

ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ

Научно-технический журнал
№ 4/2014

Издается с 1999 года
Выходит четыре раза в год

ГЛАВНЫЙ РЕДАКТОР
д.т.н., чл. кор. РАН Ю.Б. ЗУБАРЕВ

ЗАМЕСТИТЕЛИ ГЛАВНОГО РЕДАКТОРА:

д.т.н., проф. В.В. ВИТЯЗЕВ,
д.т.н., проф. В.П. ДВОРКОВИЧ

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

д.т.н., проф. Ар.С. Аджемов, д.т.н., проф. Б.А. Алпатов,
д.т.н., проф. В.Г. Бартенов, д.т.н. Ю.И. Борисов,
д.т.н., проф. Ю.А. Брюханов, д.т.н., проф. В.В. Еремеев,
д.т.н. Г.В. Зайцев, Р.В. Зубарев, А.П. Кирличников,
д.т.н., проф. М.И. Кривошеев, д.т.н., акад. РАН Н.А. Кузнецов,
В.Г. Мистюков, д.т.н., проф. С.Л. Мишенков,
д.т.н., проф. А.А. Петровский, д.т.н., проф. Ю.Н. Прохоров,
д.т.н., проф. В.Г. Санников, к.т.н., проф. В.С. Сперанский,
д.т.н., проф. Ю.С. Шинаков

Адрес редакции:

г. Москва, ул. Авиамоторная, д. 8
Научный центр МТУСИ
Тел.: 8 903 221 79 79
E-mail: rntores@mail.ru
vityazev.v.v@rsreu.ru
http://www.dsps.ru

Издатель:

Российское научно-техническое общество
радиотехники, электроники и связи им. А.С. Попова
Компьютерная верстка: И.А. Благодарова
Дизайн: М.В. Аверин

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

Подписано в печать 05.12.14 г.
Формат 60x90/8.

Гарнитура «Arial». Печать офсетная.
Бумага офсетная. Печ.л. 6,5. Тираж 700 экз.

Заказ № 7870. 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 г.

Журнал «Цифровая обработка сигналов» включен
в Перечень ведущих рецензируемых научных изданий,
в которых по рекомендации Минобрнауки РФ, должны
быть опубликованы основные научные результаты
диссертаций на соискание ученой степени доктора
и кандидата наук.

УЧРЕДИТЕЛИ: ЗАО «Инструментальные системы» ФГУП «НИИ радио»

В НОМЕРЕ:

- Клионский Д.М., Каплун Д.И.,
Вознесенский А.С., Гульванский В.В.**
*Алгоритм взвешенного перекрывающегося сложения
для обработки векторных сигналов
в задачах радиомониторинга* 2
- Минин Л.А., Насер Нихад Махмуд,
Киселев Е.А., Кургалин С.Д.**
*Метод выделения спектральных компонент
в сигналах путем интерполяции
с помощью систем целочисленных сдвигов* 9
- Мясникова Н.В., Берестень М.П.**
*Разложение на эмпирические моды
на основе экстремальной фильтрации* 13
- Иртюга В.А., Сарана Д.В.**
*Оценка частотной характеристики радиоканала
в спектральном домене с учётом данных
для OFDM сигналов* 18
- Овинников А.А.**
*Синтез низкоплотностных кодов
на основе уравновешенных неполных блок-схем* 22
- Букашкин С.А., Оглоблин А.В.**
*Перспективы создания комплексов ДКМВ радиосвязи,
использующих технологии ММО
с пространственным и поляризационным разнесением* 28
- Попов Д.И.**
Адаптивное подавление пассивных помех 32
- Новиков А.И., Юкин С.А.**
*Оценка зоны неопределенности радиоотражений
в задаче совмещения изображений с картой местности* 38
- Шеремет И.А., Рукин А.П., Лебедев В.Д.**
*Обработка изображений с помощью целочисленных
ортогональных преобразующих матриц* 45
- Вознесенская Т.В., Котов М.А., Леднов Д.А.**
Гибридный детектор речи 54
- Белобродский В.А.**
*Повышение производительности генетического алгоритма,
конструирующего цифровые фильтры
с использованием технологии CUDA* 59
- Абоелазм М.А., Мельчаков В.Н.,
Решетняк С.А., Третьяков Г.Н.**
*Исследование цифровой модели
нелинейного активного фильтра первого порядка* 62

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

АЛГОРИТМ ВЗВЕШЕННОГО ПЕРЕКРЫВАЮЩЕГОСЯ СЛОЖЕНИЯ ДЛЯ ОБРАБОТКИ ВЕКТОРНЫХ СИГНАЛОВ В ЗАДАЧАХ РАДИОМОНИТОРИНГА

Клионский Д.М., к.т.н., доцент Санкт-Петербургского государственного электротехнического университета «ЛЭТИ», e-mail: klio2003@list.ru;

Каплун Д.И., к.т.н., доцент Санкт-Петербургского государственного электротехнического университета «ЛЭТИ», e-mail: miya_kapl@front.ru;

Вознесенский А.С., аспирант Санкт-Петербургского государственного электротехнического университета «ЛЭТИ», e-mail: a-voznensky@yandex.ru;

Гульванский В.В., аспирант Санкт-Петербургского государственного электротехнического университета «ЛЭТИ», e-mail: slava-a-a@mail.ru.

Ключевые слова: радиомониторинг, векторный сигнал, алгоритм WOLA, банк фильтров, полифазная реализация, технология CUDA, графический процессор.

Введение

Понятие «мониторинг» может быть определено как систематический или непрерывный сбор информации о параметрах исследуемого объекта или системы [1-3]. Мониторинг может включать в себя как наблюдение за состоянием основных параметров системы, так и поиск отклонений в значениях этих параметров. Такие отклонения могут свидетельствовать об аномальном или преданомальном режиме функционирования, требующем принятия соответствующих оперативных мер.

В настоящее время тенденции развития алгоритмов и программно-аппаратных средств мониторинга наиболее чётко проявляются в следующих областях:

– *радиомониторинг* (частотно-временная обработка радиоизлучений, шумоподавление, обнаружение, классификация, оценивание параметров сигналов и пр.);

– *анализ вибраций* (проведение виброизмерений, оценивание параметров вибраций во временной и частотной областях, поиск резонансных частот) двигателей, турбин, механических конструкций и пр.;

– *гидроакустика* (мониторинг морской обстановки, состояния подводных и надводных объектов и пр.);

– *геофизика* (мониторинг сейсмической и геомагнитной активностей).

Данная статья посвящена рассмотрению задач радиомониторинга.

Радиомониторинг направлен на обнаружение и отслеживание сигналов и определение их свойств при наблюдении за радиозфиром в режиме реального времени

Статья посвящена разработке алгоритма взвешенного перекрывающегося сложения (*weighted overlap-add-WOLA*) для обработки векторных сигналов при проведении радиомониторинга в широком частотном диапазоне (ШЧД). Алгоритм предназначен для функционирования в аппаратуре наблюдения за радиозфиром при осуществлении обработки в режиме реального времени. Выполнено сравнение разработанного алгоритма с полифазной реализацией многоканального банка цифровых фильтров. Рассмотрена программно-аппаратная реализация многоканальной обработки и показаны преимущества применения технологии *CUDA*¹, основанной на вычислениях с использованием графических процессоров.

[1-3]. В настоящее время данная задача является актуальной как в гражданских, так и специальных приложениях. Радиомониторинг довольно часто осуществляется в ШЧД [1-3] величиной до нескольких МГц или ГГц. В таком частотном диапазоне может работать несколько десятков или сотен независимых источников радиоизлучения. Радиомониторинг в ШЧД включает в себя ряд процедур, проиллюстрированных на рис. 1.

Для решения задач радиомониторинга, как правило, используются специализированные аппаратно-программные комплексы, состоящие из аппаратуры приёма и обработки сигналов. Такие комплексы должны иметь возможность работать в различных режимах, в том числе автоматических, и одновременно решать целый ряд задач, связанных с наблюдением и контролем за радиозфиром в реальном масштабе времени. Отсюда возникает необходимость наличия технических средств для параллельной обработки, что подразумевает применение алгоритмов обработки векторных сигналов.

В рамках данной статьи рассматриваются многоканальные системы, в основу которых положена операция *цифровой фильтрации сигналов*, позволяющая определять и анализировать компоненты сигнала в частотных полосах. Операция цифровой фильтрации для выделения сигналов в конкретных частотных полосах может быть реализована путем применения *банка (ансамбля) цифровых фильтров* [4, 5]. Актуальность рассмотрения многоканальных систем обусловлена наличием в радиозфире многочисленных источников излучения с различными

¹ англ. – Compute Unified Device Architecture

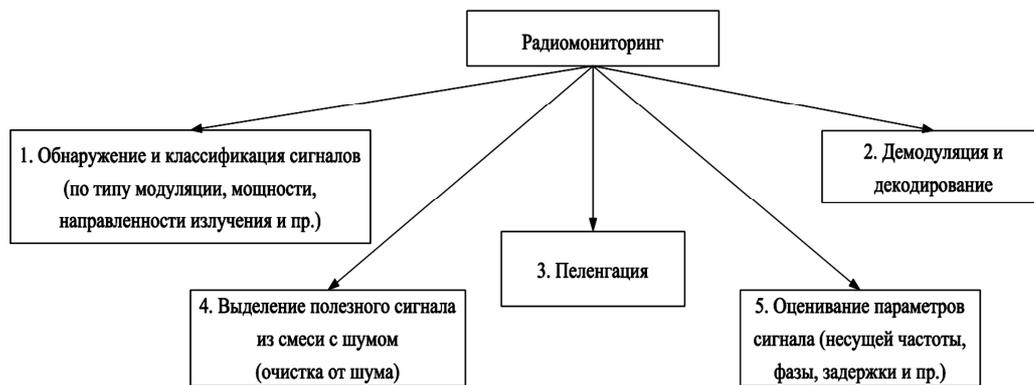


Рис. 1. Основные процедуры радиомониторинга

параметрами, что приводит к необходимости обработки векторных сигналов и проектирования многоканальных банков фильтров. Данные системы активно используются в аппаратуре радиомониторинга для решения задач радиопеленгации и обнаружения сигналов с последующей их классификацией и оцениванием параметров.

В современных системах связи находят все более широкое применение широкополосные сигналы и сигнально-кодовые конструкции, позволяющие организовывать надежные каналы радиосвязи при минимальных мощностях излучения. По этой причине особенно актуальными становятся задачи обнаружения, пеленгования и перехвата сигналов при различных отношениях сигнал/шум. К многоканальным банкам фильтров, разрабатываемым для задач радиомониторинга, предъявляется ряд требований, обусловленных необходимой точностью выделения канальных сигналов, а также программно-аппаратными затратами.

При синтезе ФНЧ-прототипа, параметры которого будут определять параметры канала банка фильтров, необходимо обеспечить малую (не более 1 дБ) неравномерность АЧХ в полосе пропускания, большое (не менее 80 дБ) подавление в полосе задерживания (другими словами, фильтр-прототип должен иметь высокую избирательность) и высокий коэффициент прямоугольности АЧХ.

Поскольку ШЧД может достигать нескольких МГц или ГГц, для организации одновременной обработки всего исследуемого частотного диапазона необходимо иметь радиоприёмные устройства с соответствующими полосами. Возможности современных радиоприёмных устройств ограничены применяемой аппаратной базой, поэтому для решения задач радиомониторинга используют комплексы, состоящие из нескольких широкополосных радиоприёмных устройств, на выходе которых формируются цифровые потоки с высокой частотой дискретизации для последующей обработки. Непосредственная обработка таких потоков крайне затруднена в силу значительных аппаратных затрат, поэтому такие потоки, как правило, проходят через несколько каскадов фильтров-дециматоров для понижения частоты (до 100-500 кГц), а затем при помощи банка фильтров разбиваются на параллельные каналы, ширина которых зависит от конкретной задачи. В частности, при радиомониторинге для конечной обработки используют канал с шириной приблизительно 3.1 кГц, т.к. это ширина спектра стандартного те-

лефонного канала и большинство сигналов, присутствующих в эфире, удовлетворяют этому требованию.

Таким образом, можно сформулировать требования к синтезируемому многоканальному банку фильтров и ФНЧ-прототипу для применения в системе радиомониторинга:

- границы рассматриваемого ШЧД: $[0; 100]$ кГц – для обнаружения, классификации и оценивания параметров широкополосных сигналов в радиоэфире;
- число каналов банка фильтров для каждого одномерного (одноканального) сигнала: $K = 32$;
- ширина полосы канала: $\Delta f = 3125$ Гц;
- подавление сигнала на границе полосы пропускания ФНЧ-прототипа: 1 дБ;
- подавление сигнала на границе полосы задерживания ФНЧ-прототипа: 120 дБ;
- неравномерность АЧХ ФНЧ-прототипа в полосе пропускания: 0.01 дБ;
- коэффициент прямоугольности АЧХ ФНЧ-прототипа: 1.25;

В силу проведения обработки в режиме реального времени, необходимо выполнять реализацию алгоритмов на высокопроизводительной аппаратно-программной базе, а для повышения эффективности реализации использовать вычислители с параллельной структурой. Такие вычислители могут быть реализованы на программируемых логических интегральных схемах (ПЛИС), а также устройствах обработки с технологией CUDA (программно-аппаратная архитектура для выполнения параллельных вычислений с использованием графических процессоров).

При вычислении сигналов на выходах каналов допустимо использование различных реализаций ДПФ-модулированного банка фильтров, включая реализацию с полной модуляцией, полифазную реализацию, а также реализацию на основе блочной обработки сигнала, составляющей суть алгоритма WOLA.

Многоканальный ДПФ-модулированный банк фильтров и алгоритм WOLA для обработки векторных сигналов

Теория построения банков цифровых фильтров развивается на протяжении более 35 лет, и за это время в данной области получены важные научно-практические результаты. Полный и систематизированный обзор и анализ основных этапов развития методов многоскорост-



стной обработки сигналов (МОС) и методов синтеза банков цифровых фильтров представлен в [6-8], где также приводится информация о первых работах в области методов МОС и методов построения банков цифровых фильтров, в частности, [9-11]. В работе [12] впервые предложено полифазное представление цифровых фильтров, используемое в рамках данной статьи. Кроме того, также уделялось внимание вопросам аппаратно-программного проектирования многоскоростных систем, в частности, с использованием цифровых сигнальных процессоров [8].

Вопросам многоканальной (векторной) фильтрации сигналов также уделялось внимание в ряде работ [13-14]. В частности, разработаны модели векторных цифровых фильтров во временной области, в том числе с учетом эффектов конечной разрядности, методы расчета разрядности коэффициентов и операционных устройств вещественных векторных цифровых фильтров, алгоритмы моделирования векторных цифровых фильтров каскадной структуры и т.д.

Алгоритм WOLA для обработки векторных сигналов может рассматриваться как обобщение соответствующего алгоритма для обработки одномерных сигналов [4], при этом предложенная в статье модификация позволит вычислительно эффективно обработать векторный входной сигнал, а также осуществить программно-аппаратную реализацию на основе высокопроизводительной аппаратно-программной базы.

По результатам вычисления сигналов на выходах каналов осуществляется их *субполосная обработка* (каждому каналному сигналу соответствует своя частотная полоса), включающая спектральный анализ, частотно-временной анализ, статистический анализ во временной области, демодуляцию и пр.

Определим входной векторный сигнал $x(n) = [x_0(n) \dots x_i(n) \dots x_{S-1}(n)]^T$ в виде набора одномерных сигналов $x_i(n)$, $i = 0, \dots, S-1$; $n = 0, \dots, N-1$; при этом векторный сигнал $x(n)$ может быть представлен прямоугольной матрицей размерностью $S \times N$, где N - длина (число отсчетов) сигнала $x_i(n)$; S - количество одномерных сигналов.

Структурная схема многоканального ДПФ-модулированного банка фильтров² (далее – многоканального банка фильтров) на основе полифазного представления показана на рис. 2.

В данной схеме использованы следующие обозначения:

- S – размерность векторного сигнала $x(n)$ на входе банка фильтров (число одномерных сигналов);
- M – коэффициент децимации $M = 1 \dots K$, где K – число каналов банка фильтров (для каждой полифазной реализации, соответствующей каждому одномерному сигналу $x_i(n)$);
- $g_i(m)$ и $e_i(m)$ - импульсная характеристика (ИХ)

i -го полифазного фильтра в прямой и обратной ветвях соответственно.

В случае *критической децимации*, когда $M = K$, вычисление выходных сигналов полифазного банка фильтров эквивалентно применению алгоритма WOLA, но при этом алгоритм WOLA ориентирован на *лоблочный* анализ.

На вход алгоритма WOLA подается векторный сигнал $x(n) = [x_0(n) \ x_i(n) \ \dots \ x_{S-1}(n)]$, $n = 0, \dots, N-1$. Кроме того, должна быть известна ИХ $h(n)$ ФНЧ-прототипа (вектор-строка размерностью $1 \times N_h$), используемая в качестве окна анализа для взвешивания сигнала. Для применения эффективных вычислительных алгоритмов на основе БПФ необходимо, чтобы длина N была целой степенью числа 2, т.е. $N = 2^u$, где u - натуральное число. Если данное условие не выполнено, все одномерные сигналы $x_i(n)$, $i = 0, \dots, S-1$ дополняются требуемым количеством нулей. ФНЧ-прототип определяет характеристики всего многоканального банка фильтров.

Взвешивание векторного сигнала $x(n)$ окном может быть записано как

$$\tilde{x}_{mi}(n) = h(mM - n)x_i(n), \quad i = 0, \dots, S-1; \quad n = 0, \dots, N-1, \quad (1)$$

где m – номер блока длины N_h . Общее число блоков P длины N_h с перекрытием $N_h - M$ определяется как

$$P = 1 + \left\lfloor \frac{N - N_h}{M} \right\rfloor, \quad \text{где символ } \lfloor \cdot \rfloor \text{ начает округление до}$$

ближайшего целого в сторону уменьшения. В результате выполнения (1) формируется вектор $\tilde{x}(n)$ взвешенных отсчетов сигнала

$$\tilde{x}(n) = [\tilde{x}_0(n) \ \tilde{x}_1(n) \ \dots \ \tilde{x}_{S-1}(n)], \quad (2)$$

где $n = 0, \dots, N-1$. Далее необходимо выполнить разбиение блоков длины N_h взвешенного векторного сигнала на неперекрывающиеся сегменты длины K и суммирование данных сегментов. Общее количество сегментов Q длины K в пределах блока длины N_h определяется как $Q = \lceil N_h / K \rceil$, где символ $\lceil \cdot \rceil$ означает округление до ближайшего целого в сторону увеличения. В результате получаем:

$$z_{0,m}(r) = \sum_{l=-\infty}^{\infty} \tilde{x}_{0,m}(r + lK); \quad z_{1,m}(r) = \sum_{l=-\infty}^{\infty} \tilde{x}_{1,m}(r + lK); \quad \dots, \\ z_{S-1,m}(r) = \sum_{l=-\infty}^{\infty} \tilde{x}_{S-1,m}(r + lK), \quad (3)$$

где $m = 0, \dots, P-1$, $r = 0, \dots, K-1$. На основе выражений (3) может быть введена матрица Z , состоящая из подматриц, каждая из которых имеет P строк, соответствующих блокам длины N_h каждого одномерного сигнала $x_i(n)$, $i = 0, \dots, S-1$. С учетом сказанного, матрица Z имеет размерность $(P \cdot S) \times K$:

²англ. – DFT-modulated filter bank

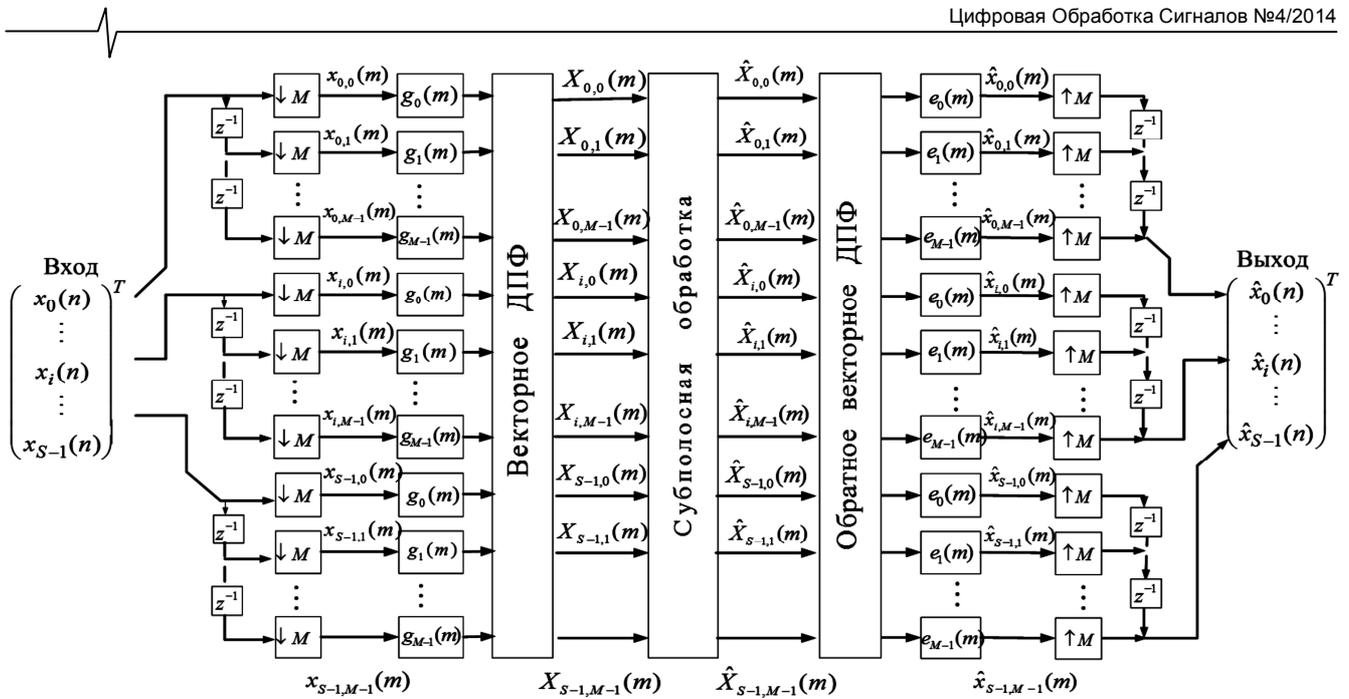


Рис. 2. Структурная схема многоканального ДПФ-модулированного банка фильтров (полифазная реализация)

$$Z = \begin{bmatrix} z_{00}(0) & z_{00}(1) & \dots & z_{00}(K-1) \\ \dots & \dots & \dots & \dots \\ z_{0,p-1}(0) & z_{0,p-1}(1) & \dots & z_{0,p-1}(K-1) \\ \dots & \dots & \dots & \dots \\ z_{S-1,0}(0) & z_{S-1,0}(1) & \dots & z_{S-1,0}(K-1) \\ \dots & \dots & \dots & \dots \\ z_{S-1,p-1}(0) & z_{S-1,p-1}(1) & \dots & z_{S-1,p-1}(K-1) \end{bmatrix} \quad (4)$$

После суммирования всех блоков каждого одномерного сигнала необходимо применить ДПФ к матрице Z . Данная операция может быть реализована на основе алгоритмов **векторного ДПФ**

$$Y = VDFT \{Z\}, \quad (5)$$

где Y - результирующая матрица ДПФ векторного сигнала, $VDFT$ - оператор векторного ДПФ.

В настоящее время в основном применяются два подхода к вычислению векторного ДПФ [15]. Первый, тривиальный, предполагает применение одномерного ДПФ к каждой строке матрицы Z (в общей сложности, потребуется вычислить $(P \cdot S)$ одномерных ДПФ размерности K). Второй подход базируется на вычислении одного одномерного ДПФ размерности $P \cdot S \cdot K$:

$$y = DFT \times \underbrace{\left\{ \left(z_{00}(0) z_{00}(1) \dots z_{00}(K-1) \right) \dots \left(z_{S-1,p-1}(0) z_{S-1,p-1}(1) \dots z_{S-1,p-1}(K-1) \right) \right\}}_{P \cdot S \cdot K}, \quad (6)$$

где DFT - оператор вычисления одномерного ДПФ. При вычислении одномерного ДПФ матрица Z «выпрямляется» и записывается в виде последовательности своих строк (после данного преобразования матрица Z превращается в вектор-строку).

В данной статье выбран подход на основе «выпрямления» матрицы по следующим причинам:

- алгоритм расчета векторного ДПФ напрямую влияет на суммарный объем и организацию вычислительной памяти, а также общее число коммутаторов параллельно-поточного БПФ-процессора. Сведение векторного

ДПФ к одномерному позволяет минимизировать общий объем вычислительной памяти и, следовательно, оптимизировать структуру ППБПФ-процессора;

- вычисление одномерного ДПФ размерности $P \cdot S \cdot K$ позволяет проще реализовать вычислительную процедуру в ППБПФ-процессоре, по сравнению с вычислением $(P \cdot S)$ различных ДПФ размерности K .

После вычисления ДПФ вектора y размерности $P \cdot S \cdot K$ осуществляется распределение его элементов по строкам матрицы. При этом размерность матрицы Y будет совпадать с размерностью матрицы Z . Матрица Y , как и матрица Z , будет состоять из подматриц

$$Y(r) = [Y_0(r) Y_1(r) \dots Y_{S-1}(r)]^T. \quad (7)$$

Каждая подматрица будет иметь вид

$$\begin{bmatrix} y_{i,0}(0) & y_{i,0}(1) & \dots & y_{i,0}(K-1) \\ \dots & \dots & \dots & \dots \\ y_{i,p-1}(0) & y_{i,p-1}(1) & \dots & y_{i,p-1}(K-1) \end{bmatrix}, i = 0, \dots, S-1. \quad (8)$$

Завершающим этапом является формирование матрицы $\tilde{Y}(r)$ путем умножения всех строк на поворачивающий множитель

$$\tilde{Y}_i(r) \Big|_{k\text{-я строка}} = Y_i(r) \Big|_{k\text{-я строка}} W_K^{-krM}, \quad r = 0, \dots, K-1; k = 0, \dots, P-1, \quad (9)$$

где поворачивающий множитель W_K^{-krM} определяется как $W_K^{-krM} = e^{-j[2\pi krM/K]}$.

Схема алгоритма WOLA для обработки векторных сигналов показана на рис. 3.

Таким образом, алгоритм WOLA для обработки векторных сигналов может быть записан в виде следующей последовательности шагов:

- 1) взвешивание векторного сигнала $x(n)$ с помощью окна для получения сигнала $\tilde{x}(n)$ в соответствии с (1);
- 2) разбиение взвешенного векторного сигнала (2) на неперекрывающиеся сегменты длины K и суммирова-

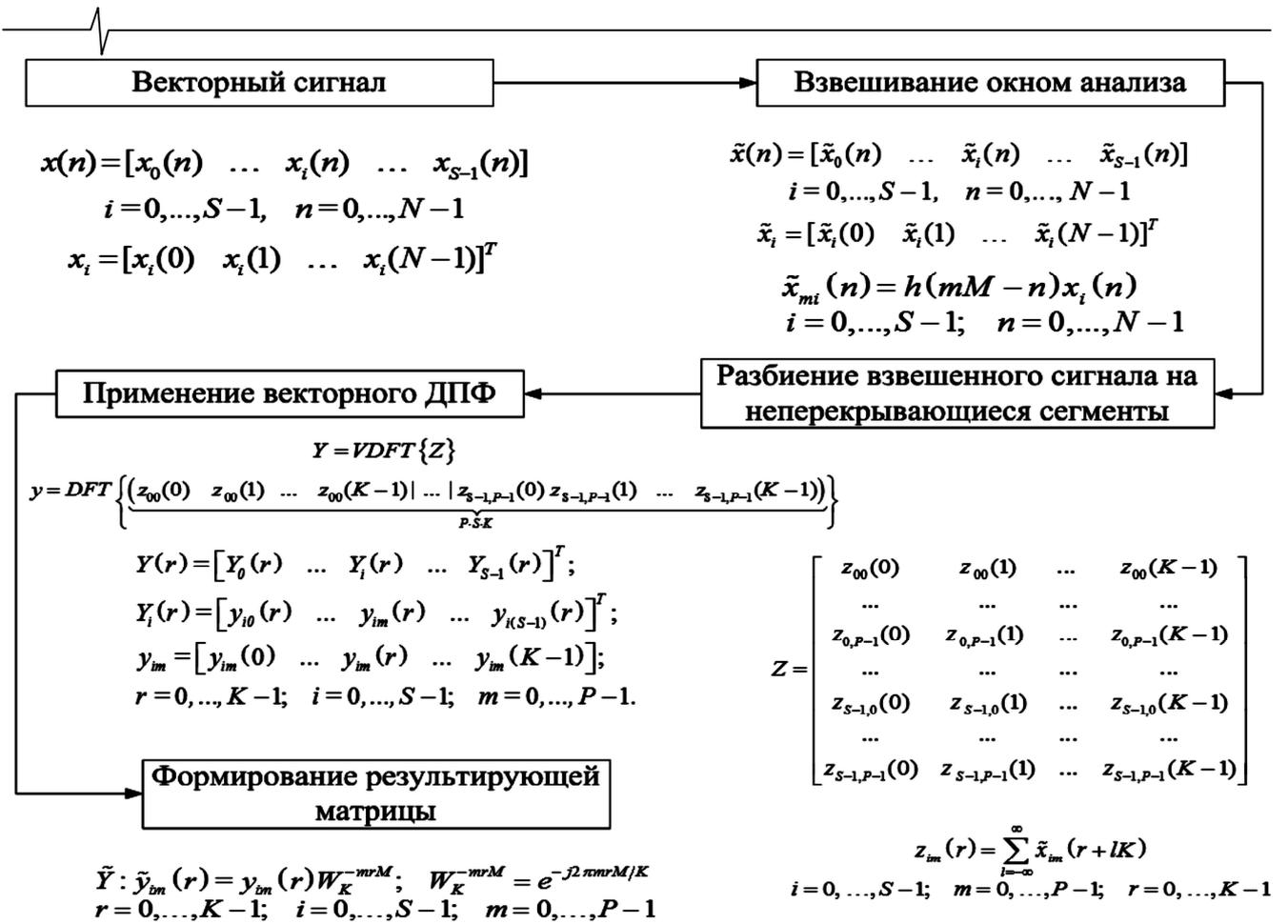


Рис. 3. Структурная схема алгоритма WOLA для обработки векторных сигналов

ние сегментов каждого одномерного сигнала в соответствии с (3);

3) применение векторного ДПФ (5) к матрице Z с помощью одномерного ДПФ (6) размерности $P \cdot S \cdot K$ и получение матрицы Y в виде (7) и (8);

4) формирование результирующей матрицы \tilde{Y} в соответствии с (9).

Алгоритм WOLA-синтез является дуальным по отношению к алгоритму WOLA-анализ, представленному в данной статье.

Оценивание вычислительной сложности алгоритма WOLA и алгоритмов-аналогов для обработки векторных сигналов

Сравним с точки зрения вычислительной сложности алгоритм WOLA для обработки векторных сигналов и многоканальный полифазный банк фильтров. При сравнении используем число операций на один отсчет входного сигнала и будем учитывать *только операции умножения*, реализация которых является более сложной, чем операций сложения.

Полифазная структура для каждого одномерного сигнала представляет собой M параллельных ветвей, и при поступлении одного отсчета в каждую ветвь требуется $\left(\frac{N_h}{M} \cdot 1\right)$ операций умножения (за счет линейной свертки с импульсной характеристикой полифазного фильтра). Для всех M ветвей число операций D_{PF} приближенно определяется как

$$D_{PF} \approx \frac{N_h}{M} \cdot M = N_h. \quad (10)$$

Для многоканального полифазного банка фильтров, предназначенного для обработки S одномерных сигналов, требуется приближенно $N_h \cdot S$ операций умножения.

При поступлении отсчетов на блок вычисления векторного ДПФ (рис. 2) необходимо $SM \log M$ операций при сведении векторного ДПФ к одномерному и применении алгоритмов БПФ.

Таким образом, общее число операций D_{PF_mult} для многоканального полифазного банка фильтров определяется как

$$D_{PF_mult} \approx N_h S + SM \log M = S(N_h + M \log M). \quad (11)$$

$$\text{Для случая критической децимации } M = K \text{ получаем } D_{PF_mult} \approx S(N_h + K \log K). \quad (12)$$

При реализации банка фильтров с использованием алгоритма WOLA рассмотрим вначале число операций для одного блока длины N_h . Взвешивание окном требует N_h операций умножения. При выполнении ДПФ размерности K для блока длины K сложность составляет $O(K \log K)$. Таким образом, общее число операций D_{WOLA} приближенно определяется как

$$D_{WOLA} \approx S(N_h + K \log K). \quad (13)$$

Если осуществлять вычисление векторного ДПФ, необ-

ходимо вначале выполнить N_h операций умножения для взвешивания каждого одномерного сигнала. Далее применяется векторное ДПФ к «выпрямленному» сигналу длины $P \cdot S \cdot K$ и осуществляется остановка алгоритма после выполнения $\log K$ операций. Число операций на отсчет совпадает с результатом, определяемым выражением (13).

На основе данных рассуждений можно сделать вывод о том, что многоканальный полифазный банк фильтров и алгоритм WOLA для обработки векторных сигналов являются эквивалентными по вычислительной сложности при критической децимации ($M = K$).

Вместе с тем, необходимо отметить, что WOLA может применяться при произвольном соотношении между M и K , что является несомненным преимуществом данного алгоритма на практике, в то время как полифазные банки фильтров используются в основном при критической децимации $M = K$. Таким образом, область применения алгоритма WOLA шире по сравнению с полифазными банками фильтров.

Программно-аппаратная реализация алгоритма WOLA для обработки векторных сигналов

Определяющим фактором вычислительной сложности при программно-аппаратной реализации многоканального банка фильтров является ФНЧ-прототип (КИХ-фильтр), формирующий АЧХ желаемой формы для одного канала банка фильтров. Порядок ФНЧ-прототипа определяется исходя из параметров, задаваемых при разработке системы радиомониторинга в ШЧД.

В эксперименте синтез ФНЧ-прототипа осуществлялся с параметрами, приведенными в начале статьи. Фильтр был синтезирован методом окон (окно Кайзера). Банк фильтров предназначен для обработки *двухканальных* сигналов. По результатам синтеза был получен ФНЧ-прототип порядка 6245.

Для повышения эффективности программно-аппаратной реализации в статье использовались вычислители с параллельной структурой на основе технологии CUDA[16]. Суть данной технологии заключается в использовании набора параллельно работающих графических процессоров (GraphicsProcessingUnit – GPU) для решения неграфических задач. GPU – специализированное вычислительное устройство, которое:

- является сопроцессором к центральному процессору (CPU);
- обладает собственной памятью;
- дает возможность параллельно выполнять большое количество отдельных потоков данных (под «поток» понимаются параллельно выполняемые части одной программы).

К преимуществам технологии CUDA относятся кросс-платформенность, наличие набора готовых библиотек, использование расширенной версии языка «Си» (англ. - C) с дополнительными средствами для параллельного программирования и создания многопоточных приложений на CUDA для написания программ. Кроме того, не применяется графический интерфейс программирования приложений (API), имеющий ряд ограничений для эффективной организации многопоточных вычислений.

Эксперимент был проведен на персональном компьютере со следующими характеристиками: процессор IntelCorei7-3630QM (IvyBridge) 2.4 ГГц; оперативная память DDR3 16 Гб; операционная система Windows 7 64 бит; видеокарта NVidia GeForce GT650M (384 ядра GPU, частота ядра 850 МГц, память видеокарты 2 Гб).

В табл. 1 приведены оценки времени выполнения алгоритма WOLA с использованием технологии CUDA, а также CPU. При использовании технологии CUDA также учитывалось время на пересылку данных между оперативной памятью персонального компьютера и памятью видеокарты.

Анализ результатов, приведенных в табл. 1, позволяет сделать следующие выводы:

- применение технологии CUDA позволяет добиться существенного сокращения времени вычислений, по сравнению с использованием CPU;
- увеличение объема данных приводит к росту выигрыша во временных затратах и, следовательно, чем больше размер сигнальной выборки, тем целесообразнее применение технологии CUDA.

Таблица 1. Временные затраты на обработку сигналов различной длины с использованием CPU и CUDA

Размер сигнала (отсчеты/Мбайт)	CUDA	CPU
3000000/ 11 Мб	0.586 с	92.782 с
8000000 / 30 Мб	0.640 с	245.312 с
16000000/ 61 Мб	0.692 с	493.668 с

Необходимо отметить, что время на пересылку данных между оперативной памятью и памятью видеокарты может в ряде случаев достигать половины от общего времени обработки, что является определенным недостатком технологии CUDA.

Заключение

В статье представлен разработанный алгоритм WOLA для обработки векторных сигналов в режиме реального времени с использованием векторного ДПФ. Данный алгоритм может быть применен для пеленгации, обнаружения сигналов в радиоэфире, их классификации и последующего оценивания характеристик во временной и частотной областях. Выполнено сравнение алгоритма с многоканальным полифазным банком фильтров и отмечено, что алгоритм WOLA имеет более широкую область применения за счет произвольного соотношения между числом каналов банка фильтров и коэффициентом децимации сигнала в каналах. Показано, что при программно-аппаратной реализации многоканальной обработки с использованием технологии CUDA время обработки может быть значительно сокращено, по сравнению с обработкой на CPU.

Данная статья и проведенная научная работа поддержаны в рамках проекта «Мой первый грант» (соглашение №14-07-31250/14), а также работы в СПбГЭТУ по проекту Министерством образования и науки Российской Федерации в рамках договора № 02.G25.31.0058 от 12.02.2013.

Литература

1. Рембовский А.М., Ашихмин А.В., Козьмин В.А. Радиомониторинг: задачи, методы, средства // 2-е изд., перераб. и доп., М.: изд. «Горячая линия - Телеком», 2010.
2. W. Hirt Ultra-Wideband Radio Technology: Overview and Future Research, *Comp. Commun.*, vol. 26, no. 1, Jan. 2003, pp. 46–52.
3. Каплун Д.И., Клионский Д.М., Олейник А.Л., Вознесенский А.С., Жукова Н.А., Гульванский В.В., Петровский А.А. Применение полифазных банков фильтров в задачах мониторинга широкого частотного диапазона // *Известия вузов России. Радиоэлектроника*. 2013. Вып. 3. С. 38-43.
4. R.E. Crochiere, L.R. Rabiner Multirate digital signal processing // Prentice Hall, 1983.
5. P.P. Vaidyanathan Multirate Systems and Filter Banks // Prentice Hall. EnglewoodCliffs.- NJ, 1993.
6. Витязев В.В., Витязев С.В., Зайцев А.А. Многокросстная обработка сигналов: ретроспектива и современное состояние (часть 1) // *Цифровая обработка сигналов*, № 1, 2008, С. 12-21.
7. Зайцев А.А. Методы построения банков цифровых фильтров: тематический обзор // *Цифровая обработка сигналов*. 2003. № 1. С. 2-10.
8. Витязев В.В. Цифровая частотная селекция сигналов. М.: Радио и связь, 1993. 240 с.
9. Bellanger M.G., Daguet J.L., Hecagnol G.P. Interpolation, extrapolation and reduction of computation speed in digital filter // *IEEE Trans. Acoust., Speech and Signal Processing*. V. ASSP-22. Aug., 1974. P. 231-235.
10. Mitra S.K. Digital Signal Processing: a computer-based approach. McGraw-Hill. Comp. Inc., 1998.
11. Rabiner L.R., Crochiere R.E. A novel implementation for narrowband FIR digital filters // *IEEE Trans. Acoust., Speech and Signal Processing*. V. ASSP-23. Oct., 1975. P. 457-464.
12. Bellanger M.G., Bonnerot G., Coudreuse M. Digital filtering polyphase network: Application to sample rate alteration and filter banks // *IEEE Trans. Acoust., Speech and Signal Processing*. V. ASSP-24. Apr., 1976. P. 109-114.
13. Гадзиковский В.И. Метод расчёта шумов квантования векторных цифровых фильтров // *Цифровая обработка сигналов*, 2005, № 4. С.24-28.
14. Гадзиковский В.И. Комплексные и векторные цифровые фильтры // *Сборник результатов научных исследований сотрудников РТФ УГТУ–УПИ «Новые методы передачи и обработки информации»*. — Екатеринбург: ГОУ ВПО «УГТУ–УПИ», 2003. С.141-155.
15. Петровский Ал.А., Станкевич А.В., Петровский А.А. Быстрое проектирование систем мультимедиа от прототипа // Минск, изд-во «Бестпринт», 2011, – 412 с.
16. J. Sanders, E. Kandrot CUDA by Example: An Intro-

duction to General-Purpose GPU Programming // Addison-Wesley Professional, 2010.

MULTICHANNEL WOLA ALGORITHM FOR PROCESSING VECTOR SIGNALS IN RADIO MONITORING TASKS

Klionskiy D.M., Kaplun D.I., Voznesenskiy A.S., Gulvanskiy V.V.

The present paper is devoted to the development of multichannel weight overlap add algorithm (WOLA algorithm) for processing vector signals in wideband radio monitoring tasks. The algorithm is intended for being integrated in equipment used for monitoring airwaves and signal processing in real time. The advantages of the algorithm are demonstrated in comparison with the direct form and polyphase form of filter bank implementations. The characteristics of the suggested algorithms and signal reconstruction accuracy versus signal-to-noise ratio (SNR) are studied in the paper. Advantages of hardware implementation of multichannel processing using the CUDA technology are shown.

Wide band monitoring is broadly applied in the following areas:

- radio monitoring (time-frequency processing of radiations, noise suppression, signal detection, signal classification, estimation of signal parameters, etc.);
- vibration analysis (vibration measurement, estimation of vibrational parameters in the time and frequency domains, resonance search);
- hydroacoustics (monitoring of marine environment, condition monitoring of surface and underwater objects, etc.);
- geophysics (monitoring of seismic and geomagnetic activity).

Radio monitoring is frequently performed in a wide frequency band, which can amount to several MHz or GHz. Several tens or hundreds of independent radiation sources can function in such a band simultaneously.

False alarm probability and correct detection probability are estimated for different SNR and it is shown in the paper that the best results are obtained for SNR exceeding -20 dB. Complexity estimation of the developed algorithm confirms gain in the total number of operations for large N in comparison with the direct implementation form.

The paper illustrates the fact that hardware implementations using CUDA make it possible to dramatically reduce (10 times and more) processing time in comparison with CPU applications. The more the size of data is the greater gain can be achieved. Experiments were done on a personal computer with the following characteristics: processor: Intel Core i7 3630QM (Ivy Bridge); RAM DDR3 8 Gbytes; OS Windows XP SP2 32 bits; videocard NVidia GeForce GT650M.

УДК 519.6:517+519.711.3

МЕТОД ВЫДЕЛЕНИЯ СПЕКТРАЛЬНЫХ КОМПОНЕНТ В СИГНАЛАХ ПУТЕМ ИНТЕРПОЛЯЦИИ С ПОМОЩЬЮ СИСТЕМ ЦЕЛОЧИСЛЕННЫХ СДВИГОВ

Минин Л.А., Воронежский государственный университет, e-mail: mininla@mail.ru;

Насер Нихад Махмуд, Университет Диала, г. Баакуба, e-mail: nihadalnasir1@yahoo.com;

Киселев Е.А., Воронежский государственный университет, e-mail: evg-kisel2006@yandex.ru;

Кургалин С.Д., Воронежский государственный университет, e-mail: kurgalin@bk.ru.

Ключевые слова: сигнал, спектр, интерполяция, алгоритм, целочисленный сдвиг, лоренцевский контур.

Введение

Проблема выделения простых составляющих (пиков) в сложных спектрах сигналов возникает в самых разных областях исследований: физико-химических, медицинских, радиотехнических и др. Ключевой здесь является задача о разделении двух пиков, близких по форме к гауссовым. Согласно классическому правилу Рэлея для пиков одинаковой амплитуды разделение возможно, если в анализируемой кривой между двумя максимумами возникает четко выраженный минимум, причем разность между максимумом и минимумом должна составлять не менее 20% от максимума. Рис. 1 иллюстрирует применение критерия Рэлея к распределению интенсивности I в интерферометре в зависимости от разности фаз δ двух лучей [1].

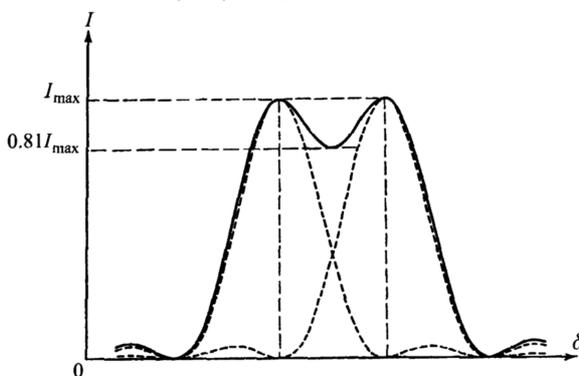


Рис. 1. Критерий Рэлея разделения двух пиков одинаковой амплитуды. I – интенсивность сигнала, δ – разность фаз двух лучей

Критерий Рэлея носит эмпирический характер и может в равной степени применяться и в задачах спектрального анализа. Математическое ограничение разрешающей способности обусловлено тем, что задача о разделении близких пиков, начиная с некоторого момента, становится некорректной. При наличии дополнительных предположений о характере спектра сигнала задача о разделении пиков снова может иметь устойчивое решение. За счет этого можно улучшить разрешающую способность алгоритмов анализа сигналов и их спектров. Решающую роль играет информация о форме спектральных линий. В зависимости от конкретной задачи здесь могут использоваться различные модели. Рас-

Разработан метод выделения отдельных компонент в сложных спектрах с помощью интерполяции по целочисленным сдвигам функции Лоренца, которая часто используется при моделировании различных процессов и явлений. Проведена оценка точности метода и границ его применимости.

смотрим модели формы спектральных линий в случае, когда изучаются атомные спектры. Спектральные линии даже у дискретных спектров испускания или поглощения не являются строго монохроматичными. Действие различных механизмов уширения линий спектра приводит к образованию некоторого спектрального распределения интенсивности вблизи частоты ω_0 квантового перехода в атоме или молекуле. Контур спектральной линии определяется механизмом уширения. При ударном или радиационном уширениях, в случаях, когда эффект Доплера мал, форма линий спектров описывается лоренцевским контуром [2, 3] с распределением интенсивности

$$g_L(\omega) = \frac{\Gamma}{2\pi} \frac{1}{(\omega - \omega_0 - \Delta)^2 + \Gamma^2/4}. \quad (1)$$

здесь Γ – параметр ширины спектральной линии, Δ – сопровождающий уширение сдвиг линии.

При доплеровском уширении возникает гауссов контур:

$$g_G(\omega) = \frac{1}{\Delta\omega_D \sqrt{\pi}} \exp\left(-\left(\frac{\omega - \omega_0}{\Delta\omega_D}\right)^2\right), \quad (2)$$

где $\Delta\omega_D = \omega_0 \frac{v_0}{c}$ – полуширина спектральной линии,

$$v_0 = \sqrt{\frac{2kT}{M}}$$

– наиболее вероятная скорость атомов, c – скорость света, M – масса атома, k – постоянная Больцмана, T – температура излучающего вещества.

Сложные спектры состоят из множества линий, которые могут накладываться друг на друга, затрудняя тем самым анализ. В общем случае задача формулируется как выделение отдельных компонент заданной формы в исследуемом сигнале. При анализе спектра ключевой является информация о расположении спектральных линий и величине их амплитуды. Кроме того, реальный спектр представляет собой оцифрованный сигнал и, как правило, задается на равномерной сетке точек. Следовательно, математически задача может быть сформулирована как разложение сигнала по системе сдвигов некоторой функции, вид которой определяется формой отдельной спектральной линии.



В настоящее время разработано множество методов аппроксимации сигналов при помощи сдвигов функции Гаусса (см., например, [4]). В то же время, вопрос о разложении сигнала по функциям Лоренца (1) в литературе практически не рассматривался. Поэтому целью данной работы является разработка соответствующего математического аппарата на основе интерполяции системами целочисленных сдвигов функции Лоренца, который будет иметь важное применение при анализе спектров.

Математическая постановка задачи

Требуется представить исследуемый спектр (спектральную плотность) $F(\omega)$ в виде линейной комбинации равномерных сдвигов $g(\omega - k\Delta\omega)$, $k \in Z$, порожденных заданной функцией $g(\omega)$. При расчетах удобно работать с системой целочисленных сдвигов, т.е. перейти от сетки с шагом $\Delta\omega$ к сетке с единичным шагом.

Сделаем замену переменной $x = \omega / \Delta\omega$ и введем новые функции $\phi(x) = g(\omega)$, $f(x) = F(\omega)$. Тогда

$$g(\omega - k\Delta\omega) = g(\Delta\omega(x - k)) = \phi(x - k). \quad (3)$$

Система функций $\phi(x - k)$ не является ортогональной. Поскольку исходная функция $F(\omega)$ в реальных задачах задана дискретно, то естественно использовать интерполяцию по узлам сетки отсчетов.

Задача интерполяции формулируется следующим образом. Требуется построить функцию

$$\tilde{f}(x) = \sum_{j=-\infty}^{\infty} p_j \cdot \phi(x - j), \quad (4)$$

совпадающую с $f(x)$ в узлах сетки, т.е. $\tilde{f}(m) = f(m)$, $m \in Z$. Обычно данная задача решается с помощью узловой функции, которая определяется следующим образом: функция

$$\tilde{\phi}(x) = \sum_{k=-\infty}^{\infty} d_k \cdot \phi(x - k), \quad (5)$$

называется узловой, если для нее выполнена система равенств $\tilde{\phi}(m) = \delta_{0m}$, $m \in Z$, где δ_{0m} – символ Кронекера. Тогда

$$\tilde{f}(x) = \sum_{n=-\infty}^{\infty} f(n) \cdot \tilde{\phi}(x - n). \quad (6)$$

Нахождение функции $\tilde{\phi}(x)$ равносильно решению линейной системы бесконечного числа уравнений с бесконечным же числом неизвестных d_k :

$$\sum_{k=-\infty}^{\infty} d_k \cdot \phi(m - k) = \delta_{0m}, \quad m \in Z. \quad (7)$$

Набор соотношений (7) представляет собой систему типа свертки. Для ее решения используются вспомогательные ряды Фурье [5]

$$D(t) = \sum_{k=-\infty}^{\infty} d_k \cdot e^{-ikt}, \quad \Phi(t) = \sum_{k=-\infty}^{\infty} \phi(k) \cdot e^{-ikt}, \quad (8)$$

с помощью которых система уравнений (7) записывается в эквивалентной форме в виде функционального равенства:

$$D(t) \cdot \Phi(t) = 1. \quad (9)$$

Следовательно, для нахождения коэффициентов d_k нужно разложить в ряд Фурье функцию $1 / \Phi(t)$.

В качестве функции $\phi(x)$, порождающей систему целочисленных сдвигов, выбираем лоренцевский контур (1). Выпишем саму функцию, используя более удобную по сравнению с (1) нормировку, и отвечающий ей ряд Фурье $\Phi(t)$:

$$\phi(x) = \frac{\sigma^2}{\sigma^2 + x^2}, \quad \Phi(t) = \sum_{k=-\infty}^{\infty} \frac{\sigma^2}{\sigma^2 + k^2} \cdot e^{-ikt}. \quad (10)$$

Этот ряд с помощью формулы суммирования Пуассона вычисляется аналитически, а для коэффициентов d_k получается следующее выражение:

$$d_k = \frac{(-1)^k \cdot \text{sh}(\sigma\pi)}{\sigma\pi^2} \int_0^{\pi} \frac{\cos(kt)}{\text{ch}(\sigma t)} dt. \quad (11)$$

Более подробно получение формулы (11) с полным обоснованием всех математических преобразований представлено в статье [6], а в данной работе мы остановимся на вычислительных особенностях интерполяции.

Результаты

В табл. 1 приведены значения нескольких рассчитанных нами коэффициентов d_k при различных параметрах σ (мы указываем три значащие цифры для удобства восприятия). Нахождение интеграла в выражении (11) осуществлялось численно с помощью квадратурных формул. Случай $k < 0$ отдельно не рассматривался, поскольку из формулы (11) очевидно, что $d_{-k} = d_k$.

Таблица 1. Значения коэффициентов узловой функции при разных параметрах σ

Параметр σ	d_0	d_1	d_{10}
0,5	1,08	-0,200	$-8,52 \cdot 10^{-4}$
1,0	1,74	-0,783	$-9,96 \cdot 10^{-4}$
1,5	3,89	-2,49	$-7,68 \cdot 10^{-4}$
2,0	10,6	-8,06	$7,30 \cdot 10^{-3}$
2,5	32,8	-27,3	0,122
3,0	110	-96,1	1,17
5,0	$2,11 \cdot 10^4$	$-2,01 \cdot 10^4$	$1,82 \cdot 10^3$
7,0	$5,77 \cdot 10^6$	$-5,63 \cdot 10^6$	$1,21 \cdot 10^6$

Как видно из табл. 1, с ростом σ возрастают и коэффициенты узловой функции. Значения d_k становятся большими по абсолютной величине и медленнее убывают при увеличении порядкового номера k по сравнению со случаем, когда $\sigma = 1$. По этой причине ряды для $\tilde{\phi}(x)$ и $\tilde{f}(x)$ сходятся медленно и для устойчивых вычислений параметров p_j формулы (4) требуется больше коэффициентов d_k , чем при $\sigma = 1$. Как результат, происходит потеря точности интерполяционной формулы (4), количественную оценку которой мы дадим ниже.

Из результатов статьи [6] и табл. 1 следует, что максимально допустимым значением параметра σ с точки зрения устойчивости интерполяции является $\sigma = 7$. Поясним, что это значит на практике. Дело в том, что построить с хорошей точностью интерполяционную фор-

мулу (4) означает, что мы будем способны разделить две лоренцевские спектральные линии с точностью до одного шага сетки отсчетов. С ростом ширины линий, мерой которой является параметр σ , они становятся практически неотличимыми друг от друга, а их общий контур выглядит как одна спектральная линия. Поэтому значение σ , равное 6-7, является естественной границей разрешающей способности, присущей самой задаче разделения спектральных линий.

На рис. 2 и 3 изображены две лоренцевские спектральные линии одинаковой амплитуды A , находящиеся на расстоянии одного шага сетки друг от друга при значении $\sigma = 1$ (рис. 2) и при $\sigma = 7$ (рис. 3).

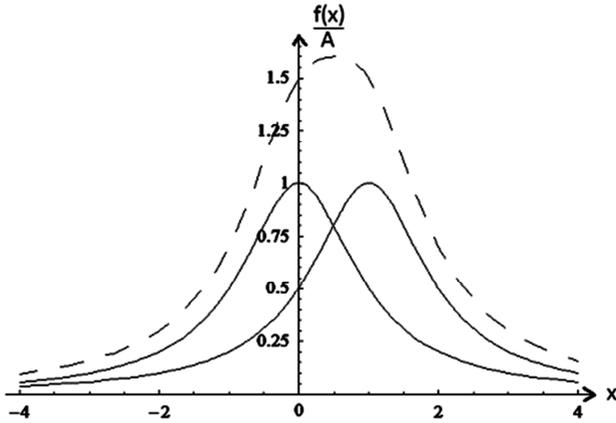


Рис. 2. Контурсы спектральных линий, находящихся на расстоянии одного шага сетки друг от друга, и их огибающая (пунктирная линия) при $\sigma = 1$.

A – амплитуда пиков,

$x = \omega / \Delta\omega$ – безразмерная переменная,

$f(x)/A$ – относительная спектральная плотность сигнала

При указанном значении $\sigma = 7$ для обеспечения приемлемой точности интерполяции необходимо учесть в формуле (5) порядка 200 коэффициентов d_k , поэтому требуется наличие 7-8 значащих цифр в экспериментальных данных. Как правило, эксперименты обеспечивают точность не более 3-4 значащих цифр, вследствие чего практической границей разрешающей способности является значение σ порядка 3-4.

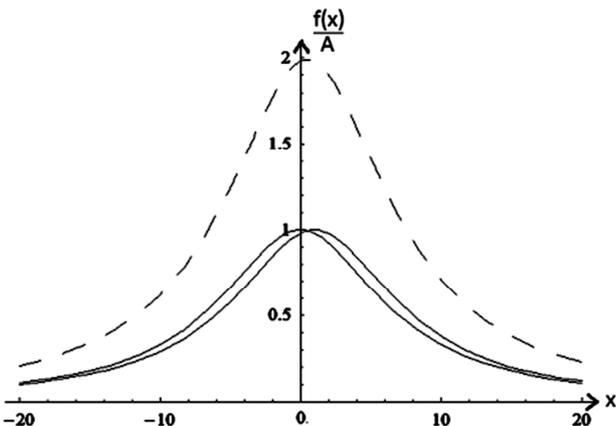


Рис. 3. Контурсы спектральных линий, находящихся на расстоянии одного шага сетки, и их огибающая (пунктирная линия) при $\sigma = 7$. A – амплитуда пиков,

$x = \omega / \Delta\omega$ – безразмерная переменная,

$f(x)/A$ – относительная спектральная плотность сигнала

Оценим, насколько эффективна процедура интерполяции при разделении спектральных линий в области математической устойчивости ($\sigma \leq 7$). Тестовый сигнал $f(x)$ задается в виде суперпозиции N спектральных линий с положениями m_k , амплитудами A_k , а также случайного шума $\varepsilon(x)$:

$$f(x) = \sum_{k=1}^N A_k \cdot \frac{\sigma^2}{\sigma^2 + (x - m_k)^2} + \varepsilon(x). \quad (12)$$

С помощью формул (4), (5) и (11) строится интерполирующая функция $\tilde{f}(x)$. Критериями эффективности работы алгоритма являются, во-первых, качество восстановления сигнала $f(x)$ с помощью $\tilde{f}(x)$, показателями которого являются визуальные отличия исходного и восстановленного сигнала на графике и соотношение максимальной по модулю величины погрешности и амплитуды полезного сигнала. Во-вторых, критерием качества восстановления сигнала будет являться величина соответствия коэффициентов интерполяции p_j амплитудам A_k .

Проведенные нами расчеты показали следующие результаты. При $\sigma \leq 7$ алгоритм дает хорошее качество восстановления сигнала: погрешность составляет величину порядка 0,1% от амплитуды сигнала. На рис. 4 показан пример восстановления тестового сигнала при $\sigma = 7$. Графики исходного и восстановленного сигнала изображены вместе (наложены друг на друга). Поскольку эти графики визуально практически не различаются, то в нижней части рисунка для наглядности изображена погрешность восстановления сигнала, увеличенная в 1000 раз.

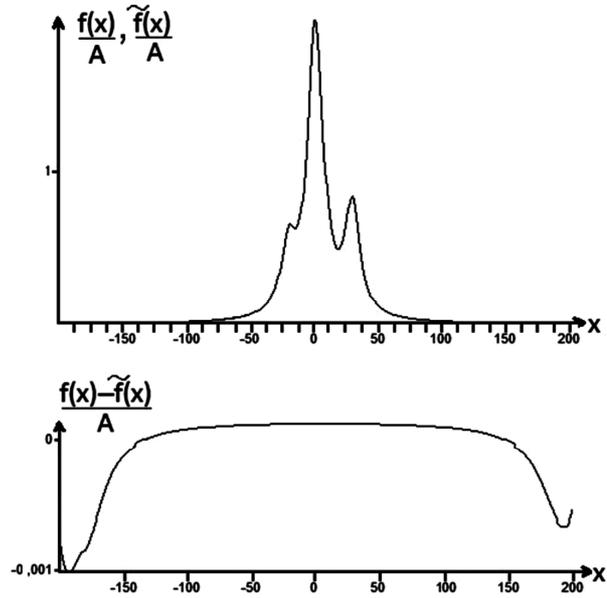


Рис. 4. Исходный сигнал $f(x)$,

интерполирующая функция $\tilde{f}(x)$ и их разность при $\sigma = 7$.

A – максимальная амплитуда отдельного пика в составе сигнала

Таким образом, при отсутствии шума метод достаточно надежно восстанавливает исходный сигнал и позволяет находить амплитуды его компонент с точностью 0,1% вплоть до значений $\sigma = 7$.

Иначе дело обстоит с зашумленными сигналами. Поскольку при больших значениях σ коэффициенты d_k убывают медленно, алгоритм выделения пиков в сигнале становится чувствительным к шуму как будет показано далее.

В табл. 2 приведены значения исходных и найденных с помощью интерполяции амплитуд A_k для двух случаев: а) при $\sigma = 7$ в отсутствие шума; б) при $\sigma = 2$ и наличии 5%-го относительно амплитуды полезного сигнала случайного шума. Наличие во втором случае шума приводит к понижению границы разрешающей способности до значений $\sigma \sim 2$. Интересно отметить, что сам сигнал восстанавливается с хорошей точностью при $\sigma \sim 5-6$ и в этом случае, хотя найденные амплитуды не соответствуют исходным (отличаются на несколько порядков). Отсюда следует, что интерполяция плохо подходит для анализа зашумленных сигналов, поскольку она опирается исключительно на значения функции в узлах, которые обычно на практике искажены шумом. Поэтому интерполяция воспроизводит и полезный сигнал, и шум, не позволяя тем самым получить информацию о компонентах полезного сигнала.

Таблица 2. Значения исходных амплитуд A_k и вычисленных с помощью интерполяции: а) при $\sigma = 7$ в отсутствие шума; б) при $\sigma = 2$ и наличии 5%-го случайного шума

Положение спектральной линии, m_k	Исходная амплитуда, A_k	Расчетная амплитуда в отсутствие шума ($\sigma=7$)	Расчетная амплитуда при наличии шума ($\sigma=2$)
-21	0	$-4,27 \cdot 10^{-4}$	0,121
-20	3	3,00	2,97
-19	0	$-4,48 \cdot 10^{-4}$	-0,221
-1	0	$-8,41 \cdot 10^{-4}$	$-3,01 \cdot 10^{-2}$
0	7	7,00	6,89
1	5	5,00	5,19
2	0	$-1,19 \cdot 10^{-3}$	-0,244
3	0	$-8,54 \cdot 10^{-4}$	0,310
4	0	$-1,18 \cdot 10^{-3}$	-0,324
5	2	2,00	2,21
6	0	$-1,18 \cdot 10^{-3}$	$-9,32 \cdot 10^{-2}$
29	0	$-4,99 \cdot 10^{-4}$	-0,183
30	5	5,00	5,09
31	0	$-5,33 \cdot 10^{-4}$	$-6,21 \cdot 10^{-2}$

Таким образом, перед применением интерполяционного алгоритма необходимы дополнительные меры для обеспечения устойчивости вычислений. Например, может применяться очистка сигнала от шума, или проведение специальной регуляризации задачи о выделении пиков. Одним из способов предварительной обработки сигнала может быть, например, его сглаживание с помощью всплесков с компактным носителем, описанное в работе [7]. В статье [7] было показано, что использование всплесков Добеши для целей сглаживания обеспечивает достаточно высокую степень очистки сигнала от шума, а пики в сигнале полностью сохраняют свое положение и амплитуду.

Заключение

В настоящей работе предложен метод выделения пиков лоренцевской формы в исследуемом сигнале путем его интерполяции целочисленными сдвигами. Рассчитаны коэффициенты узловой функции, и с их помощью проведены вычисления амплитуд составляющих компонент для тестовых сигналов. Показано, что интерполяционный метод в отсутствие шума дает хорошее качество восстановления исходного сигнала (погрешность порядка 0,1%) при значениях параметра $\sigma < 7$. При наличии 5%-го случайного шума, характерного для прикладных задач, границей применимости метода становятся значения $\sigma < 2$. Отсюда следует, что при работе с зашумленными сигналами необходимы дополнительные меры по обеспечению устойчивости вычислений. Одним из перспективных направлений исследований, по нашему мнению, является совместное применение интерполяционного метода и алгоритма очистки сигнала от шума с помощью всплесков Добеши [7].

Литература

1. Борн М. Принципы оптики / М. Борн, Э. Вольф. – М.: Наука, 1973. – 720 с.
2. Грим Г. Уширение спектральных линий в плазме. – М.: Мир, 1978. – 492 с.
3. Дробышев А.И. Основы атомного спектрального анализа. – СПб: Санкт-Петербургский университет, 1997. – 200 с.
4. Maz'ya V. Approximate approximations / V. Maz'ya, G. Schmidt // AMS Mathematical Surveys and Monographs. – 2007. – V. 141. – 350 p.
5. Новиков И.Я. Теория всплесков / И.Я. Новиков, В.Ю. Протасов, М. А. Скопина. – М.: Физматлит, 2006. – 616 с.
6. Kiselev E. A., Minin L. A., Novikov I. Ya., Sitnik S. M. On Evaluation of Riesz Constants for Systems of Shifted Gaussians, arXiv:1308.2649, 2013.
7. Сравнительный анализ эффективности сжатия сигнала ЭКГ с помощью всплесков Добеши и дискретного косинусного преобразования / С.Д. Кургалин, Л.А. Минин, Е.А. Киселев, Нихад Насер // Системы управления и информационные технологии. – 2011. – № 3.1(45). – С. 177-180.

METHOD OF EXTRACTION OF SPECTRAL COMPONENTS IN SIGNALS BY INTERPOLATION VIA SYSTEMS OF INTEGER SHIFTS

Minin L.A., Nihad Nasir, Kiselev E.A., Kurgalin S.D.

The algorithm of extraction of individual components in a complex spectrums using interpolation by integer shifts of Lorentz line shape that is often used in modeling of different processes and phenomena is developed. Estimation of the method's precision and its limits of application are carried out.

УДК 681.31

РАЗЛОЖЕНИЕ НА ЭМПИРИЧЕСКИЕ МОДЫ НА ОСНОВЕ ЭКСТРЕМАЛЬНОЙ ФИЛЬТРАЦИИ

*Мясникова Н.В., д.т.н., профессор, профессор кафедры «Автоматика и телемеханика»
ФГОУ ВПО «Пензенский государственный университет», e-mail: avitel@pnzgu.ru, genok123@mail.ru;
Берестень М.П., к.т.н., доцент кафедры «Автоматика и телемеханика»
ФГОУ ВПО «Пензенский государственный университет», e-mail: avitel@pnzgu.ru, beresten@sura.ru.*

Ключевые слова: экстремальная фильтрация, знакопеременные составляющие, разложение на эмпирические моды.

Введение

Метод декомпозиции на эмпирические моды (EMD) является одним из новых в области цифровой обработки сигналов. Этот подход обладает высокой степенью адаптации к исследуемым сигналам, что позволяет проводить точное оценивание реальных свойств процессов, в том числе и нестационарных. Ключевым моментом является использование «декомпозиции на эмпирические моды», с помощью которой любой сложный сигнал может быть разложен на конечное и часто довольно малое число «эмпирических мод», каждая из которых содержит определенную информацию об исследуемом процессе.

В настоящее время EMD-разложение не имеет такой стройной и глубоко изученной теоретической базы, которой наделен классический анализ Фурье или вейвлет-анализ. Алгоритм имеет эмпирическую природу [1-3].

Авторы уже неоднократно указывали на сходство EMD-разложения и экстремальной фильтрации, позволяющей разложить многоэкстремальный сигнал на знакопеременные составляющие [4-6]. Причем в последних статьях алгоритм экстремальной фильтрации излагали таким образом, чтобы подчеркнуть сходство с EMD-разложением [5], а затем и преобразованием Гильберта-Хуанга (ННТ) [7]. Мы хотим представить теоретическое обоснование метода экстремальной фильтрации, результатом которой является разложение на знакопеременные составляющие.

Первоначально метод экстремальной фильтрации (ЭФ) тоже позиционировался как эмпирический, т.е. не имеющий теоретического обоснования. Теоретического обоснования структуры фильтров и их коэффициентов не было. Авторами предпринимались попытки найти оптимальные структуры фильтров и значений их коэффициентов.

В дальнейшем авторам удалось теоретически обосновать метод разложения на знакопеременные составляющие и доказать, что это единственно возможные коэффициенты. Приведем это доказательство.

Приведено описание метода экстремальной фильтрации. Метод основан на последовательном выделении высокочастотных составляющих, локально определенных экстремумами сигнала. Показано сходство с разложением на эмпирические моды и преимуществу по сравнению с ним. Приведено теоретическое обоснование метода экстремальной фильтрации.

Теоретическое обоснование экстремальной фильтрации

Экстремальная фильтрация основана на параллельном выполнении двух процедур – низкочастотной и высокочастотной фильтрации:

– по экстремальным значениям $y_{эi}$ в моменты $t_{эi}, i = 1..K$ производится сглаживание оператором вида

$$y_{ci} = 0,25y_{эi-1} + 0,5y_{эi} + 0,25y_{эi+1}, \quad (1)$$

– выделяется знакопеременная составляющая

$$y_{pi} = -0,25y_{эi-1} + 0,5y_{эi} - 0,25y_{эi+1}. \quad (2)$$

На последующих итерациях уже из сглаженной составляющей y_{ci} , полученной на предыдущем шаге, извлекается следующая высокочастотная составляющая и производится ее сглаживание.

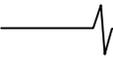
В основе метода лежит итерационная процедура выделения из сигнала составляющей с известной частотой, локально определяемой временными интервалами между экстремумами, что соответствует пропуску сигнала через адаптивный фильтр высоких частот (2), параллельно происходит и адаптивная низкочастотная фильтрация (сглаживание(1)).

Удаление известных составляющих из сигнала (временного ряда) описано в [8]. Если имеется последовательность $y_i, i = 0, 1, ..N$, то новая последовательность y_{ci} , уже не содержащая колебательной составляющей частоты ω_1 , может быть получена с помощью операции свертки, осуществляющей узкополосную фильтрацию исходной последовательности

$$y_{ci} = \sum_{k=0}^2 c_k \cdot y_{i-k},$$

где $c_0 = 1, c_1 = -2 \cdot \exp(-\alpha \cdot \Delta t) \cdot \cos(\omega_1 \cdot \Delta t), c_2 = \exp(-2\alpha \cdot \Delta t), \alpha$ – коэффициент затухания, Δt – шаг дискретизации.

Информацию о высокочастотной составляющей несут экстремумы процесса. Будем исходить из гипотезы, что в сигнале сложной формы содержится колебательная со-



ставляющая с частотой $f_1 = \frac{1}{2 \cdot \Delta}$, где Δ – минимальное расстояние между экстремумами сигнала. С учетом того, что мы выбрали только экстремумы, получим коэффициенты $c_0 = 1$, $c_1 = 2 \cdot \exp(-\alpha \cdot \Delta)$, $c_2 = \exp(-2\alpha \cdot \Delta)$. Можно выбрать небольшой участок процесса, на котором затухание процесса слабо выражено. Тогда получим $c_0 = 1$, $c_1 = 2$, $c_2 = 1$, и для удаления высокочастотной составляющей необходимо выполнить действие

$$\bar{y}_{ci} = y_{\bar{y}_{i-1}} + 2y_{\bar{y}_i} + y_{\bar{y}_{i+1}}.$$

С учетом нормирования получим сглаженную последовательность

$$y_{ci} = \frac{1}{4}y_{\bar{y}_{i-1}} + \frac{1}{2}y_{\bar{y}_i} + \frac{1}{4}y_{\bar{y}_{i+1}}.$$

А самая высокочастотная знакопеременная составляющая может быть выделена следующим образом

$$y_{pi} = y_{i-1} - y_{ci} = -\frac{1}{4}y_{\bar{y}_{i-1}} + \frac{1}{2}y_{\bar{y}_i} - \frac{1}{4}y_{\bar{y}_{i+1}}.$$

Удобней трактовать действия как осреднение i -ой точки по соседним, поэтому «сместим» отсчеты на единицу и будем пользоваться формулами (1) и (2).

Применение преобразования (1) соответствует пропуску данных через цифровой фильтр нижних частот с передаточной функцией, определяемой отношением спектров входного и выходного сигналов фильтра

$$G(f) = \frac{Y_c(f)}{Y_s(f)} = 0,25(e^{j2\pi f(t_{\bar{y}_{i-1}} - t_{\bar{y}_i})} + 2 + e^{j2\pi f(t_{\bar{y}_{i+1}} - t_{\bar{y}_i})}).$$

Этот фильтр исключает из сигнала самую высокочастотную составляющую. Положим для простоты, что экстремумы равноудалены друг от друга, тогда передаточная функция примет вид

$$G(f) = 0,25(e^{-j2\pi f \Delta} + 2 + e^{j2\pi f \Delta}) = 0,5(1 + \cos(2\pi f \Delta)).$$

Преобразование (2) реализует цифровой фильтр высоких частот с передаточной функцией

$$G(f) = \frac{Y_p(f)}{Y_s(f)} = 0,25(-e^{j2\pi f(t_{\bar{y}_{i-1}} - t_{\bar{y}_i})} + 2 - e^{j2\pi f(t_{\bar{y}_{i+1}} - t_{\bar{y}_i})}). \quad (3)$$

При тех же допущениях, что были сделаны для первого фильтра, получим выражение для передаточной функции

$$G(f) = 0,25(-e^{-j2\pi f \Delta} + 2 - e^{j2\pi f \Delta}) = 0,5(1 - \cos(2\pi f \Delta)) \quad (4)$$

Отметим, что формулы (3), (4) описывают передаточную функцию фильтра высоких частот. Однако это верно лишь по отношению к частотному диапазону

$$\left[0, \frac{1}{\Delta}\right] \quad (\Delta = \min(t_{\bar{y}_{i+1}} - t_{\bar{y}_i}, t_{\bar{y}_i} - t_{\bar{y}_{i-1}}), \text{ где } t_{\bar{y}_i} - \text{ абсцисса } i\text{-го экстремума), а т.к. при выделении экстремумов из временного ряда происходит их прореживание, а значит – и сужение частотного диапазона последовательности, то по отношению к частотному диапазону } \left[0, \frac{1}{\Delta t}\right] \text{ описываемый фильтр – адаптивный полосовой с центральной частотой } f_y = \frac{1}{2 \cdot \Delta}.$$

$$f_y = \frac{1}{2 \cdot \Delta}.$$

Адаптивность фильтра обеспечивается тем, что минимальное расстояние между экстремумами определяется лишь частотным составом сигнала: на первой итерации величина Δ определяется самой высокочастотной составляющей, на следующих итерациях – более низкочастотными составляющими, т.е. Δ увеличивается от итерации к итерации, а количество итераций определяется количеством мод.

Очевидно сходство процедур EMD и ЭФ:

- производится выделение экстремумов;
- при вычислении по формуле (1) по экстремумам оценивается среднее между огибающими, представленными экстремумами – соседние экстремумы всегда разноименные (минимумы и максимумы, принадлежащие, соответственно, нижней и верхней огибающей);
- вычисление по формуле (2) соответствует извлечению деталей (при этом в алгоритме не используются сложные алгоритмы построения огибающих с помощью сплайн-интерполяции);
- алгоритм последовательно применяют к сглаженной последовательности экстремумов y_{ci} .

Доказательство знакопеременности составляющих

К модам предъявляется два требования:

- общее число экстремумов равняется общему числу нулей с точностью до 1;
- полусумма значений верхней огибающей, интерполирующей локальные максимумы, и нижней огибающей, интерполирующей локальные минимумы, близка к нулю.

При фильтрации по экстремальным значениям первое требование вырождается в требование знакопеременности соседних экстремумов полученных последовательностей $y_{pi}^{(j)}$ (j – номер составляющей). То есть, соседние элементы последовательности всегда имеют разные знаки, поэтому составляющую можно легко отнести к определенной полосе частот. Докажем это положение.

Для доказательства рассмотрим составляющую сигнала, представленную последовательностью своих экстремумов. Признаком знакопеременности является выполнение условия

$$y_{pi}^{(j)} \cdot y_{pi+1}^{(j)} < 0 \quad (5)$$

для всех значений i .

Не нарушая общности доказательства, предположим, что $y_{\bar{y}_i}^{(j)}$ – максимум. Тогда $y_{\bar{y}_{i-1}}^{(j)}$, $y_{\bar{y}_{i+1}}^{(j)}$, $y_{\bar{y}_{i+2}}^{(j)}$ можно представить в виде соотношений:

$$\begin{aligned} y_{\bar{y}_{i-1}}^{(j)} &= y_{\bar{y}_i}^{(j)} - s_1, \\ y_{\bar{y}_{i+1}}^{(j)} &= y_{\bar{y}_i}^{(j)} - s_2, \\ y_{\bar{y}_{i+2}}^{(j)} &= y_{\bar{y}_{i+1}}^{(j)} + s_3, \end{aligned} \quad (6)$$

причем величины s_1, s_2, s_3 всегда положительны. Подставив (6) в (2) и проведя соответствующие преобразования, получим

$$\begin{aligned} y_{pi}^{(j)} &= \frac{1}{4}s_1 + \frac{1}{4}s_2, \\ y_{pi+1}^{(j)} &= -\frac{1}{4}s_2 - \frac{1}{4}s_3. \end{aligned}$$

Так как значения s_1, s_2, s_3 всегда положительны, то $y_{pi}^{(j)} > 0$, а $y_{pi+1}^{(j)} < 0$. Следовательно, неравенство (5) справедливо. Аналогично можно провести доказательство из предположения, что $y_{pi}^{(j)}$ – минимум.

Следовательно, в отличие от EMD-разложения, в котором выделяют «кандидата в моды» в методе экстремальной фильтрации всегда выделяется «мода».

Второе требование к модам обеспечивается тем фактом, что знакопеременная последовательность является результатом вычитания скользящего среднего из сигнала на первом этапе или из сглаженной составляющей на последующих этапах.

Так как основным авторским применением экстремальной фильтрации были задачи диагностики и идентификации, то дополнительное центрирование знакопеременных составляющих (как в алгоритме EMD) не применялось. Однако расширение сферы применения экстремальной фильтрации привело к необходимости совершенствования алгоритма экстремальной фильтрации. Необходимость центрирования обусловлена методической погрешностью при выделении знакопеременных составляющих – затухание при обосновании значений коэффициентов полагалось равным нулю.

Для каждой выделенной составляющей вычисляется скользящее среднее по формуле (1). Это среднее затем вычитается из составляющей $y_{pi}^{(j)}$, вычисленной по формуле (2). Результат совершенствования – уменьшение дисперсии экстремумов «чистого» тона, более точное выделение низкочастотных составляющих.

Очевидно что, выбирая или, наоборот, отбрасывая выделенные составляющие, можно легко реализовать традиционные виды фильтров (нижних частот, верхних частот, полосовой, заграждающий).

Подчеркнем, что предлагаемая фильтрация как нельзя лучше подходит к экспресс-анализу, т.к. коэффициенты фильтра $-0,5=2^{-1}$ и $0,25=2^{-2}$ допускают использование вместо умножения сдвига вправо на 1 или 2 разряда, соответственно, что более экономично с точки зрения вычислений. Этот же факт может стать решающим при использовании вычислительных средств малой мощности с ограниченным набором команд.

Представление результатов фильтрации

Форма представления результатов многократной адаптивной фильтрации зависит от решаемой задачи и определяется следующими факторами:

- необходимостью восстановления исходного сигнала и точностью этого восстановления;
- необходимостью обеспечить заданную степень сжатия исходного сигнала;
- необходимостью обеспечить заданные требования по точности и быстродействию вычисления характеристик исходного сигнала.

В зависимости от этого можно предложить две предельные формы представления результатов разложения сигнала с помощью адаптивной фильтрации: сжатая (фиксируются только экстремумы составляющих, а их привязка ко времени игнорируется) и полная (фиксируются абс-

циссы и ординаты экстремумов всех составляющих).

Распределение по шкале времени экстремумов в разных составляющих имеет важное свойство: самая высокочастотная составляющая содержит экстремумы, абсциссы которых определяют положение экстремумов всех более низкочастотных составляющих. Этот факт может быть использован для дополнительного сжатия данных без потерь информации об абсциссах экстремумов.

Полная форма представления результатов экстремальной фильтрации будет иметь вид:

$$\begin{pmatrix} t_{\varepsilon 1}^{(1)} & t_{\varepsilon 2}^{(1)} & t_{\varepsilon 3}^{(1)} & t_{\varepsilon 4}^{(1)} & t_{\varepsilon 5}^{(1)} & t_{\varepsilon 6}^{(1)} & t_{\varepsilon 7}^{(1)} & t_{\varepsilon 8}^{(1)} & \dots & t_{\varepsilon K_1}^{(1)} \\ t_{\varepsilon 1}^{(2)} & & t_{\varepsilon 2}^{(2)} & & & t_{\varepsilon 3}^{(2)} & & t_{\varepsilon 4}^{(2)} & \dots & t_{\varepsilon K_2}^{(2)} \\ \dots & \dots \\ t_{\varepsilon 1}^{(p)} & & & & & t_{\varepsilon 2}^{(p)} & & & \dots & t_{\varepsilon K_p}^{(p)} \end{pmatrix} \quad (7)$$

$$\begin{pmatrix} y_{\varepsilon 1}^{(1)} & y_{\varepsilon 2}^{(1)} & y_{\varepsilon 3}^{(1)} & y_{\varepsilon 4}^{(1)} & y_{\varepsilon 5}^{(1)} & y_{\varepsilon 6}^{(1)} & y_{\varepsilon 7}^{(1)} & y_{\varepsilon 8}^{(1)} & \dots & y_{\varepsilon K_1}^{(1)} \\ y_{\varepsilon 1}^{(2)} & & y_{\varepsilon 2}^{(2)} & & & y_{\varepsilon 3}^{(2)} & & y_{\varepsilon 4}^{(2)} & \dots & y_{\varepsilon K_2}^{(2)} \\ \dots & \dots \\ y_{\varepsilon 1}^{(p)} & & & & & y_{\varepsilon 2}^{(p)} & & & \dots & y_{\varepsilon K_p}^{(p)} \end{pmatrix} \quad (8)$$

В (7) и (8) представлены временные и амплитудные параметры составляющих. Каждая строка матрицы соответствует определенной составляющей. Подчеркнем, что в матрице (7) все значимые элементы в каждом столбце одинаковы.

Матрицы (7), (8) служат основой сжатия-восстановления сигналов, а матрица (8) с привязкой по времени (7) – это фактически время-частотная характеристика, т.к. каждая строка соответствует определенной частоте.

Матрица (7) содержит данные для оценивания частот $f_l = \frac{K_l}{2T}$ составляющих (T – длительность интервала анализа), а матрица (8) – для оценивания их амплитуд A или дисперсий σ^2 .

Параметры $\{f_l, \sigma_l^2\}$, $l = 1, \dots, p$ описывают модель быстропеременного процесса (БПП) суммой узкополосных составляющих. Модель БПП отражает свойства объекта: по изменению параметров составляющих можно судить об изменении состояния объекта, т.е. разложение – основа для экспресс-идентификации или экспресс-диагностики (экспресс-распознавания).

Применение предложенной модели БПП оправдано тем фактом, что в измерительных системах контролируются сложные технические объекты. В них, как правило, наблюдаются вибрации, акустические шумы, пульсации давления и т.д. Кроме того в системах действуют импульсные помехи, возбуждающие собственные колебания. Такие объекты описываются универсальной моделью – сумма узкополосных составляющих смешанных с шумом. Параметры модели отображают свойства исследуемого объекта, что позволяет использовать эти данные в системах диагностики, контроля и мониторинга параметров объекта.

Параметры разложения частота и амплитуда (мощность) каждой составляющей сигнала могут быть интерпретированы как текущий спектр. Изменение этих же параметров в скользящем окне может интерпретировано



как ВЧР, имеющее сходство с преобразованием Гильберта-Хуанга [7].

Свойства разложения на основе экстремальной фильтрации

Итак, мы получаем разложение, в котором моды представлены своими экстремумами. Если есть необходимость восстановить значения в дискретные моменты времени, то можно воспользоваться процедурой аппроксимации колокольными составляющими [7] или интерполировать «полукосинусами».

Адаптивный базис, основанный на выделении из данных эмпирических мод, удовлетворяет традиционным требованиям для базиса. Он является финитным, сходящимся, ортогональным и единственным.

Разложение на знакопеременные составляющие при использовании экстремальной фильтрации обладает следующими свойствами:

– всегда дает единственное разложение, вследствие жесткого детерминированного алгоритма;

– является сходящимся, так как алгоритм основан на выделении самой высокочастотной составляющей сначала из сигнала (процесса), а потом из сглаженной составляющей, а количество значимых составляющих в сигнале (процессе) всегда конечно;

– так как моды определены лишь в точках экстремумов и ортогональность можно проверить лишь на интерполированных данных, то следует говорить об ϵ -ортогональности.

Авторами предложен метод разложения на знакопеременные составляющие, имеющий сходство с разложением на эмпирические моды, при этом преимуществом метода является низкая трудоемкость, что позволяет использовать его в системах реального времени или в вычислителях малой мощности. Можно использовать сжатое представление данных – экстремумами процесса.

Применение результатов экстремальной фильтрации

Перечислим основные применения результатов фильтрации:

1. Экспресс-идентификация – определение количества составляющих, их типа (колебательные, инерционные), и параметров этих составляющих [10].
2. Экспресс-оценка спектра [4,5].
3. Интерпретация полученных результатов фильтрации как время-частотного распределения [7].
4. Адаптивная фильтрация (НЧ, ВЧ и т.д.) [4].
5. Использование параметров как диагностических признаков [10, 11].
6. Упрощение определения параметров сигнала

сложной формы $y_i = \sum_{j=1}^q A_j \cdot \exp(-\alpha_j \cdot t_i) \cdot \cos(2\pi f_j t_i + \phi_j)$

за счет предварительного разложения на моды [6].

Приведем пример разложения сигнала сложной формы широко известным методом декомпозиции на эмпирические моды (рис. 1) и с применением экстремальной фильтрации (рис. 2). Строго говоря, на втором рисунке мы получаем разложение только в точках эк-

тремумов, т.е. каждая из знакопеременных составляющих определена в точках своих экстремумов. Непрерывные кривые – результат интерполяции.

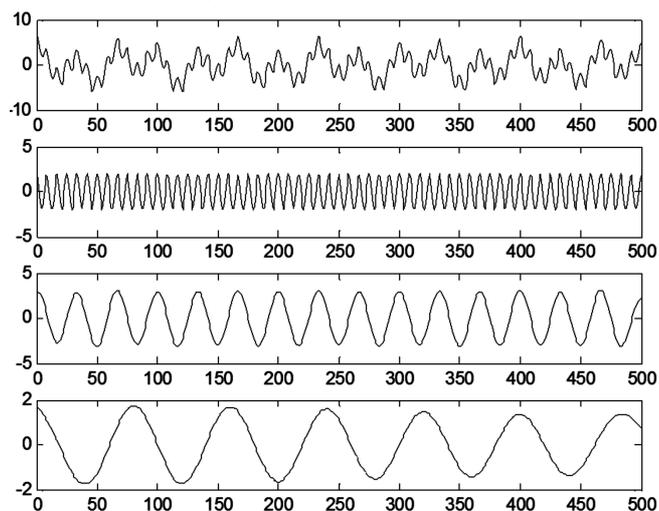


Рис. 1. Разложение сигнала на эмпирические моды

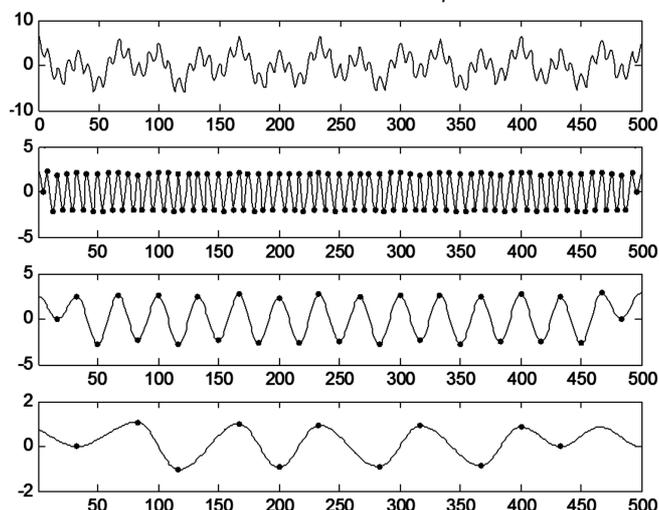


Рис. 2. Разложение на составляющие экстремальным фильтром

Из этих рисунков видно, что результаты практически одинаковы, а алгоритм разложения значительно проще и его вычислительная трудоемкость существенно меньше. Последние два фактора делают его эффективным инструментом экспресс-анализа в технических и информационных системах.

Заключение

Экстремальная фильтрация – не модификация известного алгоритма разложения на эмпирические моды, а независимое направление, развиваемое на протяжении ряда лет. Предлагаемый алгоритм осуществляет разложение на знакопеременные составляющие на основе адаптивной фильтрации сигнала по его экстремальным значениям. Такая фильтрация для краткости была названа экстремальной фильтрацией.

Литература

1. Клионский, Д.М. Декомпозиция на эмпирические моды в современной цифровой обработке сигналов /

Д.М. Клионский // 10-я Междунар. конф. и выставка «Цифровая обработка сигналов и ее применение». Москва. 26-28 марта 2008. Труды Росс. науч. техн. общества радиотехники, электроники и связи им. А.С. Попова. Сер. Цифровая обработка сигналов и ее применение. Вып. X-1. – М., 2008. – С. 188-190.

2. Клионский, Д.М. Декомпозиция на эмпирические моды и ее использование при анализе дробного броуновского движения/ Д.М. Клионский, Н.И. Орешко, В.В. Геппенер // Цифровая обработка сигналов. 2008. № 3. С. 37-45.

3. Клионский, Д.М. Декомпозиция на эмпирические моды с параболической интерполяцией огибающих в задачах очистки сигналов от шума/ Д.М. Клионский, Н.И. Орешко, В.В. Геппенер // Цифровая обработка сигналов. 2011. № 2. С. 51-60.

4. Мясникова, Н. В. Экстремальная фильтрация и ее приложения / Н. В. Мясникова, М. П. Берестень // Датчики и системы. – 2004. – № 4. – С. 8–11.

5. Мясникова, Н.В. Методы разложения сигналов на основе экстремальной фильтрации / Н.В. Мясникова, М.П. Берестень, Л.А. Долгих // Датчики и системы. – 2011. – № 2. – С. 8–12.

6. Мясникова, Н.В. Применение разложения по эмпирическим модам в задачах цифровой обработки сигналов / Н.В. Мясникова, Л.А. Долгих, М.Г. Мясникова // Датчики и системы. – 2011. – № 5. – С. 8-10.

7. Мясникова, Н.В. Время-частотное распределение на основе экстремальной фильтрации в цифровой обра-

ботке сигналов/ Н.В. Мясникова, М.П. Берестень // Датчики и системы. – 2013. – № 10. – С. 9-12.

8. Марпл.-мл., С.Л. Цифровой спектральный анализ и его приложения: Пер. с англ. – М.: Мир, 1990. – 584 с.

9. Мясникова, Н.В. Аппроксимация многоэкстремальных функций и ее приложения в технических системах/ Н.В. Мясникова, М.П. Берестень, М.П. Строганов// Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 2. – С. 113-119.

10. Мясникова, Н.В. Экспресс-анализ в технических и информационных системах / Н. В. Мясникова, М.П. Берестень // Инновационные информационные технологии. – 2013. – Т. 2. – № 2. – С. 328-333.

11. Зенов, А.Ю. Концепция организации обработки информации в системах диагностики и распознавания / А.Ю. Зенов, М.П. Берестень // Инженерный вестник Дона. – 2013. – № 1. – URL: <http://ivdon.ru/magazine/archive/n1y2013/1568> (дата обращения: 10.04.2013).

THEORETICAL FUNDAMENTALS OF EXTREME FILTERING

Myasnikova N.V., Beresten M.P.

Extremal filtration methods are considered. The method proposed is based on successive selection of the high-frequency components which are locally determined by signal's extrema. It is shown that this method is similar to Empirical Mode Decomposition and has some improvements compared with it. Theoretical basis of an extremal filtration method are given.

У в а ж а е м ы е к о л л е г и !

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

Журнал «Цифровая обработка сигналов» издается с 1999 года. Выходит ежеквартально, тиражом – 700 экз. Распространяется по подписке через агентство «Роспечать» в России (индекс 82185), СНГ и странах Балтии (индекс 20630), а также на Конференции: «Цифровая обработка сигналов и ее применение – DSPA'» и Выставках: «ExpoElectronica», «СвязьЭкспокомм», «ЭЛЕКТРОНИКА: компоненты, оборудование, технологии» (г. Москва) и др.

Научно-технический журнал «Цифровая обработка сигналов» включен в Список изданий, рекомендуемый ВАК РФ для публикации результатов научных исследований соискателями ученой степени доктора и кандидата технических наук в области радиотехники, связи, вычислительной техники, электроники, информационно-измерительных и управляющих систем.

Планируемые сроки издания отдельных номеров журнала:

№ 1 – март 2015 г. Тематический выпуск: «ЦОС в информационно-измерительных системах»

№ 2 – июнь 2015 г. Тематический выпуск: «Теория и методы цифровой обработки сигналов» (по материалам международной научно-технической конференции «Цифровая обработка сигналов и ее применение – DSPA' 2015»).

№ 3 – сентябрь 2015 г. Тематический выпуск: «Цифровая обработка изображений».

№ 4 – декабрь 2015 г. Тематический выпуск: «ЦОС в радиотехнике и системах телекоммуникаций».

Ориентировочная стоимость рекламных услуг:

4-я (внешняя) страница цветной обложки – 25 тысяч рублей.

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

1\2 цветной внутренней полосы – 8 тысяч рублей.

Ждем Ваших предложений.

С наилучшими пожеланиями, зам. главного редактора
д.т.н., профессор Витязев Владимир Викторович, телефон 8-903-834-81-81.

Предложения прошу направлять по адресу: E-mail: vityazev.v.v@rsreu.ru или info@dspa.ru

ОЦЕНКА ЧАСТОТНОЙ ХАРАКТЕРИСТИКИ РАДИОКАНАЛА В СПЕКТРАЛЬНОМ ДОМЕНЕ С УЧЁТОМ ДАННЫХ ДЛЯ OFDM СИГНАЛОВ

Иртюга В.А., к.т.н., начальник отдела разработки систем цифровой передачи данных ООО НПФ «САД-КОМ», Москва, e-mail: virtuga@mail.ru

Сарана Д.В., ведущий инженер ООО НПФ «САД-КОМ», Москва, e-mail: sarana@nircom.ru

Ключевые слова: OFDM, оценка канала, data-aided estimation, MAP, EM, Винеровская фильтрация.

Введение

В статье рассматривается алгоритм оценки комплексной частотной характеристики (ЧХ) канала с селективным частотным замиранием канала для OFDM сигнала. Оценка доплеровского смещения выходит за рамки этой статьи. На приёмной стороне расположение пилотных несущих и тип модуляции считаются известными.

Модель канала

Канал полагается многолучевым с селективным частотным замиранием и произвольной временной динамикой. Внутри полосы пропускания ЧХ канала изменяется плавно (динамика ограничена количеством лучей). На протяжении одного символа ЧХ канала считается постоянной. От символа к символу динамика изменения ЧХ зависит от скорости движения приёмника и считается заранее неизвестной.

Модель канала полагается линейной. Шум на приёмной стороне – комплексным (в спектральном домене) аддитивным белым гауссовым с постоянной мощностью для всех поднесущих и внутри интервала анализа (суперадра).

Сигнал в спектральном домене

Рассматривается обработка данных на приёмной стороне после этапов синхронизации, удаления защитного интервала и перехода в спектральный домен для каждого символа. Алгоритм требует некоторого накопленного количества символов N_s , содержащих по N_C несущих, из которых N_p являются пилотными, т.е. априори известными на приёмной стороне. Для каждого символа $k = [1..N_s]$ сигнал на приёмной стороне равен

$$\mathbf{Y}_k = \text{diag}(\mathbf{S}_k) \mathbf{H}_k + \mathbf{V}_k, \quad (1)$$

где \mathbf{Y}_k – измеряемый сигнал в частотной области (вектор-столбец размера $N_C \times 1$ для k -го символа);

\mathbf{S}_k – сигнал на стороне передатчика (вектор-столбец размера $N_C \times 1$ для k -го символа), часть из элементов которого являются известными пилотными значениями, часть – данными, подлежащими определению. Количество допустимых значений данных зависит от типа модуляции. Т.е. для 16-QAM, это 16 возможных значений для

Представлен итеративный алгоритм с низкой вычислительной нагрузкой для оценки частотной характеристики канала для OFDM-сигналов в частотном домене. Алгоритм использует начальную оценку канала, вычисляемую на основе пилотных несущих, одномерную адаптивную Винеровскую фильтрацию по временной оси и оценку Максимальной Апостериорной Вероятности (Maximum A-posteriori Probability, MAP), использующую как пилотные несущие, так и несущие с данными. На завершающем этапе оценки используется упрощённый вариант итеративного алгоритма Максимального Правдоподобия (Expectation-Maximization, EM), при этом производится только одна итерация для уменьшения общей вычислительной сложности.

каждого $\mathbf{S}_k(i)$, $i = 1..N_C$ (не считая пилотных несущих);

\mathbf{H}_k – неизвестная комплексная частотная характеристика канала (вектор-столбец размера $N_C \times 1$ для k -го символа). Временная стационарность и гладкость в частотной области зависят от скорости движения приёмника и количества отражений радиосигнала в случае многолучевого приёма;

\mathbf{V}_k – неизвестный комплексный аддитивный белый Гауссов шум (вектор-столбец размера $N_C \times 1$ для k -го символа) с нулевым средним и диагональной ковариационной матрицей. Как правило, значения элементов главной диагонали равны, т.е. энергия шума не зависит от индекса несущей.

Задачей исследования является требование максимально точно оценить комплексную частотную характеристику радиоканала $\mathbf{H} = [\mathbf{H}_1.. \mathbf{H}_{N_s}]$, используя доступные данные о расположении пилотов, значения сигнала $\mathbf{Y} = [\mathbf{Y}_1.. \mathbf{Y}_{N_s}]$, измеренные на приёмнике, а также допущения о гладкости \mathbf{H} по времени и частоте.

Предлагаемый алгоритм

Предварительная оценка по пилотным несущим

Грубая оценка \mathbf{H} доступна в пилотных точках (для k -го символа и l -й поднесущей):

$$\tilde{H}_{kp,lp} = \frac{Y_{kp,lp}}{S_{kp,lp}} \quad (2)$$

Поскольку на оценку в каждой из пилотных точек влияет неизвестный аддитивный шум $V_{kp,lp}$, можно несколько снизить его влияние, произведя усреднение в каждой из пилотных точек по нескольким соседним значениям $\tilde{H}_{kp,lp}$. Окно для усреднения (по времени и частоте)

требует знания о динамике \mathbf{H} по временному и частотному направлениям. После усреднения можно произвести двумерную интерполяцию между пилотными точками.

Более сложные адаптивные подходы к оценке двумерной частотной характеристики радиоканала сталкиваются с проблемой высоких вычислительных затрат. Как показано в [1], [2] и [3], двумерные фильтры успешно можно заменить на пару каскадно соединённых одномерных фильтров без сколько-нибудь заметного снижения точности оценки, при этом построение адаптивного одномерного фильтра требует значительно меньших вычислительных ресурсов. Можно произвести оценку сначала по частоте, т.е. независимо внутри каждого символа, а потом полученные результаты сгладить по времени одномерным адаптивным фильтром. При этом первый каскад можно сделать более гибким, рассматривая не часть пилотов в окрестности исследуемой точки, а все значения пилотов в символе совместно. Поскольку предполагается, что характеристика канала меняется вдоль несущих в символе гладко (количество минимумов АЧХ канала не может превышать количество лучей в модели многолучевого канала), то для описания канала достаточно небольшого набора параметров (например, для 8-лучевой модели канала 8-12 параметров будет достаточно). Пусть $\boldsymbol{\theta}$ – вектор таких параметров размером $N_0 \times 1$. Тогда обобщённую линейную модель канала для k -го символа можно записать как

$$\mathbf{H}_k = \mathbf{A}\boldsymbol{\theta}_k \quad (3)$$

где \mathbf{A} – фиксированная матрица преобразования вектора параметров $\boldsymbol{\theta}_k$ в вектор частотной характеристики канала \mathbf{H}_k . Размер матрицы $N_C \times N_0$.

Интерпретация физического смысла матрицы \mathbf{A} и вектора $\boldsymbol{\theta}_k$ может быть разной. Например, в качестве \mathbf{A} можно использовать первые N_0 столбцов обратного дискретного преобразования Фурье размером $N_C \times N_C$. Тогда элементы $\boldsymbol{\theta}_k$ будут представлять собой первые (низкочастотные) компоненты спектра \mathbf{H}_k , т.е. спектр спектра. Высокочастотные компоненты полагаются равными нулю и соответствующие столбцы матрицы \mathbf{A} отсутствуют [5]. В этом случае при любых конечных значениях параметров $\boldsymbol{\theta}_k$ частотная характеристика модели канала $\mathbf{H}_k = \mathbf{A}\boldsymbol{\theta}_k$ автоматически будет содержать только низкочастотные компоненты, что полностью соответствует допущению о гладкости \mathbf{H}_k . В предложенном алгоритме матрица \mathbf{A} составляется из первых N_0 полиномов Лежандра, а комплексные коэффициенты $\boldsymbol{\theta}_k$ – коэффициенты полиномиальной аппроксимации. При этом также любые ограниченные значения параметров модели $\boldsymbol{\theta}_k$ приводят к гладкой частотной характеристике, рассчитанной по этим параметрам в соответствии с формулой (3).

Проанализируем только один символ. Для простоты в дальнейшем опустим индекс символа k . Пусть \mathbf{Y} – значения, измеренные на приёмной стороне, \mathbf{S} – диагональная матрица, с главной диагональю, составленной из вектора переданного сигнала (известны только значения в пилотных точках $S_{p,lp}$), $\tilde{\mathbf{H}} = \mathbf{A}\boldsymbol{\theta}$ – искомая оценка частотной характеристики канала. Ожидаемый сигнал $\tilde{\mathbf{Y}}$, исходя из оценки канала $\tilde{\mathbf{H}}$, можно записать как:

$$\tilde{\mathbf{Y}} = \mathbf{S}\tilde{\mathbf{H}} = \mathbf{S}\mathbf{A}\boldsymbol{\theta} \quad (4)$$

Для Гауссового шума оценка по методу Максимального Правдоподобия (Maximum Likelihood, ML) приводит к минимизации среднеквадратичной нормы ошибки оценки сигнала:

$$E = \|\mathbf{W}(\mathbf{S}\mathbf{A}\boldsymbol{\theta} - \mathbf{Y})\|^2, \quad (5)$$

где \mathbf{W} – диагональная матрица размера $N_C \times N_C$ с диагональю весов с коэффициентами, равными 1 для пилотных несущих и 0 для остальных. Получаем искомый вектор параметров модели канала:

$$\boldsymbol{\theta} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{W}(\mathbf{S}\mathbf{A}\boldsymbol{\theta} - \mathbf{Y})\|^2 \quad (6 \text{ а})$$

$$\boldsymbol{\theta} = (\mathbf{A}^H \mathbf{S}^H \mathbf{W}^H \mathbf{W} \mathbf{S} \mathbf{A})^{-1} \cdot \mathbf{A}^H \mathbf{S}^H \mathbf{W}^H \mathbf{Y} \quad (6 \text{ б})$$

После нахождения $\boldsymbol{\theta}$, оптимального в смысле минимизации среднеквадратичной ошибки E , вычисляется оценка канала $\tilde{\mathbf{H}} = \mathbf{A}\boldsymbol{\theta}$ для каждого символа, и после низкочастотной фильтрации по времени – оценка $\tilde{\mathbf{H}}$ для всего блока данных

Использование межсимвольной корреляции

После оценки ЧХ канала на каждом символе независимо рассмотрим одну несущую независимо на последовательности из нескольких символов. Обозначим оцененную ЧХ как u , истинную ЧХ как s , ошибку оценки ЧХ как n . Тогда

$$u = s + n. \quad (7)$$

Межсимвольная корреляция ошибки n будет отсутствовать, поскольку оценка производилась для каждого символа независимо. Корреляция ошибки n с истинным значением s будет пренебрежимо мала, поскольку для оценки использовались все несущие. Поэтому можно считать s и n некоррелированными. Для оценки s можно использовать адаптивную Винеровскую фильтрацию [4]. Оценим s с помощью взвешенной суммы $2L+1$ значений u , т.е. вектора $\mathbf{u}(t) = [u(t-L) \dots u(t+L)]^T$, свернутого с весовым вектором \mathbf{w} :

$$\tilde{s}(t) = \mathbf{w}^H \mathbf{u}(t). \quad (8)$$

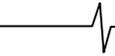
Оптимальные в смысле минимальной среднеквадратичной ошибки оценки s веса \mathbf{w} находятся с помощью решения уравнения Винера-Хопфа:

$$\mathbf{w}_{opt} = \mathbf{R}_{uu}^{-1} \mathbf{r}_{us}. \quad (9)$$

Корреляционная матрица \mathbf{R} оценивается по доступным наборам векторов \mathbf{u} . Вектор корреляции \mathbf{r}_{us} можно оценить, учтя некоррелированность s и n :

$$\mathbf{r}_{us} = \mathbf{r}_{u(u-n)} = \mathbf{r}_{uu} - \mathbf{r}_{(s+n)n} = \mathbf{r}_{uu} - \mathbf{r}_{nn}, \quad (10)$$

где вектора \mathbf{s} и \mathbf{n} строятся аналогично вектору \mathbf{u} : $\mathbf{s}(t) = [s(t-L) \dots s(t+L)]^T$ и $\mathbf{n}(t) = [n(t-L) \dots n(t+L)]^T$. В силу отсутствия межсимвольной корреляции n , можно считать, что неизвестная \mathbf{r}_{nn} содержит только одно ненулевое значение, а именно – центральное, т.е. мощность сигнала ошибки, а, следовательно, и в \mathbf{r}_{us} известно только одно центральное значение. Его можно оценить, проведя сплайновую интерполяцию центрального значения \mathbf{r}_{uu} . После этого оптимальные веса \mathbf{w} на-



ходятся из формулы (9), где в качестве \mathbf{r}_{us} выступает \mathbf{r}_{us} с интерполированным центральным значением. Для прямого решения уравнения Винера матрица \mathbf{R}_{us} является плохо обусловленной, однако при использовании достаточного объёма данных для набора статистики и высоких уровнях шумов (а, следовательно, и ошибки оценки s) решение является стабильным.

Набиралась общая статистика для всех поднесущих на интервале N_S символов. Дополнительно, данные взвешивались окном Хемминга (длиной N_S). Моделирование для 8-лучевого канала [7, 8] со скоростями движения приёмника 2, 60 и 150 км/ч показало устойчивость решения в широком диапазоне соотношения сигнал-шум от 0 до 30 дБ. При более высоких соотношениях сигнал-шум для надёжного декодирования достаточного первого этапа оценки канала.

Учёт несущих с данными

Для уточнения оценки канала есть возможность использовать информацию несущих с данными, учтя тот факт, что значения в точках данных на передающей стороне могут принимать ограниченное количество значений (16 значений для 16-QAM).

Пусть \tilde{h}_l – оценка канала, полученная после фильтра Винера для поднесущей l (здесь рассматривается один символ k , индекс которого опущен). Учитывая, что шум распределён гауссово с параметрами распределения $(0, \sigma^2)$, можно получить взвешенную оценку сигнала на передающей стороне для l -й поднесущей следующим образом:

$$\tilde{s}_l = \frac{\sum_j s_j P(s_j | \tilde{h}_l, y_l)}{\sum_j P(s_j | \tilde{h}_l, y_l)}, \quad (11)$$

где j – индекс суммирования по всем возможным значениям сигнала s_j на передающей стороне ($j=1..16$ для 16-QAM), а $P(s_j | \tilde{h}_l, y_l)$ – гауссова функция плотности вероятности с параметрами (y_l, σ^2) , вычисленная для точки $s_j \tilde{h}_l$, где y_l – измеренный сигнал на приёмной стороне в l -й поднесущей:

$$P(s_j | \tilde{h}_l, y_l) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|y_l - s_j \tilde{h}_l|^2}{2\sigma^2}\right). \quad (12)$$

Теперь можно включить в процедуру оценки канала помимо пилотных точек, точки с данными, считая известными оцененные значения \tilde{s}_l :

$$E = \|\mathbf{W}(\mathbf{S}\mathbf{A}\boldsymbol{\theta} - \mathbf{Y})\|^2 / \sigma^2 + \boldsymbol{\theta}^H \boldsymbol{\Sigma}_{00}^{-1} \boldsymbol{\theta}. \quad (13)$$

Здесь, в отличие от (5), главная диагональ матрицы \mathbf{S} вместо нулей, соответствующих несущим с данными, содержит значения \tilde{s}_l , а матрица весов \mathbf{W} становится единичной. Дополнительно, вместо ML оценки в (5) и (6) используется оценка по методу Максимальной Апостериорной Вероятности (Maximum A Posteriori Probability, MAP), учитывающая статистику параметров модели $\boldsymbol{\Sigma}_{00}$, которая может быть получена на предыдущих дан-

ных. Параметр распределения σ^2 легко может быть оценен по пилотным точкам после оценки канала с помощью фильтра Винера.

Минимизируя E , получим уточнённую оценку параметров модели канала:

$$\boldsymbol{\theta} = \left(\mathbf{A}^H \mathbf{S}^H \mathbf{W}^H \mathbf{W} \mathbf{S} \mathbf{A} + \sigma^2 \boldsymbol{\Sigma}_{00}^{-1}\right)^{-1} \cdot \mathbf{A}^H \mathbf{S}^H \mathbf{W}^H \mathbf{Y}. \quad (14)$$

После этого произведём переоценку ЧХ канала по формуле (3). Итеративное вычисление (11), (14) и (3) есть ни что иное, как EM алгоритм оценки параметров модели со скрытыми переменными \tilde{s}_l .

Алгоритм требует значительных вычислительных затрат и на практике малореализуем в такой форме. Поэтому была сделана попытка значительно уменьшить вычислительные затраты алгоритма, сохранив приемлемую точность оценки канала. Во-первых, вместо оценки ожидаемого значения \tilde{s}_l как взвешенной суммы возможных значений, выбирается наиболее вероятное

$$\tilde{s}_l = \arg \min_{s_j} \|y - \tilde{h}_l s_j\|^2. \quad (15)$$

Во-вторых, веса на главной диагонали матрицы весов \mathbf{W} вычисляются как произведение меры достоверности оценки \tilde{s}_l и меры доверия к оценке \tilde{h}_l в условиях шумов с дисперсией σ^2 :

$$w_l = \exp\left(-\frac{\|y - \tilde{h}_l \tilde{s}_l\|^2}{2\sigma^2}\right) \cdot \text{erf}\left(\frac{\|\tilde{h}_l\|^2}{\sigma^2}\right). \quad (16)$$

В третьих, эксперимент показал, что для приемлемой точности оценки канала достаточно одной итерации. При увеличении количества итераций мера достоверности гарантированно растёт (как и положено для EM-алгоритма), а рост точности оценки ЧХ канала после первой итерации не гарантируется. Это вызвано избыточностью степеней свободы модели канала на некоторых символах, и, соответственно, включением в параметры модели шумовых компонент.

После переоценки канала описанным выше способом для каждого символа независимо снова проводилась адаптивная винеровская фильтрация по временному направлению с параметром $L=7$.

Результаты

Для эксперимента использовалась многолучевая модель канала с 8 лучами в городе с движением приёмника 60 км/ч (канал #8) [6]. Аддитивный шум задавался отдельно. Т.о. «идеальная» характеристика радиоканала была известна, что позволило вычислить точность оценки канала для различных соотношений сигнал/шум. Точность вычислялась в децибелах по формуле:

$$SER = 10 \cdot \log_{10} \left(\frac{\sum_{l=1}^{N_C} \sum_{k=1}^{N_S} \|\tilde{h}_{k,l}\|^2}{\sum_{l=1}^{N_C} \sum_{k=1}^{N_S} \|\tilde{h}_{k,l} - \tilde{h}_{k,l}\|^2} \right). \quad (17)$$

Параметры сигнала, согласно стандарту РАВИС [7], [8]:
– длина символа во временной области 1800 отсчё-

тов, длина защитного интервала $1/8$, ширина полосы пропускания 245.3 кГц;

– число несущих $N_c=553$, из них 45 пилотных, с рассеянным расположением пилотных несущих;

– длина суперкадра, используемого для анализа $N_S=246$;

– тип модуляции – QPSK.

Результаты точности оценки канала в зависимости от соотношения сигнал/шум для разных методов, описанных выше, представлены в табл. 1. Первый столбец – результат оценки канала последовательной ML интерполяцией (IIa) по частотному направлению и адаптивной винеровской фильтрацией (IIb) по временному направлению. Второй столбец – уточнение оценки канала с использованием точек с данными (IIc) в частотном направлении и адаптивной винеровской фильтрацией (IIb) во временном направлении. Третий столбец – теоретический предел оценки ЧХ в шумах (при полностью известных данных на приёмной стороне).

Таблица 1.

SNR, dB	SER, dB		
	Только пилотные несущие	Предлагаемый алгоритм	Теоретический предел
4	21,72	25,26	30,34
6	23,62	27,11	31,72
8	25,54	28,93	32,98
10	27,37	30,58	34,12
12	29,12	31,97	35,0
14	30,75	33,18	35,77
16	32,22	34,16	36,31

Заключение

Рассмотренный алгоритм реализует значительный (до 3.5 дБ) выигрыш в точности оценки канала по сравнению с оценкой только на основе пилотных точек (в рабочих для 16-QAM диапазонах SNR) при приемлемых вычислительных ресурсах.

Литература

1. Hoeler P., Kaiser S., and Robertson P., «Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering», in Proc. IEEE International Conference on

Acoustics, Speech, and Signal Processing (ICASSP), Munich, Germany, Apr. 1997, pp. 1845–1848.

2. Necker M., Sanzi F., Speidel J., «An Adaptive Wiener-Filter for Improved Channel Estimation in mobile OFDM-Systems», International Symposium of Signal Processing and Information Technology, IEEE 28 – 30 December 2001, pp. 213–216.

3. Stephan Sand, Armin Dammann, Gunter Auer, «Adaptive Pilot Symbol Aided Channel Estimation for OFDM Systems», ISBN: 1-4020-1837-1 In book: Multi-Carrier Spread-Spectrum, For Future Generations Wireless Systems, Publisher: Kluwer Academic Publishers, Editors: null Fazel, K. Kaiser, St Source: DLR

4. Auer, G.; DoCoMo Euro-Labs, Munich, Germany; Karipidis, E., «Pilot Aided Channel Estimation for OFDM: a Separated Approach for Smoothing and Interpolation», Communications, 2005. ICC 2005. 2005 IEEE International Conference, Volume:4

5. Huang M, Chen X., Xiao L., Zhou S., and authors, «Kalman-filter-based channel estimation for orthogonal frequency-division multiplexing systems in time-varying channels», Communications, IET (Volume:1, Issue:4)

6. ETSI ES 201 980 V4.1.1 (2014-01). Digital Radio Mondiale (DRM); System Specification. (Annex B.2)

7. Report ITU-R BT.2295-0 (12/2013). Digital terrestrial broadcasting systems.

8. GOST R 54309–2011. Realtime audiovisual information system (RAVIS). Framing structure, channel coding and modulation for digital terrestrial narrowband broadcasting system for VHF band. Technical specification.

FREQUENCY DOMAIN DATA-AIDED CHANNEL ESTIMATION FOR OFDM SIGNALS

Irtuga V.A., Sarana D.V.

An iterative low-complexity channel estimation algorithm for OFDM signals in frequency domain is presented. Algorithm uses initial channel estimation based on pilot subcarriers, 1-D adaptive Wiener filtering in the time direction and Maximum A-posteriori Probability (MAP) estimation that uses both pilot and weighted data subcarriers. The final step is the simplified version of an iterative Expectation-Maximization (EM) algorithm that uses only single iteration to reduce total computational complexity.

Уважаемые коллеги!

Для тех, кто не успел оформить подписку на второе полугодие 2014 года через ОАО «Роспечать», сохраняется возможность приобретения журналов непосредственно в редакции по адресу: г. Москва, ул. Авиамоторная, дом 8, Научный Центр МТУСИ, ком. 612. Российское научно-техническое общество радиотехники, электроники и связи им. А.С. Попова, метро «Авиамоторная», или оформить заказ в соответствии с требованиями, выставленными на сайте журнала: www.dsra.ru.

Справки по телефонам: (8-495) 362-42-75; (+7 903) 221-79-79 (Алексеева Любовь Ильинична).

E-mail: rntores@mail.ru

СИНТЕЗ НИЗКОПЛОТНОСТНЫХ КОДОВ НА ОСНОВЕ УРАВНОВЕШЕННЫХ НЕПОЛНЫХ БЛОК-СХЕМ

*Овинников А.А., научный сотрудник кафедры телекоммуникаций и основ радиотехники
Рязанского государственного радиотехнического университета, e-mail: ovinnikovalexey@gmail.com*

Ключевые слова: помехоустойчивое кодирование, итеративное декодирование, низкоплотностные коды, уравновешенные неполные блок-схемы, системы Штейнера.

Введение

В настоящее время особое значение приобретают задачи повышения надёжности и достоверности передаваемой информации. Эффективным средством решения такой задачи является использование теории и практики помехоустойчивого кодирования.

На сегодняшний день существует огромное множество различных методов и алгоритмов канального кодирования и декодирования, которые отличаются друг от друга по сложности реализации, энергетическому выигрышу, а также другим показателям. Однако, доминирующее положение с позиции энергетической эффективности, широты внедрения, а также перспективы дальнейшего развития теории и практики помехоустойчивого кодирования занимают итеративно-декодируемые конструкции, к которым относятся турбо и низкоплотностные (LDPC) коды. Среди этих двух ансамблей LDPC коды, впервые рассмотренные Р. Галлагером [1], обладают эффективным алгоритмом декодирования, сложность которого с ростом длины кодового слова (N) пропорциональна $O(N)$. Кроме того, они могут применяться для исправления как одиночных, так и пакетных ошибок без необходимости применения процедуры перемежения, что существенно сказывается на задержке, присутствующей при обработке информации.

Существенной проблемой теории низкоплотностных кодов является отсутствие достаточного количества алгоритмов синтеза, позволяющих получать коды в большом диапазоне кодовых длин и скоростей, обладающих алгебраически стройной структурой, которая с лёгкостью может быть перенесена на современную элементную базу.

В данной работе рассматривается задача построения эффективных с точки зрения энергетического выигрыша декодирования алгебраических кодовых конструкций для высокоскоростных систем передачи информации с малой избыточностью. Её решение основано на теории комбинаторики и смежных подразделах. Эффективность полученных кодов оценивается с помощью имитационного моделирования по методу Монте Карло, а также путём вычисления ряда принципиальных параметров, присущих LDPC кодам.

Классификация кодов, полученных на основе УНБС

Конструктивные методы построения LDPC кодов [2]

Рассматривается задача построения эффективных с точки зрения энергетического выигрыша декодирования алгебраических кодовых конструкций низкоплотностных кодов для высокоскоростных систем передачи информации с малой избыточностью. Её решение основано на теории комбинаторики и смежных подразделов. Эффективность полученных кодов оценивается с помощью имитационного моделирования по методу Монте Карло, а также путём вычисления ряда принципиальных параметров, присущих LDPC кодам.

всегда привязаны к некоторым строго определённым математическим объектам. В частности, в настоящей работе рассматриваются коды, синтезированные на основе так называемых уравновешенных неполных блок-схем (УНБС, BIBD – balance incomplete block design) [3]. Уравновешенной неполной блок-схемой называется такое размещение v различных элементов по b блокам, что каждый блок содержит ровно k различных элементов, каждый элемент появляется точно в r различных блоках и каждая пара различных элементов a_i и a_j появляется точно в λ блоках. Таким образом, любая УНБС может быть задана вектором из пяти связанных друг с другом чисел – (v, b, r, k, λ) , которые в свою очередь имеют между собой детерминированную связь вида:

$$\begin{aligned} b \cdot k &= v \cdot r; \\ r(k-1) &= \lambda(v-1). \end{aligned} \quad (1)$$

Существование зависимостей (1) позволяет использовать сокращённый способ записи УНБС с тремя независимыми параметрами: (v, k, λ) , где v – количество элементов, k – число блоков в УНБС и λ – количество поэлементных сочетаний между блоками. Графически любую блок-схему можно представить в виде набора блоков или матрицей инцидентности, в частности, для $v=b=7$, $r=k=3$, $\lambda=1$ существует единственная УНБС вида:

$$\begin{aligned} B_0 &= (0, 1, 3); \\ B_1 &= (1, 2, 4); \\ B_2 &= (2, 3, 5); \\ B_3 &= (3, 4, 6); \\ B_4 &= (4, 5, 0); \\ B_5 &= (5, 6, 1); \\ B_6 &= (6, 0, 2); \end{aligned} \quad A = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix},$$

где $A = (a_{ij})$, $i = 1, \dots, v$, $j = 1, \dots, b$; причём

$$a_{ij} = \begin{cases} 1, & \text{если } a_i \in B_j, \\ 0, & \text{если } a_i \notin B_j \end{cases}, \text{ где } a_{ij} \text{ и } B_j \text{ – элементы и бло-$$

ки УНБС соответственно. Если внимательно присмот-

реться к предлагаемому примеру, то не трудно обнаружить, что укороченная на 3 последних строки матрица инцидентности A представляет собой не что иное, как проверочную матрицу самого короткого кода Хэмминга. Аналогичным образом на основе выбранной математической абстракции для больших значений параметров (v, k, λ) можно получать и другие классы канальных кодов, в том числе и низкоплотностные с проверочной матрицей (H) . Любая матрица инцидентности соответствующая блок-схеме с параметрами (v, k, λ) обладает следующими свойствами:

- любая строка или столбец содержит r и k ненулевых элементов соответственно;
- любая пара строк или столбцов имеет максимум λ общих ненулевых позиций.

Исходя из свойств УНБС, гарантировать отсутствие циклов длины 4 в графах Таннера LDPC [2] кодов возможно лишь при $\lambda=1$. Таким образом, появляется первое существенное ограничение, накладываемое на УНБС, которое приводит к сужению множества блок-схем до подмножества так называемых систем Штейнера $(v, k, 1)$ [3]. Среди всех систем Штейнера особый интерес представляет подмножество циклических блок-схем с различным количеством элементов $v=(3,4,5,\dots)$. В целом используемую в работе классификацию УНБС можно изобразить в форме вложенных подмножеств, представленных на рис. 1. Цифрами отмечено количество элементов v в каждой из блок-схем подмножества.



Рис. 1. Представление УНБС в виде вложенных множеств

Обобщённый алгоритм синтеза LDPC кодов

Учитывая свойства блок-схем, а также структуру рассмотренных выше матриц инцидентности, не трудно сформулировать обобщённый алгоритм синтеза LDPC кодов на базе используемых в данной работе математических абстракций. Процедура получения проверочной матрицы низкоплотностного кода на базе УНБС изображена в форме последовательного набора операций, представленных блоками на рис. 2. Первый и последний блок в представленном алгоритме являются наиболее трудоёмкими и вариативными. На первом этапе выполнения процедуры (рис. 2.) ставится задача получить УНБС с заданными параметрами. При этом могут использоваться различные комбинаторные объекты и структуры [3].

Стоит отметить, что далеко не все возможные блок-схемы подходят для синтеза LDPC кодов. К выбору параметров следует подходить с особой тщательностью. Рассматриваемая на рис. 2. процедура формирования кодов обладает достаточно большой вариативностью,

которая позволяет создавать проверочные матрицы H (4) в широком диапазоне кодовых длин и скоростей. При этом для одного и того же набора параметров (v, k, λ) может быть получено множество различных LDPC кодов. Алгоритм создания УНБС может существенно отличаться в зависимости от выбора величины λ , однако, чаще всего её принимают равной единице, что существенно упрощает 3-ю задачу алгоритма – декомпозицию.



(1) – УНБС (v, k, λ) , (2) – A , (3) – H

Рис. 2. Обобщённый алгоритм синтеза LDPC кодов на основе комбинаторных блок-схем

Системы Штейнера, как математическая основа построения LDPC кодов, и их свойства

Рассмотрим общие свойства, присущие ансамблю LDPC кодов, которые получены из систем Штейнера. Известно [3], что матрицы инцидентности этого математического множества имеют максимально возможное число блоков b для всех УНБС. Это приводит к тому, что LDPC код, сформированный на базе такой системы, обладает максимально возможной скоростью среди всех кодов с охватом графа Таннера $g_0=6$, весом столбца $d_s=k$ и числом проверочных уравнений $M=b$. Ранг проверочной матрицы H может оказаться неполным, что дополнительно увеличивает кодовую скорость, которая ограничена снизу величиной [4]:

$$R \geq 1 - \frac{k(k-1)}{v-1} \tag{2}$$

Кроме того, число независимых строк в матрице инцидентности систем Штейнера характеризуется следующей нижней границей [4]:

$$\text{Rank}_2(H) \geq (k-1)\sqrt{(r-1)r/k} \tag{3}$$

тогда кодовая скорость ограничена сверху, как:

$$R \leq 1 - \frac{k(k-1)\sqrt{(v-1)(v-k)/k}}{v(v-1)} \tag{4}$$

Графически зависимость кодовой длины от скорости для реально существующих квазициклических кодов на основе систем Штейнера представлена на рис. 3. Параметр k для каждой УНБС варьируется от 3-х до 5-ти, однако может быть и большим. Нетрудно заметить, что скорость кодирования крайне быстро стремится к 1 с ростом длины кода.

Важным подклассом систем Штейнера являются так называемые циклически-разрешимые системы, для которых характерно то, что общее количество блоков b можно разделить на r подгрупп (дубликатов) таким образом, что исключение одной из них уменьшает величину r для всех v элементов одинаково. В силу того, что асимптотический анализ ансамблей регулярных LDPC кодов [5] показывает, что для всех скоростей кодирования наибольшим энергетическим выигрышем в ОСШ в канале с белым шумом показывают коды с $d_s=3$ далее будет предложен обобщённый алгоритм их формирования на основе соответствующих систем Штейнера.

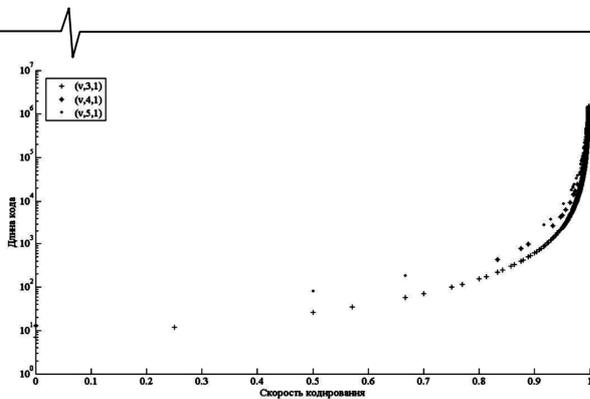


Рис. 3. Зависимость кодовой длины от скорости кодирования для LDPC кодов построенных на базе различных систем Штейнера

Блок-схема с $k=3$ и $\lambda=1$ вполне естественно называется системой троек Штейнера. Параметры этой УНБС удовлетворяют следующим равенствам, полученным из (1):

$$b = \frac{v(v-1)}{6}, \quad r = (v-1)/2. \quad (5)$$

Необходимым и достаточным условием существования троек Штейнера является выполнение тождества вида

$$v \equiv 1, \quad 3 \pmod{6}. \quad (6)$$

Таким образом [3], если $v = 6t + 1$ или $v = 6t + 3$, при целом значении числа t , существует система троек Штейнера порядка v . Некоторые УНБС рассматриваемого подмножества представлены в табл. 1 и 2. Разделение троек Штейнера на 2 группы не случайно, т.к. алгоритмы их получения в том случае, когда базовые блоки b формируют циклические группы, оказываются различными.

Обобщённый алгоритм синтеза LDPC кодов на основе систем Штейнера

Задача формирования подмножества циклических и одновременно разрешимых троек Штейнера оказывается далеко не тривиальной. Однако в работах [6] было показано, что на основе различных комбинаторных последовательностей можно сформировать абсолютно все тройки из ансамбля $(v, 3, 1)$. Принимая во внимание этот факт, можно предложить обобщённый алгоритм формирования проверочных матриц LDPC кодов на базе описанных выше блок-схем (рис. 4). Фактически алгоритм позволяет по заданным параметрам подсистемы канального кодирования (K, N) получить проверочную матрицу H регулярного низкоплотного кода с квазициклической структурой. Предлагаемая на рис. 4. процедура состоит из 3-х этапов, первый из которых заключается в получении некоторых комбинаторных последовательностей и их модификаций, получивших своё название в честь авторов – Т. Skolem и А. Rosa [6]. Длина получаемой последовательности S аналитически связана с входными параметрами кодирования следующим образом:

$$n = (K-1)/6, \quad (7.1)$$

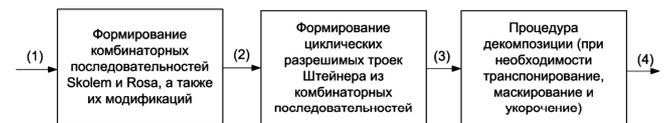
$$n = (K-3)/6, \quad (7.2)$$

$$n = \max\left(\frac{K-3}{6} - 1, \frac{K-1}{6} + 1\right), \quad (7.3)$$

причём конкретное значение величины n определяется вычислением каждого из представленных выражений, с приоритетом выбора от 7.1 к 7.3. Главное условие, чтобы n

являлось целым числом. Таким образом, в случае если первое и второе выражения дадут дробные значения, искомая величина размерности комбинаторной последовательности будет найдена по формуле (7.3), как максимальное целое значение в диапазоне указанном в скобках. На втором этапе алгоритма набор чисел S будет использован для синтеза циклически разрешимых троек Штейнера с помощью известных в математике [3, 6] процедур.

В итоге получается система Штейнера с известной матрицей инцидентности, результатом декомпозиции которой является искомый LDPC код. Может получиться такая ситуация, когда параметры кода, полученного на выходе обобщённого алгоритма синтеза (рис. 4, (1)), не совпадают с входными. В этом случае в 3-ем блоке схемы предусмотрена процедура укорочения проверочной матрицы H .



- (1) – параметры кода LDPC (K, N)
- (2) – комбинаторная последовательность S
- (3) – циклическая разрешимая тройка Штейнера $(v, 3, 1)$
- (4) – проверочная квазициклическая матрица LDPC кода

Рис. 4. Обобщённый алгоритм синтеза проверочных матриц LDPC кодов на базе циклических разрешимых троек Штейнера

Результаты имитационного моделирования LDPC кодов

Оценим энергетическую эффективность помехоустойчивых LDPC кодов, полученных с помощью обобщённого алгоритма, представленного на рис. 4. В силу того, что кодовая скорость R для выбранного ансамбля быстро стремится к единице с ростом параметра N (рис. 3), выбрать подходящие для сравнения конструкции Таннера не представляется возможным. Поэтому в качестве конкурирующих решений были взяты наиболее распространённые алгоритмы синтеза псевдослучайных кодов (Mac, PEG) [7, 8]. Каждый из них способен генерировать проверочные матрицы произвольной длины и скорости, регулярной и нерегулярной структуры с заданными весовыми функциями $\lambda(x)$ и $\rho(x)$. Отличия между выбранными алгоритмами кроются лишь в том, что второй – PEG, максимизирует локальный обхват графа Таннера, а первый – Mac, пытается как можно точнее воспроизвести заданное весовое распределение с учётом $g_0=6$. Процедура оценки помехоустойчивости кодов выполнялась по методу Монте Карло в канале с аддитивным белым гауссовским шумом и двоичной фазовой модуляцией. Декодирование осуществлялось по алгоритму BP [7], причём максимальное количество итераций фиксировалось значениями 10 и 50. Целесообразность такого ограничения заключается в необходимости оценить возможный дополнительный энергетический выигрыш декодера при кратном увеличении вычислительных затрат. Кроме того, для оценки потенциальной близости выбранных конструкций к пределу Шеннона [9] каждой кодовой длине и скорости было поставлено в соответствие минимальное значение отношения сигнал-шум (σ_{min}), для которого возможен безошибочный приём. Для численного моделирования были выбраны шесть различных кодов, параметры которых представлены в табл. 3.

Имитационное моделирование выполнялось с шагом равным 0.5 дБ, причём для получения достоверной точки необходимо было накопить более 1000 ошибок на каждом шаге ОСШ. Результаты эксперимента представлены зависимостями вероятности ошибки на бит (p_b) от отношения сигнал-шум (E_b/N_0) на рис. 5-10. Разница в энергетической эффективности оценивается по уровню $p_b=10^{-6}$. Для каждого кода при заданном числе итераций декодирования вычислено значение превышения предельного уровня ОСШ

σ_{min} , выраженное в децибелах и отражено в табл. 4 (ΔS).

Кроме того, для каждой пары кодов найдены значения относительных энергетических выигрышей $\Delta a(i)-a(j)=E_b/N_0(a(i)) - E_b/N_0(a(j))$ для 10 и 50 итераций декодирования, где $a(i)$ и $a(j)$ – наименование соответствующего алгоритма синтеза низкоплотностного кода. Параметр $\Delta S-a(i)=\sigma_{min} - E_b/N_0(a(i))$ показывает на сколько энергетическая эффективность выбранного кода отличается от предельно возможной.

Таблица 1. Первые десять троек Штейнера порядка $v = 6t + 1$

t	1	2	3	4	5	6	7	8	9	10
v	7	13	19	25	31	37	43	49	55	61
r	2	4	6	8	10	12	14	16	18	20
b	7	26	57	100	155	222	301	392	495	610

Таблица 2. Первые десять троек Штейнера порядка $v = 6t + 3$

t	1	2	3	4	5	6	7	8	9	10
v	9	15	21	27	33	39	45	51	57	63
r	4	7	10	13	16	19	22	25	28	31
b	12	35	70	117	176	247	330	425	532	651

Таблица 3. Параметры LDPC кодов, используемых в имитационном моделировании

№	R	M	N	σ_{min}
1	2/3	19	57	1,059
2	0,7	21	70	1,275
3	0,833	37	222	2,361
4	0,842	39	247	2,457
5	0,9	61	610	3,199
6	0,903	63	651	3,247

Таблица 4. Результаты имитационного моделирования LDPC кодов

M	$\Delta_{Mac-PEG}$ 10/50 ит., дБ	$\Delta_{Mac-STC}$ 10/50 ит., дБ	$\Delta_{PEG-STC}$ 10/50 ит., дБ	Δ_{S_Mac} , дБ	Δ_{S_PEG} , дБ	Δ_{S_STC} , дБ
19	2,41/2,09	2,51/2,14	0,1/0,05	7,581	5,541	5,491
21	2,62/2,21	2,56/2,01	-0,06/-0,2	7,355	5,125	5,325
37	2/1,45	2/1,52	0/0,07	4,899	3,449	3,379
39	1,39/1	1,47/1,05	0,08/0,05	4,353	3,353	3,303
61	1,13/0,7	1,18/0,75	0,05/0,05	3,211	2,511	2,461
63	0,88/0,67	0,99/0,7	0,11/0,03	3,153	2,483	2,453

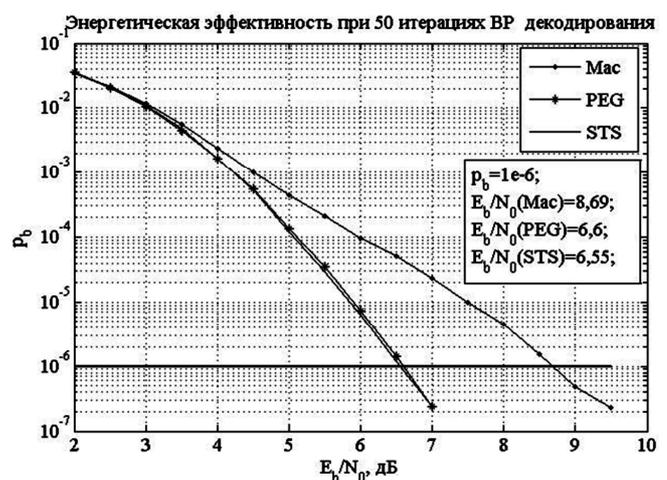
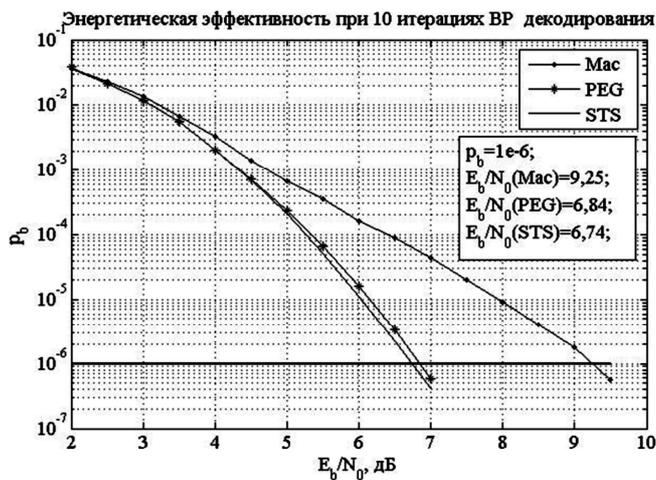


Рис. 5. Энергетическая эффективность низкоплотностных кодов с параметрами $M=19, N=57$ при различном числе итераций декодирования

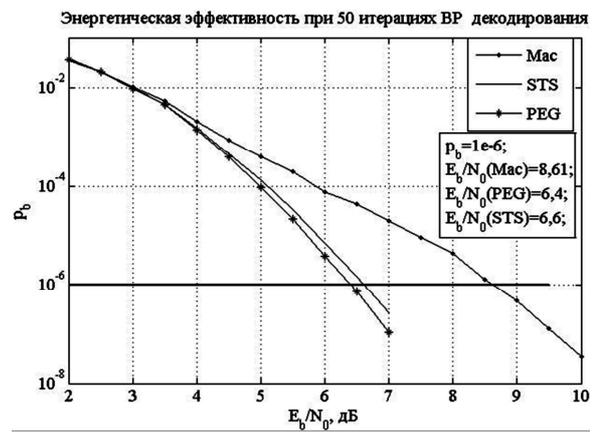
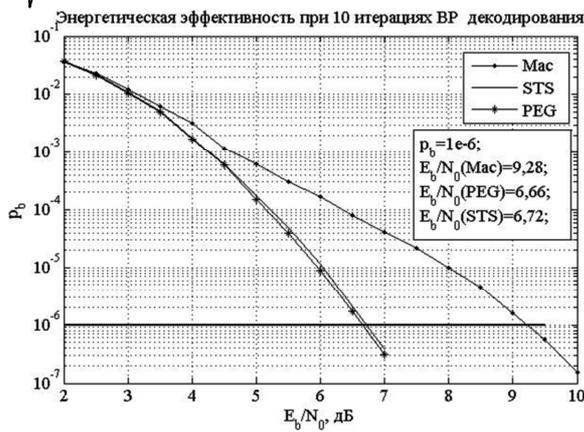


Рис. 6. Энергетическая эффективность низкоплотных кодов с параметрами $M=21$, $N=70$ при различном числе итераций декодирования

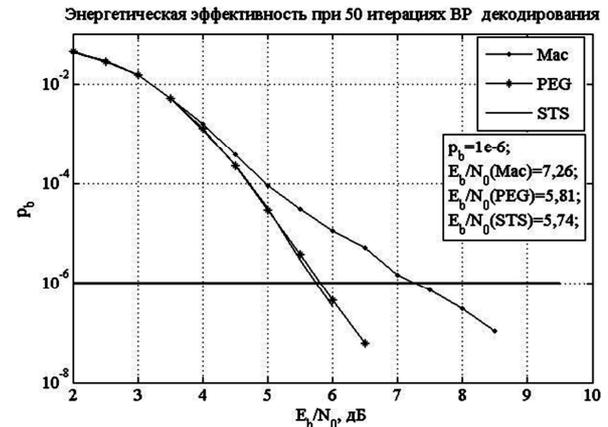
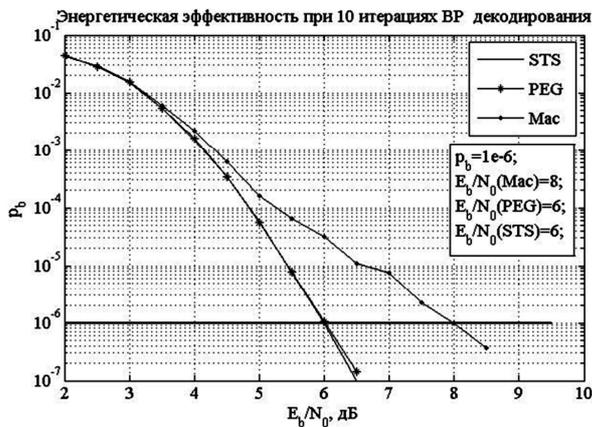


Рис. 7. Энергетическая эффективность низкоплотных кодов с параметрами $M=37$, $N=222$ при различном числе итераций декодирования

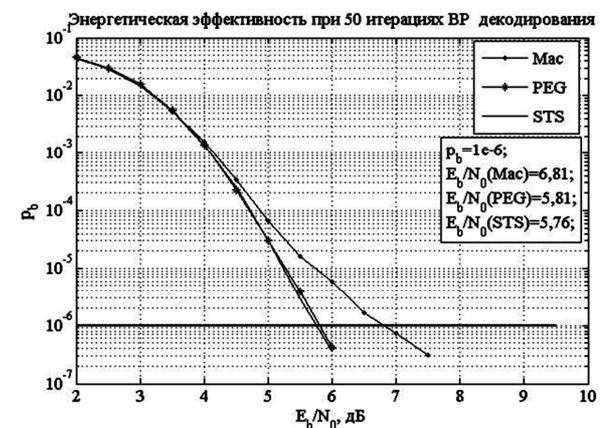
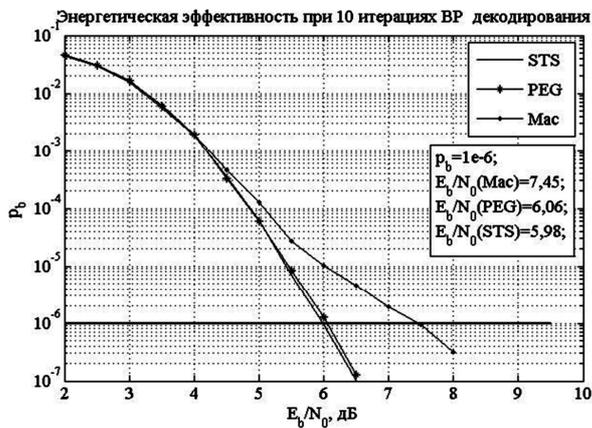


Рис. 8. Энергетическая эффективность низкоплотных кодов с параметрами $M=39$, $N=247$ при различном числе итераций декодирования

Таким образом, судя по результатам, представленным на рис. 5-10, а также в табл. 4, коды, синтезированные на основе комбинаторных последовательностей, показывают значительно лучшие результаты по сравнению с псевдослучайными кодами Маккая (Mac). Численное значение выигрыша в ОСШ для предлагаемых конструкций варьируется в диапазоне от 0,7 до 2,56 дБ и зависит исключительно от длины кода и числа итераций декодирования. Такое преимущество получено не случайно, а связано с тем, что для высоких скоростей кодирования, а также малых длин $N < 100$ рассматриваемые псевдослучайные алгоритмы синтеза не могут создать граф Таннера, не имеющих циклов минимальной длины, следовательно, $g_0=4$.

Это, в свою очередь, приводит к значительной деградации помехоустойчивости и делает неоспоримым преимущество предлагаемых квазициклических LDPC кодов по сравнению с Mac конструкцией. Однако, при сравнении с кодами PEG, преимущество в энергетической эффективности оказывается не столь очевидным, варьируется в пределах 0,03-0,1 дБ, и в одном эксперименте (рис. 6) даже наблюдается незначительный проигрыш, до 0,2 дБ. Но даже при столь малой разнице в помехоустойчивости синтезированный по обобщенному алгоритму ансамбль LDPC кодов обладает одним неоспоримым преимуществом – квазициклической структурой, которая крайне удобна при реализации подсистемы кодирования на аппаратном уровне.

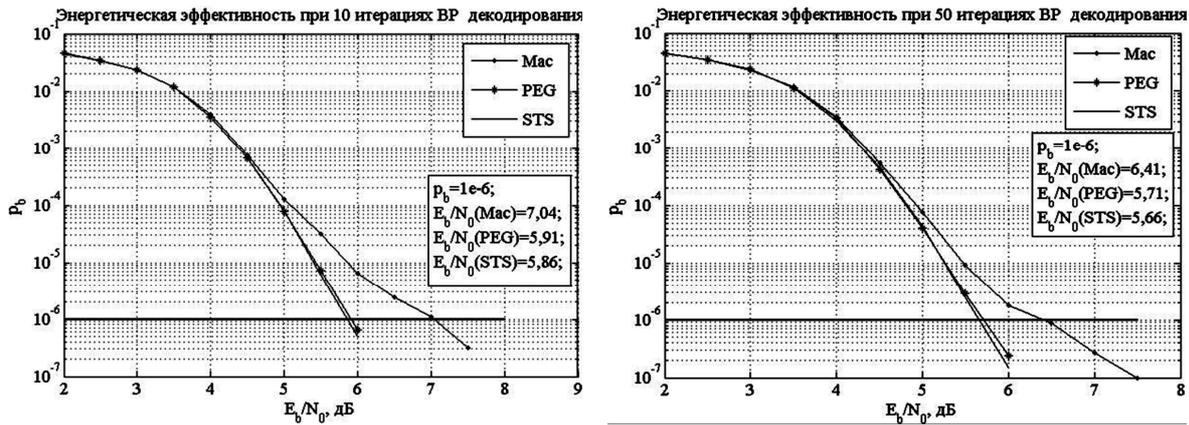


Рис. 9. Энергетическая эффективность низкоплотностных кодов с параметрами $M=61$, $N=610$ при различном числе итераций декодирования

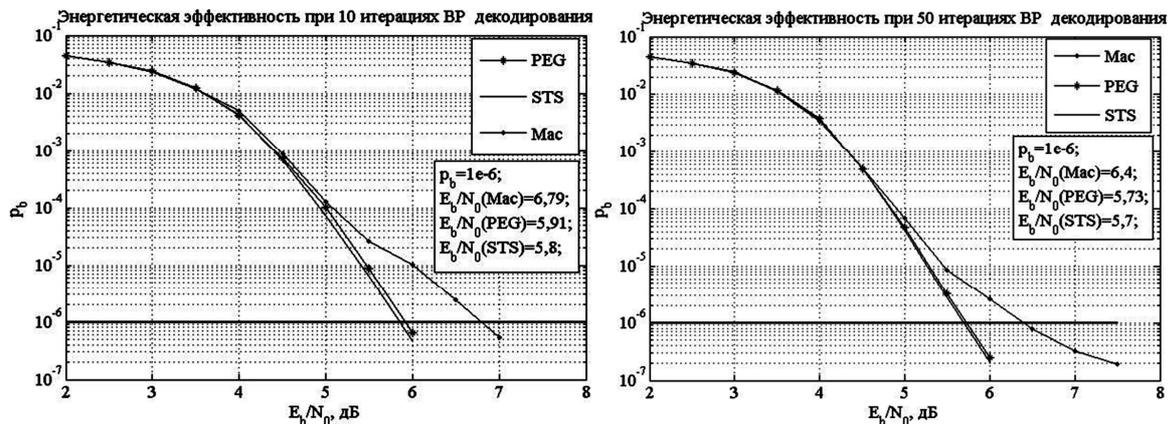


Рис. 10. Энергетическая эффективность низкоплотностных кодов с параметрами $M=63$, $N=651$ при различном числе итераций декодирования

Заключение

В работе проведено имитационное моделирование низкоплотностных кодов, синтезированных на основе уравновешенных неполных блок схем, которые в свою очередь базируются на комбинаторных последовательностях Сколема и Роса. Сформированные коды показали хорошие результаты по помехоустойчивости в сравнении с подклассом псевдослучайных кодов (Mac); полученный энергетический выигрыш декодирования лежит в диапазоне от 0,7 до 2,56 дБ. Обобщенный алгоритм синтеза (рис. 4), предлагаемый в данной работе, позволяет получать ансамбли квазициклических LDPC кодов с различными кодовыми длинами и скоростями, что потенциально расширяет границы применения подобных конструкций. Одним из направлений дальнейших исследований является поиск наиболее эффективных кодов в рамках синтезированного ансамбля.

Исследование выполнено за счет гранта Российского научного фонда (проект 14-19-01263) в Рязанском государственном радиотехническом университете.

Литература

1. Gallager R.G. Low-Density Parity-Check Codes. Cambridge MA: MIT Press, 1963.
2. Ryan W.E. and Lin S. «Channel Codes. Classical and Modern», Cambridge University Press, 2009.
3. Холл М. Комбинаторика, Издательство «МИР», М. 1970.
4. Johnson S.J. and Weller S.R. Regular low-density parity-check codes from combinatorial designs. In Proc. IEEE Information

Theory Workshop (ITW2002), p. 90-92, Cairns, Australia, September 2001.

5. Richardson T.J., Shokrollahi M.A, and Urbanke R.L. Design of capacity approaching irregular low-density parity-check codes. IEEE Trans. Inform. Theory, 47(2):619–637, February 2001.

6. Colbourn C.J. and Rosa A. Triple Systems. Oxford University Press, 1999.

7. MacKay D.J.C. Good error-correcting codes based on very sparse matrices. IEEE Trans. Inform. Theory, 45(2):399–431, March 1999.

8. Arnold D.M., Eleftheriou E., and Hu X.Y., «Progressive edgegrowth Tanner graphs», in Proc. IEEE Global Telecommun. Conf., San Antonio, TX, Nov. 2001, vol. 2, pp. 995–1001.

9. Shannon C.E. A mathematical theory of communication. Bell Sys. Tech. J., 27:379–423, 623–656, July–Oct. 1948.

BALANCE INCOMPLETE BLOCK DESIGNS BASED LDPC CODES

Ovinnikov A.A.

In this paper the problem of constructing efficient algebraic Low-Density Parity Check codes for high data transmission with low redundancy is considered. Code design algorithm is based on the theory of combinatorics and related subsections. The effectiveness of codes, which are used in the research, is evaluated using Monte Carlo simulations, as well as by calculating the number of fundamental parameters inherent to LDPC codes.

ПЕРСПЕКТИВЫ СОЗДАНИЯ КОМПЛЕКСОВ ДКМВ РАДИОСВЯЗИ, ИСПОЛЬЗУЮЩИХ ТЕХНОЛОГИИ МИМО С ПРОСТРАНСТВЕННЫМ И ПОЛЯРИЗАЦИОННЫМ РАЗНЕСЕНИЕМ

*Букашкин С.А., д.т.н., генеральный директор ОАО «Концерн «Автоматика», e-mail: fgup@nii.ru;
Оглоблин А.В., заместитель командира войсковой части, e-mail: homealeks@mail.ru.*

Ключевые слова: ДКМВ радиосвязь, технологии МИМО, многолучевое распространение, пространственное, поляризационное разнесение, антенные системы, программный комплекс.

Рассмотрены существующие технологии МИМО с пространственным и поляризационным разнесением. Проанализированы достигнутые результаты. Сформулированы основные задачи и пути создания комплексов ДКМВ радиосвязи с использованием технологий МИМО.

Введение

Роль радиосвязи в диапазоне ДКМВ уже ни у кого не вызывает сомнений, поэтому в настоящее время наметилась тенденция по интенсивному совершенствованию и развитию этого вида радиосвязи. К несомненным достоинствам ДКМВ радиосвязи относятся высокая гибкость и мобильность, относительно слабая зависимость от инфраструктуры, возможность организации связи в кратчайшие сроки и при минимальных затратах, в том числе из неподготовленных районов, высокая живучесть, способность функционировать в условиях частичных отказов, включая условия боевых действий и чрезвычайные ситуации [1, 2].

Недостатки ДКМВ радиосвязи также общеизвестны [1, 2]. Это, прежде всего, относительная узкополосность (а следовательно низкая пропускная способность) и относительно невысокое качество отдельного ДКМВ радиоканала, связанное с нестабильностью физических параметров ионосферы и сложной помеховой обстановкой, особенно в условиях постановки преднамеренных помех.

Проблема повышения пропускной способности и помехозащищенности ДКМВ радиосвязи при ограниченности частотного ресурса в настоящее время является одной из самых острых и актуальных. Применение цифровой обработки сигналов позволяет перейти к передаче информации в дискретной форме в сочетании с использованием помехозащищенных сигнально-кодовых конструкций, в связи с чем становится возможным внедрение технологии автоматизированной оперативной адаптации радиолиний к изменению параметров среды распространения [2].

Данная статья посвящена обсуждению путей разработки методики создания современных быстро разворачиваемых полевых комплексов ДКМВ радиосвязи с применением новых технологий, обеспечивающих повышение надежности и устойчивости радиосвязи.

Одним из перспективных путей повышения пропускной способности и помехозащищенности ДКМВ радиосвязи является использование технологии множественной передачи и приема (Multi Input Multi Output – MIMO). Исследования различных вариантов реализации техно-

логии ведутся достаточно давно и описаны в литературе. В частности, рассмотрены возможности реализации быстро разворачиваемых комплексов, использующих схему Аламоути и пространственное разнесение передатчиков [3, 4].

Дальнейшее развитие этого направления может быть связано с использованием увеличенного (более двух) числа передатчиков и приемников в сочетании с различными видами разнесения (пространственное, поляризационное, комбинированное). Реализация этого подхода связана с решением ряда частных задач, в числе которых следует выделить разработку математических моделей канала МИМО для различных типов трасс, видов разнесения, типов антенных систем; выбор (разработку) антенных систем для различных вариантов построения комплексов оборудования МИМО; разработку решений по построению комплексов различного назначения, включая быстро разворачиваемые, с учетом основных тактико-технических и технико-экономических факторов (требований).

Технология МИМО является методом кодирования сигнала, позволяющим повысить скорость передачи данных путём использования нескольких приемных или передающих систем. Таким образом, технология МИМО предполагает изначально многолучевое распространение сигнала. При этом необходимым условием является низкий уровень корреляции между информацией, передаваемой по разным каналам.

Анализ литературы выявил, что исследования технологии МИМО в настоящее время сосредоточены в области СВЧ. В первую очередь эта технология применяется на частотах мобильной радиосвязи в составе технологий WiFi и LTE [5, 6].

Следует отметить, что наиболее часто вышеупомянутые технологии применяются в городских условиях, где множественные переотражения приводят к многолучевому распространению сигнала естественным образом. В этих условиях, с целью повышения декорреляции сигнала, целесообразным оказывается применение нескольких антенных систем. Сравнительно малая рабо-

чая длина волны делает особенно удобным использование пространственного разнесения каналов. Применение же поляризационного разнесения ограничено.

Вместе с тем, на технологию MIMO не накладываются непосредственно ограничения по частотному диапазону. Однако, оценивая возможности и перспективы применения MIMO в ДКМВ диапазоне, следует учитывать особенности трасс распространения радиоволн данных частот.

Как известно [1, 7], ДКМВ-радиосвязь на значительные расстояния осуществляется по одно- или многоскачковым ионосферным трассам, а на относительно небольшие расстояния – по ионосферным трассам зенитного излучения или земной волной. В последнем случае дальность связи невелика, однако, при деструкции ионосферы в результате боевых действий этот вариант может оказаться единственно доступным. Следует отметить, что в реальных условиях работы возможен и часто наблюдается прием «паразитного» зенитного либо приземного излучения.

С учетом основных обстоятельств, трассы распространения характеризуются [7] нестационарностью (ввиду нестационарного характера электрофизических характеристик ионосферы), возникновением многолучевости (ввиду наличия отражений от различных слоев ионосферы или наложении волн различных трасс) и существенным изменением вида и характеристик поляризации (за счет вращения плоскости поляризации при отражении от слоев ионосферы, суперпозиции различных, нормальных и аномальных отраженных волн и т.п.), причем различные лучи характеризуются существенно различными значениями времени задержки и характеристиками деполяризации.

С учетом последнего, а также опубликованных результатов экспериментальных исследований [8], декорреляция при использовании различных видов поляризации для различных парциальных каналов распространения представляется достаточно существенной, что в свою очередь обуславливает перспективность использования в данном случае MIMO с поляризационным и комбинированным (пространственно-поляризационным) разнесением.

При этом может быть обеспечена гибкость системных решений: в качестве передающих и приемных антенных систем могут использоваться в различных сочетаниях одиночные вибраторы, би- и триортогональные излучатели [9, 10], системы из нескольких разнесенных в пространстве подобных излучателей, кольцевые решетки и т.п. В зависимости от назначения и тактических особенностей конкретного комплекса: стационарный для радицентра, стационарный для ограниченной площадки, бортовой, быстро разворачиваемый и т.п. – может быть выбран наиболее приемлемый вариант.

Укрупненная схема фрагмента радиолинии, использующей технологию MIMO, приведена на рис. 1.

Рабочий сигнал от передающей части оконечного модемного оборудования поступает на вход цифровой передающей каналообразующей системы (КОС ПРД), которая по соответствующему алгоритму кодирует информацию для передачи и формирует ансамбль из m

(по числу передатчиков) парциальных сигналов MIMO.



Рис. 1. Укрупненная схема фрагмента радиолинии с MIMO

С выходов КОС ПРД сигналы поступают на входы передатчиков (ПРД), а с их выходов – на входы m -входовой передающей антенной системы (АС ПРД) и излучаются.

Принятые n -входовой приемной антенной системой (АС ПРМ) сигналы поступают на входы приемников (ПРМ), а с их выходов – на входы цифровой приемной каналообразующей системы (КОС ПРМ), которая по соответствующему алгоритму декодирует информацию и восстанавливает рабочий сигнал.

Выбор способа кодирования/декодирования сигналов должен быть темой отдельного рассмотрения.

Антенные системы АС ПРД и АС ПРМ, как было отмечено выше, могут представлять собой комплексы различных антенных устройств, реализующие пространственное, поляризационное или комбинированное разнесение.

Совокупность АС ПРД, АС ПРМ и ансамбля парциальных каналов распространения в дальнейшем имеет смысл рассматривать как некий единый канал MIMO.

Действительно, полагая, что между каждой парой антенн существует парциальный двухлучевой канал (рис. 2), можем рассматривать канал MIMO на рис. 1 как многополюсник.

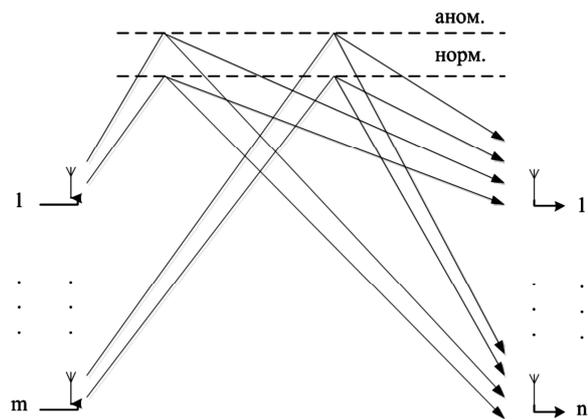
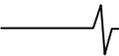


Рис. 2. Парциальные каналы в составе канала MIMO

Лучи в составе канала, вообще говоря, отличаются величиной задержки и видом поляризации благодаря эффектам вращения плоскости поляризации при отражении слоями ионосферы, суперпозиции различных, нормальных и аномальных отраженных волн, суперпозиции волн, пришедших по различным трассам (например, комбинированный тип канала, в котором распространяется земная и пространственная волна) и т.п.

В конечном итоге, в результате такого подхода канал MIMO может быть описан прямоугольной канальной



матрицей \mathbf{H} размером $m \times n$, в которой каждый элемент $h_{m,n}$ представляет собой комплексный коэффициент передачи в парциальном канале.

В случае канала ММО 3x3 с поляризационным разнесением канальную матрицу можно представить в виде [11, 12]:

$$\mathbf{H} = \begin{pmatrix} h^{xx} & h^{xy} & h^{xz} \\ h^{yx} & h^{yy} & h^{yz} \\ h^{zx} & h^{zy} & h^{zz} \end{pmatrix}, \quad (1)$$

где h^{xx}, h^{yy}, h^{zz} – комплексные коэффициенты передачи в каналах с одинаковой поляризацией, а прочие представляют собой коэффициенты передачи в кросс-поляризационных каналах. Обычно последний тип распространения считается паразитным и нежелательным. В идеальном случае канальная матрица должна быть единичной.

После определения канальной матрицы становится возможной следующая оценка качества канала (скорости передачи) ММО [13]:

$$S = \log_2 \det \left(\mathbf{I} + \frac{1}{D_N} \mathbf{Z}_{\text{RX}}^{-1} \mathbf{H} \mathbf{Z}_{\text{TX}} \mathbf{R} \mathbf{Z}_{\text{TX}}^H \mathbf{H}^H (\mathbf{Z}_{\text{RX}}^{-1})^H \right), \quad (2)$$

где D_N – мощность шума, $\mathbf{Z}_{\text{RX}}, \mathbf{Z}_{\text{TX}}$ – взаимный импеданс приемной и передающей антенн соответственно, \mathbf{H} – матрица канальных коэффициентов, \mathbf{R} – корреляционная матрица входного сигнала, символом (H) обозначена операция эрмитова сопряжения.

Скорость передачи в канале ограничивается не только канальной матрицей и отношением сигнал/шум, но и взаимным импедансом приемных и передающих антенн. Как и в случае с канальной матрицей, идеальным было бы отсутствие взаимовлияния в антенных системах. При пространственном виде разнесения сигналов это достигается увеличением расстояния между элементами системы, в случае же поляризационного разнесения необходим учёт влияния опор и подстилающей поверхности, поскольку последние способны значительно исказить взаимный импеданс и, как следствие, ухудшить кросс-поляризационную развязку.

В литературе оценка влияния взаимного импеданса антенн на скорость передачи проведена лишь для случаев с пространственным разделением сигнала. В контексте данной работы её следует дополнить оценкой для поляризационного и смешанного типа разнесения.

С практической точки зрения интересна не только максимальная скорость передачи информации, но и вероятность, с которой канал сможет её обеспечить. Обычно предполагается, что коэффициенты передачи в канале могут быть описаны случайными величинами с гауссовым распределением. Тогда, в предположении о независимости коэффициентов передачи и равенстве коэффициентов передачи в кросс-поляризационных каналах некоторому значению X , можно получить следующую функцию распределения [11]:

$$F(y) = (1 - e^{-y})^3 (1 - e^{-yX})^6. \quad (3)$$

Вероятность снижения реализуемой скорости передачи ниже установленного порогового значения S можно оценить следующим образом:

$$P(C < S) = \int_0^{\eta} W(y) dy, \quad (4)$$

где $W(y)$ – плотность вероятности случайной величины, построенной над матрицей канальных коэффициентов, η – значение коэффициента передачи, при котором достигается S в (2).

В составе АС ПРД и АС ПРМ, в зависимости от назначения оборудования и конкретных требований к нему, могут применяться самые различные антенны и антенные решетки. Однако, применительно к антеннам ДКМВ диапазона, следует отметить два существенных момента. Во-первых, конструкции излучателей и антенных систем на их основе таковы, что, как правило, антенны могут с высокой степенью точности моделироваться как проволочные структуры [14]. Во-вторых, в обязательном порядке должно быть учтено влияние земли, причем как на характеристики направленности антенных систем, так и на импеданс антенн. Последнее соображение особенно существенно при использовании би- и триортогональных излучателей [9, 10], поскольку влияние земли на кроссполяризационные характеристики весьма значительно и его игнорирование неизбежно приведет к грубым ошибкам при расчете $\mathbf{Z}_{\text{RX}}, \mathbf{Z}_{\text{TX}}$ в формуле (2).

Учёт проводников в составе излучающих систем, подстилающей поверхности и опор антенных систем должен выполняться путём электродинамического анализа. Поскольку, как уже было отмечено выше, антенны в данном случае описываются проволочными моделями, в качестве метода электродинамического моделирования предполагается использовать метод, основанный на интегральном уравнении Фредгольма 1-го рода, имеющем вид [14]:

$$E_0(l) = \int_L K(l, l') I(l') dl', \quad (5)$$

где $K(l, l')$ – ядро интегрального уравнения, $I(l, l')$ – ток, а $E_0(l)$ – напряжённость электромагнитного поля.

В качестве контура, на котором локализованы линейные токи, в (5) фигурирует вся совокупность проводников анализируемой системы, т.е. в нашем случае – все проводники, входящие в АС ПРМ или АС ПРД.

Уравнение (5) обычно решается в тонкопроволочном приближении методом Бубнова-Галеркина в кусочно-синусоидальном базисе [14]:

$$b_k(l) = \begin{cases} 0, & l \notin [l_{k-1}, l_{k+1}] \\ \sin(\beta(l - l_{k-1})) / \sin(\beta |l_k - l_{k-1}|), & l_{k-1} \leq l \leq l_k, \\ \sin(\beta(l_{k+1} - l)) / \sin(\beta |l_{k+1} - l_k|), & l_k \leq l \leq l_{k+1} \end{cases} \quad (6)$$

$$k = 1, 2, \dots, N.$$

Решением данного уравнения является распределение напряжённости электрического поля (\mathbf{E}), которое связано с токами в вибраторах (\mathbf{I}) системой уравнений Кирхгофа:

$$\mathbf{E} = \mathbf{Z} \mathbf{I}. \quad (7)$$

Полученный таким образом взаимный импеданс \mathbf{Z} может быть использован для оценки качества канала (2).

Для реализации методики создания современных

комплексов ДКМВ радиосвязи с применением технологии MIMO а также для оценки достижимых характеристик MIMO канала предполагается использовать программный комплекс SAMANT, разработанный в ОАО «Концерн «Автоматика» [15]. Однако, его возможности должны быть расширены с целью учета влияния подстилающей поверхности на характеристики импеданса антенных систем. Существует несколько возможных путей решения указанной проблемы. Первый из них заключается в совершенствовании метода включением некоего дополнительного (эквивалентного) источника, описывающего влияние земли с конечной электропроводностью. Второй путь состоит в учёте электрофизических свойств подстилающей поверхности в виде сеточной структуры в рамках прежней электродинамической модели.

В итоге решения электродинамической задачи можно определить матрицы Z_{RX} , Z_{TX} и \mathbf{H} , с учетом которых могут быть выполнены необходимые расчеты по формулам (2) – (4), и по их результатам осуществлен выбор вариантов для различных комплексов по критерию сложность/эффект.

Применение описанных подходов и методик открывает возможности создания новых, более совершенных комплексов ДКМВ радиосвязи, которые будут использовать технологии MIMO, что позволит увеличить скорость передачи информации, надежность, стойкость и помехозащищенность ДКМВ радиосвязи.

Литература

1. Головин О.В., Простов С.П. Системы и устройства коротковолновой радиосвязи / Под ред. профессора О.В. Головина. – М.: Горячая линия – Телеком, 2006.
2. Минкин М.А. Проблемы и перспективы модернизации и развития систем ДКМВ радиосвязи // Вестник СНИИР. 2006. №4(14). С. 4-10.
3. Бузов А.Л., Сухарев А.С. Вопросы создания универсальных быстроразворачиваемых комплексов технических средств ДКМВ радиосвязи // Вестник СНИИР. – 2006. – №2(12). – С.12.
4. Сухарев А.С. Оценка эффективности применения пространственно-временного кодирования OFDM для сдвоенной передачи данных по ДКМВ радиоканалу // Радиотехника (журнал в журнале). – 2006. – №10.
5. Erceg V.; Soma P.; Baum D.S.; Catreux S., «Multiple-input multiple-output fixed wireless radio channel measurements and modeling using dual-polarized antennas at 2.5 GHz», // *Wireless Communications, IEEE Transactions on*, vol.3, no.6, pp.2288,2298, Nov. 2004
6. Erceg V.; Soma P.; Baum D.S.; Paulraj A.J., «Capacity obtained from multiple-input multiple-output channel measurements in fixed wireless environments at 2.5 GHz», // *Communications, 2002. ICC 2002. IEEE International Conference on*, vol.1, no., pp.396,400, 2002
7. Долуханов М.П. Распространение радиоволн; изд. 4-е. – М: Связь, 1972. – 336 с.
8. Salous S.; Feeney S.M.; Warrington E.M.; Gunashekar S.D.; Abbasi N.M., «Experimental investigations of MIMO in the HF band», // *Ionospheric Radio Systems and Techniques (IRST 2012), 12th IET International Conference on*, vol., no., pp.1,4, 15-17 May 2012
9. Кольчугин И.Ю. Излучающая система кольцевых и многокольцевых антенных решеток на основе биортогональных излучателей // *Радиотехника*. – 2014. – №4. – С. 60-63.
10. Капишев А.Н., Красильников А.Д., Невский А.В. Разработка комплекса активных приемных ДКМВ антенн с управляемыми пространственными и поляризационными характеристиками // *Антенны*. – 2012. – № 6. – С. 57-63.
11. Habib, A, «Multiple polarized MIMO with antenna selection», // *Communications and Vehicular Technology in the Benelux (SCVT), 2011 18th IEEE Symposium on*, vol., no., pp.1,8, 22-23 Nov. – 2011
12. Mansoor Shafi and Min Zhang and Aris L. Moustakas and Peter J. Smith and Andreas F. Molisch and Fredrik Tufvesson and Steven H. Simon, «Polarized MIMO Channels in 3D: Models, Measurements and Mutual Information» // *IEEE J. Select. Areas Commun on*, vol.24, pp. 514, 527, 2006
13. Паршин Ю. Н., Комиссаров А. В. Пропускная способность MIMO телекоммуникационной системы в условиях изменяющейся пространственной структуры радиотракта с искусственной многолучевостью // *Цифровая обработка сигналов*. - 2012. - № 1. - С. 50-55.
14. Бузова М.А., Юдин В.В. Проектирование проволочных антенн на основе интегральных уравнений: Учебное пособие для ВУЗов. – М.: Радио и связь, 2005. – 172 с.
15. Программный комплекс «SAMANT» / ОАО «Концерн «Автоматика». – Свидетельство о государственной регистрации № 2013614026 от 23.04.2013.

PERSPECTIVES OF HF-BAND RADIO COMMUNICATION MIMO-ENABLED COMPLEXES, USING SPATIAL AND POLARISATION SEPARATION

Bukashkin S. A., Ogloblin A. V.

Existing MIMO technologies with spatial and polarization separation are listed. Results are analyzed. Main problems and ways to create HF-band radio communication MIMO-enabled complexes are stated.

АДАПТИВНОЕ ПОДАВЛЕНИЕ ПАССИВНЫХ ПОМЕХ

Попов Д.И., д.т.н., профессор кафедры радиотехнических систем Рязанского государственного радиотехнического университета, e-mail: adop@mail.ru.

Ключевые слова: автокомпенсация, адаптация, адаптивные режекторные фильтры, алгоритмы оценивания, обучающая выборка, пассивные помехи, точность оценивания.

Введение

При выделении радиолокационных сигналов высокой скважности на фоне пассивных помех одной из основных операций является подавление (режектирование) помехи. Эффективная реализация данной операции затрудняется в условиях априорной неопределенности спектрально-корреляционных характеристик помехи, а также их неоднородности в зоне обзора и нестационарности во времени. Преодоление априорной неопределенности на основе методов адаптации приводит к построению адаптивных режекторных фильтров (АРФ) с комплексными весовыми коэффициентами [1, 2], что при реализации данных АРФ в цифровом виде предполагает использование комплексных множителей (цифровых двумерных фазовращателей), число которых пропорционально порядку фильтра. При этом существенно усложняется структура АРФ, особенно высоких порядков, и повышаются требования к быстродействию арифметических операций для выполнения обработки в реальном масштабе времени. Избежать указанных трудностей можно путем предварительной компенсации доплеровского сдвига фазы помехи, обусловленного взаимным перемещением источника мешающих отражений и носителя радиолокатора. В работе [3] синтезированы алгоритмы оценивания и предложены принципы построения и структурные схемы автокомпенсаторов доплеровской фазы пассивных помех. Режектирование «остановленной» помехи теперь может быть осуществлено фильтром с действительными весовыми коэффициентами, адаптирующимися к корреляционным свойствам помехи на выходе автокомпенсатора [4-6]. Представляют интерес синтез оптимального и построение на его основе квазиоптимального (упрощенного) алгоритмов оценивания и измерителей коэффициентов межпериодной корреляции помехи на выходе автокомпенсатора, а также анализ эффективности АРФ в зависимости от выбора алгоритма оценивания и измерителя коэффициентов корреляции помехи, объема обучающей выборки и корреляционных свойств помехи.

Функция правдоподобия

В автокомпенсаторе доплеровской фазы пассивной помехи исходные цифровые отсчеты $U_{jl} = x_{jl} + iy_{jl} = u_{jl} e^{i(j\varphi + \varphi_0)}$ (где j и l – номера периода повторения и элемента разрешения по дальности соответственно, φ –

Синтезирован оптимальный и на его основе построен квазиоптимальный алгоритмы оценивания и измерители коэффициентов межпериодной корреляции помехи на выходе автокомпенсатора доплеровской фазы пассивной помехи. Проведен сравнительный анализ точности данных алгоритмов оценивания и эффективности адаптивного подавления пассивной помехи с использованием оптимальных и квазиоптимальных оценок коэффициентов корреляции помехи.

доплеровский сдвиг фазы помехи за период повторения T , φ_0 – начальная фаза) комплексной огибающей входных данных подвергаются двумерному повороту на угол $-j\hat{\varphi}$ [3]. При этом на выходе автокомпенсатора образуются отсчеты $\tilde{U}_{jl} = U_{jl} e^{-ij\hat{\varphi}} = u_{jl} e^{ij(\varphi - \hat{\varphi})}$, не содержащие с точностью до погрешности автокомпенсации $\Delta\varphi = \varphi - \hat{\varphi}$ доплеровских сдвигов фазы помехи. Затем в АРФ с действительными весовыми коэффициентами m -го порядка производится межпериодная обработка последовательности $m+1$ цифровых отсчетов \tilde{U}_{jl} , $j = \overline{1, m+1}$, следующих через период повторения T и образующих в l -м элементе разрешения по дальности вектор-столбец $\tilde{U}_l = \{\tilde{U}_{jl}\}^T = \{\tilde{U}_{1l}, \dots, \tilde{U}_{m+1,l}\}^T$, $l = \overline{1, n+1}$.

Пассивная помеха, создаваемая отражениями от протяженных объектов, располагается во временном строге и образует в пределах $n+1$ смежных элементов разрешения по дальности обучающую выборку в виде совокупности $\tilde{U} = \{\tilde{U}_l\} = \{\tilde{U}_1, \dots, \tilde{U}_{n+1}\}$. Так как адаптивная обработка на основе поступающих оценок осуществляется после предварительной задержки исходных данных для среднего элемента разрешения в строге [4-6], то целесообразно соответствующий этому элементу вектор \tilde{U}_l ($l = n/2 + 1$) исключить из обучающей выборки. Тогда в случае сигнала, соизмеримого по величине с помехой, или разрывной помехи при обработке элемента разрешения, содержащего сигнал, исключается возможность ослабления или подавления сигнала за счет его влияния на используемые оценки.

Пассивная помеха, как правило, создается множественными отражателями и поэтому является случайным узкополосным процессом гауссовского типа. Помеха в различных элементах разрешения по дальности ввиду полной смены элементарных отражателей является статистически независимой. Заметим, что статистические свойства пассивной помехи сохраняются и после автокомпенсации ее доплеровских сдвигов фазы.

Полагаем помеху в пределах рассматриваемого временного строга однородной. При этом в каждом элементе разрешения по дальности данного строга помеха описы-

вается корреляционной матрицей $\mathbf{R}_l = \overline{\tilde{\mathbf{U}}_l \tilde{\mathbf{U}}_l^{*T}} / 2 = \mathbf{R}$. Параметры корреляционной матрицы \mathbf{R} априори неизвестны и являются предметом оценивания в задачах адаптивного режектирования пассивных помех. Зависимость выходных отсчетов автокомпенсатора в виде совокупности $\tilde{\mathbf{U}} = \{\tilde{\mathbf{U}}_l\} = \{\tilde{\mathbf{U}}_1, \dots, \tilde{\mathbf{U}}_{n+1}\}$ от матрицы \mathbf{R} описывается функцией правдоподобия (ФП)

$$P(\tilde{\mathbf{U}} / \mathbf{R}) = (2\pi)^{-(m+1)n} \det^{-n} \text{Rexp} \left[-\frac{1}{2} \sum_{\substack{l=1 \\ l \neq \frac{n}{2}+1}}^{n+1} \tilde{\mathbf{U}}_l^{*T} \mathbf{R}^{-1} \tilde{\mathbf{U}}_l \right], \quad (1)$$

где \mathbf{R}^{-1} – матрица, обратная \mathbf{R} , элементы которой при симметричном спектре помехи $R_{jk} = \sigma_j \sigma_k \rho_{jk} e^{i(j-k)\Delta\phi} + \sigma_{\text{ш}}^2 \delta_{jk}$; $\sigma_i^2 = \sigma_{\text{ш}}^2$, $\sigma_{\text{ш}}^2$ – дисперсии пассивной помехи и собственного шума приемного устройства; ρ_{jk} – коэффициенты межпериодной корреляции помехи, δ_{jk} – символ Кронекера.

Стационарность структуры матрицы \mathbf{R} , являющейся эрмитовой и теплицевой, позволяет ограничиться оцениванием коэффициентов корреляции ρ_{1k} ($k = \overline{2, m}$), соответствующих первой строке матрицы \mathbf{R} , что при $m \geq 2$ соответствует числу оцениваемых коэффициентов корреляции, необходимых для адаптации весовых коэффициентов АРФ, равному $m-1$ [1, 2]. Для оценивания каждого коэффициента ρ_{1k} необходимы данные двух соответствующих периодов повторения. Тогда образующие совокупность $\tilde{\mathbf{U}} = \{\tilde{\mathbf{U}}_l\}$ векторы $\tilde{\mathbf{U}}_l = \{\tilde{U}_{jl}, \tilde{U}_{j+k-1, l}\}^T$, $j = \overline{1, m-k+1}$, $l = \overline{1, n+1}$. При этом без учета собственного шума приемного устройства ввиду его малости по сравнению с пассивной помехой ($\sigma_{\text{ш}}^2 \ll \sigma_i^2$), что соответствует реальной ситуации, ФП (1) принимает вид

$$P(\tilde{\mathbf{U}} / \mathbf{a}, \psi) = (2\pi \sigma_j \sigma_{j+k-1})^{-2n} (1 - \rho_{1k}^2)^{-n} \times \exp \left\{ \frac{1}{2(1 - \rho_{1k}^2)} \left(\frac{\rho_{1k} (V_{1k} e^{-i\psi} + V_{1k}^* e^{i\psi})}{\sigma_j \sigma_{j+k-1}} - \frac{a_1}{\sigma_j^2} - \frac{a_k}{\sigma_{j+k-1}^2} \right) \right\},$$

где $\mathbf{a} = \{\rho_{1k}, \sigma_j^2, \sigma_{j+k-1}^2\}$ – вектор искомых параметров помехи, $\psi = (k-1)\Delta\phi$,

$$V_{1k} = \sum_{\substack{l=1 \\ l \neq \frac{n}{2}+1}}^n \tilde{U}_{jl}^* \tilde{U}_{j+k-1, l},$$

$$a_1 = \sum_{\substack{l=1 \\ l \neq \frac{n}{2}+1}}^n |\tilde{U}_{jl}|^2, \quad a_k = \sum_{\substack{l=1 \\ l \neq \frac{n}{2}+1}}^n |\tilde{U}_{j+k-1, l}|^2.$$

Фазовый сдвиг ψ в данном случае является неинформационным (мешающим) параметром, который подлежит исключению из ФП. Классическое решение этой проблемы состоит в усреднении ФП по ψ . Другое решение основывается на максимизации ФП по ψ . Учитывая, что

$$V_{1k} e^{-i\psi} + V_{1k}^* e^{i\psi} = 2 |V_{1k}| \cos(\psi - \arg V_{1k}),$$

то при любом \mathbf{a} максимуму ФП по ψ соответствует $\cos(\psi - \arg V_{1k}) = 1$. При этом условии параметр ψ исключается из ФП:

$$P(\tilde{\mathbf{U}} / \mathbf{a}) = (2\pi \sigma_j \sigma_{j+k-1})^{-2n} (1 - \rho_{1k}^2)^{-n} \times \exp \left\{ \frac{1}{1 - \rho_{1k}^2} \left(\frac{\rho_{1k} |V_{1k}|}{\sigma_j \sigma_{j+k-1}} - \frac{a_1}{2\sigma_j^2} - \frac{a_k}{2\sigma_{j+k-1}^2} \right) \right\}.$$

Заметим, что при усреднении ФП по ψ , приводящем к модифицированной функции Бесселя нулевого порядка, с учетом свойств этой функции при больших значениях аргумента достигается аналогичный результат.

Синтез алгоритмов оценивания

При оценивании векторного параметра \mathbf{a} уравнение правдоподобия имеет вид

$$\nabla_{\mathbf{a}} \ln P(\tilde{\mathbf{U}} / \mathbf{a}) \Big|_{\mathbf{a}=\hat{\mathbf{a}}} = 0,$$

где $\nabla_{\mathbf{a}} = \left\{ \frac{\partial}{\partial \rho_{1k}}, \frac{\partial}{\partial (\sigma_j^2)}, \frac{\partial}{\partial (\sigma_{j+k-1}^2)} \right\}$ – оператор градиента, соответствующий вычислению частных производных по компонентам вектора \mathbf{a} .

В результате логарифмирования, дифференцирования и несложных алгебраических преобразований получим

$$\left. \begin{aligned} \frac{\partial \ln P}{\partial \rho_{1k}} &= 2n(1 - \rho_{1k}^2) + \frac{(1 + \rho_{1k}^2) |V_{1k}|}{\rho_{1k} \sigma_j \sigma_{j+k-1}} - \frac{a_1}{\sigma_j^2} - \frac{a_k}{\sigma_{j+k-1}^2} = 0, \\ \frac{\partial \ln P}{\partial (\sigma_j^2)} &= 2n(1 - \rho_{1k}^2) + \frac{\rho_{1k} |V_{1k}|}{\sigma_j \sigma_{j+k-1}} - \frac{a_1}{\sigma_j^2} = 0, \\ \frac{\partial \ln P}{\partial (\sigma_{j+k-1}^2)} &= 2n(1 - \rho_{1k}^2) + \frac{\rho_{1k} |V_{1k}^{(j)}|}{\sigma_j \sigma_{j+k-1}} - \frac{a_k}{\sigma_{j+k-1}^2} = 0. \end{aligned} \right\}$$

Данная система уравнений позволяет найти искомые оценки максимального правдоподобия (ОМП). Из совместного решения уравнений системы находим

$$\hat{\rho}_{1k} = |V_{1k}| / 2n \sigma_j \sigma_{j+k-1}, \quad k = \overline{2, m}.$$

Решение второго и третьего уравнений системы с учетом полученного выражения для $\hat{\rho}_{1k}$ приводит соответственно к оценкам

$$\hat{\sigma}_j^2 = a_1 / 2n, \quad \hat{\sigma}_{j+k-1}^2 = a_k / 2n.$$

Окончательно для ОМП коэффициентов корреляции имеем

$$\hat{\rho}_{1k} = |V_{1k}| / (a_1 a_k)^{1/2}, \quad k = \overline{2, m} \quad (2)$$

Алгоритм оценивания (2) является оптимальным. Соответствующая структурная схема оптимального измерителя оценки $\hat{\rho}_{12}$ приведена на рис. 1. В блоке комплексного сопряжения (*) инвертируется знак мнимых проекций задержанных в запоминающем устройстве ЗУ_T на период повторения T комплексных отсчетов \tilde{U}_{jl} , которые затем перемножаются с исходными отсчетами $\tilde{U}_{j+1, l}$ в комплексном перемножителе ($\dot{\times}$). В первом и втором накопителях (H_1 и H_2) осуществляется скольз-

щее суммирование (накопление) поступающих данных с $n+1$ смежных элементов разрешения по дальности за исключением среднего элемента с номером $l = n/2 + 1$. Накопитель H_1 состоит из двух квадратурных накопителей, каждый из которых выполнен аналогично накопителю H_2 . В блоках объединения (БО) производится суммирование квадратов проекций, а последующие операции в соответствии с алгоритмом (2) приводят к вычислению на выходе делителя (Д) оценки $\hat{\rho}_{12}$.

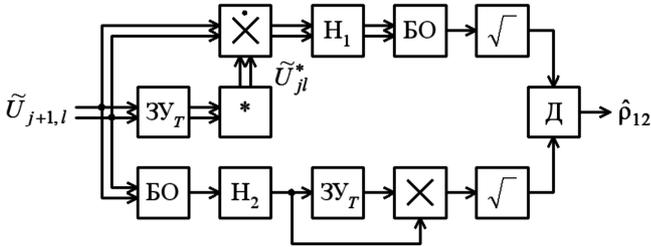


Рис. 1. Структурная схема оптимального измерителя коэффициента корреляции помехи

Представим величину V_{1k} с учетом погрешности автокомпенсации $\Delta\varphi = \varphi - \hat{\varphi}$ в виде

$$V_{1k} = |V_{1k}| e^{i(k-1)\Delta\varphi} = \text{Re}V_{1k} + i \text{Im}V_{1k} = |V_{1k}| \cos[(k-1)\Delta\varphi] + i |V_{1k}| \sin[(k-1)\Delta\varphi]$$

При малой величине погрешности $\Delta\varphi = \varphi - \hat{\varphi}$ можно упростить алгоритм (2), ограничиваясь вычислением действительной проекции $\text{Re}V_{1k}$ и пренебрегая мнимой проекцией $\text{Im}V_{1k} \cong 0$, полагая $|V_{1k}| = \sqrt{(\text{Re}V_{1k})^2 + (\text{Im}V_{1k})^2} \cong \text{Re}V_{1k}$. Квазиоптимальный (упрощенный) алгоритм оценивания принимает вид

$$\hat{r}_{1k} = \text{Re}V_{1k} / (a_1 a_k)^{1/2}, \quad k = \overline{2, m}, \quad (3)$$

где $\text{Re}V_{1k} = \sum_{l \neq \frac{n}{2}+1}^n (x_{jl} x_{j+k-1, l} + y_{jl} y_{j+k-1, l})$. Структурная схема

квазиоптимального измерителя оценки \hat{r}_{12} в соответствии с алгоритмом (3) изображена на рис. 2.

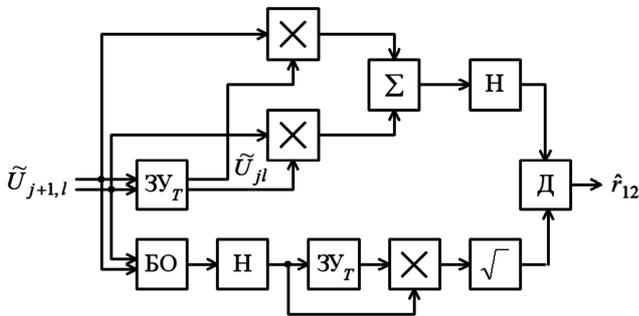


Рис. 2. Структурная схема квазиоптимального измерителя коэффициента корреляции помехи

Как видим, по сравнению с оптимальным измерителем на рис. 1 сократилось число скалярных перемножителей, по одному блоку объединения и блоку извлечения квадратного корня, а также упростилось выполнение двухканального накопителя. Измерители оценок \hat{r}_{1k} и

\hat{r}_{1k} отличаются от рассмотренных измерителей величиной задержки в ЗУ, равной $(k-1)T$.

Анализ алгоритмов оценивания

Точность оценивания зависит от корреляционных свойств выходных отсчетов \tilde{U}_j автокомпенсатора, описываемых при $\sigma_{ij}^2 = \sigma_{nk}^2 = \sigma_n^2$ коэффициентами корреляции

$$\tilde{\rho}_{jk} = \overline{\tilde{U}_j \tilde{U}_k^*} / 2\sigma_n^2 = \overline{U_j U_k^* \exp[-i(j-k)(\hat{\varphi} - \varphi)]} / 2\sigma_n^2.$$

Так как в автокомпенсаторе при определении оценки $\hat{\varphi}$ осуществляется суммирование независимых данных с n элементов разрешения по дальности, то взаимная корреляция оценки $\hat{\varphi}$ и отсчетов U_j практически отсутствует. Тогда, учитывая нормальный закон двумерного распределения отсчетов U_j, U_k и асимптотическую нормальность распределения оценки $\hat{\varphi}$ со средним φ и дисперсией $\sigma_{\hat{\varphi}}^2$, для коэффициентов корреляции отсчетов на выходе автокомпенсатора получаем

$$\tilde{\rho}_{jk} = \rho_{jk} \exp[-i(j-k)(\hat{\varphi} - \varphi)] + \lambda \delta_{jk} = \rho_{jk} \exp[-(j-k)^2 \sigma_{\hat{\varphi}}^2 / 2] + \lambda \delta_{jk},$$

где $\lambda = \sigma_n^2 / \sigma_{\hat{\varphi}}^2$ – отношение шум/помеха, $\sigma_{\hat{\varphi}}^2$ – дисперсия оценки $\hat{\varphi}$, определяемая в зависимости от типа автокомпенсатора соответствующим выражением работы [3].

Найдем характеризующую точность оценивания дисперсию $\sigma_{\hat{\rho}}^2$ оптимальной оценки $\hat{\rho} = \hat{\rho}_{12}$. Ввиду асимптотической эффективности ОМП коэффициента корреляции $\hat{\rho}$ для этой цели используем выражение Крамера–Рао, в соответствии с которым:

$$\sigma_{\hat{\rho}}^2 = - \left[\partial^2 \ln P(\{\tilde{U}_{1l}, \tilde{U}_{2l}\} / \tilde{\rho}) / \partial \tilde{\rho}^2 \right]^{-1}, \quad (4)$$

где

$$P(\{\tilde{U}_{1l}, \tilde{U}_{2l}\} / \tilde{\rho}) = (2\pi)^{-2n} \det^{-n} \|\tilde{\rho}_{jk}\| \exp \left\{ -\frac{1}{2} \sum_{l=1}^n \sum_{j,k=1}^2 w_{jk} \tilde{U}_{jl}^* \tilde{U}_{kl} \right\} \quad (5)$$

функция правдоподобия, $\tilde{\rho} = \tilde{\rho}_{12}$, w_{jk} – элементы матрицы $\|\tilde{\rho}_{jk}\|$, обратной матрице $\|\tilde{\rho}_{jk}\|$.

После соответствующих вычислений в (4) получим

$$\sigma_{\hat{\rho}}^2 = \frac{[(1+\lambda)^2 - \tilde{\rho}^2]^2}{2n[(1+\lambda)^2 + \tilde{\rho}^2]} = \frac{[(1+\lambda)^2 - \rho^2 \exp(-\sigma_{\hat{\varphi}}^2)]^2}{2n[(1+\lambda)^2 + \rho^2 \exp(-\sigma_{\hat{\varphi}}^2)]}. \quad (6)$$

Как видим, дисперсия $\sigma_{\hat{\rho}}^2$ зависит от точности компенсации доплеровского сдвига фазы помехи, характеризующейся величиной $\sigma_{\hat{\varphi}}^2$, что приводит к снижению точности оценивания коэффициента корреляции $\tilde{\rho}$.

Рассмотрим погрешности квазиоптимальной оценки $\hat{r}_{12} = \hat{r}$, которая является действительной проекцией оптимальной оценки комплексного коэффициента корреляции, т. е. $\hat{r} = \hat{\rho} \cos \Delta\varphi$. Оценка $\hat{\rho}$ и угол $\Delta\varphi$ являются случайными асимптотически нормальными случайными величинами

нами со средними значениями соответственно $\tilde{\rho} = r$ и 0. Учитывая вытекающую из свойств огибающей и фазы узкополосного случайного процесса статистическую независимость величин $\hat{\rho}$ и $\Delta\varphi$, найдем математическое ожидание оценки \hat{r} :

$$\overline{\hat{r}} = \overline{\hat{\rho} \cos \Delta\varphi} = \overline{\hat{\rho}} \cdot \overline{\cos \Delta\varphi} = r \exp(-\sigma_{\Delta\varphi}^2 / 2).$$

Так как угол $\Delta\varphi = \varphi - \hat{\varphi}$ линейно связан с оценкой $\hat{\varphi}$, определяемой по совокупности отсчетов $\{\tilde{U}_{jl}, \tilde{U}_{j+1,l}\}$, $l = \overline{1, n+1}$, то его дисперсия может быть найдена по соответствующей функции правдоподобия с помощью выражения Крамера–Рао в аналогичном работе [3] виде

$$\sigma_{\Delta\varphi}^2 = \frac{(1 + \lambda)^2 - r^2}{2nr^2}.$$

Таким образом, квазиоптимальная по алгоритму (3) оценка имеет среднее смещение

$$\begin{aligned} \Delta r = \overline{\hat{r}} - r &= \overline{\hat{r}} - r = r \left[\exp\left(-\frac{\sigma_{\Delta\varphi}^2}{2}\right) - 1 \right] = \\ &= r \left[\exp\left(-\frac{(1 + \lambda)^2 - r^2}{4nr^2}\right) - 1 \right]. \end{aligned} \quad (7)$$

Из выражения (7) следует, что при $n \rightarrow \infty$ $\Delta r \rightarrow 0$ и, следовательно, оценка \hat{r} является асимптотически несмещенной, однако при конечной величине n смещение Δr отлично от нуля.

Дисперсию $\sigma_{\hat{r}}^2$ квазиоптимальной оценки \hat{r} найдем с помощью выражения Крамера–Рао для смещенной оценки

$$\sigma_{\hat{r}}^2 = -(1 + \partial\Delta r / \partial r)^2 \left[\partial^2 \ln P(\{\tilde{U}_{1l}, \tilde{U}_{2l}\}/r) / \partial r^2 \right]^{-1}, \quad (8)$$

где $P(\{\tilde{U}_{1l}, \tilde{U}_{2l}\}/r)$ – функция правдоподобия, аналогичная (5) при $\tilde{\rho} = r$.

В результате соответствующих вычислений в (8) получим

$$\sigma_{\hat{r}}^2 = \sigma_{\hat{\rho}}^2 \exp\left(-\frac{(1 + \lambda)^2 - r^2}{2nr^2}\right) \left(1 + \frac{(1 + \lambda)^2}{2nr^2}\right)^2,$$

где $\sigma_{\hat{\rho}}^2$ – дисперсия оптимальной оценки $\hat{\rho}$, определяемая в соответствии с выражением (6).

Суммарное рассеяние оценки \hat{r} относительно истинного значения r характеризуется величиной

$$d_{\hat{r}}^2 = \sigma_{\hat{r}}^2 + (\Delta r)^2. \quad (9)$$

Сравним точность оценивания по оптимальному (2) и квазиоптимальному (3) алгоритмам. На рис. 3 приведены зависимости $\sigma_{\hat{\rho}}$ (кривые 1) и $d_{\hat{r}}$ (кривые 2) от объема обучающей выборки n для гауссовской функции корреляции помехи

$$\rho_{jk} = \exp\{-[\pi\beta_n(j-k)]^2 / 2,8\}$$

при $\beta_n = 0,1$ и $\lambda \leq 10^{-4}$, где $\beta_n = \Delta f_n T$ – нормированная ширина спектра помехи.

Сплошные кривые соответствуют теоретическим расчетам на основе выражений (6) и (9), штриховые – результа-

там моделирования алгоритмов оценивания (2) и (3) на ЭВМ. Из кривых следует, что упрощение оптимального алгоритма оценивания (2) и соответствующего ему измерителя приводит к незначительным и убывающим с ростом n потерям в точности оценивания. Достаточно близкое совпадение расчетных кривых с результатами моделирования при $n > 4$ подтверждает правомерность теоретических результатов и свидетельствует об асимптотической эффективности получаемых оценок.

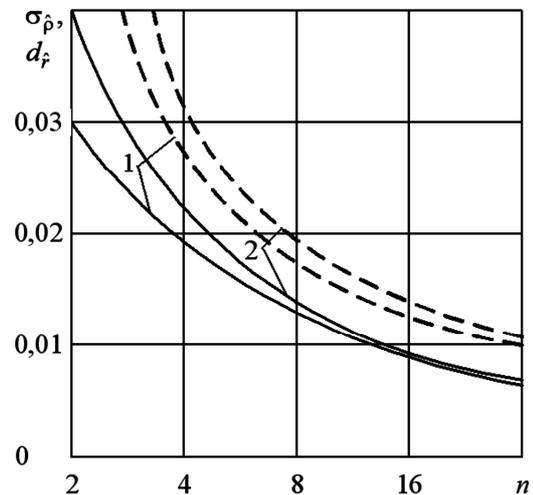


Рис. 3. Зависимости точности оценивания от объема обучающей выборки

Анализ эффективности АРФ

Режектирование «остановленной» помехи осуществляется фильтром с действительными весовыми коэффициентами \hat{g}_k , адаптирующимися к оценкам коэффициентов межпериодной корреляции в соответствии с алгоритмами работы [2]. При использовании оптимальной оценки $\hat{\rho}$ для РФ 2-го порядка ($m = 2$) имеем

$$\hat{g}_0 = \hat{g}_2 = g_0 = g_2 = 1, \quad \hat{g}_1 = -2\hat{\rho}.$$

Для РФ 3-го порядка необходимо использовать априорную информацию о форме огибающей функции корреляции помехи. Например, для гауссовской функции ($\hat{\rho}_{1k} = \hat{\rho}^{(k-1)^2}$) при $m = 3$ получаем

$$\hat{g}_0 = -\hat{g}_3 = g_0 = -g_3 = 1, \quad \hat{g}_1 = -\hat{g}_2 = -(\hat{\rho} + \hat{\rho}^2 + \hat{\rho}^3).$$

Вычисление оценок коэффициентов корреляции на основе усреднения независимых данных с n элементов разрешения приводит к отсутствию взаимной корреляции отсчетов \tilde{U}_j и получаемых оценок и, следовательно, отсчетов \tilde{U}_j и весовых коэффициентов \hat{g}_k , что упрощает последующий анализ АРФ.

Представляя выходную величину нерекурсивного АРФ m -го порядка в виде

$$Z = \sum_{k=0}^m \hat{g}_k \tilde{U}_{m-k+1},$$

для подавления помехи в АРФ найдем

$$\left(\frac{\sigma_{\hat{\rho}}^2}{\sigma^2}\right)_n = \frac{\overline{ZZ^*}}{2\sigma_n^2} = \sum_{j,k=0}^m \overline{\hat{g}_j \hat{g}_k} \tilde{\rho}_{jk} =$$

$$= \sum_{j,k=0}^m \hat{g}_j \hat{g}_k \rho_{jk} \exp[-(j-k)^2 \sigma_{\hat{\rho}}^2 / 2] + \lambda \sum_{j=0}^m \overline{\hat{g}_j^2}. \quad (10)$$

При известной форме огибающей функции корреляции помехи достаточно оценивать величину $\hat{\rho}_{12} = \hat{\rho}$. Весовые коэффициенты АРФ в этом случае определяются в результате функциональных преобразований $\hat{g}_k = f_k(\hat{\rho})$, $k = \overline{1, m-1}$. Используя асимптотические свойства ОМП коэффициента корреляции $\hat{\rho}$, произведем соответствующие усреднения в (10). Для этого рассмотрим линейную аппроксимацию зависимостей $\hat{g}_k = f_k(\hat{\rho})$ в окрестности $\bar{\rho} = \tilde{\rho}$ в виде

$$\hat{g}_k = g_k + (\hat{\rho} - \tilde{\rho}) g'_k,$$

где $g_k = f_k(\tilde{\rho})$, $g'_k = f'_k(\tilde{\rho}) = \partial f_k(\tilde{\rho}) / \partial \tilde{\rho}$.

В частности, при $m=2$ имеем $g_0 = g_2 = 1$, $g_1 = -2\tilde{\rho}$ и $g'_1 = -2$, а при $m=3$ — $g_0 = -g_3 = 1$, $g_1 = -g_2 = -(\tilde{\rho} + \tilde{\rho}^2 + \tilde{\rho}^3)$ и $g'_1 = -g'_2 = -(1 + 2\tilde{\rho} + 3\tilde{\rho}^2)$.

Учитывая асимптотическую нормальность распределения ОМП коэффициента корреляции $\hat{\rho}$ со средним $\tilde{\rho}$ и дисперсией $\sigma_{\hat{\rho}}^2$, найдем

$$\overline{\hat{g}_j \hat{g}_k} = [g_j + (\hat{\rho} - \tilde{\rho}) g'_j][g_k + (\hat{\rho} - \tilde{\rho}) g'_k] = g_j g_k + \sigma_{\hat{\rho}}^2 g'_j g'_k. \quad (11)$$

Выражение (10) с учетом соотношения (11) теперь принимает вид

$$\left(\frac{\sigma_{\hat{\rho}}^2}{\sigma^2} \right)_{\Pi} = \sum_{j,k=0}^m g_j g_k \rho_{jk} \exp[-(j-k)^2 \sigma_{\hat{\rho}}^2 / 2] + \lambda \sum_{j=0}^m g_j^2 + \sigma_{\hat{\rho}}^2 \left\{ \sum_{j,k=1}^{m-1} g'_j g'_k \rho_{jk} \exp[-(j-k)^2 \sigma_{\hat{\rho}}^2 / 2] + \lambda \sum_{j=1}^{m-1} g_j'^2 \right\}. \quad (12)$$

Выражение (12) определяет эффективность подавления пассивной помехи в АРФ в зависимости от ее корреляционных свойств, а также от входящих в выражение (6) для дисперсии $\sigma_{\hat{\rho}}^2$ объема обучающей выборки n и погрешностей автокомпенсации доплеровской фазы помехи $\sigma_{\hat{\rho}}^2$ на входе АРФ.

Анализ АРФ с квазиоптимальными измерителями коэффициентов корреляции помехи в значительной степени основывается на проведенном анализе АРФ с оптимальными измерителями. Адаптация весовых коэффициентов АРФ g_k при известной форме корреляционной функции помехи осуществляется по оценке \hat{r} в соответствии с адаптивными алгоритмами [2], реализующими функциональные преобразования $\hat{g}_k = f_k(\hat{r})$, $k = \overline{1, m-1}$. Используя линейную аппроксимацию данных преобразований в окрестности r в виде $\hat{g}_k = g_k + (\hat{r} - r) g'_k$, где $g_k = f_k(r)$, $g'_k = f'_k(r) = \partial f_k(r) / \partial r$, и учитывая асимптотическую нормальность распределения оценки \hat{r} , после соответствующих вычислений для подавления помехи в АРФ получаем аналогичное (12) соотношение, в котором вместо $\sigma_{\hat{\rho}}^2$ используется соответствующая выражению (9) величина d_r^2 ,

учитывающая помимо объема обучающей выборки и погрешностей автокомпенсации ошибки, обусловленные смещенным характером оценки \hat{r} .

В целом эффективность АРФ характеризуется усредненным по доплеровской фазе сигнала коэффициентом улучшения отношения сигнал/помеха [2], который с учетом ошибок адаптации в соответствии с выражением (12) или аналогичным для АРФ с квазиоптимальными измерителями коэффициентов корреляции помехи имеет вид

$$\mu = \sum_{j=0}^m g_j^2 / \left(\frac{\sigma_{\hat{\rho}}^2}{\sigma^2} \right)_{\Pi}.$$

Используя полученные соотношения, рассмотрим теперь обусловленные погрешностями оценивания неизвестных параметров помехи потери $\Delta\mu$ по отношению к предельной эффективности АРФ. На рис. 4 приведены зависимости потерь от объема обучающей выборки n для АРФ 2-го и 3-го порядков ($m=2$ и 3), гауссовской функции корреляции помехи, $\beta_{\Pi} = 0,1$ и $\lambda \leq 10^{-6}$. Кривые 1 соответствуют АРФ с оптимальными измерителями, а кривые 2 — АРФ с квазиоптимальными измерителями. Как видим, использование квазиоптимальных измерителей приводит к дополнительному увеличению общих потерь, обусловленных ошибками адаптации. Однако величина дополнительных потерь достаточно мала и не превосходит 0,5 дБ при $n > 4$. Малость дополнительных потерь в сочетании с аппаратным упрощением измерителей, число которых при неизвестной форме корреляционной функции должно быть равно $m-1$ [2], позволяют сделать вывод о технико-экономической целесообразности использования в адаптивных режекторных фильтрах рассматриваемых классов квазиоптимальных измерителей коэффициентов корреляции пассивной помехи.

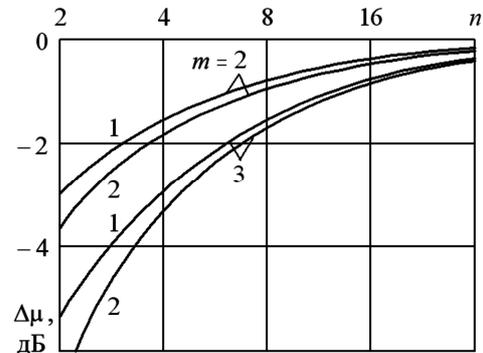


Рис. 4. Зависимости потерь в эффективности АРФ от объема обучающей выборки

Заключение

Синтезированный алгоритм оценивания коэффициентов межпериодной корреляции пассивной помехи на выходе автокомпенсатора доплеровской фазы помехи допускает построение на его основе квазиоптимального (упрощенного) алгоритма оценивания и соответствующего ему измерителя.

Сравнительный анализ точности оптимального и квазиоптимального алгоритмов оценивания показал, что упрощение оптимального алгоритма приводит к незначительным и убывающим с ростом объема обучающей

выборки потерям в точности оценивания. Правомерность данного вывода подтверждают результаты моделирования на ЭВМ алгоритмов оценивания.

Предложенный метод анализа АРФ с действительными весовыми коэффициентами позволяет учитывать корреляционные свойства пассивной помехи, тип используемого (оптимального или квазиоптимального) измерителя коэффициентов корреляции помехи и погрешности автокомпенсации доплеровской фазы помехи и адаптации весовых коэффициентов АРФ, обусловленные конечным объемом обучающей выборки.

Анализ потерь в эффективности АРФ, обусловленных ошибками автокомпенсации и адаптации, показал технико-экономическую целесообразность использования в АРФ квазиоптимальных измерителей коэффициентов корреляции помехи.

Литература

1. А. с. 934816 СССР, МПК6 G 01 S 7/36, G 01 S 13/52. Режекторный фильтр / Д.И. Попов; опубл. 27.11.1998. Бюл. № 33. – 20 с.
2. Попов Д.И. Адаптация нерекурсивных режекторных фильтров // Известия вузов. Радиоэлектроника. – 2009. – Т. 52, № 4. – С. 46-55.

3. Попов Д.И. Автокомпенсация доплеровской фазы пассивных помех // Цифровая обработка сигналов. – 2009. – № 2. – С. 30-33.

4. А. с. 875960 СССР, МПК6 G 01 S 7/36, G 01 S 13/52. Устройство для подавления пассивных помех / Д.И. Попов; опубл. 27.11.1998, Бюл. № 33. – 11 с.

5. А. с. 1015757 СССР, МПК6 G 01 S 7/36. Устройство подавления пассивных помех / Д.И. Попов; опубл. 27.11.1998, Бюл. № 33. – 12 с.

6. А. с. 1098399 СССР, МПК6 G 01 S 7/36. Устройство адаптивной режекции пассивных помех / Д.И. Попов; опубл. 20.12.1998, Бюл. № 35. – 16 с.

ADAPTIVE CLUTTER SUPPRESSION

Popov D.I.

The optimum estimation algorithm and meter of clutter interperiod correlation coefficients on output auto-compensator of clutter dopler phase are synthesized and on their basis quasi-optimum estimation algorithms and meter are constructed. The comparative analysis of the given estimation algorithms accuracy and of adaptive suppression efficiency of clutter with use optimum and quasi-optimum clutter correlation coefficients estimator is carried out.

Рязанский государственный радиотехнический университет проводит

2-й Тренинг «МНОГОЯДЕРНЫЕ ЦИФРОВЫЕ СИГНАЛЬНЫЕ ПРОЦЕССОРЫ TMS320C66xx ФИРМЫ TEXAS INSTRUMENTS» в апреле 2015 года

Цель Тренинга: приобретение навыков практической работы с платформой Keystone I многоядерных сигнальных процессоров TMS320C66xx фирмы Texas Instruments, включая создание и конфигурацию проектов в среде Code Composer Studio v6, разработку и оптимизацию программных кодов на языках низкого и высокого уровней, организацию параллельного выполнения программ на нескольких ядрах с использованием инструментариев OpenMP и IPC, программирование контроллера обработки и передачи данных в многоядерной системе Multicore Navigator с задействованием механизмов OpenEM.



Программа Тренинга, описание мероприятия и условия участия изложены в информационном письме на сайте: www.dsps.ru/workshops/ws.php.

Ждем заявок на участие!

ОЦЕНКА ЗОНЫ НЕОПРЕДЕЛЕННОСТИ РАДИООТРАЖЕНИЙ В ЗАДАЧЕ СОВМЕЩЕНИЯ ИЗОБРАЖЕНИЙ С КАРТОЙ МЕСТНОСТИ

*Новиков А.И., к.э.н., доцент Рязанского государственного радиотехнического университета,
e-mail: novikovanatoly@yandex.ru;*

*Юкин С.А., к.т.н., доцент Рязанского государственного радиотехнического университета,
e-mail: yukin-s@yandex.ru.*

Ключевые слова: совмещение, радиолокационные изображения подстилающей поверхности, зона неопределенности.

Введение

Современные летательные аппараты (ЛА) оснащены разнообразными бортовыми системами технического зрения (СТЗ), работающими не только в различных спектральных диапазонах (тепло- и телекамеры), но и на различных принципах (лидар, телекамера, радиолокатор). В связи с многообразием сенсоров стоит задача объединения формируемой информации – приведение ее к сравнимому виду для дальнейшего совмещения. Подобные задачи совмещения возникают в различных областях техники [1]. При этом возникают трудности, обусловленные различием в системах координат, относительно которых формируются изображения от различных СТЗ, а также отсутствие возможности распознавать объекты на этих изображениях по прямым признакам.

Невозможность совмещения разнородных изображений по прямым признакам (например, по форме), приводит к поиску косвенных методов, основанных на привязке изображений от различных СТЗ к цифровой карте местности (ЦКМ). Для этого необходимо синхронизировать получаемые изображения по пространственному положению, т.е. провести пространственное совмещение всех обрабатываемых изображений, причем совмещение должно выполняться автоматически и в режиме реального времени. Подобные подходы применяются при выявлении изменений на однородных изображениях (в задачах взаимной компенсации с целью выявления расхождений – обнаружения новых объектов) [2].

Формируемые по маршруту движения ЛА изображения невозможно сразу совмещать с картой, это связано с ошибками как навигационной системы ЛА так и с ошибками датчиков, определяющих пространственное положение СТЗ (например, ракурс). В этом случае для автоматического совмещения, помимо прочего, необходимо провести коррекцию возникающих геометрических искажений, вызванных неточностью измерений навигационной системы. При совмещении изображений необходимо учитывать жесткие требования к точности совмещения разнородных изображений, изложенные в [3].

Один из наиболее надежных и апробированных способов совмещения радиолокационных изображений

Предложена математическая модель зоны неопределенности на подстилающей поверхности, соответствующей зоне обзора бортовой радиолокационной станции, работающей в передней полусфере. Приведены результаты исследования модели для различных вариантов ошибок датчиков навигационных параметров. Показано, что для задач совмещения картографической информации с формируемым радиолокационным изображением необходимо расширить зону поиска с учетом зоны неопределенности в зависимости от ракурса обзора, приводящего к несимметричным искажениям (расширениям предполагаемой зоны обзора) относительно центра сектора сканирования.

(РЛИ) и изображений, синтезированных по ЦКМ, заключается в переборе всех возможных положений РЛИ относительно ЦКМ. Для решения этой задачи необходимо иметь оценки возможного пространственного положения сектора обзора РЛС на подстилающей поверхности. Теоретически таких положений – неограниченное множество, но фактически диапазон этих положений определяется точностью навигационной системы и датчиков СТЗ (для РЛС – точностью ориентации диаграммы направленности антенны в пространстве).

В [4] выполнены исследования зоны обзора в плоскости Земли для видеоизображений, получаемых на борту ЛА. Принципы формирования видеоизображения и РЛИ имеют существенные отличия, что, в конечном итоге, проявляется как в форме, так и в параметрах зон обзора в плоскости Земли для этих видов сенсоров СТЗ. Вместе с тем, некоторые элементы математических моделей процессов формирования изображений этими способами имеют общие основания. В частности, это относится к учету углов позиционирования ЛА в пространстве и ошибок по ним навигационной системы.

Цель работы – нахождение возможной (предельной) зоны обзора РЛС на подстилающей поверхности и ее геометрических искажений, обусловленных ошибками навигационной системы ЛА и датчиков СТЗ.

Постановка задачи

Необходимо построить и исследовать математические модели РЛИ для различных условий работы БРЛС, а именно:

- при отсутствии ошибок навигационной системы;
- при наличии угловых ошибок навигационной системы:
 - а) без учета стробирования по дальности;
 - б) с учетом стробирования;
- с учетом всех видов ошибок.

Для построения и исследования математической модели зоны обзора радиолокационного изображения в плоскости Земли необходимо определить набор параметров БРЛС и их модельные значения.

Параметры носителя РЛС:

h – высота носителя БРЛС, м, (в модельных расчетах $h = 1000$ м).

Сектор обзора подстилающей поверхности:

β_v – угол наклона верхней части ДНА по углу места (вверх (+), вниз (-)), град, (в модельных расчетах $\beta_v = 5^\circ$);

β_n – угол наклона нижней части ДНА по углу места (вверх (+), вниз (-)), град, (в модельных расчетах $\beta_n = 35^\circ$);

α – ширина сектора сканирования по азимуту, град, (в модельных расчетах $\alpha = 40^\circ$);

α_0 – угол поворота сектора сканирования по азимуту относительно оси летательного аппарата (ЛА), град.

Во введенных обозначениях β_v и β_n определяют сектор обзора РЛС в угломестной плоскости, если обзор ведется одной строкой, то соответствующие величины находятся исходя из ширины ДНА и ее наклона относительно строительной оси ЛА.

Параметры подстилающей поверхности: в первом приближении считаем, что поверхность ровная.

Параметры навигационной системы (рис. 1):

$\Delta\theta = 0,1^\circ$ – точность датчиков по тангажу, град;

$\Delta\gamma = 0,1^\circ$ – точность датчиков по крену, град;

$\Delta\psi = 0,1^\circ$ – точность датчиков по углу рысканья, град;

$\Delta h = 20$ – точность датчиков высоты, м;

$\Delta x = 50$ – точность определения координат ЛА по оси Ox ;

$\Delta y = 50$ – точность определения координат ЛА по оси Oy ;

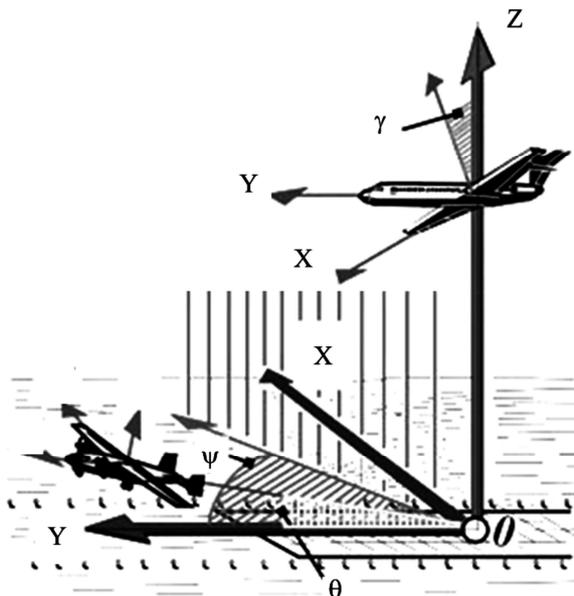


Рис. 1. Пояснения к введенным обозначениям

Параметры антенной системы:

$\Delta\alpha_{datch} = 0,02^\circ$ – точность электронного позиционирования ДНА, град;

$\Delta pr = 0,2^\circ$ – точность позиционирования антенной системы за счет приводов, град.

1. Модель РЛИ при отсутствии ошибок навигационной системы

Примем следующие допущения:

– дальность действия РЛС не больше максимальной наклонной дальности, которую определим как расстояние от носителя РЛС (ЛА) до подстилающей поверхности при угле наклона верхней части луча ДНА $\beta_v = 5^\circ$; обозначим эту величину R_{max} (рис. 2);

– минимально анализируемая дальность – расстояние от носителя РЛС (ЛА) до подстилающей поверхности при угле наклона нижней части луча ДНА $\beta_n = 35^\circ$; обозначим эту величину R_{min} .

Используя отношения между элементами на рис. 2, получаем:

$$R_{max} = h / \sin \beta_v, \quad R_{min} = h / \sin \beta_n; \tag{1}$$

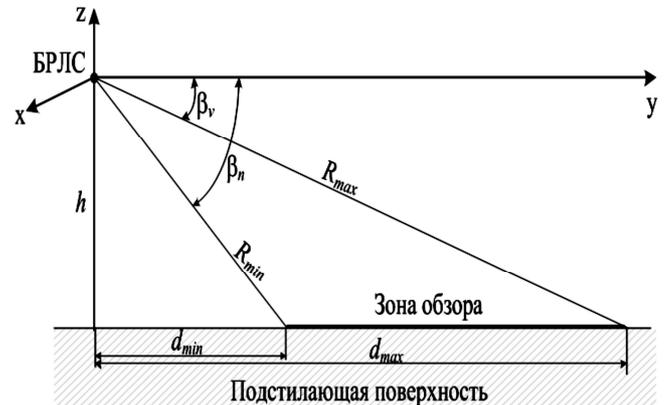


Рис. 2. Формируемая зона обзора относительно ЛА

Расстояния d_{max} , d_{min} – соответственно до переднего и заднего фронтов зоны обзора в плоскости Земли и ширину d этой зоны найдем по формулам:

$$\begin{aligned} d_{max} &= R_{max} \cdot \cos \beta_v, \\ d_{min} &= R_{min} \cdot \cos \beta_n, \\ d &= d_{max} - d_{min} \end{aligned} \tag{2}$$

Зона обзора в плоскости Земли при $\alpha_0 = 0$, т.е. при условии, что центр сектора обзора совпадает с осью ЛА, и при отсутствии ошибок в определении параметров навигационной системы имеет вид усеченного кругового сектора $A_1 B_1 B_2 A_2$ (рис. 3).

Координаты угловых точек сектора обзора в плоскости Земли при названных ограничениях найдем по формулам:

– точек A_1 и A_2 :

$$x_{A1} = d_{min} \sin \frac{\alpha}{2}, \quad y_{A1} = d_{min} \cos \frac{\alpha}{2}; \tag{3}$$

$$x_{A2} = -x_{A1}, \quad y_{A2} = y_{A1};$$

– точек B_1 и B_2 :

$$x_{B1} = d_{\max} \sin \frac{\alpha}{2}, \quad y_{B1} = d_{\max} \cos \frac{\alpha}{2}; \quad (4)$$

$$x_{B2} = -x_{B1}, \quad y_{B2} = y_{B1}.$$

Сформируем из них матрицу

$$X = \begin{pmatrix} x_{A1} & y_{A1} \\ x_{B1} & y_{B1} \\ x_{B2} & y_{B2} \\ x_{A2} & y_{A2} \end{pmatrix}.$$

Повернем теперь полученное изображение усеченного сектора на угол α_0

$$X_1 = X \cdot P \quad (5)$$

где $P = \begin{pmatrix} \cos \alpha_0 & \sin \alpha_0 \\ -\sin \alpha_0 & \cos \alpha_0 \end{pmatrix}$ – матрица преобразования.

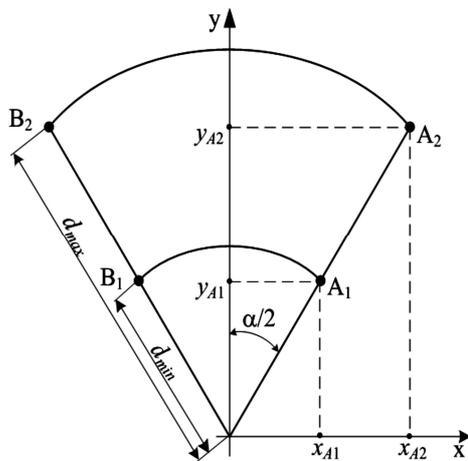


Рис. 3. Форма зоны обзора в плоскости Земли

Выполним вычисления в соответствии с формулами (1) – (5) и с исходными данными, приведенными в постановке задачи. В результате получим зону обзора, приведенную на рис. 4.

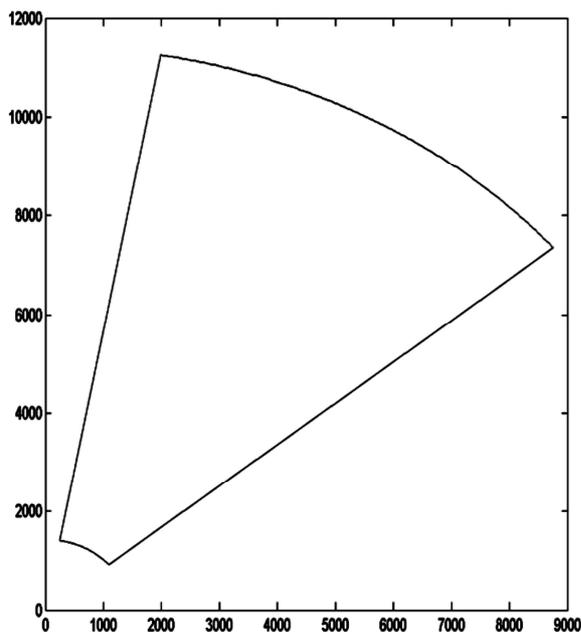


Рис. 4. Зона обзора при отсутствии ошибок

Параметры зоны обзора, приведенной на рис. 4:
– максимальная и минимальная наклонные дальности $R_{\max} = 11470 \text{ м}$, $R_{\min} = 1743,4 \text{ м}$;

– расстояния от точки O' – начала плоской системы координат в плоскости Земли – до переднего и заднего фронтов зоны обзора соответственно $d_{\max} = 11430 \text{ м}$, $d_{\min} = 1428,1 \text{ м}$;

– глубина зоны обзора $d = d_{\max} - d_{\min} = 10002 \text{ м}$;

– координаты угловых точек A_1, B_1, B_2, A_2 :

$$A_1(1094,0; 918,0), \quad B_1(8756,0; 7347,1),$$

$$B_2(1984,8; 11256,0), \quad A_2(248,0; 1406,5).$$

2. Модель РЛИ с учетом угловых ошибок навигационной системы

а) Без учета стробирования по дальности. Для построения зоны обзора в плоскости Земли с учетом ошибок в определении положения ЛА в пространстве реализуем следующий алгоритм:

– найдем орты векторов $\overline{OA_1}, \overline{OB_1}, \overline{OB_2}, \overline{OA_2}$;

– осуществим поворот этой системы векторов в плоскостях тангажа, крена и курса (рысканья) на заданные углы; в результате получим новые единичные векторы $\overline{OA'_1}, \overline{OB'_1}, \overline{OB'_2}, \overline{OA'_2}$;

– составим уравнения прямых, проходящих через точку – начало подвижной системы координат (в ней находится ЛА), – и найдем координаты точек A'_1, B'_1, B'_2, A'_2 пересечения этих прямых с плоскостью Земли, т.е координаты угловых точек зоны обзора, построенной с учетом угловых ошибок.

Орты векторов $\overline{OA_1}, \overline{OB_1}, \overline{OB_2}, \overline{OA_2}$ найдем по стандартной формуле

$$\overline{q}_i = \left(\frac{x_i}{|\overline{q}_i|}, \frac{y_i}{|\overline{q}_i|}, \frac{-h}{|\overline{q}_i|} \right) \quad i = \overline{1,4}.$$

Здесь $(x_i; y_i; -h)$ – координаты векторов $\overline{OA_1}, \overline{OB_1}, \overline{OB_2}, \overline{OA_2}$ соответственно в трехмерном пространстве.

Введем обозначения $l_i = \frac{x_i}{|\overline{q}_i|}$, $m_i = \frac{y_i}{|\overline{q}_i|}$, $n_i = \frac{-h}{|\overline{q}_i|}$ и сформируем из векторов \overline{q}_i матрицу

$$Q = \begin{pmatrix} l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \\ l_3 & m_3 & n_3 \\ l_4 & m_4 & n_4 \end{pmatrix}.$$

Для нахождения ортов векторов $\overline{OA'_1}, \overline{OB'_1}, \overline{OB'_2}, \overline{OA'_2}$ необходимо умножить матрицу Q на матрицу

$$T(\tilde{\Delta}\gamma, \tilde{\Delta}\theta, \tilde{\Delta}\psi) = T_{\tilde{\Delta}\gamma} \cdot T_{\tilde{\Delta}\theta} \cdot T_{\tilde{\Delta}\psi} \quad (6)$$

реализующую поворот единичных векторов $\overline{q}_i, i = \overline{1,4}$ на углы $\tilde{\Delta}\psi, \tilde{\Delta}\theta, \tilde{\Delta}\gamma$ в плоскостях курса, тангажа и крена

соответственно. Величины $\tilde{\Delta}\psi, \tilde{\Delta}\theta, \tilde{\Delta}\gamma$ являются суммарными угловыми ошибками, а именно:

$$\begin{aligned} \tilde{\Delta}\psi &= \Delta\psi + \mu_1 \cdot \Delta\alpha_{datch} + \mu_2 \cdot \Delta pr, \\ \tilde{\Delta}\theta &= \Delta\theta + \nu_1 \cdot \Delta\alpha_{datch} + \nu_2 \cdot \Delta pr, \\ \tilde{\Delta}\gamma &= \Delta\gamma + \lambda_1 \cdot \Delta\alpha_{datch} + \lambda_2 \cdot \Delta pr. \end{aligned} \quad (7)$$

В формулах (7) $\mu_i, \nu_i, \lambda_i, i = \overline{1,2}$ – весовые коэффициенты, принимающие значения из отрезка $[0,1]$. Они призваны учитывать вклад угловых ошибок $\Delta\alpha_{datch}$ и Δpr позиционирования ДНА по азимуту и позиционирования антенной системы за счет приводов соответственно.

Матрицы $T_{\tilde{\Delta}\psi}, T_{\tilde{\Delta}\theta}, T_{\tilde{\Delta}\gamma}$ в составе формулы (7) имеют следующий вид

$$\begin{aligned} T_{\tilde{\Delta}\psi} &= \begin{pmatrix} \cos \tilde{\Delta}\psi & -\sin \tilde{\Delta}\psi & 0 \\ \sin \tilde{\Delta}\psi & \cos \tilde{\Delta}\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\ T_{\tilde{\Delta}\theta} &= \begin{pmatrix} \cos \tilde{\Delta}\theta & 0 & \sin \tilde{\Delta}\theta \\ 0 & 1 & 0 \\ -\sin \tilde{\Delta}\theta & 0 & \cos \tilde{\Delta}\theta \end{pmatrix}, \\ T_{\tilde{\Delta}\gamma} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \tilde{\Delta}\gamma & -\sin \tilde{\Delta}\gamma \\ 0 & \sin \tilde{\Delta}\gamma & \cos \tilde{\Delta}\gamma \end{pmatrix}, \end{aligned}$$

Пусть матрица $T(\tilde{\Delta}\gamma, \tilde{\Delta}\theta, \tilde{\Delta}\psi)$ найдена. Тогда скорректированные единичные векторы $\overline{q'_1} = \overline{OA'_1}$, $\overline{q'_2} = \overline{OB'_1}$, $\overline{q'_3} = \overline{OB'_2}$, $\overline{q'_4} = \overline{OA'_2}$ будут элементами соответственно первой, второй, третьей и четвертой строк матрицы

$$Q' = Q \cdot T(\tilde{\Delta}\gamma, \tilde{\Delta}\theta, \tilde{\Delta}\psi) \quad (8)$$

Осталось реализовать последний шаг алгоритма – найти координаты угловых точек зоны обзора в плоскости Земли с учетом угловых ошибок.

Из уравнений прямых $OA'_1, OB'_1, OB'_2, OA'_2$, проходящих через точку O , в которой находится ЛА, с заданными направляющими векторами $\overline{q'_i} = (l'_i; m'_i; n'_i)$, $i = \overline{1,4}$:

$$L_i: \frac{x'_i}{l'_i} = \frac{y'_i}{m'_i} = \frac{z'_i}{n'_i}, \quad i = \overline{1,4}$$

и уравнения плоскости земли: $z = -h$ находим координаты i -й точки в плоскости Земли

$$\begin{cases} x'_i = -h \frac{l'_i}{n'_i}, \\ y'_i = -h \frac{m'_i}{n'_i}. \end{cases} \quad (9)$$

Выполним расчеты параметров зоны обзора в плоскости Земли с учетом только угловых ошибок навигаци-

онной системы.

Рассмотрим случай наиболее неблагоприятного сочетания угловых ошибок, а именно примем в формулах (7) значения всех весовых коэффициентов, равными 1: $\mu_1 = \mu_2 = \nu_1 = \nu_2 = \lambda_1 = \lambda_2 = 1$.

1. Находим орты векторов $\overline{OA_1}, \overline{OB_1}, \overline{OB_2}, \overline{OA_2}$. Они записаны построчно в указанном порядке в матрице Q

$$Q = \begin{pmatrix} 0,6275 & 0,5265 & -0,5736 \\ 0,7631 & 0,6403 & -0,0872 \\ 0,1730 & 0,9811 & -0,0872 \\ 0,1422 & 0,8067 & -0,5736 \end{pmatrix}.$$

2. Находим скорректированные с учетом угловых ошибок по формуле (8) орты векторов $\overline{OA'_1}, \overline{OB'_1}, \overline{OB'_2}, \overline{OA'_2}$. Результаты такого преобразования представлены в матрице Q'

$$Q' = \begin{pmatrix} 0,6277 & 0,5333 & -0,5671 \\ 0,7600 & 0,6451 & -0,0793 \\ 0,1680 & 0,9825 & -0,0807 \\ 0,1409 & 0,8107 & -0,5683 \end{pmatrix}.$$

3. Находим координаты угловых точек A'_1, B'_1, B'_2, A'_2 зоны обзора, найденной с учетом угловых ошибок (рис. 5):

$$\begin{aligned} A'_1(1107,0; 940,3), \quad B'_1(9582,0; 8133,1), \\ B'_2(2081,0; 12173,0), \quad A'_2(248,0; 1426,5) \end{aligned}$$

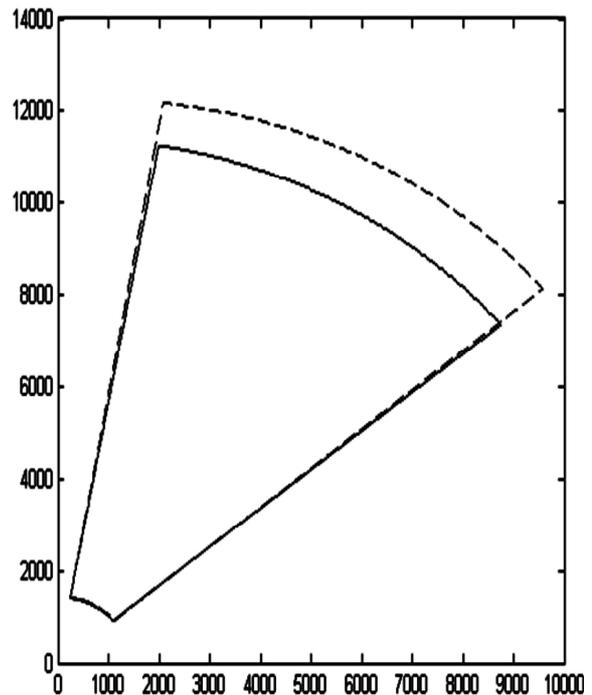


Рис. 5. Зоны обзора в плоскости Земли: сплошная линия - без учета ошибок; пунктирная линия - с учетом угловых ошибок:

$$\tilde{\Delta}\psi = \tilde{\Delta}\theta = \tilde{\Delta}\gamma = 0,32^\circ.$$

Для сравнения приведем повторно координаты угловых точек A_1, B_1, B_2, A_2 , найденные в пункте 1 без учета

ошибок и расстояния между соответствующими точками $A_1(1094, 0; 918, 0)$, $B_1(8756, 0; 7347, 1)$, $B_2(1984, 8; 11256, 0)$, $A_2(248, 0; 1406, 5)$
 $\rho(A_1, A'_1) = 25,8 \text{ м}$; $\rho(A_2, A'_2) = 20,0 \text{ м}$;
 $\rho(B_1, B'_1) = 1140,2 \text{ м}$; $\rho(B_2, B'_2) = 922,0 \text{ м}$.

б) С учетом стробирования по дальности. Зона обзора в плоскости Земли, полученная в пункте 2 а с учетом угловых ошибок навигационной системы (изображена пунктирной линией на рис. 5) в режиме стробирования по дальности будет существенно скорректирована. Режим стробирования означает, что наклонная и соответственно горизонтальная дальности РЛИ будут ограничены величинами R_{\max} и d_{\max} соответственно. В общем виде схема получения скорректированной зоны обзора в плоскости Земли в режиме стробирования по дальности такова.

1. Находятся точки пересечения прямых OB'_1 , OB'_2 с поверхностью сферы

$$S_{\max} : x^2 + y^2 + z^2 = R_{\max}^2,$$

а прямых OA'_1 , OA'_2 с поверхностью сферы

$$S_{\min} : x^2 + y^2 + z^2 = R_{\min}^2.$$

2. Полученные точки «проецируются» на плоскость Земли: $z = -h$ по дуге окружности, получающейся в результате пересечения поверхности сферы S_{\max} или S_{\min} с плоскостью, перпендикулярной плоскости Земли и содержащей соответствующую прямую (OB'_1 или OB'_2 для S_{\max} и OA'_1 или OA'_2 для S_{\min}).

Искомый результат можно получить и по другой схеме: сначала найти точки A'_1, B'_1, B'_2, A'_2 пересечения прямых $OA'_1, OB'_1, OB'_2, OA'_2$ с плоскостью Земли, а затем – точки $A''_1, B''_1, B''_2, A''_2$ пересечения ортогональных проекций этих прямых в плоскости Земли с окружностями $x^2 + y^2 = R_{\max}^2 - h^2$ и $x^2 + y^2 = R_{\min}^2 - h^2$.

Получим формулы для нахождения координат точек $A''_1, B''_1, B''_2, A''_2$ согласно описанной схеме. Решив последовательно четыре системы уравнений

$$\begin{cases} \frac{x'_i}{l'_i} = \frac{y'_i}{m'_i} = \frac{z'_i}{n'_i}, & i = 1, 4, \\ z'_i = -h \end{cases}$$

в составе которых первое уравнение является уравнением одной из прямых $OA'_1, OB'_1, OB'_2, OA'_2$, а второе уравнение – уравнением плоскости земли, найдем координаты точек A'_1, B'_1, B'_2, A'_2 (формула (9)):

$$\begin{cases} x'_i = -h \frac{l'_i}{n'_i}, \\ y'_i = -h \frac{m'_i}{n'_i}. \end{cases}$$

Выразим из этих формул y'_i через x'_i :

$$y'_i = \frac{m'_i}{l'_i} x'_i \quad (10)$$

В результате пересечения поверхностей сфер S_{\max} и

S_{\min} с плоскостью Земли получаются окружности $x^2 + y^2 = R_{\max}^2 - h^2$ и $x^2 + y^2 = R_{\min}^2 - h^2$ соответственно. Заменяя в этих уравнениях y правой частью выражения (10), после несложных преобразований, получим формулы для нахождения абсцисс точек B''_1, B''_2 в плоскости Земли

$$x''_i = \frac{l'_i}{\sqrt{(l'_i)^2 + (m'_i)^2}} \sqrt{R_{\max}^2 - h^2} \quad (11)$$

и

$$x''_i = \frac{l'_i}{\sqrt{(l'_i)^2 + (m'_i)^2}} \sqrt{R_{\min}^2 - h^2} \quad (12)$$

– абсцисс точек A''_1, A''_2 . Ординаты точек найдем по формулам (10).

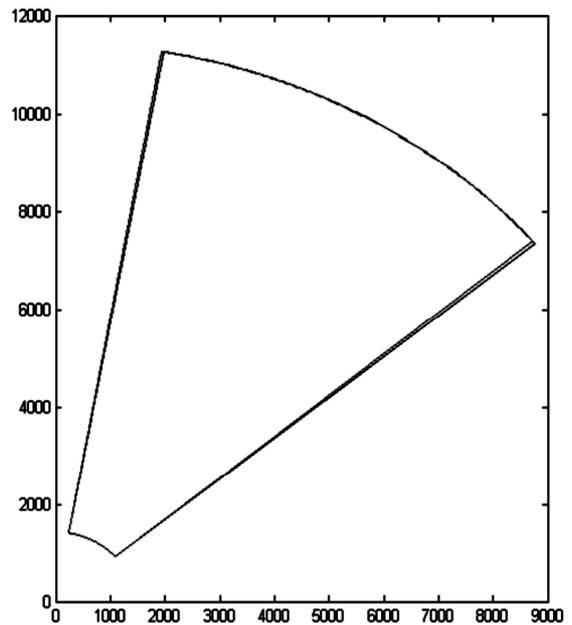


Рис. 6. Зоны обзора в плоскости Земли (рис. 5) в режиме стробирования по дальности

На рис. 6 приведены зоны обзора в плоскости Земли без учета угловых ошибок (черным цветом) и с учетом угловых ошибок, но в режиме стробирования по дальности (серым цветом).

Сравним координаты угловых точек зоны обзора без учета ошибок

$$A_1(1094, 0; 918, 0), \quad B_1(8756, 0; 7347, 1), \\ B_2(1984, 8; 11256, 0), \quad A_2(248, 0; 1406, 5)$$

и отвечающих им угловых точек зоны обзора, найденных в режиме стробирования по дальности с учетом угловых ошибок $\tilde{\Delta}\psi = \tilde{\Delta}\theta = \tilde{\Delta}\gamma = 0,32^\circ$:

$$A''_1(1188, 4; 924, 6), \quad B''_1(8714, 2; 7396, 5), \\ B''_2(1926, 0; 11267, 0), \quad A''_2(244, 5; 1407, 1)$$

Найденные ранее расстояния между соответствующими угловыми точками зон обзора без учета ошибок и с учетом угловых ошибок в режиме стробирования по дальности значительно меньше:

$$\rho(A_1, A_1'') = 8,7 \text{ м}; \quad \rho(A_2, A_2'') = 3,6 \text{ м};$$

$$\rho(B_1, B_1'') = 64,8 \text{ м}; \quad \rho(B_2, B_2'') = 59,8 \text{ м}.$$

Таким образом, влияние угловых ошибок в режиме стробирования по дальности существенно уменьшается.

3. Модель РЛИ с учетом всех видов ошибок навигационной системы.

Дополним множество ошибок навигационной системы ошибками в определении координат ЛА в пространстве, а именно – ошибками Δx , Δy определения координат ЛА по осям Ox и Oy соответственно и ошибкой Δh определения высоты ЛА. В этом случае ЛА будет находиться уже не в начале – точке O – подвижной системы координат $Oxyz$, а в точке O' с координатами $O'(\Delta x; \Delta y; \Delta h)$. Процесс формирования зоны обзора в плоскости Земли в режиме стробирования по дальности в данной ситуации в целом совпадает с процессом, описанным в пункте 2 б). Определенным образом изменяются лишь уравнения прямых $O'A_1'$, $O'B_1'$, $O'B_2'$, $O'A_2'$ и сфер S_{\max} и S_{\min} :

$$L_i: \frac{x_i' - \Delta x}{l_i'} = \frac{y_i' - \Delta y}{m_i'} = \frac{z_i' - \Delta h}{n_i'}, \quad i = \overline{1,4}$$

измененные уравнения прямых $O'A_1'$, $O'B_1'$, $O'B_2'$, $O'A_2'$; $O'A_2''$;

$$S_{\max}: (x - \Delta x)^2 + (y - \Delta y)^2 + (z - \Delta z)^2 = R_{\max}^2$$

$$S_{\min}: (x - \Delta x)^2 + (y - \Delta y)^2 + (z - \Delta z)^2 = R_{\min}^2$$

уравнения сфер переднего и заднего фронтов РЛИ.

Повторив вывод формул (11), (12) координат угловых точек зоны обзора в плоскости Земли в режиме стробирования по дальности, получим аналоги этих формул при наличии ошибок навигационной системы в определении координат ЛА в пространстве $x_i'' = \Delta x + \frac{l_i'}{\sqrt{(l_i')^2 + (m_i')^2}} \times$

$$\times \sqrt{R_{\max}^2 - (h + \Delta h)^2}, \quad - \text{ абсциссы точек } B_1'', B_2'',$$

$$x_i'' = \Delta x + \frac{l_i'}{\sqrt{(l_i')^2 + (m_i')^2}} \sqrt{R_{\min}^2 - (h + \Delta h)^2}, \quad - \text{ абсциссы}$$

точек A_1'', A_2'' . Ординаты точек найдем по формулам

$$y_i' = \Delta y + \frac{m_i'}{l_i'}(x_i' - \Delta x), \quad i = \overline{1,4}.$$

На рис. 7 приведены зоны обзора в плоскости Земли без учета ошибок (черным цветом) и с учетом всех видов ошибок в режиме стробирования по дальности (серым цветом).

По аналогии с предыдущим пунктом сравним координаты угловых точек зоны обзора без учета ошибок

$$A_1(1094,0; 918,0), \quad B_1(8756,0; 7347,1),$$

$$B_2(1984,8; 11256,0), \quad A_2(248,0; 1406,5)$$

и отвечающих им угловых точек зоны обзора,

$$A_1''(1127,6; 965,4), \quad B_1''(8765,5; 7447,6),$$

$$B_2''(1976,3; 11318,0), \quad A_2''(292,1; 1443,1),$$

найденных в режиме стробирования по дальности с учетом всех типов ошибок, а именно: $\tilde{\Delta}\psi = \tilde{\Delta}\theta = \tilde{\Delta}\gamma = 0,32^\circ$, $\Delta x = \Delta y = 50 \text{ м}$, $\Delta h = 20 \text{ м}$.

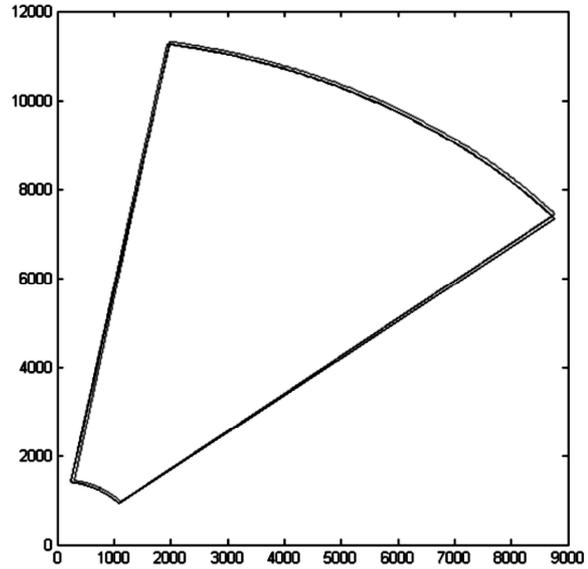


Рис. 7. Зоны обзора в плоскости Земли в режиме стробирования по дальности в условиях ошибок: $\tilde{\Delta}\psi = \tilde{\Delta}\theta = \tilde{\Delta}\gamma = 0,32^\circ$, $\Delta x = \Delta y = 50 \text{ м}$, $\Delta h = 20 \text{ м}$

Расстояния между соответствующими угловыми точками зон обзора без учета ошибок и с учетом ошибок в режиме стробирования по дальности таковы:

$$\rho(A_1, A_1'') = 58,1 \text{ м}; \quad \rho(A_2, A_2'') = 56,7 \text{ м};$$

$$\rho(B_1, B_1'') = 100,1 \text{ м}; \quad \rho(B_2, B_2'') = 56,7 \text{ м}.$$

Изменим теперь знак ошибки по координате x на противоположный ($\Delta x = -50 \text{ м}$), а значения остальных ошибок сохраним неизменными. В этом случае зона обзора в плоскости Земли в режиме стробирования по дальности будет иметь вид (рис. 8).

Координаты угловых точек зоны обзора без учета ошибок останутся неизменными, а координаты угловых точек зоны обзора, найденных в режиме стробирования по дальности с учетом всех типов ошибок ($\tilde{\Delta}\psi = \tilde{\Delta}\theta = \tilde{\Delta}\gamma = 0,32^\circ$, $\Delta x = -50 \text{ м}$, $\Delta y = 50 \text{ м}$, $\Delta h = 20 \text{ м}$) будут такими:

$$A_1''(1027,6; 965,4), \quad B_1''(8665,5; 7447,6),$$

$$B_2''(1876,3; 11318,0), \quad A_2''(192,1; 1443,1)$$

Расстояния между соответствующими угловыми точками зон обзора без учета ошибок и с учетом ошибок в режиме стробирования по дальности в этом случае увеличатся:

$$\rho(A_1, A_1'') = 81,6 \text{ м}; \quad \rho(A_2, A_2'') = 66,8 \text{ м};$$

$$\rho(B_1, B_1'') = 135,2 \text{ м}; \quad \rho(B_2, B_2'') = 125,0 \text{ м}.$$

Изменим еще и знак ошибки по координате y на противоположный ($\Delta y = -50 \text{ м}$). В этом случае



($\tilde{\Delta}\psi = \tilde{\Delta}\theta = \tilde{\Delta}\gamma = 0,32^\circ$, $\Delta x = \Delta y = -50$ м, $\Delta h = 20$ м) координаты угловых точек зоны обзора, найденные в режиме стробирования по дальности будут такими:

$$A_1''(1027,6; 865,4), \quad B_1''(8665,5; 7347,6),$$

$$B_2''(1876,3; 11218,0), \quad A_2''(192,1; 1343,1).$$

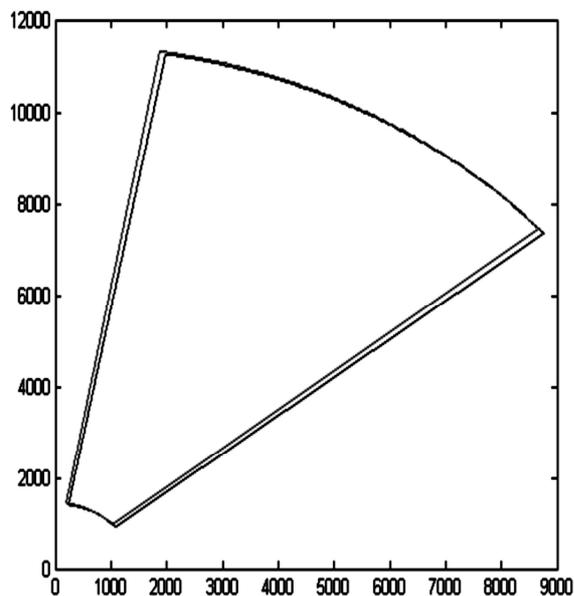


Рис. 8. Зоны обзора в плоскости Земли в режиме стробирования по дальности в условиях ошибок $\tilde{\Delta}\psi = \tilde{\Delta}\theta = \tilde{\Delta}\gamma = 0,32^\circ$, $\Delta x = -50$ м, $\Delta y = 50$ м, $\Delta h = 20$ м

Расстояния между соответствующими угловыми точками зон обзора без учета ошибок и с учетом ошибок в режиме стробирования по дальности изменятся следующим образом:

$$\rho(A_1, A_1'') = 120,0 \text{ м}; \quad \rho(A_2, A_2'') = 66,9 \text{ м};$$

$$\rho(B_1, B_1'') = 90,5 \text{ м}; \quad \rho(B_2, B_2'') = 115,0 \text{ м}.$$

Заключение

При неблагоприятном сочетании ошибок навигационных параметров по положению ЛА в пространстве – ошибки $\Delta x, \Delta y, \Delta h$ – и по ориентации ЛА в пространстве – ошибки $\Delta\psi, \Delta\theta, \Delta\gamma$ – искажения зоны обзора РЛИ в плоскости Земли, измеренные в отклонениях угловых точек относительно неискаженного изображения, могут достигать 100 – 135 м при заданных параметрах носителя РЛС и навигационной системы. Это соответствует приблизительно 1 % от максимальной горизонтальной дальности (расстояние до переднего фронта зоны обзора) и около 7 % от минимальной горизонтальной дальности (расстояние до заднего фронта зоны обзора).

Наибольшие искажения зоны обзора в плоскости Земли, как в области переднего фронта, так и заднего фронта обусловлены ошибками $\Delta x, \Delta y, \Delta h$ в определении координат ЛА в пространстве. Угловые ошибки $\Delta\psi, \Delta\theta, \Delta\gamma$ в наибольшей степени влияют на точки изображения в области переднего фронта.

Предложенная модель, разумеется, относится к простейшим (не учитывающим законы распределения ошибок и их изменения при сочетании ошибок), но в то же время является достаточной для оценки зоны неопределенности радиотражений от подстилающей поверхности для РЛС, работающей в режиме реального луча, и может использоваться в задачах, связанных с автоматическим совмещением радиолокационных изображений с картографической информацией

Литература

1. Аксенов О.Ю. Совмещение изображений // Цифровая обработка сигналов, 2005, №3.
2. R.G. White Change detection in SAR imagery // International Journal of Remote Sensing. vol.12, Issue 2, Feb. 1991.
3. Руководство Р-315 «По минимальным стандартам характеристик авиационных систем (MASPS) для систем улучшенного видения, систем искусственного видения, комбинированных систем искусственного видения и бортовых систем увеличения дальности видения» Авиационного Регистра Межгосударственного Авиационного Комитета (АР МАК), 2008. 86 с.
4. Герман Е.В., Муратов Е.Р., Новиков А.И. Математическая модель формирования зоны неопределенности в задаче совмещения изображений // Вестник РТРТУ №4 (выпуск 46, часть 2), 2013 г. – с. 11-16.

EVALUATION OF UNCERTAINTY OF RADIO REFLECTIONS IN COMBINING IMAGES WITH DIGITAL TERRAIN MAPS

Novikov A.I., Ukin S.A.

A mathematical model of the zone of uncertainty in the underlying surface, corresponding to onboard radar station field of view operating in the forward hemisphere, is proposed. The research results of the model for different sensor error of navigation parameters are presented. It is shown that for the problems of combining cartographic information with formed by the radar image it is necessary to expand the search area, taking into account areas of uncertainty depending on the angle of view, resulting in an asymmetric distortion (extensions of presumed field of view) relative to the center of scanning sector.

УДК 681.3.053

ОБРАБОТКА ИЗОБРАЖЕНИЙ С ПОМОЩЬЮ ЦЕЛОЧИСЛЕННЫХ ОРТОГОНАЛЬНЫХ ПРЕОБРАЗУЮЩИХ МАТРИЦ

Шеремет И.А., д.т.н., профессор, e-mail: sheremet_ia@apr.gov.ru;

Рукин А.П., старший преподаватель Института иностранных языков им. М. Тореза, e-mail: nto@mniti.ru;

Лебедев В.Д., к.т.н., ведущий инженер ЗАО «МНИТИ», e-mail: @mniti.ru.

Ключевые слова: обработка изображений, ортогональные преобразующие матрицы, целочисленные, матрицы Уолша-Адамара, преобразование Хаара.

В настоящей статье рассмотрены подходы к решению задачи обработки цифровых изображений с предварительным преобразованием и последующем восстановлением изображения с использованием нового класса целочисленных ортогональных преобразующих матриц.

Введение

В ЗАО «МНИТИ» были разработаны методы и алгоритмы устранения пространственной корреляционной избыточности цифровых изображений, использующие целочисленные преобразования, которые возможно применять к изображениям произвольного формата. Эти преобразования имеют единообразную форму генерации преобразующих матриц, предполагающую их многовариантность при сохранении декоррелирующих свойств, и как следствие, эффективное использование параллельной обработки при аппаратной реализации. Данные семейства матриц являются новыми и, следовательно, требующими апробации по самым разным направлениям [1-3].

Упомянутые матричные методы апробированы, в частности, по критерию степени сжатия без последующего квантования на цифровых монохромных тестовых изображениях, различающихся по размеру и структурной сложности. Ряд новых матричных методов, по степени сжатия после преобразования без последующего квантования, показал результаты не хуже, а по отдельным параметрам и лучше, чем у классических методов (ДКП, преобразования из стандарта H.264 и т.п.) [4].

В настоящей статье рассматривается возможность применения новых методов обработки (улучшения) цифровых изображений путём переноса их в область преобразования, а после проведения обработки – восстановление с применением операции обратного преобразования.

В работе [3] было показано, что генерация больших (по размеру) целочисленных ортогональных декоррелирующих матриц (ЦОДМ) из малых может производиться двумя способами, которые различаются объединяющей матричной операцией.

Если такой объединительной операцией является тензорное произведение, тогда имеем целочисленные ортогональные декоррелирующие матрицы, являющиеся обобщением матриц Уолша-Адамара. Обобщение касается размерности декоррелирующей матрицы, которая будет являться произвольной, а не натуральной степенью числа два, как для матриц Адамара.

Если же объединяющей является кратно-масштабная операция, то в качестве результата получаем обобщение матриц преобразований Хаара для случая, когда

размерность декоррелирующей матрицы произвольна и не равна исключительно натуральной степени числа два, что характерно для матриц Хаара [3, 5].

Ещё недавно популярные методы Адамара и Хаара перестают использоваться из-за их низкой эффективности. Из созданных на их структурах девяти матриц применяется лишь одна [6]. Эффективность преобразования обычно сравнивают с преобразованием Кархунена-Лоэва [7], которым достаточно точно аппроксимируют собственные векторы матрицы Тёплица, что и приводит к использованию чётного косинусного преобразования при обработке фотографических изображений.

Это действительно так. Но, с другой стороны, метод и его обобщения – не тождественны и последние, как раз и создаются, чтобы исправить недостатки их прототипа, сохранив по возможности его положительные свойства. Основными положительными свойствами методов Адамара и Хаара являются их целочисленное представление и возможность генерации больших по размеру матриц из меньших.

Оценка любого метода может быть только многопараметрической. Близкий к идеальному метод устранения пространственной корреляции Кархунена-Лоэва практически не реализуем, благодаря его сверхвысокой вычислительной сложности, которая сводит на нет все его достоинства [7]. Требуется компромисс.

Одним из таких компромиссов являются методы на основе ДКП [6]. Эти методы, в основном, вещественные, работающие в плавающей арифметике. Методы декорреляции, основанные на ЦОДМ, являются другим вариантом компромисса.

Дискуссии о преимуществах целочисленной арифметики перед вещественной, являются беспредметными, если не оговаривается вопрос аппаратной реализации алгоритма. Ещё недавно преимущества целочисленной арифметики казались очевидными. Считалось, что одна элементарная операция (сложение или умножение) в плавающей арифметике равнозначна, примерно, 10-ти таким же операциям в арифметике целочисленной. И для целочисленных процессоров, в которых вещественные операции реализованы программно в качестве эму-

ляции, это действительно так. [8] Но сегодня мы имеем резкий рост мощности современных процессоров, что постепенно сглаживает антагонизм арифметик – целочисленной и плавающей [9].

Проблема перерастает в иную плоскость: мощные, универсальные, почти не зависящие от сложности алгоритма, и, следовательно, дорогие, и простые процессоры, дешёвые, не универсальные, направленные на решение какой-то одной конкретной задачи (например, мобильный мониторинг какой-либо местности или объекта) и требующие максимального упрощения алгоритма, что приводит к использованию целочисленной арифметики. Такова цена вопроса. В рамках указанной ситуации можно предполагать, что целочисленные матричные алгоритмы декорреляции цифровых изображений найдут свою нишу и обретут право на жизнь [10].

Всё выше сказанное касалось в большей степени матричных преобразований декорреляции цифрового изображения с целью его последующего квантования и энтропийного сжатия.

Обработка изображения с целью его улучшения – это другая задача. Она может решаться, как без переноса изображения в другую область представления, так и с переносом и последующим восстановлением. При этом для такого переноса иногда достаточны самые простые методы преобразования.

Как показано в работах Е.И. Травиной (ЗАО «МНИТИ») [11], использование для декомпозиции исходного изображения нескольких модифицированных преобразований Хаара с последующей реконструкцией после обработки (улучшения) преобразованного изображения, даёт значительный положительный результат.

Возникает интерес с точки зрения определения возможности обработки (улучшения) цифрового изображения при использовании матриц ЦОДМ.

Ниже приводятся простейшие примеры генерации ЦОДМ по первому и второму указанным направлениям.

Пусть заданы элементарные матрицы ЦОДМ, размерами 2x2 и 3x3.

$$H2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \text{ и } H3 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -2 & 1 \end{bmatrix}, \quad (1)$$

которые являются ортогональными и декоррелирующими [3, 12]. Ни тензорное произведение, ни кратно-масштабное объединение не являются коммутативными операциями. Поэтому на основании приведённых двух элементарных матриц ЦОДМ, можно построить две пары декоррелирующих матриц размерами 6x6, одна пара по «адамаровскому» типу, а вторая – по «хааровскому».

Приведём пример построения больших (по размеру) матриц из малых по первому, «адамаровскому», направлению.

$$Hd6_1' = H2 \otimes H3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 0 & -1 & 1 & 0 & -1 \\ 1 & 0 & -1 & -1 & 0 & 1 \\ 1 & -2 & 1 & 1 & -2 & 1 \\ 1 & -2 & 1 & -1 & 2 & -1 \end{bmatrix},$$

$$Hd6_2' = H3 \otimes H2 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 0 & 0 & -1 & -1 \\ 1 & -1 & 0 & 0 & -1 & 1 \\ 1 & 1 & -2 & -2 & 1 & 1 \\ 1 & -1 & -2 & 2 & 1 & -1 \end{bmatrix}. \quad (2)$$

После переупорядочения строк полученных матриц по критерию возрастания их частот [13], получаем:

$$Hd6_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 0 & -1 & -1 & 0 & 1 \\ 1 & 0 & -1 & 1 & 0 & -1 \\ 1 & -2 & 1 & 1 & -2 & 1 \\ 1 & -2 & 1 & -1 & 2 & -1 \end{bmatrix}, \quad (3)$$

$$Hd6_2 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 & -1 \\ 1 & 1 & -2 & -2 & 1 & 1 \\ 1 & -1 & -2 & 2 & 1 & -1 \\ 1 & -1 & 0 & 0 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix}.$$

Генерация двух матриц «хааровского» типа выглядит так [5]:

$$H6_1 = \begin{bmatrix} H2 & \otimes & [1 & 1 & 1] \\ I2 & \otimes & [1 & 0 & -1] \\ I2 & \otimes & [1 & -2 & 1] \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}, \quad (4)$$

$$H6_2 = \begin{bmatrix} H3 & \otimes & [1 & 1] \\ I3 & \otimes & [1 & -1] \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 & -1 \\ 1 & 1 & -2 & -2 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \quad (5)$$

где $I2$ – единичная матрица размером 2x2, а $I3$ – единичная матрица размером 3x3.

Рассмотрим следующие графические сопоставления.

Ниже на рис. 1 сравниваются три ортогональные преобразующие (декоррелирующие) матрицы. Две матрицы ЦОДМ: $H6_1$, $H6_2$ и вещественная матрица ДКП размером 6x6. На рис. 1 соответственно приведены 3 группы функций конечного дискретного ортогонального базиса, каковыми являются строки преобразующих матриц.

Сопоставляемые матрицы имеют вид:

$$Hd6_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 0 & -1 & -1 & 0 & 1 \\ 1 & 0 & -1 & 1 & 0 & -1 \\ 1 & -2 & 1 & 1 & -2 & 1 \\ 1 & -2 & 1 & -1 & 2 & -1 \end{bmatrix};$$

$$Hd6_2 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 & -1 \\ 1 & 1 & -2 & -2 & 1 & 1 \\ 1 & -1 & -2 & 2 & 1 & -1 \\ 1 & -1 & 0 & 0 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix} \quad (6)$$

$$ДКП_{6 \times 6} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0.965 & 0.707 & 0.258 & -0.258 & -0.707 & -0.965 \\ 0.866 & 0 & -0.866 & -0.866 & 0 & 0.866 \\ 0.707 & -0.707 & -0.707 & 0.707 & 0.707 & -0.707 \\ 0.5 & -1 & 0.5 & 0.5 & -1 & 0.5 \\ 0.258 & -0.707 & 0.965 & -0.965 & 0.707 & -0.258 \end{bmatrix}$$

Строки расположены по возрастанию частот от 0 до 5-ти [13]. Из рисунка видно, что все три системы близки, и между их соответствующими строками существует ярко выраженная корреляция. Отличие первых двух систем от третьей заключается в их целочисленном представлении.

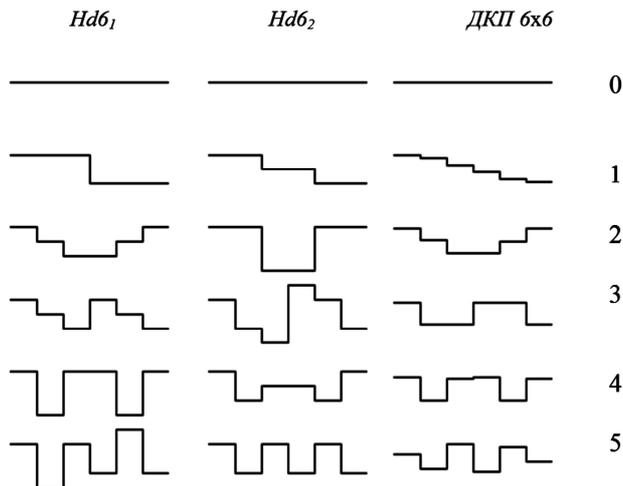


Рис. 1. Графическое сопоставление ортогональных систем ЦОДМ и ДКП размером 6x6

На рис. 2 показаны зависимости строк двух ортогональных преобразующих матриц. (Матрица ЦОДМ $H5$ размером 5×5 и классическая матрица ДКП размером 5×5).

Сопоставляемые матрицы имеют вид:

$$H5 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & -4 & 1 & 1 \\ 1 & -1 & 0 & 1 & -1 \\ 1 & -1 & 0 & -1 & 1 \end{bmatrix}, \quad (7)$$

$$ДКП_{5 \times 5} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0.951 & 0.587 & 0 & -0.587 & -0.951 \\ 0.809 & -0.309 & -1 & -0.309 & 0.809 \\ 0.587 & -0.951 & 0 & 0.951 & -0.587 \\ 0.309 & -0.809 & 1 & -0.809 & 0.309 \end{bmatrix}$$

О генерации матрицы типа $Hd5$ подробно излагается в работе [1].

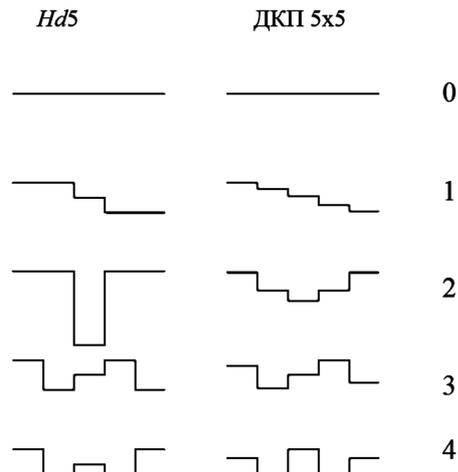


Рис. 2. Графическое сопоставление ортогональных систем ЦОДМ и ДКП размером 5x5

Из графиков (рис. 2) видно, что общая тенденция в поведении строк соответствующей частоты сохраняется.

Если провести непрерывные огибающие для дискретных функций, представленных на рис 1-2, то получатся зависимости, напоминающие полиномы Чебышева (рис. 3) с поправкой на знак и коэффициент усиления по оси Y или же тригонометрические функции тех же частот на том же участке, хотя и с большими поправками. При этом, для графиков одинаковых по частоте чётко прослеживается соответствие между экстремумами и точками перегибов.

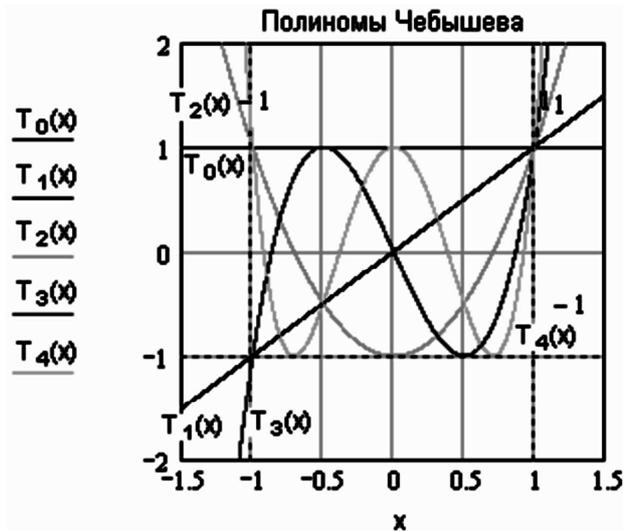


Рис.3. Полиномы Чебышева. Полином Чебышева на интервале $(-1, 1)$ ограничен значениями $(-1, 1)$, а за пределами этого интервала он растет по абсолютной величине быстрее любого другого полинома той же степени, ограниченного тем же условием [14]

К сожалению, в рамках одной статьи трудно привести примеры улучшения цифровых изображений с помощью всего семейства новых матриц или даже нескольких из них. Остановимся на одной. Выбор преобразующей матрицы $Hd6_2$ в каком-то смысле субъективен.

С одной стороны, метод декорреляции на основе этой матрицы показал один из лучших результатов по критерию степени сжатия без квантования, оставив позади даже вещественную декоррелирующую матрицу ДКП, размером 8x8 [4]. Это, как было сказано, относится к другой задаче, но поскольку проводится апробация новых матричных методов, то и этот результат на этапе первоначального поиска необходимо учитывать.

С другой стороны, хотелось бы полнее раскрыть такое свойство нового семейства матриц, как универсальность по отношению к формату изображения. Исходное тестируемое изображение (см. ниже) – это цифровое монохромное изображение размерами 246x246 пикселей. Число 246 не делится ни на 8, ни на 4, следовательно, его невозможно, без изменения формата изображения точно преобразовать методами, основанными на матрицах указанных размерностей, но оно легко преобразуется матрицей размером 6x6. С этой точки зрения было бы нежелательно использовать такие «классические» размеры, как 8x8 или 4x4, внося неизбежные искажения в исходное изображение, а применить именно матрицу размерами 6x6.

И ещё, матрицы $H2$ и $H3$, на основе которых генерируются ЦОДМ размером 6x6, в определённом смысле являются целочисленными представлениями матриц ДКП тех же размеров.

Матрица ДКП, размером 2x2 (назовём её $DCT_{2 \times 2}$) имеет следующий вид:

$$DCT_{2 \times 2} = \begin{bmatrix} \cos 0 & \cos 0 \\ \cos \frac{\pi}{4} & -\cos \frac{\pi}{4} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}; \quad (8)$$

Матрица ДКП, размером 3x3 (назовём её $DCT_{3 \times 3}$), имеет вид:

$$DCT_3 = \begin{bmatrix} \cos 0 & \cos 0 & \cos 0 \\ \cos \frac{\pi}{6} & \cos \frac{\pi}{2} & -\cos \frac{\pi}{6} \\ \cos \frac{\pi}{3} & \cos \pi & \cos \frac{\pi}{3} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ \frac{\sqrt{3}}{2} & 0 & -\frac{\sqrt{3}}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} \end{bmatrix}. \quad (9)$$

Если разделить величины элементов второй строки сверху матрицы ДКП размером 2x2 на величину $\sqrt{2}/2$, а значения элементов второй строки сверху матрицы ДКП размером 3x3 на величину $\sqrt{3}/2$, а третью строку этой же матрицы умножить на 2, то мы получим следующие ортогональные декоррелирующие матрицы, которые можно считать целочисленным представлением матриц ДКП:

$$H2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \text{ и } H3 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -2 & 1 \end{bmatrix}. \quad (10)$$

Объединение декоррелирующих матриц ДКП (в том числе и в целочисленном представлении) с помощью тензорного произведения в большую по размерам декоррелирующую ортогональную матрицу с целью изучения её свойств – это ещё одно направление для исследования. Таковы основания, почему для последующего тести-

рования была всё-таки выбрана матрица $Hd6_2$.

Ортогональное матричное преобразование является разложением конечной дискретной функции по конечному ортогональному дискретному базису. Функция в данном случае – это участок строки равный по длине (в пикселях) размерности декоррелирующей матрицы. Отсюда необходимое требование: размерность преобразующей ЦОДМ должна являться делителем длины строки преобразуемого изображения. Поскольку допустима генерация ЦОДМ произвольной размерности [3], декоррелирующую матрицу необходимой величины всегда можно подобрать.

Механизм разложения дискретной функции по ортогональному базису кратко можно представить следующим образом [15]:

Пусть R -преобразующая матрица, размерностью $n \times n$. Каждая i -я строка матрицы R обозначается как $r(i)$, где $i=0, \dots, n-1$. Пусть дискретная преобразуемая функция (участок строки или столбца) представлена в виде вектора:

$$f = (f_0, f_1, f_2, \dots, f_{n-1}). \quad (11)$$

Разложение функции f по базису R будет иметь вид:

$$f = \lambda_0 r(0) + \lambda_1 r(1) + \lambda_2 r(2) + \dots + \lambda_{n-1} r(n-1). \quad (12)$$

Задача преобразования, таким образом, сводится к нахождению вектора

$$\lambda = (\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_{n-1}). \quad (13)$$

Для нахождения λ_i вычислим скалярное произведение вектора $r(i)$ и вектора f .

Из уравнения (12) имеем следующее:

$$(r(i), f) = \lambda_0 (r(i), r(0)) + \lambda_1 (r(i), r(1)) + \lambda_2 (r(i), r(2)) + \dots + \lambda_{n-1} (r(i), r(n-1)). \quad (14)$$

Поскольку матрица R – ортогональна, то для неё характерно свойство:

$$(r_{(i)}, r_{(j)}) = \begin{cases} |r_{(i)}|^2, & i = j \\ 0, & i \neq j \end{cases}, \quad (15)$$

на основе которого, уравнение (14) сведётся к равенству

$$(r(i), f) = \lambda_i |r(i)|^2, \quad (16)$$

откуда $\lambda_i = (r(i), f) / |r(i)|^2$.

В уравнении (16) $(r(i), f)$ – скалярное произведение i -й строки матрицы R , а f – участок строки (столбца) величиной в n пикселей, а $|r(i)|^2$ – скалярный квадрат i -й строки матрицы R . Поскольку матрица R не вырождена, скалярный квадрат любой её строки – величина больше нуля, поэтому деление в уравнении всегда осуществимо. Очевидно, что обратное преобразование осуществляется с помощью уравнения (12). Операция деления при обратном преобразовании не применяется.

Первоначальный результат одномерного преобразования (по строке или столбцу) представляет собой совокупность подмножеств коэффициентов преобразования, количество которых в подмножестве равно размерности декоррелирующей матрицы. Первый элемент подмножества – коэффициент нулевой частотности, его яркост-

ная величина примерно соответствует пикселю в пространственном представлении, остальные коэффициенты – это более высокие гармоники частотного преобразования.

После матричного преобразования гармоники одинаковых частот группируются в отдельные частотные подблоки. Число таких подблоков будет, очевидно, равно количеству строк декоррелирующей матрицы. Процесс перегруппировки коэффициентов преобразования показан на рис. 4.

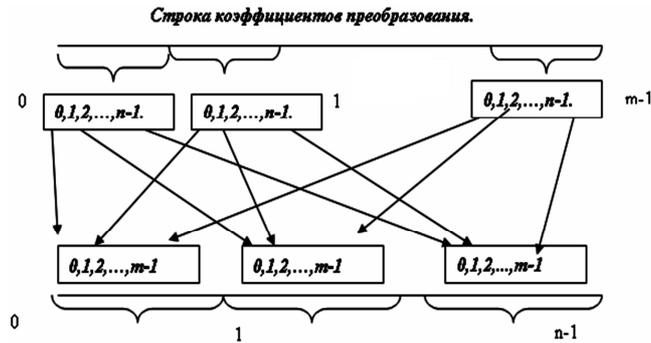


Рис. 4. Перегруппировка коэффициентов преобразования по частотным подблокам

Вслед за преобразованиями по строкам производится преобразование по столбцам, которое полностью аналогично. Единственное возможное отличие – декоррелирующая матрица может быть других размеров в зависимости от высоты столбца изображения в пикселях. В частном случае преобразующие матрицы для декорреляции по строкам и по столбцам могут совпадать.

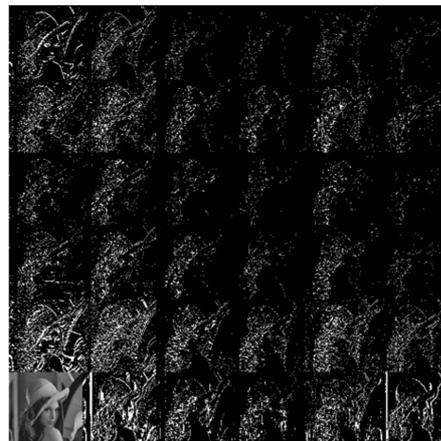
Результат полного (двумерного) преобразования представляет собой набор блоков, количество которых по строкам равно размерности декоррелирующей матрицы по строкам, а количество по столбцам равно размерности декоррелирующей матрицы по столбцам. Если обе матрицы совпадают, то количество блоков одинаково как по строкам, так и по столбцам, и равно размерности декоррелирующей матрицы. Данные блоки располагаются в двумерной частотной системе координат (двумерный спектр), а коэффициенты преобразования внутри каждого блока располагаются в пространственной двумерной системе координат. Подробно алгоритм двумерного матричного преобразования изображения изложен в [1, 4].

Результат действия матричного преобразования приведён на рис. 5. Исходное изображение – цифровое монохромное размерами 246x246 пикселей. Внизу от него на рисунке представлено преобразованное изображение. В левом нижнем углу отчётливо различается блок нулевой частотности, являющийся уменьшенной копией исходного изображения.

Ниже в виде матрицы размером 6x6 показана система двумерных частотных координат, согласно которой располагаются частотные блоки изображения на рис. 5, размером 41x41 пикселей (41x6=246), полученные в результате преобразования исходного изображения с помощью ЦОДМ Hd_6 .



Исходное изображение



Результат преобразования

Рис. 5. Преобразование с помощью ЦОДМ Hd_6

(5, 0)	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)
(4, 0)	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)
(3, 0)	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)
(2, 0)	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)
(1, 0)	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)
(0, 0)	(0, 1)	(0, 2)	(0, 3)	(0, 4)	(0, 5)

Первая координата – номер строки, вторая – номер столбца этой матрицы.

Обработка изображения заключается в умножении всех коэффициентов преобразования каждого частотного блока на соответствующий каждой двумерной частотной области множитель. При этом, множитель при нулевой частотной области (с координатами 0,0) отвечает за увеличение или снижение яркости, а множители при высокочастотных областях отвечают за детальность и шум, причём, чем выше частота, тем более мелкие детали и элементы шума может усиливать или подавлять соответствующий множитель. Всю систему множителей, необходимых для обработки изображений, можно представить в виде матрицы размером 6x6. Если пространство, в котором расположено изображение, анизотропно, то есть направления по осям OX и OY неравноправны, множители указанной матрицы будут ассиметричны относительно диагонали (0,0)-(5,5), в противном случае они будут симметричны.



Помимо этого следует иметь в виду, что пиксельная система сама по себе анизотропна. Прямые или отрезки, не являющиеся параллельными осям OX и/или OY , а также прямым, расположенным под 45° к последним, представлены ломаными линиями, что при обработке изображения часто приводит к нежелательным артефактам («ступенчатости»). Природа этих артефактов заключается как раз в анизотропности пиксельной системы – в её привязанности к прямоугольной (декартовой) сетке координат.

Обработка в частотной области (особенно, когда она затрагивает высокочастотные гармоники) связана прежде всего с детальностью. Размеры деталей, качество видимости которых регулируется усилением или подавлением коэффициентов соответствующей гармоники, обратно пропорциональны частоте гармоники. Чем выше частота (которая, разумеется, не должна превышать частоты Найквиста), тем более мелкие детали могут стать различимыми при работе с данной гармоникой. Это для одномерного случая. Двумерная система частотных координат отчасти позволяет бороться с анизотропностью деталей. Частота по горизонтали обратно пропорциональна длине, а частота по вертикали высоте детали. Подбирая нужные двумерные частотные координаты и усиливая соответствующие им гармоники, мы повышаем различимость деталей с заданной ориентацией в пространстве, избегая, при правильном подборе коэффициентов усиления, нежелательных геометрических искажений.

Вычислительная сложность обработки преобразованного изображения добавляет всего одну целочисленную операцию. То есть, если мы имеем дело с усилением (яркости, детальности), эта целочисленная операция – умножение, которую можно отнести к элементарным, дополнительные затраты вычислений будут минимальными. Если речь идёт о подавлении коэффициентов преобразования, расположенных в подблоке с одними и теми же двумерными частотными координатами, то эта операция будет делением нацело с округлением. Хотя она и не элементарна, но её часто можно свести к совокупности элементарных операций (умножение, побитовый сдвиг). После усиления или подавления коэффициентов преобразования соответствующих двумерных частотных областей (блоков) производится операция обратного преобразования (восстановления, реконструкции), в результате чего формируется обработанное исходное изображение, яркость и различные уровни детальности которого изменены согласно целям поставленной задачи.

Иногда операцию контрастирования более целесообразно осуществлять после восстановления изображения с изменёнными яркостью и детальностью различных частотных уровней. При этом значение каждого пикселя вычисляется по известной несложной формуле:

$$y'_{ij} = k \cdot (y_{ij} - mid) + mid, \quad (17)$$

где y_{ij} – значение яркости исходного пикселя, mid – среднее значение яркостей всех пикселей изображения, k – коэффициент контрастирования, y'_{ij} – значение яркости пикселя после контрастирования. При этом необ-

ходимо следить, чтобы полученные значения пикселей обработанного изображения не выходили за границы шкалы яркостей (0-255).

Ниже на рис. 6 показан пример обработки изображения указанной матрицей ЦОДМ $Hd6_2$.



Исходное изображение



Обработанное изображение

Рис. 6. Обработка изображения, преобразованного матрицей ЦОДМ Hb_2

Матрица множителей при частотных блоках (см. рис. 5) для обработки изображения, показанной на рис. 6, имеет следующий вид:

$$\begin{bmatrix} 2 & 4 & 4 & 4 & 4 & 4 \\ 2 & 4 & 4 & 4 & 4 & 4 \\ 2 & 4 & 4 & 4 & 4 & 4 \\ 2 & 4 & 4 & 4 & 4 & 4 \\ 2 & 4 & 4 & 4 & 4 & 4 \\ 1 & 2 & 2 & 2 & 2 & 2 \end{bmatrix}$$

Коэффициент контрастирования равен двум. На рис. 6. (внизу) показано обработанное изображение (исходное расположено сверху). Явно визуально наблюдается повышение контрастности и детальности (по всем высокочастотным блокам).

Кроме того наблюдаются некоторые слегка заметные артефакты: «ступенчатость» на линиях, не являющихся параллельными осям OX и/или OY , а также прямым, расположенным к ним под 45° . Природа этих артефактов заключается в анизотропности пиксельной системы – в её привязанности к прямоугольной (декартовой) сетке координат.

Для устранения этих артефактов следует уменьшать

величины ряда высокочастотных элементов матрицы множителей и одновременно частично жертвовать полученным усилением детальности.

Пример обработки того же исходного изображения, но при другой матрице множителей показан на рис.7. Новая матрица множителей при частотных блоках (рис. 5), показанной на рис. 7 обработки изображения, следует сразу же за рисунком.



Исходное изображение



Обработанное изображение

Рис. 7. Обработка изображения при другой матрице множителей

$$\begin{bmatrix} 3 & 3 & 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 & 3 & 3 \\ 1 & 3 & 3 & 3 & 3 & 3 \end{bmatrix}$$

Коэффициент контрастирования остаётся прежним равным двум. На рис. 7. (внизу) показано обработанное изображение, а сверху, как и в предыдущем случае, исходное. Наблюдается повышение контрастности и детальности (по всем высокочастотным блокам), хотя и несколько меньше, чем на рис. 6, но артефакты практически отсутствуют. Следовательно, матрица множителей подобрана правильно.

На практике обработка изображений необходима прежде всего в тех случаях, когда исходное изображение визуально трудно воспринимается. Например, пейзаж в тумане или слишком затемнённые изображения и т.п.

Пример такого изображения представлен на рис. 8.



Исходное изображение



Преобразованное изображение

Рис. 8. Преобразование слабо различимого изображения с помощью ЦОДМ $Hd62$

На рис. 9 показано обработанное в частотной области с помощью ЦОДМ $Hd6_2$ изображение, представленное на рис. 8. Матрица множителей для усиления (подавления) коэффициентов преобразования на рис. 8 имеет вид:

$$\begin{bmatrix} 2 & 4 & 4 & 4 & 4 & 4 \\ 2 & 4 & 4 & 4 & 4 & 4 \\ 2 & 4 & 4 & 4 & 4 & 4 \\ 2 & 4 & 4 & 4 & 4 & 4 \\ 2 & 4 & 4 & 4 & 4 & 4 \\ 1 & 2 & 2 & 2 & 2 & 2 \end{bmatrix}$$



Исходное изображение



Обработанное изображение

Рис. 9. Обработка изображения с плохой видимостью

Необходимо отметить: то, что не вполне подходит для обработки изображений одного типа, вполне приемлемо для обработки изображений другого. Коэффициент контрастирования равен трём.

Кроме обработки в частотной области, усиливающей детальность и контрастирование после восстановления, обработка изображения, показанная на рис. 8, включает в себя повышение яркости. Поэтому яркость обработанного изображения в таком случае следует уменьшить. Это достигается двукратным уменьшением последнего слагаемого в уравнении (17).

$$y_{ij}^* = k \cdot (y_{ij} - mid) + mid / 2, \quad (18)$$

Деление на два, как правило, реализуется побитовым сдвигом вправо на 1 бит, то есть эта процедура является элементарной операцией и требует минимума вычислительных затрат.

Заключение

Разработанная в ЗАО «МНИТИ» система генерации ряда семейств целочисленных декоррелирующих матриц произвольных размеров как неортогональных (ЦНОДМ) [12], так и ортогональных (ЦОДМ), показала хорошие результаты при решении задачи сжатия цифровых изображений [1, 4].

Показано, что преобразования целочисленными матрицами действительно позволяют изменять яркость, контрастность и детальность цифрового монохромного изображения.

Подбор основных параметров обработки – матрицы множителей, коэффициента контрастирования и яркостной составляющей производится эмпирически.

Выбор преобразующей матрицы – «полуэмпирический». Проведено графическое сопоставление дискретных конечных функций разложения по ортогональному базису, которыми являются строки матрицы ЦОДМ, с другими известными методами, на основании чего, сделан выбор в пользу матрицы $Hd6_2$.

Большой спектр полученных после преобразования гармоник открывает широкие возможности для дальнейших исследований, которые вероятно помогут установить точные взаимосвязи между структурой изображения и параметрами его матричной обработки.

Литература

1. Евстигнеев В.Г., Бондаренко А.В., Кошарновский А.Н., Лебедев В.Д., «Дихотомические и недихотомические методы де-

корреляции цифровых изображений», «Системы и средства связи, телевидения и радиовещания», М., 2010, выпуск 1, 2, 6 с.

2. Вилкова Н.Н., Евстигнеев В.Г., Лебедев В.Д., Устройство для декорреляции цифровых изображений с использованием целочисленных ортогональных декоррелирующих матриц в системах компрессии видеоизображений, Патент РФ на изобретение № 2430419, Приоритет 10 июня 2010 г, Бюл. № 27., 12 с.

3. Бондаренко А.А., Евстигнеева О.В., Кошарновский А.Н., Лебедев В.Д., Способ формирования целочисленных ортогональных декоррелирующих матриц и устройство для его осуществления, Патент РФ на изобретение № 2509364, Приоритет 15 мая 2012 г., Бюл. № 17, – 9 с.

4. Евстигнеев В.Г., Лебедев В.Д., Куликова Е.И., НИР «Кедр», «Исследование и разработка новых принципов и алгоритмов декорреляции и восстановления цифровых телевизионных изображений произвольных форматов на основе нового класса целочисленных ортогональных декоррелирующих матриц произвольных размеров, новых методов и алгоритмов их генерации», М, ЗАО «МНИТИ», 2010 г., – 64 с.

5. Дагман Э.Е., Кухарев Г.А. «Быстрые дискретные ортогональные преобразования», издательство «Наука», Новосибирск, 1983 г., – 230 с.

6. Дворкович В.П., Дворкович А.В., «Цифровые видеоинформационные системы (Теория и практика)», Техносфера, 2012, – 1008 с.

7. Сэломон Д. «Сжатие данных, изображений и звука», ТЕХНОСФЕРА, Москва, 2004 г., – 365 с.

8. Поспелов Д.А. «Арифметические основы вычислительных машин дискретного действия», «Высшая Школа», М.: – 1970.

9. Папернов А.А. «Логические основы цифровых машин и программирования», «НАУКА», М., 1968., – 591 с.

10. Вилкова Н.Н., Евстигнеев В.Г., Сухачев А.Б., Шапиро Б.Л., Лебедев В.Д., «Некоторые вопросы повышения эффективности передачи видовой информации в комплексах управления беспилотными летательными аппаратами», М., «ВЕГА», 2012 г., – 13 с.

11. Травина Е.И. «Учтение ситуационной осведомленности оператора в бортовых системах наблюдения», М.: ФНПЦ ОАО «Концерн радиостроения «Вега», 2011, – С.53.

12. Лебедев В.Д. О неортогональных преобразованиях цифрового изображения, Техника средств связи, Серия Техника телевидения, М., 2013, – С. 98-108.

13. Хармут Х.Ф. «Передача информации ортогональными функциями», М., издательство «Связь», 1975., – 272 с.

14. Корн Г., Корн Т. Справочник по математике. – М.: Наука, 1974., – 720 с.

15. Головина Л.И. Линейная алгебра и некоторые её приложения», М., «Наука», 1979., – 392 с.

IMAGES IMPROVEMENT WITH INTEGER ORTHOGONAL TRANSFORMING MATRICES

Sheremet I.A., Full D., Rykin A.P., Lebedev V.D.

In this paper are described approaches for solving the task of improvement of digital images by transforming and reconstruction them with new class of integer orthogonal transforming matrices.

УДК 534.441

ГИБРИДНЫЙ ДЕТЕКТОР РЕЧИ

Вознесенская Т.В., к.ф.-м.н. доцент Департамента больших данных и информационного поиска Факультета компьютерных наук НИУ ВШЭ, e-mail: tvoznesenskaya@hse.ru;

Котов М.А., руководитель научно-технического департамента ООО «Стел-Компьютерные Системы», e-mail: kotov@stel.ru;

Леднов Д.А., к.т.н., с.н.с, научный консультант научно-технического департамента ООО «Стел-Компьютерные Системы», e-mail: lednov@stel.ru.

Ключевые слова: детектор речи, гибридный, звуковой поток, теория обнаружения, фрейм, линейчатый спектр, тестирование.

Введение

Детектор речи предназначен для выделения из входного звукового потока, состоящего из смеси полезного (речевого) сигнала и шума, последовательности сегментов, каждый из которых содержит фразу или слово. Задача обнаружения речи сходна с задачей, решаемой в рамках классической теории обнаружения стохастических сигналов и описанной во многих классических публикациях (см. например [1, 2]). Основное положение этой теории состоит в том, что риск принять неправильное решение минимален, если решение принимается в виде

$$\delta = \begin{cases} 1, & \text{если } \lambda > \mu \\ 0, & \text{если } \lambda \leq \mu \end{cases} \quad (1)$$

где решение $\delta = 1$, соответствует решению, что сигнал содержит речь, а решение $\delta = 0$, соответствует решению, что сигнал содержит только шум. В формуле (1) введены обозначения:

$$\mu = \frac{(1 - p_1)K(0, \delta_1)}{p_1 K(1, \delta_0)},$$

$$\lambda = \frac{p(y|1)}{p(y|0)}.$$

Отношение λ как правило, называют отношением правдоподобия, y – наблюдаемый сигнал, $p(y|\theta)$ – условная плотность распределения вероятности (ПРВ) наблюдаемого сигнала в зависимости от случайной величины $\theta = \{0, 1\}$, p_1 – априорная вероятность наличия полезного сигнала в наблюдаемом процессе, $K(\theta, \delta)$ – положительно определенная функция потерь, выбираемая эмпирически, как правило, в виде

$$K(\theta, \delta) = K_{\text{ос}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Реализация данной схемы для детектирования речи приводит к тому, что необходимо иметь представление о виде ПРВ $p(y|\theta)$. Поскольку знанием об истинной форме ПРВ речи и шума мы не обладаем, то для решения этой проблемы обычно предполагается, что данную

Рассматривается гибридный детектор речи, который построен из двух последовательно соединенных детекторов, обладающих различными принципами работы. Показано, что он способен выделять речь на фоне нестационарных шумов, обладающих сплошным спектром, при низких соотношениях сигнал/шум. Однако, такой детектор не способен отличить речь от речеподобных сигналов, обладающих линейчатыми спектрами, например, от музыки. Для решения этой задачи был использован известный детектор, основанный на вычислении отношения правдоподобия статистических моделей музыки и речи, полученных в процессе обучения. Приводятся экспериментальные результаты исследования работы гибридного детектора.

ПРВ можно аппроксимировать гауссовой смесью [3] вида

$$p(y|\theta) = \sum_{i=1}^n \alpha_i p_i(y|\Phi_i, \theta),$$

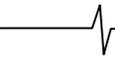
где $p_i(y|\Phi_i, \theta)$ – нормальная ПРВ с параметрами Φ_i , в качестве параметров используются ковариационная матрица и вектор математических ожиданий, α_i – априорная вероятность нормальной ПРВ, n – количество элементов смеси.

Следующий шаг, это вычисление параметров $\{\Phi_i|\theta\}$ гауссовой смеси. Для этого необходимо создать два множества записей – одно множество будет содержать образцы речи различных дикторов, а другое множество образцы различных типов шумов, а затем на основе собранного материала, используя известный EM-алгоритм [3], найти неизвестные параметры.

Сложность описанного алгоритма состоит в том, что нет практической возможности собрать образцы всех возможных типов неречевых сигналов, чтобы построить их статистические модели (плотности распределения вероятностей). В свою очередь это приводит к росту числа ошибок в работе детектора речи, как только требуется выделить речь на фоне подобных неречевых сигналов.

Возникает вопрос: может ли быть построен детектор, работа которого основана на некотором устойчивом признаке речи, отличающим ее от всех прочих шумов?

В настоящей работе в качестве такого устойчивого признака используем то, что речь обладает вокализованными интервалами, т.е. при произнесении каждого слога произносится гласный звук, спектр которого обладает признаками линейчатости [4]. Предположительно, этот подход позволит нам выделить речь на фоне широ-



кого класса шумов обладающих сплошным спектром, при этом шумы могут иметь даже нестационарный характер.

Бесспорно, признаком линейчатости спектров не обладает «шепотная» речь, и в противовес, этим признаком обладают музыкальные произведения. Что касается «шепотной» речи, то она не представляет практического значения и не встречается в сообщениях, требующих обработки. Что касается музыкальных произведений, то они будут являться источником ошибок ложных вызовов такого детектора. Однако, опыт практического использования детектора, основанного на статистических признаках, показывает, что он успешно справляется с задачей детектирования музыки. Это наводит на мысль о создании гибридного детектора речи, включающего в себя использование, как статистических, так и детерминированных характеристик речи.

В следующих разделах будут последовательно описаны методы выделения детерминированных характеристик и принципы построения детектора речи, основанного на этих характеристиках, а так же принципы построения гибридной модели детектора. В заключении будут приведены экспериментальные данные работы гибридного детектора.

Метод выделения линейчатой части спектра вокализованного звука

В качестве модели вокализованного звука используем следующее представление [4]:

$$x(t) = \int d\rho \sum_{i=1}^m K(\omega_i, \rho) \cos(\rho t + \phi(\rho)), \quad (1)$$

где в качестве ядра модели выбрано выражение

$$K(\omega_i, \tau_i, \rho) = A_i \exp\left(-\frac{(\omega_i - \rho)^2}{\tau_i^2}\right), \quad (2)$$

где τ_i ω_i – ширина и частота i -ой гармоники частоты основного тона, соответственно; $\phi(\rho)$ – фаза частоты ρ .

Такая модель сигнала позволяет нам ввести понятие ширины i -го обертона линейчатого спектра τ_i , которая всегда наблюдается при измерении спектра реального речевого сигнала. Причины возникновения ширины обертона могут быть различные: с одной стороны, это может быть изменение частоты обертона за время проведения измерений, а с другой стороны, длительность измерений может быть не кратна периоду обертона.

Для анализа спектра, заданного представлением (1), (2), воспользуемся вейвлет-преобразованием в частотной области [5]:

$$L(\omega, \tau, \{\omega_i, \tau_i\}) = \sum_{i=1}^m \theta(\omega, \tau, \omega_i, \tau_i) = \sum_{i=1}^m \int_0^{\Omega} \psi(\rho, \omega, \tau) K(\rho, \omega_i, \tau_i) d\rho \quad (3)$$

где $[0, \Omega]$ – диапазон спектра, $S(\rho, \{\omega_i\})$ – спектр Фурье вокализованного звука со своим множеством гармоник, $\psi(\rho, \omega, \tau)$ – вейвлет-функция, в качестве которой вы-

брана модифицированная «мексиканская шляпа»

$$\psi(\rho, \omega, \tau) = \frac{1.031}{\sqrt{2}\tau^{3/2}} \exp\left(-\frac{(\omega - \rho)^2}{\tau^2}\right) \left(1 - 2\frac{(\omega - \rho)^2}{\tau^2}\right), \quad (4)$$

где τ – масштаб вейвлет-функции.

Численное интегрирование показывает, что если спектр состоит из одного обертона ω_1 , ширина которого τ_1 , то максимум интеграла

$$(\omega_1, \tau_1) = \underset{\omega, \tau}{\text{ind max}} \theta(\omega, \tau, \omega_1, \tau_1)$$

соответствует положению и ширине обертона.

Этот вывод позволяет нам выполнить следующие операции:

- в диапазоне частот спектра $[\Omega_1, \Omega_2]$ для каждого значения частоты и каждого значения ширины обертона вычислить значения интегралов $\theta(\omega, \tau)$ (диапазон значений ширин обертонов выбирается из следующих соображений: ширина обертона не может быть менее, чем четверть минимальной частоты основного тона, и более, чем четверть максимальной частоты основного тона);

- для каждого значения частоты основного тона ω_0 рассчитать значение суммы

$$F(\omega_0) = \sum_{i=1}^m I(i)\theta(i\omega_0), \quad (5)$$

где индикаторная функция

$$I(i) = \begin{cases} 1, & \text{if } \theta(i\omega_0) > 0 \text{ and } (\theta((i-1)\omega_0) > 0 \text{ or } \theta((i+1)\omega_0) > 0), \\ 0, & \text{else} \end{cases}$$

применение которой позволяет избежать эффектов принятия за частоту основного тона единичных спектральных всплесков, и найти величину

$$\theta(i\omega_0) = \max_{\tau} \theta(i\omega_0, \tau);$$

- разделить диапазон частот основного тона (90 Гц-450 Гц) на три непересекающиеся области (90 Гц-179 Гц, 180 Гц-359 Гц и 360 Гц-450 Гц) и для каждой области найти максимальное значение суммы (5) и частоту основного тона, которая доставила сумме этот максимум, т.е.

$$F(\omega^*) = \max_{\omega_0} F(\omega_0),$$

$$\omega^* = \arg \max_{\omega_0} F(\omega_0).$$

Таким образом, результатом выполненных операций будет множество параметров $\{\omega_i^*, F(\omega_i^*), \theta(\omega_i^*)\}_{i=1, \dots, 3}$, вычисленное для каждого диапазона частоты основного тона в диапазоне спектра $[\Omega_1, \Omega_2]$.

Сделаем пояснения к изложенной последовательности операций. Известно, что методы выделения мгновенной частоты основного тона страдают ошибками ее удвоения. Этот эффект возникает при условии, что амплитуды четных обертонов многократно превосходят амплитуды нечетных обертонов основного тона [6]. Эффект можно ослабить, если разделить диапазон частот основного тона на три непересекающиеся области (границы которых были указаны выше) и проводить обработку речи

в каждой из выделенных областей независимо. Поскольку эффект удвоения частоты является неустойчивым во времени, совместная обработка некоторой последовательности фреймов, каждый из которых обработан с учетом множества диапазонов частот основного тона, может снизить этот эффект. Области выбраны так, чтобы каждая из них не содержала удвоенных частот.

Пусть найдены множества параметров для последовательности из M фреймов. Введем меру того, что частота основного тона ω_{it}^* , полученная в i -ом диапазоне t -ого фрейма, имеет свое продолжение в j -ом диапазоне $(t+1)$ -ого фрейма

$$\mu_{ijt} = \ln(F(\omega_{jt+1}^*)) - \frac{(\omega_{it}^* - \omega_{jt+1}^*)^2}{\sigma^2}, \quad (6)$$

где σ – допустимое изменение частоты основного тона от фрейма к фрейму (параметр модели). Тогда мера того, что за M фреймов частота основного тона двигалась по траектории по диапазонам с известными номерами $\{i_1, \dots, i_t, \dots, i_M\}$, будет равна

$$\beta(i_1, \dots, i_t, \dots, i_M) = \sum_{i=1}^{M-1} \mu_{i_{i+1}t}. \quad (7)$$

Задача состоит в том, чтобы среди всех возможных траекторий найти траекторию с максимальной мерой (7). Для решения этой задачи можно использовать метод динамического программирования, подробно описанный в [7]. Заметим, что последовательность номеров диапазонов основного тона, которые определены методом динамического программирования для максимума меры (7) при условии, что значение этого максимума превышает некоторый порог Q (параметр модели), т.е.

$$\max_{i_1, \dots, i_t, \dots, i_M} \beta(i_1, \dots, i_t, \dots, i_M) > Q, \quad (8)$$

Определяет саму частоту основного тона и множество амплитуд ее обертонов $\{\omega, \theta\}$. Если же максимум меньше порога, то принимается решение, что основного тона в данной последовательности фреймов нет.

Итак, выполнение неравенства (8) приводит к тому, что принимается решение о наличии в каждом фрейме из последовательности из M фреймов вокализованного звука. Далее происходит смещение интервала анализа последовательности, состоящей из M фреймов на один фрейм, и вновь полученная последовательность проходит анализ в соответствии с формулами (6-8). Здесь интересны два случая: 1) до некоторого последнего фрейма в последовательности с номером n_t неравенство (8) не выполнялось, а в следующий момент времени n_{t+1} неравенство (8) уже выполняется; 2) (случай обратный первому) до некоторого последнего фрейма в последовательности с номером n_t неравенство (8) выполняется, а в следующий момент времени n_{t+1} неравенство (8) не выполняется. Первый случай соответствует событию начала вокализованного звука – в качестве этого момента времени нами будет приниматься момент времени соответствующего началу фрейма с номером n_{t+1-M} . Второй случай соответствует событию окончания вокализованного звука – в качестве этого момента

времени нами будет приниматься момент времени соответствующий окончанию фрейма с номером n_t .

По сути, описанный метод дает нам значение длительности вокализованных звуков.

Структура гибридной модели

Известно, что каждый слог речи обязательно включает в себя вокализованный звук. В работе [8] подробно описаны характерные для речи частоты огибающих слогов. Если использовать низкочастотный фильтр с частотой среза равной средней частоте огибающей слогов при нормальном темпе речи (4 Гц), то можно сегментировать исходный звук на интервалы, предполагая, что каждый такой интервал включает в себя слог.

В качестве первичной обработки слога используем детектор, основанный на детерминированных характеристиках, который позволяет получить длительность вокализованной части слога внутри полученного интервала. Вычисляя отношение длительности вокализованной части d_v к длительности интервала D_i , мы определяем правдоподобность предположения, что внутри полученного интервала лежит слог. Если выполняется неравенство $\frac{d_v}{D_i} > Q_i$, где Q_i – параметр модели, то принимается решение о наличии слога на полученном интервале.

Вторичная обработка интервала производится только в том случае, если на первом этапе было принято положительное решение о звучании слога. Вторичная обработка состоит в вычислении отношения правдоподобия вида

$$\lambda = \prod_{i=1}^R \frac{p(x_i | speech)}{p(x_i | music)},$$

где R – количество фреймов в полученном интервале, $p(x_i | speech)$, $p(x_i | music)$ – условные вероятности того, что полученные в момент времени t характеристики фрейма x_t порождены речью или музыкой соответственно. В качестве характеристик фрейма нами были выбраны известные коэффициенты (mel-frequency cepstral coefficients) MFCC [9, 10].

Если выполняется неравенство $\lambda > Q_\lambda$, то принимается решение, что в интервале звучит речь, в противном случае – музыка.

Экспериментальные результаты

Прежде чем описывать экспериментальные результаты приведем технические характеристики системы, условия ее подготовки к экспериментам, а так же опишем тестирующую базу звуков.

Входной сигнал был оцифрован с частотой 8 кГц и имел разрядность – 16 бит. Для вычисления быстрого преобразования Фурье, которое как часть входит и в процедуру обнаружения линейчатых спектров и в вычисление коэффициентов MFCC, использовалось окно длительностью 64 мс, величина смещения окна во времени составляла 10 мс. В качестве фильтра низкой частоты использовался фильтр Баттерворта 5-го порядка. Частота среза фильтра ω_c была параметром модели. Для по-

Соотношение сигнал/шум (дБ)	Значение равновероятной ошибки необучаемого детектора(%)	Значение равновероятной ошибки обучаемого, статистического детектора при распознавании речи и музыки (%)
10-15	9.86	1.08
6-10	19.435	4.43
2-6	29.185	9.12

строения статистических моделей речи и музыки использовались гауссовы смеси, состоящие из 16-ти элементов.

Для поиска параметров детерминированной модели, основанной на выделении вокализованных звуков, а так же для ее тестирования была создана речевая база, содержащая речь на фоне шумов КВ-канала радиоприемника, промышленных шумов (работа отбойного молотка, экскаватора, шум проезжающих машин, шум в вагоне метро и т.д.), бытовых и естественных шумов (шаги, работа вентилятора, шум листвы деревьев и т.д.). База была размечена, т.е. были установлены границы интервалов, включающих в себя речь.

Речевая база была разделена на две части, одна из которых использовалась для оптимизации параметров модели, а другая для тестирования. Длительность базы, которая использовалась для оптимизации параметров, составляла 1 час, она включала в себя 36 минут речи. Критерием оптимизации было минимальное значение суммы ошибок пропуска цели и ложного срабатывания. Приведем значения параметров, полученных в результате оптимизации: σ (допустимое изменение частоты основного тона от фрейма к фрейму) = 40 Гц; Q (порог меры (7)) = 0,31; Q_1 (отношение длительности вокализованной части слога к длительности слога) = 0.21.

Для тестирования использовалась база длительностью около 6 часов, которая включала в себя около 2-х часов речи. Выделенная детерминированным детектором из этой базы речь использовалась для создания ее модели, которая используется статистическим детектором. Для обучения статистической модели музыки был создан музыкальный корпус, включающий в себя около 8-и часов фрагментов музыкальных произведений различных темпов и стилей. Для тестирования статистической части детектора использовались размеченные записи, содержащие речевые и музыкальные фрагменты, каждый из таких фрагментов идентифицировался. Общая длительность записей составляла около полутора часов, из них около 20 минут музыки и около 30 минут речи.

Во втором столбце табл. 1 показаны значения равновероятной ошибки обнаружения речи в зависимости от соотношения сигнал/шум в поданных на вход необучаемого детектора записях. В третьем столбце табл. 1 показаны значения равновероятной ошибки классификации речевых и музыкальных фрагментов в зависимости от соотношения сигнал/шум в поданных на вход статистического детектора записях.

Заключение

По сути, созданный гибридный детектор, это смесь детерминированного детектора и системы распознава-

ния. В настоящей работе были использованы только два класса распознавания: речь и музыка, но в общем случае количество классов может быть увеличено. В качестве классов могут быть, например, использованы голоса дикторов, которые нас не интересуют, или же языки, которые нас не интересуют и т.д.

Точность, полученная при высоких соотношениях сигнал/шум, сравнима с точностью, с которой одну и ту же запись сегментируют два эксперта. Поскольку каждый файл, использованной для тестирования базы, был обработан одним экспертом, то мы не могли собрать достоверную статистику в попытке ответить на вопрос, насколько наш детектор отличается от средней экспертной разметки. Анализ разницы между экспертной и детектированной разметками в основном состоит из разницы в определении границ слов и фраз, а также в расстановке пауз между словами. Эксперт имел склонность выделять слитную фразу целиком в отличие от автомата, который аккуратно расставлял паузы между словами.

При разработке детерминированного детектора было сделано допущение, которое существенным образом повлияло на его работу. В формуле (6) явным образом предполагается, что частота основного тона не должна изменяться от фрейма к фрейму, именно такая ситуация приведет к максимуму функционал (7), хотя практика показывает, что частота основного тона не является устойчивой. Это в свою очередь говорит о том, что должна быть разработана более совершенная модель, которая учитывает естественную неустойчивость частоты основного тона.

Литература

1. Большаков И.А. Статистические проблемы выделения потока сигналов из шума // Изд-во «Советское радио», 1969.
2. Харкевич А.А. Борьба с помехами, Изд. Второе, Из. «Наука», М. 1965.
3. Jeff A. Bilmes A Gentle Tutorial of the EM algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models// ICSI Technical Report-021, April 1998.
4. Аграновский А.В., Леднов Д.А. Репалов С.А. Метод текстнезависимой идентификации диктора на основе индивидуальности произношения гласных звуков // Акустика и прикладная лингвистика. Ежегодник РАО. Выпуск 3. М., 2002, стр. 103-115
5. Котов М.А., Леднов Д.А. и др. Способ определения параметров линейчатых спектров вокализованных звуков и система для его реализации № 2007148606/09(053252) от 27.12.2007

6. Babkin A.A. LPC Speech Coder AT 1000-1200 BPS // In Proceedings of DSPA-2000

7. Моттль В.В., Мучник И.Б. Скрытые Марковские модели в структурном анализе сигналов. – М.: ФИЗМАТ-ЛИТ, 1999

8. Steven Greenberg, Hannah Carvey, Leah Hitchcock, Shawn Chang Temporal properties of spontaneous speech – a syllable-centric perspective // Journal of Phonetics 2002, 31, pp.465-485.

9. Mermelstein P. (1976), Distance measures for speech recognition, psychological and instrumental in Pattern Recognition and Artificial Intelligence, C.H. Chen, Ed., pp. 374–388. Academic, New York.

10. Davis S.B., and Mermelstein P. (1980), Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences, in IEEE Transactions on Acoustics, Speech, and Signal Processing, 28(4), pp. 357–366.

HYBRID VOICE ACTIVITY DETECTOR

Voznesenskya T.V., Kotov M.A., Lednov D.A.

This paper focuses on a hybrid voice activity detector (VAD) which is built from two detectors connected sequentially with different operating principles. The first detector, developed by the authors, is based pitch detection. This detector uses wavelet analysis applied in the spectral domain. It is shown that this detector is capable to separate speech from the non-stationary background noise, which has a continuous spectrum, at low signal/noise ratio. The experimental results of detector investigation are shown. However, such a detector is not able to distinguish speech from the speech-like signals that have a line spectrum e.g., music. For this purpose a well-known detector is used, based on the likelihood ratio of statistic models of music and speech. The experimental results of hybrid VAD investigation are shown.

НОВЫЕ КНИГИ



Солонина А.И.

ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ И MATLAB:

учеб. пособие / А.И. Солонина, Д.М. Клинский, Т.В. Меркучева, С.Н. Перов. – СПб.: БХВ-Петербург, 2013. – 512 с. (Учебная литература для вузов) Москва: Техносфера, 2013. – 528 с.

Описываются базовые методы и алгоритмы цифровой обработки сигналов и средств их компьютерного моделирования в системе MATLAB. Даны основы алгоритмического языка MATLAB. Рассматриваются дискретные сигналы, линейные дискретные системы, дискретное преобразование Фурье с использованием алгоритмов БПФ, синтез и анализ КИХ- и БИХ-фильтров, в том числе с фиксированной точкой, спектральный анализ сигналов, многоскоростная обработка сигналов и адаптивная цифровая фильтрация.

Технология обучения в процессе компьютерного моделирования на основе созданных авторами программ или графического интерфейса пользователя MATLAB расширяет теоретические знания и позволяет понять многие важные проблемы и аспекты практического применения методов и алгоритмов ЦОС. На прилагаемом к

книге CD хранятся обучающие программы и таблицы исходных данных.

Предназначена для студентов, аспирантов и преподавателей вузов, а также специалистов в области цифровой обработки сигналов.

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:

1. Текст - текстовый редактор Microsoft Word.
2. Таблицы и рисунки должны быть пронумерованы. На все рисунки, таблицы и библиографические данные указываются ссылки в тексте статьи.
3. Объем статьи до 12 стр. (шрифт 12). Для заказных обзорных работ объем может быть увеличен до 20 стр.
4. Название статьи на русском и английском языках.
5. Рукопись статьи сопровождается:
 - краткой аннотацией на русском и английском языках;
 - номером УДК;
 - сведениями об авторах (Ф.И.О., организация, должность, ученая степень, телефоны, электронная почта);
 - ключевыми словами;
 - актом экспертизы (при наличии в вашей организации экспертной комиссии).

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, jpg.

ПОВЫШЕНИЕ ПРОИЗВОДИТЕЛЬНОСТИ ГЕНЕТИЧЕСКОГО АЛГОРИТМА, КОНСТРУИРУЮЩЕГО ЦИФРОВЫЕ ФИЛЬТРЫ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИИ CUDA

Белобродский В.А., аспирант факультета компьютерных наук Воронежского государственного университета, e-mail: belobrodsky@yandex.ru

Ключевые слова: параллельные вычисления, технология CUDA, математическое моделирование, алгоритмы обработки данных, биомедицинские сигналы.

Введение

Несмотря на впечатляющий рост вычислительных мощностей персональных компьютеров, однопоточные программы могут требовать значительного времени, особенно если требуется обработать большой объем входных данных, который свойственен многим зарегистрированным биомедицинским сигналам. Очевидно, что использование такого количества времени на ожидание результата работы неприемлемо, и этот факт делает актуальной проблему поиска наиболее «узких» и ресурсоемких мест алгоритма, с проведением дальнейшей их оптимизации с целью уменьшения времени вычисления. Эффективным способом уменьшения времени выполнения генетического алгоритма является его распараллеливание. Для осуществления этой цели очевидным кажется применение параллельного кластера для проведения вычислений, однако данный подход требует значительных организационных мероприятий и определенной технической подготовки и подчас серьезных материальных вложений.

Начиная с 2006 года появилась возможность ускорить время выполнения не графических вычислений, используя вычислительные мощности видеокарт, за счет предложенной компанией Nvidia технологии CUDA (Compute Unified Device Architecture) [1, 3, 4, 7-9]. В данной работе приводится анализ производительности двух версий программ, реализующих генетический алгоритм, одна из которых использует технологию CUDA.

Метод создания цифровых фильтров на основе генетического алгоритма

Рассмотрим теоретические положения, которые положены в основу предлагаемого метода конструирования цифровых фильтров для анализа биомедицинских сигналов с использованием идеологии генетических алгоритмов [2]. Пусть имеется поколение функций фильтров $Q_m^n(t)$ – его импульсных характеристик, где верхним индексом n обозначен номер поколения, нижним индексом m – номер фильтра внутри поколения, t – дискретное время. При этом фильтр $Q_m^n(t)$ имеет нулевое среднее и единичную норму. Выполним операцию

Разработана компьютерная программа, реализующая генетический алгоритм для конструирования цифровых фильтров с применением технологии CUDA. Проведены серии вычислительных экспериментов для экспериментального определения производительности созданной программы, которая является функцией нескольких аргументов. Приведены экспериментальные результаты замера производительности двух версий программ, которые показывают преимущество версии, использующей технологию CUDA, до 5-6 раз по сравнению с программой, не использующей данную технологию. Под производительностью в настоящей статье понимается количество популяций цифровых фильтров, генерируемых программой за 1 час работы.

свёртки сигнала $E_x^k(t)$ (с известными свойствами) с фильтром $Q_m^n(t)$:

$$S_m^n(\tau, x, k) = \int Q_m^n(t) E_x^k(t - \tau) dt. \quad (1)$$

Здесь $S_m^n(\tau, x, k)$ – коэффициенты свёртки сигнала $E_x^k(t)$ с фильтром $Q_m^n(t)$; x – порядковый номер сигнала в k -ой группе.

Для простоты в дальнейшем будем рассматривать две группы k сигналов ($k = k_1, k_2$). Свернем все сигналы каждой из этих двух групп k_1, k_2 со всеми фильтрами $Q_m^n(t)$. Усреднив максимальные модули разницы коэффициентов сверток среди всевозможных пар сигналов, принадлежащих разным группам, можно рассчитать величину Φ_m^n , которую будем называть «эффективностью» или «параметром приспособленности» фильтра Q_m^n (в литературе также встречается термин «фитнес» [5]):

$$\Phi_m^n = \frac{\sum_{i=1}^{x^{\max}} \sum_{j=1}^{x^{\max}} \text{MAX} \left| (S_m^n(\tau, x^i, k_1) - S_m^n(\tau, x^j, k_2)) \right|}{(x^{\max})^2}. \quad (2)$$

Здесь операция MAX означает, что из всего одномерного массива, составленного из модулей разницы сверток двух сигналов $\left| (S_m^n(\tau, x^i, k_1) - S_m^n(\tau, x^j, k_2)) \right|$, принадлежащих разным группам, берется только максимальное значение. В дальнейшем будет находиться среднее значение максимального модуля разницы среди всех пар, общее число которых $(x^{\max})^2$, так как здесь используется две группы сигналов по x^{\max} сигналов в каждой. Таким образом, согласно формуле (2), каждому фильтру текущего поколения Q_m^n ставится в соответствие определенное число Φ_m^n . Очевидно, что фильтры, позволяющие выявить наибольшие различия между исследуемыми сигналами, – это те, которые имеют макси-

мальные значения Φ^n_m . Они и должны быть допущены к «скрещиванию». «Скрещивание» происходит во временной области по формуле

$$Q_m^{n+1}(t) = (Q_{m_1}^n(t) + Q_{m_2}^n(t)) / 2. \quad (3)$$

С помощью «скрещивания» создается часть следующего поколения фильтров Q^{n+1} . В наших экспериментах – это 50% от всей «популяции» фильтров, а остальные 50% «популяции» заполняются лучшими фильтрами из предыдущего поколения без каких-либо их изменений. Другими словами, $(n+1)$ -ое поколение фильтров на 50% состоит из фильтров предыдущего поколения, демонстрирующих наилучшие значения величины Φ^n , и на 50% – из их потомков. Таким образом, «удачный» фильтр будет сохраняться во всём ряду «популяции», что позволяет избежать ситуации, когда фильтр, дающий высокие значения Φ^n в процессе «скрещивания», потеряет свои свойства. После формирования вышеприведенным образом нового поколения, запускается процедура, определяющая необходимость мутации с помощью логической функции

$$F(Q^n) = \begin{cases} 1, \eta > 0,5 \\ 0, \eta \leq 0,5 \end{cases}, \quad (4)$$

где η – доля отсчетов дискретных фильтров, в которых коэффициент вариации V_i , определяемый ниже по формуле (5), меньше определенного значения (в наших экспериментах в качестве этого значения мы взяли 0,2).

Коэффициент вариации определяется как

$$V_i = \frac{\sigma(Q^n(i))}{Q^n(i)}, \quad (5)$$

где $\sigma(Q^n(i))$ – среднеквадратическое отклонение i -го отсчета всех фильтров n -го поколения; $\overline{Q^n(i)}$ – среднее значение i -го отсчета всех фильтров n -го поколения.

В случае, если логическая функция (4) указывает на необходимость «мутации» (возвращает значение «истина»), то у каждого фильтра, созданного в текущем поколении, а не перенесенного из предыдущего, есть определенная вероятность претерпеть одну из следующих «мутаций»: «кроссинговер»; «кроссинговер» и случайный сдвиг фазы; добавление фликкер-шума; добавление фликкер-шума и случайный сдвиг фазы; добавление сине-зеленого шума; добавление сине-зеленого шума и случайный сдвиг фазы; случайный сдвиг фазы. В дальнейшем работа генетического алгоритма представляет собой итерационный процесс, который продолжается до тех пор, пока не сформируется заданное число поколений или не выполнится критерий остановки. Одним из возможных критериев остановки работы алгоритма может быть следующее условие:

$$|Max(\Phi^n) - Max(\Phi^{n+1})| < \varepsilon, \quad (6)$$

где ε – порог остановки алгоритма, задаваемый пользователем; $Max(\Phi^n)$ – максимальное значение параметра приспособленности в поколении n .

О выборе критерия для сравнения производительности

Одной из важнейших характеристик вычислительно-

го устройства является его быстродействие, которое для математических расчетов обычно измеряется в количестве операций с плавающей точкой в секунду (flops) [3].

Несмотря на то, что в современных языках программирования, например С#, вводится упрощенный подход для создания асинхронных программ [6], как правило, в обычной практике, программа не сможет продолжать свои вычисления без завершения других процедур (нитей), так как для их продолжения нужны входные данные, которые еще рассчитываются. Поэтому конечному пользователю важен интегральный показатель быстродействия ресурсоемких программ, как правило, время, измеряемое в секундах, необходимое до завершения вычислительного процесса какой-либо задачи. Программе может понадобиться априори неизвестное количество времени на завершение работы, особенно ввиду того, что неизвестно, сколько поколений потребуется сгенерировать, чтобы сработало условие выхода из итерационного процесса генерации. А каждая конкретная итерация будет длиться в зависимости от исходных данных, а именно от длины сигналов N , длины фильтров M , количестве сигналов и фильтров (K и L соответственно). Стоит отметить, что количество фильтров в каждом поколении сохраняется постоянным на протяжении всей работы программы. В связи с этим, в настоящей статье приводится сравнительная характеристика и показывается преимущество применения технологии CUDA в единицах производительности программного обеспечения, реализующего генетический алгоритм W , под которым будем понимать количество генерируемых «популяций» за 1 час работы. Очевидно, что количество генерируемых за 1 час работы поколений (производительность W) является функцией от нескольких переменных – таких, как длина сигналов N , длина фильтров M , количество сигналов K и количество фильтров L .

Результаты исследований

Разработанная методика конструирования значений цифровых фильтров, опирающаяся на методологию генетических алгоритмов была реализована на языке программирования Visual C# без использования каких-либо механизмов распараллеливания. Первые вычислительные эксперименты, проведенные с помощью разработанной программы, привели к удовлетворительным результатам, но потребовали значительное время ожидания (≈ 1 суток).

Как было упомянуто выше, для увеличения значений функции производительности $W(N, M, K, L)$ была создана вторая версия программного обеспечения, реализующая данный генетический алгоритм и использующая технологию CUDA для вычислений математической свертки, которая в указанном во введении примере запускалась $100 \cdot 100 \cdot 50 = 500000$ раз в каждом поколении, так как на вход программы подавались две группы по 100 сигналов в каждой, при заданном числе фильтров в каждом поколении, равном 50.

Для проведения сравнительного анализа производительности созданных программ было произведено 4 серии вычислительных экспериментов, в каждом из которых варьировался один из 4 параметров (N, M, K, L), от которых зависит функция производительности $W(N, M,$

K, L), а остальные 3 параметра принимались за константу. Значение функции производительности, выраженное в количестве поколений, создаваемых за 1 час работы программы, определялось эмпирическим путем по формуле $W(N, M, K, L) = t^{-1}$, где t – время, за которое программа генерирует одно поколение фильтров.

Это время определялось средствами ОС Windows во времени создания выходных файлов, в которых приводится детальная информация каждого из поколений сразу после их создания. На рис. 1-4 приведены зависимости функции $W(N, M, K, L)$ от значений ее параметров.

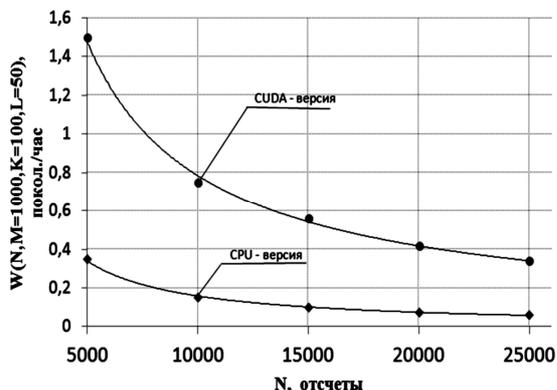


Рис. 1. Зависимость функции производительности $W(N, M = 1000, K = 100, L = 50)$ от длины сигналов N

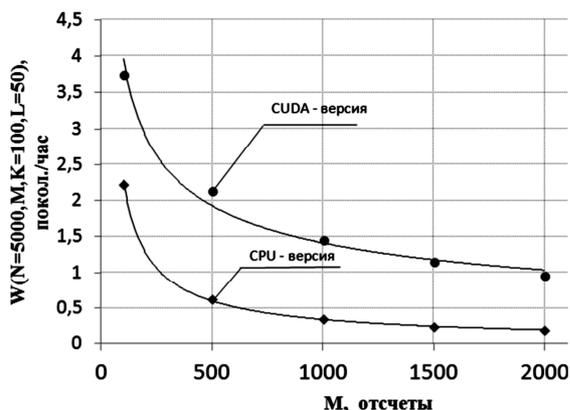


Рис. 2. Зависимость функции производительности $W(N = 5000, M, K = 100, L = 50)$ от длины фильтров M

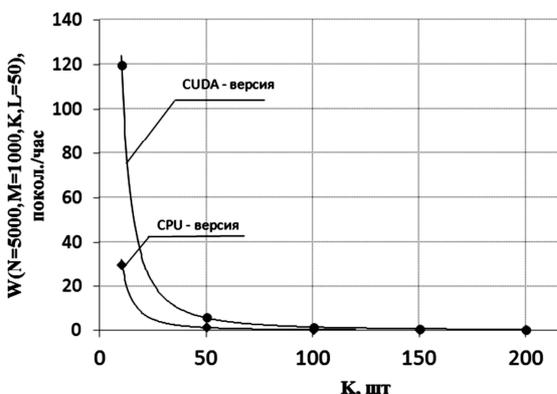


Рис. 3. Зависимость функции производительности $W(N = 5000, M = 1000, K, L = 50)$ от количества сигналов в каждой группе K

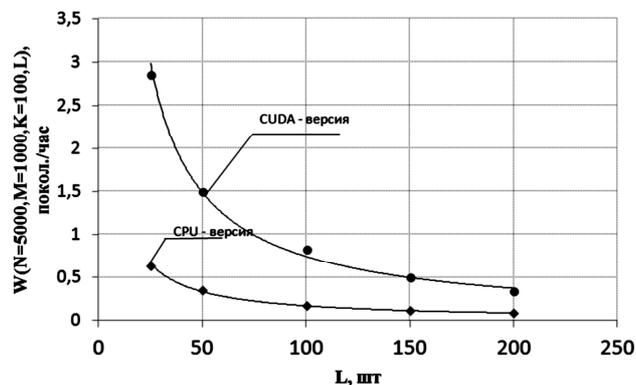


Рис. 4. Зависимость функции производительности $W(N = 10000, M = 1000, K = 100, L)$ от количества фильтров в каждой «популяции» L

Во всех приведенных примерах аппроксимирующие кривые (сплошные черные линии) описывались степенными функциями типа $f = Ax^b$, где показатель степени b был близок к минус единице и принимал значения от -0,989 до -1,098

Стоит отметить, что приведенные результаты имеют некоторую неучтенную систематическую погрешность, которая связана с тем, что в данной реализации генетического алгоритма, проводятся дополнительные преобразования типов, необходимые для запуска процедуры вычисления сверток, такие, как преобразование из типа коллекции `List<float>` в массив `float[]` и обратно после вычислений и т.п. Очевидно, что эти неопределенности прямо пропорциональны длине сигнала и количеству сигналов, и увеличивают время выполнения программы использующей CUDA, но несмотря на это, из рис. 1-4 видно её преимущество до 5 раз по величине производительности. Так же рис. 1-4 подтвердили, что наиболее ресурсоемкое место программы было определено верно, так как достигается 5-кратное увеличение производительность CUDA версии.

Заключение

Анализ производительности двух реализаций генетического алгоритма, предназначенного для генерации значений цифровых фильтров, наглядно продемонстрировал практическое преимущество внедрения технологии CUDA.

Стоит отметить, что полученное увеличение производительности работы алгоритма соответствует применению технологии CUDA только в одном, самом ресурсоемком месте кода – осуществлении процедуры математической свертки, которая при определенных значениях входных параметров (а именно, 200 – общее число файлов, 200 – число фильтров в каждой популяции) выполняется до $2 \cdot 10^6$ раз в каждой «популяции».

С другой стороны, по своей сути генетический алгоритм легко может быть подвержен дальнейшему распараллеливанию, например, из-за независимости параметров оптимизации одних фильтров от других можно производить независимые вычисления этих показателей друг от друга. Таким образом, в перспективе остается потенциал для дальнейшего увеличения производительности за счет переноса других расчетных модулей

генетического алгоритма с CPU на GPU вычисления, таких, как «скрещивание», генерация «мутаций» и т.п.

Автор выражает особую благодарность Кургалину С.Д., Туровскому Я.А., Вахтину А.А.

Литература

1. Вахтин А.А., Туровский Я.А. Реализация численного вейвлетного преобразования на графических адаптерах архитектуры NVIDIA CUDA // Вестник Воронежского Государственного Университета. Серия: Системный анализ и информационные технологии. – 2012. – №1. – С. 69-72.

2. Туровский А.Я. Создание фильтров для анализа ЭЭГ-состояний на основе генетических алгоритмов // Программная инженерия. – 2014. – №6. – С. 23-28.

3. Боресков А.В., Харламов А.А. Основы работы с технологией CUDA. М.: ДМК Пресс, 2010. – 232 с.: ил.

4. Электронный ресурс: <http://www.nvidia.ru/object/cuda-parallel-computing-ru.html>

5. Holland J.H. Adaptation in Natural and Artificial System. – MIT Press, 1992. – 205 с.

6. Snell M., Powers L. Microsoft Visual Studios C#.

7. Казённов А.М. Основы технологии CUDA.

8. Берилло А. NVIDIA CUDA – неграфические вычисления на графических процессорах (<http://www.ixbt.com/video3/cuda-1.shtml>)

9. Сандерс Дж., Кэндрот Э. Технология CUDA в примерах: введение в программирование графических процессоров. – М.: ДМК Пресс, 2011. – 232 с.

IMPROVING PERFORMANCE OF GENETIC ALGORITHM THAT CONSTRUCTS DIGITAL FILTERS BY USING OF CUDA TECHNOLOGY

Belobrodskiy V.A.

This article describes a computer program that implements a genetic algorithm for designing digital filters using the technology of CUDA. A series of computational experiments to determine the performance of the software, which is a function of several arguments, is performed.

The article shows the experimental results of measuring the performance of two version of software, which show that the version which uses the CUDA technology is 5-6 times more productive. By performance the author understands the number of digital filters populations generated by the program an hour.

AUTEX Ltd.



17-я Международная Конференция

ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ И ЕЕ ПРИМЕНЕНИЕ

The 17th International Conference DIGITAL SIGNAL PROCESSING AND ITS APPLICATIONS

Москва, 25 марта – 27 марта 2015 года

Уважаемые коллеги!

ПРИГЛАШАЕМ ВАС ПРИНЯТЬ УЧАСТИЕ В РАБОТЕ КОНФЕРЕНЦИИ

ОРГАНИЗАТОРЫ:

- | | |
|---|---|
| <ul style="list-style-type: none"> • Российское научно-техническое общество радиотехники, электроники и связи им. А.С. Попова • Институт радиотехники и электроники РАН • Компания AUTEX Ltd. (ЗАО «АВТЭКС») • Российская секция IEEE | <ul style="list-style-type: none"> • IEEE Signal Processing Society • Институт проблем управления РАН • Институт проблем передачи информации РАН • Московский научно-исследовательский телевизионный институт (ЗАО МНИТИ) |
|---|---|

РАБОТА КОНФЕРЕНЦИИ ПЛАНИРУЕТСЯ ПО 10 СЕКЦИЯМ:

(укажите соответствующей № секции в заявке к докладу)

- | | |
|--|---|
| <ol style="list-style-type: none"> 1. Теория сигналов и систем 2. Теория и методы ЦОС 3. Обработка сигналов в системах телекоммуникаций 4. Обработка сигналов в радиотехнических системах 5. Обработка и передача изображений | <ol style="list-style-type: none"> 6. Обработка и передача измерительной информации 7. Проектирование и техническая реализация систем ЦОС 8. Цифровое телерадиовещание 9. Цифровая обработка многомерных сигналов 10. Нейрокомпьютерная обработка сигналов и изображений |
|--|---|

СРОКИ ПРЕДСТАВЛЕНИЯ ДОКЛАДОВ:

(информация о регистрации обновляется на сайте: <http://www.rntores.ru>)

Реквизиты для перечисления взносов: РНТОРЭС имени А.С. Попова, ИНН 7702021967, КПП 770201001, БИК 044525225, Р/сч. 40703810038090105080 Московский банк Сбербанка России ОАО, г. Москва.

Кор/счет 3010181040000000225

Назначение платежа: «Целевой взнос на конференцию ДСПА-2015», НДС не облагается.

ИССЛЕДОВАНИЕ ЦИФРОВОЙ МОДЕЛИ НЕЛИНЕЙНОГО АКТИВНОГО ФИЛЬТРА ПЕРВОГО ПОРЯДКА

Аболазм М.А., аспирант Московского государственного технического университета радиотехники, электроники и автоматики (МГТУ МИРЭА);

Мельчаков В.Н., старший преподаватель МГТУ МИРЭА;

Решетняк С.А., д.ф.-м.н., профессор МГТУ МИРЭА, e-mail: reshets@bk.ru;

Третьяков Г.Н., к.т.н., доцент МГТУ МИРЭА, e-mail: gennady.tretyakov@rambler.ru.

Ключевые слова: отношение сигнал/шум, нелинейные фильтры низких и высоких частот.

Введение

В последнее время повышенный интерес вызывают процессы прохождения слабых сигналов и шума через нелинейные системы. Он вызван аномальным поведением выходного отношения сигнал/шум (S/N) и, в частности, открытием явления стохастического резонанса [1-3] и стохастической фильтрации сигналов [4, 5]. С практической точки зрения наиболее интересен эффект стохастической фильтрации сигналов, позволяющий достичь в радиотехнических системах выходные отношения S/N , превышающие входные значения [6, 7].

В основе исследований указанных процессов лежат стохастические дифференциальные уравнения, называемые ланжевеновскими. С целью получения аналитического решения используют подход, в котором анализируются эквивалентные им уравнения Фоккера-Планка для плотности вероятности [8, 9]. Однако уравнения в частных производных не всегда поддаются решению, поэтому данный подход имеет весьма ограниченную область применения. В спектральном методе [10] осуществляется переход от ланжевеновских уравнений к нелинейной системе алгебраических уравнений для спектральных компонент выходного сигнала, решение которой удаётся получить только в первом порядке метода итераций.

В общем случае прохождение сигнала и шума через нелинейные системы можно проанализировать на основе численного решения стохастических дифференциальных уравнений. По существу дела, численный анализ можно объединить с проведением прямых экспериментов в радиотехнических устройствах. В отличие от существующих в литературе работ по данной проблеме здесь численный анализ применяется для исследования эффекта стохастической фильтрации в нелинейных активных фильтрах как низких, так и высоких частот. Полученные выходные отношения S/N сравниваются с аналогичными отношениями на входе фильтров и указывают на наличие эффекта стохастической фильтрации в определённой области параметров фильтра, сигнала и шума.

В работе исследовались нелинейные фильтры низких и высоких частот (ФНЧ и ФВЧ), схемы которых изображены на рис. 1.

При устранении нелинейного элемента они представляют собой обычные линейные радиотехнические фильтры указанных частот. Нелинейный элемент содержит два одинаковых туннельных диода с источниками напряжения смещения E , соединённые параллельно. Левые полюса этих фильтров –

Развит метод численного исследования отношения сигнал/шум в нелинейном активном фильтре первого порядка. Показано, что при действии гармонического сигнала и шума на входе фильтра выходное отношение сигнал/шум превосходит входное при определённых параметрах как в фильтре низкой, так и в фильтре высокой частоты. Результаты анализа удовлетворительно согласуются с экспериментальными данными.

входные, а правые – выходные. На вход фильтров подавалась смесь гармонического сигнала и шума, а на их выходе анализировалось напряжение между полюсами.

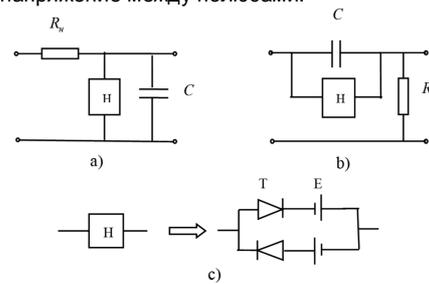


Рис. 1. Схемы ФНЧ и ФВЧ представлены соответственно на а) и б). Вид нелинейного активного элемента с двумя одинаковыми туннельными диодами (ТД) и постоянными источниками напряжения E смещения представлен на с)

Если через u обозначить напряжение на ёмкости, то для ФНЧ и ФВЧ справедливо уравнение

$$R_n C \frac{du}{dt} + R_n I(u) + u = a_c \cos(\omega_c t) + \xi(t), \quad (1)$$

где ω_c и a_c – угловая частота и амплитуда гармонического сигнала, $\xi(t)$ – стационарный гауссовый случайный процесс с конечным временем корреляции, $I(u)$ – ток через нелинейный элемент или его вольт-амперная характеристика.

Простой анализ показывает, что

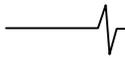
$$I(u) = i_o(E + u) - i_o(E - u),$$

где $i_o(u)$ – вольт-амперная характеристика отдельного туннельного диода. Поскольку $I(u)$ – нечётная функция переменной u , то её ряд по степеням u имеет вид

$$I(u) = Au + Bu^3 + \dots, \quad (2)$$

$$\text{где } A = \left. \frac{dI}{du} \right|_{u=0} = 2 \left. \frac{di_o}{du} \right|_{u=E}.$$

Выбираем смещение E таким, чтобы оказаться в области отрицательных сопротивлений туннельного диода. При этом коэффициент A также принимает отрицательные значения. Прямые



измерения $I(u)$ показывают [11], что в достаточно широкой области напряжений вольт-амперная характеристика нелинейного элемента описывается двумя первыми членами ряда (2).

От величины сопротивления R_n зависит вид уравнения (1). В данной работе рассматривался критический случай, когда $R_n = |A|^{-1}$, и после введения масштабов напряжения $u_M = (R_n B)^{-1/2}$ и времени $t_M = R_n C$ безразмерное уравнение (1) принимает вид

$$\frac{du}{dt} + u^3 = h \cos(\omega_c t) + \zeta(t), \quad (3)$$

где u и t – нормированные на свои масштабы напряжение и время, $h = a_c / u_M$, $\zeta = \xi / u_M$, а под безразмерной частотой ω_c следует понимать величину $\omega_c t_M$.

В численном анализе уравнения (3) использовался метод Эйлера. Пусть t_n – точки отсчётов времени анализируемого процесса с шагом $\Delta t = t_n - t_{n-1}$, а $u_n = u(t_n)$. Тогда из (3) следует, что

$$u_n = u_{n-1} + \int_{t_{n-1}}^{t_n} f(u, t') dt', \quad f(u, t) = -u^3 + e(t), \quad (4)$$

где $e(t) = h \cos(\omega_c t) + \zeta(t)$ – напряжение на входе фильтра.

Оценка интеграла в (4) по методу прямоугольников приводит к следующему рекуррентному соотношению:

$$\begin{aligned} u_n &= u_{n-1} + \Delta t \cdot f(u_{n-1}, t_{n-1}) = \\ &= u_{n-1} + \Delta t [-u_{n-1}^3 + e(t_{n-1})]. \end{aligned} \quad (5)$$

Отметим, что в (5) выходной сигнал на n -ом шаге u_n определяется своими значениями на предыдущих шагах по времени, поэтому при заданном начальном условии $u_0 = u(0)$ все u_n легко находятся.

Оценка интеграла в (4) по методу трапеций приводит к более точной формуле:

$$\begin{aligned} u_n &= u_{n-1} + \frac{\Delta t}{2} [f(u_n, t_n) + f(u_{n-1}, t_{n-1})] = S \\ &= u_{n-1} + \frac{\Delta t}{2} [-u_n^3 - u_{n-1}^3 + e(t_n) + e(t_{n-1})]. \end{aligned} \quad (6)$$

Нелинейное уравнение (6) удобно решать, используя итерационную схему. На первом шаге в правой части (6) значения u_n^3 вычисляются на основе (5). Полученное решение затем используется на втором шаге и т.д. Расчёты показали, что решение на первом шаге итераций отличается от решения на третьем шаге менее чем на 1%, т.е. для данного типа нелинейности уже в первом приближении достигается высокая степень точности.

В работе были выбраны параметры нелинейных фильтров, соответствующие эксперименту [11]. Приведём их значения: $A = -1.54 \cdot 10^{-2} \text{ Ом}^{-1}$, $B = 0.06 (B^2 \text{ Ом})^{-1}$ (для двух одинаковых туннельных диодов АИ301В), $R_n = |A|^{-1} = 65 \text{ Ом}$, $C = 15 \text{ нФ}$. Этим значениям соответствуют масштабы времени $t_M \cong 10^{-6} \text{ с}$ и напряжения $u_M = 0.5 \text{ В}$. На вход фильтров поступали гармонический сигнал и шум. Последний представлял собой стационарный случайный гауссовый процесс с нулевым средним значением, среднеквадратическим отклонением $\sigma = (\overline{\xi^2})^{1/2}$ от среднего значения и временем корреляции $\tau_\kappa = 10^{-7} \text{ с}$. Пара-

метр σ характеризует интенсивность шума.

Спектр анализируемых процессов находился с помощью быстрого преобразования Фурье выходного сигнала u в случае ФНЧ и $u_{R_n} = e(t) - u$ в случае ФВЧ. Шаг по времени выбирался равным τ_κ с числом точек отсчёта $N = 2^{14} = 16384$, что соответствует экспериментальной длительности анализируемых процессов $T \cong 1.6 \text{ мс}$ и разности $\Delta f = T^{-1}$ между дискретными частотами в спектре. В каждой точке отсчёта по времени включался датчик случайных чисел, распределённых по нормальному закону с параметром шума σ . Анализ амплитудного спектра входного процесса показал, что в нём содержится компонента сигнала $|a_s^{(ex)}|$ с номером S , определяемым частотой сигнала, а амплитуды шума практически равномерно распределены вплоть до граничной частоты τ_κ^{-1} .

Локальное в спектре отношение S/N на входе фильтра определялось, как и в работе [11]:

$$(S/N)_{ex} = |a_s^{(ex)}|^2 / |\xi_s^{(ex)}|^2,$$

$$|\xi_s^{(ex)}|^2 = \sum_{\Delta M=-20}^{20} |\xi_{s+\Delta M}^{(ex)}|^2 / 40,$$

где $\Delta n \neq 0$, а $|\xi_s^{(ex)}|^2$ – средняя мощность гармоник шума вблизи сигнальной частоты, которая с точностью до постоянного множителя $1/\Delta f$ совпадает со спектральной плотностью шума.

Аналогичным образом находилось отношение $(S/N)_{вых}$ на выходе фильтра. Как нетрудно показать, при данном определении S/N коэффициент передачи

$$q = (S/N)_{вых} / (S/N)_{ex}$$

в случае линейных фильтров всегда равен единице. Поэтому в работе основное внимание уделялось установлению областей параметров нелинейных фильтров, в которых $q > 1$.

Для разных реализаций входного шума результаты вычислений незначительно отличались друг от друга, поэтому найденные отношения S/N усреднялись по количеству K пусков шума. Для определения числа K проверялось свойство эргодичности шума или равенство его средних значений по времени и по ансамблю реализаций шума в отсутствии сигнала:

$$\overline{\xi^2} = \frac{1}{T} \int_0^T \xi^2(t) dt = \frac{1}{N} \sum_{n=0}^{N-1} (\xi_n^{(i)})^2 \cong \langle \xi^2 \rangle = \frac{1}{K} \sum_{i=1}^K (\xi_n^{(i)})^2,$$

где $\xi_n^{(i)}$ – значение шума в произвольный момент времени t_n на его i -ой реализации. Как показала проверка, это равенство при $K = 200$ выполняется с относительной погрешностью (1-3)%. Отсюда полученные отношения S/N усреднялись по 200 реализациям входного случайного процесса.

Представленные выше формулы легли в основу численного анализа, результаты которого представлены на рис. 2-4.

На рис. 2 изображены зависимости коэффициента передачи q от амплитуды h сигнала при фиксированном и достаточно большом параметре шума σ . Все кривые указывают, что коэффициент передачи принимает значения $q > 1$ при амплитудах сигнала, превышающих определённые пороговые значения. Как было показано в работе [10], данный эффект возникает в результате подавления шума сигналом. Обозначенная точками кривая на рис. 2 отражает результаты эксперимента [11] для ФНЧ. Видно, что результаты этого численного анализа ФНЧ и [11] незначительно отличаются друг от друга, что обусловлено отличием реальной вольт-амперной характеристики от теоретической при больших амплитудах сигнала. Падение q с ростом h связано с насыщением выходного отношения S/N в области больших значений h . Для выбранных частоты сигнала и

параметра шума σ эффект стохастической фильтрации в случае ФВЧ проявляется более ярко по сравнению с ФНЧ.

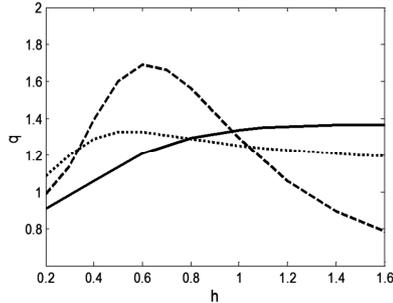


Рис. 2. Коэффициент q передачи по отношению сигнал/шум в зависимости от безразмерной амплитуды $\bar{h} = a_c / u_m$ сигнала при $f=100$ кГц и $\sigma / u_m = 0.4$. В случае ФНЧ данные численного расчёта изображены сплошной кривой, кривая с точками соответствует экспериментальным данным [11]. В случае ФВЧ численные данные для q представлены пунктирной линией

На рис. 3 приведены результаты численного анализа коэффициентов передачи q как функции от частоты сигнала для фиксированных \bar{h} и σ .

Кривые имеют локальные максимумы, соответствующие оптимальным для фильтрации частотам сигнала. В случае линейных фильтров (при отсутствии нелинейного элемента в схемах на рис. 1) граница между областями низких и высоких частот определяется характерной величиной $(R_n C)^{-1}$. Как следует из этих зависимостей, оптимальные частоты сигналов в нелинейных фильтрах существенно меньше $(R_n C)^{-1}$. В области высоких частот сигнала нелинейный член в уравнении (3) играет подчиненную роль и фильтр становится близким к линейному. В результате коэффициент передачи с ростом частоты сигнала асимптотически стремится к единице.

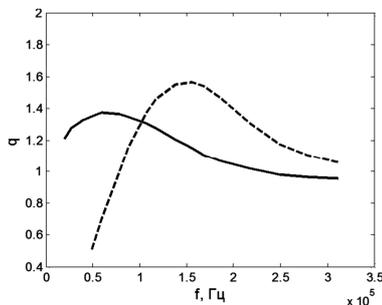


Рис. 3. Результаты численного анализа зависимости q от частоты f сигнала при $\bar{h}=1$ и $\sigma / u_m = 0.4$. Сплошная кривая – результаты для ФНЧ, а пунктирная – для ФВЧ

На рис. 4 представлена зависимость q от безразмерного параметра шума σ для фиксированных значений частоты и амплитуды сигнала. Эти кривые также имеют локальные максимумы и указывают на то, что наилучшие условия фильтрации гармонических сигналов в исследованных фильтрах возникают при амплитудах сигнала \bar{h} , превышающих параметр шума σ в (2-3) раза.

В результате проведенных исследований можно сделать следующие выводы. В работе развит эффективный метод численного анализа взаимодействия сигнала и шума в нелинейных фильтрах первого порядка. Результаты численного анализа по ФНЧ удовлетворительно согласуются с экспериментальными данными. Они также показывают, что эффект стохастической фильтрации реализуется не только в ФНЧ, но и в ФВЧ. При определенных параметрах ФВЧ, сигнала и шума коэффициент передачи по отношению S/N выше, чем коэффициент передачи у ФНЧ. Численный анализ занимает небольшое время счёта на обычных персональ-

ных компьютерах и при дальнейшей разработке его можно использовать в исследованиях фильтров высоких порядков.

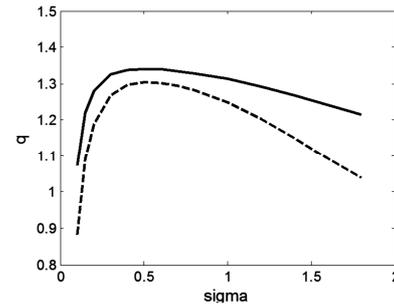


Рис. 4. Результаты численного анализа зависимости q от безразмерной интенсивности шума σ / u_m (σ) для частоты сигнала 100 кГц с амплитудой $\bar{h} = 1$. Сплошная и пунктирная кривые – соответственно результаты для ФНЧ и ФВЧ.

Литература

1. Gammaitoni L., Hanggi P., Jung P., Marchesoni F. Stochastic resonance. Rev. of Mod. Phys., 1998, v.70, №1, p. 223-287.
2. Анищенко В.С., Нейман А.Б., Мосс Ф., Шиманский-Гайер Л. Стохастический резонанс как индуцированный шумом эффект увеличения степени порядка. УФН, 1999, т.169, №1, с. 7-38.
3. Климонтович Ю.Л. Что такое стохастическая фильтрация и стохастический резонанс? УФН, 1999, т.169, №1, с. 39-47.
4. Hanggi P., Inchiosa M.E., Fogliatti D., Bulsara A.R. Nonlinear stochastic resonance: The saga of anomalous output-input gain. Phys. Rev. E., 2000. v.62, №5, p. 6155-6163.
5. Решетняк С.А., Третьяков Г.Н., Щеглов В.А. Аномальный коэффициент передачи отношения сигнал- шум при стохастическом резонансе. Кр. сообщ. по физике. ФИАН, 2000, №5, с.12-17.
6. Домбровский А.Н., Решетняк С.А. Стохастический резонанс и фильтрация сигналов в нелинейной электрической системе второго порядка. Радиотехника, 2007, №9, с.19-25.
7. Домбровский А.Н., Решетняк С.А. О стохастической фильтрации сигналов в нелинейных электрических системах. Радиотехника и электроника, 2009, т.54, №11, с.1369-1371.
8. Hu G., Haken H., Ning C.Z. A study of stochastic resonance without adiabatic approximation. Phys. Lett. A, 1992, v.172, №12, p.21-28.
9. Решетняк С.А., Щеглов В.А. О стохастическом резонансе с точки зрения фильтрующих свойств бистабильной системы. Квантовая электроника, 2003, т.33, №2, с.142-148.
10. Решетняк С.А., Третьяков Г.Н. Теоретическое исследование эффекта стохастической фильтрации. Радиотехника и электроника, 2013, т. 58, №4, с. 360-366.
11. Домбровский А.Н., Решетняк С.А. Экспериментальная проверка теорий стохастического резонанса и фильтрации сигналов в электрической цепи с двумя туннельными диодами. Научный вестник МИРЭА, 2007, №2(3), с.21-27.

NUMERICAL MODELING OF INTERACTION OF A SIGNAL AND NOISE IN THE NONLINEAR FILTER OF THE FIRST ORDER

Aboelazm M.A., Melchakov V.N., Reshetnyak S.A., Tretyakov G.N.

The method of numerical research of the signal-to-noise ratio in the nonlinear active filter of the first order is developed. It is shown that under the effect of a harmonic signal and noise at the entrance of filter the output signal-to-noise ratio exceeds the input ratio with the defined parameters both in the filter low and in the filter of high frequency. The results of analysis will satisfactory agree with experimental data.