

ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ

4/2009

НАУЧНО - ТЕХНИЧЕСКИЙ ЖУРНАЛ

ЦОС в системах телекоммуникаций

тематический выпуск



ISSN 1684-2634



9 771684 263005 >

ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ

Научно-технический журнал
№ 4/2009

Издается с 1999 года
Выходит четыре раза в год

ГЛАВНЫЙ РЕДАКТОР
Ю.Б. ЗУБАРЕВ

ЗАМЕСТИТЕЛИ ГЛАВНОГО РЕДАКТОРА:
В.В. ВИТЯЗЕВ, В.П. ДВОРКОВИЧ

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

Б.А. Бабаян, Ю.И. Борисов,
С.А. Грибачев, Г.В. Зайцев, Р.В. Зубарев,
А.П. Кирпичников, М.И. Кривошеев,
Н.А. Кузнецов, М.С. Куприянов, А.А. Ланнэ,
В.Г. Мистюков, С.Л. Мищенко,
А.А.Петровский, Ю.Н. Прохоров,
А.Н. Соловьев, Ю.Г. Сосулин,
В.В. Шахильдян, Ю.С. Шинаков

Адрес редакции:

107031 Москва, Рождественка, 6/9/20, стр. 1
Тел.: (495) 621-71-08, (495) 621-16-39
Факс: (495) 621-16-39
E-mail: nto.popov@mtu-net.ru
tor@rgta.ryazan.ru
http://www.dspsa.ru

Для писем:

129090 Москва, а/я 48.

Издатель:

Российское научно-техническое общество
радиотехники, электроники и связи им. А.С. Попова
Компьютерная верстка: В.В. Андряхина
Дизайн: Н.С. Осипова

Подписной индекс по каталогу
ОАО «Роспечать» – **82185**

Подписано в печать 30.12.09
Формат 60x90/8.

Гарнитура «Arial». Печать офсетная.
Бумага офсетная. Печ.л. 6,5. Тираж 500 экз.

Заказ № 6213. Отпечатано в
ООО НПЦ «Информационные технологии»
Рязань, ул. Островского, д. 21/1
тел.: (4912) 98-69-84

Издание зарегистрировано в Министерстве
Российской Федерации по делам печати,
телерадиовещания и средств
массовых коммуникаций.
Свидетельство о регистрации ПИ № 77-1488
от 14.01.2000

УЧРЕДИТЕЛИ:

ЗАО «Инструментальные системы»
ФГУП «НИИ радио»

В НОМЕРЕ:

Шилин А.Н., Крутякова О.А.

Анализ методов синтеза цифровых фильтров по
аналоговым моделям 3

Никитин Д.А.

Приложения алгоритма синтеза рекурсивных цифровых
фильтров по импульсной характеристике 8

Безруков А.В.

Анализ характеристик двумерных цифровых фильтров с
ортогональным ограничением протяженности спектра
выходного сигнала 16

**Султанов Б.В., Шутков С.Л., Румянцева Н.Б.,
Зефиоров С.Л., Щербаков М.А.**

Применение алгоритма Левинсона при реализации
метода прямой оценки данных в радиомодемах 21

Тарасов Г.А., Тарасов А.Г.

Адаптивная пространственно-временная обработка
ФМ-сигналов в многолучевых каналах связи на основе
алгоритма Калмана 27

**Андреанов М.Н., Бумагин А.В., Гондарь А.В.,
Калашников К.С., Прудников А.А., Пучков Г.А.,
Стешенко В.Б.**

Повышение помехоустойчивости передачи данных по
цифровым каналам связи в условиях релеевских замираний .. 33

Головкин И.В., Назаров Л.Е.

Разработка и реализация алгоритмов итеративного
декодирования блоковых турбо-кодов 37

Кравченко А.Н.

Методы и аппаратура кодирования и декодирования
систематического нерегулярного кода повторения-
накопления (IRA) для DVB-S2 и DVB-T2 демодуляторов 41

Петровский А.А.

Перцептуальный кодер звука на базе быстрого вейвлет
преобразования с динамической трансформацией
частотно-временного плана 48

Манелис В.Б., Каюков И.В.

Адаптация пилот структуры и длины защитного
интервала OFDM сигнала к изменяющимся канальным
условиям 59

Бочаров К.Ю., Демидов В.М.

Метод визуализации изображений с широким диапазоном
яркостей 65

Подписной индекс по каталогу
ОАО «Роспечать» – **82185**



Уважаемые коллеги, дорогие читатели, подошел к концу юбилейный год нашего журнала ...

Ровно 10 лет назад, по итогам прошедшей 1-й Международной научно-технической конференции и выставки «Цифровая обработка и ее применение – DSPA'98» (Москва, 30 июня – 2 июля, 1998 г.), было принято решение об издании нового общероссийского научно-технического журнала «Цифровая обработка сигналов».

Идею издания нового журнала поддержали многие известные ученые и специалисты в области информатики, вычислительной техники, электроники, радиотехники и связи. В их числе академики РАН: Ю.В. Гуляев (ИРЭ РАН), Н.А. Кузнецов (ИППИ), И.А. Мизин (ИПИ РАН), А.И. Савин (ЦНПО «Комета»), член-корр. РАН Б.А. Бабаян (НИИ ВТ) и др., а также Министерство науки и технологий РФ и ряд ведущих российских промышленных предприятий, организаций и университетов. Инициаторами издания и учредителями журнала стали ФГУП «НИИ радио» и ЗАО «Инструментальные системы», при поддержке Отделения информатики, вычислительной техники и автоматизации РАН.

В своем обращении к читателям нового журнала академик РАН В.А. Котельников отметил, что: «... издание журнала является очень своевременным и актуальным, имеющим важное значение для ученых и специалистов в области теории и техники цифровой обработки сигналов». Особое удовлетворение началом издания в России профессионального журнала в области цифровых технологий реального времени выразила Президент Общества по обработке сигналов Международного института электротехники и радиоэлектроники (IEEE Signal Processing Society) профессор Лия Джеймисон (США): «Общество по обработке сигналов с большим интересом ожидает участия в своей работе членов Российского отделения и надеется на плодотворное сотрудничество с российскими учеными, занимающимися обработкой сигналов».

За прошедшие 10 лет было издано 37 номеров журнала. Основным источником опубликованных материалов – ежегодная научно-техническая конференция «Цифровая обработка сигналов и ее применение – DSPA». Публикация материалов поддерживалась по пяти основным разделам: теория и методы обработки сигналов и изображений, современные технологии ЦОС и их применение, схемотехника обработки сигналов и изображений, инструментальные средства проектирования и техническая реализация систем и устройств ЦОС, школа информационных технологий реального времени. Ежегодно формировались тематические выпуски, посвященные обработке и передаче изображений, обработке сигналов в системах телекоммуникаций и радиолокации.

В соответствии с решением ВАК Минобрнауки России журнал «Цифровая обработка сигналов» включен в Перечень ведущих рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени доктора и кандидата технических наук. И это не случайно - в состав редколлегии журнала входят один академик и три член-корр. РАН, четыре члена редколлегии являются членами Экспертного Совета ВАК РФ.

Большой вклад в подготовку материалов статей, их рецензирование и регулярный выпуск журнала внесли заместители Главного редактора В.В. Витязев и В.П. Дворкович, члены редколлегии: А.А. Ланнэ, А.А. Петровский, Ю.Н. Прохоров, Ю.Г. Сосулин, Ю.С. Шинаков.

Издание журнала было бы невозможным без постоянной финансовой поддержки со стороны его учредителя – ЗАО «Инструментальные системы» (Генеральный директор Р.В. Зубарев), а также участия ряда других спонсирующих фирм и организаций: ЗАО «АВТЭКС», НТЦ «Модуль», ЗАО «НИИР-КОМ», ЗАО «Скан Инжиниринг Телеком», НПП «Цифровые решения» и др.

Уверен, что журнал «Цифровая обработка сигналов» будет и в дальнейшем содействовать систематизации знаний в области современных методов и средств обработки сигналов и обогащению кругозора широкой аудитории отечественных разработчиков, инженеров, преподавателей вузов, аспирантов и студентов.

С пожеланиями творческих успехов и плодотворного сотрудничества!

Главный редактор, член-корр. РАН Ю.Б. ЗУБАРЕВ

УДК 681.513.2

АНАЛИЗ МЕТОДОВ СИНТЕЗА ЦИФРОВЫХ ФИЛЬТРОВ ПО АНАЛОГОВЫМ МОДЕЛЯМ

Шилин А. Н., заведующий кафедрой электротехники Волгоградского государственного технического университета, д.т.н., профессор, e-mail: eltech@vstu.ru.

Крутякова О. А., старший преподаватель кафедры электротехники Волгоградского государственного технического университета, e-mail: eltech@vstu.ru.

Ключевые слова: аналоговая модель, цифровой фильтр, метод синтеза, передаточная функция, z-форма, частотная характеристика, устойчивость, численное решение.

Введение.

Постановка задачи

В настоящее время в связи широким внедрением цифровой обработки информации в различных отраслях техники возникает задача синтеза цифровых устройств по аналоговым моделям – электрическим и электронным схемам или по их передаточным функциям. Это связано с тем, что в работающих схемах используется большое количество информации, накопленной в результате настройки и эксплуатации аналоговых устройств. В качестве теоретической основы численного моделирования аналоговых динамических систем в настоящее время широко используются методы z-форм [1-2], основанные на использовании аппарата z- преобразования. Метод z-форм позволяет привести аналоговую модель системы в области изображений к квазидискретной, а затем по изображению функции сравнительно просто определяется рекуррентное уравнение.

В методе z-форм используются различные виды аппроксимаций при разложении в ряд функции $z = e^{pT}$:

$$z = e^{pT} \approx 1 + pT \quad (\text{прямая разность или метод Эйлера}),$$

$$z = e^{pT} \approx 1/(1 - pT) \quad (\text{обратная разность}),$$

$$z = e^{pT} \approx \frac{1 + pT/2}{1 - pT/2} \quad (\text{метод трапеций или Тустена}).$$

Использование методов аппроксимации позволяет получить формулы для перехода от непрерывной передаточной функции $W(p)$ к дискретной передаточной функции $W(z)$, а именно:

$$p = \frac{z-1}{T}; p = \frac{z-1}{zT}; p = \frac{2}{T} \left[\frac{z-1}{z+1} \right].$$

Рассмотренные выше методы позволяют моделировать задачи с нулевыми начальными условиями, представленные в форме передаточных функций. Для моделирования систем, использующих в качестве моделей электрические схемы замещения с ненулевыми начальными условиями, разработан операторно-дискретный метод анализа (ОДМ)[3]. Для решения конкретных задач возникают вопросы выбора метода (z-форм или

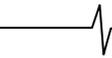
Приведен анализ методов синтеза цифровых фильтров по аналоговым моделям. В качестве моделей использованы передаточные функции фильтров. Для перехода от аналоговых моделей к импульсным моделям применены методы z-форм. Результаты анализа позволяют обоснованно выбирать метод по критериям устойчивости численного решения и минимального искажения частотной характеристики.

ОДМ) и параметров численных моделей. Необходимо отметить, что в литературных источниках приводятся рекомендации по применению методов моделирования, которые основаны на опытах их практического использования [1-2] и имеют общий характер. Поэтому для обоснованного выбора метода определения параметров численных моделей и учета ограничений на их применение необходимо проведение сравнительного анализа всех основных методов.

В предлагаемой статье приведены результаты исследования влияния основных параметров дифференциальных уравнений и значений периода дискретизации на точность и устойчивость моделирования, а именно: значений периода дискретизации, частоты и периода затухания динамического звена.

Конформные отображения

При моделировании с использованием различных методов z-форм основным вопросом является устойчивость численной модели, а именно каким образом область устойчивости из P -плоскости переходит в Z -плоскость. Хорошо известные преобразования выполняются либо численно, либо в прямоугольной системе координат [4-5]. Однако для анализа конформного отображения целесообразно использовать полярную систему координат, поскольку устойчивость характеризуется разностью между модулем z и радиусом единичной окружности. Известно, что при стремлении периода дискретизации T к нулю $|z| \rightarrow 1$, поэтому аналитическое представление в полярной системе координат позволяет получить выражение скорости приближения модуля z – полюсов к границе устойчивости при уменьшении периода дискретизации ($T \rightarrow 0$). Кроме того, в источнике [6] приводится оценка искажений частотных характеристик в цифровой форме. Эта оценка осуществляется визуально по неравномерности частотных сеток в Z -плоскости. Исследование в полярной системе координат позволяет аналитически определить частотные искажения для различных Z -форм.



Кроме известных методов аппроксимации первого порядка, таких как прямая разность, обратная разность и метод трапеций, из литературных источников известен метод преобразования без потерь [6]. Это выражение было получено и нами при выводе z -формы, эквивалентной операторно-дискретному методу, использующему комбинацию прямой и обратной разности, при равенстве нулю параметра затухания в передаточной функции звена второго порядка.

Рассмотрим соответствие между P и Z -плоскостями при преобразовании $z = e^{pT}$, а также использовании различных видов аппроксимации. Поскольку исследовать отображение областей полюсов из P -плоскости в Z -плоскость в операторно-дискретном методе невозможно, проведем подобное исследование для Z -преобразования без потерь.

Преобразование плоскости, осуществляемое аналитической функцией $z = f(p)$, обладает следующим важным свойством в окрестности точки p , для которой $z' \neq 0$. Бесконечно малые векторы всех направлений, выходящие из этой точки, увеличиваются (или уменьшаются) по своей длине в одно и то же число раз, равное $|z'|$, и поворачиваются на один и тот же угол, равный $\arg z'$. Таким образом, фигуры в бесконечно малой области преобразуются в себе подобные – сохраняют форму. Такое преобразование является конформным отображением. Фигуры конечных размеров искажаются, но углы между двумя кривыми сохраняются. Таким образом, с помощью аналитических функций можно получить множество прямоугольных систем криволинейных координат. Обратно, для любого конформного отображения существует некоторая ортогональная сетка кривых, которая преобразуется в прямоугольную декартову сетку [7].

Известно, что при дискретизации непрерывной функции с ограниченным спектром (ω_c – высшая гармоническая составляющая сигнала или частота среза) P -плоскость делится на бесконечное число полос. Определим предельные соотношения для основной полосы в области полюсов.

Пусть переменная в P -плоскости имеет выражение $p = \delta + j\omega$, и поскольку для устойчивой системы $\delta < 0$, рассмотрим левую полуплоскость. Частота и период дискретизации определяется с помощью теоремы Котельникова, а именно $\omega = 2\omega_c$ или $T = \pi/\omega_c$. По-

этому основная полоса расположена в диапазоне частот от $-\pi/T$ до π/T . Поскольку период дискретизации T гораздо меньше постоянной времени затухания $\tau = 1/|\delta|$, то $|\delta|$ не может быть больше $1/T$. Таким образом, в основной полосе область полюсов находится в прямоугольнике, ограниченном прямыми $\delta = 0$, $\delta = -1/T$, $\omega = \pm\pi/T$ (рис.1, а). Область p -полюсов отображается преобразованием $z = e^{pT}$ в круг с центром в начале координат и единичным радиусом, как показано на рис.1, б. Точки правой половины P -плоскости отображаются в область вне окружности единичного радиуса на Z -плоскости. Эта область будет областью неустойчивости.

При неограниченном увеличении затухания $\delta \rightarrow -\infty$ устойчивая область полюсов преобразуется в единичный круг.

Проводя исследование преобразования P -плоскости на Z -плоскость с помощью различных видов аппроксимации, введем новые обозначения: $\delta T = \delta'$ – приведенный коэффициент затухания, $\omega/\omega_c = \omega'$ – приведенная частота. При этом δ' принадлежит интервалу $[-1; 0]$, а $\omega' \in [-1; 1]$.

Метод Эйлера. Используя формулу перехода $p = (z-1)/T$, выразим z -переменную через p -переменную

$$z = 1 + \delta T + j\omega T. \quad (1)$$

Пусть $z = x + jy$, тогда из (1) с учетом новых обозначений следует:

$$x = 1 + \delta T = 1 + \delta', \quad y = \omega T = \pi\omega'; \quad (2)$$

Найдем модуль и аргумент z :

$$|z| = \sqrt{(\delta' + 1)^2 + (\pi\omega')^2}, \quad \varphi = \arctg \frac{\pi\omega'}{\delta' + 1}. \quad (3)$$

Определим отображение области p -полюсов на Z -плоскость в полярных координатах (рис. 2, а). Из отображения следует, что p -область по методу Эйлера преобразуется в прямоугольник, который выходит за пределы единичного круга и ограничен в пределах $0 \leq x \leq 1$, $-\pi \leq y \leq \pi$. Следовательно, при некоторых параметрах численной модели возможно неустойчивое решение. При преобразовании прямые линии, параллельные осям координат, в плоскости P переходят в прямые линии на плоскости Z .

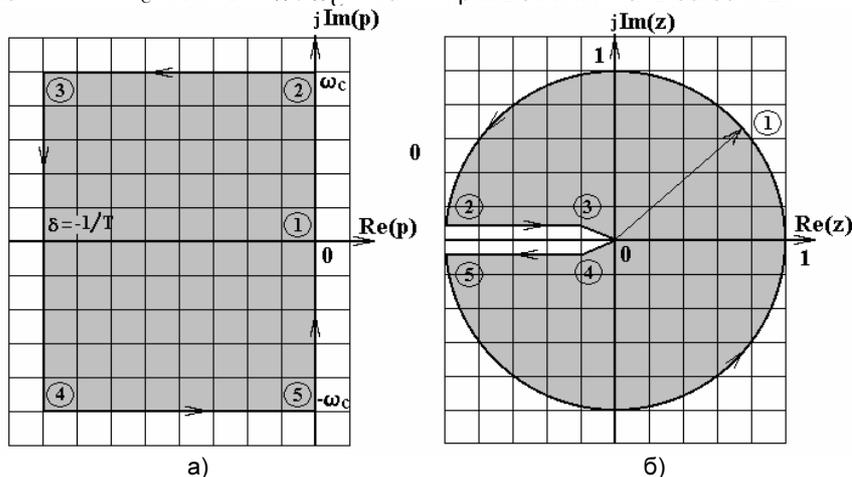


Рис. 1. Отображение области устойчивости p -плоскости на z -плоскость.

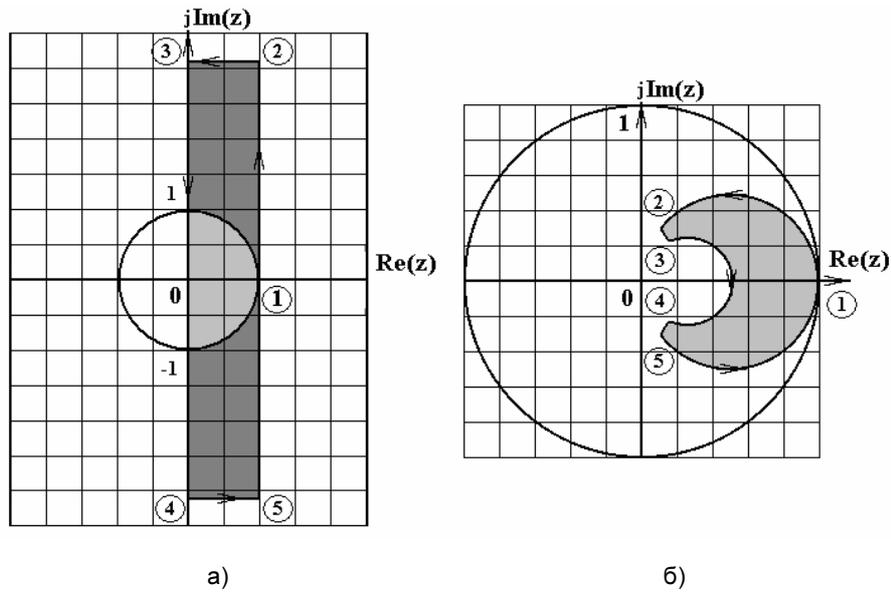


Рис. 2. Отображение области устойчивости p -плоскости на z -плоскость с помощью: а) метода Эйлера, б) метода обратных разностей.

Метод обратных разностей. Используя формулу перехода $p = (z-1)/(zT)$, выразим z -переменную через p -переменную

$$z = \frac{1 - \delta T}{(1 - \delta T)^2 + (\omega T)^2} + j \frac{\omega T}{(1 - \delta T)^2 + (\omega T)^2}, \quad (4)$$

или

$$x = \frac{1 - \delta T}{(1 - \delta T)^2 + (\omega T)^2}, \quad y = \frac{\omega T}{(1 - \delta T)^2 + (\omega T)^2}. \quad (5)$$

С учетом обозначений (2), получим:

$$x = \frac{1 - \delta'}{(1 - \delta')^2 + (\pi\omega')^2}, \quad y = \frac{\pi\omega'}{(1 - \delta')^2 + (\pi\omega')^2}. \quad (6)$$

Определим модуль и аргумент z :

$$|z| = \frac{1}{\sqrt{(1 - \delta')^2 + (\pi\omega')^2}}, \quad \varphi = \arctg \frac{\pi\omega'}{1 - \delta'}. \quad (7)$$

Отображение по формулам (7) области полюсов в P -плоскости на Z -плоскость показано на рис. 2, б, согласно которому при всех параметрах численное решение устойчиво, причем с некоторым запасом. На рисунке видно, что рассматриваемое преобразование отображает левую полуплоскость переменной p в круг половинного радиуса с центром в точке $\text{Re}(z) = 1/2$.

Если один из полюсов системы расположен близко к точке (0,0) в P -области, то необходимо учитывать погрешность нахождения этого полюса, чтобы система не вышла из области устойчивости при применении данного метода. Это единственная точка, находящаяся на границе области устойчивости при данном преобразовании.

Метод Тустена. Используя формулу перехода $p = 2(z-1)/((z+1)T)$, выразим z -переменную через p -переменную

$$z = \frac{4 - (\delta T)^2 - (\omega T)^2}{(\delta T - 2)^2 + (\omega T)^2} + j \frac{4\omega T}{(\delta T - 2)^2 + (\omega T)^2}, \quad (8)$$

где $x = \frac{4 - (\delta T)^2 - (\omega T)^2}{(\delta T - 2)^2 + (\omega T)^2}$,

$$y = \frac{4\omega T}{(\delta T - 2)^2 + (\omega T)^2}. \quad (9)$$

Тогда с учетом новых переменных (2) получим:

$$x = \frac{4 - \delta'^2 - \pi^2\omega'^2}{(\delta' - 2)^2 + (\pi\omega')^2}, \quad y = \frac{4\pi\omega'}{(\delta' - 2)^2 + (\pi\omega')^2}. \quad (10)$$

Определим модуль и аргумент z :

$$|z| = \frac{\sqrt{(4 - \delta'^2 - \pi^2\omega'^2)^2 + 16\pi^2\omega'^2}}{(\delta' - 2)^2 + (\pi\omega')^2}, \quad \varphi = \begin{cases} \arctg\left(\frac{4\pi\omega'}{4 - \delta'^2 - \pi^2\omega'^2}\right), & \text{если } 4 - \delta'^2 - \pi^2\omega'^2 \geq 0 \\ \arctg\left(\frac{4\pi\omega'}{4 - \delta'^2 - \pi^2\omega'^2}\right) + \pi, & \text{если } 4 - \delta'^2 - \pi^2\omega'^2 < 0 \end{cases}. \quad (11)$$

С помощью формул (11) выполним отображение P -области в Z -область (рис.2, в). Как видно из рисунка, отображением P -области на Z -плоскость является часть области единичного круга, и поэтому при всех параметрах численное решение устойчиво, однако в сравнении с предыдущим методом имеется меньший запас по устойчивости. При данном преобразовании границей устойчивости является отрезок 1 - 5 в P -плоскости ($\delta = 0$), который преобразуется в часть единичной окружности.

Дискретное интегральное преобразование без потерь. Данное преобразование также является алгебраическим преобразованием плоскости комплексной переменной P на плоскость Z при подстановке:

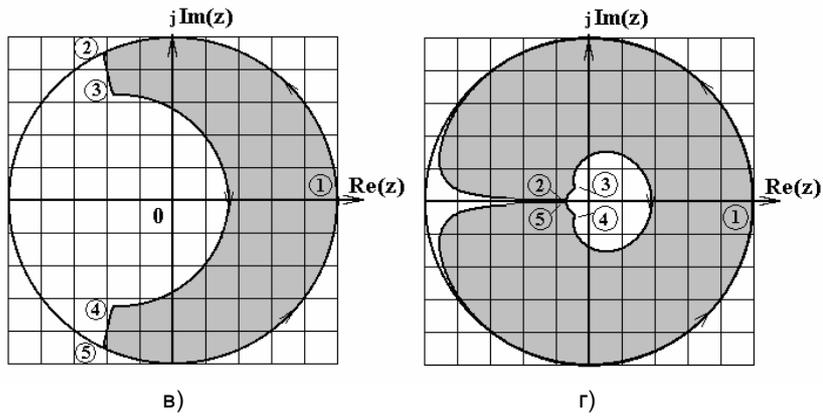


Рис. 2. Отображение области устойчивости p -плоскости на z -плоскость с помощью: в) метод Тустена, г) метод преобразования без потерь.

$$p = \frac{1 - z^{-1}}{z^{-1/2}T}$$

Выразим z -переменную через p -переменную, зная, что $p = \delta + j\omega$:

$$z_{12} = 1 + \frac{pT}{2} (pT \pm \sqrt{p^2 T^2 + 4}) \quad (12)$$

Выполним преобразование выражения (12), используя показательную форму записи комплексных чисел и введя новые переменные (2). Получим

$$z_{12} = \frac{1}{2} (\delta'^2 - \pi^2 \omega'^2 + 2 \pm P\sqrt{D} \cos \gamma + j(2\delta'\pi\omega' \pm P\sqrt{D} \sin \gamma)) \quad (13)$$

где

$$P = \sqrt{\delta'^2 + \pi^2 \omega'^2}, \quad D = \sqrt{(\delta'^2 - \pi^2 \omega'^2 + 4)^2 + (2\delta'\pi\omega')^2}$$

$$\gamma = \alpha + \frac{1}{2}\beta, \quad \alpha = \arctg\left(\frac{\pi\omega'}{\delta'}\right) + \pi,$$

$$\beta = \begin{cases} \arctg\left(\frac{2\delta'\pi\omega'}{\delta'^2 - \pi^2 \omega'^2 + 4}\right), & \text{если } \delta'^2 - \pi^2 \omega'^2 + 4 \geq 0 \\ \arctg\left(\frac{2\delta'\pi\omega'}{\delta'^2 - \pi^2 \omega'^2 + 4}\right) - \pi, & \text{если } \delta'^2 - \pi^2 \omega'^2 + 4 < 0 \end{cases} \quad (14)$$

Определив модуль и аргумент z_1 , найдем отображение P -области в Z -область (рис. 2, г). Аналогичное исследование проведено для z_2 , но в этом случае полученное отображение значительно превышает область круга единичного радиуса на Z -плоскости.

Определим чувствительность методов z -форм на границе устойчивости:

$$\Delta|z| \approx \frac{d|z|}{dT} \Delta T$$

Определим $\lim_{T \rightarrow 0} \left(\frac{d|z|}{dT}\right)$ для каждого метода:

Прямая разность: $\lim_{T \rightarrow 0} \left(\frac{2\delta(T\delta+1) + 2T\omega^2}{2\sqrt{(T\delta+1)^2 + T^2\omega^2}}\right) = \delta,$

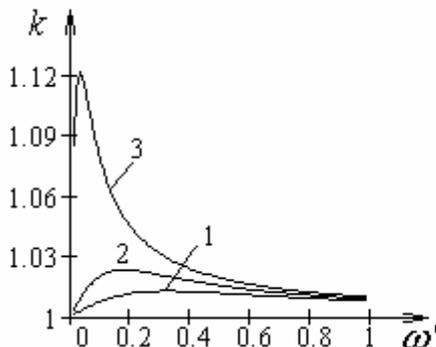
Обратная разность: $\lim_{T \rightarrow 0} \left(\frac{\delta(1-T\delta)}{\sqrt{((1-T\delta)^2 + T^2\omega^2)^3}}\right) = \delta,$

Метод трапеций:

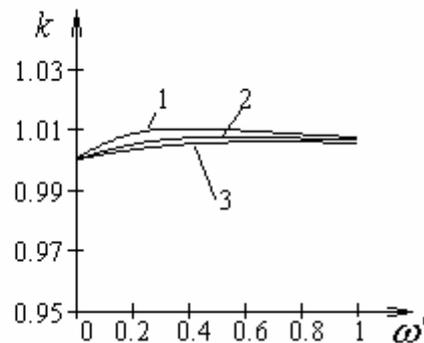
$$\lim_{T \rightarrow 0} \left(\frac{-4\delta(T^2\delta^2 + T^2\omega^2 - 4)}{(T^2\delta^2 - 4\delta T + T^2\omega^2 + 4)\sqrt{16 - 8\delta^2 T^2 + 8T^2\omega^2 + \delta^4 T^4 + 2\delta^2\omega^2 T^4 + T^4\omega^4}}\right) = \delta$$

В связи с громоздкостью выражения (13) аналитическое определение предела затруднено, и поэтому, в случае необходимости, для конкретной задачи определяется его численное значение с помощью известных пакетов математических программ.

Таким образом, при использовании z -форм только в методе прямой разности могут возникнуть случаи, когда устойчивая система может перейти в область неустойчивости. В остальных случаях преобразование из P -области в Z -область происходит с некоторым запасом устойчивости.



а)



б)

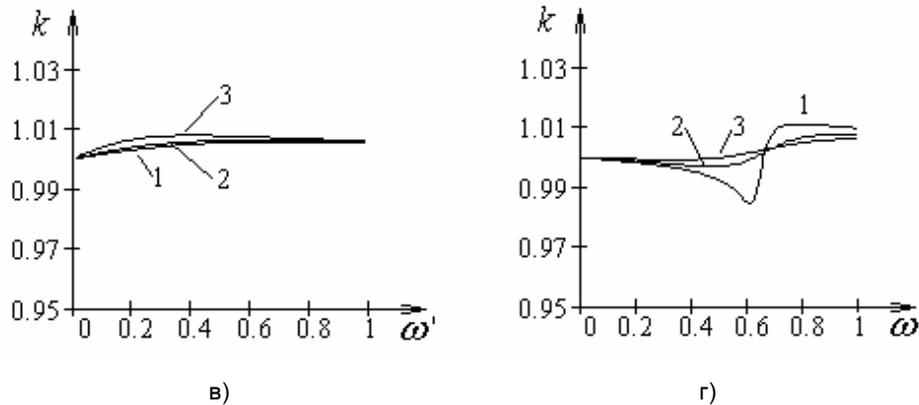


Рис. 3. Зависимости коэффициента линейности от приведенной частоты для следующих методов аппроксимации: а) прямая разности, б) обратная разность, в) метод трапеций, г) метод преобразования без потерь.

Коэффициент линейности

Известно, что при равномерной дискретизации частоты в P – области при точном преобразовании $z = e^{pT}$ осуществляется равномерная круговая дискретизация по окружности единичного радиуса. При использовании различных методов z -форм нарушается равномерность круговой дискретизации, поэтому проведем количественную оценку искажений полос частот для каждой z -формы, определив коэффициент линейности по формуле:

$$k = \frac{\varphi(\omega')/\varphi(\omega' + \Delta\omega)}{\omega'/(\omega' + \Delta\omega)}, \quad (13)$$

где $\Delta\omega$ – шаг изменения ω' , в данном случае $\omega' = 0,01$.

На рис. 3 представлены графические зависимости коэффициента линейности от приведенной частоты для следующих видов аппроксимации: а) прямая разность, б) обратная разность, в) метод трапеций, г) метод преобразования без потерь, причем номеру графика соответствуют следующие значения приведенного коэффициента затухания: 1 - $\delta_1 = -0,1$, 2 - $\delta_2 = -0,5$, 3 - $\delta_3 = -0,9$.

Заключение

Из проведенного анализа следует, что наименьшие искажения распределения частот при отображении области p – полюсов на Z - плоскость вносят метод обратной разности и метод трапеций.

Литература

1. Куо, Б. Теория и проектирование цифровых систем управления: Пер. с англ. - М.: Машиностроение, 1986.

2. Шипилло, В.П. Операторно-рекуррентный анализ электрических цепей и систем. - М.: Энергоатомиздат, 1991.
3. Шилин, А. Н. Операторно-дискретный метод анализа электрических цепей / Приборы и системы. Управление, контроль, диагностика. - 2000. - № 7. - С. 50-56.
4. Шилин, А. Н. Анализ точности моделирования аналоговых САУ / А. Н. Шилин, К. Е. Кострюков - Приборы и системы. Управление, контроль, диагностика. - 2006. - № 7. - с. 18-21.
5. Шилин, А. Н. Исследование методов цифрового моделирования аналоговых САУ / А. Н. Шилин, Е. Г. Зенина, С. А. Бедкин – Приборы и системы. Управление, контроль, диагностика. - 2001. - № 7. - с. 46-50.
6. Мулявка, Я. Схемы на операционных усилителях с переключаемыми конденсаторами: Пер. с польск. - М.: Мир, 1992. – 416 с.
7. Бронштейн, И. Н. Справочник по математике / И. Н. Бронштейн, К. А. Семендяев - М.: Наука, 1964. – 608 с.

THE ANALYSIS OF METHODS OF DIGITAL FILTERS SYNTHESIS ON ANALOG MODELS

A. N. Shilin, O. A. Krutyakova

The article gives the analysis of methods of digital filters synthesis on analog models. As models transfer functions of filters are used. Methods of Z – forms are applied to transition from analog models to pulse models. Results of the analysis allow to choose a method by criteria of stability of the numerical decision and the minimal distortion of the frequency characteristic.

ПРИЛОЖЕНИЯ АЛГОРИТМА СИНТЕЗА РЕКУРСИВНЫХ ЦИФРОВЫХ ФИЛЬТРОВ ПО ИМПУЛЬСНОЙ ХАРАКТЕРИСТИКЕ

*Никитин Д.А., старший преподаватель кафедры безопасности информационных технологий СибГАУ,
e-mail: ksili57@gmail.com.*

Ключевые слова: алгоритм синтеза, цифровой фильтр, импульсная характеристика, функциональная зависимость, обнаружение сигналов, сжатие данных, генерация массивов.

Введение

В представленной ранее работе [1] описан алгоритм синтеза БИХ-фильтров по импульсной характеристике, работоспособный при условии, что она определяется элементарной математической функцией или линейной комбинацией таких функций. Предложенный алгоритм был назван А1. Алгоритм построен так, что сначала находит минимально возможный порядок фильтра для заданной импульсной характеристики, затем строит систему линейных алгебраических уравнений соответствующего порядка, решение этой СЛАУ дает коэффициенты искомого ЦФ. В [1] даны предпосылки для разработки такого алгоритма, приведена его блок-схема, дан пример работы алгоритма, описано множество функций, для которых алгоритм всегда находит решение и, как пример, приведено доказательство существования и единственности решения для одной функции из этого множества. Функции, входящие в это множество названы функциями известными алгоритму. Здесь для краткости будем называть его множеством Φ . В данной статье будут даны дополнения, поясняющие алгоритм и расширяющие область его применения, а также описаны возможные приложения алгоритма, о которых в вышеупомянутой работе было сказано совсем мало.

Алгоритм А1 успешно работает в случае, когда исходная импульсная характеристика, по которой следует отыскивать цифровой фильтр, определяется любой функцией из множества Φ . «Определяется» означает, что отсчеты импульсной характеристики H являются значениями такой функции f , взятыми с постоянным шагом: $h_i = f(i)$, $i = 0, 1, 2, \dots, n$. Хотя импульсная характеристика БИХ-фильтра бесконечна, алгоритму достаточно лишь некоторое количество её начальных отсчетов, которое зависит от сложности функции f . Грубым критерием сложности функции в данном случае может считаться число параметров, требуемых для ее описания. Найденный фильтр называется фильтром, соответствующим исходной последовательности.

Множество Φ известных алгоритму функций было описано следующим образом:

- полиномиальные функции, $L_{min} = 2(n+1)$, где n – порядок полинома;
- показательные функции, $L_{min} = 4$;
- синусоиды (косинусоиды), $L_{min} = 6$;
- произвольные периодические функции, при условии, что период функции кратен частоте дискретизации, $L_{min} = 2T$, где T – количество отсчетов на период;

Рассматривается использование ранее предложенного алгоритма синтеза БИХ-фильтров для поиска как точных, так и приближённых функциональных зависимостей, обнаружения сигналов, сжатия данных, упрощения математических выражений, быстрой генерации массива значений заданной функции. Приведены вероятные пути развития алгоритма.

- любая линейная комбинация (конечная сумма) указанных функций (примеры L_{min} даны в [1]), где L_{min} — минимально необходимое для расчёта количества начальных отсчетов импульсной характеристики.

Поскольку любая числовая последовательность, полученная табуляцией любой из этих функций (кроме показательных с отрицательным показателем в степени), не будет являться сходящейся к нулю последовательностью, то фильтр с такой импульсной характеристикой не будет устойчивым. Кроме того, цифровые фильтры обычно рассчитывают, исходя из другой информации. Поэтому о практическом использовании алгоритма именно в приложениях цифровой фильтрации пока говорить сложно. Однако алгоритм может найти применение в приложениях, не требующих устойчивости отыскиваемых фильтров. Ниже будет описано несколько таких приложений, в которых алгоритм отыскания БИХ-фильтра, соответствующего некоторой числовой последовательности, является лишь одним из первых шагов методики.

Поиск точных функциональных зависимостей в числовых последовательностях

Задача ставится следующим образом. Имеется числовая последовательность. Необходимо обнаружить все участки, которые точно описываются любой функцией из некоторого множества функций $\Phi' \subseteq \Phi$, и установить такую функцию с точностью до свободного члена для каждого из участков.

Предполагается, что система поиска обладает некоторой базой знаний о функциональных зависимостях, которые она может обнаруживать. Такая база знаний должна содержать:

- информацию, необходимую для идентификации определённой зависимости;
- правила для нахождения значений параметров функции из значений коэффициентов соответствующего фильтра.

В качестве идентифицирующей информации для «простых» функций выступают точные значения коэффициентов обратной связи соответствующего фильтра, а для «сложных» функций — значения и соотношения между коэффициентами ЦФ. Для всех функций множества Φ (кроме последнего пункта) автором выполнено доказательство существования БИХ-фильтра соответствующего порядка. Примеры конкретных значений коэффициентов ЦФ для некоторых функций приведены в табл. 1 и [1, 2].

Таблица 1.

Коэффициенты обратной связи ЦФ, соответствующие некоторым функциям

Зависимость	Формула	Коэффициенты обратной связи ЦФ
Линейная	$y_i = s_0 \cdot i + s_1$	$[-2 \ 1]$
Квадратичная	$y_i = s_0 \cdot i^2 + s_1 \cdot i + s_2$	$[-3 \ 3 \ -1]$
Кубическая	$y_i = \sum_{k=0}^3 s_k \cdot i^k$	$[-4 \ 6 \ -4 \ 1]$
Полином 4-й степени	$y_i = \sum_{k=0}^4 s_k \cdot i^k$	$[-5 \ 10 \ -10 \ 5 \ -1]$
Полином 5-й степени	$y_i = \sum_{k=0}^5 s_k \cdot i^k$	$[-6 \ 15 \ -20 \ 15 \ -6 \ 1]$
Периодическая, с периодом T отсчётов	произвольная	$\underbrace{[0 \ 0 \ \dots \ 0 \ 0 \ -1]}_{T-1 \text{ нулей}}$

Для обнаружения показательной и синусоидальной функций используются следующие правила (a_i – коэффициенты обратной связи соответствующего ЦФ):

$$\begin{cases} a_2 > 0 \\ a_1 = -a_2 - 1 \end{cases} \quad \text{для показательной функции;}$$

$$\begin{cases} 0 < |a_2| < 3 \\ a_1 = -a_2 \\ a_3 = -1 \end{cases} \quad \text{для синусоидальной функции.}$$

На первом этапе методики поиска выполняется алгоритм А1. То есть для начала исходной последовательности выполняется синтез соответствующего фильтра. Поскольку методика может использоваться для поиска разных функциональных зависимостей, им могут соответствовать фильтры разных порядков. Таким образом, для того, чтобы проверить в последовательности наличие всех функций из базы знаний, надо в худшем случае совершить N попыток синтеза БИХ-фильтра, где N — количество разных порядков ЦФ, соответствующих функциям из базы знаний. Для нахождения каждого такого фильтра необходимо $2k$ первых элементов исходной последовательности, где k — порядок фильтра. Во избежание этого алгоритм А1, построен так, что непосредственное вычисление БИХ-фильтра каждого порядка не выполняется. Сначала находится минимальный порядок ЦФ, соответствующего данной последовательности, при котором он является единственным, а затем вычисляются коэффициенты БИХ-фильтра.

Процесс перебора порядков фильтров можно реализовать по-разному. В своих реализациях автор использовал перебор от больших порядков к меньшим, т.к. фильтры более высоких порядков позволяют обнаружить более сложные функции, охватывая при этом на этапе вычисления коэффициентов ЦФ большее количество элементов исходной последовательности. Может показаться, что такой перебор не совсем рационален, так как согласно ему первыми бу-

дут выполняться вычислительно более сложные действия (решения систем уравнений больших порядков), и если на самом деле элементы последовательности подчиняются одной из наиболее простых функций, либо вообще никакой не подчиняются, то большое количество вычислений будет произведено впустую. Однако процесс спуска к меньшим порядкам удалось оптимизировать, что позволило не перебирать все порядки фильтров. Кроме того, при переборе порядков от высших к низшим, при формировании каждой следующей системы уравнений, не используются новые элементы исходной последовательности (используется только их часть на текущем шаге). Поэтому — с учётом архитектуры большинства существующих вычислительных систем — можно утверждать, что с большой долей вероятности все данные, требуемые для формирования новой СЛАУ, уже будут в кэш-памяти. А значит, матрицы всех систем уравнений, кроме первой, будут формироваться очень быстро. Разумеется, возможны различные способы оптимизации, зависящие от специфики конкретного приложения. Например, если известно, что в анализируемых данных будет преобладать определённая функциональная зависимость, то имеет смысл всегда начинать поиск с вычисления фильтра того порядка, который соответствует этой функции.

После вычисления фильтра некоторого порядка k нужно по его коэффициентам проверить наличие всех функций, которым соответствует фильтр такого порядка. Если для текущего порядка фильтра проверка не дала положительного результата (то есть полученные коэффициенты не соответствуют ни одному правилу из базы знаний), следует проверить фильтры меньших порядков.

Если в результате проверки оказалось, что начальные элементы последовательности не подчиняются ни одной функциональной зависимости из базы знаний, надо отбросить первый рассматриваемый элемент и повторить проверку для оставшейся части последовательности.



Когда найден БИХ-фильтр, коэффициенты которого говорят о том, что элементы последовательности, используемые при его вычислении, описываются функцией известной алгоритму, следует определить, на сколько следующих членов последовательности распространяется найденная функциональная зависимость. Суть такой проверки сводится к вычислению импульсной характеристики найденного фильтра и её поэлементному сравнению с рассматриваемой последовательностью. Первые $2k$ элементов совпадают по определению, так как они использовались при вычислении ЦФ как первые отсчёты импульсной характеристики.

После того, как длина участка последовательности, подчиняющегося найденной зависимости, определена, можно сохранить всю полученную информацию об этом участке и повторить проверку для оставшейся части последовательности. В качестве сохраняемой информации могут выступать индекс первого элемента участка, длина участка (количество членов последовательности), идентификатор типа найденной функции, коэффициенты соответствующего фильтра. Коэффициенты фильтра нужны для вычисления значений параметров функции, поэтому вместо них можно хранить сами параметры.

Таким образом, после подобного анализа будут известны участки исходной последовательности, в точности подчиняющиеся полностью определённым функциям одного переменного.

Стремление найти ЦФ минимального порядка обусловлено желанием найти функцию с наименьшим числом параметров, описывающую элементы последовательности. При этом поиск не предлагается начинать с меньших порядков ЦФ из-за вероятности обнаружения более простых функций, которые будут действовать на коротких участках исходной последовательности, и при этом они не будут давать возможность обнаружить «более сложную» функцию, действующую на более длинном участке. Например, вместо длинного участка некоторой периодической функции будет обнаружено много коротких участков, подчиняющихся различным показательным или полиномиальным зависимостям. Если же никакой «сложной» функции на данном участке нет, а действительно имеются только более простые функции (т.е. с меньшим числом параметров), то при спуске от больших порядков ЦФ к меньшим они всё равно будут найдены.

Важным свойством методики является возможность изменения множества искомых функций. Таким образом, можно модифицировать методику для конкретной решаемой задачи. Процесс выбора конкретной модификации выглядит следующим образом.

1. Определяется множество искомых функциональных зависимостей.
2. В базу знаний вносятся правила для определения каждой функции из множества искомых.
3. Определяется множество порядков фильтров, соответствующих искомым функциям.
4. Далее в процессе работы перебираются порядки ЦФ только из заданного множества.

Например, при помощи ЦФ третьего порядка можно обнаружить параболы, синусоиды и все периодические функции, у которых период функции в 3 раза превосхо-

дит период дискретизации (другими словами, последовательность отсчётов такой функции состоит из одинаковых подпоследовательностей длиной в три отсчёта). Однако не все эти функции могут быть интересны для конкретного приложения. Допустим, нужно искать только синусоиды. Тогда в базе знаний для фильтров третьего порядка будет присутствовать только правило (условия), которому должны удовлетворять коэффициенты ЦФ, соответствующего синусоиде. Если бы правил было больше, то фильтр такого порядка для каждого участка исходной последовательности всё равно вычислялся бы только один раз. Но после вычисления его коэффициенты проверялись бы на соответствие нескольким правилам. Вообще, такие проверки очень просты по сравнению с затратами на само вычисление фильтра, а значит, количество искомых функций не сильно влияет на скорость поиска. Гораздо больше на производительность влияет множество порядков фильтров. Например, заметно ускорить работу можно, отказавшись от поиска всех функций, которым соответствуют ЦФ одного определённого порядка. Тогда фильтры этого порядка уже не нужно вычислять. Разумеется, большой эффект увеличения производительности даёт отказ от фильтров более высоких порядков. Эти рассуждения приводят нас к интересному свойству методики: вычислительные затраты на поиск нескольких различных функций в исходных данных не зависят линейно от количества искомых функций (как это было бы, например, при использовании МНК). Линейной зависимостью будет в худшем (очень редком) случае, а чаще же рост вычислительных затрат при росте числа искомых функций будет гораздо медленнее.

В соответствии с описанной методикой — для её проверки и демонстрации возможностей — была разработана реализация [3]. В ней осуществлялся поиск синусоид, показательных и линейных функций, полиномов 2-й, 3-й и 4-й степеней, константных зависимостей (т.е. последовательностей одинаковых элементов), суммы двух синусоид, а также периодических подпоследовательностей с периодами 2, 3, 4 и 5 отсчётов. Соответственно, для этого вычислялись БИХ-фильтры 2-го, 3-го, 4-го и 5-го порядков.

Поиск приближенных функциональных зависимостей

Исследования показали, что описанная методика поиска пригодна не только для обнаружения функций, проходящих точно через заданные точки (фактически интерполяция), но и для обнаружения аппроксимирующей функции. Это очень важно, так как на практике в любом физическом сигнале присутствуют помехи. Обнаружение аппроксимирующей функции отличается от описанной выше методики тем, что правила используемые после вычисления коэффициентов фильтра немного модифицируются. Модификация основана на том, что если форма сигнала близка к некоторой идеальной форме, то полученные коэффициенты фильтра тоже будут близки к коэффициентам, соответствующим функции, описывающей эту идеальную форму. Например, сигналу с формой параболы соответствует фильтр с коэффициентами обратной связи, равными $-3, 3, -1$

(это, фактически, соответствующее правило из базы знаний). Таким образом, если для имеющейся последовательности получен фильтр с коэффициентами ОС, равными $-3.04, 3.11, -0.97$, то можно утверждать, что она аппроксимируется квадратичной функцией.

В соответствии с этим, если требуется обнаружение приближённых функций, все правила из базы знаний модифицируются с учётом того, что все коэффициенты теперь должны принимать не фиксированные значения, а удовлетворять некоторым неравенствам. Также изменятся и условия, отражающие взаимосвязь между коэффициентами ЦФ, соответствующих «сложным» функциям.

К сожалению, пока не установлена закономерность, связывающая, например, отклонения значений коэффициентов ЦФ от идеальных и среднеквадратичное отклонение членов последовательности на отрезке от идеальной функции. Поэтому значения допустимых отклонений коэффициентов в зависимости от требуемой точности пока представляется возможным получать только опытным путём.

Возможность поиска приближённых функциональных зависимостей была опробована применительно к демодуляции телеметрической информации (ТМИ) с коррекцией искажений. При передаче ТМИ, как правило, сигнал модулируется синусоидой. В зависимости от скорости передачи при модуляции нулей и единиц скачкообразно может изменяться либо только фаза, либо и частота и фаза. Различной может быть и длительность передачи одного бита. Но в любом случае модулированный сигнал состоит из участков синусоид некоторой фазы и частоты. Причины неидеальности сигнала на принимающей стороне при этом известны: несущая на передающей стороне генерируется не идеально, всевозможные помехи в канале передачи, шумы квантования при оцифровке сигнала на принимающей стороне и др.

При этом в случае невозможности 100%-ной корректной демодуляции, бывает необходимо сохранять все записи сеансов связи в исходном виде. Хранение необработанной информации весьма накладно из-за больших размеров (например, запись минутного сеанса одного канала может занимать десятки мегабайт).

Созданный программный продукт воплощает такие преимущества методики как возможность настройки на обнаружение конкретных функциональных зависимостей, а также возможность использовать методику для обнаружения зашумлённых функциональных зависимостей и, как следствие, для сжатия с потерей информации.

Реализация была модифицирована для обнаружения только синусоид. Таким образом, вычислялись ЦФ только третьего порядка, что значительно ускорило работу программы. Для обнаружения зашумлённой синусоиды использовалось следующее правило:

$$\begin{aligned} 0 < a_1 < 3, \\ |a_1 + a_2| &\leq e_1, \\ |a_3 + 1| &\leq e_2, \end{aligned}$$

где a_1, a_2, a_3 — коэффициенты обратной связи ЦФ. Значения $e_1 = e_2 = 0.2$ были получены опытным путём (при обнаружении идеальной синусоиды они должны быть равны нулю). Чем они больше, тем больше исходные

данные могут отличаться от идеальной синусоиды.

После обнаружения участка зашумлённой синусоиды её можно заменить на идеальную синусоиду, а значит хранить вместо неё только 4 коэффициента (a_1, b_0, b_1, b_2). Сохранение обработанной (демодулированной) информации сделано в двух вариантах.

1. Сохраняется информация о распознанных участках синусоид, фазовых скачках и нераспознанных участках. Этой информации будет достаточно для отображения в программе визуализации, чтобы её последующую обработку мог закончить специалист (в части нераспознанных участков).

2. Если нераспознанных участков нет, — значит, найдены все фазовые скачки, и все нули и единицы модулированного цифрового потока могут быть однозначно идентифицированы и подвергнуты окончательному декодированию в соответствии с применяемой на передающей стороне схемой кодирования.

Обнаружение сигналов известной формы и интеллектуальный анализ данных

Два предыдущих приложения создают возможность использования алгоритма для обнаружения сигналов известной формы. Существуют области техники, где есть необходимость обнаружения коротких сигналов известной формы в непрерывно поступающих зашумлённых данных. Например, в приложениях радиолокации необходимо обнаруживать искажённые копии лоцирующих сигналов. Такие сигналы, как правило, имеют несложную форму, которую можно описать функцией (функциями), известной предлагаемому алгоритму.

Преимуществами данной методики обнаружения перед аналогами являются:

- лёгкая настраиваемость на одновременное обнаружение нескольких видов сигналов или сигналов, описываемых некоторой кусочной функцией;
- возможность быстрой аппаратной реализации.

Второе преимущество вытекает из того, что в методике используется цифровая фильтрация, для которой есть эффективные аппаратные решения в виде DSP-процессоров или реализаций на ПЛИС. При этом вне зависимости от сложности искомых функций, при работе алгоритма выполняются только самые простые операции (сложение, вычитание, умножение, сравнительно небольшое количество делений, сдвиговые регистры небольшой ёмкости). Например, при обнаружении синусоидальной функции вычисление синуса не требуется, равно как и при обнаружении полинома высокой степени ничего возводить в степень не надо. Ещё большее преимущество в быстрой реализации можно получить при одновременном обнаружении нескольких видов сигналов, т.к. выше было указано, что вычислительные затраты растут медленно при увеличении количества различных обнаруживаемых функций.

Другой большой и ещё достаточно новой областью, где также требуется поиск приближённых функциональных зависимостей и обнаружение интересующих зависимостей, является интеллектуальный анализ данных (Data Mining). В настоящее время в таких приложениях для обнаружения и прогнозирования часто используются такие методы как нейросети и генетические алгоритмы.

мы. При этом известны такие недостатки этих методов как долгое время работы и необходимость начальной настройки, требующей специальных знаний. Поэтому предложенная методика может удачно дополнить арсенал методов, используемых в приложениях Data Mining.

Сжатие данных

Выше было сказано, что вместо обнаруженного участка некоторой функции можно хранить коэффициенты соответствующего фильтра. Этот факт можно использовать для сжатия данных. Причём, как вытекает из вышесказанного, предложенный алгоритм можно использовать для сжатия, как с потерями информации, так и без них. При сжатии без потерь информации должны исходить только функции, точно описывающие отрезки сжимаемой последовательности. А при сжатии с потерями можно искать и аппроксимирующие функции.

В общем случае метод сжатия, основанный на предложенном алгоритме, будет действовать по такой схеме. Просматривая с самого начала сжимаемые данные, архиватор ищет участки, описываемые (точно или приближенно) какой-нибудь из известных ему функций. Вместо каждого такого участка в архиве будет сохраняться его длина, идентификатор обнаруженной функциональной зависимости и некоторое количество коэффициентов цифрового фильтра. Возможны и другие сочетания, необходимые для восстановления сжатого блока данных, однако данное сочетание было выбрано как оптимальное с точки зрения как занимаемого размера, так и скорости восстановления данных.

Поясним, например, почему было решено сохранять коэффициенты фильтра вместо параметров самой функции. Цель, ради которой они хранятся, одна — это восстановление отсчётов найденной функции, т.е., например, для участка данных, описываемого полиномом второй степени

$$y_i = \sum_{k=0}^2 s_k \cdot i^k,$$

$i = 0, 1, \dots, L$, надо сохранить 3 параметра s_k . При этом ему соответствует БИХ-фильтр третьего порядка с шестью коэффициентами (три из них — коэффициенты обратной связи). На первый взгляд хранить коэффициенты невыгоднее. Однако выше уже упоминалось, что определённым типам функций соответствуют конкретные коэффициенты обратной связи ЦФ (для «простых» функций) или некоторые соотношения между коэффициентами ЦФ (для «сложных» функций). Именно на этом и основана возможность обнаружения функциональных зависимостей. Поэтому часть коэффициентов можно не хранить, так как потом их можно будет восстановить, зная тип функциональной зависимости (в частности для этого сохраняется идентификатор функциональной зависимости). Например, для того же полинома второй степени не нужно хранить коэффициенты ОС, так как они будут иметь для такой функции фиксированное значение: $-3, 3, -1$. Таким образом, необходимо сохранять только оставшиеся три коэффициента — столько же, сколько и параметров функции. При этом их количество не зависит от длины L сжимаемого участка данных. Поэтому чем больше L , тем больше степень сжатия данно-

го участка. Для других видов функций, в том числе и «сложных», получается аналогично — количество коэффициентов фильтра, которые необходимо сохранить, равно количеству параметров функции с учётом свободного члена.

Для восстановления данных, можно использовать два разных способа: восстановление как отсчётов некоторой функции, взятой с равным шагом, и восстановление исходных данных как импульсной характеристики ЦФ. При этом второй способ, использующий именно коэффициенты соответствующего фильтра, вычислительно эффективнее. Поясним это на примере. Для полинома 4-й степени $f(i) = a \cdot i^4 + b \cdot i^3 + c \cdot i^2 + d \cdot i + e$ с целью восстановления каждого члена исходной последовательности требуется 10 умножений и 4 сложения.

Если же восстанавливать данные как импульсную характеристику ЦФ, то вычислительные затраты будут следующими. Полиному 4-й степени соответствует ЦФ 5-го порядка с 5-ю коэффициентами a_k и 5-ю коэффициентами b_k :

$$y(n) = \sum_{k=0}^4 b_k x(n-k) - \sum_{k=1}^5 a_k y(n-k)$$

Поскольку $x(n)$ в нашем случае является единичным импульсом, то первая сумма имеет ненулевое значение только при вычислении первых пяти значений $y(n)$ и равно по сути b_n , то есть для вычисления не требует даже умножения. Таким образом, восстановление каждого члена исходной последовательности в худшем случае потребует 5 умножений и 4 сложения. При этом не происходит возведения в большую степень, поэтому вероятность переполнения или потери значимости гораздо меньше, чем в классическом случае. Поэтому, если не требуется знать точных выражений функций, описывающих участки исходных данных, (а сжатие — именно такой случай), то не стоит тратить время на вычисление их параметров.

Вышесказанное говорит о том, что при сжатии можно вообще не определять, какой зависимости соответствует найденный ЦФ и не вычислять её параметры. Это хорошо с точки зрения скорости, так как эти действия могут быть достаточно трудоёмки, особенно для «сложных» функций. Основная задача алгоритма при сжатии — отыскание БИХ-фильтра, импульсная характеристика которого совпадает (при сжатии без потерь информации) или описывает с допустимой ошибкой (при сжатии с потерями) некоторый участок исходных данных. Определение вида соответствующей функции и её параметров при этом становится задачей вторичной и даже необязательной.

Этот приём использовался при создании пробной реализации архиватора [4], использующего данный алгоритм синтеза БИХ-фильтров по импульсной характеристике. Чтобы эффективно сжимать данные архиватор должен иметь как можно большую базу знаний коэффициентов обратной связи для «простых» функций и базу знаний правил определения «сложных» функций по коэффициентам ЦФ. Эти знания определяют список функций, известных алгоритму. Каждой такой функции поставлен в соответствие уникальный идентификатор. Размер идентификатора был принят равным 1 байту,

т.е. всего максимально такой архиватор мог бы «знать» 256 функций. При обнаружении использовались порядки ЦФ от 2 до 7. С учётом перечисленного в начале статьи стандартного множества Φ (без учёта последнего пункта), при помощи фильтров таких порядков можно обнаружить гораздо меньше функций, чем 256. Поэтому было решено расширить множество известных архиватору функций за счёт линейных комбинаций уже известных функций. Однако таких комбинаций очень много, и вычислять для каждой правила, по которым её можно обнаружить, нерационально. Вместо этого было сделано следующее. Алгоритмом обрабатывались наборы тестовых данных, и затем анализировались коэффициенты всех найденных фильтров, импульсные характеристики которых описывали достаточно длинные участки. Фильтры с наиболее часто встречающимися коэффициентами ОС или некоторыми зависимостями, связывающими коэффициенты фильтра, заносились в базу знаний алгоритма (архиватора). Некоторые из этих фильтров были позже проанализированы, и оказалось, что их импульсные характеристики действительно точно совпадают со значениями функций, являющихся линейными комбинациями других известных алгоритму функций. Делать это же для всех найденных фильтров не было необходимости. Главное, что в исходных данных такие фильтры описывали участки данных такого размера, что при их замене на коэффициенты фильтра достигается сжатие. Подобным образом было найдено несколько десятков функций (фильтров). Тем не менее, в авторской реализации задействовано менее половины идентификаторов. При этом было видно, что при внесении в базу знаний новых правил для обнаружения, тестовые файлы, по которым были обнаружены эти правила, после этого сжимаются лучше.

Так как при сжатии может встретиться функция, не входящая в список известных алгоритму, но, тем не менее, описывающая достаточно длинный участок исходной последовательности, то предусмотрена возможность сохранения и блока неизвестной зависимости. Разумеется, при этом приходится сохранять все коэффициенты фильтра.

Данный архиватор был сделан как универсальный, то есть он выполняет сжатие без потерь информации. Получившийся коэффициент сжатия хуже, чем у распространённых архиваторов. В скорости сжатия данная реализация также уступает аналогам, а скорость декомпрессии получилась очень хорошей, так как при этом в соответствии с уравнением цифровой фильтрации выполняются только самые простые команды — сложение и умножение. Основные пути улучшения сжатия:

- использование ЦФ более высоких порядков;
- расширение множества функций, известных алгоритму (приоритетными являются нерасходящиеся и медленно расходящиеся функции);
- использование сжатия с потерями информации.

Вообще использование данного метода сжатия для универсального архиватора на практике является сомнительным. Данная реализация была сделана, прежде всего, для проверки и отработки алгоритма. Более перспективным представляется использование алгоритма в областях, где, во-первых, в данных ожидается наличие

зависимостей, выражаемых математической формулой, а, во-вторых, допустимо сжатие с потерей информации. Такой пример был представлен выше — это алгоритм демодуляции телеметрической информации с коррекцией искажений. Полученная в том случае степень сжатия была гораздо выше, чем при сжатии без потерь. Кроме того, было использовано другое свойство алгоритма: возможность настройки на поиск только определённых функций, что послужило значительному ускорению работы.

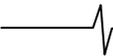
Упрощение алгебраических выражений

Другим интересным возможным приложением для предлагаемого алгоритма синтеза БИХ-фильтров является упрощение алгебраических выражений. При работе в современных математических пакетах с символьными выражениями они могут предоставлять возможность упрощения таких выражений. Как известно, справляются они с этой задачей не всегда успешно, иногда даже предлагая в качестве результата более сложную функцию, или такую же, но записанную в менее читаемом виде. При этом для разных видов упрощаемых выражений разработаны разные методы, и в разных математических пакетах могут быть реализованы разные сочетания этих методов [5].

Решение такой задачи с использованием алгоритма А1 может быть выполнено по довольно простой методике. Пусть имеется алгебраическая запись функции от одного переменного $f(x)$. Приложение вычисляет последовательность значений этой функции с постоянным шагом, а затем выполняет действия, описанные в первом пункте данной статьи — «Поиск точных функциональных зависимостей в числовых последовательностях». Если окажется, что полученной последовательности соответствует некоторая функция из базы знаний и она отличается от исходной, то задачу упрощения можно считать решённой. Вероятность того, что найденная функция будет сложнее, чем исходная, практически отсутствует, так как методика построена так, что ищет самую простую функцию, проходящую через заданный набор точек (в отличие от распространённых способов отыскания интерполирующих функций). А точнее — функцию с наименьшим количеством параметров.

Например, для последовательности $\{1, 1, 1, 1, 1, 1, \dots\}$ будет найдена функция $y=1$, а не синусоида, или многочлен степени, равной длине последовательности минус один. Аналогично, для последовательности $\{0, -1, 0, 3, 8, \dots\}$ будет найдена парабола $y=x^2-1$. А не какая-либо функция с большим числом параметров, которых на самом деле бесконечно много.

При этом надо учитывать тот факт, что методика может ошибиться в случае, если исходная функция (а значит и искомая) — периодическая, а шаг, с которым брали её значения, — больше половины периода (следует из теоремы Котельникова). Если шаг недостаточно мал или, более того, кратен периоду исходной функции, то будет найдена неверная функция. Однако такой ошибки несложно избежать. Надо всего лишь после того, как найдена некоторая функция, протабулировать исходную функцию с меньшим шагом и ещё раз осуще-



ствить поиск. Если найдена другая функция, значит, на предыдущем шаге значения функции $f(x)$ были взяты со слишком большим шагом. Более того, если упрощение выполняется в полуавтоматическом режиме, когда пользователь может указывать некоторые параметры для функции упрощения, то ему часто уже заранее может быть известно о наличии периодичности и значении периода анализируемой функции, и он сможет его указать, с тем чтобы ускорить поиск.

Разумеется, такая методика будет успешно упрощать выражения только до известных ей функций. Поэтому, как и в вышеприведённых приложениях, можно указать на необходимость расширения множества функций, входящих в базу знаний. Для решения задач упрощения алгебраических выражений и поиска функциональных зависимостей будет весьма полезно внесение в базу знаний сочетания не только линейных комбинаций уже известных функций, но и более сложных комбинаций, например, произведений таких функций.

Быстрая генерация подряд идущих значений заданной функции

В разных приложениях ЦОС могут потребоваться массивы значений различных функций, например синуса. Предложенный алгоритм может стать основой для процедуры быстрой генерации значений заданной функции одного переменного, равноотстоящих по оси абсцисс. Для этого до генерации (один раз) должен быть найден БИХ-фильтр, соответствующий нужной функции. Затем, вычисляя L первых отсчётов импульсной характеристики фильтра, получаем значения требуемой функции для значений аргумента $0, 1, 2, 3, \dots, L-1$. Если требуются значения функции для другого диапазона и с другим, но постоянным шагом, эту задачу легко можно привести к решаемой при помощи замены переменной. При этом функция не изменит свой вид, а изменятся только её параметры (по крайней мере, это выполняется для всех функций, входящих сейчас во множество функций, известных алгоритму, а значит, также выполняется и для их любой линейной комбинации). Например, значения функции $y_1(x) = k_1 \sin(w_1 x + \varphi_1) + c$, взятые с равным шагом s на интервале $[t_1, t_2]$, можно вычислить как значения другой функции $y_2(x) = k_2 \sin(w_2 x + \varphi_2) + c$, взятые на интервале $[0, t_3]$ с единичным шагом.

Почему данный подход имеет преимущество в скорости перед непосредственным вычислением значений некоторой функции, было дано выше в пункте о сжатии. Причина кроется в разнице количества выполняемых операций и в том, что при вычислении импульсной характеристики используются только простейшие операции сложения и умножения.

В частности, был проведён вычислительный опыт для вычисления массива значений синуса. Измерялась скорость заполнения массива в тысячи значений синуса в диапазоне $(0, 2\pi)$ с равномерным шагом. Один массив заполнялся с использованием стандартной функции вычисления синуса на языке Си. Второй — при помощи предложенного метода. Вычисления производились на процессорах популярной архитектуры x86. На вычислительных системах разных конфигураций (с процессора-

ми разных поколений — от Pentium II до Pentium-4) скорость авторского метода превосходила скорость вычисления стандартным путем в 5-25 раз. Из-за того, что при генерации каждого члена импульсной характеристики БИХ-фильтра используются значения предыдущих членов, в вычислениях с плавающей запятой имеется накапливаемая погрешность. В данном случае погрешность могла быть оценена сравнением со значениями первого массива. Она составила для первых членов массива — порядка 10^{-18} , для последних — порядка 10^{-12} . В подавляющем большинстве приложений это можно считать приемлемой точностью. К тому же, некоторыми алгоритмическими методами можно уменьшить погрешность для последних членов и, соответственно, среднюю ошибку. Кроме того, погрешность имеет место только для определённых целевых функций, когда фильтр имеет нецелые коэффициенты, а в остальных случаях фильтр и его импульсная характеристика получаются целочисленными.

Кроме того, стоит упомянуть, что если для работы приложения необходим только один массив, то не имеет большого смысла оптимизировать этот процесс. В этих случаях такой массив обычно вычисляется в начале работы (или просто прошивается в постоянной памяти), а затем используется выборка из массива значений. Предложенную методику стоит использовать при наличии каких-то из указанных ниже условий: таких массивов может быть несколько, целевые функции заранее точно неизвестны, и массивы будут генерироваться во время работы приложения; недостаточно памяти для хранения предварительно вычисленных массивов значений, либо значения целевой функции будут нужны всего один-два раза и при этом в порядке возрастания аргумента.

Перспективы развития алгоритма

Наиболее востребованными направлениями развития предложенного алгоритма синтеза БИХ-фильтров и его приложений, являются следующие:

1. *Расширение на многомерный случай.* На данный момент все сделанные реализации приложений алгоритма являются одномерными. При сжатии данные сжимаются как одномерный поток чисел, при обнаружении — ищутся только одномерные сигналы (функции) и т.д. Однако в приложениях сжатия, обнаружения и Data Mining может найти применение многомерная модификация алгоритма. Например, двухмерная — для сжатия изображений.

2. *Поиск новых функций, для которых существуют соответствующие фильтры.* Выше уже не раз упоминалось о необходимости расширения множества известных алгоритму функций. Это положительно повлияет на потенциальные возможности всех приложений алгоритма. Кроме линейных комбинаций было бы очень полезно определить применимость алгоритма к более сложным комбинациям известных алгоритму функций, например, произведениям функций. Во время написания статьи было обнаружено несколько видов функций, не входящих во множество Φ , для которых алгоритм A1 работоспособен. Среди них есть функции, сходящиеся к нулю, функции, стабилизирующиеся со временем к пе-

риодическим функциям, и функции, являющиеся произведением функций уже известных алгоритму.

3. *Расширение множества известных алгоритму функций при помощи разложения в ряд Тейлора.* Тот факт, что описанная выше методика поиска функциональных зависимостей способна обнаруживать много члены любых порядков, наталкивает на мысль, что можно использовать разложения в ряд Тейлора других функций, неизвестных алгоритму, для их обнаружения с некоторой допустимой ошибкой. Например, для функций логарифма или обратных тригонометрических функций не удалось аналитическим путём доказать существование соответствующего БИХ-фильтра или определить возможный порядок такого фильтра. Однако известны разложения этих функций в бесконечный степенной ряд Тейлора, являющиеся, по сути, полиномами бесконечной степени. Использование рядов Тейлора для приближённого вычисления значений функций известно давно. Точность вычисления зависит от количества взятых первых членов разложения функции в ряд. При этом, чем больше слагаемых, тем меньше ошибка вычисления, но больше вычислительных затрат. Поэтому в ресурсоёмких приложениях и системах реального времени следует использовать настолько короткий ряд, насколько позволяет значение допустимой ошибки вычисления. Допустимая ошибка может быть определена из требуемой точности приложения или предельной точности используемых вычислительных средств. После того, как выбрано допустимое значение ошибки, можно, основываясь на разложении интересующей функции (в виде остаточного члена ряда, наличии/отсутствии знакопеременности членов ряда) определить минимальное количество членов ряда, которые необходимо вычислять.

4. *Синтез целочисленных фильтров.* Известно, что использование целочисленных фильтров и DSP-процессоров с фиксированной точкой в ряде случаев выгоднее, чем фильтров с вещественными коэффициентами и процессоров с плавающей точкой. Операции с целыми числами выполняются быстрее, целые числа часто занимают меньше памяти. Поэтому, если будут найдены классы сходящихся к нулю функций, для которых предложенным алгоритмом можно найти соответствующий фильтр, либо будут найдены методы отыскания таких функций, то алгоритм можно будет использовать для синтеза устойчивых БИХ-фильтров. При этом было обнаружено, что если члены исходной импульсной характеристики являются целыми числами, то соответствующий ей фильтр в большинстве случаев тоже целочисленный. Исключение составляют случаи, когда в функции, описывающей члены импульсной характеристики, в качестве слагаемого присутствует синус (косинус). Возможность алгоритма работать в целых числах была использована в пробной реализации архиватора.

Заключение

Таким образом, предложенный ранее алгоритм синтеза БИХ-фильтров по импульсной характеристике, описываемой функцией из определённого множества, находит применение в ряде областей теории и практики ЦОС.

Поскольку в приложениях предложенного алгоритма одним из этапов является вычисление импульсной характеристики (т.е. фактически — процесс фильтрации единичного импульса), то хорошим известием можно считать объявление фирмы Intel, о том, что в 2009 году процессоры этой фирмы будут дополнены расширением команд AVX (Advanced Vector eXtensions), содержащим команды для реализации БИХ-фильтрации [6]. Это сделает возможным создание ещё более эффективной реализации алгоритма и его приложений на ПК.

Литература

1. Никитин Д.А., Ханов В.Х. Синтез рекурсивных цифровых фильтров по импульсной характеристике, определяемой элементарной математической функцией // Цифровая обработка сигналов. 2008, № 3, С. 10-14.
2. Ханов В.Х., Никитин Д.А. Алгоритм анализа числовых последовательностей // Вестник СибГАУ. — Красноярск, 2006. — № 6 (13). — С. 10-14.
3. Никитин Д.А. Программа поиска функциональных зависимостей в числовых последовательностях «FunSearch 1.0» — М.: ВНИИЦ, 2007. — № 50200700388, рег. № ОФАП 7729.
4. Никитин Д.А. Сжатие временных рядов с использованием блочной интерполяции // Информационные технологии моделирования и управления. Науч.-технич. журнал. — Воронеж: Научная книга, 2007. — № 1 (35). — С. 85–89.
5. M. Monagan, P. Lisonek, H. Bauck, Simplification of Algebraic Expressions. MITACS Research Report: <http://www.cecm.sfu.ca/CAG/products.html>, November 2000.
6. Астахов И. Реализация БИХ-фильтров с помощью набора инструкций Intel AVX для комплексных чисел с плавающей точкой. — <http://softwarecommunity-rus.intel.com/articles/rus/3783.htm>.

AN ALGORITHM OF IIR-FILTERS SYNTHESIS ON THE PULSE RESPONSE APPLICATIONS

Nikitin D. A.

How algorithm may be used for searching of exact or approximate functional dependencies, signals detection, data compression, simplification of mathematical expressions, fast generation of prescribed function values array is described in the article. Also possible ways of algorithm evolution are given.

АНАЛИЗ ХАРАКТЕРИСТИК ДВУМЕРНЫХ ЦИФРОВЫХ ФИЛЬТРОВ С ОРТОГОНАЛЬНЫМ ОГРАНИЧЕНИЕМ ПРОТЯЖЁННОСТИ СПЕКТРА ВЫХОДНОГО СИГНАЛА

Безруков А. В., к.т.н., ООО «СЦС Совинтел», e-mail: alexander_b75@mail.ru

Введение

Проектирование двумерных цифровых фильтров для аппаратуры связи различного назначения и, особенно, для систем передачи и обработки двумерных и многомерных массивов данных (ТВ сигналы, видеоинформационные сигналы, пространственно распределённые структуры информационных отсчётов уровня радиосигналов, характеристик, преобразованных в двумерный вид одномерных сигналов и т. д.) реализуют в настоящее время без необходимого теоретического обоснования требований к выбору характеристик и параметров. Это приводит к уменьшению эффективности работы соответствующего оборудования и программного обеспечения, используемого при регистрации и передачи двумерной и многомерной информации.

Результаты, полученные при анализе одномерных цифровых фильтров [1-9], чаще всего не могут быть однозначно трансформированы в двумерный и тем более в многомерный случаи. При этом следует учитывать также общую специфику построения и качественных отличий характеристик и параметров двумерных цифровых фильтров по отношению к одномерному случаю. С другой стороны, двумерная цифровая фильтрация является необходимым первичным этапом в разработке теории и практики многомерной цифровой параллельно-последовательной обработки (преобразований) информационных сигналов, которая по существу и является ближайшей перспективой эффективного совершенствования телекоммуникационного оборудования.

В данной работе дан анализ специфики характеристик двумерных цифровых фильтров с осевой симметрией и ортогональным ограничением протяжённости спектра выходного сигнала в реальном и в частотном пространствах, которые (из-за простоты реализации) находят все более широкое применение.

Анализ характеристик двумерных цифровых фильтров низких частот

В качестве одного из базовых вариантов определим вариант характеристики коэффициента передачи цифрового фильтра низких частот, который реализует в двумерном частотном пространстве идеализированные ограничение и дискретизацию двумерного спектра сигнала с ортогональным распределением и заданным периодом следования отсчётов, следующим образом:

$$K_H(\omega_x, \omega_y) = \sum_{\alpha=-m}^m \delta(\omega_x - \beta\omega_{x1}) \sum_{\beta=-n}^n \delta(\omega_y - \beta\omega_{y1}), \quad (1)$$

Проведён анализ симметричного вида характеристик двумерных цифровых фильтров, обеспечивающих ортогональное ограничение протяжённости дискретного спектра выходного сигнала в реальном и частотном пространствах. Получены расчётные соотношения для вариантов с относительным изменением ориентации структуры дискретизации по отношению к положению функции ограничения.

Показана целесообразность согласования (в областях пропускания и затухания) общих особенностей спектров исходных информационных сигналов с параметрами колебательных изменений уровня в сечениях двумерной характеристики коэффициента передачи цифровых фильтров, которые осуществлены в частотном пространстве по заданным направлениям.

В реальном пространстве соотношению (1) соответствует импульсная характеристика в виде результата двумерного преобразования Фурье:

$$G(x, y, \omega_{x1}, \omega_{y1}) = \frac{1}{4\pi^2} \sum_{\alpha=-m}^m \sum_{\beta=-n}^n \exp(jx\alpha\omega_{x1} + jy\beta\omega_{y1}) \quad (2)$$

Специфика функции, представленной соотношением (2), позволяет осуществить следующее преобразование:

$$G(x, y, \omega_{x1}, \omega_{y1}) = \frac{1}{4\pi^2} \left(1 + \sum_{\alpha=1}^m 2 \cos x\alpha\omega_{x1}\right) \cdot \left(1 + \sum_{\beta=1}^n 2 \cos y\beta\omega_{y1}\right) \quad (3)$$

Учитывая известное равенство

$$\sum_{\xi=0}^p \cos \xi \cdot z - 1 = 1 + 2 \cos z + 2 \cos 2z + 2 \cos 3z + \dots + 2 \cos (p-1)z + 2 \cos pz = \frac{\sin \frac{2p+1}{2}z}{\sin \frac{z}{2}}$$

приведём соотношение (3) к следующему виду:

$$G(x, y, \omega_{x1}, \omega_{y1}) = D_m(x) \cdot D_n(y) = \frac{1}{4\pi^2} \frac{\sin \frac{2m+1}{2} \cdot \omega_{x1} \cdot x}{\sin \frac{\omega_{x1} \cdot x}{2}} \cdot \frac{\sin \frac{2n+1}{2} \cdot \omega_{y1} \cdot y}{\sin \frac{\omega_{y1} \cdot y}{2}} \quad (4)$$

Выражение (4) фактически соответствует произведению функций ядер Дирихле –

$$D_p(z) = \frac{1}{2\pi} \cdot \frac{\sin \frac{2p+1}{2} \cdot z}{\sin \frac{z}{2}}$$

Максимальная амплитуда данной функции периодически определяется величиной

$$D_p(0) = \frac{2p+1}{2\pi}.$$

С увеличением значения p возрастает число осцилляций уровня функции $D_p(z)$, укладываемых в интервале между её смежными максимумами амплитуды, возникающими по оси z при значениях $z/2$, кратных величине π . При относительно больших значениях p функции ядер Дирихле для суммируемых (интегрируемых) функций $f(z)$ обладают свойствами δ -функции. Вместе с тем нормы функции ядра Дирихле не ограничены и, следовательно, имеются отдельные функции, для которых фильтрующее свойство δ -функции нарушается, и интеграл

$$\lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} f(z) D_n(z) dz \text{ не существует, т.е. не выполняется}$$

$$\text{свойство } \lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} f(z) D_n(z) dz = f(0).$$

Согласно выражению (4),

$$G(0,0, \omega_{x1}, \omega_{y1}) = D_m(0) \cdot D_n(0) = \frac{1}{4\pi^2} (2m+1) \cdot (2n+1)$$

и, следовательно, нормированная импульсная характеристика, соответствующая в реальном пространстве идеализированной периодической дискретизации и ограничению протяжённости спектра входного сигнала, имеет следующий вид:

$$G_0(x, y, \omega_{x1}, \omega_{y1}) = D_{0m}(x) \cdot D_{0n}(y) = \frac{1}{(2m+1) \cdot (2n+1)} \frac{\sin \frac{2m+1}{2} \cdot \omega_{x1} \cdot x}{\sin \frac{\omega_{x1} \cdot x}{2}} \cdot \frac{\sin \frac{2n+1}{2} \cdot \omega_{y1} \cdot y}{\sin \frac{\omega_{y1} \cdot y}{2}}, \quad (5)$$

$$= \frac{1}{(2m+1) \cdot (2n+1)} \frac{\sin \frac{\omega_{mx1} \cdot x}{2}}{\sin \frac{\omega_{x1} \cdot x}{2}} \cdot \frac{\sin \frac{\omega_{ny1} \cdot y}{2}}{\sin \frac{\omega_{y1} \cdot y}{2}}$$

где $\omega_{mx1} = (2m+1) \cdot \omega_{x1}, \omega_{ny1} = (2n+1) \cdot \omega_{y1}$.

Осуществим симметричное ограничение пространства локализации функции полученной идеализированной импульсной характеристики $G_0(x, y, \omega_{x1}, \omega_{y1})$:

$$G_{01}(x, y, \omega_{x1}, \omega_{y1}) = G_0(x, y, \omega_{x1}, \omega_{y1}) \cdot \text{rect}\left(\frac{x}{x_{11}}\right) \cdot \text{rect}\left(\frac{y}{y_{11}}\right). \quad (6)$$

Второй сомножитель здесь - прямоугольная функция

$$R(x, y) = \text{rect}\left(\frac{x, y}{x_{11}, y_{11}}\right) = \text{rect}\left(\frac{x}{x_{11}}\right) \cdot \text{rect}\left(\frac{y}{y_{11}}\right) = \begin{cases} 1 \text{ npu } |x| \leq x_{11}, |y| \leq y_{11} \\ 0 \text{ npu } |x| > x_{11}, |y| > y_{11} \end{cases}$$

спектр которой определяется известным выражением

$$S_r(\omega_x, \omega_y) = 4x_{11} \cdot y_{11} \frac{\sin \omega_x \cdot x_{11}}{\omega_x \cdot x_{11}} \cdot \frac{\sin \omega_y \cdot y_{11}}{\omega_y \cdot y_{11}}. \quad (7)$$

Произведение реальных функций, представленное в соотношении (6), отражается в частотном пространстве свёрткой соответствующих спектров. С учётом этого получим выражение для характеристики коэффициента передачи цифрового фильтра с ограничением идеализированной импульсной характеристики по площади в пространстве её локализации функцией "окна" в виде прямоугольного параллелепипеда:

$$K_0(\omega_x, \omega_y) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{\alpha=-m}^m \delta(\omega_x - \alpha \cdot \omega_{x1} - \xi_x) \sum_{\beta=-n}^n \delta(\omega_y - \beta \cdot \omega_{y1} - \xi_y) 4x_{11} y_{11} \frac{\sin x_{11} \cdot \xi_x}{x_{11} \cdot \xi_x} \cdot \frac{\sin y_{11} \cdot \xi_y}{y_{11} \cdot \xi_y} d\xi_x d\xi_y = \frac{x_{11} \cdot y_{11}}{\pi^2} \sum_{\alpha=-m}^m \frac{\sin(x_{11} \cdot (\omega_x - \alpha \cdot \omega_{x1}))}{x_{11} \cdot (\omega_x - \alpha \cdot \omega_{x1})} \sum_{\beta=-n}^n \frac{\sin(y_{11} \cdot (\omega_y - \beta \cdot \omega_{y1}))}{y_{11} \cdot (\omega_y - \beta \cdot \omega_{y1})} \quad (8)$$

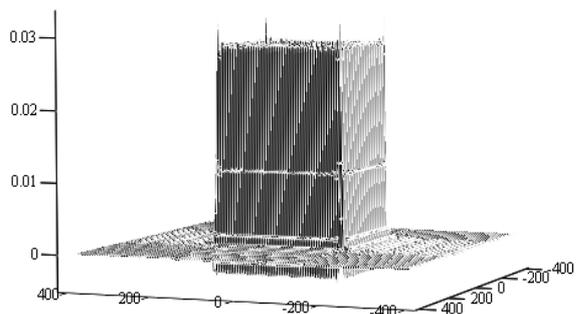
Полученное выражение (8) позволяет осуществлять анализ и выявить специфику зависимости двумерной функции характеристики коэффициента передачи от величин параметров $m, n, a_{x1}, a_{y1}, x_{11}, y_{11}$. На рис. 1а,б,в показана двумерная характеристика передачи и её сечения, которая соответствует импульсной характеристике, представленной произведением (6), и была вычислена с использованием соотношения (8) при $m = 24, n = 24$,

$$\omega_{x1} = \frac{2\pi}{1,0}, \omega_{y1} = \frac{2\pi}{1,0}, x_{11} = 0,5, y_{11} = 0,5.$$

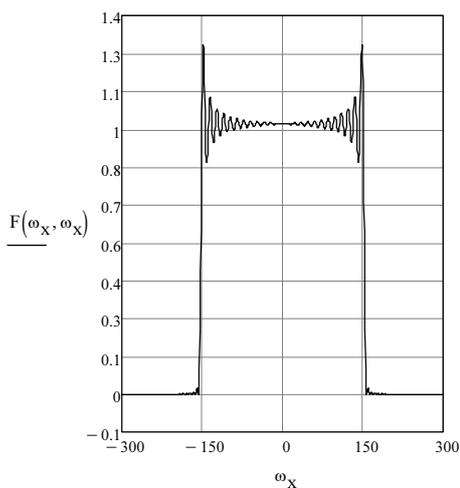
Следовательно, общее число отсчётов спектра составляет в данном случае величину $N_0 = 49 \cdot 49 = 2401$.

Как показали проведённые вычисления, увеличение протяжённости (площади области пропускания) исходного коэффициента передачи (идеализированного типа), которое имеет место при возрастании значений параметров m и n , сопровождается сопутствующим изменением масштаба двумерной функции коэффициента передачи цифрового фильтра, увеличением частоты (числа) колебательных изменений его уровня в пределах области пропускания и возрастанием крутизны падения уровня данной функции в участках переходов от двумерной области пропускания к полосе подавления.

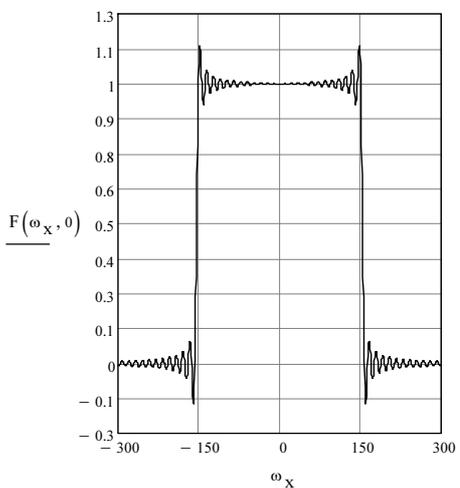
Существенное влияние на трансформацию формы функции коэффициента передачи двумерного фильтра оказывает протяжённость "окна", которая определяется величинами параметров x_{11}, y_{11} . С уменьшением протяжённости "окна" (значений параметров x_{11}, y_{11}) связано снижение частоты колебательных изменений уровня в пределах двумерной области пропускания и падение крутизны спада в участках переходов двумерной функции характеристики коэффициента передачи от области пропускания к области подавления. Последнее обусловлено увеличением протяжённости (площади) двумерной функции спектра "окна" и соответственно увеличением его относительного интегрирующего действия в соотношении (8). Форма сечений плоскостями в пространстве $\omega_x K_0 \omega_y$ локализации двумерной функции коэффициента передачи зависит от заданной ориентации такой плоскости.



а)



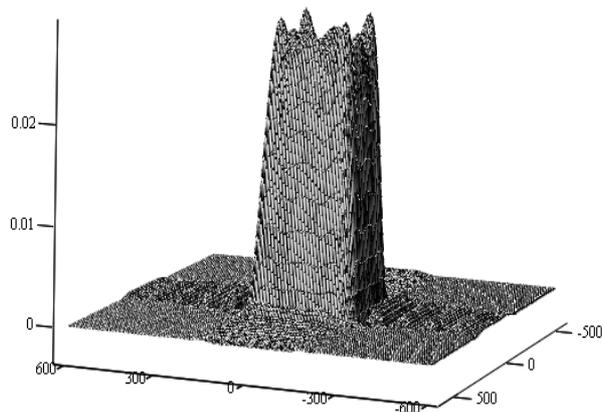
б)



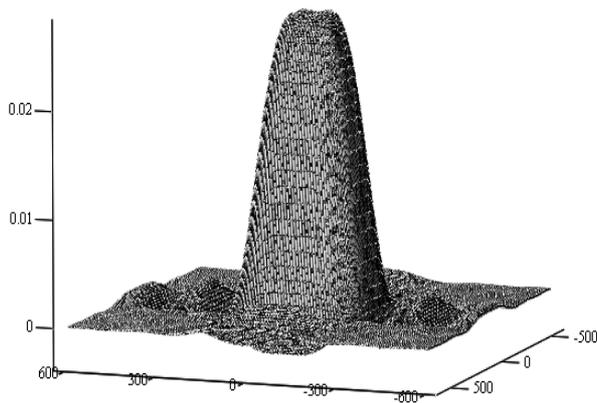
в)

Рис.1а,б,в. Двумерная характеристика передачи (а) и её сечения: б- в горизонтальном направлении, в - под углом в сорок пять градусов.

Полученные сечения свидетельствуют о существенной анизотропии колебательных изменений уровня в зависимости от положения плоскости сечения двумерной характеристики коэффициента передачи. В диагональных направлениях в рассмотренном случае сочетается возрастание уровня колебательных изменений в области пропускания с резким снижением уровня и протяжённости их локализации в полосе подавления характеристики коэффициента передачи цифрового фильтра.



а)



б)

Рис.2. Двумерные характеристики коэффициента передачи при $m = 24$, $n = 24$, $\omega_{x1} = \frac{2\pi}{1,0}$, $\omega_{y1} = \frac{2\pi}{1,0}$ и изменении протяжённости функции "окна" – $x_{11} = 0,07$, $y_{11} = 0,07$ (а) и $x_{11} = 0,03$, $y_{11} = 0,03$ (б).

На рис.2а показана специфика изменений характеристики двумерной функции коэффициента передачи, полученная с использованием выражения (8) при $m = 24$, $n = 24$,

$$\omega_{x1} = \frac{2\pi}{1,0}, \quad \omega_{y1} = \frac{2\pi}{1,0}, \quad x_{11} = 0,07, \quad y_{11} = 0,07, \quad \text{а на рис. 2б -}$$

$$m = 24, \quad n = 24, \quad \omega_{x1} = \frac{2\pi}{1,0}, \quad \omega_{y1} = \frac{2\pi}{1,0}, \quad x_{11} = 0,03, \quad y_{11} = 0,03$$

Специфика двумерных функций, представленных на рис.2, показывает, что, за счёт снижения значений параметров x_{11} и y_{11} можно добиться практического устранения колебательного процесса двумерной характеристики передачи в полосе пропускания. Однако в этом случае имеет место существенное падение крутизны спада характеристики в участках перехода двумерной функции характеристики коэффициента передачи от области пропускания к области подавления. Существенно снижается в данном случае частота колебательных изменений и возрастает в частотном пространстве протяжённость (площадь) их локализации, что определяет в конечном итоге вероятность увеличения в реальном пространстве интенсивности составляющих, проникающих в двумерный выходной сигнал из области подавления.

Другим вариантом ортогонального ограничения протяжённости спектра двумерной совокупности отсчётов спектра является вариант, отличающийся поворотом (например, на 45 градусов) структуры отсчётов исходного спектра и ориентации минимального периода их следования в диагональном направлении. Определим для такого случая идеализированный коэффициент передачи:

$$K_{ид}(\omega_x, \omega_y) = \sum_{\alpha=-2m}^{2m} \delta(\omega_x - \alpha \cdot \omega_{x1}) \sum_{\beta=-2n}^{2n} \delta(\omega_y - \beta \omega_{y1}) - \sum_{\alpha_1=-m}^m \delta(\omega_x - \alpha_1 \cdot 2\omega_{x1}) \cdot \sum_{\beta=-2n}^{2n} \delta(\omega_y - \beta \omega_{y1}) - \sum_{\alpha=-2m}^{2m} \delta(\omega_x - \alpha \cdot \omega_{x1}) \sum_{\beta_1=-n}^n \delta(\omega_y - \beta_1 \cdot 2\omega_{y1}) + 2 \cdot \sum_{\alpha_1=-m}^m \delta(\omega_x - \alpha_1 \cdot 2 \cdot \omega_{x1}) \sum_{\beta_1=-n}^n \delta(\omega_y - \beta_1 \cdot 2 \cdot \omega_{y1}) \quad (9)$$

Согласно соотношениям (1-6), двумерная импульсная характеристика в данном случае может быть выражена через вычитание и суммирование соответствующих произведений функций Дирихле:

$$G(x, y, \omega_{x1}, \omega_{y1}) = \left(\frac{\sin \frac{4m+1}{2} \cdot \omega_{x1} \cdot x}{\sin \frac{\omega_{x1} \cdot x}{2}} \cdot \frac{\sin \frac{4n+1}{2} \cdot \omega_{y1} \cdot y}{\sin \frac{\omega_{y1} \cdot y}{2}} - \frac{\sin \frac{2m+1}{2} \cdot 2\omega_{x1} \cdot x}{\sin \frac{2 \cdot \omega_{x1} \cdot x}{2}} \cdot \frac{\sin \frac{4n+1}{2} \cdot \omega_{y1} \cdot y}{\sin \frac{\omega_{y1} \cdot y}{2}} - \frac{\sin \frac{4m+1}{2} \cdot \omega_{x1} \cdot x}{\sin \frac{\omega_{x1} \cdot x}{2}} \cdot \frac{\sin \frac{2n+1}{2} \cdot 2 \cdot \omega_{y1} \cdot y}{\sin \frac{2 \cdot \omega_{y1} \cdot y}{2}} + 2 \cdot \frac{\sin \frac{2m+1}{2} \cdot 2 \cdot \omega_{x1} \cdot x}{\sin \frac{2 \cdot \omega_{x1} \cdot x}{2}} \cdot \frac{\sin \frac{2n+1}{2} \cdot 2 \cdot \omega_{y1} \cdot y}{\sin \frac{2 \cdot \omega_{y1} \cdot y}{2}} \right) \quad (10)$$

С учётом соотношения (9), аналогичным образом, получим выражение для характеристики коэффициента передачи двумерного цифрового фильтра с ограничением идеализированной импульсной характеристики (соответствующей выражению (10)), по площади в пространстве её локализации, функцией "окна" в виде прямоугольного параллелепипеда:

$$K_{0Д}(\omega_x, \omega_y) = \frac{x_{11} \cdot y_{11}}{\pi^2} \cdot \left(\sum_{\alpha=-2m}^{2m} \frac{\sin(x_{11}(\omega_x - \alpha \cdot \omega_{x1}))}{x_{11}(\omega_x - \alpha \cdot \omega_{x1})} \cdot \sum_{\beta=-2n}^{2n} \frac{\sin(y_{11}(\omega_y - \beta \cdot \omega_{y1}))}{y_{11}(\omega_y - \beta \cdot \omega_{y1})} - \sum_{\alpha_1=-m}^m \frac{\sin(x_{11}(\omega_x - \alpha_1 \cdot \omega_{x1}))}{x_{11}(\omega_x - \alpha_1 \cdot \omega_{x1})} \cdot \sum_{\beta=-2n}^{2n} \frac{\sin(y_{11}(\omega_y - \beta \cdot \omega_{y1}))}{y_{11}(\omega_y - \beta \cdot \omega_{y1})} - \sum_{\alpha=-2m}^{2m} \frac{\sin(x_{11}(\omega_x - \alpha \cdot \omega_{x1}))}{x_{11}(\omega_x - \alpha \cdot \omega_{x1})} \cdot \sum_{\beta_1=-n}^n \frac{\sin(y_{11}(\omega_y - \beta_1 \cdot \omega_{y1}))}{y_{11}(\omega_y - \beta_1 \cdot \omega_{y1})} + 2 \cdot \sum_{\alpha_1=-m}^m \frac{\sin(x_{11}(\omega_x - \alpha_1 \cdot \omega_{x1}))}{x_{11}(\omega_x - \alpha_1 \cdot \omega_{x1})} \cdot \sum_{\beta_1=-n}^n \frac{\sin(y_{11}(\omega_y - \beta_1 \cdot \omega_{y1}))}{y_{11}(\omega_y - \beta_1 \cdot \omega_{y1})} \right) \quad (11)$$

Проведённые вычисления свидетельствуют о зависимости параметров коэффициента передачи от поворота структуры отсчётов исходного спектра и ориентации минимального периода их следования. На рис.3 показаны сечения коэффициента передачи, соответствующие выражению (11), для горизонтального и диагонального направлений (для варианта с поворотом структуры дискретизации при неизменной ориентации двумерной "оконной" функции), вычисленные при

$$m=17, n=17, \omega_{x1} = \frac{2\pi}{1,414}, \omega_{y1} = \frac{2\pi}{1,414}, x_{11}=0,5, y_{11}=0,5.$$

Общее число отсчётов спектра составляет в данном случае, согласно (9), величину $N_{0Д} = 69 \cdot 69 + 2 \cdot 35 \cdot 35 - 2 \cdot 69 \cdot 35 = 2381$, т.е. относительно несущественно, по величине данного параметра, отличается от рассмотренного ранее варианта ($N_0 = 49 \cdot 49 = 2401$).

Сравнение расчётных графиков рис.1 и 3 показывает, что относительное фазовое изменение (поворот) ориентаций структуры отсчётов спектра и оконной функции не сопровождается общими изменениями формы сечений двумерной функции коэффициента передачи, которая определяется выбранной функцией "окна", но заметно отражается на отдельных её параметрах. Имеет место, в частности при реализации поворота, общее относительное падение максимальной амплитуды колебательных изменений в сечениях функции коэффициента передачи. При этом в горизонтальном направлении величина падения $\cong 17\%$, а в диагональном $\cong 40\%$. Уменьшается также величина перепада в амплитуде колебательных изменений в области высоких и низких частот. В связи с последним возрастает эффективность компенсационных преобразований (из-за колебательных изменений уровня характеристики передачи) спектра сигнала. Последнее позволяет более эффективно обрабатывать в двумерном цифровом фильтре сигналы с более гладкой (в среднем) структурой спектра исходного сигнала.

Следует отметить [2], что существенная минимизация уровня колебательного процесса в полосе пропускания и затухания одномерных цифровых фильтров достигается при использовании варианта функции "окна" гауссовского типа с полной компенсацией ступени, появление которой обусловлено усечением протяжённости указанной функции. Этот вариант, безусловно, как отмечено в [2], целесообразно использовать и при реализации двумерных цифровых фильтров. Однако следует учитывать, что параметры функции ограничения протяжённости "окна" гауссовского типа в двумерном пространстве должны быть оптимизированы в каждом конкретном случае (по отношению к характеристикам выбранной структуры дискретизации спектра информационного сигнала) и по форме, и по ориентации. Последнее, естественно, справедливо при выборе анизотропного варианта "оконной" функции.

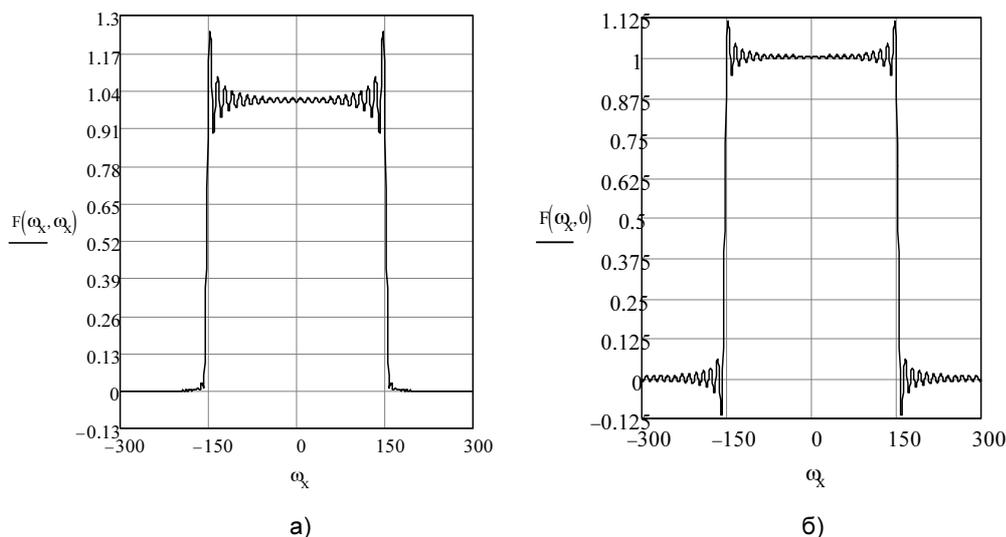


Рис.3а,б. Сечения двумерной характеристики передачи для варианта с фазовым изменением ориентации “окна”. а- сечение в горизонтальном направлении, б- сечение по диагонали.

Заключение

Таким образом, проведённый анализ специфики частных (с осевой симметрией) характеристик двумерных цифровых фильтров с ортогональным ограничением протяжённости спектра выходного сигнала в реальном и в частотном пространствах показал:

- увеличение протяжённости области пропускания двумерного цифрового фильтра подобного типа в частотном пространстве сопровождается увеличением частоты (числа) колебательных изменений его коэффициента передачи в пределах области пропускания и возрастанием крутизны падения уровня данной функции в участках переходов от двумерной области пропускания к полосе подавления;

- уменьшение протяжённости двумерного “окна” в реальном пространстве определяет снижение частоты колебательных изменений уровня в пределах двумерной области пропускания и падение крутизны спада в участках переходов двумерной функции характеристики коэффициента передачи от области пропускания к области подавления;

- разработанная методика анализа и полученные расчётные соотношения позволяют конкретизировать тонкую специфику зависимостей структуры характеристик двумерных цифровых фильтров от параметров структуры дискретизации информационного сигнала и параметров, определяющих функцию “окна”.

Литература

1. Безруков А.В. Анализ искажений амплитудно-частотной характеристики цифрового фильтра с симметричной импульсной характеристикой // Радиотехника.- 2005, №10, с.115-119.
2. Безруков А.В. Анализ и выбор параметров одномерных цифровых нерекурсивных фильтров с характеристиками ограниченной протяжённости // Радиотехника.- 2000, №10, с.35-41.
3. Хемминг Р.В. Цифровые фильтры.- М.: Сов.радио.-1980г.- 224с.

4. Рабинер Л., Голд Б. Теория и применение цифровой обработки сигналов.-М:Мир.-1978.-848с.
5. Khalid Saeed. On the realization of digital filters // Proceedings of the 1st International Conference “DSP and its applications”.-М.:v.1-1998.-С.141-143.
6. Карташев В.Г. Основы теории дискретных сигналов и цифровых фильтров.-М.:В.Ш, 1982.-109с.
7. Smith S.W. The scientist and Engineer’s Guide to Digital Signal Processing.-: California Technical Publishing, ISBN: 0-9660176-3-3, 1997.
8. I.W.Selesnick, C.S.Burrus. Maximally flat low-pass FIR filters with reduced delay// IEEE Trans. Circuits Syst. II, vol.45, Jan.1998, pp.53-68
9. F.Candocia, J.C.Principe Comments on “Sync interpolation of discrete periodic signals”//IEEE Trans.Signal Processing.- v.46,Jul.1998, pp.2044-2047.

AN ANALYSIS OF PARTICULAR SYMMETRICAL FORM CHARACTERISTICS OF TWO-DIMENSIONAL DIGITAL FILTERS

A.V.Bezrukov

The research contains an analysis of particular symmetrical form characteristics of two-dimensional digital filters for an output discrete signal spectrum orthogonal limitation in both real and frequency domains. Reference ratios were calculated for the cases of discretization structure orientation relative change against the limitation function state.

By using developed analysis procedure and calculated reference ratios it’s possible to specify two-dimensional digital filters thin structure changes based on informational spectrum signal discretization parameters and “window” function basic parameters. The expediency of balancing of common properties of spectrums of initial informational signals with parameters of level oscillatory changes in cuts of two-dimensional characteristic of carryover factor of digital filters realized in frequency domain in given direction is shown in areas of passing and attenuation.

УДК 621.396

ПРИМЕНЕНИЕ АЛГОРИТМА ЛЕВИНСОНА ПРИ РЕАЛИЗАЦИИ МЕТОДА ПРЯМОЙ ОЦЕНКИ ДАННЫХ В РАДИОМОДЕМАХ

*Султанов Б.В., д.т.н., профессор кафедры ИБСТ Пензенского государственного университета (ПГУ),
bv_sultanov@mail.ru*

Шутов С.Л., с.н.с. ПГУ, sshutov58@mail.ru

Румянцева Н.Б., студентка ПГУ

Зефиоров С.Л., к.т.н., доцент, зав. кафедрой ИБСТ ПГУ, ibst@pnzgu.ru

Щербаков М.А., д.т.н., профессор, проректор по научной работе ПГУ, mash@sura.ru

Ключевые слова: алгоритм, оценка данных, метод, радиомодем, символ, передача, фрейм, система линейных уравнений, теплицева матрица.

Введение

В ряде практических задач цифровой обработки сигналов, например, таких как линейное предсказание и адаптивная фильтрация [1], спектральный анализ [2], оценивание параметров канала и сигналов [3], помехоустойчивое кодирование [4] и др. возникает проблема решения систем линейных уравнений высокого порядка, сводимых к матричному уравнению вида

$$Af = g, \quad (1)$$

где A представляет собой теплицевую матрицу (т.е. матрицу, у которой элементы, расположенные параллельно главной диагонали, являются одинаковыми) размером $n \times n$;

f и g – векторы-столбцы с компонентами f_i и g_i , $i \in \overline{0, n-1}$.

Стандартные методы решения уравнения (1), разработанные для матриц общего вида, как прямые (метод Гаусса, метод Холецкого и др.), так и итерационные (метод Зейделя и др.) имеют высокую вычислительную сложность, пропорциональную числу n^3 операций с плавающей точкой, т.е. порядка $O(n^3)$. При этом специфические свойства теплицевой матрицы, а так же определенные виды её симметрии, возникающие вследствие особенностей конкретных практических приложений, делают возможным разработку более эффективных методов решения данной задачи с существенно меньшей вычислительной сложностью.

Один из первых наиболее известных методов решения матричного уравнения (1) с симметричной теплицевой матрицей с вычислительной сложностью $O(n^2)$ является метод Левинсона [2,4]. В дальнейшем были разработаны другие быстрые (характеризующиеся вычислительной сложностью порядка $O(n^2)$) алгоритмы, позволяющие решать как частные (метод Левинсона-Дурбина [1,4,5], предполагающий, что последний столбец матрицы A в (1) совпадает с записанным в обратном порядке вектором g), так и более общие (метод Тренча [4], не требующий симметричности матрицы A) задачи. К настоящему времени предложен ряд супербыстрых алгоритмов [6], позволяющих решать уравнение (1) в общем случае с вычислительной сложностью меньшей $O(n^2)$. Вместе с

Рассмотрены особенности использования алгоритма Левинсона в задаче многократного решения матричного уравнения высокого порядка, возникающей при реализации метода прямой оценки данных в последовательных радиомодемах, предназначенных для работы в КВ диапазоне частот, и предложены решения, позволяющие существенно сократить её вычислительную сложность.

тем существует важное практическое приложение, в котором, в силу его специфики, применение алгоритма Левинсона с учетом возникающих при этом особенностей, по-прежнему, остаётся привлекательным. Таким приложением является метод прямой (непосредственной) оценки данных в последовательных радиомодемах, предназначенных для работы в коротковолновом диапазоне частот [3] (в дальнейшем так называемых КВ-модемах). В настоящей статье анализируются особенности использования алгоритма Левинсона применительно к данной задаче, формулируются и обосновываются предложения, позволяющие повысить вычислительную эффективность её решения.

Метод прямой (непосредственной) оценки данных

Метод непосредственной оценки данных в настоящее время является одним из наиболее эффективных подходов к построению последовательных КВ-модемов и обеспечивает возможность получения относительно высоких скоростей передачи данных при наличии фединга и многолучёвости. Достаточно сказать, что стандарты НАТО, определяющие технические характеристики модемов данного класса, регламентируя структуру передаваемого сигнала, предполагают использование именно этого метода [7,8]. Данная структура базируется на использовании многократной фазовой или квадратурно-амплитудной модуляции и объединении модулированных символов данных в так называемые блоки решения (decision blocks). Как показано на рис. 1, блок решения содержит транспонированные (с индексом T): вектор $A_1^T = [a_0, a_1, \dots, a_{N-1}]$, компонентами которого являются N известных символов; вектор $B_1^T = [b_0, b_1, \dots, b_{M-1}]$, включающий M неизвестных символов, и вектор $A_2^T = [a_N, a_{N+1}, \dots, a_{2N-1}]$, состоящий из N известных символов. Сигнал передаётся в виде последовательных фреймов, каждый из которых включает обучающий и информационный блоки, состоящие соответственно из N известных и M неизвестных символов.

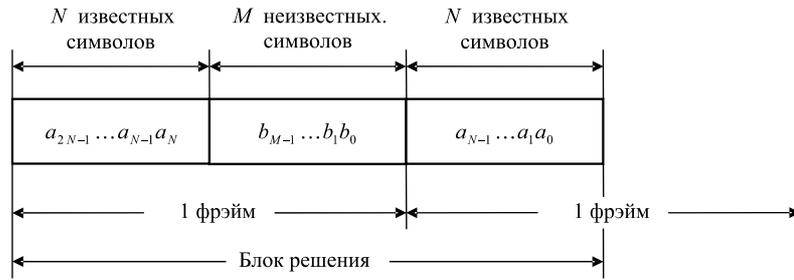


Рис.1

Рассматривая в качестве модели канала нерекурсивный фильтр с импульсной характеристикой, содержащей $L+1$ ненулевых отсчетов ($L \leq N$) и описываемой вектором $h^T = [h_L, h_{L-1}, \dots, h_0]$, в предположении постоянства h^T на интервале следования символов блока решения, вектор отсчетов выходного сигнала канала r , обусловленных воздействием входных символов a_i и b_i (т.е. не включающий отсчеты, являющиеся реакцией только на обучающие символы a_i), можно определить выражением:

$$r^T = M_1 B + M_2 A_3 + M_3 A_4 + n^T, \quad (2)$$

где M_1 , M_2 , M_3 – матрицы размера $(M+L) \times M$, $(M+L) \times L$ и $(M+L) \times L$ соответственно вида

$$M_1 = \begin{bmatrix} h_0 & 0 & 0 & 0 \\ h_1 & h_0 & 0 & 0 \\ \vdots & h_1 & \vdots & 0 \\ h_L & \vdots & h_0 & \\ 0 & h_L & h_1 & \\ 0 & 0 & \vdots & \\ 0 & 0 & 0 & h_L \end{bmatrix}; \quad M_2 = \begin{bmatrix} h_L & h_{L-1} & \dots & h_1 \\ 0 & h_L & \dots & h_2 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & h_L \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}; \quad (3)$$

$$M_3 = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 \\ h_0 & 0 & 0 & \dots & 0 \\ h_1 & h_0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h_{L-1} & h_{L-2} & h_{L-3} & \dots & h_0 \end{bmatrix};$$

$$A_3^T = [a_{N-L}, a_{N-L-1}, \dots, a_{N-1}]; \quad A_4^T = [a_N, a_{N+1}, \dots, a_{N+L-1}];$$

$n = [n_{N-L}, n_{N-L+1}, \dots, n_{M+N-1}]$ – вектор отсчетов

аддитивного шума.

После того, как будет накоплен набор выходных символов канала, входящих в состав блока решения, и с использованием известных значений обучающих символов тем или иным способом получена оценка канальной импульсной характеристики h^T , которая в соответствии с (3) однозначно определяет матрицы M_1 , M_2 и M_3 , на основании (2) с учетом (3) можно определить вектор C

$$C = r^T - M_2 A_3 - M_3 A_4 = M_1 B + n^T. \quad (4)$$

Минимизация по отношению к оценке \hat{B} вектора переданных данных B среднеквадратической ошибки, определяемой как

$$\varepsilon^2 = (M_1 \hat{B} - C)^H (M_1 \hat{B} - C),$$

где индекс « H » означает эрмитово сопряжение матрицы (включающее транспонирование и комплексное сопряжение), приводит к результату [3]:

$$\hat{B} = (M_1^H M_1)^{-1} M_1^H C. \quad (5)$$

Однако вычисление величины \hat{B} непосредственно по формуле (5) предполагает выполнение чрезвычайно трудоёмкой (особенно при больших значениях M , порядка нескольких сотен) процедуры обращения комплексно-сопряженной по отношению к R_M матрицы

$$R_M^* = M_1^H \cdot M_1, \quad (6)$$

При этом задаваемое выражением (5) значение \hat{B} является решением матричного уравнения, аналогичного (1) и имеющего вид:

$$R_M^* \cdot \hat{B}_M^T = G_M^T,$$

где $G_M^T = M_1^H \cdot C$.

Подставив M_1 из (3) в (6), получим:

$$R_M = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{M-1} \\ \gamma_1^* & \gamma_0 & \gamma_1 & \dots & \gamma_{M-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{M-1}^* & \gamma_{M-2}^* & \gamma_{M-3}^* & \dots & \gamma_0 \end{bmatrix},$$

где

$$\gamma_0 = \sum_{i=0}^L |h_i|^2; \quad \gamma_1 = \sum_{i=0}^{L-1} h_{i+1} h_i^*; \quad \dots; \quad \gamma_L = h_L h_0^*; \quad \gamma_t = 0 \text{ для } t > L. \quad (7)$$

Таким образом, матрица R_M является эрмитовой, теплолицевой. Также можно показать, что она положительно определена. Кроме того, данная матрица является ленточной с шириной $2L-1$ (т.е. $(2L-1)$ -диагональной [7]), поскольку в соответствии с (7) её элементы с индексами, превышающими значение L , являются нулевыми.

Как показано в [3], при реализации рассмотренного метода получения вектора оценки \hat{B} переданных данных, вследствие влияния межсимвольной интерференции, дисперсия различных его компонент $\hat{b}_0, \hat{b}_1, \dots, \hat{b}_{M-1}$ является неодинаковой и минимальна для крайних составляющих \hat{b}_0 и \hat{b}_{M-1} . Данное обстоятельство делает целесообразной реализацию поэтапной оценки блока переданных данных. На каждом из этапов описанным выше способом определяется вектор \hat{B} , однако окончательное решение о значениях переданных символов принимается лишь относительно двух крайних компонент. Принятые решения приписываются к блокам известных символов A_3 и A_4 , увеличивая размерность каждого из них на 1, а размерность оцениваемого блока B при этом сокращается на 2. Затем соответствующим образом пересчитываются входящие в

уравнение (5) матрицы R_{M-2} и G_{M-2} , и оно вновь решается относительно вектора уменьшенной размерности \hat{B}_{M-2} . Данная процедура повторяется $M/2$ раз до тех пор, пока не будут получены все максимально достоверные оценки $\hat{b}_0, \hat{b}_1, \dots, \hat{b}_{M-1}$.

Рассмотренный метод называется методом нелинейной непосредственной оценки данных (Nonlinear-Data-Directed-Estimator (NDDE))[3]. Позволяя обеспечить высокую помехоустойчивость, он характеризуется большой вычислительной сложностью, причём основные реализационные проблемы связаны с необходимостью многократного решения уравнения (5). Однако отмеченные выше особенности матрицы R_M позволяют существенно сократить ожидаемые вычислительные затраты, эффективно применив в данных условиях «несупербыстрый» алгоритм Левинсона.

Алгоритм Левинсона для решения систем линейных уравнений с тридиагональной эрмитовой матрицей

Следуя методике, приведённой в [4] (где этот алгоритм рассмотрен в поле действительных чисел), получим соотношения, определяющие алгоритм Левинсона для решения уравнения (1), в котором матрица A является эрмитовой тридиагональной. При этом оно принимает вид:

$$\begin{bmatrix} a_0 & a_1^* & a_2^* & \dots & a_{n-2}^* & | & a_{n-1}^* \\ a_1 & a_0 & a_1^* & \dots & a_{n-3}^* & | & a_{n-2}^* \\ a_2 & a_1 & a_0 & \dots & a_{n-4}^* & | & a_{n-3}^* \\ \vdots & \vdots & \vdots & \dots & \vdots & | & \vdots \\ a_{n-2} & a_{n-3} & a_{n-4} & \dots & a_0 & | & a_1^* \\ - & - & - & - & - & - & - \\ a_{n-1} & a_{n-2} & a_{n-3} & \dots & a_1 & | & a_0 \end{bmatrix} \begin{bmatrix} f_0^{(n)} \\ f_1^{(n)} \\ f_2^{(n)} \\ \vdots \\ f_{n-2}^{(n)} \\ - \\ f_{n-1}^{(n)} \end{bmatrix} = \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ \vdots \\ g_{n-2} \\ - \\ g_{n-1} \end{bmatrix}. \quad (8)$$

Разделение матрицы и векторов на блоки здесь сделано для того, чтобы проиллюстрировать сущность метода окаймливания [2], заложенного в основу рассматриваемого итеративного алгоритма. На r -й итерации ($1 \leq r \leq n$) вычисляется решение r -го усечения задачи (описываемого соотношением (8) при $n = r$; верхний индекс у составляющих вектора-столбца $f^{(r)}$ характеризует номер итерации), причём это вычисление основывается на использовании полученного на предыдущей итерации ($r-1$ -го усечения). Алгоритм Левинсона при начальном условии $a_0 \cdot f_0^{(0)} = g_0$ рекурсивно модернизирует $f^{(r)}$ таким образом, чтобы $f^{(n)}$ равнялся решению исходной задачи.

Помимо вектора $f^{(n)}$ в итерациях участвуют несколько рабочих переменных: скалярные переменные α_r, β_r и γ_r и вектор $t^{(r)}$, содержащий r компонент: $t^{(r)} = [t_0^{(r)}, t_1^{(r)}, \dots, t_{r-1}^{(r)}]$. Эти переменные выбираются так, чтобы на r -м шаге выполнялись следующие матричные равенства:

$$A_r \cdot t^{(r)} = \begin{bmatrix} \alpha_r \\ 0_{r-1} \end{bmatrix}; \quad (9)$$

$$A_{r+1} \cdot \begin{bmatrix} f_0^{(r)} \\ \vdots \\ f_{r-1}^{(r)} \\ 0 \end{bmatrix} = \begin{bmatrix} g_0 \\ \vdots \\ g_{r-1} \\ \gamma_r \end{bmatrix}; \quad (10)$$

$$A_{r+1} \cdot \begin{bmatrix} t^{(r)} \\ 0 \end{bmatrix} = \begin{bmatrix} \alpha_r \\ 0_{r-1} \\ \beta_r \end{bmatrix}, \quad (11)$$

где $A_r = \begin{bmatrix} a_0 & a_1^* & \dots & a_{r-1}^* \\ a_1 & a_0 & \dots & a_{r-2}^* \\ \dots & \dots & \dots & \dots \\ a_{r-1} & a_{r-2} & \dots & a_0 \end{bmatrix}$; 0_{r-1} – вектор столбец,

содержащий $(r-1)$ нулей.

Используя обменную матрицу

$$J_r = \begin{bmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad (\text{размера } r \times r)$$

и последующее комплексное сопряжение, соотношение (10) можно записать в виде:

$$[J_r A_r J_r]^* \cdot [J_r t^{(r)}]^* = \left[J_r \begin{bmatrix} \alpha_r \\ 0_{r-1} \end{bmatrix} \right]^*. \quad (12)$$

Поскольку $J_r A_r J_r = A_r^*$, из (12) следует:

$$A_r \cdot \begin{bmatrix} t_{r-1}^{*(r)} \\ t_{r-2}^{*(r)} \\ \vdots \\ t_0^{*(r)} \end{bmatrix} = \begin{bmatrix} 0_{r-1} \\ \alpha_r^* \end{bmatrix}. \quad (13)$$

Аналогичное преобразование соотношения (11) приводит к результату:

$$A_{r+1} \cdot \begin{bmatrix} 0 \\ t_{r-1}^{(r)} \\ \vdots \\ t_1^{(r)} \\ t_0^{(r)} \end{bmatrix} = \begin{bmatrix} \beta_r^* \\ 0 \\ \vdots \\ 0 \\ \alpha_r^* \end{bmatrix}. \quad (14)$$

Положим, что на $(r+1)$ -й итерации векторы $t^{(r+1)}$ и $f^{(r+1)}$ можно определить как

$$\begin{bmatrix} t_0^{(r+1)} \\ \vdots \\ t_{r-1}^{(r+1)} \\ t_r^{(r+1)} \end{bmatrix} = k_1 \begin{bmatrix} t_0^{(r)} \\ \vdots \\ t_{r-1}^{(r)} \\ 0 \end{bmatrix} + k_2 \begin{bmatrix} 0 \\ t_{r-1}^{*(r)} \\ \vdots \\ t_1^{*(r)} \\ t_0^{*(r)} \end{bmatrix} \quad (15)$$

$$\text{и} \quad \begin{bmatrix} f_0^{(r+1)} \\ \vdots \\ f_{r-1}^{(r+1)} \\ f_r^{(r+1)} \end{bmatrix} = \begin{bmatrix} f_0^{(r)} \\ \vdots \\ f_{r-1}^{(r)} \\ 0 \end{bmatrix} + k_3 \begin{bmatrix} 0 \\ t_r^{*(r+1)} \\ \vdots \\ t_1^{*(r+1)} \\ t_0^{*(r+1)} \end{bmatrix}, \quad (16)$$

где k_1, k_2 и k_3 – подлежащие дальнейшему вычислению коэффициенты.

Умножив обе части равенства (15) на A_{r+1} и приняв во внимание соотношения (9), (11), (14), получим:

$$\begin{bmatrix} \alpha_{r+1} \\ 0_r \end{bmatrix} = k_1 \begin{bmatrix} \alpha_r \\ 0_r \\ \beta_r \end{bmatrix} + k_2 \begin{bmatrix} \beta_r^* \\ 0_r \\ \alpha_r^* \end{bmatrix}, \quad (17)$$

откуда следует, что k_1 и k_2 должны выбираться так, чтобы выполнялись равенства

$$0 = k_1 \beta_r + k_2 \alpha_r^*; \quad (18)$$

$$\alpha_{r+1} = k_1 \alpha_r + k_2 \beta_r^*. \quad (19)$$

Умножив обе части соотношения (16) на A_{r+1} с учетом равенств (8) (при $n = r$), (10) и (13) (с заменой r на $r + 1$), получим выражение, завершающее $(r + 1)$ -ю итерацию:

$$\begin{bmatrix} g_0 \\ \vdots \\ g_{r-1} \\ g_r \end{bmatrix} = \begin{bmatrix} g_0 \\ \vdots \\ g_{r-1} \\ \gamma_r \end{bmatrix} + k_3 \begin{bmatrix} 0_r \\ \alpha_{r+1}^* \end{bmatrix}. \quad (20)$$

Для выполнения (20) необходимо, чтобы k_3 выбиралось из условия

$$g_r = \gamma_r + k_3 \alpha_{r+1}^*.$$

Таким образом,

$$k_3 = (g_r - \gamma_r) / \alpha_{r+1}^*. \quad (21)$$

Выражение (18) устанавливает соотношение между коэффициентами k_1 и k_2 , но не даёт основания для их однозначного выбора. В работе [4] предлагается принять $k_1 = \alpha_r^*$ (в [4] рассматривается не эрмитова, а симметричная действительная теплицева матрица A и там предлагается выбрать $k_1 = \alpha_r$, поскольку знак сопряжения в аналоге соотношения (18) отсутствует). Тогда в соответствии с (18) $k_2 = -\beta_r$ и на основании (19)

$$\alpha_{r+1} = |\alpha_r|^2 - |\beta_r|^2. \quad (22)$$

Однако такой выбор коэффициентов k_1 и k_2 является неудачным, поскольку при решении систем уравнений высокого порядка (при $n > 10$) приводит к появлению в процессе выполнения итераций экстремально больших или малых (выходящих, например, даже за пределы представимых в универсальном компьютере максимальных и минимальных по модулю чисел) значений α_r , β_r , $t_i^{(r)}$ и k_3 . В зависимости от вида матрицы A , при вытекающем из (11) соотношении

$$\beta_r = \sum_{i=1}^r a_i \cdot t_{r-i}^{(r)}$$

и соответствующем (15) правиле $t_i^{(r+1)} = \alpha_r^* t_i^{(r)} - \beta_r t_{r-i}^{(r)}$, вычисление по формуле (22) последующих значений α_{r+1} приводит к их экспоненциальному росту или уменьшению (в зависимости от задаваемого значением a_0 начального условия) параметра α_r . В ту же сторону и с такой же скоростью изменяются β_r и компоненты вектора t_r , а значение k_3 в соответствии с (21) изменяется аналогичным образом в обратном направлении. В результате при больших n алгоритм становится практически неработоспособным.

Ситуация меняется, если в (18) положить $k_1 = 1$. В этом случае

$$k_2 = -\beta_r / \alpha_r^*, \quad (23)$$

и в соответствии с (17)

$$\alpha_{r+1} = \alpha_r - |\beta_r|^2 / \alpha_r^*. \quad (24)$$

Отметим, что при начальном условии $\alpha_1 = 1$ все последующие, вычисляемые по формуле (24), значения α_r являются действительными и

$$\alpha_r^* = \alpha_r. \quad (25)$$

Выражения (11), (10), (15), (16), (21), (23), (25) и определяют алгоритм Левинсона, изображенный в виде блок-схемы, представленной на рис. 2.

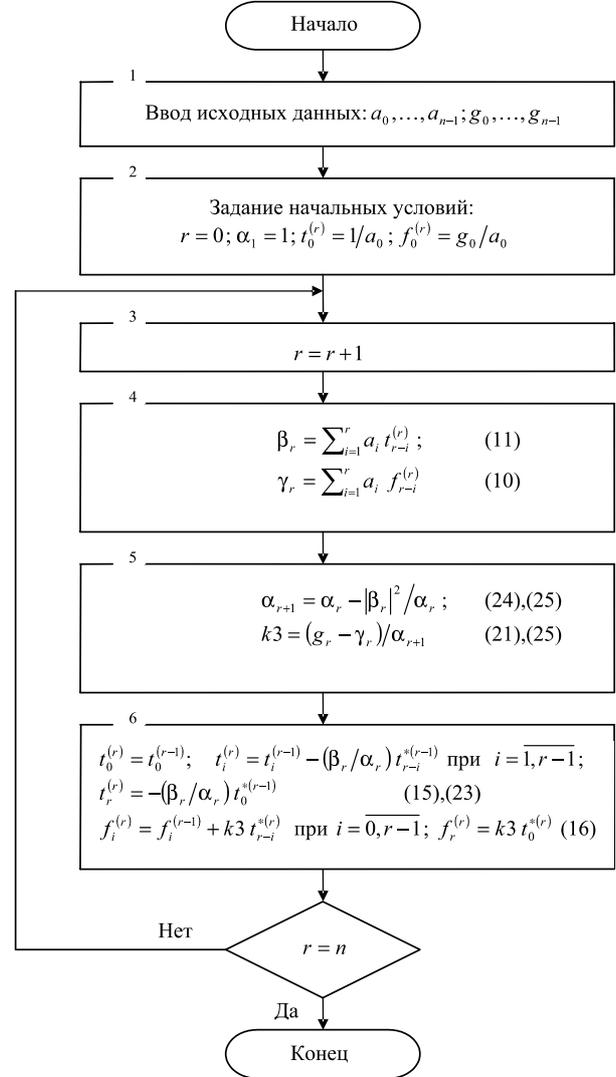


Рис. 2

Особенности использования алгоритма Левинсона при реализации метода нелинейной непосредственной оценки данных NDDE

Как отмечалось выше, возникающая при реализации метода NDDE необходимость многократного решения матричного уравнения (5) обуславливает высокую вычислительную сложность данной задачи. Однако её специфика и некоторые свойства рассмотренного алгоритма Левинсона обеспечивают возможности существенного повышения эффективности вычислений.

Одна из таких возможностей заключается в следующем. Как видно из уравнений (11), (15), (23) и (24), при решении матричного уравнения (1) методом Левинсона все промежуточные значения компонент вектора $t^{(r)}$, а также скалярных величин α_r и β_r полностью определяются элементами матрицы A и не зависят от вектора

g , представляющего правую часть (1). При этом отмеченные выше особенности матрицы R_m , фигурирующей в уравнении (5), приводят к тому, что при осуществлении в методе NDDE сокращения её размера путем удаления верхней и нижней строк и правого и левого боковых столбцов и последующем решении уравнения (5) меньшей размерности, получаемые в ходе выполнения итераций Левинсона значения $t^{(r)}$, α_r и β_r оказываются точно такими же, как и их одноимённые аналоги, полученные в ходе выполнения тех же итераций на предыдущем этапе (уменьшается лишь их количество, что связано с уменьшением общего числа итераций на данном этапе). Таким образом, при реализации метода NDDE полный объём вычислений, предусматриваемый алгоритмом, изображенным на рис.2, необходимо осуществить лишь на первом этапе, когда матрица R_m имеет максимальный размер. При этом все значения компонент вектора $t^{(r)}$ и переменных α_r и β_r , получаемые в ходе промежуточных итераций, следует занести в память вычислителя с тем, чтобы использовать их при выполнении последующих $(M/2 - 1)$ этапов. В результате на этих этапах при каждом новом решении уравнения (5) уменьшенной размерности в процессе выполнения итераций Левинсона необходимо заново вычислять лишь γ_r (формула (10)), k_3 (формула 21) и компоненты вектора решения $f^{(r)}$ (формула (16)), что значительно снижает общую вычислительную сложность реализации метода NDDE.

Дополнительного заметного сокращения вычислительных затрат можно добиться, если использовать еще

одно интересное свойство, которым обладает алгоритм Левинсона. Это свойство заключается в том, что если теплицева эрмитова матрица A в (1) является ленточной и положительно определённой (а именно такими свойствами и обладает фигурирующая в (5) матрица R_m), то начальные компоненты вектора решения $(f_0, f_1, \dots, f_m, m \ll M)$ с хорошей точностью определяются задолго до окончания итерационного процесса. На рис.3а и 3б показан характер изменения действительной части $\text{Re}(f_0^{(r)})$ значения $f_0^{(r)}$ (мнимая часть $\text{Im}(f_0^{(r)})$ изменяется аналогичным образом) в ходе выполнения итераций для случаев, когда матрица A является положительно определённой (рис.3а) и не обладает этим свойством (рис.3б). Размер $n \times n$ матрицы A , участвующей в иллюстрируемых этими рисунками процессах, а также количество её диагоналей (ширина ленты) $tt = 2L - 1$ выбраны исходя из практически важной ситуации, возникающей при реализации метода NDDE в соответствии с требованиями, задаваемыми стандартом [9], а именно: $n = M = 256$, $L = 16$ и $tt = 31$. В обоих случаях по окончании итераций $f_0^{(n)}$ принимает значение, равное верному решению (на рис.3а и 3б $\text{Re}(f_0^{(256)}) = -0,707$). При положительной определённости матрицы A с увеличением номера итерации наблюдается тенденция уменьшения (стремления к нулю) модулей реальной и мнимой частей участвующих в процессе переменных β_r и старших компонент $t_r^{(r)}$ вектора $t^{(r)}$; значение переменной α_r при этом стремится к некоторой const.

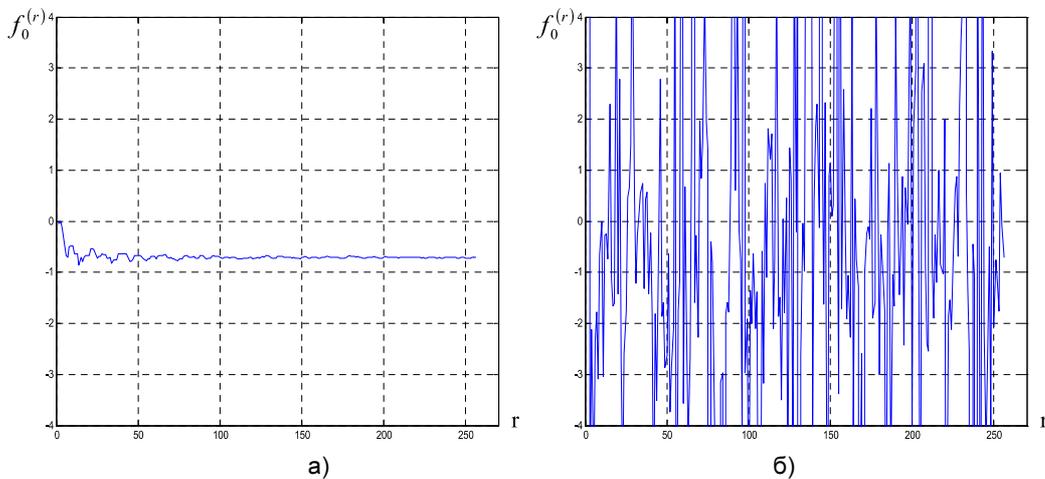


Рис. 3

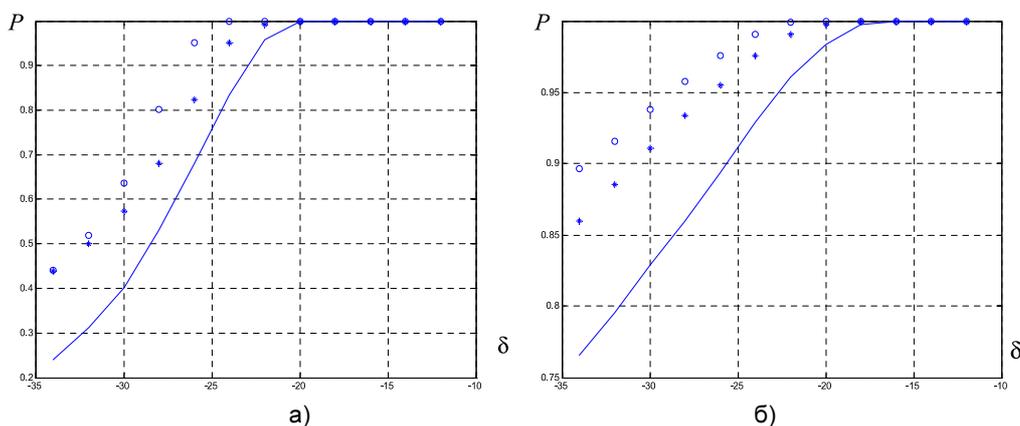


Рис. 4



Погрешность оценки значения начальной компоненты $f_0^{(r)}$ вектора решения $f^{(r)}$, получаемой при усечении числа итераций, является случайной величиной и зависит от конкретного вида матрицы A и вектора g в правой части уравнения (1). Количественные характеристики этой погрешности представлены на рис. 4а и 4 б, где изображены экспериментально полученные (путём моделирования на компьютере) графики её интегральной функции распределения $P(\delta)$. Аргумент δ этой функции представляет собой относительную погрешность оценки, выраженную в дБ:

$$\delta = 20 \lg \Delta / f_0,$$

где $\Delta = \hat{f}_0 - f_0$; \hat{f}_0 и f_0 соответственно оценка и истинное значение начальной компоненты вектора решения.

Поскольку величины \hat{f}_0 и f_0 являются комплексными, при расчетах и построении графиков использовались их либо действительные, либо мнимые части. Получающиеся при этом результаты оказались практически одинаковыми.

В качестве матрицы A в этих экспериментах использовалась матрица R_m , определяемая выражениями (6) и (3) (матрица M_1), причем необходимая для её формирования импульсная характеристика канала h задавалась в соответствии с требованиями, предъявляемыми стандартом [10] к КВ радиоканалу с умеренными (moderate) и плохими (poor) характеристиками. Рис.4а соответствует каналу moderate, а 4б – каналу poor. При этом значения M и L (а, следовательно, и tt) оставались такими же как и в процессе, иллюстрируемом рис.3а. Вектор g формировался в соответствии с уравнением (1), в котором в качестве вектора f использовался набор из n комплексных чисел, действительные и мнимые части которых с равной вероятностью принимали значения $\pm 0,7071$. (На практике это соответствует набору отсчетов сигнала с двукратной фазовой модуляцией). Основной линией на рис. 4а и 4б показаны зависимости, характеризующие точность оценки, получаемой на 48-й, линией из звёздочек – на 64-й и линией из круглых точек – на 80-й итерациях. Графики показывают, что с очень большой вероятностью при усечении числа итераций алгоритма Левинсона точность получаемой оценки начальной составляющей вектора решения является достаточно высокой.

Как отмечалось выше, при реализации метода NDDE на каждом его этапе необходимо определение как начальной f_0 , так и финальной f_{n-1} компонент вектора решения. Нетрудно показать, что если в уравнении (1) с треугольной эрмитовой матрицей A вектор $g^T = [g_0, g_1, \dots, g_{n-1}]$ заменить вектором $g_1^T = [g_{n-1}, g_{n-2}, \dots, g_0]$, то начальная компонента

решения f_1 вновь образованного матричного уравнения $Af_1 = g_1$ совпадёт с оригинальной компонентой f_{n-1} вектора решения уравнения (1). Поэтому рассмотренное свойство алгоритма Левинсона может быть использовано для определения обеих необходимых в методе NDDE составляющих вектора решения и обеспечивает возможность существенного сокращения вычислительной сложности его реализации.

Заключение

В работе выявлены особенности алгоритма Левинсона, позволяющие существенно сократить вычислительную сложность одного из наиболее эффективных методов построения радиомодемов – метода нелинейной прямой оценки данных.

Литература

1. Солонина А.И. и др. Основы цифровой обработки сигналов. Санкт-Петербург, «БВХ-Петербург», 2005
2. Марпл-мл. Цифровой спектральный анализ и его приложения / Пер. с англ. – М.: Мир, 1990.
3. Hsu Frank M. Data directed estimation techniques for single-tone HF modems. Military Communications Conference, 1985. MILCOM 1985. IEEE, volume:1, pp. 271-280.
4. Р. Блейхут. Быстрые алгоритмы цифровой обработки сигналов / Пер. с англ. – М.: Мир, 1989.
5. Петров О.А. Использование алгоритма Левинсона-Дурбина для решения системы линейных уравнений и «Инфокоммуникационные технологии», том 1, №1, 2003г.
6. Петров О.А. Быстрый алгоритм решения систем уравнения с треугольной матрицей. «Инфокоммуникационные технологии», том 4, №1, 2006г.
7. Воеводин В.В., Кузнецов Ю.А. Матрицы и вычисления. – М.:Наука, 1984. – 320с.
8. STANAG 4285 Characteristics of 1200/2400/3600 Bit Per Second Single Tone Modulators/ Demodulators for HF Radio Links.
9. STANAG4539. Technical standards for Non-Hopping HF Communications Waveforms.
10. ITU_R F.520-2. Use of high frequency ionospheric channel simulators.

APPLICATION OF LEVINSON'S ALGORITHM AT REALIZATION DATA DIRECTED ESTIMATION METHOD IN RADIO MODEMS

Features of Levinson's algorithm are considered at the repeated decision of the matrix equation of the high order during realization of data direct estimation for single-tone HF modems, and the decisions allowing essentially to reduce its computing complexity are offered.

УДК 621.396

АДАПТИВНАЯ ПРОСТРАНСТВЕННО-ВРЕМЕННАЯ ОБРАБОТКА ФМ-СИГНАЛОВ В МНОГОЛУЧЕВЫХ КАНАЛАХ СВЯЗИ НА ОСНОВЕ АЛГОРИТМА КАЛМАНА

Тарасов Г.А., к.т.н., доцент, в.н.с. ФГНУ ГИИТЦ «Наука», tarasov_ga@mail.ru

Тарасов А.Г., к.т.н., в.н.с. ФГНУ ГИИТЦ «Наука»

Ключевые слова: канал связи, многолучевой, адаптивная обработка, алгоритм Калмана, антенная система, разнесенный прием, фазомодулированный сигнал.

Введение

Привлекательность адаптивных антенных систем (ААС) применительно к прикладным задачам радиоприема состоит в том, что ААС имеют потенциальную возможность фокусировки на полезный сигнал и формирования глубоких нулей диаграммы направленности в направлении прихода станционных помех [7]. Кроме того, при использовании диаграммообразующих устройств (ДОУ) на линиях задержки (ЛЗ) с отводами ААС имеют потенциальную возможность раздельного приема лучей [7, 13]. Особенную актуальность эти потенциальные возможности ААС приобретают в задачах приема ФМ сигналов в многолучевых каналах связи.

Значительное внимание, уделяемое фирмами производителями системам связи с фазовой модуляцией в последние десятилетия, обусловлено целым рядом обстоятельств. Структура ФМ сигналов обеспечивает удобство выделения тактовых импульсов [6] и позволяет выделять когерентную несущую с помощью легко реализуемых процедур восстановления несущей по принимаемым решениям [6]. Кроме того, для ФМ сигналов разработаны эффективные алгоритмы компенсации межсимвольной интерференции (МСИ) [6], которые позволяют отслеживать относительно быстрые изменения характеристик канала во времени. Указанные достоинства ФМ сигналов обусловили их широкое распространение в современных системах связи. Вместе с тем, ФМ сигналы весьма уязвимы к эффектам многолучевого распространения радиоволн. Тенденция повышения скорости передачи привела к тому, что для ФМ сигналов длительность информационной посылки зачастую оказывается значительно меньше задержки между лучами многолучевого сигнала. В этих условиях весьма актуальной является задача повышения качества приема ФМ сигналов в многолуче-

Рассмотрен прикладной вопрос использования адаптивных антенных систем для повышения качества приема фазомодулированных (ФМ) сигналов в многолучевых каналах связи. Предложенный подход обеспечивает адаптивное формирование оптимальной диаграммы направленности, компенсацию межсимвольной интерференции и демодуляцию в виде единой процедуры пространственно-временной обработки с адаптацией по критерию минимума среднего квадрата ошибки на выходе демодулятора.

вых каналах связи за счет использования потенциальных возможностей ААС.

Адаптивная пространственно-временная обработка ФМ-сигналов

Структурная схема ААС для приема ФМ сигналов на основе критерия минимума среднего квадрата ошибки (МСКО) представлена на рис. 1. Сигналы с выходов антенных элементов (АЭ) A_1, A_2, \dots, A_N подаются на входы ДОУ, где они взвешиваются и суммируются. Управление весовыми коэффициентами ДОУ осуществляется с помощью адаптивного процессора, который выполняет оценивание весовых коэффициентов по сигналам с выходов АЭ и сигналу ошибки с выхода демодулятора [7].

В общем случае электрические сигналы на выходах АЭ содержат аддитивную смесь сигналов, соответствующих различным лучам многолучевого сигнала, лучам многолучевой помехи, и шумов. Соответственно, вектор наблюдения на выходах АЭ можно представить в следующем виде [7]:

$$X(i) = \sum_{m=1}^M C_m(i) + \sum_{k=1}^K \Pi_k(i) + H(i), \quad (1)$$

$$\text{где } C_m(i) = \sqrt{p_{cm}} c_m(i) V_{cm}(i); \quad (2)$$

$$c_m^*(i) c_m(i) = 1; \quad V_{cm}^+(i) V_{cm}(i) = N; \quad (3)$$

$$\Pi_k(i) = \sqrt{p_{nk}} n_k(i) V_{nk}(i); \quad (4)$$

$$n_k^*(i) n_k(i) = 1; \quad V_{nk}^+(i) V_{nk}(i) = N; \quad (5)$$

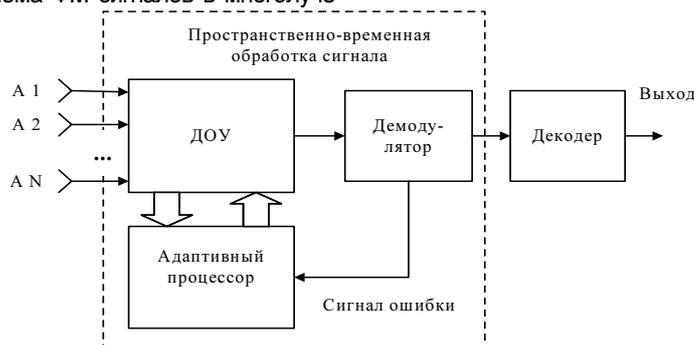


Рис. 1. Структурная схема ААС на основе критерия МСКО для приема ФМ сигналов.

Здесь: $C_m(i)$ - комплексный вектор m -го луча сигнала на выходах АЭ; $c_m(i)$ - комплексный нормированный сигнал m -го луча в фазовом центре АР; P_{cm} - средняя мощность m -го луча сигнала; $V_{cm}(i)$ - вектор запаздывания m -го луча сигнала на раскрые АР; M - число лучей многолучевого сигнала; $\Pi_k(i)$ - комплексный вектор k -го луча помехи на выходах АЭ; $n_k(i)$ - комплексный нормированный сигнал k -го луча помехи в фазовом центре АР; P_{nk} - средняя мощность k -й помехи; $V_{nk}(i)$ - вектор запаздывания k -го луча помехи на раскрые АР; K - число лучей многолучевой помехи; $H(i)$ - комплексный вектор шумов (шум в различных каналах приема предполагается взаимно некоррелированным с одинаковой мощностью); i - номер временного отсчета; N - число АЭ; * - знак комплексного сопряжения; + - знак эрмитова сопряжения.

Векторы запаздывания каждого из лучей сигнала (2) и помехи (4) можно представить следующим образом:

$$V(\varphi, \Delta, i) = [v_1(\varphi, \Delta, i), \dots, v_n(\varphi, \Delta, i), \dots, v_N(\varphi, \Delta, i)]^T, \quad (6)$$

$$\text{где } v_n(\varphi, \Delta, i) = |v_n(i)| \exp\{j[\theta_n(\varphi, \Delta) + \tilde{\theta}_n(i)]\}; \quad (7)$$

$$\theta_n(\varphi, \Delta) = \alpha_0 \tau_n; \quad (8)$$

φ - азимут принимаемого луча ($0^\circ \leq \varphi < 360^\circ$); Δ - угол места принимаемого луча ($0^\circ \leq \Delta < 90^\circ$); $v_n(\varphi, \Delta, i)$ - компонента вектора запаздывания, соответствующая n -му АЭ; $|v_n(i)|$ - множитель, учитывающий амплитудные замирания в n -м АЭ; $\theta_n(\varphi, \Delta)$ - регулярная составляющая фазы принимаемого луча в n -м АЭ; $\tilde{\theta}_n(i)$ - флюктуационная составляющая фазы принимаемого луча в n -м АЭ; α_0 - средняя частота спектра сигнала; τ_n - время запаздывания принимаемого луча в n -м АЭ относительно фазового центра антенной решетки (начала координат); T - знак транспонирования.

Структурная схема ДООУ ААС с комплексным взвешиванием сигналов с отводов линии задержки показана на рис. 2.

Как видно из рис. 2, отсчеты комплексных сигналов с выходов АЭ $x_1(i), \dots, x_n(i), \dots, x_N(i)$ поступают на входы линий задержки с L отводами. При дискретной обработке временная задержка между сигналами на соседних отводах равна периоду частоты дискретизации. Вы-

ходной сигнал ААС получают путем умножения сигналов с выходов ЛЗ на комплексный весовой вектор [7]. В векторно-матричных обозначениях выходной сигнал ААС $y(i)$ определяется выражением

$$y(i) = W^T(i)X(i), \quad (9)$$

где:

$$W^T(i) = [w_{11}, w_{12}, \dots, w_{1L}, w_{n1}, w_{n2}, \dots, w_{nL}, w_{N1}, w_{N2}, \dots, w_{NL}]^T, \quad (10)$$

$$X^T(i) = [x_{11}, x_{12}, \dots, x_{1L}, x_{n1}, x_{n2}, \dots, x_{nL}, x_{N1}, x_{N2}, \dots, x_{NL}]^T.$$

ДООУ с комплексным взвешиванием сигналов с отводов линии задержки (рис. 2) объединяет в себе функции пространственного и частотно-временного фильтра [7], что делает эту структуру весьма привлекательной для борьбы с эффектами многолучевого распространения радиоволн.

Проблеме синтеза оптимальных алгоритмов адаптации антенных систем посвящено значительное количество публикаций [3, 4, 5, 7, 8, 11, 13, 14]. Для синтеза привлекаются самые разнообразные математические методы. Широкое распространение получили методы стохастической аппроксимации, методы случайного поиска, метод неопределенных множителей Лагранжа, метод непосредственного, либо рекуррентного обращения выборочной корреляционной матрицы [7]. В настоящей работе рассмотрен алгоритм адаптации ААС, основанный на использовании редуцированного фильтра Калмана [7, 11].

Для постоянной СПО уравнения алгоритма Калмана могут быть получены путем рекуррентного обращения выборочной корреляционной матрицы (ВКМ) в уравнении Винера-Хопфа [7]. В научной литературе эта процедура носит название "алгоритм РНК" (рекурсивный наименьших квадратов). Алгоритм РНК может быть получен непосредственно из уравнения Винера-Хопфа с использованием матричного тождества Вудбери [7, 14].

Как известно, оптимальный ВВК по критерию МСКО на k -м шаге на L отсчетах определяется из уравнения Винера-Хопфа:

$$W_L(k) = R_L^{-1}(k)P_L(k), \quad (11)$$

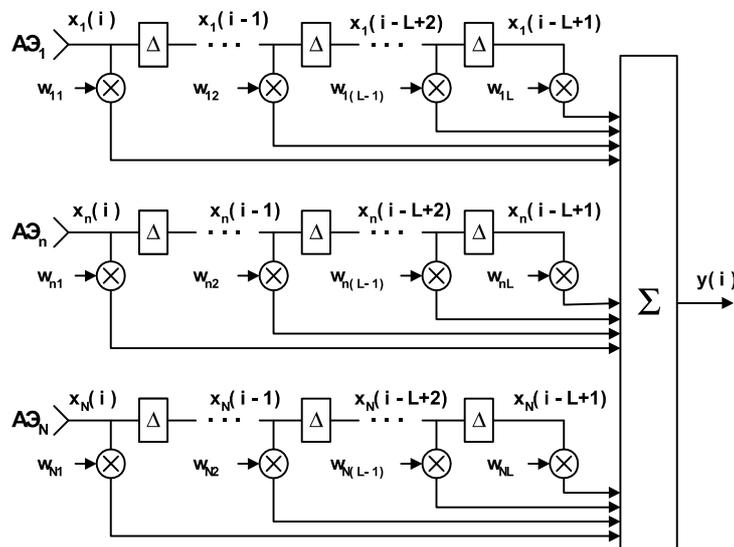


Рис. 2. Структурная схема ДООУ ААС с комплексным взвешиванием сигналов с отводов линии задержки.

где $R_L(k) = \sum_{j=k-L+1}^k X^*(j)X^T(j)$ (12)

выборочная корреляционная матрица (ВКМ) на k -м шаге на L отсчетах; $*$ – знак комплексного сопряжения;

$P_L(k) = \sum_{j=k-L+1}^k X^*(j)z_{эм}(j)$ (13)

вектор взаимной корреляции с эталонным сигналом $z_{эм}(j)$.

Для упрощения дальнейших выкладок в выражениях (12), (13) опущен несущественный нормирующий множитель $1/L$. Корреляционную матрицу на k -м шаге на $L+1$ отсчетах можно представить следующим образом [7]

$R_{L+1}(k) = \sum_{j=k-L}^k X^*(j)X^T(j) = \sum_{j=k-L}^{k-1} X^*(j)X^T(j) + X^*(k)X^T(k) = (14)$
 $= R_L(k-1) + X^*(k)X^T(k)$.

Аналогично для корреляционного вектора получаем [7]

$P_{L+1}(k) = \sum_{j=k-L}^k X^*(j)z_{эм}(j) = P_L(k-1) + X^*(k)z_{эм}(k)$. (15)

Для нахождения матрицы обратной (12) воспользуемся матричным тождеством Вудбери [2, 7]

$[aR + bX^*X^T]^{-1} = \frac{1}{a}R^{-1} - \frac{b}{a^2} \left[I + \frac{b}{a}X^TR^{-1}X^* \right]^{-1} R^{-1}X^*X^TR^{-1}$. (16)

Для случая, когда $a = b = 1$, формула (16) упрощается и принимает следующий вид

$[R + X^*X^T]^{-1} = R^{-1} - [I + X^TR^{-1}X^*]^{-1} R^{-1}X^*X^TR^{-1}$. (17)

С учетом (17) выражение (14) можно преобразовать к виду [7]

$R_{L+1}^{-1}(k) = [R_L(k-1) + X^*(k)X^T(k)]^{-1} = (18)$
 $= \left[R_L^{-1}(k-1) - \frac{R_L^{-1}(k-1)X^*(k)X^T(k)R_L^{-1}(k-1)}{1 + \bar{X}^T(k)R_L^{-1}(k-1)\bar{X}(k)} \right]$

С учетом (11), (12) и (14) получаем выражение для оптимального весового вектора на k -м шаге на $L+1$ -й выборках [7]

$\bar{W}_{L+1}(k) = R_{L+1}^{-1}(k)P_{L+1}(k) = (19)$
 $= \left[R_L^{-1}(k-1) - \frac{R_L^{-1}(k-1)X^*(k)X^T(k)R_L^{-1}(k-1)}{1 + v(k-1)} \right] [P_L(k-1) + X^*(k)z_{эм}(k)] =$
 $= W_L(k-1) + K_L(k)[z_{эм}(k) - X^T(k)W_L(k-1)]$,

где $v(k-1) = X^T(k)R_L^{-1}(k-1)X^*(k)$, (20)

$K_L(k) = \frac{R_L^{-1}(k-1)X^*(k)}{1 + X^T(k)R_L^{-1}(k-1)X^*(k)}$. (21)

Итак, при добавлении новой выборки данных получаем следующие уравнения [7, 14]

$R_{L+1}^{-1}(k) = R_L^{-1}(k-1) - K_L(k)X^T(k)R_L^{-1}(k-1)$, (22)

$W_{L+1}(k) = W_L(k-1) + K_L(k)[z_{эм}(k) - X^T(k)W_L(k-1)]$, (23)

$K_L(k) = \frac{R_L^{-1}(k-1)X^*(k)}{1 + X^T(k)R_L^{-1}(k-1)X^*(k)}$, (24)

Аналогично, при отбрасывании старой выборки данных получаем следующие уравнения [7, 14]:

$R_L^{-1}(k) = R_{L+1}^{-1}(k) + K_{L+1}(k)X^T(k-L)R_{L+1}^{-1}(k)$, (25)

$W_L(k) = W_{L+1}(k) - K_{L+1}(k)[z_{эм}(k-L) - X^T(k-L)W_{L+1}(k)]$, (26)

$K_{L+1}(k) = \frac{R_{L+1}^{-1}(k)X^*(k-L)}{1 - X^T(k-L)R_{L+1}^{-1}(k)X^*(k-L)}$. (27)

Таким образом, оценивание ВВК ААС может быть выполнено путем решения векторно-матричных уравнений (22) – (27). В начальный момент времени вычисляются $R_L^{-1}(0)$ и $W_L(0)$. Затем вычисляется вектор $K_L(1)$ (24), который используется при добавлении новой выборки данных (22), (23). Полученное значение $R_{L+1}^{-1}(1)$ используется для вычисления вектора $K_{L+1}(1)$ (27), который, в свою очередь, используется при вычислении значений $R_L^{-1}(1)$ (25) и $W_L(1)$ (26) на первом шаге. Далее процедура повторяется. Достоинство алгоритма (22) – (27) состоит в том, что оценка ВВК зависит лишь от конечного числа предыдущих выборок данных. В рассматриваемом случае размер выборки (окна) равен L . Параметр L решающим образом определяет скорость сходимости рассматриваемого алгоритма и качество пространственно-временной обработки сигналов и помех в установившемся режиме.

В рамках настоящей работы предлагается подход, основанный на допущении постоянства параметров СПО на коротких временных интервалах, сопоставимых со скоростью сходимости алгоритма Калмана. Под постоянством СПО понимается, прежде всего, постоянство разностных амплитудно-фазовых параметров на выходах АЭ. Для постоянной СПО на интервале длительностью L отсчетов уравнения дискретного алгоритма Калмана определяются выражениями (22) – (27). Эти уравнения описывают процедуру оценивания оптимального ВВК ААС на скользящем окне длительностью L отсчетов. Параметр k определяет момент времени, для которого рассчитывается оптимальное значение ВВК. Суть предлагаемой процедуры адаптивного оценивания параметров СПО сводится к адаптивному оцениванию параметра L_{opt} для заданного момента времени k .

Математически предлагаемая процедура адаптации формулируется как задача минимизации целевой функции следующего вида:

$Min_{K,L} \left\{ \varepsilon^2(k, K, L) \right\} = Min_{K,L} \left\{ \frac{1}{K} \sum_{j=k-K+1}^k \varepsilon^2(j, L) \right\}$, (28)

$L_{min} \leq L \leq L_{max}$, $k = const$

где: $\varepsilon^2(k, K, L)$ - средний квадрат ошибки в k -й момент времени для фильтра Калмана с окном L отсчетов, полученный на K отсчетах; $\varepsilon^2(j, L)$ - мгновенное значение квадрата ошибки для фильтра Калмана с окном L отсчетов в j -й момент времени; $L_{min} \dots L_{max}$ - диапазон изменения размера окна алгоритма Калмана.

Параметр K , входящий в выражение (28), определяет инерционность цепей оценивания параметров СПО. Процедуру поиска минимума целевой функции (28) целесообразно разбить на два этапа. На первом этапе решается задача оптимизации размера окна L . На втором этапе решается задача оптимизации значения параметра K .

На первом этапе предположим, что параметр K фиксирован. После фиксирования значения параметра K двухпараметрическая целевая функция (28) становится однопараметрической, и задача поиска ее минимума существенно упрощается.

Ввиду того, что целевая функция (28) вычисляется на конечном временном интервале в условиях шумов, в общем случае она имеет "зашумленный" вид и может оказаться многомодальной. Для этого общего случая процедура поиска предполагает два этапа. На первом этапе выполняется грубый поиск локальных экстремумов методом "оптимальных покрытий" [1] в узлах сетки:

$$L_1 = L_{min}, L_2, \dots, L_{M-1}, L_M = L_{max}, \quad (29)$$

где M - число узлов сетки, в которых осуществляется вычисление значений целевой функции (28). После вычисления значений $\hat{\varepsilon}_m$ целевой функции (28) в узлах сетки L_m выполняется поиск локальных экстремумов, из которых выбирается глобальный. После этого задача поиска глобального экстремума многомодальной функции сводится к одномодальному случаю, и в окрестности найденного глобального экстремума делается более точный поиск экстремального значения целевой функции [1].

Окончательный ответ о многомодальности или унимодальности целевой функции и о выборе конкретной процедуры поиска ее экстремумов может быть дан лишь на основании детального исследования ее характерных особенностей [1, 9, 10]. В случае с рассматриваемой целевой функцией в ходе проведенных экспериментальных исследований было установлено, что она является унимодальной и достаточно хорошо аппроксимируется квадратичной функцией при использовании логарифмических осей по оси абсцисс и оси ординат. Типичный характер целевой функции (28) при усреднении на блоке передачи показан на рис. 3.

Процедура квадратичной аппроксимации является весьма эффективной процедурой поиска экстремумов унимодальных функций и позволяет существенно сократить объем вычислений благодаря тому, что после выполнения аппроксимации по значениям в узлах сетки, экстремум может быть найден расчетным путем без перехода к более мелкой сетке. В основе процедуры аппроксимации лежит сглаживание по критерию наименьших квадратов. Поэтому второе достоинство квадратичной аппроксимации состоит в том, что за счет сглаживания происходит ослабление влияния "зашумленности" целевой функции (28) на точность оценивания оптимального размера окна L_{opt} .

Средний квадрат ошибки аппроксимации целевой функции (28) в узлах сетки определяется выражением [12]:

$$\varepsilon^2 = \sum_{i=1}^M \left[\hat{\varepsilon}_i^2 - k_1 (\lg L_i)^2 - k_2 (\lg L_i) - k_3 \right]^2, \quad (30)$$

где $\hat{\varepsilon}_i$ - значения целевой функции (28) в узлах сетки L_i ; k_1, k_2, k_3 - параметры аппроксимирующей параболы.

Взяв от (30) частные производную по каждому из параметров k_1, k_2, k_3 и приравняв их к нулю [12], получаем следующую систему линейных уравнений:

$$\begin{cases} k_1 \sum_{m=1}^M (\lg L_i)^4 + k_2 \sum_{m=1}^M (\lg L_i)^3 + k_3 \sum_{m=1}^M (\lg L_i)^2 = \sum_{m=1}^M (\lg x_i)^2 \hat{\varepsilon}_i \\ k_1 \sum_{m=1}^M (\lg L_i)^3 + k_2 \sum_{m=1}^M (\lg L_i)^2 + k_3 \sum_{m=1}^M (\lg L_i) = \sum_{m=1}^M (\lg x_i) \hat{\varepsilon}_i \\ k_1 \sum_{m=1}^M (\lg L_i)^2 + k_2 \sum_{m=1}^M (\lg L_i) + k_3 = \sum_{m=1}^M \hat{y}_i \end{cases} \quad (31)$$

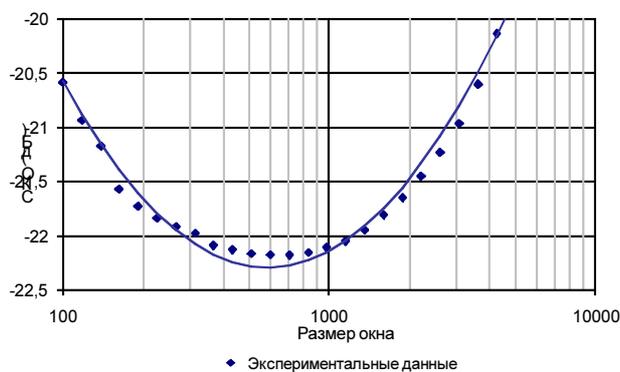


Рис. 3. Типичный вид целевой функции (28) при усреднении на блоке передачи.

Решив систему линейных уравнений (31), можно найти параметры сглаживающей параболы k_1, k_2, k_3 . Оптимальное значение размера окна L_{opt} , обеспечивающее минимизацию целевой функции, совпадает с координатой минимума сглаживающей параболы:

$$L_{opt} = -\frac{k_2}{2k_1}. \quad (32)$$

Предложенный способ пространственно-временной обработки сигналов в многолучевых каналах связи был реализован и испытан на реальных каналах связи применительно к сигналам с тройной фазовой модуляцией. Один из типичных случаев приема сигнала с тройной фазовой модуляцией в условиях многолучевости проиллюстрирован на рис. 4.

Слева на рис. 4 а-d приведены взаимно корреляционные функции сигналов с выходов АЭ с сигналом синхровставки. Как видно из этих рисунков, в первом и третьем каналах присутствуют по три явно выраженных луча и один-два более слабых луча, во втором и четвертом каналах – по два явно выраженных луча и один-два более слабых луча.

Справа на рис. 4 а-d показаны сигнальные созвездия на выходах демодуляторов при одноканальном приеме для тех же моментов времени. Взаимно корреляционная функция выходного сигнала ААС и соответствующее сигнальное созвездие приведено на рис. 4 е. Как видно из рис. 4 е, выходной сигнал ААС содержит лишь один луч.

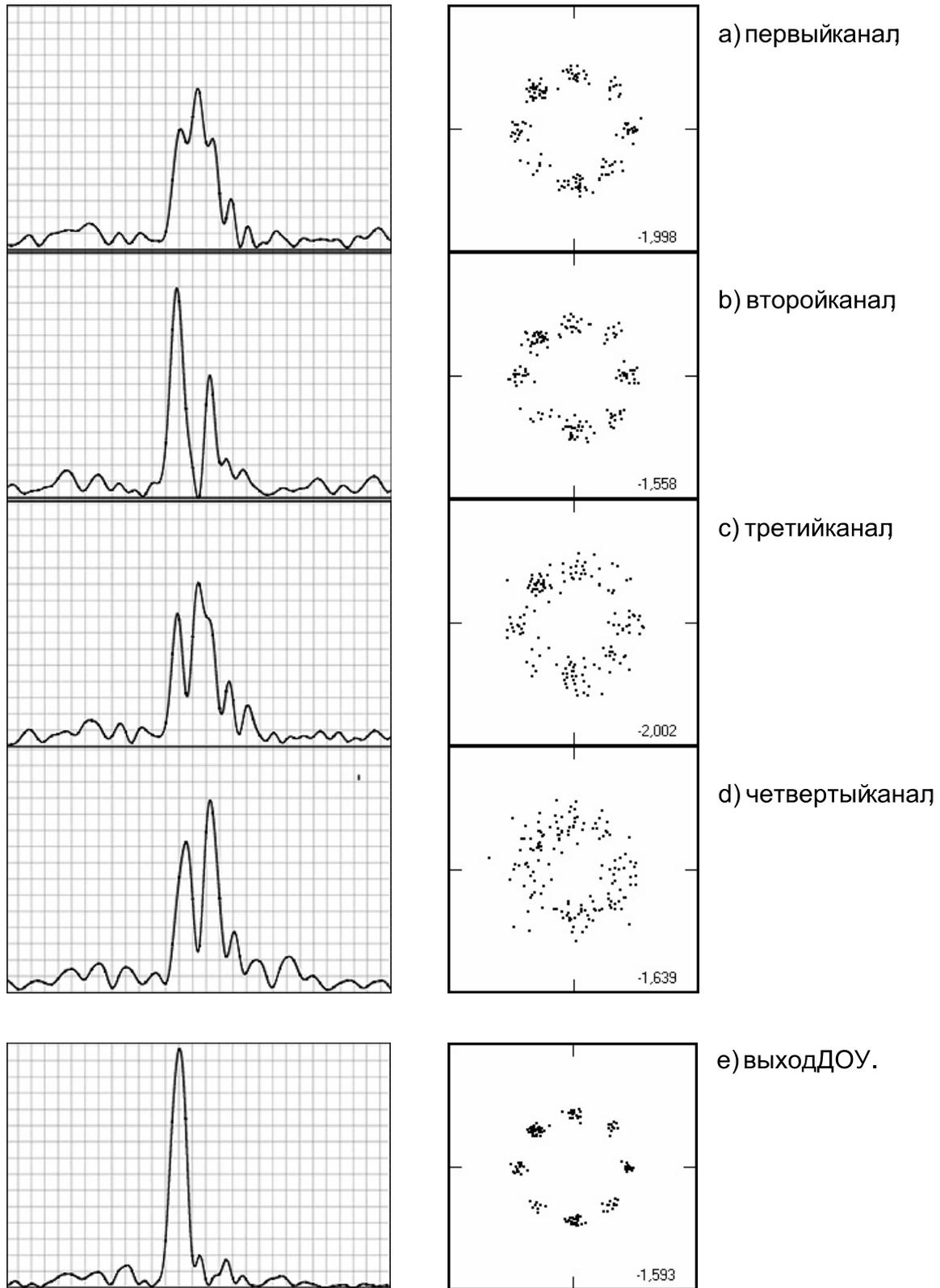


Рис. 4. Выходные сигналы коррелятора и сигнальные созвездия в различных каналах (a-d) и на выходе ДОУ (e).

Из сравнения сигнальных созвездий в каналах (рис. 4 a-d) и на выходе ААС (рис. 4 e) видно, что адаптивная пространственно-временная обработка сигналов с выходов АЭ позволила существенно снизить СКО на выходе демодулятора.

Таким образом, проведенные испытания на реальных сигналах показали, что предложенный

способ пространственно-временной обработки ФМ сигналов ВЧ диапазона с адаптацией по критерию МСКО на выходе демодулятора, обладает потенциальной возможностью разнесения лучей и повышения эффективности разнесенного приема за счет когерентного взвешивания разделенных лучей.

Заключение

Итак, в статье рассмотрен прикладной вопрос использования адаптивных антенных систем для повышения качества приема ФМ сигналов в многолучевых каналах связи. Главная особенность предложенного способа повышения качества приема ФМ сигналов состоит в том, что адаптивное формирование оптимальной ДН, компенсация МСИ и демодуляция выполняются в виде единой процедуры пространственно-временной обработки с адаптацией по критерию МСКО на выходе демодулятора. Для оценивания ВВК ААС предложен редуцированный алгоритм Калмана с адаптивным оцениванием размера окна в соответствии с изменяющейся СПО.

Экспериментальная часть исследования выполнена на реальных сигналах. Результаты испытаний показали, что предложенный способ пространственно-временной обработки ФМ сигналов обеспечивает повышение качества приема за счет когерентного взвешивания сигналов с разнесенных антенн и обладает потенциальной возможностью раздельного приема лучей многолучевых сигналов.

Статья выполнена при поддержке Гранта Президента РФ № МК - 1.2008.10.

Литература

1. Батищев Д.И. Методы оптимального проектирования.- М.: Радио и связь, 1984.- 248 с.
2. Беллман Р. Введение в теорию матриц/ Пер. с англ. Под ред. В.Б.Лидского.- М.: Наука, 1976.- 351 с.
3. Гейбриел В. Введение в теорию адаптивных антенных решеток// ТИИЭР, 1976, т.64, №2.- С. 55-95.
4. Журавлёв А.К., Хлебников В.А., Родимов А.П. и др. Адаптивные радиотехнические системы с антенными решетками.- Л.: Изд. ЛГУ, 1991.- 544 с.
5. Кловский Д.Д., Соيفер В.А. Обработка пространственно-временных сигналов.- М.: Связь, 1976.- 207 с.
6. Куреши Ш.У.Х. Адаптивная коррекция// ТИИЭР.- 1985.- Т. 73.- № 2.- С. 5-49.

7. Монзинго Р.А., Миллер Т.У. Адаптивные антенные решетки: Введение в теорию/ Пер. с англ. Под ред. В.А. Лексаченко. - М.: Радио и связь, 1986.- 446 с.
8. Поповский В.В. Адаптивные антенные решетки.- Л.: ВАС, 1986.- 60 с.
9. Реклейтис Г., Рейвиндран А., Регсдел К. Оптимизация в технике. Т. 1, М.: Мир, 1986, 349 с.
10. Реклейтис Г., Рейвиндран А., Регсдел К. Оптимизация в технике. Т. 2, М.: Мир, 1986, 318 с.
11. Родимов А.П. и др. Исследование эффективности и чувствительности алгоритмов пространственно-временной обработки сигналов и помех, синтезированных по методу переменных состояния// Вопросы кибернетики (теория чувствительности и ее применение).- М.: АН СССР, 1981.- Вып. 74.- С. 171-191.
12. Тихонов В.И. Статистическая радиотехника.- М.: Радио и связь, 1982.- 622 с.
13. Уидроу Б. и др. Адаптивные антенные системы.- ТИИЭР, 1967, т.55, №12.- С. 78-95.
14. Уидроу Б., Стирнз С. Адаптивная обработка сигналов: Пер. с англ.- М.: Радио и связь, 1989.- 440 с.

ADAPTIVE TIME-SPATIAL PROCESSING OF PM SIGNALS IN MULTIPATH CHANNELS BASED ON KALMAN'S ALGORITHM

Tarasov G.A., Tarasov A.G.

The article contemplates the applied problem of using of adaptive antennas systems to raise quality of reception of phase-modulated (PM) signals in multipath channels. Introduced approach provides adaptive optimal beam-forming, compensation of intersymbol interference and demodulation in single procedure of time-spatial processing minimizing error mean square in output signal of demodulator.

Уважаемые авторы!

Редакция научно-технического журнала "Цифровая обработка сигналов" просит Вас соблюдать следующие требования к материалам, направляемым на публикацию:

1) Требования к текстовым материалам и сопроводительным документам:

- Текст - текстовый редактор Microsoft Word.
- Таблицы и рисунки должны быть пронумерованы. На все рисунки, таблицы и библиографические данные указываются ссылки в тексте статьи.
- Объем статьи до 12 стр. (шрифт 12). Для заказных обзорных работ объем может быть увеличен до 20 стр.
- Название статьи на русском и английском языках.
- Рукопись статьи сопровождается:
 - краткой аннотацией на русском и английском языках;
 - номером УДК;
 - сведениями об авторах (Ф.И.О., организация, телефоны, электронная почта).

2) Требования к иллюстрациям:

Векторные (схемы, графики) - желательно использование графических редакторов Adobe Illustrator или Corel DRAW.

- Растровые (фотографии, рисунки) - М 1:1, разрешение не менее 300dpi, формат tiff.

УДК 621.391.82

ПОВЫШЕНИЕ ПОМЕХОУСТОЙЧИВОСТИ ПЕРЕДАЧИ ДАННЫХ ПО ЦИФРОВЫМ КАНАЛАМ СВЯЗИ В УСЛОВИЯХ РЕЛЕЕВСКИХ ЗАМИРАНИЙ

Андреанов М.Н., в.н.с. НПП «Цифровые решения»

Бумагин А.В., к.т.н., начальник сектора ОАО «Российские космические системы»

Гондарь А.В., инженер ОАО «Российские космические системы»

Калашиников К.С., инженер ОАО «Российские космические системы»

Прудников А.А., инженер ОАО «Российские космические системы»

Пучков Г.А., ведущий инженер ОАО «Российские космические системы»

Стещенко В.Б., к.т.н., начальник отдела ОАО «Российские космические системы», steshenko@steshenko.ru

Ключевые слова: помехоустойчивость, замирания, релейский канал, разнесенный прием, сигнал, многолучевое распространение.

Введение

Особую сложность при построении современных систем связи представляет организация приема сигнала в городских условиях при многолучевом распространении сигнала и релейских замираниях. Многолучевое распространение сигнала приводит к его значительному искажению и усложняет задачу синхронизации и демодуляции, снижая тем самым помехоустойчивость системы.

В работе исследован вопрос влияния замираний на помехоустойчивость цифровых систем связи в условиях обобщенного релейского канала, а также проанализированы известные методы разнесенного приема с точки зрения повышения помехоустойчивости.

Влияние замираний на помехоустойчивость цифровых систем связи

Сигналы с полосой менее полосы когерентности канала на выходе многолучевого канала, вследствие медленных релейских замираний, имеют более низкую помехоустойчивость относительно помехоустойчивости сигналов в гауссовом шуме. На рис.1 представлены сравнительные зависимости вероятностей ошибок приема сигналов в гауссовом шуме от отношения энергии сигнала к спектральной плотности шума и при релейских замираниях от среднего значения указанного отношения при когерентном и некогерентном приемах. Сигналы имеют двухпозиционную модуляцию (ФМ-2, ОФМ-2). Формулы (1), (2) описывают вероятность ошибки когерентного и некогерентного приема в гауссовом шуме, а формулы (3), (4) - при релейских замираниях

$$p_1 = \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha\gamma}), \quad (1)$$

$$p_2 = \frac{1}{2} \exp(-\alpha\gamma), \quad (2)$$

Исследуются вопросы воздействия замираний на помехоустойчивость цифровой связи в обобщенном релейском канале, что особенно актуально при организации связи в городских условиях, когда имеет место многолучевое распространение сигналов. Оценивается выигрыш в помехоустойчивости от использования известных методов разнесенного приема.

$$p_3 = \frac{1}{2} \cdot \left(1 - \frac{1}{\sqrt{1 + \frac{1}{\alpha\gamma_0}}} \right), \quad (3)$$

$$p_4 = \frac{1}{2} \cdot \frac{1}{1 + \alpha\gamma_0}, \quad (4)$$

$$\text{где } \operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \exp(-t^2) dt, \quad (5)$$

γ – отношение сигнал/шум,

γ_0 – среднее отношение сигнал/шум.

Из анализа кривых, представленных на рис.1, следует, что помехоустойчивость когерентного и некогерентного приема сигналов при медленных релейских замираниях значительно ниже помехоустойчивости приема сигналов в гауссовом шуме. С увеличением отношения сигнал/шум (ОСШ) в канале связи это различие растет экспоненциально.

В обобщенном релейском канале связи помимо диффузионной составляющей сигнала присутствует регулярная компонента. Плотность распределения вероятности (ПРВ) отношения сигнал-шум в таком канале описывается выражением (6), вывод которого приведен в приложении.

$$f_\gamma = \frac{1}{\gamma_0 - \gamma_a} \cdot \exp\left(-\frac{\gamma + \gamma_a}{\gamma_0 - \gamma_a}\right) \cdot I_0\left(\frac{2 \cdot \sqrt{\gamma \cdot \gamma_a}}{\gamma_0 - \gamma_a}\right). \quad (6)$$

где γ_0 – среднее ОСШ в обобщенном релейском канале, γ_a – ОСШ для регулярной составляющей сигнала. На рис.2 представлены вычисленные плотности вероятности при $\gamma_a = 9$ и $\gamma_0 = 10, 12, 20, 30, 40$. Для соблюдения условия ($a \gg \sigma$) γ_a необходимо выбирать значительным (не вблизи нуля) и $\gamma_a < \gamma_0$.

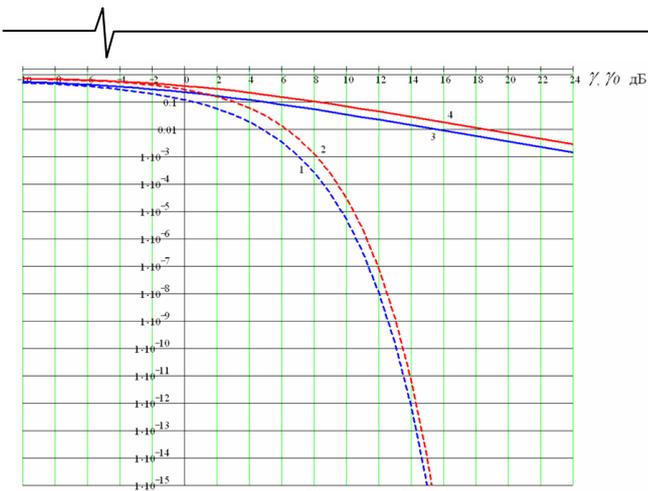


Рис. 1. Вероятности ошибок когерентного и некогерентного приема сигналов в гауссовом шуме от отношения сигнал/шум (γ) и при релейевских замираниях от среднего отношения сигнал/шум (γ_0).

Из графиков видно, что с увеличением γ_0 при фиксированном γ_a увеличивается дисперсионная составляющая сигнала, вероятность его пребывания в области низких ОСШ возрастает. При приближении γ_a к γ_0 дисперсионная составляющая сигнала относительно регулярной снижается, он становится менее случайным, и при $\gamma_a = \gamma_0$ сигнал полностью детерминирован, плотность вероятности превращается в δ -функцию.

Усредняя вероятности ошибочного некогерентного (7) и когерентного приема (8) сигналов ОФМ-2 и ФМ-2 для канала с АБГШ по статистике полученных замираний (6), определим вероятности ошибочного приема в обобщенном релейевском канале, усредняя соответственно по всем значениям γ и $\gamma \geq \gamma_0$:

$$P_{nc}(\gamma_0) = \frac{1}{2} \int_0^{\infty} \exp(\alpha\gamma) \frac{1}{\gamma_0 - \gamma_a} \cdot \exp\left(-\frac{\gamma + \gamma_a}{\gamma_0 - \gamma_a}\right) \cdot I_0\left(\frac{2 \cdot \sqrt{\gamma \cdot \gamma_a}}{\gamma_0 - \gamma_a}\right) d\gamma, \quad (7)$$

$$P_c(\gamma_0) = \frac{1}{2} \int_0^{\infty} \text{erfc}(\sqrt{\alpha\gamma}) \frac{1}{\gamma_0 - \gamma_a} \cdot \exp\left(-\frac{\gamma + \gamma_a}{\gamma_0 - \gamma_a}\right) \cdot I_0\left(\frac{2 \cdot \sqrt{\gamma \cdot \gamma_a}}{\gamma_0 - \gamma_a}\right) d\gamma, \quad (8)$$

Полученные вероятности ошибок при отношениях γ_a/γ_0 , равных 0,7 и 0,9, представлены соответственно на рис.3, 4.

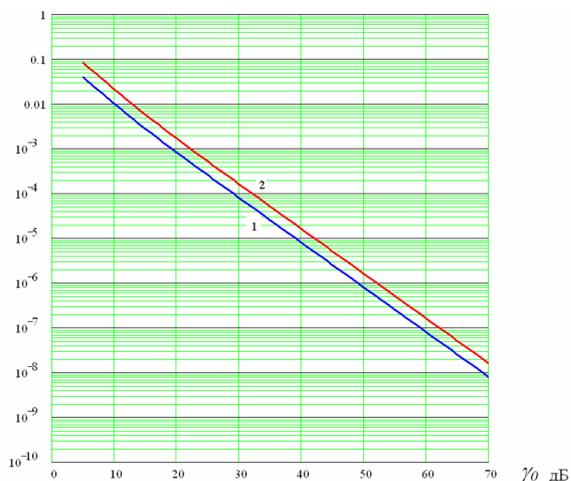


Рис. 3. Вероятности ошибок когерентного и некогерентного приема сигнала (кривые 1, 2) в канале с обобщенными релейевскими замираниями при отношении γ_a/γ_0 равном 0,7.

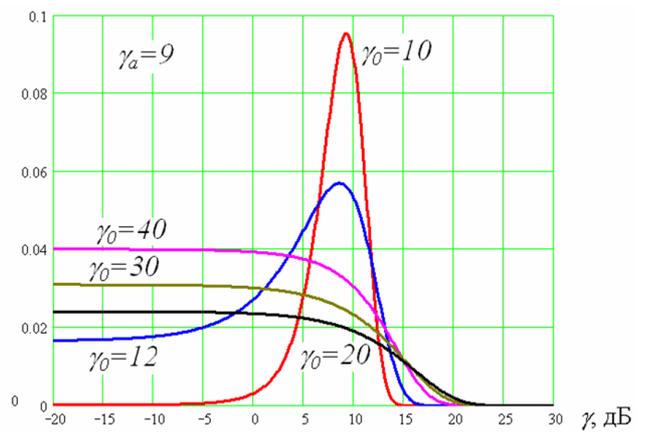


Рис.2. Плотности вероятности ОСШ в канале с обобщенным Релеем при $\gamma_a = 9$ и $\gamma_0 = 10, 12, 20, 30, 40$.

При возрастании γ_a относительно γ_0 помехоустойчивость радиолинии возрастает, снижается вероятность ошибки, например, вероятность ошибки 10^{-7} в первом случае достигается при среднем ОСШ, равном 58 дБ при когерентном приеме, а во втором случае указанная вероятность получается уже при 35 дБ. Это особенно актуально при организации связи в городских условиях, где присутствуют замирания сигналов из-за переотражений, однако мест, в которых сигнал распределен по закону Релея, не так много. Как правило, присутствует значительная величина регулярной составляющей относительно дисперсионной, и среднее ОСШ при этом, вследствие малых расстояний и незначительного затухания сигнала, велико. При нулевом значении регулярной составляющей, как правило в зоне тени, канал становится чисто релейевским, имеющим только дисперсионную составляющую сигнала, при нулевом значении дисперсионной составляющей замирания отсутствуют, сигнал принимается в условиях гауссова δ -коррелированного шума.

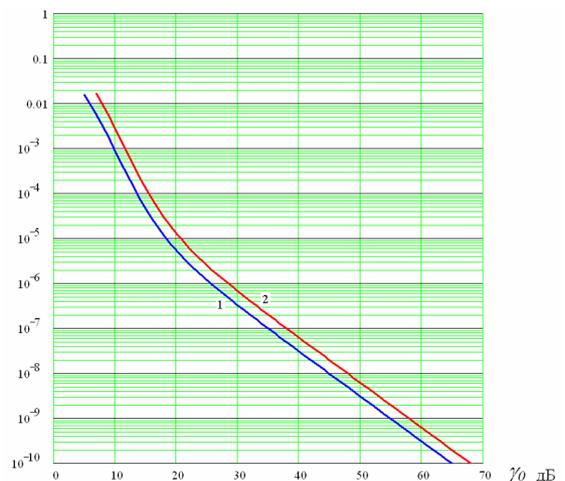


Рис. 4. Вероятности ошибок когерентного и некогерентного приема сигнала (кривые 1, 2) в канале с обобщенными релейевскими замираниями при отношении γ_a/γ_0 равном 0,9.

Повышение помехоустойчивости системы за счет разнесенного приема сигнала

Для снижения вероятности ошибок при коррелированных замираниях сигналов применяют различные способы. Среди них наиболее распространено разнесение сигналов. Разнесение может быть классифицировано как: пространственное, временное, частотное, поляризационное, многолучевое. При пространственном разнесении приемные антенны разносят на расстояние примерно 7-10 длин волн для получения значительной декорреляции сигналов. В технологии разнесенного приема большое значение имеют алгоритмы объединения ветвей разнесения. Выбор способа объединения ветвей определяется сложностью реализации и преимуществами того или иного способа объединения ветвей. При теоретическом анализе обычно сравнивают три способа объединения: оптимальное сложение, максимизирующее отношение сигнал-шум (СМОСШ), линейное сложение с равными весами и автовыбор. При этом предполагается, что сигналы флуктуируют согласно закону Рэлея [2].

Оптимальное сложение, максимизирующее отношение сигнал-шум, обеспечивает наилучший в статистическом смысле эффект ослабления замираний по сравнению с любым другим способом объединения ветвей. Например, при требуемом уровне надежности связи равном 99% и статистической независимости ветвей двукратное разнесение обеспечивает выигрыш 11,5дБ, а четырехкратное - 19дБ. По сравнению со способом автовыбора это соответствует улучшению показателей качества на 1,5 и 3дБ соответственно. Обычно автовыбор применяют в тех случаях, когда считается, что СМОСШ реализовать сложно и дорого. В настоящее время в связи с совершенствованием элементной базы и одновременном снижении ее стоимости, а так же с учетом того, что в радиоканалах, как правило, статистическая независимость ветвей не обеспечивается, представляется целесообразным применять оптимальное сложение. Использование линейного сложения практически оказывается не менее сложным по отношению к оптимальному из-за необходимости обеспечения равенства коэффициентов передачи и шума в приемных ветвях. При невыполнении условия равенства резко увеличиваются проигрыши по сравнению с оптимальным сложением.

Находят применение также комбинированные спосо-

бы объединения ветвей, когда при одних уровнях сигналов используется автовыбор, а при других оптимальное сложение. Проигрыш при таком комбинированном способе объединения ветвей по сравнению с СМОСШ практически несущественен.

Виды разнесения сигналов, связанные с пространственным положением антенн или их элементов (особенно в сочетании с СМОСШ) энергетически более выгодны, и их использование обеспечивает значительные преимущества СПС. Однако их применение осложняется ухудшением массогабаритных характеристик аппаратуры радиосвязи и увеличением энергопотребления. Эти осложнения усиливаются при увеличении кратности разнесения. Некоторые виды пространственного разнесения, например поляризационное, не позволяют организовать больше двух ветвей. Использование больше двух пространственно разнесенных ветвей резко ухудшает массогабаритные характеристики, а получаемый выигрыш с увеличением кратности снижается.

Аналитическая зависимость вероятности ошибки от среднего ОСШ при когерентном (оптимальном) сложении сигналов с некогерентной демодуляцией описывается выражением (9)

$$P_{\gamma_0} = \frac{1}{2} \cdot \left(\frac{1}{1 + \alpha\gamma_0} \right)^M, \quad (9)$$

где M – число ветвей разнесения.

Указанные зависимости вероятности ошибки от среднего ОСШ при различных M представлены на рис.5.

Из анализа кривых, представленных на рис. 5, следует, что вклад в снижение вероятности ошибки от увеличения ветвей разнесения уменьшается с ростом количества самих ветвей разнесения. Например, при вероятности ошибки 10^{-3} сдвоенный разнесенный прием эффективнее одиночного примерно на 10 дБ, а разница в энергетической эффективности между семью и восьмью ветвями разнесения при той же вероятности ошибки (10^{-3}) составит всего 1 дБ. При этом, как было уже указано, с возрастанием числа ветвей разнесения существенно возрастает сложность приемной аппаратуры. Поэтому на практике часто ограничиваются сдвоенным разнесенным приемом.

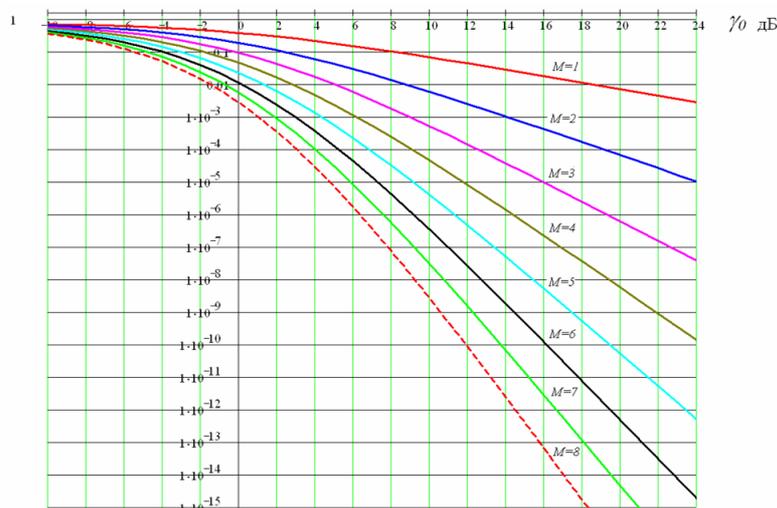


Рис.5. Зависимость вероятности ошибки для медленных релейских замираний от среднего ОСШ при разнесенном приеме, при оптимальном сложении различного числа ветвей разнесения (M), с некогерентной демодуляцией.

Заключение

Рассмотренные в работе методы разнесенного приема сигналов позволяют существенно повысить помехоустойчивость систем цифровой связи в условиях города. Из этих методов наиболее целесообразным в настоящее время является методы на основе оптимального сложения, максимизирующего отношение/шум.

Приложение: Вывод выражения для ПРВ ОСШ в обобщенном релейском канале.

Плотность вероятности распределения амплитуды сигнала x (П.1) в данном канале определяется выражением:

$$f_x = \frac{x}{\sigma^2} \cdot \exp\left(-\frac{x^2 + a^2}{2\sigma^2}\right) \cdot I_0\left(\frac{ax}{2\sigma^2}\right), \quad (\text{П.1})$$

причем при отсутствии регулярной составляющей ($a=0$) указанная плотность вероятности преобразуется в релейскую, где присутствует только диффузионная составляющая. Преобразование формулы (П.1) в аналитическую зависимость плотности вероятности распределения ОСШ в обобщенном релейском канале целесообразно вычислять при условии $a \gg \sigma$, поскольку при a соизмеримом или меньшем σ модель канала приближается к релейской, а значение сигнала представлено преимущественно дисперсионной составляющей. В этом случае математическое ожидание и дисперсия канала с обобщенным Релеем будут выражены формулами (П.2), (П.3) [1]:

$$m_x = a \cdot \left(1 + \frac{\sigma^2}{2a^2}\right), \quad (\text{П.2})$$

$$\sigma_x^2 = \sigma^2 \cdot \left(1 - \frac{\sigma^2}{4a^2}\right), \quad (\text{П.3})$$

а средний квадрат составит (П.4):

$$\langle x_0^2 \rangle = \sigma_x^2 + m_x^2 = 2\sigma^2 + a^2. \quad (\text{П.4})$$

Учитывая выражения (П.5) - (П.7) и

$$f_\gamma = f_x \cdot \frac{dx}{d\gamma}, \quad (\text{П.5})$$

$$x = \sqrt{\frac{N\gamma}{T}}, \quad (\text{П.6})$$

где N – спектральная плотность шума, T – длительность сигнала, а также

$$\frac{dx}{d\gamma} = \frac{1}{2} \cdot \sqrt{\frac{N}{\gamma T}}, \quad (\text{П.7})$$

получим аналитическую зависимость плотности вероятности распределения ОСШ в обобщенном релейском канале:

$$f_\gamma = \frac{1}{\gamma_0 - \gamma_a} \cdot \exp\left(-\frac{\gamma + \gamma_a}{\gamma_0 - \gamma_a}\right) \cdot I_0\left(\frac{2 \cdot \sqrt{\gamma \cdot \gamma_a}}{\gamma_0 - \gamma_a}\right). \quad (\text{П.8})$$

$$\int_0^\infty \frac{1}{\gamma_0 - \gamma_a} \cdot \exp\left(-\frac{\gamma + \gamma_a}{\gamma_0 - \gamma_a}\right) \cdot I_0\left(\frac{2 \cdot \sqrt{\gamma \cdot \gamma_a}}{\gamma_0 - \gamma_a}\right) d\gamma = 1. \quad (\text{П.9})$$

При этом соблюдаются условия нормировки (П.9) для плотности вероятности распределения γ .

Литература

1. Заездный А.М. Основы расчетов по статистической радиотехнике. Издание 2-е. – М.: Связь, 1981. – 448 с.
2. Киселев И.Г., Андрианов М.Н. О выборе видов разнесения сигналов и способов объединения ветвей приема в мобильных системах связи // Мобильные системы. – 2006. – №3. – С. 45-48.

IMPROVEMENT OF NOISE STABILITY OF DATA TRANSMISSION OVER THE DIGITAL COMMUNICATION CHANNELS IN PRESENCE OF RAYLEIGH FADING.

Andrianov M.N., Bumagin A.V., Gondar A.V., Kalashnikov K.S., Prudnikov A.A., Puchkov G.A., Steshenko V.B.

Presence of Rayleigh fading is especially actual for communication systems working in urban conditions. In this paper we present an analysis of fading influence on noise stability of digital communications over the Rayleigh channel and improvement of system performance by use of known diversity reception methods.

УДК 621.391.01

РАЗРАБОТКА И РЕАЛИЗАЦИЯ АЛГОРИТМОВ ИТЕРАТИВНОГО ДЕКОДИРОВАНИЯ БЛОКОВЫХ ТУРБО-КОДОВ

Головкин И. В., аспирант ФИРЭ им. В.А.Котельникова РАН

Назаров Л. Е., д.ф.-м.н., в.н.с. Института радиотехники и электроники им. В.А. Котельникова РАН (Фрязинский филиал), nazarov@ire.rssi.ru

Ключевые слова: декодирование, турбо-код, алгоритм, протокол, передача информации, апостериорная вероятность, итерация, интегральная схема.

Введение

В общем множестве кодовых конструкций под общим названием "турбо-коды" выделяется класс блочных турбо-кодов [1]. Данные турбо-коды являются базовыми для ряда интернациональных протоколов передачи информации [1], например, для протоколов IESS-15, IESS-315, IESS-316e, разработанных для спутниковой системы связи Intelsat, а также в виде опционального применения в протоколе IEEE 802.16, разработанного для высокоскоростных беспроводных широкополосных сетей передачи в городских сетях.

Широкое применение блочных турбо-кодов в системах связи различного назначения определяет интерес исследователей к разработке эффективных алгоритмов декодирования и их реализации средствами цифровой вычислительной техники.

В работе [2] рассмотрены особенности реализации алгоритма декодирования блочных турбо-кодов средствами цифровых сигнальных процессоров (ЦСП) с использованием формата чисел с плавающей запятой. Основу данного алгоритма составляет алгоритм оптимального вычисления апостериорных символьных вероятностей.

В данной статье рассматривается вторая процедура декодирования блочных турбо-кодов, ориентированная на реализацию средствами программируемых логических интегральных схем (ПЛИС) с использованием формата целых чисел. Основу данной процедуры составляет алгоритм подоптимального вычисления апостериорных символьных вероятностей.

По отношению к первой процедуре декодирования рассматриваемая процедура характеризуется меньшей эффективностью – энергетические потери достигают 0.4 дБ при наличии в канале аддитивного белого гауссовского шума. Для канала с Релеевскими замираниями потери более значительны и достигают 1.5 дБ. Вместе с тем устройства кодирования и декодирования на основе ПЛИС обеспечивают существенно большие скорости передачи, что определяет перспективность использования рассматриваемой процедуры декодирования в высокоскоростных системах связи.

Рассмотрены алгоритмы итеративного декодирования блочных турбо-кодов средствами ПЛИС с использованием формата целых чисел, обеспечивающие существенно большие скорости передачи информации. Основу алгоритмов составляет подоптимальное вычисление символьных апостериорных вероятностей с базовыми операциями "сложение-вычитание-сравнение". Для ряда блочных турбо-кодов приведены характеристики реализованных декодеров средствами ПЛИС с различной степенью параллельности.

Постановка задачи

Будем использовать обозначения, введенные в работе [2]: $C_1(n_1, k_1)$, $C_2(n_2, k_2)$ - составляющие двоичные блочные коды в систематическом виде, кодовые слова которых являются соответственно строками и столбцами двумерной матрицы размером $n_1 \times n_2$. Двумерная матрица эквивалентна кодовым словам блочного турбо-кода с параметрами $n = n_1 \cdot n_2$, $k = k_1 \cdot k_2$, $R = k/n$. Здесь n , k , R - длительность кодовых слов, размерность кода и кодовая скорость.

Пусть $\vec{A} = (a_{ij}; 0 \leq i < k_1; 0 \leq j < k_2)$ - последовательность информационных символов, образующих матрицу в составе двумерной матрицы $\vec{B} = (b_{ij}; 0 \leq i < n_1; 0 \leq j < n_2)$ блочного турбо-кода; $\vec{Y} = (y_{ij}; 0 \leq i < n_1; 0 \leq j < n_2)$ - реализация с выхода демодулятора сигналов;

$$L(y_{ij}|b_{ij}) = \ln \left(\frac{p(y_{ij}|b_{ij}=0)}{p(y_{ij}|b_{ij}=1)} \right) - \text{отношение правдоподобия}$$

условных плотностей вероятностей отсчетов y_{ij} ;

$$L(b_{ij}) = \ln \left(\frac{\Pr(b_{ij}=0)}{\Pr(b_{ij}=1)} \right) - \text{отношение априорных}$$

символьных вероятностей. Функция правдоподобия $p(\vec{Y}|\vec{B})$ полагается известной, используется канал передачи без памяти, передаваемые кодовые слова равновероятны.

В работе [2] приведено описание двухэтапной процедуры итеративного декодирования блочных турбо-кодов. На первом этапе m -ой итерации вычисляются приращение отношений апостериорных вероятностей $L^{(1,m)}(b_{ij}|\vec{Y}_i^{(1)}, L^{(2,m)}(b_{ij}))$ для кодовых символов b_{ij} , $j = 0, 1, \dots, n_1 - 1$ для i -го кодового слова $\vec{B}_i^{(1)}$ составляющего блочного кода C_1

$$L^{(1,m)}(b_{ij}) = L^{(1,m)}(b_{ij}|\vec{Y}_i^{(1)}, L^{(2,m)}(b_{ij})) - (L(y_{ij}|b_{ij}) + L^{(2,m)}(b_{ij})). \quad (1)$$

Здесь

$$L^{(1,m)}(b_{ij}|\bar{Y}_i^{(1)}, L^{(2,m)}(b_{ij})) = \ln \left(\frac{\Pr(b_{ij} = 0|\bar{Y}_i^{(1)}, L^{(2,m)}(b_{ij}))}{\Pr(b_{ij} = 1|\bar{Y}_i^{(1)}, L^{(2,m)}(b_{ij}))} \right);$$

$\bar{Y}_i^{(1)}$ - реализация в составе \bar{Y} , соответствующая кодовому слову $\bar{B}_i^{(1)}$. Для первой итерации ($m=1$) верно условие $L^{(2,1)}(b_{i,j}) = L(b_{i,j})$.

На втором этапе m -ой итерации подобные вычисления производятся для вычисления приращения апостериорных символьных вероятностей кодовых слов $\bar{B}_j^{(2)}$ кода C_2

$$L^{(2,m)}(b_{ij}) = L^{(2,m)}(b_{ij}|\bar{Y}_j^{(2)}, L^{(1,m)}(b_{ij})) - (L(y_{ij}|b_{ij}) + L^{(1,m)}(b_{ij})). \quad (2)$$

Величины $L^{(2,m)}(b_{ij})$ используются в качестве априорной символьной информации для первого этапа последующей ($m+1$)-ой итерации, то есть $L^{(1,m+1)}(b_{ij}) = L^{(2,m)}(b_{ij})$.

На последней итерации принимаются решения относительно кодовых символов b_{ij} : если $L^{(2,m)}(b_{ij}|\bar{Y}_i^{(2)}, L^{(1,m)}(b_{ij})) > 0$, принимается решение $b_{ij} = 0$, в противном случае $b_{ij} = 1$.

Ниже приведена процедура вычисления величин $L(b_{ij}|\bar{Y}, L(b_{ij}))$, ориентированная на реализацию средствами ПЛИС с использованием формата целых чисел.

Алгоритм вычисления отношения апостериорных символьных вероятностей

Точное выражение для отношения апостериорных символьных вероятностей $L(b_i|\bar{Y}, L(b_i))$ блочного кода с параметрами (n, k) имеет вид

$$L(b_i|\bar{Y}, L(b_i)) = \ln \left(\frac{\sum_{\bar{B}_m: b_{im}=0} p(\bar{Y}|\bar{B}_m) \Pr(\bar{B}_m)}{\sum_{\bar{B}_m: b_{im}=1} p(\bar{Y}|\bar{B}_m) \Pr(\bar{B}_m)} \right). \quad (3)$$

Рассматриваемый алгоритм подоптимальной оценки $L(b_i|\bar{Y}, L(b_i))$ основан на приближении к (3), полученном в [3]

$$L(b_i|\bar{Y}, L(b_i)) = \ln \left(\frac{\max_{\bar{B}_m: b_{im}=0} \{p(\bar{Y}|\bar{B}_m) \Pr(\bar{B}_m)\}}{\max_{\bar{B}_m: b_{im}=1} \{p(\bar{Y}|\bar{B}_m) \Pr(\bar{B}_m)\}} \right). \quad (4)$$

Пусть условия

$$\max_{\bar{B}_m: b_{im}=0} \{p(\bar{Y}|\bar{B}_m) \Pr(\bar{B}_m)\}, \max_{\bar{B}_m: b_{im}=1} \{p(\bar{Y}|\bar{B}_m) \Pr(\bar{B}_m)\}$$

справедливы для кодовых слов $\bar{B}_h(i)$ и $\bar{B}_l(i)$ соответственно. Для канала с аддитивным белым гауссовским шумом и при условии равновероятности кодовых слов соотношение (4) имеет вид [3]

$$L(b_i|\bar{Y}, L(b_i)) = C \cdot \left(y_i + \sum_{t=1, t \neq i}^n y_t \cdot (-1)^{b_t} \cdot d_t \right). \quad (5)$$

Здесь $d_t = 0$, если $b_{h,t} = b_{l,t}$ и $d_t = 1$ в противном случае, C - энергетический параметр, постоянный для всех кодовых слов.

В соответствии с формализмом итеративного декодирования второе слагаемое в (5) эквивалентно приращению отношений апостериорных символьных вероят-

ностей $L^{(1,m)}(b_{ij})$, $L^{(2,m)}(b_{ij})$, задаваемых соотношениями (1), (2) для составляющих кодов C_1 и C_2 .

Процедура поиска кодовых векторов $\bar{B}_h(i)$ и $\bar{B}_l(i)$, $i = 0, 1, \dots, n-1$, основана на использовании алгоритма Чейза, его реализация требует выполнения следующих шагов [3]:

1) принимаются бинарные ("жесткие") решения относительно кодовых символов переданного слова \bar{B} (возможно с ошибками) на основе отсчетов реализации \bar{Y} ;

2) формируется множество векторов-ошибок $\{\bar{T}\}$, порождающая с их помощью пробные слова $\bar{S}_j = \bar{B} \oplus \bar{T}_j$, $j = 1, 2, \dots, 2^p$. Множество $\{\bar{T}\}$ состоит из 2^p слов, содержащих возможные комбинации ошибок, расположенных в p наименее достоверных позициях слова \bar{B} . Здесь \oplus - операция посимвольного сложения в поле GF(2);

3) каждый из ненулевых пробных слов \bar{S}_j декодируется с помощью алгоритма алгебраического декодирования для блочных кодов C_1 или C_2 , и формируется множество кодовых слов $\{D\}$ с объемом, не превышающим 2^p ;

4) определяется кодовое слово с наибольшей корреляцией между двоичными сигналами, которые соответствуют словам из $\{D\}$, и \bar{Y} , которое принимается в качестве $B_h(i)$, подмножество кодовых слов $\{\bar{D}\}$ объемом $u \leq 2^p$ с максимальными коэффициентами корреляции, не содержащее $B_h(i)$, соответствует $B_l(i)$ для символов b_i .

Для символов b_i , для которых существует $\bar{B}_l(i)$ в составе множества $\{\bar{D}\}$, вычисляется приращение $L^{(\xi,m)}(b_i)$ ($\xi = 1, 2$) с использованием соотношения (5).

Для символов b_i , для которых не существует $\bar{B}_l(i)$ в составе множества $\{\bar{D}\}$, вычисление $L^{(\xi,m)}(b_i)$ осуществляется с использованием правила [3]

$$L^{(\xi,m)}(b_i) = \beta \cdot (-1)^{b_{hi}}, \xi = 1, 2. \quad (6)$$

Здесь b_{hi} - i -ый символ вычисленного на шаге 4 кодового слова $B_h(i)$; β - параметр алгоритма итеративного декодирования турбо-кодов. В литературе известно несколько методов его задания. Одно из рекомендуемых аналитических выражений имеет вид

$$\beta = (E_1 - E_2) / p. \quad (7)$$

Здесь E_1, E_2 - минимальное и максимальное значения коэффициентов корреляции между реализацией \bar{Y} и двоичными словами, которые соответствуют словам из множества $\{\bar{D}\}$.

Вопросам реализации и оптимизации рассматриваемого алгоритма итеративного декодирования посвящен ряд работ [3, 4, 5, 6, 7]. В [4] приведено описание производительного метода реализации шагов 1-4 при поиске кодовых слов $\bar{B}_h(i)$, $\bar{B}_l(i)$. Его основу составляет упорядочение по определенному закону векторов ошибок $\{\bar{T}\}$, позволяющее рекуррентно исполнить требуемые шаги для текущего вектора \bar{T}_i , используя результаты для предыдущего слова \bar{T}_{i-1} , что дает возможность существенного сокращения требуемых вычислительных операций.

Использование нормализованных (взвешенных) величин приращений $\alpha^{(m)} \cdot L^{(\xi,m)}(b_i)$ ($\xi = 1, 2$) при реализации приведенного алгоритма итеративного декодиро-

вания обуславливает приближение его вероятностных характеристик к характеристикам, получаемым с использованием процедуры оптимального вычисления соответствующих приращений апостериорных вероятностей [3]. В общем случае коэффициенты нормализации $\alpha(m)$ зависят от номера итерации m . Результаты моделирования показывают, что для значений β , вычисляемых с использованием соотношения (7), коэффициенты нормализации могут принимать постоянные значения $\alpha(m) = 0.5$ [6,8].

Реализация алгоритма итеративного декодирования блочных турбо-кодов

Ниже приведены условия реализации алгоритма итеративного декодирования (1)-(7), обусловленные требованиями на разрядность q_1 входных отсчетов \bar{Y} , на разрядность q_2 промежуточных вычислений, на требуемую память V , на объем 2^P множества векторов ошибок $\{\bar{T}\}$, на объем подмножества кодовых слов $\{\bar{D}\}$ и на количество итераций I [7].

Требуемый объем памяти V определяется в основном входной реализацией \bar{Y} длительностью $n = n_1 n_2$ с разрядностью отсчетов q_1 и множеством величин приращений $\alpha \cdot L^{(\xi, m)}(b_i)$ длительностью $n = n_1 n_2$ с разрядностью отсчетов q_2 и равен $V = (q_1 + q_2) n_1 n_2$.

Моделирование показало, что разрядность входных отсчетов $q_1 = 5$, число итераций $I = 8$ и объем подмножества кодовых слов $\{\bar{D}\}$, равный $u = 5$, определяют достижение практически предельных вероятностных характеристик рассматриваемого алгоритма итеративного декодирования блочных турбо-кодов [7,8].

При выполнении данных исследований максимальная разрядность q_2 в соответствии с соотношением (5) определяется условием

$$q_2 = q_1 + \lceil \log_2(I \cdot (p + (d-1)/2)) \rceil. \quad (8)$$

Здесь d - минимальное расстояние Хэмминга для составляющего кода, $\lceil x \rceil$ - наименьшее целое число, большее x . Вместе с тем моделирование показало, что выбор разрядности $q_2 = q_1 + 3$ приводит к незначительным энергетическим потерям (не превышающим 0.1 дБ)

по отношению к (8).

На рис. 1 в качестве примера приведены результаты моделирования с целью исследования влияния разрядностей q_1, q_2 на вероятностные характеристики рассматриваемого алгоритма итеративного декодирования для блочного турбо-кода на основе расширенного кода Хэмминга (128,120,4): длительность кодовых слов $n = 16384$, объем слов $k = 14400$. При моделировании применялись 8 итераций, объем подмножества кодовых слов $\{\bar{D}\}$ равен $u = 5$. Кривая 1 соответствует применению процедуры оптимального вычисления приращений апостериорных вероятностей $L^{(\xi, m)}(b_i)$ с использованием формата чисел с плавающей запятой [2], кривые 2 - 5 соответствуют применению рассматриваемой процедуры подоптимального вычисления приращений апостериорных вероятностей $L^{(\xi, m)}(b_i)$ и 8-ми, 6-ти, 5-ти и 4-разрядным отсчетам реализации \bar{Y} (разрядности промежуточных вычислений соответственно равны 10, 8, 7, 6).

Видно, что кривые 2 и 3 практически эквивалентны, то есть 6-ти разрядные отсчеты входных данных и 8-ми разрядные значения промежуточных вычислений определяют достижение практически предельных вероятностных характеристик рассматриваемого алгоритма итеративного декодирования данного турбо-кода.

Подобные результаты получены и для ряда блочных турбо-кодов на основе расширенных кодов Хэмминга и их модификаций. Отметим, что при этих разрядностях и параметрах рассматриваемого алгоритма декодирования данного турбо-кода его энергетические потери по отношению к алгоритму с использованием процедуры оптимального вычисления приращений апостериорных вероятностей $L^{(\xi, m)}(b_i)$ достигают 0.4 дБ для вероятностей на бит $P_b = 10^{-6}$.

В таблице 1 приведены характеристики декодеров для ряда блочных турбо-кодов, реализованных авторами средствами ПЛИС. Следует отметить, что структура кодовых слов блочных турбо-кодов в виде двумерной матрицы дает возможность исполнения приведенной процедуры итеративного декодирования с различной степенью параллельности.

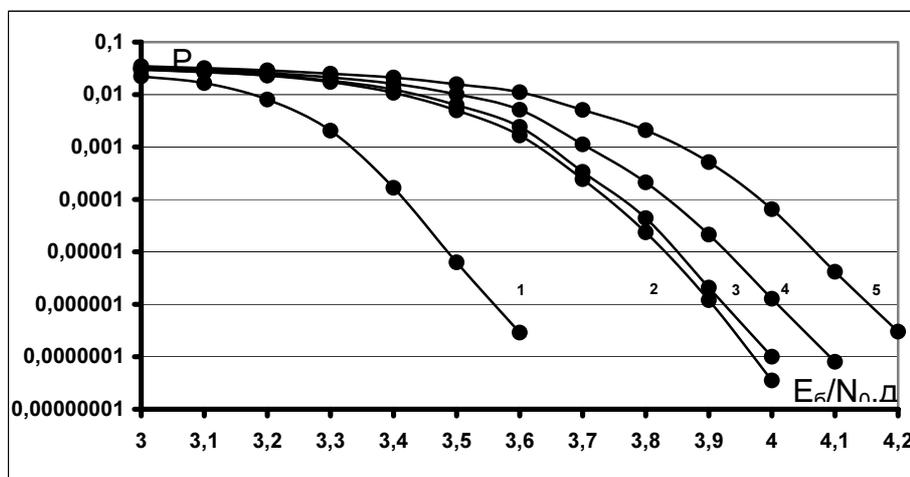


Рис.1. Вероятностные кривые декодеров турбо-кода на основе расширенного кода Хэмминга (128,10,4) (8 итераций, объем множества $\{\bar{D}\}$ равен $u = 5$).

Характеристики реализованных декодеров для блочных турбо-кодов.

№	Тип составляющего кода	Длина кодового блока	Кодовая скорость	E_b/N_0 дБ, $P_b = 10^{-5}$	Скорость Мбит/сек	Микросхема (с минимальными требуемыми параметрами)
1	(128,120)	16384	0.88	3.9	120	XC4VLX40 (55% внутреннего ОЗУ, 91% логики, тактовая частота 160 МГц)
2	(128,113)	16384	0.78	3.3	5	XC4VLX15 (66% внутреннего ОЗУ, 87% логики, тактовая частота 110 МГц)
3	(32,21)	1024	0.43	2.6	2.5	
4	(27,21)	729	0.6	3.25	4.6	XC2S400E-6 (60% внутреннего ОЗУ 50% логики, тактовая частота 100 МГц)

При реализации декодера для турбо-кода №1 с кодовой скоростью 0.88, формируемого на основе расширенного кода Хэмминга (128,120,4), степень параллельности равна 8, то есть одновременно происходит обработка отсчетов входной реализации \vec{Y} , относящихся к 8 строкам или 8 столбцам двумерной матрицы размером 128×128 . Применение этого метода обеспечивает достижение информационной скорости для данного турбо-кода до 120 Мбит/сек. Предельное Шенноновское значение отношения E_b/N_0 для кодовой скорости $R = 0.88$ при использовании двоичных сигналов и дискретно-непрерывного канала равно 3.2 дБ. Из рис.1 видно, что вероятность ошибки $P_b = 10^{-5}$ при использовании рассматриваемого турбо-кода достигается при отношении $E_b/N_0 = 4$ дБ, это значение лишь на 0.8 дБ отличается от предельного значения.

Примерами исполнения со степенью параллельности, равной 1, являются реализованные декодеры для турбо-кодов №2, №3, №4, формируемых соответственно на основе блочного БЧХ кода (128,113,6), на основе блочного БЧХ кода (32,21,6) и на основе укороченного кода Хэмминга с дополнительной проверкой на четность (27,21,4). Реализованные декодеры для этих турбо-кодов обеспечивают информационные скорости передачи, не превышающие 5 Мбит/сек.

Заключение

Рассмотрен алгоритм итеративного декодирования блочных турбо-кодов, реализованный средствами ПЛИС с использованием формата целых чисел. Его основу составляет подоптимальное вычисление символьных апостериорных вероятностей с использованием алгоритма Чейза и алгоритма синдромного декодирования составляющих блочных кодов с базовыми операциями "сложение-вычитание-сравнение-пересылки". По отношению к алгоритму итеративного декодирования на основе оптимального вычисления апостериорных вероятностей рассматриваемый алгоритм характеризуется меньшей эффективностью - энергетические потери достигают 0.4 дБ для канала с аддитивным белым гауссовским шумом.

Для ряда блочных турбо-кодов с различными информационными объемами и кодовыми скоростями реализованы соответствующие декодеры средствами микросхем ПЛИС с различной степенью параллельности.

Литература

1. Solemani M.R., Gao Y., Vilaipornsawai U. Turbo coding for satellite and wireless communications. New York. Kluwer Academic Publishers. 2002. 214 p.
2. Назаров Л.Е., Головкин И.В. Реализация алгоритмов итеративного приема блочных турбо-кодов. // Цифровая обработка сигналов. 2009. №2. С. 2-6.
3. Pundiah R.M. Near-optimum decoding of product codes: block turbo-codes. // IEEE Transactions on Communication. 1998. V.46. N8. P.1003-1010.
4. Hirst S.A., Honary B., Markarian G. Fast Chase algorithm with an application in turbo decoding. // IEEE Transactions on Communications. 2001. V.49. N10. P.1693-1699.
5. Argon C., McLaughlin S.W. An efficient Chase decoder for turbo product codes. // IEEE Transactions on Communications. 2004. V.52. N6. P.896-898.
6. Changlong X., Ying-Chang L., Wing S.L. A low complexity decoding algorithm for extended turbo product codes.// IEEE Transactions on Wireless Communications. 2008. V.7. N1. P.43-47.
7. Adde P., Pyndiah R., Raoul O., Nisan J. Block turbo decoder design.// Proceedings of International Symposium on Turbo Codes. 1997. France. Brest. P. 166-169.
8. Назаров Л.Е., Головкин И.В. Исследование вероятностных характеристик турбо-кодов на основе блочных кодов.// Труды 9-ой Международной конференции "Цифровая обработка сигналов и ее применение". Москва, 28-30 марта 2007. С.20-22.

THE DESIGN AND IMPLEMENTATION OF ITERATIVE DECODING ALGORITHMS FOR TURBO BLOCK CODES

Nazarov L.E., Golovkin I.V.

The implementation of iterative decoding procedures using programmable gate arrays (FPGA Xilinx circuits) for block turbo codes has been investigated in the paper. The base of these iterative decoding procedures is Chase-Pundiah algorithm used for suboptimal evaluation of a posteriori code symbol probabilities. The encoding-decoding scheme showed exceptional performance for high code rates. The coding gains were close to the theoretical Shannon's coding gain in additive Gaussian noise channel. This paper presents results of implementation with FPGA Xilinx circuits for several block turbo-codes.

УДК 681.391

МЕТОДЫ И АППАРАТУРА КОДИРОВАНИЯ И ДЕКОДИРОВАНИЯ СИСТЕМАТИЧЕСКОГО НЕРЕГУЛЯРНОГО КОДА ПОВТОРЕНИЯ-НАКОПЛЕНИЯ (IRA) ДЛЯ DVB-S2 И DVB-T2 ДЕМОДУЛЯТОРОВ

Кравченко А. Н., к.т.н., alexander.kravchenco@thomson.net

Ключевые слова: кодирование, декодирование, низкоплотностной код, демодулятор, метод, алгоритм, граф, матрица.

Основные понятия

LDPC коды (низкоплотностные коды) были предложены Галлагером в 1962 году. Отличительной особенностью LDPC кода является существование случайной разреженной проверочной $m \times n$ матрицы H , т.е. такой матрицы, у которой число ненулевых элементов пренебрежимо мало по сравнению с ее размерами.

Проверочная матрица может быть представлена в виде графа Таннера, изображенного на рис. 1. Граф Таннера является двудольным неориентированным графом, имеющим n символьных и m проверочных узлов, причем i -ый символьный узел соединен с j -ым проверочным узлом тогда и только тогда, когда $H = [h_{i,j}]_{m \times n} = 1$. Ребра в графе Таннера задают систему связей между символьными и проверочными узлами. Степень узла определяют число ребер, выходящих из узла. (C, R)-LDPC код называется *регулярным*, если степени (r) всех проверочных узлов одинаковы. Доля ребер в графе Таннера, соединенных с символьными узлами, в этом случае равна (r). Степени всех символьных узлов (c) также одинаковы, и доля ребер в графе Таннера, соединенных с проверочными узлами, в этом случае равна (c).

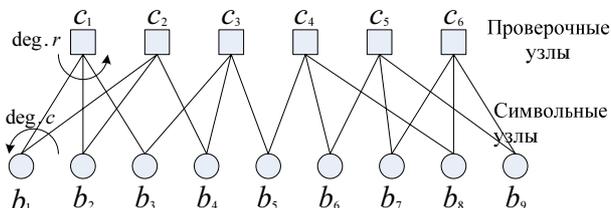


Рис. 1. Граф Таннера двоичного регулярного (2, 3)-LDPC кода

Минимальная кодовая скорость определяется как $R = 1 - m/n$. Число ребер в соответствующем двудольном графе - вес Хэмминга $w(H)$ проверочной матрицы H $w(H) = nc = mr$, и плотность определяется, как $d(H) = w(H)/mn = c/m = r/n$. Минимальное кодовое расстояние этих кодов увеличивается линейно с увеличением длины кодового слова n для данной кодовой скорости R и степеней символьных ($\text{deg.}c$) и про-

Рассмотрена методика конструирования проверочных матриц систематического нерегулярного кода повторения-накопления (IRA) для DVB-S2 и DVB-T2 демодуляторов на основе таблиц (Parity Bit Tables) стандарта, специфицирующих IRA коды с различными кодовыми скоростями. Приводится методика конструирования структурированного ("architecture-aware") LDPC кода из IRA кода. Показано, что полученная конструкция LDPC кода является циклической. Описана архитектура декодера, реализующего "Novel" алгоритм декодирования, который имеет высокую сходимость декодирования и, следовательно, более высокую пропускную способность, а также уменьшенные требования к блоку памяти. Представлены результаты моделирования эффективности декодера LDPC кода длиной 16200 при различных кодовых скоростях в канале с АБГШ для случая двоичной кодовой модуляции.

верочных узлов ($\text{deg.}r$). В нерегулярном LDPC коде (C, R)-LDPC степени символьных ($\text{deg.}c$) и проверочных узлов ($\text{deg.}r$) определяются совокупностями R и C , соответственно. Обычно регулярные коды легче кодируются и имеют простые архитектуры декодеров чем нерегулярные, однако последние достигают более высокого выигрыша от кодирования.

LDPC код для демодулятора DVB-S2

Систематические нерегулярные коды повторения-накопления (IRA)

IRA коды [1], разработанные на основе RA кодов [2], используются в DVB-S2 стандарте. Систематический IRA код может быть представлен с помощью графа Таннера с m проверочных узлов и n символьных узлов. Символьные узлы n могут быть разделены на $k = n - m$ информационных узлов (IN) и m кодовых узлов (PN). Информационные узлы представляют собой биты информации кодового слова в то время как кодовые узлы представляют кодовые биты, соответственно. Проверочные узлы (CN) отображают проверочные уравнения. Все проверочные узлы соединены с информационными узлами через перемешиватель (permutation), как показано на рис. 2.

Информационные узлы имеют различные степени распределения - $f[\text{deg.}c^{\max}, \dots, 3, 2]$. Каждый проверочный узел связан с " a " информационными и двумя кодовыми узлами (принимается во внимание тот факт, что самый первый проверочный узел связан только с одним кодовым узлом).

Кодовая скорость систематического IRA кода определяется как [3]:

$$R = a * k / (a * k + m) \quad (1)$$

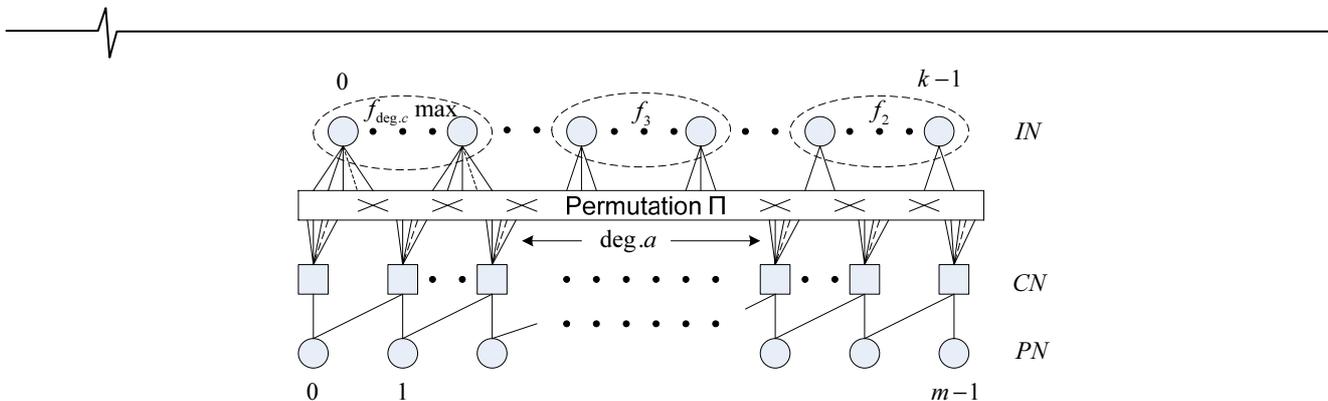


Рис. 2. Граф Таннера IRA кода

Кодирование IRA кода

Проверочная матрица систематического IRA кода для DVB-S2 стандарта имеет следующий вид [4]:

$$H = (\Pi | T) = (H_1 | H_2 | T), \quad (2)$$

где T - $m \times m$ диагональная матрица принимает вид рис 3.

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & \dots & 0 & 0 & 0 \\ \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & 1 \end{bmatrix}$$

Рис. 3. Структура диагональной матрицы

Матрица T устанавливает связь между проверочными и кодовыми узлами. Каждая строка соответствует одному проверочному узлу, а каждый столбец - одному кодовому узлу. Матрица Π размерностью $m \times k$ устанавливает связи между информационными и проверочными узлами. Каждая строка матрицы Π имеет одинаковый вес "а" для всех длинных кодов и частично для коротких кодов DVB-S2 стандарта. Вес столбцов зависит от степени распределения символьных узлов. Проверочные матрицы (Π -части) DVB-S2 стандарта специфицируются таблицами [4]. Например, таблица 1 (Parity Bit Accumulators Table) специфицирует короткий код с кодовой скоростью 3/5:

2765 5713 6426 3596 1374 4811 2182 544 3394 2840 4310 771
 4951 211 2208 723 1246 2928 398 5739 265 5601 5993 2615
 210 4730 5777 3096 4282 6238 4939 1119 6463 5298 6320
 4016

4167 2063 4757 3157 5664 3956 6045 563 4284 2441 3412 6334
 4201 2428 4474 59 1721 736 2997 428 3807 1513 4732 6195
 2670 3081 5139 3736 1999 5889 4362 3806 4534 5409 6384
 5809
 5516 1622 2906 3285 1257 5797 3816 817 875 2311 3543 1205
 4244 2184 5415 1705 5642 4886 2333 287 1848 1121 3595 6022
 2142 2830 4069 5654 1295 2951 3919 1356 884 1786 396 4738

Табл. 1.

0	2161 2653	9 266 4878
1	1380 1461	10 4913 3247
2	2502 3707	11 4763 3937
3	3971 1057	12 3590 2903
4	5985 6062	13 2566 4215
5	1733 6028	14 5208 4707
6	3786 1936	15 3940 3388
7	4292 956	16 5109 4556
8	5692 3417	17 4908 4177

Согласно таблице 1, Π -часть матрицы H состоит из двух частей: H_1, H_2 .

Этот код имеет следующие параметры:

кодовая скорость=3/5, $n=16200, m=6480, deg.c1=12, deg.c2=3, deg.c3=2, deg.r=11, bitgr1=27, bitgr2=9, bitgr3=9$ (каждая группа включает 360 бит), $q=18$ ($q=m/360$). На рис. 4 показана структура проверочной матрицы кода.

LDPC кодер кодирует систематически информационный блок размером $k, i = (i_0, i_1, \dots, i_{k-1})$ в кодовое слово размерности $n, c = (i_0, i_1, \dots, i_{k-1}, p_0, p_1, \dots, p_{n-k-1})$. Передача кодового слова начинается в определенном порядке с i_0 , и заканчивается с p_{n-k-1} .

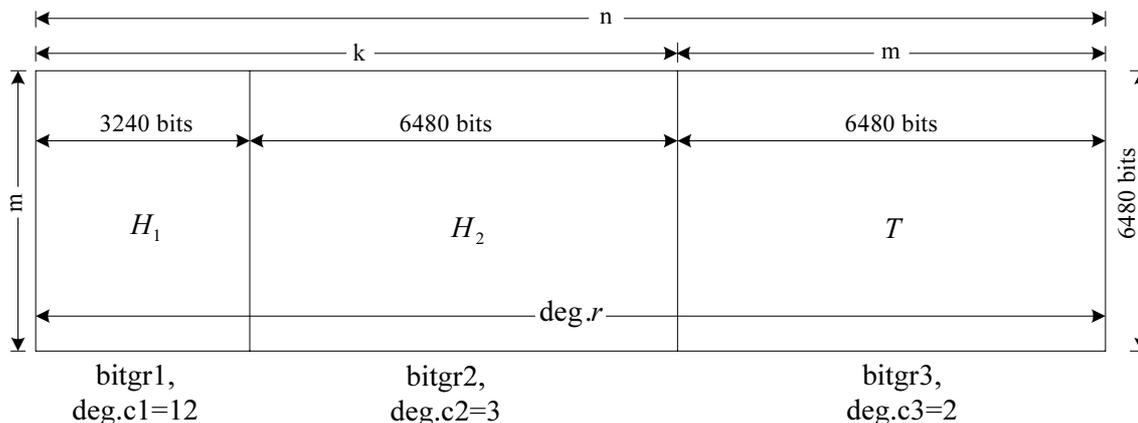


Рис. 4. Структура проверочной матрицы короткого IRA кода с кодовой скоростью равной 3/5.

В соответствии с методом, предложенном в [4], кодирование осуществляется следующим образом (пример для IRA кода с кодовой скоростью $R=3/5$):

1. Инициализация $P_0 = P_1 = \dots = P_{n-k-1} = 0$.

2. Суммировать первый информационный бит i_0 с информацией, содержащейся в соответствующих адресах кодовых бит, указанных в первой строке таблицы 1.

$$\begin{aligned}
 P_{2765} &= P_{2765} + i_0 & P_{2182} &= P_{2182} + i_0 \\
 P_{5713} &= P_{5713} + i_0 & P_{544} &= P_{544} + i_0 \\
 P_{6426} &= P_{6426} + i_0 & P_{3394} &= P_{3394} + i_0 \\
 P_{3596} &= P_{3596} + i_0 & P_{2840} &= P_{2840} + i_0 \\
 P_{1374} &= P_{1374} + i_0 & P_{4310} &= P_{4310} + i_0 \\
 P_{4811} &= P_{4811} + i_0 & P_{771} &= P_{771} + i_0
 \end{aligned}$$

3. Для следующих 359 бит информации, i_m , $m = 1, 2, \dots, 359$ суммировать i_m бит с информацией, содержащейся в соответствующих адресах кодовых бит $\{x + m \bmod 360 \times q\} \bmod (n - k)$, где x обозначает адреса кодовых бит, вычисленных в соответствии с первым битом, а q является кодовой константой. Так, продолжая пример, для информационного бита i_1 выполняются следующие операции:

$$\begin{aligned}
 P_{2783} &= P_{2783} + i_1 & P_{2200} &= P_{2200} + i_1 \\
 P_{5731} &= P_{5731} + i_1 & P_{562} &= P_{562} + i_1 \\
 P_{6444} &= P_{6444} + i_1 & P_{3412} &= P_{3412} + i_1 \\
 P_{3614} &= P_{3614} + i_1 & P_{2858} &= P_{2858} + i_1 \\
 P_{1392} &= P_{1392} + i_1 & P_{4329} &= P_{4329} + i_1 \\
 P_{4828} &= P_{4828} + i_1 & P_{789} &= P_{789} + i_1
 \end{aligned}$$

4. Для 361 информационного бита адреса кодовых бит приведены во второй строке таблицы 1. В аналогичном порядке адреса кодовых бит для следующих 359 бит информации, $m = 361, 362, \dots, 719$ получены с помощью формулы $\{x + (m \bmod 360) \times q\} \bmod (n - k)$, где x обозначает адреса кодовых бит, вычисленных в соответствии с информационным битом i_{360} , т.е. позиции во второй строке таблицы 1.

5. В аналогичном порядке, для каждой группы из 360 новых информационных бит, новая строка в таблице 1 используется для поиска адресов кодовых бит.

После обработки последней строки из таблицы 1 для каждого информационного бита известны все адреса кодовых бит, накапливающих/суммирующих этот бит.

Основываясь на этой информации, могут быть определены адреса информационных бит, вовлеченных в вычисление каждого кодового бита. По сути дела адреса информационных бит есть адреса "1" в строке проверочной матрицы (П-часть). Такое представление проверочной матрицы обычно называют "alist format".

Таблица 2 иллюстрирует часть проверочной матрицы $H = (\Pi | T)$ короткого LDPC кода с кодовой скоростью $= 3/5$.

Таблица 2.

3	908	1202	2308	3121	3218	3240	4181	6569	9720	0
445	721	1716	2049	2198	3014	3480	3600	9488	9720	9721
698	729	857	2430	2926	3191	3960	6067	9107	9721	9722

Схема кодера представлена на рис. 5. Кодер можно рассматривать как генератор проверочных узлов, основанный на проверочной матрице и рекурсивном сверточном кодере с кодовой скоростью 1.

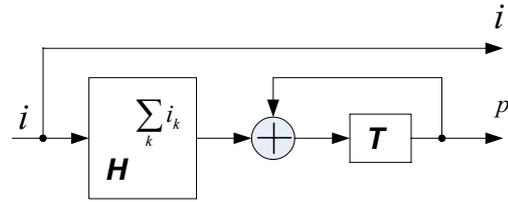


Рис. 5. Схема кодера IRA кода

Кодер суммирует (по модулю 2) все биты информации в соответствии с позициями "1" в проверочной матрице H (П-часть), и результат подается на сверточный кодер. В свою очередь сверточный кодер вычисляет кодовые биты.

Расчет кодового бита 0 (p_0) это особый случай. Бит p_0 имеет позицию 9720 и вычисляется в соответствии с алгоритмом:

```
for(k=0; k<deg.r-2; ++)
```

$$[9720]=[3]+[908]+[1202]+[2308]+[3121]+[3218]+[3240]+[4181]+[6569];$$

Следующие кодовые биты вычисляются в соответствии с алгоритмом:

```
for(k=0; k<deg.r-1; ++)
```

$$[9721]=[445]+[721]+[1716]+[2049]+[2198]+[3014]+[3480]+[3600]+[9488]+[9720];$$

$$[9722]=[698]+[729]+[857]+[2430]+[2926]+[3191]+[3960]+[6067]+[9107]+[9721]; \text{ и т.д.}$$

Декодирование

Декодирование IRA кода можно свести к процедуре декодирования структурированного ("architecture-aware" LDPC (AA-LDPC) кода [5].

Как известно [5], применение AA-LDPC кодов при проектировании LDPC декодера, позволяет решить проблемы, связанные с внутренними межузловыми связями между вычислительными элементами, уменьшить требования к памяти, увеличить скорость сходимости декодирующего алгоритма, а также решить проблему масштабируемости, ассоциированную с LDPC декодером.

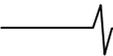
Построение AA-LDPC кодов

AA-LDPC код описывается "блочной" матрицей, состоящей из $B \times D$ подматриц (блоков), где каждый блок - двоичная $S \times S$ подматрица, формируемая по вектору кодонезависимых параметров S .

Каждая подматрица должна удовлетворять следующим двум условиям: каждый столбец имеет одну 1, и каждая строка - одну 1. Подматрицы, составляющие блочную матрицу должны быть только трех видов:

1. Подматрицы, включающие только нули (*all-zero matrix*);
2. Единичные подматрицы (identity matrix, I^0);
3. Перестановочные матрицы (permutation matrix, I^x).

$S \times S$ перестановочная матрица получается из единичной подматрицы либо путем случайной перестановки рядов (идентично столбцов), либо путем циклического сдвига единичной подматрицы в соответствии с индексом сдвига x , при этом матрица обозначается как I^x . Регуляр-



ный (C, R)-LDPC код должен иметь D подматриц в столбце и B подматриц в строке базисной матрицы соответственно. Регулярный AA-LDPC код имеет длину $n=SB$, и кодовую скорость $R \geq 1-D/B$.

Проектирование "AA-LDPC" кода из IRA кода для DVB-S2 стандарта

Так как линейные блочные коды обладают свойствами коммутативности, то в процессе декодирования можно изменить порядок строк матрицы H .

Из алгоритма кодирования, предложенного в [4], можно видеть, что он обрабатывает группы, каждая из которых равна 360 бит. Поэтому имеет смысл разбить проверочную матрицу на подматрицы размерностью 360×360 . Кроме того, имеет смысл переупорядочить проверочные уравнения и сгруппировать их в группы проверочных узлов следующим образом:

Проверочные узлы: $0, q, 2 \times q, 3 \times q, \dots, 359 \times q$ формируют группу проверочных узлов 0 (CNG0).

Проверочные узлы: $1, 1+q, 1+2 \times q, 1+3 \times q, \dots, 1+359 \times q$ формируют группу проверочных узлов 1 (CNG1).

Проверочные узлы: $2, 2+q, 2+2 \times q, 2+3 \times q, \dots, 2+359 \times q$ формируют группу проверочных узлов 2 (CNG2).

Проверочные узлы: $q-2, q-2+q, q-2+2 \times q, q-2+3 \times q, \dots, q-2+359 \times q$ формируют группу проверочных узлов $q-2$ (CNG $q-2$).

Проверочные узлы: $q-1, q-1+q, q-1+2 \times q, q-1+3 \times q, \dots, q-1+359 \times q$ формируют группу проверочных узлов $q-1$ (CNG $q-1$).

Дополнительно каждые 360 символьных узлов группируются в группы символьных узлов.

Символьные узлы: $0, 1, 2, 3, \dots, 359$ формируют группу символьных узлов 0 (BNG0)

Символьные узлы: $0+360, 1+360, 2+360, 3+360, \dots, 359+360$ - группа символьных узлов 1 (BNG1).

Символьные узлы: $0+2 \cdot 360, 1+2 \cdot 360, 2+2 \cdot 360, 3+2 \cdot 360, \dots, 359+2 \cdot 360$ - группа символьных узлов 2 (BNG2).

Символьные узлы: $0+n360, 1+n360, 2+n360, 3+n360, \dots, 359+n360$ - группа символьных узлов n (BNG n), где максимальное значение параметра n для длинного кода равно $n_{\max} = 179 - q$ и для короткого кода $n_{\max} = 44 - q$ соответственно.

Результирующая переупорядоченная проверочная матрица может быть записана следующим образом:

$$H = \begin{bmatrix} H(0,0) & H(0,1) & \dots & H(0,44) \\ H(1,0) & H(1,1) & \dots & H(1,44) \\ \vdots & \vdots & \vdots & \vdots \\ H(q,0) & H(q,1) & \dots & H(q,44) \end{bmatrix}$$

Обозначим $H(m,n)$ как ненулевую подматрицу переупорядоченной проверочной матрицы H , соответствующей группе проверочных узлов m , и группе символьных узлов n .

Для строки n из таблицы 1 может быть вычислена группа подматриц, соответствующих группе символьных узлов n .

Для данного значения k в строке n , соответствующая

группа проверочных узлов есть $m = k \bmod q$, индекс сдвига равен $x = \lfloor k/q \rfloor$, и соответствующая перестановочная подматрица записывается как $H(m,n) = I^{(x)}$.

Например, для первой строки таблицы 1
2765 5713 6426 3596 1374 4811 2182 544 3394 2840 4310 771

соответствующие значения перестановочных подматриц:

$$\begin{aligned} 2765: H(11,0) &= I^{153} & 2182: H(4,0) &= I^{121} \\ 5713: H(7,0) &= I^{317} & 544: H(4,0) &= I^{30} \\ 6426: H(0,0) &= I^{357} & 3394: H(10,0) &= I^{188} \\ 3596: H(14,0) &= I^{199} & 2840: H(14,0) &= I^{157} \\ 1374: H(6,0) &= I^{76} & 4310: H(8,0) &= I^{239} \\ 4811: H(5,0) &= I^{267} & 771: H(15,0) &= I^{43} \end{aligned}$$

Перестановочные подматрицы для следующих строк из таблицы 1 должны рассчитываться по тому же пути. Полученные таким образом перестановочные подматрицы удовлетворяют условиям, предъявляемым к AA-LDPC коду.

"Novel" Алгоритм декодирования

"Novel" алгоритм декодирования был предложен в работе [6] для декодирования AA-LDPC кода в демодуляторе DVB-S2. В других публикациях этот алгоритм известен также как "layered" [7], "staggered" [8].

Этот алгоритм декодирования обладает определенными преимуществами по сравнению с другими [9, 10, 11, 12]:

1. Декодер, спроектированный при использовании этого алгоритма, требует меньшего объема памяти, чем стандартный декодер, реализующий стандартный двухфазный метод декодирования - "сумма-произведение" (two-phase message-passing, TPMP algorithm).

2. Скорость сходимости алгоритма существенно выше, чем у "TPMP" (алгоритм выполняет меньше итераций для достижения равной эффективности декодирования).

При описании алгоритма воспользуемся метрикой, называемой логарифмическим отношением функций правдоподобия (log-likelihood ratio - LLR) [13], которая определяется выражением.

$$LLR(x|y) = \ln \left[\frac{p(y|x=0)}{p(y|x=1)} \right] \quad (3)$$

Пусть $x = [x_1, x_2, \dots, x_n]$ - кодовое слово, которое модулируется при использовании двоичной фазовой модуляции, и модулированные значения x передаются по каналу с белым аддитивным гауссовым шумом (АБГШ).

Входную последовательность принятых сигналов (символов) обозначим, как $y = [y_1, y_2, \dots, y_n]$. Демодулятор принимает входную последовательность сигналов и вычисляет соответствующие LLR значения:

$$\text{For } j = 1, 2, \dots, n \quad \lambda_j = LLR(y_j|x_j) = \ln \frac{p(y_j|x_j=0)}{p(y_j|x_j=1)} \quad (4)$$

Обычно при двоичной биполярной передаче по каналу LLR значения вычисляются как

$$\lambda_j = \frac{2}{\sigma^2} * y_j, \tag{5}$$

где σ^2 - дисперсия АБГШ.

При описании алгоритма воспользуемся следующими обозначениями:

λ_j - "llr" значение принятого j -го символа, $R_{ij}[k]$ - внешние (extrinsic) "llr" значения, $Q_{ij}[k]$ - внутренние (intrinsic) "llr" значения, Λ_j - апостериорные "llr" значения, $R[i]$ - индексы символьных узлов, вовлеченных в вычисление i -го проверочного узла.

"Novel" алгоритм описывается следующим образом:

1. Инициализация:

$$R_{ij}^0 = 0, \Lambda_j^0 = \lambda_j \tag{6}$$

2. Горизонтальное сканирование (check node update rule):

$$Q_{ji}^{(k)} = \Lambda_j^{(k-1)} - R_{ij}^{(k-1)}; \tag{7}$$

$$R_{ij}^{(k)} = \prod_{j \in R(i) \setminus j} \text{sign}(Q_{ji}^{(k)}) * \{ \text{MIN}_{j \in R(i) \setminus j} | Q_{ji}^{(k)} | \}; \tag{8}$$

$$\Lambda_j^{(k)} = Q_{ji}^{(k)} + R_{ij}^{(k)}. \tag{9}$$

Архитектура LDPC декодера

Центром архитектуры декодера, реализующего "Novel" алгоритм декодирования, является вычислительный блок [6], изображенный на рис. 6. Блок вычисляет внешние "llr" и апостериорные "llr" значения, используя априорные "llr" значения, поступающие на вход блока. Декодер включает 360 вычислительных блоков (проблема масштабируемости в данной работе не рассматривается), что соответствует размеру группы символьных узлов (BNG). $\text{Deg}_r * Q_{ji}$ - число внутренних "llr" значений вычисляется путем вычитания R_{ij} из последовательно поступающих значений Λ_j . Вычисленные внутренние "llr" значения запоминаются в FIFO элементе и одновременно в блоке вычисления внешних "llr" величин (Check Node Update Unit) для дальнейшего вычисления новых внешних "llr" значений.

Апостериорные "llr" значения вычисляются путем сложения внутренних величин с выхода FIFO элемента, с новыми внешними "llr" значениями, поступающими с выхода блока вычисления внешних величин.

Архитектура декодера представлена на рис. 7. Де-

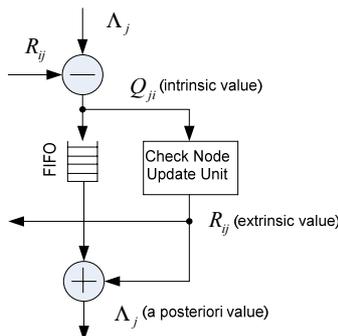


Рис. 6. Схема вычислительного блока

кодер включает следующие блоки:

1. Конфигурирующий блок памяти (Configuration ROM);
 2. Блок памяти апостериорных "llr" значений (Lambda memory);
 3. Два сдвигающих регистра (barrel shifters);
 4. Блоки памяти внешних "llr" значений (R memories).
- Конфигурирующий блок памяти содержит адреса групп символьных узлов в блоке памяти апостериорных "llr" значений, а также сдвигающие индексы для соответствующих групп символьных узлов.

Блок памяти апостериорных "llr" значений используется для запоминания "llr" значений с выхода демодулятора (λ_j), а также для записи апостериорных "llr" значений с выходов вычислительных блоков (Λ_j).

Сдвиговые регистры выполняют циклический сдвиг соответствующих символьных групп в соответствии с их индексами сдвига.

Блоки памяти внешних "llr" значений предназначены для записи и чтения внешних "llr" значений в процессе декодирования.

На рис. 8 представлен алгоритм декодирования, включающий следующие блоки и операции.

1. Счетчик групп проверочных узлов (k), счетчик групп символьных узлов (i), счетчик итераций, указатели адреса записи и чтения в R блоках памяти (read_ptr, write_ptr) и счетчик llr_word's (j) обнуляются, кодовое слово с выхода демодулятора запоминается в "Lambda" блоке памяти.
2. Для всех групп символьных узлов, принадлежащих текущей группе проверочных узлов (старт $k = 0$), вычисляются внутренние "llr" значения, вычисленные на основе априорных "llr" значений, считанных из "Lambda" блока памяти.
3. Deg_r внешних "llr" значений вычисляются в каждом "Check Node Update Unit" блоке.
4. Используя внутренние "llr" значения, считанные из FIFO's, и внешние "llr" значения, считанные из "Check Node Update Unit" блоков, вычисляются апостериорные "llr" значения в каждом вычислительном блоке, которые запоминаются в "Lambda" блоке памяти.

Шаги 2-4 повторяются до тех пор, пока не будет выполнено условие $k=q-1$.

Если это условие удовлетворено, выполняется шаг 5.

5. Проверка условия: $x \otimes H^T = 0$

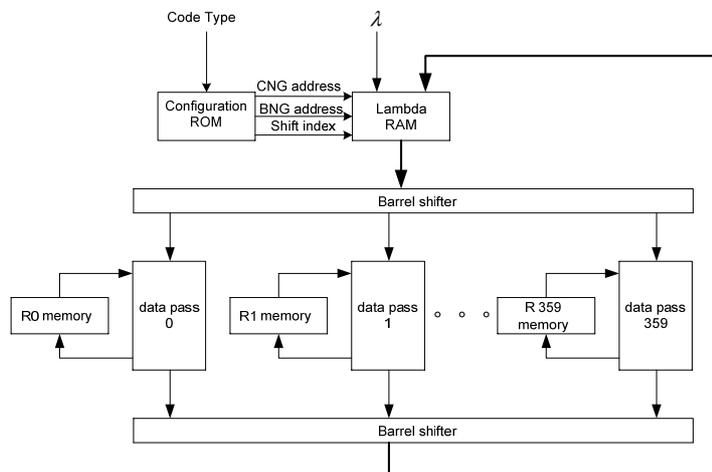


Рис. 7. Архитектура декодера

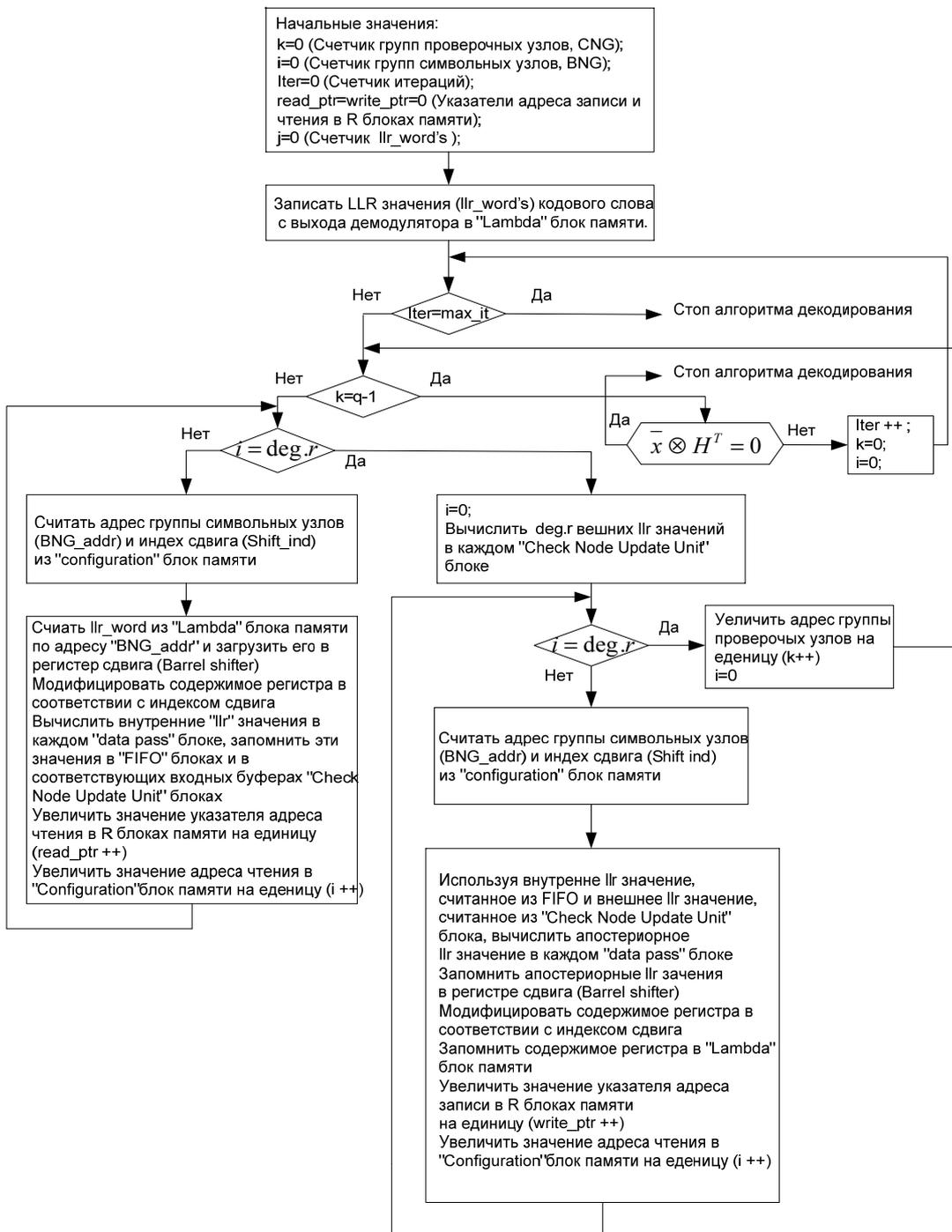


Рис. 8. Алгоритм декодирования

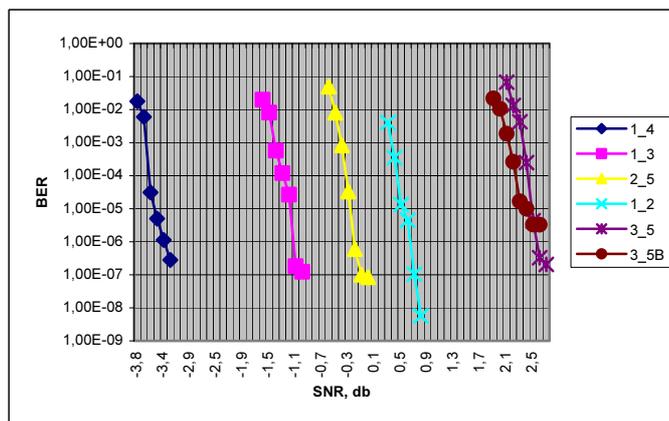


Рис. 9. Эффективность декодера для кодовых скоростей - 1/4, 1/3, 2/5, 1/2, 3/5, 3/5B

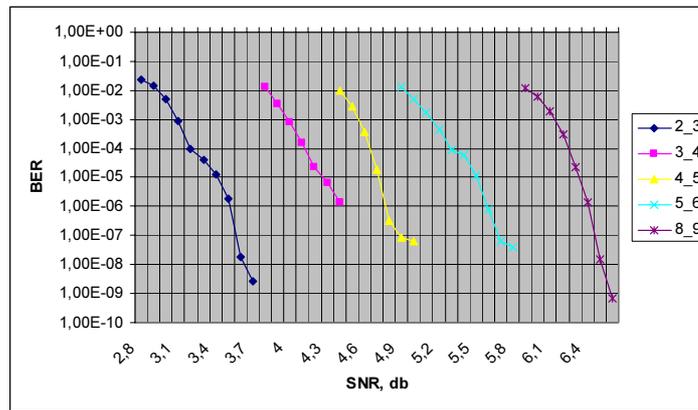


Рис. 10. Эффективность декодера для кодовых скоростей – 2/3, 3/4, 4/5, 5/6, 8/9

Если это условие удовлетворено, то декодирование завершается, декодированный бинарный вектор x передается к "BCH" декодеру и алгоритм декодирования переходит к шагу 1. Если это условие не удовлетворено, то счетчик итераций увеличивается на единицу, и алгоритм декодирования переходит к шагу 2.

На рис. 9, 10 представлены оценки эффективности декодера LDPC кода длиной 16200 при различных кодовых скоростях в канале с АБГШ для случая двичной фазовой модуляции.

При моделировании декодера использовалась арифметика с фиксированной запятой. Для представления мягких решений (LLR значений) демодулятора и сообщений внутри декодера использовалась прямой код с числом разрядов $b=6$.

Литература

1. D. Divsalar, H. Jin, and R. McEliece. Coding theorems for turbo-like codes. Proc. 36th Annual Allerton Conf. on Comm., Control, and Computing, Sept. 1998, pp. 201-210.
2. H. Jin, A. Khandekar, and R. McEliece. Irregular repeat-accumulate codes. Proc. 2nd. Int. Symp. on Turbo Codes and Related Topics, Brest, France, Sept. 2000, pp. 1-8.
3. Mohammad M. Mansour. High-Performance Decoders for Regular and Irregular Repeat-Accumulate Codes. IEEE Communication Society Globecom, 0-7803-8794-5/04/\$20.00 © 2004 IEEE, pp. 2583-2588.
4. European Telecommunications Standards Institute. Digital video broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications. DRAFT EN 302 307 (v.1.1.2 06.2006).
5. Mohammad M. Mansour. Architecture-Aware Low-Density Parity-Check Codes. 0-7803-7761-3/03/\$17.00 © 2003 IEEE, pp. 57-60.
6. J. Dielissen, A. Hekstra, and V. Berg. Low cost LDPC decoder for DVB-S2. In Proc. 2006 Design, Automation and Test in Europe (DATE '06), Munich, Germany, Mar. 2006.
7. D. Hecovar. A Reduced Complexity Decoder Architecture via Layered Decoding of LDPC Codes. In IEEE Workshop on Signal Processing Systems, SISAP 2004, 2004, pp 107-112.
8. Mohammad M. Mansour. Unified Decoder Architectures for

Repeat-Accumulate and LDPC Codes. 0-7803-8622-1/04/\$20.00 © 2004 IEEE, pp. 527-531.

9. Овечкин Г. В., Чикин А.В. Архитектура и реализация декодера LDPC кодов для демодулятора DVB-S2. ТРУДЫ НИИР, ВЫПУСК 3, 2008.
10. Ефимов А.Г. Об аппаратной реализации декодеров LDPC-кодов. ВОПРОСЫ ПЕРЕДАЧИ И ЗАЩИТЫ ИНФОРМАЦИИ, Сборник статей под редакцией д.т.н Е. А. Крука, Санкт-Петербург, 2006.
11. F.Kienle, T. Brack, and N. Wehn. A synthesizable IP core for DVB-S2 LDPC code decoding. In IEEE Conference on Design Automation and Test Europe (DATE), 2005.
12. P. Urard, E. Yeo, L. Paumier, P. Georgelin, T. Michel, V. Lebars, E. Lantreibecq, B. Gupta. A 135Mb/s DVB-S2 Compliant Codec Based on 64800b LDPC and BCH Codes. 2005 IEEE International Solid-State Circuits Conference, 0-7803-8904-2/05/\$20.00 ©2005 IEEE.
13. Бернадт Сяляр. Цифровая связь. Теоретические основы и практическое применение. М.: Издательский дом "Вильямс", 2003, 1104 с.

METHODS AND THE APPARATUS FOR CODING AND DECODING SYSTEMATIC IRREGULAR REPEAT ACCUMULATE CODE (IRA) OF DVB-S2 AND DVB-T2 DEMODULATORS.

Kravtchenko A.N.

Design method of parity check matrixes of systematic irregular repeat accumulate code is proposed for DVB-S2 and DVB-T2 demodulators based on the Parity Bit Tables, which specify IRA codes with different code rates. Scheme of encoder is presented. Method of the design of structural AA-LDPC code from IRA code is suggested. Designed code has the cycle code property. Decoder architecture for AA-LDPC codes is described based on a novel decoding algorithm which has a faster convergence rate and hence a throughput advantage and also the memory overhead problem is reduced over the standard decoding algorithm (TPMP). Interconnect complexity problem in the decoder is mitigated by using the AA-LDPC codes. Result of performance simulation of the decoder for 16200 bits code words with different code rates over AWGN channel is presented.

ПЕРЦЕПТУАЛЬНЫЙ КОДЕР ЗВУКА НА БАЗЕ БЫСТРОГО ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ С ДИНАМИЧЕСКОЙ ТРАНСФОРМАЦИЕЙ ЧАСТОТНО-ВРЕМЕННОГО ПЛАНА

Петровский А. А., к.н.т., доцент Белорусского государственного университета информатики и радиоэлектроники, email: petrovsky@bsuir.by

Ключевые слова: кодер звука, перцептуальный, вейвлет-преобразование, фрейм, частотно-временной план, психоакустический анализ, порог маскирования, алгоритм, структура, вычислительная платформа.

Введение

Структуры большинства кодеров звука на основе психоакустики сходны и могут быть представлены обобщенной схемой (рис. 1) [2-4].

Входной звуковой сигнал $x(n)$ разбивается на квазистационарные фреймы длительностью от 2 до 50 миллисекунд в зависимости от алгоритмов обработки и методов кодирования. Блок частотно-временного анализа аппроксимирует временные и спектральные параметры звука для каждого анализируемого фрейма с учетом шкалы критических частотных полос. В блоке психоакустического анализа оценивается энергия маскирующего сигнала (уровни маскирующих порогов) на базе психоакустической модели. При этом определяются максимальные искажения, возникающие в каждой точке частотно-временной плоскости в процессе квантования и кодирования частотно-временных оценок без введения искусственного артефакта слышимости при восстановлении сигнала. Следовательно, психоакустический анализатор вычисляет частотно-временной параметр невосприимчивости акустической информации слушателем, который затем передается в блок квантования и кодирования. В процессе психоакустического кодирования необходимо: во-первых, установить вид маскирующего сигнала, во-вторых, вычислить соответствующие пороги, затем полученную информацию использовать для того, чтобы расположить спектр шума кодирования ниже порога маскирования, известный как порог едва различимых искажений JND (just noticeable distortion) [5].

Значения психоакустических порогов маскирования неявно показывают результат кодирования звука, а точнее вероятный шаг квантования. Минимальное число

Предлагается перцептуальный кодер звука на основе адаптируемого под сигнал и заданный вычислительный ресурс быстрого вейвлет-преобразования с динамической трансформацией частотно-временного плана [1].

бит, необходимое для кодирования воспринимаемого слушателем сигнала, оценивается перцептуальной энтропией PE (perceptual entropy) [6]:

$$PE = \frac{1}{N} \sum_{f=f_l}^{f_h} \max(0, \log_2 \frac{|signal(f)|}{threshold(f)}), \quad (1)$$

где N – число частотных компонент в частотном диапазоне f_l и f_h ; f_l – нижняя частота диапазона; f_h – верхняя частота диапазона, например $f_h = 22050$ Гц; $|signal(f)|$ – амплитуда частотной компоненты f ; $threshold(f)$ – оценка порога маскирования на частоте f . На практике, PE в форме (1) часто называют функцией Джонсона и вычисляют на основе субполосного анализа звука [6]:

$$PE = \sum_{i=1}^{25} \sum_{\omega=bl_i}^{bh_i} \log_2 \left(2 \left| \text{nint} \left(\frac{Re(\omega)}{\sqrt{6 \cdot T_i / k_i}} \right) \right| + 1 \right) + \log_2 \left(2 \left| \text{nint} \left(\frac{Im(\omega)}{\sqrt{6 \cdot T_i / k_i}} \right) \right| + 1 \right), \quad (2)$$

где i – индекс критической полосы; bl_i и bh_i – нижнее и верхнее значение частоты i -ой критической полосы; k_i – количество компонентов преобразования в i -ой критической полосе; T_i – значение порога маскирования в критической полосе i ; $\text{nint}(\)$ – операция округления до ближайшего целого значения. Следовательно, та часть сигнала, которая может быть изменена (в общем случае отброшена) и при этом не вносится дополнительных искажений при его восстановлении, является перцептуально избыточной, а та часть сигнала, отражающая слышимую акустическую информацию человеком, изменяется и кодируется.

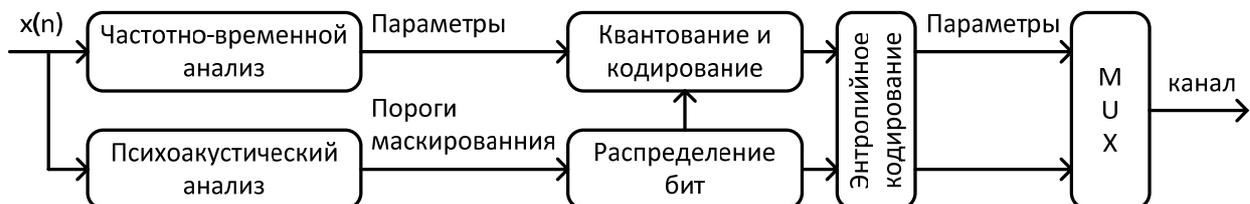


Рис. 1. Общая структура перцептуального кодера звука

Выбор метода частотно-временного анализа неразрывно взаимосвязан с точностью вычисления психоакустической модели и общей концепцией построения системы кодирования сигналов звука. При этом выбор методологии частотно-временного анализа всегда обуславливает решение фундаментального вопроса: обеспечения требуемого разрешения как в частотной так и временной областях. Развитие этих работ в настоящее время идет по пути построения перцептуальных кодеров звука на основе подходящего частотно-временного преобразования, в частности, банков цифровых фильтров и наиболее гибкого, в смысле управления частотно-временным разрешением, быстрого вейвлет преобразования (пакет дискретного вейвлет преобразования (ПДВП)) с динамической трансформацией дерева алгоритма.

ПДВП есть обобщение диадического вейвлет преобразования, которое позволяет получить множество структур путем его соответствующих декомпозиций [7]. Структура ПДВП больше согласуется с сигналом, чем вейвлет преобразование [8], и характеризуется следующими свойствами: малая вычислительная сложность процедуры декомпозиции звукового сигнала в выбранном базисе (процедура анализа); малая вычислительная сложность процедуры реконструкции сигнала (синтеза) в выбранном базисе; конвейерность вычислительного процесса процедур анализа и синтеза, что способствует организации поточных и параллельно-поточных структур процессоров реального времени; гибкое изменение частотно-временного разрешения, обеспечивающее локализацию нестационарностей в сигнале; единственность преобразования, т.е. в ограниченном числе структур ПДВП имеется одна структура, идентифицирующая соответствующие компоненты сигнала [9].

Психоакустический анализ
Аппроксимация критического частотно-временного плана

Пусть $\{\varphi_n(t): n \in \mathbb{Z}\}$ определяет множество структур деревьев ПДВП, и пусть $E \subset \{(l, n): 0 \leq l \leq L, 0 \leq n \leq 2^l\}$ представляет собой узлы дерева ПДВП [9], тогда отрезок $[0,1)$ разделяется на диадические интервалы

$$I_{l,n} = [n2^{-l}, (n+1)2^{-l}), \tag{3}$$

которые соответствуют специфическому множеству узлов E . В частности,

$$\{\varphi_{l,n,k}(t): (l, n) \in E, k \in \mathbb{Z}\}, \tag{4}$$

где $\varphi_{l,n,k}(t) \triangleq 2^{-\frac{l}{2}}\varphi_{l,n}(2^{-l}t - k)$ является базовой формой в пространстве сигнала $\overline{\text{span}}\{\varphi_{0,0}(t - k): k \in \mathbb{Z}\}$. Узел $(l, n) \in E$ дерева ПДВП ассоциируется с частотной полосой, у которой центральная частота и полоса пропускания приблизительно задаются следующими соотношениями:

$$f_{l,n} = 2^{-l}(GC^{-1}(n) + 0.5) \cdot \frac{f_s}{2}, \tag{5}$$

$$\Delta f_{l,n} = 2^{-l} \cdot \frac{f_s}{2}, \tag{6}$$

где GC^{-1} – обратный код перестановок Грея, f_s – частота дискретизации сигнала.

Аппроксимация критической шкалы частот на основе ПДВП осуществляется таким образом, чтобы расстояние между центральными частотами $z(f)$ полос пропускания было размером в один барк. Соотношение между шкалой частот в герцах и барках приведено в [10].

На рис. 2 показано дерево ПДВП (Critical Band Wavelet Packet Decomposition (CB – WPD)), полученное эмпирически, которое осуществляет разделение частотного интервала звукового сигнала на полосы, согласно критической шкале частот [11]:

$$CB - WPD: (l, n) \in E_{CB}, l = \overline{0, 8}, \tag{7}$$

где E_{CB} – обозначает множество узлов дерева ПДВП, соответствующего CB – WPD.

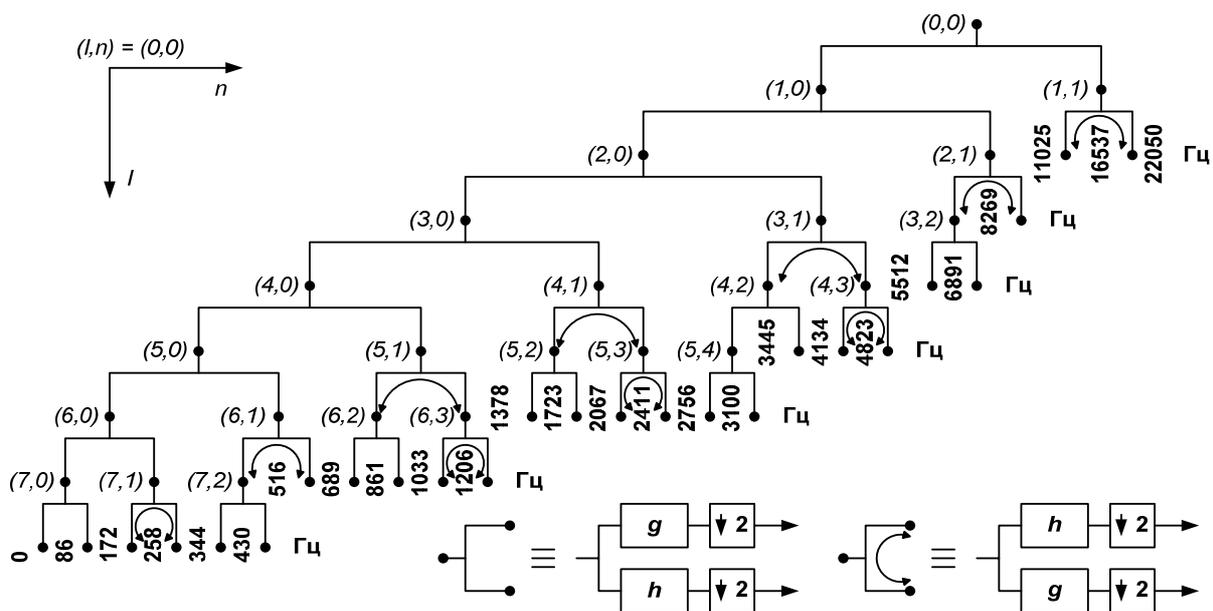


Рис.2. Дерево ПДВП $(l, n) \in E_{CB}$

На рис. 3 и рис. 4 показаны аппроксимации центральной частоты и ширины каждой частотной полосы критической шкалы частот соответственно деревом ПДВП, структура которого приведена на рис. 2.

Дерево $CB - WPD$ делит частотный диапазон $[0 - 22,05 \text{ кГц}]$ на 25 неравных полос $CBW(f)$ согласно [10], т.е. на 25 барков. Корневой узел $(l, n) = (0, 0)$ данного дерева соответствует всему частотному диапазону сигнала. Каждый внутренний узел дерева $(l, n) \in E$, названный узлом предка, делится на два потомка: 1-й потомок и 2-й потомок, ассоциируемые соответственно с высокочастотной и низкочастотной фильтрацией, выходные сигналы (вейвлет коэффициенты) которых децимируются в соотношении 2:1:

$$X_{l,n,k}(t) = \langle x(t), \varphi_{l,n,k}(t) \rangle, (l, n) \in E_{CB}, k \in \mathbb{Z}. \quad (8)$$

где l - номер масштабного уровня преобразования ($0 \leq l \leq 8$), n - номер узла масштабного уровня преобразования, k - число вейвлет коэффициентов в полосе (узле (l, n) дерева E). Банк вейвлет фильтров ($CB - WPD: (l, n) \in E_{CB}$), согласованный с критической шкалой частот восприятия акустической информации человеком, является предельной структурой для метода перцептуального кодирования сигнала звука [11]. Ширина каждой клеточки частотно-временного плана есть длина фрейма, которая определяется как $F_l = 2^l (F_{min} = 2 \text{ отсчетов и } F_{max} = 256 \text{ отсчетов})$. Следовательно, длина анализируемого окна равна $W = (P - 1)(F_l - 1) + 1$ отсчетов. Для первого уровня $l = 1$ преобразования определяющей является область верхних частот, и длина окна $W = 40$ отсчетов при длине фильтра прототипа $P = 40$. Для уровня $l = 8$ преобразования наибольшая частотная разрешающая способность в области нижних частот, а окно $W = 9946$ отсчетов.

Расчет порогов маскирования в области вейвлет-коэффициентов

Порог маскирования $T_{l,n}$ в каждой критической полосе $(l, n) \in E_{CB}$ вычисляется на основе спектральных компонент сигнала, попавших в полосу. Значение маскирующего порога непосредственно зависит от уровня

давления звука, частоты маскира, тональности маскира и маскируемых маскиров. Для точного определения маскирующих порогов нужен учет совместного влияния как частотного, так и временного маскирования. Причем, пост-маскирование, т.е. маскирование вперед, имеет более важное значение перед маскированием назад в силу своей продолжительности по времени.

Процедура расчета порогов маскирования в вейвлет области для кодера звука на базе ПДВП, согласованного с критической шкалой частот, показана в [12].

Динамическая трансформация частотно-временного плана на основе перцептуального критерия

Критерий оптимизации структуры ПДВП

Поиск структур деревьев ПДВП базируется на известном утверждении [9]: любая комбинация целых индексов $(l, n, k) \in \mathbb{Z}$, для которых вейвлеты сконцентрированы на двоичных интервалах $[n2^{-l}, (n + 1)2^{-l}]$ из диапазона $[0, \infty)$, соответствует ортогональным базисам $\psi_{l,n,k}(t), \varphi_{l,n,k}(t)$ из пространства $L^2(k)$. Утверждение доказывает существование множества структур ПДВП. Причем, ПДВП ассоциируется с алгоритмом выбора лучшей структуры E преобразования из множества структур, путем изменения и минимизации определенной меры качества (стоимостной функции) J , зависящей от обрабатываемых данных $\min J(Ax)$, где A - структура ПДВП, x - входной сигнал. В перцептуальном кодировании звукового сигнала, в первую очередь, интересна такая декомпозиция ПДВП, при которой минимизируется скорость передачи с сохранением соответствующего высокого качества восприятия человеком декодированного сигнала. Таким образом, оптимизация - это итеративный процесс, и «лучшее» дерево преобразования выбирается как можно ближе к шкале барков [13]. Конструкция алгоритма выбора лучшей декомпозиции E ПДВП следующая [9]. Для каждого узла $(l, n) \in E$ дерева ПДВП вычисляется стоимостная функция J . Далее рост структуры дерева осуществляется в соответствии со следующей процедурой:

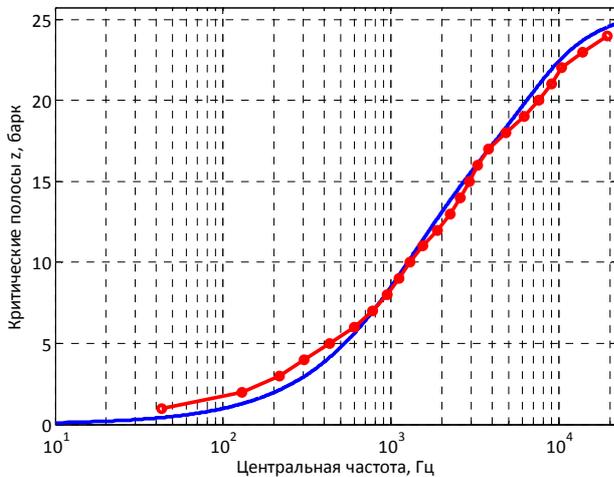


Рис. 3 Аппроксимация центральных частот $CB - WPD: (l, n) \in E_{CB}, l = 0,8$

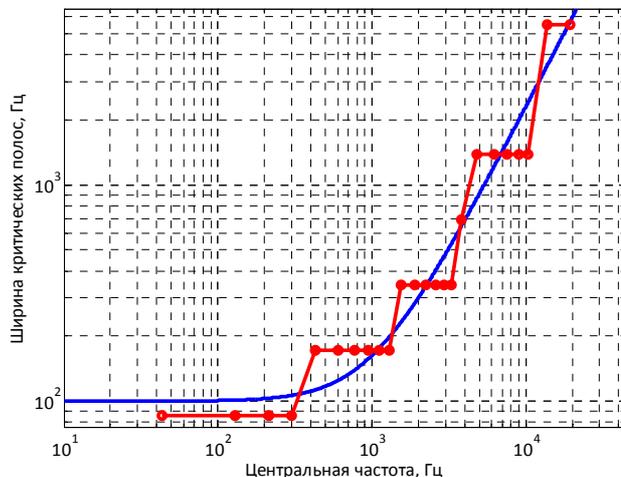


Рис. 4 Аппроксимация ширины критических частотных полос $CB - WPD: (l, n) \in E_{CB}, l = 0,8$

Процедура роста структуры дерева ПДВП:

Пусть $PE_{\text{предок}}$ - мера качества узла предка дерева ПДВП,

$PE_{\text{потомки}}$ - сумма мер качества соответствующих узлов потомков,

Тогда,

если $PE_{\text{предок}} > PE_{\text{потомки}}$,

то $PE_{\text{предок}} = PE_{\text{потомки}}$,

иначе отсечь узлы потомков.

Структура банка фильтров (дерева преобразования ПДВП) определяет вычислительную сложность алгоритма процедур анализа и синтеза на каждом фрейме анализа. Стратегия выбора структуры дерева ПДВП должна базироваться под заданный вычислительный ресурс (вычислительный ресурс – реконфигурируемая вычислительная среда: поточные, параллельно-поточные структуры процессоров). При этом декомпозиция дерева преобразования осуществляется в процессе работы ПДВП кодера, т.е. «на ходу». Преимущество данного подхода в том, что кодер учитывает локальные изменения спектра звукового сигнала. Любая структура дерева E может быть выражена конечными его узлами (l, n) :

$$E_j = \{(l, n) : l = \overline{1, L}; n = \overline{0, 2^l - 1}\}, \quad (9)$$

где j – номер структуры дерева из множества E ; L – максимальное значение масштабного уровня j -ой структуры дерева ПДВП.

Множество структур E_j образуют пространство структур деревьев ПДВП. В силу того, что в структуре дерева ПДВП каждый узел может является предком двух потомков (двух новых узлов) при переходе с уровня l на новый уровень $l + 1$, то структуры ПДВП зависимы друг от друга:

$$E_j \in E_{j+1}. \quad (10)$$

По мере обработки входного фрейма сигнала решение задачи предлагается разбить на два этапа:

1. На основании стоимостных функций для каждого узла дерева преобразования $(l, n) \in E$, определяемых перцептуальной энтропией, осуществляется декомпозиция данных узлов, при которой будет минимизировано требуемое количество бит кодирования вейвлет коэффициентов $X_{l,n,k}$, а шумы квантования не воспринимаются слухом человека. Таким образом, осуществляется рост дерева преобразования.

2. Оценивается информативность новой структуры дерева преобразования, т.е. как точно новое частотно-временное разрешение банка фильтров анализа локализовало нестационарности сигнала, а также определяется требуемый вычислительный ресурс.

Стоимостная функция декомпозиции узлов $(l, n) \in E$ дерева ПДВП выбирается следующим образом:

$$PE_{l,n} = \sum_{\forall X_{l,n,k}} \log_2(2[\text{nint}(SMR_{l,n,k})] + 1), \quad (11)$$

$(l, n) \in E, k \in \mathbb{Z}$,

где $SMR_{l,n,k}$ – отношение сигнала к порогу маскирования в полосе узла (l, n) дерева E . Соотношение (11) вычисляется для каждой частотной полосы (узла дерева ПДВП) и представляет собой отношение среднеквадратического значения вейвлет коэффициентов в узле (l, n)

к среднеквадратическому значению шума квантования, который будет внесен в реконструированный сигнал. Максимально возможный уровень шума, не воспринимаемый на слух человеком, очевидно, является значением порога маскирования. Таким образом, весовая функция $PE_{l,n}$ является индикатором необходимого числа битов для кодирования звука.

В качестве меры информативности дерева ПДВП может быть выбрана энтропия. Исходя из свойств энтропии [9], в частности, характеризующих среднюю неопределенность выбора, применительно в ПДВП кодеру, предлагается конструировать меру количества информативности ПДВП (неопределенности) некой структуры дерева преобразования из множества структур в виде следующей стоимостной функции:

$$J_{l,n} = \sum_{\forall (l,n) \in E_j} |X_{l,n,k}|^2 \log \frac{1}{|X_{l,n,k}|^2}, \quad (l, n) \in E, k \in \mathbb{Z}. \quad (12)$$

С учетом определения энтропии несложно показать, что

$$H(X_{l,n,k}) = \frac{J_{l,n}}{\|X_{l,n,k}\|^2} + \log \|X_{l,n,k}\|^2, \quad (l, n) \in E_i, k \in \mathbb{Z}, \quad (13)$$

т.е. минимизация стоимостной функции $J_{l,n}$ ведет к минимизации энтропии $H(X_{l,n,k})$, $(l, n) \in E_j$.

Динамическая трансформация алгоритма ПДВП

Как отмечалось выше, декомпозиция дерева ПДВП должна осуществляться в процессе работы ПДВП кодера, т.е. «на ходу», а структура дерева CB – WPD является предельной структурой преобразования. Таким образом, стоит задача оценки SMR не в критических частотных полосах восприятия акустической информации человеком, а в частотных полосах соответствующих узлов (l, n) текущего дерева E_j преобразования (см. [11]). Отношение среднеквадратического значения вейвлет коэффициентов $X_{l,n,k}$ в полосе узла (l, n) дерева E_j к соответствующему маскирующему порогу $T_{l,n}$, равномерно распределенному между $K_{l,n}$ коэффициентами $X_{l,n,k}$, $k = \overline{1, K_{l,n}}$ узла (l, n) , определяется следующим образом [11]:

$$SMR_{l,n,k} = \frac{|X_{l,n,k}|}{\sqrt{12 \cdot T_{l,n} / K_{l,n}}}, \quad (14)$$

где знаменатель $\sqrt{12 \cdot T_{l,n} / K_{l,n}}$ представляет собой максимальный шаг квантователя $\Delta_{l,n}$ вейвлет-коэффициентов в узле $(l, n) \in E_j$, а величина $SMR_{l,n,k}$ задает минимальное количество уровней квантования.

Следовательно, стоимостная функция $PE_{l,n}$ декомпозиции узлов $(l, n) \in E_j$ дерева ПДВП (роста структуры ПДВП) определяется как перцептуальная энтропия узла $(l, n) \in E_j$ и показывает требуемое число двоичных разрядов для кодирования звукового сигнала в частотной полосе, определяемой узлом (l, n) :

$$PE_{l,n} = \sum_{k=0}^{K_{l,n}-1} \log_2(2[\text{nint}(SMR_{l,n,k})] + 1), \quad \left[\frac{\text{бит}}{(l,n)} \right], \quad (15)$$

$(l, n) \in E_j, k \in \mathbb{Z}$.

Функция $PE_{l,n}(l, n) \in E_j$ представляет собой функцию перцептуальной энтропии Джонстона (1), однако вычисляемую для действительных коэффициентов и в вейвлет области для текущего дерева E_j ПДВП. Декомпози-

ция ПДВП, т.е. «рост» дерева преобразования или другими словами динамическая трансформация частотно-временного плана, может осуществляться на основании следующего алгоритма:

Пусть решение о декомпозиции узла (l, n) дерева E_j ПДВП будет обозначаться как $split(l, n)$, где l – уровень декомпозиции, т.е. масштабный уровень преобразования, а n есть n -й узел на уровне l . Пусть текущий узел (предок) будет (l, n) , а его потомки определяются как $(l + 1, 2n)$ и $(l + 1, 2n + 1)$, где $l = 0, 1, 2, 3, \dots, n = 0, 1, 2, 3, \dots$

Шаг 1. Пусть $l = 0$, $split(l, n) = YES$, т.е. задан корневой узел $(0, 0)$ дерева преобразования $E_0 = E_j$, где $j = 0$ – входной фрейм звукового сигнала, перцептуальная энтропия которого равна $PE_{0,0}$ и информативность дерева равна H_{E_0}

Шаг 2. Осуществляется декомпозиция узлов предков (l, n) входного сигнала на основе ячейки – банка из двух ортонормальных вейвлет фильтров.

Шаг 3. $l = l + 1, j = j + 1$.

ЕСЛИ $l - 1 >$ максимального масштабного уровня предельного дерева $CB - WPD$,

ТОГДА STOP – конец роста дерева ПДВП.

Шаг 4. Вычисляется перцептуальная энтропия $PE_{l,2n}$ и $PE_{l,2n+1}$ в узлах декомпозиции $(l, 2n)$ и $(l, 2n + 1)$ соответственно для всех n .

Шаг 5. Для каждого узла n уровня l рост дерева E_j ПДВП осуществляется следующим образом:

ЕСЛИ $PE_{l,n} \geq PE_{l+1,2n} + PE_{l+1,2n+1}$,

ТОГДА $split(l, n) = YES$

ИНАЧЕ $split(l, n) = NO$

Шаг 6. Оценивается информативность H_{E_j} дерева E_j :

ЕСЛИ $H_{E_j} > H_{E_{j-1}}$,

ТОГДА STOP – конец роста дерева, результирующее дерево ПДВП E_{j-1} .

Шаг 7. Переход к шагу 2.

Достоинством данного алгоритма является то, что рост дерева осуществляется сверху вниз, без возвратов на меньшие масштабные уровни преобразования и необходимости построения полного дерева ПДВП [1]. Данный алгоритм роста дерева ПДВП позволяет определить субоптимальную структуру декомпозиции ПДВП при минимальном числе бит на отсчет звукового сигнала без воспринимаемых на слух искажений, вносимых в процессе кодирования входного сигнала.

Эффективность данного алгоритма динамической трансформации частотно-временного плана иллюстрируется на 5. На 5б для входного сигнала 5а показаны этапы работы алгоритма роста дерева ПДВП и формирование частотно-временного плана.

Как видно, по мере увеличения декомпозиции дерева ПДВП порог маскирования – порог едва различимых искажений приближается к такому же порогу, который был построен для предельного дерева преобразования $CB - WPD$ (рис. 6).

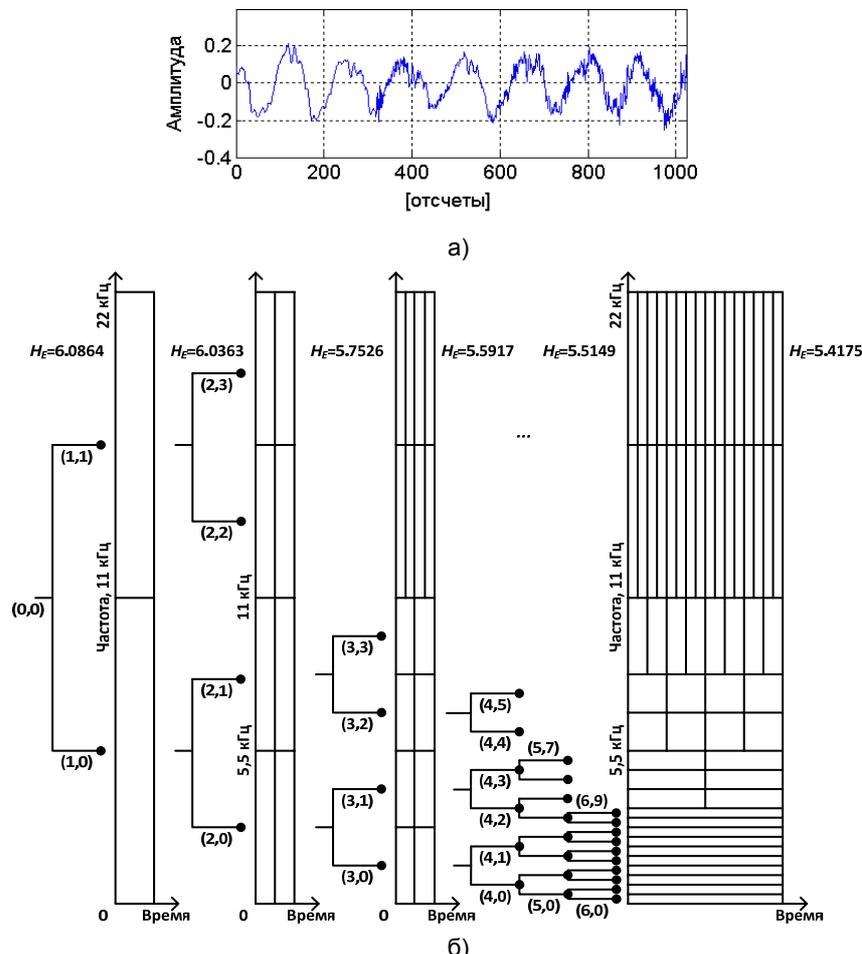


Рис.5. Формирование дерева ПДВП на основании алгоритма динамической трансформации частотно-временного плана

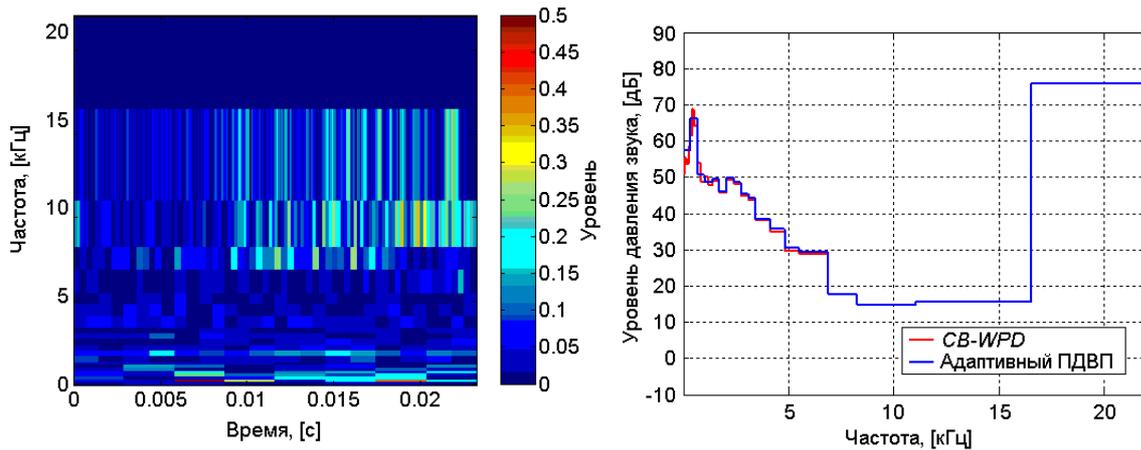


Рис.6. Спектрограмма и порог маскирования звукового сигнала

Структура кодера звука на базе динамической трансформации алгоритма ПДВП

На рис. 7 представлена структура перцептуального кодера звукового сигнала, ядром которой является блок «WPD», выполняющий реализацию адаптивного ПДВП. В темпе обработки сигналов для каждого текущего дерева E_i вычисляются пороги маскирования $T_{l,n}$, в соответствии с [12], значения перцептуальной энтропии $PE_{l,n}$, а также энтропии H_{E_i} структуры дерева ПДВП E_i . На основании данной информации в блоке «Split» рассчитываются параметры реконфигурации дерева ПДВП $r_{l,n}$ [1]. Данный процесс осуществляется поступательно, без возвратов на меньшие масштабные уровни преобразования. Следовательно, весь вычислительный процесс идеально ложится на архитектуру параллельно-поточных процессоров [14]. Обработка звукового сигнала

ла в ПДВП кодере выполняется кадрами, и ввиду того, что преобразование осуществляется над каждым последующим кадром в новой структуре дерева преобразования, то для устранения фазовых разрывов, соседние кадры сигнала перекрываются на 1/8 своей длины. Следующим этапом работы кодера является квантование и кодирование вейвлет коэффициентов $X_{l,n,k}: (l, n) \in E, k \in \mathbb{Z}$, результирующего субоптимального дерева ПДВП E . Управление данным процессом осуществляется в блоке «Алгоритм размещения бит». Кодирование структуры дерева ПДВП E выполняется в блоке «Кодирования структуры дерева ПДВП». Принимая во внимание тот факт, что рост дерева осуществляется поступательно и изменчивость сигнала во времени более инерционна, чем время обработки, кодируются только изменения структуры дерева от фрейма к фрейму.

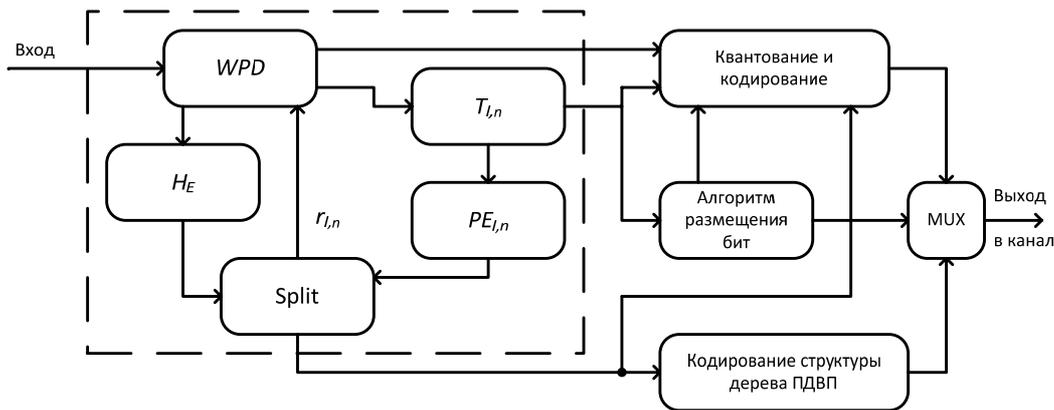


Рис. 7. Структура ПДВП кодера звука

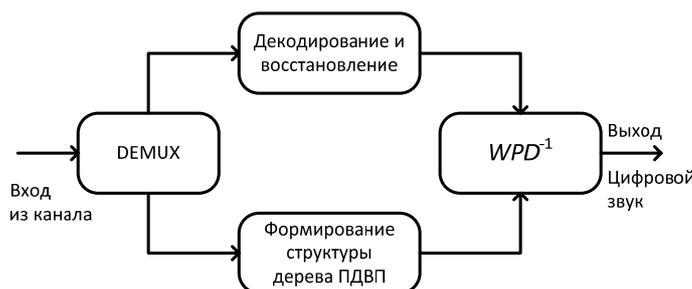


Рис. 8. Структура ПДВП декодера звука

Структура ПДВП декодера звука показана на рис. 8. Работа декодера выполняется в следующем порядке: входная информация разделяется на два потока данных в блоке «DEMUX»: один содержит кодированные вейвлет коэффициенты, а второй – код структуры дерева ПДВП; формирование структуры дерева ПДВП происходит в блоке «Формирование структуры дерева ПДВП» и осуществляется путём суммирования кодовой последовательности, описывающей изменение структуры дерева ПДВП, с текущей кодовой последовательностью структуры дерева ПДВП. В случае передачи первого фрейма, структура дерева ПДВП кодируется полностью в кодере. Реконструкция звукового сигнала выполняется в блоке « WPD^{-1} » на основе обратного адаптивного ПДВП над декодированными и восстановленными вейвлет коэффициентами (блок «Декодирования и восстановления») в соответствии со структурой дерева ПДВП. Синтезированные кадры реконструированного сигнала умножаются на модифицированное треугольное окно и складываются для формирования выходной последовательности $\hat{x}(n)$.

Вычислительная платформа системы кодирования

Структура вычислительной платформы

Структура вычислительной платформы кодера звука на основе ДТА-подхода включает управляющий процессор цифровой обработки сигналов (ЦПОС) и ПДВП процессор с реконфигурируемой архитектурой. На ЦПОС возложены следующие задачи: обработка вейвлет коэффициентов $X_{l,n,k}$ в частотных полосах (l, n) , соответствующих текущей структуре дерева E_i ; вычисление психоакустической модели дерева $(l, n) \in E_i: T_{l,n}$ и $PE_{l,n}$; получение вектора реконфигурации ПДВП процессора $r_{l,n}, (l, n) \in E_i$; формирование и разборка данных передачи $[code(X_{l,n,k}, (l, n) \in E_i, k \in \mathbb{Z})]$ и приема $\hat{X}_{l,n,k}(code), (l, n) \in E_i, k \in \mathbb{Z}$, если используется дуплексный режим работы кодера звука. ПДВП процессор реализуется на параллельно-поточной архитектуре с динамической реконфигурацией. Выбор данной архитектуры обусловлен алгоритмом роста дерева ПДВП. Процессор работает в дуплексном режиме: конвейер кодера $X_{l,n,k} = WPD_{l,n}(x_j(t)), (l, n) \in E_i, x_j(t) - j$ -й фрейм звукового сигнала и конвейер декодера $\hat{x}_j(t) = WPD_{l,n}^{-1}(\hat{X}_{l,n,k}), (l, n) \in E_i, \hat{x}_j(t) - j$ -й фрейм реконструированного звукового сигнала. Длина конвейера параллельно-поточного

процессора определяется предельной структурой дерева ПДВП ($CB - WPD$) (рис. 2). Анализ сложности реализации базовой операции ПДВП (пара КИХ фильтров нижних и верхних частот с децимацией), модулей коммутации и буферизации и требуемого аппаратного ресурса ПДВП процессора дает следующий результат: объём КЛБ для построения 8-ми модулей базовой операции ПДВП на MAC архитектуре – 2048, и для схем буферизации и переключения – 504. Расчетное число КЛБ с учётом 20% надбавки, обусловленной накладными расходами при реализации на конкретной ПЛИС, равно 3062 КЛБ.

Структурная схема вычислительной платформы представлена на рис. 9. Управляющий процессор реализован на ЦПОС TMS320C6713, а ПДВП процессор синтезирован на ПЛИС Xilinx семейства Spartan-3 XC3S2000 со структурой FPGA. На ПЛИС реализован также внешний цифровой интерфейс аппаратной платформы. Внешняя память SDRAM общим объёмом 8 Мбайт и ПЛИС подключаются к TMS320C6713 через 32-разрядный интерфейс внешней памяти EMIF. Интерфейс внешней памяти EMIF имеет четыре отдельные адресные области CE0-CE3. Flash память общим объёмом 2 Мбайта предназначена для хранения программ. Процессор использует Flash память в загрузочном режиме. После начальной инициализации модуля программа из Flash памяти автоматически, с помощью начального загрузчика, загружается во внутреннюю память управляющего процессора. Для работы со стерео звуком используется кодек TLC320AIC23. Кодек TLC320AIC23 подключается к ЦПОС с помощью двух последовательных каналов: McBSP0 – для доступа к конфигурационным регистрам, McBSP1 – для приема и передачи данных. Период частоты синхронизации может настраиваться за счет использования блоков DMC (Digital Clock Manager) ПЛИС. Для реализации внешнего интерфейса системы используются порты ввода/вывода ПЛИС.

Разработка проекта на ПЛИС реализована с помощью САПР Xilinx ISE. В данном случае конфигурационная последовательность для ПЛИС генерируется с помощью утилиты iMPACT пакета и загружается через интерфейс JTAG в конфигурационную Flash память платформы. Каждый раз после включения питания ПЛИС считывает эту последовательность из Flash памяти.

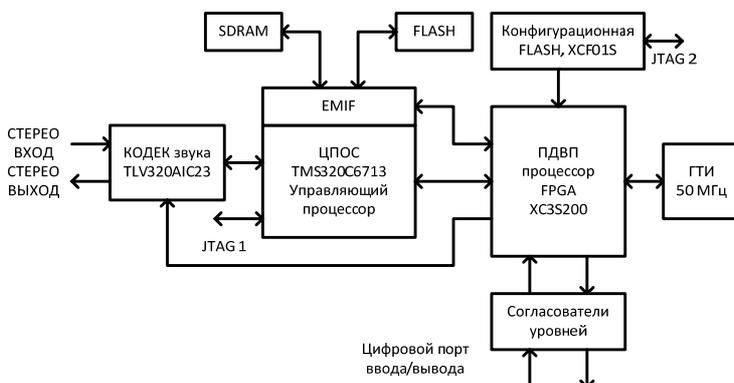


Рис. 9. Структурная вычислительной платформы



Рис. 10. Модуль цифровой обработки сигналов с реконфигурируемой логикой

Архитектурное решение вычислительной платформы позволяет получить следующие особенности: мощное вычислительное ядро с плавающей точкой для обработки данных с производительностью для 2400/1800 MIPS/MFLOPS; возможность распараллеливания задач обработки в двух-процессорной системе; реконфигурируемость платформы путем изменения конфигурационной последовательности ПЛИС и программ ЦПОС; возможность встраивания платформы в разнообразное оборудование за счет гибкой настройки внешнего интерфейса с помощью ПЛИС; использование стандартных средств конфигурирования и программирования. На рис. 10 представлен внешний вид модуля ЦОС с реконфигурируемой логикой.

Организации вычислительного процесса

Процесс кодирования звукового сигнала можно представить в виде последовательности следующих действий: прием очередного кадра сигнала длительностью 23 мс (1024 отсчета при частоте дискретизации 44.1 кГц); передача в ПЛИС вектора реконфигурации ; анализ входного сигнала; расчет параметров психоакустической модели на ЦПОС; формирование вектора реконфигурации ПДВП процессора задачи анализа; кодирование вейвлет коэффициентов на управляющем процессоре; формирование пакетов передачи коди-

рованной информации через цифровой порт.

Следует отметить, что работа кодера звука в реальном масштабе времени обуславливает достаточно жесткую синхронизацию вычислительных процессов управляющего процессора и ПДВП процессора. Так, перестройка ПДВП процессора осуществляется на основе формирования вектора реконфигурации согласно алгоритму динамической трансформации дерева ПДВП. При этом для фреймов, обрабатываемых согласно текущей структуре дерева на ступенях ПДВП процессора, управляющий процессор осуществляет выполнение процедур расчета порогов маскирования, оценки перцептуальной энтропии и информативности структуры дерева ПДВП. Распределение вычислительного ресурса между процедурами в управляющем процессоре показано на рис. 11. Время выполнения процедуры 1 (алгоритм вычисления порогов маскирования (рис. 12)) и процедуры 2 (оценка перцептуальной энтропии (рис. 12)) не постоянно, оно зависит от числа ступеней конвейера ПДВП процессора (кодер), задействованных в обработке текущего фрейма входного сигнала (см рис. 13). Этой же причиной обусловлено непериодическое выполнение процедур 4 и 5 (квантование и кодирование вейвлет коэффициентов).

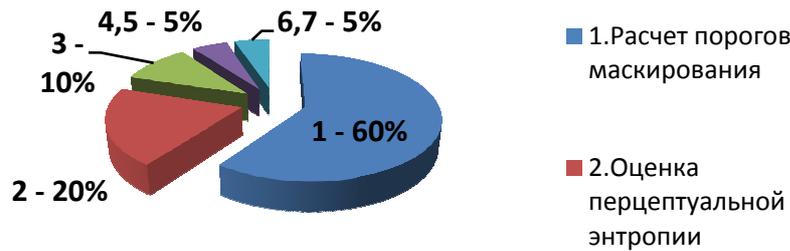


Рис. 11. Распределение вычислительного ресурса между процедурами в управляющем процессоре

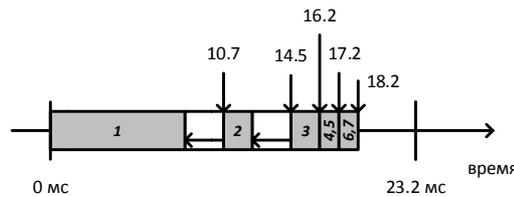


Рис. 12. Время выполнения процедур в управляющем процессоре

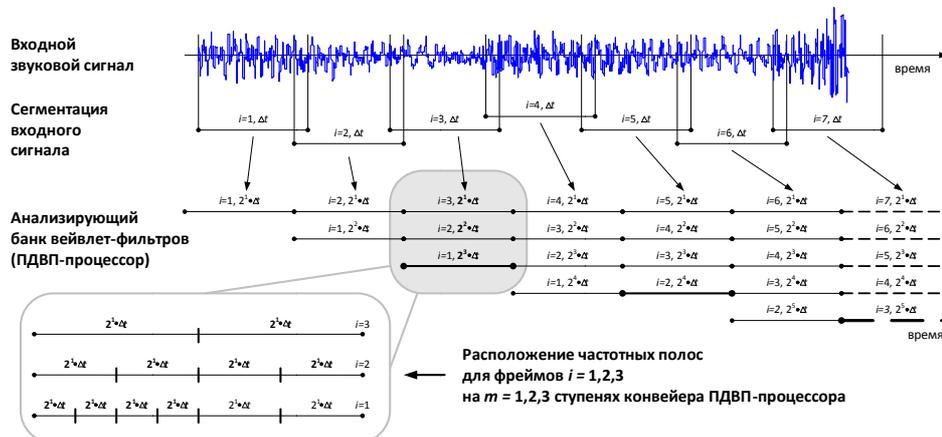


Рис. 13. Анализа входного сигнала в ПДВП процессоре

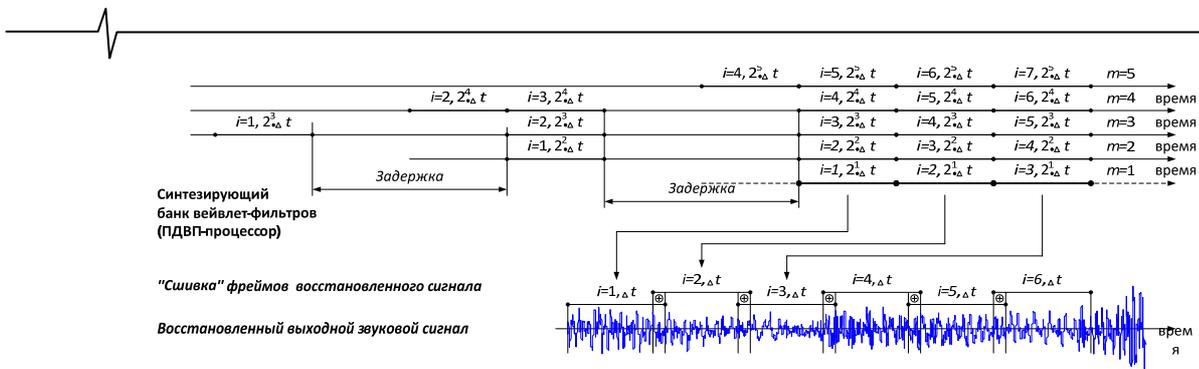


Рис. 14. Синтез выходного сигнала в ПДВП процессоре

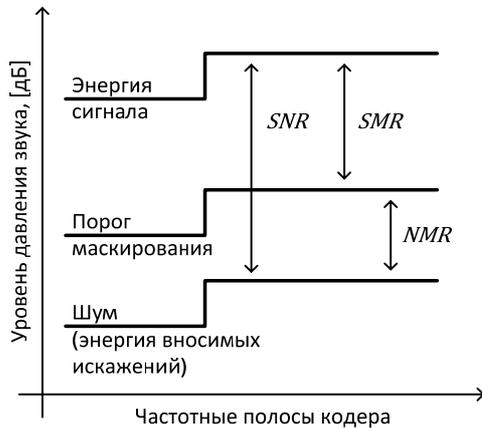


Рис. 15. Объективные оценки качества сигнала

В режиме декодирования управляющий процессор декодирует и восстанавливает фреймы из принимаемой кодовой последовательности, выполняя процедуры 6,7 (декодирование и восстановление), загружает восстановленный фрейм i на соответствующую ступень m конвейера ПДВП процессора (декодера). Переход фрейма i на следующую ступень процессора выполняется тогда, когда управляющий процессор примет следующий фрейм $i + 1$, который необходимо будет загрузить уже на задействованную ступень конвейера ПДВП процессора (декодера) (см. рис.14). Для согласования работ, выполняемых на каждой ступени конвейера процессора, необходимо ввести задержку, кратную времени обработки одного фрейма на одной ступени, тем самым будет обеспечиваться ритмичная работа параллельно-поточной структуры ПДВП процессора (декодер).

Результаты тестирования

Ниже представлены результаты исследований данного субполосного перцептуального ПДВП кодера звука. Для разных сигналов (шумовых, тональных, ритмичных) оценивалась возможная минимальная скорость передачи (кбит/с), максимальная степень компрессии, среднее количество бит на отсчет входного звукового сигнала, а также определялись объективные оценки качества: NMR_{total} , NMR_{SEG} – общее и сегментное соотношение шума к порогу маскирования, показывающее границу порога восприятия артефакторов на слух, SB - ширина полосы частот реконструированного сигнала, MF - процент искаженных фреймов реконструированного сигнала (см. рис.15) [15].

В таблице 1 приведены результаты экспериментов в сравнении с популярным кодером звука MPEG-1 уровень III [16], минимальная скорость передачи которого 64 кбит/с. Анализ полученных данных показывает, что минимальная скорость передачи для ПДВП кодера звука меняется от 36 кбит/с для тональных, ритмичных звуковых сигналов до 45 кбит/с для шумовых сигналов. При этом, оценка NMR_{total} меняется от -4,037 дБ до -8,857 дБ, т.е. граница порога восприятия минимум в два раза ниже порога маскирования. Для сравнения у MPEG-1 данная оценка позитивная, т.е. энергия воспринимаемых шумов не менее чем в два раза больше порога маскирования и, соответственно, практически каждый фрейм реконструированного сигнала содержит искажения и, $MF_{max} \cong 99,78\%$. Это обусловлено скорее тем, что ширина частотной полосы SB реконструированного сигнала достаточно мала: $SB_{min} = 11,89$ кГц, против 22,05 кГц в оригинальном сигнале (см. рис.16 – рис.18).

Таблица 1.

Оценки перцептуального ПДВП кодера звука и MPEG-1 уровень III

Тестовый материал $f_s = 44.1$ кГц, 16 бит ИКМ	Тип кодера	Мин. скорость передачи, [кбит/с]	Макс. степень компрессии	Среднее кол., [бит/отсчет]	Ширина частотной полосы SB, [кГц]	NMR_{total} , [дБ]	NMR_{SEG} , [дБ]	МГ, % искаженных фреймов
Boney-M 15с.	ПДВП-кодер	45	15.68	1.02	18.49	-8.297	-0.8699	27.04
	MPEG-1 ур. III	64	11.02	1.45	15.13	4.430	0.4076	99.85
Classic 10с.	ПДВП-кодер	42	16.8	0.95	17.62	-6.495	-0.7987	43.45
	MPEG-1 ур. III	64	11.02	1.45	12.58	3.284	0.3145	99.78
ABBA "Take a chance on me" 2с.	ПДВП-кодер	36	19.7	0.81	18.58	-4.037	-0.5844	66.25
	MPEG-1 ур. III	64	11.02	1.45	15.39	3.720	0.3312	98.70

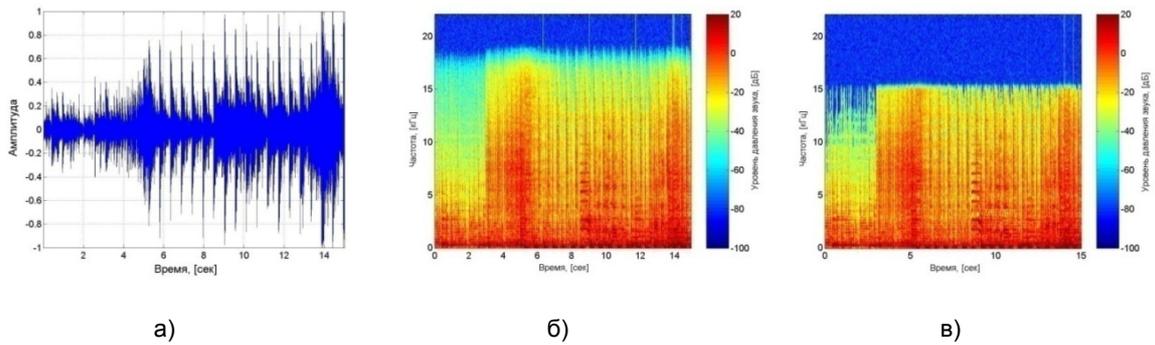


Рис. 16. Звуковой сигнал «VoiceM-2000»: а) оригинальный сигнал; б) спектрограмма реконструированного сигнала (ПДВП-кодер); в) спектрограмма реконструированного сигнала (MPEG-1 уровень III)

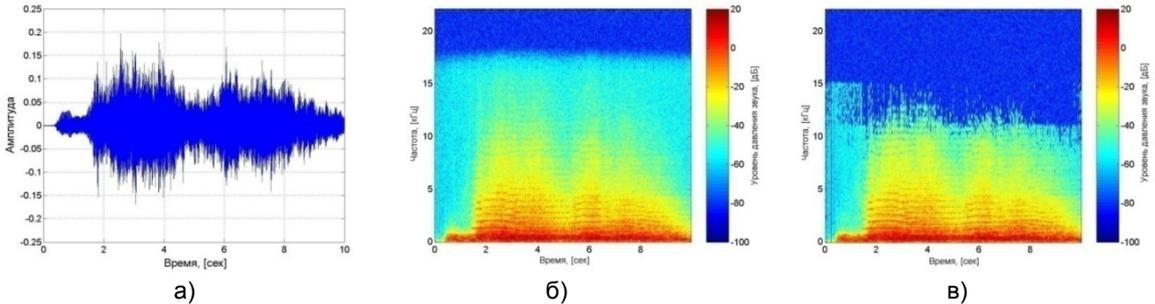


Рис. 17. Звуковой сигнал «Classic»: а) оригинальный сигнал; б) спектрограмма реконструированного сигнала (ПДВП кодер); в) спектрограмма реконструированного сигнала (MPEG-1 уровень III)

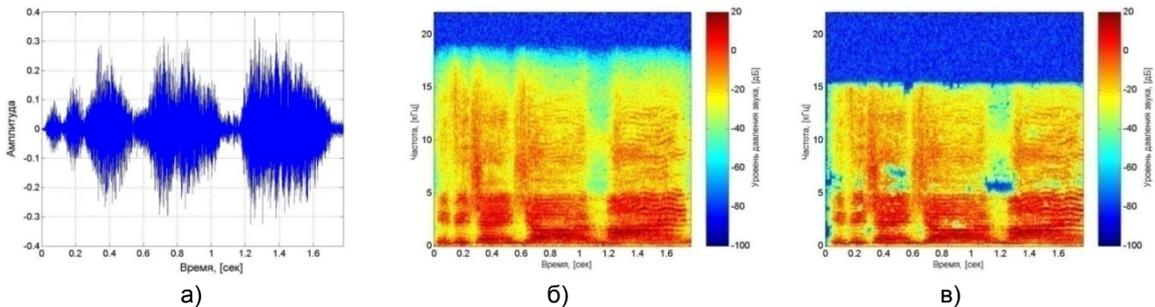


Рис.18. Звуковой сигнал «ABBA "Take a chance on me"»: а) оригинальный сигнал; б) спектрограмма реконструированного сигнала (ПДВП кодер); в) спектрограмма реконструированного сигнала (MPEG-1 уровень III)

Заключение

Таким образом, объективные оценки, полученные при экспериментальных исследованиях перцептуального ПДВП кодера звуковых сигналов, показывают, что реконструированный сигнал имеет достаточно хорошее качество, соответствующее требованиям стандарта ITU-R PEAQ [17] при высокой степени компрессии в 15 и более раз. Причём, для NMR_{total} меньше -10 дБ считается, что восстановленный сигнал не содержит слышимых артефактов. Для данного субполосного перцептуального ПДВП кодера звука данный показатель ≈ -9 дБ, но для схемы измерения NMR_{total} , главным достоинством которой, а также оценок MF и SB является то, что в отличие от [15], они вычисляются в вейвлет области и при тестировании кодеров звука на базе адаптивных ПДВП дают минимальную погрешность вследствие отсутствия взаимных частотно-вейвлетных преобразований [18,19].

Список литературы

1. Al.Petrovsky, A.Petrovsky, Dynamic algorithm transforms for reconfigurable real-time audio coding processor // The proc. of the Intern. Conference on Parallel Computing in Electrical Engineering, PARELEC'02, Warsaw, Poland, Sep. 22-25, 2002, - IEEE Press, Los Alamitos, California, 2002. – pp.422-424.
2. A. Spanias, T. Painter, V. Atti, Audio signal processing and coding. - Hoboken, NJ: John Wiley & Sons, Inc., 2007. – 464 p.
3. M. Bosi, R. Goldberg, Introduction to digital audio coding and standards. – Springer, Inc., 2003. – 434 p.
4. J.A. Kovalgin, D.A. Hitrov, M.V. Zyrianov. Advanced perceptual digital audio coding algorithm // In the proc. of the AES 17th intern. Conference "High-quality audio coding", Florence, Italy. – 1999. – pp.233-243.
5. N. Jayant, P. Noll. Digital coding of waveforms: principles and applications to speech and video. – Englewood Cliffs, NJ: Prentice-Hall, 1984. – 964 p.
6. J.D. Johnston. Transform coding of audio signals using perceptual noise criteria. IEEE Trans. on Select. Areas Commun. – Feb. 1988, – vol. 6. – pp. 314-323.
7. D. Sinha, A. Tewfik. Low bit rate transparent audio compression using adapted wavelets. IEEE Trans. on Signal Processing.- Dec. 1993, vol. 41 - pp. 3463-3479.
8. M. Vetterli, J. Kovacevic. Wavelets and subband coding. - Englewood Cliffs, NJ, Prentice-Hall, 1995. – pp. 488.

9. M.V. Wickerhauser. Adaptive Wavelet Analysis from Theory to Software. – A.K. Peters Ltd., Massachusetts, 1994. – 486 p.
10. E. Zwicker, H. Fastl. Psychoacoustics: facts and models. – Berlin, Springer-Verlag, 1990. – 380 p.
11. Al.Petrovsky, D.Krahe, A.A.Petrovsky. Real-time wavelet packet-based low bit rate audio coding on a dynamic reconfiguration system // AES, Convention paper 5778, 114th Convention, 2003 March 22-25, Amsterdam, The Netherlands. – 22 p.
12. Ал.А. Петровский. Построение психоакустической модели в области вейвлет-коэффициентов для перцептуальной обработки звуковых и речевых сигналов. – Речевые технологии. - № 4, 2008.- Москва. – С.61-71.
13. P.Srinivasan, L.H.Jamieson. High-quality audio compression using an adaptive wavelet packet decomposition and psychoacoustic modeling. IEEE Trans. on Signal Processing, vol. 46, No 4, April 1998. - pp.1085-1093.
14. Al.A.Petrovsky, A.A.Petrovsky. Reconfigurable processing: the solution to audio coding in real-time // The proc. of the 9th International conference "Mixed design of integrated circuits and systems", MIXDES'2002, 20-22 June 2002, Wroclaw, Poland. – pp.629-634.
15. K. Brandenburg, T. Sporer. "NMR" and "Masking Flag": evaluation of quality using perceptual criteria // In the proc. of the 11th intern. AES conference "Test and measurement conference", Portland, Oregon. – 1992. – pp.169-179.
16. Ю.А. Ковалгин, Э.И. Вологдин. Цифровое кодирование звуковых сигналов. Санкт-Петербург, Корона принт, 2004. - 240 с.
17. C. Colomes, C. Schmidmer, T. Thiede, W. Treurniet. Perceptual quality assessment for digital audio: PEAQ – the new ITU standard for objective measurement of the perceived audio quality // In the proc. of the AES 17th Intern. Conference "High-quality audio coding", Florence, Italy. – 1999. – pp.337 – 351.
18. Ал.А. Петровский. Объективная оценка качества восстановленного аудиосигнала перцептуальным ПДВП-кодером на базе периферийной модели уха человека // Сборник трудов 5-й Международной научной конференции "Цифровая обработка сигналов и её применение" (DSPA'2003), т.2, Москва, Россия, 2003. – С. 123-126.
19. Ал.А. Петровский. Субполосное перцептуальное кодирование звуковых и речевых сигналов.- Минск, «Беспринт», 2008. – 218 с.

Уважаемые коллеги!

Предлагаем вам принять участие в формировании тематических выпусков журнала «Цифровая обработка сигналов» и размещению рекламы продукции (услуг) Вашей фирмы на его страницах. В случае положительного решения просим представить в редакцию журнала Ваши предложения по плановому размещению информационных материалов и макет рекламы продукции (услуг) Вашей фирмы с указанием желаемого её месторасположения: обложка (2-я, 3-я или 4-я стр.), цветная внутренняя полоса (объем полосы).

В 2010 году планируется выпуск 4-х номеров журнала (тираж до 700 экз.). Журнал распространяется по подписке через агентство «Роспечать» в России, СНГ и странах Балтии (индекс 82185), а также на отраслевых всероссийских и международных Выставках.

Размещение рекламы Вашей фирмы на страницах журнала «Цифровая обработка сигналов» на плановой основе (не менее 2-х полных или 4-х половинчатых рекламных полос в течение года) предоставит Вам следующие возможности и права:

1. Первоочередное право расположения рекламных материалов на всех обложках (кроме 1-й) и страницах журнала.
2. Публикация представленных Вами рабочих (рекламных) материалов (статей) объемом до 6 полос в каждом очередном номере (в счет оплаченной рекламы).
3. Установка баннера Вашего сайта (или логотипа вашей организации) на 1-й странице сайта журнала «Цифровая обработка сигналов» (www.dsps.ru) в течение всего года, что привлечет внимание к продукции (услугам) Вашей фирмы новых участников на рынке DSP-технологий (ежедневно фиксируется до 100 и более посещений сайта www.dsps.ru).
4. Предоставление до 10 экз. очередного выпуска журнала.

Ориентировочная стоимость рекламных услуг:

4-я (внешняя) страница цветной обложки - 20 тысяч рублей.

2-я и 3-я (внутренние) страницы цветной обложки - 15 тысяч рублей.

1/2 цветной внутренней полосы - 7 тысяч рублей.

1/2 черно-белой внутренней полосы – 1 тысяча рублей.

Ждем Ваших предложений.

С наилучшими пожеланиями, зам. главного редактора
д.т.н., профессор Витязев Владимир Викторович

Предложения прошу направлять по адресу: E-mail: tor@rgta.ryazan.ru или info@dsps.ru

УДК 621.391

АДАПТАЦИЯ ПИЛОТ СТРУКТУРЫ И ДЛИНЫ ЗАЩИТНОГО ИНТЕРВАЛА OFDM СИГНАЛА К ИЗМЕНЯЮЩИМСЯ КАНАЛЬНЫМ УСЛОВИЯМ

Манелис В.Б., к.ф.-м.н., в.н.с. ООО «АРК-Радио», vldman@mail.ru.

Каюков И.В., к.т.н., руководитель лаборатории ООО «АРК-Радио», ikaioukov@gmail.com

Ключевые слова: OFDM, адаптация, пилот структура, защитный интервал, спектральная эффективность.

Введение

В последнее время широкое распространение получили многочастотные системы (OFDM – Orthogonal Frequency Division Multiplexing) [1]. В OFDM системах передаваемый поток данных разделяется на несколько низкоскоростных потоков, которые передаются на различных поднесущих. При этом можно увеличивать скорость передачи данных, не уменьшая длительность символа и сохраняя межсимвольную помеху на приемлемо низком уровне. OFDM системы имеют также и другие преимущества по сравнению с традиционными одночастотными системами: устойчивость к частотно-селективным замираниям, простота цифровой реализации и др.

Эти системы часто используют когерентный прием, который требует информацию о частотном отклике канала распространения. Оценка канала на приемной стороне выполняется с помощью известных (пилот) символов, некоторым образом распределенных в частотно-временной области. Качество оценки канала в значительной степени определяется структурой пилот сигнала (числом пилот символов и их расположением в частотно-временной области) и является ключевым фактором эффективной работы OFDM системы. Поэтому оптимизация структуры пилот сигнала в зависимости от скорости изменений во времени и частотно-селективных свойств канала распространения является важной задачей при проектировании OFDM систем. С увеличением числа пилот символов улучшается качество оценки канала, однако увеличивается также непроизводительный ресурс системы. Оптимальная пилот структура OFDM системы представляет собой компромисс между этими факторами.

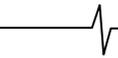
Для устранения межсимвольных помех каждый OFDM символ содержит защитный интервал (CP - cyclic prefix), который представляет собой последние L_{CP} отсчетов этого символа. Длина защитного интервала должна выбираться равной длительности импульсного отклика канала $h(t)$. Если длина защитного интервала меньше, появляются межсимвольные помехи, снижающие качество приема; если больше – необоснованно увеличивается доля неинформационной части сигнала. Обычно при проектировании OFDM системы величина защитного интервала фиксируется в соответствии с максимально ожидаемой длиной импульсного отклика канала. При этом она может достигать 1/8 или даже 1/4 длительности OFDM символа. То есть непроизводительный ресурс системы из-за защитного интервала может быть достаточно большим.

Предложен алгоритм адаптации длины защитного интервала и плотности пилот символов во временной и частотной областях OFDM сигнала. Адаптация основана на информации о частоте фединга и длине импульсного отклика канала пространства. Показано, что по сравнению с фиксировано установленными пилот структурой и длиной защитного интервала их адаптация снижает неинформативную долю сигнала и обеспечивает существенный выигрыш в спектральной эффективности OFDM системы

OFDM системы могут функционировать в различных канальных условиях. Для достижения высокой пропускной способности и надежности передачи данных они должны адаптивно подстраиваться к состоянию канала. Традиционными адаптивно подстраиваемыми параметрами являются способ модуляции, скорость кодирования, а также мощность сигнала. Кроме того, основываясь на информации о частоте фединга и длине канала, передатчик может устанавливать также такие пилот структуру и длину защитного интервала, которые максимизируют спектральную эффективность системы. Информацию о частоте фединга и длине канала в режиме частотного дуплекса передатчик получает с приемной стороны, а в режиме временного дуплекса осуществляет их измерение самостоятельно.

Цель данной работы – разработка и анализ алгоритма адаптации пилот структуры и длины защитного интервала к изменяющимся канальным условиям.

Оптимизация пилот структуры OFDM систем рассматривалась в [2-7]. В некоторых работах оптимизация пилот структуры выполнялась только в частотной области для постоянных во времени или квазистатических каналов [2,3]. В результате минимизации среднего квадрата ошибки оценки импульсного отклика канала получено оптимальное число пилот поднесущих, определяемое длиной канала. Поскольку схема предполагала изменение канала лишь в одном измерении, она не могла быть использована для двумерного случая, когда нужно учитывать изменения канала как в частотном, так и во временном направлении. Оптимизация пилот структуры в частотно-временной области для изменяющегося во времени частотно-селективного канала в [4,5] выполнялась при условии фиксированной плотности пилот символов. При быстрых изменениях канала во временном и частотном направлениях такой подход приводит к существенному ухудшению точности оценки канала и, как следствие, значительному снижению пропускной способности OFDM системы. Кроме того, в [5] требуется точное знание моментов доплеровского спектра и профиля многолучевости. Получение такой информации часто бывает затрудни-



тельными. В [6] анализ пилот структуры выполнялся для алгоритма оценки канала на основе винеровской двумерной интерполяции. Такая оценка канала сложна для реализации. Известно [7], что при фиксированной сложности реализации оценка канала с использованием двух последовательных 1-D интерполяторов (в частотной и временной области) предпочтительнее, чем на основе двумерной 2-D интерполяции.

В данной работе мы следуем этой 2x1-D концепции. Принято эквидистантное расположение пилот символов как в частотной, так и во временной области в соответствии со структурой, представленной на рис. 1, часто используемой в OFDM системах [1,5,7,8]. Интерполяция во временной области выполняется в соответствии с байесовским алгоритмом, в частотной области – в соответствии с алгоритмом максимального правдоподобия, не требующим знания профиля многолучевости [9,10].

Адаптация пилот структуры и защитного интервала выполняется на основе результатов максимизации спектральной эффективности OFDM системы [11,12] и заключается в выборе длины защитного интервала и расстояния между пилот символами во временной и частотной областях из некоторого набора их возможных значений в зависимости от доплеровской частоты и длины канала.

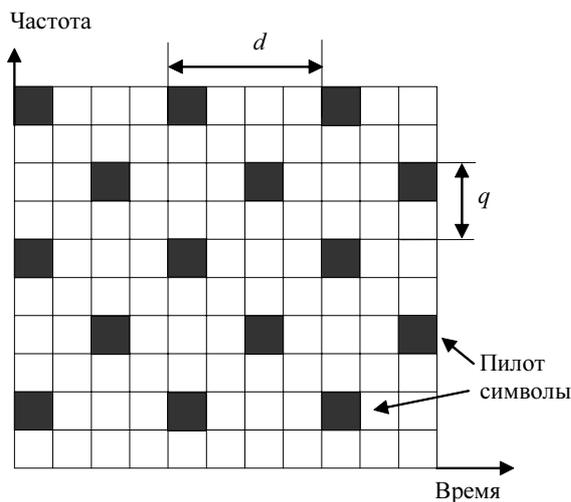


Рис. 1. Структура пилот символов OFDM сигнала

Анализ спектральной эффективности OFDM системы

Рассмотрим OFDM систему, использующую N поднесущих для параллельной передачи символов данных (рис. 2).

Над каждым передаваемым блоком N модулированных символов осуществляется обратное дискретное преобразование Фурье (ОДПФ). В результате формируется N -мерный вектор отсчетов во временной области, расстояние между отсчетами – $T_s = 1/F$, F – ширина полосы сигнала. Для устранения межсимвольных помех к началу полученного вектора добавляется защитный интервал. Сформированный таким образом OFDM символ передается в канал. На приемной стороне осуществляют последовательно-параллельное преобразование отсчетов принимаемого OFDM символа. При этом отсчеты защитного интервала удаляются, а оставшиеся отсчеты подвергаются дискретному преобразованию Фурье (ДПФ) размерностью N . В результате для принимаемого сигнала на n -ой поднесущей k -го OFDM символа имеем

$$X_{n,k} = b_{n,k} H_{n,k}^{(0)} + w_{n,k}, \quad (1)$$

$H_k^{(0)} = \{H_{n,k}^{(0)}\}$ – частотный отклик канала, который представляет собой ДПФ от вектора отсчетов импульсного отклика канала

$$\mathbf{h}_k = \{h_{0,k}, h_{1,k}, \dots, h_{L-1,k}\}^T, \quad n = \overline{0, N-1},$$

L – длина импульсного отклика канала в отсчетах, $b_{n,k}$ – переданный информационный или пилот символ, $w_{n,k}$ – белый гауссовский шум.

Оценка канала (канальных параметров) $P_{n,k}$ для пилот символов, вставленных в известные позиции частотно-временной области (n, k) , равна

$$P_{n,k} = X_{n,k} / b_{n,k}, \quad (2)$$

где $b_{n,k}$ – передаваемый пилот символ.

Когерентный прием данных требует оценки частотно-отклика канала $H_{n,k}$ для информационных символов. Эта оценка осуществляется посредством интерполяции по величинам (2). При этом в соответствии с концепцией 2x1-D интерполяция выполняется отдельно во временной и в частотной областях.

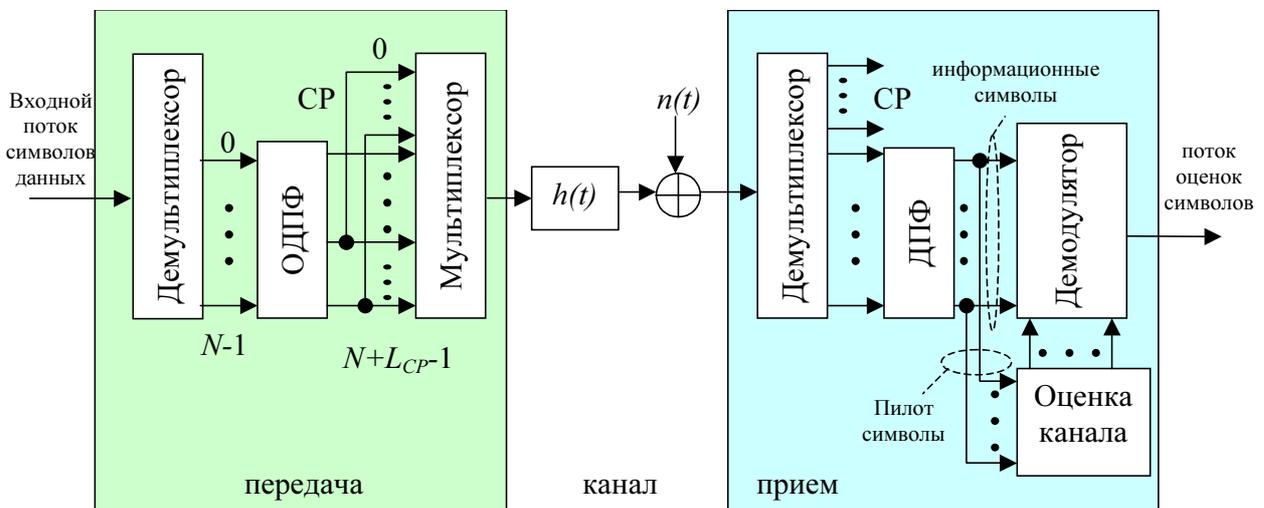


Рис. 2. Структурная схема OFDM системы

Причем сначала осуществляется интерполяция во временной области для тех поднесущих, которые содержат пилот символы [8]. Интерполяция выполняется в соответствии с байесовским алгоритмом (по минимуму среднего квадрата ошибки). Она требует информацию о доплеровской частоте f_D и отношении сигнал-шум по мощности Z и может быть записана как

$$Y_k = \mathbf{VQ}^{-1}\mathbf{P}, \quad (3)$$

где Y_k – результат интерполяции для k -го OFDM символа рассматриваемой поднесущей, номер которой для простоты опущен,

$$\mathbf{P} = (P_{k+l-ld}, \dots, P_{k+l-d}, P_{k+l}, P_{k+l+d}, \dots, P_{k+l+(l-1)d})$$

– вектор канальных коэффициентов пилот символов (2), используемых для интерполяции, $2l$ – число этих пилот символов, d – расстояние между соседними пилот символами поднесущей, l – расстояние от интерполируемого k -го OFDM символа до ближайшего прошедшего пилот символа,

$$\mathbf{V} = (J_0[2\pi f_D T(l-ld)], \dots, J_0[2\pi f_D Tl], \dots, J_0[2\pi f_D Tl+(l-1)d])$$

– вектор кросс-корреляции, определяемый расстоянием между интерполируемым информационным и опорными пилот символами, $J_0(\cdot)$ – функция Бесселя нулевого порядка первого рода,

$$\mathbf{Q} = \{Q_{i,m} = J_0[2\pi f_D Td(i-m)] + \delta_{i,m}/Z, \quad i, m = \overline{1, 2l}\}$$

– автокорреляционная матрица пилот символов, определяемая расстояниями между ними, $\delta_{i,m}$ – символ Кронекера.

Затем для каждого OFDM символа по опорным тонам, полученным в результате первой интерполяции, осуществляется интерполяция в частотной области. Она выполняется в соответствии с методом максимального правдоподобия [9,10]. Этот метод не требует знания профиля многолучевости и при больших отношениях сигнал-шум близок по характеристикам к байесовскому методу. Интерполяция в частотной области для OFDM символа, номер которого для простоты опущен, может быть представлена как

$$\mathbf{H} = \mathbf{G}(\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \mathbf{Y}, \quad (4)$$

где \mathbf{Y} – вектор опорных тонов, позиции которых $(i-1)q$, $i = \overline{1, N_p}$ соответствуют тем поднесущим, где выполнена интерполяция во временной области (где есть пилот символы), q – период пилот поднесущих,

$N_p = \text{floor}(N/q)$ – число опорных пилот тонов OFDM символа,

$$\mathbf{B} = \{\exp(-j2\pi(i-1)q(m-1)/N), \quad i = \overline{1, N_p}, m = \overline{1, L}\},$$

$$\mathbf{G} = \{\exp(-j2\pi(n-1)(k-1)/N), \quad n = \overline{1, N}, k = \overline{1, L}\}.$$

Оценка канала (4) используется при демодуляции данных.

В [11] выполнялась оптимизация параметров плотности пилот символов во временной d и в частотной области q при различных видах модуляции, различных значениях длины канала τ , частоты Доплера f_D и различных отношениях сигнал-шум. Исследование показало, что существует область параметров плотности пилот символов

$$f_D T d = 0.2 \div 0.4, \quad \tau F_s q = 0.2 \div 0.4, \quad (5)$$

в пределах которой значения спектральной эффективности близки к максимуму при любых рассмотренных в работе условиях. В (5) $T = T_s(N + L_{CP})$ – длительность OFDM символа, $F_s = F/N$ – разность частот соседних поднесущих. В пределах приведенной области проигрыш максимальному значению спектральной эффективности не превышает 5%. При $f_D T d > 0.4$ или $\tau F_s q > 0.4$ введенных пилот символов становится недостаточно, качество оценки канала резко падает, и наблюдается существенное снижение спектральной эффективности. В то же время заметим, что оптимальная по спектральной эффективности пилот структура может не обеспечивать высокую точность оценки канала, если для этого требуется очень большое число пилот символов. Это иллюстрируется на рис. 3, где показана зависимость вероятности пакетной ошибки PER и спектральной эффективности W от отношения сигнал-шум для OFDM системы с $N=128$, $L_{CP}=16$, скоростью кодирования $R=1/2$, модуляцией 16-QAM в двухлучевом канале, $f_D T = 0.025$, $\tau F = 12$. Приведены кривые для двух различных наборов параметров плотности пилот символов q и d : $q=4, d=16$ и $q=1, d=2$. Видно, что для первого (оптимального) набора из-за ошибок оценки канала вероятность пакетной ошибки больше. Тем не менее спектральная эффективность для этого случая оказывается выше. Например, при $Z=20$ дБ для первого, более разреженного набора $W=1.65$ бит/с/Гц, а для второго – $W=0.86$ бит/с/Гц.

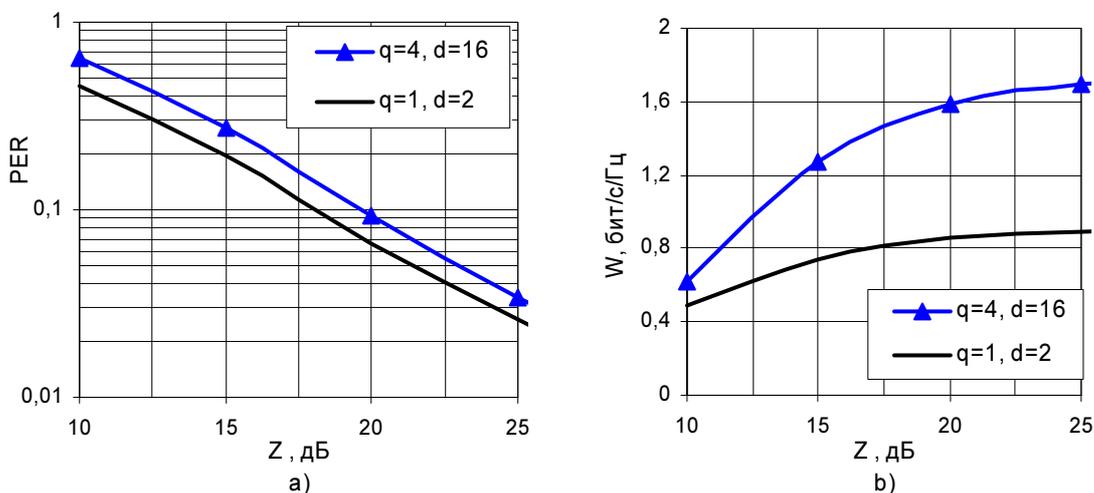


Рис. 3. Зависимость вероятности пакетной ошибки и спектральной эффективности от отношения сигнал-шум

Алгоритм адаптации параметров OFDM системы и анализ его эффективности

Изменяющиеся каналные условия ставят вопрос о текущей оценке параметров канала и необходимости адаптации к новым условиям эффективной передачи OFDM символов.

Поскольку оценка параметров канала τ , f_D может выполняться с ошибками, при адаптации параметров d , q будем ориентироваться на середины оптимальных интервалов (5).

В [11] показано, что максимальная спектральная эффективность достигается, когда длина защитного интервала равна длине канала. Поэтому при адаптации защитного интервала выбирается число его отсчетов L_{CP} , наиболее близкое к величине $F\tau$.

Полагаем, что при проектировании OFDM системы максимальные значения частоты Доплера $f_D^{(\max)}$ и длины канала $\tau^{(\max)}$, при которых она должна функционировать, известны. Пусть каждый адаптируемый параметр может принимать одно из четырех возможных значений, которые представляют собой степень числа два и сгруппированы в следующие множества:

$$L_{CP} \in \{2^n, 2^{n-1}, 2^{n-2}, 2^{n-3}\}, \quad q \in \{2^k, 2^{k+1}, 2^{k+2}, 2^{k+3}\},$$

$$d \in \{2^m, 2^{m+1}, 2^{m+2}, 2^{m+3}\}, \quad (6)$$

$$\text{где } n = \text{round} \left[\log_2(F\tau^{(\max)}) \right], \quad k = \text{round} \left[\log_2 \left(\frac{0.3N}{F\tau^{(\max)}} \right) \right],$$

$$m = \text{round} \left[\log_2 \left(\frac{0.3}{f_D^{(\max)} T_s (N + 2^n)} \right) \right], \quad \text{round}(\cdot) -$$

обозначает ближайшее целое. Первые элементы множеств (6) соответствуют адаптируемым параметрам при максимальных значениях частоты Доплера и длины канала. Такое задание множеств (6) обеспечивает возможность изменения адаптируемых параметров в широком динамическом диапазоне при небольшом числе возможных вариантов. Для сообщения об используемом варианте требуется всего 6 бит (по два бита на каждый адаптивный параметр).

Правило принятия решения о параметрах L_{CP} , q , d по измеренным величинам τ , f_D с учетом вышеизложенного заключается в определении элементов множеств (6), наиболее близких к величинам $F\tau$, $0.3N/F\tau$ и $0.3/f_D T_s$ соответственно. Учитывая, что эти элементы представляют собой степень числа два, получим

$$L_{CP} = \begin{cases} 2^n, & \text{round}[\log_2(F\tau)] > n \\ 2^{\text{round}[\log_2(F\tau)]}, & n \geq \text{round}[\log_2(F\tau)] \geq n-3 \\ 2^{n-3}, & \text{round}[\log_2(F\tau)] < n-3 \end{cases}, \quad (7)$$

$$q = \begin{cases} 2^k, & \text{round} \left[\log_2 \left(\frac{0.3N}{F\tau} \right) \right] < k \\ 2^{\text{round} \left[\log_2 \left(\frac{0.3N}{F\tau} \right) \right]}, & k+3 \geq \text{round} \left[\log_2 \left(\frac{0.3N}{F\tau} \right) \right] \geq k \\ 2^{k+3}, & \text{round} \left[\log_2 \left(\frac{0.3N}{F\tau} \right) \right] < k+3 \end{cases}, \quad (8)$$

$$d = \begin{cases} 2^m, & \text{round} \left[\log_2 \left(\frac{0.3}{f_D T_s (N + L_{CP})} \right) \right] < m \\ 2^{\text{round} \left[\log_2 \left(\frac{0.3}{f_D T_s (N + L_{CP})} \right) \right]}, & m+3 \geq \text{round} \left[\log_2 \left(\frac{0.3}{f_D T_s (N + L_{CP})} \right) \right] \geq m \\ 2^{m+3}, & \text{round} \left[\log_2 \left(\frac{0.3}{f_D T_s (N + L_{CP})} \right) \right] > m+3 \end{cases}. \quad (9)$$

Качество представленного алгоритма адаптации (7) - (9) будем характеризовать спектральной эффективностью, т. е. числом верно принятых бит в единицу времени на единицу выделенной полосы пропускания. Рассматривается пакетная передача данных с использованием помехоустойчивого сверточного кодирования. В неверно принятом пакете все биты полагаются принятыми неверно. Спектральную эффективность при фиксированных параметрах канала τ , f_D можно записать как

$$W(Z, K, \tau, f_D) = KR [1 - \text{PER}(Z, K, \tau, f_D)] \left(\frac{1}{1 + L_{CP}/N} \right) \left(1 - \frac{1}{dq} \right), \quad (10)$$

где K – количество бит, передаваемых в одном символе одной поднесущей и определяемое видом модуляции, R – скорость кодирования, $\text{PER}(Z, K, \tau, f_D)$ – вероятность пакетной ошибки, зависящая от отношения сигнал-шум Z , параметров канала τ , f_D и от вида модуляции. В (10)

$$\left(\frac{1}{1 + L_{CP}/N} \right) - \text{отношение числа информативных}$$

отсчетов к общему числу отсчетов OFDM символа,

$$\left(1 - \frac{1}{dq} \right) - \text{доля информационных символов в}$$

общем ресурсе OFDM сигнала.

Спектральная эффективность в условиях меняющихся параметров канала распространения может быть записана как

$$W(Z, K) = \int_0^{\tau^{(\max)}} \int_0^{f_D^{(\max)}} W(Z, K, \tau, f_D) p(\tau) p(f_D) d\tau df_D, \quad (11)$$

где $p(\tau)$ – плотность распределения длины канала, $p(f_D)$ – плотность распределения частоты Доплера, а $W(Z, K, \tau, f_D)$ определяется выражением (10).

Аналитически получить характеристики (10) или (11) затруднительно из-за сложности анализа помехоустойчивости квазикогерентной демодуляции в сочетании со сверточным декодированием в условиях многолучевого нестационарного канала распространения. Поэтому расчет вероятности пакетной ошибки и спектральной эффективности производился методом компьютерного моделирования OFDM системы со следующими основными параметрами: размерность ДПФ $N=128$, скорость кодирования $R=1/2$, модуляция – QPSK ($K=2$) и 16-QAM ($K=4$), размер пакета – 128 и 256 бит для каждого вида модуляции соответственно. Принимаемый сигнал представлял собой федингующий четырехлучевой OFDM сигнал, наблюдаемый на фоне гауссовского шума. Замирания сигналов лучей – независимые и соответствуют широко используемой модели Джейкса. Они расположены на интервале $[0, \tau]$ эквидистантно, их относительная мощность убывает по линей-

норму закону, так что сигнал четвертого луча слабее сигнала первого луча на 3 дБ. При моделировании предполагалось, что параметры канала τ, f_D известны точно.

Для случая различной длины канала рассматривались следующие плотности распределения длины канала: равномерная

$$p(\tau) = \begin{cases} 1/\tau^{(\max)}, & \tau \in [0, \tau^{(\max)}] \\ 0, & \tau \notin [0, \tau^{(\max)}] \end{cases} \quad (12)$$

и усеченная экспоненциальная

$$p(\tau) = \begin{cases} C \exp(-3\tau/\tau^{(\max)}), & \tau \in [0, \tau^{(\max)}] \\ 0, & \tau \notin [0, \tau^{(\max)}] \end{cases} \quad (13)$$

$C \approx 3.157$ – нормировочный множитель. Плотность распределения частоты Доплера выбиралась усеченно экспоненциальной

$$p(f_D) = \begin{cases} C \exp(-3f_D/f_D^{(\max)}), & f_D \in [0, f_D^{(\max)}] \\ 0, & f_D \notin [0, f_D^{(\max)}] \end{cases} \quad (14)$$

На рис. 4 представлена зависимость спектральной эффективности (11) при адаптивно выбираемых параметрах L_{CP}, q, d в соответствии с правилом (7)-(9) от отношения сигнал-шум при различных характеристиках канала. Кривая 1 соответствует распределению длины канала (13) с $\tau_{\max} F = 16$ и распределению частоты Доплера (14) с $f_D^{(\max)} T_s N = 0.025$, кривая 2 – распределению длины канала (13) с $\tau_{\max} F = 32$ и распределению частоты Доплера (14) с $f_D^{(\max)} T_s N = 0.05$, кривая 3 – распределению длины канала (12) с $\tau_{\max} F = 32$ и распределению частоты Доплера (14) с $f_D^{(\max)} T_s N = 0.05$. Из двух рассмотренных видов модуляции (QPSK, 16QAM) здесь и на последующих графиках выбиралась та модуляция, которая обеспечивает более высокую спектральную эффективность. Установлено, что QPSK модуляция предпочтительнее при отношении сигнал-шум приблизительно $Z \leq 10$ дБ, а 16QAM – при $Z > 10$ дБ. Видно, что с увеличением средней длины канала и частоты Доплера (что эквивалентно увеличению скорости изменения канала соответственно в частотной и временной области) спектральная эффективность OFDM системы несколько уменьшается.

На рис. 5 и рис. 6 для различных характеристик канала приведены кривые спектральной эффективности предложенного адаптивного алгоритма, а также для сравнения кривые спектральной эффективности при фиксировано выбранных параметрах L_{CP}, q, d . Рассмотрены два варианта фиксированного выбора: первый, при котором параметры соответствуют максимальным величинам $\tau^{(\max)}, f_D^{(\max)}$ (кривые «без адаптации, 1»):

$$L_{CP} = 2^{\text{round}[\log_2(F\tau^{(\max)})]}, \quad q = 2^{\text{round}[\log_2(\frac{0.3N}{F\tau^{(\max)}})]},$$

$$d = 2^{\text{round}[\log_2(\frac{0.3N}{f_D^{(\max)}T_s(N+L_{CP})})]},$$

и второй, при котором параметры соответствуют половине максимальных величин

$\tau^{(\max)}/2, f_D^{(\max)}/2$ (кривые «без адаптации, 2»):

$$L_{CP} = 2^{\text{round}[\log_2(F\tau^{(\max)}/2)]}, \quad q = 2^{\text{round}[\log_2(\frac{0.6N}{F\tau^{(\max)}})]},$$

$$d = 2^{\text{round}[\log_2(\frac{0.6N}{f_D^{(\max)}T_s(N+L_{CP})})]}.$$

Видно, что адаптация длины защитного интервала и пилот структуры позволяет повысить спектральную эффективность по сравнению с фиксированным заданием этих параметров. Причем, выигрыш может достигать 50%. Он растет с увеличением области возможных для OFDM системы значений длины канала и частоты Доплера, т. е. величин $\tau^{(\max)}, f_D^{(\max)}$.

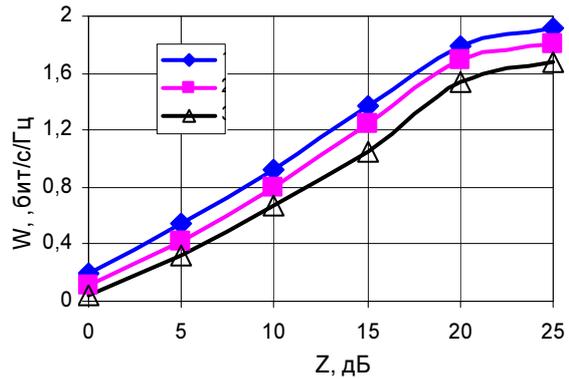


Рис. 4. Зависимость спектральной эффективности от отношения сигнал-шум при использовании адаптивного алгоритма

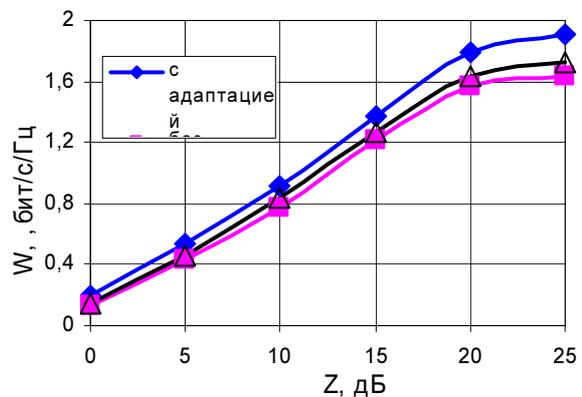


Рис. 5. Зависимость спектральной эффективности от отношения сигнал-шум, $\tau_{\max} F = 16, f_D^{(\max)} T_s N = 0.025$

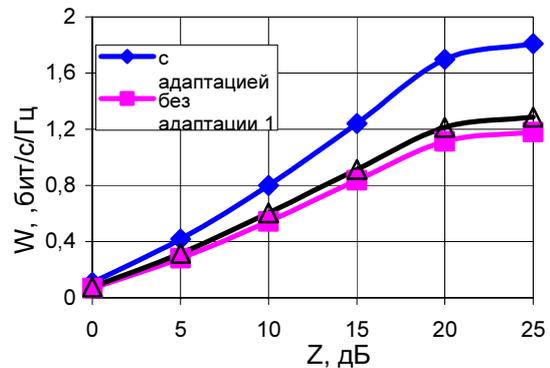


Рис. 6. Зависимость спектральной эффективности от отношения сигнал-шум, $\tau_{\max} F = 32, f_D^{(\max)} T_s N = 0.05$

Заключение

Таким образом, предложен алгоритм адаптации длины защитного интервала и плотности пилот символов во временной и частотной областях, позволяющий существенно повысить спектральную эффективность OFDM системы. По сравнению с фиксировано устанавливаемыми параметрами выигрыш может достигать 50%. При этом необходимый объем служебной информации возрастает несущественно. Для сообщения об используемом наборе адаптируемых параметров требуется всего 6 бит (по два бита на каждый адаптивный параметр). Полученные результаты позволяют рекомендовать процедуру адаптации длины защитного интервала и плотности пилот символов при проектировании OFDM системы связи для повышения ее спектральной эффективности.

Литература

1. R. van Nee, R. Prasad, OFDM Wireless Multimedia Communications, Artech House, Boston-London, 2000.
2. R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," IEEE Trans. Consum. Electron., vol. 44, no. 3, pp. 1122–1128, Aug. 1998.
3. S. Ohno and G. Giannakis, "Capacity maximizing pilots and precoders for wireless OFDM over rapidly fading channels," in Proc. Int. Symp. Signals, Systems and Electronics, Tokyo, Japan, Jul. 2001, pp. 246–249.
4. M. Dong, L. Tong, and B. M. Sadler, "Optimal pilot placement for channel tracking in OFDM," in Proc. Military Communications Conf. (MILCOM), Anaheim, CA, Oct. 2002, pp. 602–606.
5. J. Choi, and Y. Lee, "Optimum pilot pattern for channel estimation in OFDM systems," IEEE Trans. on Wireless Communications, vol. 4, no. 5, pp. 2083–2088, Sept. 2005
6. P. Hoeher, S. Kaiser, and P. Robertson, "Two-dimensional pilot-symbol aided channel estimation by Wiener filtering," in Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), Munich, Germany, Apr. 1997, pp. 1845–1848.
7. R. Nilsson, O. Edfors, M. Sandell, and P. O. Borjesson, "An analysis of two-dimensional pilot-symbol assisted modulation for OFDM," in Proc. Int. Conf. Personal Wireless Communications (ICPWC), Bombay, India, Dec. 1997, pp. 71–74.
8. Kaioukov I. V., Manelis V. B., Cleveland J. R., Channel Estimation for MIMO-OFDM Systems in Rapid Time-Variant Environments Based On Channel Statistics Estimation, GLOBECOM 2006– IEEE Global Telecommunications Conference, vol. 25, no. 1, November 2006, pp. 2752–2756
9. M. Morelli, U. Mengali A Comparison of Pilot-Aided Channel Estimation Methods for OFDM Systems IEEE, Transactions on Signal Processing, vol.49, no.12 December 2001 pp. 3065–3073
10. Каюков И. В., Манелис В. Б., Оценка канала в мобильных системах связи OFDM // Мобильные системы, 2005. – № 10, С.20-24.
11. Манелис В.Б., Каюков И.В. Оптимизация пилот структуры OFDM сигнала по максимуму спектральной эффективности// Радиолокация, навигация, связь, 14-я международная научно-техническая конференция, Воронеж, 2008, Т.2, С.847-854.
12. Манелис В.Б., Каюков И.В. Адаптация защитного интервала OFDM сигнала к изменяющимся канальным условиям // Радиолокация, навигация, связь, 14-я международная научно-техническая конференция, Воронеж, 2008, Т.2, С.855-860

OFDM SIGNAL PILOT PATTERN AND GUARD INTERVAL ADAPTATION TO VARIABLE CHANNEL CONDITIONS

V.B. Manelis, I.V. Kaioukov

Adaptation algorithm of OFDM signal guard interval and density of pilot pattern in time and frequency domains was proposed. Adaptation is based on information about Doppler frequency and length of propagation channel impulse response. It is shown, that proposed adaptation decreases overhead and provides substantial growth of the OFDM system spectral efficiency in comparison with the fixed pilot pattern and guard interval. The gain can achieve 50%.

Уважаемые коллеги!

Для тех, кто не успел оформить подписку на первое полугодие 2010 года через ОАО «Роспечать», сохраняется возможность приобретения журналов непосредственно в редакции по адресу: 107031, г. Москва, Рождественка, 6\9\20, стр. 1, Российское научно-техническое общество радиотехники, электроники и связи им. А.С. Попова, или оформить Заказ в соответствии с требованиями, выставленными на сайте журнала: www.dsra.ru.

Справки по телефонам: (495) 621-71-08, 621-06-10.

Факс: (495) 621-16-39.

E-mail: nto.popov@mtu-net.ru

tor@rgta.ryazan.ru

УДК 621.397

МЕТОД ВИЗУАЛИЗАЦИИ ИЗОБРАЖЕНИЙ С ШИРОКИМ ДИАПАЗОНОМ ЯРКОСТЕЙ

Бочаров К.Ю., Научно-исследовательский институт системных исследований РАН, mpostboxx@yandex.ru
Демидов В.М., к.т.н., ОАО «ЦНИИ «Циклон», demidovvova@yandex.ru

Ключевые слова: тепловизионное изображение, гистограмма, распределение уровней, яркость, контрастность, функция отображения.

Алгоритмы распределения уровней

Распределение уровней выполняется в несколько этапов: сбор статистических данных (гистограммы), построение функции отображения путем анализа собранных статистических данных, применение функции отображения к изображению.

Функция отображения представляется в виде таблицы преобразования входного сигнала в выходной. При автоматической регулировке функции отображения используется алгоритм автоматического регулирования уровней (АРУ).

Линейный метод

Самым простым методом построения функции отображения является линейный метод [1, 2]. На основании распределения уровней сигнала определяется отображаемый интервал. Обозначим крайние точки интервала S_a и S_b , нижнюю и верхнюю границы получаемых оттенков C_a и C_b , соответственно. Для восьмиразрядного выходного изображения можно принять $C_a = 0$, $C_b = 255$. Функция отображения сигнала S в оттенок C имеет вид:

$$C(S) = \begin{cases} C_a & , \text{при } S \leq S_a \\ K_l S + B & , \text{при } S_a < S < S_b \\ C_b & , \text{при } S_b \leq S \end{cases} \quad (1)$$

где $K_l = \frac{C_b - C_a}{S_b - S_a}$, $B = C_a - S_a K_l$.

Наиболее сложной задачей при использовании линейного метода является выбор границ отображаемого интервала – оттенков S_a и S_b . Как правило, строится гистограмма распределения входного сигнала, и границы выбираются с отсечением определенного числа крайних пикселей (самых ярких и самых темных). Отсечение крайних значений необходимо для исключения влияния выбросов (дефектных пикселей) на выходное изображение.

Если бы не надо было считать количество отсекаемых пикселей, то можно было бы и не строить гистограммы, а определять минимальное и максимальное значение входного сигнала. Задачу сбора гистограммы при необходимости можно оптимизировать по используемой памяти, поскольку для работы алгоритма не требуется вся гистограмма, а нужно лишь определенное число крайних значений.

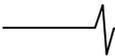
Рассматриваются методы преобразования изображения для согласования его уровней сигнала и палитры устройства отображения. Предлагается новый, усовершенствованный алгоритм автоматического регулирования уровней.

Сужение границ применяется и для уменьшения числа оттенков, выделяемых на объекты, которые предположительно нас не интересуют (очень яркие и очень темные либо напротив, недостаточно яркие или темные, как правило – фон). Например, при наблюдении горячих объектов целесообразно число отсекаемых ярких пикселей сделать как можно меньшим (достаточным для корректной работы), а число отсекаемых темных всего в несколько раз меньше площади фона. Для наблюдения общей картины сцены имеет смысл производить отсечение очень горячих и холодных объектов. Сами эти объекты будут видны как соответственно яркие или темные области, но средние температуры будут отображены более подробно.

Управление граничными значениями отображаемого диапазона может использоваться для регулирования параметров изображения. Так, при наблюдении ярких объектов целесообразно уменьшить число отсекаемых ярких пикселей и увеличить число отсекаемых темных, то есть сместить обе границы в направлении увеличения. Таким способом регулируется яркость изображения (выше уровни границ – ниже яркость и наоборот). Для повышения или понижения контрастности изображения границы можно сближать и раздвигать. Такая регулировка может проводиться автоматически, без необходимости в ручной коррекции. Для исключения мерцания изображения в случае быстро меняющейся сцены при использовании автоматического определения границ вводится накопление границ во времени.

Применение алгоритма поиска границ интервала для каждого кадра в отдельности в некоторых случаях дает нежелательный эффект. Последовательность изображений быстро меняющейся сцены может значительно отличаться по яркости и контрасту. Это приводит к мерцанию изображения и может значительно затруднить работу оператора. Для устранения этого явления вводится накопление по времени границ интервала. Алгоритм и параметры накопления зависят от характера наблюдаемых сцен и решаемых задач.

Линейный метод достаточно прост и часто дает приемлемый результат. Его основной недостаток заключается в том, что во всем отображаемом интервале оттенки распределены равномерно. Наблюдается эффект «засветки»: при контрастном по отношению к фону объекте детали на выходном изображении становятся неразличимыми.



Модернизацией линейного метода является кусочно-линейный метод, при котором увеличивают число исследуемых интервалов. При его использовании возникают вопросы: как определить число интервалов; если взять несколько интервалов, то в каком соотношении распределить оттенки по интервалам. Эти вопросы делают реализацию кусочно-линейного метода либо зависимой от очень большого числа настраиваемых параметров, либо применимой для наблюдения исключительно определенной по характеру сцены. Сложная логика выбора границ и наличие резкого перехода от одного интервала к другому повышают вероятность «неустойчивой» работы алгоритма. Под «неустойчивой» работой здесь понимается резкое изменение палитры между кадрами, что проявляется в виде «вспышки» – резкого изменения яркости изображения, либо «подмаргивания» – небольшого, но периодического изменения яркости изображения некоторого объекта. Надо отметить, что подобные эффекты могут проявляться не только при заметном изменении сцены, но и при ее относительном постоянстве (при незначительных ее изменениях или даже при влиянии шума). В связи с этим достаточно гибкий и надежный кусочно-линейный метод трудно реализовать на практике.

Гистограммный метод

Другой метод преобразования изображения – гистограммный, связанный с так называемым выравниванием гистограммы (histogram equalization) [1-3]. Метод сводится к вычислению функции отображения непосредственно по значениям гистограммы входного сигнала.

Функция отображения в интервале [0; 1] вычисляется в соответствии с выражением:

$$F_i = \frac{\sum_{j=0}^i H_j}{\sum_{j=0}^{n-1} H_j}, \quad (2)$$

где F – таблица отображения (палитра); H – гистограмма входного изображения; i – рассматриваемая градация; j – индекс градации в гистограмме; n – размер гистограммы.

В простейшем случае, если расчет гистограммы проводился без каких-либо предварительных операций над входным видеосигналом, сумма в знаменателе равна числу пикселей в изображении.

Суммирование в числителе рационально проводить итерационно, последовательно определяя каждый элемент палитры. Используя вышесказанное и перейдя от интервала получаемых оттенков [0; 1] к заданному интервалу, определенному границами C_a и C_b , представим выражение (2) в следующем виде:

$$F_i = K_g \cdot Hs_i + C_a, \quad (3)$$

$$\text{где } K_g = \frac{C_b - C_a}{\sum_{j=0}^{n-1} H_j},$$

$$Hs_0 = H_0, \quad Hs_i = Hs_{i-1} + H_i, \quad i = 1 \dots n-1.$$

Заметим, что если в линейном методе можно было и не хранить палитру, вычисляя оттенок непосредственно для каждого пиксела, то здесь значительно рациональнее вычислять палитру сразу целиком, за один проход.

Гистограммный метод хорошо работает для изображения, где присутствует большое число объектов с разнообразной яркостью. Всем оттенкам на полученном изображении соответствует примерно одинаковое число пикселей. Следовательно, на изображении отображается большинство мелких деталей. С другой стороны, на большие однотонные объекты выделяется большое число оттенков. Как результат, резко проявляется шум.

Еще один недостаток проявляется в отсутствии резкой границы между двумя объектами с различным уровнем сигнала. Это связано с отсутствием выделения промежуточных оттенков, лежащих между яркостями этих двух объектов.

Модернизированный гистограммный метод

Предлагается достаточно простой метод обработки изображений с широким диапазоном яркостей, базирующийся на рассмотренных методах, но позволяющий в значительной мере преодолеть основные недостатки каждого из них.

Сначала рассмотрим вопрос о том, как уменьшить влияние шума и снизить влияние больших однотонных зон, а затем – как обеспечить контраст между двумя объектами с различной яркостью. Для решения этих задач в гистограммный метод добавляется алгоритм преобразования гистограммы входного сигнала перед ее применением для построения функции отображения.

Гистограммный метод выделяет тем больший перепад в оттенках, чем большую часть изображения занимает сигнал:

$$F_i - F_{i-1} = K_g \cdot (Hs_i - Hs_{i-1}) = K_g \cdot H_i. \quad (4)$$

Большие значения величин в гистограмме, т.е. большое число пикселей с одинаковым оттенком, приводит к выделению неиспользуемых промежуточных оттенков и к чрезмерному повышению контрастности (повышение уровня шума). Для устранения этого эффекта значения H_i гистограммы входного сигнала, превышающие определенный уровень, ограничиваются сверху значением H_{max} :

$$H'_i = \begin{cases} H_i & , \text{при } H_i < H_{max} \\ H_{max} & , \text{при } H_i \geq H_{max} \end{cases}$$

Как показали эксперименты, задавать ручную максимальный уровень H_{max} нельзя, поскольку его выбор сильно зависит от контрастности входного изображения. Исходя из того, что отсекается больших значений гистограммы позволяет уменьшить влияние шума, а уровень шума на входном изображении можно принять постоянным, вводится ограничение на значение перепада оттенков:

$$F_i - F_{i-1} \leq \Delta C,$$

где ΔC – эмпирическая величина, при которой шум имеет приемлемые значения. Поиск H_{max} выполняется исходя из этого условия.

Поскольку коэффициент K_g в (4) зависит от гистограммы, уменьшение значений гистограммы влечет к увеличению K_g . Для выбора уровня H_{max} можно использовать следующее ограничение:

$$\frac{C_b - C_a}{\sum_{j=0}^{n-1} H'_j} \cdot H_{max} \leq \Delta C \quad (5)$$

Введем обозначение $\Omega_{H_{\max}} = \sum_{j=0}^{n-1} H_j'$.

Тогда для исключения деления неравенство (5) можно представить в виде:

$$(C_b - C_a) \cdot H_{\max} \leq \Delta C \cdot \Omega_{H_{\max}}. \quad (6)$$

Поскольку величина $\Omega_{H_{\max}}$ постоянно меняет свое значение, зависимость между предыдущим и последующим ее значениями при уменьшении H_{\max} на единицу можно представить в следующем виде:

$$\Omega_{h-1} = \Omega_h - G_h,$$

где G_h – количество столбцов гистограммы, равных по высоте h .

Напомним, что при анализе Ω_h имеем дело с гистограммой с ограниченными значениями, то есть столбцов выше h нет. Следовательно, можно рассчитать гистограмму G на основании гистограммы H .

При уменьшении значения H_{\max} на единицу производится перерасчет $G_h := G_h + G_{h+1}$, а затем $G_{h+1} := 0$. В программе обнуление можно не производить, а суммирование осуществлять, добавляя G_{h+1} к некоторой переменной, исполняющей роль текущего элемента, без изменения значений самой гистограммы G . Последовательное уменьшение H_{\max} происходит от максимального значения гистограммы, пока не будет выполняться условие (6). При выполнении этого условия текущее значение G_h равно искомому H_{\max} .

Для обеспечения контраста между объектами различной яркости надо поднять уровень гистограммы между зонами с большими ее значениями. Проведение корректного анализа зон с малым значением гистограммы и выделение на них соответствующих оттенков является достаточно сложной задачей. Наиболее простым решением является добавление ко всей гистограмме на определенном интервале интенсивности $[S_a, S_b]$

константы $P = \frac{L}{S_b - S_a}$, где L – эмпирически выбранная

величина.

Соотношение между L и $\Omega_{H_{\max}}$ показывает, насколько велик вклад линейной составляющей в функции отображения по отношению к гистограммной

составляющей. Иными словами, параметр L регулирует пропорцию применения гистограммного и линейного метода. При более детальном анализе становится видно, что параметры ΔC и L задают ограничения наклона функции отображения (контраста) сверху и снизу.

Реализация

При использовании модернизированного гистограммного алгоритма для регулирования доступны несколько параметров. Настройка двух из них – ΔC и L связана с субъективным восприятием изображения наблюдателем. Поэтому важно обеспечить возможность изменять эти параметры непосредственно в процессе работы с прибором.

При реализации устройства для ручной регулировки был выбран параметр ΔC , ограничивающий число оттенков на гистограмме. Его увеличение приводит к возрастанию количества оттенков, выделяемых на элементы изображения наибольшей площади. Вместе с тем, повышается уровень шума на таких элементах. Подобный режим удобен для сцен, в которых есть крупные яркие объекты на относительно равномерном фоне. В этом случае сохраняется детализация и объекта, и фона, но становятся более заметными шумы. Уменьшение параметра ΔC приводит к выравниванию гистограммы, как следствие – к более равномерному распределению оттенков. Функция отображения приближается к соответствующей функции линейного метода, со всеми его плюсами и минусами.

Альтернативой является регулирование параметра L . Его увеличение приведет к повышению общей контрастности и выделению мелких деталей на переходных участках изображения.

В ходе проведенных экспериментов модифицированный гистограммный метод показал хорошие результаты для большинства наблюдавшихся реальных сцен. По визуальным оценкам качества выходного изображения в большинстве случаев он превосходит гистограммный алгоритм и имеет значительное преимущество перед линейным методом. Примеры изображений, преобразованных с использованием рассмотренных алгоритмов, приведены на рис. 1.

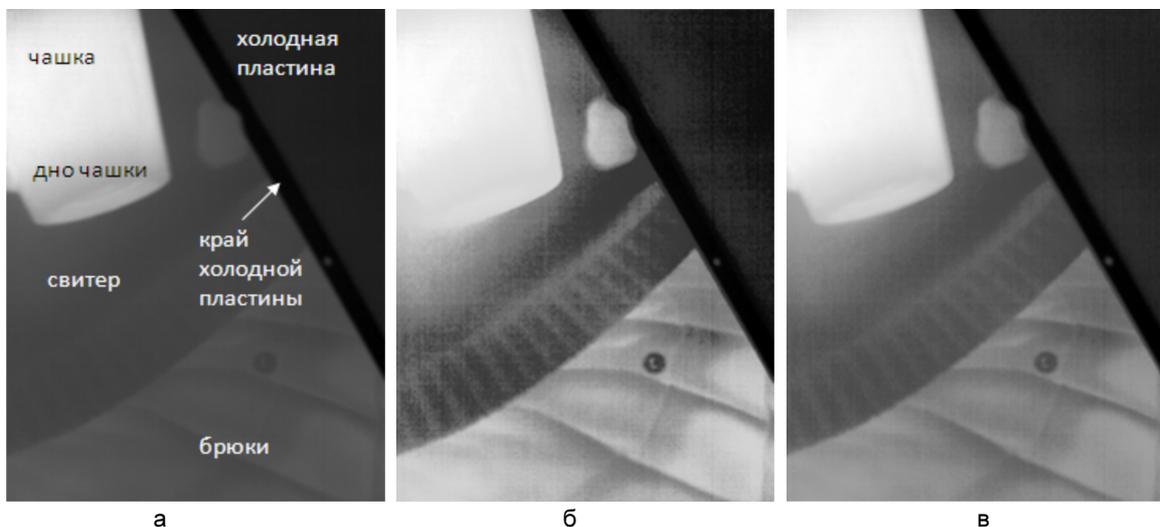
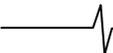


Рис. 1. Примеры визуализации тепловизионных изображений, полученных с использованием линейного метода (а), гистограммного метода (б) и модифицированного гистограммного метода (в).



Заключение

Таким образом, модифицированный гистограммный метод достаточно прост в реализации и не требует значительных вычислительных ресурсов, что позволяет реализовать его в современных системах цифровой обработки видеоизображения в режиме реального времени. Регулирование параметров преобразования позволяет использовать алгоритм в различных условиях, оптимизируя функцию преобразования для той или иной сцены. Предложенный алгоритм позволяет получить изображение для восприятия человеком лучше или, по крайней мере, не хуже, чем при использовании линейного или гистограммного методов обработки изображений.

Литература

1. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М.: Техносфера, 2005.
2. Методы компьютерной обработки изображений / Под. ред. В.А. Сойфера. 2-е изд., испр. – М.: ФИЗМАТЛИТ, 2003.
3. Шапиро Л., Стокман Дж. Компьютерное зрение. – М.: БИНОМ. Лаборатория знаний, 2006.

VISUALIZATION TECHNIQUE FOR WIDE INTENSITY RANGE IMAGES

Bocharov Konstantin, Demidov Vladimir

In this paper image processing methods for matching levels of input signal to display device palette are considered. New efficient algorithm for automatic level adjustment is proposed.

Новости Университетской программы Texas Instruments

Фирма Texas Instruments Inc. (TI) продолжает вести активную поддержку российских вузов в рамках Университетской программы. Руководители программы хотят обратить внимание всех заинтересованных лиц – преподавателей, научных работников и ученых, аспирантов и студентов – на следующие наиболее интересные события, касающиеся развития Университетской программы, произошедшие в 2009 году.

Одним из важнейших пунктов в развитии программы следует считать открытие раздела русскоязычного сайта фирмы TI, посвященного Университетской программе (<http://www.ti.com/ww/ru/universities.htm>) с возможностью регистрации на русском языке. Это событие произошло еще в конце 2008 – начале 2009 года. В 2009 году открылся еще один раздел этого сайта – организована страница форумов по различным тематикам, касающимся продуктов TI, в том числе форум по Университетской программе на русском языке (<http://e2e.ti.com/ru/forums/default.aspx>). Ответственные за программу специалисты TI, разработчики и работники образования готовы к диалогу в on-line-режиме.

Подготовлены обучающие материалы по сигнальным контроллерам F28335 семейства “Delfino”, пока только на английском языке. Эти материалы являются переработкой обучающего компакт-диска (Teaching ROM) по контроллеру F2812 и в настоящий момент проходят тестирование. Желающие принять участие в тестировании данных материалов могут отправить свою полную контактную информацию (на английском языке) на электронный адрес: j-hodgkiss@ti.com.

Руководители Университетской программы предлагают обратить внимание на следующие новые аппаратные средства фирмы TI, которые могут оказаться полезными в развитии учебных и исследовательских лабораторий высших учебных заведений. Это, во-первых, полноценное отладочное средство для сигнального процессора TMS320C5505 стоимостью всего \$49 (<http://focus.ti.com/docs/toolsw/folders/print/tmdx5505ezdsp.html>). Отладочный модуль TMS320VC5505 eZdsp выполнен в формате USB-брелка. В комплект входят полноценный эмулятор и отладочная среда CCStudio V.4. Низкая стоимость делает доступным профессиональный инструмент разработки устройств на базе DSP исследователям, студентам и просто увлеченным людям. Комплект ориентирован на разработку портативных устройств, таких как: аудиоплееры, диктофоны, IP-телефоны, портативные медицинские приборы и другие.

Второе отладочное средство – это набор разработчика на базе новых энергоэффективных сигнальных процессоров OMAP-L138 и TMS320C6748, C6746, C6742 (<http://focus.ti.com/docs/toolsw/folders/print/tmdxl138logicexp.html>). Стоимость комплекта – \$149. Осуществляется поддержка открытым сообществом. Комплект построен на базе процессора OMAP-L138, имеет 64 МВ памяти mDDR, комплектуется операционной системой Linux, драйверами DSP/BIOS™ и может быть использован в разработках устройств на любом из четырех вышеуказанных процессоров, ориентированных на промышленное и коммуникационное оборудование.

В заключение кратко расскажем о новых DSP-лабораториях, открытых в российских вузах и странах СНГ при поддержке Университетской программы TI в 2009 году. В Чувашском государственном университете открыта новая учебная лаборатория. Базовым оборудованием фирмы TI здесь являются отладочные комплекты C28027USB “Piccolo Sticks”. Они используются для обучения основам DSP и индивидуальной работы студентов. В Самарском государственном университете развернута лаборатория на базе микроконтроллеров MSP430 и малопотребляющих радиочастотных устройств. Таганрогский технологический институт «Южного федерального университета» открыл новую лабораторию на базе микроконтроллеров MSP430. Черниговский государственный технологический университет (Украина) развернул учебную лабораторию на базе стартовых наборов разработчика на процессорах C64x и C67x.

Пожелаем успехов российским преподавателям и научным работникам в разработке учебных программ и курсов, позволяющих передавать студентам знания и навыки работы на современном техническом и научном уровне!