ТУРБО КОДОВ

Представлен кодер и декодер новых составных Турбо конструкций. Приводится выведенная аналитическая граница вероятности ошибки бита. Оценивается энергетическая эффективность составных Турбо конструкций с пределом Шеннона и Турбо кодами. Производятся измерения эффективности Турбо конструкций с многопозиционными временными сигналами на реальных каналах тональной частоты.

ВВЕДЕНИЕ

В многопозиционных временных сигналах (МВС) информация о передаваемом символе заложена не в значениях отдельных отсчетов, а в местах нахождения значащих моментов модуляции (ЗММ) на интервале формирования сигнальных конструкций [1]. В конструкциях на основе многопозиционных временных сигналов признаками кодового слова являются число переходов, заданная длительность кодового слова, начало отсчета. В качестве базового элемента при МВС выбирается временной интервал Δ , длительность которого меньше найквистового интервала t_0 для цифровых сигналов в S раз $\Delta = t_0/S$. Для уменьшения межсимвольных искажений расстояния между ЗММ выбирается длительность сигнала $t_c \geq S\Delta$. На рисунке 1 показаны вышеизложенные принципы формирования МВС. Представленные на рисунке кодовые слова МВС сформированы на фиксированном временном интервале $T_k = mt_0$.

Кодовые слова МВС с постоянной длительностью представляют собой сигнальные конструкции с количеством базовых элементов $i = mS$ (m – длительность кодовой комбинации в единичных элементах $m = T_k/t_0$).

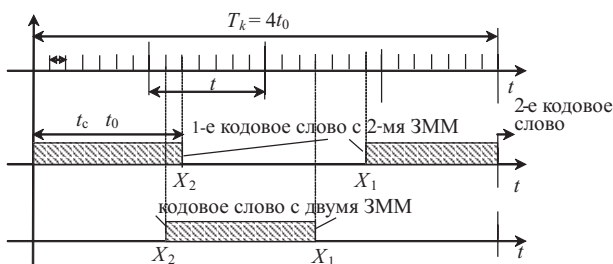


Рисунок 1 – Формирование МВС

© Топалов В. В., 2006

Скорость передачи в таких сигналах выше в сравнении с цифровыми сигналами за счет возможности получения большого числа реализаций (N_p) – $N_p > 2^m$ на интервале $T_k = mt_0$. Общее число реализаций при заданном i – количестве переходов и заданной длительности сигнала [2]:

$$N_p = C_{ms-i}^i \quad (1)$$

На рисунке 1 отображены возможные варианты переходов при фиксированной длительности $T_k = 4t_0$ сигнальной конструкции кодового слова с двумя переходами при $S = 7$.

Выбор Δ определяется вероятностью смещения значащих моментов времени (ЗМВ) кодового слова $P_{см}$ на величину $\geq \Delta/2$, количеством переходов в кодовом слове j , среднеквадратическим отклонением δ смещений ЗММ и законом распределения их смещения. При нормальном законе распределения ошибок в канале Δ определяется:

$$\Delta = 2\delta * \Psi^{-1}(j\sqrt{(1-P_{см})}), \quad (2)$$

где Ψ^{-1} – обратная функция Крампа.

$$\Psi^{-1} = 1/\Psi(t) = 1/\left(\frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt + 0,5\right) \quad (3)$$

1 ПОСТАНОВКА ЗАДАЧИ

Как показано в работах [3, 4], для каналов с вероятностью ошибки смещения от 10^{-2} до 10^{-3} добиться вероятности ошибки приема бита 10^{-8} при скорости передачи $> 2/3\Delta F$ с помощью известных многопозиционных временных кодов невозможно. Автором предлагается для повышения помехоустойчивости формировать конструкции МВС на основе избыточной цифровой последовательности. Предлагается к рассмотрению новая составная турбо конструкция из рекурсивных сверточных кодов и блочных кодов. Рассматриваемая конструкция – Турбо-блок коды, представляющие собой последовательное соединение двух параллельных

сверточных систематических рекурсивных кодов с двумя параллельными или последовательными блочными кодами. В качестве блочных кодов рассматриваются коды Боуза-Чоудхури-Хоквингема (БЧХ).

2 КОДЕР И ДЕКОДЕР ТУРБО-БЛОК КОДА

Последовательное соединение двух параллельных сверточных рекурсивных кодов с двумя параллельными БЧХ кодерами представлено на рисунке 2.

В данной конструкции рассматриваются линейные систематические блочные коды C_1 с параметрами (n_1, k_1, d_1) и C_2 с параметрами (n_1, k_1, d_2) , где n_i, k_i, d_i ($i = 1, 2$) длина блока, количество информационных элементов и минимальное расстояние Хэмминга, соответственно. Вторая часть Турбо конструкции состоит из линейных систематических сверточных кодов C_3 (m_3, g_3) и C_4 (m_4, g_4), где m_j, g_j ($j = 3, 4$) количество элементов памяти декодера и порождающий полином, соответственно. Используются симметричные перемежители $[I]$ с псевдослучайным распределением. Схема Турбо-блок декодера, для данной конструкции представлена на рисунке 3.

В схеме Турбо-блок декодера используется базовый декодер с вероятностным входом и вероятностным выходом (ВВВВ). На вход ВВВВ декодера поступают с канала связи не округленные значения до (1,0), а непрерывные значения, зависящие от отношения сигнал/шум. На выходе такого декодера формируются «мягкие» решения, т. е. опять непрерывные значения [5].

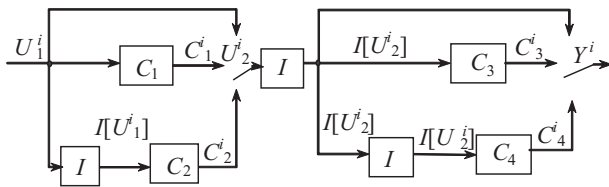


Рисунок 2 – Схема Турбо-блок кодера параллельных конструкций

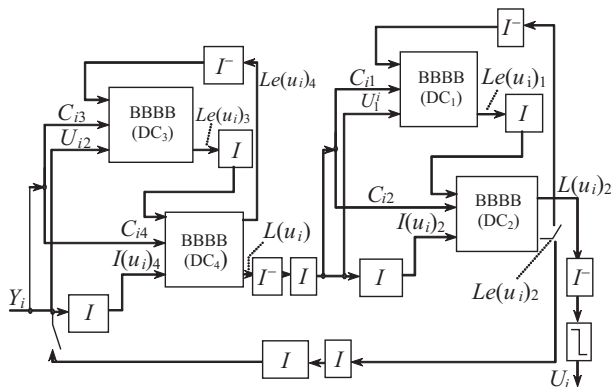


Рисунок 3 – Общая схема Турбо-блок декодера

На вход Турбо-блок декодера подаются непрерывные значения элементов Y^i , принятые из канала связи, которые разделяются на информационные u^i и проверочные C^i элементы перед поступлением на ВВВВ декодер DC_3 . Третий декодер DC_3 формирует априорную информацию $Le(u^i)_3$, поступающую на четвертый декодер DC_4 . DC_4 на следующем шаге формирует также априорную информацию $Le(u^i)_4$, которая является входной величиной для DC_3 . После N итераций четвертый декодер формирует вероятностное решение $L(u^i)_4$, являющееся входной последовательностью для БЧХ декодеров (C_1, C_2) с вероятностным входом и выходом. Для БЧХ декодеров производится также разделение входной последовательности на информационные и проверочные элементы. Работа декодеров DC_1, DC_2 аналогична DC_3, DC_4 . После операций дегермежевания $[I]$ и «оценки» $[L]$ элементы вероятностного решения $L(u^i)_2$ преобразуются в информационные элементы U_j . Для декодирования применяется алгоритм ЛМАВ (log-MAP [6]), но с дополнительной нормализацией выходных последовательностей после параллельных сверточных кодов и параллельных БЧХ кодов. Данная нормализация обусловлена выполнением условия равновероятных величин, поступающих на ВВВВ декодер [6]. Так как значения выходных последовательностей после БЧХ и сверточных декодеров отличаются, то требуется их нормализация. Нормализация заключается в уменьшении разброса среднего значения вероятностных величин выходных элементов в соответствии со входными. Коэффициент нормализации выходной последовательности равен отношению суммы абсолютных значений массива информационных элементов (u_j^i) к сумме абсолютных значений массива выходной последовательности ($|L(u^i)_j|$) на длине последовательности q :

$$\vartheta = \frac{\sum_{j=1 \dots q} u_j^i}{\sum_{j=1 \dots q} |L(u^i)_j|} \quad (4)$$

Полученный в результате данного соответствия коэффициент нормализации, используется для коррекции выходной последовательности $L'(u^i)_j$.

$$L'(u^i)_j = L(u^i)_j / \vartheta \quad (5)$$

3 АНАЛИТИЧЕСКАЯ ГРАНИЦА ВЕРОЯТНОСТИ ОШИБКИ БИТА

Аналитическая граница вероятности ошибки бита линейной конструкции Турбо-блок кода определяется несколькими составляющими: границей для Турбо кода и границей для Турбо блочного кода. Аналитическая граница вероятности ошибки бита при аддитивном

белом гауссовском шуме для Турбо кода согласно [7] определяется как:

$$P_b \leq \sum_{d=d_{\min}}^n \frac{\omega_d A_{w,d}}{k} Q\left(\sqrt{d \frac{2rE_b}{N_0}}\right), \quad (6)$$

где d_{\min} – минимальное кодовое расстояние, n – длина кодового слова, ω_d – средний вес $A_{w,d}$ сообщений, $A_{w,d}$ – число кодовых слов веса d , Q – функция Крампа,

$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{t^2}{2}} dt$, r – скорость кода, E_b – энергия на бит, k – количество информационных элементов, N_0 – спектральная мощность шума.

Считая, что после перемежителей поток ошибок описывается гауссовским распределением, из выражения (7) можем найти отношение сигнал/шум выходной последовательности

$$h = Q^{-1}\left(0,5 - \sum_{d=d_{\min}}^n \frac{\omega_d A_{w,d}}{k} Q\left(\sqrt{d \frac{2rE_b}{N_0}}\right)\right) / \sqrt{2}. \quad (7)$$

После Турбо декодера данная последовательность декодируется параллельным Турбо блочным декодером.

Аналитическая граница вероятности ошибки бита для линейного Турбо блочного кода [8]:

$$P \leq \sum_{d_{pc}} A_{d_{pc}}^c * Q\left(\frac{d_{pc}}{2\sigma}\right), \quad (8)$$

где $A_{d_{pc}}^c$ – количество кодовых слов с евклидовым расстоянием d_{pc}

$d_{pc} A_{d_{pc}}^c = \sum_{l \geq d_{\min}} \frac{A_l^0 * A_{l,d_{pc}}^i}{(N/l)}$, A_l^0 – количество кодовых слов после кодирования с весом Хэмминга

равным l , $A_{l,d_{pc}}^i$ – количество кодовых слов до кодирования с весом Хэмминга равным l и выходным евклидовым весом d_{pc} , N – длина перемежителя, σ – среднеквадратическое отклонение шума.

Учитывая, что среднеквадратическое отклонение шума для гауссовского канала $\sigma = 1/(\sqrt{2}h)$, то выражение для аналитической границы вероятности ошибки бита Турбо-блок кода будет иметь вид:

$$P \leq \sum_{d_{pc}} A_{d_{pc}}^c * Q\left(\frac{d_{pc}h}{\sqrt{2}}\right) \geq \sum_{d_{pc}} A_{d_{pc}}^c * Q \times \left(d_{pc} Q^{-1}\left(0,5 - \sum_{d=d_{\min}}^n \frac{\omega_d A_{w,d}}{k} Q\left(\sqrt{d \frac{2rE_b}{N_0}}\right)\right) / 2\right). \quad (9)$$

4 ОЦЕНКА ЭФФЕКТИВНОСТИ ТУРБО-БЛОК КОДА ПРИ МАТЕМАТИЧЕСКОМ МОДЕЛИРОВАНИИ

С целью проверки точности полученных выражений аналитической границы и оценки энергетической эффективности были выбраны сверточные кодеры C_1, C_2 с порождающим полиномом, предложенным Дивсаларом (33/37), $k = 5$ [9] и БЧХ кодеры C_3, C_4 с параметрами $n = 127, m = 120, d = 1$ и полиномом $G = 211$. Скорость Турбо-блок кода составляет в данной реализации 0,45. Длина блока перемежителя бралась равной 16000 и 64000. На рисунке 4 представлена энергетическая эффективность Турбо-блок кода с выше перечисленными параметрами при моделировании на гауссовском канале связи (кривые 3, 4). Для оценки энергетической эффективности на рисунке 4 представлен предел Шеннона (кривая 1), аналитическая граница Турбо-блок кода с выше указанными параметрами кода и длиной перемежителя 64000 элементов (кривая 2) и значения Турбо кода с порождающим полиномом (33/37) и $k = 5$ (кривая 5).

На рисунке 4 видно, что в отличие от Турбо кодов данная конструкция не обладает порогом эффективности до вероятности ошибки 10^{-11} . Ввод блочных кодов в Турбо конструкцию позволил получить прирост по энергетической эффективности 0,3–0,4 дБ. Минимальное расстояние до предела Шеннона, полученное при данном моделировании, и вероятности ошибки бита $10^{-8} - 0,4$ дБ для Турбо-блок кода с длиной перемежителя 64000.

5 ЭФФЕКТИВНОСТЬ МВС С ТУРБО КОНСТРУКЦИЯМИ ПРИ ПЕРЕДАЧЕ ЧЕРЕЗ КАНАЛЫ ТОНАЛЬНОЙ ЧАСТОТЫ

Нижеследующие результаты представляют измерения потока данных на основе МВС после Турбо-блок

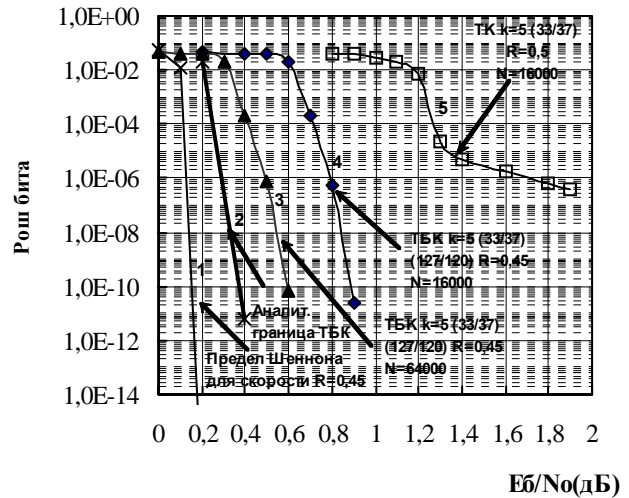


Рисунок 4 – Энергетическая эффективность Турбо-блок кодов при гауссовском распределении ошибки

кодера на реальном канале ТЧ с ограничением полосы пропускания и дальнейшую их обработку Турбо-блок декодером (таблица 1). Был выбран Турбо-блок код с параметрами сверточного кода $k = 5$, $G = (33/37)$ и параметрами блочного кода $n = 127$, $k = 120$, $d = 1$ и полиномом $G = 211$, длиной перемежителя $N = 64000$ и общей скоростью ТБК $R = 0,45$. Многопозиционные сигналы были выбраны с параметрами $S = 8$, $m = 2$ и $j = 1$.

Таблица 1 – Результаты измерения и декодирования данных на каналах ТЧ

В, бод	100	200	700
F, Гц	700	700	700
C, бит/с	67,5	135	472,5
Вероятность дробления	6,90E-04	7,98E-04	1,30E-02
Вероятность увел. число переходов	2,40E-04	2,20E-04	1,25E-02
Вероятность умен. число переходов	4,50E-04	5,78E-04	5,12E-04
Вероятность смещения >10%	6,96E-03	8,38E-03	1,50E-02
Σ Вероятности ошибки МВК	7,65E-03	9,18E-03	2,80E-02
Вероятность ошибки бита до ТБК	6,00E-03	8,20E-03	2,18E-02
Вероятность ошибки бита после ТБК	5,3E-12	2,67E-11	2,16E-08

Использование Турбо-блок кода в синтезе с МВС сигналами позволило достичь вероятности ошибки 10^{-8} при передаче многопозиционных временных сигналов в канале ТЧ.

ВЫВОДЫ

Кодовые конструкции на основе синтеза Турбо блочных и Турбо кодов позволяют получить высокие результаты энергетической эффективности, сравнимые с Турбо кодами для гауссовского канала связи, но без порога эффективности. Использование данных кодовых конструкций вместе с МВС показало высокую эффективность на реальных каналах ТЧ с группированием ошибок.

УДК 691.385

Е. Я. Швец, Л. Л. Вережкин

МИКРОКОНТРОЛЛЕРНОЕ УСТРОЙСТВО ДЛЯ ЛЕЧЕНИЯ АМБЛИОПИИ

Представлены результаты разработки нового устройства, предназначенного для восстановления зрения амблиопического глаза, перспективность которого заключается в возможности проведения комплекса лечебных мероприятий для детей различных возрастных групп.

© Швец Е. Я., Вережкин Л. Л., 2006

ПЕРЕЧЕНЬ ССЫЛОК

1. Захарченко Н. В. Выбор средств связи АСУ: Учебное пособие. – Одесса: УГАС, 1975. – 110 с.
2. Расчет эффективности совместного использования РЦК и МВС: Учебное пособие / Под ред. Н. В. Захарченко. – Одесса: УГАС, 1996. – 68 с.
3. Захарченко М. В., Ульєв О. П., Топалов В. В. Ймовірність помилкового прийому сигнальної конструкції надлишкового багатопозиційного часового коду при і-кратному повторенні // Радіотехніка. – Випуск № 112. – 1999. – С. 52–55.
4. Горохов С. М., Захарченко В. М., Топалов В. В. Ефективність використання широтно-імпульсної модуляції при передачі дискретної інформації // Збірник наукових праць УДАЗ за 1-е півріччя. – 1999. – С. 44–48.
5. Berro C., Glavieux A., Thitimajshima P. Near Shannon limit error correcting coding and decoding // Turbo codes. Proc. International Conference on Communications. – May 1993. – Pp. 1064–1070.
6. Robertson P., Hoeher P., Villebrun E. Optimal and suboptimal maximum a posteriori algorithms suitable for turbo decoding // European Trans. on Telecommun. – Vol. 8. – Mar./Apr. 1997. – Pp. 119–125.
7. Matthew C. Valenti. Iterative Detection and Decoding for Wireless Communications // Dissertation for the degree of Doctor of Philosophy in Electrical Engineering. – Blacksburg-Virginia, 1999. – P. 217.
8. Tolhuizen, S. Baggen and E. Hekstra-Nowacka. Union bounds on the performance of product codes // Proc. IEEE International Symposium on Information Theory. – MIT, Mass, Boston, August 1998. – P. 267.
9. Pyndiah R., Glavieux A., Picart A., Jasq S. Near optimum decoding of product code // Proc. IEEE Globecom Conf. – San Francisco, November 1994. – Pp. 339–343.

Надійшла 2.03.06
Після доробки 17.04.06

Відображується принципи формування кодера та декодера нових складених Турбо конструкцій. Приведена виведена аналітична границя ймовірності помилки для нових конструкцій. Оцінюється енергетична ефективність складених Турбо конструкцій з межою Шенона та Турбо кодами. Проводяться вимірювання ефективності нових складених Турбо конструкцій з багатопозиційними часовими сигналами на реальних каналах тональної частоти.

The coder and the decoder of new compound Turbo constructions is presented in article. We show the deduced analytical border of error probability for new constructions. Power efficiency of compound Turbo constructions with Shannon limit and the Turbo codes is estimated. Measurements of efficiency of the Turbo constructions with multiposition time signals on real voice-frequency channels are made.

ВВЕДЕНИЕ

Амблиопия – понижение зрения, обусловленное функциональными расстройствами зрительного анализатора,

возникающими, в основном, в период развития зрительной системы.

Все методы лечения амблиопии можно разделить на две большие группы. К первой группе относятся методы, требующие активного участия больного (метод засветки с определением феномена последовательного образа, метод на основе занятий с локализатором-корректором и т. д.). Ко второй группе принадлежат методики, которые не предполагают сознательного участия пациента (метод окклюзии, метод засветки красным светом, метод на основе занятий с «Амблиоспеклом» и др.) [1]. Каждая из групп имеет свои преимущества и недостатки. Наиболее эффективно лечение, когда пациент сознательно пытается сохранить в памяти последовательный образ, при этом включаются резервные механизмы в коре и подкорке головного мозга, которые снимают процессы торможения в амблиопичном глазе. Однако такие методы применимы для детей старшего возраста, так как маленький ребенок не всегда может сообразить, как фиксировать последовательный образ. Пассивное воздействие на амблиопичный глаз – раздражение сетчатки красным светом или наблюдение зерен на «Амблиоспекле» – применимо для детей младшего возраста. Но участие у таких детей высших корковых связей при этом снижается, что ухудшает качество лечения.

ПОСТАНОВКА ЗАДАЧИ

Целью настоящей работы была разработка специализированного электронного устройства, в котором объединяются преимущества существующих методик лечения амблиопии и пациенту предлагаются оптические образы с переменным цветом свечения и регулируемой длительностью предъявления тест-объектов (оптотипов) во времени. Дополнительным условием эффективности разрабатываемого устройства являлось обеспечение возможности регулирования яркости опто типа в соответствии с индивидуальными особенностями зрения пациента.

ТЕОРЕТИЧЕСКИЙ ПОДХОД К РАЗРАБОТКЕ МЕТОДИКИ

Световой поток, сформированный источником излучения в изображении, т. е. содержащий зрительную информацию, является раздражителем рецепторов сетчатки и нейронов. При этом возбуждающее действие на детекторные элементы вызывает ответную реакцию в виде зрительного образа. Одним из способов раздражения сетчатки является наблюдение светового поля через сканирующие его отверстия (рис. 1). Если расположить отверстия 1 во вращающемся непрозрачном диске 2 по спирали Архимеда, а за диском предъявлять подсвеченные тест-объекты 3, то будут последова-

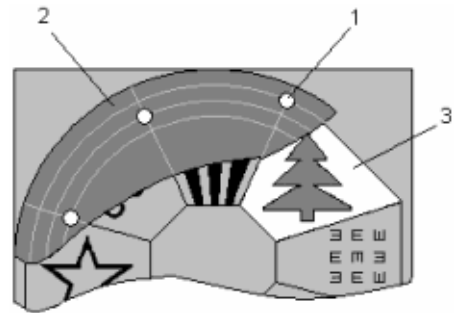


Рисунок 1 – Размещение вращающегося диска с отверстиями относительно панели с подсвеченным опто типом

Равномерное сканирование объекта достигается тем, что все отверстия на диске движутся с одинаковой угловой скоростью, проходя разные расстояния на своих окружностях за один и тот же промежуток времени. Кроме того, области сканирования освещенного сектора отверстиями с постоянной линейной скоростью смещаются к центру диска. При этом каждая область сканирования тест-объекта запоминается зрительным анализатором и дает возможность слить изображение в единое целое. Таким образом в связи с инертностью зрительного восприятия возникает эффект последовательного образа.

Распознавание глазом различных геометрических фигур способствует развитию форменного зрения.

При восстановлении зрения амблиопичного глаза эффективны упражнения на локализацию наблюдаемого объекта. Разрабатываемая методика предусматривает сканирование отверстиями восьми областей предъявления тест-объектов и распознавание пациентом программно предъявляемого опто типа в одной из них.

За счет изменения скорости вращения диска с отверстиями считывание тест-объектов для амблиопичного глаза происходит с нагрузкой. Уменьшение скорости вращения позволяет одновременно с локализацией проводить упражнения для тренировки верзионных движений глаза (переноса взора с одного предмета на другой, находящийся на таком же удалении) производимых по дуге, образованной при переносе взора с уходящего за область сектора отверстия на входящее в него.

Расположение панели с тест-объектами на некотором удалении от вращающегося диска приводит к развитию глубинного зрения. При этом тренировка аккомодационного аппарата глаза происходит за счет различия расстояний при наблюдении фрагмента опто типа, видимого сквозь отверстие в поверхности вращающегося диска.

Применение в качестве подсветки программно выбранного сектора красного, зеленого и белого цветов эффективно как реализация методики засветки сетчатки цветовыми полями. Переключение подсветки сектора

с определенной частотой и разным цветом свечения позволяет выполнить раздражение сетчатки модулированным световым потоком.

Изменение времени предъявления пациенту различных тестовых объектов приводит к развитию статической остроты зрения.

Все рассмотренные методики в настоящее время реализованы в отдельных приборах. Их отличает сложность конструкции, дороговизна, они применимы только в стационарных условиях и кабинетах охраны зрения детей. Перспективным является объединение наиболее эффективных методик в одном устройстве.

РАЗРАБОТКА УСТРОЙСТВА ДЛЯ ЛЕЧЕНИЯ АМБЛИОПИИ

С целью повышения эффективности лечения амблиопии пациентов всех возрастных групп разработано микроконтроллерное устройство «Плеофор» (рис. 2), которое создает пассивную нагрузку на сетчатку амблиопичного глаза и реализует активное сознательное участие в лечении пациента даже младшего возраста [2].

Задачей разрабатываемого устройства является повышение остроты зрения амблиопичного глаза до уровня 0,3–0,4, при котором возможно слияние изображений, что необходимо для проведения мероприятий по лечению косоглазия и восстановлению бинокулярного зрения.

Устройство позволяет получить на экране последовательный образ, производить засветку амблиопичного глаза слепящими полями белого, красного и зеленого цвета свечения, производить тренировку аккомодационного аппарата, фузионных резервов, распознавать тест-объекты с различной степенью нагрузки на орган зрения.

Устройство состоит из непрозрачного диска 1 с отверстиями, расположенными по спирали Архимеда. Перед диском 1 установлен прозрачный экран 2. За диском размещена сменная панель 4 с оптотипами, расположенными в восьми секторах. Диск 1 вращается электродвигателем 3. Оптотипы на панели 4 через па-

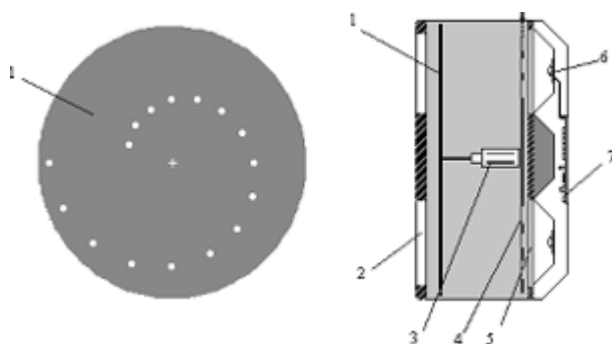


Рисунок 2 – Устройство для восстановления зрения амблиопичного глаза «Плеофор»

нель 5 подсвечиваются светодиодами 6 в определенном порядке. Порядок включения светодиодов и скорость вращения диска управляется микроконтроллером 7 по специально разработанной программе. Во время вращения диска 1 пациент последовательно видит фрагменты подсвеченного тест-объекта на панели 4 и должен, подсознательно объединяя их, получить полное изображение оптотипа, подсвечиваемого светодиодами 6 через панель 5 [3].

Способность слить зрительным анализатором промежуточные фрагменты тест-объекта (рис. 3), видимые через отверстия в непрозрачном диске, в непрерывное изображение, близка к способности сохранять последовательный образ, так как в основе обоих феноменов находится эффект инертности зрительного восприятия.

Получение полного изображения оптотипа, в результате слияния его фрагментов в единое целое, для ребенка намного легче, чем удержание последовательного образа. Это происходит подсознательно и пациент даже не замечает, как он получил полное изображение оптотипа. Скорость вращения диска с отверстиями изменяется регулировкой с пульта управления устройства и контролируется показаниями на дисплее. Использование управляемых светодиодов обеспечивает возможность предъявлять группы оптотипов в различных местах панели 4. Благодаря этому имеет место упражнение на локализацию.

Когда пациент наблюдает оптотип через подвижные отверстия диска 1, то, благодаря спиральному расположению отверстий, процесс предъявления фрагментов оптотипа происходит по окружности и к центру. Если использовать радиальную щель, то процесс слияния идет только по кругу, в направлении вращения. Таким образом, использование отверстий, расположенных по спирали, дает более эффективную нагрузку на амблиопичный глаз, чем использование радиальной щели.

Разработанная схема обеспечивает управление скоростью вращения диска с отверстиями, программное переключение позиции предъявляемых оптотипов, изменение яркости и цвета подсветки оптотипов, визуальный контроль выполнения программы. На функциональной

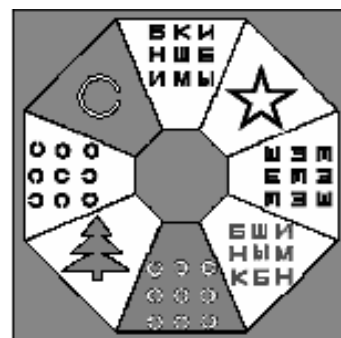


Рисунок 3 – Сменная панель с оптотипами

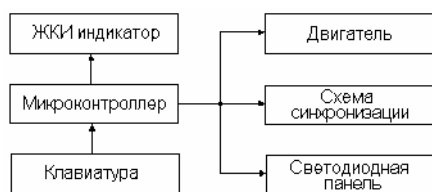


Рисунок 4 – Функциональная схема устройства для лечения амблиопии «Плеофор»

схеме (рис. 4) дисплей (ЖК индикатор) предназначен для контроля режимов лечебных мероприятий, проводимых на устройстве «Плеофор».

С клавиатуры поступает команда на микроконтроллер о начале выполнения программы, изменении режимов, занесении данных о времени проведения сеанса и режимах в память, остановке программы. Микроконтроллер осуществляет общее управление устройством в соответствии с выбранной пользователем программой.

Схема синхронизации обеспечивает включение на светодиодной панели заданного контроллером сектора с оптотипами при прохождении мимо него верхнего отверстия вращающегося диска.

Светодиодная панель состоит из восьми секторов, каждый из которых подсвечивается одним из трех светодиодов белого, красного или зеленого цвета свечения. Схема управления светодиодной панелью позволяет менять яркость свечения.

Схема электрическая принципиальная устройства для лечения амблиопии представлена на рис. 5.

Микроконтроллер AT90S2313 предназначен для управления процессом функционирования устройства.

Жидкокристаллический индикатор МТС – S16204XRGHS-01 отображает информацию о режимах программного обеспечения в текстовом формате, что удобно при выполнении лечебных мероприятий.

Оптопара DA1 в схеме синхронизации формирует сигнал и передает его на соответствующий вход мультиплексора DD3 при прохождении синхронизирующим отверстием вращающегося диска того сектора, адресный код которого задан микроконтроллером. Этот код передается и на дешифратор DD2, который выбирает группу подключаемых светодиодов в секторе. Получив информацию от микроконтроллера, мультиплексор, в соответствии с заданной программой, определяет цвет свечения сектора и посредством дешифратора DD1 подключает один из коммутирующих светодиодов транзисторов VT2–VT4. Светодиоды VD1–VD9 красного цвета свечения, VD10–VD17 белого цвета свечения, VD18–VD25 зеленого цвета свечения.

Постоянное напряжение 12 В от внешнего источника питания необходимо для обеспечения работы электродвигателя. Стабилизатор интегральный DA3 преобразует это напряжение в 5 В для питания интегральных микросхем и внутренних узлов схемы.

Двигатель постоянного тока имеет возможность менять скорость вращения при изменении потенциала коллектора транзистора VT1. Режим изменения скорости задается с выхода PWB/OC1 микроконтроллера, в соответствии с выбранной на клавиатуре программой.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Методика проведения тренировки следующая: ребенка сажают перед аппаратом на расстоянии 33 см от опто типа так, чтобы глаза пациента находились приблизительно на уровне центра диска. За диском с отверстиями предьявляют различные опто типы. Пациент должен назвать фигуру, буквы, направление разрыва в кольцах Ландольта и т. п. Тренировки проводят в течение 15 минут. Если у ребенка амблиопия с неправильной фиксацией, то сначала делают засвечивание большим безрефлексным офтальмоскопом, а затем проводят тренировку на разработанном устройстве «Плеофор».

Эффективность методики была подтверждена во второй детской больнице города Запорожье у двадцати пяти больных с амблиопией слабой и средней степени. Из них десять больных возрастом пять – шесть лет. Среди этих детей у пяти была амблиопия низкой степени, у трех средней и у двух высокой. Больные с низкой степенью амблиопии получали лечение только разработанной методикой на устройстве «Плеофор». Больные со средней и высокой степенью амблиопии сначала получали засветы методом Кюперса, а затем удержание последовательного образа было заменено считыванием опто типа по разработанной методике. Повышение остроты зрения наблюдалось в среднем на две десятых.

У тринадцати больных с разными формами астигматизма острота зрения на оба глаза была 0,3–0,4. После подбора очков острота зрения повысилась до 0,4–0,5. Не ожидая месячного срока ношения очков, после сужения зрачков, сразу был назначен десятидневный курс лечения новым методом, острота зрения в очках поднялась до 0,8–0,9. Эти дети быстро адаптировались к ношению очковой коррекции.

ВЫВОДЫ

Существенным достоинством предлагаемого аппарата является возможность одновременного выполнения формирования эффекта последовательного образа, изменения времени предьявления пациенту разных тестовых объектов для развития статической остроты зрения, распознавание глазом различных геометрических фигур для развития форменного зрения. Панель с тест-объектами располагается на расстоянии от вращающегося диска, что обеспечивает возможность тренировки глубинного зрения. За счет применения различных по цвету излучателей появляется возможность

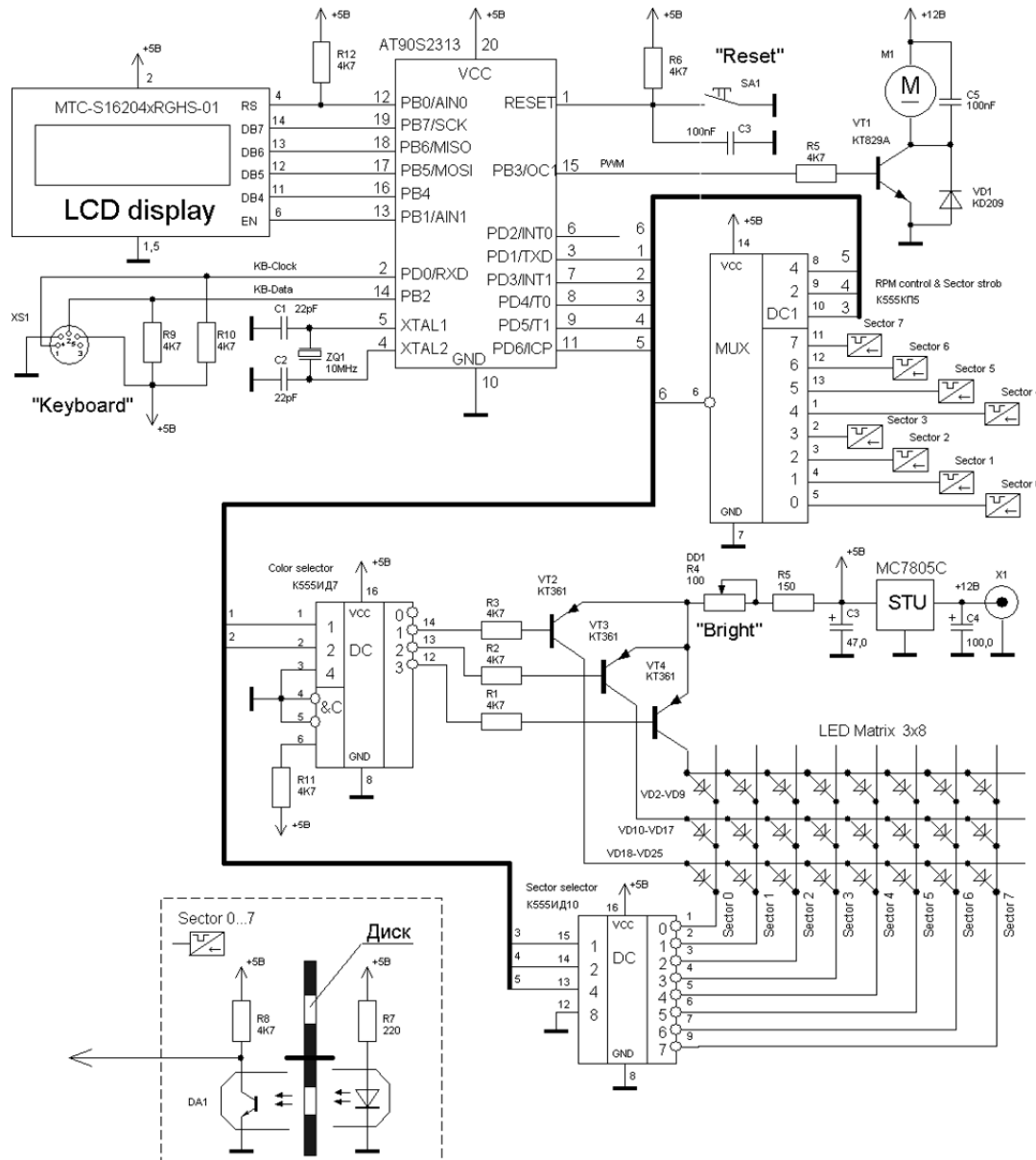


Рисунок 5 – Схема электрическая принципиальная устройства «Плеофор»

осуществлять терапию с разными цветами и яркостью предъявляемых оптоотипов. Микроконтроллерная система управления позволяет автоматически регулировать частоту предъявления оптоотипов, автоматически изменять цвет и позицию оптоотипа, а также применить элементы биоадаптивной обратной связи в процессе восстановления бинокулярного зрения.

ПЕРЕЧЕНЬ ССЫЛОК

1. Офтальмология: Учебник / Под ред. Е. И. Сидоренко. – М.: ГЭОТАР-МЕД, 2002. – С. 320–322.
2. Декларационный патент. Украина. Апарат для лікування амбліопії «Плеофор» / Поправка О. М., Вережкин Л. Л., Швець Є. Я., Червоний І. Ф., Сидоренко М. Г., Посунь-

- ко О. П., Багаев Р. А. (Україна). – № 61226; Заявл. 30.09.04; Опубл. 15.04.05; Бюл. № 4. – 6 с.
3. Вережкин Л. Л., Посунько А. П., Слюсаревская А. А. Апарат для лечения амблиопии // Праці ІХ науково-технічної конференції викладачів та студентів ЗДІА. – Частина 3. – Запоріжжя: ЗДІА. – 2004. – С. 61–64.

Надійшла 15.03.06

Представлені результати розробки нового пристрою, призначеного для відновлення зору амбліопічного ока, перспективність якого полягає в можливості проведення комплексу лікувальних заходів для дітей різних вікових груп.

The results of development of device of the sight of amblyopia eye perspective of which consists in possibility of lead through of complex of medical measures for the children of different age-dependent groups intended for renewal are represented.

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

MATHEMATICAL AND COMPUTER MODELLING

УДК 539.3

С. Н. Гребенюк, С. И. Гоменюк, Е. Л. Мизерная, В. В. Киричевский

МОДЕЛИРОВАНИЕ НАПРЯЖЕННО-ДЕФОРМИРОВАННОГО СОСТОЯНИЯ КОМПОЗИТОВ И КОНСТРУКЦИЙ НА ИХ ОСНОВЕ

Предложена методика решения некоторых задач механики композитов. Рассмотрены задачи вязкоупругого деформирования и деформирования с трещиной. Решение получено на основе метода конечных элементов. Данная методика реализована в виде пакета прикладных программ «МИРЕЛА+».

Композиционные материалы получили широкое распространение в технике благодаря высокой прочности и жесткости при наименьшей материалоемкости, что достигается изменением структуры армирования. Так как композиты создаются искусственно, то это позволяет варьировать их свойства. При проектировании конструкций необходимым является проведение прочностного расчета, что требует определения напряженно-деформированного состояния конструкции или ее составляющих. Разработка эффективных систем автоматизированного проектирования при создании конструкций из композиционных материалов позволяет значительно облегчить труд инженеров и исследователей, работающих в этой области.

Значительные математические трудности при расчете появляются из-за анизотропии свойств материала, различных особенностей конструкции, ее несовершенств, таких как поры, трещины, разрезы. Так как для очень многих композитов в качестве матрицы используется резина, то на сложность расчета влияют еще вязкоупругие свойства материала. Кроме того, конструкции часто имеют сложную геометрическую форму. Это приводит к тому, что использовать аналитические методы расчета становится невозможным. В последнее время очень широко применяются численные методы расчета с использованием компьютерной техники.

В настоящее время нет универсальной методики определения напряженно-деформированного состояния композиционных конструкций сложной геометрической формы в условиях вязкоупругого деформирования. Использование разработанных на данный момент теорий, описывающих вязкоупругое поведение композиционных материалов, возможно лишь для конструкций простой геометрической формы. Широкое их при-

© Гребенюк С. Н., Гоменюк С. И., Мизерная Е. Л., Киричевский В. В., 2006

менение сдерживается из-за громоздкости и сложности математических моделей и невозможности их решения большинством существующих методов или низкой точностью получаемых результатов.

В работе [1] изложен метод операторных цепных дробей, который позволяет в наиболее полной форме при соблюдении высокой степени точности получать решения многих задач линейной теории вязкоупругости для анизотропных тел как с фиксированными, так и подвижными границами. С помощью данного метода также исследовано вязкоупругое деформирование полубесконечного ортотропного тела при действии жесткого штампа [2].

С помощью метода операторных цепных дробей в работе [3] получено решение задачи о напряженном состоянии вязкоупругой ортотропной полуплоскости, нагруженной силой, нормальной к границе полуплоскости. На основе принципа Вольтерра и решения соответствующей задачи теории упругости определено поле напряжений в полуплоскости.

На основе наследственной теории Больцмана-Вольтерра получено аналитическое решение задачи о напряженно-деформированном состоянии полого ортотропного цилиндра под действием внутреннего давления, защемленного по наружной поверхности [4]. Выведены соотношения для перемещений, деформаций и напряжений в зависимости от времени.

В процессе изготовления либо эксплуатации в композиционных материалах возникают различного рода трещины, которые существенно влияют на распределение напряженно-деформированного состояния у вершины, а также на долговечность работы конструкции в целом [5].

Данная работа является развитием методики, предложенной в статьях [5–7]. Так как композиционный материал, как правило, состоит из матрицы и армирующих волокон, а частота армирования достаточно большая, то учесть влияние каждого волокна на напряженно-деформированное состояние всей конструкции не представляется возможным. Поэтому при моделировании напряженно-деформированного состояния композиционных материалов используют различные гипотезы, упрощающие расчет. Композит представляется в виде однородного анизотропного тела, а его упругие характеристики выражаются через упругие постоянные матрицы и армирующего материала в зависимости от их объемных долей.

Для описания вязкоупругих свойств матрицы и армирующих волокон воспользуемся наследственной теорией Больцмана-Вольтерра [8], согласно которой модуль упругости матрицы и модуль упругости армирующих волокон принимают вид

$$\tilde{E}_R = E_R - \int_0^t R_R(t-\tau)E_R(\tau)d\tau, \quad (1)$$

$$\tilde{E}_C = E_C - \int_0^t R_C(t-\tau)E_C(\tau)d\tau, \quad (2)$$

где $R_R(t-\tau)$ – разностное ядро релаксации материала матрицы, E_R – модуль упругости материала матрицы, $R_C(t-\tau)$ – разностное ядро релаксации материала армирующих волокон, E_C – модуль упругости материала армирующих волокон.

Связь между напряжениями и деформациями для анизотропного материала запишется на основе закона Гука:

$$\sigma^{ij} = C^{ijkl}\varepsilon_{kl},$$

где σ^{ij} – компоненты тензора напряжений, C^{ijkl} – компоненты тензора упругих постоянных, ε_{kl} – компоненты тензора деформаций.

Тензор упругих постоянных материала определяется через упругие постоянные структурных составляющих композита [9]. Заменяем интегральным оператором Вольтерра (1), (2) модуль упругости армирующих волокон и модуль упругости матрицы [8].

Согласно правилу смесей [10] продольный модуль упругости в системе координат армирования (рис. 1) рассчитывается по формуле

$$\tilde{E}_1 = \tilde{E}_C\psi_C + \tilde{E}_R(1-\psi_C),$$

где $\psi_C = \frac{\pi d_C^2}{4h_0}i_C$ – коэффициент армирования, характеризующий относительное объемное содержание волокон, d_C – диаметр волокон, h_0 – толщина армированного слоя, i_C – частота армирования.

Поперечные модули упругости \tilde{E}_2 и \tilde{E}_3 на основе наследственной теории Больцмана-Вольтерра определяются соотношениями

$$\tilde{E}_2 = \tilde{E}_3 = \frac{\tilde{E}_C\tilde{E}_R}{\psi_C\tilde{E}_R + (1-\psi_C)\tilde{E}_C}.$$

В системе армирования модули сдвига \tilde{G}_{12} и \tilde{G}_{13} принимают вид

$$\tilde{G}_{12} = \tilde{G}_{13} = \frac{\tilde{G}_C\tilde{G}_R}{\psi_C\tilde{G}_R + (1-\psi_C)\tilde{G}_C},$$

здесь \tilde{G}_C и \tilde{G}_R – модули сдвига материала волокон и матрицы:

$$\tilde{G}_C = \frac{\tilde{E}_C}{2(1+\nu_C)}, \quad \tilde{G}_R = \frac{\tilde{E}_R}{2(1+\nu_R)}, \quad (3)$$

где ν_C и ν_R – коэффициенты Пуассона материала волокон и матрицы, соответственно.

Соотношение для модуля сдвига \tilde{G}_{23} с учетом (3) примет вид

$$\tilde{G}_{23} = \tilde{G}_R \frac{K_R + \psi_C + (1 - \psi_C) \frac{\tilde{G}_R}{\tilde{G}_C}}{\psi_R K_R + (1 + \psi_C K_R) \frac{\tilde{G}_R}{\tilde{G}_C}},$$

где $K_R = 3 - 4\nu_R$.

Одним из наиболее универсальных методов расчета конструкций является метод конечных элементов. Данный метод позволяет рассчитывать конструкции сложной геометрической формы при любых условиях нагружения и закрепления конструкции.

Воспользовавшись конечно-элементным подходом на основе вариационного принципа Лагранжа [6], получим систему разрешающих уравнений с интегральными операторами. Для нахождения решения проведем дискретизацию по времени. На основе модифицированного метода Ньютона-Канторовича предложен алгоритм решения задачи механики композитов в условиях вязкоупругого деформирования.

Данная методика реализована в виде пакета прикладных программ «МРЕЛА+» [11, 12] для РС IBM на языке Фортран и позволяет при минимальных знаниях языка программирования и механики деформируемого твердого тела производить расчет напряженно-деформированного состояния конструкций из композиционных материалов в условиях вязкоупругого деформирования. В качестве ядер релаксации и ползучести могут использоваться ядра Работнова, Ржаницина и Колтунова.

Задача 1. Рассмотрим шарнирно закрепленную пластину из композиционного материала. Размеры пластины: $0,2 \times 0,05 \times 0,02$ м (рис. 1). Физические постоянные металлических волокон $E_C = 1277,5$ МПа, $\nu_C = 0,3$ и резиновой матрицы $E_R = 4,4$ МПа, $\nu_R = 0,4$. Диаметр волокон $d_C = 7 \cdot 10^{-4}$ м, частота армирования $i_C = 2900$ волокон/м, интенсивность нагрузки $\sigma = 10^{-2}$ МПа. Время принимается равным $t = 1$ с.

Так как резина относится к материалам с наследственными свойствами, а для металлов, в сравнении с резиной, наследственные свойства не столь ярко выражены, то расчет будем проводить в предположении отсутствия вязкоупругих свойств металла. В качестве ядра релаксации резины используем ядро Ю. Н. Работнова с реологическими характеристиками $\alpha = -(0,6)$, $\beta = 1,06$, $\gamma = 0,58$. Расчет проводился при различных углах армирования.

На рис. 2 показана зависимость максимальных напряжений от угла армирования. Как видно, наименьшее напряжение в упругом случае наблюдается при угле 30° , а в вязкоупругом – 45° , затем напряжение резко увеличивается и становится максимальным при угле 60° и снова уменьшается.

На рис. 3 показано распределение максимальных перемещений в зависимости от угла армирования. Сначала перемещения плавно возрастают и принимают наибольшее значение при угле 30° , затем резко уменьшаются и принимают минимальное значение при угле 90° .

Можно сделать следующие выводы. Наибольшее отклонение вязкоупругого решения от упругого наблюдается при угле армирования 0° и составляет 31–32 %. Минимальное отклонение наблюдается при угле 90°

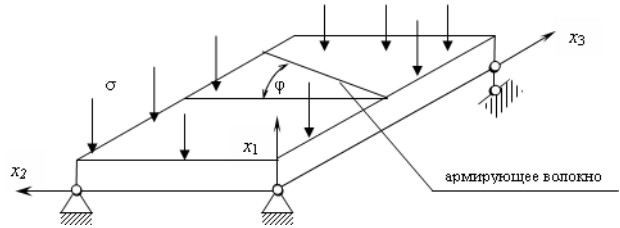


Рисунок 1 – Расчетная схема композитной пластины

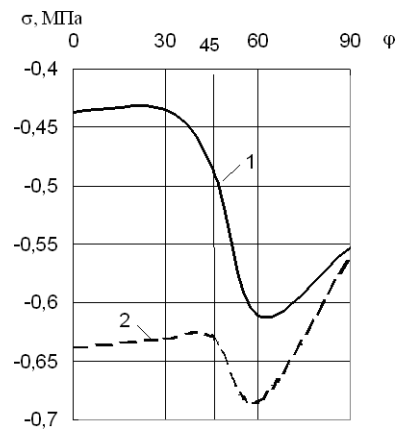


Рисунок 2 – Распределение максимальных напряжений в зависимости от угла армирования материала:

1 – упругое, 2 – вязкоупругое решение

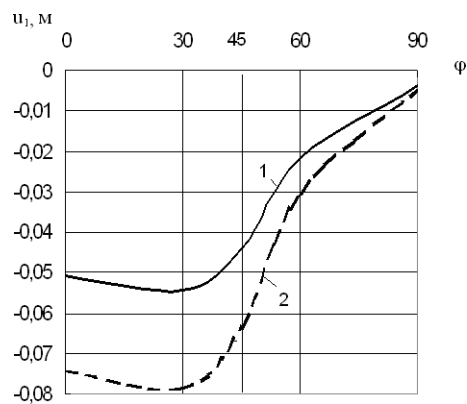


Рисунок 3 – Распределение максимальных перемещений в зависимости от угла армирования материала:

1 – упругое, 2 – вязкоупругое решение

и составляет для напряжений 2 %, для перемещений 18 %.

Другой важной задачей механики композитов является исследование напряженно-деформированного состояния конструкций вблизи вершины трещины. Для того чтобы более точно определить распределение на-

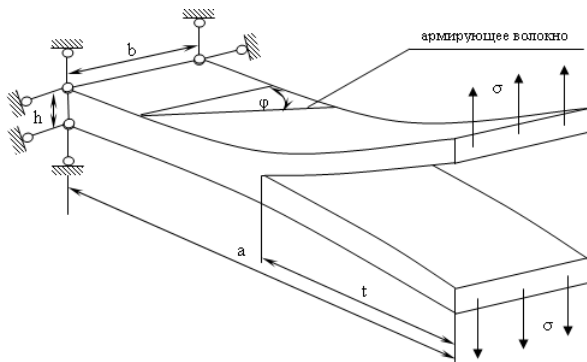


Рисунок 4 – Расчетная схема композитной балки

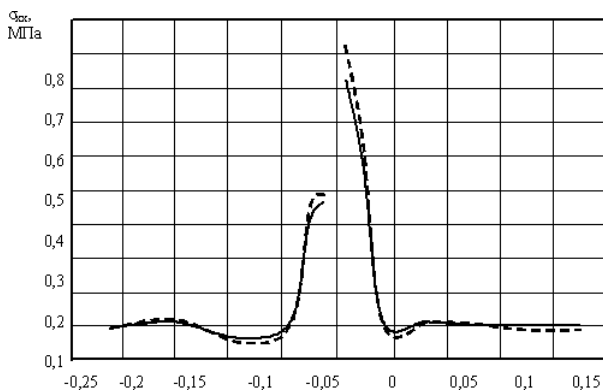


Рисунок 5 – Распределение напряжений по длине пластины при $\varphi = 0^\circ$:

----- упругое, ——— вязкоупругое решение

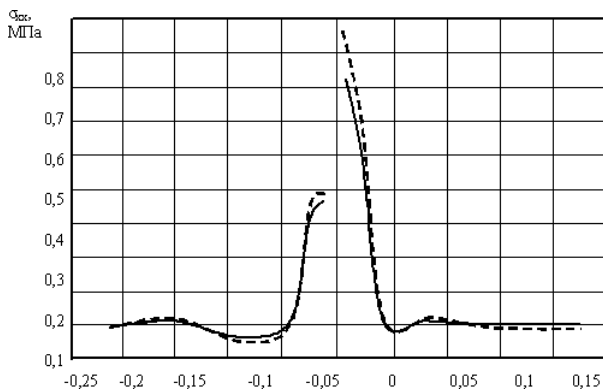


Рисунок 6 – Распределение напряжений по длине пластины при $\varphi = 90^\circ$:

----- упругое, ——— вязкоупругое решение

пряжений и деформаций вблизи вершины трещины, используют либо сгущение сетки трещины, либо сингулярный конечный элемент. Данную методику можно применить в рамках пакета прикладных программ «МРЕЛА+».

Задача 2. Рассмотрим ортотропную композиционную балку с трещиной. Расчетная схема приведена на рисунке 4. Размеры балки: длина $a = 0,5$ м; ширина $b = 0,06$ м; толщина $h = 0,065$ м, длина трещины $t = 0,25$ м.

Упругие постоянные материалов: материала матрицы – модуль упругости $E_R = 1277,5$ МПа, коэффициент Пуассона $\nu_R = 0,4$; материала волокон – модуль упругости $E_C = 3,4$ МПа, коэффициент Пуассона $\nu_C = 0,3$. Диаметр волокон $d_C = 7 \cdot 10^{-4}$ м, частота армирования $i_C = 2900$ волокон/м, приложенная нагрузка $\sigma = 0,0154$ МПа. Задача решалась в упругой и вязкоупругой постановке. В качестве ядра релаксации резины используем ядро Ю. Н. Работнова с реологическими характеристиками $\alpha = -0,6$, $\beta = 1,06$, $\gamma = 0,58$.

Напряженно-деформированное состояние для углов армирования и около вершины трещины приведены на рисунках 5 и 6 соответственно.

Таким образом, около вершины трещины растягивающие вязкоупругие напряжения превышают упругие приблизительно на 14 % при угле армирования 0° и приблизительно на 20 % при угле армирования 90° .

ПЕРЕЧЕНЬ ССЫЛОК

1. Каминский А. А. Исследование деформирования анизотропных вязкоупругих тел // Прикладная механика. – 2000. – Т. 36, вып. 11. – С. 39–63.
2. Каминский А. А., Черноиван Ю. А. Вязкоупругое деформирование ортотропного тела (полуплоскости) под действием жесткого штампа // Прикладная механика. – 2000. – Т. 36, вып. 7. – С. 81–91.
3. Каминский А. А., Черноиван Ю. А. Напряженное состояние вязкоупругой ортотропной полуплоскости, нагруженной сосредоточенной силой // Прикладная механика. – 2000. – Т. 36, вып. 2. – С. 124–130.
4. Киричевский В. В., Гребенюк С. Н., Мизерная Е. Л. Напряженно-деформированное состояние полого ортотропного цилиндра в условиях вязкоупругого деформирования // Новые материалы и технологии в металлургии и машиностроении. – 2005. – № 2. – С. 76–80.
5. Гребенюк С. Н., Киричевский В. В. Напряженно-деформированное состояние композитных конструкций с трещинами // Вісник Східноукраїнського національного університету. – 2006. – № 3 (97). – С. 53–58.
6. Гребенюк С. Н., Киричевский В. В., Толоч В. А. Определение напряженно-деформированного состояния композиционных материалов в условиях вязкоупругого деформирования // Труды двенадцатого симпозиума «Проблемы шин и резинокордных композитов», 15–19.10.2001. – Изд-во НИИШП, Москва, 2001. – Т. 1. – С. 147–151.
7. Киричевський В. В., Гребенюк С. М. Розрахунок багатошарової композитної пластины в умовах в'язкопружної деформації // Вісник Східноукраїнського національного університету. – 2002. – № 3 (49). – С. 95–100.
8. Киричевский В. В. Метод конечных элементов в механике эластомеров. – К.: Наук. думка, 2002. – 655 с.
9. Киричевский В. В., Дохняк Б. М., Карпушин А. Д. Матрица жесткости пространственного конечного элемента для исследования конструкций из композиционных

- матеріалов // Вісник Східноукраїнського державного університету. – 1999. – № 3 (18). – С. 109–116.
10. Композиционные материалы. Справочник // Под общ. ред. Д. М. Карпиноса. – К.: Наукова думка, 1985. – 592 с.
 11. Метод конечных элементов в вычислительном комплексе «MIRELA+» / Киричевский В. В., Дохняк Б. М., Козуб Ю. Г., Гоменюк С. И., Киричевский Р. В., Гребенюк С. Н. / Под ред. В. В. Киричевского. – К.: Наукова думка, 2005. – 408 с.
 12. Гребенюк С. Н., Киричевский В. В., Гоменюк С. И. Вязкоупругое деформирование конструкций из композиционного материала // Вісник Східноукраїнського національного університету. – 2003. – № 12 (70). – С. 226–231.

Надійшла 2.06.06

Запропоновано методику розв'язання деяких задач механіки композитів. Розглянуто задачі в'язкоупругого деформування та деформування з тріщиною. Розв'язок отримано на основі метода скінченних елементів. Дану методику реалізовано у вигляді пакету прикладних програм «MIRELA+».

The method of the solution of some problems of a mechanics of composites is offered. The problems of viscoelastic deformation and deformation with a crack are reviewed. The solution is obtained on the basis of a finite element method. The given method realised as an application package «MIRELA+».

УДК 539.3

Р. В. Киричевский, Е. С. Решевская, В. М. Тархова, Е. В. Прокопенко,
В. В. Киричевский

ИССЛЕДОВАНИЕ ЗАДАЧ МЕХАНИКИ ДЕФОРМИРУЕМОГО ТВЕРДОГО ТЕЛА В ВЫЧИСЛИТЕЛЬНОМ КОМПЛЕКСЕ «MIRELA+»

Описываются классы задач механики деформируемого твердого тела, решение которых можно получить при помощи вычислительного комплекса «MIRELA+», основанного на методе конечных элементов (МКЭ). Решены задачи контакта эластомерных виброизоляторов, задачи определения напряженно-деформированного состояния прямоугольной плиты с трещиной. Вычислены коэффициенты интенсивности напряжений вдоль фронта трещины.

ВВЕДЕНИЕ

Для решения задач механики деформируемого твердого тела одним из широко используемых методов является метод конечных элементов. Применение этого метода позволяет производить расчет конструкций на прочность и жесткость со сложной геометрией, различными схемами нагружения, а также учитывать специфические свойства материалов деталей.

В связи с развитием численных методов решения различных классов задач теории упругости появилась потребность в универсальных системных комплексах, реализующих данные методы на высоком и эффективном уровне. Такие системы должны обладать гибкостью, позволять пользователю вносить изменения на уровне метода расчета. Среди первых программных комплексов, разработанных для исследования металлических конструкций можно выделить: ASKA [1], TITUS [2], GENESYS [3], SESAM-69 [4], NASTRAN [5], TURBAN [6], ПРОЧНОСТЬ [7], МИРАЖ [8] и др. Вычислительных систем, предназначенных для опреде-

ления напряжения и деформаций эластомерных конструкций, сравнительно немного: SARMAS [9], EFE-SYS [10], ADINA [11], UPRUG [12], MIRELA+ [13].

В программном комплексе «MIRELA+» реализован расчет эластомерных конструкций на основе метода конечных элементов, развитого применительно к особенностям эластомеров. Процесс расчета состоит из трех взаимосвязанных этапов. На первом этапе реализуется работа препроцессора: задается конечно-элементная дискретизация расчетной схемы, топология и граничные условия исследуемого объекта, физические характеристики материала. На втором этапе работа процессора включает в себя: расчет матрицы жесткости конечных элементов, построение разрешающей системы уравнений и ее решение. На третьем этапе реализуется работа постпроцессора: находится вектор узловых перемещений, на основе которого определяются поля деформаций и напряжений и их значения заносятся в файл.

Вычислительный комплекс «MIRELA+» состоит из нескольких подсистем, решающих задачи теории упругости различных классов. В нем реализованы задачи линейной и нелинейной теории упругости, вязкоупругости и теплопроводности, задачи нелинейной механики разрушения, определения температуры диссипативного разогрева, линейной механики разрушения, расчета конструкций из композитов, динамики, долговечности и задач контактной механики.

© Киричевский Р. В., Решевская Е. С., Тархова В. М., Прокопенко Е. В., Киричевский В. В., 2006

**РЕШЕНИЕ КОНТАКТНЫХ ЗАДАЧ
В МЕХАНИКЕ ЭЛАСТОМЕРОВ
В СИСТЕМЕ «МІРЕЛА+»**

Развитие машиностроения, автомобильного транспорта и других отраслей промышленности приводит к необходимости создания принципиально новых конструкций, в которых существенно важными являются требования минимизации энерго- и металлоемкости машин, а также повышения надежности и долговечности техники. Вместе с тем, режимы работ современных конструкций приводят к увеличению усталостных эффектов, износу. Для решения возникающих проблем можно внедрять новые конструкции машин и технологии, применять новые материалы, однако, неизменной остается на сегодняшний день концепция введения в структурную схему машин полимерных и эластомерных звеньев, как-то виброизоляторы, износостойкие покрытия, упругие прокладки, демпферы и т. д. [14].

По высокоэластичным и упругим свойствам механические характеристики резины подразделяются на равновесные, действующие при установившемся стационарном процессе, и кинетические, описывающие релаксационные процессы. Практически считается, что резина – несжимаемый материал и ее коэффициент Пуассона $\nu = 0,5$. Гипотеза о несжимаемости материала не может быть применена для определенных классов задач, в случае, если они сформулированы для эластомерных материалов – это исследование больших деформаций эластомеров, расчет тонкослойных резино-металлических элементов, задачи учета контакта и др.

Для учета особенностей эластомеров в системе «МІРЕЛА+» используется моментная схема конечного элемента. Используемый вариант МКЭ приводит к матрицам жесткости КЭ, которые учитывают жесткие смещения, эффект ложного сдвига, слабую сжимаемость эластомеров. Это достигается путем сопоставления раз-

ложений полей перемещений, деформаций и функции изменения объема (функции объемной деформации). Так, к примеру, коэффициенты разложения линейных деформаций в выражении для разложения функции изменения объема вносят существенные погрешности в решения. Отбрасывая или минимизируя данные члены разложения, тем самым обходим эффекты, связанные с вырождением задачи и замедлением сходимости МКЭ. Порядок разложения деформаций и функций изменения объема находится в строгом соответствии с порядком аппроксимации функции перемещений и выбирается с таким расчетом, чтобы исключить все компоненты деформаций, реагирующие на жесткие смещения и эффект ложного сдвига, и все компоненты функции изменения объема, реагирующие на слабую сжимаемость эластомера [15].

В частности, используя средства приведенного вычислительного комплекса можно найти решение контактных задач для эластомерных амортизаторов с различной формой поверхности нагружения. Расчетные схемы и условия нагружения приведены для исследуемых виброизоляторов ВН-100 (см. рис. 1, а) и ВРВ-100 (см. рис. 1, б).

Модель контактного взаимодействия описывалась при помощи условия непроникновения, полученная величина зоны контакта уточняется итерационно [16].

Алгоритм решения контактной задачи заключается в следующем. На первом шаге нагружения вычисляется матрица жесткости $[K]$. Текущее значение параметров нагрузки определяется следующим образом:

$$P_n = P_{n-1} + \Delta P_n. \quad (1)$$

Из системы линейных алгебраических уравнений

$$[K]\{u_{n-1}\} = \{P_{n-1}\} \quad (2)$$

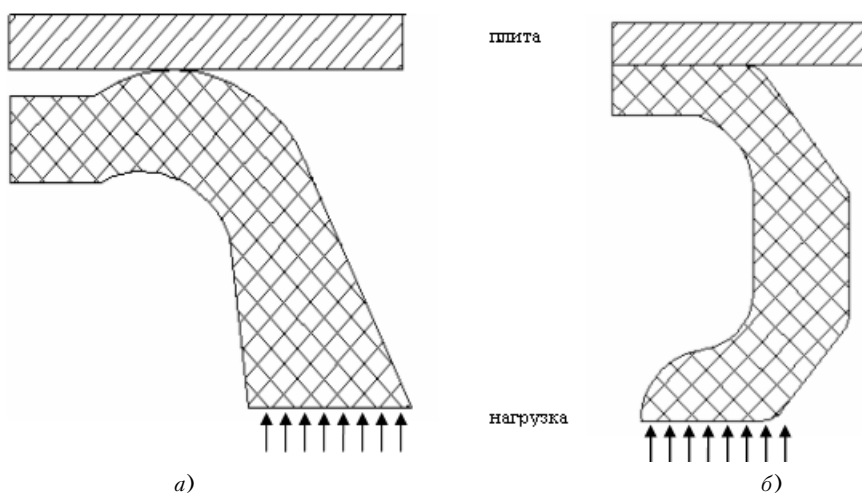


Рисунок 1

находится вектор перемещений. Начальное приближение искомого решения $\{\Delta u_{n(1)}\}$ находится с помощью экстраполяционной формулы

$$\{\Delta u_{n(1)}\} = \{u_{n-1}\} \frac{\Delta P_n}{\Delta P_{n-1}}. \quad (3)$$

Затем задаются краевые условия непроникания контактирующих тел. К узловым точкам, которые пересекают границу контакта после нагружения амортизатора, прикладывается вектор дополнительной нагрузки, который определяется новым вектором перемещений

$$\{\Delta u_{n(2)}\} = \{\Delta u_{n(1)}\} - \{\Delta v_{n(1)}\}, \quad (4)$$

где $\{\Delta u_{n(1)}\}$ – вектор перемещения точек без учета граничных условий, $\{\Delta u_{n(2)}\}$ – новый вектор перемещений, $\{\Delta v_{n(1)}\}$ – вектор перемещений от дополнительной загрузки, который определяется по формуле

$$\{\Delta v_{n(1)}\} = \{\Delta u_{n(1)}\} \sqrt{\frac{(c - xn_2)^2 \left(\left(\frac{xn_3 - x_3}{xn_2 - x_2} \right)^2 + 1 \right)}{(xn_3 - x_3)^2 + (xn_2 - x_2)^2}}, \quad (5)$$

где прямая $x_2 = c$ – граница раздела двух тел, x_2, x_3 – координаты узловых точек конструкции, xn_2, xn_3 – координаты узловых точек после нагружения.

Приближенное значение перемещений $\{\Delta u_{n(2)}\}$ подставляется в линейные уравнения, и определяется вектор узловых невязок $\{R_{n(2)}\}$, численной характеристикой которого является сумма квадратов компонент узловых невязок $\{R^*\}$, позволяющая судить о сходимости получаемых решений.

Результаты исследований контактного взаимодействия эластомерных виброизоляторов с плитами различной массы представлены в таблице 1 и рис. 2 и 3.

Таблица 1

ВН-100		ВРВ-100	
Масса плиты, кг	Величина зоны контакта, мм	Масса плиты, кг	Величина зоны контакта, мм
150	6,487	100	29,7
300	10,614	200	38,4
450	14,969	300	39,7

В системе «МИРЕЛА+» имеется возможность графического представления, как исходных данных, так и результатов расчета, в виде трехмерных изображений. Распределения перемещений (см. рис. 4) и напряжений (см. рис. 5) при контактной нагрузке на конечно элементной модели амортизаторов изображены ниже.

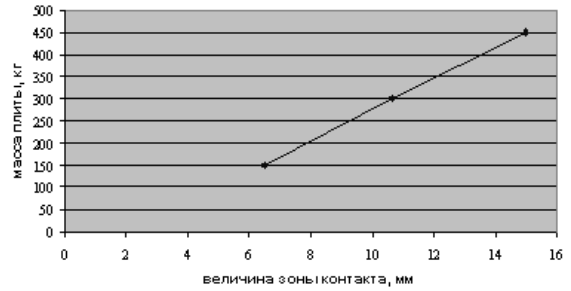


Рисунок 2

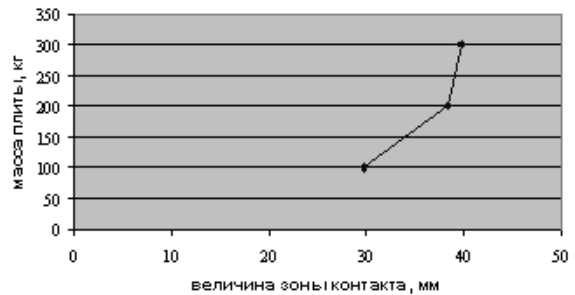


Рисунок 3

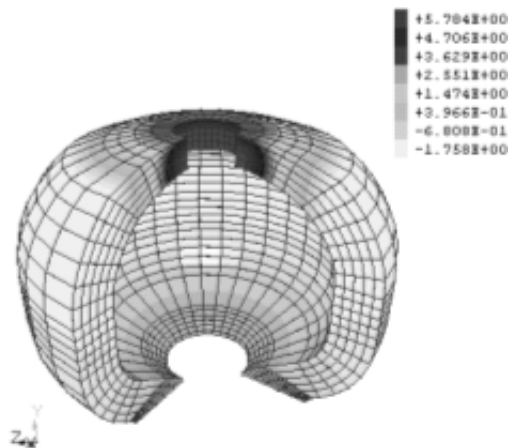


Рисунок 4

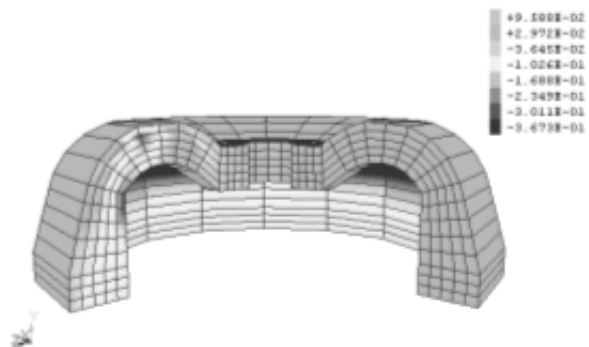


Рисунок 5

Как видно из рисунков 2 и 3, получены линейные зависимости величины зоны контакта от действующих нагрузок, что подтверждает предположение о том, что рассматриваемые задачи являются задачами теории упругости. Отличия величин зон контакта может быть объяснено различием форм виброизоляторов и соответствующими им режимами нагружения.

**ОПРЕДЕЛЕНИЕ КОЭФФИЦИЕНТОВ
ИНТЕНСИВНОСТИ НАПРЯЖЕНИЙ
В ПРЯМОУГОЛЬНОЙ ПЛИТЕ С ТРЕЩИНОЙ
В РАМКАХ ВЫЧИСЛИТЕЛЬНОГО
КОМПЛЕКСА «МІРЕЛА+»**

Задание трещины при решении задач механики разрушения методом конечных элементов можно осуществлять несколькими способами. Трещина может моделироваться при помощи специальных элементов, у которых закладывается часть объема с нулевыми упругими характеристиками по сравнению с материалом конструкции.

При решении симметричных задач, когда трещина лежит в плоскости симметрии, удобно задавать трещину с помощью краевых условий. При этом для узлов, лежащих перед фронтом трещины, перемещения в направлении, перпендикулярном плоскости симметрии, запрещаются. А узлы, лежащие на берегах трещины, могут перемещаться в этом направлении.

При решении несимметричных задач трещина задается в виде вырезанной полости нулевого объема, при этом узлам, которые лежат на берегах трещины, присваиваются различные порядковые номера, а узлам, которые лежат за фронтом трещины, одинаковые порядковые номера – такая область называется вырожденной. При моделировании продвижения трещины эти параметры изменяются [17].

Одним из основных методов определения коэффициентов интенсивности напряжений является прямой метод – непосредственно с помощью формул:

$$\begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \frac{\sqrt{2r}}{\mu\sqrt{\pi}} \times \begin{bmatrix} K_{III} \sin \frac{\theta}{2} \\ \frac{K_I}{8} \begin{bmatrix} (5-8\nu) \cos \frac{\theta}{2} \cos \frac{3\theta}{2} \\ (7-8\nu) \sin \frac{\theta}{2} \sin \frac{3\theta}{2} \end{bmatrix} + \frac{K_{II}}{8} \begin{bmatrix} (9-8\nu) \sin \frac{\theta}{2} \sin \frac{3\theta}{2} \\ (-3+8\nu) \cos \frac{\theta}{2} \cos \frac{3\theta}{2} \end{bmatrix} \end{bmatrix}, \quad (6)$$

$$\begin{Bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{13} \\ \sigma_{12} \end{Bmatrix} = \frac{1}{\sqrt{2\pi r}} \times \begin{bmatrix} K_I \begin{bmatrix} 2\nu \cos \frac{\theta}{2} \\ \cos \frac{\theta}{2} \left(1 - \sin \frac{\theta}{2} \sin \frac{3\theta}{2}\right) \\ \cos \frac{\theta}{2} \left(1 + \sin \frac{\theta}{2} \sin \frac{3\theta}{2}\right) \\ \sin \frac{\theta}{2} \cos \frac{\theta}{2} \cos \frac{3\theta}{2} \end{bmatrix} + K_{II} \begin{bmatrix} -2\nu \sin \frac{\theta}{2} \\ -\sin \frac{\theta}{2} \left(2 + \cos \frac{\theta}{2} \cos \frac{3\theta}{2}\right) \\ \sin \frac{\theta}{2} \cos \frac{\theta}{2} \cos \frac{3\theta}{2} \\ \cos \frac{\theta}{2} \left(1 - \sin \frac{\theta}{2} \sin \frac{3\theta}{2}\right) \end{bmatrix} \\ K_{III} \begin{bmatrix} \cos \frac{\theta}{2} \\ -\sin \frac{\theta}{2} \end{bmatrix} \end{bmatrix}, \quad (7)$$

где r, θ – полярные координаты в плоскости, перпендикулярной фронту трещины.

Можно использовать методику определения коэффициентов интенсивности напряжений, предложенную в [18], согласно которой коэффициенты интенсивности напряжений вычисляются в определенных точках и затем усредняются. При этом большое значение имеет форма и размеры области, в которой берутся эти точки, а также степень сгущения сетки разбиения.

В рамках вычислительного комплекса «МІРЕЛА+» [13] определено напряженно-деформированное состояние плиты с угловой трещиной (см. рис. 6, а) и плиты с поверхностной трещиной (см. рис. 6, б).

Приведем решение задачи о растяжении плиты с угловой трещиной (см. рис. 7), размеры которой соответственно равны $l = 8$ м, $m = 8$ м, $h = 16$ м. Размеры трещины $a = 0,055$ м, $b = 0,04$ м; на концах плиты приложена растягивающая нагрузка 0,1 МПа; модуль Юнга $E = 2,1$ МПа; коэффициент Пуассона $\nu = 0,3$.

Распределение напряжений σ_{33} вдоль оси x^3 приведено на рис. 8.

На рис. 9 приведено распределение коэффициентов интенсивности напряжений K_I вдоль фронта трещины.

Аналитическая формула вычисления коэффициентов интенсивности напряжений имеет вид [19]

$$K_I = \frac{\sigma\sqrt{\pi b}}{E(k)} F_s \left[\frac{b}{a}, \frac{b}{t}, \frac{a}{W}, \varphi \right],$$

$$E(k) = \left[1 + 1,464 \left(\frac{b}{a} \right)^{1,65} \right]^{\frac{1}{2}},$$

$$F_s = \left[M_1 + M_2 \left(\frac{b}{t} \right)^2 + M_3 \left(\frac{b}{t} \right)^4 \right] g f_\phi f_W,$$

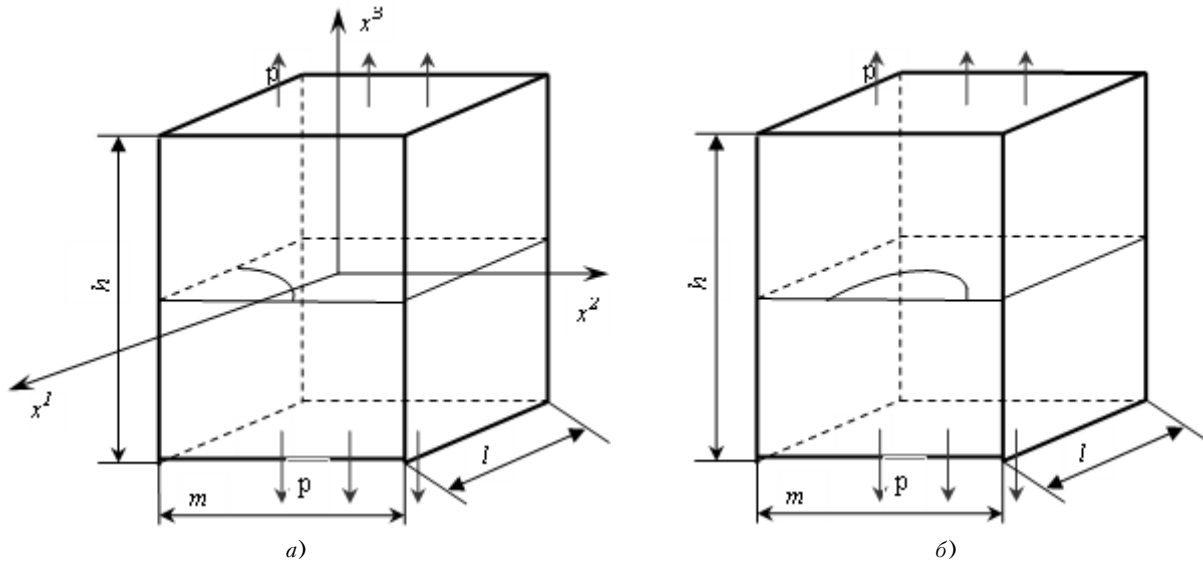


Рисунок 6

$$\begin{aligned}
 M_1 &= 1,13 - 0,09 \frac{b}{a} \text{ при } \frac{b}{a} \leq 1, \\
 M_2 &= -0,54 + 0,89 \left(0,2 + \frac{b}{a}\right)^{-1} \text{ при } \frac{b}{a} \leq 1, \\
 M_3 &= 0,5 + \left(0,65 + \frac{b}{a}\right)^{-1} + 14 \left[1 - \frac{b}{a}\right]^{24} \text{ при } \frac{b}{a} \leq 1, \quad (8)
 \end{aligned}$$

$$g = 1 + \left(0,1 + 0,35 \left(\frac{b}{l}\right)^2\right) (1 - \sin \varphi)^2,$$

$$f_\varphi = \left[\left(\frac{b}{a}\right)^2 \cos^2 \varphi + \sin^2 \varphi\right]^{\frac{1}{4}} \text{ при } \frac{b}{a} \leq 1,$$

$$\begin{aligned}
 f_W &= \left[\sec\left(\frac{\pi a}{2W}\right) \sqrt{\frac{b}{t}}\right]^{\frac{1}{2}}, \\
 (t = l, W = m).
 \end{aligned}$$

Угол φ , входящий в (8), определяется согласно рис. 10.

Далее в рамках вычислительного комплекса «МИРЕЛА» [13] определено напряженно-деформированное состояние плиты с поверхностной трещиной (см. рис. 11). Приведем решение задачи о растяжении плиты, размеры которой соответственно равны $l = 8$ м, $m = 8$ м, $h = 16$ м. Размеры трещины $a = 0,055$ м, $b = 0,04$ м; на концах плиты приложена растягивающая нагрузка $0,1$ МПа; модуль Юнга $E = 2,1$ МПа; коэффициент Пуассона $\nu = 0,3$.

На рис. 12 приведено распределение коэффициентов интенсивности напряжений K_I вдоль фронта трещины.

Аналитическая формула вычисления коэффициентов интенсивности напряжений имеет вид [19]

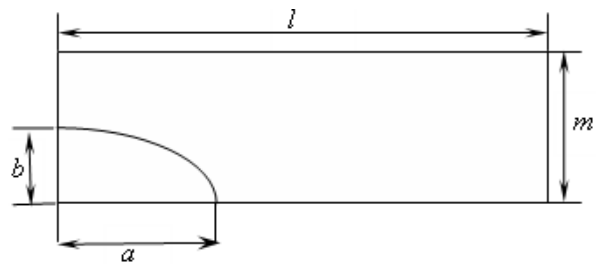


Рисунок 7

$$K_I = \frac{\sigma \sqrt{\pi b}}{E(k)} F_e \left[\frac{b}{a}, \frac{b}{t}, \varphi\right],$$

$$E(k) = \left[1 + 1,464 \left(\frac{b}{a}\right)^{1,65}\right]^{\frac{1}{2}},$$

$$F_c = \left[M_1 + M_2 \left(\frac{b}{t}\right)^2 + M_3 \left(\frac{b}{t}\right)^4\right] g_1 g_2 f_\varphi,$$

$$M_1 = 1,08 - 0,03 \frac{b}{a} \text{ при } \frac{b}{a} \leq 1,$$

$$M_2 = -0,44 + 1,06 \left(0,3 + \frac{b}{a}\right)^{-1} \text{ при } \frac{b}{a} \leq 1,$$

$$M_3 = -0,5 + 0,25 \frac{b}{a} + 14,8 \left[1 - \frac{b}{a}\right]^{15} \text{ при } \frac{b}{a} \leq 1, \quad (9)$$

$$g_1 = 1 + \left(0,08 + 0,4 \left(\frac{b}{t}\right)^2\right) (1 - \sin \varphi)^3,$$

$$g_2 = 1 + \left(0,08 + 0,15 \left(\frac{b}{t}\right)^2\right) (1 - \cos \varphi)^3,$$

$$f_\varphi = \left[\left(\frac{b}{a}\right)^2 \cos^2 \varphi + \sin^2 \varphi\right]^{\frac{1}{4}} \text{ при } \frac{b}{a} \leq 1, \quad (t = l, W = m).$$

Угол φ , входящий в (9), определяется согласно рис. 10.

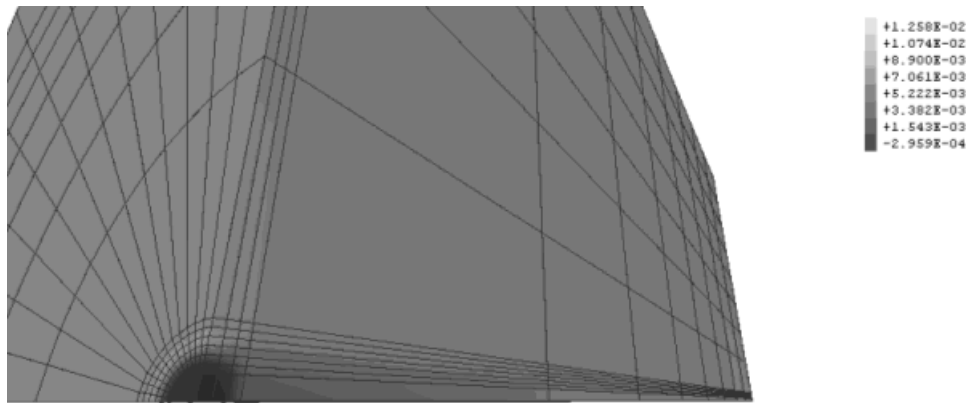


Рисунок 8

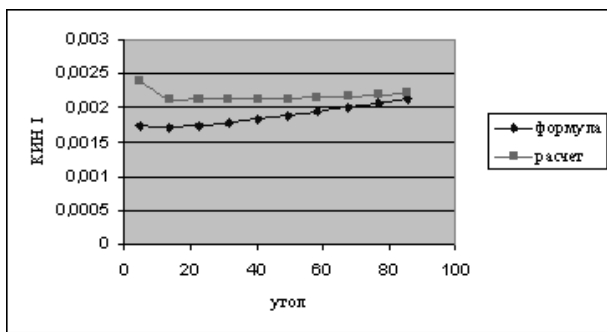


Рисунок 9

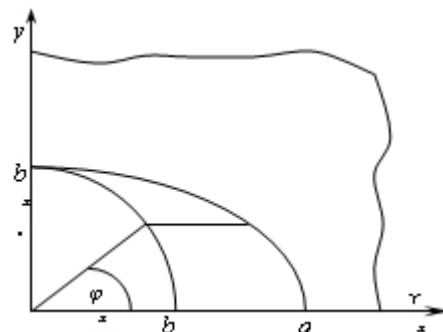


Рисунок 10

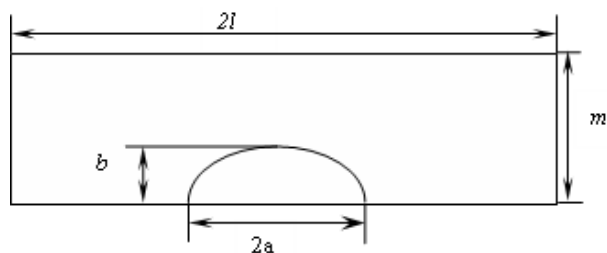


Рисунок 11

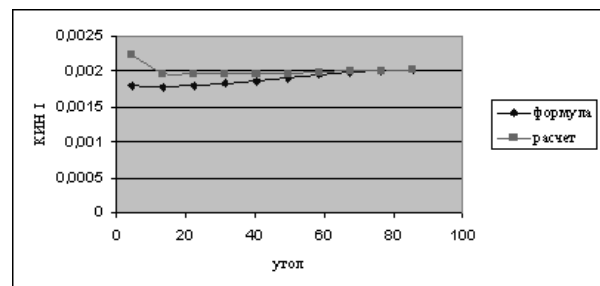


Рисунок 12

ВЫВОДЫ

По расчетным моделям контактных задач в процессе вычисления в программном комплексе «МИРЕЛА+» получены линейные зависимости величины зоны контакта от действующих нагрузок, что подтверждает предположение о том, что рассматриваемые задачи являются задачами теории упругости. Отличия величин зон контакта может быть объяснено различием форм виброизоляторов и, соответственно, площадью поверх-

ности, вступающей в контакт, а также режимами нагружения.

Результаты расчета задач механики разрушения показывают, что данная методика адекватно описывает распределение напряжений и деформаций у вершины трещины и позволяет определить коэффициенты интенсивности напряжений в задаче о центральной эллиптической трещине в прямоугольной плите с достаточной точностью.

ПЕРЕЧЕНЬ ССЫЛОК

1. Скрим Э., Рой Дж. Р. Автоматическая система кинематического анализа // Расчет упругих конструкций с использованием ЭВМ / Пер. с англ. Под ред. А. П. Филина: В 2-х т. – Л.: Судостроение, 1974. – Т. 2. – С. 36–37.
2. Launay P. et al. The three-dimensional thermoelastic computer code «TITUS» // Prepr. 1 st. Int. Conf. Struct. Mech. React. Technol. Berlin, 1971. – Amsterdam: Amsterdam e. a., 1971. – 5. – P. M5-4/1–M5-4/21.
3. Ruoff G., Stein E. The development of general purpose programs and systems // Structural Mechanics Computer Programs. Symposium held at the University of Maryland, June, 1974. – Univ. Press of Virginia Charlottesville, 1974. – P. 703–719.
4. Araldsen P. O. The application of the superelement method in analysis and design of ship structure and machinery components // National Symp. On Computerized structural Analysis and Design, Norway, March. – 1972. – P. 2–93.
5. Butler T. G., Michel D. NASTRAN. A summary of the functions and capabilities of the NASA structural analysis computer system. – Washington, 1971. – 22 p.
6. Mlejer H. P., Mai M. M. TURBAN. Version 1.0. Ein Speziale Programmiersystem zur linearen elastischen Analyse von rotierender Schalentragwerken. Anleitung für den Benutzer // ISD. Ber. – 1970. – № 3. – S. 174–179.
7. Комплекс прикладных программ расчета оболочечных конструкций. Листинг. ПРОЧНОСТЬ-75. Система математического обеспечения расчетов пространственных конструкций / Гончаренко И. Е., Завьялов Г. Г., Киричевский В. В. и др. – К.: Республиканский фонд алгоритмов и программ АН УССР, 1975. – Т. 5. – 357 с.
8. Городецкий А. С. К расчету комбинированных систем методом конечных элементов // Сопротивление материалов и теория сооружений. – 1972. – Вып. 16. – С. 123–125.
9. Gupta B. P., Finney R. H. Application of finite element method to the analysis of high-capacity laminated elastomeric parts // Exp. Mech. – 1980. – 2, № 3. – P. 18–31.
10. Dunger R. EFESYS – an engineering finite element system // Eng. Soft-ware Proc. 1-st. Int. Conf., Southampton. – 1979. – P. 94–119.
11. Bathe K. I. Nonlinear finite element analysis and ADINA // 3-Comput. And Struct. – 1983. – 17, № 5–6. – P. 625–631.
12. Хархурим И. Я. Метод конечных элементов в нелинейной механике эластомеров. Механика эластомеров. Науч. труды Кубан. гос. ун-та. – 1980. – Т. 3. – С. 13–23.
13. Метод конечных элементов в вычислительном комплексе «МИРЕЛА+» / Киричевский В. В., Дохняк Б. М., Козуб Ю. Г., Гоменюк С. И., Киричевский Р. В., Гребенюк С. Н. – К.: Наук. думка, 2005. – 403 с.
14. Дырда В. И., Чижик Е. Ф. Резиновые детали в машиностроении. – Днепропетровск: Полиграфист, 2000. – 581 с.
15. Киричевский В. В. Метод конечных элементов в механике эластомеров. – К.: Наук. думка, 2002. – 655 с.
16. Киричевский В. В., Гребенюк С. Н., Тархова В. М. Решение контактной задачи механики эластомеров с учетом условия проскальзывания // Вісник дніпропетровського університету. – 2005. – Вып. 9. – С. 112–118.
17. Метод конечных элементов: теория, алгоритмы, реализация / В. А. Толок, В. В. Киричевский, С. И. Гоменюк, С. Н. Гребенюк, Д. П. Бувайло. – К.: Наук. думка, 2003. – 316 с.
18. Баженов В. А., Гуляр А. И., Сахаров А. С., Топор А. Г. Полуаналитический метод конечных элементов в механике деформированных тел. – К.: НДІБМ, 1993. – 376 с.
19. Справочник по коэффициентам интенсивности напряжений: В 2-х т. / Пер. с англ.; под ред. Ю. Мураками. – Т.1 – М.: Мир, 1990. – 448 с.

Надійшла 24.04.06
Після доробки 28.07.06

Наведені класи задач механіки деформованого твердого тіла розв'язок яких можна одержати за допомогою обчислювального комплексу «МИРЕЛА+», заснованого на методі скінчених елементів (МСЕ). Наведено задачі контакту еластомірних віброізоляторів, задачі визначення пружно-деформованого стану прямокутної плити із тріщиною. Обчислені коефіцієнти інтенсивності напруг уздовж фронту тріщини.

Classes of problems of the mechanics of the deformed solid are described. Decision of these problems are possible to resolve with the help of the computer complex «МИРЕЛА+». It is based on a method of final elements (FEM). Problems of contact of elastomeric vibration isolators, calculation of pressure-deformed state of a rectangular plate with a crack are given. Factors of stress intensity along front edge of a crack are calculated.

УДК 621.3

М. А. Новотарський

МЕРЕЖІ ДЛЯ МОДЕЛЮВАННЯ СКЛАДНИХ СИСТЕМ

В роботі дано короткий огляд сучасного стану мереж, що орієнтовані на формальний опис складних систем з метою побудови інформативних імітаційних моделей. Обґрунтовано необхідність подальшого розвитку мережних засобів для підвищення адекватності моделей складних систем. Розглянуто статичні характеристики мереж для формального опису систем з асинхронною взаємодією процесів. Опис динаміки даного типу мереж виконано з застосуванням апарату алгебри процесів та дано означення допустимих видів взаємодії. Наведено приклади опису процесів і їх взаємодії.

ВСТУП

Широке застосування мереж як засобів абстрактного опису моделей бере свій початок від мереж Петрі [1]. Їх успіх базується на тому факті, що представлення об'єкта моделювання множиною станів і алгоритмів обробки уніфікує процес переходу від конкретної моделі до абстрактної, тобто такої, яка дозволяє застосувати формальні процедури аналізу. В результаті накопичення досвіду застосування мереж Петрі поряд з безумовними перевагами поступово проявилися також і їх недоліки, основний з яких полягає у використанні фік-

сованих правил спрацювання переходів. Незважаючи на те, що доведена функціональна повнота цих правил, застосування їх до опису роботи реальних об'єктів призводить до невиправдано громіздких моделей. Цей факт став причиною виникнення як численних модифікацій мереж Петрі, так і нових мереж. Очевидно, що все розмаїття навколишнього світу неможливо вкласти в прокрустове ложе обмеженої множини алгоритмів, що і є внутрішнім стимулом розвитку теорії мережного моделювання.

Якщо об'єктом дослідження є деякі системи, які реалізують паралельну або розподілену обробку інформації, то вони можуть розглядатися як динамічні системи з дискретними подіями [2]. Моделі таких систем, як правило, містять опис елементів структури та алгоритмів їх функціонування. Наведений підхід дуже нагадує правила побудови мережних моделей, що стало причиною виникнення ідеї застосування їх при реалізації механізму моделювання дискретними подіями [3]. У даному випадку формальний опис мережі задає стан елементів об'єкта моделювання та структуру зв'язків між ними, а динаміка функціонування мережі представлена послідовностями подій, описаних у термінах процесів [4]. Безумовна перевага такого підходу полягає в можливості задавати як завгодно складні алгоритми функціонування елементів системи, не спираючись на фіксований набір правил спрацювання переходів. Принципи опису процесів, започатковані в [5, 6], дозволяють успішно розв'язати ще одну проблему, яка є актуальною для мереж Петрі і полягає в необхідності уникнення тупикових ситуацій при моделюванні асинхронної взаємодії елементів системи.

Дана робота присвячена подальшому розвитку інструментів опису паралельних асинхронно взаємодіючих процесів з метою використання їх при створенні мережних моделей. Мережі, які використовують ієрархічні можливості опису структури, характерні для PRO-мереж, і реалізують динаміку асинхронної взаємодії, дістали назву APRO-мережі.

1 ОБ'ЄКТИ APRO-МЕРЕЖІ

Об'єкти мережі традиційно можуть бути представлені в графічному або аналітичному вигляді. Фрагмент APRO-мережі, який містить вхідну позицію p_1 , вихідну позицію p_2 , перехід t_1 , вхідне ребро (p_1, t_1) та вихідне ребро (t_1, p_2) , показано на рис. 1.

Означення 1. Формальним описом APRO-мережі є кортеж об'єктів:

$$\Phi = (\mathbf{P}, \mathbf{T}, \mathbf{F}, \mathbf{M}, \mathbf{V}), \quad (1)$$

де $\mathbf{P} = \{p_i\}_{i=1}^n$ – множина позицій, яка описує ступінь деталізації станів мережі; $\mathbf{T} = \{t_j\}_{j=1}^m$ – множина пе-

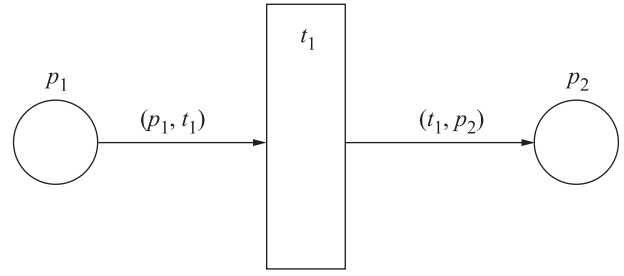


Рисунок 1 – Тривіальна APRO-мережа

реходів, яка задає динаміку функціонування за допомогою алгоритмів обробки інформації, структурованої у елементах множини \mathbf{P} ; $F = P \times T \cup T \times P$ – множина ребер, елементи якої (p_i, t_j) і (t_j, p_i) задають зв'язки між відповідними елементами множин \mathbf{P} і \mathbf{T} ; $\mathbf{M} = \{\mu_k\}_{k=1}^r$ – множина допустимих міток; $\mathbf{V} = \{v_j\}_{j=1}^m$ – множина допустимих протоколів.

Довільна позиція APRO-мережі $p_i = \{c_i, q_i\}$ містить множину параметрів c_i , яка складається з підмножини атрибутивних констант і підмножини атрибутивних змінних, та інформаційну структуру даних q_i . Атрибутивні константи задають ті характеристики позиції, які відображають особливості об'єкта абстрактного опису, а тому не можуть бути змінені. Атрибутивні змінні визначають поточний стан даного об'єкта і змінюються в ході функціонування мережі в межах допустимих значень.

Довільний перехід $t_j = \{\chi_j, \text{Process}_j\}$ містить множину атрибутивних параметрів χ_j і процес переходу Process_j . Множина ребер \mathbf{F} може бути задана матрицею інцидентності \mathbf{H} з елементами:

$$\mathbf{H}(p_i, t_j) = \begin{cases} -1, & (p_i, t_j) \in \mathbf{F}, \\ +1, & (t_j, p_i) \in \mathbf{F}, \\ 0, & (p_i, t_j) \notin \mathbf{F}, (t_j, p_i) \notin \mathbf{F}, \end{cases} \quad 1 \leq i \leq n, \quad 1 \leq j \leq m, \quad (2)$$

де (p_i, t_j) – вхідне ребро, яке сполучає позицію p_i з переходом t_j при $\mathbf{H}(p_i, t_j) = -1$, (t_j, p_i) – вихідне ребро, яке сполучає перехід t_j з позицією p_i при $\mathbf{H}(p_i, t_j) = 1$.

Такий спосіб задавання ребер мережі є найпростішим, але не єдиним з можливих. Формальний опис APRO-мережі допускає використання довільної однозначної системи задавання зв'язків між переходами та позиціями.

Інформаційні потоки APRO-мережі структуровані у вигляді множини міток \mathbf{M} , кожен елемент якої $\mu_k = \{\lambda_k, A_k\}$ складається з підмножини параметрів λ_k і множини атрибутів A_k . Атрибути мітки інтерпретують як дані, які необхідно передавати від одного об'єкта мережі

до іншого, а параметри містять службову інформацію про структуру атрибутів, час їх створення та ін.

Представлення об'єктів у вигляді множин структурних та інформаційних елементів зумовлене використанням процесів, у розвитку та взаємодії яких полягає динаміка функціонування АPRO-мережі.

2 ОПИС ПРОЦЕСІВ У АPRO-МЕРЕЖАХ

Функціонування АPRO-мережі будемо розглядати як деяку послідовність зміни її станів, яка характеризується наступними часовими вимірами:

1. Мить – значення мережного часу, яке відповідає унікальному стану мережі.
2. Інтервал – тривалість часу між двома послідовними митями.
3. Період – послідовність з одного і більше інтервалів.

Стан мережі складається зі станів об'єктів, які входять у мережу. Стан кожного об'єкта визначається певними числовими значеннями його атрибутивних параметрів. Активністю називатимемо стан об'єкта моделі на протязі одного інтервалу.

Подія відбувається миттєво і є причиною зміни стану об'єкта шляхом ініціювання нової активності. Подія може бути детермінованою у випадку, коли умова її виникнення прямо залежить від величини системного часу, або недетермінованою, коли умови її виникнення не піддаються чіткому визначенню. Отже активність об'єкта – це його стан між двома послідовними подіями. Послідовність активностей об'єкта, ініційованих подіями на протязі деякого періоду, утворюють процес. З кожним об'єктом мережі може бути асоційовано кілька процесів, тому не всі події, що відбуваються в об'єкті, і не всі активності обов'язково впливають на розвиток того чи іншого процесу. Множину подій, яка спричиняє розвиток даного процесу, називають його алфавітом подій, а множину активностей – алфавітом активностей. Для математично строгого опису процесів та їх взаємодії спиратимемось на формальний апарат алгебри процесів [6], який задає алфавіт активностей у вигляді:

$$A = \Lambda \cup \bar{\Lambda} \cup I, \tag{3}$$

де $\Lambda, \bar{\Lambda}$ – взаємно однозначні множини зовнішніх активностей; I – множина внутрішніх активностей. Зовнішніми активностями $\lambda \in \Lambda$ і $\bar{\lambda} \in \bar{\Lambda}$ називають активності, які забезпечують відповідно вивід та ввід інформації і пов'язані між собою залежністю: $\lambda = \bar{\bar{\lambda}}$. Множина $I = \emptyset$ представляє послідовність активностей обробки інформації процесом без взаємодії з зовнішнім середовищем.

Існування довільної активності $a(\alpha, t) \in A$ залежить від стартової події α та стартового часу t , який ви-

значається станом локального лічильника часу процесу на момент виникнення події α .

На множині активностей задамо функцію $\delta: A \rightarrow R^+$, яка визначає тривалість кожної з тих активностей, що входять до алфавіту A :

$$\delta(a) = \begin{cases} \delta(\lambda) = \delta(\bar{\lambda}) = \varphi(R^+ \cup \{0\}), \lambda \in \Lambda, \bar{\lambda} \in \bar{\Lambda}, \\ \delta(i) = \varphi(R^+ \cup \{0\}), i \in I, \end{cases} \tag{4}$$

де $\varphi(*)$ – функція, яка задає скінченну підмножину.

Нехай s_k – деякий допустимий k -й процес. Очевидно, що завжди повинна існувати активність a , яка відбувається першою серед послідовності активностей даного процесу. Таку активність називатимемо префіксом процесу, а префіксний запис процесу матиме вигляд: $a.s_k$. Загальний процес s може бути представлений з різним ступенем деталізації:

$$s \Leftarrow a.s_k \Leftarrow a.b.nil. \tag{5}$$

Активність b у лінійній послідовності (5) створює подію запуску активності nil , яка не виконує ніяких дій і має нульову тривалість $\delta(nil) = \{0\}$.

Означення 2. Тривіальний процес s , заданий на АPRO-мережі Φ , складається з двох активностей, одна з яких входить до алфавіту даної мережі $A|_{\Phi}$, а другою завжди є активність зупинки nil .

Якщо необхідно описати циклічну роботу тривіального процесу, то використовують рекурсивний запис:

$$s \Leftarrow a.s, \tag{6}$$

який задає процес з нескінченною кількістю активностей a : $\underbrace{a\dots a\dots a.a}_{\infty}.nil$.

Будемо користуватися таким записом також для задавання лінійної послідовності нетривіальних процесів s_k : $s \Leftarrow s_k.s$.

Виходячи з визначення процесу, його загальна тривалість визначається тривалістю активностей:

$$\delta(s_k) = \sum_n \delta(\lambda_n) + \sum_m \delta(i_m), \tag{7}$$

де n, m – відповідно кількість входжень в процес зовнішніх і внутрішніх активностей.

Якщо в системі припустимо існування K процесів $\{s_k\}_{k=1}^K$, кожен з яких може визначати поведінку системи в залежності від стану навколишнього середовища, то узагальнений процес s задають виразом: $\pi \Leftarrow \pi_1 + \dots + \pi_k + \dots + \pi_K$, а операцію «+» називають операцією вибору. Тривалість узагальненого процесу визначають за формулою:

$$\delta(s) = \delta(s_1) \cup \dots \cup \delta(s_k) \cup \dots \cup \delta(s_K).$$

Операція композиції « $\|$ » описує ситуацію, коли в системі одночасно існують K процесів, а $s \in$ композицією дій всіх цих процесів: $s \leftarrow s_1 \| s_2 \| \dots \| s_k \| \dots \| s_{K-1} \| s_K$. Тривалість узагальненого процесу, сформованого в результаті операції композиції, задають множиною:

$$\delta(s) = \{\max(l_1, \dots, l_K) \mid l_1 \in \delta(s_1), \dots, l_K \in \delta(s_K)\},$$

де l_1, \dots, l_K – допустимі скінченні тривалості розгортання процесів s_1, \dots, s_K .

Операцію виключення « \setminus » використовують у випадку, коли потрібно виключити активність або деяку множину активностей з узагальненого процесу: $s = s_k \setminus D$. Тривалість узагальненого процесу $\delta(s)$ після операції виключення:

$$\delta(s) = \delta(s_k) - \delta(D),$$

де $\delta(D)$ – загальна тривалість всіх активностей, виключених з процесу s_k .

Нехай $D = \Lambda \cup \bar{\Lambda}$, тоді узагальнений процес s містить тільки внутрішні активності I .

Операція « $[f]$ », яку називають операцією перейменування, полягає в застосуванні функції $f: A \rightarrow A$ з властивостями $a = -f(a)$, $\bar{f}(\bar{a}) = -f(\bar{a})$ до кожної активності процесу: $s \leftarrow s_k [f]$. Перейменування процесу не впливає на його тривалість, тому $\delta(s) = \delta(s_k)$.

Префіксація, вибір, композиція, виключення та перейменування є операціями класичної алгебри процесів. Вони призначені для опису таких процесів, активності яких не мають прямого взаємного впливу.

Спільне функціонування компонентів паралельних систем ґрунтується на взаємодії процесів. Говорять, що взаємодія відбулася, якщо деяка порція інформації була передана від одного процесу до іншого. Сам факт передачі інформації завжди потребує узгодження між процесом-передавачем та процесом-приймачем. Таке узгодження і визначає спосіб взаємодії.

Якщо активності процесів, які утворюють паралельну систему, запускаються спільною зовнішньою подією, то така подія виступає в ролі згаданого узгодження, і процес-передавач може безпосередньо передавати дані процесові-приймачеві. Даний спосіб взаємодії будемо називати синхронною взаємодією. Операція синхронної взаємодії « $\|$ » задає композицію синхронних процесів, об'єднаних узагальненим процесом s :

$$s \leftarrow s_1 \| s_2 \| \dots \| s_k \| \dots \| s_K.$$

У випадку, коли кожен з процесів є керованим власним потоком подій, необхідно застосовувати додаткові засоби узгодження. Для цього введемо спеціальні синхронізуючі процеси σ , що складаються з од-

нієї активності $e(\beta, t_e)$, яка рекурсивно запускається подією β в момент часу t_e і виконує роль синхронізуючих імпульсів [7]: $\sigma \leftarrow e \cdot \sigma$. Для узгодження в часі виникнення відповідних активностей взаємодії вводять операцію синхронізації $[s_k] \sigma(s_l)$, яка затримує тривалість активності взаємодії до моменту виникнення синхронізуючої активності. Спосіб взаємодії, який використовує додатковий синхронізуючий процес, будемо називати синхронізованою взаємодією.

Засобом узгодження може також виступати деяке зовнішнє середовище, яке готове в будь-який момент часу прийняти інформацію від процесу-передавача і видати її процесові-приймачеві. Таке середовище представимо у вигляді спеціального процесу u з можливістю паралельного виконання двох активностей: $\lambda_k \bar{\lambda}_l$, $k \neq l$. Спосіб взаємодії з використанням згаданого середовища будемо називати асинхронною взаємодією. Операція асинхронної взаємодії « $\langle \rangle$ » задає композицію асинхронних процесів з узагальненим процесом s :

$$s \leftarrow s_1 \langle \rangle s_2 \langle \rangle \dots \langle \rangle s_K.$$

Семантику операцій алгебри процесів [8] задамо у вигляді дробу:

$$\text{операція} \Rightarrow \frac{\text{дія}}{\text{правило}} (\text{додаткова умова}).$$

Чисельник дробу вказує на дію, а знаменник задає правило виконання тої чи іншої операції під час цієї дії. Дія $s_k \xrightarrow{a(\alpha, t), \delta(a)} s_k$ може бути означена таким чином: процес s_k бере участь у виконанні активності a , яка ініційована подією α в момент локального часу t , а по закінченні періоду $\delta(a)$ поводить як процес s_k . Якщо додаткова умова не задана, то мають на увазі, що вона завжди справджується. Правила у знаменнику означають виконання заданої операції.

Розглянемо базові правила алгебри процесів, орієнтованої на опис процесів у АPRO-мережах.

$$\text{Префіксація: } a.s_k \Rightarrow \frac{s_k \xrightarrow{a(\alpha, t), \delta(a)} s_k}{a.s_k \xrightarrow{a} s_k}. \text{ Ця операція}$$

вказує на виконання префіксованої активності без зміни самого процесу s_k . Якщо в алфавіті процесу s_k входить активність a і в момент часу t відбулася подія α , то процес s_k вважають ініційованим префіксованою активністю a після закінчення періоду $\delta(a)$.

$$\text{Вибір: } s_k + s_l \Rightarrow \left\langle \frac{s_k \xrightarrow{a(\alpha, t), \delta(a)} s'_k}{s_k + s_l \xrightarrow{a} s'_k}, \frac{s_l \xrightarrow{a(\alpha, t), \delta(a)} s'_l}{s_k + s_l \xrightarrow{a} s'_l} \right\rangle.$$

Операція вибору має два семантичні правила, які вказують на той факт, що процес-спадкоємець визначається заданою в системі дією.

Композиція:

$$s_k | s_l \Rightarrow \left\langle \frac{s_k \xrightarrow{a(\alpha, t), \delta(a)} s'_k, s_l \xrightarrow{a(\alpha, t), \delta(a)} s'_l}{s_k | s_l \xrightarrow{a} s'_k | s'_l}, \frac{s_k | s_l \xrightarrow{a} s'_k | s'_l}{s_k \xrightarrow{a(\alpha, t), \delta(a)} s'_k, s_l \xrightarrow{a(\alpha, t), \delta(a)} s'_l} \right\rangle.$$

Ці семантичні правила нагадують правила операції вибору і задають результат операції для випадку, коли дія відбувається тільки над одним з паралельних процесів.

Виключення: $s_k \setminus D \Rightarrow \frac{s_k \xrightarrow{a(\alpha, t), \delta(a)} s'_k}{s_k \setminus D \xrightarrow{a} s'_k \setminus D} (a, \bar{a} \notin D).$

Правило виключення містить додаткові умови, які забороняють використовувати ті активності, що входять в множину D . Таким чином виключають вплив тих чи інших активностей на розвиток процесу.

Перейменування: $s_k[f] \Rightarrow \frac{s_k \xrightarrow{a(\alpha, t), \delta(a)} s'_k}{s_k[f] \xrightarrow{f(a)} s'_k[f]}.$ Правило

перейменування відрізняється від базової дії тільки тим, що операція виконується над перейменованими активностями.

Правило формування змінної процесу:

$$a.s \Rightarrow \frac{s_k \xrightarrow{a(\alpha, t), \delta(a)} s'_k}{s \xrightarrow{a} s'_k}.$$

Якщо загальний процес s складається з процесу s_k , то процес s поводить ідентично процесу s_k .

Базові правила описують дії, що стосуються тільки одного з процесів, незалежно від кількості процесів, які одночасно можуть існувати в системі. Правила, які включають дії над двома і більше процесами, будемо називати правилами взаємодії [9].

Синхронна взаємодія. Якщо взаємно однозначна дія існує для обох процесів, то можемо говорити про синхронну взаємодію процесів. Правило синхронної взаємодії має вигляд:

$$s_k || s_l \Rightarrow \frac{s_k \xrightarrow{\lambda(\alpha, t), \delta(\lambda)} s'_k, s_l \xrightarrow{\bar{\lambda}(\alpha, t), \delta(\bar{\lambda})} s'_l}{s_k || s_l \xrightarrow{i(\alpha, t), \delta(i)} s'_k || s'_l},$$

де $i \in I$ – внутрішня активність.

Умовою виникнення синхронної взаємодії є існування деякої зовнішньої події α , яка відбувається в момент часу t і запускає множину активностей: $\alpha \Rightarrow \{\lambda, \bar{\lambda}, i\}$.

Синхронізована взаємодія:

$$[s_k] \sigma(s_l) \Rightarrow \frac{s_k \xrightarrow{\lambda(\alpha, t), \delta(\lambda)} s'_k, s_l \xrightarrow{\bar{\lambda}(\alpha, t), \delta(\bar{\lambda})} s'_l}{[s_k] \sigma(s_l) \xrightarrow{s(\beta, t_s), \delta(s)} s'_k | s'_l}.$$

Якщо процес s_k потребує обміну даними з процесом s_l , то тривалість відповідної активності процесу s_k подовжується до настання моменту готовності процесу s_l .

Після акту обміну даними в системі виникає паралельна композиція процесів $s_k | s'_l$.

Асинхронна взаємодія:

$$s_k \langle s_l \Rightarrow \frac{s_k \xrightarrow{\lambda_k(\alpha_k, t_k), \delta(\lambda_k)} s'_k, s_l \xrightarrow{\bar{\lambda}_l(\alpha_l, t_l), \delta(\bar{\lambda}_l)} s'_l}{[s_k] \sigma(u) \xrightarrow{i_k(\alpha_k, t_k), \delta_i(\alpha_k)} s'_k | u} (t_k \leq t_l),$$

$$[s_l] \sigma(u) \xrightarrow{i_l(\alpha_l, t_l), \delta_l(\alpha_l)} s'_l | u$$

Процес s_k передає дані за допомогою зовнішньої активності λ_k , ініційованої подією α_k в момент часу t_k , а процес s_l виконує прийом даних за допомогою зовнішньої активності $\bar{\lambda}_l$, ініційованої подією α_l в момент часу t_l при додатковій умові: $t_k \leq t_l$. Оскільки в загальному випадку $\lambda_k \neq \bar{\lambda}_l$, то проміжним середовищем для зберігання даних виступає процес u .

Однією з важливих властивостей АPRO-мереж є ієрархічність. Завдяки цій властивості стає можливим формальний опис елементів об'єкта моделювання з різним ступенем деталізації. Успішне функціонування моделі за таких умов базується на уніфікованому механізмі взаємодії між процесами. Всі процеси, які функціонують на АPRO-мережі, мають на меті реалізацію цільової функції моделювання, тому можуть бути об'єднані в єдиний глобальний процес.

Означення 3. Глобальний процес \mathbf{S} , заданий на АPRO-мережі Φ , завжди визначений на алфавіті активностей даної мережі $A|_{\Phi}$ і містить підмножину процесів позицій $\{\pi_m\}_{m=1}^M$ та підмножину процесів переходів $\{\pi_n\}_{n=1}^N$:

$$\mathbf{S} = \{\pi_m\}_{m=1}^M \cup \{\pi_n\}_{n=1}^N, \quad (8)$$

де M, N – відповідно загальна кількість позицій та переходів мережі Φ .

Мінімальний алфавіт $A|_{\Phi_{\min}}$ тривіальної АPRO-мережі Φ_{\min} складається з такої множини активностей:

$$A|_{\Phi_{\min}} = \Lambda \cup \Lambda \cup I = \{\lambda_p, \lambda_t\} \cup \{\bar{\lambda}_p, \bar{\lambda}_t\} \cup \{i_p, i_t\}. \quad (9)$$

На рис. 2 показано зв'язок цих активностей із структурними елементами.

Наведемо приклад опису активностей, характерних для тривіальної АPRO-мережі Φ_1 , яка показана на рис. 3. Мережа складається із вхідної позиції p_1 , вихідної позиції p_2 і переходу t_1 , які оперують таким алфавітом активностей:

$$A_1 = \Lambda_1 \cup \Lambda_1 \cup I_1 = \{\lambda_p, \lambda_t\} \cup \{\bar{\lambda}_p, \bar{\lambda}_t\} \cup \{w_p, r_p, w_t, r_t, g_t\}, \quad (10)$$

де $\bar{\lambda}_p$ – прийом мітки на позицію; $\bar{\lambda}_t$ – прийом мітки на перехід; λ_p – видача мітки з позиції; λ_t – видача мітки

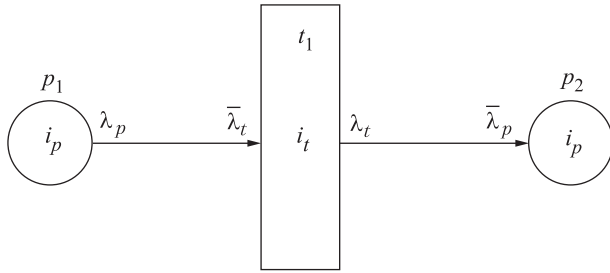


Рисунок 2 – Зв'язок між активностями та структурними елементами APRO-мережі

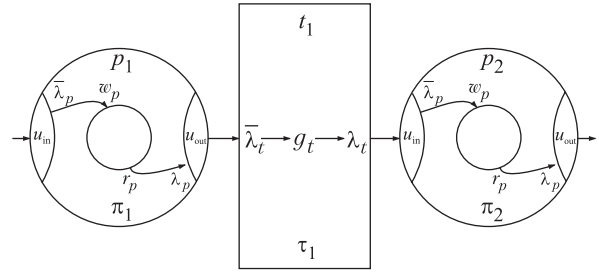


Рисунок 3 – Активності тривіальної APRO-мережі

з переходу; w_p – запис мітки у внутрішню структуру даних позиції; r_p – зчитування мітки з внутрішньої структури даних позиції; w_t – активація процесу обробки даних в переході; r_t – формування результатів обробки даних; g_t – активність обробки даних переходом.

Кожна позиція APRO-мережі виконує циклічно таку послідовність дій:

1. Аналізує наявність на вхідному ребрі готової мітки.
2. За наявності вільного місця зчитує мітку і розміщує її у внутрішній структурі даних.
3. Зчитує з внутрішньої структури мітку, яка має найвищий пріоритет.
4. Видає інформацію про готовність мітки до передачі на вихідне ребро.

В термінах алгебри процесів дії позиції представляють у вигляді відповідних активностей, що дозволяє записати узагальнений процес π_i у вигляді терму:

$$\pi_i \Leftarrow \bar{\lambda}_p \cdot w_p \cdot r_p \cdot \lambda_p \cdot \pi_i. \quad (11)$$

Перехід APRO-мережі також циклічно виконує наступні дії:

1. Аналізує наявність на вхідному ребрі готової мітки.
2. За наявності вільного ресурсу зчитує мітку і виконує її обробку.
3. Видає готову мітку на вихідне ребро.

Відповідний терм узагальненого процесу переходу має такий вигляд:

$$\tau_j \Leftarrow \bar{\lambda}_t \cdot g_t \cdot \lambda_t \cdot \tau_j. \quad (12)$$

Згідно з означенням 3 глобальний процес S_1 тривіальної APRO-мережі (рис. 3) представимо такою конфігурацією процесів:

$$S_1 \Leftarrow \pi_1 \langle \rangle \tau_1 \langle \rangle \pi_2.$$

Нехай процес S_1 забезпечує зчитування даних з позиції p_1 , обробку їх на переході t_1 та запис результатів

на позицію p_2 . Тоді, ввівши рекурсивні послідовності активностей

$$\pi_1 \Leftarrow r_p \cdot \lambda_p \cdot \pi_1, \quad \pi_2 \Leftarrow \bar{\lambda}_p \cdot w_p \cdot \pi_2, \quad \tau \Leftarrow \bar{\lambda}_t \cdot g_t \cdot \lambda_t \cdot \tau_2,$$

отримаємо:

$$S_1 \Leftarrow (r_p \cdot \lambda_p \cdot \pi_1) \langle \rangle (\bar{\lambda}_t \cdot g_t \cdot \lambda_t \cdot \tau_2) \langle \rangle (\bar{\lambda}_p \cdot w_p \cdot \pi_2).$$

Оскільки процеси переходів і позицій APRO-мережі характеризуються асинхронною взаємодією, на кожній з позицій введено вхідний u_{in} та вихідний u_{out} узгоджувальні процеси. Процес S_1 з урахуванням згаданих узгоджувальних процесів матиме вигляд:

$$S_1 \Leftarrow (r_p \cdot \lambda_p \cdot \pi_1) | u_{out} | (\bar{\lambda}_t \cdot g_t \cdot \lambda_t \cdot \tau_2) | u_{in} | (\bar{\lambda}_p \cdot w_p \cdot \pi_2).$$

ВИСНОВОК

В роботі розглянуто APRO-мережі, які призначені для формального опису складних дискретних систем з метою їх імітаційного моделювання. Для формального опису динаміки таких мереж запропоновано застосування апарату алгебри процесів. В термінах операцій алгебри процесів означені три основні види взаємодії між процесами: синхронна взаємодія, синхронізована взаємодія та асинхронна взаємодія. Наведено приклад опису процесів та їх взаємодії, характерних для тривіальної APRO-мережі.

ПЕРЕЛІК ПОСИЛАНЬ

1. Petri C. A. Introduction of general net theory // Lecture Notes in Computer Science. – Berlin: Springer-Verlag, 1980. – Vol. 84. – Pp. 1–26.
2. Ho Y. C. Dynamics of Discrete Event Systems // Proceedings of the IEEE. – 1989. – Vol. 77, No. 1. – Pp.3–6.
3. Котов В. Е. Сети Петри. – М.: Наука. – 1984. – 160 с.
4. Нестеренко Б. Б., Новотарский М. А. Имитационные модели параллельных асинхронных методов математической физики. – Киев, 1997. – 52 с. – (Препр./НАН України. Ін-т математики; 97.14).
5. Хоар Ч. Взаимодействующие последовательные процессы. – М.: Мир, 1989. – 264 с.
6. Milner R. Calculus of Communicating Systems // Lecture Notes in Computer Sciences. – 1980. – Vol. 92. – P. 1–260.

7. Nicollin X., Sifakis J. The algebra of timed processes ATP: Theory and application // Information and Computation. – 1994. – Vol. 104, No. 1. – Pp. 131–178.
8. Plotkin G. D. A structural approach to operational semantics. – Technical Report DAIMI-FN-19, Computer Science Department, University of Aarhus, Denmark, 1981. – 133 p.
9. Milner R. Calculus of synchrony and asynchrony // Theoretical Computer Science. – 1983. – Vol. 25. – Pp. 267–310.

Надійшла 17.09.05
Після доробки 17.03.06

В работе дан краткий обзор современного состояния сетей, ориентированных на формальное описание сложных систем с целью построения информативных имитационных моделей. Обоснована необходимость дальнейшего развития сетевых средств для повышения адекватности моделей сложных систем. Рассмотрены статические характеристики сетей для формального описания систем с асинхронным взаимодействием процессов. Описание

динамики данного типа сетей выполнено с применением аппарата алгебры процессов, даны определения допустимых видов взаимодействия. Приведены примеры описания процессов и их взаимодействия.

In paper the brief review of modern development of networks focused on the formal description of complex systems with the purpose of construction of informative simulations is given. Necessity of the further development of network means for increase of adequacy simulations of complex system is proved. Static characteristics of networks for the formal description of systems with asynchronous interaction of processes are considered. The description of dynamics of the given type of networks is executed with process algebra applications; definitions of allowable kinds of interaction are given. Examples of the description of processes and their interactions are resulted.

УДК 681.3.06:518

С. Н. Сердюк, К. Х. Ибрагим, Ж. К. Каминская

УСОВЕРШЕНСТВОВАНИЕ МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЭКСПЕРТНЫХ ПРОБЛЕМНЫХ ЗНАНИЙ

Проведен анализ семантико-синтаксической структуры предложений русского языка (СССПРЯ) с целью усовершенствования моделей представления экспертных проблемных знаний (МПЭПЗ), позволяющих автоматизировать процесс извлечения знаний из экспертов в более узкой области – по предметным знаниям. Процесс формализации проблемных знаний представлен двумя взаимосвязанными моделями: прикладной моделью представления знаний экспертов и моделью представления этих знаний с помощью направленных графов.

ПОСТАНОВКА ПРОБЛЕМЫ

В жизни современного социума электронная информация приобретает все более значимую роль. В мировых информационных хранилищах накоплены огромные объемы информации на естественном языке, обработать которые человеку без использования специализированных технических решений, практически, невозможно.

Одной из основных задач при проектировании интеллектуальных систем является задача автоматизированного извлечения знаний из экспертов в конкретной предметной области (ПдО). Процесс перехода от экспертных представлений, выраженных в естественноречевой форме, к сгенерированному системой множеству вопросов, служащих для интерактивного извлечения знаний о ПдО, представляет собой многоступенчатое преобразование. Такое преобразование можно представить следующими последовательными этапами:

- извлечение знаний из эксперта, владеющего более общими знаниями – знаниями о проблемной области (ПрО);
- этап формализации проблемных знаний;
- этап построения множества вопросов экспертам в ПдО;
- автоматизированный опрос эксперта в ПдО и пополнение базы знаний и данных (БЗД).

Извлечение знаний из эксперта в ПрО – это подготовительный, неавтоматизированный этап извлечения знаний. Формализация проблемных знаний обеспечивается семантико-синтаксическим анализом введенной информации и представлением ее с помощью формальных моделей. На этом этапе выявляются метазнания, которые определяют структуру знаний низшего уровня, что позволит автоматизировать опрос эксперта в ПдО путем генерации множества вопросов к данным экспертам.

В статье приводится построение формальных моделей, описывающих проблемные экспертные знания на основании проведенного семантико-синтаксического анализа предложений русского языка.

АНАЛИЗ СОСТОЯНИЯ ВОПРОСА

Модели представления проблемных знаний предназначены для поэтапного представления экспертных проблемных знаний, выраженных через предложения ес-

© Сердюк С. Н., Ибрагим К. Х., Каминская Ж. К., 2006

тественного языка, в виде формально-языковых конструкций, позволяющих получить описание исследуемой области знаний в виде, пригодном для дальнейшей автоматизированной обработки.

Базисные концепции и допущения, определяющие номенклатуру типов предложений русского языка и общие требования, необходимые для построения МПЭПЗ, следуют из проведенного СССПРЯ и изложены в работе [1].

Одной из составляющих изложенной концепции является разделение предложения на группу подлежащего и группу сказуемого. Любое глагольное действие предполагает наличие субъекта, производящего действие, представляемого группой подлежащего. В случае пространственного двусоставного предложения в группе сказуемого предполагается также наличие объекта – актанта, на которого направлено действие. В общем случае группа подлежащего выражается именным словосочетанием. Основным носителем смысла в таких словосочетаниях является, как правило, первое слева существительное, для которого типичной является форма именительного падежа существительного. Группа сказуемого, в общем случае, выражается глаголом, кратким прилагательным или кратким причастием с определяющими или дополняющими их словами и словосочетаниями.

Ограничением вышеупомянутой концепции является строго определенное положение этих групп относительно друг друга, что снижает область применения разработанных в [1] МПЭПЗ. В данной статье рассмотрен подход, устраняющий данный недостаток МПЭПЗ.

АНАЛИЗ СЕМАНТИКО-СИНТАКСИЧЕСКОЙ СТРУКТУРЫ ПРЕДЛОЖЕНИЙ РУССКОГО ЯЗЫКА

Русский язык – язык со свободным порядком слов в предложении. Из чего следует, что предложения «В двигателе есть подсистема вырыска» и «Подсистема вырыска есть в двигателе» являются, с точки зрения синтаксиса, правильно построенными предложениями, а с точки зрения семантики – абсолютно идентичными. В приведенном примере, используя только положения определенные выше, невозможно однозначно определить, какая из сущностей выступает субъектом, а какая – объектом. Проведенный анализ синтаксической структуры предложений русского языка показал, что существует возможность устранения проблем, связанных с неоднозначностью положения групп подлежащего и сказуемого, путем введения в МПЭПЗ понятия предлог. Предлоги могут входить как в группу, описывающую субъект, так и группу, характеризующую объект высказывания. Рассмотрев классификацию предлогов [2], можно выделить те классы, которые целесообразно включить в МПЭПЗ вследствие того, что они позво-

ляют определить, которая из сущностей является отправной точкой (источник действия, направленного на объект), а что из сущностей выступает в роли получателя действия (объект). Поэтому для формализации выбраны лишь предлоги пространственных отношений, исключая предлоги распространения действия по пространству, т. к. они не указывают на субъект или объект, а указывают на место протекания события. Включение в МПЭПЗ объектных предлогов обосновано тем, что их цель обозначить объект. Предлоги, не подлежащие формализации в модели представления знаний допустимы, но их необходимо рассматривать как составляющие групп субъекта или объекта. Таким образом сформулированы дополнительные требования к МПЭПЗ. В результате МПЭПЗ должна предусматривать следующие средства описания:

- сложных предложений сочинительного типа, представленных в виде простых предложений, соединенных логическими связками;
- простых предложений, подлежащих формализации;
- актантов предложения субъектов и объектов, находящихся в определенном отношении друг с другом;
- словосочетаний и местоименных слов, которые в МПЭПЗ должны представляться в виде следующих слов «ЛЮБОЙ», «НЕКОТОРЫЙ», «НЕСКОЛЬКО»;
- предлогов, подлежащих формализации классов.

ПОСТРОЕНИЕ МОДЕЛЕЙ ПРЕДСТАВЛЕНИЯ ПРОБЛЕМНЫХ ЗНАНИЙ

Процесс перехода от экспертных представлений, выраженных в естественно-языковой форме, к сгенерированному системой множеству вопросов, служащих для интерактивного извлечения знаний о ПдО, представляет собой многоступенчатое преобразование. Каждому этапу этого преобразования должна соответствовать своя модель, формально его описывающая. Целесообразно представить процесс формализации проблемных знаний двумя взаимосвязанными моделями: прикладной моделью представления знаний экспертов и моделью представления этих знаний в форме направленных графов, что позволяет на этапе формализации однозначно указать объект и субъект отношений.

Прикладная модель представления знаний определяется кортежем:

$$M1 = \langle K_{\text{пр}}, K_{\text{пр}}^{\text{фг}} \rangle,$$

где $K_{\text{пр}}$ – неформализованные знания экспертов о ПрО; $K_{\text{пр}}^{\text{фг}}$ – знания экспертов о той же ПрО, представленные в синтаксисе формальной грамматики.

Модель представления знаний в форме направленных графов определяется кортежем

$$M2 = \langle K_{пр}^{фг}, K_{пр}^{\Gamma} \rangle,$$

где $K_{пр}^{\Gamma}$ – знания экспертов, представленные в форме направленных графов.

Определим последовательность построения этих моделей, одновременно отражая способ представления компонентов прикладной модели – моделью в форме направленных графов. Проблемные знания, извлеченные инженером по знаниям (ИПЗ), представляют собой множество простых и сложных предложений. Сложные предложения с сочинительной связью могут быть представлены в виде совокупности простых. Простым повествовательным предложениям в модели $M1$ соответствуют элементарные высказывания. В соответствие с проведенным в [1] обоснованием следует различать следующие типы элементарных высказываний (ТЭВ).

ТЭВ1. Двусоставному нераспространенному предложению соответствуют высказывания вида (1) и (2) в зависимости от положения субъекта относительно предиката отношения.

$$V = \text{sub}\langle S_1_S_2_ \dots_ S_n \rangle \text{Rel}, \quad (1)$$

где $\text{sub}\langle S_1_S_2_ \dots_ S_n \rangle$ – субъект высказывания; $S_1_S_2_ \dots_ S_n$ – составное имя субъекта; $S_i (i = \overline{1, n})$ – элементы составного имени объекта, поясняющие его смысл; Rel – отношение, соответствующее сказуемому.

$$V = \text{Rel sub}(S_1_S_2_ \dots_ S_n), \quad (2)$$

В модели $M2$ высказываниям вида (1) и (2) соответствует ветвь графа:

$$\Gamma = (H, \text{Rel}), \quad (3)$$

где Γ – ветвь графа; H – множество вершин вида (4), создающееся на основании анализа ТЭВ,

$$H = \text{sub} \cup \text{obj}, \quad (4)$$

Rel – множество характеристик ребер между вершинами графа которое задается кортежем вида (5).

$$\text{Rel} = \langle \text{sub}, \text{obj}, \text{tr}, \text{col}, \text{alt} \rangle, \quad (5)$$

где sub – вершина, описывающая субъект действия; obj – вершина, указывающая на объект действия; tr – признак истинности ($\text{tr} = 1$) или ложности ребра ($\text{tr} =$

$= 2$); col – указывает на наличие местоименных слов при сущностях; alt – альтернативность ребер.

Пустая вершина – это особая вершина графа, которая имеет связи с другими вершинами, но при этом ее название не указывается. Такие вершины в дальнейшем либо подлежат уточнению в диалоге с экспертом, либо их определение будет найдено при дальнейшем анализе текста. Пустые вершины не могут принимать участие в формировании связей между подграфами графа, а до тех пор пока они не уточнены из них строится новый граф.

ТЭВ 1.1. Двусоставному нераспространенному предложению с местоименными словами в модели $M1$ соответствуют высказывания вида (6) и (7).

$$V = \text{sub}(W_S_1_S_2_ \dots_ S_n) \text{Rel}, \quad (6)$$

где W – одно из формализуемых местоименных слов.

$$V = \text{Rel sub}(W_S_1_S_2_ \dots_ S_n). \quad (7)$$

В зависимости от местоименного слова W в модели $M2$ получают различные формулы. Если местоименное слово W – «НЕКОТОРЫЙ» или «НЕСКОЛЬКО», то получают формулу (8), если W – «ЛЮБОЙ», то получают формулу (9).

$$\exists x P(x, \text{sub}) = 1: \Gamma = (H, \text{Rel}), H = x \cup \text{obj}, \quad (8)$$

где x – предметная переменная, соответствующая субъекту или объекту; $P(x, \text{sub})$ – предикат, указывающий на принадлежность предметной переменной x субъекту.

$$\forall x P(x, \text{sub}) = 1: \Gamma = (H, \text{Rel}), H = x \cup \text{obj}. \quad (9)$$

ТЭВ 2. Двусоставному распространенному предложению в модели $M1$ соответствует высказывание вида (10).

$$V = \text{Pred1Sush1}(S_1_S_2_ \dots_ S_n) \text{Rel} \times \\ \times \text{Predl2Sush2}(P_1_P_2_ \dots_ P_n), \quad (10)$$

где Predl1 , Predl2 – предлоги, разделяющие сущности на класс объектов и субъектов, если предлогов нет, то принимается, что левая относительно отношения Rel сущность – субъект, а правая – объект; $\text{Sush1}(S_1_S_2_ \dots_ S_n)$, $(P_1_P_2_ \dots_ P_n)$ – сущности, одна из которых относится к группе субъектов, а вторая – объектов; $P_1_P_2_ \dots_ P_n$ – составное имя субъекта; $P_i (i = \overline{1, n})$ – элементы составного имени субъекта, поясняющие его смысл.

В частном случае отсутствия предлогов высказывание (10) принимает вид

$$V = \text{sub}(S_1_S_2_ \dots_ S_n) \text{Rel obj}(P_1_P_2_ \dots_ P_n). \quad (11)$$

В модели $M2$ формула (10) может приобретать различную форму (в зависимости от положения и наличия предлогов Predl1 , Predl2) и представляется в виде:

$$\begin{aligned} \text{Prim_pred}(\text{Predl1}) = 1 \wedge \text{Slave_pred}(\text{Predl2})=1 \Rightarrow \\ \Rightarrow \text{Sush1}(S_1_S_2_ \dots _S_n) \equiv \text{sub}, \text{Sush2} \times \\ \times (P_1_P_2_ \dots _P_n) \equiv \text{obj}, \end{aligned} \quad (12)$$

$$\begin{aligned} \text{Prim_pred}(\text{Pred2})=1 \wedge \text{Slave_pred}(\text{Pred1})=1 \Rightarrow \\ \Rightarrow \text{Sush1}(S_1_S_2_ \dots _S_n) \equiv \text{obj}, \text{Sush2} \times \\ \times (P_1_P_2_ \dots _P_n) \equiv \text{sub}, \end{aligned} \quad (13)$$

$$\begin{aligned} \Gamma = (H, \text{Rel}), H = \text{sub} \cup \text{obj}, \text{Rel}= \\ = \langle \text{sub}, \text{obj}, \text{tr}, \text{col}, \text{alt} \rangle, \end{aligned} \quad (14)$$

где Prim_pred , Slave_pred – предикаты, разделяющие предлоги на группу источников или приемников действия.

В модели $M2$ высказывание (2.11) можно представить формулой (14).

ТЭВ 2.1 Двусоставному распространенному предложению с местоименными словами в модели $M1$ соответствуют высказывания вида (15–18)

$$\begin{aligned} V = \text{Predl1Sush1}(W_S_1_S_2_ \dots _S_n)\text{Rel} \times \\ \times \text{RelPredl2Sush2}(P_1_P_2_ \dots _P_n), \end{aligned} \quad (15)$$

$$\begin{aligned} V = \text{Predl1Sush1}(S_1_S_2_ \dots _S_n)\text{Rel} \times \\ \times \text{Predl2Sush2}(W_P_1_P_2_ \dots _P_n), \end{aligned} \quad (16)$$

$$\begin{aligned} V = \text{sub}(W_S_1_S_2_ \dots _S_n)\text{Rel} \times \\ \times \text{obj}(P_1_P_2_ \dots _P_n), \end{aligned} \quad (17)$$

$$\begin{aligned} V = \text{sub}(S_1_S_2_ \dots _S_n)\text{Rel} \times \\ \times \text{obj}(W_P_1_P_2_ \dots _P_n). \end{aligned} \quad (18)$$

В зависимости от местоименного слова W для высказывания (15) в модели $M2$ можно построить формулы (19–22).

$$\begin{aligned} \text{Prim_pred}(\text{Predl1})=1 \wedge \text{Slave_pred}(\text{Predl2})= \\ = 1 \Rightarrow \text{Sush1}(W_S_1_S_2_ \dots _S_n) \equiv \text{sub}, \\ \text{Sush2}(P_1_P_2_ \dots _P_n) \equiv \text{obj}; \\ \exists xP(x, \text{sub}) = 1: \Gamma = (H, \text{Rel}), \\ H = x \cup \text{obj}; \end{aligned} \quad (19)$$

$$\begin{aligned} \forall xP(x, \text{sub}) = 1: \Gamma = (H, \text{Rel}), \\ H = x \cup \text{obj}; \end{aligned} \quad (20)$$

$$\begin{aligned} \text{Prim_pred}(\text{Predl2}) = 1 \wedge \text{Slave_pred}(\text{Predl1}) = \\ = 1 \Rightarrow \text{Sush1}(W_S_1_S_2_ \dots _S_n) \equiv \text{obj}, \\ \text{Sush2}(P_1_P_2_ \dots _P_n) \equiv \text{sub}; \\ \exists xN(x, \text{obj}) = 1: \Gamma = (H, \text{Rel}), \\ H = x \cup \text{obj}, \end{aligned} \quad (21)$$

где $N(x, \text{obj})$ – предикат, указывающий на принадлежность предметной переменной x к объекту;

$$\forall xN(x, \text{obj}) = 1: \Gamma = (H, \text{Rel}), H = \text{sub} \cup \text{obj}. \quad (22)$$

В зависимости от местоименного слова W для выражения (16) в модели $M2$ можно построить формулы (23)–(26).

$$\begin{aligned} \text{Prim_pred}(\text{Predl1}) = 1 \wedge \text{Slave_pred}(\text{Predl2}) = \\ = 1 \Rightarrow \text{Sush1}(S_1_S_2_ \dots _S_n) \equiv \text{sub}, \\ \text{Sush2}(W_P_1_P_2_ \dots _P_n) \equiv \text{obj}; \\ \exists xN(x, \text{obj}) = 1: \Gamma = (H, \text{Rel}), H = x \cup \text{sub}; \end{aligned} \quad (23)$$

$$\forall xN(x, \text{obj}) = 1: \Gamma = (H, \text{Rel}), H = \text{sub} \cup \text{obj}; \quad (24)$$

$$\begin{aligned} \text{Prim_pred}(\text{Predl2}) = 1 \wedge \text{Slave_pred}(\text{Predl1}) = \\ = 1 \Rightarrow \text{Sush1}(W_S_1_S_2_ \dots _S_n) \equiv \text{obj}, \\ \text{Sush2}(W_P_1_P_2_ \dots _P_n) \equiv \text{sub}; \end{aligned}$$

$$\exists xP(x, \text{sub}) = 1: \Gamma = (H, \text{Rel}), H = x \cup \text{obj}; \quad (25)$$

$$\forall xP(x, \text{sub}) = 1: \Gamma = (H, \text{Rel}), H = \text{sub} \cup \text{obj}. \quad (26)$$

В зависимости от местоименного слова W выражению (17) в модели $M2$ соответствуют формулы (19) и (27).

$$\exists xP(x, \text{sub}) = 1: \Gamma = (H, \text{Rel}), H = \text{sub} \cup \text{obj}. \quad (27)$$

В зависимости от местоименного слова W для выражения (18) в модели $M2$ соответствуют формулы (23) и (24).

Формулы (19), (21), (23), (25), (27) применимы, если используется местоименное слово «НЕКОЛЬКО» или «НЕКОТОРЫЙ», формулы (20), (22), (24), (26) применимы, если используется местоименное слово «ЛЮБОЙ».

Таким образом, выше приведены разработанные формальные представления типов простых повествовательных предложений в прикладной и формальной моделях. Помимо элементарных высказываний, при описании проблемных знаний допустимы также следующие типы сложных высказываний образованных из ТЭВ [1]: конъюнкция, дизъюнкция, исключаящая дизъюнкция, отрицание, условные высказывания.

ВЫВОДЫ

Расширение номенклатуры средств описания формализуемых в моделях представления экспертных проблемных знаний, за счет предлогов выделенных классов, позволило расширить применение этих моделей для формального описания простых и сложных предложений русского языка со свободным порядком слов в предложении. Приведенные в данной статье, а также в работе [1] результаты позволяют формализовать полученные от экспертов проблемные знания, что делает их пригодными для дальнейшей машинной обработки с целью автоматизации процесса приобретения знаний из экспертов в предметных областях.

ПЕРЕЧЕНЬ ССЫЛОК

1. Сердюк С. М. Моделі представлення експертних проблемних знань. // Вісник ДУ «Львівська політехніка». Радіоелектроніка та телекомунікації. – 2000. – № 387. – С. 387–391.

2. Гужва Ф. К. Морфология русского языка: пособие для преподавателя. – К.: Рад. шк., 1987. – 223 с.

Надійшла 17.09.05
Після доробки 19.05.06

Проведено дослідження семантико-синтаксичної структури речень російської мови з метою удосконалення моделей представлення експертних проблемних знань, що дозволяють автоматизувати процес здобуття знань про більш вузьку область – із експерта по предметним знанням. Процес формалізації проблемних знань поданий двома взаємозалежними моделями: прикладною моделлю представлення знань експертів і моделлю представлення цих знань за допомогою спрямованих графів.

The analysis of Russian language sentence for semantic and syntactic structure is undertaken targeting to enhance models of expert objective knowledge. The models are engaged to automate the retrieval process to extract knowledge from experts specifically objected to a certain knowledge area. The knowledge subject's formalisation process is driven by two interconnected models: one is the application expert knowledge retrieval model and the model for that knowledge itself, presented by directional graphs.

УДК 681.324.01

А. Я. Складар, И. А. Макрушан

МОДЕЛИРОВАНИЕ ТРАНСПОРТНЫХ КОММУНИКАЦИЙ КОРПОРАТИВНЫХ КОМПЬЮТЕРНЫХ СЕТЕЙ: РАСПРЕДЕЛЕНИЕ ТРАНЗАКЦИЙ ПО МАРШРУТАМ

В данной статье рассмотрены вопросы моделирования транспортных коммуникаций корпоративных компьютерных сетей. Предлагается два подхода решения задачи: построение модели распределения потоков данных на основе методов максимизации энтропии, а также использование гравитационной модели распределения транзакций. В соответствии с предлагаемыми подходами представлены гравитационная и энтропийная модели распределения потоков данных. Показана практическая значимость предложенных подходов.

ВВЕДЕНИЕ

При проектировании систем управления распределенными корпоративными информационными технологиями, для того чтобы оценить интенсивность потоков данных (порождаемых задачами управления) между различными хостами сети, а также для моделирования распределения потоков данных по участкам транспортной сети, обычно используется некоторая математическая модель [1]. Эта модель, как правило, должна состоять из четырех подмоделей, которые связаны с генерацией трафика, распределением запросов по комму-

никациям, расщеплением типов коммуникаций (маршрутов) и размещением общих сетевых ресурсов сети. В этой статье вопросы, связанные с назначениями транзакций, не рассматриваются; здесь предполагается, что некоторые оценки транзакций в конечных пунктах заранее известны. Некоторые результаты статистического моделирования позволяют делать определенные выводы по размещению информационных ресурсов, но нас в основном будут интересовать распределение транзакций и расщепление типов коммуникаций [2]. Для решения этих задач будут применяться методы максимизации энтропии [3].

Цель статьи заключается в том, чтобы показать главное преимущество методов максимизации энтропии в области транспортных задач, которое состоит в том, что они помогают строить модели, описывающие довольно сложные явления, а также в том, что они допускают прямую интерпретацию полученных уравнений.

В связи с этим возникают следующие задачи:

- 1) подробно рассмотреть вопросы, связанные с гравитационной моделью распределения транзакций [2];

© Складар А. Я., Макрушан И. А., 2006

2) показать, как необходимо строить модели распределения потоков данных на основе методов максимизации энтропии;

3) привести ряд выводов, существенных для интерпретации полученных результатов моделирования.

1 ГРАВИТАЦИОННАЯ МОДЕЛЬ РАСПРЕДЕЛЕНИЯ ПОТОКОВ ДАННЫХ

Задача, которая ставится перед любой моделью распределения транзакций, заключается в оценке числа транзакций между каждой парой «точек» [4, 5]. Под «точкой» в данном случае понимается некоторый общий сетевой ресурс компьютерной сети. Предполагается, что рассматриваемая корпоративная сеть некоторым подходящим образом разбита на подсети [6].

Для простоты рассмотрим в качестве примера формирование и передачу запросов к файл-серверам корпоративной сети. Пусть T_{ij} – число транзакций (обращений к сетевому ресурсу), а C_{ij} – затраты на передачу из сети i в сеть j ; пусть Q_i – полное число отправленных запросов из сети i к файл-серверам других сетей, а D_j – полное число принятых запросов файл-сервером сети j . Модель распределения транзакций оценивает T_{ij} как функцию от Q_i , D_j , C_{ij} . Эти величины сами также могут зависеть от других независимых переменных.

Наиболее простая модель такого типа – это так называемая гравитационная модель, разработанная по аналогии с ньютоновским законом, связывающим силу притяжения F_{ij} между двумя массами m_i и m_j , расположенными на расстоянии d_{ij} друг от друга [2]

$$F_{ij} = \gamma \frac{m_i m_j}{d_{ij}^2}, \quad (1)$$

где γ – некоторая константа. По аналогии транспортная гравитационная модель может быть представлена в виде:

$$T_{ij} = k \frac{Q_i D_j}{c_{ij}^2}, \quad (2)$$

где k – некоторая константа, а затраты на передачу c_{ij} выступают в качестве «расстояния». Согласно уравнению (2) величина T_{ij} пропорциональна Q_i и D_j и обратно пропорциональна квадрату «расстояния» между ними. Но у этого уравнения имеется, по меньшей мере, один очевидный недостаток: если удвоить заданные значения Q_i и D_j , то число транзакций между этими сетями в соответствии с (2) учетверится, а естественно ожидать, что оно лишь удвоится. Или более точно: величины T_{ij} всегда должны удовлетворять следующим

ограничениям (число отправленных запросов из всех сетей в сеть j должно быть равно количеству прибывших запросов в сеть j):

$$\sum_i T_{ij} = D_j, \quad (3)$$

$$\sum_j T_{ij} = Q_i, \quad (4)$$

а уравнение (1) этого не обеспечивает. Это означает, что суммы по строкам и столбцам матрицы транзакций должны совпадать с числом запросов из каждой сети, и с числом полученных запросов в каждой сети соответственно. Этим ограничениям можно удовлетворить, если ввести наборы констант A_i и B_j , связанные соответственно с сетями источниками и сетями назначения. Иногда их называют балансирующими множителями. Кроме того, нет оснований считать, что расстояние играет в уравнении (2) такую же роль, что и в ньютоновской физике, поэтому используем более общую форму функции расстояния $f(c_{ij})$. Модифицированная гравитационная модель, при этом, будет иметь следующий вид:

$$T_{ij} = A_i B_j Q_i D_j f(c_{ij}), \quad (5)$$

где

$$A_i = \left[\sum_j B_j D_j f(c_{ij}) \right]^{-1}, \quad (6)$$

$$B_j = \left[\sum_i A_i Q_i f(c_{ij}) \right]^{-1}. \quad (7)$$

Уравнения для A_i и B_j решаются итерационными методами, и можно утверждать, что они гарантируют, что величины T_{ij} из уравнения (5) удовлетворяют ограничениям (3) и (4). Отметим также, что величины c_{ij} в такой модели могут служить общей мерой «сопротивления» передач данных между сетями i и j , в качестве которой могут выступать: географическое расстояние, время передач, затраты на передачи или, что часто бывает более эффективным, взвешенная комбинация этих факторов, которую иногда называют «обобщенными затратами».

Рассмотрим другой способ получения гравитационной модели транспортной системы с помощью максимизации энтропии [3], используя вначале другую физическую аналогию. Для этого введем дополнительное к (3) и (4) ограничение на T_{ij} . Это ограничение имеет вид:

$$\sum_i \sum_j T_{ij} c_{ij} = C. \quad (8)$$

Тогда наиболее вероятным распределением транзакций будет матрица $\{T_{ij}\}$, максимизирующая энтропию

$$\ln W(\{T_{ij}\}) = \ln T! - \sum_i \sum_j \ln T_{ij}!, \quad (9)$$

где T – полное число транзакций при ограничениях (3), (4), (8). Если учесть, что $W(\{T_{ij}\})$ – полное число состояний системы, соответствующих распределению $\{T_{ij}\}$, то число всех состояний системы можно оценить по формуле [3]

$$w = \sum W(\{T_{ij}\}), \quad (10)$$

где суммирование проводится по всем T_{ij} , удовлетворяющим (3), (4) и (8). При этом оказывается, что максимальное значение $W(\{T_{ij}\})$ настолько доминирует над всеми остальными членами суммы, что распределение $\{T_{ij}\}$, соответствующее максимуму энтропии, является существенно наиболее вероятным распределением.

Для получения набора T_{ij} , максимизирующего $\ln W(\{T_{ij}\})$ из уравнения (9) при ограничениях (3), (4) и (8), следует максимизировать лагранжиан L , равный

$$L = \ln W + \sum_i \lambda_i^{(1)} \left(Q_i - \sum_j T_{ij} \right) + \sum_j \lambda_j^{(2)} \left(D_j - \sum_i T_{ij} \right) + \beta \left(C - \sum_i \sum_j T_{ij} c_{ij} \right), \quad (11)$$

где $\lambda_i^{(1)}$, $\lambda_j^{(2)}$ и β – множители Лагранжа. Поскольку предполагается, что количество транзакций T_{ij} достаточно велико, то можно воспользоваться формулой Стирлинга [3] ($\ln T_{ij}! = T_{ij} \ln T_{ij} - T_{ij}$ – эта замена справедлива для больших значений T_{ij}), согласно которой из (9) получим

$$\ln W = - \sum_i \sum_j T_{ij} \ln T_{ij}.$$

Значения T_{ij} , которые доставляют максимум L и, следовательно, являются наиболее вероятным распределением транзакций, представляют собой решение системы уравнений $\frac{\partial L}{\partial T_{ij}} = 0$ совместно с ограничениями (3), (4) и (8).

Будем иметь

$$\frac{\partial L}{\partial T_{ij}} = \ln T_{ij} - \lambda_i^{(1)} - \lambda_j^{(2)} - \beta c_{ij}.$$

Это выражение будет равно нулю, когда

$$T_{ij} = \exp(-\lambda_i^{(1)} - \lambda_j^{(2)} - \beta c_{ij}). \quad (12)$$

Подставляя (12) в (3) и (4), получим выражения, используя которые можно определить $\lambda_i^{(1)}$ и $\lambda_j^{(2)}$:

$$\exp(-\lambda_i^{(1)}) = Q_i \left[\sum_j \exp(-\lambda_j^{(2)} - \beta c_{ij}) \right]^{-1}, \quad (13)$$

$$\exp(-\lambda_j^{(2)}) = D_j \left[\sum_i \exp(-\lambda_i^{(1)} - \beta c_{ij}) \right]^{-1}. \quad (14)$$

Чтобы представить окончательный результат в более привычном виде, запишем

$$A_i = \frac{\exp(-\lambda_i^{(1)})}{Q_i}, \quad (15)$$

$$B_j = \frac{\exp(-\lambda_j^{(2)})}{D_j}. \quad (16)$$

Отсюда

$$T_{ij} = A_i B_j Q_i D_j \exp(-\beta c_{ij}), \quad (17)$$

и в соответствии с уравнениями (13)–(16) имеем

$$A_i = \left[\sum_j B_j D_j \exp(-\beta c_{ij}) \right]^{-1}, \quad (18)$$

$$B_j = \left[\sum_i A_i Q_i \exp(-\beta c_{ij}) \right]^{-1}. \quad (19)$$

Таким образом, наиболее вероятное распределение транзакций является таким же, как в рассмотренной ранее гравитационной модели, которая определена уравнениями (5)–(7). Причем функция f была заменена на экспоненту с отрицательным показателем степени. Таким образом, статистическая теория утверждает [5, 7–9], что при заданных величинах запросов данных в удаленную сеть и приема их в каждой сети назначения и однородной цели запросов, при заданных затратах на передачу данных между сетями и фиксированных полных затратах на транспортировку, в корпоративной сети существует наиболее вероятное распределение транзакций между локальными сетями, и это распределение совпадает с тем, которое задается гравитационной моделью.

2 ИНТЕРПРЕТАЦИЯ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Особенность статистической механики состоит в том, что параметры, появляющиеся в уравнении, которое описывает наиболее вероятное распределение, имеют определенный физический смысл. То же самое справедливо и здесь в нашем случае. Величины Q_i , D_j и c_{ij} были определены ранее, а выражение $\exp(-\beta c_{ij})$ появляется в этой формулировке как явная форма зависимости от расстояния, причем параметр β определяется теоретически через уравнение (11). Но он имеет свою обычную интерпретацию. Он тесно связан со средним расстоянием между парами соединений: чем больше β , тем меньше среднее расстояние. Этот факт очевидным образом связан с величиной C в уравнении (8). Если увеличить C , то увеличатся затраты на передачи, и среднее расстояние передач данных возрастает; анализ левой части (8) показывает, что β при этом уменьшится.

Осталось интерпретировать A_i и B_j . Пусть одна из величин D_j изменяется, например, D_1 . Тогда величины

$$T_{i1} = A_i B_1 Q_i D_1 \exp(-\beta c_{i1}), \quad (20)$$

характеризующие запросы из каждой сети i в сеть 1, изменятся пропорционально. Величины A_i , как следует из уравнения (18), будут мало изменяться, поскольку выражение, включающее D_1 в каждом A_i – это только один из членов суммы. Величины B_1 возможно изменятся еще меньше, так как любое изменение в них будет вызвано изменениями в A_i .

Таким образом, роль A_i заключается (предполагая, что величина D_1 увеличилась) в небольшом сокращении числа всех запросов, которое компенсирует увеличение числа запросов в сеть 1. Следовательно, под A_i можно понимать некий конкурирующий параметр, сокращающий большинство запросов вследствие роста привлекательности одной сети. Знаменатель A_i также обычно используется как мера доступности, и можно сказать, что увеличение D_1 увеличивает доступность сети 1 для каждого пользователя корпоративной сети, хотя обычно мы будем пользоваться такой интерпретацией относительно изменений c_{ij} . Таким образом, этот анализ позволяет интерпретировать A_i в терминах конкурентной доступности сетевых ресурсов. Аналогичную роль играют величины B_j , которые связаны с основными изменениями в Q_i , а не в D_j . Изменение c_{ij} или несколько одновременных изменений Q_i и D_j приведет к сложному процессу перестройки A_i и B_j .

ВЫВОДЫ

Научная новизна. Проведенный анализ показал, что гравитационная модель имеет прочную основу и может быть применена для статистической оценки потоков

данных в магистральных каналах корпоративных компьютерных сетей. Однако следует заметить, что весь анализ проводился для одноцелевых транзакций и для однородной группы запросов. Транзакции в компьютерных сетях не являются идентичными в том смысле, в котором идентичны частицы в физике, поэтому вряд ли можно ожидать, что для реальной корпоративной сети можно получить достаточно точные результаты моделирования. Тем не менее, рассмотренный в работе подход может быть использован для предварительной оценки загрузки магистральных каналов передачи данных, и даже дать хорошие результаты, если удастся классифицировать транзакции по цели и по типу передач так, чтобы добиться разумной однородности потоков данных.

Практическая значимость. Предложенный подход может быть использован для предварительной статистической оценки потоков данных в корпоративной компьютерной сети. Полученные статистические оценки, в свою очередь, целесообразно использовать при построении общих моделей потоков, предусматривающих различия по множеству типов запросов, а также учитывающих различия в типах коммуникаций, соединяющих различные сети с использованием маршрутизаторов и шлюзов.

ПЕРЕЧЕНЬ ССЫЛОК

1. *Вишневецкий В. М.* Теоретические основы проектирования компьютерных сетей. – М.: «Техносфера», 2003. – 512 с.
2. *Иммельбаев Ш. С., Шмульян Б. А.* Анализ стохастических коммуникационных систем с применением термодинамического подхода // Автоматика и телемеханика. – 1977. – № 5. – С. 26–31.
3. *Хакен Г.* Информация и самоорганизация: макроскопический подход к сложным системам: Пер. с англ. – М.: Мир, 1991. – 240 с.
4. *Гуржий А. Н., Коряк С. Ф., Самсонов В. В., Скляр А. Я.* Контроль и управление корпоративными компьютерными сетями: инструментальные средства и технологии. – Харьков: «Компания СММТ», 2003. – 664 с.
5. *Клейнрок Л.* Коммуникационные сети: Перев. с англ. – М.: «Наука», 1975. – 256 с.
6. *Скляр А. Я., Макрушан И. А.* Математическая модель синтеза иерархических структур систем управления с последовательным применением основных методов декомпозиции // АСУ и приборы автоматики. – 2005. – Вып. 131. – С.69–73.
7. *Советов Б. Я., Яковлев С. А.* Построение сетей интегрального обслуживания. – Л.: Машиностроение. Ленинград. отделение, 1990. – 332 с.
8. *Шеннон К.* Работы по теории информации и кибернетике: Пер. с англ. под ред. Р. Л. Добрушина и О. Б. Лупанова, с предисловием А. Н. Колмогорова. – М.: «Изд. иностранной литературы», 1963. – 829 с.
9. *Стратонович Р. Л.* Теория информации. – М.: Сов. радио, 1985. – 486 с.

Надійшла 23.03.06
Після доробки 3.07.06

У даній статті розглянуті питання моделювання транспортних комунікацій корпоративних комп'ютерних мереж. Пропонується два підходи рішення задачі: побудова моделі розподілу потоків даних на основі методів максимізації ентропії, а також використання гравіта-

ційної моделі розподілу транзакцій. Дано інтерпретацію отриманих результатів.

In the article the problems of modeling of transport communications of corporate network have been considered.

Two approaches to the problems of solution have been proposed: the construction of a model of data flow distribution on the basis of entropy maximization methods, as well as the usage of gravitation model of transactions distribution. The interpretation of received results has been given.

УДК 519.713

Е. Е. Сыревич

ВЕРИФИКАЦИЯ ВЫСОКОУРОВНЕВЫХ МОДЕЛЕЙ ЦИФРОВЫХ УСТРОЙСТВ

Предлагается стратегия верификации моделей цифровых устройств, описанных с помощью языков описания аппаратуры. Основная идея лежит в генерации различающих псевдоисчерпывающих тестов для отдельных функциональных элементов, суперпозиции этих тестов и в интерактивном вычислении эталонных реакций.

ВВЕДЕНИЕ

В современной САПР РЭА основным способом описания устройств являются языки описания аппаратуры [1, 2]. С помощью языков описания аппаратуры (ЯОА), например VHDL или Verilog, проще и быстрее ввести и проверить проект в рамках САПР. 10 строками ЯОА можно описать как 1, так и 100000 триггеров. Описание устройств на ЯОА обладает рядом преимуществ. Проект на ЯОА – это объединение структуры цифрового устройства (ЦУ) и алгоритма его функционирования, одновременное описание ЦУ в статике и динамике. ЯОА – модель не требует дополнительного описания на человеческом языке и в виде схем, обладает высокой надежностью. Проект на ЯОА без труда переносится на другую элементную базу, например Altera FPGA или ASIC. Схема под конкретную элементную базу живет 2–4 года, а корректно написанный код на ЯОА может жить десятилетия. С помощью ЯОА можно описывать ЦУ на разных уровнях иерархии: алгоритмическом, структурном, регистровых передач и потоков данных, логическом и аналоговых схем.

Мировые компании по производству цифровых систем вынуждены постоянно сокращать время до попадания их продукции на рынок. А по оценкам производителей верификация, в том числе и функциональная, занимает до 80 % трудоемкости в цикле проектирования. Существует спрос на средства функциональной верификации моделей цифровых устройств на стадии их описания на поведенческом уровне с помощью конструкций языка VHDL.

Несмотря на обилие научных работ, связанных с верификацией и диагностикой цифровых устройств, в на-

стоящее время весьма актуальны и востребованы рынком инструментальные средства автоматической генерации тестов для функциональных моделей сложных ЦУ. Дефицит автоматических генераторов нового поколения испытывают и известные фирмы: Aldec, Altera, Actel, Xilinx, Synopsis.

Таким образом, цель данной работы – разработка стратегии верификации моделей ЦУ на языках описания аппаратуры, которая позволит уменьшить временные затраты на проектирование ЦУ путем использования методов сужения области возможных тестовых значений при автоматической генерации тестов.

1 ПОСТАНОВКА ЗАДАЧИ

Исходя из поставленной цели, для ее достижения решаются следующие задачи:

- 1) разработка внутренней модели цифрового устройства, представленного на ЯОА VHDL в целях последующей верификации;
- 2) разработка метода сужения количества подаваемых тестовых наборов;
- 3) разработка стратегии функциональной верификации моделей цифровых устройств на этапе ввода и моделирования проекта;
- 4) разработка методов оценивания качества верификации;
- 5) разработка информационного и методического обеспечения автоматического генератора тестов, поддерживающего стандарт VHDL.

Объектом исследования являются модели цифровых систем, проектируемые с помощью языков описания аппаратуры высокого уровня.

2 ВНУТРЕННЯЯ МОДЕЛЬ ДЛЯ ВЕРИФИКАЦИИ

Процесс построения тестов основывается на двух аспектах: внутренняя модель, позволяющая выполнять

процедуры прямой и обратной импликации, и, собственно, алгоритм генерации тестов.

Предлагается композиционная модель устройства для верификации на основании описания устройства на ЯОА-коде, например VHDL. Код на ЯОА должен быть откомпилирован и ориентирован на синтез. VHDL-модель представляется двумя графами. Первый – информационный I-граф – описывает поток данных и их преобразование (подобно операционному автомату в классической композиционной модели с микропрограммным управлением) без учета условных ветвей. Второй граф соответствует цепочке условий.

Информационный I-граф содержит два типа вершин: операнды и функции. Типы функций ограничены синтезируемым подмножеством VHDL. Дуги соединяют вершины следующим образом: вершина-источник соединяется с функциональной вершиной, затем дуга выходит из функциональной вершины и входит в вершину-приемник. Дуги, которые входят в вершины-приемники, могут быть условные и безусловные. Условные дуги соответствуют операторам, которые находятся внутри условных конструкций VHDL. Условные дуги содержат метки. В свою очередь, второй граф (управляющий C-граф) содержит условные конструкции (например, *case*, *if... then...*, *with ... select*) из исходного описания цифрового устройства. Каждый предикат условия – подграф, с определенной меткой.

3 ПРИНЦИПЫ ВЕРИФИКАЦИИ

На сегодняшний день самый распространенный подход к верификации состоит в следующем [3]. Имеются модели M1 и M2, эталонная и верифицируемая соответственно. Генерируются тесты для модели M1, и в результате моделирования M1 на полученных тестовых наборах определяются эталонные значения (аналогично процессу тестового диагностирования). Затем эти же тесты подаются на M2, их также моделируют и получают экспериментальные ответные реакции. Реакции моделей M1 и M2 сравниваются. Модель M1 эквивалентна M2, если реакции совпадают, и содержит ошибку в противном случае. Обычно модель M1 – это модель высокого уровня, а M2 – более низкого уровня, чем M1 (ближе к аппаратной реализации). M2 также может быть реальным физическим устройством [4].

В современных технологиях проектирования вместо реального физического устройства используется его представление на ЯОА. То есть в качестве M1 выступает спецификация, в качестве M2 – описание на ЯОА. Данные о разрабатываемом устройстве в спецификации могут быть представлены разными способами, чаще всего неформальными или частично формализованными. Автоматизировать процесс получения эталонных реакций по спецификации можно формализовав процесс получения информации из спецификации.

Это возможно с помощью языков описания спецификаций. Следовательно, верификация M2 сводится к формальному доказательству эквивалентности двух языковых моделей, что есть трудоемкой и не всегда однозначной задачей. Необходимо использовать другие способы верификации. Спецификация, представляющая собой частично формализованное задание закона функционирования устройства, содержит эталонные реакции в неявном виде, а иногда не содержит их вовсе. Разработчики утверждают, что именно такое задание спецификации наиболее распространено [4, 5].

Таким образом, при любом способе верификации по неформальному описанию основной оказывается задача вычисления явных эталонных реакций. Рассмотрим один из возможных способов вычисления эталонных реакций, связанный с уточнением спецификации.

Предполагается, что инженер-разработчик для любых входных данных может вычислить реакции разрабатываемого устройства по спецификации, если данные соответствуют описанным в спецификации состояниям.

Суммируя все вышесказанное, получим следующий обобщенный шаг алгоритма верификации:

1. Получить от инженера-разработчика откомпилированный ЯОА-код, описывающий устройство или его логически (функционально) законченный блок.
2. Выполнить построение тестов по полученному ЯОА-коду путем генерации последовательностей, различающих заданный оператор. Задать контрольные точки.
3. Возвратить полученные тесты инженеру-разработчику для вычисления соответствующих выходных значений и/или значений в заданных контрольных точках.
4. Выполнить моделирование сгенерированных тестов на ЯОА-коде в целях получения экспериментальных значений.
5. Сравнить значения, полученные при моделировании ЯОА-кода, и значения, полученные от инженера-разработчика.
6. При несовпадении сделать вывод об ошибках проектирования в ЯОА-коде.

4 ПОСТРОЕНИЕ ТЕСТОВ

Тест строится активизацией многоуровневых путей в I-графе от внешних входов до соответствующих операторов обработки данных и от них до внешних выходов. В качестве элементарных тестов для операторов обработки данных будем использовать предварительно вычисленные различающие тестовые последовательности [6, 7].

Учитывая, что операторы языка (примитивы) не содержат внутри себя ошибок, очевидно, что подача на примитивы тестов, проверяющих закон их функционирования, является нецелесообразной. Поэтому смысл

тестирования примитива состоит не в проверке функционирования, а идентификации его типа. Таким образом, на примитив необходимо подать такие тестовые наборы, чтобы после анализа реакций на них можно было идентифицировать тип (функцию) примитива и отличить его от других примитивов. Такие различающие последовательности позволяют найти ошибки связанные с заменой операторов в ЯОА-коде. Причем замену логических операторов логическими, арифметическими арифметическими, так как перекрестную замену можно выявить, как правило, в процессе компиляции. Алгоритм генерации тестов состоит в построении различающих тестовых последовательностей для определения данного функционального элемента среди остальных из заданного подмножества, подачи тестов на входы этого элемента и транспортировании результатов до внешних выходов графовой модели или до ближайших контрольных точек.

Вводится понятие наблюдаемой точки, которая отличается от контрольной тем, что точное значение сигнала или переменной в этой точке вычислять нет необходимости. Контрольная точка же, во-первых, наблюдаема, а во-вторых, значения в контрольных точка вычисляются для некоторой последовательности тестов. Наблюдаемая точка – это выход каждого тестируемого оператора, а контрольная точка задается инженером-верификатором.

На внутренней модели тесты строятся по I-графу. На каждую функциональную вершину необходимо подать различающую последовательность и провести реакции этой вершины к ближайшей наблюдаемой точке. Такой вершиной может быть внутренняя точка (переменная или сигнал) или выходная вершина. Таким образом, проверяется каждая функциональная вершина графа: на нее подается тест, а реакции проводятся до наблюдаемой точки. Подавая тесты на каждую вершину и сравнивая реакции в контрольных точках, заданных верификатором, можно, в случае несовпадения реакции, указать на вершину (вершины), в которых происходит искажения данных, вызванные неправильным применением оператора. А так как каждая функциональная вершина соответствует оператору языка, то для разработчика будет указан неправильный оператор или фрагмент кода.

Чтобы различить операцию «плюс» из подмножества арифметических операций {плюс, минус, умножить, разделить}, необходимо на один из входов примитива подать ноль, а на другой вход – значение больше единицы. Далее разбиваем на подмножества {плюс, минус} и {умножить, разделить} таким образом: если результат на предыдущем шаге равен 0, то отнесем тестируемый функционал в подмножество {умножить, разделить}; если не равно 0, то {плюс, минус}. Подаем

на входы одинаковые значения, большие единицы. Если значение не равно 0, то примитив – операция сложения.

Если строить тесты полным перебором всех данных на всех входных сигналах [2], то общая длина минимального полного теста составляет 2^n , где n – общая размерность множества входных сигналов, при условии подачи стабильных 0 и 1. Кроме того, при ориентации на синтез необходимо учитывать то, что на целочисленные сигналы выделяется 32 разряда, если не указан их диапазон. Кроме этого, необходимо организовать пути продвижения значений от функциональных примитивов до внешних выходов, что дает разрастание количества тестовых векторов. Путем использования методики генерации различающих векторов достигается эффект сужения области тестовых последовательностей [8].

5 ОЦЕНКА КАЧЕСТВА ВЕРИФИКАЦИИ

На основе перечня функциональных режимов строится граф режимов, обход которого и дает представление о качестве верификации. Спецификация на устройство задается в неформальном виде, словесной форме. Из спецификации можно выделить функциональные режимы. Каждый набор тестовых векторов также соответствует режиму работы ЦУ. Таким образом, последовательность наборов входных векторов, полученных после построения тестов, позволяет задать путь обхода графа.

Определим количественную оценку качества теста [1].

$$Q = \frac{f_{\text{код}}^{\text{пров}}}{f_{\text{спец}} + f_{\text{доп}}} * 100\%,$$

где $f_{\text{спец}}$ – режимы из спецификации, $f_{\text{доп}}$ – внесенные в спецификацию дополнительные режимы, $f_{\text{код}}^{\text{пров}}$ – проверенные режимы в коде. Применение подобной оценки позволит дать количественную оценку качества проведенной верификации.

ЗАКЛЮЧЕНИЕ

Научная ценность и новизна предложенной методики состоит в разработке нового подхода к функциональной верификации. Генерация тестов проводится для примитивов, которые в данном случае представляют собой операторы языка описания аппаратуры.

Практическая польза от предложенной стратегии состоит в том, что она позволяет верифицировать описание ЦУ на ЯОА-коде и сделать вывод о его соответствии спецификации на функциональном уровне.

ПЕРЕЧЕНЬ ССЫЛОК

1. S. Tasiran, K. Keutzer Coverage Metrics For Functional Validation Of Hardware Designs // Ieee Design & Test Of Computers. – 2001. – July–August. – P. 36–45.
2. Иан Соммервилл. Инженерия программного обеспечения, 6-е издание.: Пер. с англ. – М.: Издательский дом «Вильямс», 2002. – 624 с.
3. F. Corno Et Al, Automatic Test Bench Generation For Validation Of Rt–Level Descriptions: An Industrial Experience IN PROC. – IEEE DATE, 2000. – PP. 385–389.
4. Головач В., Бельшкин А. Проектирование интерфейса как часть разработки ТЗ // Мир ПК. – 2000. С№ 7. – С. 18–24.
5. Боуэн Джонатан П., Хинчи Майкл Дж. Десять заповедей формальных методов // Мир ПК. – 1997. – № 9. – С. 45–48.
6. Рустинов В. А., Сыревич Е. Е., Сыревич А. В. Интервальный метод представления многоразрядных операндов для выполнения импликации при синтезе тестов верификации // Всеукраинский межведомственный НТС «Автоматизированные системы управления и приборы автоматки». Выпуск 122. – Харьков, 2003. – С. 96–103
7. Рустинов В. А., Сыревич Е. Е., Сыревич А. В., Чегликов Д. И. Процедуры импликации на арифметических операциях при синтезе тестов верификации // Радиэлектроника и информатика. Выпуск 130. – Харьков, 2005. – С. 21–26.
8. Krivulya G., Syrevitch Yev., Karasev A., Krasovskaya A. Hdl–Models Verification Strategy // Proceedings of the International conference TCSET'2006, February 28–March 4, 2004, Lviv–Slavsko, Ukraine. – С. 570–573.

Надійшла 23.03.06
Після доробки 12.05.06

Пропонується стратегія верифікації моделей цифрових пристроїв, описаних за допомогою мов опису апаратури. Основна ідея лежить у генерації розрізняючих тестів для окремих функціональних елементів, їхньої суперпозиції у інтерактивному обчисленні еталонних реакцій.

Verification strategy of digital devices models, which are represented with the help of hardware description languages. The main idea stays in distinguishing sequence generation for separate functional elements, their superposition, and interactive ethalon calculation.

УДК 519.713:681.326

В. И. Хаханов, А. Н. Парфентий, Хассан Ктейман, Ваде Гриби

HEDEFS – ПРОГРАММНО-АППАРАТНАЯ СИСТЕМА МОДЕЛИРОВАНИЯ НЕИСПРАВНОСТЕЙ

Предлагается программно-аппаратная реализация дедуктивно-параллельного топологического метода моделирования неисправностей HEDEFS – Hardware Embedded DEductive Fault Simulation, который использует ad hoc технологии анализа сходящихся разветвлений, обратного моделирования дефектов, тестопригодного преобразования схем, ориентированного на обработку цифровых проектов большой размерности вентильного и регистрового уровней описания. Представлены структурные аппаратные решения для реализации метода при оценке качества генерируемых тестов.

ВВЕДЕНИЕ

Необходимость существенного повышения быстродействия средств моделирования для анализа качества тестов определяется возрастающей сложностью проектируемых цифровых систем на кристаллах, насчитывающих миллионы вентилей. Существующие средства моделирования неисправностей ведущих фирм мира, таких как: Cadence, Mentor Graphics, Synopsys, затрачивают на анализ цифровой системы, размерностью порядка миллионов вентилей, несколько часов (рабочая станция с частотой и оперативной памятью в 0,5 ГГц и 0,5 Гбайт). Для проектировщика такие затраты являются весьма значительными. Предлагаемые модели, алгоритмические и аппаратные решения ориентированы на существенное повышение быстродействия средств

моделирования. Они могут быть конкурентоспособными на рынке средств генерации тестов и оценки их качества для цифровых устройств большой размерности.

Объект тестирования – цифровая система, имплементируемая в кристаллы ASIC или PLD, спецификация которой представлена на языке VHDL.

Цель исследования – аппаратная реализация быстродействующего дедуктивного метода моделирования константных неисправностей HEDEFS – Hardware Embedded DEductive Fault Simulation для оценки качества синтезируемых тестов цифровых систем, имплементируемых в кристаллы, содержащие миллионы вентилей.

Задачи исследования: 1) разработка модели дедуктивно-параллельного анализа сходящихся разветвлений (СР); 2) реконфигурирование модели устройства для эффективного обратного моделирования неисправностей; 3) разработка модели дедуктивно-топологического анализа неисправностей древовидных структур (ДС); 4) структурная аппаратная реализация дедуктивно-параллельного метода моделирования неисправностей сходящихся разветвлений; 5) Тестирование и верификация программно-аппаратного продукта HEDEFS.

В основу предлагаемого метода положены: работы, связанные с повышением быстродействия моделирования дефектов [1–3], дедуктивная модель транс.портирова-

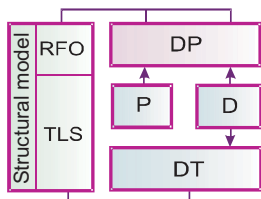


Рисунок 1 – Взаємодія моделей і методів

ния неисправностей [4, 5], параллельный метод обработки списков дефектов функционального элемента [4, 6] и алгоритм обратного моделирования примитивов [7] при обработке древовидных структур дедуктивной модели цифровой системы. На рис. 1 представлена структура взаимодействия методов моделирования неисправностей для обработки моделей цифровой системы, состоящей из двух частей. Для моделирования сходящихся разветвлений (RFO) применяется дедуктивно-параллельный алгоритм [8, 9], P – параллельный, D – дедуктивный. Для анализа древовидных структур (TLS) используется разработанный дедуктивно-топологический метод (DT)

Такое разделение функций обработки цифровых схем большой размерности позволяет на порядок повысить быстродействие моделирования неисправностей по сравнению с дедуктивным и параллельным методами.

1 МОДЕЛЬ ДЕДУКТИВНО-ПАРАЛЛЕЛЬНОГО АНАЛИЗА НЕИСПРАВНОСТЕЙ

Предлагается модель дедуктивно-параллельного синхронного анализа неисправностей, которая позволяет за одну итерацию обработки схемы вычислять все дефекты, проверяемые на двоичном тест-векторе. Она основана на решении следующего уравнения [3, 4]:

$$L = T \oplus F, \tag{1}$$

где $F = (F_{m+1}, F_{m+2}, \dots, F_i, \dots, F_n)(i = \overline{m+1, n})$ – совокупность функций исправного поведения устройства; m – число его входов; $Y_i = F_i(X_{i1}, \dots, X_{ij}, \dots, X_{in_i})$ – n_i -входовой i -й элемент схемы, реализующий F_i для определения состояния линии (выхода) Y_i на тест-векторе T_t , где X_{ij} – j -й вход i -го элемента; двоичный тест $T = (T_1, T_2, \dots, T_t, \dots, T_k)$ – упорядоченная совокупность двоичных векторов, доопределенная в процессе исправного моделирования на множестве входных, внутренних и выходных линий, где тест-вектор задается в виде

$$T_t = (T_{t1}, T_{t2}, \dots, T_{ti}, \dots, T_{tn}), \tag{2}$$

а его n -входная координата вычисляется моделированием функции $T_{ti} = Y_i = F_i(X_{i1}, \dots, X_{ij}, \dots, X_{in_i})$ на

тест-векторе T_t ; $L = (L_1, L_2, \dots, L_t, \dots, L_k)$ – множество дедуктивных схем или моделей, определяемых выражением (1), где $L_t = (L_{t1}, L_{t2}, \dots, L_{ti}, \dots, L_{tn})$;

$$L_{ti} = T_t \oplus F_i \tag{3}$$

– дедуктивная функция (ДФ) параллельного моделирования неисправностей на тест-векторе T_t , соответствующая исправному элементу F_i , которая дает возможность вычислять список входных неисправностей, транспортируемых на выход элемента F_i [10].

Понятие синхронности введенной модели (1) определяется условием: $\Delta t = (t_{j+1} - t_j) \gg \tau \gg \tau_i$, когда интервал времени между сменой входных наборов $(t_{j+1} - t_j)$, подаваемых на схему, много больше максимальной задержки схемы τ и элемента τ_i . Это позволяет исключить время как несущественный параметр [10], что используется в технологиях моделирования и синтеза тестов.

В общем случае, когда функция устройства представлена таблицей истинности, применение формулы (1) позволяет получить для заданного тест-вектора T_t таблицу транспортирования неисправностей, по которой можно записать ДФ моделирования дефектов.

С учетом разбиения теста на составляющие векторы уравнение (1) получения ДФ для $T_t \in T$ принимает следующий вид: $L_t = T_t \oplus F$. Если функциональное описание цифрового устройства представлено компонентами (примитивами), формирующими состояния всех линий схемы, то в качестве формулы преобразования исправной модели примитива F_i на тест-векторе T_t в дедуктивную функцию L_{ti} выступает выражение

$$L_{ti} = T_t \oplus F_i = f_{ti}[(X_{i1} \oplus T_{t1}), (X_{i2} \oplus T_{t2}), \dots, (X_{ij} \oplus T_{tj}), \dots, (X_{in_i} \oplus T_{tn_i})] \oplus T_{ti}, \tag{4}$$

которое является основой дедуктивного анализа цифровых устройств [3, 6].

Пример 1. Получить дедуктивные функции параллельного моделирования неисправностей на исчерпывающем тесте для базиса логических элементов И, ИЛИ, НЕ.

Используя (4), выполняется построение ДФ для логического элемента И:

$$\begin{aligned} L[T = (00, 01, 10, 11), F = (X_1 \wedge X_2)] &= \\ &= L\{(\bar{x}_1 \bar{x}_2 \vee \bar{x}_1 x_2 \vee x_1 \bar{x}_2 \vee x_1 x_2) \wedge \\ &\wedge [(X_1 \oplus T_{t1} \wedge X_2 \oplus T_{t2}) \oplus T_{t3}]\} = \\ &= (\bar{x}_1 \bar{x}_2)\{[X_1 \oplus 0] \wedge (X_2 \oplus 0) \oplus 0\} \vee \\ &\vee (\bar{x}_1 x_2)\{[X_1 \oplus 0] \wedge (X_2 \oplus 1) \oplus 0\} \vee \\ &\vee (x_1 \bar{x}_2)\{[X_1 \oplus 0] \wedge (X_2 \oplus 0) \oplus 0\} \vee \\ &\vee (x_1 x_2)\{[X_1 \oplus 0] \wedge (X_2 \oplus 1) \oplus 1\} = \\ &= (\bar{x}_1 \bar{x}_2)(X_1 \wedge X_2) \vee (\bar{x}_1 x_2)(X_1 \wedge \bar{X}_2) \vee \\ &\vee (x_1 \bar{x}_2)(\bar{X}_1 \wedge X_2) \vee (x_1 x_2)(X_1 \wedge X_2). \end{aligned}$$

Аналогичные преобразования для логического элемента ИЛИ дают следующий результат:

$$L[T = (00, 01, 10, 11), F = (X_1 \wedge X_2)] = (\bar{x}_1 \bar{x}_2)(X_1 \wedge X_2) \vee (\bar{x}_1 x_2)(\bar{X}_1 \wedge X_2) \vee (x_1 \bar{x}_2)(X_1 \wedge \bar{X}_2) \vee (x_1 x_2)(X_1 \wedge X_2).$$

Здесь $T_t = (T_{t1}, T_{t2}, T_{t3})$, $(t = \overline{1, 4})$ – тест-вектор, имеющий 3 координаты, где последняя определяет состояние выхода элемента И (ИЛИ).

В следующем преобразовании для инвертора тест-вектор имеет 2 координаты: #, где последняя – состояние выхода элемента:

$$L[T = (0, 1), F = \bar{X}_1] = L\{(\bar{x}_1 \vee x_1)[(\bar{X}_1 \oplus T_{t1}) \oplus T_{t2}]\} = \bar{x}_1 \wedge [(\bar{X}_1 \oplus 0) \oplus 1] \vee x_1 [(\bar{X}_1 \oplus 1) \oplus 0] = \bar{x}_1 \bar{X}_1 \vee x_1 \bar{X}_1 = \bar{x}_1 X_1 \vee x_1 X_1.$$

Данное выражение иллюстрирует несущественность операции инверсии на выходе элемента для транспортирования дефектов. Поэтому данная функция не фигурирует на выходах дедуктивных элементов.

2 АППАРАТНАЯ СТРУКТУРА ДЕДУКТИВНО-ПАРАЛЛЕЛЬНОГО МОДЕЛИРОВАНИЯ

Аппаратная реализация ДФ для двухвходовых элементов И, ИЛИ на исчерпывающем тесте представлена на рис. 2 схемой дедуктивно-параллельного анализа неисправностей. В симуляторе представлены булевы (x_1, x_2) и регистровые (X_1, X_2) переменные, сигнал V выбора типа исправной функции: $V = 0$ (AND), $V = 1$ (OR), выходная регистровая переменная Y . Состояния двоичных входов x_1, x_2 и V выбирают одну из четырех дедуктивных функций для получения вектора Y проверяемых неисправностей.

Работа данной схемы представлена в таблице параллельного моделирования входных 4-разрядных векторов неисправностей в целях получения на выходе Y множества проверяемых дефектов для логических элементов 2И, 2ИЛИ:

V	$x_1 x_2$	X_1	X_2	Y
0	00	0111	1011	0011
1	00	0111	1011	1111
0	11	0101	0110	0111
1	11	1101	0111	0101

Применение такого симулятора дает возможность трансформировать вентильную модель F исправного поведения схемы в дедуктивную L , которая инвариантна в смысле универсальности тестовым наборам и не

предполагает в процессе моделирования использовать модель F . Поэтому симулятор, как аппаратная модель ДФ, ориентирован на создание встроенных средств дедуктивно-параллельного моделирования, повышающих быстродействие анализа в 10–1000 раз по сравнению с программной реализацией. Но при этом соотношение объемов вентильных (после синтеза) моделей исправного моделирования и анализа неисправностей составляет 1:16. Подход аппаратного анализа неисправностей направлен на расширение функциональных возможностей встроенных средств исправного моделирования (HES™ – Hardware Embedded Simulator) фирмы Aldes [11]. Вычислительная сложность обработки цифрового устройства, состоящего из n вентилях, равна $Q = (2n^2\tau)/W$, где τ – время выполнения регистровой операции (AND, OR, NOT); W – разрядность регистра.

Другое интересное направление применения симулятора связано с анализом сходящихся разветвлений, как первой и времязатратной фазы дедуктивно-топологического метода, рассмотренного ниже. Для аппаратной реализации дедуктивно-параллельного моделирования на основе предложенного симулятора используется структура, представленная на рис. 3. Особенность

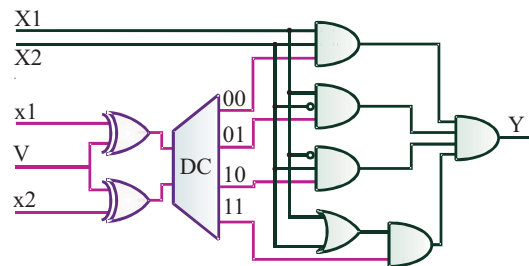


Рисунок 2 – Симулятор неисправностей

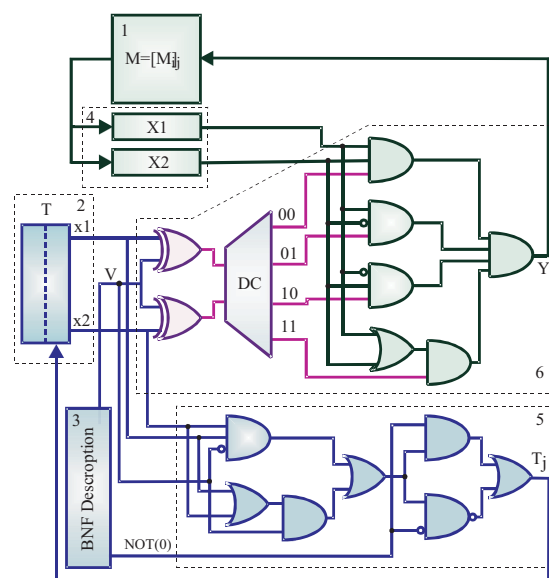


Рисунок 3 – HEDEFS-структура

схемной реализации заключается в совместном выполнении двух операций: однобитовой – для эмуляции элементов И, ИЛИ и параллельной – для обработки многоэлектрических векторов неисправностей. Функциональное назначение основных блоков: 1) $M = [M_{ij}]$ – прямоугольная матрица моделирования неисправностей, где $i = 1, p$ – параметр количества сходящихся разветвлений; $j = 1, q$ – общее число линий в обрабатываемой схеме; 2) векторы сохранения состояний исправного моделирования, определенные в моменты времени $t - 1$ и t , необходимые для формирования дедуктивных функций примитивов; 3) модуль памяти для хранения схемного описания в виде структуры логических элементов; 4) буферные регистры, размерностью p , для хранения операндов и выполнения регистровых параллельных операций над векторами неисправностей, считанных из матрицы M ; 5) блок исправного моделирования для определения булевого состояния выхода очередного обрабатываемого логического элемента; 6) дедуктивно-параллельный симулятор, обрабатывающий за один такт две регистровых переменных $X1$ $X2$ в целях определения вектора дефектов, транспортируемых на выход логического элемента Y .

Достоинство предложенной структуры моделирования неисправностей заключается в следующем: 1) существенное уменьшение количества моделируемых дефектов, определяемое только числом сходящихся разветвлений, которое составляет до 20 % от общего числа линий; 2) снижением объема памяти, необходимого для хранения матрицы моделируемых неисправностей; 3) достаточно простой реализацией HEDEFs-структуры в аппаратном исполнении, что позволяет на порядок увеличить быстродействие моделирования неисправностей; 4) использованием HEDEFs в качестве первой фазы дедуктивно-топологического метода, который основывается на результате обработки сходящихся разветвлений для быстродействующего анализа древовидных структур.

3 ДЕДУКТИВНО-ТОПОЛОГИЧЕСКИЙ МЕТОД МОДЕЛИРОВАНИЯ

Использованию метода обратного моделирования неисправностей [7] без фазы дедуктивно-параллельного анализа препятствует невозможность выполнения одномерной активизации дефектов, которая может быть проиллюстрирована двумя известными вариантами некорректностей, возникающих в процессе анализа схемы (рис. 4) на тест-векторе 101.

В левой схеме константная неисправность 2^1 ложно проверяется на одномерном пути 2–5–6 при выполнении обратного моделирования. Однако при этом не учитывается тот факт, что данный дефект изменяет состояние линии 4 с 1 на 0, что создает условия запрета транспортирования неисправности на выход схемы.

Поэтому неисправность на линии $a-2^1$ является ложно проверяемой на наборе (101) при рассмотрении ее транспортирования по одномерному пути активизации. Правая схема является примером альтернативной ситуации – к выходу 6 нет одномерного пути транспортирования дефектов с внешних входов и, в частности, от линии b . Тем не менее, неисправность 2^1 , поскольку ее присутствие создает кратный дефект на линиях 4 и 5, который изменяет состояние выхода. В обоих случаях имеется существенная некорректность, которая не позволяет использовать обратное моделирование дефектов путем рассмотрения одномерных путей

Таким образом, некорректность метода связана исключительно с одномерной активизацией неисправностей сходящихся разветвлений. Выходом из данной ситуации является моделирование на первой стадии только сходящихся разветвлений, которые далее исключаются из обработки. Вторая стадия заключается в обратном одномерном моделировании древовидной структуры схемы, где вершинами деревьев являются внешние выходы и уже обработанные сходящиеся разветвления.

Далее рассматривается дедуктивно-топологический (ДТ) метод обратного моделирования неисправностей в привязке к топологии дедуктивной модели цифрового устройства. Основные стадии алгоритма обработки представлены на рис. 5.

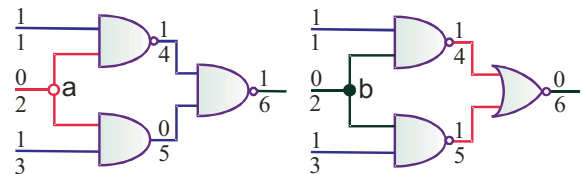


Рисунок 4 – Ложная проверка и непроверка дефектов

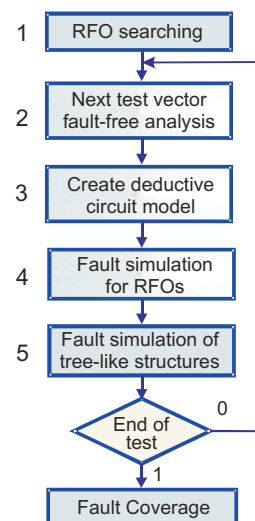


Рисунок 5 – Стадии ДТ-моделирования

1. Препроцессор структурного анализа модели схемы – осуществляет поиск СР в схеме цифрового устройства. Вычислительная сложность данной графо-теоретической задачи $Q_r = n^2$, но она является разовой и поэтому практически не влияет на быстродействие моделирования входных наборов в целом.

2. Фаза исправного моделирования цифровой схемы. Предназначена для определения реакции всех невходных линий устройства на входной набор $T_t \in T = [T_{ti}]$.

3. Преобразование схемы на текущем тест-векторе в дедуктивную модель по правилам (4).

4. Моделирование неисправностей сходящихся разветвлений по дедуктивной модели схемы, количество которых значительно меньше остальных линий.

5. Обратное одномерное моделирование дефектов древовидных подграфов по дедуктивной модели устройства.

Теоретическим обоснованием применения данного пункта является доказательство следующих теорем и формулировка следствий из них.

Лемма. Сходящиеся разветвления в комбинационной схеме являются причиной появления кратных дефектов на входах примитивов.

Теорема 1. Для древовидной структуры дедуктивной схемы инверсный вход примитива И: $\bar{X}_i \in L$ запрещает транспортирование всех неисправностей, принадлежащих линиям-предшественникам.

Доказательство. Во-первых, инверсия на входе может иметь место на примитиве, имеющем две и более входные линии, во-вторых, она означает вычитание списка неисправностей, принадлежащего данному входу, учитывая изоморфизм $\bar{X}_i X_j \approx X_j / X_i$. Но согласно лемме, пересечение двух списков линий-предшественников, относящихся ко входам одного примитива, равно пустому множеству (7). Следовательно, вычитание списка линий-предшественников никогда не уменьшит мощность любого другого списка, относящегося к неинверсному входу рассматриваемого элемента. Таким образом, для древовидных структур комбинационных схем всегда будет выполняться условие:

$$f^*(X_i) / f^*(X_j) = f^*(X_i) \leftarrow f^*(X_i) \cap f^*(X_j) = \emptyset.$$

Теорема 2. Если в древовидной структуре дедуктивной схемы L существует примитив И, имеющий более одного неинверсного входа $X_i X_j$, то такой элемент запрещает транспортирование всех неисправностей, принадлежащих линиям-предшественникам всех его входов.

Доказательство. Наличие двух и более неинверсных, прямых входов в дедуктивном элементе означает пересечение списков линий, являющихся предшественниками для рассматриваемых входов. Но согласно лемме, входы одного примитива ДС не имеют общих предшественников, то отсюда следует, что пересечение упо-

мянутых списков будет всегда равно пустому множеству:

$$f^*(X_i) \cap f^*(X_j) = \emptyset \leftarrow X_i X_j.$$

Если же дедуктивный элемент имеет входы с инверсией, то по теореме 1 вычитание из полученного пустого множества предшественников любого непустого, принадлежащего входу с инверсией, дает также пустой результат.

Следовательно, дедуктивный элемент И, имеющий более одного неинверсного входа, не пропустит через себя неисправности линий предшественников.

4 АЛГОРИТМ АНАЛИЗА ДРЕВОВИДНЫХ СТРУКТУР

1. Обработка примитивов. Все входы дедуктивного примитива И определяются инверсными, если он имеет более одного неинверсного входа. Сходящееся разветвление отмечается знаком инверсии, если его неисправность не проверяется на тест-векторе. Сходящееся разветвление следует рассматривать как наблюдаемый выход схемы, если его неисправность проверяется на тест-векторе.

2. Обработка структур. Инверсный вход дедуктивного элемента И запрещает активизацию всех неисправностей линий-предшественников (отмечаются знаками инверсии), относящихся к данному входу. Дедуктивный элемент И пропустит неисправности линий-предшественников только по единственному неинверсному входу. Дедуктивный элемент ИЛИ не может иметь инверсных входов.

3. Формирование результата топологического моделирования: неисправности линий, не отмеченные знаками инверсий на дедуктивной модели схемы, проверяются.

В качестве иллюстрации основных шагов алгоритма анализа древовидных структур предлагаются два примера.

Пример 2. Дано цифровое устройство (рис. 6, первая схема), содержащее 3 сходящихся разветвления. Входной вектор: 1011111. Определить проверяемые на нем дефекты константного типа. Состояния линий исправного моделирования представлены в скобках. Результат преобразования исправной модели устройства в дедуктивную и параллельное моделирование неисправностей линий СР зафиксирован на второй схеме. Для моделирования был использован дедуктивно-параллельный метод, который сформировал следующую матрицу.

М	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16
02	.	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1
13	1	1	1	1

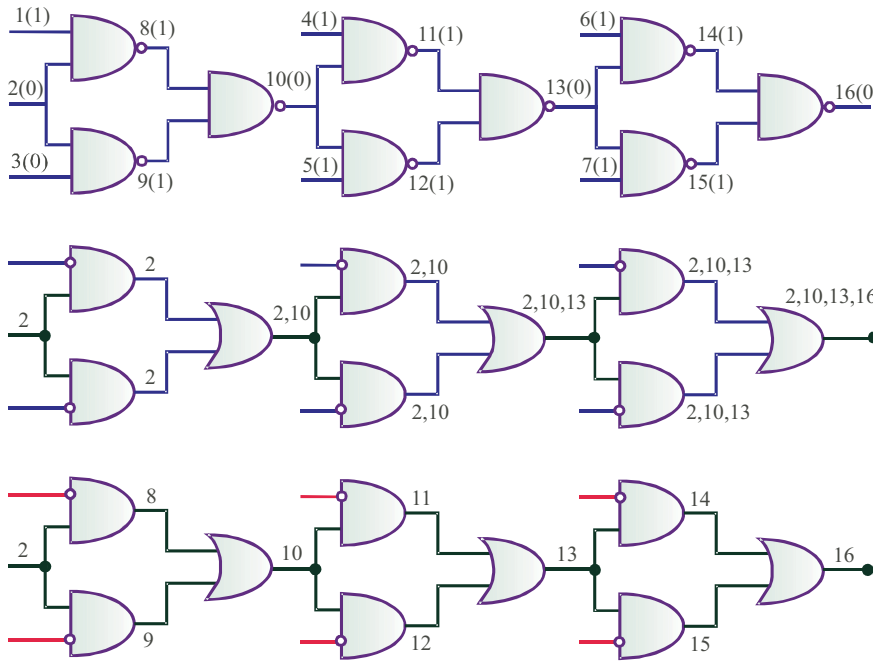


Рисунок 6 – Моделирование схемы с разветвлениями

Все дефекты разветвлений (2, 10, 13), инверсные по отношению к исправному состоянию этих линий, являются проверяемыми. Этот факт отмечен на схеме черными кружками.

Алгоритм анализа древовидных структур в целях определения списка проверяемых дефектов заключается в идентификации непроверяемых входных линий знаками инверсии (прозрачные кружки). При этом черные кружки есть проверяемые линии разветвлений, которые имеют статус наблюдаемых выходов (см. рис. 6, третья схема). Таким образом, неисправности линий {2, 8, 9, 10, 11, 12, 13, 14, 15, 16}, не отмеченные знаками инверсии, проверяются на заданном тест-векторе.

Пример 3. Определить список проверяемых на тест-векторе 10111 неисправностей для схемы, представленной на рис. 7, содержащей 2 сходящихся разветвления.

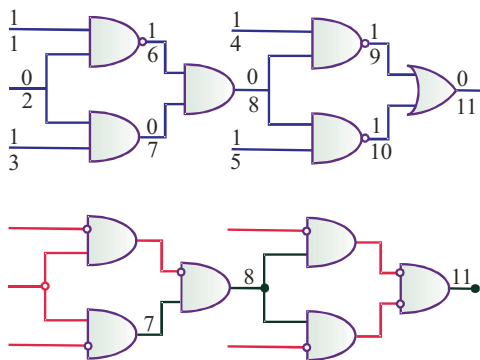


Рисунок 7 – Схема с двумя разветвлениями

Здесь неисправности линий 9 и 10 не проверяются. Тем не менее дефекты линий 7 и 8, обнаруживаются, благодаря наличию проверяемого сходящегося разветвления 8, дефект которого транспортируется на внешний выход 11.

ЗАКЛЮЧЕНИЕ

Предложенный программно-аппаратный дедуктивно-топологический метод моделирования неисправностей ориентирован на обработку цифровых систем на кристаллах, содержащих миллионы вентилях. На рис. 8 представлена структура системы HEDEFS, ориентированная на раздельную обработку RFO- и TLS-моделей сложных цифровых систем. Hardware часть состоит из собственно симулятора, модели схемы для обработки сходящихся разветвлений и тестовых последовательностей, хранящихся с специальной памяти. Более детализированная структура аппаратной части была приведена ранее (см. рис. 3). Software часть состоит из дедуктивно-топологического (DT) симулятора, TLS-модели, ориентированной на обработку древовидных структур и тестовых наборов. Вся система работает под управлением MS Windows, системы моделирования Riviera 7.1 фирмы Aldec и PLI модуля, ориентированного на отладку структур данных в процессе подготовки моделей. Hardware часть соединяется с Software-симулятором посредством PCI-интерфейса. На стадии подготовки аппаратной RFO-модели цифрового устройства выполняется ввод, отладка, синтез и созда-

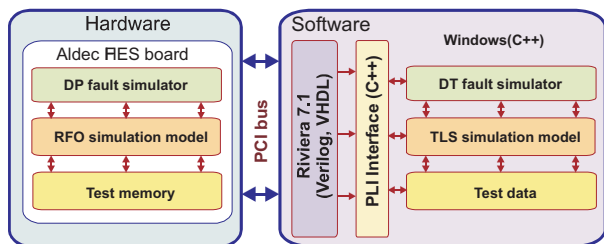


Рисунок 8 – Структура Hardware-Software моделирования дефектов

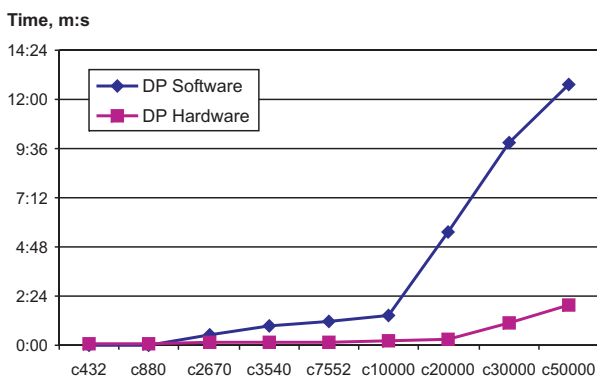


Рисунок 9 – Анализ быстродействия средств моделирования

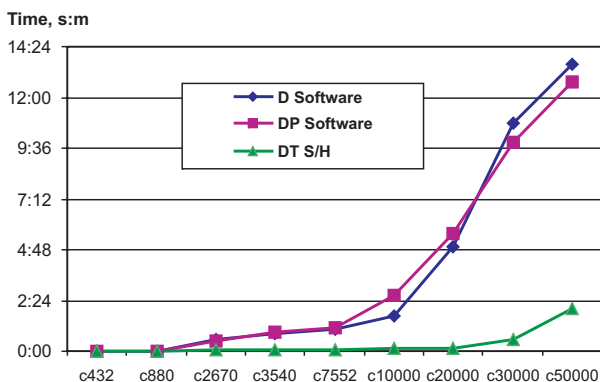


Рисунок 10 – Анализ быстродействия методов моделирования

ние файла конфигурации для программирования FPGA с помощью средств: Riviera 7.1, Snplify фирмы Sinplify, ISE фирмы XILINX, работающих под управлением пакета DVM фирмы Aldec [12–14].

Тестовые эксперименты и верификация программно-аппаратной реализации метода на цифровых комбинационных и последовательностных схемах дали хорошие результаты по быстродействию по сравнению с тради-

ционными алгоритмами параллельного и дедуктивного моделирования. Примеры сравнительного анализа быстродействия аппаратной и программной реализации ДП-метода (обработка тест-примеров на 1000 входных последовательностей, IBM PC 500 МГц, 256 Мбайт) показаны на рис. 9. Ускорение моделирования составляет один два порядка. На рис. 10 представлены результаты анализа быстродействия программно-аппаратной реализации ДТ-метода моделирования цифровых систем по сравнению с дедуктивным алгоритмом на одном и том же компьютере при обработке 1000 векторов. Также показано преимущество ДТ-метода перед дедуктивно-параллельным [8, 9]. Выигрыш в быстродействии более существенен для схем большой размерности. Число сходящихся разветвлений в тест-схемах в среднем составляет 20 % от общего количества линий.

Таким образом, основным результатом данной работы является усовершенствование дедуктивного метода моделирования неисправностей цифровых систем, заключающееся в:

- 1) предложении стратегии раздельного моделирования неисправностей сходящихся разветвлений и древовидных структур, что дает возможность существенно (в десятки раз) повысить быстродействие средств анализа качества тестов для проектируемых цифровых систем;
- 2) создании обобщенной модели процесса дедуктивно-параллельного анализа цифровых схем, имеющей вычислительную сложность n^2 от числа линий схемы для моделирования неисправностей сходящихся разветвлений;
- 3) разработке топологического алгоритма обратного моделирования неисправностей для древовидной структуры цифрового устройства, имеющего линейную вычислительную сложность от числа линий схемы;
- 4) аппаратной реализации дедуктивно-параллельного анализа сходящихся разветвлений, который является наиболее времязатратным компонентом моделирования в системе HEDEFs.

ПЕРЕЧЕНЬ ССЫЛОК

1. Wang X., Hill F. G., Mi Zh. A sequential circuit fault simulation by surrogate fault propagation // Proc. IEEE International Test Conference. IEEE Computer society. – 1989. – P. 9–18.
2. Reduced fault simulator / Nishida T., Miyamoto S., Kozawa T., Satoh K. RFSIM // IEEE Transactions on computer-aided design. – 1987. – Vol. CAD-6, No. 3. – P. 392–402.
3. Hahanov V. I., Babich A. V., Hyduke S. M. Test Generation and Fault Simulation Methods on the Basis of Cubic Algebra for Digital Devices // Proceedings of the Euro-micro Symposium on Digital Systems Design DSD2001. – Warsaw. – 2001. – P. 228–235.
4. Хаханов В. И., Хак Х. М. Джахирул, Масуд М. Д. Мехеди. Модели анализа неисправностей цифровых систем на основе FPGA, CPLD // Технология и конструирование в электронной аппаратуре. – 2001. – № 2. – С. 3–11.
5. Levendel Y. H., Menon P. R. Comparison of fault simulation methods – Treatment of unknown signal values // Journal of digital systems. – 1980. – Vol. 4. – P. 443–59.

6. Abramovici M., Breuer M. A. and Friedman A. D. Digital systems testing and testable design.– Computer Science Press. – 1998. – 652 p.
7. Убар Р. Р. Анализ диагностических тестов для комбинационных цифровых схем методом обратного прослеживания неисправностей // Автоматика и телемеханика. – 1977. – № 8. – С. 168–176.
8. Hahanov V. I., Sysenko I. Y., Skvortsova O. B. Test Generator Used Genetic Algorithms and Reconfigurable Deductive-Parallel Fault Simulation Method for Digital Devices // 9-th International Conference. Mixed Design of Integrated Circuits and Systems. – Wroclaw (Poland). – 2002. – P. 549–554.
9. V. Hahanov, R. Ubar and S. Hyduke. Back-Traced Deductive-Parallel Fault Simulation for Digital Systems // Proceedings of Euromicro Symposium on Digital Systems Design.– Belek-Antalya (Turkey). – 2003. – P. 370–377.
10. Автоматизированное проектирование цифровых устройств / С. С. Бадулин, Ю. М. Барнаулов и др./ Под ред. С. С. Бадулина. М.: Радио и связь. – 1981. – 240 с.
11. Active-HDL User's Guid. Second Edition // Aldec Inc.– 1999. – 213p.
12. Riviera 2006.02, DVM. Режим доступа: <http://www.aldec.com/products/hes/pages/designverificationmanager>. – Загл. с экрана.
13. Synplify, Simplicity. Режим доступа: <http://www.synplicity.com/products/synplifypremier/index.html>. – Загл. с экрана.
14. ISE XILINX. Режим доступа: http://www.xilinx.com/ise/design_tools/leadership.htm. – Загл. с экрана.

Надійшла 23.03.06

Представлено швидкодійний метод моделювання несправностей, який використовує переваги дедуктивного та паралельного алгоритмів моделювання несправностей і орієнтований на обробку цифрових схем, представлених на вентильному рівні або рівні регістрових передач. Запропонований метод дозволяє підвищити швидкість моделювання несправностей у десять разів порівняно з алгоритмами, що використовують методи прямого розповсюдження сигналів.

Fast fault simulation method integrated the advantages of deductive and parallel fault simulation algorithms and oriented on evaluation of digital circuit represented on gate or RTL description level is offered. The speed up of backward fault simulation is better on 10 times than methods with forward propagation algorithms.

НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

НЕЙРОИНФОРМАТИКА И ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

NEUROINFORMATICS AND INTELLIGENT SYSTEMS

УДК 519.7:007.52

Е. В. Бодянский, Е. А. Винокурова, Н. С. Ламонова

ДВОЙНОЙ ВЭЙВЛЕТ-НЕЙРОН И АЛГОРИТМ ЕГО ОБУЧЕНИЯ

В статье предложена архитектура двойного вэйвлет-нейрона, который является модификацией стандартного вэйвлет-нейрона, и алгоритм обучения его параметров. Предложенная архитектура позволяет улучшить аппроксимирующие свойства вэйвлет-нейрона. Проведено имитационное моделирование предложенного двойного вэйвлет-нейрона и алгоритма его обучения с использованием нестационарного хаотического временного ряда.

ВВЕДЕНИЕ

Последнее время в задачах анализа и обработки нестационарных сигналов произвольной природы в условиях неопределенности все шире применяются методы вычислительного интеллекта, а именно гибридные нейронные сети. Одной из важных задач, связанных с обработкой сигнала, является прогнозирование и эмуляция состояния динамических нестационарных систем в будущие моменты времени.

Для решения такого рода задач используется большое количество архитектур нейронных сетей, в том числе и гибридных структур, однако эти системы либо громоздки по своей архитектуре (например, многослойный персептрон), либо плохо приспособлены для обучения в реальном времени. В большинстве случаев ак-

тивационными функциями таких сетей являются гипер-базисные функции, сплайны, полиномы и радиально-базисные функции.

На сегодня широкое распространение получила также теория вэйвлет-анализа [1–3], которая позволяет с высокой точностью выявлять локальные особенности нестационарных сигналов. На стыке этих двух подходов и возникли так называемые гибридные вэйвлет-нейронные сети [4–18] благодаря своим высоким аппроксимирующим свойствам и чувствительностью к изменениям характеристик анализируемых процессов.

Весьма привлекательной с точки зрения технической реализации, обеспечиваемой точности и простоты обучения является так называемый вэйвлет-нейрон, предложенный и описанный в [19–21]. При этом вэйвлет-функции реализованы либо на уровне синаптических весов, либо на выходе нейрона, а для обучения используется градиентный алгоритм обучения с постоянным шагом. Для улучшения аппроксимирующих свойств и ускорения процесса обучения в настоящей работе введена конструкция, названная нами двойным вэйвлет-нейроном, и алгоритм обучения, обладающий как сглаживающими, так и следящими свойствами.

1 ВЭЙВЛЕТ-АКТИВАЦИОННАЯ ФУНКЦИЯ

В качестве активационных функций двойного вэйвлет-нейрона можно использовать различные виды аналитических вэйвлетов. Одними из наиболее интересных по своим свойствам являются два семейства: POLYWOG-вэйвлеты и RASP-вэйвлеты.

Семейство вэйвлетов RASP – вэйвлеты на основе рациональных функций (Rational functions with Second-order Poles – RASP), связанные с теоремой о вычетах комплексных переменных [4].

На рис. 1 представлены два типичных представителя материнских вэйвлетов RASP, которые описываются выражениями

$$\phi_{ji}^1(x_i(k)) = \frac{\beta^1 \cos(x_i(k))}{x_i^2(k) + 1}, \quad \beta^1 = 2,7435, \quad (1)$$

$$\phi_{ji}^2(x_i(k)) = \frac{\beta^2 \sin(\pi x_i(k))}{x_i^2(k) - 1}, \quad \beta^2 = 0,6111. \quad (2)$$

Эти вэйвлеты являются вещественными нечетными функциями с нулевым средним.

Еще одно достаточно широкое семейство вэйвлетов можно получить из полиномиальных оконных гауссовых функций (POLYNomials Windowed with Gaussians type of function – POLYWOG) [4]. Интересно заметить, что производные от этих функций также являются вэйвлетами POLYWOG и могут использоваться в качестве материнских вэйвлетов.

На рис. 2 представлено несколько типичных вэйвлетов из семейства POLYWOG, описываемых выражениями вида

$$\begin{aligned} \phi_{ji}^1(x_i(k)) &= \mu^1 x_i(k) \exp\left(\frac{-x_i^2(k)}{2}\right), \\ \mu^1 &= \exp\left(\frac{-1}{2}\right), \end{aligned} \quad (3)$$

$$\begin{aligned} \phi_{ji}^2(x_i(k)) &= \mu^2 (x_i^3(k) - 3x_i(k)) \exp\left(\frac{-x_i^2(k)}{2}\right), \\ \mu^2 &= 0,7246, \end{aligned} \quad (4)$$

$$\begin{aligned} \phi_{ji}^3(x_i(k)) &= \mu^3 (x_i^4(k) - 6x_i^2(k) + 3) \exp\left(\frac{-x_i^2(k)}{2}\right), \\ \mu^3 &= \frac{1}{3}, \end{aligned} \quad (5)$$

$$\phi_{ji}^4(x_i(k)) = (1 - x_i^2(k)) \exp\left(\frac{-x_i^2(k)}{2}\right). \quad (6)$$

Некоторые вэйвлеты семейства POLYWOG могут быть получены с помощью простых генераторов. Так, в частности, вэйвлеты этого семейства могут быть сгенерированы с учетом свойств эрмитовости производной полинома и функции Гаусса.

2 СТРУКТУРА ДВОЙНОГО ВЭЙВЛЕТ-НЕЙРОНА

Введем в рассмотрение структуру двойного вэйвлет-нейрона, приведенную на рис. 3. Как видно, двойной вэйвлет-нейрон достаточно близок по конструкции к *n*-входовому вэйвлет нейрону [19–21], однако содержит нелинейные вэйвлет-функции на уровне синаптических весов, так и на выходе структуры.

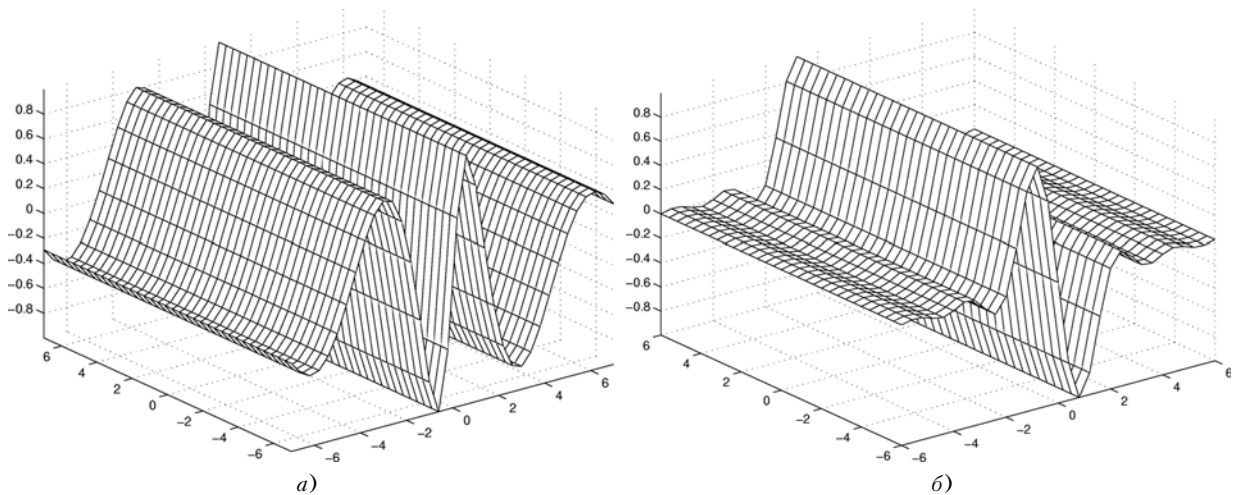


Рисунок 1 – Представители семейства вэйвлетов RASP

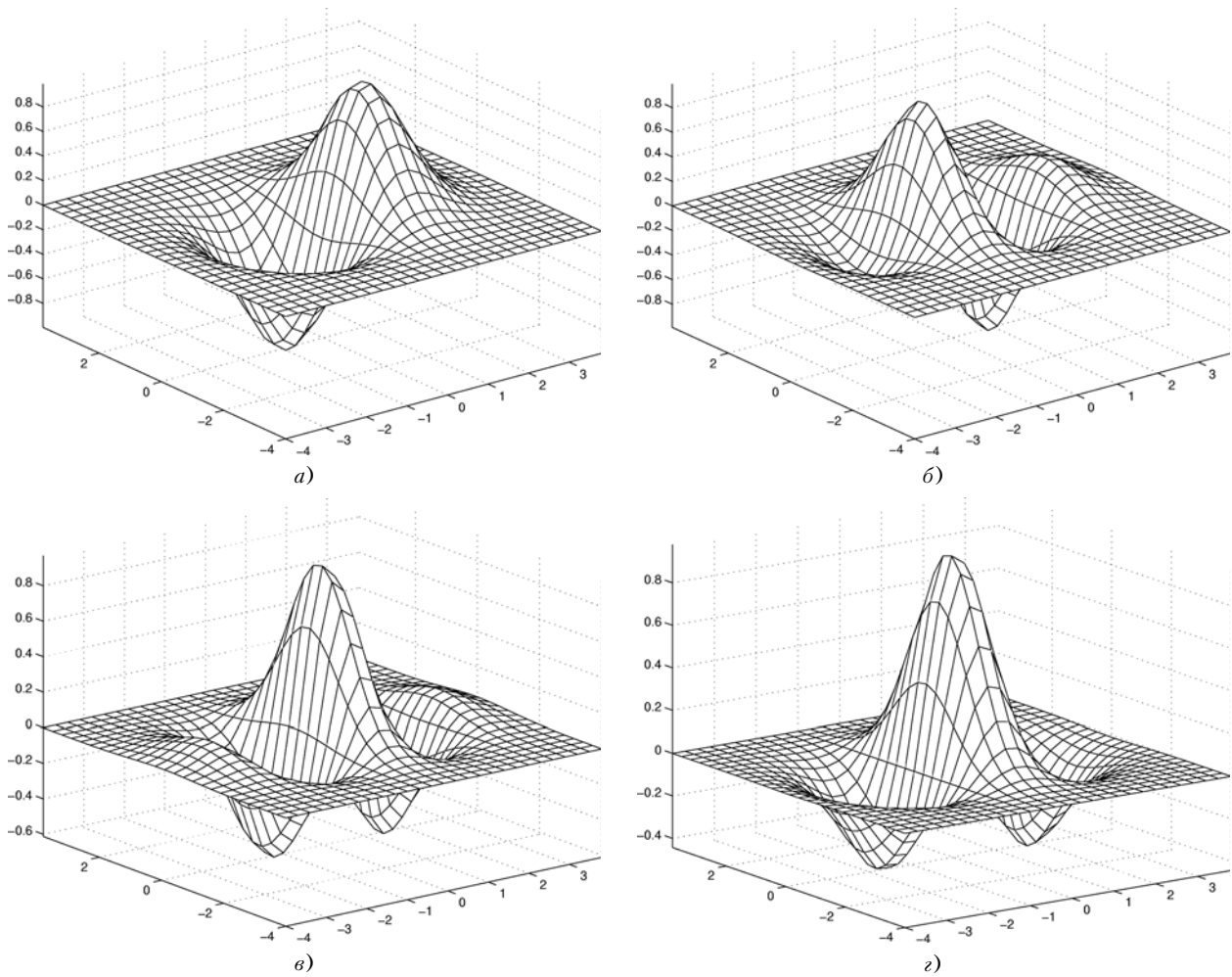


Рисунок 2 – Представители семейства вэйвлетов POLYWOG

При подаче на вход двойного вэйвлет-нейрона, приведенного на рис. 4, векторного сигнала $x(k) = (x_1(k), x_2(k), \dots, x_n(k))^T$ (здесь $k = 0, 1, 2, \dots$ – номер наблюдения в обучающей выборке или текущее дискретное время) на его выходе появляется скалярный сигнал вида

$$\begin{aligned}
 y(k) &= f_0 \left(\sum_{i=1}^n f_i(x_i(k)) \right) = f_0(u(k)) = \\
 &= \sum_{l=0}^{h_2} \varphi_{l0} \left(\sum_{i=1}^n \sum_{j=0}^{h_1} \varphi_{ji}(x_i(k)) \omega_{ji}(k) \right) \omega_{j0} = \\
 &= \sum_{l=0}^{h_2} \varphi_{l0}(u(k)) \omega_{l0}(k), \quad (7)
 \end{aligned}$$

определяемый как настраиваемыми синаптическими весами $\omega_{ji}(k)$, ω_{l0} , так и значениями используемых

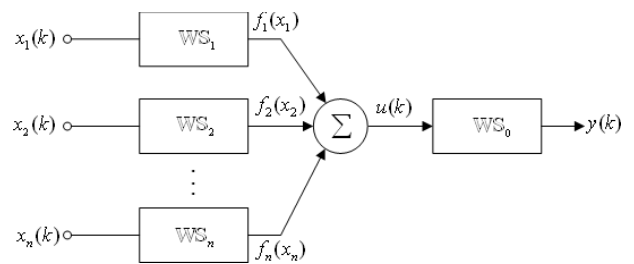


Рисунок 3 – Обобщенная структура двойного вэйвлет-нейрона

вэйвлет-функций $\varphi_{ji}(x_i(k))$, $\varphi_{l0}(u(k))$, при этом полагается $\varphi_{00}(\bullet) = \varphi_{0i}(\bullet) \equiv 1$.

Двойной вэйвлет-нейрон состоит из двух слоев: скрытого слоя, в котором n вэйвлет-синапсов по h_1 вэйвлет-функций в каждом, и выходного слоя, состоящего из одного вэйвлет-синапса с h_2 вэйвлет-функциями.

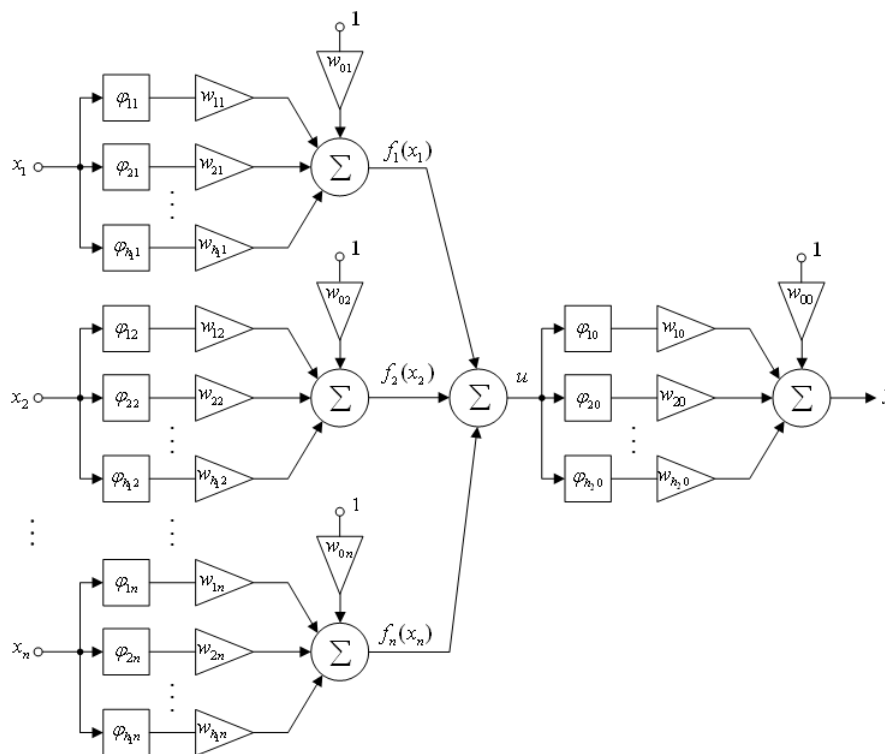


Рисунок 4 – Архитектура двойного вэйвлет-нейрона с нелинейными вэйвлет-синапсами

В каждом вэйвлет-синапсе реализованы вэйвлеты, отличающиеся между собой параметрами растяжения и смещения.

3 СИНТЕЗ АЛГОРИТМА ОБУЧЕНИЯ ДВОЙНОГО ВЭЙВЛЕТ-НЕЙРОНА

Для обучения выходного слоя двойного вэйвлет-нейрона будем использовать критерий вида

$$E(k) = \frac{1}{2}(d(k) - y(k))^2 = \frac{1}{2}e^2(k), \quad (8)$$

где $d(k)$ – внешний обучающий сигнал.

Алгоритм обучения выходного слоя двойного вэйвлет-нейрона на основе градиентного подхода имеет вид

$$w_{j0}(k+1) = w_{j0}(k) + \eta_0(k)e(k)\varphi_{j0}(u(k)), \quad (9)$$

или в векторной форме

$$w_0(k+1) = w_0(k) + \eta_0(k)e(k)\varphi_0(u(k)), \quad (10)$$

где $w_0(k) = (w_{10}(k), w_{20}(k), \dots, w_{h_0}(k))^T$ – вектор си-наптических весов, $\varphi_0(k) = (\varphi_{10}(k), \varphi_{20}(k), \dots, \varphi_{h_0}(k))^T$ – вектор вэйвлет-активационных функций, $e(k)$ – ошиб-

ка обучения, $\eta_0(k)$ – шаг обучения, подлежащий опре-делению.

Для увеличения скорости сходимости процесса обу-чения следует перейти от градиентных процедур к алго-ритмам второго порядка, наибольшее распространение среди которых получил алгоритм Левенберга-Марк-вардта.

После несложных преобразований [6] можно по-лучить алгоритм обучения вида

$$\begin{cases} w_0(k+1) = w_0(k) + \frac{e(k)\varphi_0(u(k))}{\gamma_i^{w_0}(k)}, \\ \gamma_i^{w_0}(k+1) = \alpha\gamma_i^{w_0}(k) + \|\varphi_0(u(k+1))\|^2, \end{cases} \quad (11)$$

где α – параметр забывания устаревшей информации ($0 < \alpha < 1$).

Обучение скрытого слоя проводится аналогичным образом на основе обратного распространения ошиб-ки с использованием того же критерия, записанного в форме

$$E(k) = \frac{1}{2}(d(k) - f_0(u(k)))^2 = \frac{1}{2} \left(d(k) - f_0 \left(\sum_{i=1}^n \sum_{j=0}^{h_1} \varphi_{ji}(x_i(k))w_{ji}(k) \right) \right)^2. \quad (12)$$

Алгоритм обучения скрытого слоя двойного вэйвлет-нейрона на основе градиентной оптимизации имеет вид

$$w_{ji}(k+1) = w_{ji}(k) + \eta(k)e(k)f_0'(u(k))\varphi_{ji}(x_i(k)), \quad (13)$$

или в векторной форме

$$w_i(k+1) = w_i(k) + \eta(k)e(k)f_0'(u(k))\varphi_i(x_i(k)), \quad (14)$$

где $w_i(k) = (w_{1i}(k), w_{2i}(k), \dots, w_{h_i i}(k))^T$ – вектор си-
наптических весов, $\varphi_i(k) = (\varphi_{1i}(k), \varphi_{2i}(k), \dots, \varphi_{h_i i}(k))^T$ –
вектор вэйвлет-активационных функций, $e(k)$ – ошиб-
ка обучения, $\eta(k)$ – шаг обучения, подлежащий опре-
делению.

По аналогии с (11) можно ввести процедуру

$$\begin{cases} w_i(k+1) = w_i(k) + \frac{e(k)f_0'(u(k))\varphi_i(x_i(k))}{\gamma_i^{w_i}(k)}, \\ \gamma_i^{w_i}(k+1) = \alpha\gamma_i^{w_i}(k) + \|\varphi_i(x_i(k+1))\|^2, \end{cases} \quad (15)$$

где α – параметр забывания устаревшей информации
($0 < \alpha < 1$).

4 ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ

Эффективность работы предложенного двойного вэйвлет-нейрона и алгоритма его обучения (11), (15) исследовалась в процессе решения задачи прогнозирования и эмуляции хаотического поведения нелинейной динамической системы вида [19]

$$x_{n+1} = \frac{5x_n}{1+x_n^2} - 0,5x_n - 0,5x_{n-1} + 0,5x_{n-2} \quad (16)$$

с начальными значениями $x_0 = 0,2$, $x_1 = 0,3$, $x_2 = 1,0$.

Структура предиктора с элементами задержек на основе двойного вэйвлет-нейрона приведена на рис. 5.

Обучающая выборка содержала 10000 значений, а проверочная – 500 значений. Двойной вэйвлет-нейрон имел 5 синапсов в скрытом слое, соответствующих 5 входам $x(k-4)$, $x(k-3)$, $x(k-2)$, $x(k-1)$, $x(k)$, ($n = 5$) по 20 вэйвлетов в каждом синапсе ($h_i = 20$, $i = 1 \dots 5$). Выходной слой состоял из 5 вэйвлетов в синапсе WS_0 . Начальные значения синаптических весов генерировались случайным образом от $-0,1$ до $+0,1$.

В качестве оценки качества прогноза использовались несколько критериев:

– среднеквадратичная ошибка (RMSE)

$$RMSE = \frac{1}{N} \sum_{k=1}^N (x(k) - \hat{x}(k))^2;$$

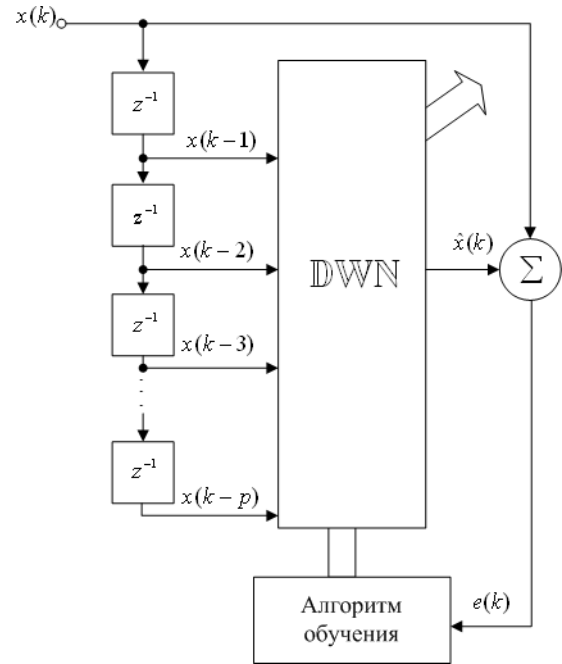


Рисунок 5 – Структура предиктора на основе двойного вэйвлет-нейрона

– Trefferquote [22, 23], представляющая собой процентное отношение правильно спрогнозированных направлений по отношению к фактическому направлению сигнала

Trefferquote =

$$= \frac{N - \frac{1}{2} \sum_{k=1}^N |\text{sign}(\hat{x}(k) - x(k-1)) - \text{sign}(x(k) - x(k-1))|}{N} \times 100 \%;$$

– Wegstrecke [22, 23], представляющий собой оценку качества прогнозирующей модели (при этом значение $+1$ соответствует оптимальной прогнозирующей модели, а значение -1 – неверному прогнозу) и описываемый выражением

$$Wegstrecke = \frac{\sum_{k=1}^N \text{signal}(k)(x(k) - x(k-1))}{\sum_{k=1}^N |x(k) - x(k-1)|},$$

где $\text{signal}(k)$ – сигнум-функция вида

$$\text{signal}(k) = \begin{cases} 1, & \text{если } \hat{x}(k) - x(k-1) > 0, \\ -1, & \text{если } \hat{x}(k) - x(k-1) < 0, \\ 0 & \text{в других случаях,} \end{cases}$$

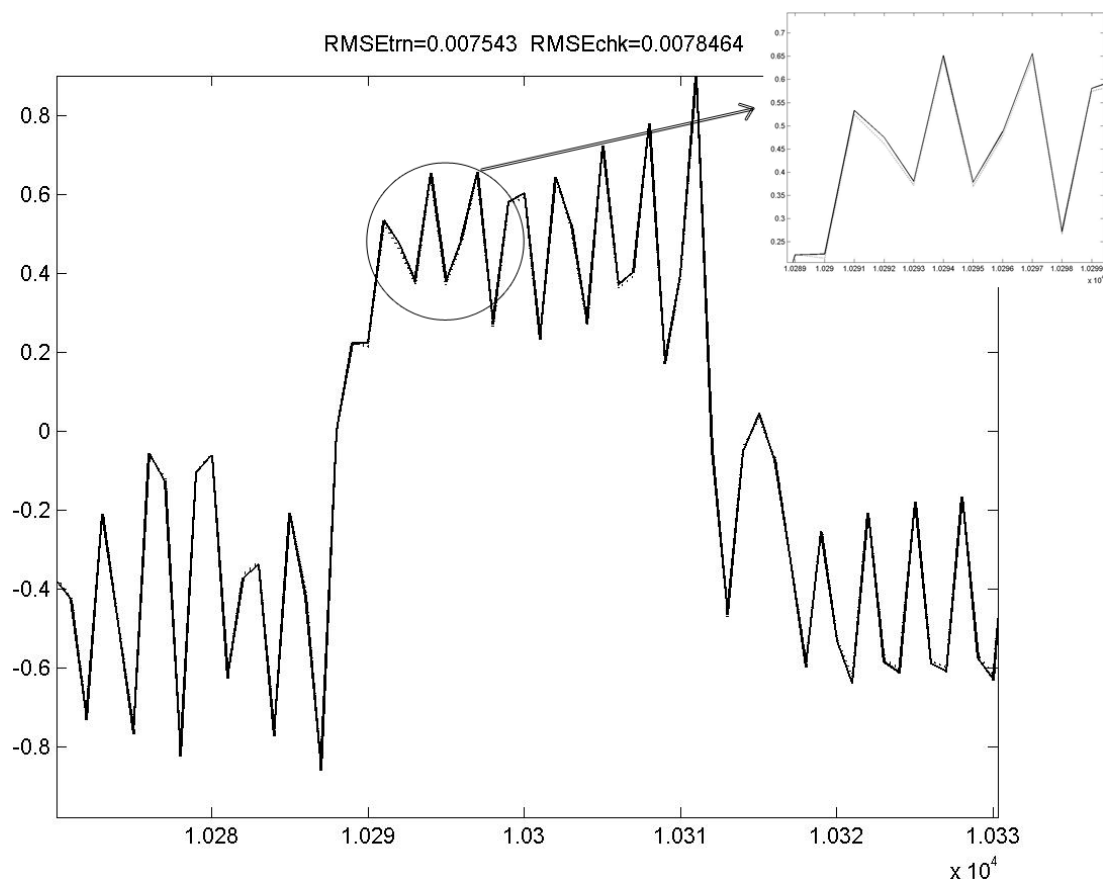


Рисунок 6 – Прогнозирование поведения хаотической динамической системы с помощью двойного вэйвлет-нейрона

$x(k)$ – фактическое значение прогнозируемого процесса, $\hat{x}(k)$ – прогноз, N – длина обучающей выборки.

На рис. 6 представлены результаты прогнозирования данных из тестового множества после 10 эпох обучения с параметром $\alpha = 0,99$.

Результаты прогнозирования на основе двойного вэйвлет-нейрона сравнивались с результатами прогнозирования на основе стандартного вэйвлет-нейрона с градиентным алгоритмом обучения, радиально-базис-

ной нейронной сетью и многослойным персептроном. Результаты приведены в таблице 1.

Таким образом, как видно из результатов эксперимента, при практически одинаковом количестве настраиваемых параметров предложенный двойной вэйвлет-нейрон с алгоритмом обучения (11), (15) обеспечивает более высокое качество прогноза и высокую скорость обучения по сравнению с традиционными архитектурами.

Таблица 1 – Результаты прогнозирования временного ряда

Нейронная сеть / Алгоритм обучения	Количество настраиваем. параметров	Критерии		
		RMSE	Wegstrecke	Trefferquote
Двойной вэйвлет-нейрон / Предложенный алгоритм обучения параметров вэйвлет синапсов (11) (15)	105	0,0078	1	99,8 %
Вэйвлет-нейрон / Градиентный алгоритм обучения параметров вэйвлет-синапсов с постоянным шагом	100	0,0101	0,98	98,8 %
Радиально базисная нейронная сеть / РМНК	100	0,5774	0,4883	55,2 %
Многослойный персептрон / Градиентный алгоритм обучения	115	0,6132	0,5882	75,5 %

ВЫВОДЫ

Предложена структура двойного вейвлет-нейрона и алгоритм его обучения, позволяющий настраивать все параметры сети. Алгоритм прост в численной реализации, обладает высокой скоростью сходимости и дополненными следящими и сглаживающими свойствами.

ПЕРЕЧЕНЬ ССЫЛОК

1. Chui C. K. An Introduction to Wavelets. – New York: Academic, 1992. – 264 p.
2. Daubechies I. Ten Lectures on Wavelets. Philadelphia, PA: SIAM, 1992. – 228 p.
3. Meyer Y. Wavelets: Algorithms and Applications. Philadelphia, PA: SIAM, 1993. – 133 p.
4. Lekutai G., van Landingham H. F. Self-tuning control of nonlinear systems using neural network adaptive frame wavelets // Proc. IEEE Int. Conf. on Systems, Man and Cybernetics. – Piscataway, N. J. – 1997. – 2. – P. 1017–1022.
5. Bodyanskiy Ye., Lamonova N., Pliss I., Vynokurova O. An adaptive learning algorithm for a wavelet neural network // Expert Systems. – 2005. – 22. – № 5 – P. 235–240.
6. Bodyanskiy Ye., Kolodyazhnyi V., Pliss I., Vynokurova O. Learning wavelet neuron based on the RASP-function // Radio Electronics. Computer Science. Control. – 2004. – № 1. – P. 118–122.
7. Бодянский Е. В., Винокурова Е. А., Ламонова Н. С. Адаптивная гибридная вейвлет-нейронная сеть для решения задачи прогнозирования и эмуляции // Сб. науч. трудов 12-й международной конференции по автоматическому управлению «Автоматика 2005», Т. 3. – Харьков: Изд-во НТУ «ХПИ». – 2005. – С. 40–41.
8. Бодянский Е. В., Винокурова Е. А. Треугольный вейвлет и формальный нейрон на его основе // Сб. науч. труды 3-й Международной научно-практической конференции «Математичне та програмне забезпечення інтелектуальних систем» (MPZIS-2005) – Дніпропетровськ: ДНУ. – 2005. – С. 14–15.
9. Billings S. A., Wei H.-L. A new class of wavelet networks for nonlinear system identification // IEEE Trans. on Neural networks. – 2005. – 16. – № 4. – P. 862–874.
10. Szu H. H., Telfer B., Kadambe S. Neural network adaptive wavelets for signal representation and classification // Opt. Eng. – 1992. – 31. – P. 1907–1916.
11. Zhang Q. H., Benveniste A. Wavelet networks // IEEE Trans. on Neural Networks. – 1992. – V. 3. – № 6. – P. 889–898.
12. Dickhaus H., Heinrich H. Classifying biosignals with wavelet networks // IEEE Eng. Med. Biol. Mag. – 1996. – 15. – № 5. – P. 103–111.

13. Cao L. Y., Hong Y. G., Fang H. P., He G. W. Predicting chaotic time series with wavelet networks // Phys. D. – 1995. – 85. – P. 225–238.
14. Oussar Y., Dreyfus G. Initialization by selection for wavelet network training // Neurocomputing. – 2000. – 34. – P. 131–143.
15. Zhang J., Walter G. G., Miao Y., Lee W. N. W. Wavelet neural networks for function learning // IEEE Trans. on Signal Process. – 1995. – 43. – № 6. – P. 1485–1497.
16. Zhang Q. H. Using wavelet network in nonparametric estimation // IEEE Trans. on Neural Networks. – 1997. – 8. – № 2. – P. 227–236.
17. Casdagli M. Nonlinear prediction of chaotic time series // Phys. D. – 1989. – 35. – P. 335–356.
18. Soltani S. On the use of wavelet decomposition for time series prediction // Neurocomputing. – 2002. – 48. – P. 267–277.
19. Yamakawa T., Uchino E., Samatu T. Wavelet neural networks employing over-complete number of compactly supported non-orthogonal wavelets and their applications // IEEE Int. Conf. on Neural Networks, Orlando, USA. – 1994. – P. 1391–1396.
20. Yamakawa T., Uchino E., Samatu T. The wavelet network using convex wavelets and its application to modeling dynamical systems // The Trans. on the IEICE. – 1996. – J79-A. – № 12. – P. 2046–2053.
21. Yamakawa T. A novel nonlinear synapse neuron model guaranteeing a global minimum – Wavelet neuron // Proc. 28 th IEEE Int. Symp. On Multiple-Valued Logic. – Fukuoka, Japan: IEEE Comp. Soc., 1998. – P. 335–336.
22. Baumann M. Nutzung neuronaler Netze zur Prognose von Aktienkursen. – Report Nr. 2/96, TU Ilmenau, 1996. – 113 s.
23. Fueser K. Neuronale Netze in der Finanzwirtschaft. – Wiesbaden: Gabler, 1995. – 437 s.

Надійшла 22.02.06

В статті запропоновано архітектуру подвійного вейвлет-нейрону, що є модифікацією стандартного вейвлет-нейрону та алгоритм навчання його параметрів. Ця архітектура дозволяє поліпшити апроксимуючі властивості вейвлет-нейрону. Проведено імітаційне моделювання запропонованого подвійного вейвлет-нейрону та алгоритму його навчання з використанням нестационарної хаотичної часової послідовності.

In this paper a new double wavelet neuron architecture obtained by modification of standard wavelet neuron, and its learning algorithm for its parameters is proposed. Offered architecture allows to improve the approximation properties of wavelet-neuron. Double wavelet neuron and its learning algorithm are examined for predicting non-stationary chaotic time series.

УДК 004.932.001.57

А. О. Дранкова, Е. В. Ткаченко

ИССЛЕДОВАНИЕ СЕТИ ХОПФИЛДА ДЛЯ РАСПОЗНАВАНИЯ ЦИФРО-ПОДОБНЫХ ОБРАЗОВ

Исследованы функциональные возможности дискретной сети Хопфилда для распознавания цифро-подобных образов при наличии шумов различной природы и интенсивности. Результаты компьютерного моделирования позволяют рекомендовать данные сети для распознавания

стационарных образов при достаточно большом процентном содержании помех.

В настоящее время все более распространенным подходом к распознаванию цифро-подобных образов

при наличии помех различной интенсивности является подход, основанный на применении модели Хопфилда. Модель Хопфилда является наиболее успешной реализацией автоассоциативной памяти. Реализация автоассоциативной памяти включает в себя восстановление сохраненных образов при предъявлении памяти неполных или зашумленных их версий [1].

Данная статья описывает компьютерное моделирование и тестирование дискретной сети Хопфилда, выступающей в качестве ассоциативной памяти ограниченной емкости. В эксперименте использовалась сеть, состоящая из $N = 900$ нейронов и, следовательно, имеющая $N^2 - N = 809100$ синоптических связей. Эта сеть обучалась для извлечения 10 бинарных цифро-подобных образов, приведенных на рисунке 1. Каждый из этих рисунков содержит 30×30 пикселей (элементов). Для входов применяемых в сети предполагалось, что белому цвету, образа соответствует значение -0 , а черному -1 . Поэтому в качестве функций активации использовалась линейная функция, ограниченная 0 и 1. Каждый из этих образов использовался как ячейка фундаментальной памяти на фазе запоминания (обучения) сети Хопфилда для создания матрицы синаптических весов W .

В работе ассоциативной памяти можно выделить 2 фазы:

- фаза запоминания, которая соответствует процессу обучения сети на исходных образах;
- фаза вспоминания, соответствующая извлечению запомненного образа в ответ на представление в сеть зашумленной или искаженной версии образа [2].

Фаза вспоминания осуществлялась в асинхронном режиме.

Во время первой части фазы вспоминания на вход сети подавались ячейки фундаментальной памяти. Таким образом, выполнялась проверка способности их корректного восстановления из информации, сохраненной в матрице синаптических весов. В каждом из случаев желаемый образ восстанавливался сетью после одной итерации. Данная фаза практически соответствует случаю, когда отношение сигнал/шум и емкость сети M не превышает критическое значение, определяемое по формуле: $M = 0,14N$ [3].

Во второй части фазы вспоминания рассматриваемый образ преднамеренно искажался шумами раз-



Рисунок 1 – Исходные цифро-подобные образы



а)



б)



в)

Рисунок 2 – Искаженные цифро-подобные образы при отношении сигнал / шум = 2:

а – нормальный гауссовский шум; б – импульсная помеха; в – мультипликативный шум



а)



б)



в)

Рисунок 3 – Результаты симуляции искаженных цифро-подобных образов при отношении сигнал / шум = 2:

а – нормальный гауссовский шум; б – импульсная помеха; в – мультипликативный шум

личного типа и интенсивности (рисунок 2), чтобы продемонстрировать способность сети Хопфилда корректировать ошибки.

Результат симуляции сети Хопфилда на искаженных образах показан на рисунке 3.

Исследования показали, что при отношении сигнал / шум = 10 : 1 вероятность ошибки не превышает 5–10 % для нормального гауссовского шума, 1 % для мультипликативного шума и 40 % для импульсной помехи. Результаты моделирования приведены в таблице 1.

Таблица 1 – Вероятность ошибки распознавания в процентах

Отношение сигнал/шум	10	9	8	7	6	5	4	3	2	1
Вид шума										
Нормальный гауссовский шум	0	0	0	0	0	0	0,03	0,25	0,86	6,44
Импульсная помеха	0	0	0	0	0,05	4,76	11,35	18,48	27,12	40,06
Мультипликативный шум	0	0	0	0	0	0	0,005	0,011	0,023	0,037

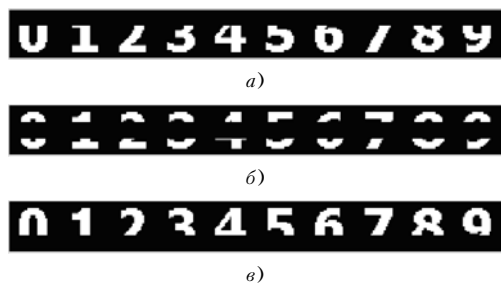


Рисунок 4 – Искаженные цифро-подобные образы:

а – при удалении верхней трети образа; б – при удалении средней трети образа; в – при удалении нижней трети образа

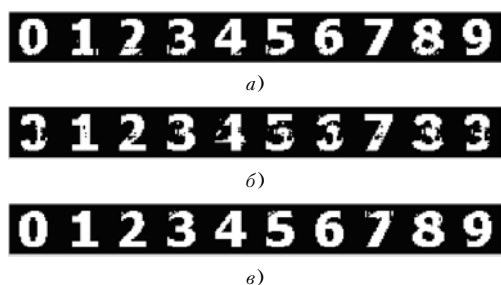


Рисунок 5 – Результаты симуляции искаженных цифро-подобных образов:

а – при удалении верхней трети образа; б – при удалении средней трети образа; в – при удалении нижней трети образа

Третья часть фазы вспоминания исследует способность сети Хопфилда к восстановлению образа по его части: удаление верхней трети образа, удаление средней трети образа, удаление нижней трети образа (рисунок 4).

УДК 681.3.07

А. И. Липчанский, Механна Сами, Хабис А. А. Зидат

ИСПОЛЬЗОВАНИЕ ПРОДУКЦИОННЫХ СИСТЕМ И БИНАРНЫХ ОТНОШЕНИЙ В ЭКСПЕРТНЫХ СИСТЕМАХ ДИАГНОСТИКИ КОМПЬЮТЕРНОЙ ТЕХНИКИ

Рассматривается возможность использования продукций для создания экспертных систем, способных поставить диагноз возможному состоянию компьютерной техники. Показано, что эта цель может быть достигнута с помощью использования бинарных отношений, заданных на множестве состояний объекта диагностирования. Эффективность предлагаемого подхода иллюстрируется на конкретных примерах.

© Липчанский А. И., Механна Сами, Хабис А. А. Зидат, 2006

Удаление верхней либо нижней трети образа позволяет обеспечить вероятность ошибки не более 2 %. Как показали, исследования средняя часть цифро-подобного образа является наиболее информативной, и ее удаление повышает вероятность ошибки до 7 %. Результат симуляции сети Хопфилда при удалении различных частей образа представлен на рисунке 5.

Несмотря на высокий процент вероятности правильного распознавания цифро-подобных образов, все еще открытым остается вопрос емкости сети Хопфилда, которая зависит от размерности исходных образов и их коррелированности, а также от отношения сигнал/шум.

ПЕРЕЧЕНЬ ССЫЛОК

1. Осовский С. Нейронные сети для обработки информации / Пер. с польского. – М.: Финансы и статистика, 2004. – 344 с.
2. Саймон Хайкин. Нейронные сети: полный курс, 2-е издание.: Пер. с англ. – М.: издательский дом «Вильямс», 2006. – 1104 с.
3. Медведев В. С., Потемкин В. Г. Нейронные сети. МАТЛАБ 6/ Под общ. ред. к. т. н. В. Г. Потемкина. – М.: ДИАЛОГ-МИФИ, 2002. – 496 с.

Надійшла 11.03.06
Після доробки 10.07.06

Досліджені функціональні можливості дискретної мережі Хопфілда для розпізнавання цифро-подібних образів при наявності шумів різної природи та інтенсивності. Результати комп'ютерного моделювання дозволяють рекомендувати данні мережі для розпізнавання стаціонарних образів при достатньо великій процентній наявності завод.

Functional potential of discrete Hopfield network for recognition number-similar characters in the presence of different nature and energy noise is analyzed. The computer modeling results allow to recommend this kind of network for fixed characters recognition big enough proportion of noise.

1 АКТУАЛЬНОСТЬ ИССЛЕДОВАНИЯ

Технология экспертных систем (ЭС) является одним из направлений искусственного интеллекта (ИИ). Исследования в области ИИ направлены на разработку программ, решающих задачи, с которыми сейчас лучше справляется человек. Поэтому «разработку

программы, которая будет выполнять сложную статистическую обработку данных, нельзя рассматривать как исследование в области ИИ, какие бы сложные алгоритмы в ней не использовались» [1].

ЭС содержит знания в определенной предметной области (ПО), накопленные в результате практической деятельности человека, и использует их для решения проблем, специфичных для этой области. Этим ЭС отличаются от прочих, «традиционных» систем, в которых предпочтение отдается более общим и менее связанным с ПО теоретическим методам, чаще всего математическим.

Диагностические ЭС (ДЭС) предназначены для обнаружения источников неисправностей по результатам наблюдений за поведением контролируемой системы. ДЭС выполняют диагностирование, используя описание ситуаций, характеристики поведения или знания о конструкции компонент, чтобы установить вероятные причины неправильного функционирования диагностируемой системы [2].

ДЭС часто являются консультантами, которые не только ставят диагноз, но также помогают в отладке. Они могут взаимодействовать с пользователем, чтобы оказать помощь при поиске неисправностей, а затем предложить порядок действий по их устранению. В медицинской ПО разработано больше ДЭС, чем в любой другой отдельно взятой ПО. Однако в настоящее время многие ДЭС разрабатываются для приложений к инженерному делу и компьютерным системам [1–3].

Например, анализаторы глобальной компьютерной сети (ГКС) проверяют и анализируют соглашения об уровне сервиса, статистику и проблемы за длительный период, а также случаи перегрузки каналов. Для оценки и поиска ошибок анализатор ГКС, состоящий, как правило, из аппаратной и программной частей, должен среди прочего иметь функцию ДЭС реального времени. База знаний для диагностики в реальном времени с предложением возможных решений упрощает поиск ошибок и тем самым повышает готовность сети. Отображение экспертных сообщений имеет смысл для представления критичных аномалий глобальной сети [4–6].

Центральным вопросом построения ЭС является выбор формы представления знаний – способа формального выражения знаний о ПО. Форма представления знаний оказывает существенное влияние на характеристики и свойства ЭС, поэтому представление знаний является одной из наиболее важных проблем, характерных для ЭС.

Одним из старейших методов представления знаний о ПО в ЭС является подход, основанный на правилах (продукциях). Как один из наиболее естественных он широко используется в коммерческих и экспериментальных ЭС [7–12].

При диагностике компьютерной техники (КТ) экспертами часто используется продукционная система

представления знаний, которые можно сформулировать в виде «если ..., то ...». Например:

– «если объект имеет неисправные элементы и для наблюдаемого выхода выполняется условие ..., тогда множество ... содержит хотя бы один неисправный элемент» [13];

– «если после повторного выполнения сеанса анализа протоколов ошибка на другое устройство не перешла ..., проверьте соответствующую абонентскую область ...» [14].

Важным недостатком продукционных ДЭС (ПДЭС) является то, что существующие методы их построения не гарантируют классификации каждого исследуемого ПДЭС объекта. В ПДЭС, как правило, не рассматривается задача полной классификации объекта исследования. Их главной задачей является фиксация знаний эксперта в определенной ПО. Эти знания позволяют классифицировать некоторые состояния объекта, оставляя другие возможные состояния за пределами разрабатываемой ПДЭС.

Очевидно, что сама постановка задачи построения полной классификации возможна лишь тогда, когда осуществлена предварительная структуризация ПО (определение множества признаков, описывающих различные состояния объекта исследования, и бинарных отношений между ними) [15–18].

Целью данной работы является решение задачи полной классификации состояний объекта диагностирования, используя знания эксперта о ПО, представленные в виде продукционной системы и бинарных отношений, заданных на множестве состояний объекта диагностирования.

2 ПОСТАНОВКА ПРОБЛЕМЫ

Дано: знания о ПО «Диагностика КТ», находящиеся в памяти человека-эксперта или на материальных носителях знаний (в учебниках, монографиях, статьях, методиках и на других носителях профессиональных знаний).

Требуется: разработать ДЭС, которая для любого состояния объекта диагностирования способна указать, какой неисправностью обладает объект в данном состоянии, если таковая имеется.

Разработка программных комплексов ЭС до сих пор находится на уровне скорее искусства, чем науки. Процесс разработки промышленной ЭС практически для любой ПО можно разделить на шесть этапов [8]:

- выбор проблемы;
- разработка прототипа ЭС;
- доработка до промышленной ЭС;
- оценка ЭС;
- стыковка ЭС;
- поддержка ЭС.

Выбор подходящей проблемы начинается с определения ПО и задачи: ПО – «Компьютерная техника», задача – диагностика неисправностей КТ. Потом необходимо найти эксперта, желающего сотрудничать при решении проблемы, и назначить коллектив разработчиков ЭС: эксперт – специалист по ремонту персональных компьютеров (ПК) или опытный администратор компьютерной сети (КС); в коллектив разработчиков кроме эксперта в ПО КТ должен входить инженер по знаниям, извлекающий знания из эксперта, программист, реализующий ДЭС, и пользователь, для которого создается ДЭС. После этого надо определить предварительный подход к решению проблемы: представление знаний в виде продукций и бинарных отношений. Наконец, необходимо подготовить подробный план разработки ЭС, в котором, в частности, проанализировать расходы и прибыли: этот этап выходит за рамки данной работы.

Технология быстрого прототипирования ЭС включает в себя следующие шесть стадий [8]:

- идентификация проблемы;
- получение знаний;
- структурирование;
- формализация;
- реализация прототипа;
- тестирование.

На стадии идентификации определяются необходимые ресурсы (время, люди, оборудование и т. п.), источники знаний, имеющиеся аналогичные ЭС, цели и классы решаемых ЭС задач. Извлечение знаний – это получение наиболее полного из возможных представлений о ПО и способах принятия решений в ней. Затем выявляется структура полученных знаний о ПО: терминология, список основных понятий и их атрибутов, отношения между понятиями, структура входной и выходной информации, стратегия принятия решений. После этого строится формализованное представление концепций ПО на основе выбранного языка представления знаний. Наконец, создается прототип ЭС, включающий базу знаний, оценивается и проверяется его работа с целью приведения в соответствие с реальными запросами пользователей.

В данной работе основное внимание уделяется этапу формализации.

Переход от прототипа к промышленной ЭС включает следующие этапы [8]:

- демонстрационный прототип ЭС;
- исследовательский прототип ЭС;
- действующий прототип ЭС;
- промышленная ЭС;
- коммерческая ЭС.

Здесь основная работа заключается в существенном расширении базы знаний, разработки и адаптации интерфейсов. На первом этапе ЭС должна решать малую часть задач, демонстрируя жизнеспособность подхода.

На втором этапе ЭС должна уметь решать большинство задач, но может быть неустойчива в работе и не полностью проверена. На третьем этапе ЭС должна надежно решать все задачи на реальных примерах, хотя решения сложной задачи будет требовать много времени и памяти. На четвертом этапе ЭС должна обеспечивать высокое качество решений при минимизации требуемого времени и памяти, что может потребовать ее переписывания с использованием более эффективных средств представления знаний. Наконец, на последнем пятом этапе ЭС должна быть пригодна к продаже (быть хорошо документированной и снабженной сервисом).

ЭС оцениваются главным образом для того, чтобы проверить точность работы программы и ее полезность. К тестированию широко привлекаются другие эксперты с целью апробирования работоспособности системы на различных примерах.

Под стыковкой ЭС подразумевается разработка связей с другими программными средствами в среде, в которой ЭС будет работать, и обучение людей, которых она будет обслуживать.

Если ЭС сохраняет все знание о ПО и это знание не будет изменяться в ближайшем будущем, то в поддержке ЭС нет необходимости. Но если ЭС создана именно из-за того, что ПО изменяется, то необходимо поддерживать ЭС в ее инструментальной среде разработки.

3 ИСПОЛЬЗОВАНИЕ ПРОДУКЦИОННЫХ СИСТЕМ

В системах продукций знания представляются с помощью наборов правил вида: «если А, то В. Здесь А и В могут пониматься как «ситуация – действие», «причина – следствие», «условие – заключение» и т. п. Множество продукционных правил образует базу правил, каждое из которых представляет обособленный фрагмент знаний о ПО.

Например, при поиске неисправностей в сети Token Ring используются блок-схемы, из которых можно извлечь следующее правило [14]:

P_1 – «Если кольцевая станция зависает, то необходимо проверить выполняется ли на данной кольцевой станции одно и то же приложение, использует ли она на каком-либо конкретном файл-сервере один и тот же каталог или файл в момент проявления этого симптома?».

Используя только правило P_1 диагноз поставить невозможно. Поэтому необходимо извлекать и использовать некоторые другие правила, например:

P_2 [14] – «Если на данной кольцевой станции выполняется одно и то же приложение или она использует на каком-либо конкретном файл-сервере один и тот же каталог или файл в момент проявления симптома зависания, то необходимо проверить, нет ли каких либо

проблем в конкретному приложенні файл-сервера или в его соответствующем каталоге/файле?»;

P_3 [14] – «Если сетевой файл-сервер испытывает проблемы с какой-либо кольцевой станцией при доступе к тому или иному конкретному приложению/группе приложений или определенным каталогам/файлам либо при их использовании, то проверьте сетевой файл-сервер и кольцевую станцию на следующие параметры и требования программно-аппаратной конфигурации: (приводиться соответствующий список)»;

D_1 [14] – «Если имеются какие-то поддающиеся идентификации проблемы с установкой аппаратной или программной конфигурации для сетевого файл-сервера, то примите необходимые меры для решения обнаруженной проблемы с конфигурацией».

При помощи цепочки продукций:

$$P_1 \rightarrow P_2 \rightarrow P_3 \rightarrow D_1 \quad (1)$$

можно поставить диагноз и решить проблему, если конечно она связана с конфигурацией файл-сервера.

Использовать знания, представленные продукциями такого вида, непосредственно в ПДЭС достаточно трудно. Гораздо удобнее представлять их в виде триплета: «объект – атрибут – значение» [7]. При этом продукции примут более структурированный вид:

P_1 – «Если (объект – кольцевая станция, атрибут – состояние, значение – зависит), то (объект – файл-сервер, атрибут – приложение, значение – используется)»;

P_2 – «Если (объект – файл-сервер, атрибут – приложение, значение – Application-1), то (объект – приложение, атрибут – доступ, значение – нет)»;

P_3 – «Если (объект – приложение, атрибут – доступ, значение – нет), то (объект – файл-сервер, атрибут – конфигурация, значение – определенные параметры)»;

D_1 – «Если (объект – файл-сервер, атрибут – конфигурация, значение – определенные параметры), то (объект – файл-сервер, неисправность – проблемы с конфигурацией)».

Рассмотрим теперь формальное описание процесса постановки этого диагноза. Задана система S – сеть Token Ring. В состав системы S входят подсистемы S_1 – кольцевая станция, S_2 – файл-сервер, S_3 – приложение. У подсистемы S_1 есть атрибут a_1^1 – состояние, у подсистемы S_2 – атрибуты a_1^2 – приложение и a_2^2 – конфигурация, у S_3 – a_1^3 – доступ. Атрибут a_1^1 может иметь значение $v_{1,1}^1$ – зависит, атрибут a_1^2 – значение $v_{1,1}^2$ – Application-1, атрибут a_2^2 – $v_{2,1}^2$ – определенные параметры, a_1^3 – $v_{1,1}^3$ – нет. Наконец, у подсистемы S_2

может быть неисправность f_1^2 – проблемы с конфигурацией.

При таком описании продукции примут следующий формальный вид:

$$P_1: (S_1, a_1^1, v_{1,1}^1) \rightarrow (S_2, a_1^2, v_{1,1}^2),$$

$$P_2: (S_2, a_1^2, v_{1,1}^2) \rightarrow (S_3, a_1^3, v_{1,1}^3),$$

$$P_3: (S_3, a_1^3, v_{1,1}^3) \rightarrow (S_2, a_2^2, v_{2,1}^2),$$

$$D_1: (S_2, a_2^2, v_{2,1}^2) \rightarrow (S_2, f_1^2).$$

Процесс постановки диагноза выглядит теперь следующим образом:

$$(S_1, a_1^1, v_{1,1}^1) \rightarrow (S_2, a_1^2, v_{1,1}^2) \rightarrow (S_3, a_1^3, v_{1,1}^3) \rightarrow (S_2, a_2^2, v_{2,1}^2) \rightarrow (S_2, f_1^2).$$

Наконец, рассмотрим формальное описание процесса постановки диагноза в общем виде.

Задана диагностируемая система S . Система S состоит из множества подсистем:

$$\text{Sub} = (S_1, S_2, \dots, S_i, \dots, S_l).$$

У каждой подсистемы S_i есть множество атрибутов:

$$A_i = \{a_1^i, a_2^i, \dots, a_j^i, \dots, a_{j(i)}^i\}.$$

У каждого атрибута a_j^i есть множество значений:

$$V_j^i = \{v_{j,1}^i, v_{j,2}^i, \dots, v_{j,k}^i, \dots, v_{j,K(i,j)}^i\}.$$

У каждой подсистемы S_i есть множество неисправностей $F_i = \{f_1^i, f_2^i, \dots, f_m^i, \dots, f_{M(i)}^i\}$.

Диагноз системе S ставится с помощью множества продукций

$$P = \bigcup_{\substack{(i_1, j_1, k_1) \in I_1 J_1 K_1 \\ (i_2, j_2, k_2) \in I_2 J_2 K_2}} \{(S_{i_1}, a_{j_1}^{i_1}, v_{j_1, k_1}^{i_1}) \rightarrow (S_{i_2}, a_{j_2}^{i_2}, v_{j_2, k_2}^{i_2})\},$$

отражающих промежуточные рассуждения при постановке диагноза и множества продукций

$$D = \bigcup_{\substack{(i_3, j_3, k_3) \in I_3 J_3 K_3 \\ m \in M}} \{(S_{i_3}, a_{j_3}^{i_3}, v_{j_3, k_3}^{i_3}) \rightarrow (S_j, f_m^j)\},$$

отражающих постановку окончательного диагноза.

Вернемся к примеру диагностирования сети Token Ring. В процессе диагностирования установлено, что эта сеть в данный момент находится в следующем состоянии:

- кольцевая станция находится в состоянии зависания ($a_1^1 = v_{1,1}^1$);
- при этом на файл-сервере используется приложение Application-1 ($a_1^2 = v_{1,1}^2$);
- к этому приложению нет доступа ($a_1^3 = v_{1,1}^3$);
- в конфигурации файл-сервера есть определенные параметры $a_2^2 = v_{2,1}^2$.

Сети в данном состоянии поставлен диагноз: проблемы с конфигурацией файл-сервера f_1^2 .

Формально состояние системы S описывается вектором $(v_{1,1}^1, v_{1,1}^2, v_{1,1}^3, v_{2,1}^2)$. И это состояние было отнесено к классу неисправности f_1^2 .

В общем виде можно говорить о текущем состоянии $c_i = (c_1^i, c_2^i, \dots, c_j^i, \dots, c_{J(i)}^i)$ подсистемы S_i , где $c_j^i \in V_j^i$. Тогда текущее состояние системы S характеризуется вектором $c = (c_1, c_2, \dots, c_i, \dots, c_I)$. Это состояние c в результате диагностирования может быть отнесено к некоторому классу неисправностей из множества $F = \bigcup_{i=1}^I F_i$. Состояние c может быть отнесено к нескольким классам из множества F , если в системе S присутствует несколько неисправностей. Если система S в состоянии c исправна, то это состояние не может быть отнесено к какому-либо классу из множества F . Формально можно ввести класс f_0 – система S находится в исправном состоянии. Тогда результатом диагностирования будет отнесение текущего состояния c системы S к некоторым классам из множества $F_0 = F \cup \{f_0\}$. Естественно, если состояние c принадлежит к классу f_0 , то к другому классу оно принадлежать не может.

Количество N всех гипотетически возможных состояний системы S определяется формулой

$$N = \prod_{i=1}^I \left(\prod_{j=1}^{J(i)} K(i, j) \right).$$

Пусть в состав системы S входит 10 подсистем, и каждая подсистема имеет по 10 атрибутов. Если каждый атрибут имеет по 10 значений, то количество всех гипотетически возможных состояний системы S равно:

$$N = \prod_{i=1}^{10} \left(\prod_{j=1}^{10} 10 \right) = \prod_{i=1}^{10} \left(\frac{10 \times \dots \times 10}{10} \right) = \prod_{i=1}^{10} (10^{10}) = 10^{100}.$$

Вряд ли эксперт за всю свою жизнь сталкивался со всеми $N = 10^{100}$ гипотетически возможными состояни-

ями системы S . Даже если предположить невозможное, что эксперт диагностировал 1 состояние за 1 секунду, то за 1 минуту он диагностировал 60 состояний, за 1 час – 3 600 состояний, за 1 сутки – 86 400, за 1 год – 31 536 000. Пусть эксперт способен увеличить продуктивность своей работы до 10 000 000 000 состояний в год. Тогда 10^{100} состояний он будет диагностировать 10^{10} лет! Абсолютно невозможно...

Таким образом, с помощью множества продукций P и D , отражающих знания и опыт практической деятельности эксперта в ПО «Диагностика КТ», можно разработать только демонстрационный прототип ПДЭС, способный решать лишь очень малую часть задач. Уровень исследовательского прототипа ПДЭС может достичь, только если она будет уметь решать большинство задач. Очевидно, для построения ПДЭС такого уровня для представления знаний эксперта необходимо использовать продукционные системы в сочетании с другими методами.

4 ИСПОЛЬЗОВАНИЕ БИНАРНЫХ ОТНОШЕНИЙ

В своей практической деятельности эксперт пользуется не только знаниями и опытом, полученным в результате решения реальных задач, но и тем, что принято называть «интуицией». Используя «интуицию» эксперт способен успешно решать даже такие сложные задачи, с которыми он раньше на практике не встречался. Знания, которые отражают «интуицию» эксперта, являются, возможно, самыми ценными. И если даже поверхностные знания эксперта, которыми он пользуется при решении стандартных задач, выявить достаточно проблематично, то выявление «глубинных» знаний эксперта, которые объясняют явления и могут использоваться для прогнозирования поведения объектов, является весьма сложной проблемой. Современные ЭС работают в основном с поверхностными знаниями [8]. Это связано с тем, что на данный момент нет универсальных методик, позволяющих выделять глубинные структуры знаний и работать с ними.

В работах [15–18] для представления глубинных «интуитивных» знаний эксперта предлагается использовать аппарат теории отношений. При этом предполагается, что:

- разные значения атрибутов имеют разные степени характеристики для диагностируемых неисправностей;
- для каждого атрибута эксперт может упорядочить его значения по их характеристике для соответствующей неисправности и этот порядок не зависит от значений других атрибутов.

Вернемся к примеру диагностирования сети Token Ring. В процессе диагностирования установлено, что проблемы с конфигурацией файл-сервера влияли на

доступ к приложению Application-1. Формально это означает, что значение $v_{1,1}^2$ атрибута a_1^2 характерно для неисправности f_1^2 .

Пусть на файл-сервере кроме приложения Application-1 используется также приложение Application-2. Предположим, эксперт знает, что отсутствие доступа к приложению Application-2 более характерно для неисправности «проблемы с конфигурацией файл-сервера», чем отсутствие доступа к приложению Application-1. Формально это означает, что у атрибута a_1^2 кроме значения $v_{1,1}^2$ (Application-1) есть еще значение $v_{1,2}^2$ – Application-2. При этом значение $v_{1,2}^2$ более характерно для неисправности f_1^2 , чем значение $v_{1,1}^2$. Этот факт можно записать так:

$$R(a_1^2, f_1^2): v_{1,2}^2 \rightarrow v_{1,1}^2. \quad (2)$$

Покажем, как можно эффективно использовать это бинарное отношение характерности. Ранее состояние $v = (v_{1,1}^1, v_{1,1}^2, v_{1,1}^3, v_{2,1}^2)$ системы S было отнесено к классу неисправности f_1^2 . Рассмотрим состояние $w = (v_{1,1}^1, v_{1,2}^2, v_{1,1}^3, v_{2,1}^2)$, которое отличается от состояния v только значением второго компонента. Можно сказать, что состояние w более характерно для неисправности f_1^2 , чем состояние v . Действительно, значение второго компонента состояния w более характерно для неисправности f_1^2 , чем значение второго компонента состояния v . При этом значения всех остальных компонентов состояний w и v соответственно равны. Формально это означает, что отношение $R(a_1^2, f_1^2)$, заданное на множестве значений атрибута a_1^2 , индуцирует следующее отношение $R(f_1^2)$, заданное на множестве всех гипотетически возможных состояний системы S :

$$R(f_1^2): w \rightarrow v. \quad (3)$$

Из этого следует, что если состояние v было отнесено с помощью продукций к классу f_1^2 , то и состояние w , как более характерное для неисправности f_1^2 , тоже должно быть отнесено к классу f_1^2 автоматически, без использования каких-либо продукций.

Таким образом, используя цепочку продукций (1), отражающих «поверхностные» знания эксперта, можно поставить диагноз только состоянию v . Привлекая же «глубинные» знания эксперта, которые выражаются в отношении (2), индуцирующем отношении (3),

можно с помощью цепочки (1) ставить диагноз и состоянию w .

В общем виде предполагается, что для каждого атрибута a_j^i эксперт может упорядочить множество V_j^i его значений по характерности для каждой неисправности f_m^q в виде отношения $R(a_j^i, f_m^q)$. Эти отношения индуцируют отношение $R(f_m^q)$, упорядочивающее по характерности для неисправности f_m^q все гипотетически возможные состояния системы S .

Отношения $R(f_m^q)$ вместе с множеством продукций P и D можно использовать для построения ЭС, способной диагностировать все гипотетически возможные состояния системы S . Таким образом, появляется принципиальная возможность создания ЭС коммерческого уровня, обеспечивающего высокое качество постановки диагноза.

5 ПРОЦЕСС ФОРМАЛИЗАЦИИ

Рассмотрим простой пример процесса построения формализованного описания концепций ПО «Диагностика КТ» на основе представления знаний эксперта в виде продукционных систем и бинарных отношений.

При поиске неисправностей в сети Token Ring используются блок-схемы, из которых можно извлечь следующие знания, общие для всех сетевых операционных систем (СОС) [14].

Задана диагностируемая система S – сеть Token Ring. Система S состоит из множества подсистем:

$$\text{Sub} = \{S_1, S_2\},$$

где S_1 – сетевой файл-сервер, S_2 – проблемные компьютеры.

У подсистемы S_1 есть множество атрибутов

$$A_1 = \{a_1^1\},$$

где a_1^1 – источник проблемы.

У подсистемы S_2 есть множество атрибутов

$$A_2 = \{a_1^2\},$$

где a_1^2 – параметры и требования программно-аппаратной конфигурации.

У атрибута a_1^1 есть множество значений

$$V_1^1 = \{v_{1,1}^1, v_{1,2}^1, v_{1,3}^1\},$$

где $v_{1,1}^1$ – кольцевая станция, $v_{1,2}^1$ – группа кольцевых станций, $v_{1,3}^1$ – сеть в целом.

У атрибута a_1^2 есть множество значений

$$V_1^2 = \{v_{1,1}^2, v_{1,2}^2, v_{1,3}^2, v_{1,4}^2, v_{1,5}^2\},$$

где $v_{1,1}^2$ – убедиться в том, что программы сетевой оболочки СОС, драйверы сетевых плат, config.sys, autoexec.bat и все необходимые файлы СОС отвечают специализированной конфигурации соответствующей СОС и иерархической структуре каталогов СОС; $v_{1,2}^2$ – проверить, насколько структура каталогов СОС отвечает доступу кольцевой станции и всех соответствующих сетевых периферийных устройств; $v_{1,3}^2$ – проверить параметры системы обеспечения безопасности СОС; $v_{1,4}^2$ – убедиться в том, что на сетевом диске и дисках кольцевых станций установлены все необходимые каталоги (файлы) для всех приложений СОС; $v_{1,5}^2$ – убедиться в том, что в файл-сервере (кольцевых станциях) установлена, правильно сконфигурирована и нормально функционирует вся необходимая аппаратура.

У подсистемы S_2 есть множество неисправностей

$$F_2 = \{f_1^2\},$$

где f_1^2 – проблемы с конфигурацией.

Диагноз системе S ставится с помощью множества продукций

$$P = \bigcup_{i=1}^3 \bigcup_{j=1}^5 \{P_{i,j}: (S_1, a_1^1, v_{1,i}^1) \rightarrow (S_2, a_1^2, v_{1,j}^2)\},$$

отражающих промежуточные рассуждения при постановке диагноза и множества продукций

$$D = \bigcup_{j=1}^5 \{D_j: (S_2, a_1^2, v_{1,j}^2) \rightarrow (S_2, f_1^2)\},$$

отражающих постановку окончательного диагноза.

Граф продукций показан на рисунке 1.

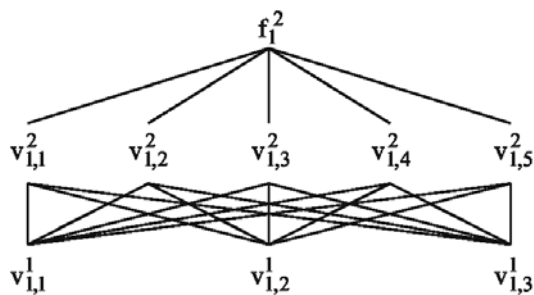


Рисунок 1 – Граф продукций

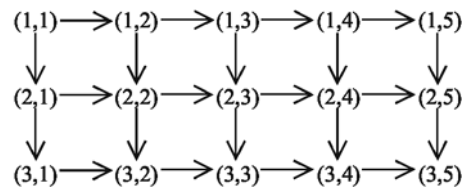


Рисунок 2 – Граф доминирования состояний

Из той последовательности, в которой упоминаются значения атрибута a_1^1 в работе [14, с. 272–273], можно предположить, что эксперт может упорядочить множество V_1^1 его значений по характерности для неисправности f_1^2 в виде отношения

$$R(a_1^1, f_1^2): v_{1,1}^1 \rightarrow v_{1,1}^1 \rightarrow v_{1,3}^1.$$

Аналогично, можно предположить, что

$$R(a_1^2, f_1^2): v_{1,1}^2 \rightarrow v_{1,2}^2 \rightarrow v_{1,3}^2 \rightarrow v_{1,4}^2 \rightarrow v_{1,5}^2.$$

Эти отношения индуцируют отношение $R(f_1^2)$, упорядочивающее по характерности для неисправности f_1^2 все гипотетически возможные состояния системы S . Граф этого отношения показан на рис. 2.

На рис. 2 для простоты вместо значений атрибутов указан их порядок в отношениях характерности. Например, вместо $(v_{1,2}^1, v_{1,3}^2)$ указано просто (2, 3).

Покажем, как можно эффективно использовать бинарное отношение характерности $R(f_1^2)$.

С помощью продукций $P_{3,5}$ и D_5 системе S можно поставить диагноз f_1^2 , используя цепочку продукций $(S_1, a_1^1, v_{1,3}^1) \rightarrow (S_2, a_1^2, v_{1,5}^2) \rightarrow (S_2, f_1^2)$, которая относит состояние $(v_{1,3}^1, v_{1,5}^2)$ системы S к классу f_1^2 . Тогда и все состояния системы S , более характерные для неисправности f_1^2 , тоже должны быть отнесены к этому же классу. Но все такие состояния описываются отношением $R(f_1^2)$.

Таким образом, с помощью всего 2-х продукций $P_{3,5}$ и D_5 и 2 простых отношений $R(a_1^1, f_1^2)$ и $R(a_1^2, f_1^2)$, индуцирующих отношение $R(f_1^2)$, можно ставить диагноз, для постановки которого ранее требовалось 20 продукций из множеств P и D .

6 НАУЧНЫЕ И ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ

Основным недостатком ДЭС, основанных на представлении знаний в виде продукций, является невоз-

возможность классификации всех гипотетически возможных состояний диагностируемого объекта. Точнее, классификация всех состояний требует такого количества продукций, которое сравнимо с количеством этих состояний. В приведенном выше примере для классификации 15 состояний потребовалось 20 продукций. Поэтому для сложного объекта диагностирования, который может находиться в миллионах различных состояний, представление знаний эксперта в виде продукций является неэффективным подходом по следующим основным причинам [7]:

- процесс вывода имеет низкую эффективность, так как при большом числе продукций значительная часть времени затрачивается на непроизводительную проверку условий применения правил;
- проверка непротиворечивости системы продукций становится весьма сложной из-за недетерминированности выбора выполняемой продукции из конфликтного множества.

Предлагаемый в данной работе подход, основанный на представлении «поверхностных» знаний эксперта в виде продукций и «глубинных» знаний – в виде бинарных отношений, обладает следующими преимуществами:

- характерное для рассматриваемой ПО «Диагностика КТ» представление знаний эксперта в виде продукций;
- существенное уменьшение количества продукций, необходимых для классификации всех гипотетически возможных состояний объекта диагностирования, за счет использования бинарных отношений характеристики (в рассмотренном выше примере для классификации 15 состояний потребовалось всего 2 продукции и 2 линейных отношения);
- возможность проверки полученной формальной системы представления знаний на непротиворечивость, используя методы, описанные в работах [15–18], для проверки непротиворечивости представления знаний в виде бинарных отношений.

7 СРАВНЕНИЕ С ЛУЧШИМИ АНАЛОГАМИ

Одним из эффективных методов построения ДЭС является метод КЛАСС [15, 16]. Основная идея этого метода состоит в предъявлении эксперту для классификации некоторых специальным образом выбранных состояний объекта диагностирования.

Существенным недостатком такого подхода является большая размерность вектора, описывающего состояние сложного объекта. По мнению большинства психологов [19], процессы принятия решений происходят в кратковременной памяти человека. Объем этой памяти ограничен примерно 7 элементами (7 ± 2). Поэтому классификация состояний даже относительно простого объекта, описание которого содержит 10 атрибутов,

опытным экспертом может в результате оказаться ошибочной и противоречивой.

Предлагаемый в данной работе подход требует от эксперта умения работать одновременно или с двумя атрибутами (при формировании продукции), или с двумя значениями (при сравнении в процессе формирования отношения характерности). В этом и состоит его несомненное преимущество.

Методы ДИФКЛАСС, СТЕПКЛАСС и КЛАНШ [15], существенно более эффективные по числу обращений к эксперту, чем метод КЛАСС, отличаются друг от друга стратегией предъявления состояний эксперту, и тоже обладают описанным выше недостатком метода КЛАСС.

ВЫВОДЫ

Продукционные системы являются естественным формализмом для представлений знаний экспертов в ПО «Диагностика КТ». Однако на их основе нельзя построить работоспособную ДЭС, классифицирующую все возможные состояния объекта диагностирования.

Бинарные отношения, задающие характерность состояний объекта диагностирования для его возможных неисправностей, позволяют строить ДЭС, осуществляющих полную классификацию состояний. Однако методы, построенные на их основе, требуют умения классифицировать состояния, которые в ПО «Диагностика КТ» превосходят возможности даже очень опытных экспертов.

Предлагаемый в данной работе подход, основанный на комбинировании продукционных систем и бинарных отношений, позволит существенно упростить процесс создания ДЭС в ПО «Диагностика КТ».

Перспективным для дальнейших исследований является вопрос минимизации количества продукций, гарантирующих классификацию любого состояния, при заданных бинарных отношениях характерности.

ПЕРЕЧЕНЬ ССЫЛОК

1. Джексон П. Введение в экспертные системы. – М.: «Вильямс», 2001. – 624 с.
2. Люгер Д. Ф. Искусственный интеллект: стратегии и методы решения сложных проблем. – М.: «Вильямс», 2003. – 864 с.
3. Уотермен Д. Руководство по экспертным системам. – М.: Мир, 1989. – 388 с.
4. Кривуля Г. Ф., Липчанский А. М., Механна Сами, Зидат Хабис. Диагностика компьютерных сетей с использованием экспертных систем // Вестник ХГТУ. 2004. – № 1(19) – С. 11–16.
5. Эзер Д. Шерлок Холмс на канальном уровне // Журнал сетевых решений/LAN. – 2005. – Март – С. 44–48.
6. Zhongmin Cai, Xiaohong Guan, Ping Shao, Qinke Peng, Guoji Sun A rough set theory based method for anomaly intrusion detection in computer network systems // Expert Systems. November – 2003. – Vol. 20, No. 5. – P. 251–259.
7. Бондарев В. Н., Аде Ф. Г. Искусственный интеллект. – Севастополь: Изд-во СевНТУ, 2002. – 615 с.

8. Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем. – СПб.: Питер, 2000. – 384 с.
9. Черноруцкий И. Г. Методы принятия решений. – СПб.: БХВ-Петербург, 2005. – 416 с.
10. Джонс М. Т. Программирование искусственного интеллекта в приложениях. – М.: ДМК Пресс, 2004. – 312 с.
11. Бакаев А. А., Гриценко В. И., Козлов Д. Н. Экспертные системы и логическое программирование. – Киев: Наук. думка, 1992. – 220 с.
12. Корнеев В. В., Гареев А. Ф., Васютин С. В., Райх В. В. Базы данных. Интеллектуальная обработка информации. – М.: «Нолидж», 2000. – 352 с.
13. Хаханов В. И. Техническая диагностика элементов и узлов персональных компьютеров. – К.: ИСМО, 1997. – 308 с.
14. Нессер Д. Дж. Оптимизация и поиск неисправностей в сетях. К.: «Диалектика», 1996. – 384 с.
15. Ларичев О. И. Теория и методы принятия решений. – М.: Логос, 2002. – 392 с.
16. Ларичев О. И., Мечитов А. И., Мошкович Е. М., Фуремс Е. М. Выявление экспертных знаний (процедуры и реализации). – М.: Наука, 1989. – 128 с.
17. Ларичев О. И., Мошкович Е. М. Качественные методы принятия решений. Вербальный анализ решений. – М.: Наука. Физматлит, 1996. – 208 с.
18. Розен В. В. Цель – оптимальность – решение (математические модели принятия оптимальных решений). – М.: Радио и связь, 1982. – 168 с.
19. Солсо Р. Когнитивная психология. – СПб.: Питер, 2002. – 592 с.

Надійшла 23.03.06

Розглядаються проблеми створення експертних систем для діагностики станів комп'ютерної техніки. Для формалізації експертних знань пропонується використання комбінації продукційних систем і бінарних відносин. Ефективність запропонованого підходу показана на прикладі діагностики локальної комп'ютерної мережі.

The problems of expert systems making for diagnosis of every possible states of computer equipment are considered. In order to formalize knowledge of experts, using of production systems in combination with binary relations is proposed. The efficiency of suggested approach is shown on example of local computer network diagnosis.

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

ПРОГРЕССИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

PROGRESSIVE INFORMATION TECHNOLOGIES

УДК 681. 3. + 681. 5. 007

Н. В. Алипов, Н. М. Кораблев, М. И. Хиль, М. В. Гусятин

СТРУКТУРА ЦИФРОВОГО АВТОМАТА С ПСЕВДОСЛУЧАЙНЫМИ ПЕРЕХОДАМИ ИЗ НАЧАЛЬНОГО СОСТОЯНИЯ В ОДНО И ТО ЖЕ КОНЕЧНОЕ СОСТОЯНИЕ

В статье впервые описана структура цифрового автомата с псевдослучайными переходами для систем защиты информации при ее передаче.

ВВЕДЕНИЕ

Известны цифровые автоматы Мура и Миля, основной особенностью которых является то, что каждое состояние автомата связано тем или иным образом с выходным сигналом. Однако из всего многообразия подобных автоматов можно выделить такие автоматы (автоматы Мура), у которых выходной сигнал зависит только от конечного состояния автомата.

Рассмотрим такие автоматы. Первый из них автомат поиска данных по ключу (предполагается, что записи упорядочены по возрастанию либо по его убыванию ключа).

В подобной ситуации поиск осуществляется по такому алгоритму: диапазон ключей, расположенных по возрастанию их значений, разбивают на две равные части, затем сравнивают значения искомого ключа, со

значением ключа, расположенного на середине диапазона. При этом могут возникнуть такие исходы:

$$\text{а) } K_{\text{и}} \leq K_n; \text{ б) } K_{\text{и}} > K_n,$$

где $K_{\text{и}}$ – значение искомого ключа; K_n – значение ключа, расположенного на середине диапазона $[K_{\text{min}}, K_{\text{max}}]$; K_{min} – минимальное значение ключа; K_{max} – максимальное значение ключа.

Для исходов а) и б) соответственно будут сформированы такие новые интервалы неопределенности.

$$K_{\text{и}} \in \left\{ K_{\text{min}}, K_{\text{min}} + \left] \frac{K_{\text{max}} - K_{\text{min}}}{2} [\right\};$$

$$K_{\text{и}} \in \left\{ K_{\text{min}} + \left] \frac{K_{\text{max}} - K_{\text{min}}}{2} [, K_{\text{max}} \right\}.$$

В дальнейшем каждый из диапазонов уменьшаем снова в два раза и процесс продолжается тех пор, пока будет истинным такой предикат

$$P\{K_{\text{и}} = K_n\}.$$

© Алипов Н. В., Кораблев Н. М., Хиль М. И., Гусятин М. В., 2006

Если предположить, что в процессе сравнения K_{ii} и K_n действуют случайные возмущения, то результат поиска будет недостоверным, (автомат приходит в конечное состояние не соответствующее искомой записи). В таких ситуациях у автомата наблюдается псевдослучайные блуждания по внутренним его состояниям.

Другим примером подобных автоматов является устройство управления аналого-цифрового преобразователя (АЦП).

В таких автоматах основной операцией является сравнения:

$$x + \xi(t_j) \leq x_a^j,$$

где x – значение аналогового сигнала; $\xi(t)$ – случайное возмущение, накладываемое на x либо x_a^j ; x_a^j – эталонное значение, вырабатываемое устройством управления на j -м такте его работы. Не трудно заметить, что если $\xi(t_j) = 0$, то случайное блуждание отсутствует. Если же $\xi(t_j) \neq 0$, то возникает неправильный переход автомата на j -м такте его работы, которое никакими известными методами никаким образом невозможно скорректировать.

Следующим примером подобных автоматов является устройство защиты информации, реализующее случайные подстановки. Случайные блуждания автомата поиска по ключу и устройства управления аналого-цифрового преобразования устраняется помехоустойчивыми специальными алгоритмами их функционирования [4].

Генерация псевдослучайных перестановок для символов открытого текста в настоящее время невозможна: не разработаны структуры ЦАПП и принципы формирования виртуальных последовательностей.

Целью исследований является разработка структуры ЦАПП и принципов формирования виртуальных последовательностей.

СТРУКТУРА ЦАПП

Генерация псевдослучайных подстановок для символов входного алфавита (x_1, x_2, \dots, x_n) осуществляется следующим образом: первоначальный алфавит кодируют десятичными числами, принадлежащими множеству $\{1, 2, 3, \dots, N\}$; затем заменяют символы открытого текста их числовыми эквивалентами; потом организуют преобразование кода каждого символа открытого текста x_j^1 устройством, реализующим помехоустойчивое аналого-цифровое преобразование. Для организации блуждания этого устройства по внутренним состояниям в структуру устройства вводят блок, формирующий нерегулярную либо регулярную виртуальную последовательность, амплитуда которой на j -м такте равна x_j^2 .

Затем на вход устройства, реализующего помехоустойчивый алгоритм аналого-цифрового преобразования подаются аддитивную смесь

$$NZ = (x_j^1 + x_j^2). \quad (1)$$

Поскольку x_j^2 является псевдослучайной величиной, то NZ изменяется от такта к такту работы устройства преобразования открытого текста в шифротекст и этим самым обеспечивается псевдослучайное блуждание устройства преобразования по его внутренним состояниям.

Алгоритм функционирования описанного устройства формирования для символов входного алфавита псевдослучайных подстановок синтезируется таким образом, чтобы из одного начального его состояния в одно и тоже конечное (каждое конечное состояние соответствует одному символу входного алфавита) существовало бы множество маршрутов; все маршруты были бы различной длины; маршруты выбирались псевдослучайно.

Структура цифрового автомата с псевдослучайными переходами (ЦАПП) для системы защиты информации приведена на рисунке 1.

На рис. 1 использованы такие функциональные узлы: Σ – двоичный сумматор; ЦК – цифровой компаратор; ЦА – цифровой автомат; ДВП – датчик виртуальной последовательности (организует псевдослучайное блуждание по внутренним состояниям ЦА); A, L, H – параметры виртуальной последовательности (A – максимальная амплитуда случайного выброса; L – максимальная длительность выброса виртуальной последовательности; H – минимальный интервал времени между двумя соседними выбросами виртуальной последовательности); x_j^2 – двоичный код амплитуды выброса виртуальной последовательности на j -м такте работы

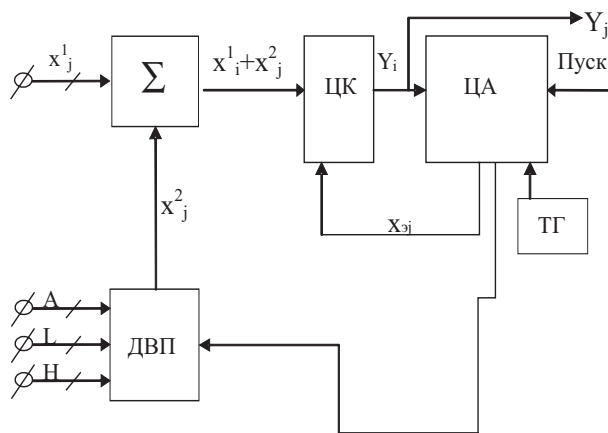


Рисунок 1 – Структурная схема организации псевдослучайных переходов цифрового автомата

структуры; x_j^1 – двоичный код кодируемого символа; Y_j – выходной двоичный сигнал ЦК и выходной двоичный сигнал ЦАПП; а) $x_{эj}$ – эталонный сигнал, выработывает ЦА на j -м шаге его работы.

Выразим параметры виртуальной последовательности через безмерные величины [1]:

$$a = A_{\max}/\delta; l = L_{\max}/\Delta t; h = H_{\min}/\Delta t,$$

где δ – дискретность преобразования по уровню; Δt – длительность работы ЦАПП.

В работе [1] рассматриваются двуполярные, однополярные импульсные последовательности вида $A_2(a, l, h)$, $A_1(a, l, h)$. Эти последовательности генерируются программно и схематически. В дальнейшем будем рассматривать только их программную реализацию.

В таблице 1 приведены возможные виды таких последовательностей.

Таблица 1

№ п/п	Параметры			Условное обозначение последовательности
	a	l	h	
1	+	+	+	$A_{1,1}; A_{2,1}$
2	+	+	–	$A_{1,2}; A_{2,2}$
3	+	–	+	$A_{1,3}; A_{2,3}$
4	+	–	–	$A_{1,4}; A_{2,4}$
5	–	+	+	$A_{1,5}; A_{2,5}$
6	–	+	–	$A_{1,6}; A_{2,6}$
7	–	–	+	$A_{1,7}; A_{2,7}$
8	–	–	–	$A_{1,8}; A_{2,8}$

В таблице знак «+» означает, что соответствующий параметр является случайной величиной; знак «–» означает, что соответствующий параметр не является случайной величиной.

АЛГОРИТМ ФОРМИРОВАНИЯ

Как следует из структуры цифрового автомата с псевдослучайными переходами необходимо, чтобы проектируемое программное средство (ПС) реализовывало бы следующие функции:

- функцию FUN1 формирования виртуальной последовательности;
- функцию FUN2 суммирования двоичного кода символов входного текста $x_j^1 (i = \overline{1, N_1}; N_1$ – размерность массива входного текста (данных) $\|MWD\|_{N_1}$) и дво-

ичного кода амплитуды выброса виртуальной последовательности x_j^2 ;

- функцию FUN3 организации псевдослучайных переходов ЦА.

Наиболее сложными с точки зрения реализации являются функции FUN1 и FUN3. Каждая из этих функций реализуется отдельным ПС. В дальнейшем условно будем их обозначать следующим образом:

- ПС1 – ПС, реализующее функцию FUN1;
- ПС2 – ПС, реализующее функции FUN2 и FUN3.

Взаимодействие этих ПС можно организовать двумя способами:

- а) ЦАПП переходит в новое состояние по мере формирования новых значений амплитуды выброса виртуальной последовательности:

- вновь сформированное функцией FUN1 значение амплитуды виртуальной последовательности передается функции FUN2;

- сформированное новое значение суммы $(x_j^1 + x_j^2)$ передается функции FUN3, которая формирует новое эталонное значение $x_{эj}$ и сравнивает его с суммой $(x_j^1 + x_j^2)$.

- б) первоначально ПС1 формируются значения амплитуд выбросов виртуальной последовательности (реализуется FUN1), которые записываются в массив $\|MWP\|_N$, затем передается управление ПС2, реализующего функции FUN2 и FUN3.

Если при выполнении функции FUN2 использованы все элементы массива $\|MWP\|_N$, то переходят к повторному выполнению функции FUN1, программным средством ПС1. В этом случае выполнение функций FUN2 и FUN3 задерживается до момента завершения процесса формирования массива $\|MWP\|_N$. Затем передается управление ПС2, которое реализует функции FUN2 и FUN3. Этот процесс продолжается до тех пор, пока не будет получен шифротекст для последнего символа открытого текста, записанного в массиве $\|MWD\|_N$.

Недостатком первого способа является то, что структура ПС1 должна включать и структуру ПС2. Эта особенность первого способа усложняет процесс взаимодействия ПС1 и ПС2. Этому недостатка лишен второй способ взаимодействия ПС1 и ПС2.

На рисунке 2 приведена структура программного комплекса реализующего взаимодействие ПС1 и ПС2 согласно способу б). Для описания схемы организации взаимодействия программных средств ПС1 и ПС2 использованы следующие массивы и переменные:

- $\|MWD\|_{N_1}$ – массив входного текста;
- $\|MWP\|_N$ – массив формируемых значений амплитуд выбросов виртуальной последовательности;
- $\|MWK\|_{N_2}$ – массив шифротекста, получаемого в результате замены символов входного текста их подстановками;

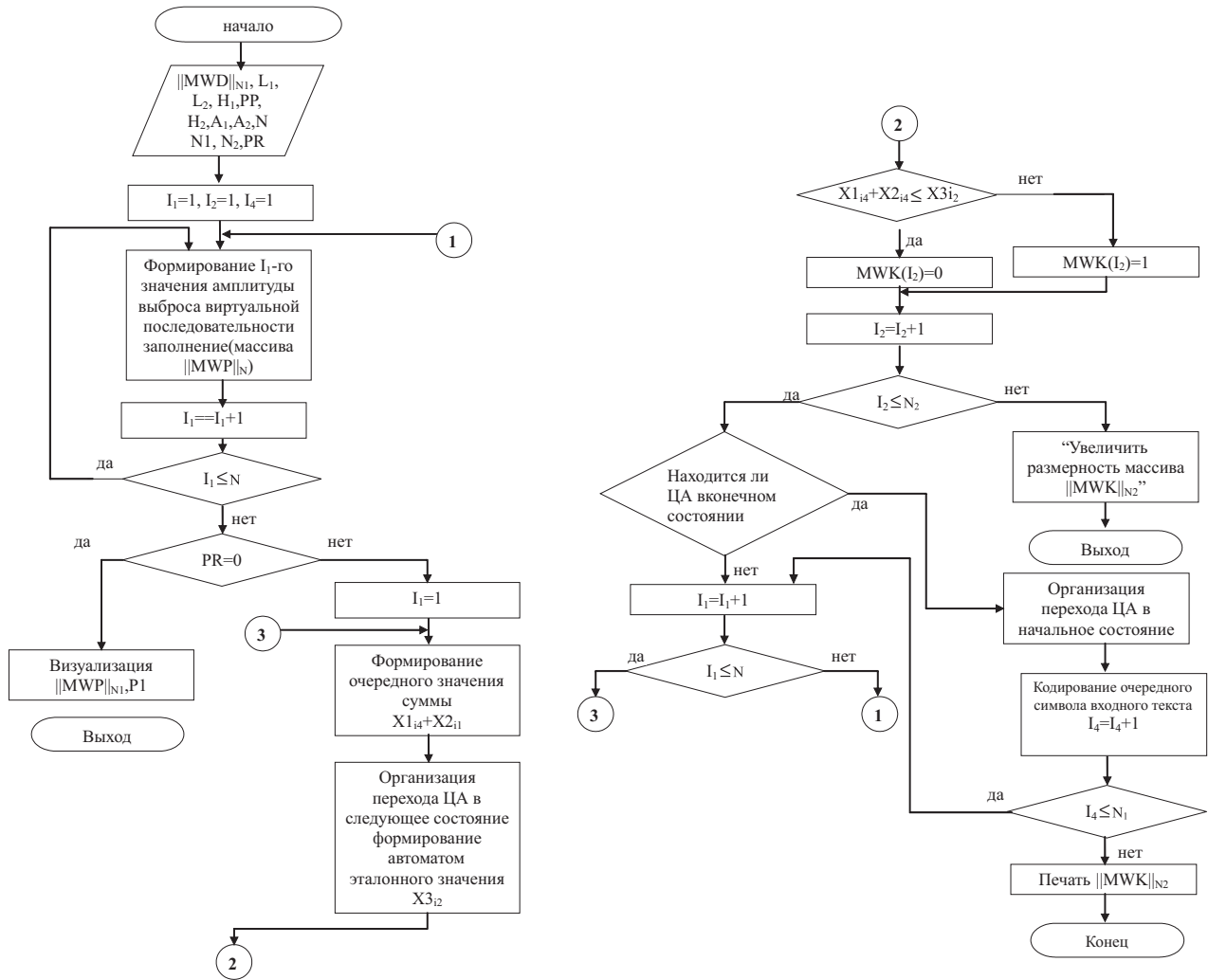


Рисунок 2 – Структурная схема взаимодействия ПС1 и ПС2

- I_1 – индекс массива $\|MWP\|_N$;
 - I_2 – индекс массива; $\|MWK\|_{N_2}$;
 - I_3 – индекс массива; $\|MWD\|_{N_1}$;
 - (L_1, L_2) – минимальное и максимальное значения длительности вопроса виртуальной последовательности;
 - (H_1, H_2) – минимальное и максимальное значения паузы – (интервал времени между двумя соседними выбросами);
 - (A_1, A_2) – минимальное и максимальное значения амплитуду выбросов (импульсов) виртуальной последовательности;
 - X_{3I_2} – эталонные значения, сформированное ЦА на I_2 -м шаге его работы.
- Введен флаг, для которого

$$PR = \begin{cases} 0, & \text{ПС реализует только FUN1;} \\ 1, & \text{ПС реализует функции FUN1, FUN2, FUN3.} \end{cases}$$

Одним из важных элементов структуры ЦАПП является датчик виртуальных последовательностей. В дальнейшем мы остановимся на его организации.

При формировании виртуальной последовательности необходимо каким-то образом выбирать ее начало: пауза или выброс; первый или i -й временной квант времени паузы (выброса), $i = \overline{1, N}$.

Для достаточно продолжительной регулярной последовательности ($H_1 = H_2$; $L_1 = L_2$; $A_1 = A_2$) вероятность появления паузы определяется соотношением:

$$P = \frac{H_1}{H_1 + L_1} = \frac{H_1}{TO},$$

где TO – период следования выбросов в последовательности ($TO = H_1 + L_1$).

В том случае, когда $H_1 \neq H_2, L_1 \neq L_2$ вероятности появления паузы (выброса) будем определять, используя следующие соотношения:

$$T_0 = (H_1 + \lfloor \frac{H_2 - H_1}{2} \rfloor) + (L_1 + \lfloor \frac{L_2 - L_1}{2} \rfloor),$$

$$P_1 = \frac{H_{cp}}{T_0},$$

где $H_{cp} = H_1 + \lfloor \frac{H_2 - H_1}{2} \rfloor$.

Если предположить, что распределение пауз в последовательности распределено по равномерному закону, то выбор начала последовательности будем осуществлять такой процедурой:

– выбор начала последовательности:

$$R = \text{RND};$$

если $R \leq P_1$, то последовательность начинается с паузы, в противном случае – с выброса;

– выбор кванта времени паузы (выброса), с которого начинается последовательность, осуществляется таким образом:

$$R = \text{RND};$$

$$X_2 = H_1 + \lfloor R(H_2 - H_1) \rfloor;$$

где $\lfloor \rfloor$ – целая часть числа;

$$R = \text{RND};$$

$$X_1 = L_1 + \lfloor R(L_2 - L_1) \rfloor;$$

$$R = \text{RND};$$

$$NC_2 = 1 + \lfloor R(X_2 - 1) \rfloor;$$

$$R = \text{RND};$$

$$NC_1 = 1 + \lfloor R(X_1 - 1) \rfloor.$$

Для организации процесса формирования виртуальной последовательности введем дополнительно следующие переменные:

P_1 – вероятность появления паузы в виртуальной последовательности;

PP – вероятность появления в последовательности выброса положительной полярности;

Z_1 – переменная знака амплитуды выброса виртуальной последовательности;

Y_1 – переменная значения амплитуды формируемого выброса последовательности;

x_1, x_2 – вспомогательные переменные соответственно для длительности выброса и паузы виртуальной последовательности;

NC_1 – счетчик временных квантов длительности текущего выброса последовательности, которые необходимо еще сформировать;

NC_2 – счетчик временных квантов длительности текущей паузы виртуальной последовательности, которые необходимо еще сформировать;

R_1 – флаг, указывающий на то, что формируемая виртуальная последовательность начинается с паузы ($R_1 = 1$) либо с выброса ($R_1 = 0$).

Значения индексов I_1, I_2, I_3, I_4 , изменяются в таких пределах:

$$I_1 = \begin{cases} I_1 + 1, & \text{если } I_1 + 1 \leq N; \\ 1 & \text{в противном случае;} \end{cases} \quad I_2 = \overline{1, N_2};$$

$$I_3 = \begin{cases} I_3 + 1, & \text{если } I_3 + 1 < K_3; \\ 0 & \text{в противном случае;} \end{cases} \quad I_4 = \overline{1, N_1};$$

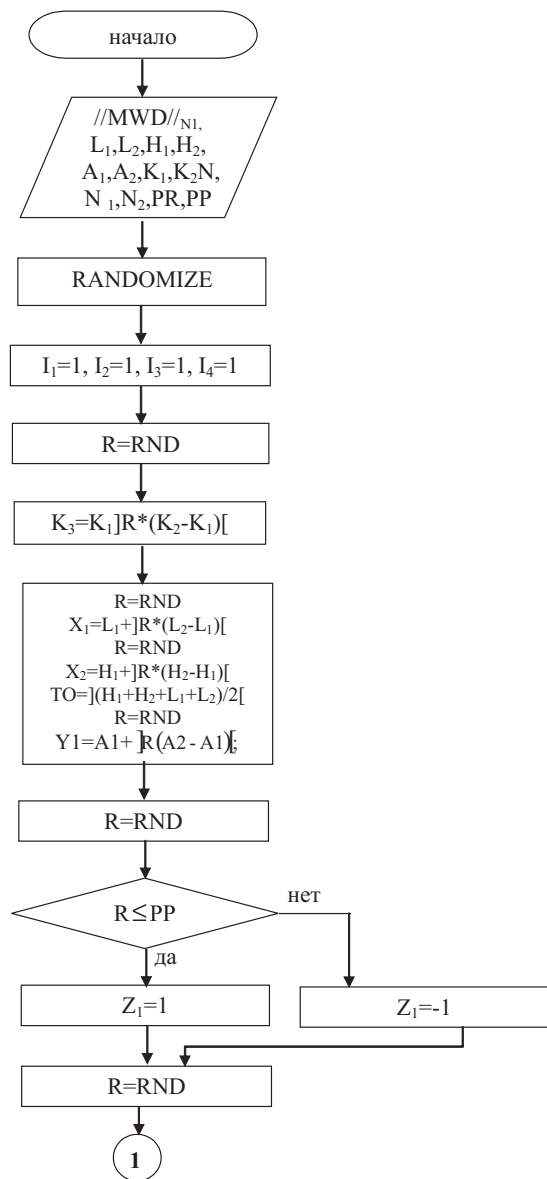
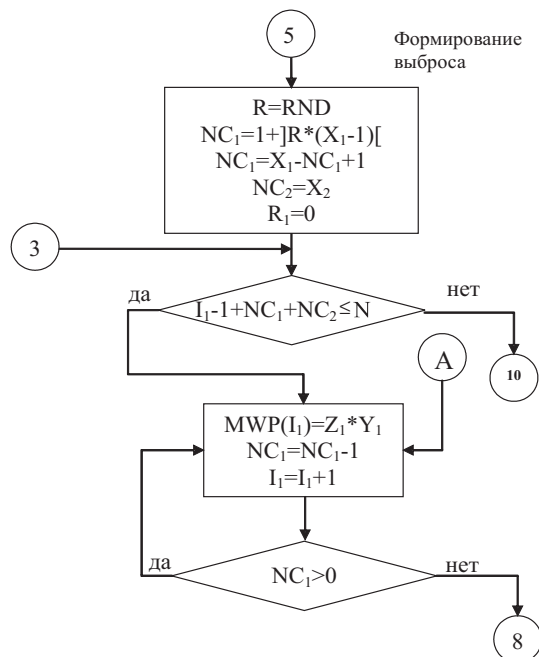
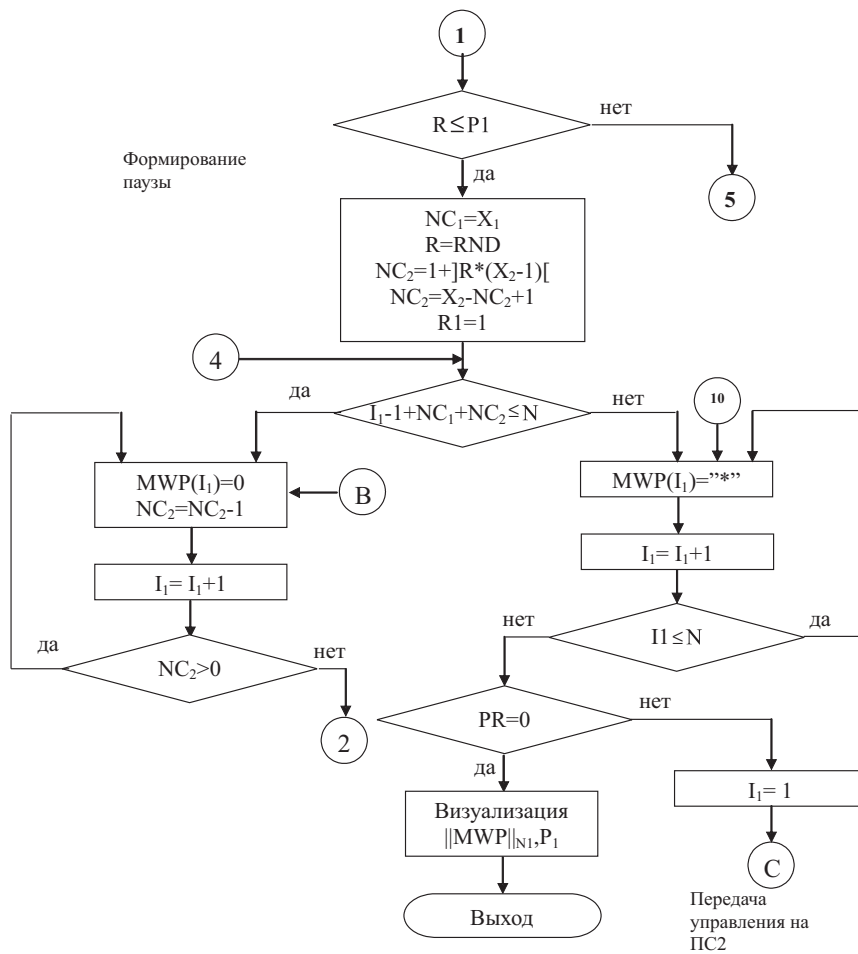
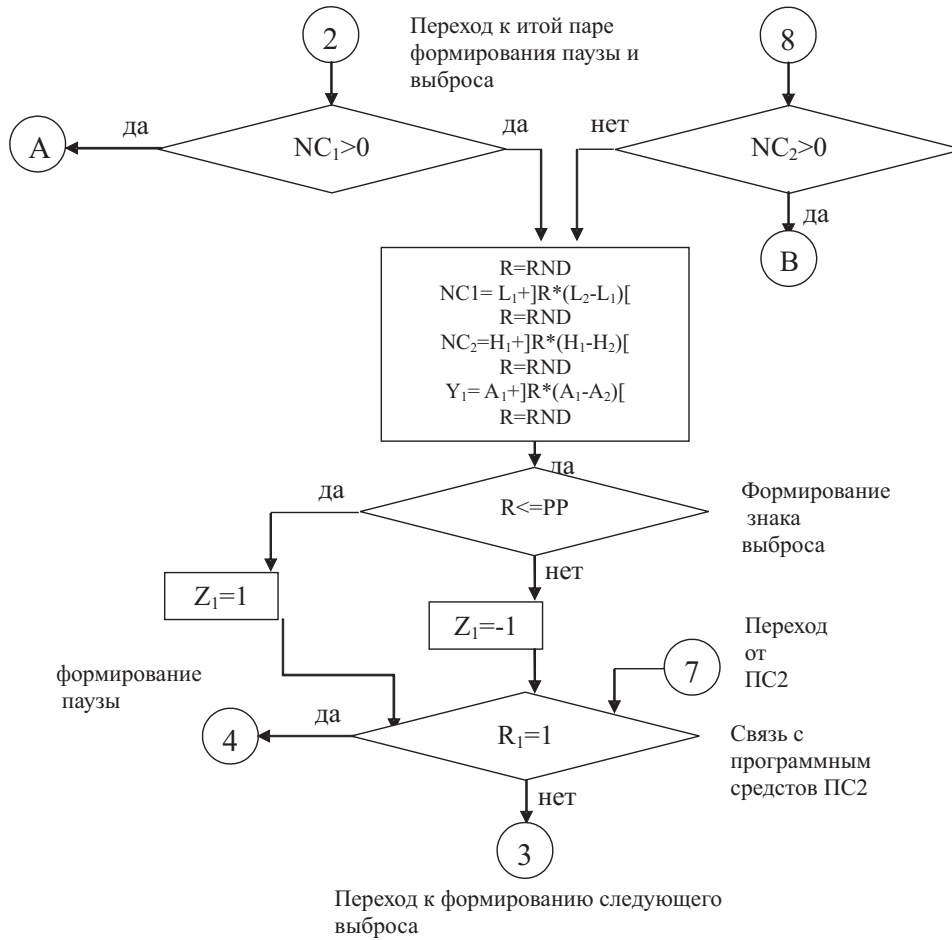


Рисунок 3 – Схема алгоритма формирования нерегулярных симметричных виртуальных последовательностей



Продолжение рисунка 3



Продолжение рисунка 3

Исходные значения индексов:

$$I_1 = 1, I_2 = 1, I_3 = 1, I_4 = 1.$$

Основные соотношения, используются при формировании виртуальной последовательности:

$$\begin{aligned} R &= \text{RND}, x_1 = L_1 + IR(L_2 - L_1); \\ R &= \text{RND}, x_2 = H_1 + IR(H_2 - H_1); \\ TO &= \lfloor (H_1 + H_2 + L_1 + L_2) / 2 \rfloor; \\ R &= \text{RND}, \\ Y_1 &= A_1 + IR(A_2 - A_1); \\ R &= \text{RND}, \\ NC_1 &= 1 + \lfloor R(x_1 - 1) \rfloor; \\ R &= \text{RND}, \\ NC_2 &= 1 + \lfloor R(x_2 - 1) \rfloor. \end{aligned}$$

Формирование знака амплитуды выброса последовательности осуществляется на основании такой проце-

дуры: $R = \text{RND}$; если $R \leq PP$, то выброс положительной полярности ($Z_1 = 1$), в противном случае выброс отрицательный ($Z_1 = -1$).

Принятые соглашения, правила формирования переменных и индексов позволили определить схему алгоритма формирования виртуальной нерегулярной симметрической последовательности (см. рисунок 3).

ЗАКЛЮЧЕНИЕ

Итак, разработанные нами принципы организации случайного блуждания цифрового автомата по внутренним его состояниям, структура цифрового автомата с псевдослучайными переходами, правила и схема алгоритма формирования виртуальных нерегулярных последовательностей, структура программного комплекса для формирования псевдослучайных шифров замены позволяют решить задачу генерации псевдослучайных шифров замены в системах защиты информации.

ПЕРЕЧЕНЬ ССЫЛОК

Надійшла 23.03.06

1. Альсведе Р., Вегенер И. Задачи поиска. – М.: Мир, 1982. – 368 с.
2. Алипов Н. В. Дискретные автоматы с псевдослучайными переходами и подстановочные методы защиты информации на их основе // Радиоэлектроника и информатика. – 2001. – № 4. – С. 95–98.
3. Алипов Н. В., Алипов И. Н., Ребезюк Л. Н. и др. Датчики виртуальной последовательности, используемые для организации функционирования дискретных автоматов в системах защиты информации // Радиотехника. – 1999. – Вып. 11. – С. 33–39.
4. Алипов Н. В. Синтез оптимальных помехоустойчивых вопросников для угадывания числа с ложными ответами // Проблемы бионики. – 1987. – Вып. 38. – С. 108–117.

Розроблено принцип організації блукання цифрового автомата по внутрішніх станах, структура нашого автомата і правила і схема алгоритму формування віртуальних нерегулярних послідовностей. Такі автомати дозволяють вирішувати задачу генерації псевдоімовірних шифрів заміни в системах захисту інформації.

The principles of roaming of digital automata by internal states, rules and algorism scheme of forming virtual irregular sequences were developed. Proposed digital automates allow to solve the problem of generating of pseudorandom replacement codes in information protection system.

УДК 65.012.8: 004.492

Г. Г. Грездов

НОВЫЙ СПОСОБ ОЦЕНКИ ВЕЛИЧИНЫ ОСТАТОЧНОГО РИСКА ПРИ ФОРМИРОВАНИИ ЭКОНОМИЧЕСКИ ЭФФЕКТИВНОЙ КОМПЛЕКСНОЙ СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ В АВТОМАТИЗИРОВАННЫХ СИСТЕМАХ

Предлагается методика оценки вероятности проявления угроз информации. Рассматриваются два класса угроз информации.

ВВЕДЕНИЕ

В настоящее время актуальна проблема разработки эффективных систем защиты информации. При этом следует рассматривать следующие аспекты эффективности указанных систем: во-первых, система защиты информации должна эффективно противодействовать угрозам, которые могут нанести ущерб защищаемой информации (проблема эффективности механизмов защиты информации). Во-вторых, процесс защиты информации в автоматизированной системе (далее – АС) можно рассматривать как процесс распределения ресурсов, выделяемых на защиту информации (проблема экономической эффективности). При этом особое значение приобретает проблема оценки величины рисков информации.

Существующие подходы к решению задачи оценки величины рисков

В настоящее время широкое распространение получили следующие методики оценки информационных рисков:

– механизм оценивания рисков на основе нечеткой логики [1];

– методика SANS/GIAC (SANS Global Information Assurance Certification – программа профессионального обучения и сертификации, учрежденная институтом SANS) [2];

– метод ожидаемых потерь [3, 4].

Рассмотрим особенности указанных подходов решения задачи оценки величины рисков.

Механизм оценивания рисков на основе нечеткой логики, по существу, является экспертной системой, в которой базу знаний составляют правила, отражающие логику взаимосвязи входных величин и риска. При этом выполняются следующие требования:

– для входных величин и риска заданы трехуровневые шкалы, на которых определены нечеткие термы, соответствующие «большому», «среднему» и «низкому» значениям переменных;

– логика связи входных величин и риска соответствует «табличному» механизму оценки риска, представленному в рекомендациях NIST 800-30;

– значимость всех логических правил вывода одинакова (все весовые коэффициенты продукционных правил равны единице).

Таблица 1 – Оценка риска по трехуровневым шкалам

Вероятность	Ущерб		
	Большой	Средний	Низкий
Большая	Б	С	Н
Средняя	С	С	Н
Низкая	Н	Н	Н

Методика SANS/GIAC. Формула для определения уровня серьезности атаки выглядит следующим образом:

$$S = (C + L) - (SC + NC), \quad (1)$$

где S – величина риска, связанного с осуществлением атаки; C – величина возможного ущерба определяется критичностью ресурсов, против которых направлена атака; L – вероятность успешного осуществления атаки; SC – эффективность контрмер системного уровня; NC – эффективность контрмер сетевого уровня.

В рамках методики используется балльная шкала оценки указанных выше величин в зависимости от многих факторов.

Метод ожидаемых потерь основан на вычислении убытков, которые может понести компания, в сравнении с вложениями, которые должны предотвратить нарушения политики безопасности. Указанный метод основан на практическом опыте организаций, эксплуатирующих АС, обрабатывающих информацию, которая составляет коммерческую тайну. Метод оперирует понятием «финансовой выгоды» – сбережениями, которые могут быть получены при использовании системы информационной безопасности.

$$AS = ALE \cdot E - AC, \quad (2)$$

где AS – ежегодные сбережения (Annual Saving); ALE – показатель ожидаемых потерь (Annualized Loss Expectancy); E – эффективность используемой системы защиты информации; AC – ежегодные затраты на безопасность (Annual Cost).

Недостатки современных подходов решения задачи оценки величины рисков

1. Методы нестрогой математики создают предпосылки для решения задачи, но не гарантируют ее эффективного решения [5]. Такая гарантия может быть обеспечена рациональными действиями людей, которые используют нечеткие алгоритмы. Из этого следует, что организация функционирования систем с высоким уровнем неопределенности должна включать в себя подготовку персонала к решению соответствующих задач с использованием методов нестрогой математики.

2. Методика SANS/GIAC не позволяет учитывать возможности различных категорий противника по реализации деструктивных воздействий.

3. Метод ожидаемых потерь рассматривает риски как математическое ожидание потерь. Эта методика не учитывает многих факторов, оказывающих влияние на безопасность информации [6]. Например, не были учтены потери, которые могла понести АС вследствие применения механизмов защиты информации [7].

ПОСТАНОВКА ЗАДАЧ ИССЛЕДОВАНИЯ

В соответствии с изложенным выше, задачи исследования могут быть сформулированы таким образом:

1. Разработка общей модели процесса выбора средств защиты информации.

2. Разработка методики формирования адекватной системы защиты информации АС, которая позволит должным образом учесть характерные для АС угрозы информации.

ОБЩАЯ МОДЕЛЬ ПРОЦЕССА ФОРМИРОВАНИЯ ЭКОНОМИЧЕСКИ ЭФФЕКТИВНОЙ КОМПЛЕКСНОЙ СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ В АВТОМАТИЗИРОВАННОЙ СИСТЕМЕ

Под общими моделями процессов защиты информации в компьютерных системах будем понимать модели, которые позволяют оценивать общие характеристики указанных систем и процессов. Основное назначение указанных моделей состоит в создании предпосылок для объективной оценки общего состояния компьютерной системы с точки зрения меры уязвимости или уровня защищенности информации в ней. Необходимость в таких оценках обычно возникает при анализе общей ситуации с целью выработки стратегических решений при организации защиты информации.

На рис. 1 приведена общая модель процесса выбора средств защиты информации Системы¹. Отметим, что порядок работ по формированию состава комплексной системы защиты информации (далее – КСЗИ) АС подробно изложен в [8]. Таблица 2 содержит обозначения переменных, используемых в общей модели процесса выбора средств защиты информации.

Сформулируем задачи, которые должны быть решены для выбора средств защиты информации Системы [4, 7, 8].

1. Разработать модель функционирования Системы и модель использования ее ресурсов. Результатом должна быть технологическая схема функционирования Системы ($\{TS\}$) и параметры использования ресурсов Системы ($\{MR\}$). Указанный этап необходим для адекватной оценки Системы, определения функциональных участков Системы, классификации обслуживающего персонала, описания правил доступа к информации и т. д.

2. Построить модель вероятного противника (внутреннего и внешнего), оценить его возможности. Результатом построения указанной модели должны стать сведения о категориях противника, его возможностях по реализации атак, а также времени на реализацию атаки, которым обладает атакующая сторона (T).

¹ Здесь и далее под термином «Система» будем понимать объект защиты – АС класса 1 и 2.

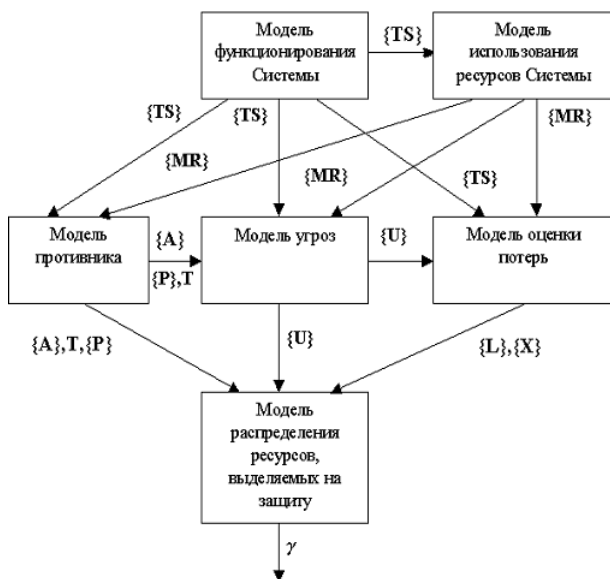


Рисунок 1 – Общая модель процесса выбора средств защиты информации

Таблица 2 – Обозначения переменных, используемых в общей модели процесса выбора средств защиты информации

Переменные	Обозначения переменных
{A}	Множество средств реализации атак на АС
{AS}	Множество параметров, описывающих объект защиты
C	Финансовые средства, которыми располагает противник
G	Сведения об эффективности существующих средств защиты информации по противодействию угрозам
{L}	Множество количественных оценок потерь Системы в случае успешной реализации угроз информации
{MR}	Множество параметров использования ресурсов Системы
{P}	Множество нарушителей
T	Время, которым располагает атакующая сторона для реализации угроз
{TS}	Технологическая схема функционирования Системы
{U}	Множество угроз информации
{X}	Множество количественных оценок потерь Системы в случае применения средств защиты информации
{Z}	Множество средств защиты информации
Ψ	Сведения о возможностях средств реализации угроз информации
γ	Вектор использования средств защиты информации в Системе

3. Разработать модель угроз информации Системы. В качестве исходных данных для построения этой модели необходимы: технологическая схема функционирования Системы ($\{TS\}$), модель использования ее ресурсов ($\{MR\}$), а также время, которым располагает атакующая сторона для реализации своих действий (T). Результатами модели должен стать список угроз информации. При этом для каждой угрозы должны быть учтены частота ее возможного проявления, источник угрозы, компоненты Системы, на которые она направлена, а также описание угрозы.

4. Разработать модель оценки потерь. В качестве исходных данных модели должны быть заданы перечень угроз информации ($\{U\}$), модель использования ресурсов Системы ($\{MR\}$), а также технологическая схема функционирования Системы ($\{TS\}$). Указанная модель должна учитывать как возможные потери, вызванные успешными атаками, так и потери от применения средств защиты информации.

5. Исходя из возможностей противника, а также на основании модели угроз, необходимо построить вероятную модель распределения ресурсов на защиту информации. Исходными данными рассматриваемой модели будут: множество угроз информации Системы ($\{U\}$), вектор значений возможных потерь в случае успешной реализации угроз ($\{L\}$), а также время, которым располагает атакующая сторона (T). Результатом этой модели должен стать вектор использования средств защиты информации (γ) в Системе.

Исходя из вышеизложенного, рассмотрим формальное описание каждой из перечисленных моделей. Обозначения переменных, используемых в моделях, приведены в таблице 2.

Модель функционирования системы может быть формально представлена в виде функции:

$$F_{MF}(AS) \rightarrow \{TS\}.$$

Модель использования ресурсов Системы представляет собой функцию:

$$F_{IR}(\{AS\}, \{TS\}) \rightarrow \{MR\}.$$

Модель противника представляет следующую функцию:

$$F_{MP}(\{AS\}, \{TS\}, \{MR\}) \rightarrow (\{P\}, C, T, \{A\}).$$

Модель угроз описывает следующий функционал:

$$F_{MU}(\{MR\}, \{TS\}) \rightarrow \{U\}.$$

Модель оценки возможных потерь описывается следующим образом:

$$F_{ML}(\{TS\}, \{MR\}, \{U\}) \rightarrow (\{L\}, \{X\}).$$

Модель распределения ресурсов, выделяемых на защиту информации:

$$F_MZ(\{U\}, \{L\}, \{X\}, \{A\}, \{P\}, \{Z\}, T, C) \rightarrow (\gamma).$$

Рассмотрим порядок взаимодействия моделей. На первом этапе производится разработка модели функционирования Системы. Результаты этого этапа будут использованы во всех последующих моделях процессов защиты информации. Итогом моделирования должна стать технология функционирования Системы, которая описана формально.

После получения описания технологии функционирования Системы необходимо определить ресурсы Системы, то есть определить, какие ресурсы используются Системой для решения задач по обработке информации. Далее, необходимо формально описать схему использования ресурсов. В дальнейшем эта схема понадобится для определения возможных объектов атак злоумышленников, понадобится при формировании каналов утечки информации и т. д.

Следующим этапом является разработка модели противника. Для этого понадобятся результаты предыдущих этапов: технология функционирования и схема использования ресурсов. Указанные данные помогут классифицировать возможного противника, что в свою очередь позволит в дальнейшем построить адекватную систему защиты информации. Результатом построения модели противника должно стать множество возможных категорий злоумышленников, а также объем финансовых средств и возможностей, которыми они обладают.

Когда работы предыдущего этапа будут закончены, можно будет приступать к работам по формированию модели угроз информации. Исходными данными для моделирования будут технология функционирования системы и схема использования ресурсов. Результатом этого этапа моделирования должен быть перечень угроз информации системы, а также способов их реализации. При этом каждый элемент перечня должен содержать информацию о категориях злоумышленников, которые могут реализовать указанную угрозу. Кроме того, должны быть указаны свойства информации, которые будут нарушены в случае успешной реализации угрозы.

После получения указанного перечня можно приступать к построению модели оценки потерь. Для этого необходимо для каждой угрозы из перечня провести работы по подсчету оценки возможных потерь, которые может понести Система. Исходными данными для построения такой модели будут технология функционирования системы, схема использования ресурсов, а также перечень угроз информации системы. Результатом моделирования будет выступать перечень возможных потерь.

После завершения всех перечисленных выше этапов можно приступать к решению задачи выбора средств защиты информации. Результатом решения указанной задачи должен стать бинарный вектор применения известных средств защиты информации.

Задача выбора средств защиты информации является многовариантной, так как одна и та же угроза информации может быть нейтрализована с помощью различных способов: организационных мер, программных или аппаратных средств. Очевидно, что применение средств защиты информации сопряжено с некоторыми затратами, размер которых тоже необходимо учитывать. Кроме того, различные средства защиты информации обладают различными показателями эффективности по противодействию угрозам информации Системы. Таким образом, задача выбора необходимых услуг безопасности можно рассматривать как задачу оптимизации по многим критериям [7, 8].

ФОРМИРОВАНИЕ КСЗИ АС ДЛЯ ЗАЩИТЫ ИНФОРМАЦИИ, СОСТАВЛЯЮЩЕЙ КОММЕРЧЕСКУЮ ТАЙНУ

Используя результаты предыдущих этапов моделирования, решить задачу выбора средств защиты информации для защиты Системы.

Минимизировать

$$f(\gamma) = \sum_{i=1}^N L_i \left(R_i - \sum_{j=1}^M G_{ij} \cdot \gamma_j \right) \quad (3)$$

при следующих ограничениях:

$$\sum_{j=1}^M \gamma_j \cdot (C_j + X_j) \leq C_d. \quad (4)$$

Таблица 3 содержит описание переменных, используемых в модели формирования КСЗИ АС.

Рассмотрим более подробно суть предлагаемого решения задачи распределения средств, выделяемых на защиту информации. В роли целевой функции выступает риск Системы. Согласно [10], риском называется функция вероятности реализации определенной угрозы, вида и величины нанесенного ущерба. Указанная формула рассматривает суммарный риск от угроз информации всей Системы. Для этого вычисляется сумма потерь (L), которые понесет Система в случае реализации каждой из угроз.

Методика определения указанной величины состоит в следующем. На первом этапе определяется размер потерь, которые понесет Система в случае реализации отдельной угрозы (в финансовом эквиваленте). Полученное значение (размерность величины – денежный

Таблица 3 – Параметры переменных, используемых в модели формирования КСЗИ АС

Обозначения переменных	Значения переменных	Ограничения переменных	Размерности переменных
N	Число угроз информации	$N > 0$	–
L_i	Оценка стоимости потерь в случае реализации i -й угрозы	$L_i > 0$	Гривны
R_i	Вероятность реализации i -й угрозы	$0 \leq R_i \leq 1$	–
M	Число существующих средств защиты	$M > 0$	–
G_{ij}	Эффективность j -го механизма защиты информации по нейтрализации i -й угрозы	$0 \leq G_{ij} \leq 1$	–
γ_i	Признак использования i -го механизма защиты информации в составе КСЗИ АС (равен 1, если механизм задействован в составе КСЗИ, в противном случае равен нулю)	$\gamma_i \in (0;1)$	–
C_d	Средства, которые могут быть выделены на защиту информации в АС	$C_d > 0$	Гривны
C_j	Затраты на приобретение (разработку) и использование j -го механизма защиты информации	$C_j > 0$	Гривны
X_j	Размер потерь АС, вызванных использованием j -го механизма защиты информации в составе КСЗИ АС	$X_j > 0$	Гривны

эквивалент) умножается на вероятность проявления угрозы, которая будет получена на втором этапе.

При определении вероятности проявления дестабилизирующих факторов необходимо учесть следующие обстоятельства:

1. Неизвестно, какие средства, которыми располагает противник, могут быть использованы для нанесения ущерба Системе (вектор R в обозначениях модели).

2. Применение механизмов защиты информации уменьшает риск проявления угроз, однако необходимо определить состав средств, используемых для защиты информации в Системе (в обозначениях модели – вектор γ).

Решение задачи состоит в отыскании значений вектора γ . Задача выбора средств защиты информации является задачей линейного программирования и может быть решена по методу решения транспортной задачи [7, 9].

ОПРЕДЕЛЕНИЕ ВЕРОЯТНОСТИ ПРОЯВЛЕНИЯ i -Й УГРОЗЫ (R_i)

В настоящее время одним из важнейших требований к КСЗИ АС является ее адекватность реальным условиям [4]. Для построения адекватной системы защиты информации необходимо реально оценить возможности вероятного противника.

При определении вероятностей проявления угроз информации будем полагать, что все множество угроз информации АС формируется из множества активных и пассивных угроз. Причиной возникновения пассивных угроз будем считать уязвимости и особенности компонентов Системы, активных – действия вероятного

противника. Пример множества классов угроз информации Системы приведен в Приложении А.

Будем полагать, что вероятность любой угрозы информации определяется через вероятность «активной» (R_a) и «пассивной» (R_i) составляющей угрозы.

Значение R_i может быть получено статистическими методами, или выведено с помощью метода экспертных оценок [5].

При оценке возможных действий противника сложно определить, какой именно из доступных способов будет им выбран для нанесения ущерба объекту защиты [3, 10].

Таблица 4 содержит порядок средств, выделяемых на реализацию атак различными категориями злоумышленников.

Таблица 4 – Сравнительный анализ возможностей вероятного противника

Категории вероятного противника	Средства, выделяемых для реализации атак (доллары США)
Одиночки	100
Группы хакеров	1000
Мелкие преступные группы	100000
Крупные преступные группы	1000000
Транснациональные преступные организации, спецслужбы иностранных государств	100000000

Поэтому при получении составляющей R_a будем исходить из наихудших предположений о возможностях противника [11]. Значение $R_a = 1$, если финансовые возможности противника превышают стоимость хотя бы одного из средств нападения, способного вызвать

дестабилизирующий фактор. В противном случае элемент вектора будет равен 0.

Формально эта зависимость может быть описана таким образом:

$$R_{a_i} = 1, \text{ если } \min(C(A_j)) \leq C_{\text{прот}},$$

для которых $\Psi_{ij} = 1$,

где A_j – j -е средство реализации угрозы информации; $C(A_j)$ – стоимость j -го средства реализации угроз информации; $C_{\text{прот}}$ – размер финансовых средств, которыми располагает противник; Ψ_{ij} – возможности j -го средства по реализации i -й угрозы информации (коэффициент равен 1, если j -е средство способно создать i -ю угрозу информации; в противном случае коэффициент равен 0).

Вероятность проявления i -й угрозы может быть получена следующим образом:

$$R_i = \max(R_a; R_n). \quad (5)$$

ВЫВОДЫ ИЗ ИССЛЕДОВАНИЯ И ПЕРСПЕКТИВЫ ДАЛЬНЕЙШИХ РАЗРАБОТОК

1. Разработана модель распределения ресурсов, выделяемых на защиту информации в АС. В отличие от существующих, модель использует новые параметры, учитывающие потери, которые может понести объект защиты вследствие применения механизмов защиты.

2. В работе предложена методика оценки вероятности проявления угрозы информации. В отличие от существующих, для решения указанной задачи вероятность проявления угрозы информации для АС, рассматривается как комбинация активной и пассивной составляющей угроз.

Отметим направления перспективных разработок для решения задач формирования КСЗИ АС:

– совершенствование методов моделирования, позволяющих более качественно моделировать бизнес-процессы АС различного класса и назначения;

– процесс защиты информации в АС может быть рассмотрен как бесконечная антагонистическая игра (взаимодействия средств реализации атак и средств КСЗИ АС).

ПЕРЕЧЕНЬ ССЫЛОК

1. Балашов П. А., Безгузиков В. П., Кислов Р. И. Оценка рисков информационной безопасности на основе нечеткой логики. – Электрон. дан. – Режим доступа: http://www.active.net.ru/document/ocenka_riskov.htm. – Загл. с экрана.
2. Астахов Е. Актуальные вопросы выявления сетевых атак. – Электрон. дан. – Режим доступа: <http://sphe-rix.jeka.ru/index.php?page=2&prnt=871>. – Загл. с экрана.

3. Баутов А. Эффективность защиты информации. – Электрон. дан. – Режим доступа: <http://www.citforum.ru/security/articles/eff>. – Загл. с экрана.
4. Петренко С. А., Терехова Е. М. Обоснование инвестиций в безопасность. // Научно-технический журнал «Защита информации. INSIDE». – 2005. – № 1. – С. 49–53.
5. Герасименко В. А. Защита информации в автоматизированных средствах обработки данных. – Книги 1, 2. – М.: Энергоатомиздат, 1994. – 400 с. и 176 с.
6. Медведовский И. Современные методы и средства анализа и контроля рисков информационных систем компаний. – Электрон. дан. – Режим доступа: <http://www.Citforum.ru/Products/Dsec/itrisk>. – Загл. с экрана.
7. Грездов Г. Г. Способ решения задачи формирования комплексной системы защиты информации для автоматизированных систем 1 и 2 класса: Препр. / НАН Украины. Отделение гибридных моделирующих и управляющих систем в энергетике ИПМЭ им. Г. Е. Пухова. – К.: 2005. – № 1. – 66 с.
8. НД ТЗІ 3.7-001-99. Методичні вказівки щодо розробки технічного завдання на створення комплексної системи захисту інформації в автоматизованій системі. – Киев: ДСТЗИ СБ України, 1999. – 14 с.
9. Демьянов В. Ф., Малоземов В. Н. Введение в мини-макс. – М.: Наука, 1972. – 368 с.
10. НД ТЗІ 1.1-003-99. Термінологія в галузі захисту інформації в комп'ютерних системах від несанкціонованого доступу. – Киев: ДСТЗИ СБ України, 1999. – 30 с.
11. Гайкович В. Основы безопасности информационных технологий. – Электрон. дан. – Режим доступа: <http://bezpeka.mk.ua/articles/zi.htm> – Загл. с экрана.

ПРИЛОЖЕНИЕ А. КЛАССЫ УГРОЗ ИНФОРМАЦИИ АВТОМАТИЗИРОВАННЫХ СИСТЕМ

В описании классов угроз информации используются следующие термины:

1. Количественная недостаточность – физическая нехватка одного или нескольких компонентов Системы для обеспечения требуемой защищенности информации.
2. Качественная недостаточность – несовершенство одного или нескольких компонентов Системы, в силу чего не обеспечивается требуемая защита информации.
3. Отказ – утрата способности Системы или ее компонента выполнять определенную функцию.
4. Сбой – временное нарушение работоспособности компонента Системы, следствием чего может быть неправильное выполнение этим элементом своих функций.
5. Ошибка – неправильное (одноразовое или систематическое) выполнение элементом Системы одной или нескольких функций.
6. Умышленное действие¹ – действия людей, специально направленные на нарушение защищенности информации.
7. Побочное явление – явление, сопутствующее выполнению компонентом Системы своих основных функций, следствием которого может быть нарушение защищенности информации.

¹ Ошибки, сбои и отказы могут быть следствием двух типов угроз (умышленные и неумышленные действия), однако умышленные и неумышленные действия сами по себе представляют угрозу информации Системы.

Таблица А.1 – Классы пассивных угроз информации Системы

Типы угроз информации	Информация	Персонал	Аппаратное обеспечение	Программное обеспечение	Помещения
Количественная недостаточность	1.1	1.2	1.3	1.4	1.5
Качественная недостаточность	2.1	2.2	2.3	2.4	2.5
Побочные явления	3.1	3.2	3.3	3.4	3.5
Отказы, вызванные умышленными действиями	4.1	4.2	4.3	4.4	4.5
Сбои, вызванные умышленными действиями	5.1	5.2	5.3	5.4	5.5
Ошибки, вызванные умышленными действиями	6.1	6.2	6.3	6.4	6.5
Отказы, вызванные неумышленными действиями	7.1	7.2	7.3	7.4	7.5
Сбои, вызванные неумышленными действиями	8.1	8.2	8.3	8.4	8.5
Ошибки, вызванные неумышленными действиями	9.1	9.2	9.3	9.4	9.5

Таблица А.2 – Классы активных угроз информации Системы

	Информация	Персонал	Аппаратное обеспечение	Программное обеспечение	Помещения
Хищение	10.1	10.2	10.3	10.4	–
Подмена	11.1	11.2	11.3	11.4	–
Считывание	12.1	12.2	12.3	12.4	12.5
Нарушение функционирования	13.1	13.2	13.3	13.4	13.5

С помощью предложенной методики может быть получено 45 возможных классов пассивных угроз информации для Системы.

Рассмотрим возможные классы активных угроз информации Системы (Таблица А.2 содержит 18 классов указанных угроз.)

Надійшла 28.02.06
Після доробки 11.07.06

Пропонується методика оцінки імовірності прояву загроз інформації. Розглядаються два класи загроз інформації.

The estimations method for probability of data threats is offered. Two classes of data threats are considered.

УДК 519.71

В. И. Дубровин, Е. Н. Федорченко

ДИАГНОСТИКА НА ОСНОВЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ

Исследуются генетические алгоритмы Canonical, Simple, Genitor, Gibrid и Island при обучении нейронной сети решению задачи диагностики. Для решения тестовой задачи диагностики создана автоматизированная система подбора весов нейронной сети. Приведены результаты экспериментов, позволяющие выявить наилучший генетический алгоритм для решения задач диагностики.

ВВЕДЕНИЕ

Разработка автоматизированных систем диагностики и прогнозирования предполагает создание математического обеспечения для сбора, хранения, обработки и анализа диагностической информации.

При исследовании недостаточно изученных процессов на первый план выдвигается задача отыскания на-

иболее информативных признаков, влияющих на качественные показатели процесса. Знание информативных признаков дает возможность сделать математическое описание процесса достаточно простым при максимуме подобия. В то же время, если из множества информативных признаков в математическую модель не будет включен хотя бы один признак (и данный признак будет стабилизирован во время исследований), то полученная модель будет представлять собой лишь сечение в факторном пространстве, позволяющее найти только частный экстремум целевой функции. Если же признак будет изменяться произвольным образом, то это вызовет существенный рост уровня шума. В обоих случаях главный экстремум не может быть достигнут. Таким образом, возникает задача оценки сравнитель-

ного влияния большого числа входных признаков на целевую функцию и исключения из дальнейшего рассмотрения неинформативных признаков [1].

Для решения оптимизационных задач, связанных с анализом диагностических данных при наличии случайных и непредсказуемых воздействий, перспективно использовать математический аппарат генетических алгоритмов (ГА).

Целью данной работы является анализ наиболее часто применяемых на практике генетических алгоритмов: канонического, простого, джениторного, гибридного и островного по скорости и точности распознавания, а также создание автоматизированной системы для проведения исследований ГА.

1 МОДЕЛИ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ

Схема работы обобщенного ГА представлена на рис. 1.

Суть ГА заключается в циклической замене одной популяции следующей, более приспособленной. Каждая особь оценивается мерой ее приспособленности согласно тому, насколько хорошо соответствующее ей решение задачи. Приспособленность определяется как вычисленная целевая функция для каждой из хромосом. Правила отбора стремятся оставить только те точки, где имеет место максимум целевой функции. Наиболее приспособленные особи получают возможность воспроизводить потомков с помощью перекрестного скрещивания с другими особями популяции. Это приводит к появлению новых особей, которые сочетают в себе некоторые характеристики, наследуемые ими от родителей. Наименее приспособленные особи с меньшей вероятностью смогут воспроизвести потомков, так что те свойства, которыми они обладали, будут постепенно исчезать из популяции в процессе эволюции. Это означает, что гены от высоко адаптированных или приспособленных особей будут распространяться в увели-

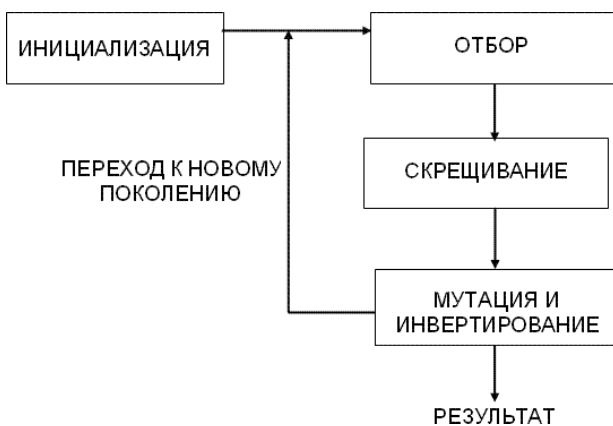


Рисунок 1 – Схема обобщенного ГА

чивающемся количестве потомков на каждом последующем поколении.

Таким образом, из поколения в поколение хорошие характеристики распространяются по всей популяции. Комбинация хороших характеристик от различных родителей иногда может приводить к появлению супер-приспособленного потомка, чья приспособленность больше, чем приспособленность любого из его родителя. Скрещивание наиболее приспособленных особей приводит к тому, что исследуются наиболее перспективные участки пространства поиска. В конечном итоге популяция будет сходиться к оптимальному решению задачи [2–4].

После скрещивания иногда происходят мутации и инвертирования, или спонтанные изменения в генах, которые случайным образом разбрасывает точки по всему множеству поиска. Процесс сходимости за счет отбора должен быть более выражен в сравнении с разбросом точек за счет мутации и инвертирования. Иначе сходимость к экстремумам не будет иметь места, сходимость за счет отбора не должна быть слишком быстрой, иначе все точки могут собраться вблизи локального экстремума, а глобальный экстремум, так и не будет найден.

После скрещивания, мутации и инвертирования размер популяции увеличивается. Однако для последующих преобразований необходимо сократить число хромосом текущей популяции, поэтому снова применяется оператор отбора.

Дальнейшая работа генетического алгоритма представляет собой итерационный процесс выполнения генетических операторов.

Рассмотрим основные модели ГА.

Канонический (классический) ГА. Алгоритм имеет следующие характеристики:

- фиксированный размер популяции;
- фиксированная разрядность генов;
- пропорциональный отбор;
- особи для скрещивания выбираются случайным образом;
- односточечный кроссовер и односточечная мутация.

Следующее поколение формируется из потомков текущего поколения без «элитизма». Потомки занимают места своих родителей [5, 6].

Простой (Simple GA) ГА. Он отличается от канонического тем, что использует либо рулеточный, либо турнирный отбор.

Дженитор (Genitor) ГА. Алгоритм отличается от классического ГА следующими тремя свойствами:

- на каждом шаге только одна пара случайных родителей создает только одного потомка;
- этот потомок заменяет не родителя, а одну из худших особей популяции;
- отбор особи для замены производится по ее рангу (рейтингу), а не по приспособленности.

Вначале популяция инициализируется и ее особи оцениваются, дальше выбираются случайным образом две особи, скрещиваются, причем получается только один потомок, который оценивается и занимает место наименее приспособленной особи. Дальше снова производится выбор двух особей случайным образом, и получается потомок. На каждом шаге в популяции обновляется только одна особь [4–6].

Идея гибридного алгоритма (hybrid algorithm) основывается на использовании генетического алгоритма с некоторым другим методом поиска, который подходит для решения выбранной задачи. На каждом поколении каждый полученный потомок оптимизируется этим методом, после чего производятся обычные для ГА действия. При использовании этой модели получается, что каждая особь достигает локального максимума, вблизи которого она находится, после чего подвергается отбору, скрещиванию и мутации [2].

Островной алгоритм (island algorithm) – представляет собой модель параллельного генетического алгоритма [2, 3, 7, 8]. Алгоритм имеет следующие характеристики:

- наличие нескольких популяций, как правило одинакового фиксированного размера;
- фиксированная разрядность генов.

Любые комбинации стратегий отбора и формирования следующего поколения в каждой популяции. Можно сделать так, что в разных популяциях будут использоваться разные комбинации стратегий, хотя даже один вариант дает разнообразные решения на различных «островах».

Ограничений на тип кроссовера и мутации нет.

2 АВТОМАТИЗИРОВАННАЯ СИСТЕМА ГЕНЕТИЧЕСКОГО ПОДБОРА ВЕСОВ НЕЙРОННОЙ СЕТИ

Для исследования свойств ГА была создана автоматизированная система. Поскольку существует большое число различных ГА и их операторов, то целесообразно для обеспечения возможности развития и наращивания методов обработки информации создать систему, которая, имеет модульную структуру.

Разделение на подсистемы позволяет системе быстрее обрабатывать информацию, а также в случае необходимости позволяет облегчить подключение нового модуля в систему.

Подсистема генерации задач содержит процедуру, реализующую генерацию выборок данных с помощью псевдослучайных данных. В связи с тем, что функции ГА часто оперируют случайными величинами, реализация унифицированной процедуры генерации случайных чисел как отдельной процедуры является рациональной. Эта процедура генерирует случайные числа

в необходимом диапазоне, опционально возвращает целые или дробные значения.

Подсистема предобработки данных производит сортировку и вычисление выходов нейронной сети.

В подсистеме реализации настройки весов нейронной сети реализована работа генетических операторов: отбора, скрещивания, мутации и инверсии. Операция отбора производится на основании значений выходов нейронной сети (НС), вычисленных в подсистеме предобработки.

Подсистема интерфейса обеспечивает два вида взаимодействия:

- между системой и внешней средой;
- между подсистемами предобработки и реализации ГА.

При начале работы с системой пользователь должен определить следующие параметры:

- обучающую выборку;
- операторы ГА (заданы их вероятностные характеристики);
- структуру обучающейся нейронной сети (количество слоев и количество нейронов в каждом слое).

При определении пользователем обучающей выборки нужно задать директорию, в которой находятся файлы с обучающими выборками.

При определении операторов ГА задается критерий останова и при необходимости файл, содержащий начальную популяцию.

После останова работы ГА результаты выводятся на экран. По желанию пользователя оптимальный набор признаков и/или популяции последнего поколения могут быть записаны в файл.

Система позволяет пользователю:

- самостоятельно задавать начальную популяцию и структуру НС;
- использовать в качестве обучающей выборки файлы различных типов с различными способами хранения информации;
- самостоятельно выбирать операторы ГА, которые будут использованы в программе, и задавать их вероятностные характеристики;
- визуально контролировать ход работы ГА;
- определять различные способы критериев останова;
- останавливать работу ГА в любой момент времени;
- сохранять конечную популяцию в файле для дальнейшего ее использования в качестве начальной популяции при необходимости продолжения поиска наилучшего сочетания весовых коэффициентов.

Входными данными системы являются:

- файлы с обучающей выборкой;
- структура нейронной сети;
- файлы с начальной популяцией;
- характеристики операторов ГА.

Выходными данными системы являются:

- настроенные весовые коэффициенты нейронной сети;
- файлы с выходной популяцией.

3 ГЕНЕТИЧЕСКИЕ АЛГОРИТМЫ ПРИ ОБУЧЕНИИ НЕЙРОННОЙ СЕТИ

Процесс обучения нейронной сети с помощью генетического алгоритма разбивается на два этапа.

На первом этапе с помощью генетического алгоритма осуществляется поиск общих параметров нейронной сети, а именно: количество скрытых слоев и нейронов в каждом слое.

На втором этапе генетический алгоритм используется для поиска значения связей между нейронами и функцией активации.

Целевая функция задается для множества примеров и представляет собой максимальное относительное отклонение от эталонного значения, выраженное в процентах. Цель применения генетического алгоритма – минимизация функции

$$C = \max_{j=1}^k \frac{(y'_{ij} - y_{ij})}{y_{ij}} \cdot 100 \%,$$

где k – количество примеров; y'_{ij} – значение i -го выхода нейронной сети для j -го примера, $y'_{ij} > 0$.

Генетические алгоритмы для подстройки весов скрытых и выходных слоев используются следующим образом. Каждая хромосома (решение, последовательность, индивидуальность, «родитель», «потомок») представляет собой вектор из весовых коэффициентов (веса считываются с нейронной сети в установленном порядке – слева направо и сверху вниз).

Начальная популяция выбирается случайно, значения весов лежат в промежутке $[-1.0, 1.0]$. Для обучения сети к начальной популяции применяются простые операции: селекция, скрещивание, мутация, – в результате чего генерируются новые популяции.

Алгоритм прекращает работу в одном из следующих случаев:

- найдено решение;
- истекло установленное время работы либо число поколений;
- популяция длительное время не прогрессирует.

В результате работы ГА получается популяция, которая содержит особь, гены которой лучше генов других особей соответствуют требуемым условиям. Данная особь и будет являться найденным с помощью ГА решением. Следует отметить, что найденное решение может и не быть наилучшим, однако оно может быть близко к оптимальному. Отличительной особенностью ГА является то, что вычислительная сложность алгоритма мало зависит от сложности задачи.

4 ЭКСПЕРИМЕНТЫ И РЕЗУЛЬТАТЫ

Модели ГА исследовались при решении тестовой задачи классификации лопаток газотурбинных авиадвигателей.

При исследовании эффективности обучения нейронной сети различными методами генетических алгоритмов, в качестве тестовых примеров рассматривался однослойный персептрон с 32 входными сигналами и одним выходным сигналом 0 и/или 1.

На вход НС подавалось обучающее множество, состоящее из 100 элементов.

Выполнялось обучение персептрона с помощью алгоритма Левенберга-Марквардта и с помощью рассмотренных генетических алгоритмов.

Результаты работы ГА и алгоритма Левенберга-Марквардта приведены в таблице 1.

Генетические алгоритмы стохастичны, поэтому при разных запусках популяция может сходиться к разным решениям (хотя все они в некоторой степени «хорошие»). Островная модель позволяет запустить алгоритм сразу несколько раз и (из полученных решений выбрать наилучшее решение одного острова).

Классический ГА хорош для понимания работы генетических алгоритмов, однако он не является наиболее эффективным из них.

Использование гибридного алгоритма позволяет объединить преимущества ГА с преимуществами классических методов.

Таблица позволяет сделать следующие выводы.

Генетические алгоритмы значительно выигрывают по сравнению с алгоритмом Левенберга-Марквардта для решаемой задачи, генерирует лучшие векторы весовых коэффициентов за меньшее количество времени, при этом точность распознавания выше 80 % (в зависимости от модели ГА), а гибридный ГА показывает нулевую погрешность работая (см. рис. 2–5) при этом в 3–4 раза быстрее.

По результатам экспериментов наиболее быстрыми оказался канонический и дженитор алгоритмы, следующим идет простой и гибридный, затем островной алгоритм (его достоинства при параллельной обработке популяций могут быть использованы в полную силу лишь на многопроцессорном сервере) и алгоритм Левенберга-Марквардта.

Исследования позволили получить следующие результаты погрешности распознавания: погрешность распознавания моделей, построенных на основе гибридного ГА – 0 %, канонического – 5 %, островного – 7 %, простого – 9 %, дженитор – 14 %, алгоритма Левенберга – Марквардта – 0 %.

Проведенные эксперименты позволили выделить канонический генетический алгоритм как самый оптимальный по скорости распознавания и точности.

Таблица 1 – Сравнение результатов работы ГА и алгоритма Левенберга-Марквардта

ГА	Канонический	Простой	Дженитор	Гибридный	Островной	Алгоритм Левенберга-Марквардта	
Популяции из 5 особей						0	1,03
Погрешность	15,6	25	18,9	0	17,2		
Время, сек.	0,2	0,4	0,1	0,2	0,4		
Популяции из 10 особей							
Погрешность	9,4	17,5	15,6	0	10,9		
Время, сек.	0,1	0,3	0,1	0,3	0,4		
Популяции из 20 особей							
Погрешность	9,4	20,3	17,2	0	7,8		
Время, сек.	0,1	0,3	0,1	0,5	0,3		
Популяции из 50 особей							
Погрешность	7,8	17,2	17,2	0	9,4		
Время, сек.	0,2	0,2	0,2	0,7	0,4		
Популяции из 100 особей							
Погрешность	6,5	9,4	15,6	0	7,3		
Время, сек.	0,22	0,46	0,67	0,8	0,76		
Популяции из 250 особей							
Погрешность	7,8	9,4	21,9	0	9,4		
Время, сек.	0,8	1,1	1,1	1,1	1,0		
Популяции из 500 особей							
Погрешность	9,4	9,4	18,8	0	7,3		
Время, сек.	4,2	4,9	6,9	4,5	4,4		
Популяции из 1000 особей							
Погрешность	9,3	20,2	10,9	0	6,2		
Время, сек.	7,1	11,8	4,7	7,3	17,9		

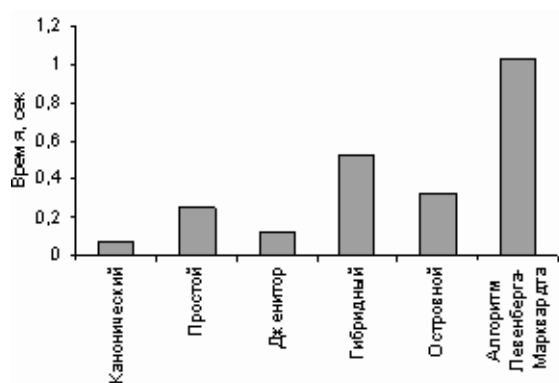


Рисунок 2 – Сравнение результатов работы генетических алгоритмов и алгоритма Левенберга-Марквардта по скорости обучения (10 особей в популяции)

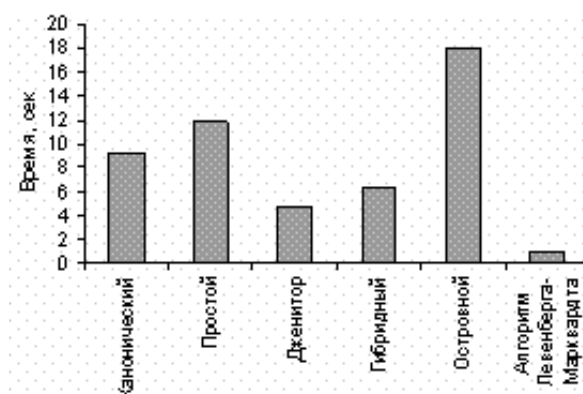


Рисунок 3 – Сравнение результатов работы генетического алгоритмов и алгоритма Левенберга-Марквардта по скорости обучения (1000 особей в популяции).



Рисунок 4 – Залежність часу розпізнавання від розміру популяції на прикладі канонічного алгоритма

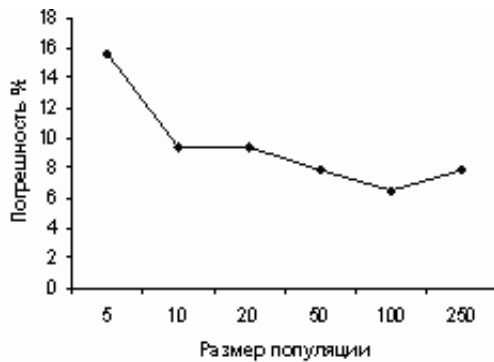


Рисунок 5 – Помилка розпізнавання в залежності від кількості особин в популяції на прикладі канонічного алгоритма

Для підвищення ефективності навчання ІНС необхідно вирішити наступні задачі:

- збільшити швидкість знаходження рішення;

- забезпечити можливість видалення неінформативних і малоінформативних ознак при навчанні ІНС;

- підвищити адаптивні властивості алгоритма.

ПЕРЕЧЕНЬ ССЫЛОК

1. Биргер И. А. Техническая диагностика. – М.: Машиностроение, 1978. – 240 с.
2. Кисляков А. В. Генетические алгоритмы: операторы скрещивания и мутации // Информационные технологии. – 2000. – № 12. – С. 2–24.
3. Кисляков А. В. Генетические алгоритмы: математический анализ некоторых схем репродукции. // Информационные технологии. – 2001. – № 1. – С. 29–34.
4. Назаров А. В., Лоскутов А. И. Нейросетевые алгоритмы прогнозирования и оптимизации систем: – СПб.: Наука и техника, 2003. – 384 с.
5. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы: Пер. с польск. – М.: Горячая линия – Телеком, 2004. – 452 с.
6. Симанков В. С., Частикова В. А. Генетические алгоритмы и поиск оптимальных решений // Автоматизация и современные технологии. – 2003. – № 6. – С. 39–45.
7. Beasley D., Bull D. R., Martin R. R., An Overview of Genetic Algorithms: Part I, Fundamentals // University Computing. – 1993. – № 2. – P. 58–69.
8. Вороновский Г. К., Махотило К. В., Петрашев С. Н., Сергеев С. А. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности. – Харьков: Основа, 1997. – 112 с.

Надійшла 7.06.06

Досліджуються генетичні алгоритми Canonical, Simple, Genitor, Gibrid i Island при навчанні нейронної мережі рішення завдання діагностики. Для рішення тестового завдання діагностики створена автоматизована система підбору ваг нейронної мережі. Наведено результати експериментів, що дозволяють виявити найкращий генетичний алгоритм для рішення завдань діагностики.

The genetic algorithms Canonical, Simple, Genitor, Gibrid and Island are investigated at neural network training for the decision of a diagnostics problem. For the decision of a test problem of diagnostics the automated system of selection of weights of neural networks is created. The results of experiments are given, allowing to reveal the best genetic algorithm for the decision of diagnostics problems.

УДК 681.51:57

А. С. Кулик, А. Ю. Завгородний

ДИАГНОСТИРОВАНИЕ ДАННЫХ КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМ НА ОСНОВЕ ЕСТЕСТВЕННОЙ ИЗБЫТОЧНОСТИ

Представлен подход к диагностированию данных корпоративных информационных систем на основе естественной избыточности, базирующийся на концепции диагностических моделей. Последовательно решены задачи обнаружения ошибки, поиска ошибочных кортежей, ошибочных атрибутов и определения вида ошибки. Представлены оценки вычислительной сложности решенных задач.

ВВЕДЕНИЕ

Повышение конкурентоспособности современной организации не представляется возможным без внедрения в ее деятельность информационных систем управления. Использование организацией такой системы

© Кулик А. С., Завгородний А. Ю., 2006

позволяет не только систематизировать труд административно-управленческого персонала, обеспечить контроль над административно-финансовой деятельностью организации, сократить трудоемкость операций, но и получить мощные инструменты анализа имеющихся данных в целях принятия управленческих решений [1].

В современных условиях, когда объемы перерабатываемой информации огромны, деятельность руководителя без применения таких инструментов анализа будет чрезвычайно непродуктивна. Однако, эти инструменты неэффективны, в случае если данные, накопленные системой управления, неактуальны или недостоверны. Причины искажения данных разнообразны, но особенно следует отметить человеческий фактор – даже профессиональные операторы в силу своих психологических и физических особенностей ошибаются в среднем 1 раз на каждые 100 введенных символов [2]. Таким образом, создание эффективных средств борьбы с искажениями становится необходимым условием для успешного функционирования информационных систем.

В работах, посвященных проблеме качества данных [2–7], достаточно подробно освещены причины, классификация возникающих ошибок, методы обнаружения некоторых классов ошибок. Следует также выделить два сформировавшихся в этой области направления: создание методов повышения актуальности и достоверности данных на основе искусственной избыточности, когда одни и те же данные дублируются в нескольких источниках, и сформировать заключение о достоверности таких данных возможно путем применения мажоритарного принципа, и методы, базирующиеся на естественной избыточности – специфических особенностях данных и их сочетаниях. Не умаляя достоинств методов созданных в рамках первого направления, следует отметить существенное ограничение сферы применения

таких методов вытекающее из необходимости наличия в системе дублирующейся информации. Вместе с тем анализ работ, посвященных созданию средств, базирующихся на естественной избыточности [2, 5, 7], позволяет сделать вывод об отсутствии на сегодняшний день системного подхода к решению такой задачи.

1 ЗАДАЧИ СИСТЕМ ПОВЫШЕНИЯ КАЧЕСТВА ДАННЫХ В КОРПОРАТИВНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Рассмотрим задачи, стоящие перед системой повышения качества данных. В соответствии с современным пониманием категории качества, повышение качества является не событием, а процессом [3]. Упрощенно такой процесс и место процедур повышения качества данных в нем можно представить следующим образом (см. рис. 1).

Пользователи операционных подсистем в соответствии с поступившими документами вводят детализированные данные. Именно на этом этапе данные могут быть искажены ошибками оператора. Все данные, вводимые пользователями, попадают в базы данных операционных подсистем. Далее, по мере накопления связанных между собой данных, в различных подсистемах выполняется их извлечение, преобразование в единый формат и интеграция. По окончании этого этапа получаем универсальное отношение $\tilde{R}'_{\Delta}[k]$ с некоторой, определяемой предметной областью структурой. В общем виде такое отношение можно описать как совокупность заголовка, т. е. множества атрибутов вида $A'_i:T_i$, где A'_i – имя атрибута, а T_i – домен атрибута, определяемый как множество допустимых значений, и тела, т. е. множества m кортежей вида $A'_i:v_i$, где v_i – значение атрибута A'_i , $v_i \in T_i$ [8]. Для упрощения даль-

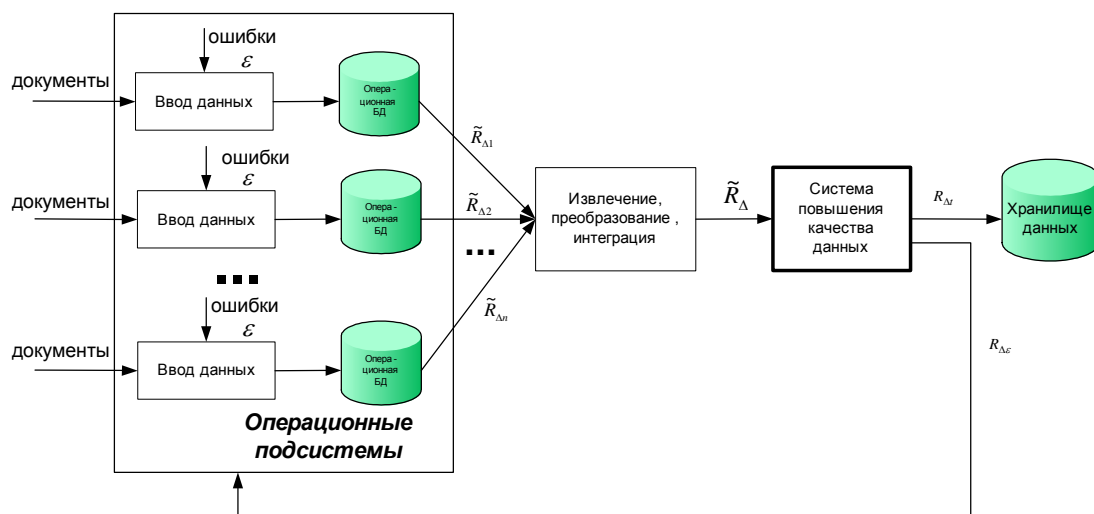


Рисунок 1 – Процесс повышения качества данных на основе естественной избыточности

нейших математических рассуждений определим отношение $\tilde{R}_\Delta[k]$ получаемое из $\tilde{R}'_\Delta[k]$ на бинарном домене $T_{01} = \{0,1\}$ следующим образом. Определим заголовков отношения $\tilde{R}_\Delta[k]$ как множество атрибутов $A = T_1 \cup T_2 \cup \dots \cup T_n$ определенных на домене T_{01} . Таким образом, любой возможной паре $A'_i:v_i$ из отношения $\tilde{R}'_\Delta[k]$ будет соответствовать ровно один атрибут в новом отношении $\tilde{R}_\Delta[k]$. Количество кортежей в отношениях $\tilde{R}'_\Delta[k]$ и $\tilde{R}_\Delta[k]$ будет одинаковым, причем атрибут A_i в кортеже r_k из $\tilde{R}_\Delta[k]$ будет равен 1 тогда, когда в кортеже r'_k из $\tilde{R}'_\Delta[k]$ встречается соответствующая A_i пара $A'_i:v_i$ и будет равен 0 в противоположном случае.

Следует отметить, что, несмотря на некоторые возможности по обнаружению некачественных данных на этапе извлечения, преобразования и интеграции, результирующее отношение $\tilde{R}_\Delta[k]$ все еще содержит ошибки, внесенные на этапе ввода данных.

$$\tilde{R}_\Delta[k] = \hat{R}_\Delta[k] \oplus E[k], \quad (1)$$

где $\hat{R}_\Delta[k]$ – эталонное, т. е. безошибочное отношение; $E[k]$ – отношение, определенное таким же образом, как и отношение $\tilde{R}_\Delta[k]$ и отражающее каждую внесенную ошибку единицами в атрибутах соответствующих неправильно выбранному значению, и значению, которое должно было быть внесено; \oplus – операция, подразумевающая сложение по модулю 2, всех соответствующих друг другу атрибутов отношений $\hat{R}_\Delta[k]$ и $E[k]$.

Тогда задачей системы повышения качества данных является анализ отношения $\tilde{R}_\Delta[k]$, поступившего на вход системы в некоторый момент времени $t[k]$ с целью выявления всех ошибочных атрибутов, нахождения предполагаемого правильного значения (т. е. таким образом, находятся значения отношений $R_\Delta[k]$ и $E[k]$), чтобы затем вернуть данные в операционные подсистемы для исправления. Кроме того, задачей системы является также формирование объяснения пользователю, в чем именно он ошибся. Подытоживая сказанное выше, задача системы повышения качества данных – диагностирование поступивших на вход данных с глубиной до вида отказа.

Информационным ресурсом, позволяющим решать такие задачи, являются знания о конкретных специфических свойствах данных и их сочетаний. Представим знания о данных как множество правил $F = \{f_1, f_2, \dots, f_n\}$. Каждое правило, в свою очередь, можно представить в виде тройки $f = (i, H, C)$, где i – условный номер правила в базе знаний; H, C – множества атрибутов

($H \subset A, C \subset A$ и $H \cap C = \emptyset$), описывающие посылку и следствие правила соответственно. Отражая, таким образом, зависимости между атрибутами, определяемые предметной областью, каждое правило позволяет сформировать двузначный предикат, отражающий не противоречат ли атрибуты кортежа r данному правилу:

$$f_i(r) = ((r_{h_1} = 1) \wedge (r_{h_2} = 1) \wedge \dots \wedge (r_{h_{H_{cnt}}} = 1)) \supset ((r_{c_1} = 1) \vee (r_{c_2} = 1) \vee \dots \vee (r_{c_{C_{cnt}}} = 1)), \quad (2)$$

где $f_i(r) = 0$ – атрибуты r противоречат правилу f_i , а $f_i(r) = 1$ – атрибуты r соответствуют правилу f_i или не противоречат ему вследствие не выполнения условий в посылке правила; $h_1, h_2, \dots, h_{H_{cnt}}, c_1, c_2, \dots, c_{C_{cnt}}$ – все атрибуты, составляющие посылку и следствия правила. Следует отметить, что подобное представление знаний не является единственно возможным [9, 10]. Тем не менее, естественная модульность правил, наглядность и простота их создания [10], наличие методов, позволяющих автоматически находить правила предметной области и, как следствие, упрощать работу эксперта [9–11] обусловили выбор именно такой модели представления знаний о данных.

2 ПОДХОД К ДИАГНОСТИРОВАНИЮ ДАННЫХ НА ОСНОВЕ ЕСТЕСТВЕННОЙ ИЗБЫТОЧНОСТИ

Возникновение в $\tilde{R}_\Delta[k]$ любой ошибки является событием неопределенным с точки зрения момента k его появления, ошибочного кортежа из $\tilde{R}_\Delta[k]$, атрибута в котором она проявится, конкретного вида ошибки. В такой ситуации применение «наивного» подхода, заключающегося в последовательной проверке каждого кортежа из $\tilde{R}_\Delta[k]$ на соответствие каждому правилу из F , будет чрезвычайно неэффективным.

Вместе с тем, подобная проблема при решении задачи диагностирования далеко не впервые возникает перед исследователями. Так, например, в работе [12] представлен системный подход к диагностированию систем управления (СУ), подразумевающий последовательное снятие неопределенности о техническом состоянии. Общая задача диагностирования в данном случае, исходя из заданного множества характерных для объекта диагностирования отказов, разбивается на ряд таких подзадач, как обнаружение отказа, поиск места, установление класса отказа и определение его вида. Такая декомпозиция позволяет существенно сократить среднее время необходимое для проведения процедуры диагностирования, поскольку в случае, если СУ функционирует в номинальном режиме, то будет решена лишь задача обнаружения отказа, после чего сделано соответствующее заключение о наличии отказа в системе.

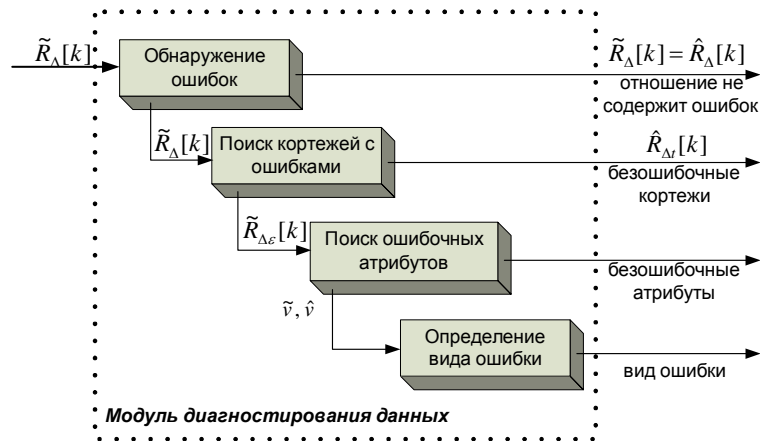


Рисунок 2 – Последовательность этапов диагностирования

Ключевым положением этого подхода является новая концепция диагностических моделей. Диагностические модели, т. е. математические конструкции, связывающие косвенные диагностические признаки с прямыми, строятся для каждой из упомянутых задач диагностирования, формируя, таким образом, иерархию диагностических моделей. Диагностирование по косвенным признакам, позволяющим однозначно судить о наличии соответствующих им прямых, дает возможность не использовать прямые диагностические признаки, вычисление/измерение, которых может быть невозможно или затруднено.

Принципы, которые легли в основу изложенного подхода могут быть применены не только для систем управления, но и для объектов различной природы. Так, например, рассмотренный выше подход к диагностированию СУ стал базовым при формировании подхода к повышению качества данных на основе искусственной избыточности, предложенного в [1, 13]. Упомянутые подходы стали базовыми при создании подхода к диагностированию данных корпоративных информационных систем на основе естественной избыточности. Сформулируем основные принципы предлагаемого подхода:

1) в связи с неопределенностью момента и места возникновения, вида ошибки эффективным по быстродействию является последовательное снятие неопределенности, т. е. следующее друг за другом решение задач обнаружения ошибки в отношении, поиска ошибочных кортежей, поиска ошибочных атрибутов и только затем определение вида отказа;

2) выполнение требования об обнаружении любого отказа или ошибки, зачастую принципиально невозможно или же приводит к неоправданному повышению стоимости и усложнению системы в целом. Поэтому целесообразным является диагностирование исходя из

заданного множества наиболее часто встречающихся, характерных видов ошибочных данных;

3) вычисление прямых признаков отказа не представляется возможным или является алгоритмически сложным, поэтому эффективным является диагностирование по косвенным признакам;

4) для каждой из основных задач в терминах соответствующих прямых признаков ошибки строятся диагностические модели, связывающие прямые и косвенные признаки;

5) оценка качества полученных диагностических моделей производится путем анализа их диагностируемости, т. е. возможности однозначного установления факта наличия соответствующего прямого признака ошибки по вычисленным косвенным признакам.

Руководствуясь данными принципами мы должны последовательно решать задачи обнаружения ошибки в $\tilde{R}_\Delta[k]$, поиска ошибочных кортежей и атрибутов в $\tilde{R}_{\Delta\epsilon}[k]$, формирование эталонных значений и только затем определения вида ошибки. Такая последовательность действий отражена на рис. 2.

Рассмотрим подробнее решение каждой из задач.

3 ОБНАРУЖЕНИЕ ОШИБОК В УНИВЕРСАЛЬНОМ ОТНОШЕНИИ

В соответствии с вышеприведенной схемой диагностирования данных, первой решаемой задачей является обнаружение ошибок в отношении $\tilde{R}_\Delta[k]$. Результатом решения такой задачи должно стать заключение о наличии или отсутствии ошибок в $\tilde{R}_\Delta[k]$. Прямым признаком наличия ошибки в таком случае будет

$$S_{\text{det}} = \{(\exists r_i \in \tilde{R}_\Delta[k])(\exists f_i \in F)(f_j(r_i) = 0)\}. \quad (3)$$

Предикат в (3) интерпретируется следующим образом: «Существует хотя бы один кортеж $r_i \in \tilde{R}_\Delta[k]$, противоречащий, хотя бы одному правилу $f_j \in F$ ». Кванторы существования в предикате (3) подразумевают необходимость применения каждого правила из F к каждому кортежу из $\tilde{R}_\Delta[k]$ для решения задачи обнаружения ошибки в универсальном отношении. В худшем случае сложность такого подхода составит $O(|\tilde{R}_\Delta[k]| * |F|)$, где $|\tilde{R}_\Delta[k]|$ – кардинальность отношения $\tilde{R}_\Delta[k]$, а $|F|$ – количество правил в базе знаний о данных, что, в свою очередь, подтверждает предположение о неэффективности диагностики данных по прямым признакам.

Построим диагностическую модель (ДМ) обнаружения отказа по косвенным признакам. Введем для этого ряд определений. Пусть $Q_{supp}(R, Z)$ – функция поддержки множества атрибутов Z в отношении R :

$$Q_{supp}(R, Z) = |R|_{z_1 = 1 \wedge z_2 = 2 \wedge \dots \wedge z_n = 1} \quad (4)$$

где R – отношение определенное на бинарном домене (см. п. 1); z_1, z_2, \dots, z_n – множество атрибутов отношения R составляющих в совокупности исходный набор Z ; $R|_{\text{логическое выражение}}$ – оператор реляционной алгебры обозначающий выборку из отношения R по заданному условию. Используя функцию (4), определим функцию $d_{det}(R, F)$ следующим образом.

$$d_{det}(R, F) = \sum_{i=1}^{|F_{ruleset}|} Q_{supp}(R, f_{i \ ruleset}) - \sum_{i=1}^{|F_{hyp}|} Q_{supp}(R, f_{i \ hyp}), \quad (5)$$

где $F_{ruleset}$ и F_{hyp} – множества наборов атрибутов, определяемые на основании множества правил о данных следующим образом.

$$F_{hyp} = \prod_{i, H} F, \\ F_{ruleset} = \prod_{i, RS} (EXT F \ ADD(H \cup C) \ AS \ RS),$$

где $\prod_{\text{атрибут}_1, \dots, \text{атрибут}_n} R$ – оператор реляционной алгебры, обозначающий проекцию отношения R по заданному множеству атрибутов; $EXT \ R \ ADD(\text{выражение}) \ AS \ \text{ИмяАтрибута}$ – оператор расширения отношения R , добавляющий новый атрибут *ИмяАтрибута*, вычисляемый для каждого кортежа как указано в *выражении*; $f_{i \ ruleset}, f_{i \ hyp}$ – i -е множество атрибутов из $F_{ruleset}$ и F_{hyp} соответственно.

Тогда ДМ обнаружения ошибок в отношении $\tilde{R}_\Delta[k]$ опишем следующим образом:

$$\Delta d_{det} = d_{det}(\tilde{R}_\Delta([k], F)) - d_{det}(\hat{R}_\Delta([k], F)), \quad (6)$$

где Δd_{det} – косвенный признак наличия ошибки.

Качество полученной ДМ в отношении диагностируемости было проанализировано путем доказательства следующих утверждений.

Утверждение 1. Если R_1 – отношение, содержащее только один кортеж r , а F – множество правил то значение функции $d_{det}(R_1, F)$ будет равным нулю тогда и только тогда, когда кортеж r не противоречит ни одному из правил F , и будет меньше нуля тогда и только тогда, когда кортеж r противоречит хотя бы одному из правил F , т. е. $d_{det}(R_1, F) = 0 \Leftrightarrow (\forall f_i \in F) f_i(r) = 1$ и $d_{det}(R_1, F) < 0 \Leftrightarrow (\exists f_j \in F) f_j(r) = 0$.

Утверждение 2. Отношение R является безошибочным, т. е. ни один из кортежей R не противоречит ни одному из правил из F тогда и только тогда, когда значение функции $d_{det}(R, F)$ будет равно 0, т. е. $(\forall r_1 \in R) (\forall f_1 \in F) f_1(r_1) = 1 \Leftrightarrow d_{det}(R, F) = 0$.

Таким образом доказанное утверждение 3 позволяет, во-первых, упростить предлагаемую в (6) диагностическую модель:

$$\Delta d_{det} = d_{det}(\tilde{R}_\Delta([k], F)) \quad (7)$$

поскольку $d_{det}(\hat{R}_\Delta([k], F)) \equiv 0$ и, во-вторых, говорить о возможности однозначного установления в ней факта наличия прямого признака (3) по косвенному признаку (7).

Для вычисления Q_{supp} поддержки в $|\tilde{R}_\Delta[k]|$ всех множеств атрибутов составляющих $F_{ruleset}$ и F_{hyp} используется известный метод, впервые предложенный в [11] и базирующийся на использовании предварительно построенного хэш-дерева. Вычислительная сложность такого метода составляет $O(|\tilde{R}_\Delta[k]|)$, следовательно вычислительная сложность решения задачи обнаружения ошибки с использованием ДМ (7), также составит $O(|\tilde{R}_\Delta[k]|)$.

4 ПОИСК КОРТЕЖЕЙ С ОШИБКАМИ В УНИВЕРСАЛЬНОМ ОТНОШЕНИИ

В случае, если в результате решения предыдущей задачи было сформировано заключение о наличии ошибок в $\tilde{R}_\Delta[k]$, необходимо найти все кортежи, содержащие ошибочные атрибуты. На практике, результатом решения задачи поиска кортежей с ошибками должны

стать отношения $\hat{R}_{\Delta t}[k]$ и $\tilde{R}_{\Delta \varepsilon}[k]$, такие что $\tilde{R}_{\Delta}[k] = \hat{R}_{\Delta t}[k] \cup \tilde{R}_{\Delta \varepsilon}[k]$, $\hat{R}_{\Delta t}[k] \cap \tilde{R}_{\Delta \varepsilon}[k] = \emptyset$, $\hat{R}_{\Delta t}[k]$ – отношение содержащее все безошибочные кортежи из $\tilde{R}_{\Delta}[k]$; $\tilde{R}_{\Delta \varepsilon}[k]$ – отношение содержащее все ошибочные кортежи из $\tilde{R}_{\Delta}[k]$; \cup , \cap – соответственно операции соединения и пересечения реляционной алгебры; \emptyset – пустое отношение. Прямым признаком наличия ошибки в кортеже r будет

$$S_{t \text{ fnd}} = \{(\exists f_j \in F)f_j(r) = 0\}. \quad (8)$$

Высказывание, составляющее предикат из (8), трактуется как «Существует хотя бы одно правило $f_j \in F$ которому противоречит данный кортеж». Так же как и при решении предыдущей задачи использование такого признака подразумевает применения каждого правила из F к каждому кортежу из $\tilde{R}_{\Delta}[k]$, что является неэффективным и приводит к необходимости создания диагностической модели, которая позволила бы решать задачу поиска по косвенным признаком ошибочности кортежа.

Как и ранее, прежде чем представить соответствующую ДМ, введем ряд определений. Пусть A_r множество всех атрибутов кортежа r , равных 1:

$$A_r = \{a | r_a = 1\}, \quad (9)$$

тогда, используя (9) определим функцию $T_{\text{sup } p}(r, M)$:

$$T_{\text{sup } p}(r, M) = |\{x | x \in M \wedge x \subseteq A_r\}|, \quad (10)$$

где M – некоторое множество, элементами которого в свою очередь являются множества атрибутов; A_r – множество атрибутов, определяемое для заданного кортежа r как в (9). Здесь следует также отметить, что $T_{\text{sup } p}(r, M) \equiv Q_{\text{sup } p}(R_1, Z)$, в случае, если R_1 – отношение, содержащее только один кортеж r , а множества M и Z равны. Тогда функцию $d_{t \text{ fnd}}$ определим как

$$d_{t \text{ fnd}}(r, F) = T_{\text{sup } p}(r, F_{\text{hyp}}) - T_{\text{sup } p}(r, F_{\text{ruleset}}), \quad (11)$$

где F_{ruleset} и F_{hyp} – множества наборов атрибутов определяемые так же как в (5). Функция, определенная в (11) позволяет сформировать следующую ДМ поиска кортежей с ошибками:

$$\Delta d_{t \text{ fnd}} = d_{t \text{ fnd}}(\tilde{r}, F) - d_{t \text{ fnd}}(\hat{r}, F). \quad (12)$$

Как и в модели обнаружения ошибки (6) разность из (12) является косвенным диагностическим признаком, позволяющим, тем не менее, однозначно судить о наличии прямого признака ошибочности кортежа.

Диагностируемость ДМ (12) вытекает из доказанного ранее утверждения 2. Оттуда же следует, что $d_{t \text{ fnd}}(\hat{r}, F) \equiv 0$ и, следовательно, $\Delta d_{t \text{ fnd}} = d_{t \text{ fnd}}(\tilde{r}, F)$. Теперь, определив предварительно отношение $\tilde{R}_{\Delta \text{ ext}}[k]$ на основании исходного $\tilde{R}_{\Delta}[k]$ как $\tilde{R}_{\Delta \text{ ext}}[k] = \text{EXT } \tilde{R}_{\Delta}[k] \text{ ADD}(d_{t \text{ fnd}}(\tilde{r}, F)) \text{ AS } \Delta d_{t \text{ fnd}}$ получим, в качестве результатов решения с использованием ДМ (12) задачи поиска кортежей с ошибками, следующие отношения:

$$\hat{R}_{\Delta t}[k] = R_{\Delta \text{ ext}}[k] |_{\Delta d_{t \text{ fnd}} = 0}, \quad (13)$$

$$\hat{R}_{\Delta \varepsilon}[k] = R_{\Delta \text{ ext}}[k] |_{\Delta d_{t \text{ fnd}} < 0}. \quad (14)$$

На практике значения $T_{\text{sup}}(r, F_{\text{ruleset}})$ и $T_{\text{sup}}(r, F_{\text{hyp}})$ для каждого кортежа могут быть рассчитаны одновременно со значением функции $d_{\text{det}}(\tilde{R}_{\Delta}[k], F)$ при решении задачи обнаружения ошибки. В этом случае вычислительная сложность получения отношений $\hat{R}_{\Delta t}[k]$ и $\hat{R}_{\Delta \varepsilon}[k]$ составит $O(|\tilde{R}_{\Delta}[k]|)$.

5 ПОИСК ОШИБОЧНЫХ АТТРИБУТОВ

В результате решения предыдущей задачи мы получаем $\tilde{R}_{\Delta \varepsilon}[k]$ – отношение, содержащее все ошибочные кортежи из $\tilde{R}_{\Delta}[k]$. Теперь, для каждого кортежа из $\tilde{R}_{\Delta \varepsilon}[k]$ необходимо найти те атрибуты, которые стали причиной того, что кортеж был признан ошибочным. Решение этой задачи производится в два этапа следующим образом. На первом этапе для каждого кортежа из $\tilde{r} \in \tilde{R}_{\Delta \varepsilon}[k]$ на основании правил из F выполняется прямой логический вывод [9]. Результатом такого вывода становится кортеж \tilde{r}^n атрибуты которого равны 1 если они равны 1 в \tilde{r} или существуют правила, в соответствии с которыми они должны быть равны 1. На втором этапе для каждого атрибута A'_i из исходного не преобразованного отношения $\tilde{R}_{\Delta}[k]$ формируется множество $A_{\tilde{r}^n - i}$ соответствующих ему значений атрибутов A_i (см. раздел 1) из \tilde{r}^n . Тогда, определив функцию $d_{a \text{ fnd}}$ как

$$d_{a \text{ fnd}}(r, W) = |\{x | x \in W, r_x = 1\}| \quad (15)$$

равную количеству единиц в подмножестве атрибутов W кортежа r . Получим ДМ обнаружения ошибочного атрибута:

$$\Delta d_{a \text{ fnd}} = d_{a \text{ fnd}}(\tilde{r}^n, A_{\tilde{r}^n - i}) - d_{a \text{ fnd}}(\hat{r}, A_{\tilde{r}^n - i}). \quad (16)$$

Также как и ранее сравнительно легко показывает, что для безошибочного атрибута A'_i значение функции $d_{a\text{ fnd}}(r, A'_{r-i})$ будет равным, а по косвенному признаку $\Delta d_{a\text{ fnd}}$ можно однозначно судить о ошибочности атрибута A'_i .

6 ОПРЕДЕЛЕНИЕ ВИДА ОШИБКИ

Результаты, полученные при решении предыдущей задачи, а также правила, по которым отношение $\tilde{R}_\Delta[k]$ определялось на бинарном домене $\{0,1\}$ (см. раздел 1), позволяют для каждого ошибочного атрибута из $\tilde{R}_\Delta[k]$ получить соответствующее значение \tilde{v} , которое было введено пользователем и значение \hat{v} которое, предположительно, должно было быть введенным. Как и в любой знаниеориентированной системе, прежде чем передать пользователю вывод об ошибочности атрибута, следует также сформировать объяснение полученного вывода [9]. Одной из составляющих такого объяснения и является вид ошибки, которая была допущена.

Здесь следует отметить, что в зависимости от предусмотренного в программном обеспечении способа ввода оператором конкретного значения, могут существенно отличаться виды возможных ошибок. Рассмотрим в данной работе наиболее распространенный случай, когда значение вводится полностью с клавиатуры, и может быть представлено как текстовая строка. В работе [2] представлены результаты исследований процесса ввода данных оператором, которые позволяют сформировать множество характерных для данного случая видов ошибочных данных (см. табл. 1).

В соответствии с работой [2] любой вид ошибки, можно определить как минимальную последовательность операций редактирования необходимых для преобразования строки \hat{v} в строку \tilde{v} . Следовательно можно решить задачу диагностирования, вычислив массив расстояний Левенштейна между строками (сложность $O(|\hat{v}|*|\tilde{v}|)$, где $|\hat{v}|, |\tilde{v}|$ – длина эталонной и введенной строки соответственно), и затем след наименьшей цены (сложность $O(|\hat{v}| + |\tilde{v}|)$), получив таким образом последовательность операций редактирования [14].

В то же время применение принципа диагностирования исходя из заданного множества наиболее часто

встречающихся, характерных видов ошибочных данных позволяет эффективнее решить такую задачу. Сформируем, исходя из определения каждого вида из табл. 1, прямые признаки наличия определенного вида отказа.

Однократные транскрипции:

$$S_{ts1} = \{(|\tilde{v}| = |\hat{v}|) \wedge ((\exists i \in [1, |\hat{v}|] \tilde{v}[i] \neq \hat{v}[i]) (\forall j \in [1, |\hat{v}|] i \neq j), \tilde{v}[j] = \hat{v}[j])\}, \quad (17)$$

где $v[n]$ – n -й символ строки.

Добавление символа:

$$S_{add} = \{(|\tilde{v}| - 1 = |\hat{v}|) \wedge (\exists i \in [1, |\tilde{v}|]) : ((\forall j \in [1, i - 1]), \tilde{v}[j] = \hat{v}[j]) \wedge ((\forall j \in [i, |\tilde{v}|]), \tilde{v}[j + 1] = \hat{v}[j])\}. \quad (18)$$

Пропуск символа:

$$S_{del} = \{(|\tilde{v}| + 1 = |\hat{v}|) \wedge (\exists i \in [1, |\hat{v}|]) : ((\forall j \in [1, i - 1]), \tilde{v}[j] = \hat{v}[j]) \wedge ((\forall j \in [i, |\tilde{v}|]), \tilde{v}[j] = \hat{v}[j + 1])\}. \quad (19)$$

Транспозиция смежных символов:

$$S_{tra} = \{(|\tilde{v}| = |\hat{v}|) \wedge ((\exists i \in [1, |\hat{v}|] \tilde{v}[i] \neq \tilde{v}[i + 1] \wedge \tilde{v}[i + 1] \neq \hat{v}[i] \wedge \tilde{v}[i] \neq \hat{v}[i + 1]) \times (\forall j \in [1, |\hat{v}|] (j = i) \wedge (j = i + 1)), \tilde{v}[j] = \hat{v}[j])\}. \quad (20)$$

Двукратная транскрипция:

$$S_{ts2} = \{(|\tilde{v}| = |\hat{v}|) \wedge ((\exists i \in [1, |\hat{v}|] \tilde{v}[i] \neq \hat{v}[i]) (\exists k \in [1, |\hat{v}|] k \neq j \tilde{v}[k] \neq \hat{v}[k]) \times (\forall j \in [1, |\hat{v}|] i \neq j), \tilde{v}[j] = \hat{v}[j])\}. \quad (21)$$

Совокупность косвенных признаков, также определим как множество двузначных предикатов:

$$Z_1 = \{|\tilde{v}| = |\hat{v}|\}, \quad (22)$$

$$Z_2 = \{Eq(\tilde{v}, \hat{v}) = 2\}, \quad (23)$$

где $Eq(v1, v2)$ – функция от двух строк $v1, v2$ равной длины, значение которой равно количеству позиций, в которых символы из $v1$ и $v2$ не равны.

Таблица 1 – Наиболее распространенные орфографические ошибки человека

Виды ошибок	Удельный вес	Примеры
Однократная транскрипция	До 76 %	245.75-> 295.75
Добавление символа	До 15 %	професор->профцесор
Пропуск символа	До 12 %	к. т. н.-> к.. н.
Транспозиция смежных символов	До 6 %	Пешков-> Пекшов
Двукратная транскрипция	До 3 %	731.25->832.25

$$Z_3 = \{(\exists i \in [1, |\hat{v}|] \tilde{v}[i] \neq \hat{v}[i+1] \wedge \tilde{v}[i+1] \neq \hat{v}[i] \wedge \tilde{v}[i] \neq \hat{v}[i+1]) (\forall j \in [1, |\hat{v}|] (j = i) \wedge (j = i+1)), \tilde{v}[j] = \hat{v}[j])\}; \quad (24)$$

$$Z_4 = \{Eq(\tilde{v}, \hat{v}) = 1\}; \quad (25)$$

$$Z_5 = \left\{ \min(|\tilde{v}|, |\hat{v}|, \text{Cmp}(\tilde{v}, \hat{v}) + \text{Cmp}(\tilde{\bar{v}}, \hat{\bar{v}})) = \frac{|\tilde{v}| + |\hat{v}| - 1}{2} \right\}, \quad (26)$$

где $\text{Cmp}(v_1, v_2)$ – функция от двух строк, значение которой равно $n - 1$, где n – номер первой позиции, в которой строки не равны; \bar{v} – строка, являющаяся зеркальным отображением строки v , т. е. последний символ v является первым в \bar{v} и т. д.

$$Z_6 = \{|\tilde{v}| - |\hat{v}| = 1\}. \quad (27)$$

Далее построим следующие ДМ для каждого вида ошибки, связывающие прямые признаки (17)–(21) с косвенными (22)–(27).

$$S_{ts1} = Z_1 \wedge \neg Z_2 \wedge Z_4; \quad (28)$$

$$S_{add} = \neg Z_1 \wedge Z_5 \wedge Z_6; \quad (29)$$

$$S_{del} = \neg Z_1 \wedge Z_5 \wedge \neg Z_6; \quad (30)$$

$$S_{tra} = Z_1 \wedge Z_2 \wedge Z_3; \quad (31)$$

$$S_{ts2} = Z_1 \wedge Z_2 \wedge \neg Z_3. \quad (32)$$

Так же, как и при решении других задач диагностирования, для каждой из построенной в (28)–(32) ДМ была оценена диагностируемость построенной модели. Доказательства диагностируемости каждой модели в данной работе не приводятся в связи с ограниченностью объема статьи, а также их тривиальностью.

Тогда процедура диагностирования на этапе определения вида ошибки, с использованием предложенных ДМ, может быть представлена как дихотомическое дерево (рис. 3).

Оценим алгоритмическую сложность предлагаемой процедуры. Сложность вычисления значения предикатов: $Z_1 - O(1)$, $Z_2 - O(|\tilde{v}|)$, $Z_3 - O(|\tilde{v}|)$, $Z_4 - O(|\tilde{v}|)$, $Z_5 - O(|\tilde{v}|)$, $Z_6 - O(1)$. Тогда в наихудшем случае сложность процедуры диагностирования в целом составит $O(2 * |\tilde{v}| + 1)$, что позволяет утверждать о превосходстве подхода с использованием ДМ относительно

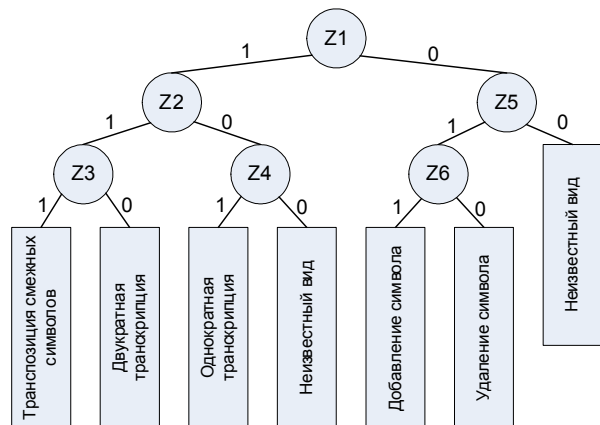


Рисунок 3 – Дихотомическое дерево определения вида ошибки

подхода предполагающего расчет следа наименьшей цены.

ЗАКЛЮЧЕНИЕ

Таким образом, на базе системного подхода к диагностированию объектов различной природы был сформирован новый подход к диагностированию данных на основе естественной избыточности. Основными положениями подхода стали принципы последовательного снятия неопределенности и диагностирования по косвенным признакам исходя из множества, характерных для данных вводимых пользователем ошибок. Для каждой задачи диагностирования были построены диагностические модели, позволяющая однозначно установить соответствующую ошибку. Кроме того, на каждом этапе аналитически показано превосходство предлагаемых решений относительно имеющихся альтернатив. Вместе с тем, остается открытым вопрос диагностируемости базы данных в целом, т. е. какова вероятность того, что пользователь допустит ошибку, обнаружение которой на основании имеющихся знаний о данных не представляется возможным.

ПЕРЕЧЕНЬ ССЫЛОК

1. Информационно-аналитические модели управления техническими высшими учебными заведениями / Гуржий А. Н., Кривцов В. С., Кулик А. С. и др. – Х.: Нац. аэрокосм. ун-т «Харьк. авиац. ин-т», 2004. – 387 с.
2. Литвинов В. А., Крамаренко В. В. Контроль достоверности и восстановление информации в человеко-машинных системах. – К.: Техніка, 1986. – 200 с.
3. W. W. Eckerson. Data Quality and the Bottom Line. Achieving Business through a Commitment to High Quality Data. – The Data Warehouse Institute, 2002. – 40 p.
4. Maletic, J., Marcus, A. Data Cleansing: Beyond Integrity Analysis // Proceedings of The Conference on Information Quality (IQ2000). – Massachusetts Institute of Technology, Boston, MA, USA. – 2000. – P. 200–209. Challenges in Comprehensive Data Cleansing. – Technical Report HUB-IB-164, Humboldt University, Berlin, 2003.

6. P. Oliveira, F. Rodrigues, P. Henriques, H. Galhardas. A Taxonomy of Data Quality Problems // In Proceedings of the 2nd International Workshop on Data and Information Quality (em conjunto com a conferência CAiSE'05), Porto, Portugal, Junho de 2005. – P. 219–233

7. J. Hipp, U. Guntzer, U. Grimmer. Data Quality Mining – Making a Virtue of Necessity // In Proc. of the 6th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD 2001), Santa Barbara, California, 2001. – P. 52–57.

8. Дейт, К., Дж. Введение в системы баз данных, 7-е издание. – М.: Издательский дом «Вильямс», 2001. – 1072 с.

9. Уэно Х., Исидзука М. Представление и использование знаний – М.: Мир, 1989. – 220 с.

10. Сироджа И. Б. Математическое и программное обеспечение интеллектуальных и компьютерных систем. – Х.: Харьковский авиационный институт, 1992. – 101 с.

11. Agrawal R., Imielinski T., Swami A. Mining. Association rules between Sets of Items in Large Databases // Proceedings of ACM SIGMOD International Conference on Management of Data, Washington D. C. – 1993. – P. 207–216.

12. Кулик А. С. Сигнально-параметрическое диагностирование систем управления. – Х.: Гос. аэрокосм. ун-т «ХАИ». Бизнес Информ, 2000. – 260 с.

13. Чухрай А. Г. Методы и средства повышения качества данных в автоматизированных системах организационного управления: Дис...канд. техн. наук: 05.13.06. – Х., 2004. – 173 с.

14. Graham A. Stephen String searching algorithms. – World scientific Publishing Co, 1994. – 243 p.

Надійшла 25.04.06
Після доробки 31.07.06

Представлено підхід до діагностування даних корпоративних інформаційних систем на основі природної надмірності, що базується на концепції діагностичних моделей. Послідовно вирішено задачі виявлення помилки, пошуку помилкових кортежів, помилкових атрибутів та визначення виду помилки. Приведено оцінки обчислювальної складності вирішення кожної задачі.

Approach to data diagnosis in enterprise information systems based on diagnostic models conception is presented. Systematic solution of error detection, erroneous tuples and attributes search, and kind of error determination tasks is proposed. Estimations of computational complexity of every diagnostic task are also presented.

УДК 681.3.013

А. И. Рязанцев, В. С. Кардашук

СИСТЕМА ЭКОЛОГИЧЕСКОГО МОНИТОРИНГА ОКРУЖАЮЩЕЙ СРЕДЫ

Предлагается архитектура и структура системы экологического мониторинга окружающей среды с использованием современных средств измерительной и вычислительной техники. Приведены функции, структура системы, ее состав, технические характеристики.

ВВЕДЕНИЕ

Рост промышленного производства, ввод в эксплуатацию новых предприятий неизбежно влечет за собой решение вопросов связанных с загрязнением окружающей среды. Особенно это актуально для регионов где сосредоточены предприятия нефтегазовой, химической, металлургической промышленности.

В настоящее время вопросы экологической безопасности, загрязнения окружающей среды приобрели особую актуальность в связи с возросшими требованиями как государственных так и международных экологических организаций.

Международные стандарты, распространяющиеся на управление окружающей средой, предназначены для обеспечения предприятий и организаций элементами эффективной системы управления окружающей средой, которые могут быть объединены с другими элементами административного управления с тем, чтобы со-

действовать в деле достижения экологических и экономических целей [1]. Мониторинг – это наблюдение за состоянием окружающей среды (атмосферы, гидросферы, почвенно-растительного покрова, а также техногенных систем) с целью ее контроля, прогноза и охраны (рис. 1).



Рисунок 1 – Модель системы управления окружающей средой согласно стандарту ISO-14001:2004

© Рязанцев А. И., Кардашук В. С., 2006

Мониторинг состояния воздуха ранее производился в ручном режиме, но с развитием микроэлектроники, вычислительной и измерительной техники все активнее используются современные средства, построенные на базе современных контроллеров и персональных компьютеров.

ПОСТАНОВКА ЗАДАЧИ

Для контроля состояния воздуха необходимо определить перечень вредных веществ, которые могут присутствовать в атмосфере данного региона, населенного пункта для выбора оптимального решения построения системы мониторинга, т. к. стоимость системы может составлять от десятка до сотен тысяч долларов.

Применительно к Луганскому региону, где сосредоточены флагманы нефтяной и химической промышленности в воздухе могут присутствовать Cl_2 , CO_2 , SO_2 , NH_3 , NO_2 , CO_2 и ряд других вредных для здоровья человека в большой концентрации химических веществ и соединений. Основная задача системы мониторинга – контроль за состоянием окружающего воздуха в пределах допустимых величин концентрации вышеперечисленных химических соединений. В основу концепции экологического мониторинга положен интегрированный подход, что означает создание систем выполняющих весь комплекс необходимых операций, включая:

- первичные измерения;
- сбор, передачу, накопление и обработку измерительных данных;
- анализ экологической ситуации;

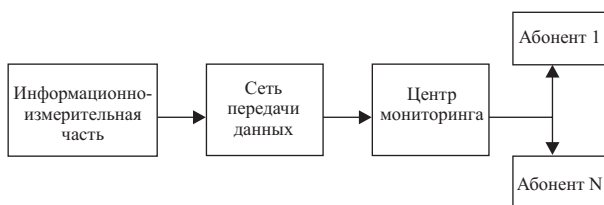


Рисунок 2 – Архитектура системы экологического мониторинга

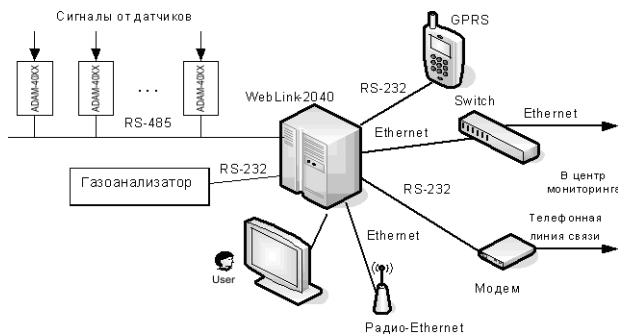


Рисунок 3 – Структурная схема измерительной части

- поддержку принятия решений по управлению экологической обстановкой;
- распределение результатов мониторинга между пользователями.

Следует отметить, что разрабатываемая система несет в себе только измерительно-информационные функции за состоянием атмосферы. С появлением новых технических средств и технологий построение систем экологического мониторинга значительно упрощается. Исчезают дополнительные преобразователи с интерфейса на интерфейс, расширяется номенклатура модулей УСО перекрывающая потребности заказчиков и т. д., улучшается достоверность и качество передаваемой информации на значительные расстояния за счет использования современных средств передачи данных. Что касается измерительных приборов, то большинство из них в настоящее время имеют в своем составе аналоговые выходы 4...20 мА (0...5 мА) и интерфейс RS-232, а некоторые фирмы включают в состав прибора интерфейс RS-485 или цифровой выход для подключения непосредственно к контроллерам напрямую, исключая дополнительные устройства.

АРХИТЕКТУРА СИСТЕМЫ

Архитектура системы экологического мониторинга представлена на рис. 2.

Информационно-измерительная часть представляет собой набор датчиков, газоанализатора и контроллера для съема информации о состоянии воздуха в контролируемой точке [2]. Информационно-измерительная сеть, объединяющая точки контроля, может быть сосредоточена в пределах одной территории (предприятие, завод и т. д.), так и распределена на значительном удалении друг от друга (город, район и т. д.).

Сеть передачи данных обеспечивает передачу измерительной информации в центр мониторинга для дальнейшей обработки результатов измерений. Сеть передачи данных может быть построена как с использованием проводных, так и беспроводных устройств связи.

Центр мониторинга представляет собой объединенные в локальную сеть Ethernet компьютеры, выполняющие функции приема, обработки, отображения, накопления и распределения полученной информации среди абонентов, решающих вопросы контроля экологической обстановки в реальном масштабе времени. На основе центра мониторинга может быть создан Web-узел для пользователей сети Internet.

СТРУКТУРА СИСТЕМЫ

Измерительная часть (рис. 3) включает в себя набор датчиков и газоанализатора, подключенных к контроллеру для ввода аналоговых сигналов 4...20 мА

(0...5 mA). Если измерительный прибор не имеет в своем составе аналогового выхода, то он при наличии порта RS-232 подключается к последовательному порту контроллера. Следует отметить, что двух последовательных портов контроллера при большом количестве газоанализаторов или измерительных приборов недостаточно. В этом случае необходимо использовать многоканальные блоки ввода-вывода, имеющие в своем составе буферы FIFO. Количество многоканальных блоков ввода-вывода зависит от числа газоанализаторов или измерительных приборов. Многоканальные блоки позволяют сократить время приема информации от газоанализаторов за счет уменьшения количества прерываний центрального процессора и буфера FIFO. Основу таких мнoportовых блоков ввода-вывода составляют универсальные асинхронные приемники-передатчики фирмы Texas Instruments.

Исследование функциональных характеристик контроллеров (Уникон [3], MicroPC, Advantech) показали, что по соотношения цена/функциональность предпочтительно использовать в данной системе Web-сервер WEBLINK-2040 фирмы Advantech.

Web-сервер включает в себя:

- процессор Intel 80486 – 66 МГц;
- контроллер VGA;
- Compact Flash;
- 12 (6+6) каналов дискретного ввода-вывода;
- 3 последовательных порта: 2 – RS-232, 1- RS-485;
- порты: 1 – IDE, 1 – FDD, 1 – LPT;
- предустановленную ОС Windows CE и программное обеспечение Scada-сервера Advantech Studio.

Подключение датчиков и измерительных приборов с аналоговым выходом 4...20 mA к Web-серверу WEBLINK-2040 осуществляется через модули ADAM-4000 фирмы Advantech. Такой вариант значительно расширяет диапазон подключаемых приборов и датчиков (температура, влажность, направление и скорость ветра и т. д.).

Модули серии ADAM-4000 представляют собой распределенные компактные интеллектуальные устройства обработки сигналов датчиков разработанные для применения в промышленности. Наличие встроенных микропроцессоров позволяет им осуществлять нормализацию сигналов, операции аналогового и дискретного ввода-вывода и их передачу по интерфейсу RS-485. Все модули серии ADAM-4000 имеют гальваническую развязку по цепям питания и интерфейса RS-485, программную установку параметров, командный протокол ASCII и сторожевой таймер. Диапазон рабочих температур модулей серии ADAM-4000 от –10 до +70 °С. Структурная схема измерительной системы приведена на рис. 3. Длина линии связи при подключении 32 абонентов по интерфейсу RS-485 может составлять до 1,2 км, скорость передачи в этом случае составляет 93,75 кбит/с, а с использованием повторителей рас-

стояние передачи данных составляет 4,8 км при скорости передачи 93,75 кбит/с.

От датчиков или измерительных приборов токовые сигналы 4...20 mA (0...5 mA) поступают на модули серии ADAM-4000. Модули монтируются на крепежную рейку и объединяются в локальную сеть по интерфейсу RS-485. В зависимости от их количества выбирается источник питания. Питание модулей осуществляется нестабилизированным источником питания 10...30 В постоянного тока. Объединенные в локальную сеть модули серии ADAM-4000 подключаются по двухпроводной витой паре к порту RS-485 Web-сервера WEBLINK-2040. На концах линии связи интерфейса RS-485 устанавливаются нагрузочные резисторы номиналом 120 Ом для согласования линии связи и исключения отражения сигналов.

Дальнейшим развитием использования современных средств изменения в измерительных системах является использование «интеллектуальных» измерительных приборов, предназначенных для преобразования текущих концентраций газа в выходной сигнал 4...20 mA, в цифровой сигнал, передаваемый по интерфейсу RS-485 [3, 4].

Передача информации в центр мониторинга может осуществляться как по телефонным каналам связи (выделенным или коммутируемым) через модемы по интерфейсу RS-232, так и по локальной сети Ethernet со скоростью передачи 100 Мбит/с (при передаче на небольшие расстояния). Здесь необходимо учитывать, что максимальное расстояние при таком способе передачи без дополнительного оборудования (усилители, коммутаторы) не превышает 100 м.

При передаче на значительные расстояния в качестве альтернативного способа передачи информации по радиоканалам может быть использован радио-Ethernet (устройство, подключаемое к порту контроллера Ethernet и осуществляющее беспроводную передачу информации на расстояние до 10 км).

Подсистема связи с центром мониторинга может быть реализована по радиоканалу с помощью радиостанции M-120 фирмы Motorola и контроллера пакетной радиосвязи KPC-3 фирмы Kontronics, используя УКВ-диапазон 144–174 МГц, что обеспечивает устойчивую передачу данных на расстояние до 40 км. Обмен между центром мониторинга и удаленными измерительными комплексами ведется на одной частоте по протоколу AX.25. Контроллер пакетной радиосвязи подключается к последовательному порту RS-232.

Спутниковый Интернет и GPRS (General Packet Radio Service) – современные беспроводные технологии связи позволяют создать скоростной канал доступа в центр мониторинга и одновременно решить проблему подключения удаленных абонентов. Основным преимуществом GPRS является мобильность. Данный сервис обеспечивает реальную пропускную способность свыше

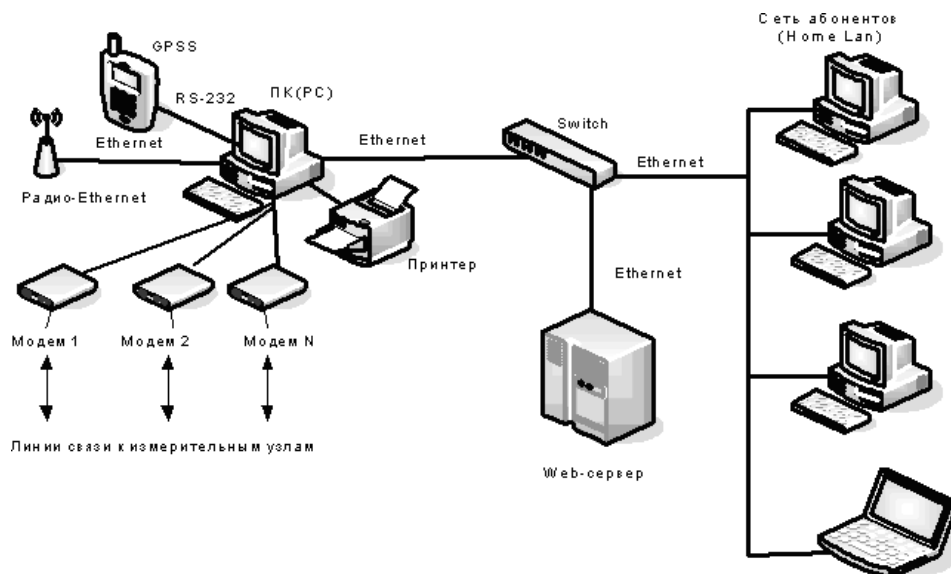


Рисунок 4 – Структурная схема центра мониторинга

48 кбит/с – это примерно те же скорости, что и в случае хороших наземных соединений через модем. Подключение осуществляется через порт RS-232.

В последнее время широко используются, так называемые, «домашние сети» (Home Lan), к которым подключаются абоненты через контроллеры Ethernet, находящиеся на значительном удалении друг от друга. Используя IP-адресацию, протокол TCP/IP и прикладное программное обеспечение абоненты получают доступ практически к любому сетевому ресурсу такой системы, в том числе и к Web-серверу WEBLINK-2040 напрямую.

Информационные функции центра мониторинга (рис. 4) обеспечиваются следующими подсистемами:

- сбора и первичной обработки информации о концентрациях вредных веществ в контролируемых точках;
- отображение информации;
- регистрации истории изменения концентрации вредных веществ в контролируемых точках.

Подсистема сбора и первичной обработки производит непрерывный циклический опрос во всех точках измерения и фиксирует текущее состояние загазованности.

Одновременно такую же информацию о состоянии конкретного узла измерения можно получить, подключив средство отображения (монитор, технологический пульт) к VGA-контроллеру Web-сервера WEBLINK-2040 непосредственно в точке контроля.

В качестве центрального узла системы используется персональный компьютер (ПК) или рабочая станция (РС) с выше перечисленными функциями. Выбор ПК (РС) зависит от выбора операционной системы (ОС) и количества выполняемых задач под управлением ОС. Следует отметить, что мощности современных ПК

вполне хватает для реализации поставленной задачи. Здесь основное внимание при выборе следует обратить на объем оперативной памяти (ОЗУ), т. к. при недостаточном объеме ОЗУ и увеличении количества задач выполняемых в системе увеличивается нагрузка на дисковую подсистему ПК, что в конечном итоге сказывается на производительности. Для увеличения пропускной способности запросов к дисковой подсистеме следует использовать ПК, имеющие в своем составе SATA-контроллер, как наиболее дешевая альтернатива дорогим SCSI-контроллерам.

Обычно в составе ПК имеется два последовательных COM-порта. Если рассматривать использование такого компьютера в офисном варианте, то такого количества COM-портов вполне достаточно, но если же на базе ПК организован коммуникационный сервер для подключения удаленных абонентов, то такого количества COM-портов явно недостаточно.

Наиболее простым способом решения проблемы увеличения COM-портов является использование многопортовых коммуникационных контроллеров (МКК). Количество портов с применением только одного такого МКК может возрасти до 8. МКК могут комплектоваться различными интерфейсами RS-232, RS-422, RS-485. Выбор интерфейса осуществляется переключением соответствующих перемычек на плате МКК (по умолчанию установлен интерфейс RS-232).

Если компьютер занимается обслуживанием сетевого обмена и количество требуемых портов невелико (4–8 портов), то здесь можно применить многопортовый коммуникационный контроллер PCI-1610A/1610B/1620A/1620B фирмы Advantech со следующими характеристиками:

- 4 порта RS-232 (8 портов для PCI-1610A) поддержка технологии plug-and-play;
- коммуникационная микросхема 16PCI954 – UART с буфером FIFO на 128 байт;
- скорость передачи до 921 кбит/с;
- защита от импульсных помех до 3000 В (1610В/1620В);
- программная поддержка в среде Windows.

Если же компьютер недостаточно производителен или на него возложены еще функции по защите intranet-сети от внешнего вмешательства, то в таком варианте лучше всего использовать интеллектуальную коммуникационную плату, которая сможет производить подготовку данных в соответствующем формате для обмена по системной шине. Примером такой интеллектуальной коммуникационной плат может служить PCI-контроллер фирмы Advantech PCL-747+/PCI-1630A со следующими характеристиками:

- от 8 до 32 портов RS-232/422;
- встроенный RISC-процессор TMS320C52;
- скорость обмена до 460 кбит/с;
- программная поддержка в среде Windows.

Наряду с фирмой Advantech аналогичные многопортовые коммуникационные контроллеры выпускают фирмы Cyclades, Digi, Моха.

Вспомогательные функции центра мониторинга обеспечивают подготовку системы к работе, настройку мнемосхем, начальный запуск и перезапуск системы в нестандартных ситуациях.

Подсистема отображения и связи с абонентами обеспечивает вызов мнемосхемы на экран ПК (PC), отображение на мнемосхемах текущих значений измерений, выдачу сообщений о превышениях предельно-допустимых концентраций (ПДК) в воздухе.

С помощью средств системы накопления истории процесса измерений можно отследить показания приборов за любой промежуток времени (прошедшие сутки, неделю, месяц).

ЗАКЛЮЧЕНИЕ

Научная ценность и новизна предложенной структуры состоит в том, что для ее построения использованы последние достижения измерительной, вычислительной

техники и средств коммуникаций. Система открыта для расширения и позволяет включать в свой состав дополнительные модули в зависимости от решаемых задач. Она может быть дополнена не только информационно-измерительными, но и управляющими функциями.

Практическая польза от предложенной структуры состоит в том, что она позволяет создавать гибкую систему мониторинга окружающей среды с минимальным набором технических средств.

Дальнейшие исследования в этой области продолжаются по следующим направлениям:

- исследование сети передачи данных на базе интерфейса RS-485;
- разработка программ обслуживания измерительных датчиков и газоанализаторного прибора;
- разработка программы связи центра мониторинга и измерительной части;
- разработка загрузочного модуля контроллера и программного обеспечения для связи абонентов.

ПЕРЕЧЕНЬ ССЫЛОК

1. ГОСТ Р ИСО 14001-98. Системы управления окружающей средой. Требования и руководство по применению; Введ. 21.10.98. – М., 1998. – С. 5.
2. Колтыгин С. И., Петрулевич А. А. Автоматизированные системы экологического контроля: интегрированный подход // Современные технологии автоматизации. – 1997. – № 1. – С. 28–32.
3. Якимчук П. С., Пономаренко В. А. Автоматизированная система экологического контроля производственных помещений // Экологические системы и приборы. – 2001. – № 3. – С. 3–7.
4. Болдуринов Б. А., Михайлов А. А. Газоаналитическая система СКВА-01 для контроля газовой среды во взрывоопасных зонах класса 1 // Промышленные АСУ и контроллеры. – 2004. – № 4. – С. 47–49.

Надійшла 21.12.05

Після доробки 10.05.06

Пропонується архітектура і структура системи екологічного моніторингу навколишнього середовища з використанням сучасних засобів вимірювальної і обчислювальної техніки. Приведені функції, структура системи, її склад, технічні характеристики.

It is proposed architecture and structure ecological monitoring system of environment with using of modern means of measuring and computer engineering techniques. Functions, system structure, its composition, technique characteristics were adduced in this work.

ТЕОРІЯ І МЕТОДИ АВТОМАТИЧНОГО УПРАВЛІННЯ

ТЕОРИЯ И МЕТОДЫ АВТОМАТИЧЕСКОГО УПРАВЛЕНИЯ

THEORY AND METHODS OF AUTOMATIC CONTROL

УДК 62 – 506

А. К. НОВИКОВ

СИНТЕЗ ОПТИМАЛЬНЫХ УПРАВЛЕНИЙ НА ОСНОВЕ АНАЛИТИЧЕСКОГО РЕШЕНИЯ ЛИНЕЙНОЙ СТАЦИОНАРНОЙ ДВУХТОЧЕЧНОЙ КРАЕВОЙ ЗАДАЧИ

Часть 1. Постановка и решение задачи. Базовый алгоритм

Рассматривается построение аналитического решения краевой задачи для линейных стационарных систем без ограничений на управляющие функции. В основу предлагаемого подхода положено, как и в классической теории управления, задание распределения корней управляемой системы. Решение задачи представлено в виде универсального (базового) алгоритма.

Не решай сложную задачу,
не решив простую (принцип простоты).

Первая заповедь математического моделирования [1]

ВВЕДЕНИЕ

Задачи, рассматриваемые в теории управления, делятся на два класса. К первому из них относятся задачи стабилизации движения относительно программных траекторий или заданных стационарных состоя-

ний, ко второму – задачи отыскания этих программных траекторий на основе решения краевых задач. Все успехи в развитии теории управления связаны, в основном, с решением задач первого класса. Что касается задач второго класса, то здесь достижения являются намного менее впечатляющими. Основные методы построения оптимальных программных траекторий (принцип максимума и метод динамического программирования) были сформулированы около пятидесяти лет назад. За это время с использованием указанных методов были решены многие важные частные задачи, однако какого-либо общего конструктивного подхода не найдено до сих пор [2, 3]. Как и полвека назад, непреодолимым препятствием на пути широкого применения метода динамического программирования лежит пресловутое «проклятие размерности». Высокая чувствительность

краевой задачи к погрешностям задания начальных значений сопряженных переменных в принципе максимума сильно затрудняет построение оптимальных траекторий методами «проб и ошибок» даже без учета ограничений на управляющие функции. Так как в любой реальной задаче указанные ограничения всегда присутствуют, подход с использованием принципа максимума оказывается здесь и вовсе бесперспективным. Все это дает основания говорить о наличии определенного кризиса в развитии науки об управлении [2].

По мнению автора, одна из гипотез о возможных причинах данного кризиса может быть сформулирована следующим образом. Множества методов решения задач каждого класса (и внутри классов) являются весьма специфичными и не имеют общих элементов. При этом в каждом множестве существуют задачи, для которых внутри самого этого множества методы решения отсутствуют. Положение, как говорится, патовое. Между тем, история науки знает немало примеров, когда новые идеи и даже новые области знаний возникали как раз на стыке наук. Опираясь на этот факт, в данной работе сделана попытка построения нового подхода к проблеме решения краевых задач, органически сочетающего в себе основные принципы классической теории управления, принципа максимума и метода динамического программирования. Основной особенностью данного подхода является аналитическое решение краевой задачи для специального класса линейных стационарных систем без ограничений на управляющие функции.

Предлагаемый в работе подход существенно опирается на возможность проведения символических преобразований в средах современных систем программирования MATLAB, Maple V и др. Для понимания работы требуются лишь минимальные знания в области матричного анализа, поэтому она вполне доступна широкому кругу инженеров, занимающихся разработкой систем управления.

1 ПОСТАНОВКА ЗАДАЧИ

Чем проще модель, тем реже она обманет (принцип надежности).

Четвертая заповедь математического моделирования

Рассматривается задача терминального управления в следующей максимально простой постановке. Пусть дана полностью управляемая линейная стационарная система

$$\begin{aligned} \dot{x} &= Fx + Gu; \quad x \in R^n, \quad u \in R^m, \\ F &\in R^{n \times n}, \quad G \in R^{n \times m}. \end{aligned} \tag{1}$$

Движение системы рассматривается на интервале времени $t \in [t_0, t_f]$. Начальный t_0 и конечный t_f моменты времени считаются заданными. Граничные условия для системы (1) задаются соотношениями

$$x(t_0) = c_0, \tag{2}$$

$$Mx(t_f) = c_f. \tag{3}$$

Постоянные векторы $c_0 \in R^n$, $c_f \in R^k$ и постоянная матрица $M \in R^{k \times n}$ ранга $0 < k \leq n$ считаются заданными. В качестве критерия оптимальности перевода системы (1) из состояния (2) в состояние (3) принимается классический функционал вида

$$J = \frac{1}{2} \int_{t_0}^{t_f} (x^T Qx + u^T Ru) dt, \tag{4}$$

где $Q \in R^{n \times n}$, $R \in R^{m \times m}$; $Q = \text{diag}(q_1, q_2, \dots, q_n) \geq 0$, $R > 0$. В соответствии с формализмом принципа максимума [4] составляется гамильтониан $H = \frac{1}{2}(x^T Qx + u^T Ru) + p^T (Fx + Gu)$, где $p \in R^n$ – сопряженный вектор, и записываются необходимые условия оптимальности

$$\dot{x} = \frac{\delta H}{\delta p} = Fx + Gu; \tag{5}$$

$$\dot{p} = -\frac{\delta H}{\delta x} = -F^T p - Qx; \tag{6}$$

$$0 = \frac{\delta H}{\delta u} = Ru + G^T p. \tag{7}$$

Уравнения (5)–(7) являются уравнениями краевой задачи. Граничные условия для нее задаются соотношениями (2), (3) и выражениями

$$p(t_0) = \xi_0; \tag{8}$$

$$p(t_f) = M^T \xi_f, \tag{9}$$

вытекающими из условий трансверсальности. В этих выражениях $\xi_0 \in R^n$ и $\xi_f \in R^k$ – неопределенные множители Лагранжа, которые должны выбираться так, чтобы выполнялись условия (2), (3).

Задача формулируется следующим образом: построить аналитическое решение полученной линейной стационарной двухточечной краевой задачи.

2 РЕШЕНИЕ ЗАДАЧИ

Из уравнения (7) найдем выражение для оптимального управления

$$u = -R^{-1}G^T p, \quad (10)$$

подставим его в уравнение (5) и запишем полученное соотношение и уравнение (6) в виде следующей расширенной системы:

$$\dot{z} = Az, \quad (11)$$

где

$$A = \begin{bmatrix} F & -GR^{-1}G^T \\ -Q & -F^T \end{bmatrix}; \quad z = \begin{bmatrix} x \\ p \end{bmatrix}. \quad (12)$$

Пусть $t_1, t_2 \in [t_0, t_f]$ – два любых момента времени и $t = t_2 - t_1$ есть соответствующий им интервал. Тогда решение однородной линейной стационарной системы (11) может быть представлено в виде [5]

$$z(t_2) = \Phi(t_2, t_1)z(t_1), \quad (13)$$

где $\Phi(t_2, t_1) = \Phi(t_2 - t_1) = \Phi(t) = e^{At}$ – переходная матрица системы. Построение переходной матрицы является первым и наиболее трудным этапом решения краевой задачи. Второй этап, собственно решение этой задачи с учетом краевых условий (2), (3) и (8), (9), особых трудностей не представляет. Рассмотрим оба эти этапа последовательно.

Построение переходной матрицы. Точное выражение для переходной матрицы стационарной системы (11) дается формулой [5]

$$e^{At} = L^{-1}[(sI - A)^{-1}], \quad (14)$$

где символом $L^{-1}[\cdot]$ обозначено обратное преобразование Лапласа. При практическом применении этого преобразования основной трудностью является разложение знаменателя резольвенты $(sI - A)^{-1}$ матрицы A на элементарные множители, то есть решение полной проблемы собственных значений для матриц произвольного вида. Гамильтонова матрица $A \in R^{2n \times 2n}$ имеет $2n$ собственных значений (с учетом их кратности), из которых всегда ровно половина имеют отрицательные вещественные части, а вторая половина – положительные, причем чисто мнимых собственных значений нет [6]. Следовательно, ее определитель может быть записан следующим образом:

$$\det(A) = (s^2 - \lambda_1^2)(s^2 - \lambda_2^2) \dots (s^2 - \lambda_n^2), \quad (15)$$

или в развернутом виде

$$\det(A) = (s^2)^n - S_2(\lambda_1^2, \dots, \lambda_n^2)(s^2)^{n-1} + S_4(\lambda_1^2, \dots, \lambda_n^2)(s^2)^{n-2} \dots \pm S_{2n}(\lambda_1^2, \dots, \lambda_n^2), \quad (16)$$

где $S_{2k}(\lambda_1^2, \dots, \lambda_n^2)$, $k = 1, \dots, n$ – элементарные симметрические функции, являющиеся суммами всех C_n^k k -членных произведений различных собственных значений, выбираемых среди $\lambda_1^2, \dots, \lambda_n^2$ [7]. С другой стороны, в матрице A для каждого $k = 1, \dots, 2n$ существует C_{2n}^k различных главных миноров порядка k , суммы которых обозначаются через $E_k(A)$ [7]. Напомним, что главным минором порядка k называется определитель подматрицы, расположенной на пересечении k строк и столбцов с одинаковыми множествами номеров. Вследствие гамильтоновости матрицы A все суммы $E_k(A)$ с нечетными индексами оказываются равными нулю, так что ее определитель, выраженный через указанные суммы, принимает вид

$$\det(A) = (s^2)^n - E_2(A)(s^2)^{n-1} + E_4(A)(s^2)^{n-2} \dots \pm E_{2n}(A). \quad (17)$$

Сравнивая в уравнениях (16) и (17) выражения при одинаковых степенях s^2 , можем записать следующие соотношения:

$$S_{2k}(\lambda_1^2, \dots, \lambda_n^2) = E_{2k}(A), \quad k = 1, \dots, n, \quad (18)$$

существование которых свидетельствует о тесной взаимосвязи между элементарными симметрическими функциями собственных значений матрицы A и суммами ее главных миноров. Считая соотношения (18) алгебраическими уравнениями относительно неизвестных $\lambda_1, \dots, \lambda_n$ с правыми частями, зависящими от элементов матрицы A , мы приходим к одной из труднейших задач численного анализа – классической задаче на собственные значения в ее традиционной постановке. Имеется много подходов к решению этой задачи, однако ни один из них не является абсолютно надежным и универсальным. Вычисление собственных значений непосредственно из уравнения $\det(sI - A) = 0$ «...является одной из хорошо известных глупостей в численном анализе» [8]. Приведение матриц к диагональному виду не всегда возможно, так как далеко не всякая матрица является диагонализуемой. Процесс приведения матрицы к форме Жордана является и технически сложным, и исключительно неустойчивым. Малейшие изменения в матрице могут приводить к большим изменениям элементов жордановой матрицы и даже ее

структуры. Обеспечение численной устойчивости этого процесса – дело безнадежное [7]. Даже применение одного из самых эффективных алгоритмов вычислительной математики – QR-алгоритма приведения матриц к форме Шура – далеко не всегда обеспечивает требуемую точность вычисления собственных значений и собственных векторов матриц общего вида, особенно при наличии у них кратных или близких по модулю собственных значений [6, 7].

Таким образом, в настоящее время не существует надежных способов преодоления самой первой и, по-видимому, самой главной трудности в решении краевой задачи – трудности решения полной проблемы собственных значений. Единственный корректный выход из сложившегося положения – вообще не решать данную проблему, а считать, как и в классической теории управления, все собственные значения заданными. Такое простое, на первый взгляд, решение меняет ситуацию самым радикальным образом. Все то, что для традиционного подхода было неразрешимой проблемой, с принятием нового подхода осуществляется легко и просто. Появляется возможность, например, осуществлять точное вычисление матриц преобразования к диагональной или жордановой форме, а также приведение матрицы A к форме Шура. Причем, во многих случаях необходимые для этого вычисления могут быть доведены до конца даже в символьном (!) виде. В этом и состоит основная идея рассматриваемого в работе подхода. Соотношения (18) при этом, разумеется, по-прежнему должны выполняться. Чтобы при новом подходе это стало возможным, в правых частях указанных соотношений должны существовать некоторые свободные параметры, выбором которых можно было бы обеспечить выполнение этих соотношений. В краевой задаче такие параметры существуют, и как раз в нужном количестве, – это суть n диагональных элементов q_1, \dots, q_n матрицы Q . Принимая это во внимание, преобразуем уравнения (18) к виду

$$E_{2k}(q_1, \dots, q_n) = S_{2k}(\lambda_1^2, \dots, \lambda_n^2), \quad (19)$$

где неизвестными являются теперь не параметры $\lambda_1, \dots, \lambda_n$, а параметры q_1, \dots, q_n . Эти параметры входят в левые части уравнений (19) в виде линейных комбинаций, так называемых, полилинейных функций, т. е. p -членных произведений различных величин, выбираемых среди q_1, \dots, q_n . Здесь $p = 1, \dots, m$, где m – размерность вектора управления. В общем случае, уравнения (19), будучи записанными в форме

$$f_k(q) = E_{2k}(q_1, \dots, q_n) - S_{2k}(\lambda_1^2, \dots, \lambda_n^2) = 0, \quad k = 1, \dots, n, \quad (20)$$

легко решаются относительно компонент вектора $q = (q_1, \dots, q_n)^T$ любым численным методом (например, итерационным методом Ньютона). Вычисление требующегося в методе Ньютона якобиана $\delta f / \delta q^T$ облегчается тем, что выражения для $f(q)$ могут быть получены в символьном виде. Последующее дифференцирование этих функций также легко осуществляется в символьной среде MATLAB. В частном случае, когда управление в задаче является скалярным ($m = 1$), уравнения (20) оказываются линейными, и их решение легко может быть получено в символьном виде.

Процесс построения переходной матрицы может осуществляться несколькими способами. Можно, например, представить детерминант матрицы A в форме (15) и выражения для элементов переходной матрицы получать непосредственно из уравнения (14). Символьное представление резольвенты $(sI - A)^{-1}$ матрицы A в этом случае легко вычисляется с использованием алгоритма Леверье [5]. Кстати, получаемые в ходе реализации этого рекуррентного алгоритма выражения для коэффициентов a_{2k} разложения детерминанта матрицы A по степеням s^2 , совпадают с выражениями для коэффициентов E_{2k} . Таким образом, применение алгоритма Леверье можно рассматривать в качестве еще одного способа формирования уравнений (20). Подобное непосредственное вычисление элементов переходной матрицы является, пожалуй, единственным способом, позволяющим, как говорится, «воочию» увидеть выражения для этих элементов в символьном виде. Однако на этом достоинства этого способа и заканчиваются. Указанные выражения в большинстве случаев оказываются такими невообразимо громоздкими, что никакого практического значения этот способ не имеет и применять его, вообще говоря, не следует (за исключением, может быть, самых простых случаев).

Гораздо более конструктивным является другой подход к построению переходной матрицы, основанный на методах теории подобных преобразований. Суть этого подхода состоит в представлении матрицы A в следующем факторизованном виде [5]:

$$A = TYT^{-1}, \quad (21)$$

где T – матрица подобного преобразования (трансформирующая матрица), Y – одна из матричных канонических форм. Выражение (14) для переходной матрицы принимает при этом следующий вид:

$$e^{At} = TL^{-1}[(sI - Y)^{-1}]T^{-1}. \quad (22)$$

В зависимости от выбранного распределения корней и способа приведения матрицы A к виду (21), матрица Y может оказаться диагональной, жордановой матрицей

или треугольной формой Шура. Смысл введения представления (22) заключается в том, что, по сравнению с матрицей A , матрица Y имеет более простую структуру, благодаря чему обратное преобразование Лапласа в формуле (22) легко вычисляется. При этом выражение для переходной матрицы принимает вид

$$\Phi(t) = T e^{Yt} T^{-1}. \quad (23)$$

Если в выбранном распределении $\lambda_1, \dots, \lambda_n$ нет кратных корней, символьные представления матриц T и T^{-1} формируются в соответствии с алгоритмом [5]. Столбцами матрицы T являются $2n$ правых собственных векторов матрицы A , строки матрицы T^{-1} формируются из соответствующих левых собственных векторов матрицы A . Поскольку собственные векторы всегда определяются неоднозначно, для полученных таким способом матриц вместо условия $T^{-1}T = I$ выполняется обычно лишь условие $T^{-1}T = D$, где D – диагональная матрица. Поэтому на этапе численных расчетов вместо матриц T и T^{-1} , вычисленных по их символьным представлениям, должны использоваться матрицы $\hat{T} = TD^{-\frac{1}{2}}$ и $\hat{T}^{-1} = D^{\frac{1}{2}}T^{-1}$. Если среди корней $\lambda_1, \dots, \lambda_n$ имеются кратные, то формируется только символьное представление матрицы T . Матрица T^{-1} вычисляется затем численным методом уже на этапе численных расчетов. При заданном распределении корней и неизменных параметрах управляемой системы матрицы T и T^{-1} являются постоянными. Таким образом, при вычислении переходной матрицы в соответствии с выражением (23) для каждого нового значения t заново приходится вычислять только матрицу простой структуры e^{Yt} .

Как всякая переходная матрица, матрица Φ , определяемая уравнением (23), является симплектической, то есть удовлетворяет условию [6]

$$\Phi^T J \Phi = J, \quad (24)$$

где $J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$ – матричный аналог мнимой единицы.

Указанное свойство значительно облегчает обращение матрицы Φ , а равенство (24) может служить хорошей проверкой правильности ее вычисления.

Хотя в данной статье рассматриваются, в основном, способы *аналитического* конструирования переходной матрицы, следует отметить, что существует также способ *численного* решения указанной задачи, основанный на использовании в качестве канонической формы Y треугольной формы Шура S [6, 7]. Вычисления в этом

способе реализуются следующим образом. Вначале для заданных значений $\lambda_1, \dots, \lambda_n$ и элементов матрицы R вычисляются значения элементов матрицы Q , обеспечивающие сохранение инвариантов подобия, и формируется (в численном виде) матрица A . Затем эта матрица приводится к верхней форме Хессенберга H . Наконец, с использованием QR -алгоритма (со сдвигами) выполняется приведение матрицы H к верхней форме Шура S . В данном случае все сдвиги известны точно – это суть заданные нами собственные значения $\lambda_1, \dots, \lambda_n, -\lambda_1, \dots, -\lambda_n$. Вследствие этого, теоретически бесконечный итерационный процесс приведения, характерный для стандартного QR -алгоритма, в рассматриваемом случае превращается в конечный – приведение к форме Шура осуществляется всего за $(2n - 1)$ шагов. Данная модификация QR -алгоритма обладает многими привлекательными свойствами. Прежде всего, этот алгоритм одинаково хорошо работает при любом распределении корней и любых значениях элементов матрицы R . Кроме того, сам алгоритм является идеально обусловленным, т. к. матрицы преобразования к форме Хессенберга и форме Шура – всегда ортогональны. Наконец, так как сдвиги могут задаваться сразу в требуемом порядке, исключается операция переформирования трансформирующих матриц и самой формы Шура, обычно присутствующая в стандартном QR -алгоритме. Недостаток у этого алгоритма всего один, но довольно существенный. При любом изменении параметров модели, значений корней распределения или элементов матрицы R вычисления с использованием этого довольно громоздкого алгоритма должны выполняться заново, начиная с вычисления новых значений элементов матрицы Q , формирования матрицы A и предварительного приведения ее к форме Хессенберга. В связи с этим, применение формы Шура в ряде случаев может оказаться затруднительным.

Решение краевой задачи. Разобьем переходную матрицу на блоки

$$\Phi(t_2, t_1) = \begin{bmatrix} \Phi_{11}(t_2, t_1) & \Phi_{12}(t_2, t_1) \\ \Phi_{21}(t_2, t_1) & \Phi_{22}(t_2, t_1) \end{bmatrix}$$

и запишем соотношение (13) в виде двух уравнений

$$\begin{aligned} x(t_2) &= \Phi_{11}(t_2, t_1)x(t_1) + \Phi_{12}(t_2, t_1)p(t_1); \\ p(t_2) &= \Phi_{21}(t_2, t_1)x(t_1) + \Phi_{22}(t_2, t_1)p(t_1). \end{aligned}$$

Эти уравнения справедливы при любых значениях $t_1, t_2 \in [t_0, t_f]$. В частности, полагая в первом из них $t_2 = t_0, t_1 = t_f$, можем записать

$$x(t_0) = \Phi_{11}(t_0, t_f)x(t_f) + \Phi_{12}(t_0, t_f)p(t_f). \quad (25)$$

Полагая во втором уравнении $t_2 = t_f, t_1 = t_0$, получим

$$p(t_f) = \Phi_{21}(t_f, t_0)x(t_0) + \Phi_{22}(t_f, t_0)p(t_0). \quad (26)$$

Для сокращения записи введем обозначения

$$\Phi(t_f, t_0) = U = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}; \quad (27)$$

$$\Phi(t_0, t_f) = V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}. \quad (28)$$

Симплектичность матриц U и V позволяет обойтись вычислением по уравнению (23) только одной (любой) из них. Если, например, вычислена матрица V , то обратная ей матрица U может быть получена из соотношений

$$U = V^{-1} = \begin{bmatrix} V_{22}^T & -V_{12}^T \\ -V_{21}^T & V_{11}^T \end{bmatrix}. \quad (29)$$

Выразим $x(t_f)$ из уравнения (25) и, с учетом граничных условий (2), (9), запишем

$$x(t_f) = V_{11}^{-1}(c_0 - V_{12}M^T\xi_f).$$

Умножим это уравнение на матрицу M слева

$$Mx(t_f) = MV_{11}^{-1}c_0 - MV_{11}^{-1}V_{12}M^T\xi_f \quad (30)$$

и введем обозначение

$$H = MV_{11}^{-1}V_{12}M^T.$$

Тогда из уравнения (30), с учетом граничного условия (3), получим

$$\xi_f = H^{-1}(MV_{11}^{-1}c_0 - c_f). \quad (31)$$

Перепишем уравнение (26), с учетом (8), (9), в виде

$$M^T\xi_f = U_{21}c_0 + U_{22}\xi_0.$$

Отсюда с использованием соотношений (28), (29) получим окончательное выражение для множителя Лагранжа $\xi_0 = U_{22}^{-1}(M^T\xi_f - U_{21}c_0)$ или, с учетом (27)–(29), (31),

$$\xi_0 = V_{11}^{-T}[(M^TH^{-1}MV_{11}^{-1} + V_{21}^T)c_0 - M^TH^{-1}c_f]. \quad (32)$$

Выражения для вычисления в момент $t \in [t_0, t_f]$ программных значений переменных краевой задачи принимают при этом следующий вид:

$$\begin{aligned} x^*(t) &= \Phi_{11}(t, t_0)c_0 + \Phi_{12}(t, t_0)\xi_0; \\ p^*(t) &= \Phi_{21}(t, t_0)c_0 + \Phi_{22}(t, t_0)\xi_0. \end{aligned} \quad (33)$$

Как правило, исходные физические уравнения управляемой системы являются нелинейными, их вид определяется соответствующими законами природы. Опыт успешных применений методов оптимизации к решению практических задач [10, 11, 12] показывает, что для решения задач оптимизации использовать эти уравнения в таком виде никогда не следует. Для постановки оптимизационной задачи на основе исходных уравнений всегда должна разрабатываться более простая линейная стационарная модель управляемой системы. Если управление $u(t)$ в этой линейной модели соответствует физическому управлению в исходной модели, то уравнение

$$u^*(t) = -R^{-1}G^T p^*(t) \quad (34)$$

представляет собой искомое аналитическое решение краевой задачи в форме программного управления. Если положить в приведенных выше формулах t_0 равным текущему моменту времени t , то решение краевой задачи, с учетом (32), может быть получено в форме закона управления с обратной связью

$$u(t) = -K(t)x(t) + K_f(t)c_f,$$

где $K(t) = R^{-1}G^T V_{11}^{-T}(M^TH^{-1}MV_{11}^{-1} + V_{21}^T)$, $K_f(t) = R^{-1}G^T V_{11}^{-T}M^TH^{-1}$.

Размерность линейной модели не обязательно должна совпадать с размерностью исходной системы. Если используется расширенная модель системы, управление $u(t)$ в которой не связано с физическим управлением в исходной задаче, решение задачи программного управления определяется соотношением

$$u^*(t) = x_k^*(t), \quad (35)$$

где k – номер координаты в линейной системе, связанной с физическим управлением в исходной задаче.

Для компенсации внешних возмущений и погрешностей модели суммарное управление в линейной системе должно формироваться в виде

$$u(t) = u^*(t) + \delta u(t), \quad (36)$$

где $\delta u(t)$ – некоторая поправка, обеспечивающая стабилизацию движения объекта относительно программной траектории $x^*(t)$. Вычисление величины этой

поправки может осуществляться как методами классической теории регулирования, так и с использованием базового алгоритма.

Анализ обусловленности задачи. Решение краевой задачи происходит в два этапа и на каждом из них существует опасность встретиться с плохо обусловленными вычислениями. Из сказанного выше, уже ясно, что для реализации вычислений на первом этапе – этапе построения переходной матрицы – не существует алгоритма, наилучшего во всех отношениях. Вычисления с использованием формы Шура всегда идеально обусловлены, и в этом смысле их применение является весьма желательным. Однако эти вычисления очень громоздки и во многих случаях попытки их реализации в адаптивных системах в реальном масштабе времени могут натолкнуться на серьезные трудности. Напротив, вычисления по аналитическим формулам очень просты и, в этом смысле, идеально вписываются в концепцию построения адаптивных систем. Однако, если не предпринимать никаких специальных мер, эти же вычисления оказываются, как правило, и чрезвычайно плохо обусловленными. Связано это с тем, что трансформирующие матрицы преобразования (21) формируются из собственных векторов, а плохая обусловленность вычислений с собственными векторами – дело обычное [6].

Как показывают результаты моделирования, характер процессов в задаче терминального управления слабо зависит от выбранного распределения корней и элементов матрицы R , однако сам процесс вычисления переходной матрицы при неудачном выборе этих параметров может оказаться чрезвычайно плохо обусловленным. Поэтому начинать нужно, прежде всего, с выбора таких значений указанных параметров, при которых как сама матрица A краевой задачи, так и трансформирующие матрицы преобразования (21) имели бы как можно меньшее число обусловленности (напомним, что число обусловленности – это тот коэффициент, на который умножаются ошибки в начальных данных при формировании конечного результата). Использование этого подхода позволяет в некоторых случаях добиться даже идеальной обусловленности. Например, если матрица A принадлежит к классу так называемых нормальных матриц, то есть, если для нее выполняется условие ($AA^T = A^T A$), то, независимо от наличия у нее кратных собственных значений, она всегда диагонализуема, причем диагонализация всегда осуществляется унитарным преобразованием [7]. Таким образом, нормальная матрица A – это тот идеал, к которому следует всячески стремиться в ходе решения краевой задачи. В некоторых случаях этого идеала действительно можно достичь путем рационального выбора распределения корней характеристического уравнения и элементов матрицы R . Например, если система (1) представляет собой n -мерный интегратор, наилучшим оказывается

выбор $R = 1$ и распределение корней по типу распределения Баттерворта [5]

$$\lambda_k = \operatorname{rexp}\left[\left(\frac{2k-1}{n} + 3\right)\frac{\pi i}{2}\right]; \quad \lambda_{n+k} = -\lambda_k, \quad k = 1, \dots, n \quad (37)$$

с радиусом круга распределения $r = 1$. При таком выборе все диагональные элементы матрицы Q , кроме q_1 , становятся равными нулю, q_1 оказывается равным 1, матрица A переходит в класс нормальных, а матрица преобразования T превращается в унитарную с числом обусловленности, равным 1.

Если бы мы выбрали другое распределение корней, например, биномиальное, число обусловленности матрицы T в рассматриваемом примере составило бы величину порядка $5 \cdot 10^8$. Это свидетельствует о том, что обусловленность вычислений в краевой задаче сильно зависит от распределения корней и элементов матрицы R . Поэтому вопросам выбора этих параметров необходимо уделять самое серьезное внимание и не жалеть для этого никаких усилий. Даже если всегда не удастся найти такое идеальное их сочетание, как в рассмотренном выше примере, всегда можно выбрать указанные параметры таким образом, чтобы существенно улучшить обусловленность вычислений по сравнению с их обусловленностью при каком-либо случайном выборе данных параметров.

Необходимо стремиться также к хорошей обусловленности еще одной матрицы, входящей в разложение (21), а именно, матрицы e^{Yt} . Ее сингулярное число обусловленности можно оценить величиной e^{2at} , где $a = \max[\operatorname{abs}(\operatorname{Re}(\lambda_k))]$, $k = 1, \dots, n$, (для диагональных матриц эта оценка является точной). Величина a должна выбираться в зависимости от величины интервала времени t таким образом, чтобы обеспечивалось согласование чисел обусловленности матрицы e^{Yt} и трансформирующей матрицы T .

Таким образом, требование наилучшей обусловленности вычислений накладывает довольно жесткие ограничения на свободу выбора распределения корней и элементов матрицы R . Можно сказать даже, что указанным требованием эти параметры определяются почти однозначно. Отсюда следует, что в рассматриваемом подходе матрицы R и Q более не являются свободными параметрами и им невозможно придать тот физический смысл, который обычно приписывается этим параметрам в классической теории оптимальных систем. Теперь R – это просто некоторая матрица коэффициентов, совместно с корнями характеристического уравнения обеспечивающая наилучшую численную обусловленность процесса построения переходной матрицы, а Q – зависящая от них матрица, отвечающая за

сохранение инвариантов подобия. Некоторая свобода выбора элементов матрицы R появляется лишь в тех случаях, когда имеющейся разрядности оказывается достаточно для проведения вычислений и можно особо не заботиться о проблеме обусловленности.

Подводя итог, можно сказать, что при использовании на первом этапе вычислений по аналитическим формулам, за счет выбора распределения корней и элементов матрицы R всегда можно (если не радикально, то, по крайней мере, существенно) улучшить обусловленность вычислений. В критических случаях можно попытаться использовать описанную выше модификацию QR-алгоритма. Другими словами, обусловленность вычислений на первом этапе является в той или иной мере управляемой.

Совсем иная ситуация складывается на втором этапе решения краевой задачи – этапе формирования значений адаптивных управлений. В классической и современной теории управления для обеспечения устойчивости движения в качестве полюсов замкнутой системы всегда выбираются только те n характеристических чисел, которые расположены в левой полуплоскости. Для рассматриваемого в данной статье подхода характерно, что в решении краевой задачи участвуют все $2n$ характеристических чисел матрицы A расширенной системы (11), половина из которых расположена в левой полуплоскости, а вторая половина является их зеркальным отражением относительно мнимой оси, т. е. расположена в правой полуплоскости. Это означает что как матрица A , так и связанная с ней краевая задача являются неустойчивыми. К тому же, в силу специфики самой задачи, в ходе ее решения приходится оперировать не с самой переходной матрицей (как единым объектом), а с ее блоками. Даже в том случае, когда выбором распределения корней и элементов матрицы R удастся добиться идеальной обусловленности самой переходной матрицы, ее внедиагональные блоки могут оказаться крайне плохо обусловленными. Это связано с тем, что характеристические числа матрицы A в выражения для вычисления элементов этих блоков всегда входят парами – наряду с числом λ_k используется и число с противоположным знаком $-\lambda_k$. Это приводит к тому, что выражения для вычисления элементов указанных матриц оказываются знакопеременными, вследствие чего итоговые значения этих элементов вычисляются всегда как малые разности близких (но отнюдь не малых) чисел.

Вот типичный пример подобной ситуации, когда система (1) является n -мерным интегратором ($n = 8$). Для этой системы при $t = 0,025$ с и распределении корней в соответствии с (37) число обусловленности переходной матрицы U в уравнении (27) равно 1,05. Если бы вычисления проводились с самой этой матрицей, столь близкое к единице число обусловленности, без сомнения, свидетельствовало бы об очень хо-

рошей обусловленности вычислений. Однако, как мы уже убедились, в решении краевой задачи участвует не сама матрица U , а ее блоки. И если диагональные блоки U_{11} и U_{22} обусловлены так же хорошо, как и сама матрица U , то обусловленность внедиагональных блоков в данном случае оказывается просто катастрофически плохой. Число обусловленности блока U_{12} , например, характеризуется величиной $1,25 \cdot 10^{30}$. Вдумайтесь только – это число более чем в миллиард миллиардов (!) раз больше числа обусловленности матрицы U . Причину столь плохой обусловленности можно понять, если детально рассмотреть процесс вычисления какого-либо элемента этой матрицы, например, элемента $U_{12}(1, 1)$. В выражениях для его вычисления положительные и отрицательные слагаемые чередуются самым регулярным образом. Вычисленная средствами MATLAB сумма положительных слагаемых равна

0,3203292768698897028728241250754498983474819975,

сумма отрицательных

-0,3203292768698897028728241250754798976352843760,

точное решение

0,0000000000000000000000000000000000000007121976215.

Легко видеть, что, если проводить вычисления с разрядностью, меньшей 36, в ответе не окажется вообще ни одного верного знака.

Во многих случаях ситуация оказывается не столь катастрофической, как в приведенном примере, однако общая тенденция всегда сохраняется. Поэтому нет ничего удивительного в том, что попытки решения краевых задач традиционными численными методами (да еще в рамках стандартной разрядности) очень часто оканчиваются неудачей. В таких случаях обычно отмечается, что задача обладает слишком высокой чувствительностью конечных значений фазовых переменных к погрешностям вычисления начальных значений сопряженных переменных [9]. Как видим, причина этой высокой чувствительности крайне проста и заключается она в недостаточной разрядности проводимых вычислений.

Для линейных краевых задач самым радикальным и единственным способом преодоления указанной трудности является проведение вычислений с повышенной разрядностью. Для нелинейных задач, по мнению автора, вряд ли помогут вычисления с любой, даже самой высокой, разрядностью, так как к этим задачам неприменим краеугольный принцип теории линейных систем – принцип суперпозиции.

Из приведенного примера видно, что требуемое повышение разрядности может иногда оказаться довольно значительным, особенно для систем высокого порядка. В этом смысле, образно можно сказать, что над методами оптимизации, основанными на принципе максимума, тяготеет некое «проклятие», которое, по аналогии со знаменитым беллмановским «проклятием размерности» в динамическом программировании, можно было бы назвать «проклятием разрядности». К счастью, кажется, преодолеть это второе «проклятие» намного легче, чем первое. Для этого достаточно разработать новые версии компиляторов, допускающих вычисления с повышенной разрядностью. Задача эта не представляется такой уж неразрешимой. По крайней мере, один успешный опыт такой разработки известен – созданный много лет назад компилятор с языка PL/1 для ЕС ЭВМ позволял проводить вычисления с разрядностью мантиссы, равной 33. Стандартной разрядностью повсеместно распространенных в настоящее время IBM PC является разрядность 16, и возможностей ее увеличения в рамках существующих систем программирования компилятивного типа, к сожалению, не существует.

Базовый алгоритм. Данный алгоритм не является вычислительным алгоритмом в обычном понимании этого слова. Наряду с чисто вычислительными этапами он содержит этапы принятия решений, которые не могут быть выполнены без участия человека. Кроме того, алгоритм самым существенным образом опирается на возможность проведения символьных преобразований, которые для получения достаточно простых конечных выражений необходимо проводить в интерактивном режиме, т. е. также с участием человека. Таким образом, можно сказать, что базовый алгоритм – это последовательность действий человека, проектирующего адаптивную систему для своего конкретного приложения и находящегося с этой целью в некоторой среде символьного программирования.

Алгоритм формулируется в виде следующей последовательности шагов.

1. На основе исходных (как правило, нелинейных нестационарных) уравнений, вытекающих из законов природы, строится линейная стационарная модель управляемого процесса. Для этого могут использоваться различные методы, среди которых особого внимания заслуживает подход, основанный на идее [10] о возможности точного представления уравнений исходной задачи в виде нескольких независимых 2-мерных интеграторов. Эта идея допускает естественное обобщение на случай интеграторов произвольного числа измерений. При этом вектор состояния может расширяться как «вверх» (введением интегралов – например, для придания системе свойств астатизма), так и «вниз» (введением высших производных). Подобное расширение во многих случаях оказывается весьма эффек-

тивным средством придания управляемому процессу новых свойств, которые невозможно получить при равенстве размеров исходной и оптимизационной моделей.

Часто в качестве оптимизационной модели может быть принято даже довольно грубое приближение к уравнениям исходной задачи. При этом параметры этой упрощенной модели необходимо выбирать таким образом, чтобы в ходе управления с обратной связью влияние сделанных допущений уменьшалось по мере приближения управляемого объекта к заданному терминальному состоянию c_f . Удачные примеры такого подхода содержатся в [11, 12].

Важным является также требование, чтобы оптимизационная задача с использованием построенной модели не содержала ограничений на управляющие функции и, тем более, на фазовые координаты. Только в этом случае крайняя задача, оставаясь линейной, может быть решена предлагаемым в данной работе способом. Если в исходной задаче ограничения на управления все-таки присутствуют, они могут быть учтены *a posteriori* в ходе управления с обратной связью.

2. Для построенной линейной модели в соответствии с выражением (12) формируется символьное представление матрицы A . При этом для параметров модели, относительно которых предполагается, что они могут изменяться, вводятся краткие символьные обозначения.

3. Принимается решение о характере распределения собственных значений $\lambda_1, \dots, \lambda_n$ матрицы A : являются ли все они различными или же среди них имеется определенное количество кратных собственных значений.

4. Для принятого характера распределения собственных значений определяется структура канонической формы Y , к которой преобразование подобия (21) приводится матрица A . Самый простой случай, когда все собственные значения – различны. Каноническая форма в этом случае является диагональной матрицей с элементами, равными этим собственным значениям, а матрицы преобразования T и T^{-1} состояются из правых и левых собственных векторов, соответствующих этим собственным значениям.

Если в выбранном распределении присутствуют кратные собственные значения, в число столбцов и строк матриц преобразования входят соответствующие этим значениям обобщенные собственные векторы. Канонической формой в этом случае является жорданова матрица, структура которой определяется числом и размерами жордановых блоков. Для того чтобы определить эти параметры, необходимо для каждого различного λ_p определить степени $(\lambda_p I - A)^k$, где $k = 0, 1, \dots, n$, проанализировать последовательность рангов этих степеней и по результатам анализа определить размеры и число жордановых клеток, отвечающих данному собственному значению [7].

5. В соответствии с алгоритмом [5] формируется символьное представление матрицы преобразования T . Ее столбцами являются $2n$ правых собственных векторов матрицы A (включая и обобщенные). Если в выбранном распределении нет кратных корней, символьное представление строк матрицы T^{-1} формируется из соответствующих левых собственных векторов матрицы A . Для этого указанный алгоритм применяется к матрице A^T .

6. С использованием обратного преобразования Лапласа вычисляется символьное представление матрицы $e^{Yt} = L^{-1}[(sI - Y)^{-1}]$.

7. Формируются уравнения для определения элементов q_1, \dots, q_n . Получить эти уравнения можно одним из следующих способов:

- приравниванием коэффициентов при одинаковых степенях s^2 в выражениях для детерминантов матриц $(sI - A)$ и $(sI - Y)$;

- с использованием инвариантов подобия E_{2k} в уравнении (20);

- с использованием коэффициентов a_{2k} , получаемых в ходе реализации рекуррентного алгоритма Леверье [5];

- путем вычисления $(2n \times 2n)$ – матрицы $M = TY - YT$ с последующим приравниванием нулю тех n из $2n$ ее ненулевых элементов, которые образуют систему независимых уравнений.

8. В случае, если полученные уравнения оказываются линейными относительно q_k , они решаются в символьном виде. В результате находятся символьные выражения для q_1, \dots, q_n . В противном случае в символьном виде формируются система нелинейных уравнений $f(q) = 0$, где $q = [q_1, \dots, q_n]^T$, и якобиан этой системы $\frac{\delta f}{\delta q^T}$. Символьные выражения для элементов якобиана находятся с использованием соответствующих символьных средств системы MATLAB.

9. В соответствии с выбранным распределением корней задаются их численные значения

$$\lambda_k = a_k + b_k i; \lambda_{n+k} = -\lambda_k; k = 1, \dots, n.$$

10. Задаются численные значения элементов матрицы R .

11. Вычисляются значения параметров q_1, \dots, q_n по формулам п. 8 или с использованием итерационного алгоритма Ньютона

$$q_{p+1} = q_p - \left[\frac{\delta f}{\delta q^T} \right]_{q=q_p}^{-1} f(q_p), p = 0, 1, \dots$$

В ходе вычислений необходимо контролировать знаки параметров q_k ; все полученные после окончания

итераций значения q_k должны быть неотрицательными. Добиться этого можно выбором распределения корней и/или значений элементов матрицы R . В необходимых случаях пункты 9–11 приходится повторить несколько раз.

12. По символьным представлениям трансформирующих матриц T и T^{-1} определяются их численные значения, находится диагональная матрица $D = T^{-1}T$,

и вычисляются матрицы $\hat{T} = TD^{-\frac{1}{2}}$ и $\hat{T}^{-1} = D^{-\frac{1}{2}}T^{-1}$,

удовлетворяющие условию $\hat{T}^{-1}\hat{T} = I$. Если среди корней распределения имеются кратные, матрица T^{-1} вычисляется как обратная матрице T . В этом случае указанная выше балансировка строк и столбцов не производится.

13. Задается величина интервала $t = t_0 - t_f$ и по символьному представлению матрицы e^{Yt} вычисляется ее численное значение.

14. По формуле (23) вычисляется переходная матрица $V = \Phi(t_0, t_f)$ и по формуле (24) осуществляется проверка правильности ее вычисления.

15. Задаются граничные условия (2), (3) краевой задачи и по формуле (32) вычисляется начальное значение ξ_0 сопряженного вектора p .

16. Для требуемого момента времени t по формулам (33) вычисляются значения параметров $x^*(t)$ и $p^*(t)$ программной траектории и по формулам (34) или (35) – значение программного управления $u^*(t)$.

17. С использованием соответствующих алгоритмов формируется величина поправки $\delta u(t)$ и по формуле (36) вычисляется управление $u(t)$ в линейной задаче.

18. При необходимости, по управлению $u(t)$ формируется физическое управление $\tau(t)$, подаваемое на исполнительные органы исходной системы.

Возможности данного алгоритма ограничиваются только возможностями существующих пакетов символьного программирования.

ВЫВОДЫ

В работе предложен конструктивный подход к решению проблемы синтеза адаптивных линейных стационарных систем без ограничений на фазовые переменные и управляющие функции. В его основу положена идея о замене традиционного подхода с априорным заданием матриц Q и R критерия качества новым подходом, в котором априорно задаются полюса замкнутой системы. Сохранение инвариантов подобия матрицы A краевой задачи при этом обеспечивается соответствующим выбором диагональных элементов матрицы Q критерия качества.

В классической теории управления, в основе которой лежит тот же принцип априорного выбора полюсов, основным критерием для подобного выбора является *устойчивость движения* управляемой системы. Методам, основанным на принципе максимума, внутренне присущи высокие требования к разрядности вычислений («проклятие разрядности»). Поэтому в предлагаемом подходе аналогичный критерий должен основываться на понятии *обусловленности вычислений*.

ПЕРЕЧЕНЬ ССЫЛОК

1. Андриевский Б. Р., Фрадков А. Л. Элементы математического моделирования в программных средах MATLAB и Scilab. – СПб.: Наука, 2001.
2. Красовский А. А. Науковедение и состояние теории процессов управления // АИТ. – 2000. – № 4. – С. 3–19.
3. Габасов Р., Дмитрук Н. М., Кириллова Ф. М. Оптимизация многомерных систем управления с параллелепипедными ограничениями. // АИТ. – 2002. – № 3. – С. 3–26.
4. Сейдж Э. П., Уайт Ч. С., III. Оптимальное управление системами: Пер. с англ. – М.: Радио и связь, 1982. – 392 с.
5. Квакернаак Х., Сиван Р. Линейные оптимальные системы управления. – М.: Наука, 1977. – 652 с.
6. Икрамов Х. Д. Численное решение матричных уравнений. – М.: Наука, 1984. – 192 с.
7. Хорн Р., Джонсон Ч. Матричный анализ. – М.: Мир, 1989. – 656 с.

8. Хайрер Э., Нерсетт С., Ваннер Г. Решение обыкновенных дифференциальных уравнений. – М.: Мир, 1990. – 512 с.
9. Токарев В. В. О знаках импульсов в задачах оптимального управления с закрепленными концами траектории. // АИТ. – 2001. – № 8. – С. 46–55.
10. Дуайер Т. А. У., III. Точное нелинейное управление быстрыми вращениями КЛА посредством внутренней передачи количества движения // Аэрокосмическая техника. – 1987. – № 3. – С. 151–159.
11. Синха С. К., Шривастава С. К. Оптимальный закон наведения многоступенчатой ракеты-носителя на трехмерной траектории, использующий выражения в явном виде // Аэрокосмическая техника. – 1991. – № 3. – С. 74–87.
12. Чендлер Д. С., Смит И. Е. Схема итеративного управления и ее использование для различных аппаратов и космических операций // Астронавтика и ракетодинамика. – 1967. – № 8. – С. 1–24.

Надійшла 15.07.06

Розглядається побудова аналітичного розв'язку крайової задачі для лінійних стаціонарних систем без обмежень на керуючі функції. В основу запропонованого підходу покладено, як і в класичній теорії керування, задання розподілу коренів керованої системи. Розв'язок задачі подано у вигляді універсального (базового) алгоритму.

It is consider the analytical decision of edge problem for linear stationary systems without restriction of control functions. For basis of the proposed approach it was taken, as in the classical control theory, the forming of control system roots distribution. The decision it is produced in the form of universal (base) algorithm.

УДК 62 – 506

А. К. НОВИКОВ

СИНТЕЗ ОПТИМАЛЬНЫХ УПРАВЛЕНИЙ НА ОСНОВЕ АНАЛИТИЧЕСКОГО РЕШЕНИЯ ЛИНЕЙНОЙ СТАЦИОНАРНОЙ ДВУХТОЧЕЧНОЙ КРАЕВОЙ ЗАДАЧИ

Часть 2. Примеры применения базового алгоритма

На конкретных примерах рассматриваются особенности применения базового алгоритма, построенного в первой части статьи. Для систем с ограничениями на управляющие функции используется принцип поэтапного планирования метода динамического программирования.

ВВЕДЕНИЕ

Издавна в научных кругах принято считать хорошей традицией проверку любого нового метода или алгоритма на известных примерах. Не отступая от этой традиции, автор включил в данный раздел несколько известных из литературы примеров, решенных ранее

с использованием других подходов. Эти примеры охватывают круг задач терминального управления, оптимального регулирования и наблюдения, управления с эталонной моделью, управления с ограничениями в виде неравенств, наложенными на управляющие функции.

Остальные примеры, взятые из близкой автору области управления в космосе, иллюстрируют возможности применения базового алгоритма к решению задач регулирования и стабилизации, наведения, а также обеспечения плавной стыковки участков траектории в случае, когда общая задача управления разбивается

на несколько элементарных подзадач. Внимания заслуживает также пример использования базового алгоритма для решения такой нетрадиционной для теории оптимального управления задачи, как высокоточное численное интегрирование уравнений небесной механики.

Автор надеется, что приведенные примеры помогут читателям найти свои подходы к решению интересных их задач в других областях применения.

ПРИМЕРЫ

Пример 1. Рассматривается задача синтеза алгоритмов оптимального регулятора с обратной связью по выходной переменной $y = Cx$. Как известно, такой регулятор является комбинацией наблюдателя, в котором восстанавливается состояние системы, и закона управления, представляющего собой линейную функцию этого восстановленного состояния [2].

Решение задачи оптимального *регулирования* для системы $\dot{x} = Fx + Gu$ и критерия качества

$$J_R = \frac{1}{2}x^T(t_f)P_f x(t_f) + \frac{1}{2} \int_{t_0}^{t_f} (x^T Q x + u^T R u) dt$$

имеет вид закона управления с обратной связью $u(t) = -K(t)x(t)$, где $K(t) = R^{-1}G^T P(t)$, $P(t)$ – решение уравнения Риккати [2]

$$-\dot{P}(t) = F^T P(t) + P(t)F + Q - P(t)BR^{-1}B^T P(t), \\ P(t_f) = P_f, t \leq t_f.$$

В явном виде это уравнение решается в исключительно редких случаях, применение численных методов также является не вполне корректным, т. к. процесс численного интегрирования данного уравнения в прямом направлении является неустойчивым. Даже нахождение его установившегося решения наталкивается на серьезные трудности, поскольку эта задача тесно связана с решением уже упоминавшейся проблемы собственных значений.

Применение базового алгоритма позволяет обойти все эти трудности. Вычислив с использованием этого алгоритма переходную матрицу $\Phi(t, t_f)$, для регулятора с переменной настройкой можем записать [2]

$$P(t) = [\Phi_{21}(t, t_f) + \Phi_{22}(t, t_f)P_f][\Phi_{11}(t, t_f) + \Phi_{12}(t, t_f)P_f]^{-1}, t \leq t_f.$$

Матрица коэффициентов закона управления для этого случая имеет вид

$$K_R(t) = R^{-1}G^T P(t).$$

Установившееся решение уравнения Риккати находится из выражения $\bar{P} = T_{22}T_{12}^{-1}$, где T_{12} и T_{22} – блоки матрицы преобразования к канонической форме

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}. \text{ При этом матрица коэффициентов усиления оптимального регулятора с постоянной настройкой определяется соотношением } \bar{K}_R = R^{-1}G^T \bar{P}.$$

В тех случаях, когда возможностей вычислительных средств оказывается достаточно для символьного обращения матрицы T_{12} , для элементов матрицы \bar{K}_R могут быть получены даже символьные выражения, часто помогающие глубже понять суть рассматриваемой задачи. Например, для управляемой системы [2]

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ 0 & -a \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ k \end{bmatrix} \mu(t), \quad (1)$$

элементы матрицы Q , отвечающие за сохранение инвариантов подобия, определяются выражениями

$$q_1 = (\lambda_1^2 \lambda_2^2 r_1 / k^2); q_2 = (\lambda_1^2 + \lambda_2^2 - a^2) r_1 / k^2; \quad (2)$$

где r_1 – элемент скалярной матрицы R , а выражения для элементов матрицы коэффициентов усиления $\bar{K}_R = [k_1 \ k_2]$ оптимального регулятора с постоянной настройкой имеют вид

$$k_1 = \lambda_1 \lambda_2 / k; k_2 = (\lambda_1 + \lambda_2 - a) / k. \quad (3)$$

Анализ выражений (3) показывает, что коэффициенты регулятора не зависят от элементов матриц Q и R , а определяются только параметрами управляемой системы и выбранным распределением корней (ср. с [2]). На основании формул (2),(3) могут быть сделаны также обоснованные выводы о параметрах этого распределения. Ясно, например, что для значений λ_1, λ_2 должны выполняться условия

$$\lambda_1 + \lambda_2 > a; \lambda_1^2 + \lambda_2^2 > a^2. \quad (4)$$

С учетом того, что $\lambda_1 = a + bi, \lambda_2 = a - bi$ условия (4) преобразуются к виду $a > \sqrt{a^2/2 + b^2}$.

Решение задачи синтеза алгоритмов оптимального наблюдателя может быть получено на основе свойства дуальности задач оптимального регулирования и оптимального наблюдения. Как известно [2], задачу оптимального наблюдения можно представить в виде задачи оптимального регулирования для системы

$$\dot{x} = -F^T x - C^T u$$

с критерием качества

$$J_0 = \frac{1}{2}x^T(t_0)S_0x(t_0) + \frac{1}{2}\int_{t_f}^{t_0}(x^TQx + u^TRu)dt.$$

Матрица расширенной системы (12) для данной задачи имеет вид

$$A = \begin{bmatrix} -F^T & C^TR^{-1}C \\ Q & F \end{bmatrix}.$$

В соответствии с базовым алгоритмом для нее может быть вычислена переходная матрица $\Phi(t, t_0)$. С использованием этой матрицы решение матричного уравнения Риккати в задаче наблюдения

$$\begin{aligned} \dot{S}(t) &= FS(t) + S(t)F^T + Q - \\ &- S(t)C^TR^{-1}CS(t), S(t_0) = S_0 \end{aligned}$$

записывается в виде

$$\begin{aligned} S(t) &= [\Phi_{21}(t, t_0) + \Phi_{22}(t, t_0)S_0] \times \\ &\times [\Phi_{11}(t, t_0) + \Phi_{12}(t, t_0)S_0]^{-1}, t \geq t_0. \end{aligned}$$

Для нахождения установившегося решения этого уравнения используется соотношение

$$\bar{S} = T_{22}T_{12}^{-1}.$$

Оптимальные коэффициенты наблюдателя определяются при этом выражением $K_0(t) = S(t)C^TR^{-1}$, $t \geq t_0$ или $\bar{K}_0 = \bar{S}C^TR^{-1}$.

Для рассматриваемого примера (1) элементы матрицы Q в задаче наблюдения определяются выражениями

$$q_1 = (\lambda_1^2 + \lambda_2^2 - a^2)r_1; \quad q_2 = (a^4 - (\lambda_1^2 + \lambda_2^2)a^2 + \lambda_1^2\lambda_2^2)r_1.$$

Выражения для коэффициентов $\bar{K}_0 = [k_1 \ k_2]^T$ оптимального наблюдателя с постоянной настройкой имеют вид

$$k_1 = (\lambda_1 + \lambda_2) - a; \quad k_2 = a^2 - (\lambda_1 + \lambda_2)a + \lambda_1\lambda_2.$$

Отметим, что эти коэффициенты также не зависят от матриц Q и R (ср. с [2]). На основании анализа полученных выражений могут быть также сформулированы условия вида (4), которым должны удовлетворять параметры распределения корней наблюдателя.

Комбинируя полученные для регулятора и наблюдателя результаты, можем для системы (1) записать уравнения *оптимального регулятора* с обратной связью по измеренным значениям выходной переменной $\tilde{y} = Cx + v$ [2]

$$\begin{aligned} \dot{\hat{x}} &= (F - GK_R - K_0C)\hat{x} + K_0\tilde{y}, \hat{x}(t_0) = \hat{x}_0, \\ u &= -K_R\hat{x}. \end{aligned}$$

Пример 2. Решается задача оптимального управления движением мостового крана. Уравнения движения, все обозначения и величины параметров приняты в соответствии с работой [6]. Дана линейная стационарная система вида $\dot{x} = Fx + Gu$, где

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -a & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -c & 0 & 0 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 0 \\ e \\ 0 \\ e \end{bmatrix}, \quad e = b/u_m.$$

Граничные условия (2), (3) в рассматриваемом примере имеют вид

$$c_0 = [0, 0, 0, 0]^T, \quad c_f = [0, 0, s_f/l, 0]^T, \quad M = I.$$

Ограничение, наложенное на управление u , учитывается в данном случае путем выбора такой величины t_f , при которой условие $|u| \leq 1$ заведомо выполняется на всем интервале управления. Другие способы учета ограничений рассматриваются в следующих примерах.

Продемонстрируем на этом простом примере все этапы решения краевой задачи. Прежде всего, выберем в качестве корней характеристического полинома замкнутой системы корни распределения Баттерворта (37) для $n = 4$ и представим каноническую форму матрицы A в виде диагональной символьной матрицы $D = \text{diag}(\lambda_1, \dots, \lambda_4, -\lambda_1, \dots, -\lambda_4)$. Уравнения для определения параметров q_1, \dots, q_4 в данном случае имеют вид

$$(2r_1a - e^2(q_2 + q_4)) = -(\lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \lambda_4^2)r_1;$$

$$\begin{aligned} &(q_1 + q_3 - 2vq_4)e^2 + a^2r_1 = \\ &= (\lambda_2^2\lambda_4^2 + \lambda_3^2\lambda_4^2 + \lambda_1^2\lambda_2^2 + \lambda_2^2\lambda_3^2 + \lambda_1^2\lambda_3^2 + \lambda_1^2\lambda_4^2)r_1; \end{aligned}$$

$$(2q_3 - vq_4)e^2 = -(\lambda_1^2\lambda_3^2\lambda_4^2 + \lambda_1^2\lambda_2^2\lambda_4^2 + \lambda_1^2\lambda_2^2\lambda_3^2 + \lambda_2^2\lambda_3^2\lambda_4^2)r_1;$$

$$v^2e^2q_3 = \lambda_1^2\lambda_2^2\lambda_3^2\lambda_4^2r_1.$$

Для сокращения записи в этих формулах введено обозначение $v = a - c$. В результате решения этой системы линейных уравнений находим

$$q_3 = \lambda_1^2 \lambda_2^2 \lambda_3^2 \lambda_4^2 r_1 / v^2 / e^2;$$

$$q_4 = ((\lambda_1^2 \lambda_3^2 \lambda_4^2 + \lambda_1^2 \lambda_2^2 \lambda_4^2 + \lambda_1^2 \lambda_2^2 \lambda_3^2 + \lambda_2^2 \lambda_3^2 \lambda_4^2) r_1 / v / e^2 + 2q_3) / v;$$

$$q_1 = ((\lambda_2^2 \lambda_4^2 + \lambda_3^2 \lambda_4^2 + \lambda_1^2 \lambda_2^2 + \lambda_2^2 \lambda_3^2 + \lambda_1^2 \lambda_3^2 + \lambda_1^2 \lambda_4^2 - a^2) \times r_1 - q_3 + 2vq_4) / e^2;$$

$$q_2 = (\lambda_2^2 + \lambda_4^2 + \lambda_1^2 + \lambda_3^2 + 2a) r_1 / e^2 - q_4.$$

Обозначим столбцы матрицы преобразования T к канонической форме через $t_k, k = 1, \dots, 2n$. В соответствии с алгоритмом [2], символьные выражения для этих столбцов находятся путем решения $2n$ систем линейных уравнений вида

$$(A - \lambda_k I) t_k = 0, \quad k = 1, \dots, 2n. \quad (5)$$

Проще всего решение этих уравнений может быть получено следующим образом. Вначале для каждого k путем вычеркивания в символьной матрице $(A - \lambda_k I)$ строки и столбца с номером k формируется квадратная матрица B_k порядка $2n - 1$. Вычеркиванием элемента с номером k в векторе t_k вводится новый вектор v_k . Полученная в результате система уравнений с невырожденной матрицей B_k легко решается относительно v_k средствами символьного пакета MATLAB. По результатам решения формируется искомый вектор t_k , причем вначале элементу в позиции k этого вектора присваивается значение, равное 1.

Полученные при этом выражения зачастую оказываются очень громоздкими. Чтобы получить более про-

стые выражения, можно воспользоваться тем, что векторы t_k из уравнений (5) определяются неоднозначно, и умножить каждый из них на любое подходящее выражение. Например, таким способом в выражениях для элементов t_k можно избавиться от знаменателей, которые как раз и придают полученным первоначально выражениям упомянутый необозримо громоздкий вид. В рассматриваемом примере окончательные выражения для t_k могут иметь, например, вид, определяемый уравнениями

$$t_k = \begin{bmatrix} e^2 (\lambda_k^2 + v) (-q_3 + q_4 \lambda_k^2) \\ \lambda_k e^2 (\lambda_k^2 + v) (-q_3 + q_4 \lambda_k^2) \\ \lambda_k^2 ((\lambda_k^2 + a)^2 r_1 + e^2 (q_1 - \lambda_k^2 q_2)) \\ \lambda_k^3 ((\lambda_k^2 + a)^2 r_1 + e^2 (q_1 - \lambda_k^2 q_2)) \\ -\lambda_k (c (\lambda_k^2 + a) r_1 + e^2 (q_1 + v q_2)) (-q_3 + q_4 \lambda_k^2) \\ (c (\lambda_k^2 + a) r_1 + e^2 (q_1 - \lambda_k^2 q_2)) (-q_3 + q_4 \lambda_k^2) \\ -\lambda_k q_3 ((\lambda_k^2 + a)^2 r_1 + e^2 (q_1 - \lambda_k^2 q_2)) \\ -((\lambda_k^2 + a)^2 r_1 + e^2 (q_1 - \lambda_k^2 q_2)) (-q_3 + q_4 \lambda_k^2) \end{bmatrix}, \quad k = 1, \dots, 2n. \quad (6)$$

Вычисления на последующих этапах осуществляются в соответствии с базовым алгоритмом.

Результаты моделирования движения крана при $t_f = 4c$ приведены на рис. 1, 2.

Пример 3. Рассматривается задача регулирования из статьи [1]. Дана линейная система $\dot{x} = Fx + Gu$, матрицы которой имеют вид

$$F = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 1 & 0 & 0 \\ 0, 1 & -1, 02 & 0 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

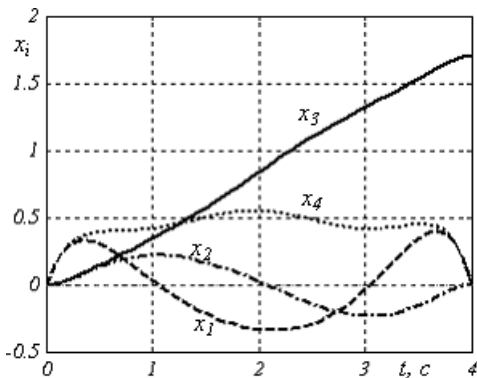


Рисунок 1

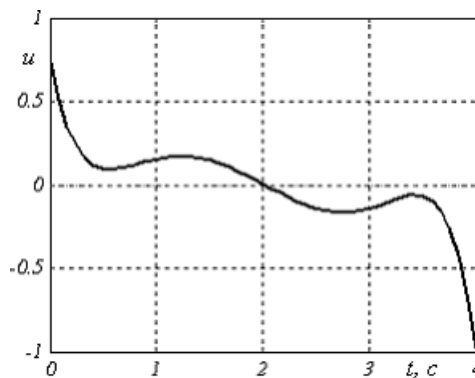


Рисунок 2

Движение системы рассматривается на интервале $t \in [0, 15]$. Граничные условия имеют вид: $c_0 = [2, 1, 7, 5]^T$; $c_f = [0, 0, 0, 0]^T$; $M = I$. В качестве канонической формы в данном примере принята диагональная матрица.

На управляющие функции $u_k (k = 1, 2, 3)$ наложены ограничения вида $u_{\min}(t) \leq u_k(t) \leq u_{\max}(t)$. Учесть эти ограничения в постановке оптимизационной задачи невозможно, т. к. краевая задача при этом станет нелинейной. Поэтому используется способ *апостериорного* учета этих ограничений, заключающийся в том, что для каждого t величины $u_k(t)$ вычисляются вначале в соответствии с базовым алгоритмом (т. е. так, как будто никаких ограничений нет), а затем они подвергаются нелинейному преобразованию

$$\bar{u}_k(t) = \min\{u_{\max}(t), \max[u_{\min}(t), u_k(t)]\}, \quad (7)$$

в результате чего формируются величины управлений $\bar{u}_k(t)$, фактически подаваемые на исполнительные органы. Результаты моделирования для случая $u_{\min} = -1, u_{\max} = 1$ приведены на рис. 3–6.

Пример 4. Рассматривается задача построения алгоритмов автопилота с эталонной моделью, обеспечивающего управление боковым движением летательного аппарата с полной развязкой каналов крена и рыс-

кания [7]. Угловое движение аппарата описывается уравнениями

$$\begin{aligned} \dot{\omega} &= a_{11}(t)\omega + a_{12}(t)\beta + a_{13}(t)\dot{\beta} + b_1(t)u_1; \\ \dot{\beta} &= a_{31}(t)\omega + a_{32}(t)\beta + a_{33}(t)\dot{\beta} + b_2(t)u_2, \end{aligned} \quad (8)$$

где ω – угловая скорость по крену; β – угол скольжения; u_1, u_2 – углы отклонения рулевых органов; коэффициенты a_{ij} ($i = 1, 3; j = 1, 2, 3$), b_k ($k = 1, 2$) изменяются по времени с изменением высоты и скорости полета.

В качестве эталонной модели принимается система уравнений вида

$$\dot{z} = Cz + g, \quad (9)$$

где $z = [\omega_m, \beta_m, \dot{\beta}_m]^T$; $g = [g_1, 0, g_2]^T$; g_1, g_2 – управляющие воздействия по крену и рысканию, поступающие от ручки и педали управления соответственно. Элементы матрицы

$$C = \begin{bmatrix} c_{11} & 0 & 0 \\ 0 & 0 & 1 \\ 0 & c_{32} & c_{33} \end{bmatrix}$$

являются заданными постоянными величинами.

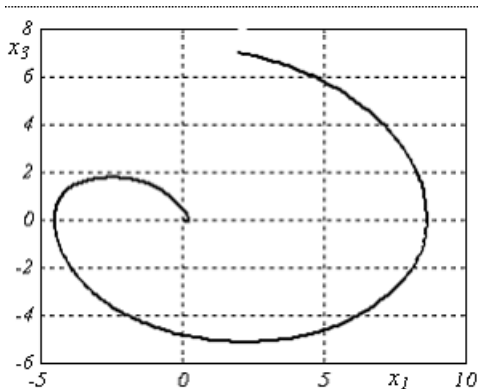


Рисунок 3

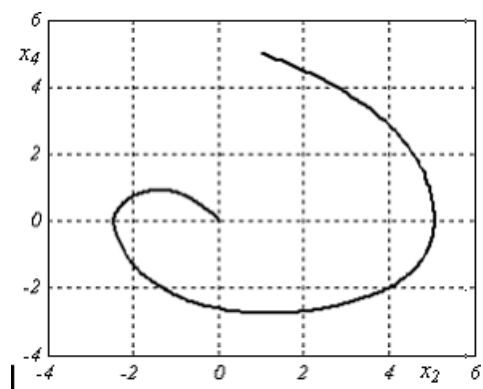


Рисунок 4

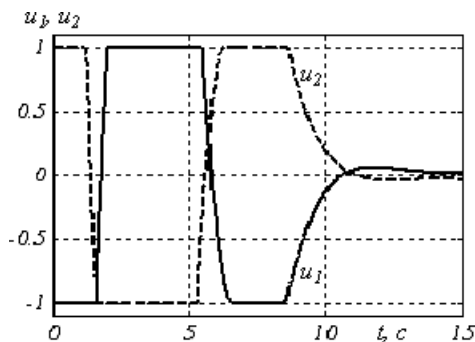


Рисунок 5

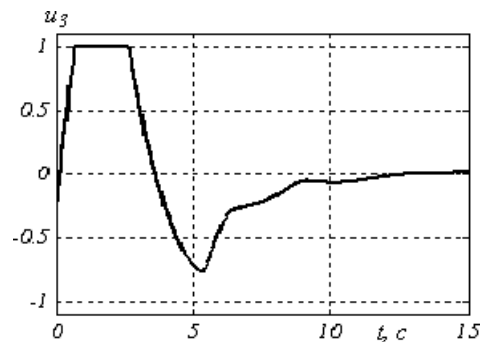


Рисунок 6

Решение уравнения (9) дается формулой Коши

$$z(t) = \Psi(t, t_0)z(t_0) + \int_{t_0}^t \Psi(t, \tau)g(\tau)d\tau. \quad (10)$$

Переходная матрица $\Psi(t, t_0)$ легко находится с использованием средств символьного пакета MATLAB

$$\Psi(t, t_0) = \frac{1}{c} \begin{bmatrix} ce_1 & 0 & 0 \\ 0 & d_2e_3 - d_3e_2 & e_2 - e_3 \\ 0 & c_{32}(e_2 - e_3) & d_2e_2 - d_3e_3 \end{bmatrix}$$

где

$$e_k = \exp[d_k(t - t_0)], \quad k = 1, 2, 3;$$

$$d_1 = c_{11}; \quad d_2 = (c_{33} + c)/2; \quad d_3 = (c_{33} - c)/2;$$

$$c = \sqrt{c_{33}^2 + 4c_{32}}.$$

Система (8) являются нестационарной, поэтому непосредственное применение к ней базового алгоритма невозможно. Мы поступим следующим образом. Разобьем время управления на циклы постоянной длины h . Величина h определяет время реакции системы на действия пилота, поэтому она должна выбираться минимально возможной (например, $h < 0,1$ с). На малом интервале времени от момента $t_0 = t$ до момента $t_h = t + h$ исходную модель управляемой системы (8) заменим более простой моделью, представляющей собой два независимых интегратора вида $\ddot{\omega} = v_1$ и $\ddot{\beta} = v_2$, где v_1, v_2 – новые управляющие переменные. В таком подходе нет ничего необычного – замена на малом интервале времени сложных моделей более простыми математическими объектами (полиномами, сплайнами и т. п.) широко используется в вычислительной математике, например, при построении алгоритмов численного интегрирования.

Представим новую расширенную модель системы в виде

$$\dot{x} = Fx + Gv,$$

где

$$x = [\omega, \dot{\omega}, \beta, \dot{\beta}]^T; \quad v = [v_1, v_2]^T;$$

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \quad G = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

В начальный момент $t_0 = t$ координаты $\dot{\omega}, \beta, \dot{\beta}$ вектора c_0 в уравнении (2) могут быть найдены по результатам измерений фактических значений параметров ω и β путем их численного дифференцирования либо получены в качестве выходных параметров некоторого наблюдателя. Координаты вектора c_f в конечный момент $t_f = t_h$ определим следующим образом. Располагая информацией о значениях параметров g_1, g_2 эталонной модели в k моментов времени $t_0 = t, t_1 = t - h, \dots, t_k = t - (k - 1)h, \dots$, мы можем вычислить экстраполированные значения этих параметров в момент $t_h = t + h$. Тогда значения параметров $\omega_h, \beta_h, \dot{\beta}_h$ вычисляются по формуле (10), а параметров $\dot{\omega}_h, \dot{\beta}_h$ – по уравнениям эталонной модели (9). Сформированные подобным образом граничные условия c_0 и c_f используются в базовом алгоритме для вычисления по формулам (33) программной траектории $x^*(t)$ перехода от фактического углового положения летательного аппарата к задаваемому эталонной моделью требуемому положению. Вычислять вектор $p^*(t)$ при этом нет необходимости – полученное с использованием этого вектора управление (34) для нас совершенно бесполезно, т. к. оно не связано ни с каким физически реализуемым управлением в исходной задаче. Поэтому в данном случае используется другой способ вычисления требуемых управляющих функций u_1 и u_2 . Заметим, что в числе программ $x^*(t)$ вычисляются функции $x_2^*(t) = \dot{\omega}^*(t)$ и $x_5^*(t) = \dot{\beta}^*(t)$. Полагая в уравнениях (8) фактической модели $\dot{\omega}(t) = \dot{\omega}^*(t), \dot{\beta}(t) = \dot{\beta}^*(t)$, значения параметров управления найдем из уравнений

$$u_1 = (\dot{\omega}^* - (a_{11}\omega + a_{12}\beta + a_{13}\dot{\beta}))/b_1;$$

$$u_2 = (\dot{\beta}^* - (a_{31}\omega + a_{32}\beta + a_{33}\dot{\beta}))/b_2.$$

При таком подходе к выбору параметров управления движение летательного аппарата осуществляется в соответствии с уравнениями эталонной модели (9), т. е. оказывается не зависящим от любых конкретных (но известных) значений параметров его модели. При этом в качестве модели летательного аппарата (8) могут быть использованы уравнения любой сложности, а не только линеаризованные уравнения.

Моделирование процесса управления проводилось для случая, когда коэффициенты модели изменяются во времени произвольным образом. В данном случае рассматривался вариант задания этих функций в виде отрезков синусоид со случайно выбранными частотами, амплитудами и фазами.

Результаты моделирования, приведенные на рис. 7, 8, свидетельствуют о том, что при описанном подходе обеспечивается практически идеальное отслеживание

эталонной траектории, задающей независимое движение летательного аппарата в каналах крена и рыскания. На рис. 9 приведены отклонения фактических значений угла скольжения от его эталонных значений в процессе маневра. Всплески погрешностей отслеживания эталонных значений на графике соответствуют точкам, в которых производная функции g_2 претерпевает разрывы, вследствие чего возрастают погрешности экстраполяции значений этой функции на моменты t_h . На рис. 10 представлен характер изменения в ходе маневра величин углов отклонения элеронов (u_1) и руля направления (u_2).

Пример 5. Рассматривается задача перевода КА с маховичной системой управления, находящегося на круговой орбите, из произвольного неориентированного положения в положение, соответствующее полету в орбитальной системе координат (ОСК). Цель данного примера – показать, каким образом базовый алгоритм (алгоритм терминального управления с фиксированным временем) может использоваться для решения задач регулирования и стабилизации, где время заранее не задается или даже может быть вообще неограниченным.

Угловое движение КА относительно ОСК описывается уравнениями

$$\begin{aligned} J\dot{\omega} &= h \times (\omega + \omega_0) + J(\omega \times \omega_0 - \dot{\omega}_0) - \tau; \\ \dot{\gamma} &= \Gamma\omega, \end{aligned} \quad (11)$$

где ω_0 – вектор орбитальной угловой скорости, остальные обозначения совпадают с [4]. Для решения задачи воспользуемся известным преобразованием [4], суть которого заключается в переходе от множества переменных $\{\omega, \gamma, \tau\}$ исходной задачи к новому множеству $\{\gamma, \dot{\gamma}, u\}$. Система (11) при этом *точно* преобразуется в эквивалентную линейную систему вида $\dot{x} = Fx + Gu$,

$$\text{где } x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \gamma \\ \dot{\gamma} \end{bmatrix}; F = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}; G = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Таким образом, система (11) представляется в виде двумерного интегратора. Граничные условия c_0 и c_f для этой новой системы задаются следующим образом. В текущий момент времени t , принимаемый за начальный t_0 , координаты вектора c_0 определяются по данным системы определения ориентации и показаниям измерителей угловой скорости. Вектор c_f задается в виде $c_f = 0$, матрица $M = I$.

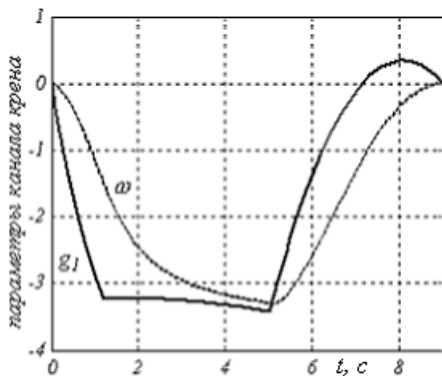


Рисунок 7:

g_1 – управляющее воздействие от ручки управления;
 ω – эталонная и фактическая угловые скорости крена

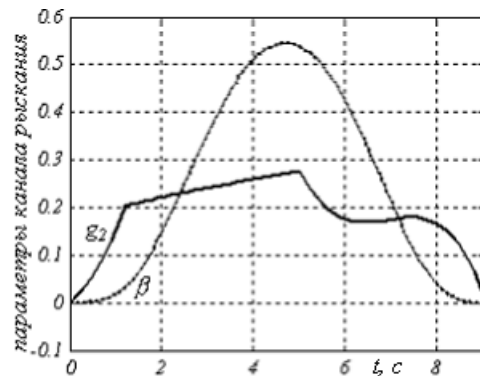


Рисунок 8:

g_2 – управляющее воздействие от педали управления;
 β – эталонный и фактический углы скольжения

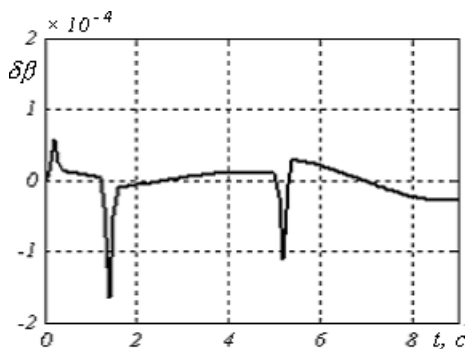


Рисунок 9

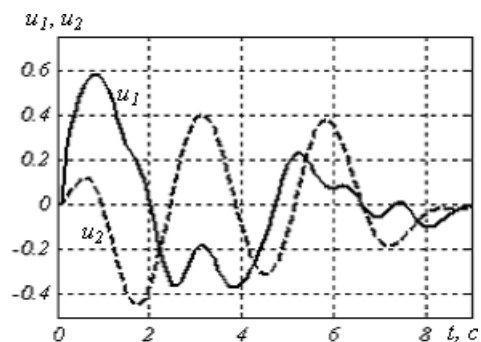


Рисунок 10

Для использования базового алгоритма требуется также задание времени t_f достижения заданного терминального состояния c_f . В задачах регулирования, где это время обычно неизвестно, значение t_f может выступать в качестве одного из параметров, регулирующих скорость протекания переходного процесса.

Возврат от управления u в линейной задаче к управлению τ в исходной задаче осуществляется по формуле, аналогичной [4]

$$\tau = J\Gamma^{-1}u + h \times (\omega + \omega_0) - 1/2\gamma_0^{-1}(\omega^T \omega)J\gamma + J(\omega \times \omega_0 - \dot{\omega}_0),$$

где

$$\gamma_0 = \sqrt{1 - x_1^T x_1}; \quad \gamma = x_1;$$

$$\omega = 2(\gamma_0 x_2 + \gamma_0^{-1} x_1 x_1^T x_2 - x_1 \times x_2).$$

Учет ограничений на величины управляющих моментов маховиков осуществляется путем использования нелинейного преобразования (7).

Результаты моделирования процесса переориентации для двух вариантов задания t_f приведены на рисунках 11, 12. Видно, что процесс регулирования, длящийся примерно 400–800 с плавно переходит в процесс стабилизации, который теоретически может продолжаться неограниченно долго.

Пример 6. Для многих управляемых процессов характерной является ситуация, когда общая задача управления разбивается на несколько элементарных подзадач. Траектория движения в этом случае оказывается состоящей из нескольких участков. При этом важной является проблема обеспечения плавной стыковки этих участков. Рассмотрим один из способов решения данной проблемы на примере задачи кадровой фотосъемки участков земной поверхности с борта КА, находящегося на околоземной орбите.

На рис. 13 приведены результаты моделирования процесса управления съемкой участка поверхности Земли размером 15×15 км (3 строки по 3 кадра, размер кадра 5 км) в предположении, что, как и в предыдущем примере, для перехода в пространство линейных параметров используется модель движения в виде совокупности независимых интеграторов. Как видно из графиков, на границах смежных участков производные угловых скоростей в каналах крена (ω_1) и тангажа (ω_2) терпят разрывы. При этом создаются неблагоприятные условия для системы стабилизации углового движения КА относительно программных траекторий. Плавная стыковка отдельных участков может быть обеспечена следующим образом. Добавим к вектору состояния эквивалентной линейной системы (6) координату $x_3 = \gamma$ и изменим соответствующим

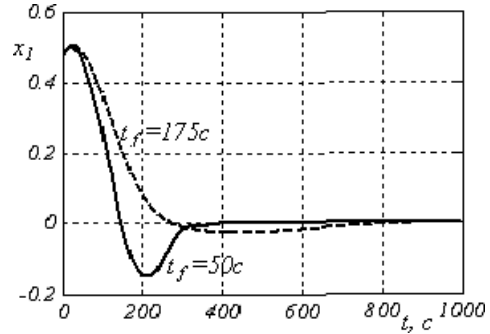


Рисунок 11

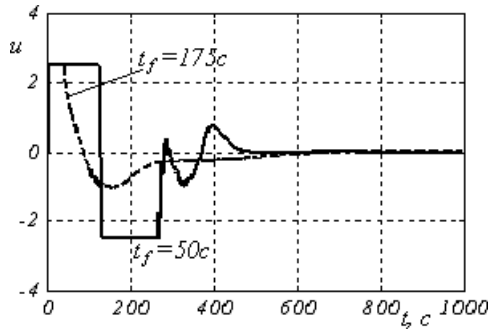


Рисунок 12

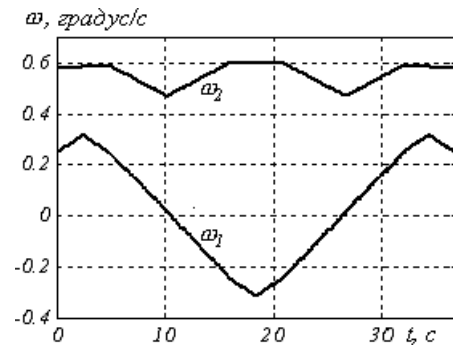


Рисунок 13

образом матрицы F и G . В исходной задаче граничные условия на эту координату не наложены, поэтому мы можем располагаться ими по своему усмотрению.

В частности, мы можем потребовать, чтобы в моменты стыковки смежных участков совпадали не только значения координат x_1, x_2 , но и значения координаты x_3 . Тем самым, будет обеспечена требуемая плавность изменения параметров ω_1 и ω_2 . Результаты моделирования для этого случая приведены рис. 14, 15 (пунктирные линии на графиках соответствуют линейной модели в виде 2-мерного интегратора, сплошные – расширенной модели в виде 3-мерного интегратора). Графики управляющих функций показаны

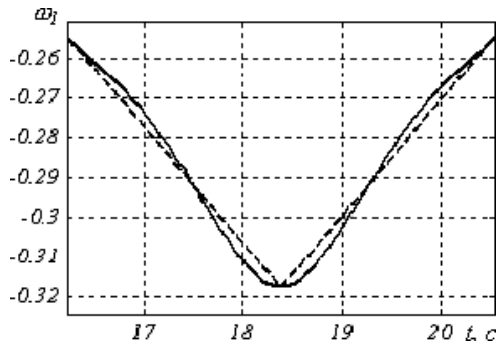


Рисунок 14

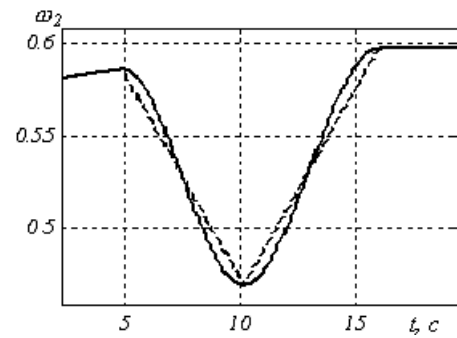


Рисунок 15

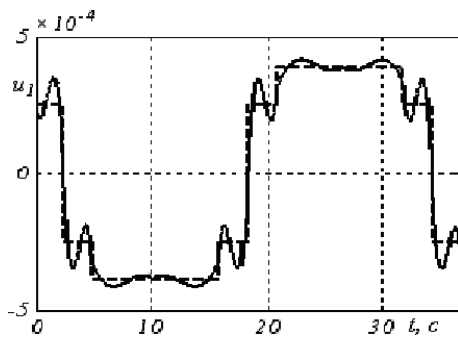


Рисунок 16

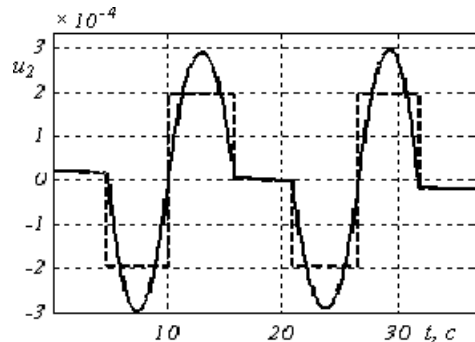


Рисунок 17

на рис. 16, 17. Из рисунков видно, что предложенный подход обеспечивает более благоприятные условия для работы системы стабилизации. При дальнейшем расширении вектора состояния может быть обеспечена любая желаемая степень гладкости.

Пример 7. Для традиционных подходов к задачам оптимального управления характерно почти полное отсутствие возможностей учета априорной информации об управляемом процессе. Между тем, зачастую, еще только приступая к решению задачи, мы уже достаточно много знаем о ней. Например, при решении задач наведения в космосе мы можем принять во внимание хорошо известный результат, полученный ранее многими исследователями, – для маневров в гравитационном поле оптимальными являются режимы полета с максимальной тягой. С учетом сказанного, рассмотрим задачу терминального управления на примере выведения КА в заданную точку эллиптической орбиты в заданное время. Для простоты будем считать, что движение происходит в центральном поле (как мы увидим далее, вид гравитационного поля особого значения не имеет). Уравнения движения КА относительно инерциального базиса имеют вид

$$\ddot{r} = ae + g, \quad (12)$$

где $e = \begin{bmatrix} \cos\theta\cos\psi \\ \sin\theta\cos\psi \\ -\sin\psi \end{bmatrix}$ – единичный вектор направления

силы тяги; $a = u_c/(\tau - t)$ – модуль ускорения силы тяги; θ – угол тангажа; ψ – угол рыскания; u_c – скорость истечения; $\tau = m_0/c$ – время полного «выгорания» топлива; m_0 – начальная масса; c – секундный расход; $g = -vr$ – вектор гравитационного ускорения; $v = \mu/|r|^3$.

Чтобы привести исходное нелинейное уравнение (12) к эквивалентной линейной системе введем новые переменные $x_1 = r$, $x_2 = \dot{r}$, $x_3 = \ddot{r}$, $x_4 = \dddot{r}$ и новое управление u . Запишем новые уравнения движения в виде трех (по числу каналов управления) независимых четырехмерных интеграторов. Уравнения движения в каждом канале управления при этом приводятся к уравнениям вида $\dot{x} = Fx + Gu$ с матрицами

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Теперь для того чтобы применить базовый алгоритм, осталось сформировать граничные условия c_0

и c_f . Будем считать начальные условия движения r_0, \dot{r}_0 известными, конечные условия r_f, \dot{r}_f и время наведения $(t_f - t_0)$ – заданными. Заданными будем считать также начальные и конечные значения углов тангажа θ_0, θ_f и рыскания ψ_0, ψ_f . Располагая этими данными, мы можем непосредственно сформировать граничные условия для переменных x_1 и x_2 . Для переменной x_3 граничные условия определяются из уравнения (12). Чтобы задать указанные условия для переменной x_4 , найдем по уравнению (12) производную

$$\dot{r} = \dot{a}e + a\dot{e} - \dot{v}r - v\dot{r},$$

где

$$\dot{e} = \dot{\theta} \begin{bmatrix} -\sin\theta\cos\psi \\ \cos\theta\cos\psi \\ 0 \end{bmatrix} - \dot{\Psi} \begin{bmatrix} \cos\theta\sin\psi \\ \sin\theta\sin\psi \\ \cos\psi \end{bmatrix}; \quad \dot{a} = a/(\tau - t); \quad \dot{v} = -3vp; \quad p = r^T \dot{r}/|r|^2.$$

Будем предполагать, что на значения параметров $\dot{\theta}, \dot{\psi}$ в моменты времени t_0, t_f никаких условий не наложено, и мы можем распоряжаться ими по своему усмотрению. Выберем их наиболее естественным способом:

$$\dot{\theta}_0 = \dot{\theta}_f = (\theta_f - \theta_0)/(t_f - t_0),$$

$$\dot{\Psi}_0 = \dot{\Psi}_f = (\Psi_f - \Psi_0)/(t_f - t_0).$$

Наконец, зададим матрицу M в уравнении (3) равной единичной матрице и сформируем векторы c_0 и c_f . В заключение, выберем корни характеристического уравнения в соответствии с распределением Баттерворта и зададим радиус круга этого распределения, равным 1.

В процессе работы базового алгоритма в числе программ $x^*(t)$ вычисляются функции $x_1^*(t) = r^*(t)$ и $x_3^*(t) = \dot{r}^*(t)$. На основании этих данных программные значения величин ускорения силы тяги $a^*(t)$ и углов ориентации вектора тяги $\theta^*(t), \psi^*(t)$ в момент t определяются следующим образом.

На основании уравнения (12) можем записать

$$x_3^*(t) = a(t)e(t) + g[x_1^*(t)].$$

Отсюда получаем

$$a^*(t) = \sqrt{[x_3^*(t) - g^*(t)]^T [x_3^*(t) - g^*(t)]}; \quad e^*(t) = [x_3^*(t) - g^*(t)]/a^*(t).$$

Окончательно имеем

$$\psi^*(t) = -\arcsin[e_3^*(t)]; \quad \theta^*(t) = \arctan[e_2^*(t)/e_1^*(t)].$$

На рис. 18, 19 приведены результаты моделирования процесса наведения для следующих значений параметров КА и участка наведения [5]:

$$t_0 = 0; \quad t_f = 387 \text{ с}; \quad m_0 = 14000 \text{ кг}; \quad c = 25 \text{ кг/с};$$

$$u_c = 4,5 \text{ км/с}; \quad \theta_0 = 47,35^\circ; \quad \theta_f = 96,23^\circ;$$

$$\psi_0 = 55,4^\circ; \quad \psi_f = 0;$$

$$r_0 = [6372, 9; -1570, 0; -299, 4] \text{ км};$$

$$\dot{r}_0 = [1, 502; 5, 477; 1, 885] \text{ км/с};$$

$$r_f = [6510, 663510927766; 1238, 442814248355; 0] \text{ км};$$

$$\dot{r}_f = [-1, 105899452214530; 10, 13460122429019; 0] \text{ км/с}.$$

При вычислениях с разрядностью мантиссы 16 десятичных знаков погрешности выведения КА составляют: по координатам $\sim 3,1 \cdot 10^{-3}$ м, по скоростям $\sim 2,9 \cdot 10^{-8}$ м/с. Вычисления с более высокой разрядностью обеспечивают еще большую точность выведения. В этом, конечно же, нет ничего удивительного, так как мы решили краевую задачу аналитически, т. е. совершенно точно. Погрешности при этом определяются только погрешностями численной реализации процесса вычисления программного движения по этим формулам и дей-

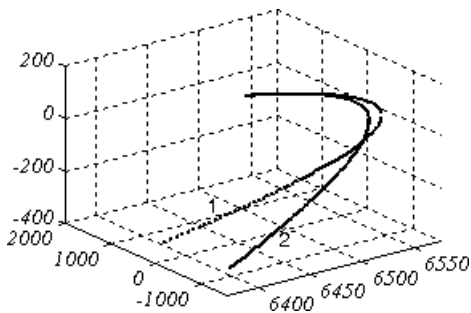


Рисунок 18:

1 – заданная орбита; 2 – траектория выведения

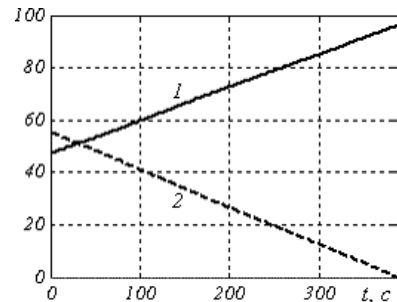


Рисунок 19:

1 – угол тангажа; 2 – угол рыскания

ствующими возмущениями. Для компенсации возмущений управление в линейной задаче должно формироваться в виде $u(t) = u^*(t) + \delta u(t)$. Поэтому КА должен быть оснащен двигателем с регулируемой в небольших пределах тягой.

Интересно сравнить рассмотренный в данном примере метод наведения с лучшим на сегодняшний день явным методом [5, 8], который впервые был применен в 60–70-х годах прошлого века для наведения ракеты-носителя *Saturn V* в ходе реализации программы полетов КК *Apollo* на Луну. В настоящее время этот, несколько модифицированный, метод применяется для наведения большинства современных ракет-носителей (*Space Shuttle*, *Arian* и др.). Главное отличие заключается в том, что с использованием явного метода обеспечивается выполнение заданных конечных условий лишь для пяти параметров движения (отсутствует управление продольной координатой). При этом время выхода на заданную орбиту и положение точки на этой орбите, в которой оказывается КА в момент конца участка выведения, зависят от возмущений и заранее неизвестны. Метод наведения с использованием базового алгоритма обеспечивает выполнение заданных конечных условий по всем шести параметрам движения КА точно в заданное время. Данное свойство позволяет, дополнительно к тем возможностям, которыми обладает явный метод наведения, решать многие качественно новые задачи, возникающие в ходе выполнения маневров в космосе. Например, при выполнении операции по спасению космонавтов с терпящего бедствие КА решающим может оказаться фактор времени. С использованием предлагаемого метода оказывается возможным непосредственное решение задачи встречи на орбите, минуя занимающие значительное время операции фазирования орбит. При этом обеспечивается столь высокая точность наведения, что практически осуществимым может оказаться выведение КА непосредственно в ту область пространства, из которой далее могут начинаться этапы причаливания и стыковки.

Кроме того, в рассматриваемом подходе крайняя задача решена точно, без каких бы то ни было упрощений исходной модели движения, которые надо было бы компенсировать дополнительным расходом топлива. Поэтому данный подход обеспечивает более высокую степень оптимальности управления по сравнению с явным методом наведения. Наконец, что тоже немало важно, алгоритмы рассматриваемого метода значительно проще, чем алгоритмы явного метода наведения.

Пример 8. Рассматривается применение базового алгоритма для высокоточного численного интегрирования уравнений небесной механики. Традиционные методы численного интегрирования являются предметом вычислительной математики. Подход к задачам интегрирования как к задачам оптимального управления позволяет по-новому взглянуть на многие проблемы, свя-

занные с численным интегрированием. Рассмотрим особенности предлагаемого подхода на примере интегрирования периодических орбит в круговой ограниченной задаче трех тел. Уравнения движения имеют вид [3]

$$\ddot{x} = 2\dot{y} + \frac{\delta\Omega}{\delta x}; \quad \ddot{y} = 2\dot{x} + \frac{\delta\Omega}{\delta y},$$

где

$$\Omega = \frac{1}{2} \left[(1-\mu) \left(\rho_1^2 + \frac{2}{\rho_1} \right) + \mu \left(\rho_2^2 + \frac{2}{\rho_2} \right) \right];$$

$$\rho_1^2 = (x + \mu)^2 + y^2; \quad \rho_2^2 = [x - (1 - \mu)]^2 + y^2;$$

$$\mu = 0,012277471.$$

Начальные условия интегрирования:

$$x_0 = 0,994; \quad \dot{x}_0 = 0;$$

$$\dot{y}_0 = 2,0317326295573368357302057924;$$

период $P = 11,124340337266085134999734047$. Орбита, соответствующая этим начальным условиям, приведена на рис. 20.

Как видно из рисунка, в окрестности точки 2 наблюдается тесное сближение движущегося тела с Луной. Именно на этом участке при численном интегрировании происходит основная потеря точности. Поэтому интегрирование подобных орбит является хорошим испытанием для любого метода интегрирования. Уравнения (13) являются нелинейными. Ясно, что непосредственно к ним базовый алгоритм применить невозможно. Поэтому мы вновь воспользуемся возможностью

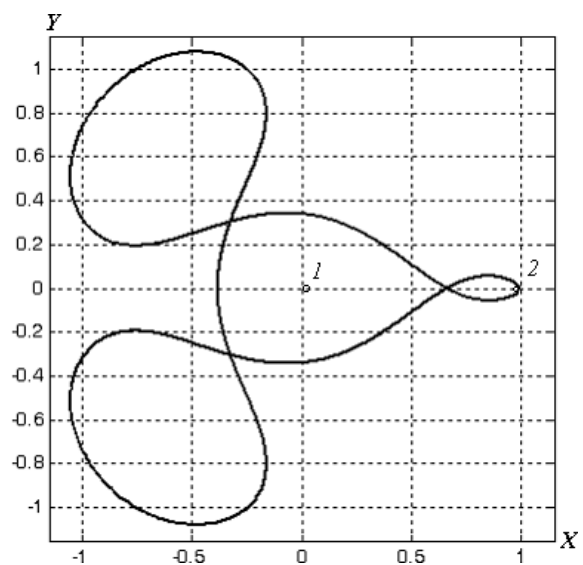


Рисунок 20:

1 – Земля; 2 – Луна

точного приведения этих уравнений к эквивалентной линейной системе. Введем новый n -мерный вектор состояния

$$z = [z_1^T, z_2^T, \dots, z_n^T]^T,$$

где $z_1 = [x, y]$; $z_2 = [\dot{x}, \dot{y}]$; \dots , $z_n = [x^{(n-1)}, y^{(n-1)}]$.

В этих новых переменных уравнения (13) представляются в виде n -мерного интегратора $\dot{z}_1 = z_2, \dot{z}_2 = z_3, \dots, \dot{z}_{n-1} = z_n, \dot{z}_n = u$, где u – некоторое фиктивное управление. Ввиду недостатка места символьные выражения для координат z_4, \dots, z_n здесь не приводятся, однако они легко могут быть получены путем дифференцирования уравнений (13) с использованием средств символьного пакета MATLAB.

Граничные условия $c_0 = z(t_0)$ в начале каждого шага интегрирования вычисляются с использованием уравнений (13) и полученных символьных выражений для высших производных. Граничные условия в конце шага интегрирования задаются следующим образом.

Поскольку в момент $t_f = t_0 + h$ требуемые значения параметров $z_1(t_f), z_2(t_f)$ неизвестны, в связи с чем граничные условия для них задать невозможно, выберем матрицу M в уравнении (3) в виде единичной матрицы порядка n с вычеркнутыми из нее двумя первыми строками. Интегрирование будем вести по обычной для неявных методов интегрирования схеме «прогноз плюс коррекция». На самом первом шаге интегрирования прогнозируемые значения $z_1(t_f), z_2(t_f)$ вычисляются с использованием разложения в ряд Тейлора

$$z_1(t_f) = z_1(t_0) + z_2(t_0)h + z_3(t_0)h^2/2 + z_4(t_0)h^3/6 + \dots;$$

$$z_2(t_f) = z_2(t_0) + z_3(t_0)h + z_4(t_0)h^2/2 + \dots$$

На последующих шагах интегрирования прогноз осуществляется с результатами использования базового алгоритма, причем в качестве ξ_0 берется его значение, полученное на предыдущем шаге.

По вычисленным таким образом значениям $z_1(t_f), z_2(t_f)$ с использованием полученных ранее символьных выражений находятся начальные приближения для оставшихся координат вектора $z(t_f)$ и формируется вектор $c_f = [z_3(t_f), \dots, z_n(t_f)]$. На этом подготовка к использованию базового алгоритма завершается.

В результате работы базового алгоритма на этапе коррекции вычисляются некоторые новые значения $z_1(t_f)$ и $z_2(t_f)$, по ним вычисляется новые значения координат вектора c_f и т. д. Как показывают численные эксперименты, процесс этот сходится за 2–3 итерации.

Моделирование процесса интегрирования проводилось на интервале времени, равном одному периоду P , для двух значений порядков расширенной системы: $n = 6$ и $n = 7$. При $n = 6$ точность интегрирования характеризуется следующими величинами погрешностей:

- по координатам $\delta r = 4 \cdot 10^{-13}$,
- по составляющим скорости $\delta \dot{r} = 6 \cdot 10^{-11}$.

Такая точность сопоставима с точностью, обеспечиваемой лучшими на сегодняшний день методами интегрирования [9]. При той же последовательности шагов интегрирования для случая $n = 7$ погрешности составляют соответственно $\delta r = 3 \cdot 10^{-18}$, $\delta \dot{r} = 5 \cdot 10^{-16}$. Как видим, увеличение порядка метода на единицу приводит к уменьшению величин погрешностей на пять порядков. Это свойство предлагаемого подхода к задачам численного интегрирования, казалось бы, указывает путь к возможности практически неограниченного повышения точности интегрирования. Однако, в настоящее время этому препятствует уже упоминавшееся выше «проклятие разрядности».

Пример 9. При рассмотрении предыдущих примеров мы убедились в том, что в случае, когда уравнения исходной задачи являются *нелинейными*, расширение вектора состояния путем включения в него нескольких высших производных позволяет построить эффективные законы управления, обладающие рядом ценных свойств. Если исходная система уравнений *уже является линейной*, к ней может быть применен точно такой же подход с расширением вектора состояния.

Пусть исходная система задана в виде системы второго порядка

$$\dot{x}_1 = f_{11}x_1 + f_{12}x_2; \quad \dot{x}_2 = f_{21}x_1 + f_{22}x_2 + gu.$$

Для моделирования в качестве такой исходной системы возьмем систему (1), уже рассматривавшуюся в примере 1. Зададим интервал управления, равным 10 с, и граничные условия в виде $c_0 = [5; -5]$; $c_f = [-1; 1]$. Результаты моделирования процесса управления для этой системы второго порядка, полученные с использованием базового алгоритма, приведены на рис. 21, 22.

Как видно из рис. 21, для процесса управления в исходной задаче характерным является приложение максимальных значений u в начале и в конце интервала управления. Между тем ясно, что такой характер изменения управляющей функции по ряду причин на практике может оказаться неприемлемым. Например, на управляющую функцию могут быть наложены ограничения вида $u_{\min} \leq u \leq u_{\max}$. При наличии подобных ограничений решение нелинейной краевой задачи получить вообще невозможно.

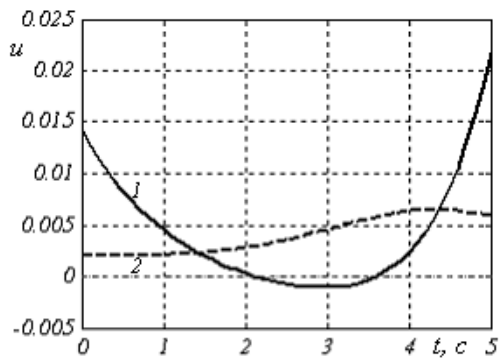


Рисунок 21:

1 – управление в исходной задаче;
2 – то же в расширенной задаче

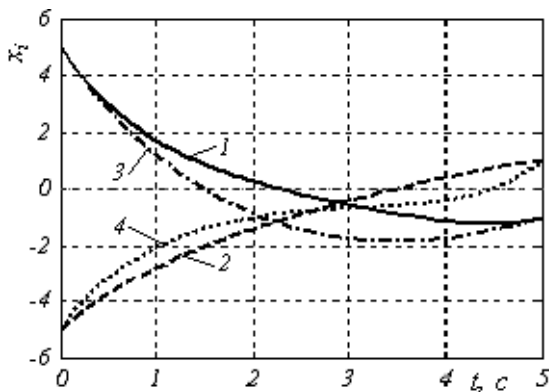


Рисунок 22:

1, 2 – координата и скорость в исходной задаче;
3, 4 – то же в расширенной задаче

Расширим вектор состояния системы (14), для чего введем обозначение $f_{23} = g$ и добавим к вектору состояния новую переменную $x_3 = u$. В исходной задаче никаких ограничений на эту переменную не наложено. Следовательно, мы можем распорядиться ее граничными значениями, а также граничными значениями нескольких ее высших производных, по своему усмотрению. Ограничиваясь для простоты первой и второй производными, запишем вместо системы (14) расширенную систему

$$\dot{x} = Fx + Gv,$$

где v – новое управление, а матрицы F и G имеют вид

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -a & k & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}; \quad G = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Пользуясь свободой выбора граничных условий для переменных x_3 и x_4 , сформируем граничные условия для этой новой модели управляемой системы в виде

$$c_0 = [5; -5; 2e-3; 0]; \quad c_f = [-1; 1; 6e-3; -5e-4].$$

Результаты моделирования процесса управления с данными граничными условиями, представленные на рис. 22, показывают, что управляющая функция $u = x_3$ имеет в данном случае гораздо более приемлемый для практической реализации вид. Это свидетельствует о том, что подход с расширением вектора состояния линейных систем в сочетании с использованием базового алгоритма позволяет оказывать активное влияние на характеристики управляемых процессов. С использованием этого подхода в большинстве случаев можно добиться существенного улучшения указанных характеристик.

ВЫВОДЫ

Сочетание основных принципов классической теории управления (свободный выбор распределения корней системы и расширение вектора состояния для улучшения свойств замкнутой системы), принципа максимума (базовый алгоритм) и метода динамического программирования (поэтапное планирование операций) в рамках единого подхода, разработанного в первой части, во многих случаях позволяет получить решение задачи синтеза оптимальных управлений как для линейных, так и для нелинейных нестационарных систем, в том числе с ограничениями на управляющие функции.

БЛАГОДАРНОСТИ

Данная работа, состоящая из двух частей, своими корнями уходит в далекие 60-е годы прошлого века. Тогда автору, только что окончившему ХАИ, выпала редкая удача – в самом начале своего профессионального пути поработать в уникальном научном коллективе на предприятии, которое теперь называется «Хартрон». В те годы на этом вновь организованном предприятии все были молоды, веселы и очень талантливы. Уже тогда они искали свои подходы к решению проблем управления, закладывая фундамент научной школы, получившей со временем всеобщее признание. Именно этим коллективом впоследствии, в числе многих уникальных проектов, была разработана система управления стратегической ракетой, известной теперь под названием «Сатана».

Общение с такими выдающимися специалистами как Я. Е. Айзенберг, А. С. Гончар, С. С. Корума, В. Н. Романенко, В. Д. Стадник и др. наложило глу-

бокий отпечаток на мировоззрение автора, круг его научных интересов и, в конечном счете, на его судьбу. Всем этим замечательным людям – живым, а также тем, кого, к величайшему сожалению, уже нет с нами, – автор выражает свою глубокую искреннюю благодарность.

ПЕРЕЧЕНЬ ССЫЛОК

1. Габасов Р., Дмитрук Н. М., Кириллова Ф. М. Оптимизация многомерных систем управления с параллельными ограничениями // *АиТ*. – 2002. – № 3. – С. 3–26.
2. Квакернаак Х., Сиван Р. Линейные оптимальные системы управления. – М.: Наука, 1977. – 652 с.
3. Хайрер Э., Нерсетт С., Ваннер Г. Решение обыкновенных дифференциальных уравнений. – М.: Мир, 1990. – 512 с.
4. Дуайер Т. А. У., III. Точное нелинейное управление быстрыми вращениями КЛА посредством внутренней передачи количества движения // *Аэрокосмическая техника*. – 1987. – № 3. – С. 151–159.
5. Синха С. К., Шривастава С. К. Оптимальный закон наведения многоступенчатой ракеты-носителя на трехмерной траектории, использующий выражения в явном виде // *Аэрокосмическая техника*. – 1991. – № 3. – С. 74–87.
6. Кабанов С. А. Управление системами на самоорганизующихся моделях // *АиТ*. – 2001. – № 7. – С. 122–128.
7. Земляков С. Д., Рутковский В. Ю. О некоторых результатах развития теории и практического применения беспилотных адаптивных систем. // *АиТ*. – 2001. – № 7. – С. 103–121.
8. Чендлер Д. С., Смит И. Е. Схема итеративного управления и ее использование для различных аппаратов и космических операций // *Астронавтика и ракетодинамика*. – 1967. – № 8. – С. 1–24.
9. Everhart E. Implicit Single Sequence Methods for Integrating Orbits // *Celest. Mech.* – 1974. – V. 10. – P. 35–55.

Надійшла 15.07.06

На конкретних прикладах розглядаються особливості застосування базового алгоритму, побудованого у першій частині статті. Для систем з обмеженнями на керуючі функції використовується принцип поетапного планування методу динамічного програмування.

It is considered on concrete examples the peculiarity of use of the base algorithm, which was synthesized in the first part of the article. For system with restrictions of the control function it is utilized the principle of stage by stage planning of dynamic programming method.

КЕРУВАННЯ У ТЕХНІЧНИХ СИСТЕМАХ

УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ СИСТЕМАХ

CONTROL IN TECHNICAL SYSTEMS

УДК 681.51

Н. В. Ефименко, Н. В. Луценко

ПОСТРОЕНИЕ РЕГУЛЯРИЗОВАННЫХ ЭЛЛИпсоИДАЛЬНЫХ ОЦЕНОК В ЗАДАЧЕ ОПРЕДЕЛЕНИЯ ВЕКТОРА СОСТОЯНИЯ ЛИНЕЙНОЙ ДИСКРЕТНОЙ СИСТЕМЫ С МНОГОМЕРНЫМ ВЫХОДОМ

Предлагается новый класс алгоритмов эллипсоидальных оценок состояния динамических объектов, у которых вектор выхода измеряется с аддитивной помехой. Компоненты вектора помех считаются ограниченными по модулю заданными константами. В отличие от известных, предлагаемые алгоритмы нечувствительны к нарушениям априорных предположений о модели объекта управления.

ВВЕДЕНИЕ

В настоящее время существуют два подхода к решению задач оценивания состояния динамических систем, отличающиеся способами задания неопределенности в математических моделях объектов управления или, другими словами, априорных предположений об объекте управления. В рамках первого из них используются теоретико-вероятностные модели неопределенности, требующие знания статистических характеристик. В литературе этот подход получил название теории стохастической фильтрации. Начало этому подходу положил Р. Калман [1]. Теория линейной фильтрации, разработанная Р. Калманом, получила широкое развитие и признание. Эта теория в настоящее время широ-

ко применяется в ракетно-космической технике, авиации, в технической диагностике, биомедицине и многих других практических областях. Однако этот метод имеет один серьезный недостаток: неточности задания модели и статистических характеристик шумов могут привести к тому, что в реальных условиях работы уровень ошибок фильтра Калмана может существенно превышать уровень, предсказываемый теорией. Фактические ошибки в этом случае могут непрерывно возрастать, несмотря на то, что текущая оценка все время корректируется по вновь поступающим измерениям. Причиной расходимости является то, что в процессе фильтрации норма матрицы коэффициентов усиления фильтра уменьшается независимо от измерений и, начиная с некоторого момента, новые измерения практически не участвуют в формировании текущей оценки. Подробно вопросы расходимости алгоритмов типа фильтра Калмана и способы их коррекции, обеспечивающие получение нерасходящихся алгоритмов, описаны в [2].

При втором подходе [3–6], получившем название гарантированного, используются теоретико-множественные модели неопределенности, при которых оцениваемые

величины, возмущения и шумы в объекте задаются множествами своих возможных значений (множествами неопределенности). Решение задачи оценивания при этом сводится к построению некоторых множеств, гарантированно содержащих неизвестный вектор состояния объекта управления. Для аппроксимации множеств неопределенности во многих работах используются многомерные эллипсоиды, в связи с чем, этот подход получил название метода эллипсоидов. Достоинством алгоритмов оценивания фазового состояния динамического, объекта, получающихся при втором подходе, является следующее:

– во-первых, для их реализации не нужно знание статистических характеристик неопределенности объекта управления;

– во-вторых, что особенно важно, матрица коэффициентов усиления этих алгоритмов зависит от текущих измерений. В связи с этим алгоритмы этого типа нечувствительны к вычислительным погрешностям и нарушениям априорных предположений об объекте управления.

Обсуждение различия и сходства указанных подходов, некоторые общие соображения по их применению, основанные на степени полноты априорной информации об объекте, можно найти в [7, 8].

Кроме перечисленных подходов существуют и другие, среди которых необходимо выделить метод регуляризации. В его основе лежит разработанная академиком А. Н. Тихоновым математическая теория решения так называемых некорректно поставленных задач [9]. К сожалению, данный подход не нашел широкого применения в задаче оценивания вектора состояния динамической системы, несмотря на то, что, согласно этой теории, задача фильтрации поставлена некорректно и ее решение необходимо искать в классе регуляризирующих алгоритмов. На возможность построения таких алгоритмов впервые указывалось в [10]. Данная статья является развитием работы [11] и посвящена разработке и исследованию способов построения регуляризованных эллипсоидальных наблюдателей для линейной системы с многомерным выходом.

ПОСТАНОВКА ЗАДАЧИ

Рассматривается линейная полностью восстанавливаемая динамическая система, которая в пространстве состояний описывается разностным уравнением

$$x_{k+1} = A_k x_k, \quad k = 0, 1, 2, 3, \dots \quad (1)$$

Уравнение канала измерений имеет вид

$$y_k = C_k x_k + \xi_k. \quad (2)$$

В уравнениях (1), (2) k – целочисленный временной параметр, $x_k \in R^n$ – вектор состояния системы

в момент времени k , $y_k \in R^m$ – измеряемый выход объекта, $\xi_k \in R^m$ – погрешность измерений, A_k, C_k – матрицы соответствующих размерностей, $A_k \in R^{n \times n}$, $C_k \in R^{m \times n}$. Символами $R^n, R^m, R^{n \times n}$ и $R^{m \times n}$ обозначаются линейные нормированные пространства, элементами которых являются вещественные векторы и матрицы. Погрешность измерений ξ_k удовлетворяет условию

$$\xi_k^T \Sigma^{-1} \xi_k \leq \delta_k^2. \quad (3)$$

В выражении (3) матрица $\Sigma = \Sigma^T > 0$ и скаляр δ_k полагаются заданными. Ограничению (3) в R^n соответствует множество

$$Y_k = \{x: (y_k - C_k x)^T \Sigma^{-1} (y_k - C_k x) \leq \delta_k^2\}. \quad (4)$$

Требуется построить алгоритм типа наблюдателя для вычисления оценки $\hat{x}_k, k = 0, 1, 2, 3$ вектора состояния системы (1), (2) при ограничении (4).

МЕТОД РЕШЕНИЯ

Свяжем с оцениваемым вектором состояния x_k эллиптическое множество вида

$$E_k = \left\{ x: (x - \hat{x}_k)^T \hat{H}_k^{-1} (x - \hat{x}_k) \leq 1 \right\}, \quad (5)$$

где \hat{x}_k – центр эллипсоида, \hat{H}_k – симметрическая положительно определенная матрица.

Пусть найдены такие \hat{x}_k и \hat{H}_k , что $x_k \in E_k$. Найдем алгоритм построения эллипсоида E_{k+1} , такого, что $x_{k+1} \in E_{k+1}$. Рассмотрим образ \tilde{E}_{k+1} множества E_k при его отображении с помощью линейного преобразования (1). Несложно убедиться, что множество \tilde{E}_{k+1} представляет собой эллипсоид вида

$$\tilde{E}_{k+1} = \left\{ x: (x - \tilde{x}_{k+1})^T \tilde{H}_{k+1}^{-1} (x - \tilde{x}_{k+1}) \leq 1 \right\},$$

параметры которого определяются соотношениями

$$\tilde{x}_{k+1} = A_k \hat{x}_k, \quad \tilde{H}_{k+1} = A_k \hat{H}_k A_k^T.$$

Будем полагать, что пересечение

$$G_{k+1} = \tilde{E}_{k+1} \cap Y_{k+1} \quad (6)$$

не пусто, т. е. $G_{k+1} \neq \emptyset$.

При формировании множественной оценки \tilde{E}_{k+1} не учитывалась информация о векторе состояния, содержащаяся в измерении. После уточнения оценки по измерению y_{k+1} параметры эллипсоида \tilde{E}_{k+1} изменятся. Соответствующий теоретико-множественный алгоритм трансформации \tilde{E}_{k+1} в E_{k+1} должен удовлетворять условию

$$E_{k+1} \supset G_{k+1}. \quad (7)$$

Будем искать уравнение поверхности эллипсоида E_{k+1} в виде

$$M_{k+1}(x) = r_{k+1}^2, \quad (8)$$

где

$$M_{k+1}(x) = \Omega_{k+1}(x) + \lambda_{k+1}(y_{k+1} - C_{k+1}x)^T \times \\ \times \Sigma^{-1}(y_{k+1} - C_{k+1}x), \quad (9)$$

$$\Omega_{k+1}(x) = (x - \tilde{x}_{k+1})^T \tilde{H}_{k+1}^{-1} (x - \tilde{x}_{k+1}). \quad (10)$$

В выражениях (8)–(10) $\lambda_{k+1} > 0$ и $r_{k+1}^2 > 0$ – некоторые скалярные переменные, подлежащие определению. Так как ниже речь будет идти лишь о $k+1$ шаге, то для удобства изложения индекс $k+1$ в промежуточных выкладках будем опускать.

Уравнение (8) представляет собой общее уравнение поверхности второго порядка в R^n . Приведем его к канонической форме, получим

$$(x - \hat{x})^T \hat{H}^{-1} (x - \hat{x}) \leq 1.$$

Здесь

$$\hat{x} = \tilde{x} + \lambda K \tilde{y}, \quad (11)$$

$$\hat{H} = \rho(\tilde{H} - \lambda \tilde{H} C^T \Psi^{-1} C \tilde{H}), \quad (12)$$

$$K = \tilde{H} C^T \Psi^{-1}, \quad (13)$$

$$\tilde{y} = y - C \tilde{x}, \quad (14)$$

$$\Psi = \Sigma + \lambda C \tilde{H} C^T, \quad (15)$$

$$\rho = -\lambda \tilde{y}^T \Psi^{-1} \tilde{y} + r^2. \quad (16)$$

Необходимым и достаточным условием существования поверхности (8) является выполнение неравенства $r^2 > \min M(x)$. Можно показать, что \hat{x} есть точка минимума функционала $M(x)$ и этот минимум равен

$$\min_{x \in R^n} M(x) = \lambda \tilde{y}^T \Psi^{-1} \tilde{y}.$$

Следовательно, для любого $r^2 \geq \lambda \tilde{y}^T \Psi^{-1} \tilde{y}$ всегда можно построить эллипсоид (5), воспользовавшись соотношениями (11)–(16). Определим r^2 следующим образом:

$$r^2 = 1 + \lambda \tilde{y}^T \Psi^{-1} \tilde{y} + f(\tilde{y}, \lambda), \quad (17)$$

где $f(\tilde{y}, \lambda)$ – некоторая положительная функция, удовлетворяющая условию $f(\tilde{y}, \lambda) \geq \lambda \delta^2 \forall (\lambda > 0)$. Подставив (17) в (16), получим $\rho = 1 + f(\tilde{y}, \lambda)$.

Рассмотрим опорную функцию эллипсоида, содержащего множество G_{k+1} . Согласно работе [5] эта функция имеет вид

$$\delta(l/G_{k+1}) = l^T \hat{x}_{k+1} + \{\chi(l^T \tilde{H} l)\}^{\frac{1}{2}}, \quad (18)$$

где

$$\tilde{H} = \tilde{H} - \lambda \tilde{H} C^T \Psi^{-1} C \tilde{H},$$

$$\chi = 1 + \lambda(\delta^2 - \tilde{y}^T \Psi^{-1} \tilde{y}) \geq 0 \forall \lambda \geq 0.$$

Из (12) имеем

$$\tilde{H} = \frac{\hat{H}}{\rho}. \quad (19)$$

Подставив (19) в (18), получим

$$\delta(l/G_{k+1}) = l^T \hat{x}_{k+1} + \left(\frac{\chi}{\rho} (l^T \hat{H} l) \right)^{\frac{1}{2}}.$$

Несложно показать, что при $f(\tilde{y}, \lambda) \geq \lambda \delta^2$ параметр $\rho \geq \chi$ и, следовательно,

$$\delta(l/G_{k+1}) \leq l^T \hat{x}_{k+1} + (l^T \hat{H} l)^{\frac{1}{2}}. \quad (20)$$

Правая часть неравенства (20) есть опорная функция эллипсоида E_{k+1} . Из (20) и свойств опорных функций [12] вытекает, что $E_{k+1} \supset G_{k+1}$. Таким образом, при сделанных допущениях относительно функции $f(\tilde{y}, \lambda)$, справедливо следующее:

Утверждение. Если параметры семейства E_{k+1} определены таким образом, что

$$\hat{x}_{k+1} = \tilde{x}_{k+1} + \lambda_{k+1} K_{k+1} \tilde{y}_{k+1},$$

$$\hat{H}_{k+1} = \rho_{k+1} (\tilde{H}_{k+1} - \lambda_{k+1} K_{k+1} C_{k+1}^T \tilde{H}_{k+1}),$$

$$K_{k+1} = \tilde{H}_{k+1} C_{k+1}^T \Psi_{k+1}^{-1},$$

$$\tilde{y}_{k+1} = y_{k+1} - C_{k+1} \tilde{x}_{k+1},$$

$$\Psi_{k+1} = \Sigma + \lambda_{k+1} C_{k+1} \tilde{H}_{k+1} C_{k+1}^T,$$

$$\rho = 1 + f(\tilde{y}_{k+1}, \lambda_{k+1}),$$

$$f(\tilde{y}_{k+1}, \lambda_{k+1}) \geq \lambda_{k+1} \delta^2 \nabla \lambda_{k+1} > 0,$$

то семейство E_{k+1} удовлетворяет соотношению $E_{k+1} \supset \supset \tilde{E}_{k+1} \cap Y_{k+1}$.

Проиллюстрируем применение предложенного алгоритма на примере оценки вектора состояния космического аппарата (КА).

Математическая модель углового движения КА относительно центра масс имеет следующий вид:

$$J_x \ddot{\phi} + 4\omega_0^2 (J_y - J_z) \phi + (J_x - J_y + J_z) \omega_0 \dot{\psi} = M_u^x,$$

$$J_y \ddot{\theta} + 3\omega_0^2 (J_x - J_z) \theta = M_u^y,$$

$$J_z \ddot{\psi} + (J_y - J_x) \omega_0^2 \psi - (J_x - J_y + J_z) \omega_0 \dot{\phi} = M_u^z.$$

В этих уравнениях ϕ – угол крена, θ – угол тангажа, ψ – угол рыскания, ω_0 – угловая скорость движения КА по орбите, J_x, J_y, J_z – осевые моменты инерции КА, M_u^x, M_u^y, M_u^z – управляющие моменты соответственно в канале крена, тангажа и рыскания.

Будем полагать, что на КА имеется магнитометр, измеряющий проекции вектора напряженности магнитного поля Земли на оси связанной с КА системы координат. В качестве модели магнитометра примем следующую систему уравнений:

$$U_{Bx} = k_x (H_{ox} + (\psi + \delta_z) H_{oy} - (\theta + \delta_y) H_{oz}) + \Delta U_x;$$

$$U_{By} = k_y (H_{oy} + (\psi + \delta_z) H_{ox} - (\phi + \delta_x) H_{oz}) + \Delta U_y;$$

$$U_{Bz} = k_z (H_{oz} + (\theta + \delta_y) H_{ox} - (\phi + \delta_x) H_{oy}) + \Delta U_z,$$

где $U_B^T = (U_{Bx}, U_{By}, U_{Bz})$ – выходной сигнал магнитометра; $\Delta U^T = (\Delta U_x, \Delta U_y, \Delta U_z)$ – погрешность измерения магнитометра; $K^T = (k_x, k_y, k_z)$ – крутизна выходной характеристики магнитометра; $\delta^T = (\delta_x, \delta_y, \delta_z)$ – погрешность установки магнитометра относительно связанной системы координат; $H_0^T = (H_{0x}, H_{0y}, H_{0z})$ – теоретические значения проекций вектора напряженности магнитного поля Земли на оси орбитальной системы координат. Предполагается, что на борту имеется информация о положении КА на орбите и модель магнитного поля Земли, по которой можно рассчитать H_0 .

Численные значения параметров модели КА были следующие: $J_x = 97 \text{ Н}\cdot\text{м}\cdot\text{с}^2$, $J_y = 103 \text{ Н}\cdot\text{м}\cdot\text{с}^2$, $J_z =$

$= 2,6 \text{ Н}\cdot\text{м}\cdot\text{с}^2$, $\omega_0 = 0,001067 \text{ с}^{-1}$, $\delta = 0,0034$, $k_i = 1,02$ ($i = x, y, z$).

По непрерывной модели КА была построена дискретная модель КА и магнитометра в пространстве состояний

$$x_{k+1} = A_k x_k + B u_k, \quad y_{k+1} = C_{k+1} x_{k+1} + v_{k+1}.$$

Вектор погрешностей измерений магнитометра ΔU моделировался случайным процессом, имеющим нормальное распределение с математическим ожиданием, равным 200 нТл, и дисперсией, равной 200 нТл.

При построении наблюдателя использовались приближенные исходные данные $\tilde{J}_x = 100 \text{ Н}\cdot\text{м}\cdot\text{с}^2$, $\tilde{J}_y = 100 \text{ Н}\cdot\text{м}\cdot\text{с}^2$, $\tilde{J}_z = 2 \text{ Н}\cdot\text{м}\cdot\text{с}^2$, $\tilde{k}_i = 1$, ($i = x, y, z$). В качестве функции $f(\tilde{y}, \lambda)$ использовалась функция вида $f(\tilde{y}, \lambda) = \lambda \tilde{y}^T \Psi^{-1} y^T$, где λ определялась из уравнения обобщенной невязки $|y - C \hat{x}(\lambda)|^2 = \delta^2$ [9].

На рис. 1, 2, 3 показаны погрешности определения углов ориентации КА. Рис. 4 демонстрирует динамику

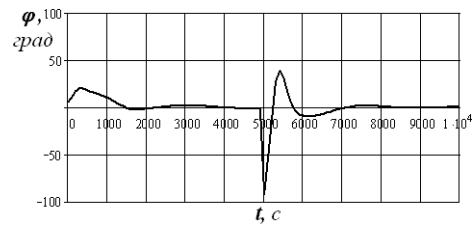


Рисунок 1 – Погрешность определения угла крена

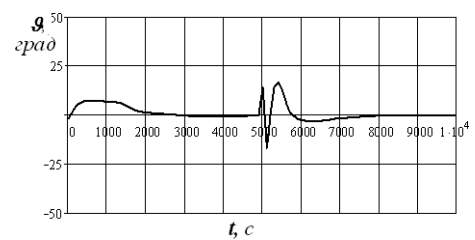


Рисунок 2 – Погрешность определения угла тангажа

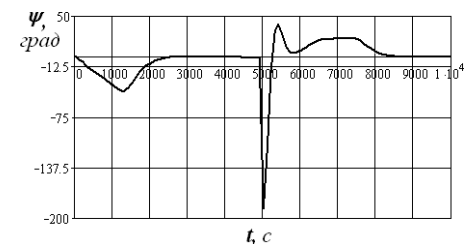


Рисунок 3 – Погрешность определения угла рыскания

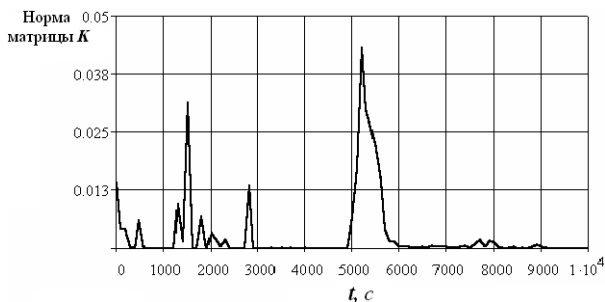


Рисунок 4 – Норма матрицы коэффициентов усиления наблюдателя

изменения нормы матрицы коэффициентов усиления наблюдателя.

Как видно из рис. 1–4, применение предложенного алгоритма наблюдения обеспечивает высокую точность определения текущих оценок и быструю сходимость многомерных эллипсоидов. На 5000 секунде моделировался сбой в работе эллипсоидального наблюдателя. В этот момент все компоненты вектора состояния наблюдателя умножались на коэффициент 100 и меняли знак на противоположный. Из рис. 4 также видно, что при этом происходит увеличение нормы матрицы коэффициентов усиления фильтра. Это объясняется тем, что в момент скачка эллипсоид $e_k^T \hat{H}_k e_k = 1$ теряет вектор состояния КА и начинает увеличиваться до тех пор, пока снова не захватит его. После этого происходит монотонное уменьшение объема.

Таким образом, предложенный алгоритм обладает свойством робастности, заключающимся в том, что он не чувствителен к сбоям в работе алгоритма оценивания.

ПЕРЕЧЕНЬ ССЫЛОК

1. Kalman R. A new approach to linear filtering and problem // Trans. ASME. Ser. D. – 1960. – 82, № 1. – Pp. 33–45.
2. Кузовков Н. Т., Карабанов С. В., Салычев О. С. Непрерывные и дискретные системы управления и методы идентификации. – М.: Машиностроение, 1978. – 222 с.

3. Бакан Г. М. Оптимизация алгоритмов теоретико-множественного оценивания // Труды международной конференции «Идентификация систем и задачи управления». – Москва, 2000. – С. 2173–2181.
4. Черноусько Ф. Л. Об оптимальном эллипсоидальном оценивании динамических систем, подверженных неопределенным возмущениям // Кибернетика и системный анализ. – 2002. – № 2. – С. 65–72.
5. Волосов В. В. Разработка и исследование робастных алгоритмов гарантированного эллипсоидального оценивания состояния многомерных линейных дискретных динамических систем. Часть I // Проблемы управления и информатика. – 1997. – № 4. – С. 31–43.
6. Schweppe F. C. Recursive state estimation; unknown but bounded errors and system input // IEEE Trans. Automat. Control. – 1968. – AC-13. – № 1. – Pp. 22–28.
7. Жуков А. А., Фурасов В. Д. Рекуррентное эллипсоидальное оценивание состояний дискретных систем // Развитие и применение методов функций Ляпунова / Под редакцией В. М. Матросова. – Новосибирск: Наука, 1992. – С. 220–224.
8. Овсеич А. И., Черноусько Ф. Л. Свойства оптимальных эллипсоидов, приближающих области достижимости с неопределенностями // Известия Академии наук. Теория и системы управления. – 2004. – № 4. – С. 8–18.
9. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. – М.: Наука, 1986. – 286 с.
10. Черноусько Ф. Л. Оценивание фазового состояния динамических систем. Метод эллипсоидов. – М.: Наука, 1988. – 320 с.
11. Ефименко Н. В., Новиков А. К. Регуляризованные эллипсоидальные наблюдатели и их применение в задаче определения ориентации космического аппарата // Проблемы управления и информатика. – 1998. – № 6. – С. 145–155.
12. Рокафелар Р. Выпуклый анализ. – М.: Мир, 1972. – 470 с.

Надійшла 14.03.06
Після доробки 22.05.06

Пропонується новий клас алгоритмів еліпсоїдального оцінювання стану динамічних об'єктів, у яких вектор виходу вимірюється з адитивними завадами. Компоненти вектора завад вважаються обмеженими по модулю заданими константами. На відміну від відомих, запропоновані алгоритми нечутливі до порушень априорних припущень щодо моделі об'єкта керування.

The new class of ellipsoidal algorithms is proposed for state estimates of dynamic objects, which output vector is measured with series-mode interference. The components of disturbance vector are supposed to be bounded by absolute values of given constants. In contrast to the known algorithms the proposed ones are insensitive to the violation of a priori assumptions on characteristics of uncertainties in object model.

АВТОРЫ НОМЕРА

Алипов Николай Васильевич

доктор технических наук, профессор
Харьковского национального университета
радиоэлектроники

Артюшенко Богдан Андреевич

аспирант Запорожского национального
технического университета

Баракат Халед

аспирант Донецкого национального
технического университета

Баркалов Александр Александрович

доктор технических наук, профессор
Донецкого национального технического
университета

Бодянский Евгений Владимирович

доктор технических наук, профессор
Харьковского национального университета
радиоэлектроники

Брайловский Владимир Васильевич

кандидат физико-математических наук,
доцент Черновецкого национального
университета

Ваде Гриби

аспирант Харьковского национального
университета радиоэлектроники

Ватаманюк Петр Петрович

кандидат физико-математических наук,
доцент Черновецкого национального
университета

Веревкин Леонид Леонидович

доцент Запорожской государственной
инженерной академии

Винокурова Елена Анатольевна

кандидат технических наук, старший
научный сотрудник Харьковского
национального университета
радиоэлектроники

Гоменюк Сергей Иванович

доктор технических наук, декан
Запорожского национального
университета

Гребенюк Сергей Николаевич

кандидат технических наук, доцент
Запорожского национального
университета

Грездов Глеб Геннадиевич

соискатель Института проблем модели-
рования в энергетике им. Г. Е. Пухова
НАН Украины, г. Киев

Грядун Виктор Иванович

старший преподаватель Запорожского
национального технического
университета

Гусятин Максим Владимирович

инженер Харьковского национального
университета радиоэлектроники

Дранкова Алла Олеговна

кандидат технических наук, доцент
Одесского национального политехни-
ческого университета

Дубровин Валерий Иванович

кандидат технических наук, доцент
Запорожского национального
технического университета

Ефименко Николай Владимирович

кандидат технических наук, начальник
отдела НПП «ХАРТРОН-КОНСАТ»,
г. Запорожье

Завгородний Андрей Юрьевич

ассистент Национального аэрокосмиче-
ского университета им. Н. Е. Жуковско-
го «Харьковский авиационный институт»

Заскоцкая Марина Владимировна

студентка Запорожского национального
технического университета

Ибрагим Карим Хакам

инженер-программист ОАО «Днепроспец-
сталь», г. Запорожье

Каминская Жанна Константиновна

инженер Запорожского национального
университета

Кардашук Владимир Сергеевич

ассистент Северодонецкого технологического института Восточнoукраинского национального университета имени Владимира Даля

Карпуков Леонид Матвеевич

доктор технических наук, профессор Запорожского национального технического университета

Киричевский Валерий Викторович

аспирант Запорожского национального университета

Киричевский Виктор Владимирович

доктор технических наук, заведующий кафедрой Запорожского национального университета

Киричевский Ростислав Викторович

кандидат технических наук, доцент Луганского национального педагогического университета имени Тараса Шевченко

Кораблев Николай Михайлович

кандидат технических наук, доцент Харьковского национального университета радиоэлектроники

Крапивной Александр Викторович

кандидат физико-математических наук, доцент Запорожского национального технического университета

Крищук Владимир Николаевич

кандидат технических наук, заведующий кафедрой Запорожского национального технического университета

Ктейман Хассан

аспирант Харьковского национального университета радиоэлектроники

Кулик Анатолий Степанович

доктор технических наук, заведующий кафедрой Национального аэрокосмического университета им. Н. Е. Жуковского «Харьковский авиационный институт»

Ламонова Наталья Сергеевна

кандидат технических наук, старший научный сотрудник Харьковского национального университета радиоэлектроники

Липчанский Алексей Иванович

кандидат технических наук, докторант Харьковского национального университета радиоэлектроники

Лихограй Василий Григорьевич

кандидат физико-математических наук, доцент Харьковского национального университета радиоэлектроники

Луценко Наталья Владимировна

старший преподаватель Запорожского национального технического университета

Макрушан Ирина Анатольевна

ассистент Харьковского национального университета радиоэлектроники

Мальчева Раиса Викторовна

кандидат технических наук, доцент Донецкого национального технического университета

Механна Сами

аспирант Харьковского национального университета радиоэлектроники

Мизерная Елена Леонидовна

аспирант Запорожского национального технического университета

Новиков Александр Кузьмич

кандидат технических наук, старший научный сотрудник НПП «Хартрон-КОНСАТ», г. Запорожье

Новотарский Михаил Анатольевич

кандидат технических наук, старший научный сотрудник Института математики НАН Украины, г. Киев

Оборжицкий Валерий Иванович

кандидат технических наук, доцент Национального университета «Львовская политехника»

Парфентий Александр Николаевич

аспирант Харьковского национального университета радиоэлектроники

Прокопенко Елена Владимировна

аспирант Запорожского национального университета

Решевская Екатерина Сергеевна

аспирант Запорожского национального университета

Рождественская Маргарита Григорьевна

кандидат физико-математических наук, доцент Черновецкого национального университета

Рязанцев Александр Иванович

кандидат технических наук, заведующий кафедрой Северодонецкого технологического института Восточноукраинского национального университета имени Владимира Даля

Сердюк Сергей Никитович

кандидат технических наук, доцент Запорожского национального технического университета

Скляр Александр Яковлевич

кандидат технических наук, доцент Харьковского национального университета радиоэлектроники

Стрельницкий Александр Александрович

младший научный сотрудник Харьковского национального университета радиоэлектроники

Стрельницкий Александр Евгеньевич

старший преподаватель Харьковского национального университета радиоэлектроники

Сыревич Евгения Ефимовна

аспирант Харьковского национального университета радиоэлектроники

Тархова Виктория Михайловна

аспирант Запорожского национального университета

Ткаченко Елена Викторовна

аспирант Одесского национального политехнического университета

Топалов Владислав Владимирович

инженер Украинской государственной академии связи им. А.С. Попова, г. Одесса

Фарафонов Алексей Юрьевич

ассистент Запорожского национального технического университета

Федорченко Евгений Николаевич

аспирант Запорожского национального технического университета

Хабис А. А. Зидат

аспирант Харьковского национального университета радиоэлектроники

Хаханов Владимир Иванович

доктор технических наук, профессор Харьковского национального университета радиоэлектроники

Хиль Михаил Иванович

кандидат технических наук, заведующий кафедрой Северодонецкого технологического института Восточноукраинского национального университета имени Владимира Даля

Цона Александр Иванович

кандидат технических наук, старший научный сотрудник Харьковского национального университета радиоэлектроники

Чумаченко Виталий Павлович

доктор физико-математических наук, заведующий кафедрой Запорожского национального технического университета

Швец Евгений Яковлевич

кандидат технических наук, первый проректор Запорожской государственной инженерной академии

Шило Галина Николаевна

кандидат технических наук, доцент Запорожского национального технического университета

Шокало Владимир Михайлович

доктор технических наук, заведующий кафедрой Харьковского национального университета радиоэлектроники

ДЛЯ СВЕДЕНИЯ АВТОРОВ

Журнал «Радиоэлектроника. Информатика. Управління» предназначен для публикации наиболее значимых научных и практических результатов исследований ученых высших учебных заведений и научных организаций.

Журнал включен в перечень научных изданий Украины, в которых могут публиковаться результаты диссертационных работ на соискание ученых степеней доктора и кандидата технических наук.

Статьи, опубликованные в журнале, реферируются в реферативных журналах и базах данных ВИНТИ (Россия) и «Джерело» (Украина). Рефераты статей размещаются на web-сайте Запорожского национального технического университета (<http://www.zntu.edu.ua/base/news2005/ric.htm>).

Журнал издается два раза в год и распространяется по подписке в Украине. Подписной индекс 22914.

Последовательность размещения материала статьи: индекс УДК, название статьи, инициалы и фамилия автора(ов), аннотация, текст, перечень ссылок. Рукопись статьи должна быть подписана всеми авторами.

Текст статьи

Принимаются статьи на русском, украинском и английском языках. Размер статьи до 0,5 авторского листа.

Структура статьи

Редакционная коллегия принимает к рассмотрению только научные статьи, которые содержат такие необходимые элементы: постановка проблемы в общем виде и ее связь с важными научными или практическими заданиями; анализ последних исследований и публикаций, в которых начато решение данной проблемы, и на которые опирается автор; выделение нерешенных ранее частей общей проблемы, которым посвящается статья; формулирование целей статьи (постановка задачи); изложение основного материала исследования с полным обоснованием полученных научных результатов, выводы по данному исследованию и перспективы дальнейших исследований в данном направлении.

Требования к оформлению

Статьи принимаются набранные в редакторе Word for Windows (v.6 и выше).

Параметры страницы:

- размер бумаги – А4 (210x297);
- ориентация – книжная;
- шрифт – Times New Roman Cyr, размер – 12 pt;
- междустрочный интервал – полуторный;
- верхнее поле – 20мм, нижнее – 20 мм, левое – 25 мм, правое – 15 мм.

Аннотация

Аннотация является обязательной и подается в начале статьи на русском, украинском (для граждан Ук-

раины) и английском языках. Объем аннотации не должен превышать 40 слов.

Рисунки

Рисунки помещаются в тексте и подаются в отдельных файлах (формат.TIF с разрешением не менее 200 dpi). Рисунки нумеруют и подписывают внизу.

Таблицы

Заголовки таблиц размещаются по центру.

Формулы

Формулы выполняются с помощью встроенного в Word for Windows редактора Microsoft Equation. Формулы нумеруют в круглых скобках справа:

$$Z(\Theta) = 101 \log \left(\frac{\bar{y}^2}{s^2} \right) \quad (3)$$

Формулы большого размера записываются в несколько строк.

Перечень ссылок

Перечень ссылок в конце рукописи подается языком оригинала и составляется в порядке упоминания ссылок в тексте и в соответствии со стандартом на библиографическое описание. Ссылки на литературу в тексте обозначаются цифрой в квадратных скобках.

В сведениях об авторах необходимо привести фамилию, имя, отчество (полностью), место работы, должность, ученую степень, рабочий и домашний адрес, телефоны, e-mail.

В редакцию журнала необходимо представить:

- 1) распечатанный текст статьи в 2-х экземплярах;
- 2) экспертное обоснование о возможности опубликования;
- 3) сведения об авторах;
- 4) сопроводительное письмо-ходатайство из организации, где была выполнена работа;
- 5) название статьи, фамилии, имена и отчества авторов на английском языке;
- 6) дискета 3,5' с текстом документа, файлами рисунков, данными о статье на английском языке и сведениями об авторах в формате Word for Windows (v.6 и выше). Файлы можно выслать электронной почтой в виде заархивированных (ZIP-архиватором) файлов.

Статьи, которые не отвечают указанным требованиям, не принимаются к рассмотрению.

Рукописи и дискеты не возвращаются.

Адрес редакции: 69063, г. Запорожье, ул. Жуковского, 64, ЗНТУ, редакция журнала "PIU"

Тел.:

(061) 7-698-2-96 – редакционно-издательский отдел

(061) 7-698-3-13 – главный редактор

(061) 7-698-2-96 – зам. главного редактора

Факс: (0612) 64-21-41

E-mail: rvv@zntu.edu.ua, TQM@zntu.edu.ua

Наукове видання

**Радіоелектроніка.
Інформатика.
Управління**

№ 2/2006

Науковий журнал

Головний редактор – д. т. н. Потапенко Є. М.
Заст. головного редактора – к. т. н. Дубровін В. І.

Комп'ютерне моделювання та верстання

Новікова К. О.

Рибалка І. С.

Оригінал-макет підготовлено у редакційно-видавничому відділі ЗНТУ

*Підписано до друку 11.12.2006. Формат 60x84/8.
Папір офс. Різогр. друк. Ум. друк. арк. 19,3.
Тираж 300 прим. Зам. № 1951.*

69063 м. Запоріжжя, ЗНТУ, друкарня, вул. Жуковського, 64