

# ЦИФРОВАЯ БРАБОТКА СИГНАЛОВ

2/2005

НАУЧНО - ТЕХНИЧЕСКИЙ ЖУРНАЛ

**ПРОСТРАНСТВЕННО-ВРЕМЕННОЕ  
КОДИРОВАНИЕ**

**ШИРОКОПОЛОСНЫЙ CELP-КОДЕР**

**СЖАТИЕ ИЗОБРАЖЕНИЙ**

**ВИДЕОСИСТЕМЫ НА КРИСТАЛЛЕ**



# ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ

Научно-технический журнал  
№ 2 (16)/2005

Издается с 1999 года  
Выходит четыре раза в год

**ГЛАВНЫЙ РЕДАКТОР**  
Ю.Б. ЗУБАРЕВ

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

**РЕДАКЦИОННАЯ КОЛЛЕГИЯ:**

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

**Адрес редакции:**  
107241 Москва, Щелковское ш., 23А  
Тел.: (095) 290-9088  
Факс: (095) 290-9085  
E-mail: editor@dspa.ru  
http://www.dspa.ru

**Для писем:**  
129090 Москва, а/я 48.

**Издатель:**  
КБ волоконно-оптических приборов  
Ответственный редактор: Д.А. Задворнов  
Дизайн и верстка: Е.Н. Юрлова

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

Подписано в печать 7. 07.2005  
Формат 60×90/8.

Гарнитура «Pragmatica C». Печать офсетная.  
Бумага мелованная. Печ.л. 7,0. Тираж 1000 экз.

Заказ № 61. Отпечатано в  
ООО «Типография Принтхаус»  
Москва, ул. Вавилова, вл. 9а, стр. 7,  
тел.: 132-7218, 105-0798.

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

© КБ ВП, 2005

## В НОМЕРЕ:

А.П. Шумов, М.Г. Бакулин, В.Б. Крейнделин.

**Методы пространственно-временного  
блочного кодирования .....2**

Н. В. Гудкова

**Адаптивные влгоритмы регулирования уровня  
дискретных сигналов .....11**

В.И. Кошелев, В.Н. Горкин

**Структура доплеровского процессора  
радиолокционного сигнал  
при переменном периоде излучения ..... 15**

М.З. Лившиц, М. Парфенюк, А.А. Петровский

**Широкополосный CELP-кодер  
с мультиполосным возбуждением  
и многоуровневым векторным квантованием  
по кодовой книге с реконфигурируемой структурой ..20**

О.Ю. Аксенов

**Задержка обработки информации  
в многопроцессорных вычислителях  
различной конфигурации .....36**

А.В. Черников, М.К. Чобану

**Сжатие изображений нв  
базе вейвлет-преобрвзования  
и иерархического алгоритма кодирования .....40**

В.В. Березин, В.М. Гатаулин

**Новый этап развития твердотельного телевидения –  
«Видеосистемы нв кристалле» .....50**

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

УДК 621.396

## Методы пространственно-временного блочного кодирования

**А.П. Шумов, М.Г. Бакулин, В.Б. Крейнделин**

### Введение

**П**ространственно-временное кодирование было впервые определено как самостоятельный вид кодирования и опубликовано в 1998 году в работе [1] применительно к способу кодирования, использующему решетчатые коды. Введенное пространственно-временное кодирование (Space-Time Coding, далее STC) представляет собой новое направление кодирования для беспроводных систем связи со многими передающими и многими приемными антеннами. Использование многоантенных систем является одним из основных направлений развития систем подвижной связи, и совершенствование методов передачи информации в многоантенных системах является важным с точки зрения расширения возможностей таких систем.

Многоантенные системы связи и пространственно-временная обработка сигналов и до того, как было предложено пространственно-временное кодирование, широко использовались в системах с разнесенной передачей и разнесенным приемом для повышения надежности связи в условиях замираний и для резкого повышения пропускной способности и скорости передачи информации. Для них и ранее существовали эффективные алгоритмы передачи и приема сигналов. Среди таких алгоритмов можно назвать передачу сигналов с разнесением по задержке, прием на несколько разнесенных антенн с автовыбором канала, прием с линейным комбинированием сигналов с нескольких приемных антенн, систему BLAST. Тем не менее, введение пространственно-временного кодирования позволило не просто объединить существовавшие алгоритмы разнесенной передачи и разнесенного приема в новый класс, оно позволило выработать критерии синтеза и синтезировать новые эффективные алгоритмы пространственно-временной обработки сигналов.

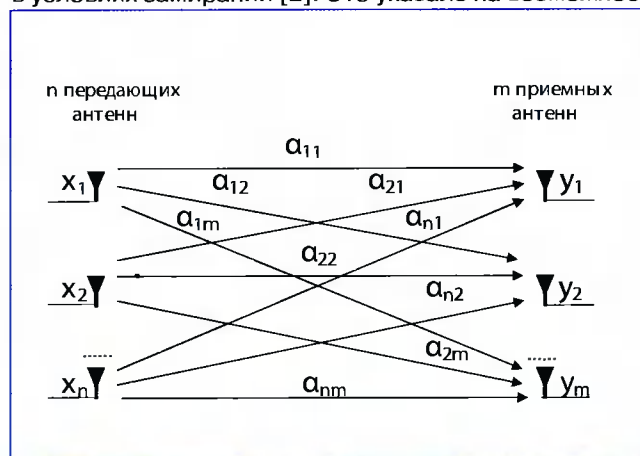
*Дан обзор построения и анализ качественных характеристик пространственно-временных кодов, используемых в многоантенных системах связи. Обсуждены вещественные и комплексные ортогональные формы для пространственно-временных блочных кодов. STBC на основе ортогональных форм могут достичь полного разнесения и имеют экстремально простой максимально правдоподобный алгоритм декодирования, использующий только линейное комбинирование в приемнике; эффективны в условиях замираний.*

Модель канала со многими передающими и многими приемными антеннами приведена на *рис. 1*. Такой канал называется каналом со многими входами и многими выходами – Multiple Input Multiple Output (MIMO) channel. Канал образован  $n$ -передающими антеннами и  $m$ -приемными антеннами. Канал описывается  $n \times m$  случайной матрицей  $H$  комплексных коэффициентов передачи

$$H = \begin{pmatrix} \alpha_{11} & \dots & \alpha_{1m} \\ \vdots & \ddots & \vdots \\ \alpha_{n1} & \dots & \alpha_{nm} \end{pmatrix}.$$

Статистические характеристики матрицы  $H$  определяются свойствами моделируемого канала.

Исследования последних лет позволили найти предельные характеристики систем с MIMO каналами в условиях замираний [2]. Это указало на возможнос-



**Рис. 1. Модель канала со многими передающими и многими приемными антеннами.**



ти резкого повышения пропускной способности и спектральной эффективности до уровней, которые не могут быть достигнуты при использовании других технологий.

Спектральная эффективность показывает, насколько эффективно система использует выделенный частотный диапазон и определяется как

$$\eta = \frac{\text{Скорость передачи}}{\text{Полоса частот}}, [\text{бит / с / Гц}] .$$

При этом предел Шеннона для спектральной эффективности равен

$$\eta_{\max} = \log\left(1 + \frac{P_c}{P_w}\right), [\text{бит / с / Гц}] ,$$

где  $P_c$  – мощность сигнала;

$P_w$  – мощность шума в канале.

### Кодирование для многоантенных систем

Рассмотрим общий случай пространственно-временного кодирования для многоантенной системы связи. Будем считать, что модель системы описывается следующими параметрами:

- $n$  – число передающих антенн;
- $m$  – число приемных антенн;
- $\alpha_{i,j}$  – коэффициент передачи пути от передатчика  $i$  к приемнику  $j, 1 \leq i \leq n, 1 \leq j \leq m$ . Полагаем, что  $\alpha_{i,j}$  являются независимыми и имеют гауссовское распределение с нулевым средним и дисперсией  $1/2$  для вещественной и мнимой частей.
- $l$  – длина блока кода по времени;
- $c_t^i$  – сигнал, переданный в дискретный момент времени  $t$  передающей антенной  $i, 1 \leq t \leq l, 1 \leq i \leq n$ ;
- $r_t^j$  – сигнал, принятый в дискретный момент времени  $t$  приемной антенной  $j, 1 \leq t \leq l, 1 \leq j \leq m$ .
- $\eta_t^j$  – аддитивный белый гауссовский шум с нулевым средним значением и дисперсией  $1/SRN$  на измерение.

Кодовое слово  $\mathbf{c}$  в системах STC описывается вектором  $\mathbf{c} = c_1^1 c_1^2 \dots c_1^n c_2^1 c_2^2 \dots c_2^n \dots c_l^1 c_l^2 \dots c_l^n$

Сигнал  $r_t^j$ , принятый антенной  $j$  в момент времени  $t$ , с учетом введенных обозначений определяется соотношением

$$r_t^j = \sum_{i=1}^n \alpha_{i,j} c_t^i + \eta_t^j .$$

Демодулятор максимального правдоподобия, дающий наилучшую оценку, для всех возможных кодовых слов должен вычислить решающую метрику

$$\sum_{t=1}^l \sum_{j=1}^m |r_t^j - \sum_{i=1}^n \alpha_{i,j} c_t^i|^2$$

и в качестве оценки переданного кодового слова выбрать то, которое данную решающую метрику минимизирует. Однако, с одной стороны, в общем случае при увеличении количества передающих и приемных антенн или скорости передачи информации алгоритм максимального правдоподобия из-за вычислительной сложности становится нереализуемым, и возможности использования MIMO-каналов становятся зависящими от наличия приемлемых по сложности алгоритмов декодирования пространственно-временных кодов. С другой стороны, необходимы способы построения пространственно-временных кодов, которые позволили бы реализовать тот или иной критерий качества системы.

В соответствии с работой [1] рассмотрим зависимость вероятности ошибки при приеме в системе с MIMO-каналом от свойств пространственно-временного кода и параметров многоантенной системы.

При неверном декодировании переданным считается кодовое слово  $\mathbf{e} = e_1^1 e_1^2 \dots e_1^n e_2^1 e_2^2 \dots e_2^n \dots e_l^1 e_l^2 \dots e_l^n$ . Вероятность такой ошибки при заданных  $\alpha_{i,j}$  хорошо аппроксимируется выражением  $P(\mathbf{c} \rightarrow \mathbf{e} | \alpha_{i,j}) \leq \exp[-d^2(\mathbf{c}, \mathbf{e})SRN/2]$ , где

$$d^2(\mathbf{c}, \mathbf{e}) = \sum_{j=1}^m \sum_{t=1}^l \left| \sum_{i=1}^n \alpha_{i,j} c_t^i - \sum_{i=1}^n \alpha_{i,j} e_t^i \right|^2$$

Эта функция является стандартной аппроксимацией для гауссовских хвостов. Обозначим  $\Omega_j = (\alpha_{1,j}, \dots, \alpha_{n,j})$  и  $A_{i,j} = (c_1^i - e_1^i, \dots, c_l^i - e_l^i) \cdot (c_1^i - e_1^i, \dots, c_l^i - e_l^i)$ .

Тогда после преобразований можно записать, что

$$d^2(\mathbf{c}, \mathbf{e}) = \sum_{j=1}^m \Omega_j A \Omega_j^*$$

Обозначим как  $\lambda_1, \dots, \lambda_m$  собственные числа матрицы  $A$  (возможно нулевые). Матрица  $A$  является эрмитовой, поэтому  $\lambda_i$  будут вещественными и неотрицательными. Обозначим через  $r$  ранг матрицы  $A$ . Усреднением верхней границы для  $P(\mathbf{c} \rightarrow \mathbf{e} | \alpha_{i,j})$  после некоторых преобразований получим

$$P(\mathbf{c} \rightarrow \mathbf{e}) \leq \left( \prod_{i=1}^r \lambda_i \right)^{-m} \left( \frac{SNR}{2} \right)^{-rm}$$

*Замечание.* Степень SNR в знаменателе последнего выражения для вероятности ошибки определяет преимущество, достигнутое за счет разнесения. Кодовое преимущество есть приближительная мера выигрыша над системой без кодирования, которая действует с тем же самым преимуществом разнесения.

С учетом этого замечания показатель  $r \cdot m$  называют выигрышем разнесения или порядком разнесения, тогда как произведение  $\left(\prod_{i=1}^r \lambda_i\right)^{1/r}$  называется кодовым выигрышем.

Определим следующую матрицу  $\mathbf{B}(\mathbf{c}, \mathbf{e})$

$$\mathbf{B}(\mathbf{c}, \mathbf{e}) = \begin{bmatrix} e_1^1 - c_1^1 & e_2^1 - c_2^1 & \dots & e_r^1 - c_r^1 \\ e_1^2 - c_1^2 & e_2^2 - c_2^2 & \dots & e_r^2 - c_r^2 \\ \vdots & \vdots & \ddots & \vdots \\ e_1^n - c_1^n & e_2^n - c_2^n & \dots & e_r^n - c_r^n \end{bmatrix}$$

Эта матрица является корнем квадратным из матрицы  $\mathbf{A}(\mathbf{c}, \mathbf{e})$ , то есть  $\mathbf{A} = \mathbf{B}\mathbf{B}^*$ , и ранги матриц  $\mathbf{A}$  и  $\mathbf{B}$  равны.

Вышеприведенное выражение верхней границы для  $P(\mathbf{c} \rightarrow \mathbf{e})$  дает два критерия для пространственно-временного кодирования [1]:

- Ранговый критерий: для достижения максимального порядка разнесения, равного  $n \cdot m$ , матрица  $\mathbf{B}(\mathbf{c}, \mathbf{e})$  должна иметь полный ранг для всех пар кодовых слов  $\mathbf{c}$  и  $\mathbf{e}$ .
- Критерий кодового выигрыша: для заданного порядка разнесения  $r \cdot m$  следует максимизировать минимум произведения  $\left(\prod_{i=1}^r \lambda_i\right)$  по всем парам кодовых слов.

В работе [1] с использованием этих критериев впервые были разработаны пространственно-временные коды, которые были названы пространственно-временные решетчатые коды (Space-Time Trellis Codes, далее STTC). Эти коды имеют высокие качественные показатели в каналах с замираниями при сравнительно небольшом количестве передающих и приемных антенн. Однако при увеличении числа антенн, а также при увеличении скорости передачи декодирование этих кодов становится чрезмерно сложным.

### Пространственно-временной код Аламаути

Рассматривая вопросы сложности декодирования STTC Аламаути [3] предложил замечательную схему пространственно-временного блочного кодирования для передачи с двумя антеннами. Эта схема поддерживает максимально правдоподобное детектирование, основанное только на линейной обработке в приемнике. Именно эта схема была позже обобщена в работе [4] на произвольное число антенн и позволила достичь полного разнесения для используемого количества передающих и приемных антенн.

Кратко об этом коде.

Матрица передачи, соответствующая коду Аламаути, имеет вид

$$\begin{pmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{pmatrix}$$

Предполагается, что сигнальное созвездие содержит  $2^b$  сигналов. В первом временном слоте  $2b$  бит поступают на кодер и выбирают два комплексных символа  $s_1$  и  $s_2$ . Эти символы передаются одновременно с первой и второй антенн соответственно. В следующем временном слоте сигналы  $-s_2^*$  и  $s_1^*$  также передаются одновременно с антенн один и два.

Максимально правдоподобное детектирование равнозначно минимизации решающей статистики

$$\sum_{j=1}^m \left( |r_1^j - \alpha_{1,j} s_1 - \alpha_{2,j} s_2|^2 + |r_2^j + \alpha_{1,j} s_2^* - \alpha_{2,j} s_1^*|^2 \right)$$

по всем возможным величинам  $s_1$  и  $s_2$ . Минимизирующие величины являются оценками  $s_1$  и  $s_2$  соответственно. Это эквивалентно минимизации решающей статистики

$$\left| \sum_{j=1}^m (r_1^j \alpha_{1,j}^* + (r_2^j)^* \alpha_{2,j}) - s_1 \right|^2 + \left( -1 + \sum_{j=1}^m \sum_{i=1}^2 |\alpha_{i,j}|^2 \right) |s_1|^2$$

для определения  $s_1$  и решающей статистики

$$\left| \sum_{j=1}^m (r_1^j \alpha_{2,j}^* + (r_2^j)^* \alpha_{1,j}) - s_2 \right|^2 + \left( -1 + \sum_{j=1}^m \sum_{i=1}^2 |\alpha_{i,j}|^2 \right) |s_2|^2$$

для декодирования  $s_2$ . Рассматриваемый алгоритм декодирования является очень простым и обеспечивает полное разнесение  $2m$  при использовании  $m$  приемных антенн. И хотя код Аламаути уступает STTC по уровню вероятности ошибки, отмеченные свойства схемы Аламаути являются очень привлекательными, и хотелось бы иметь пространственно-временные коды с такими же свойствами для числа передающих антенн больше, чем две. Анализ кода Аламаути показал, что матрица передачи кода по своим свойствам соответствует так называемым "orthogonal designs" – одному из направлений теории матриц, которое изучалось в частности в [5].

Авторам неизвестен математический термин из отечественной математической литературы, который соответствовал бы термину orthogonal design, поэтому в дальнейшем в качестве перевода orthogonal design будет введен термин «ортогональная форма».

К моменту появления кода Аламаути существовали только вещественные прямоугольные ортогональные формы – для  $n=2, 4$  и  $8$ . На основе этих ортогональных форм могли быть построены пространственно-вре-

менные коды, использующие вещественные сигнальные созвездия (например, PAM) и 2, 4 или 8 передающих антенн. Но сразу стало ясно, что необходимы коды на основе комплексных сигнальных созвездий и для произвольного числа передающих антенн. Соответствующая теория была разработана в [4].

В последующих разделах кратко рассмотрены вещественные и вновь разработанные в [4] обобщенные вещественные, комплексные и обобщенные комплексные ортогональные формы, а также пространственно-временные коды на их основе.

### Вещественные ортогональные формы и пространственно-временные коды

Вещественные ортогональные формы послужили основой для дальнейших исследований, целью которых стала разработка пространственно-временных блочных кодов (Space-Time Block Codes, далее STBC) на основе комплексных сигнальных созвездий для произвольного числа передающих антенн.

Вещественная ортогональная форма размерности  $n$  есть  $n \times n$  ортогональная матрица, чьи строки являются перестановками действительных чисел  $\pm x_1, \pm x_2, \dots, \pm x_n$ . Проблема существования для ортогональных форм известна в математической литературе как проблема Гурвица-Радоны [5]. Эта проблема была полностью определена Радоном в другом контексте в начале прошлого века. Было показано, что фактически ортогональные формы существуют, если и только если  $n=2, 4$  или 8.

Без потери общности первая строка ортогональной формы может быть определена как  $(x_1, x_2, \dots, x_n)$ . Например, вещественные ортогональные формы для  $n=2$  и  $n=4$  имеют вид

$$\begin{bmatrix} x_1 & x_2 \\ -x_2 & x_1 \end{bmatrix}, \quad \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ -x_2 & x_1 & -x_4 & x_3 \\ -x_3 & x_4 & x_1 & -x_2 \\ -x_4 & -x_3 & x_2 & x_1 \end{bmatrix}.$$

Пространственно-временные блочные коды, основанные на вещественных ортогональных формах, размерности  $n$  ( $n=2, 4$  или 8) могут быть сконструированы следующим образом. Кодер берет блок из  $nb$  бит. Для каждого  $i$   $1 \leq i \leq n$  кодер выбирает символ  $s_i$  из вещественного созвездия размерности  $2^b$ . Затем кодер использует  $s_1, \dots, s_n$  для построения ортогональной матрицы  $O(s_1, \dots, s_n)$ , основанной на вещественных ортогональных формах размерности  $n$ . В момент времени  $t$   $n$  антенн передают  $t$ -ю строку матрицы  $O(s_1, \dots, s_n)$ . Из-за ортогональности порядок разнесения этого кода равен  $n \cdot m$ . Скорость кода равна  $b$

бит/с/Гц, что является оптимальным при порядке разнесения  $n \cdot m$  для созвездия размерности  $2^b$ . Ортогональность также упрощает декодирование по правилу максимального правдоподобия. Обозначим через  $\delta_k(i)$  знак  $x_i$  в  $k$ -й строке матрицы  $O(s_1, \dots, s_n)$  и через  $\in_k(p)$  номер столбца в строке  $p$ , в котором находится  $\pm x_k$ . Следовательно, приемнику только необходимо линейно скомбинировать принятые сигналы

$$R_i = \sum_{t=1}^n \sum_{j=1}^m r_t' \alpha_{\in_k(i), j} \delta_t(i), \quad 1 \leq i \leq n,$$

и решить по  $S_i$ , так что

$$s_i = \arg \min_{s \in A} |R_i - s|^2 + \left( -1 + \sum_{k,l} |\alpha_{k,l}|^2 \right) |s|^2.$$

Вещественные ортогональные формы существуют только для  $n=2, 4$ , или 8. Путем развития обобщенных вещественных ортогональных форм могут быть получены пространственно-временные блочные коды для разного числа антенн.

До обсуждения обобщенных вещественных ортогональных форм отметим, что более доступным обобщением являются ортогональные формы с линейной обработкой, в которых элементы в ортогональных матрицах  $O(x_1, \dots, x_n)$  есть линейные комбинации из  $x_1, \dots, x_n$  (вместо перестановок,  $\pm x_1, \dots, \pm x_n$ , как перед этим). Однако применение линейной обработки не увеличивает числа возможных значений  $n$ , так чтобы ортогональные формы существовали.

### Обобщенные вещественные ортогональные формы

Обобщенная вещественная ортогональная форма размерности  $n$  есть  $n \times n$  матрица  $G$  с элементами 0,  $\pm x_1, \dots, \pm x_k$  такая, что  $G^T G = D$ , где  $D$  –  $n \times n$  диагональная матрица, в которой  $i$ -й диагональный элемент имеет форму  $l_1^2 x_1^2 + \dots + l_k^2 x_k^2$  с  $l_1, \dots, l_k$ , являющимися строго положительными целыми. Однако без потери общности можно допустить, что  $G^T G = I(x_1^2 + \dots + x_k^2)$ .

Пространственно-временные блочные коды, основанные на обобщенных вещественных ортогональных формах размерности  $n$ , могут быть построены так. Кодер берет блок из  $k \cdot b$  бит. Для каждого  $i$ ,  $1 \leq i \leq k$  кодер выбирает символ  $s_i$  из вещественного созвездия  $A$  размерности  $2^b$ . Затем декодер использует  $s_1, \dots, s_k$  для построения матрицы  $G(s_1, \dots, s_k)$ , основанной на обобщенных вещественных ортогональных формах размерности  $n$ .

Результирующий порядок разнесения равен  $n \cdot m$ . Поскольку максимальная скорость есть  $b$  бит/с/Гц, а длина блока есть  $p$ , то скорость этого кода определена как отношение между действительно переданными битами и максимальным возможным числом переданных бит  $(kb)/(pb)=k/p$ .

Определим  $A(R, n)$  как минимальное  $p$ , такое, что существует  $p \times n$  ортогональная форма со скоростью по меньшей мере  $R$ . (Если такого  $p$  не существует, полагаем  $A(R, n) = \infty$ .) Нахождение величин  $A(R, n)$  рассматривается как фундаментальный вопрос теории обобщенных ортогональных форм. Наиболее интересная часть этого вопроса есть нахождение  $A(1, n)$ , так как этот случай включает коды с оптимальными скоростями.

Было показано, что для любых  $R$ ,  $A(R, n) < \infty$ . В частности, для  $R=1$

$$A(1, n) = \min_{(c, d \in U)} 2^{4c+d}, \text{ где } U = \{(c, d): 0 \leq c, 0 \leq d \leq 4, 8c + 2^d \geq n\}.$$

Поэтому можно строить пространственно-временные коды для любого числа передающих антенн. Однако для величин  $n$ , других чем 2, 4 или 8 мы не имеем длины блока  $p$  равной числу антенн  $n$ . Опишем построение кода подробно.

Построение кода основано на семействе матриц Гурвица-Радона. Совокупность  $p \times n$  вещественных матриц  $\{B_1, \dots, B_k\}$  называется семейством матриц Гурвица-Радона размерности  $k$ , если

$$B_i^* B_j = \delta_{ij}, \quad B_i^* = -B_i, \quad 1 \leq i \leq k,$$

$$B_i B_j = -B_j B_i, \quad 1 \leq i, j \leq k.$$

Для любых  $n=2^a b$ , где  $b$  нечетно и  $a=4c+d$  при  $0 \leq d < 4$  и  $0 \leq c$ , Радон показал, что любое семейство матриц Гурвица-Радона содержит меньше, чем  $\rho(n) = 8c + 2^d \leq n$  матриц. Более того, для любого  $n$  существует семейство матриц с  $\rho(n)-1$  членами, которые являются целыми (все элементы есть 0 или  $\pm 1$ ).

### Построение пространственно-временного блочного кода длины $p$ .

Выбираем семейство целых матриц Гурвица-Радона с  $\rho(n)-1$  членами  $\{A_1, A_2, \dots, A_{\rho(n)-1}\}$ . Полагаем  $A_0 = I$  и обозначаем  $X = (x_1, \dots, x_p)$ . Строим  $p \times n$  обобщенную вещественную ортогональную форму  $G$ , определяя  $j$ -й столбец матрицы  $G$  как  $A_j X^T$ . Отсюда следует, что  $G$  имеет полный ранг и таким образом дает порядок разнесения  $n \cdot m$ , как и требовалось.

Используя этот метод, сконструированы вещественные ортогональные формы размерностей 3, 5, 6 и

7. Поэтому доступными являются блочные коды для любого числа антенн между 2 и 8.

Примеры ортогональных форм размерностей 3 и 5 приведены ниже

$$G_3 = \begin{bmatrix} x_1 & x_2 & x_3 \\ -x_2 & x_1 & -x_4 \\ -x_3 & x_4 & x_1 \\ -x_4 & x_3 & x_2 \end{bmatrix}, \quad G_5 = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ -x_2 & x_1 & x_4 & -x_3 & x_6 \\ -x_3 & -x_4 & x_1 & x_2 & x_7 \\ -x_4 & x_3 & -x_2 & x_1 & x_8 \\ -x_5 & -x_6 & -x_7 & -x_8 & x_1 \\ -x_6 & x_5 & -x_8 & x_7 & -x_2 \\ -x_7 & x_8 & x_5 & -x_6 & -x_3 \\ -x_8 & -x_7 & x_6 & x_5 & -x_4 \end{bmatrix}$$

### Комплексные ортогональные формы и STBC

Комплексная ортогональная форма  $O_c$  размерности  $n$  есть ортогональная матрица, строки которой являются перестановками комплексных чисел  $\pm x_1, \dots, \pm x_n$ , комплексно-сопряженных с ними чисел  $\pm x_1^*, \dots, \pm x_n^*$  или произведений этих чисел на  $\pm \sqrt{-1}$ .

Показано, что комплексные ортогональные формы существуют, если и только если  $n=2$ . Как и ранее, линейная обработка переданных сигналов не увеличивает набор  $n$ , для которых ортогональные формы существуют.

С учетом изложенного, код Аламаути является уникальным и не допускает развития в рамках комплексных ортогональных форм для использования при количестве передающих антенн больше двух.

Чтобы расширить набор  $n$  для комплексных ортогональных форм, были введены обобщенные комплексные ортогональные формы, определенные аналогично обобщенным вещественным ортогональным формам. Однако для комплексных сигналов блочные коды гарантированно существуют для произвольной величины  $n$  только для скорости  $R \leq 1/2$ . Для скорости  $R > 1/2$  при линейной обработке передаваемых сигналов существуют блочные коды скорости  $3/4$  для  $n=3$  и 4. Проблема комплексных ортогональных форм скорости  $R > 1/2$  остается все еще изученной недостаточно.

### Обобщенные комплексные ортогональные формы

По результатам, полученным для обобщенных вещественных ортогональных форм, в работе [4] были предложены обобщенные комплексные ортогональные формы.

Обобщенная комплексная ортогональная форма  $G_c$  размерности  $n$  есть  $p \times n$  ортогональная матрица с элементами  $(l'_1 |x_1|^2 + l'_2 |x_2|^2 + \dots + l'_k |x_k|^2)$  или их произведениями с мнимой единицей  $i$ , так что  $G_c^* G_c = D$ , где  $D_c$  есть диагональная матрица с  $(i, i)$ -ми диагональными элементами формы  $(l'_1 |x_1|^2 + l'_2 |x_2|^2 + \dots + l'_k |x_k|^2)$  и



коэффициентами  $l'_1, l'_2, \dots, l'_k$  всеми строго положительными целыми. Тогда  $G_c$  рассматривается как обобщенная ортогональная форма размерности  $n$  и скорости  $R=k/p$ .

В работе [4] показано, что без потери общности любая  $p \times n$  обобщенная комплексная ортогональная форма  $G_c$  с переменными  $\pm x_1, \dots, \pm x_m, \pm x_1^*, \dots, \pm x_n^*$  удовлетворяет условию  $G_c^* G_c = (l_1^2 + l_2^2 + \dots + l_k^2) \mathbf{I}$  после соответствующей нормировки.

Передача, использующая обобщенные комплексные ортогональные формы, подобна передаче с использованием обобщенной вещественной ортогональной формы. Максимально правдоподобное декодирование также аналогично декодированию по схеме Аламаути и может быть выполнено с использованием линейной обработки в приемнике.

Для комплексных сигнальных созвездий обобщенные комплексные ортогональные формы  $G_c$  могут быть сконструированы для любого числа передающих антенн и скорости  $R=1/2$  следующим образом:

$$G_c = \begin{bmatrix} L n(x_1, x_2, \dots, x_k) \\ L n(x_1^*, x_2^*, \dots, x_k^*) \end{bmatrix}$$

где  $L n(\dots)$  представляет собой соответствующую вещественную ортогональную форму. При этом полускоростные STBC, использующие три и четыре передающих антенны определяются следующим образом:

$$G_3^c = \begin{bmatrix} x_1 & x_2 & x_3 \\ -x_2 & x_1 & -x_4 \\ -x_3 & x_4 & x_1 \\ -x_4 & -x_3 & x_2 \\ x_1 & x_2 & x_3 \\ -x_2^* & x_1^* & -x_4^* \\ -x_3^* & x_4^* & x_1^* \\ -x_4^* & -x_3^* & x_2^* \end{bmatrix}, \quad G_4^c = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ -x_2 & x_1 & -x_4 & x_3 \\ -x_3 & x_4 & x_1 & -x_2 \\ -x_4 & -x_3 & x_2 & x_1 \\ x_1^* & x_2^* & x_3^* & x_4^* \\ -x_2^* & x_1^* & -x_4^* & x_3^* \\ -x_3^* & x_4^* & x_1^* & -x_2^* \\ -x_4^* & -x_3^* & x_2^* & x_1^* \end{bmatrix}$$

Для 3 и 4 передающих антенн в [4] также предложены следующие обобщенные комплексные ортогональные формы с линейной обработкой и скоростью  $R=3/4$ ,

$$H_3 = \begin{bmatrix} x_1 & x_2 & \frac{x_3}{\sqrt{2}} \\ -x_2^* & x_1^* & \frac{x_3}{\sqrt{2}} \\ \frac{x_3^*}{\sqrt{2}} & \frac{x_3^*}{\sqrt{2}} & \frac{x_1 - x_1^* + x_2 - x_2^*}{2} \\ \frac{x_3^*}{\sqrt{2}} & -\frac{x_3^*}{\sqrt{2}} & \frac{x_2 + x_2^* + x_1 - x_1^*}{2} \end{bmatrix}$$

$$H_4 = \begin{bmatrix} x_1 & x_2 & \frac{x_3}{\sqrt{2}} & \frac{x_3}{\sqrt{2}} \\ -x_2^* & x_1^* & \frac{x_3}{\sqrt{2}} & -\frac{x_3}{\sqrt{2}} \\ \frac{x_3^*}{\sqrt{2}} & \frac{x_3^*}{\sqrt{2}} & \frac{-x_1 - x_1^* + x_2 - x_2^*}{2} & \frac{-x_2 - x_2^* + x_1 - x_1^*}{2} \\ \frac{x_3^*}{\sqrt{2}} & -\frac{x_3^*}{\sqrt{2}} & \frac{x_2 + x_2^* + x_1 - x_1^*}{2} & \frac{-x_1 + x_1^* + x_2 - x_2^*}{2} \end{bmatrix}$$

В работах [6, 9] предложены два обобщенных комплексных ортогональных блочных кода скоростей 7/11 и 3/5 для 5 и 6 передающих антенн. Для  $n=2$  передающих антенн комплексная ортогональная форма размера  $2^l \times 2^l$  с переменными  $x_1, \dots, x_l$  может быть получена рекурсивным путем:

$$G_l(x_l) = x_l \mathbf{I}_l$$

$$G_{2^l} = \begin{bmatrix} G_{2^{l-1}}(x_1, \dots, x_l) & x_{l+1} \mathbf{I}_{2^{l-1}} \\ -x_{l+1}^* \mathbf{I}_{2^{l-1}} & G_{2^{l-1}}^H(x_1, \dots, x_l) \end{bmatrix}$$

Далее на основе полученного таким образом кода  $G_4(x_1, x_2, x_3)$  может быть построен код скорости 7/11 для 5 передающих антенн с размерностью матрицы передачи  $11 \times 5$

$$G_5^c = \begin{bmatrix} x_1 & x_2 & x_3 & 0 & x_4 \\ -x_2^* & x_1^* & 0 & x_3 & x_5 \\ x_3^* & 0 & -x_1^* & x_2 & x_6 \\ 0 & x_3^* & -x_2^* & -x_1 & x_7 \\ x_4^* & 0 & 0 & -x_7^* & -x_1^* \\ 0 & x_4^* & 0 & x_6^* & -x_2^* \\ 0 & 0 & x_4^* & x_5^* & -x_3^* \\ 0 & -x_5^* & x_6^* & 0 & x_1 \\ x_5^* & 0 & x_7^* & 0 & x_2 \\ -x_6^* & -x_7^* & 0 & 0 & x_3 \\ x_7 & -x_6 & -x_5 & x_4 & 0 \end{bmatrix}$$

и затем на основе этого кода  $G_5$  может быть построен код  $G_6$  скорости 3/5 для 6 передающих антенн с размерностью передающей матрицы  $30 \times 6$ .

### Основные результаты разработки STBC на основе комплексных ортогональных форм

На основе комплексных ортогональных форм разработаны следующие пространственно-временные коды:

- Для двух передающих антенн – пространственно-временной блочный код с полной скоростью 1.
- Для 3 и 4 передающих антенн – пространственно-временной блочный код с максимальной скоростью 3/4.



- Для 5 и 6 передающих антенн – пространственно-временной блочный код со скоростью 7/11 и 3/5.
- Для произвольного числа передающих антенн существуют пространственно-временные блочные коды, имеющие скорость 1/2.

В табл. 1 для сравнения приводятся все STBC на основе комплексных ортогональных форм, представленные как примеры в этой работе.

Формулы для декодирования пространственно-временных кодов можно найти в работе [7].

### Сравнительный анализ пространственно-временных кодов

В этом разделе приводятся и сравниваются характеристики пространственно-временных кодов,  $G_2^c$ ,  $G_3^c$ ,  $G_4^c$ ,  $H_3$  и  $H_4$  [7, 8]. Использование разных схем модуляции приводит к разным эффективным скоростям передачи, поэтому системы сравнивались на основе одинаковой эффективной скорости передачи. Кроме того, сделаны следующие предположения:

- 1) канал с общими замираниями;
- 2) приемнику точно известны коэффициенты передачи каналов.

### Характеристики схем при эффективной скорости 1 бит/с/Гц

На рис. 2 и рис.3 сравниваются характеристики пространственно-временных кодов  $G_2^c$ ,  $G_3^c$ , и  $G_4^c$  при 1 бит/с/Гц., использующие один и два приемника соответственно. BPSK модуляция была использована для пространственно-временного кода  $G_2^c$ . Поскольку пространственно-временные коды  $G_3^c$  и  $G_4^c$  и являются полускоростными кодами, то была использована QPSK модуляция. Из рис. 2 можно видеть, что при вероятности ошибки  $10^{-5}$  пространственно-временной коды  $G_3^c$  и  $G_4^c$  дают 5 dB и 7.5dB выигрыша над кодом  $G_2^c$  соответственно.

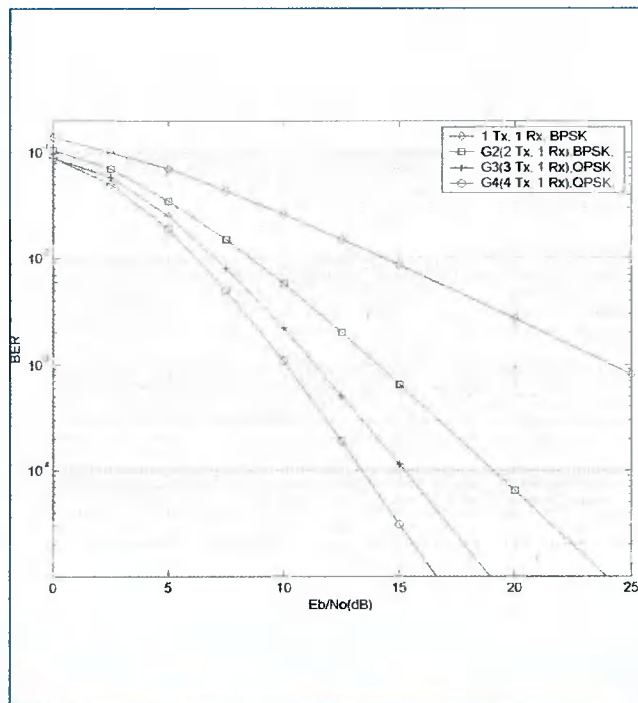


Рис. 2. Графики зависимости вероятности ошибки на бит (Bit Error Rate, далее BER) для кодов  $G_2^c$ ,  $G_3^c$ ,  $G_4^c$  при скорости 1 бит/с/Гц и одной приемной антенне.

Если число приемников возросло до 2, как показано на рис.3, то выигрыш в отношении сигнал/шум уменьшается до порядка 2 dB и 3.5dB соответственно. Причина в том, что в рассмотренных каналах с общими замираниями большая часть достижимого выигрыша разнесения уже получена за счет использования кода  $G_2^c$  и двух приемников.

### Характеристики схем при эффективной скорости 2 бит/с/Гц

На рис.4 сравниваются характеристики пространственно-временных кодов  $G_2^c$ ,  $G_3^c$ ,  $G_4^c$ ,  $H_3$  и  $H_4$  при спектральной эффективности 2 бит/с/Гц, использующих один приемник. Для кода  $G_2^c$  использована модуляция

Таблица 1.

Пространственно-временной код	Число передатчиков	Число входных символов	Кодовый интервал	Скорость кода
$G_2^c$	2	2	2	1
$G_3^c$	3	4	8	1/2
$G_4^c$	4	4	8	1/2
$H_3$	3	3	4	3/4
$H_4$	4	3	4	3/4
$G_5^c$	5	7	11	7/11

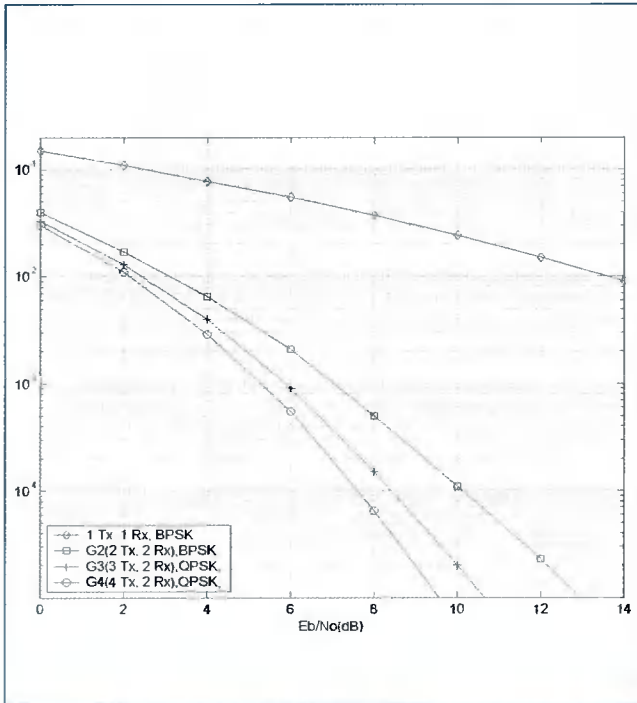


Рис. 3. Графики зависимости вероятности ошибки на бит для кодов  $G_2^c$ ,  $G_3^c$  и  $G_4^c$  при скорости 1 bits/s/Hz и двух приемных антеннах.

QPSK. Для кодов  $G_3^c$  и  $G_4^c$  применена модуляция 16QAM. Поскольку кодовая скорость  $H_3$  и  $H_4$  равна 3/4, то для достижения спектральной эффективности 2,25 бит/с/Гц была применена 8PSK модуляция. Из рис. 4 видно, что при больших вероятностях ошибок код  $G_2^c$  слегка превосходит другие. Однако, ситуация противоположна, когда система работает при малых вероятностях ошибки. При вероятности ошибки  $10^{-5}$  код  $G_4^c$  дает выигрыш разнесения над кодом  $G_2^c$  только 5 dB. 2.5 dB потерь по сравнению с 7.5 dB выигрыша, достигнутого на 1 бит/с/Гц, объясняются тем, что более уязвимая 16QAM модуляция была использована для пространственно-временного кода  $G_4^c$ . Поскольку 16QAM сигнальное созвездие более плотно упаковано по сравнению с QPSK, то оно более подвержено ошибкам. Так как пространственно-временной код  $G_3^c$  также является полускоростным, то он подвержен тем же недостаткам. По рис. 4 можно также оценить качество 3/4 скоростных пространственно-временных кодов  $H_3$  и  $H_4$ . При вероятности ошибки  $10^{-5}$  качество  $H_4$  на 0.5 dB лучше, чем кода  $G_4^c$ . Это снова объясняется тем, что более высокого порядка модуляция применена для полускоростного кода  $G_4^c$ . Отсюда качество кодов  $G_3^c$  и  $G_4^c$  хуже, чем качество кодов  $H_3$  и  $H_4$  соответственно.

### Заключение

В работе дан обзор построения и анализ качественных характеристик пространственно-временных кодов.

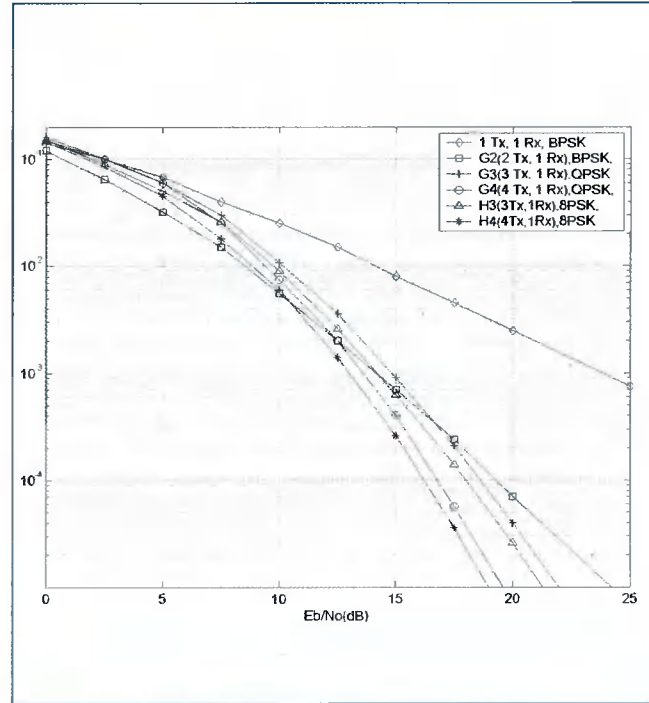


Рис. 4. Графики зависимости вероятности ошибки на бит для кодов  $G_2^c$ ,  $G_3^c$  и  $G_4^c$  при скорости 2 bits/s/Hz и одной приемной антенне.

Обсуждены вещественные и комплексные ортогональные формы для пространственно-временных блочных кодов. STBC на основе ортогональных форм могут не только достичь полного разнесения, но также имеют экстремально простой максимально правдоподобный алгоритм декодирования, использующий только линейное комбинирование в приемнике. Эти свойства все идут из ортогональности столбцов ортогональных форм.

Еще одной быстро развивающейся ветвью многоантенных систем являются системы BLAST. И хотя оказывается, что блочный код превосходит V-BLAST на большинстве отношений сигнал/шум, следует отметить, что скорости передачи, для которых тестировались системы BLAST, имеют порядок 20-40 бит/с/Гц [10-12]. Этот порядок скоростей может быть легко достигнут с использованием BLAST архитектуры, например, с 10 антеннами на приеме и со столькими же передающими антеннами, каждая из которых передает символы из 8QAM алфавита. При этом достигается скорость 3 бит/с/Гц на пространственный массив, или 30 бит/с/Гц на всю систему. Для достижения сопоставимых скоростей с использованием STBC будет требоваться нереализуемая схема модуляции, поддерживающая 30 бит на символ. Поэтому развитие реализуемых методов построения STBC с целью повышения, например, скорости передачи информации при сохранении существующей простоты декодирования является весьма важной и перспективной задачей.

## Литература

1. Tarokh, N. Seshadri, and A.R. Calderbank. Space-time codes for high data rates wireless communications: Performance criterion and code construction, *IEEE Trans. Inform. Theory*, vol. 44, pp. 744- 765, no. 2, March 1998.
2. E. Telatar. Capacity of multi-antenna Gaussian channels, AT&T-Bell Labs Internal Tech. Memo., June 1995.
3. S. M. Alamouti. A simple transmitter diversity scheme for wireless communications. *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.
4. V. Tarokh, H. Jafarkhani, and A. Calderbank. Space–time block codes from orthogonal designs. *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, July 1999.
5. A. V. Geramita and J. Seberry, *Orthogonal Designs, Quadratic Forms and Hadamard Matrices, Lecture Notes in Pure and Applied Mathematics*, vol. 43. New York and Basel: Marcel Dekker, 1979.
6. W. Su and X.G. Xia. Two generalized complex orthogonal space-time block codes of rates 7/11 and 3/5 for 5 and 6 transmit antennas. *IEEE Trans. on Information Theory*, Vol. 49, No. 1, Jan. 2003.
7. V. Tarokh, H. Jafarkhani, and A.R. Calderbank. Space-time block coding for wireless communications: Performance results. *IEEE Journal on Select Areas in Communications*, vol. 17, pp. 451-460, 1999.
8. T. H. Liew and Lajos Hanzo. Space–Time Codes and Concatenated Channel Codes for Wireless Communications. *Proceedings of the IEEE*, Vol. 90, No. 2, Feb. 2002.
9. W. Su and X.G. Xia. On Space-Time Block Codes from Complex Orthogonal Designs. *Wireless Personal Communications*, Vol. 25 No. 1, pp1-26, Apr. 2003.
10. Gerard J. Foschini, Glen D. Golden, Reinaldo A. Valenzuela, and Peter W. Wol-niansky. Simplified processing for high spectral efficiency wireless communication employing multi-element arrays. *IEEE Journal on Selected Areas in Communications*, 17(11):1841{1852, November 1999.
11. Shloma A., Bakulin M., Kreyndelin V., Tong W. Combined Space Time Coding and Spatial Multiplexing. // "International Conference on Circuits and Systems for Communications, Moscow, 2004".
12. Шлома А.М., Бакулин М.Г., Крейнделин В.Б., Панкратов Д.Ю. Квазиоптимальный алгоритм демодуляции в системе BLAST // *Научно-технические технологии*, Москва, № 11, 2004, т.5.

## Уважаемые авторы!

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

### 1) Требования к текстовым материалам и сопроводительным документам:

- Текст – текстовый редактор Microsoft Word на базе версии WINDOWS'95 или выше.
- Таблицы и рисунки должны быть пронумерованы. На все рисунки, таблицы и библиографические данные указываются ссылки в тексте статьи.
- Объем статьи до 12 стр. (шрифт 12). Для заказных обзорных работ объем может быть увеличен до 20 стр.
- Название статьи на русском и английском языках.
- Рукопись статьи сопровождается:
  - краткой аннотацией на русском и английском языках;
  - номером УДК;
  - сведениями об авторах (Ф.И.О, организация, телефоны, электронная почта).

### 2) Требования к иллюстрациям:

- Векторные (схемы, графики) – желательно использование графических редакторов Adobe Illustrator или Corel DRAW.
- Растровые (фотографии, рисунки) – М 1:1, разрешение не менее 300dpi, формат tiff.

Справки по телефонм: (095) 290-90-88, (0912)96-10-95 или по электронной почте [info@dspa.ru](mailto:info@dspa.ru), or [rgta.ryazan.ru](mailto:rgta.ryazan.ru)



УДК 681.513.6

# Адаптивные алгоритмы регулирования уровня дискретных сигналов

Н. В. Гудкова

**К**ак известно, для качественной работы цифровых систем обработки информации во многих случаях требуется нормализация уровня сигналов, поступающих в компьютер из каналов связи.

В качестве одного из вариантов решения этой задачи можно использовать алгоритм, в основу которого положен традиционный для радиотехники принцип построения аналоговых схем автоматической регулировки усиления сигналов (АРУ). Структурная схема АРУ показана на рис. 1.

Схема состоит из двух блоков перемножения БП<sub>1</sub> и БП<sub>2</sub>, выпрямителя В и фильтра низких частот ФНЧ.

Алгоритм работы устройства следующий. Выходным сигналом схемы является сигнал на выходе блока БП<sub>1</sub>, где временные отсчеты входного сигнала  $u_{1k}$  ( $k=0, 1, 2, \dots$ ) умножаются на коэффициент  $K_a$ , то есть

$$u_{2k} = u_{1k} K_a \quad (1)$$

Коэффициент  $K_a$  вычисляется в блоке БП<sub>2</sub> по формуле

$$K_a = \frac{u_{3m}}{u_{2k}^0} \quad (2)$$

где  $u_{3m}$  – эталонный постоянный сигнал, с помощью которого задается желаемый уровень выходного сигнала,

*Предлагается система регулирования уровня дискретных сигналов, построенная на принципах цифровой адаптивной фильтрации. Разработана методика настройки параметров сходимости адаптивных алгоритмов на основании требований к устойчивости, точности и быстродействию системы. Представлены результаты моделирования системы на ПК, которые демонстрируют высокое качество регулирования уровня стационарных сигналов (гармонических, импульсных т.п.). Показано, что использование канала автоматической настройки параметра сходимости регулятора обеспечивает нормализацию уровня выходного сигнала при изменении амплитуды входного сигнала в диапазоне до 60 дБ и более.*

а  $u_{2k}^0$  – среднее значение выпрямленного выходного сигнала  $u_{2k}^e = \text{abs}(u_{2k})$ , которое формируется фильтром ФНЧ. Из формулы (2) видно, что величина коэффициента  $K_a$  обратно пропорциональна сигналу  $u_{2k}^0$ , что обеспечивает автоматическое регулирование амплитуды выходного сигнала  $u_{2k}$ . Данный алгоритм прост в программной реализации и обладает достаточно высоким быстродействием. Его можно использовать для уменьшения уровня периодических сигналов. Однако он мало пригоден для усиления слабых сигналов, амплитуда которых находится на уровне помех. Кроме того, схема вносит в сигналы большие нелинейные искажения.

Этих недостатков в значительной степени лишен предлагаемый ниже алгоритм регулировки уровня дискретных сигналов, построенный на принципах цифровой адаптивной фильтрации, изложенных в работе [1].

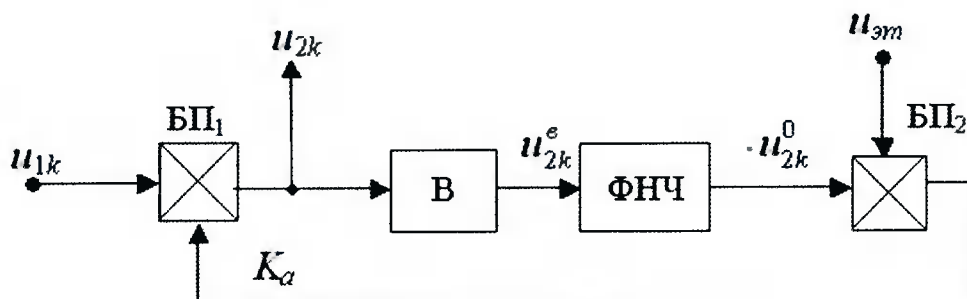


Рис. 1. Структурная схема АРУ

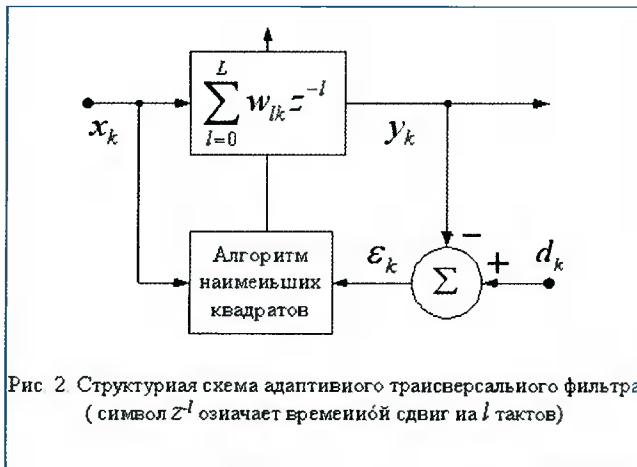


Рис 2 Структурная схема адаптивного трансверсального фильтра (символ  $z^{-l}$  означает временной сдвиг на  $l$  тактов)

Базовой структурой алгоритма является адаптивный трансверсальный фильтр, схема которого показана на рис. 2.

На схеме выходной сигнал сумматора  $y_k$  представляет собой линейную комбинацию дискретных отсчетов входного сигнала  $x_k$  и весовых коэффициентов  $w_{lk}$ , то есть

$$y_k = \sum_{l=0}^L w_{lk} x_{k-l}, \quad (3)$$

где  $l$  - номер весового коэффициента,  $k=1,2,\dots$  - номер временного отсчета.

В процессе адаптации осуществляется коррекция коэффициентов  $w_{lk}$  таким образом, чтобы выходной сигнал фильтра  $y_k$  имел наилучшее приближение к сигналу полезного отклика  $d_k$ . Для этого формируется сигнал ошибки  $\epsilon_k = d_k - y_k$ , и вектор весовых коэффициентов оптимизируется так, чтобы обеспечить минимальное значение среднеквадратического значения ошибки, т. е.  $CKO = E[\epsilon_k^2] = \min$ .

Текущие значения весовых коэффициентов вычисляются по методу наименьших квадратов

$$w_{l,k+1} = w_{lk} + 2\mu \epsilon_k x_{k-l}, \quad (4)$$

где  $\mu$  - параметр, определяющий устойчивость и скорость процесса адаптации.

Конкретное применение адаптивного сумматора определяется источником сигнала полезного отклика  $d_k$ .

На рис.3 показана структурная схема цифрового адаптивного регулятора уровня дискретных сигналов (ЦАРУ).

Схема ЦАРУ работает следующим образом. Сигнал  $u_{1k}$  поступает на вход адаптивного фильтра первого порядка с одним весовым коэффициентом  $w_k^p$ .

Выходной сигнал регулятора равен

$$u_{2k} = u_{1k} w_k^p \quad (5)$$

В качестве полезного отклика регулятора задается желаемое (эталонное) среднее значение  $u_{эм}$  амплитуды выходного сигнала  $u_{2k}$ .

Сигнал ошибки вычисляется по формуле

$$\epsilon_k = u_{эм} - u_{2k}^0, \quad (6)$$

где  $u_{2k}^0 = E[abs(u_{2k})]$  - формируемое фильтром низких частот ФНЧ<sub>2</sub> среднее значение выпрямленного детектором В<sub>2</sub> выходного сигнала  $u_{2k}$ .

В процессе адаптации коэффициент  $w_k^p$  перестраивается по формуле наименьших квадратов

$$w_{k+1}^p = w_k^p + 2\mu_p \epsilon_k u_{1k}^0. \quad (7)$$

В выражении (7)  $\mu_p$  - параметр сходимости алгоритма адаптации регулятора, а сигнал  $u_{1k}^0 = E[abs(u_{1k})]$  представляет собой формируемое фильтром низких частот ФНЧ<sub>1</sub> среднее значение выпрямленного детектором В<sub>1</sub> входного сигнала.

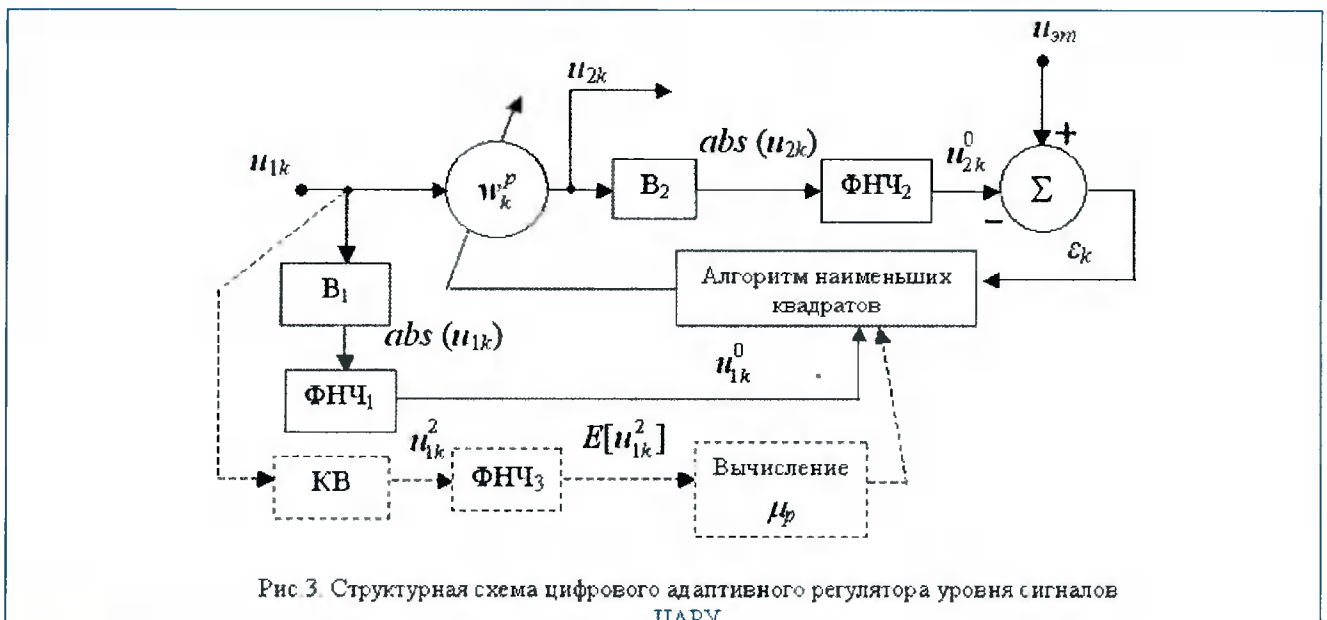


Рис 3. Структурная схема цифрового адаптивного регулятора уровня сигналов ЦАРУ

Штриховыми линиями на рис.3 обозначен дополнительный канал, предназначенный для автоматического расчета параметра  $\mu_p$ . Он состоит из блока КВ, выполняющего операцию возведения в квадрат входного сигнала, фильтра ФНЧ<sub>3</sub>, вычисляющего его мощность  $E[u_{pk}^2]$ , и блока вычисления текущего значения  $\mu_p$ .

Как видно из выражения (6), для формирования сигнала ошибки в схеме регулятора используется не сам выходной сигнал  $u_{2k}$  а его среднее значение  $u_{2k}^0$ . Анализ показал, что корректное применение адаптивного сумматора в данном приложении возможно при условии, что эта величина вычисляется с высокой точностью. В предлагаемой схеме ЦАРУ требуемая точность усреднения достигается благодаря использованию адаптивного фильтра низких частот ФНЧ<sub>2</sub>, который представляет собой адаптивный трансверсальный фильтр с одним весовым коэффициентом  $w_k^\phi$  [2].

Структурная схема адаптивного ФНЧ<sub>2</sub> показана на рис. 4. Здесь в качестве эталонного сигнала  $x_k$  используется постоянный сигнал единичной амплитуды. В этом случае выходной сигнал фильтра равен его весовому коэффициенту, т.е.  $u_{2k}^0 = w_k^\phi$  а сигнал ошибки вычисляется по формуле

$$\epsilon_k = \text{abs}(u_{2k}) - u_{2k}^0 \quad (8)$$

Весовой коэффициент определяется по рекуррентной формуле, аналогичной (7), т.е.

$$w_{k+1}^\phi = w_k^\phi + 2\mu_\phi \epsilon_k, \quad (9)$$

где  $\mu_\phi$  – параметр сходимости алгоритма адаптации фильтра ФНЧ<sub>2</sub>.

Для повышения качества работы регулятора фильтры ФНЧ<sub>1</sub> и ФНЧ<sub>3</sub> также реализуются в виде адаптивных алгоритмов, аналогичных алгоритму ФНЧ<sub>2</sub>.

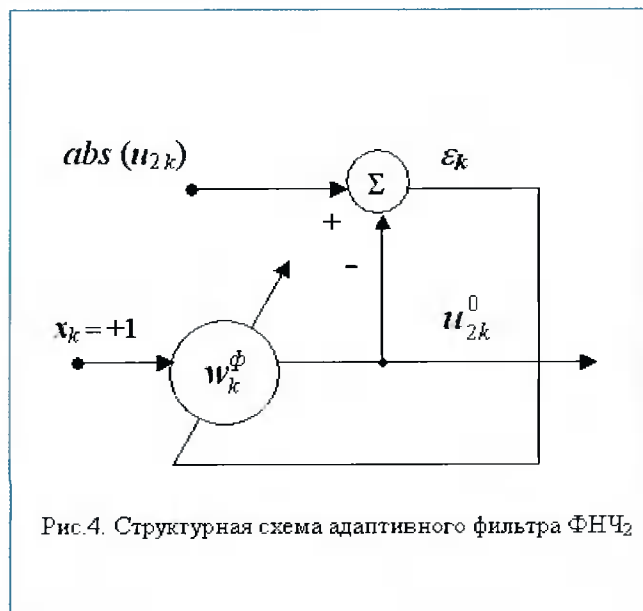


Рис.4. Структурная схема адаптивного фильтра ФНЧ<sub>2</sub>

Таким образом, в рассматриваемой системе имеются четыре адаптивных процесса:

- процесс формирования амплитуды выходного сигнала;
- процесс формирования среднего значения амплитуды входного сигнала;
- процесс формирования среднего значения амплитуды выходного сигнала;
- процесс вычисления мощности входного сигнала.

Качество работы регулятора определяется выбором параметров сходимости адаптивных процессов. Ниже приводятся некоторые рекомендации по настройке системы.

В [2] показано, что для устойчивой работы адаптивного трансверсального фильтра с одним весовым коэффициентом (рис.2) необходимо выполнение условия

$$\mu = \frac{M}{E[x_k^2]}, \quad (10)$$

где  $M$  – относительная погрешность адаптации ( $0 < M < 1$ );  $E[x_k^2]$  – мощность входного сигнала.

Время сходимости алгоритма адаптации (время сходимости СКО к минимальному значению) определяется по формуле

$$\tau_a = \frac{T_0}{\mu}, \quad (11)$$

где  $T_0$  – интервал квантования сигналов по времени.

Из приведенных соотношений видно, что чем выше требования к точности фильтра, тем меньше должен быть коэффициент  $\mu$ , что приводит к уменьшению быстродействия системы. Выбор величины  $\mu$  осуществляется путем принятия компромиссного решения.

Для определения параметра  $\mu_p$  по формуле (10) необходимо знать среднюю мощность входного сигнала  $E[u_{pk}^2]$ , которая не всегда известна. В этом случае при расчете можно вначале задать максимально допустимую мощность, которая обеспечит наибольший запас устойчивости при выбранной величине  $M$ , а затем уточнить ее значение в процессе эксплуатации.

Оптимальную настройку обеспечит использование дополнительного канала автоматического расчета текущего значения  $\mu_p$  (рис. 3).

Величины коэффициентов  $\mu_\phi$  для фильтров ФНЧ<sub>1</sub>, ФНЧ<sub>2</sub> и ФНЧ<sub>3</sub> выбираются на основании требований к полосе пропускания в соответствии с формулой [3]

$$\mu_\phi = f_\phi \pi T_0, \quad (12)$$



где  $f_\phi$  – частота, на которой мощность подавляемого сигнала уменьшается вдвое. Очевидно, что с уменьшением частоты  $f_\phi$  точность усреднения сигнала возрастает, что приводит к затягиванию переходного процесса в фильтре.

При запуске регулятора требуется задать начальное значение коэффициента  $w_p^0$ , которое может быть произвольным, но желательно, чтобы оно было положительным и не превышало единицы.

На рис.5 приведены графики, иллюстрирующие процессы адаптации в ЦАРУ с каналом автоматического вычисления параметра  $\mu_p$ .

Входной сигнал (рис.5а) представляет собой меандровые импульсы частоты 400 Гц. Амплитуда импульсов изменяется на интервале наблюдения от 2В до 0,02 В. Интервал квантования сигналов  $T_{\sigma}=4,5 \cdot 10^{-5}$  с, коэффициент  $w_p^0=0,01$  для всех трех фильтров низких частот, эталонное среднее значение выходного сигнала  $u_{zm}=0,5$ , начальное значение весового коэффициента  $w_p^0=0,01$ .

Из графика (5б) видно, что при выбранном эталонном сигнале амплитуда выходного сигнала после адаптации поддерживается на желаемом уровне  $u_{ж} \cong 1$ .

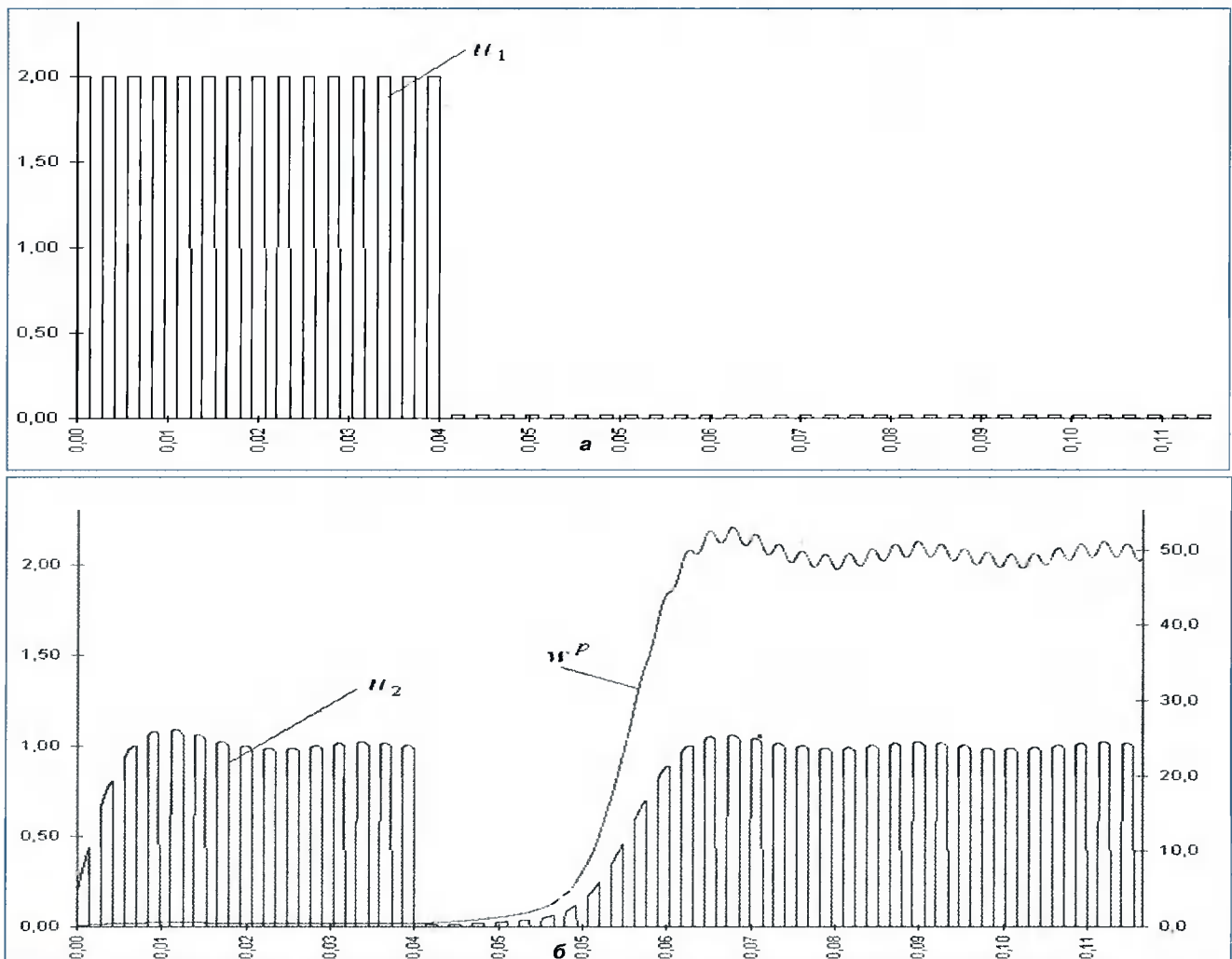
Многочисленные результаты моделирования системы на ПК показали высокое качество регулирова-

ния уровня стационарных сигналов (гармонических, импульсных и т.п.).

Использование канала  $\mu_p$  автоматической настройки параметра сходимости регулятора обеспечивает нормализацию уровня выходного сигнала при изменении амплитуды входного сигнала в диапазоне до 60дБ и более.

## Литература

1. Уидроу Б., Стирнз С., Адаптивная обработка сигналов. Пер. с англ. - М.: Радио и связь, 1989.
2. Гудков В.И., Гудкова Н.В. Программируемый адаптивный многофункциональный фильтр. В сб. «Компьютерные технологии в инженерной и управленческой деятельности». Материалы Всероссийской научно-технической конф. с международным участием. Таганрог: Изд-во ТРТУ, 2001.
3. Гудкова Н.В., Гудков В.И., и др. «Программируемый адаптивный регулятор уровня дискретных сигналов». В сб. «Проблемы современной аналоговой микросхемотехники». Международный научно-практический семинар, Шахты, 2003.



УДК 621.396.

## Структура доплеровского процессора радиолокационного сигнала при переменном периоде излучения

**В.И. Кошелев, В.Н. Горкин**

**Н**езквидистантные последовательности импульсов (временная импульсная модуляция) используются в различных радиотехнических приложениях. В радиолокации применение незквидистантной расстановки импульсов (вобуляции периода излучения или изменения скважности) зависит от типа и режима работы радиолокационной станции (РЛС) и решаемой радиолокационной задачи. В когерентно-импульсных РЛС, однозначно измеряющих дальность до лоцируемых объектов, использование вобуляции периода излучения зондирующих импульсов большой скважности позволяет улучшить скоростную характеристику систем селекции движущихся целей (СДЦ) и устранить, так называемые, слепые скорости. Соотношение периодов следования импульсов оптимизируется для улучшения равномерности скоростной характеристики. Широко используемые в данном режиме фильтры режекции пассивных помех синтезируются с учетом параметров вобуляции периода повторения импульсов. Исследованиям в данном направлении посвящены, в частности, работы [1,2].

В импульсно-доплеровских РЛС в режиме однозначного измерения радиальной скорости объектов используются зондирующие радиоимпульсы малой скважности. При этом вобуляция зондирующих импульсов позволяет устранить теневые зоны дальности и раскрыть неоднозначность ее определения.

Изменение скважности радиоимпульсов существенно влияет на спектрально-корреляционные свойства обрабатываемого эхо-сигнала. Незквидистантность расстановки импульсов на временной оси, эквивалентная их время-импульсной модуляции, приводит к появлению дополнительных спектральных составляющих. Данный эффект усиливается при увеличении глубины вобуляции. Применение классичес-

*Обоснована необходимость модификации классического радиолокационного алгоритма доплеровской фильтрации при незквидистантных импульсно-лучечных зондирующих сигналах. Выполнен статистический синтез алгоритма обработки на основе модифицированного алгоритма быстрого преобразования Фурье (БПФ), учитывающий параметры вобуляции периода повторения импульсов. Оценены пороговые сигналы при обнаружении и вычислительная эффективность доплеровского процессора.*

ких алгоритмов фильтрации для обработки незквидистантных последовательностей импульсов, не учитывающих их спектрально-временных свойств, приводит к энергетическим потерям, а также к ошибкам измерения параметров. На *рис. 1* приведены фрагменты нормированных амплитудных спектров последовательности из 10 прямоугольных видеоимпульсов при постоянной скважности  $Q=5$ , а также при двухступенчатой вобуляции периода повторения с переменной скважностью, определяемой значениями  $Q_1=4$ ,  $Q_2=6$ . Они иллюстрирует особенности спектров незквидистантных последовательностей импульсов.

Ниже рассмотрены алгоритмы обработки вобулированных последовательностей импульсов в квазинепрерывном импульсно-доплеровском режиме работы РЛС. В качестве гребенки согласованных доплеровских фильтров (МДФ) в таком режиме обычно применяются процессоры быстрого преобразования Фурье (БПФ). Если пренебречь частотным рассогласованием в пределах полосы пропускания доплеровского фильтра его амплитудно-частотная характеристика (АЧХ) примерно совпадает с модулем спектра эквидистантной пачки импульсов эхо-сигнала. При вобуляции периода повторения зондирующих импульсов в доплеровском фильтре возникают энергетические потери по сравнению с согласованным фильтром, возрастающие с ростом глубины вобуляции. Поэтому актуальной является задача синтеза алгоритмов, учитывающих параметры вобуляции

периода повторения импульсов и реализуемых на основе алгоритмов быстрых преобразований.

Рассмотрим случай, при котором отраженный сигнал представляет собой когерентную пачку импульсов с многоступенчатой вобуляцией периода повторения, а число различных периодов повторения в пачке импульсов определяется параметром  $\eta$ . Считаем, что огибающая неэквилидистантной пачки импульсов, отраженной от некоторой совокупности блестящих точек, представляет собой стационарный случайный процесс  $s(t)$  с постоянными значениями доплеровской частоты и времени задержки, обусловленной дальностью до цели, которые не изменяются за время обработки. Будем выбирать параметры вобуляции так, чтобы отношение числа импульсов в пачке сигнала  $N$  к числу различных периодов  $\eta$  было целым числом. В этом случае весь массив отсчетов сигнала можно представить в виде совокупности эквидистантных секций, сформированных с временным прореживанием по правилу:

$$s_k = \begin{cases} s\{kT\} & k=0, \eta, \dots, N-\eta \\ s\{(k+v_1)T\} & k=1, \eta+1, \dots, N-\eta+1 \\ \dots \\ s\{(k+v_{\eta-1})T\} & k=\eta-1, 2\eta-1, \dots, N-1 \end{cases}, \quad (1)$$

где  $k=0, N-1$ ,  $v=\{0, v_1, \dots, v_{\eta-1}\}^T$ ,  $v_i \in [0, 1]$  – вектор коэффициентов, определяющих сдвиг импульсов относительно среднего периода повторения  $T = \frac{1}{\eta} \sum_{m=1}^{\eta} T_m$ . При отсутствии вобуляции  $v=0$ .

Таким образом, вектор отсчетов сигнала будет представлять собой совокупность  $\eta$  секций эквидис-

тантно расположенных отсчетов  $s_1 \dots s_\eta$ , перемежающихся на временной оси и относительно сдвинутых на величину, определяемую средним периодом повторения и коэффициентами  $v_m$ . Это позволяет сформулировать принцип обработки неэквилидистантных последовательностей сигналов. Каждая из  $\eta$  секций входных данных, сформированных по правилу (1), обрабатывается с помощью алгоритма, оптимального для обработки эквидистантных импульсов, а затем результаты обработки объединяются с учетом параметров вобуляции импульсов. При синтезе алгоритма обработки будем считать эхо-сигнал дружно флюктуирующим. Выражение для достаточной статистики при обнаружении сигнала на фоне некоррелированных шумов [4] можно преобразовать к виду:

$$\tilde{\xi} = u^H w = \sum_{m=1}^{\eta} \sum_{k=0}^{N/\eta-1} u_{mk}^* w_{mk}, \quad (2)$$

где  $u = s + n$  – вектор входного процесса,  $n$  – вектор отсчетов белого гауссовского шума,  $u_1, \dots, u_h$  – векторы входного процесса, образованные из вектора входного процесса  $u$  по правилу (1),  $w_1, \dots, w_h$  – векторы обработки для соответствующих последовательностей. Каждый из векторов  $u_1, \dots, u_h$  содержит отсчеты, расположенные эквидистантно на временной оси с периодом  $\eta T$ .

При априорно неизвестном доплеровском сдвиге фазы сигнала система обработки должна быть многоканальной по частоте. При равномерной расстановке частотных каналов алгоритм (2) удобно реализовать в следующей последовательности. Исходный вектор из  $N$  отсчетов входной последовательности разбивается на секции по  $N/\eta$  отсчетов в каждой в

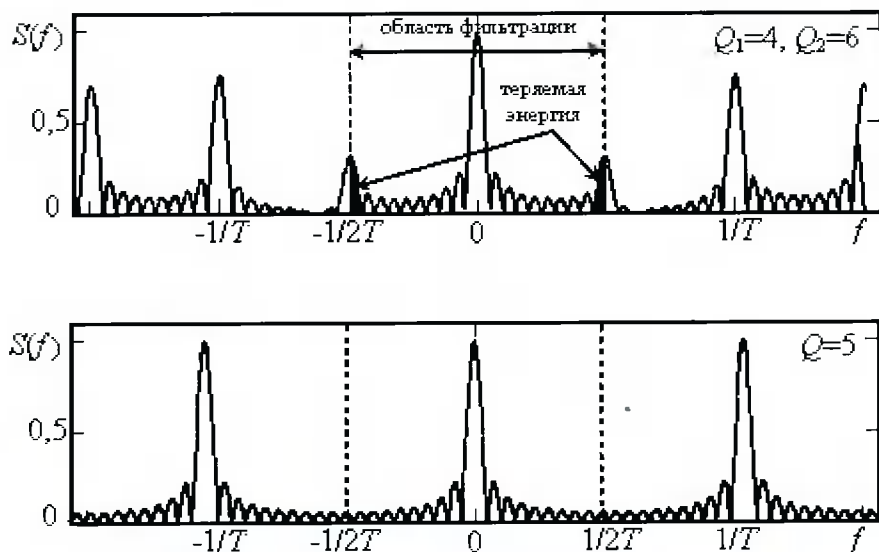


Рис. 1. Спектры сигналов



соответствии с правилом (1). Т.к. число импульсов в пачке кратно  $\eta$ , то каждая сумма  $\sum_{k=0}^{N/\eta-1} u_{nk} w_{nk}$  может быть вычислена как  $N/\eta$ -точечное дискретное преобразование Фурье (ДПФ) от соответствующих эквидистантных временных отсчетов каждой секции. При этом в качестве векторов обработки  $w_1, \dots, w_h$  выступают экспоненциальные функции. Если, кроме того, число  $N/\eta$  кратно степени 2, то алгоритмы ДПФ могут быть реализованы как  $N/\eta$ -точечные классические БПФ по основанию 2. Итоговый алгоритм многоканальной фильтрации включает операцию объединения результатов  $\eta$  секционных БПФ.

Операция разбиения исходного массива отсчетов и на секции  $u_1, \dots, u_\eta$  в соответствии с (1) аналогична разбиению входного массива в классическом алгоритме БПФ с прореживанием по времени [ ], а операция объединения промежуточных БПФ аналогична последовательности классических базовых операций алгоритма БПФ с прореживанием по времени, модифицированных с учетом параметров вобуляции периодов повторения зондирующих импульсов. Поэтому целесообразно использовать именно такой алгоритм для реализации промежуточных операций БПФ от векторов  $u_1, \dots, u_\eta$ .

Структурная схема многоканального доплеровского процессора, соответствующего алгоритму (2), представлена на рис. 2.

Входные комплексные данные запоминаются в буферном СОЗУ, затем из них формируются секции отсчетов в соответствии с правилом (1), над каждой секцией осуществляется классическое БПФ и затем результаты вычислений сводятся с помощью итоговой объединяющей операции. Стрелками показаны массивы комплексных отсчетов, размерность массива определяется числом внутри стрелки.

Рассмотрим подробнее синтез и анализ алгоритма обработки когерентной пачки сигнала для наиболее простого случая двухступенчатой вобуляции периода

повторения. Параметры, характеризующие глубину вобуляции, определим в виде следующих соотношений (полагаем  $Q1 < Q2$ ):

$$\gamma = \frac{T_2}{T_1} = \frac{Q_2}{Q_1}, \quad Q_1 = T_1/\tau, \quad Q_2 = T_2/\tau, \quad \nu = \frac{\Delta T}{2T} = \frac{T_2 - T_1}{T_2 + T_1} = \frac{Q_2 - Q_1}{Q_2 + Q_1} = \frac{\gamma - 1}{\gamma + 1},$$

где  $\tau$  – длительность импульса,  $T_1, T_2$  – периоды повторения импульсов.

Аналогично (1), при аппроксимации сигнала  $s$  в виде детерминированной экспоненциальной модели вектор сигнала можно представить в следующей форме:

$$s_k = \begin{cases} \exp\{jk\varphi_c\} & k - \text{четное} \\ \exp\{j(k+\nu)\varphi_c\} & k - \text{нечетное}, k = 0, N-1. \end{cases}$$

При таком представлении отсчетов вектора сигнала выражение для достаточной статистики (2) принимает следующий вид:

$$\tilde{z} = u^H w = \sum_{k=0}^{N/2-1} u_{1k} w_{1k} + \exp\{-j\varphi_c(1+\nu)\} \sum_{k=0}^{N/2-1} u_{2k} w_{2k}. \quad (3)$$

Первая сумма  $\sum_{k=0}^{N/2-1} u_{1k} w_{1k}$  вычисляется как  $N/2$ -точечное БПФ четных элементов входной последовательности, а сумма  $\sum_{k=0}^{N/2-1} u_{2k} w_{2k}$  – как  $N/2$ -точечное БПФ нечетных элементов входной последовательности.

Итоговая операция «сведения» результатов секционных вычислений в отличие от классического алгоритма состоит в том, что последняя так называемая «бабочка» БПФ будет учитывать параметр вобуляции  $\nu$  в величине поворачивающего множителя, вычисляемого по формуле

$$W_\nu^k = \exp\left\{-\frac{j2\pi k(1+\nu)}{N}\right\}$$

Для частотных каналов, при  $n \geq N/2$  поворачивающий множитель принимает значение  $W_\nu^k \exp\{-j\pi(1+\nu)\}$ ,

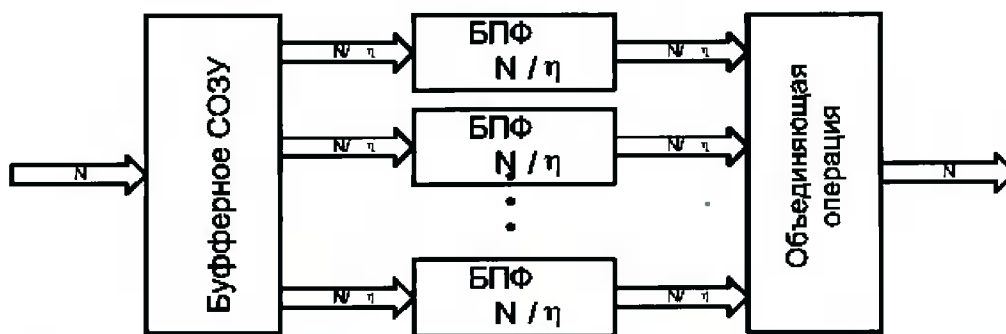


Рис. 2. Структура многоканального фильтра при вобуляции периода повторения импульсов

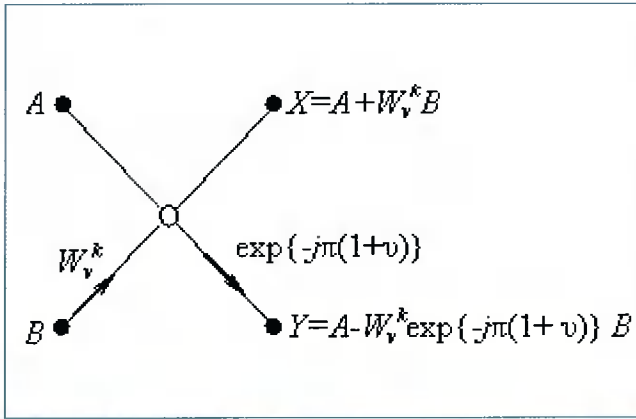


Рис.3. Модифицированная операция БПФ

вследствие чего модифицированная операция БПФ содержит дополнительную операцию умножения по сравнению с классическим алгоритмом (рис. 3), и соответственно, общее количество операций умножения по сравнению с классическим алгоритмом БПФ возрастает на величину  $N/2$ .

При соблюдении введенных условий алгоритм обработки неэквидистантного сигнала можно обобщить на случай применения произвольного ядра вобуляции  $\eta$ . Без учета тривиальных вычислительных опера-

ций и сокращения их числа при реализации алгоритма обработки для конкретных значений  $\eta$  общее количество операций умножения по сравнению с классическим алгоритмом возрастает на величину, не превышающую  $(\eta-1)N/\eta$ .

Схема, позволяющая вычислять достаточную статистику по алгоритму (3), изображена на рис. 4.

Сверхоперативное запоминающее устройство (СОЗУ) служит для хранения выборок входного сигнала, набор умножителей реализует для каждого частотного канала алгоритм взвешивания во временной области, процессор вычисляет  $N/2$ -точечное БПФ отдельно для четных и нечетных элементов входной последовательности. Кроме того, необходимо выполнить  $N$  умножений и суммирований для вычисления модифицированной операции алгоритма БПФ. Модули комплексных величин спектральных отсчетов  $|\tilde{\xi}|$  поступают на обнаружитель. В частном случае при  $T_1 = T_2$ ,  $v=0$  ( $r=1$ ), т.е. при отсутствии вобуляции алгоритм и соответствующая структурная схема совпадают с классическими.

Анализ модифицированного алгоритма показывает, что в условиях обнаружения сигнала на фоне не-

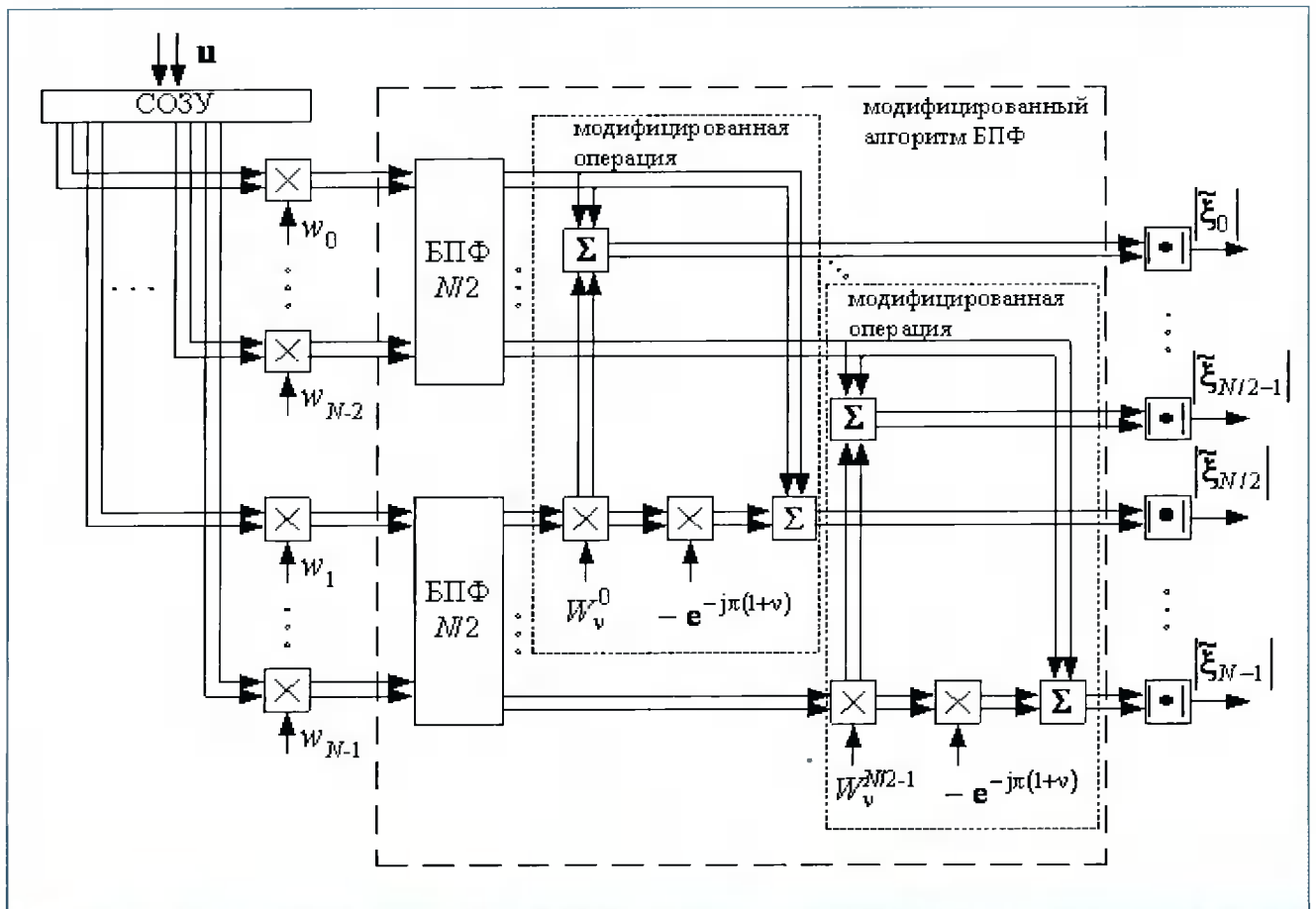


Рис. 4. Структура МДФ на базе модифицированного алгоритма БПФ

коррелированного шума и вариации параметров скважностей в диапазоне  $Q2/Q1=1,5...3$  пороговый сигнал при обнаружении удается уменьшить на величину от 1 до 10 дБ, что связано с использованием дополнительной энергии сигнала, попадающей в полосу пропускания АЧХ модифицированного доплеровского фильтра (рис. 1).

Таким образом, сформулирован общий алгоритм обработки неэквилидистантных последовательностей импульсов, реализуемый на базе алгоритмов обработки эквидистантных последовательностей, проведен статистический синтез алгоритма оптимальной обработки на основе алгоритма БПФ, получены его энергетические характеристики и проанализированы вычислительные затраты.

## Литература

1. Murakami T., Jonson R.S. Clutter suppression by use of weighted pulse trains // RCA Review.– 1971.– V. 32.– №3.– P. 402-428.
2. Лернер В.Е. Алгоритмы цифровой селекции движущихся целей при переменном периоде повторения // Радиотехника и электроника.– 1975.– Вып. 20.– №4.– С. 746-752.
3. Новосельцев Л.Я., Флягин А.Е. Обработка сигналов РЛС при вобуляции частоты повторения зондирующих импульсов // Изв. Вузов. Радиоэлектроника.– 1977.– Т. 20.– №3.– С. 40-45.
4. Кошелев В.И., Горкин В.Н. Синтез фильтров многоканальной доплеровской обработки сигналов. // Изв. вузов. Радиоэлектроника.– 2003.– Т. 46.– №1.– С. 63-68.
5. Рабинер Б., Гоулд Л. Теория и применение цифровой обработки сигналов: Пер. с англ.– М.: Мир, 1975.– 848 с.

## Уважаемые читатели!

**Вы можете подписаться на наш журнал в любом отделении связи. Подписной индекс издания по каталогу ОАО «Роспечать» – 82185.**

**Кроме того, сохраняется возможность оформить подписку непосредственно в редакции. Для этого необходимо заполнить подписной лист и отправить его по факсу (095) 290-9085 или по электронной почте [info@dspa.ru](mailto:info@dspa.ru)**

### Подписной лист

Наименование организации \_\_\_\_\_  
 ИНН \_\_\_\_\_  
 Почтовый адрес \_\_\_\_\_  
 Контактное лицо \_\_\_\_\_  
 Телефон, факс, электронная почта \_\_\_\_\_

### Банковские реквизиты

Наименование банка \_\_\_\_\_  
 Р/с \_\_\_\_\_ Город \_\_\_\_\_  
 К/с \_\_\_\_\_ БИК \_\_\_\_\_

**Справки по телефонам: (095) 290-9088, 290-9085  
 или по электронной почте: [info@dspa.ru](mailto:info@dspa.ru)**



УДК 621.391.25

## Широкополосный CELP-кодер с мультиполосным возбуждением и многоуровневым векторным квантованием по кодовой книге с реконфигурируемой структурой

**М.З. Лившиц, М. Парфенюк, А.А. Петровский**

### Введение

Телефонные сети, в общем случае, ограничивают полосу частот речевых сигналов в диапазоне 300 – 3400 Гц. Это ограничение ухудшает качество речи в части естественности, разборчивости и узнаваемости диктора. Однако цифровые сети передачи информации могут устранить данное ограничение. Подобные сети способны обеспечить высокое качество связи за счет увеличения полосы частот речевых сигналов. Определенно, к передаче широкополосной речи с диапазоном частот 50 – 7000 Гц сегодня возрастает интерес, так как она обеспечивает более высокое качество, по сравнению с телефонной полосой частот. Расширение низкочастотного диапазона вниз до 50 Гц способствует большей естественности, а высокочастотный диапазон 3400 – 7000 Гц улучшает разборчивость. Пока широкополосная речь находит применение преимущественно в конференцсвязи и видеоконференцсвязи. Ожидается, что она будет использована в ряде приложений, таких как мультимедиа службы, голосовая связь по ISDN каналам, вокодерная телефония и т.п. В большинстве этих приложений кодирование широкополосной речи играет ключевую роль в эффективности передачи и хранения широкополосных речевых сигналов, в которых большинство наиболее важных формант обычно располагается в области низких частот, что обуславливает большую концентрацию энергии в низкочастотной области. Данные свойства используются в схемах кодирования на ос-

*Описывается эффективный подход для повышения качества субполосного CELP-кодирования. Предлагается модель широкополосного CELP-кодера с реконфигурацией структуры кодовой книги на основании оценок порога маскирования, полученного из психоакустической модели. Это позволяет снизить объем передаваемой информации, сохраняя высокое перцептуальное качество реконструированного речевого сигнала. Показано, что мультиполосный принцип возбуждения в широкополосном CELP-кодере совместно с многоуровневым векторным квантованием субполосного сигнала возбуждения позволяет существенно улучшить качество реконструированной речи, а также обеспечивает возможность кодирования не только речевых данных, но и аудио сигналов данного частотного диапазона.*

нове преобразования и субполосного кодирования с целью получения высокого качества. Основной принцип схем на основе преобразования и субполосного кодирования – декомпозиция речевого сигнала на субполосы и раздельное их кодирование.

Перцептуальное кодирование аудио сигнала использует технику адаптивного окрашивания шума, в которой входной аудио сигнал подвергается декомпозиции банком фильтров или ортогональному преобразованию в частотно-временные компоненты. Затем эти компоненты адаптивно квантуются, что обеспечивает более эффективное маскирование шума. При этом требования к хорошему временному и частотному разрешению не могут быть обеспечены одновременно в одной структуре банка фильтров. Более того, так как аудиторная система человека работает в неравномерной частотной шкале, известной как шкала барков [1], равномерная субполосная декомпозиция не отражает ее частотно зависимое поведение. В частности, одна субполоса может покрывать несколько критических полос в низкочастотной области, увеличивая при этом эффекты субоптимального окрашивания шума.

В последнее время ведутся активные исследования в области широкополосного CELP-кодирования.

Современные широкополосные кодеры речи обычно строятся по встраиваемой технологии: пирамидальная структура [2], многополосная структура [3], многоуровневая структура [4]. Чаще всего компрессоры имеют не жестко фиксированную структуру, а гибкую, настраиваемую на базе психоакустической модели восприятия акустической информации человеком. При этом возникает вопрос о максимальном соответствии психоакустической модели и структуры широкополосного компрессора речевых сигналов. Одним из перспективных подходов повышения качества субполосного CELP-кодирования является техника многоуровневого векторного квантования сигнала возбуждения в частотных полосах [4]. При этом осуществляется возбуждение несколькими источниками (субполосное возбуждение) и реализуется принцип спектрального вычитания вклада субполосного сигнала возбуждения. Обработка субполос осуществляется в направлении от наиболее важных, в перцептуальном смысле, к менее значимым. Основным смыслом многоуровневого векторного квантования субполосного сигнала возбуждения заключается в том, что ошибка квантования (остаточный сигнал) после кодирования очередной субполосы квантуется на следующем уровне. Таким образом, осуществляется процесс «обеления» шума (ошибки) квантования, что приводит к улучшению степени компрессии аудио данных. Введение в состав кодера психоакустической модели и мультиполосной кодовой книги с реконфигурируемой структурой, отражающей свойства кодируемого сигнала, позволяет исключить статистическую и перцептуальную избыточность, снизив тем самым требования к объему передаваемой информации.

В настоящей работе предлагается модель широкополосного CELP-кодера с реконфигурацией структу-

ры кодовой книги по результатам работы психоакустической модели [5], реализованной на базе ДПФ с неравномерным разложением коэффициентов Фурье по частотной шкале (от англ. – Warped Discrete-Fourier Transform - WDFT) [6], которая позволяет снизить объем передаваемой информации, сохраняя высокое перцептуальное качество реконструированного речевого сигнала.

### Актуальность субполосного кодирования

Рассмотрим субполосный кодер, представленный на рис. 1, состоящий из  $k=0, 1, \dots, M-1$  каналов. Каждый канал имеет анализирующий фильтр  $H_k(z)$ , синтезирующий фильтр  $F_k(z)$ , дециматор/интерполятор с коэффициентом передискретизации  $m_k$  и  $b_k$ -битный квантователь. Входной сигнал  $x[n]$  после пропуска через анализирующие фильтры принимает значения  $x'_k[n]$ , а субполосные сигналы –  $x_k[n]$ . Субполосный кодер максимально децимирован, т.е. выполняется условие [7]:

$$\sum_{k=0}^{M-1} m_k = 1 \tag{1}$$

Пусть  $x_k$  –  $k$ -й спектральный компонент сигнала, а  $Q(x_k)$  – его  $b_k$ -битный квантованный аналог,  $Q$  – операция квантования. Тогда ошибка реконструкции  $k$ -го компонента равна  $\sigma_k = x_k - Q^{-1}(Q(x_k))$ . Другими словами – это внесенное искажение в сигнал в результате его кодирования. Среднее число бит на один спектральный компонент равно:

$$b = \frac{1}{M} \sum_{k=0}^{M-1} b_k \tag{2}$$

где  $M$  – количество спектральных компонент (каналов в банке фильтров).

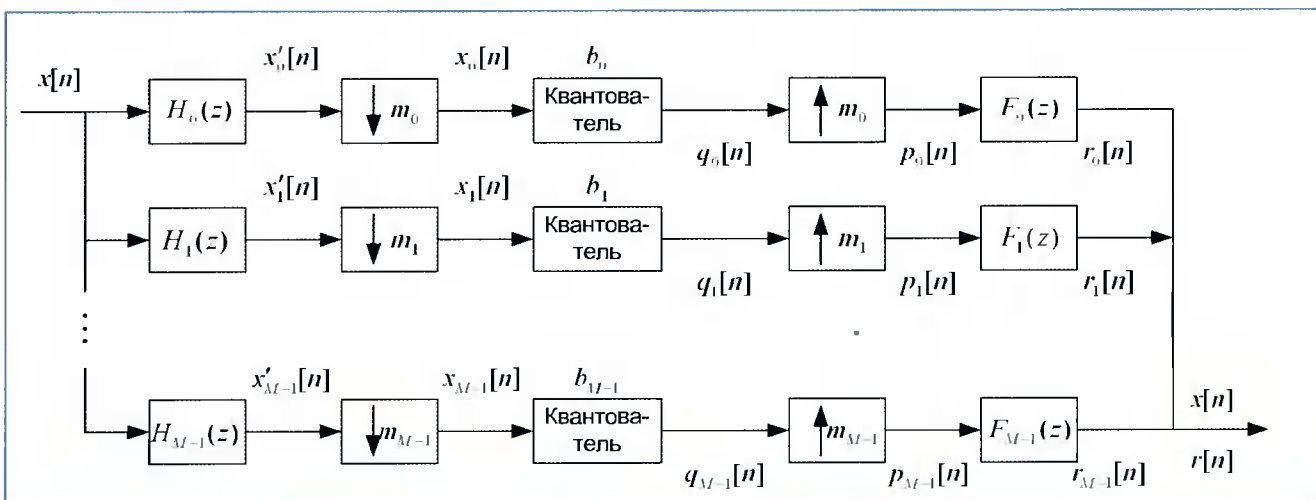


Рис. 1. Субполосный кодер



Принимая во внимание, что шум квантователя является белым [8], дисперсия внесенных искажений в сигнал в результате кодирования для ИКМ квантователя равна:

$$\sigma^2 = \frac{1}{M} \sum_{k=0}^{M-1} \left( \frac{x_k^2}{3 \cdot 2^{2b_k}} \right). \quad (3)$$

Целью оптимизации является минимизация дисперсии ошибок реконструкции  $\sigma^2$  при ограничении на общее распределение бит. Обозначим число уровней реконструкции для квантования компоненты  $k$ -го канала банка фильтров  $L_k = 2^{b_k}$ , тогда

$$b = \frac{1}{M} \sum_{k=0}^{M-1} \log_2 L_k = \frac{1}{M} \log_2 \prod_{k=0}^{M-1} L_k. \quad (4)$$

Далее,

$$2^b = \prod_{k=0}^{M-1} L_k = L_g^M, \text{ где } L_g = \left( \prod_{k=0}^{M-1} L_k \right)^{\frac{1}{M}} \quad (5)$$

является средним геометрическим значением уровней реконструкции квантователя.

Минимизация дисперсии внесенных искажений при кодировании сигнала основывается на методе множителей Лагранжа  $\lambda$ :

$$\frac{d}{dL_k} \left\{ \frac{1}{M} \sum_{k=0}^{M-1} \frac{x_k^2}{3 \cdot L_k^2} + \lambda \prod_{k=0}^{M-1} L_k \right\} = 0. \quad (6)$$

После дифференцирования и некоторых преобразований, формула оптимального распределения бит по каналам банка фильтров примет вид:

$$b_k = b + \frac{1}{2} \log_2 (x_k^2) - \frac{1}{2} \log_2 \left( \prod_{k=0}^{M-1} x_k^2 \right)^{\frac{1}{M}}. \quad (7)$$

Из выражения (7) следует, что минимальное число бит в каждом  $k$ -ом канале определяется распределением спектральной энергии в сигнале, и выигрыш в количестве бит по сравнению с однополосным банком фильтров будет только в том случае, когда среднее геометрическое значение спектральной плотности мощности сигнала будет много меньше её среднего арифметического значения. Отношение среднего геометрического значения спектральной плотности мощности сигнала к её среднему арифметическому значению есть мера пологости спектра сигнала (*Spectral Flatness Measure – SFM*) [9]:

$$SFM = \frac{\left( \prod_{k=0}^{M-1} x_k^2 \right)^{\frac{1}{M}}}{\frac{1}{M} \sum_{k=0}^{M-1} x_k^2}. \quad (8)$$

Из формулы (8) видно, что значения *SFM* варьируются в пределе от 0 до 1. Если *SFM* = 1, то подразумевается, что входной сигнал с пологим спектром и соответственно никакого увеличения компрессии нельзя получить. Пусть *SFM* = 1, тогда, согласно (7), получается, что  $b_k = b$ . Следует отметить, что *SFM* зависит не только от распределения спектральной энергии сигнала, но также и от разрешающей способности банка фильтров, т.е. от общего числа  $M$  каналов в банке фильтров.

Таким образом, мерой избыточности в сигнале является мера пологости спектра *SFM* [9]: чем более пологий спектр сигнала, тем меньше избыточности в сигнале. Малое значение *SFM* подразумевает потенциально высокую степень компрессии сигнала, которую естественно можно оценить числом бит, необходимых для кодирования сигнала без артефактов. Из выше приведенных формул может быть получено выражение, показывающее уменьшение энтропии входного сигнала за счет его разбиения банком фильтров.

В перцептуальном кодере сигналов стоит цель не только устранения информационной избыточности, но и изоляции перцептуальной избыточности акустической информации в сигнале. Это желание расположить в спектре сигнала внесенные искажения в реконструированный сигнал в результате кодирования ниже порога маскирования, т.е. порога восприятия акустической информации слушателем. Соотношение сигнал-шум SNR для квантования компонент каналов банка фильтров равно

$$SNR = 10 \log \frac{x^2}{\sigma^2}, \quad (9)$$

а соотношение сигнал к порогу маскирования  $T$  определяется следующим образом:

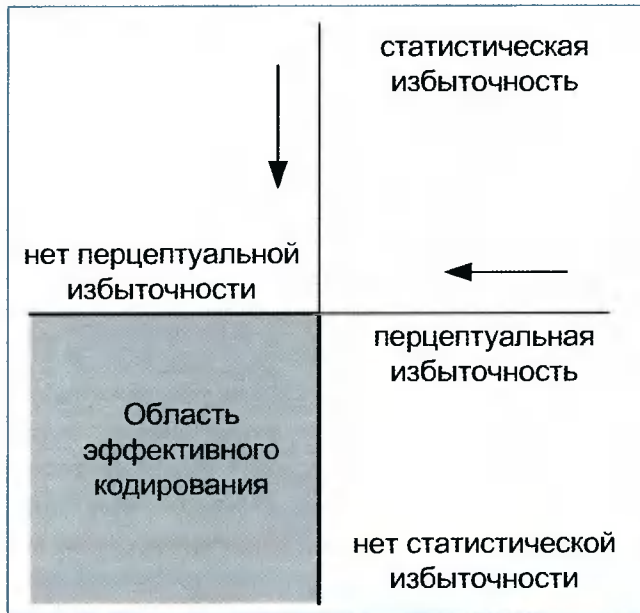
$$SMR = 10 \log \frac{x^2}{T^2}. \quad (10)$$

Далее, для компонент сигнала  $k$ -го канала, значения которых больше порога маскирования  $T_k$ , хотелось бы максимизировать разность  $SNR - SMR$ , или, что эквивалентно, минимизировать разность  $SMR - SNR$ . Для соотношения  $SMR - SNR$  с учетом дисперсии  $\sigma^2$  (3), дисперсия внесенных кодированием искажений, взвешенная маскирующим фактором, равна

$$\frac{\sigma^2}{T^2} = \frac{1}{M} \sum_{k=0}^{M-1} \left( \frac{x_k^2 / T_k^2}{3 \cdot 2^{2b_k}} \right) \quad (11)$$

где  $T_k$  – уровень порога маскирования в  $k$ -ом канале банка фильтров.





**Рис. 2.** Характеристика информационной емкости аудио сигнала в частотной области и область его эффективного кодирования (стрелками показаны направления уменьшения избыточности соответственно статистической и перцептуальной)

Минимизация данной взвешенной ошибки, аналогично варианту минимизации дисперсии ошибки реконструкции  $\sigma^2$  (3), приводит к следующей формуле оптимального распределения бит по каналам банка фильтров:

$$b_k = b + \frac{1}{2} \log_2 \left( \frac{x_k^2}{T_k^2} \right) - \frac{1}{2} \log_2 \left( \prod_{k=0}^{M-1} \frac{x_k^2}{T_k^2} \right)^{\frac{1}{M}} \quad (12)$$

Из формулы (12) следует, что мера перцептуальной избыточности [10] определяется как отношение

$$PSFM = \frac{\left( \prod_{k=0}^{M-1} \frac{x_k^2}{T_k^2} \right)^{\frac{1}{M}}}{\frac{1}{M} \sum_{k=0}^{M-1} \frac{x_k^2}{T_k^2}} \quad (13)$$

Как видно из выражения (13), PSFM зависит от распределения по частотному диапазону спектральной энергии, взвешенной энергией порога маскирования. Характеристика информационной емкости сигнала в частотной области и область его эффективного кодирования схематически показаны на рис. 2.

Таким образом, при построении банков фильтров необходимо учитывать количество субполос, на которые осуществляется декомпозиция с учетом принятой шкалы, а также перераспределение бит между субполосами на основе меры перцептуальной избыточности (13) с целью минимизации их общего количества для получения выигрыша при субполосном кодировании

## Структура широкополосного CELP-кодера с мультиполосным возбуждением и многоуровневым векторным квантованием сигнала возбуждения

### Структура кодера и декодера

На основании всего вышеизложенного, предлагается следующая структурная схема широкополосного CELP-кодера (рис. 3), ядром которой является мультиполосная кодовая книга, состоящая из субполосных кодовых книг форм и коэффициентов усиления.

Входной сигнал квантуется при помощи 16 бит/отсчет с частотой дискретизации  $f_s = 16$  кГц, после чего кодируется широкополосным кодером, структура которого представлена на рис. 3а. Кодер использует блочный LP-анализ, без интерполяции коэффициентов линейного предсказания. В качестве оптимального для полосы частот 50 – 8000 Гц, был определен порядок фильтра-предсказателя (Short-Term Predictor - STP), равный  $p=16$ . Для улучшения работоспособности взвешивающего фильтра (осуществления независимого контроля формы спектральной огибающей и формантного взвешивания) на входе используется предыскажающий фильтр с передаточной функцией [11]:

$$F(z) = 1 - \mu \cdot z^{-1} \quad (14)$$

где  $\mu$  - адаптивный параметр.

LP-анализ осуществляется по предыскаженной оригинальной речи. Инверсный фильтр имеет передаточную характеристику

$$A(z/\gamma_1) = 1 - \sum_k a_k \gamma_1^k z^{-k}, \quad (15)$$

где  $a_k$  - коэффициенты линейного предсказания,  $\gamma_1$  - коэффициент, осуществляющий расширение спектральных пиков в диапазоне 0–60 Гц (подстройка осуществляется по результатам минимизации ошибки линейного предсказания).

Взвешивающий синтезирующий фильтр определяется следующим выражением:

$$W(z/\gamma_2) = \frac{1}{1 + \sum_k a_k \gamma_2^k z^{-k}} \quad (16)$$

где  $\gamma_2 = R(1)/R(0)$  (определен экспериментально),  $R(\cdot)$  - значение функции корреляции для указанного в скобках отсчета входного сигнала.

Взвешенный входной сигнал (фрейм) разбивается на 4 субфрейма длительностью 5мс, для каждого из которых осуществляется подбор параметров долговре-

менного фильтра-предсказателя (*Long-Term Predictor – LTP*)  $P(z) = 1 - \beta \cdot z^{-D}$  по замкнутому циклу и квантование по восьмиполосной кодовой книге. Параметры *LTP*-фильтра определяются по взвешенному оригинальному широкополосному сигналу. Используется *LTP*-фильтр-предсказатель 1-го порядка с диапазоном задержек 16 – 160 отсчетов (100 – 1000Гц).

Кодер имеет восьмиуровневую структуру, т.е. количество уровней соответствует количеству обрабатываемых субполос. Поиск осуществляется последовательно по мере перцептуальной важности полос, т.е. от низкочастотных к высокочастотным. Очередной оптимальный вектор с соответствующим коэффициентом усиления пропускается через последовательно соединенные *LTP*- и взвешивающий синтезирующий фильтр. По окончании цикла минимизации ошибки из остаточного взвешенного оригинального сигнала вычитается вклад очередной субполосы (уровня). Процесс выполняется последовательно для всех восьми субполос. После обработки всех полос производится синтез речевого вектора по оптимальному восьмиполосному возбуждающему сигналу. Далее *LTP*-буфер обновляется синтезированным вектором сигнала, и осуществляется обновление памяти взвешивающего синтезирующего фильтра. Алгоритм повторяется для оставшихся субфреймов.

### Разбиение частотной оси на субполосы

В рассматриваемом широкополосном перцептуальном *CELP*-кодере с целью отражения частотно-зависимого поведения аудиторной системы человека принята шкала барков [1], и используется схема разбиения на субполосы, представленная в таб. 1.

Таблица 1 – Характеристика субполос кодера

Субполоса	Частотный диапазон, Гц	Барки
1	100-510	4
2	510-1080	4
3	1080-1720	3
4	1720-2320	2
5	2320-3150	2
6	3150-4100	1.5
7	4100-5300	1.5
8	5300-8000	3

Выбор субполос – результат информативных тестов прослушивания, а также оптимизации на основе критериев (8) и (13). Ширина полос увеличивается с частотой с целью отражения частотной избирательности аудиторной системы человека и формантной структуры речи. Для формирования субполосных обучающих множеств был использован полифазный банк фильтров с косинусной модуляцией и неравномерным разбиением на полосы [12], согласно таб. 1. АЧХ банка фильтров анализа представлена на рис. 4.

Необходимо отметить, что отфильтрованные кодовые векторы имеют конечную длительность (равную длине субфрейма), вследствие чего существует утечка «просачивание» энергии между смежными кодовыми книгами (субполосами). Однако, так как кодовые векторы разных кодовых книг близко ортогональные, последовательный поиск в каждой субполосной кодовой книге обеспечивает почти тоже качество, что и оптимальный совместный поиск во всех субполосных книгах, но со значительным снижением сложности. Более того, «просачивание» спектральной энергии *i*-й субполосы может быть компенсировано при квантовании

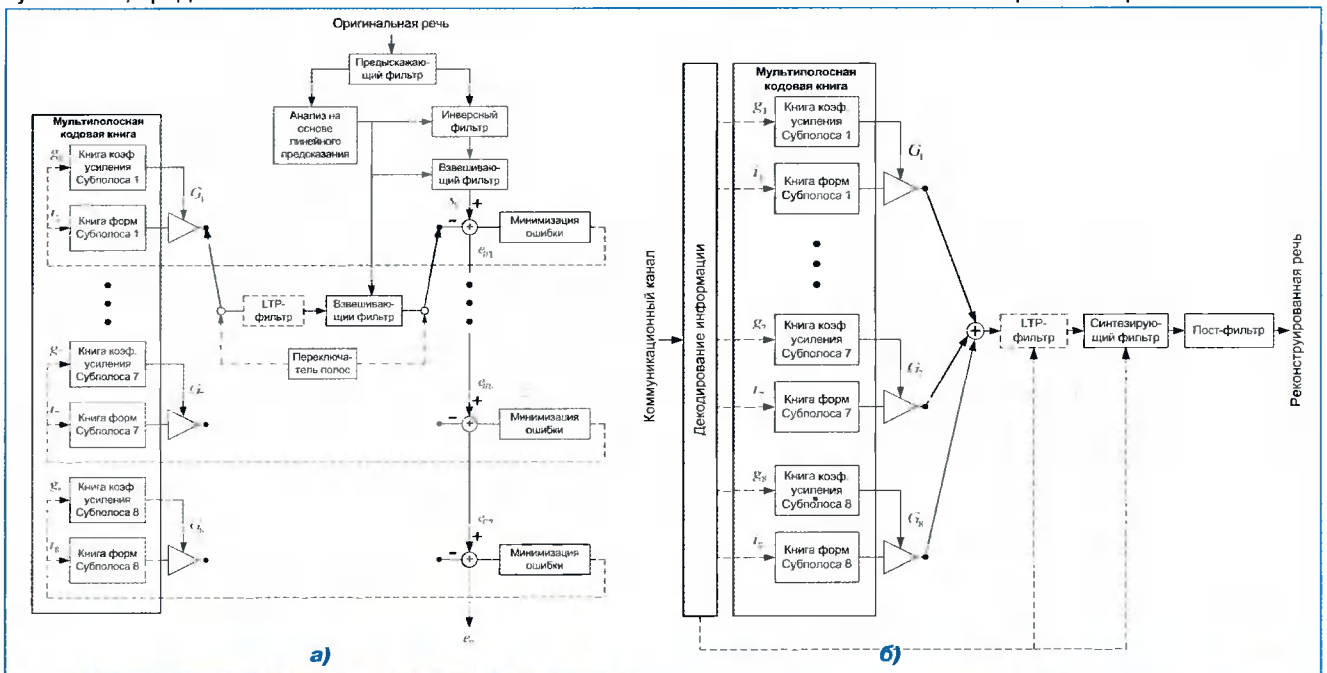


Рис. 3. Структурная схема широкополосного *CELP*-кодера с многоуровневым векторным квантованием и мультиполосным возбуждением: а) кодер; б) декодер

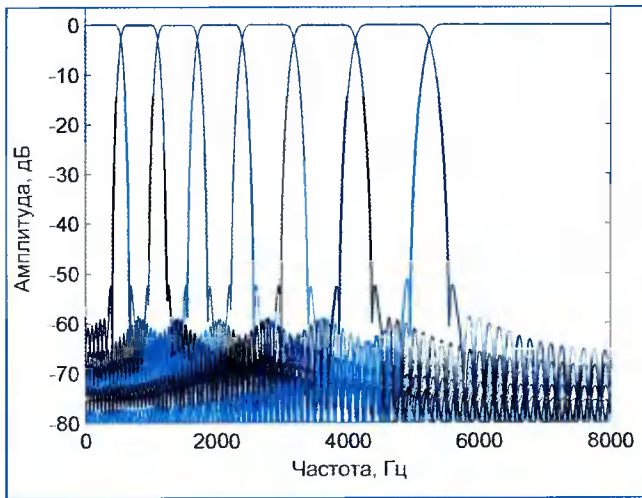


Рис. 4. АЧХ банка фильтров

$(i+1)$ -й субполосы, поэтому нет необходимости в банке фильтров с совершенной реконструкцией.

**Алгоритм работы многоуровневого векторного квантователя**

Процесс многоуровневого векторного квантования субполосного сигнала возбуждения может быть представлен в виде выражений (17):

$$\begin{aligned}
 e_{w1}(n) &= s_w(n) - G_1 \hat{s}_{w1}(n), \\
 e_{wi}(n) &= e_{w(i-1)} - G_i \hat{s}_{wi}(n), \quad 2 \leq i \leq 8, \\
 e_w(n) &= s_w(n) - \sum_{i=1}^8 G_i \hat{s}_{wi}(n),
 \end{aligned}
 \tag{17}$$

где  $s_w(n)$  – взвешенная оригинальная речь,  $\hat{s}_{wi}(n)$  – взвешенная синтезированная речь  $i$ -го уровня (субполосы),  $G_i$  – коэффициент усиления сигнала  $i$ -го уровня (субполосы),  $e_{wi}(n)$  – ошибка квантования  $i$ -го

уровня,  $e_{w(i-1)}(n)$  – остаточный сигнал после квантования предыдущего  $i-1$  уровня (ошибка квантования  $i-1$  уровня).

Коэффициенты линейного предсказания преобразуются в линейные спектральные пары (ЛСП) и кодируются при помощи расщепленного векторного квантования с разбиением на 3 субвектора размерностью 3,3,10, соответственно. Каждый субвектор квантуется по 9-битной кодовой книге. Таким образом, ЛСП представляются 27 битами, т.е. 1.6875 бит/ЛСП/фрейм. В процедуре поиска кодового слова минимизируемая норма вектора ошибки учитывает расстояние между ЛСП [13] и уменьшение частотного разрешения в области высоких частот. Индексы векторов возбуждения, коэффициенты усиления, параметры STP-фильтра (коэффициент усиления модели и ЛСП), параметры LTP-фильтра передаются в декодер (рис. 3б).

Индексы векторов и значения коэффициентов их усиления восстанавливаются, формируются субполосные сигналы возбуждения, которые суммируются и окрашиваются синтезирующим фильтром. После чего осуществляется адаптивная пост-обработка реконструированной речи и коррекция предсказания. Характеристика квантования параметров модели речеобразования представлена в таб. 2.

Анализ табл. 2 показывает, что основной вклад в суммарный поток данных вносит информация о коэффициентах усиления субполосных сигналов возбуждения и их индексах в кодовой книге. При этом требуемый объем информации, необходимый для кодирования параметров, увеличивается пропорционально количеству рассматриваемых субполос и глубине соответствующих кодовых книг.

**Таблица 2 – Характеристика квантования параметров модели речеобразования широкополосного CELP-кодера с фиксированной предельной структурой кодовой книги**

Параметр модели	Кол-во парам./фрейм	Кол-во бит/параметр	Суммарное количество бит/фрейм	Кол-во фреймов/с	Скорость потока, бит/с
LSF	16	1.6875	27	50	1350
Model Gain	1	7	7		350
LTP Delays	4	8	32		1600
LTP Gains	4	5	20		1000
Excitation Gains	8 субполос * 4 субфрейма = 32	4	128		6400
Book Indexes	8 субполос * 4 субфрейма = 32	8	256	12800	
Результирующий поток данных:					23500



Данная схема обладает рядом преимуществ: высокое качество реконструированной речи, скорость потока данных 23.5 кбит/с, отсутствие банка фильтров анализа (субполосные кодовые книги формируются на этапе проектирования кодера). Однако существенным недостатком является фиксированный характер структуры кодовой книги, который может быть устранен путем использования субполосных кодовых книг с многоуровневой структурой с разной степенью детализации (количеством векторов). Структура мультиполосной кодовой книги при этом преобразуется к виду, представленному на рис. 5. Субполосные книги могут быть как структурированными, так и нет. Под структурированностью кодовой книги будем понимать определенную организацию для ускорения поиска, например, в форме древовидной структуры.

Модифицированная структура мультиполосной многоуровневой кодовой книги (рис. 5) позволяет пользователю в зависимости от требований, предъявляемых к качеству речи и пропускной способности коммуникационного канала, выбрать количество кодируемых субполос и ограничить глубину поиска в субполосных кодовых книгах. Такая структура эффективна для коммуникационных каналов с фиксированным трафиком, однако на практике часто приходится иметь дело с переменной пропускной способностью. К тому же эта структура не учитывает свойства аудиторной системы человека и интегрирующей способности его слуха, при этом устраняется только статистическая избыточность сигнала, в то время как перцептуальная избыточность все еще присутствует. Для устранения перцептуальной избыточности предлагается использовать психоакустическую модель, согласующуюся с принятой критической частотной шкалой разбиения на субполосы и структурой многоуровневого векторного квантования субполосного сигнала возбуждения, результатом работы которой является оптимальная для рассматриваемого фрейма структура кодовой книги.



Рис. 5. Мультиполосная кодовая книга с многоуровневой структурой

## Психоакустическая модель на базе WDFТ

### Выбор подхода построения модели

В известной работе Джонстона [14] психоакустическая модель основана на ДПФ: расчет ДПФ взвешенного сегмента сигнала, группировка коэффициентов преобразования в группы, соответствующие критическим частотным полосам восприятия, и расчет энергии в критических частотных полосах. Достижение приемлемого спектрального разрешения в наименьших критических частотных полосах требует использования ДПФ с достаточно длинным временным окном. Поэтому концептуальная простота и эффективность нивелируются недостаточным временным разрешением, неприемлемым для анализа более тонкого феномена, такого как маскирование назад («pre-masking») [15]. Второй класс, использующий банки фильтров для декомпозиции сигнала на неравномерные полосы, исключает этот недостаток, однако требует значительного увеличения вычислительных затрат, особенно если необходима хорошая аппроксимация критических полос. Ни одно из решений не превосходит другое, оба из подходов находят свое применение.

Исследования, проведенные в работах [5] и [6], показали, что малоразмерное *WDFТ* может успешно заменить *ДПФ* с большой длиной выборки. Это оказалось возможным благодаря тому, что *WDFТ* позволяет разместить частотные компоненты в соответствии с распределением критических частотных полос, поэтому в психоакустической модели на базе *WDFТ* могут быть уравновешены как хорошее частотное, так и временное разрешение.

### Основные принципы WDFТ

WDFТ последовательности  $x[n]$  из  $N$  точек определяется по формуле [6]:

$$\hat{X}(z_k) = X(\hat{z}_k) = \sum_{n=0}^{N-1} x[n] \hat{z}_k^{-n}, \quad k = 0, \dots, N-1, \quad (18)$$

где  $\hat{z}_k$  – изображения равноотстоящих точек на единичной окружности в  $z$ -плоскости, получаемые из преобразования

$$z_k^{-1} = e^{-j \frac{2\pi k}{N}} \rightarrow \hat{z}_k^{-1} = A(z_k), \quad k = 0, \dots, N-1 \quad (19)$$

с произвольным порядком всепропускающей функции  $A(z)$ .

Простейший вариант WDFТ основан на всепропускающем звене первого порядка с действительным коэффициентом [6]:



$$z^{-1} \rightarrow A(z) = \frac{-a + z^{-1}}{1 - az^{-1}} \quad (20)$$

Условием стабильности является  $|a| < 1$ . В зависимости от знака  $a$ , растягивается низкочастотный ( $a > 0$ ) или высокочастотный ( $a < 0$ ) диапазон, таким образом, что оставшаяся часть единичной окружности становится сжатой. Формально это может быть выражено [5]:

$$\hat{\omega} = \omega + 2 \arctan\left(\frac{a \sin \omega}{1 - a \cos \omega}\right) \quad \text{for} \quad \begin{cases} z = e^{j\omega} \\ \hat{z} = e^{j\hat{\omega}} \end{cases}, \quad (21)$$

то есть осуществляется билинейное преобразование  $z$ -плоскости в новую искривленную  $\hat{z}$ -плоскость (рис. 6).

Таким образом, *WDFT* является обобщением *ДПФ* с частотными компонентами, расположенными неравномерно, но регулярно на единичной окружности. В матричной записи (с  $\hat{X}[k]$ , обозначающим  $\hat{X}[z_k]$ ) *WDFT* может быть представлено выражением (22):

$$\begin{bmatrix} \hat{X}[0] \\ \hat{X}[1] \\ \vdots \\ \hat{X}[N-1] \end{bmatrix} = \begin{bmatrix} 1 & A(z_0) & \dots & A(z_0)^{N-1} \\ 1 & A(z_1) & \dots & A(z_1)^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & A(z_{N-1}) & \dots & A(z_{N-1})^{N-1} \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix}. \quad (22)$$

*WDFT* наследует некоторые свойства *ДПФ*. Одним из важнейших является сопряженная симметрия для действительных данных:

$$\hat{X}[N-k] = \hat{X}^*[k] \quad (23)$$

Алгоритм преобразования, по эффективности сравнимый с *БПФ*, не может быть построен для *WDFT* из-за асимметричности *WDFT*-матрицы. Тем не менее, прямой алгоритм работы с комплексной матрицей может быть значительно оптимизирован. Наиболее усовер-

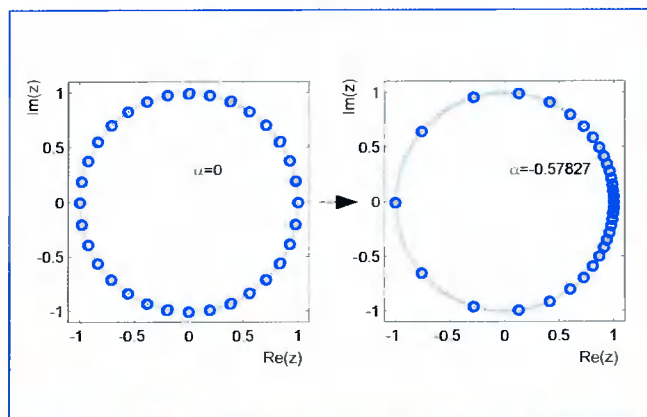


Рис. 6. Расположение частотных отсчетов *ДПФ* (слева) и *WDFT* (справа)

шенствованный алгоритм, несмотря на сложность  $O(N^2)$ , был предложен в работе [6]. Он использует факторизацию *WDFT*-матрицы в произведение действительной *ДПФ* (осуществляемой через *БПФ*) и комплексной диагональной матрицы. Данный метод хорошо подходит для обработки изображений, где все данные поступают одновременно. В обработке аудио сигналов, где отсчеты поступают последовательно один за другим, может быть использована даже прямая реализация преобразования. Учитывая выражение (23), преобразование (22) может быть записано в следующем виде:

$$\begin{bmatrix} \hat{X}[0] \\ \hat{X}[1] \\ \vdots \\ \hat{X}[N-1] \end{bmatrix} = \sum_{n=0}^{N-1} \begin{bmatrix} A(z_0)^n \\ A(z_1)^n \\ \vdots \\ A(z_{N-1})^n \end{bmatrix} x[n] \quad (24)$$

Каждый элемент в данной сумме относится только к одному входному отсчету. Он может быть рассчитан при поступлении отсчета и сохранен для получения результата после поступления  $N$  отсчетов. При этом вычислительная нагрузка на входной отсчет равна  $O(N)$ .

#### Аппроксимация шкалы барков

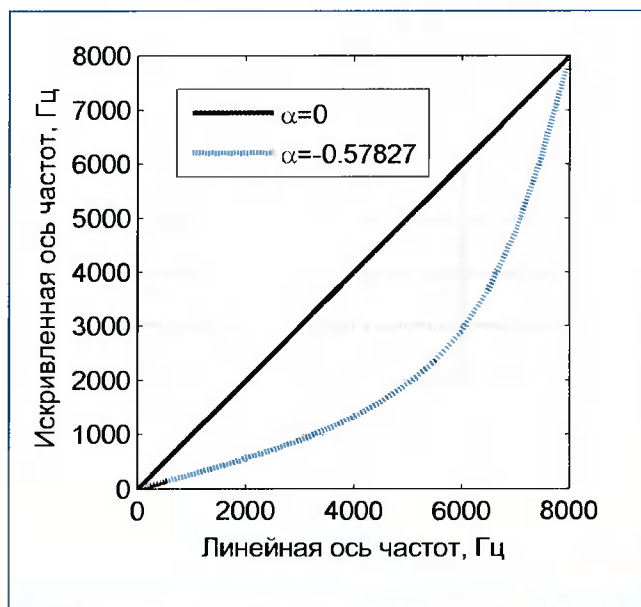
Первый шаг при использовании *WDFT* в психоакустической модели – проектирование соответствующего всепропускающего преобразования. Частотные коэффициенты  $z$  преобразования должны быть представлены равномерно в перцептуальной области. В работе [16] было показано, что всепропускающее звено первого порядка достаточно хорошо аппроксимирует перцептуальную шкалу барков, при этом значение коэффициента всепропускающего фильтра для заданной частоты дискретизации определяется по следующему выражению:

$$a_{\text{Bark}} = 0.1957 - 1.048 \cdot \left[ \frac{2}{\pi} \arctan\left(0.07212 \frac{f_s}{1000}\right) \right]^2. \quad (25)$$

Для случая  $F_s=16\text{kHz}$  коэффициент  $a_{\text{Bark}} = -0.57827$ .

На рис. 7 представлена трансформация (искривление) частотной оси согласно шкале барков.

Так как *ДПФ* имеет равномерное частотное разрешение (в то время как ширина критических частотных полос строго изменяется с их местоположением на частотной шкале), то различное количество коэффициентов преобразования ассоциируется с конкретной критической частотной полосой. В части *A таб. 3* количество коэффициентов в группах варьируется от 3 до 38, в то время как для *WDFT* той же размерности не отдается предпочтения ни одной из полос, все коэф-



**Рис. 7. Искривление частотной оси в соответствии с критической шкалой барков**

коэффициенты преобразования распределены практически равномерно (часть Б).

*WDFT* в его оригинальной форме не сохраняет энергию сигнала в соответствующих частях единичной

окружности до и после всепропускающего преобразования (рис. 8). Так как искривление приводит к растяжению одного частотного диапазона при одновременном сжатии другого, оно должно сопровождаться масштабированием *WDFT*-амплитуд с целью коррекции уровня энергии в каждой критической частотной полосе. В работе [5] предлагается следующая коррекция:

$$\hat{X}_{PC}(z) = \frac{\sqrt{1-a^2}}{1-az} \hat{X}(z) \quad (26)$$

Рассмотрим в качестве примера широкополосный аудио сигнал, который содержит голосовое сообщение (первые 3.5с) и музыкальный фрагмент с вокальным сопровождением (рис. 9 (сверху)). Построим спектрограммы данного сигнала при помощи *ДПФ* и *WDFT* (рис. 9 центр и низ, соответственно). Как видно из рис. 9, *WDFT*-решение дает достаточно четкий результат, оценка энергии более точная. Представленные факты свидетельствуют о том, что *WDFT* может действительно служить базисом для психоакустической модели.

**Таблица 3 – Отображение коэффициентов *ДПФ* и *WDFT* на критические частотные полосы [5]**

Крит. полоса	Часть А (размер FFT = 512, Fs=16 кГц)			Часть Б (размер WDFT = 512, Fs=16 кГц)		
	Диапазон коэффициентов	Кол-во	Диапазон частот, Гц	Диапазон коэффициентов	Кол-во	Диапазон частот, Гц
1	1 – 3	3	31 - 94	1 - 12	12	8 – 100
2	4 – 6	3	125 - 188	13 - 24	12	109 – 202
3	7 – 9	3	219 - 281	25 - 36	12	210 – 305
4	10 – 13	4	313 - 406	37 - 48	12	314 – 412
5	14 – 16	3	438 - 500	49 - 60	12	421 – 523
6	17 – 20	4	531 - 625	61 - 73	13	533 – 650
7	21 – 24	4	656 - 750	74 - 85	12	660 – 776
8	25 – 29	5	781 - 906	86 - 97	12	787 – 912
9	30 – 34	5	938 - 1063	98 - 110	13	923 – 1073
10	35 – 40	6	1094 - 1250	111 - 123	13	1086 - 1254
11	41 – 46	6	1281 - 1438	124 - 135	12	1269 - 1443
12	47 – 54	8	1469 - 1688	136 - 148	13	1460 - 1680
13	55 – 62	8	1719 - 1938	149 - 161	13	1700 - 1961
14	63 – 73	11	1969 - 2281	162 - 174	13	1985 - 2302
15	74 – 86	13	2313 - 2688	175 - 186	12	2331 - 2690
16	87 – 102	16	2719 - 3188	187 - 198	12	2726 - 3174
17	103 – 122	20	3219 - 3813	199 - 210	12	3220 - 3792
18	123 – 145	23	3844 - 4531	211 - 221	11	3851 - 4513
19	146 – 173	28	4563 - 5406	222 - 231	10	4588 - 5328
20	174 – 205	32	5438 - 6406	232 - 242	11	5419 - 6412
21	206 – 243	38	6438 - 7594	243 - 252	10	6520 - 7533
22	244 – 256	13	7625 - 8000	253 - 256	4	7650 - 8000

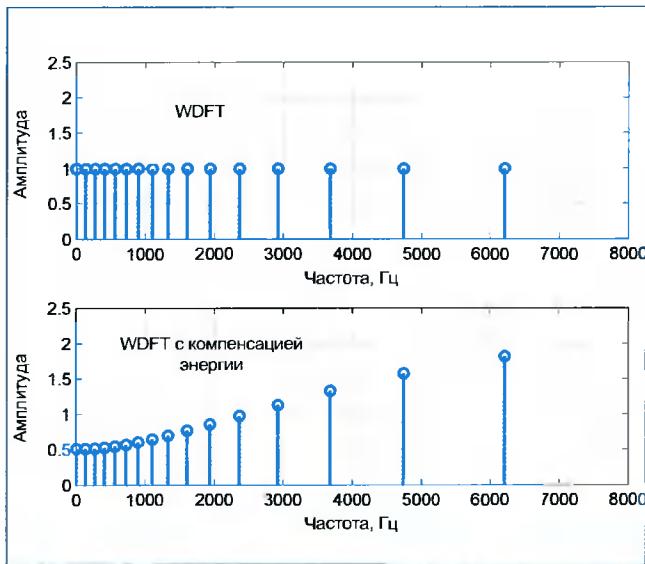


Рис. 8. Компенсация энергии WDFT ( ) единичного импульса

### Реконфигурирование структуры кодовой книги на основе субполосной перцептуальной энтропии

Оценка порога маскирования основана на общей психоакустической модели [14], предложенной Джонстоном в 1988 году. Данная процедура была подстроена под широкополосную речь с полосой 8 кГц и модифицирована для использования свойств WDFT. Основные шаги оценки порога маскирования следующие: расчет энергии критических частотных полос по WDFT-спектру мощности; свертка с функцией распро-

странения; вычитание относительного смещения порога; нормализация и сравнение с абсолютным порогом слышимости.

Энергия рассчитывается в критических частотных полосах путем суммирования компонент спектра мощности в каждой критической частотной полосе. Коэффициенты преобразования группируются в субполосы в соответствии с таблицей 1 и 3, в каждой из которых осуществляется оценка порогов маскирования. Как упоминалось выше, в случае WDFT мы имеем неравномерную концентрацию энергии в окрестностях преобразуемых центральных частот. Поэтому, для вычисления энергии, сосредоточенной в критических частотных полосах, необходимо учесть масштабирование (26).

Дискретное представление энергии критической частотной полосы может быть выражено как взвешенная сумма компонент спектра мощности:

$$E_b = \frac{1}{2\pi} \sum_{k \in K_b} \Delta\omega_k |\hat{S}(\omega_k)|^2 \quad (27)$$

где  $b$  – номер критической частотной полосы,  $K$  – количество коэффициентов преобразования, попадающих в  $b$ -ю критическую частотную полосу,  $k$  – индекс коэффициента преобразования (спектрального отсчета),  $\hat{S}(\omega_k)$  – оценка спектра оригинального речевого сигнала.

На следующем шаге вычисляется свертка энергий критических частотных полос  $E_b$  с функцией распрост-

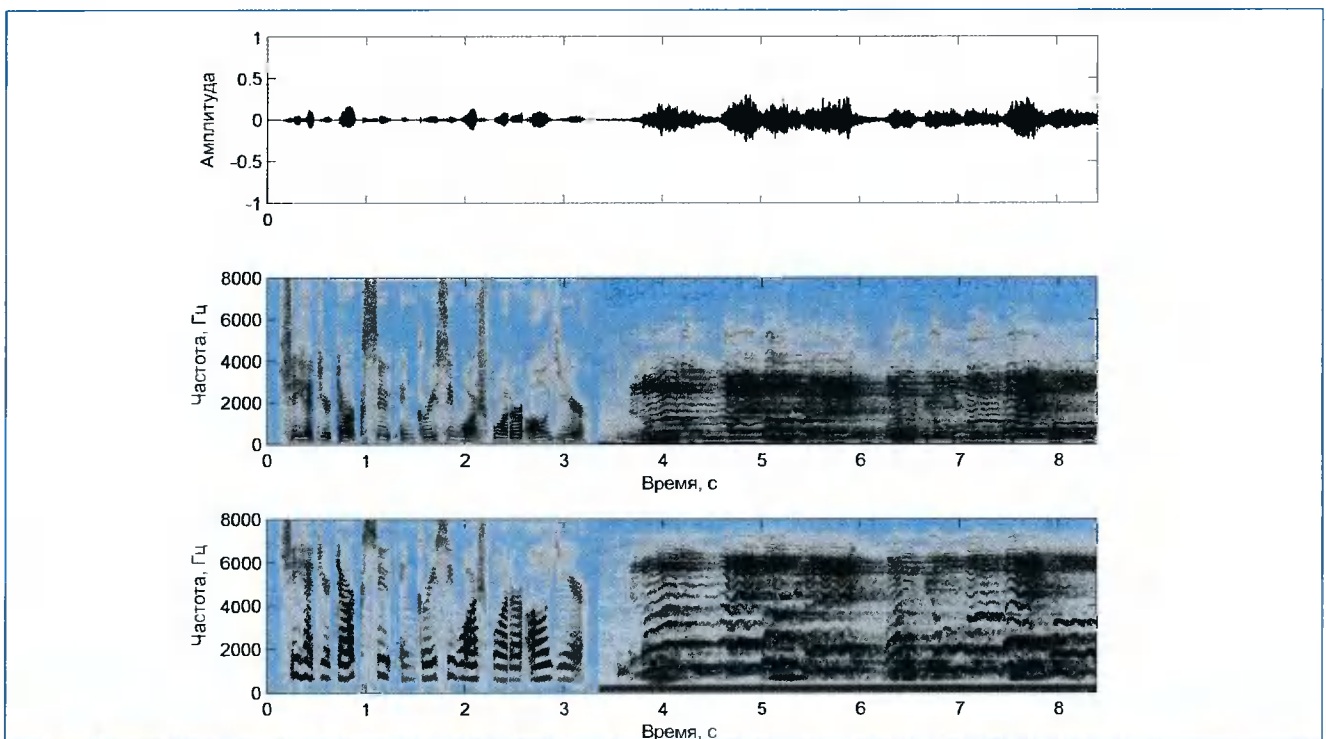


Рис. 9. Широкополосный сигнал во временной области (сверху) и его спектрограммы, полученные при помощи ДПФ (в центре) и WDFT с коэффициентом (снизу)



ранения по базиллярной мембране  $SF_b$  для учета распространения маскирования:

$$C_b = E_b * SF_b \quad (28)$$

Относительное смещение порога в каждой критической частотной полосе рассчитывается с использованием меры тональности. Для определения типа маскира (шумоподобный или тональный) используется мера спектральной пологости спектра (8), которая определяется как отношение среднего геометрического  $\mu_{gb}$  к среднему арифметическому  $\mu_{ab}$  спектральных компонент внутри критической полосы  $b$ :

$$SFM_b = 10 \log_{10} \left( \frac{\mu_{gb}}{\mu_{ab}} \right), [\text{дБ}] \quad (29)$$

Данная оценка может быть затруднена в случае обычных моделей с малым разрешением, так как количества ДПФ коэффициентов в низкочастотных полосах обычно не достаточно для эффективного расчета  $SFM$ . В подобных случаях  $SFM$  рассчитывается для всего спектра или predeterminedных тональностей для каждой критической частотной полосы. В случае  $WDFT$ , коэффициенты преобразования группируются равномерно в критических частотных полосах, следовательно,  $SFM$  может быть определена для каждой полосы отдельно.

Смещение порога маскирования  $O_b$  оценивается по следующему выражению:

$$O_b = ton_b \cdot (14.5 + b) + (1 - ton_b) \cdot 5.5 \quad [\text{дБ}] \quad (30)$$

где  $ton_b$  - индекс тональности, определяемый как

$$ton_b = \min \left( \frac{SFM_b}{-60 \text{дБ}}, 1 \right). \quad (31)$$

Для получения энергии порога маскирования относительное смещение (30) вычитается из свернутого спектра критической полосы (28):

$$T_b = 10^{(\log_{10} C_b - O_b / 10)} \quad (32)$$

а затем осуществляется нормализация и сравнение с абсолютным порогом слышимости.

По вычисленной энергии порога маскирования осуществляется определение перцептуальной важности кодируемых субполос или субполосная перцептуальная энтропия [14]:

$$SPE_b = \log_2 \left( \frac{E_b}{\sqrt{\frac{6 \cdot T_b}{bandwidth_b}}} + 1 \right), [\text{бит/полосу}] \quad (33)$$

где  $bandwidth_b$  - ширина субполосы  $b$  в спектральных отсчетах.

Для того чтобы преобразовать перцептуальную энтропию каждой субполосы  $SPE_b$  в глубину поиска в соответствующей субполосной кодовой книге  $dim_b$ , используется масштабирование вектора  $SPE$  для каждого анализируемого фрейма входного сигнала таким образом, чтобы максимальное значение в векторе  $SPE_{max}$  соответствовало максимально возможной глубине поиска в субполосной кодовой книге  $dim_{b, MAX}$ . Для

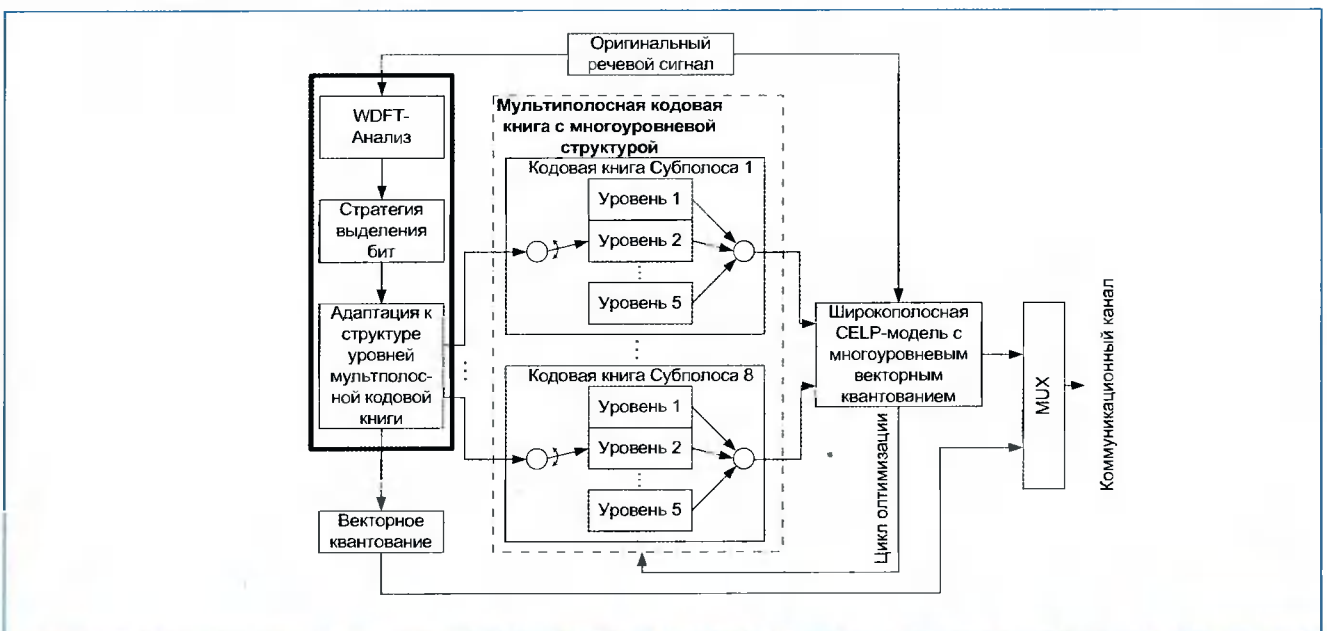


Рис. 10. Схема реконфигурации структуры кодовой книги широкополосного кодера на базе WDFT-психоакустической модели



этого вычисляется вклад каждого уровня субполосной книги  $PE\_per\_level_b$  в суммарное значение субполосной перцептуальной энтропии  $SPE_{max}$ :

$$if \quad dim_{b,MAX} \geq \max_{b \in B} (SPE_b) \\ PE\_per\_level_b = \frac{\max_{b \in B} (SPE_b)}{dim_{b,MAX}} \quad (34)$$

else

$$PE\_per\_level_b = \frac{dim_{b,MAX}}{\max_{b \in B} (SPE_b)} \\ dim_b = round (SPE_b \cdot PE\_per\_level_b)$$

где  $B$  – множество субполос,  $round$  – операция округления.

Упрощенная схема реконфигурирования структуры мультиполосной многоуровневой кодовой книги представлена на рис. 10.

Принцип изменения структуры книги заключается в следующем. Часть, отвечающая за реконфигурацию кодовой книги кодера, обведена жирным контуром (рис. 10(слева)). Оригинальный речевой сигнал (очередной кодируемый фрейм) поступает на вход блока *WDFT*-анализа, в котором осуществляется оценка субполосных порогов маскирования  $T_b$ , затем производится оценка перцептуальной энтропии каждой из субполос кодера  $SPE_b$ . Таким образом, на выходе мы имеем

своего рода стратегию выделения бит или вектор  $SPE$ , содержащий значения перцептуальной важности кодируемых субполос для текущего фрейма сигнала. Данная стратегия, согласно алгоритму (34), преобразуется в вектор, содержащий глубину поиска (номер уровня детализации векторов возбуждения) в субполосных кодовых книгах. При этом если  $SPE_b = 0$ , то данная субполоса удаляется из рассмотрения, т.е. не кодируется индекс вектора возбуждения и его коэффициент усиления. Вектор  $dim$  квантуется по 10-битной кодовой книге и передается на сторону декодера. Если обозначить  $Stage\_Size_b[i]$  размер  $i$ -го уровня субполосной кодовой книги  $b$  в векторах, то количество бит, необходимых для кодирования индексов векторов возбуждения в многоуровневой мультиполосной кодовой книге, составляет

$$Book\ Indexes = 4 \cdot \sum_{\substack{b=1 \\ SPE_b > 0}}^B \log_2 (Stage\_Size_b[dim_b]) \quad [бит/фрейм], \quad (35)$$

а для коэффициентов усиления

$$Excitation\ Gains = 4 \cdot \sum_{\substack{b=1 \\ SPE_b > 0}}^B 4, \quad [бит/фрейм]. \quad (36)$$

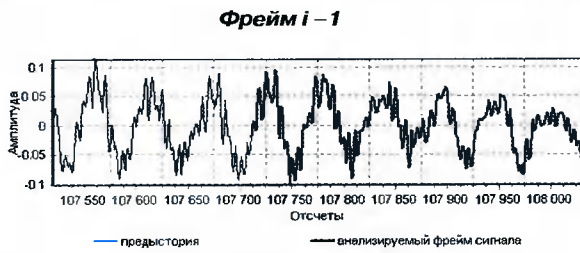
Таким образом, схема квантования параметров модели речеобразования преобразуется к виду, представленному в таб 4.

**Таблица 4 – Характеристика квантования параметра модели речеобразования широкополосного CELP-кодера с реконфигурируемой структурой кодовой книги**

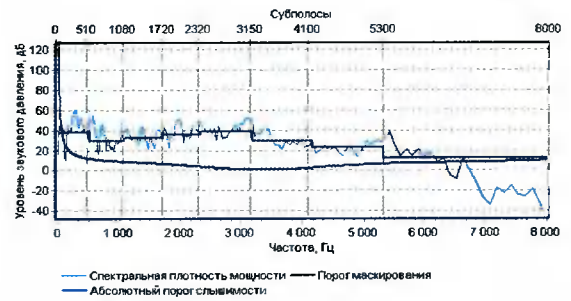
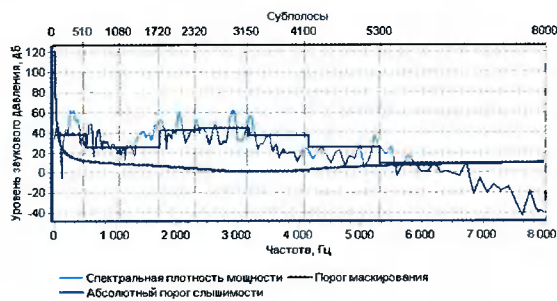
Параметр модели	Кол-во парам./фрейм	Колво бит/параметр	Суммарное количество бит/фрейм	Колво фреймов/с	Скорость потока, бит/с
LSF	16	1.6875	27	50	1350
Model Gain	1	7	7		350
LTP Delays	4	8	32		1600
LTP Gains	4	5	20		1000
dim	1 вектор из 8 элементов	1.25	10		500
Excitation Gains	Выражение 36	4	VAR		VAR
Book Indexes	Выражение 36	VAR	VAR	VAR	
Результирующий поток данных:					VAR

Процесс реконфигурации мультиполосной многоуровневой кодовой книги по этапам представлен на *рис. 11*. Буферизация анализируемого речевого сигнала (*рис. 11а*) позволяет учитывать свойства предыдущего анализируемого фрейма, при этом можно избежать резких изменений в структуре кодовой книги (*рис. 11е*). На *рис. 11б*

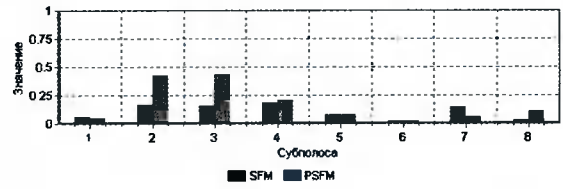
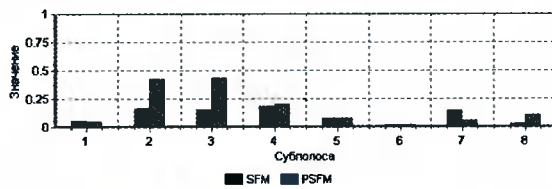
проиллюстрированы пороги маскирования в субполосах для двух последовательных фреймов. Анализ *рис. 11в,г,д*, показывает, что чем меньше значение меры пологости спектра субполосного сигнала, тем больше величина отношения сигнал-маска и, следовательно, значение субполосной перцептуальной энтропии.



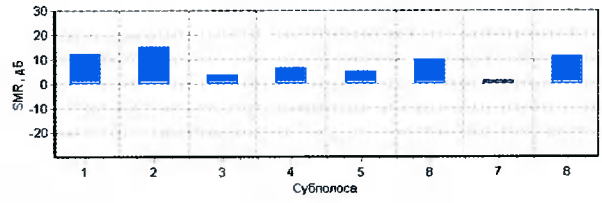
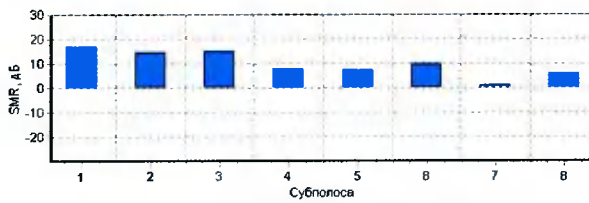
**а) буфер анализируемого речевого сигнала**



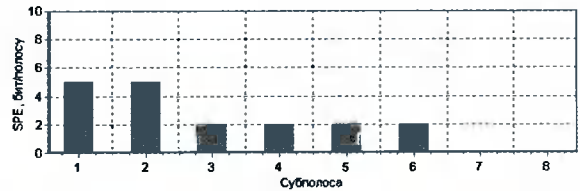
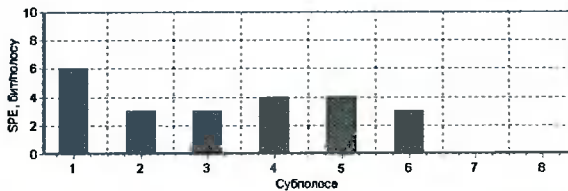
**б) пороги маскирования в субполосах**



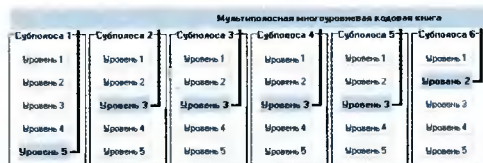
**в) оценка меры пологости спектра SFM и PSFM в субполосах**



**г) соотношении сигнал-маска SMR а субполосах**



**д) величины субполосной перцептуальной энтропии**



**е) структура мультиполосной многоуровневой кодовой книги**

**Рис. 11. Пример работы психоакустической модели с результатами реконфигурации мультиполосной многоуровневой кодовой книги**

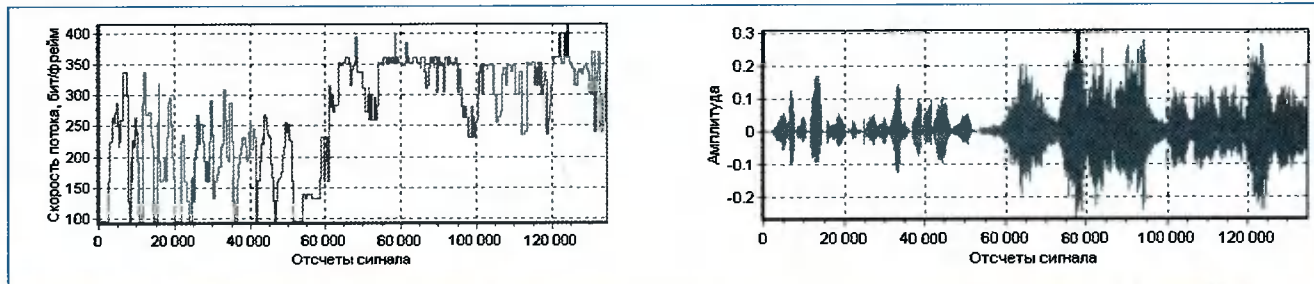


Рис. 12. Изменение скорости потока в зависимости от перцептуальной важности сигнала: огибающая сигнала (справа), поток данных (слева)

На рис. 12 приведен пример изменения скорости потока в зависимости от перцептуальной важности сигнала.

Как видно из рис. 12, предлагаемая схема психоакустической оптимизации (реконфигурации) структуры мультиполосной многоуровневой кодовой книги кодера позволяет существенно снизить вклад данных об индексах векторов возбуждения и коэффициентах их усиления в результирующий информационный поток. При этом увеличение количества бит, необходимых для кодирования информации, наблюдается лишь в перцептуально значимых областях (фрагментах) речевого сигнала.

### Оценка качества широкополосного CELP-кодера

При оценке качества кодера применялась следующая модификация книг: многоуровневые субполосные кодовые книги с 5-уровневой организацией и глубиной каждого уровня – 16;32;64;128;256 векторов, т.е. 4;5;6;7;8 бит на уровень, соответственно. В качестве тестового материала использовались речевые данные с частотой дискретизации  $f_s$  и квантованием 16 бит/от-

счет из базы данных TIMIT. Длительность речевого материала составила 5 мин. Оценка качества реконструированного речевого сигнала осуществлялась по методике, описанной в работах [17],[18], на базе двух психоакустических моделей, ядром которых являлось дерево пакета дискретного вейвлет-преобразования, разделяющее частотный интервал аудио сигнала на полосы, согласно критической шкале частот (барков) *CB-WPD*, и *WDFT*. Были оценены следующие величины: отношение сигнал-шум *SNR* и отношение энергии шума (энергии вносимых искажений) к порогу маскирования *NMR* [19]; искажение спектра барков *BSD* и искажение спектра барков, определяемое по модифицированному методу *MBSD* [20]-[23], обладающему сильно выраженной корреляцией с субъективной оценкой качества, проводимой по процедуре *MOS*. Сравнительный анализ оценки качества реконструированной речи для фиксированной и реконфигурируемой структуры кодовой книги кодера представлен в таб. 5.

Анализ таб. 5 показывает, что две различные модели оценки субъективного качества дают приблизительно одинаковые результаты, при этом качество реконструированной речи оценивается выше четырех баллов по пя-

Таблица 5 – Сравнительный анализ оценок качества реконструированного речевого сигнала для различных структур кодовой книги

Структура мультипол. кодовой книги	Скор. потока, бит/с			Модель объективной оценки субъективно предсказанного качества на базе										
				CB-WPD					WDFT					
	мин.	сред.	макс.	SNR seg, дБ	NMR seg, дБ	NMR total, дБ	BSD (MOS)	MBSD (MOS)	SNR seg, дБ	NMR seg, дБ	NMR total, дБ	BSD (MOS)	MBSD (MOS)	
Фикс. кол-во кодир. субполос	1	6700			1.72	5.61	6.11	1.76	1.69	1.85	5.22	5.96	1.82	1.80
	2	9100			4.04	3.23	4.25	2.03	1.83	4.62	3.12	3.75	2.07	2.09
	3	11500			6.12	2.15	3.10	2.77	2.79	6.20	2.02	2.56	2.83	2.78
	4	13900			8.74	0.95	2.56	3.10	3.04	8.96	0.76	1.23	3.15	3.12
	5	16300			9.80	-0.82	0.53	3.91	3.89	9.89	-0.97	0.21	3.97	3.99
	6	18700			10.50	-2.56	-3.47	4.02	4.03	10.68	-2.81	-3.82	4.09	4.06
	7	21100			11.73	-4.49	-5.32	4.15	4.07	11.95	-4.63	-5.56	4.22	4.19
	8	23500			12.35	-6.78	-8.61	4.26	4.19	12.57	-7.08	-8.79	4.35	4.31
С психоакустической оптимиз. (WDFT)	4300	9255.7	22200	10.95	-5.45	-7.24	4.08	4.11	11.07	-5.66	-7.51	4.29	4.22	



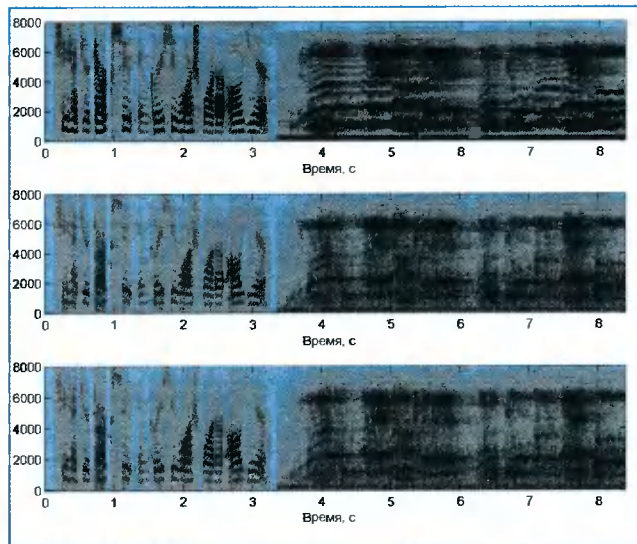
тибальной шкале *MOS*. Оценка *NMRtotal* находится в пределах от -3.47 дБ (6-полосная структура) до -8.61 дБ (8-полосная структура), т.е. граница порога восприятия минимум в 1.41 раза ниже порога маскирования. Кодер с реконфигурируемой структурой кодовой книги при средней скорости потока данных 9255.7 бит/с и максимальной – 22200 бит/с обеспечивает перцептуальный уровень качества, сопоставимый с уровнем, обеспечиваемым предельной восьмиполосной структурой.

На *рис. 13* представлены спектрограммы оригинального речевого сигнала, реконструированного для кодера с предельной структурой кодовой книги и восстановленного после кодирования кодером с реконфигурируемой кодовой книгой. Из анализа спектрограмм, представленных на *рис. 13*, видно, что качество речевого сигнала, восстановленного после кодирования предлагаемым кодером с предельной структурой кодовой книги и с реконфигурируемой, по спектральному составу практически не различается. Сравнительный анализ предлагаемой психоакустической модели и *MPEG Psychoacoustic Model 1* [24], проведенный в работе [25], показал, что предложенная схема обеспечивает лучшую степень компрессии информации (устранение перцептуальной избыточности) при отсутствии каких-либо перцептуальных различий между сравниваемыми моделями.

На *рис. 14* представлен сравнительный анализ качества предлагаемого кодера с широкополосными кодерами, реализованными по стандарту *ITU-T G.722* [26], из которого видно, что предлагаемый кодер превосходит по качеству кодер *ITU-T G.722* со скоростью 64 кбит/с.

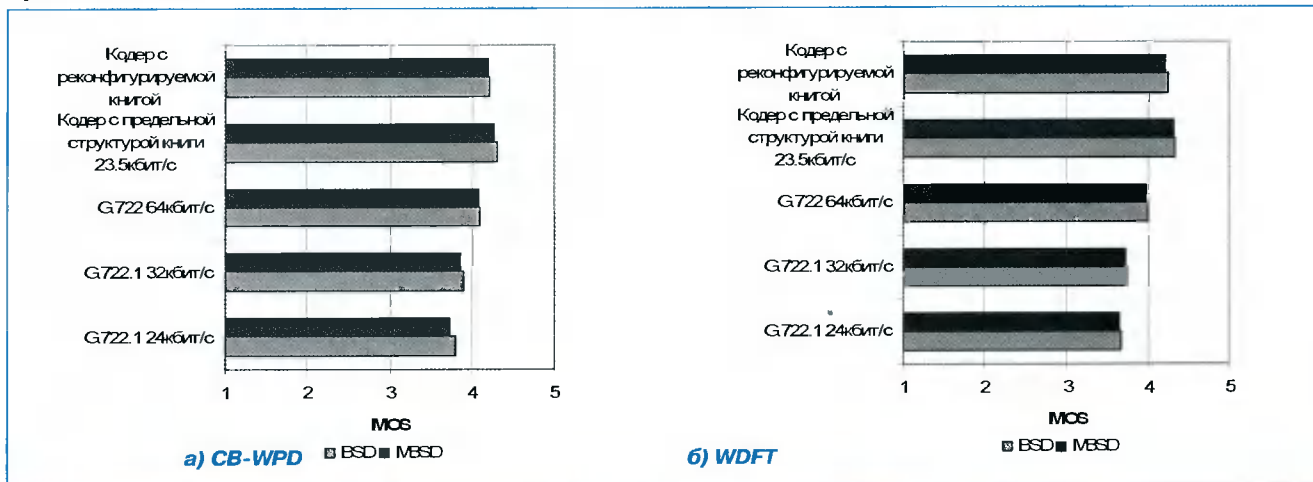
## Заключение

В настоящей работе было показано, что мультиполосный принцип возбуждения в широкополосном CELP-коде совместно с многоуровневым векторным квантованием субполосного сигнала возбуждения позволяет существен-



**Рис. 13. Спектрограммы, полученные при помощи WDFТ: оригинальный сигнал (сверху), реконструированный для кодера с предельной структурой кодовой книги (в центре), восстановленный для кодера с реконфигурируемой кодовой книгой (снизу)**

но улучшить качество реконструированной речи, а также обеспечить возможность эффективного кодирования не только речевых данных, но и аудио сигналов заданного частотного диапазона. При этом в каждой кодируемой субполосе осуществляется своего рода независимый компонентный анализ, а многоуровневое векторное квантование позволяет учитывать ошибку предыдущего уровня и кодировать ее на следующем. Таким образом, осуществляется процесс «обеления» входного кодируемого сигнала, в результате чего увеличивается мера его пологости спектра и, следовательно, устраняется статистическая избыточность. Перцептуально настраиваемая (реконфигурируемая) структура мультиполосной многоуровневой кодовой книги кодера обеспечивает достаточно высокое качество реконструированного речевого сигнала при значительном снижении скорости информационного потока, о чем свидетельствуют показатели, представленные в таблице 5, и сравнительный анализ широкополосных кодеров (*рис. 14*).



**Рис. 14. Сравнительный анализ качества широкополосных кодеров**



## Литература

1. E. Zwicker, H. Fastl, "Psychoacoustics Facts and Models", Springer-Verlag, Berlin Heidelberg, 1990.
2. Christoph Erdmann, Peter Vary, "Embedded Speech Coding Based on Pyramid CELP", IEEE Workshop on Speech Coding, pp.29-31, Tsukuba, Ibaraki, Oct.6-9, Japan, 2002.
3. Alexis Bernard, Abeer Alwan "Perceptually Based and Embedded Wideband CELP Coding of Speech", Proc. of Eurospeech, pp.1543-1546, 1999.
4. М.З. Лившиц, А.А. Петровский, "Многоуровневое векторное квантование речевого сигнала по мультиполосной кодовой книге в широкополосном CELP-коде с психоакустической мотивацией", IV Международная научная конференция "Цифровая обработка сигналов и ее применение", труды РНТОРЭС им. А.С. Попова, выпуск VI-1, стр.119-123, Москва, 2004.
5. M. Parfieniuk, A. Petrovsky, "Warped DFT as the basis for psychoacoustic model," The proc. of the IEEE International conference on Acoustic, Speech, Signal processing, ICASSP, vol. IV, pp.185-188, May 2004, Montreal, Canada.
6. A. Makur, S.K. Mitra, "Warped Discrete-Fourier Transform: Theory and Applications," IEEE Trans. Circuits Systems I, vol.48, pp.1086-1093, Sept. 2001.
7. A. Makur, "Derivation of Subband Coding Gain: The Most General Case", www.ntu.edu.sg/home/eamakur/codinggain.pdf
8. Л. Рабинер, Б. Гоулд, "Теория и применение цифровой обработки сигналов". – М.: Мир, 1978. – 848 с.
9. M. Bosi, "Filter banks in perceptual audio coding" // The Proc. of the AES 17th International Conference "High-Quality Audio Coding", pp.125-136, Florence, Italy. 1999.
10. J. Johnston, "Audio coding with filter banks," in Subband and Wavelet Transforms // A. Akansu and M. J. T. Smith – Eds: Kluwer Academic, 1996. – P. 287-307.
11. B. Bessette, R. Lefebvre, R. Salami et al. "Techniques for High-Quality ACELP coding of Wideband Speech", Eurospeech 2001, pp.1997-2000.
12. A. Petrovsky, M. Parfieniuk, K. Bielawski, "Psychoacoustically Motivated Non-uniform Cosine Modulated Polyphase Filter Bank", 2nd International Workshop on Spectral Methods and Multirate Signal Processing (SMMSP 2002), pp.95-101, Toulouse, France, September 7-8, 2002.
13. R. Laroia, N. Phamdo, N. Farvardin, "Robust and efficient quantization of speech LSP parameters using structured vector quantizers", IEEE ICASSP, vol.5, pp.641-644, 1991.
14. J.D. Johnston, "Transform coding of audio signals using perceptual noise criteria," IEEE J. Selected Areas in Comm., vol.6, pp.314-323, Feb. 1988.
15. T. Thiede, E. Kabet, "A New Perceptual Quality Measure for Bit Rate Reduced Audio," Proc. 100th AES Convention, Copenhagen, Preprint 4280, 1996.
16. J.O. Smith III, J.S. Abel, "Bark and ERB Bilinear Transforms," IEEE Trans. Speech, Audio Processing, vol.7, pp.697-708, June 1999.
17. Ал. А. Петровский, "Объективная оценка качества восстановленного аудио сигнала перцептуальным ПДВП-кодером на базе периферийной модели уха человека", V Международная научная конференция "Цифровая обработка сигналов и её применение", труды РНТОРЭС им. А.С. Попова, т.2, стр.123-126, Москва, Россия, 2003.
18. A.A. Petrovsky, D. Krahe, A.I.A. Petrovsky, "Real-Time Wavelet Packet-based Low Bit Rate Audio Coding on a Dynamic Reconfigurable System", 114th AES Convention, preprint No 5778, 22-25 May, Amsterdam, Netherlands, 2003. -22p
19. K. Brandenburg, T. Sporer, " "NMR" and "Masking Flag": Evaluation of Quality Using Perceptual Criteria", The Proc. of the 11th Int. Conv. Aud. Eng. Soc., "Test and measurement", pp.169-179, May 1992, Portland, USA.
20. W. Yang, M. Dixon, and R. Yantorno, "A modified bark spectral distortion measure which uses noise masking threshold," IEEE Speech Coding Workshop, pp. 55-56, Pocono Manor, 1997.
21. W. Yang, M. Benbouchta, and R. Yantorno, "Performance of a modified bark spectral distortion measure as an objective speech quality measure," IEEE ICASSP, pp.541-544, Seattle, 1998.
22. W. Yang and R. Yantorno, "Comparison of two objective speech quality measures: MBSD and ITU-T recommendation P861," IEEE MMSP, pp.426-431, Redondo Beach, 1998.
23. W. Yang and R. Yantorno, "Improvement of MBSD by scaling noise masking threshold and correlation analysis with MOS difference instead of MOS," IEEE ICASSP, pp. 673 -676, Phoenix, 1999.
24. ISO/IEC JTC1/SC29/WG11 NO803, MPEG, International Standard IS 13818-3 Information Technology – Generic Coding of Moving Pictures and Associated Audio: Audio, 11th November 1994.
25. М.З. Лившиц, М. Парфенюк, А.А. Петровский, "Многоуровневое векторное квантование с переменной глубиной поиска в перцептуальных кодерах речи с психоакустической моделью на основе сжатого ДПФ" // VII Международная конференция «Цифровая обработка сигналов и ее применение», труды РНТОРЭС им. А.С. Попова, выпуск VII-1, стр.187-191, Москва, 2005.
26. ITU-T Recommendation G.722 "7-kHz audio - coding within 64 kbit/s", 1993.

## Задержка обработки информации в многопроцессорных вычислителях различной конфигурации

**О.Ю. Аксенов**

Создание многопроцессорных вычислительных устройств, предназначенных для обработки непрерывных потоков информации, связано с определением информационных связей между процессорами и банками памяти, а также с выбором типа используемой памяти. Особенно это актуально для встраиваемых устройств, когда коммуникационные возможности памяти различных типов влияют не только на вычислительные возможности всего устройства, но и на его сложность, стоимость и, главное, габариты. Ниже анализируется влияние конфигурации многопроцессорного вычислительного устройства на задержку обработки информации.

Рассмотрение ведется для следующих условий.

1. Задача, решаемая при обработке потока информации, допускает распараллеливание. Применительно к задачам обработки потока изображений это означает, что площадь каждого изображения при обработке может быть разбита на  $N$  равных независимых участков, либо процедура обработки может быть разделена на  $N$  параллельных ветвей.
2. Применяемую память по характеру взаимодействия с источником и приемником информации (например, с процессором), можно отнести к одному из следующих типов:
  - память с независимым доступом (память типа  $D$ ) – обеспечивающая одновременную работу с двумя процессорами (двухпортовая);
  - разделяемая память (память типа  $S$ ) – работающая с двумя процессорами в режиме разделения времени;
  - локальная память (память типа  $L$ ) – работающая только с одним процессором.
3. На входе и на выходе рассматриваемого устройства установлена память типа  $D$  ( $D_{in}$  и  $D_{out}$  соответственно).
4. Входная информация (например, отдельный кадр видеопотока) приходит в многопроцессорное вычислительное устройство равномерно на некотором временном интервале.

*Рассматриваются встраиваемые многопроцессорные вычислители, предназначенные для осуществления в реальном масштабе времени обработки потоков информации, формируемых, например, видеосистемами. Анализируется связь конфигурации вычислителей с задержкой обработки информации. Определяются подходы к выбору конфигурации вычислителя для различных условий обработки.*

5. Многопроцессорное вычислительное устройство должно обеспечивать:

- обработку информации с темпом поступления;
- минимальную задержку обработки.

Предполагается, что требуемый темп обработки информации обеспечивается за счет одновременного использования нескольких однотипных процессоров.

Для оперативной обработки потока информации, например потока кадров, формируемого видеосистемами, можно использовать различные конфигурации устройства обработки. Ниже рассматриваются некоторые из них, обозначенные как схемы обработки информации "А", "Б", "В", "Г", "Д". Символом  $B$  на схемах обозначены процессоры, символами  $D, S, L$  – банки памяти соответствующих типов.

Схема "А" (рис. 1) соответствует однопроцессорному вычислителю. Эта схема проигрывает по производительности всем остальным, рассмотренным ниже, имеющим в своем составе  $N$  процессоров. Схема "Б" (рис. 2) представляет собой конвейер с использованием между процессорами памяти типа  $D$ . Схема "В" (рис. 3) – конвейер с использованием памяти типа  $S$ .



Рис. 1. Схема "А" обработки информации



Рис. 2. Схема "Б" обработки информации



Рис. 3. Схема "В" обработки информации

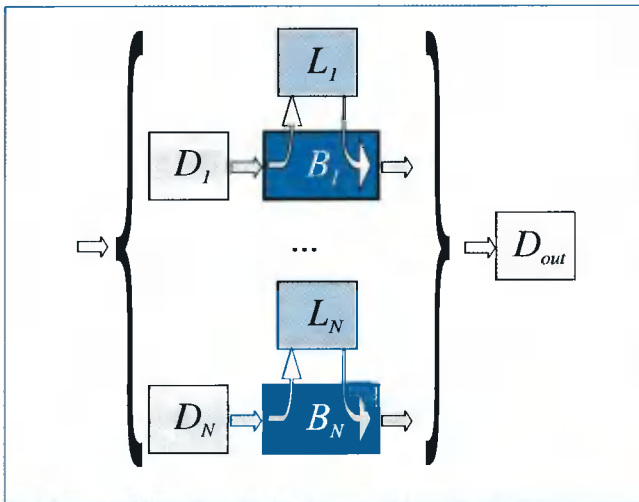


Рис. 4. Схема "Г" обработки информации

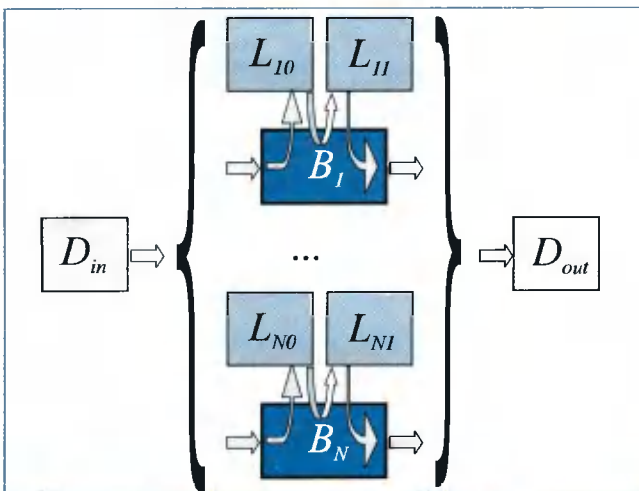


Рис. 5. Схема "Д" обработки информации

Рис. 4 представляет схему "Г" – параллельное соединение процессоров при обеспечении одновременной зачки входной информации во входные банки памяти каждого процессора ( $D_1...D_N$ ). Такая схема целесообразна, если одну и ту же информацию (например, изображение) требуется обработать по различным алгоритмам. Для каждого ( $n$ -ого) процессора этой схемы характерны два этапа работы. На первом производятся собственно вычисления. Исходная информация берется из банка памяти  $D_n$ , а результат записывается в банк  $L_n$ . Второй этап – копирование из банка  $L_n$  в выходной банк  $D_{out}$ . Этот этап осуществляется в режиме программного разделения между процессорами времени доступа к памяти.

Схему "Д" отображает рис. 5. Эта схема предполагает, что каждый ( $n$ -ый) процессор обрабатывает только  $1/N$  часть общей входной информации (кадра). При этом на первом этапе осуществляется копирование этой части информации в локальный банк памяти  $L_{n0}$ . На втором – собственно вычисления, при которых исходные данные берутся из банка памяти  $L_{n0}$ , а результат записывается в банк  $L_{n1}$ . На третьем этапе производится копирование из локального банка памяти  $L_{n1}$  в выходной банк  $D_{out}$ . Первый и третий этап тут осуществляются с программным разделением между процессорами времени доступа к памяти  $D_{in}$  и  $D_{out}$  соответственно.

Временные диаграммы работы рассмотренных схем обработки информации представлены на рис. 6–10. Используются следующие обозначения:

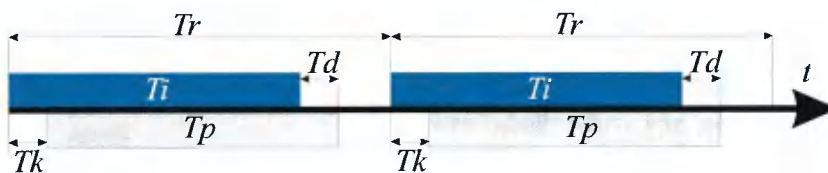


Рис. 6. Временная диаграмма обработки информации по схеме "А"

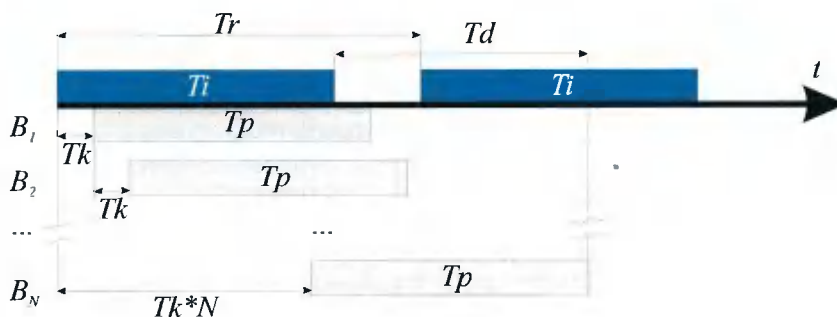


Рис. 7. Временная диаграмма обработки информации по схеме "Б"



$T_r$  – период прихода входной информации;  
 $T_i$  – время прихода входной информации;  
 $T_p$  – время на обработку кадра одним процессором  
 $T_s = N \cdot T_p$  – время на обработку кадра процессорами;  
 $T_c$  – время копирования кадра (переноса из одного банка памяти в другой);  
 $T_k$  – время прихода минимальной достаточной для начала обработки порции информации ( $T_k < T_i$ );  
 $T_d$  – общая задержка обработки.

Анализ временных диаграмм работы позволяет оценить для рассматриваемых схем допустимый период поступления информации и задержку ее обработки. При этом число процессоров –  $N$  определяется исходя из того, чтобы обеспечить требуемый темп обработки:  $N = T_s / T_p \Big|_{T_p = T_r} = T_s / T_r$ . Соответствующие выражения содержит таб. 1.

Удобно использовать относительные значения времени, приняв за единицу измерения время обработки кадра одним процессором –  $T_p$ . Кроме того, исходя из условий обеспечения обработки непрерывного потока информации, целесообразно принять  $T_i \approx T_p$ . Тогда можно ввести коэффициенты, характеризующие:

$s = T_s / T_p$  – сложность задачи ( $s \geq 1$ );

$k = T_k / T_p$  – скорость прихода минимальной порции данных ( $k \ll 1$ );

$c = T_c / T_p$  – скорость копирования ( $c < 1$ );

$i = T_i / T_p$  – скорость прихода данных ( $i \approx 1$ ).

Использование этих коэффициентов позволяет осуществить переход к относительным значениям допустимого периода поступления информации ( $\tau = T_r / T_p$ ) и задержки обработки информации ( $t = T_d / T_p$ ). В этом случае соотношения, входящие в таб. 1, примут вид, который отображает таб. 2.

Зависимость относительной задержки, обеспечиваемой различными схемами, для  $s=4$ ;  $c=0$ ;  $i=0$ , 9 от  $k$  при  $N$  иллюстрирует рис. 11, а при  $N=2$  и  $N=4$  – рис. 12. Из представленных рисунков следует, например, что при  $N=4$  и  $k > 0,3$ , наименьшую задержку обеспечивает схема “Д”, тогда как при меньших значениях  $k$  более выгодной становится схема “Б”. Ана-

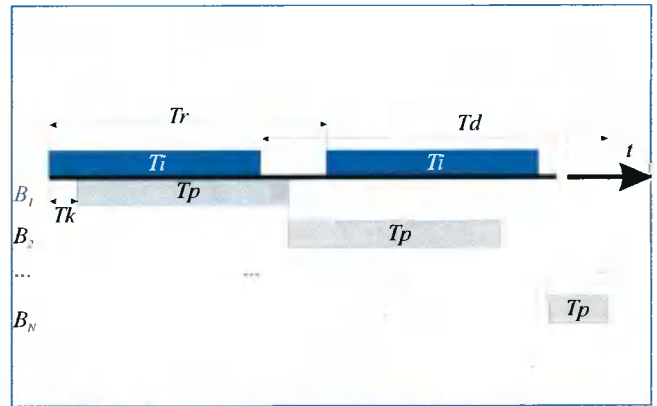


Рис. 8. Временная диаграмма обработки информации по схеме “Б”

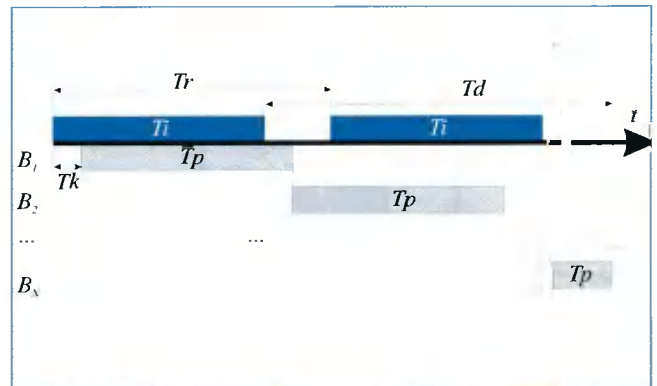


Рис. 9. Временная диаграмма обработки информации по схеме “Г”

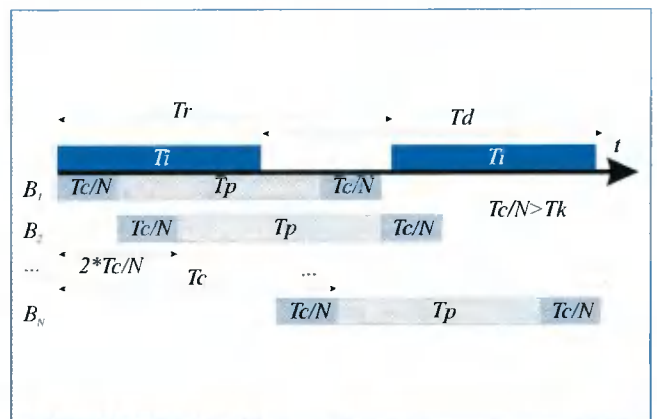


Рис. 10. Временная диаграмма обработки информации по схеме “Д”

Таблица 1

Схема	Период поступления	Задержка обработки	Сложность задачи
“А”	$T_r \geq T_k + T_p$	$T_d = T_k + T_p - T_i$	$T_s = T_p$
“Б”	$T_r \geq T_k + T_p$	$T_d = T_k \cdot N + T_p - T_i$	$T_s = N \cdot T_p$
“В”	$T_r \geq T_k + T_p$	$T_d = T_k + N \cdot T_p - T_i$	$T_s = N \cdot T_p$
“Г”	$T_r \geq T_k + T_p$	$T_d = T_k + T_p + T_c - T_i$	$T_s = N \cdot T_p$
“Д”	$T_r \geq T_p + 2 \cdot T_c / N$	$T_d = T_p + T_c \cdot (1 + 1/N) - T_i$	$T_s = N \cdot T_p$

Таблица 2

Схема	Относительный период поступления ( $\tau = T_r / T_p$ )	Относительная задержка обработки ( $t = T_d / T_p$ )
“А”	$\tau_A \geq k + s$	$t_A = k + s - i$
“Б”	$\tau_B \geq k + s/N$	$t_B = k \cdot N + s/N - i$
“В”	$\tau_B \geq k + s/N$	$t_B = k + s - i$
“Г”	$\tau_G \geq k + s/N$	$t_G = k + s/N + c - i$
“Д”	$\tau_D \geq (s + 2 \cdot c) / N$	$t_D = s/N + c \cdot (1 + 1/N) - i$



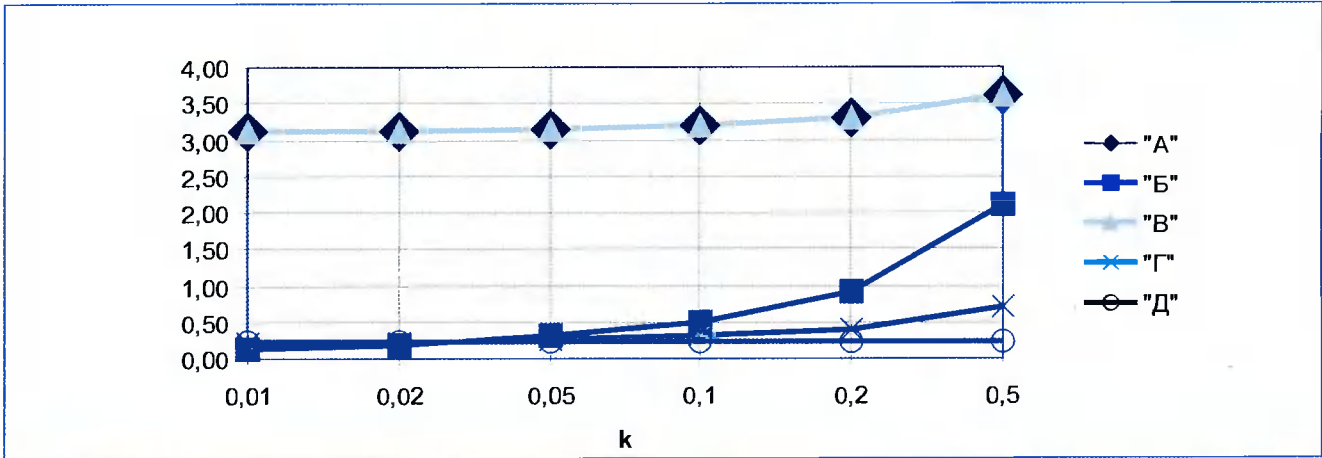


Рис. 11 Относительная задержка обработки для  $s=4$ ;  $c=0,1$ ;  $i=0,9$ ;  $N=4$ .

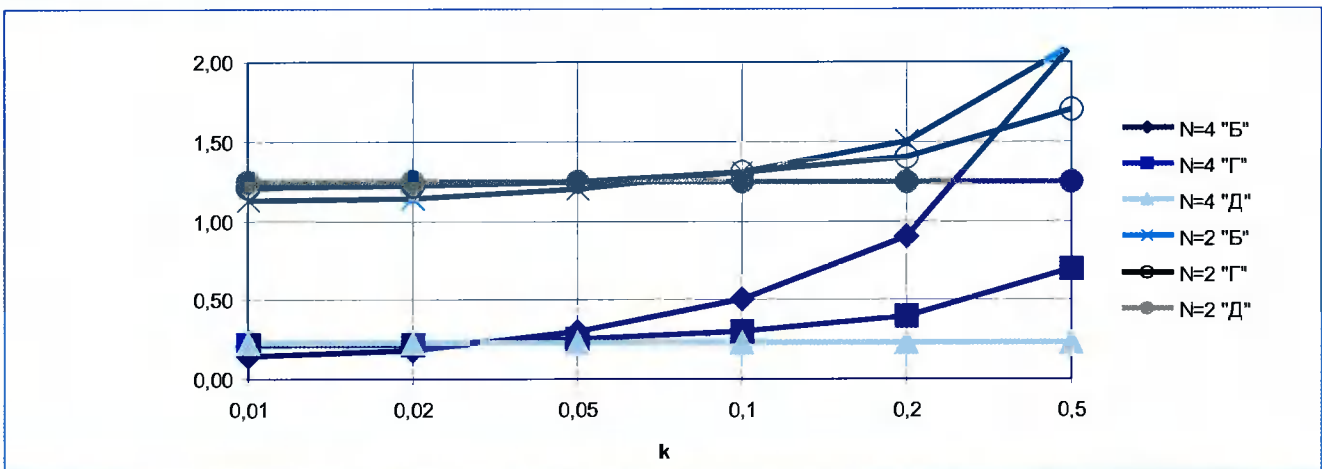


Рис. 12. Относительная задержка обработки для  $s=4$ ;  $c=0,1$ ;  $i=0,9$ .

логично для  $N=2$  пограничным значением оказывается  $k \approx 0,7$ .

Сравнение выражений, представленных в таб. 2, позволяет определить значения параметров, при которых та или иная схема обеспечивает меньшую задержку или способна функционировать при меньшем значении периода поступления информации. Эти значения соответствуют точкам пересечения графиков. Так  $\tau_B = \tau_G = \tau_D > \tau_A$  имеет место при  $k\tau > 2c/N$ . Аналогично:

- $t_B > t_G$  при  $k_{BG} > c/(N-1)$ ;
- $t_B > t_D$  при  $k_{BD} > c \cdot (N+1)/N^2$ ;
- $t_A > t_B$  при  $k_{AB} > s/N$ ;
- $t_G > t_D$  при  $k_{GD} > c/N$ .

В приведенные соотношения входит величина  $k$ , характеризующая минимальную порцию исходной информации, получив которую, можно начинать обработку. Эта величина определяется источником информации и способом обработки. Так, например, пусть требуется обработать изображение размером  $1024 \times 768$  пикселей фильтром с апертурой  $9 \times 9$  пиксе-

лов. Исходная информация поступает построчно. Значит, для начала обработки надо получить 8 строк плюс 9 первых элементов 9-ой строки. Отсюда  $Tk/Ti = (1024 \cdot 8 + 9)/(1024 \cdot 768) \approx 0,01$ . При  $i \approx 1$  это означает, что  $k \approx 0,01$ . Тогда для схем "Б", "В", "Г", "Д" и случая  $s=4$ ;  $c=0,1$ ;  $N=4$  наименьшую задержку обеспечит схема "Б". Однако, если исходная информация поступает полукадрами (сначала только четные строки, потом только нечетные), то  $Tk/Ti = (1024 \cdot 4 + 9 + 1024 \cdot 768/2)/(1024 \cdot 768) \approx 0,5$ , и при  $k \approx 0,5$  выгоднее использовать схему "Д".

Следует отметить, также, что существенным моментом, определяющим на практике выбор схемы обработки, является удобство («стоимость») аппаратной реализации памяти рассмотренных типов.

Таким образом, предложенный подход и полученные соотношения позволяют в конкретных случаях выбрать наилучшую (по критериям «наибольший темп поступления данных» или «наименьшая задержка обработки») схему обработки информации и соответственно конфигурацию многопроцессорного вычислителя.

## Сжатие изображений на базе вейвлет-преобразования и иерархического алгоритма кодирования

**А.В. Черников, М.К. Чобану**

**Б**азовое сокращение избыточности изображений можно представить в виде трех основных этапов – декоррелирующее преобразование, процедура квантования и энтропийное кодирование.

До появления кодеров на базе вейвлет-преобразования в качестве декоррелятора использовалось дискретное косинусное преобразование (DCT) [1]. Метод JPEG, например, использует DCT отдельно для каждого блока изображений размером 8x8. Кодер JPEG квантует полученные коэффициенты в соответствии с их важностью и заданной степени сжатия.

В последнее время в кодерах изображений применяют вейвлет-декомпозицию сигнала [2,3], которая предусмотрена, в частности, стандартом JPEG2000. Этим стандартом предусматриваются предварительная обработка, выполнение вейвлет-декомпозиции (на основе лифтинговой схемы), сжатие без потерь с помощью целочисленного преобразования, квантование, энтропийное кодирование, упаковка выходного потока в пакеты, схемы сжатия с потерями и без потерь и т.д. “Сердцем” JPEG2000 является алгоритм сжатия EBCOD [8,9]. Применяется также ряд не стандартизованных, но зачастую не менее эффективных алгоритмов. Одним из них является SPIHT. На его базе развиваются новые алгоритмы, и в оценках эффективности любого нового вейвлет-кодера обязательно можно найти сравнительные с ним тесты.

Все кодеры, работающие с вейвлет-декомпозицией, принято делить на два класса: межполосные (inter-band) и внутриполосные (intra-band). Алгоритм SPIHT относится к классу межполосных кодеров, т.е. он использует избыточность, связанную с корреляцией между уровнями декомпозиции. Рассмотрению этого класса кодеров и посвящена настоящая статья.

*Рассмотрены вопросы реализации методов сжатия изображений с помощью двумерного вейвлет-преобразования. Кодированная часть основана на переработанном и оптимизированном варианте метода пространственно ориентированных иерархических деревьев (SPIHT) и позволяет кодировать изображения в формате BMP (цветные и в градации серого). Степень сжатия устанавливается произвольно, а размер выходного файла можно задать с точностью до байта, что является несомненным преимуществом кодера. Проект реализован научной группой кафедры Электрофизики Московского энергетического института.*

### Вейвлет-декомпозиция

В результате выполнения одного уровня разделимого вейвлет-преобразования выделяются четыре подполосы: *LL, LH, HL, HH* (рис. 1). Где *LL* – выделенные низкая частота по горизонтали и низкая по вертикали, *LH* – низкая по горизонтали и высокая по вертикали и так далее соответственно. Аналогичным образом выполняется вейвлет-преобразование *LL* части изображения. Итерация повторяется необходимое число раз для получения заданного числа уровней разложения. На рис. 2 приведен пример вейвлет-декомпозиции изображения.

Для рассматриваемых алгоритмов сжатия большое значение имеют банк фильтров и количество уровней декомпозиции. При выборе фильтров следует исходить из конкретной задачи. Фильтры с хорошей гладкостью могут дать лучший результат по качеству сжатия, но вследствие своего большого размера увеличат время декомпозиции. Фильтры меньшего размера или с простым вычислением, например по схеме лифтинга, могут значительно сократить время вычисления, уменьшив при этом качество сжатия.

### Приоритет

В упрощенной форме алгоритм SPIHT можно рассматривать как последовательную передачу битовых полей разложения, начиная со старших битов. Это один из аспектов расстановки приоритетов: более важными являются старшие битовые поля.

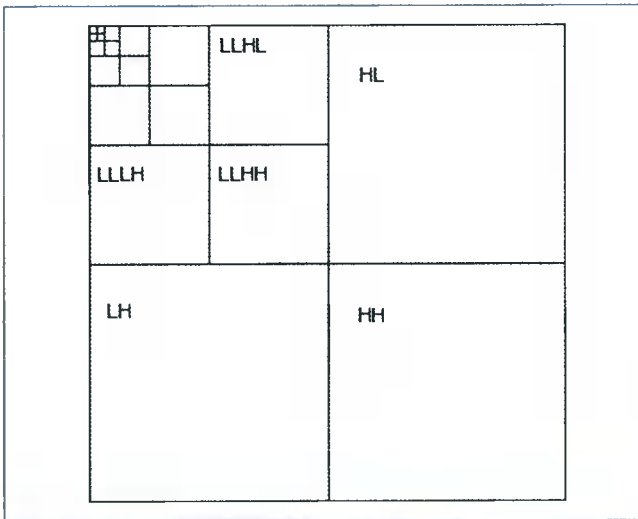


Рис. 1. Выделение подполос.

В общем случае отсчеты всех субполос разложения распределены по одному набору битовых полей, визуально это выражается примерно равной яркостью субполос на всех уровнях (см. рис. 2).

Теперь, если представить себе работу алгоритма, понятно, что в передаче старшего бита, будут на равных правах участвовать субполосы всех уровней. При восстановлении же для большинства сигналов более важными являются отсчеты низкочастотных составляющих (или старших уровней). Поэтому передавать их необходимо раньше и точнее. Простой способ установки межуровневого приоритета – нормализация фильтров. Если нормализовать фильтры банка анализа на число  $C$  большее единицы, то формально каждый уровень разложения будет умножен на число  $C^N$ , где  $N$  – номер уровня. Следовательно, старшие уровни получают больший приоритет. Нормализация же фильтров



Рис. 2: Вейвлет-декомпозиция (изображение - Lena256Gray.bmp, фильтр – 2bior2.2, 2 уроная).

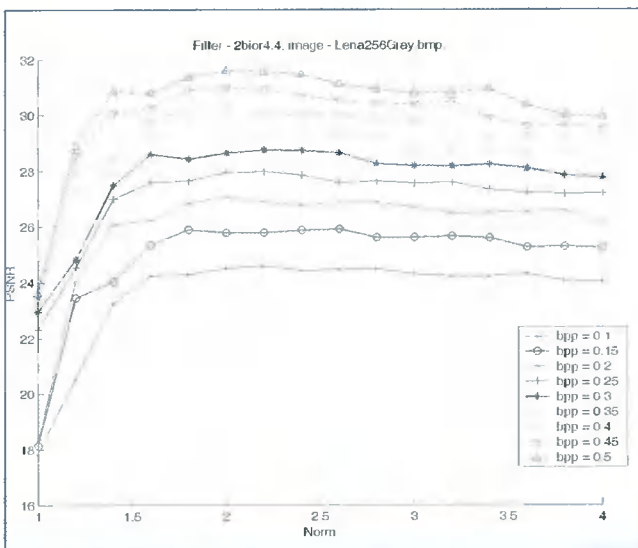


Рис. 3. Тест нормализации для фильтра 2bior4.4. Каантование

банка синтеза на число обратное  $C$  позволит сохранить свойство точного восстановления. В процессе разложения с нормализованными фильтрами отсчеты от уровня к уровню занимают все более высокие битовые поля, целая часть их значений содержит все больше знаков после запятой. В результате чего в передаче старших битовых полей участвуют высшие уровни разложения. Далее постепенно (и последовательно) в передачу битовых полей включаются младшие уровни (это второй аспект расстановки приоритетов).

Коэффициент  $C$  на практике легко найти протестировав несколько изображений в поисках максимального соотношения  $PSNR$ , задав пределы нормализации, например, от 1 до 4.  $PSNR = 20 \lg(255/MSE)$  – используемый в работе критерий качества, где  $MSE$  – средний квадрат разницы исходного и тестируемого сигналов;  $bpr$  – число бит в потоке условно приходящееся на пиксель, характеризует степень сжатия. Тест желательно сделать для каждого фильтра.

На рис. 3 приведен пример такого теста для фильтра 2bior4.4. Использовано тестовое изображение Lena размером 256\_256 в градации серого. График построен для набора  $bpr$ . Из этого теста можно сделать вывод о нормализации данного фильтра – 2bior4.4 достигает максимума  $PSNR$  на всех  $bpr$  при нормализации лежащей в



окрестности 2. Коэффициент 2 характерен для большинства исследованных фильтров и может быть взят за правило (за некоторыми исключениями). Исключения эти носят довольно специфичный характер и не нарушают общности для большинства изображений и фильтров.

## Квантование

Ввиду того, что в процессе разложения с нормализованными фильтрами отсчеты от уровня к уровню занимают все более высокие битовые поля, целая часть их значений содержит все больше знаков после запятой. Таким образом, нормализация исполняет роль нелинейного квантования с более точной передачей старших уровней. И после такой фильтрации для получения целых чисел (SPIHT работает с целыми числами) достаточно просто отбросить дробную часть. В случае же необходимости можно умножить всё разложение на некоторое число (обычно это степень двойки), выводя разряды после запятой в целую часть.

## Используемые свойства вейвлет-декомпозиции

Одним из важных свойств вейвлет-преобразования является компактность энергии. Основная энергия концентрируется в субполосе  $LL$ . Эта компактность энергии обеспечивает эффективное применение скалярных квантователей. Для квантования областей с низкой энергией отводится меньшее количество бит, что реализует сжатие. Однако такой подход не учитывает остаточную структуру, сохраняющуюся в вейвлет-коэффициентах, в особенности высокочастотных субполос. В этих субполосах имеются обычно большие области с нулевой или малой энергией. Области с высокой энергией повторяют от субполосы к субполосе свои очертания и местоположение. Медленно изменяющиеся, гладкие области исходного изображения хорошо описывают низкочастотные вейвлет-базисы, что приводит к «упаковке» энергии в малом числе коэффициентов области  $LL$ . Этот процесс примерно повторяется на всех уровнях декомпозиции, что и приводит к визуальной «похожести» различных субполос. Априорное знание о том, что изображение состоит из гладких областей, текстур и контуров, помогает учитывать эту межполосную структуру.

## Алгоритм А. Льюиса и Г. Ноулеса

Данный алгоритм базируется на построении пространственных деревьев над полем вейвлет-коэффициентов [4]. Дерево выстраивается из выбранной

точки – корня, который может быть расположен в любой субполосе. Корень дает четыре потомка в том же субдиапазоне на предыдущем уровне декомпозиции.

Площадь субдиапазонов от уровня к уровню уменьшается в 4 раза. И, следовательно, точке  $(i,j)$  на выбранном уровне ставится в соответствие 4 точки на предыдущем уровне в данном субдиапазоне, а именно  $(i*2,j*2)$ ,  $(i*2,j*2+1)$ ,  $(i*2+1,j*2)$ ,  $(i*2+1,j*2+1)$  (начало координат в левом верхнем углу в точке  $(0,0)$ ). Полное дерево  $D(i,j)$  есть множество, состоящее из корня и ветвей – деревьев выстроенных из потомков корня:

$$D(i,j) = (i,j) + D(i*2,j*2) + D(i*2,j*2+1) + D(i*2+1,j*2) + D(i*2+1,j*2+1).$$

Дерево, выстроенное из точки первого уровня декомпозиции, имеет только корень и не имеет потомков. На *рис. 4* представлен пример построения таких деревьев.

Алгоритм «квантования нулевым деревом» основан на наблюдении, что если коэффициент корня мал, его потомки на дереве зачастую тоже малы. А.Льюис и Г.Ноулес свели это предсказание к минимуму, предположив, что если какой-либо коэффициент незначимый, то все его потомки также будут незначимыми. Авторы [4] использовали следующий алгоритм квантования вейвлет-коэффициентов. Вначале каждый корневой узел квантуется квантователем, оптимальным для плотности распределения Лапласа. Если значение узла меньше некоторого порога, его потомки игнорируются. Иначе осуществляется переход к четырем отпрыскам узла, и процедура повторяется. Если узел не имеет отпрысков, начинает обрабатываться следующий корневой узел и т.д.

Данный алгоритм является эффективным в силу хорошей «упаковки» энергии вейвлет-преобразованием, а также за счет совместного кодирования нулей. Для кодирования нулей обычно применяется кодер

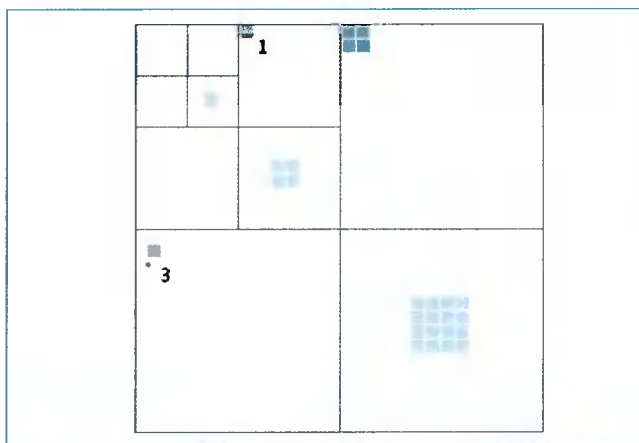


Рис. 4. Построение пространственных деревьев.



длин серий, на вход которого коэффициенты должны подаваться в определенном порядке (например, зигзагообразное сканирование). Наверное, наиболее важным вкладом этой работы была демонстрация того, что область вейвлет-коэффициентов прекрасно приспособлена для работы кодера длин серий.

Характеристики алгоритма Льюиса и Ноулеса незначительно превосходят JPEG, хотя визуальное качество изображений лучше. Недостатком алгоритма является способ порождения и распознавания нулевых деревьев. В случае, если коэффициент корня мал и обнуляются значимые коэффициенты, алгоритм Льюиса и Ноулеса ведет к большим искажениям. Идеи данного алгоритма легли в основу многих современных кодеров изображения.

### Алгоритм вложенного нулевого дерева (EZW)

Ж. Шапиро [5] разработал элегантный метод, названный алгоритмом вложенного нулевого дерева (Embedded Zerotree Wavelet coder - EZW). Данный кодер учитывает возможность значимости потомков незначимого узла. Перед кодированием выполняется квантование, для чего выбирается начальный порог (*initial threshold*)  $T_0$ :  $|a_{ij}| < 2T_0$ , для любого вейвлет-коэффициента  $a_{ij}$ . Остальные уровни квантования (*threshold levels*) находятся как  $T_i = T_{i-1}/2$ . В процессе кодирования могут формироваться четыре символа:

- 1) корень нулевого дерева (*zerotree root*);
- 2) изолированный ноль (*isolated zero*);
- 3) положительно значимый (*positive significant*);
- 4) отрицательно значимый (*negative significant*).

Вероятности появления этих символов различны, и существует некоторая избыточность, позволяющая эффективно использовать энтропийное кодирование. Если при анализе текущего вейвлет-коэффициента (для порога  $T_i$ ) кодер сформировал символ «корень нулевого дерева», то декодеру известно, что на данном этапе все дерево, выстроенное из этой точки, является незначимым, и анализ потомков проводиться не будет. Если же сформирован символ «изолированный ноль», значит текущий вейвлет-коэффициент, рассматриваемый как корень, является незначимым, но среди его потомков имеются значимые коэффициенты. Тогда декодер осуществляет анализ потомков. Точки рассматриваются в определенном порядке, поэтому дополнительной информации (координат) не требуется. Символы «положительно значимый» и «отрицательно значимый» говорят о значимости коэффициента на данном этапе.

А.Саид и В.Пирлман [7] улучшили алгоритм EZW. Их версия кодера называется «Пространственно-упо-

рядоченные иерархические деревья» (Set Partition Hierarchical Trees - SPIHT). Алгоритм получился простым и быстрым, превосходящим при этом EZW по качеству. SPIHT-кодер лег в основу данной работы.

### Побитовая передача

Начнём рассмотрение алгоритма SPIHT с разбора принципа побитовой передачи:

- старшие биты несут больше информации, значительно уменьшая неопределенность;
- точек участвующих в  $N$  битовых полях становится меньше при увеличении  $N$ . Другими словами, количество больших по модулю чисел должно быть минимальным.

Учитывая эти утверждения, построим следующий алгоритм:

Известно, что количество больших по модулю чисел минимально. Поэтому, вместо того, чтобы передавать старший бит  $n_{max}$  всего разложения, укажем точки значимые в этом бите координатами  $(i, j)$  (предварительно передав их количество). Таким образом, значительно сократим объем передачи. Декодер, получив номер старшего бита и координаты значимых точек, взведёт необходимый разряд. И в качестве предположения выставит математическое ожидание значения отсчета, рассматриваемого как случайную величину с равномерным распределением по интервалу возможных значений. Например, если отсчет значим в 7-ом бите, диапазон значений его модуля в двоичной системе 10000000 – 11111111, тогда математическое ожидание будет 10111111.

Выигрыш, естественно, будет только тогда, когда передача координат меньше по объему, нежели передача старшего бита всего изображения. Об остальных точках на этот момент известно, что значение бита номер  $n_{max}$  нулевое. Теперь координаты переданных точек необходимо запомнить, для этого формируется список значимых точек (одинаковый на передающей и приёмной сторонах). Эти точки из дальнейшего анализа значимости исключаются.

Переходим к следующему разряду  $(n_{max} - 1)$ . Ищем точки значимые в этом разряде (не забывая, исключить из анализа точки значимые в  $n_{max}$  разряде). Находим таким образом точки, у которых первая единица появляется в  $(n_{max} - 1)$ -ом разряде (считая со старших разрядов), передаем их количество, затем координаты. Теперь важный момент – уточняем точки из списка значимых, то есть передаем их  $(n_{max} - 1)$ -ый бит. И уже не важно 0 он или 1, в любом случае он несет полезную информацию, снимающую равную неопределённость. Добавляем значимые на данном этапе точ-

ки в список, исключаем из анализа незначимые. И так далее для всех разрядов. Такой метод можно рассматривать, как надстройку над простой последовательной передачей битовых полей всего разложения, только теперь точки разложения постепенно включаются в работу, и когда все они попадут в список значимости, алгоритмы совпадут полностью. Но в этой постепенности и заключается выигрыш. Не приходится передавать нули старших разрядов не значимых точек, они передаются косвенно – указывая значимые точки. Учитываем таким образом некоторую избыточность разложения и априорную информацию о том, что точки малые по модулю встречаются чаще.

### Формализованный алгоритм кодера:

1. Передача  $n_{max}$  – количество разрядов (считая от нуля) занимаемых максимальным отсчетом разложения, то есть  $n_{max} = \log_2(\max_{ij}(|a_{ij}|))$  (здесь и далее с округлением вниз);
2. Инициализация списка значимых точек – пустое множество;
3. Для  $k$  от  $n_{max}$  до 0 цикл
  - поиск по всему изображению точек, для которых  $\log_2 |a_{ij}| = k$ ;
  - передача количества найденных точек  $N_z$ ;
  - передача координат  $(i,j)$  и знака ( $sign(a_{ij})$ ) найденных точек;
  - если  $k < n_{max}$ , то передача  $k$ -го разряда всех элементов из списка значимых точек;
  - добавление переданных координат  $(i,j)$  в список значимых точек.

Конец цикла.

### Алгоритм декодера:

1. Приём  $n_{max}$ ;
2. Инициализация списка значимых точек – пустое множество;
3. Для  $k$  от  $n_{max}$  до 0 цикл
  - прием количества точек  $N_z$ ;
  - приём  $N_z$  посылок типа [координата  $(i,j)$ , знак ( $sign(a_{ij})$ )], взведение  $k$ -го разряда для каждой точки  $(i,j)$  и установка знака;
  - если  $k < n_{max}$ , то прием и установка  $k$ -го разряда для всех элементов из списка значимых точек;

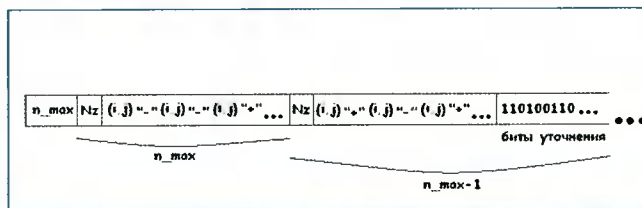


Рис. 5. Структура потока.

- добавление переданных координат  $(i,j)$  в список значимых точек.

Конец цикла.

Еще раз обратим внимание на то, что списки значимых точек формируются одинаково и уточнение производится синхронным перебором списков кодера и декодера. При этом нет необходимости помещать значимые точки для исключения их из анализа. Исключение делает сама формула  $\log_2 |a_{ij}| = k$  (с округлением вниз). Для точек ставших значимыми на предыдущих этапах  $\log_2 |a_{ij}| > k$ .

Важно отметить некоторые свойства выходного потока. Важнейшее из них – наиболее ценная информация передаётся в первую очередь. Если, к примеру, имеется какой-то фиксированный размер памяти, отведённый под изображение, можно не раздумывая заполнять его последовательно потоком кодера, то есть нет проблемы выбора важной информации – она группируется к началу потока. Другое важное свойство – возможность прервать поток в любом месте. Нет блоков, которые было бы необходимо передавать целиком, закончить можно в любой момент. Это важное свойство обеспечивает такую удобную характеристику кодера как произвольная степень сжатия. Кодер просто заполняет отведенный ему объем памяти. На рис. 5 схематично изображен выходной поток.

Этот не сложный принцип побитовой передачи с некоторой модификацией используется в SPIHT. Устраняется основной недостаток вышеописанного алгоритма. Обратим внимание на то, как указываются декодеру значимые точки – передачей их координат. На практике сразу же выявляется минус такой организации. Предположим, что размеры исходного изображения 512x512. Тогда, чтобы выбрать точку, нужно 18 бит ( $2^9=512$ ). А для того, чтобы передать битовое поле требуется 262 144 бит ( $512*512$ ). Теперь представим себе, передачу старшего разряда. Предположим, точек значимых в старшем разряде 2000 штук. На простую передачу битового поля уходят все 262 144 бита, применяя вышеописанный алгоритм – 36 000 бит ( $18*2000$ ). В этом случае алгоритм дает значительное преимущество. Но 2000 – это всего лишь около 0.7% от общего количества точек. Значимость столь малого объема возможна только в старших битах. Если оценить критичный объем значимых точек, то получается, что если на каком-то этапе нашлось хотя бы 6% значимых точек, то на передачу координат нужно потратить больше бит, нежели на передачу всего битового поля. И это ещё без учёта того, что необходимо будет уточнить текущий бит точек из списка значимых. На практике же найдется большее количество значимых точек, причём на ранних этапах кодирования.

Как итог можно сделать вывод, что вышеописанный алгоритм в чистом виде не эффективен. Однако сам подход может быть использован.

### Алгоритм пространственно ориентированных иерархических деревьев (SPIHT)

Алгоритм *SPIHT* является развитием алгоритма *EZW*. Он использует в своей работе аналогичные древовидные структуры, внося в процесс их построения динамизм. Применяя вышеизложенные принципы прогрессивной побитовой передачи, алгоритм формирует вложенный битовый поток.

Введем некоторые обозначения:

- 1)  $Q(i,j)$  – множество координат детей для точки  $(i,j)$ , рассмотренной как корень;
- 2)  $D(i,j)$  – множество координат всех потомков для точки  $(i,j)$ ;
- 3)  $H$  – множество координат корней;
- 4)  $L(i,j) = D(i,j) - Q(i,j)$ .

Термин «дети» здесь и далее обозначает четырех потомков на предыдущем уровне декомпозиции, то есть  $Q(i,j) = \{(i*2, j*2), (i*2, j*2+1), (i*2+1, j*2), (i*2+1, j*2+1)\}$ .

$L(i,j)$  состоит из потомков за исключением детей, то есть из «внуков», «правнуков» и так далее. Также следует заметить, что  $D(i,j)$  и  $L(i,j)$  не содержат корня  $(i,j)$ . На начальном этапе формируется множество корней деревьев максимального размера –  $H$ . Их координаты находятся в  $LL$  части самого старшего уровня декомпозиции. В данной работе использовался второй тип сведения деревьев. Далее формируются списковые субструктуры  $LIS$ ,  $LIP$ ,  $LSP$ . Эти структуры являются нововведением алгоритма *SPIHT* и строятся над полем вейвлет-коэффициентов и выстроенных из них деревьев.

Дадим определение этим структурам:

- 1)  $LIS$  – список незначимых множеств (list of insignificant sets);
- 2)  $LIP$  – список незначимых точек (list of insignificant pixels);
- 3)  $LSP$  – список значимых точек (list of significant pixels).

$LIP$  и  $LSP$  содержат координаты  $(i,j)$  и описывают точки, находящиеся непосредственно по этим координатам.  $LIS$  также содержит координаты  $(i,j)$ , но описывает множества, выстроенные из этих точек. Множества могут быть  $D(i,j)$  (тип А) и  $L(i,j)$  (тип В). Список  $LIS$  инициализируется элементами  $H$ , то есть все  $(i,j)$   $LL$  части старшего уровня декомпозиции. Поле типа каждого элемента  $LIS$  содержит на данном этапе значение  $A$ . В процессе кодирования элемент этого списка считается значимым, если множество, соответствующее

элементу  $(i,j)$  (типа А или В), содержит хотя бы один значащий коэффициент, в противном случае элемент считается незначимым. Если элемент  $(i,j)$  типа А значим, то он переходит в конец списка  $LIS$  со сменой типа на В, а его «дети»  $Q(i,j)$  добавляются в конец  $LIP$  или  $LSP$ , в зависимости от того значим каждый из них или нет. Если становится значимым элемент  $(i,j)$  типа В, то его «дети»  $Q(i,j)$  добавляются в конец списка  $LIS$ , как элементы типа А, а сам элемент  $(i,j)$  из  $LIS$  удаляется. Этот процесс можно рассматривать как отбрасывание предположительно незначимых элементов. Пока все элементы дерева незначимы, генерируется символ «0», что и обеспечивает сжатие. Если же дерево содержит значащий элемент можно априорно предположить, что этот элемент находится в вершине дерева (вспомним свойство упаковки энергии и нормализацию фильтров). Тогда кодер формирует символ «1» и отбрасывает верхние элементы дерева в  $LIP$  и  $LSP$  для их последующего анализа. Теперь эти точки рассматриваются изолированно от дерева. Элементы  $LIS$  переходят в конец списка для дальнейшего анализа на текущем этапе. Так будет происходить до тех пор, пока деревья не будут содержать ни одного значащего элемента. Все значимые элементы, таким образом, попадут в  $LSP$ , а не значащие, но отброшенные деревьями, попадут в  $LIP$ . В процессе обработки  $LIP$  формируется символ «1» в выходном потоке, если элемент  $(i,j)$  значим на данном этапе и «0» в противном случае (сразу после «1» следует символ знака элемента: «1» если «-» и «0» если «+»). Если элемент значим, то он удаляется из  $LIP$  и добавляется в конец списка  $LSP$ . Обработка  $LSP$  сводится к передаче текущего разряда. Для всех элементов  $LSP$  знак уже передан декодеру.

Начальный порог  $T_0$  выбирается равным  $2^{N_{max}}$ . То есть передача начинается со старшего битового поля вейвлет-разложения и продолжается до нулевого битового поля. Значимость отсчета определяется наличием единицы в соответствующем разряде. Процесс обработки списков обрамляется алгоритмом побитовой передачи, и в таком случае *SPIHT* формально

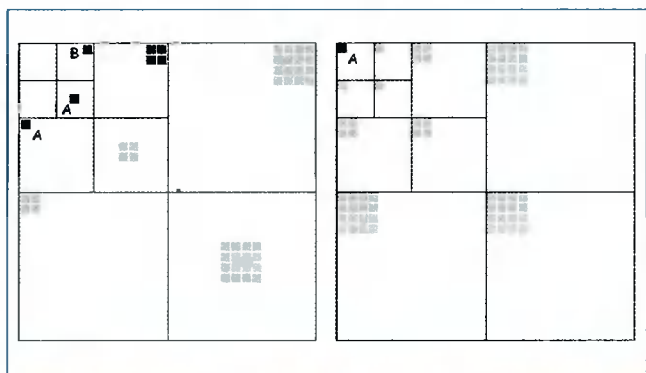


Рис. 6. Примеры элементов  $LIS$ .



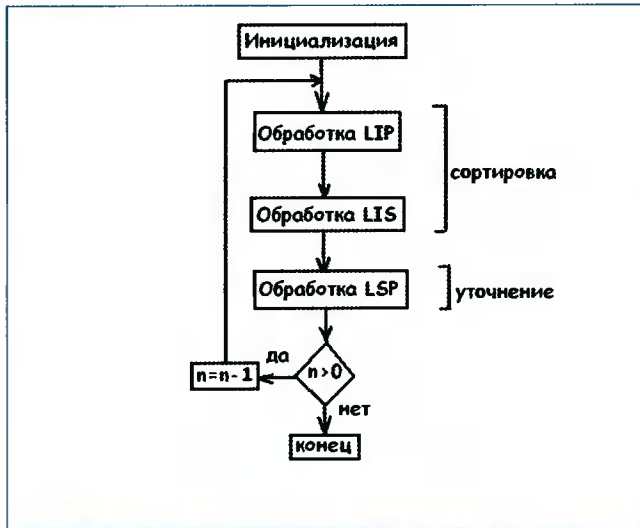


Рис. 7. Упрощенная блок схема SPIHT кодера.

можно рассматривать как смену системы координат. Сжатие происходит, как не трудно заметить, в момент формирования символа «0» в процессе анализа LIS. На рис.6 приведен пример элементов LIS (черным цветом обозначены отброшенные отсчеты).

Обозначим функцию значимости множеств как  $S_n\{T\}$ :  
 $S_n\{T\} = 1$ , если  $\max_{ij}(a_{ij}) \geq 2n$  ( $(i,j)$  принадлежат  $T$ ),  
 $S_n\{T\} = 0$ , в противном случае.

На рис. 7 приведена упрощенная схема SPIHT кодера.

### Алгоритм SPIHT кодера:

#### 1) Инициализация:

В выходной поток выводится число разрядов занимаемых максимальным по модулю отсчетом ( $n = \log_2(\max_{ij}(a_{ij}))$ ). Список значимых точек LSP инициализируется как пустой, незначимых точек LIP – все  $(i,j)$  из  $H$ , незначимых множеств LIS – все  $(i,j)$  из  $H$  с типом А.

#### 2) Проход сортировки:

2.1) Просмотр элементов списка незначимых точек LIP и вывод в поток текущей значимости ( $S_n\{(i,j)\}$ ); в случае если элемент значим, необходимо вывести его знак ( $sign(a_{ij})$ ) и поместить элемент в список значимых точек LSP.

#### 2.2) Просмотр элементов списка незначимых множеств LIS.

Для элементов типа А – вывод значимости ( $S_n\{D(i,j)\}$ ). Если элемент не значим, осуществляется переход к следующему элементу, если же значимость 1, необходимо проанализировать детей из множества  $Q(i,j)$ , выводя при этом в поток их значимость. В случае значимости потомка он помещается в LSP, и в поток выводится его знак, иначе потомок помещается в LIP. Сам же значимый элемент  $(i,j)$  помещается в конец списка незначимых множеств LIS с типом В, если  $L(i,j)$  не пусто, и удаляется в противном случае.

Для элементов типа В – вывод значимости ( $S_n\{L(i,j)\}$ ), если элемент не значим осуществляется переход к следующему элементу, если же значимость 1, все элементы множества  $Q(i,j)$  добавляются в конец списка незначимых множеств LIS с типом А, а сам элемент  $(i,j)$  из LIS удаляется.

#### 3) Проход уточнения:

для всех  $(i,j)$  из LSP вывести  $n$ -ый разряд отчета  $a_{ij}$ ;

#### 4) Переход на следующий бит:

декремент  $n$  на 1 и переход на этап 2.

Этапы 2 и 3 выполняются последовательно для всех битовых полей (от  $n_{max}$  до 0), сохраняя, таким образом, все свойства прогрессивной передачи.

Следует отметить, что элементы LSP на практике также имеют поле типа. Оно необходимо, чтобы пометить те элементы, текущий разряд которых уже передан. Это элементы, ставшие значимыми на данном этапе кодирования при обработке LIP и LIS, их текущий разряд содержит «1». Такие элементы исключаются из обработки LSP, но на следующем этапе обрабатываются на общих основаниях.

Формируемый поток, как и в случае алгоритма побитовой передачи, разбивается на участки передачи одного разряда. Внутри каждого из участков поток делится на биты сортировки и биты уточнения. Биты сортировки содержат помимо служебной информации разбиения деревьев информацию о содержании  $n$ -го разряда значимых отчетов. Возможна была бы и такая структура кодера в которой : при значимости элемента LIS типа А все «дети»  $Q(i,j)$  помещались бы в LIP, а анализ LIP следовал бы после обработки LIS. Это упростило бы обработку LIS. Но в таком случае в выходном битовом потоке существовали бы длинные серии служебных бит, генерируемых при анализе LIS, которые не содержали бы полезной информации для декодера, а лишь меняли структуру деревьев. То есть были бы возможны случаи увеличения размера выходного потока без улучшения качества изображения. Чтобы избежать этого, в обработку LIS вносится анализ значимости точек, таким образом «замешивается» служебная информация и информация, снимающая неопределенность.

### Свойства потока

Алгоритм SPIHT генерирует вложенный, иерархический код. Подобные кодеры позволяют осуществить прогрессивную передачу изображения с последовательным уточнением его на приеме. При этом изображение вначале аппроксимируется небольшим количеством бит, а потом эта аппроксимация уточняется. Вложенный код имеет то свойство, что при  $R_1 > R_2$  код для  $R_2$  будет



префиксом кода для  $R_i$ . Такие коды имеют большой практический интерес по следующим причинам:

- 1) возможность точного регулирования скорости передачи, а при записи в файл его размер может быть задан с точностью до байта;
- 2) возможность восстановления всего изображения при прекращении приема декодером бит в любой точке. При этом изображение будет максимально хорошего качества для данного числа бит. Это применимо для передачи по каналам с потерями, а также для приложений вещания. В этом случае кодер генерирует высокоскоростной высококачественный поток, который передается по каналам различной пропускной способности декодерам различной вычислительной возможности, последние выделяют из него нужные им субпотoki;
- 3) возможность быстрого просмотра изображений в удаленной базе данных. Для поиска достаточно и грубой копии, а при нахождении нужного изображения оно декодируется полностью.

### Предобработка

В процессе анализа элементов  $LIS$ , кодер должен просматривать все дерево в поисках значимых коэффициентов. Этот процесс, неоднократно повторяясь, занимает значительную часть вычислительных ресурсов. Для ускорения процесса кодирования в алгоритм была введена новая структура – матрица значимости ( $MS$ ), формируемая на этапе предобработки в процессе кодирования (для декодирования не требуется). Размер  $MS$  равен размеру вейвлет-разложения. Под каждый элемент отводится один байт памяти. Формируется  $MS$  таким образом, что значение каждого ее элемента  $(i,j)$  равно числу бит, занимаемому максимальным по модулю коэффициентом де-

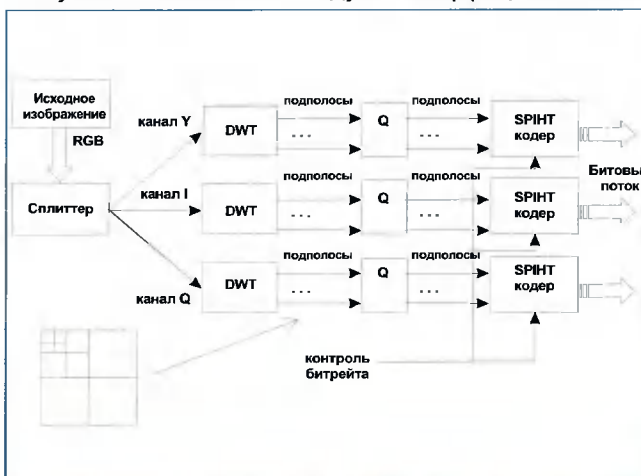


Рис. 8. Метод сжатия изображения.

рева, выстроенного из точки  $(i,j)$  с учетом корня. Имея в распоряжении такую структуру для анализа элемента типа  $A$ , достаточно рассмотреть только точки с координатами «детей», обращаясь при этом к  $MS$ . Анализируя же элементы типа  $B$ , необходимо рассмотреть только точки с координатами «внуков» в структуре  $MS$ . Это простое инженерное решение позволило значительно сократить время работы.

Матрицу  $MS$  можно сформировать за один проход, начиная обработку из правого нижнего угла изображения и двигаясь влево по строкам. При этом в  $MS$  на  $(i,j)$  позицию записывается количество бит, занимаемое отсчетом  $(i,j)$  изображения, если у отсчета нет потомков и записывается максимум из уровней значимости данного отсчета и его потомков, в случае их наличия.

### Программные решения

Алгоритм SPIHT содержит операции логарифмирования, возведения в степень, умножения и работы с вещественными числами. Операции логарифмирования и возведения в степень можно заменить логическими операциями с использованием массива степеней двойки. Большинство операций умножения можно заменить сдвиговыми операциями. Работу с вещественными числами можно полностью исключить, оперируя непосредственно битами их 32-разрядного представления.

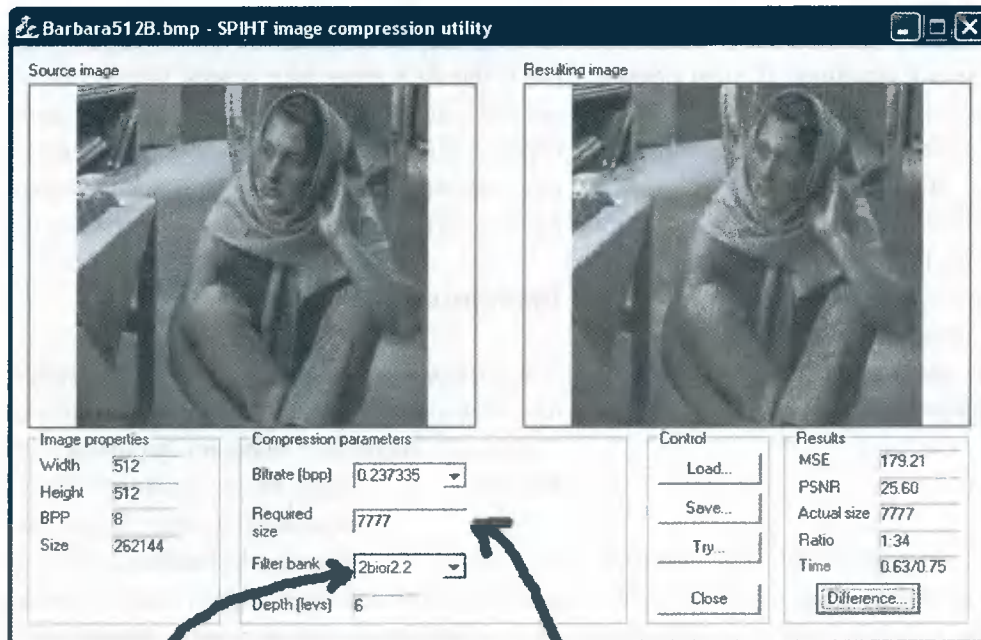
### Программный комплекс «I\_comp»

Процесс разработки метода сжатия сопровождался его программной реализацией. Ее результат – две программы под ОС Windows, которые, по сути дела, являются важным достижением всей разработки в целом, показывая применимость метода.

Разработанная программа позволяет кодировать изображения в формате BMP (цветные и в градации серого). Степень сжатия задается произвольно, можно также задать точный размер выходного файла в байтах, что является несомненным преимуществом кодера. Банк фильтров, применяемый при вейвлет-декомпозиции, можно выбрать из списка. В процессе проектирования была разработана структура файла выходного потока – стандарт SPT, который может быть декодирован и сохранен в BMP программой-декодером «SPTV». Предусмотрена возможность добавления новых фильтров к уже имеющимся; описаны соответствующие процедуры. Также есть возможность записи отсчетов фильтра в выходной файл, для того, чтобы удаленный декодер мог восстановить изображе-

ние, не имея в своей базе фильтра примененного при кодировании. Программа использует оконный интерфейс для взаимодействия с пользователем. Возможна также работа с командной строки. На сегодняшний день данный проект можно считать наиболее полной и функциональной реализацией SPIHT-кодера. Внешний вид приложения представлен на *рис.9*.

На *рис. 8* приведена блок схема процесса сжатия одного изображения. На первом этапе цветные изображения преобразуются из RGB в YIQ систему. Затем каждая из составляющих проходит дискретное вейвлет-преобразование (DWT), квантование и кодирование. Общий размер памяти, выделенный под выходной файл, делится между составляющими в соответствии с их энергией и степенью влияния на восприятие человеком.



выбор банка фильтров

точный размер выходного файла

характеристики результата

Рис. 9. Программный комплекс «I\_com».

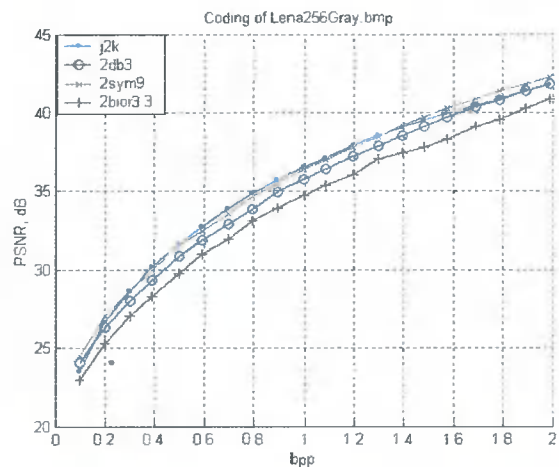
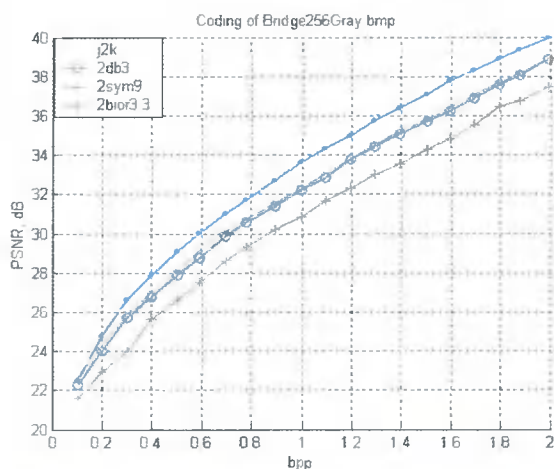


Рис. 10. Кодирование тестовых изображений JPG2000-кодером (j2k) и алгоритмом SPIHT с применением различных фильтров.

## Тесты

На рис. 10 приведен сравнительный тест разработанного алгоритма с одним из лидеров в области сжатия изображений JPG2000. На графиках видно, что работа SPIHT-кодера в значительной мере зависит от используемого банка фильтров, а тесты говорят о большей приспособленности алгоритма SPIHT к изображениям с малым числом контуров и резких переходов, которые хорошо аппроксимируются вейвлетами, например Lepa256Gray (рис 11). В этом случае SPIHT-кодер, имея на много более простую структуру, дает результат сравнимый с JPG2000. Применительно же к изображениям с большим числом контуров JPG2000 показывает некоторое преимущество, что можно увидеть на тесте для Bridge256Gray. Следует также отметить, что описываемая версия алгоритма SPIHT не содержит арифметического кодирования, введение которого позволит улучшить результат.

## Заключение

Данная статья носит методический характер и направлена на восполнение недостатка русскоязычной литературы по алгоритму SPIHT и ему подобных. Авторы надеются, что эта статья позволит инженерам с различным уровнем подготовки не только разобраться в алгоритме, но и при необходимости приступить к его реализации, используя материалы статьи. Изложение базовых подходов к компрессии изображений направлено на формирование определенной методологии для ее использования при разработке новых алгоритмов.



Рис. 11: 1. Тестовое изображение; 2. JPG2000; 3. SPIHT.  
Степень сжатия 1:30.

## Литература

- [1] Ю.Б.Зубарев, В.П.Дворкович. Цифровая обработка телевизионных и компьютерных изображений, Москва, 1997
- [2] В.И. Воробьев и В.Г. Грибунин. Теория и практика вейвлет-преобразования. ВУС, 1999.
- [3] Н.М. Астафьева. Вейвлет – основы теории и некоторые приложения. //Успехи физических наук, 1996, №11.С.1145 -1170.
- [4] Lewis A., Knowles G. "Image compression using the 2-d wavelet transform", IEEE Transactions on Image Processing, 1992, №2 . P.244-250.
- [5] Shapiro J. "Embedded image coding using zerotrees of wavelet coefficients", IEEE Trans. on Signal Processing, 1993, P.3445-3462.
- [6] J.Shapiro, An embedded hierarchical image coder using zerotrees of wavelet coefficients, in: Proceedings of the IEEE Data Compression Conference (Snowbird), 1993, pp.214 .223.
- [7] A. Said and W.A. Pearlman, "A New Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees, "IEEE Trans. on Circuits and Systems for Video Technology, vol. 6., pp. 243--250, June 1996.
- [8] David Taubman, "Embedded block coding in JPEG 2000", Signal Processing Image Communication 17 (2002)49 .72
- [9] D.Taubman, EBCOT: Embedded block coding with optimized truncation, Tech. Rep. N1020R, ISO/IECJTC1/SC29/WGI, October 1998.



## Новый этап развития твердотельного телевидения — «Видеосистемы на кристалле»

**В.В. Березин, В.М. Гатаулин**

**Н**а рубеже веков телевизионная техника переживает новый этап революционного изменения технологии. Если первый этап этой революции [1] связан с рождением твердотельного телевидения, воплощенного в матричных приборах с зарядовой связью (ПЗС) и микропроцессорах, то нынешний этап характеризуется следующими чертами:

- совершенствование технологии изготовления матричных фотоприемников позволило массово выпускать приборы высокой четкости, причем число элементов разложения измеряется мегапикселями;
- совершенствование технологии изготовления микроэлектронных устройств обработки сигналов привело к синтезу разнородных вычислительных устройств – микропроцессоров, запоминающих устройств и программируемой логики в пределах одной микросхемы, которые получили название «системы на кристалле»;
- применение технологии, развитой для построения микроэлектронных запоминающих устройств, к реализации матричных фотоприемников породило новую ветвь твердотельных устройств, объединяющих в себе фотоприемник, устройства развертки, квантования и обработки изображения.

Совокупность этих черт говорит именно о революционном, а не эволюционном характере развития телевизионной техники. Это видно хотя бы из того, что освоение все более высокой точности изготовления приборов (уже приближающейся к 0,1 мкм), после двух десятилетий работы с матрицами вещательного формата (около 0,4 мегапикселя) позволило быстро преодолеть мегапиксельный барьер, и сейчас даже в мобильных телефонах массово применяются мегапиксельные матрицы, а в цифровых фотоаппаратах – пяти и более мегапиксельные матрицы. В результате телевизионная техника уже полностью вытеснила фотографические методы в астрономии, и наблюдается

*Рассмотрены варианты архитектур нового класса устройств телевизионной техники – «видеосистема на кристалле». Выявлена тенденция объединения в рамках одной микросхемы формирователя изображения и устройств обработки.*

устойчивая тенденция к вытеснению пленочной фотографии из массового фотолюбительства.

При построении систем наблюдения для беспилотных летательных аппаратов [2] актуальным является уменьшение габаритов и энергопотребления подобных систем при сокращенном времени их проектирования. Настоящий материал посвящен решению такой задачи на основе современных тенденций в области устройств формирования изображений.

Видеоинформатика, изучающая все этапы рождения сигналов изображений, преобразования изображений и их передачи [3, 4], охватывает много разнородных этапов и подсистем (оптическая подсистема; фотоприемник, аналоговый видеопроцессор; аналого-цифровой преобразователь; цифровой видеопроцессор целевого назначения, кодек, процессор принятия решений; средства доставки изображения или решения потребителю).

Безусловно, каждый из этих этапов имеет самостоятельную ценность и многообразные варианты реализаций, однако практические требования к современной технике телевидения и технологические возможности требуют их интеграции. До наступления нынешнего этапа твердотельной революции все эти этапы разделялись, в том числе и по технологии изготовления. Происходящий в настоящее время процесс характеризуется единством технологии большинства подсистем, которое сделало возможной такую интеграцию. Подчеркнем, что она имеет целью не только уменьшение массогабаритных показателей и энергопотребления, но и перенос центра сложности обработки видеоинформации и принятия решений из внешнего вычислителя непосредственно в телекамеру. Это в корне меняет роль телекамеры в системе –





Рис. 1. «Видеосистема на кристалле» в типовой телевизионной системе.

из пассивного формирователя изображения она превращается в интеллектуальный источник видеoinформации и решений. Другими словами, телекамера и сейчас уже методологически неотделима от компьютера [4], а недалеко то время, когда телекамера будет составлять единое технологическое и методологическое целое с компьютером или процессором цифровой обработки.

Движение к такому будущему идет с двух сторон. Со стороны средств цифровой обработки информации появилось новое направление «система на кристалле», сочетающее различные парадигмы такой обработки. Так, при проектировании цифрового видеопроцессора длительное время стоял вопрос о выборе программной или аппаратной реализации алгоритмов обработки видеосигнала, острота которого была снята с появлением приборов, сочетающих в себе преимущества обоих классов [5]. Со стороны средств формирования изображения происходит объединение на одном кристалле датчика изображения и аналогового видеопроцессора. Это привело к появлению новой элементной базы – однокристалльных цифровых телевизионных камер или «видеосистем на кристалле». Такие камеры в настоящее время выпускаются многими производителями [6], число которых превышает 50 и непрерывно растет.

Технологически наиболее приспособленной к такой интеграции является КМОП технология. Основ-

ные процессы формирования и преобразования изображений приведены на рис. 1.

Структуры конкретных «видеосистем на кристалле» объединяют в различных конфигурациях, но, как правило, интегрируются блоки 2–10. Специфика функционирования телевизионной аппаратуры, например, параметры накопления, развертки, усиления и кодирования сигнала изображения задаются программным путем с помощью конфигурационных регистров. Рассмотрим характеристики и структурные схемы построения ряда однокристалльных камер с разрешением 640x480 элементов изображения [6], имеющих в своем составе полную обработку видеосигнала с цифровым выходом, основные показатели которых приведены в табл. 1. Целью такого рассмотрения является анализ технических решений и архитектур обработки сигнала для ряда присутствующих на рынке «видеосистем на кристалле» и попытка прогноза развития этого направления. Основное внимание уделяется не собственно фотоэлектрическим характеристикам преобразователя свет–сигнал [3, 7], входящего в состав «видеосистемы на кристалле», а таким особенностям аппаратно-программного окружения, которые превращают эти приборы в новый класс устройств формирования и обработки изображений. В зарубежной литературе уже появилась соответствующая аббревиатура для подобных устройств – VSoC (Video System on Chip).

Первый представитель VSoC – видеодатчик VS6552, предназначен для использования в мобильных приложениях. Его структурная схема приведена на рис. 2. Она включает в свой состав КМОП матрицу, формирователь адреса строк и столбцов, аналого-цифровой преобразователь (АЦП), буферную память на строку изображения, схемы интерфейса и управления питанием. Данная «видеосистема на кристалле» характеризуется простотой реализации и ее отличительными особенностями по сравнению с аналогами являются:

- совмещение в едином модуле объектива и формирователя изображения;

Таблица 1. Характеристики однокристалльных видеосистем

Тип прибора	VS6552	KAC-9628	MT9V11	ZDM33220	OV7640
Разрешение	644x484	648x488	640x480	840x640	640x480
Размер элемента изображения, мкм	5.6x5.6	7.5x7.5	3.6x3.6	10.6x10.6	5.6x5.6
Разрядность встроенного АЦП, бит	10	8/10/12	10	8/10	8
Максимальная	30	30	90	100	30/60
Кол-во выводов корпуса	14	48	44	52	28
Интерфейс управления	I2C	I2C	I2C	UART	SCCB

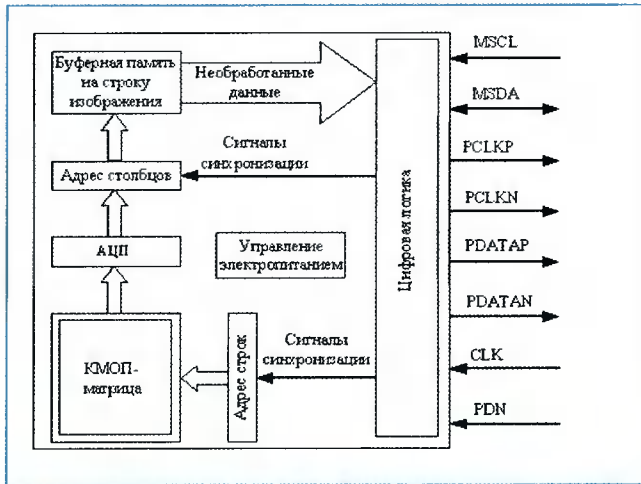


Рис. 2. «Видеосистема на кристалле» VS6552 фирмы STMicroelectronics

- применение АЦП с последовательным выходом и передачей видеoinформации по дифференциальному высокоскоростному последовательному интерфейсу.

Видеодатчик VS6552 подключается к специализированному процессору, который обеспечивает обработку информации о цвете, согласно используемого кодирующего светофильтра RGB, управление экспозицией и организацию вывода видеoinформации в форматах YCbCr, RGB или JPEG.

Наиболее широко на отечественном рынке известны однокристалльные камеры производства OmniVision Technologies [8]. В состав «видеосистемы на кристалле» OV7640/OV7141 входят (рис. 3):

- светочувствительная матрица с микролинзами для увеличения чувствительности, имеющая дополни-

тельно 6 строк и два столбца для установки уровня черного и цветовой интерполяции (полное разрешение 682x482 элемента изображения);

- схема тактового генератора, которая осуществляет формирование кадра изображения в форматах 640x480 или 320x240 элементов изображения, управление временем экспозиции, формированием внешних синхронизирующих импульсов; для своей работы схема тактового генератора использует внешнюю синхронизацию частотой до 27 МГц;
- схема аналоговой обработки с цифровым управлением, которая обеспечивает регулировку усиления сигнала изображения, баланса белого и настройку других параметров, влияющих на качество изображения;
- встроенный АЦП, который включает два 8-разрядных параллельных канала, один из которых предназначен для квантования яркостной или зеленой составляющей сигнала (Y/G), второй для квантования аналогового сигнала, несущего информацию о цвете (CbCr/BR). Применение сдвоенного АЦП совместно со схемой компенсации уровня черного является особенностью данной системы на кристалле;
- блок форматирования с цифровым видеопортом, который обеспечивают постобработку цифрового видеосигнала, например, вывод определенной зоны светочувствительной матрицы (windowing) и трансляцию выходного двоичного кода на внешние схемы;
- трехпроводной интерфейс управления SCCB и набор регистров, которые обеспечивают цифровую конфигурацию однокристалльной видеокamеры.

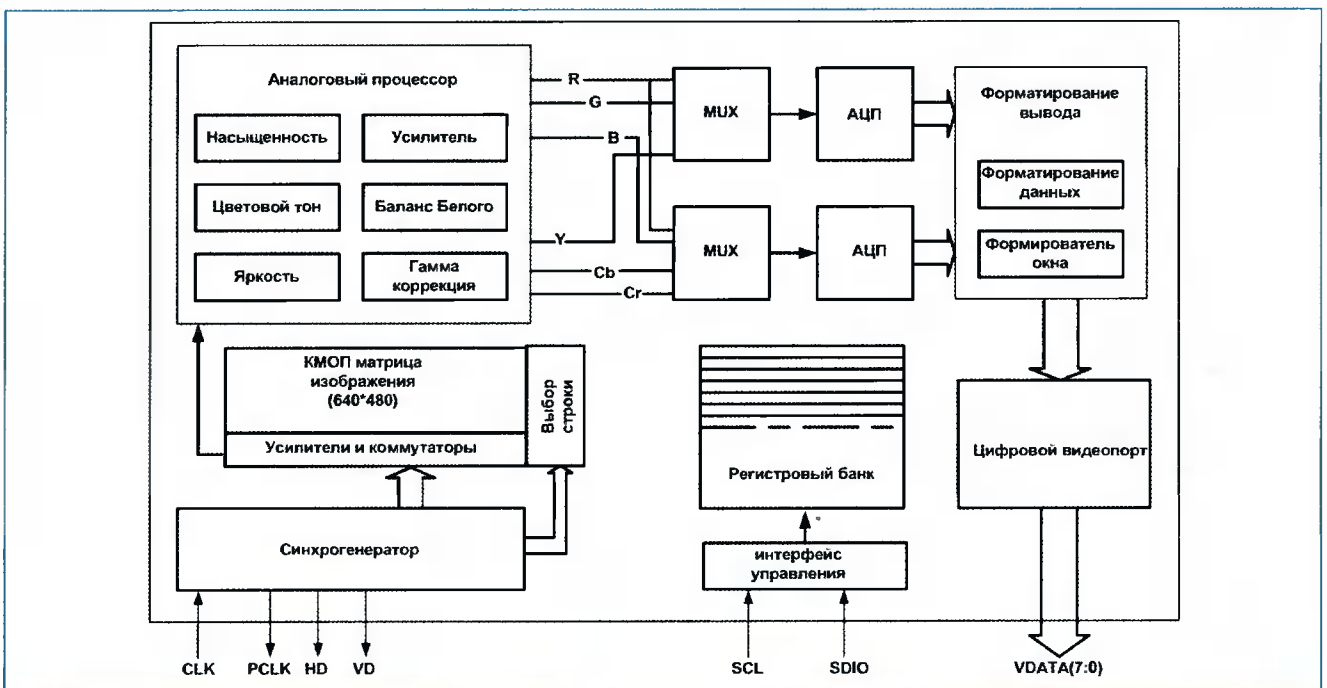


Рис. 3. «Видеосистема на кристалле» OV7640 фирмы OmniVision

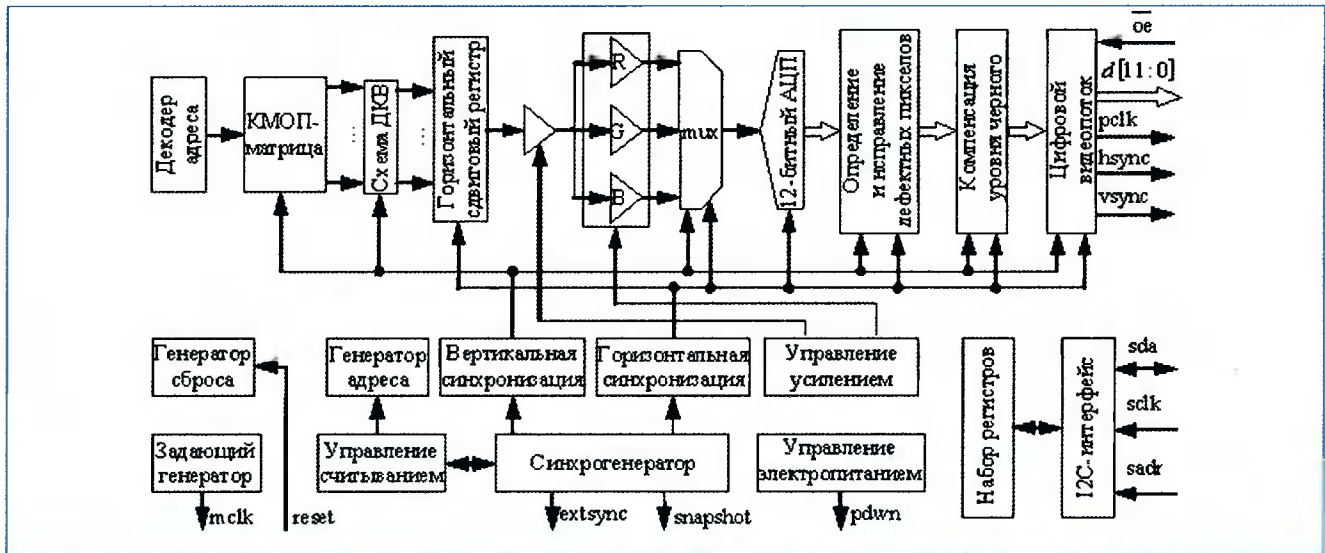


Рис. 4. «Видеосистема на кристалле» KAC-9628 фирмы KODAK

Общее число байт информации, которой управляется микросхема V7640/OV7141, превышает 80. Кроме стандартных настроек однокристалльной камеры, с помощью интерфейса управления можно получать статистику о среднем значении сигнала изображения в каналах яркости и цветности, изменять направление развертки для получения зеркального или перевернутого изображений и их вида (чересстрочная или прогрессивная), управлять энергопотреблением микросхемы. Это позволяет организовывать гибкие перестраиваемые системы получения изображения в зависимости от текущей цели аппаратуры.

Структурная схема «видеосистемы на кристалле» фирмы Kodak приведена на рис. 4. Она позиционируется как однокристалльная камера с большим динамическим диапазоном – до 110 дБ. Полный формат светочувствительной матрицы составляет 648x448 элементов. По 4 пиксела с каждой стороны матрицы используются для формирования оптического «черного». В начале каждого кадра производится очистка накопленных зарядов, после окончания времени накопления зарядовое изображение с помощью двойной коррелированной выборки преобразуется в электрический сигнал изображения. Далее аналоговый сигнал подается на два каскада усиления: общий и избирательный по цвету. Коэффициент усиления каждого из этих каскадов регулируется цифровым кодом в пределах от 0 до 15 дБ. После усиления сигнал изображения каждой цветовой составляющей через аналоговый мультиплексор поступает на вход АЦП.

Цифровой 12-разрядный сигнал изображения проходит блок коррекции дефектных пикселей, схемы регулировки уровня черного и поступает на выход кристалла в программируемом формате 8, 10, 12 бит. Блок коррекции может программно отключаться, а

логика его функционирования включает в свой состав 4 программируемых регистра пороговых уровней. Работа схемы коррекции уровня черного основана на вычитании из сигнала изображения оценки уровня. Кроме этого, отдельный программно-доступный регистр определяет смещение уровня при аналого-цифровом преобразовании.

Блок синхронизации тактируется от внешнего сигнала частотой до 48 МГц и содержит схемы вертикальной и горизонтальной развертки. Программируемые регистры определяют окно вывода информации, которое может располагаться в произвольном месте светочувствительной матрицы. Также, программным путем определяются несколько режимов субдискретизации, позволяющие прореживать пиксели при формировании цифрового изображения и режимы прогрессивного или чересстрочного разложения. Таким образом, результирующая частота кадров определяется входной тактовой частотой, размерами выводимого окна, видом субдискретизации и может регулироваться в широких пределах от стандартных значений до существенно превышающих вещательную кадровую частоту. Высокая кадровая частота, вплоть до 1 кГц и выше, весьма актуальна при классификации динамических изображений, например, в адаптивной оптике.

Один и тот же прибор «видеосистемы на кристалле» может использоваться и в сверхбыстродействующих, и в малокадровых телевизионных системах. Для реализации малокадрового режима характерно использование программирования синхронизации для покадровой съемки. Логика такого режима работы определяется входным сигналом «Запрос кадра», программируемой задержкой на несколько кадров, программируемым временем накопления и некото-



рыми другими факторами. Выходной сигнал «Захват кадра» имеет высокий активный уровень во время формирования кадра изображения, что может быть использовано для организации дополнительной подсветки или механического затвора.

Для достижения широкого динамического диапазона входных освещенностей может использоваться программно-управляемая кусочно-линейная аппроксимация свет-сигнальной характеристики. Содержимое регистров определяет точки излома и наклон преобразования для получения зависимости сигнала

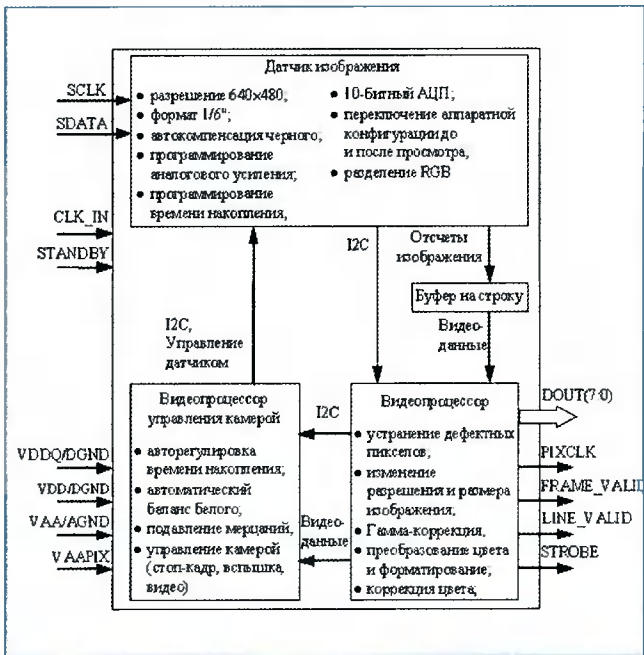


Рис. 5. «Видеосистема на кристалле» MT9V11 фирмы Micron

яркости от освещенности, близкой к зрительному восприятию или для достижения потенциальной контрастной чувствительности.

Важным узлом «видеосистемы на кристалле» является набор регистров конфигурации и интерфейс управления. Общее количество программно-доступных байт в системе составляет 43, которые доступны по широко распространенному двухпроводному интерфейсу *i2c*.

Блок-схема «видеосистемы на кристалле» фирмы Micron MT9V112 приведена на рис. 5. Эта видеосистема ориентирована на применение в составе мобильного оборудования и, как следствие, большинство функций обработки сигнала выполняются автоматически без настройки. К ним относятся компенсация уровня черного, автоматический баланс белого, подавление эффекта мелькания, вызванного люминесцентными лампами, коррекция дефектных пикселей, цветовая обработка.

Система на кристалле MT9V112 включает в свой состав 10-битный АЦП на этапе первичного квантования и внутренней обработки сигнала. На внешние выходы подаются результирующие 8 разрядов после преобразований сигнала. Такой подход иллюстрирует сочетание точности внутреннего представления сигнала с технологичностью его внешнего использования.

Синхронизация MT9V112 осуществляется от источника тактовой частоты 27 МГц, блок развертки обеспечивает прогрессивное разложение с частотой кадров 15 Гц при формате 640x480 элементов, а

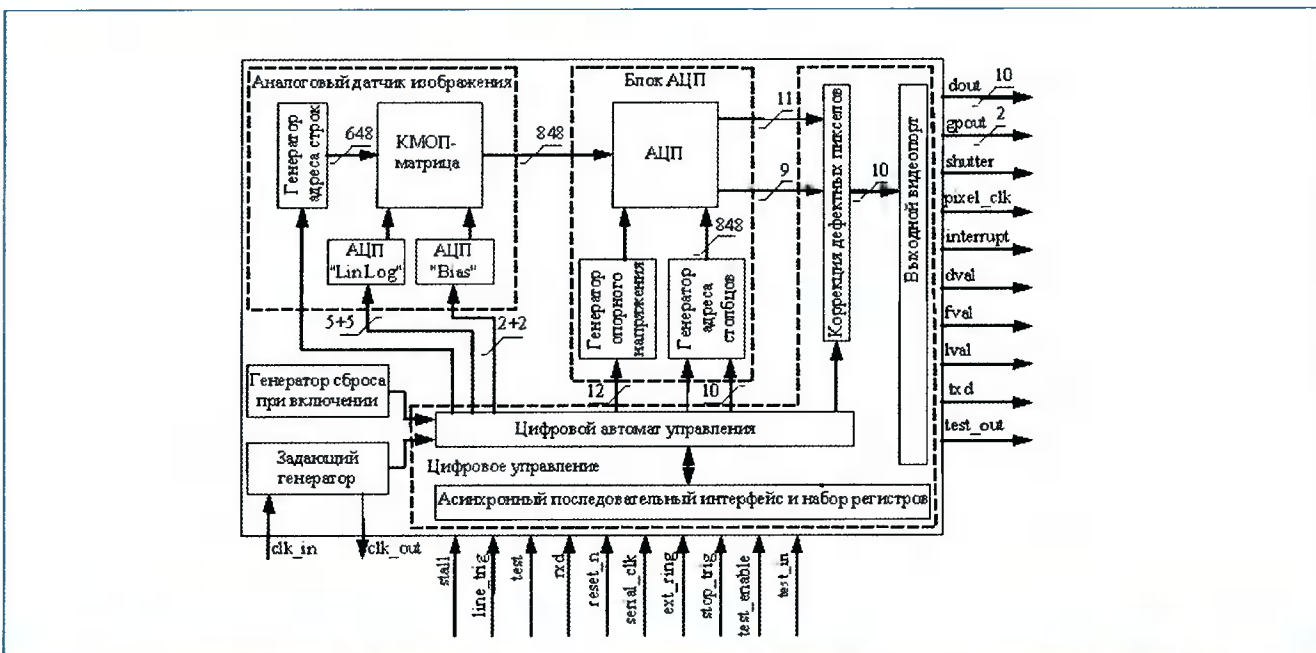


Рис. 6. «Видеосистема на кристалле» ZDM33220 фирмы ZMD AG



при частоте кадров до 90 Гц реализуется формат 320x240 элементов.

К конфигурируемым через интерфейс i2c относятся такие параметры как усиление сигнала изображения, опорное напряжение АЦП, настройка электронного затвора. Подобно микросхеме КАС-9628 данная видеосистема имеет возможность организации кадровой съемки с одновременной синхронизацией осветителя, например, лампы вспышки.

К «видеосистемам на кристалле» повышенного разрешения относится микросхема ZDM33220. От описанных выше она отличается большим размером элемента изображения и оптическим форматом  $\frac{2}{3}$  дюйма. Структурная схема ZDM33220 приведена на рис. 6.

Она включает в свой состав КМОП датчик изображения, программируемые схемы развертки (генераторы адреса строк и столбцов), цифроаналоговые преобразователи для коррекции свет-сигнальной характеристики для увеличения динамического диапазона (ЦАП «LinLog»), цифроаналоговый преобразователь установки уровня черного (ЦАП «Bias»), генератор опорного напряжения, 12-битный АЦП видеосигнала, схему коррекции дефектных пикселей, выходной видеопорт со схемой синхронизации, схему последовательного интерфейса и регистровый банк, программируемый цифровой автомат управления, схему тактового питания и схему управления энергопитанием.

Данная видеосистема находит применение не только для целей видеонаблюдения, но и для прикладных задач, поэтому имеет расширенные варианты своей программной конфигурации. Важной функцией видеосистемы является наличие до трех программно управляемых окон изображения, положение и размер которых задаются в пределах КМОП матрицы. При этом возможен обмен пространственного разрешения на временное: частота считывания такой зоны определяется частотой опроса элемента, размером зоны и может достигать нескольких килогерц.

«Видеосистемы на кристалле» приобретают новые черты как в технологии изготовления, так и в способах их использования. Например, алгоритм функционирования этих устройств задается путем программирования параметров, что позволяет проводить адаптивное управление с учетом характеристик канала связи и получателя. Правда, в некоторых случаях задание способа развертки, дающее существенное уменьшение сложности прибора, сокращает возможности проектировщика при согласовании оптического изображения с компьютерным классификатором.

Существенный пласт цифровой обработки изображений в составе уже выпускаемых однокристалльных камер отсутствует, что связано с принципом их применения, в первую очередь, для целей видеонаблюдения. Однако, можно прогнозировать, что при дальнейшем развитии «видеосистем на кристалле» в их составе обязательно появятся процессор управления конфигурацией системы на кристалле и процессор цифровой обработки, а также программируемая логика для целевой обработки изображений и принятия решений. Применение «видеосистем на кристалле» для телевизионного наблюдения с беспилотных летательных аппаратов позволило значительно сократить время проектирования такого оборудования, при значительном уменьшении сложности, габаритов и энергопотребления. Недостатком подобных систем на сегодняшний день является пониженная в 2 раза чувствительность по сравнению с ПЗС датчиками изображения.

Итак, по приведенному обзору можно сделать следующие выводы:

1. В настоящее время происходит становление нового класса телевизионных систем, знаменующее новый этап твердотельной революции в телевидении, – «видеосистемы на кристалле», являющиеся логическим развитием микроэлектронной техники, которые привели к слиянию двух направлений: твердотельных фотоприемных матриц и новых устройств цифровой обработки сигналов, получивших название «системы на кристалле».
2. Появление нового класса «видеосистем на кристалле» ведет к изменению метода проектирования телевизионных устройств, т. к. существенная часть телевизионной камеры предоставляется проектировщику готовой. Если на начальном этапе развития телевидения, согласно методологии создателя электронного телевидения В. К. Зворыкина, в проектировании преобладали методы физики, то новый этап порождает новый метод проектирования, который характеризуется высокой скоростью создания систем, резким возрастанием роли системного подхода и значимости программно-алгоритмического обеспечения.
3. Развитие «видеосистем на кристалле» превращает однокристалльную камеру в центральный (в пределе – единственный) узел системы формирования и обработки изображений, для которой становится актуальной задача распределения ресурсов между фотоприемником и подсистемой обработки для наилучшего выполнения системой стоящей перед ней цели.

## Литература

1. Хромов Л. И., Лебедев Н. В., Цыцулин А. К., Куликов А. Н. Твердотельное телевидение. М., Радио и связь, 1986.
2. Анцев Г. В., Тупиков В. А., Турнецкий Л. С. Мониторинговый комплекс с малогабаритным дистанционно пилотируемым вертолетом. Проблемы транспорта, 2004, № 10, с. 235 – 239.
3. Цифровое преобразование изображений: Учеб. пособие для вузов / Р. Е. Быков, Р. Фрайер, К. В. Иванов, А. А. Манцветов; под ред. профессора Р.Е. Быкова. – М.: Горячая линия – Телеком, 2003. – 228 с.
4. Цыцулин А. К. Телевидение и космос: Учеб. пособие. – СПб., Изд-во СПбГЭТУ «ЛЭТИ», 2003. – 228 с.
5. Березин В. В. Повышение пропускной способности реализации аппаратного стека протоколов TCP/IP /Телекоммуникации № 4 2004 г. с. 18-21.
6. [http://video-equipment.globalspec.com/SpecSearch/Suppliers/Video\\_Imaging\\_Equipment/Image\\_Sensors/CMOS\\_Image\\_Sensors.html](http://video-equipment.globalspec.com/SpecSearch/Suppliers/Video_Imaging_Equipment/Image_Sensors/CMOS_Image_Sensors.html)
7. Вахромеева О.С., Манцветов А.А., Шиманская К.А. Характеристики чувствительности телевизионных камер на матричных приборах с зарядовой связью// Изв. вузов России. Радиоэлектроника. 2004. вып. 4. С. 25-35.8
8. Ракович Н. Н. CameraChips: полная видеосистема на одном кристалле/Компоненты и технологии № 1 2004 г. с. 128 - 130

## Уважаемые коллеги !

Приглашаем вас принять участие в формировании тематических выпусков журнала «Цифровая обработка сигналов» в 2005\2006 гг.

Научно-технический журнал «Цифровая обработка сигналов» выпускается на регулярной основе с 2000 года. Цель его издания – содействие развитию и широкому применению методов цифровой обработки сигналов (ЦОС) и современных информационных технологий реального времени в разработках отечественных ученых и специалистов. Журнал «Цифровая обработка сигналов» внесен в дополнительный Перечень периодических изданий, в которых рекомендуется публикация основных результатов диссертаций на соискание ученой степени доктора технических наук.

Просим вас выразить свое отношение к данному предложению по сотрудничеству с редакцией журнала. Желательно заранее сообщить ориентировочное название статей и их предполагаемый объем (максимальный объем – до 20 страниц для обзорных статей и до 12 страниц для оригинальных работ).

**При подготовке ваших предложений просим учесть, что предпочтение отдается статьям обзорного характера и работам, посвященным решению проблем эффективного проектирования и технической реализации систем ЦОС с применением сигнальных процессоров и ПЛИС, а также различным приложениям технологий обработки сигналов и изображений в реальном времени.**

По всем вопросам, связанным с требованиями и порядком предоставления материалов статей, обращайтесь на сайт журнала: [www.dsps.ru](http://www.dsps.ru).

Предварительные заявки и материалы работ на рецензирование и публикацию принимаются по E-mail: [tor@rgta.rayzan.ru](mailto:tor@rgta.rayzan.ru)

Ждем ваших предложений.

Зам. главного редактора, профессор

В.В. Витязев