ISSN 0555-2923

РОССИЙСКАЯ АКАДЕМИЯ НАУК

# Проблемы передачи информации


ТОМ 59 ВЫП. 1

2023

РОССИЙСКАЯ АКАДЕМИЯ НАУК

# ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

**Журнал основан в январе 1965 г.** **ISSN: 0555-2923** **Выходит 4 раза в год**

**Том 59, 2023** **Вып. 1**

**Москва**

## СОДЕРЖАНИЕ

**CONTENTS**

**ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ**

**Том 59** **2023** **Вып. 1**

УДК 621.391 : 519.725

© 2023 г. **С.Г. Колесников, В.М. Леонтьев**

# СЕРИИ ФОРМУЛ ДЛЯ ПАРАМЕТРОВ БХАТТАЧАРЬИ В ТЕОРИИ ПОЛЯРНЫХ КОДОВ[1]

В теории полярных кодов для определения позиций замороженных и информационных бит используются параметры Бхаттачарьи. Они характеризуют скорость поляризации каналов $W_N^{(i)}$, $1 \leqslant i \leqslant N$, специальным образом построенных из исходного канала $W$, где $N = 2^n$ – длина кода, $n = 1, 2, \ldots$ В случае, когда $W$ – двоичный симметричный канал без памяти, приведены две серии формул для параметров $Z\big(W_N^{(i)}\big)$: при $i = N - 2^k + 1$, $0 \leqslant k \leqslant n$, и при $i = N/2 - 2^k + 1$, $1 \leqslant k \leqslant n-2$. Формулы требуют порядка $\binom{2^{n-k} + 2^k - 1}{2^k} 2^{2^k}$ операций сложения для первой серии и порядка $\binom{2^{n-k-1} + 2^k - 1}{2^k} 2^{2^k}$ для второй. Для случаев $i = 1, N/4 + 1, N/2 + 1, N$ найденные выражения для параметров удалось упростить, вычислив входящие в них суммы. Указаны возможные обобщения для значений $i$ из интервала $(N/4, N)$. Также исследуются комбинаторные свойства поляризационной матрицы $G_N$ полярного кода с ядром Арикана. В частности, установлены простые рекуррентные соотношения между строками матриц $G_N$ и $G_{N/2}$.

*Ключевые слова*: полярный код, параметр Бхаттачарьи, поляризационная матрица.

**DOI:** 10.31857/S0555292323010011, **EDN:** JDDBTP

## § 1. Введение и основные результаты

Пусть $W$ – двоичный симметричный канал без памяти с входным алфавитом $X = \{0, 1\}$, выходным алфавитом $Y = \{0, 1\}$ и переходными вероятностями

$$W(y \,|\, x) = \begin{cases} p, & \text{если } x \neq y, \\ 1 - p & \text{в противном случае.} \end{cases}$$

Через $W^N$, $N = 2^n$, $n = 1, 2, \ldots$, обозначим $N$-ю декартову степень канала $W$. Для каждого натурального числа $i$, $1 \leqslant i \leqslant N$, определим канал $W_N^{(i)} \colon X \to Y^N \times X^{i-1}$ с переходными вероятностями

$$W_N^{(i)}(y, u' \,|\, u_i) = \frac{1}{2^{N-1}} \sum_{u'' \in X^{N-i}} W^N(y \,|\, uG_N),$$

где $y \in Y^N$, $u' \in X^{i-1}$, $u_i \in X$, $u = u'u_iu''$ – конкатенация векторов $u'$, $(u_i)$ и $u''$, а $G_N$ – поляризационная матрица полярного кода с ядром Арикана $F = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$.

[1] Работа поддержана Красноярским математическим центром, финансируемым Министерством образования и науки Российской Федерации (номер соглашения 075-02-2022-876).

Тогда [1] значение параметра Бхаттачарьи $Z\big(W_N^{(i)}\big)$ определяется равенством

$$Z\big(W_N^{(i)}\big) = \sum_{y\in Y^N}\sum_{u'\in X^{i-1}} \sqrt{W_N^{(i)}(y,u'\,|\,0)\; W_N^{(i)}(y,u'\,|\,1)}. \tag{1}$$

В той же работе [1] формула (1) была упрощена до следующей:

$$Z\big(W_N^{(i)}\big) = 2^{N-1}\sum_{y\in L} \sqrt{W_N^{(i)}(y,(0,\ldots,0)\,|\,0)\; W_N^{(i)}(y,(0,\ldots,0)\,|\,1)},$$

где $L$ – подмножество в $Y^N$ мощности $2^i$, и было показано, что числа $Z\big(W_N^{(i)}\big)$ удовлетворяют рекуррентной системе равенств и неравенств

$$\begin{cases} Z\big(W_{2N}^{(2i-1)}\big) \leqslant 2Z\big(W_N^{(i)}\big) - \big[Z\big(W_N^{(i)}\big)\big]^2, \\ Z\big(W_{2N}^{(2i)}\big) = \big[Z\big(W_N^{(i)}\big)\big]^2. \end{cases} \tag{2}$$

В [2] предложены два метода аппроксимации (“сверху” и “снизу”) вероятности ошибки в поляризованных битовых каналах. Там же [2, теорема 1] показывается, что для достаточно больших $N$ сложность алгоритма аппроксимаций для всех каналов одновременно имеет порядок $O(N\mu^2\log\mu)$, где $\mu > \mu_0$, причем константа $\mu_0$ зависит от пропускной способности канала $I(W)$, скорости и вероятности блоковой ошибки кода, но не зависит от $N$.

В настоящей статье авторы ставили перед собой задачу получения точных формул для параметров Бхаттачарьи, которые в перспективе можно использовать для расчетов. Приведем основные результаты.

Т е о р е м а 1. *Справедливы равенства*

$$Z\big(W_N^{(1)}\big) = \sqrt{1-(1-2p)^{2N}}, \quad Z\big(W_N^{(N)}\big) = 2^N p^{N/2}(1-p)^{N/2}.$$

Для формулировки следующего результата нам потребуется ввести ряд обозначений. Пусть $m \in \mathbb{N}$ – произвольное натуральное число. Через $\mathbb{Z}_2^m$ обозначим $m$-ю декартову степень поля вычетов $\mathbb{Z}_2$ с операцией сложения $\oplus$. Подмножества в $\mathbb{Z}_2^m$, состоящие из векторов с четной и нечетной суммой координат, обозначим, соответственно, через $\mathbb{Z}_{2,\mathrm{e}}^m$ и $\mathbb{Z}_{2,\mathrm{o}}^m$. Для произвольного $(0,1)$-вектора $z = (z_1,\ldots,z_m)$ и произвольного целочисленного вектора $t = (t_1,\ldots,t_m)$ положим

$$r(t,z) = w(z)\frac{N}{m} + \sum_{j=1}^{m}(\bar z_j - z_j)t_j,$$

где, как обычно, $w(z) = z_1 + \ldots + z_m$ – вес Хэмминга вектора $z$, $\bar 0 = 1$, $\bar 1 = 0$. Пусть также

$$H(t,z) = p^{r(t,z)}(1-p)^{N-r(t,z)}.$$

Наконец, определим на векторах $t$ функцию $\theta(t)$, полагая

$$\theta(t) = \frac{m!}{k_1!\ldots k_\ell!},$$

где $k_1,\ldots,k_\ell$ – мощности всех подмножеств, состоящих из одинаковых элементов набора $t_1,\ldots,t_m$.

Теорема 2. *Пусть* $n, k \in \mathbb{N}$, *причем* $1 \leqslant k < n$, *и пусть* $N = 2^n$, $m = 2^k$, $s = 2^{n-k} - 1$. *Тогда*

$$Z\big(W_N^{(N-m+1)}\big) = \\ = 2 \sum_{t_1=0}^{s} \dots \sum_{t_m=t_{m-1}}^{s} \binom{s}{t_1} \dots \binom{s}{t_m} \theta(t) \sqrt{\left[\sum_{z \in \mathbb{Z}_{2,\mathrm{e}}^m} H(t,z)\right]\left[\sum_{z \in \mathbb{Z}_{2,\mathrm{o}}^m} H(t,z)\right]}.$$

*Замечание* 1. Несложно видеть, что число слагаемых во внешних суммах в совокупности равно числу сочетаний с повторениями из $s+1$ по $m$, т.е. $\binom{s+m}{m}$, а суммы под корнем содержат по $2^{m-1}$ слагаемых каждая.

*Замечание* 2. Пусть $N = 2^n$ и $m = 2^k + \ell$, где $1 \leqslant k < n$ и $0 < \ell < 2^k$. Для $j = 1, 2, \dots, 2^{k+1}$ обозначим через $z_j$ вектор, у которого на позициях с номерами $(j-1)2^{n-k-1} + 1, \dots, j2^{n-k-1}$ стоят единицы, а остальные координаты равны нулю. Допустим, что существует эффективно вычислимая функция $\varkappa_m$, принимающая значение 1, если вектор $\alpha_1 z_1 + \dots + \alpha_{2^{k+1}} z_{2^{k+1}}$, $\alpha_j \in \{0,1\}$, принадлежит подпространству, порожденному последними $m$ строками матрицы $G_N$, и равная нулю в противном случае. Тогда справедлив следующий аналог формулы из теоремы 2:

$$Z\big(W_N^{(N-m+1)}\big) = 2 \sum_{t_1=0}^{s'} \dots \sum_{t_{m'}=t_{m'-1}}^{s'} \binom{s'}{t_1} \dots \binom{s'}{t_{m'}} \theta(t) \sqrt{\left[\sum_{z} H(t,z)\right]\left[\sum_{z} H(t,z)\right]},$$

где $m' = 2^{k+1}$, $s' = 2^{n-k-1} - 1$, $t = (t_1, \dots, t_{m'})$, а векторы $z$ пробегают множество $\mathbb{Z}_2^{m'}$ и удовлетворяют условию $\varkappa_{m-1}(z) = 1$ для первой суммы и условиям $\varkappa_m(z) = 1$ и $\varkappa_{m-1}(z) = 0$ для второй.

Далее, положим

$$tz = \sum_{j=1}^{m} (-1)^{z_j} t_j$$

и

$$F(t,z) = 2^{(s+1)w(z)+tz} p^{Nw(z)/2m+tz} (1-p)^{N-Nw(z)/2m-2(s+1)(m-w(z))+tz} \times \\ \times \left(p^2 + (1-p)^2\right)^{(s+1)(m-w(z))-tz}.$$

Теорема 3. *Пусть* $n, k \in \mathbb{N}$, *причем* $1 \leqslant k < n-1$, *и пусть* $N = 2^n$, $m = 2^k$, $s = 2^{n-k-1} - 1$. *Тогда*

$$Z\big(W_N^{N/2-m+1}\big) = \\ = 2 \sum_{t_1=0}^{s} \dots \sum_{t_m=t_{m-1}}^{s} \binom{s}{t_1} \dots \binom{s}{t_m} \theta(t) \sqrt{\left[\sum_{z \in \mathbb{Z}_{2,\mathrm{e}}^m} F(t,z)\right]\left[\sum_{z \in \mathbb{Z}_{2,\mathrm{o}}^m} F(t,z)\right]}.$$

Легко видеть, что для вычислений по формуле из теоремы 3 потребуется порядка $\binom{2^{n-k-1}+2^k-1}{2^k} 2^{2^k}$ операций сложения. Также отметим, что все сказанное в замечании 2 можно с незначительными изменениями переформулировать для значений $i$ из множества $\{N/4+1, \dots, N/2-1\}$.

Для $i = N/2+1$ и $i = N/4+1$ формулы из теорем 2 и 3 удалось упростить.

Следствие 1. *Положим* $L_1 = p^2 + (1-p)^2$ *и* $L_2 = 2p(1-p)$. *Справедливы равенства*

$$Z\big(W_N^{(N/2+1)}\big) = 1 - \big(p^2+(1-p)^2\big)^{N/2} + \sqrt{\big(p^2+(1-p)^2\big)^N - \big(p^2-(1-p)^2\big)^N},$$
$$Z\big(W_N^{(N/4+1)}\big) = 1 - \big(L_1^2+L_2^2\big)^{N/4} + \sqrt{\big(L_1^2+L_2^2\big)^{N/2} - \big(L_1^2-L_2^2\big)^{N/2}}.$$

Таким образом, для фиксированного $N = 2^n$ теоремы 1–3 вместе с формулой (2) дают точные выражения для параметров Бхаттачарьи для $n^2+1$ каналов. Конечно, приведенные формулы являются не альтернативой, а дополнением к методу из [2] в тех случаях, когда расчеты по ним возможны. Также авторы выражают надежду на то, что дальнейшие исследования свойств поляризационной матрицы откроют возможности для построения новых серий, а изучение симметрий в найденных формулах позволит значительно упростить последние, в идеале – до такого вида, как в теореме 1 или следствии 1, в крайнем случае – до формул с полиномиальным числом слагаемых под корнем.

## § 2. Свойства поляризационной матрицы

Напомним, что поляризационная матрица $G_{2^n}$ является произведением перестановочной матрицы $B_N$ на $n$-ю кронекерову степень ядра Арикана $F^{\otimes n}$. В свою очередь, матрица $B_N$ раскладывается в произведение

$$B_N = R_N(E_2 \otimes R_{N/2}) \dots (E_{N/2} \otimes R_2),$$

где $R_{2^i}$, $i = 1, \dots, n$, – такая перестановочная матрица, что

$$(x_1, \dots, x_{2^i})R_{2^i} = (x_1, x_3, \dots, x_{2^i-1}, x_2, x_4, \dots, x_{2^i}),$$

$E_{2^i}$ – единичная матрица порядка $2^i$.

Обозначим через $f_{2^n}(i,j)$ и $g_{2^n}(i,j)$ элементы матриц $F^{\otimes n}$ и $G_{2^n}$ соответственно, стоящие на позициях $(i,j)$. В [3] для $i,j \in \{1,\dots,2^n\}$ доказано, что

$$f_{2^n}(i,j) = \begin{cases} 1, & \text{если } i-1 \succeq j-1, \\ 0 & \text{в противном случае,} \end{cases}$$

где условие $i-1 \succeq j-1$ выполняется, когда в $n$-разрядных двоичных записях чисел $i-1$ и $j-1$ каждый разряд числа $i-1$ больше соответствующего разряда числа $j-1$ или равен ему. Чтобы напомнить известную связь между элементами матриц $F^{\otimes n}$ и $G_{2^n}$ и установить ряд новых фактов, определим две функции, областью определения и областью значений которых является множество $\{0,\dots,2^n-1\}$. Положим

$$\mathrm{rev}_n(d_{n-1}\dots d_0) = d_0 \dots d_{n-1}$$

и

$$\mathrm{inv}_n(d_{n-1}\dots d_0) = \bar{d}_{n-1}\dots\bar{d}_0,$$

где $d_{n-1}\dots d_0$ – двоичная $n$-разрядная запись числа $k$, дополненная нулями, если для записи $k$ достаточно меньшего числа разрядов.

В [1, с. 3064] установлено равенство

$$g_{2^n}(i,j) = f_{2^n}(\mathrm{rev}_n(i-1)+1, j)$$

для любых $i,j \in \{1,\dots,2^n\}$. Имеют место также следующие утверждения.

Теорема 4. *Для любого $n \in \mathbb{N}$ матрица $G_{2^n}$ является персимметричной, т.е. симметричной относительно побочной диагонали*:

$$g_{2^n}(i,j) = g_{2^n}(2^n - j + 1, 2^n - i + 1)$$

*для любых $i, j \in \{1, \ldots, 2^n\}$.*

Доказательство. Имеем

$$g_{2^n}(i,j) = f_{2^n}(\mathrm{rev}_n(i-1) + 1, j),$$
$$g_{2^n}(2^n - j + 1, 2^n - i + 1) = f_{2^n}(\mathrm{rev}_n(2^n - j) + 1, 2^n - i + 1).$$

Используя соотношение $\mathrm{inv}_n(k) = 2^n - 1 - k$, справедливое для любого $k$ из множества $\{0, \ldots, 2^n - 1\}$, перепишем второе равенство в виде

$$g_{2^n}(2^n - j + 1, 2^n - i + 1) = f_{2^n}(\mathrm{rev}_n(\mathrm{inv}_n(j-1)) + 1, \mathrm{inv}_n(i-1) + 1).$$

Убедимся, что $\mathrm{rev}_n(i-1) \succeq j - 1$ тогда и только тогда, когда

$$\mathrm{rev}_n(\mathrm{inv}_n(j-1)) \succeq \mathrm{inv}_n(i-1).$$

Для двоичных $n$-разрядных записей $i - 1 = x_{n-1} \ldots x_0$ и $j - 1 = y_{n-1} \ldots y_0$ неравенства $x_0 \ldots x_{n-1} \succeq y_{n-1} \ldots y_0$ и $\bar{y}_0 \ldots \bar{y}_{n-1} \succeq \bar{x}_{n-1} \ldots \bar{x}_0$ эквивалентны. ▲

Следствие 2. *Для любых $n \in \mathbb{N}$ и $i, j \in \{1, \ldots, 2^n\}$ в $i$-й строке матрицы $G_{2^n}$ правее последней единицы расположено $\mathrm{rev}_n(\mathrm{inv}_n(i-1))$ нулей, а в $j$-м столбце матрицы $G_{2^n}$ выше первой единицы расположено $\mathrm{rev}_n(j-1)$ нулей.*

Доказательство. Так как $F^{\oplus n}$ является нижнетреугольной матрицей с единицами по главной диагонали, то ее $i$-я строка содержит $2^n - i$ нулей правее последней единицы. Из равенства $g_{2^n}(i,j) = f_{2^n}(\mathrm{rev}_n(i-1) + 1, j)$ заключаем, что в $i$-й строке матрицы $G_{2^n}$ правее последней единицы расположено

$$2^n - 1 - \mathrm{rev}_n(i-1) = \mathrm{inv}_n(\mathrm{rev}_n(i-1)) = \mathrm{rev}_n(\mathrm{inv}_n(i-1))$$

нулей.

Из персимметричности матрицы $G_{2^n}$ следует, что в ее $j$-м столбце выше первой сверху единицы расположено $\mathrm{rev}_n(\mathrm{inv}_n(2^n - j)) = \mathrm{rev}_n(j-1)$ нулей. ▲

Теорема 5. *Для любых $n \in \mathbb{N}$ и $i \in \{1, \ldots, 2^{n-1}\}$ строка с номером $2i$ матрицы $G_{2^n}$ представляет собой повторенную два раза строку с номером $i$ матрицы $G_{2^{n-1}}$, а строка с номером $2i - 1$ матрицы $G_{2^n}$ – строку с номером $i$ матрицы $G_{2^{n-1}}$, дополненную справа $2^{n-1}$ нулями.*

Доказательство. Ввиду соотношений

$$F^{\oplus n} = \begin{pmatrix} F^{\oplus(n-1)} & 0 \\ F^{\oplus(n-1)} & F^{\oplus(n-1)} \end{pmatrix}$$

и

$$g_{2^n}(k,j) = f_{2^n}(\mathrm{rev}_n(k-1) + 1, j), \quad k, j = 1, \ldots, 2^n,$$

достаточно для любого $i \in \{1, \ldots, 2^{n-1}\}$ доказать равенства

$$\mathrm{rev}_n(2i-1) - 2^{n-1} = \mathrm{rev}_{n-1}(i-1), \quad \mathrm{rev}_n(2i-2) = \mathrm{rev}_{n-1}(i-1).$$

Рассмотрим двоичную запись $i - 1 = d_{n-2} \ldots d_0$. Имеем

$$\mathrm{rev}_n(2i-2) = \mathrm{rev}_n(2(i-1)) = \mathrm{rev}_n(d_{n-2} \ldots d_0 0) = d_0 \ldots d_{n-2} = \mathrm{rev}_{n-1}(i-1),$$
$$\mathrm{rev}_n(2(i-1)+1) - 2^{n-1} = \mathrm{rev}_n(d_{n-2} \ldots d_0 1) - 2^{n-1} = d_0 \ldots d_{n-2} =$$
$$= \mathrm{rev}_{n-1}(i-1). \quad ▲$$

## § 3. Доказательство теорем 1–3

Зафиксируем еще несколько обозначений. Всюду далее $V$ – линейное пространство размерности $N = 2^n$ над полем $\mathbb{Z}_2$. Операцию сложения в $V$ будем обозначать так же, как и операцию сложения в поле, а именно $\oplus$. Для любых натуральных чисел $i, j$, удовлетворяющих условиям $1 \leqslant i \leqslant j \leqslant N$, через $U_i^j$ обозначим подпространство в $V$, образованное векторами, у которых равны нулю первые $i-1$ и последние $j+1$ координат. Также по определению положим $U_1^0 = U_{N+1}^N = \{0\}$. Далее, для $i, j, k \in \mathbb{N}$, таких что $1 \leqslant i \leqslant k \leqslant j \leqslant N$, через $u_i^j(k, \varepsilon)$ обозначим вектор из $U_i^j$, у которого $k$-я координата равна $\varepsilon$. Наконец, для произвольного числа $i \in \mathbb{N}$, $1 \leqslant i \leqslant N$, произвольных векторов $y \in V$, $u \in U_1^{i-1}$ и числа $\varepsilon \in \{0, 1\}$ положим

$$A(y, u, i, \varepsilon) = \sum_{u' \in U_{i+1}^N} W^N\big(y \,|\, (u \oplus u_i^i(i, \varepsilon) \oplus u')G_N\big)$$

и будем полагать $A(y, i, \varepsilon) = A(y, u, i, \varepsilon)$, если вектор $u$ нулевой.

Лемма 1. *Пусть $i \in \mathbb{N}$ и $1 \leqslant i \leqslant N$. Тогда*

$$Z\big(W_N^{(i)}\big) = 2 \sum_{y \in U_1^{i-1}} \sqrt{A(yG_N, i, 0)\, A(yG_N, i, 1)}.$$

Доказательство. Заметим, что $V = \{yG_N \,|\, y \in V\}$, поскольку матрица $G_N$ обратима; далее, $V = \{y \oplus u \,|\, y \in V\}$ для любого фиксированного вектора $u \in V$; наконец, для любых векторов $y, x, u \in V$ справедливо равенство $W^N\,(y \oplus u \,|\, x \oplus u) = W^N\,(y \,|\, x)$. Поэтому

$$\begin{aligned}
Z\big(W_N^{(i)}\big) &= \frac{1}{2^{N-1}} \sum_{y \in V} \sum_{u \in U_1^{i-1}} \sqrt{A(y, u, i, 0)\, A(y, u, i, 1)} = \\
&= \frac{1}{2^{N-1}} \sum_{y \in V} \sum_{u \in U_1^{i-1}} \sqrt{A(yG_N, u, i, 0)\, A(yG_N, u, i, 1)} = \\
&= \frac{1}{2^{N-1}} \sum_{u \in U_1^{i-1}} \left[ \sum_{y \in V} \sqrt{A((y \oplus u)G_N, u \oplus u, i, 0)\, A((y \oplus u)G_N, u \oplus u, i, 1)} \right] = \\
&= \frac{2^{i-1}}{2^{N-1}} \sum_{y \in V} \sqrt{A(yG_N, i, 0)\, A(yG_N, i, 1)}.
\end{aligned}$$

Для завершения доказательства остается заметить, что для любого вектора $u \in U_{i+1}^N$ имеем

$$A\big((y \oplus u)G_N, i, \varepsilon\big) = A(yG_N, i, \varepsilon),$$

и значит, суммирование по $V$ в последней сумме можно заменить, предварительно домножив ее на $2^{N-i}$, на суммирование по представителям фактор-пространства $V/U_{i+1}^N$, т.е. по $U_1^i$. Наконец, с учетом равенства $A((y \oplus 1_i^i)G_N, i, \varepsilon) = A(yG_N, i, \bar{\varepsilon})$ подпространство суммирования можно сократить до $U_1^{i-1}$, удвоив при этом сумму. ▲

Доказательство теоремы 1. Первая строка матрицы $G_N$ имеет вид $(1, 0, \ldots, 0)$, а остальные строки содержат четное число единиц. Поэтому множества $\{(0, u_2, \ldots, u_N)G_N\}$ и $\{(1, u_2, \ldots, u_N)G_N\}$ состоят из всех $N$-мерных векторов

с четным и нечетным числом единиц соответственно. Значит,

$$A(0,1,0)=\sum_{j=0}^{N/2}\binom{N}{2j}p^{2j}(1-p)^{N-2j},$$

$$A(0,1,1)=\sum_{j=1}^{N/2}\binom{N}{2j-1}p^{2j-1}(1-p)^{N-2j+1}.$$

Первая и вторая суммы являются, соответственно, суммами всех положительных и отрицательных слагаемых в разложении выражения $(-p+(1-p))^N$ в бином Ньютона, а обе вместе дают $(p+(1-p))^N$. Следовательно,

$$Z\big(W_N^{(1)}\big)=2\sqrt{A(0,1,0)A(0,1,1)}=$$
$$=2\sqrt{\big((1+(1-2p)^N\big)/2\big)\big((1-(1-2p)^N\big)/2\big)}=\sqrt{1-(1-2p)^{2N}}.$$

Докажем второе равенство теоремы 1. Имеем

$$A(yG_N,N,0)=W^N(yG_N\,|\,(0,\ldots,0))=p^{w(yG_N)}(1-p)^{N-w(yG_N)},$$
$$A(yG_N,N,1)=W^N(yG_N\,|\,(1,\ldots,1))=p^{N-w(yG_N)}(1-p)^{w(yG_N)}.$$

Отсюда

$$Z\big(W_N^{(N)}\big)=2\sum_{y\in U_1^{N-1}}\sqrt{A(yG_N,N,0)\cdot A(yG_N,N,1)}=$$
$$=2\cdot 2^{N-1}\sqrt{p^N(1-p)^N}=2^Np^{N/2}(1-p)^{N/2},$$

что завершает доказательство теоремы 1. ▲

В нижеследующих леммах 2–4 полагаем $N=2^n$, где $n>1$, и $m=2^k$, где $0\leqslant k<n$. Также полагаем $s=N/m-1$.

Лемма 2. *Первые $N-m$ строк матрицы $G_N$ порождают подпространство*

$$L_{N,m}=\Big\{(\underbrace{x_1,\ldots,x_{N/m-1},0}_{N/m},\ldots,\underbrace{x_{N-N/m+1},\ldots,x_{N-1},0}_{N/m})\mid x_i\in\mathbb{Z}_2\Big\}.$$

Доказательство. Поскольку матрица $G_N$ невырождена, ее первые $N-m$ строк линейно независимы и порождают подпространство размерности $N-m$. Мы докажем лемму, если установим, что в столбцах с номерами $N/m,2N/m,\ldots,mN/m$ первые сверху $N-m$ элементов равны нулю. Согласно следствию 2 для этого достаточно доказать неравенство

$$\mathrm{rev}_n(tN/m-1)\geqslant N-m=2^n-2^k\quad\text{для}\quad t=1,2,\ldots,m.$$

Зафиксируем целое число $t$, $1\leqslant t\leqslant m$, и пусть $t-1=d_{k-1}\ldots d_0$ – двоичная запись числа $t-1$. Из тождества

$$2^{n-k}(t-1)+2^{n-k}-1=t2^{n-k}-1$$

следует, что $t2^{n-k}-1=d_{k-1}\ldots d_01\ldots1$, здесь справа от $d_0$ расположено $n-k$ единиц. Отсюда

$$\mathrm{rev}_n(tN/m-1)=1\ldots1d_0\ldots d_{k-1}=(2^{n-k}-1)2^k+C=2^n-2^k+C,$$

где $C=d_0\ldots d_{k-1}\geqslant 0$. ▲

Лемма 3. *Последние $m$ строк матрицы $G_N$ порождают подпространство*

$$R_{N,m} = \{(z_1, \dots, z_1, \dots, z_m, \dots, z_m) \mid z_1, \dots, z_m \in \mathbb{Z}_2\},$$

*здесь каждая переменная $z_i$ повторяется $N/m$ раз. Последние $m-1$ строк матрицы $G_N$ порождают в $R_{N,m}$ подпространство $R^{\mathrm{e}}_{N,m}$ с четной суммой $z_1 + \ldots + z_m$. Линейные комбинации строки $N-m+1$ со строками с бóльшими номерами образуют в $R_{N,m}$ подмножество $R^{\mathrm{o}}_{N,m}$ с нечетной суммой $z_1 + \ldots + z_m$, или, иначе говоря,*

$$R^{\mathrm{o}}_{N,m} = (\underbrace{1, \dots, 1}_{N/m}, 0, \dots, 0) \oplus R^{\mathrm{e}}_{N,m}.$$

Доказательство. Когда $m = 1$ или $n = 1$, утверждение очевидно. Далее воспользуемся индукцией по $n$. Рассмотрим матрицу $G_N$, $N = 2^n > 2$, и пусть $m > 1$. По предположению индукции последние $m/2$ строк матрицы $G_{N/2}$ порождают подпространство $R_{N/2,m/2}$, векторы которого имеют вид $(z_1, \dots, z_1, \dots, z_{m/2}, \dots, z_{m/2})$, где каждая переменная $z_i$ повторяется $(N/2)/(m/2) = N/m$ раз. Применяя теорему 5, получаем, что каждая из последних $m$ строк матрицы $G_N$ содержится в $R_{N,m}$. Учитывая, что размерность пространства $R_{N,m}$ равна $m$, а последние $m$ строк матрицы $G_N$ линейно независимы, получаем, что они порождают $R_{N,m}$. По тем же соображениям последние $m-1$ строк матрицы $G_N$ порождают подпространство $R^{\mathrm{e}}_{N,m-1}$. Кратное применение теоремы 5 показывает, что строка с номером $N-m+1$ имеет вид $(1, \dots, 1, 0 \dots, 0)$, где единица повторяется $N/m$ раз. ▲

Зафиксируем произвольный вектор $y \in U_1^{N-m}$, и пусть $x = yG_N$. По лемме 2 имеем $x \in L_{N,m}$, и следовательно, $x = (x_1, \dots, x_s, 0, \dots, x_{N-s}, \dots, x_{N-1}, 0)$. Обозначим через $t_i$, $1 \leqslant i \leqslant m$, число единиц среди координат $x_{(i-1)N/m+1}, \dots, x_{(i-1)N/m+s}$ вектора $x$.

Лемма 4. *Имеют место равенства*

$$A(yG_N, N-m+1, 0) = \sum_{z \in \mathbb{Z}^m_{2,\mathrm{e}}} H(t,z), \quad A(yG_N, N-m+1, 1) = \sum_{z \in \mathbb{Z}^m_{2,\mathrm{o}}} H(t,z).$$

Доказательство. Пусть $i = N-m+1$. Зафиксируем вектор $u \in U^N_{i+1}$, и пусть $v = u^i_i(i,\varepsilon) \oplus u$, где $\varepsilon \in \{0,1\}$. По лемме 3 имеем $vG_N = (z_1, \dots, z_1, \dots, z_m, \dots, z_m)$, где $z_1, \dots, z_m \in \mathbb{Z}_2$ и каждая переменная повторяется $N/m$ раз. Для фиксированного $j$, $1 \leqslant j \leqslant m$, количество отличных от $z_j$ чисел среди $x_{(j-1)N/m+1}, \dots, x_{(j-1)N/m+s}, 0$ равно $z_j N/m + (\bar{z}_j - z_j)t_j$. Отсюда следует, что количество различных координат у векторов $yG_N$ и $vG_N$ выражается числом $r(t,z)$, где $t = (t_1, \dots, t_m)$, $z = (z_1, \dots, z_m)$, а количество одинаковых равно $N - r(t,z)$. Значит, $W^N(yG_N \,|\, vG_N) = H(t,z)$. Для завершения доказательства остается заметить, что согласно лемме 3, когда $u$ пробегает подпространство $U^N_{i+1}$ и $\varepsilon = 0$, вектор $vG_N$ пробегает подпространство $R^{\mathrm{e}}_{N,m}$, а соответствующий ему вектор $z$ пробегает множество $\mathbb{Z}^m_{2,\mathrm{e}}$. Если $\varepsilon = 1$, то $vG_N$ пробегает множество $R^{\mathrm{o}}_{N,m}$, а соответствующий ему вектор $z$ – множество $\mathbb{Z}^m_{2,\mathrm{o}}$. ▲

Доказательство теоремы 2. Очевидно, что каждое из чисел $t_1, \dots, t_m$ может изменяться от 0 до $s$. Далее, для фиксированных $t_1, \dots, t_m$ в $L_{N,m}$ существует $\binom{s}{t_1} \dots \binom{s}{t_m}$ векторов с $t_j$ единицами среди координат $x_{(j-1)N/m+1}, \dots, x_{(j-1)N/m+s}$. Поэтому из лемм 1 и 4 следует равенство

$$Z\big(W_N^{(N-m+1)}\big) = 2 \sum_{t_1=0}^{s} \dots \sum_{t_m=0}^{s} \binom{s}{t_1} \dots \binom{s}{t_m} \sqrt{\left[\sum_{z \in \mathbb{Z}^m_{2,\mathrm{e}}} H(t,z)\right]\left[\sum_{z \in \mathbb{Z}^m_{2,\mathrm{o}}} H(t,z)\right]}.$$

Упростим полученную формулу до вида, указанного в теореме 2. Заметим, что при перестановке чисел $t_1, \ldots, t_m$ не меняется произведение $\binom{s}{t_1} \ldots \binom{s}{t_m}$ и не меняются суммы, стоящие под корнем. Количество различных перестановок, которые можно получить из чисел $t_1, \ldots, t_m$, очевидно, равно $\theta(t_1, \ldots, t_m)$. В связи с этим определим на декартовой степени $\{0, 1, \ldots, s\}^m$ отношение эквивалентности $\sim$, считая $(t_1, \ldots, t_m) \sim (t'_1, \ldots, t'_m)$, если один набор можно путем перестановки компонент привести ко второму. Ввиду биективности отображения

$$\{(t_1, \ldots, t_m) \mid 0 \leqslant t_1 \leqslant \ldots \leqslant t_m \leqslant s\} \to \{0, \ldots, s\}^m / \sim$$

нижний предел в $j$-й сумме можно заменить на $t_{j-1}$ для $j = 2, \ldots, m$, что завершает доказательство теоремы 2. ▲

Прежде чем приступить к доказательству теоремы 3, отметим следующий факт, вытекающий из теоремы 5. Для любого натурального числа $i$, $1 \leqslant i \leqslant N/2$, строка с номером $i + N/2$ матрицы $G_N$ получается добавлением по одной единице справа к каждой группе отдельно стоящих единиц ее $i$-й строки. Следующие три леммы непосредственно следуют из этого замечания и лемм 2–4. До конца этого параграфа полагаем $N = 2^n$ и $n > 2$; $m = 2^k$ и $1 \leqslant k \leqslant n - 2$; $s = 2^{n-k-1} - 1$.

Лемма 5. *Первые $N/2 - m$ строк матрицы $G_N$ порождают подпространство*

$$L'_{N,m} = \big\{x = (\ldots, x_1^\ell, 0, x_2^\ell, 0, \ldots, x_{s-1}^\ell, 0, x_s^\ell, 0, 0, 0, \ldots) \mid \ell = 1, \ldots, m,\ x_j^\ell \in \mathbb{Z}_2\big\}.$$

Лемма 6. *Строки матрицы $G_N$ с номерами $N/2 - m + 1, \ldots, N/2$ порождают подпространство*

$$\widetilde{R}_{N,m} = \big\{\widetilde{z} = (\ldots, z_\ell, 0, z_\ell, 0, \ldots, z_\ell, 0, \ldots) \mid \ell = 1, \ldots, m,\ z_\ell \in \mathbb{Z}_2\big\}.$$

*Строки с номерами $N/2 - m + 2, \ldots, N/2$ порождают в $\widetilde{R}_{N,m}$ подпространство $\widetilde{R}^{\mathrm{e}}_{N,m}$ с четной суммой $z_1 + \ldots + z_m$. Линейные комбинации строки $N/2 - m + 1$ с элементами из $\widetilde{R}^{\mathrm{e}}_{N,m}$ образуют в $\widetilde{R}_{N,m}$ подмножество $\widetilde{R}^{\mathrm{o}}_{N,m}$ с нечетной суммой $z_1 + \ldots + z_m$.*

Лемма 7. *Строки матрицы $G_N$ с номерами $N/2 + 1, \ldots, N$ порождают подпространство*

$$L_{N,2} = \big\{u = \big(\ldots, u_1^\ell, u_1^\ell, u_2^\ell, u_2^\ell, \ldots, u_{s+1}^\ell, u_{s+1}^\ell, \ldots\big) \mid \ell = 1, \ldots, m,\ u_j^\ell \in \mathbb{Z}_2\big\}.$$

Обозначим через $x^\ell$ и $u^\ell$ $\ell$-й блок координат векторов $x$ и $u$ соответственно, а каждому вектору $\widetilde{z} = (\ldots, z_\ell, 0, z_\ell, 0, \ldots, z_\ell, 0, \ldots) \in \widetilde{R}_{N,m}$ поставим в соответствие вектор $z = (z_1, \ldots, z_m) \in \mathbb{Z}_2^m$. Несложно проверить, что число несовпадений координат у векторов $(x', 0)$ и $(z' \oplus u', u')$ выражается суммой $z' + u' + (-1)^{z'} u' + (-1)^{u'+z'} x'$. Отсюда вытекает следующая

Лемма 8. *Число несовпадений координат у векторов $x$ и $\widetilde{z} \oplus u$ выражается суммой*

$$\gamma(x, z, u) = w(z)\frac{N}{2m} + w(u) + \sum_{\ell=1}^{m} (-1)^{z_\ell} w(u^\ell) + \sum_{\ell=1}^{m} \sum_{j=1}^{s} (-1)^{u_j^\ell + z_\ell} x_j^\ell. \tag{3}$$

Доказательство теоремы 3. Из лемм 1 и 5–8 следует равенство

$$Z\big(W_N^{(N/2-m+1)}\big) = 2 \sum_{x \in L'_{N,m}} \sqrt{\left[\sum_{z \in \mathbb{Z}^m_{2,\mathrm{e}}} \widetilde{H}(x, z)\right]\left[\sum_{z \in \mathbb{Z}^m_{2,\mathrm{o}}} \widetilde{H}(x, z)\right]}, \tag{4}$$

где

$$\widetilde{H}(x,z)=\sum_{u\in R_{N,2}} p^{\gamma(x,z,u)}(1-p)^{N-\gamma(x,z,u)}. \tag{5}$$

Глядя на формулы (3)–(5), видим, что суммы под корнем в (4) зависят только от количества $t_\ell$ ненулевых координат вектора $x$ внутри блока $x^\ell$, но не зависят от их расположения. Поэтому формула (4) упрощается до следующей:

$$Z\big(W_N^{(N/2-m+1)}\big)=$$
$$=2\sum_{t_1=0}^{s}\dots\sum_{t_m=t_{m-1}}^{s}\binom{s}{t_1}\dots\binom{s}{t_m}\theta(t)\sqrt{\left[\sum_{z\in\mathbb{Z}_{2,\mathrm{e}}^m}\widetilde{H}(t,z)\right]\left[\sum_{z\in\mathbb{Z}_{2,\mathrm{o}}^m}\widetilde{H}(t,z)\right]}, \tag{6}$$

где $t=(t_1,\dots,t_m)$.

Упростим $\widetilde{H}(t,z)$. Обозначим через $a_\ell$ и $b_\ell$ уменьшенное в два раза число единиц в блоке $u^\ell$ вектора $u$, расположенных от первой координаты до координаты $2t_\ell$ и от $2t_\ell+1$ до $2s+2$ соответственно. Пусть $a=(a_1,\dots,a_m)$ и $b=(b_1,\dots,b_m)$. Тогда

$$\gamma(x,z,u)=\gamma(t,z,a,b)=w(z)\frac{N}{2m}+w(a)+w(b)+\sum_{j=1}^{m}(-1)^{z_j}(t_j-a_j+b_j)$$

и

$$\widetilde{H}(t,z)=\sum_{a_1=0}^{t_1}\dots\sum_{a_m=0}^{t_m}\sum_{b_1=0}^{s+1-t_1}\dots\sum_{b_m=0}^{s+1-t_m}\binom{t_1}{a_1}\dots\binom{t_m}{a_m}\times$$
$$\times\binom{s+1-t_1}{b_1}\dots\binom{s+1-t_m}{b_m}p^{\gamma(t,z,a,b)}(1-p)^{N-\gamma(t,z,a,b)}.$$

Зафиксируем вектор $z$, и пусть $z_{\alpha_1}=\dots=z_{\alpha_q}=0$, а $z_{\alpha_{q+1}}=\dots=z_{\alpha_m}=1$. Здесь $\alpha_1,\dots,\alpha_m$ – некоторая перестановка чисел $1,\dots,m$. Тогда

$$\gamma(t,z,a,b)=w(z)\frac{N}{2m}+2h+tz,$$

где $h=a_{\alpha_{q+1}}+\dots+a_{\alpha_m}+b_{\alpha_1}+\dots+b_{\alpha_q}$. Величина $\gamma(t,z,a,b)$ не зависит от групп переменных $a_{\alpha_1},\dots,a_{\alpha_q}$ и $b_{\alpha_{q+1}},\dots,b_{\alpha_m}$. Поэтому $\widetilde{H}(t,z)$ раскладывается в произведение числа $p^{w(z)N/2m+tz}(1-p)^{N-w(z)N/2m-tz}$, сумм

$$\sum_{a_{\alpha_1}=0}^{t_{\alpha_1}}\binom{t_{\alpha_1}}{a_{\alpha_1}},\ \dots,\ \sum_{a_{\alpha_q}=0}^{t_{\alpha_q}}\binom{t_{\alpha_q}}{a_{\alpha_q}},\ \sum_{b_{\alpha_{q+1}}=0}^{s+1-t_{\alpha_{q+1}}}\binom{s+1-t_{\alpha_{q+1}}}{b_{\alpha_{q+1}}},\ \dots,$$
$$\dots,\ \sum_{b_{\alpha_m}=0}^{s+1-t_{\alpha_m}}\binom{s+1-t_{\alpha_m}}{b_{\alpha_m}},$$

произведение которых равно $2^{(s+1)(m-q)+tz}$, и суммы

$$S=\sum_{a_{\alpha_{q+1}}=0}^{t_{\alpha_{q+1}}}\dots\sum_{a_{\alpha_m}=0}^{t_{\alpha_m}}\sum_{b_{\alpha_1}=0}^{s+1-t_{\alpha_1}}\dots\sum_{b_{\alpha_q}=0}^{s+1-t_{\alpha_q}}\binom{t_{\alpha_{q+1}}}{a_{\alpha_{q+1}}}\dots\binom{t_{\alpha_m}}{a_{\alpha_m}}\times$$
$$\times\binom{s+1-t_{\alpha_1}}{b_{\alpha_1}}\dots\binom{s+1-t_{\alpha_q}}{b_{\alpha_q}}p^{2h}(1-p)^{-2h}.$$

Вычислим сумму $S$. Показатели степеней $p$ и $1-p$ зависят от величины $h$, которая принимает все целые значения от 0 до $(s+1)q-tz$. Используя обобщенную свертку Вандермонда и формулу бинома Ньютона, находим

$$S = \sum_{h=0}^{(s+1)q-tz} p^{2h}(1-p)^{-2h} \times$$
$$\times \sum_{a_{\alpha_{q+1}}+\ldots+a_{\alpha_m}+b_{\alpha_1}+\ldots+b_{\alpha_q}=h} \binom{t_{\alpha_{q+1}}}{a_{\alpha_{q+1}}} \ldots \binom{t_{\alpha_m}}{a_{\alpha_m}} \binom{s+1-t_{\alpha_1}}{b_{\alpha_1}} \ldots \binom{s+1-t_{\alpha_q}}{b_{\alpha_q}} =$$
$$= (1-p)^{-2(s+1)q+2tz} \sum_{h=0}^{(s+1)q-tz} \binom{(s+1)q-tz}{h} (p^2)^h \big((1-p)^2\big)^{(s+1)q-tz-h} =$$
$$= (1-p)^{-2(s+1)q+2tz} \big(p^2+(1-p)^2\big)^{(s+1)q-tz}.$$

Таким образом,

$$\widetilde{H}(t,z) = 2^{(s+1)(m-q)+tz} p^{w(z)N/2m+tz} (1-p)^{N-w(z)N/2m-2(s+1)q+tz} \times$$
$$\times \big(p^2+(1-p)^2\big)^{(s+1)q-tz} = 2^{(s+1)w(z)+tz} p^{w(z)N/2m+tz} \times$$
$$\times (1-p)^{N-w(z)N/2m-2(s+1)(m-w(z))+tz} \big(p^2+(1-p)^2\big)^{(s+1)w(z)-tz} = H(t,z),$$

что завершает доказательство теоремы 3. ▲

## § 4. Доказательство следствия 1

Сперва докажем первую формулу. Имеем $k=n-1$, $m=N/2$, $s=1$, $i=N/2+1$. Из вида функции $r(t,z)$ следует, что величины

$$H_{\mathrm{e}}(t_1,\ldots,t_m) = \sum_{z\in\mathbb{Z}^m_{2,\mathrm{e}}} H(t,z), \quad H_{\mathrm{o}}(t_1,\ldots,t_m) = \sum_{z\in\mathbb{Z}^m_{2,\mathrm{o}}} H(t,z)$$

зависят только от количества единиц среди чисел $t_1,\ldots,t_m$, но не от их расположения, поэтому

$$Z\big(W_N^{(N/2+1)}\big) = \sum_{j=0}^{N/2} \binom{N/2}{j} \sqrt{H_{\mathrm{e}}(e_j) H_{\mathrm{o}}(e_j)},$$

где $e_j$ – $N/2$-мерный $(0,1)$-вектор с $j$ единицами в начале. Вычислим $H_{\mathrm{e}}(e_j)$, когда $j>0$. Имеем

$$r(e_j,z) = 2w(z) + (\bar{z}_1 - z_1) + \ldots + (\bar{z}_j - z_j) = j + 2(z_{j+1}+\ldots+z_{N/2}).$$

Величина $j+2(z_{j+1}+\ldots+z_{N/2})$ принимает значения $j+0$, $j+2$, …, $j+2(N/2-j) = N-j$, причем значение $j+2\ell$, $0 \leqslant \ell \leqslant N/2-j$, принимается на следующем количестве векторов $z$ при четном и нечетном $\ell$ соответственно:

$$\binom{N/2-j}{\ell}\left[\binom{j}{0}+\binom{j}{2}+\ldots\right] = \binom{N/2-j}{\ell}\left[\binom{j}{1}+\binom{j}{3}+\ldots\right] =$$
$$= 2^{j-1}\binom{N/2-j}{\ell}.$$

Следовательно,

$$H_{\mathrm{e}}(e_j) = \sum_{\ell=0}^{N/2-j} 2^{j-1}\binom{N/2-j}{\ell} p^{j+2\ell}(1-p)^{N-j-2\ell}.$$

Аналогичные рассуждения приводят к равенству $H_{\mathrm{o}}(e_j) = H_{\mathrm{e}}(e_j)$. Наконец, легко видеть, что

$$H_{\mathrm{e}}(e_0) = \sum_{\ell=0}^{N/4} \binom{N/2}{2\ell} p^{4\ell}(1-p)^{N-4\ell}, \quad H_{\mathrm{e}}(e_0) = \sum_{\ell=0}^{N/4-1} \binom{N/2}{2\ell+1} p^{4\ell+2}(1-p)^{N-4\ell-2},$$

откуда

$$H_{\mathrm{e}}(e_0) = \frac{Q_1^{N/2} + Q_2^{N/2}}{2}, \quad H_{\mathrm{o}}(e_0) = \frac{Q_1^{N/2} - Q_2^{N/2}}{2},$$

где $Q_1 = p^2 + (1-p)^2$ и $Q_2 = p^2 - (1-p)^2$. Поэтому

$$Z\big(W_N^{(N/2+1)}\big) = \sqrt{Q_1^N - Q_2^N} + \sum_{j=1}^{N/2} \sum_{\ell=0}^{N/2-j} 2^j \binom{N/2}{j}\binom{N/2-j}{\ell} p^{j+2\ell}(1-p)^{N-j-2\ell}.$$

Величина $j + 2\ell$ изменяется при указанных ограничениях на изменение индексов $j$ и $\ell$ от 1 до $N-1$. Положим $j' = j + 2\ell$. Коэффициент при $p^{j'}(1-p)^{N-j'}$ равен

$$\sum_{\ell=0}^{[(j'-1)/2]} 2^{j'-2\ell}\binom{N/2}{j'-2\ell}\binom{N/2-j'+2\ell}{\ell},$$

где $[\,\cdot\,]$ – целая часть числа. Используя метод коэффициентов [4] (авторы благодарят Г.П. Егорычева за проведенные вычисления), находим

$$\sum_{\ell=0}^{[(j'-1)/2]} 2^{j-2\ell}\binom{N/2}{j'-2\ell}\binom{N/2-j'+2\ell}{\ell} = \begin{cases} \binom{N}{j'}, & \text{если } j'-1 \text{ четно,} \\ \binom{N}{j'} - \binom{N/2}{j'/2} & \text{в противном случае.} \end{cases}$$

Поэтому

$$\begin{aligned} &Z\big(W_N^{(N/2+1)}\big) = \\ &= \sqrt{Q_1^N - Q_2^N} + \sum_{j'=1}^{N-1}\binom{N}{j'} p^{j'}(1-p)^{N-j'} - \sum_{j'=1}^{N/2-1}\binom{N/2}{j'} p^{2j'}(1-p)^{N-2j'} = \\ &= \sqrt{Q_1^N - Q_2^N} + (p+1-p)^N - (1-p)^N - p^N - \big(p^2+(1-p)^2\big)^{N/2} + p^N + \\ &+ (1-p)^N = 1 - \big(p^2+(1-p)^2\big)^{N/2} + \sqrt{\big(p^2+(1-p)^2\big)^N - \big(p^2-(1-p)^2\big)^N}. \end{aligned}$$

Теперь докажем вторую формулу следствия. Пусть $N = 2^n > 4$, $m = N/4$, $s = 1$. Так как числа $t_j$ в рассматриваемом случае равны 0 или 1, то суммы

$$\sum_{z\in\mathbb{Z}_{2,\mathrm{e}}^m} F(t,z) \quad \text{и} \quad \sum_{z\in\mathbb{Z}_{2,\mathrm{o}}^m} F(t,z) \tag{7}$$

зависят от количества чисел $t_j$, равных 1, но не от их расположения. Обозначим через $t_0$ число единиц в векторе $t = (t_1, \ldots, t_m)$ и будем считать, что они расположены в начале. Тогда

$$tz = \sum_{j=1}^{m} (-1)^{z_j} t_j = \sum_{j=1}^{t_0} (-1)^{z_j},$$

где при $t_0 = 0$ сумму считаем равной нулю. Далее, заметим, что $F(t,z) = F(t,z')$, если в первых $t_0$ координатах векторов $z$ и $z'$ расположено одинаковое число единиц и в последних $m - t_0$ координатах тоже расположено одинаковое число единиц. Обозначим количества этих единиц через $k_1$ и $k_2$ соответственно. Тогда $w(z) = k_1 + k_2$ и $tz = (-1)^0(t_0 - k_1) + (-1)^{-1}k_1 = t_0 - 2k_1$. Отсюда $(s+1)w(z) + tz = t_0 + 2k_2$, $w(z)N/2m + tz = t_0 + 2k_2$ и

$$N - w(z)N/2m - 2(s+1)(m - w(z)) + tz = t_0 + 2k_2,$$
$$(s+1)(m - w(z)) - tz = N/2 - t_0 - 2k_2.$$

Следовательно, $F(t,z) = Q_1^{N/2} Q_2^{t_0 + 2k_2}$, где $Q_1 = p^2 + (1-p)^2$ и $Q_2 = 2p(1-p)/Q_1$.

Рассмотрим выражение

$$Q_1^{N/2} Q_2^{t_0} \sum_{k_1=0}^{t_0} \sum_{k_2=0}^{m-t_0} \binom{t_0}{k_1} \binom{m-t_0}{k_2} Q_2^{2k_2}. \tag{8}$$

Если суммирование в (8) вести только по таким $k_1$ и $k_2$, что сумма $k_1 + k_2$ четна, то мы получим первую из сумм (7), а если нечетна, то вторую.

Пусть $t_0 = 0$. Тогда

$$\sum_{z \in \mathbb{Z}_{2,\mathrm{e}}^m} F(t,z) = Q_1^{\frac{N}{2}} \sum_{k_1=0}^{0} \binom{0}{2k_1} \sum_{k_2=0}^{N/8} \binom{N/4}{2k_2} Q_2^{4k_2} = \frac{\left(1+Q_2^2\right)^{\frac{N}{4}} + \left(1-Q_2^2\right)^{\frac{N}{4}}}{2} Q_1^{\frac{N}{2}},$$

$$\sum_{z \in \mathbb{Z}_{2,\mathrm{o}}^m} F(t,z) = Q_1^{\frac{N}{2}} \sum_{k_1=0}^{0} \binom{0}{2k_1} \sum_{k_2=0}^{N/8-1} \binom{N/4}{2k_2+1} Q_2^{4k_2+2} =$$

$$= \frac{\left(1+Q_2^2\right)^{\frac{N}{4}} - \left(1-Q_2^2\right)^{\frac{N}{4}}}{2} Q_1^{\frac{N}{2}}.$$

При $0 < t_0 < N/4$ имеем

$$\sum_{z \in \mathbb{Z}_{2,\mathrm{e}}^m} F(t,z) = Q_1^{\frac{N}{2}} Q_2^{t_0} \times$$

$$\times \left[ \sum_{k_1=0}^{\left[\frac{t_0}{2}\right]} \binom{t_0}{2k_1} \sum_{k_2=0}^{\left[\frac{m-t_0}{2}\right]} \binom{m-t_0}{2k_2} Q_2^{4k_2} + \sum_{k_1=0}^{\left[\frac{t_0}{2}\right]} \binom{t_0}{2k_1+1} \sum_{k_2=0}^{\left[\frac{m-t_0}{2}\right]} \binom{m-t_0}{2k_2+1} Q_2^{4k_2+2} \right] =$$

$$= Q_1^{\frac{N}{2}} Q_2^{t_0} 2^{t_0-1} \sum_{k_2=0}^{\frac{N}{4}-t_0} \binom{N/4-t_0}{k_2} Q_2^{2k_2} = Q_1^{\frac{N}{2}} Q_2^{t_0} 2^{t_0-1} \left(1+Q_2^2\right)^{\frac{N}{4}-t_0}$$

и аналогично

$$\sum_{z \in \mathbb{Z}_{2,\mathrm{o}}^m} F(t,z) = Q_1^{\frac{N}{2}} Q_2^{t_0} 2^{t_0-1} \left(1+Q_2^2\right)^{\frac{N}{4}-t_0}.$$

Наконец, если $t_0 = N/4$, то

$$\sum_{z\in\mathbb{Z}^m_{2,\mathrm{e}}} F(t,z) = \sum_{z\in\mathbb{Z}^m_{2,\mathrm{o}}} F(t,z) = Q_1^{\frac{N}{2}} Q_2^{\frac{n}{4}} 2^{\frac{N}{4}-1}.$$

Отсюда

$$\begin{aligned}
&Z\big(W_N^{(N/4+1)}\big) = Q_1^{N/2}\sqrt{\big(1+Q_2^2\big)^{N/2} - \big(1-Q_2^2\big)^{N/2}} + \\
&+ Q_1^{N/2} \sum_{t_0=1}^{N/4-1} \binom{N/4}{t_0} (2Q_2)^{t_0} \big(1+Q_2^2\big)^{N/4-t_0} + Q_1^{N/2} Q_2^{N/4} 2^{N/4} = \\
&= Q_1^{\frac{N}{2}} \left[\sqrt{\big(1+Q_2^2\big)^{\frac{N}{2}} - \big(1-Q_2^2\big)^{\frac{N}{2}}} + (1+Q_2)^{\frac{N}{2}} - \big(1+Q_2^2\big)^{\frac{N}{4}} - (2Q_2)^{\frac{N}{4}} + \right. \\
&\left. + (2Q_2)^{\frac{N}{4}}\right] = Q_1^{\frac{N}{2}} \left[\sqrt{\big(1+Q_2^2\big)^{\frac{N}{2}} - \big(1-Q_2^2\big)^{\frac{N}{2}}} + (1+Q_2)^{\frac{N}{2}} - \big(1+Q_2^2\big)^{\frac{N}{4}}\right].
\end{aligned}$$

Внося в скобки $Q_1^{N/2}$ и замечая, что $Q_1 + Q_1Q_2 = 1$, $Q_1 = L_1$ и $Q_1Q_2 = L_2$, получаем требуемую формулу. ▲

В заключение авторы выражают благодарность рецензенту за полезные ссылки и замечания, позволившие значительно упростить доказательства утверждений из § 2 и улучшить текст статьи в целом.

## СПИСОК ЛИТЕРАТУРЫ

1. *Arıkan E.* Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels // IEEE Trans. Inform. Theory. 2009. V. 55. № 7. P. 3051–3073. https://doi.org/10.1109/TIT.2009.2021379
2. *Tal I., Vardy A.* How to Construct Polar Codes // IEEE Trans. Inform. Theory. 2013. V. 59. № 10. P. 6542–6582. https://doi.org/10.1109/TIT.2013.2272694
3. *Sarkis G., Tal I., Giard P., Vardy A., Thibeault C., Gross W.J.* Flexible and Low-Complexity Encoding and Decoding of Systematic Polar Codes // IEEE Trans. Commun. 2016. V. 64. № 7. P. 2732–2745. https://doi.org/10.1109/TCOMM.2016.2574996
4. *Егорычев Г.П.* Интегральное представление и вычисление комбинаторных сумм. Новосибирск: Наука, 1977.

*Колесников Сергей Геннадьевич*
*Леонтьев Владимир Маркович*
Институт математики и фундаментальной информатики
Сибирского федерального университета, Красноярск,
кафедра алгебры и математической логики
sklsnkv@mail.ru
v.m.leontiev@outlook.com

Поступила в редакцию
23.08.2022
После доработки
01.02.2023
Принята к публикации
08.02.2023

**ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ**

**Том 59** **2023** **Вып. 1**

УДК 621.391 : 519.72

© 2023 г. **М. Фернандес[1], Г.А. Кабатянский[2], С.А. Круглик[3], И. Мяо[4]**

# КОДЫ ДЛЯ ТОЧНОГО НАХОЖДЕНИЯ НОСИТЕЛЯ РАЗРЕЖЕННОГО ВЕКТОРА ПО ОШИБОЧНЫМ ЛИНЕЙНЫМ ИЗМЕРЕНИЯМ И ИХ ДЕКОДИРОВАНИЕ

Построены коды, позволяющие точно находить носитель неизвестного разреженного вектора, у которого модули всех ненулевых координат примерно равны, по результатам линейных измерений в присутствии шума с ограниченной сверху $\ell_p$-нормой. Предложен алгоритм декодирования, имеющий асимптотически минимальную сложность.

*Ключевые слова*: сжатие измерений, носитель разреженного вектора, групповое тестирование, поиск фальшивых монет, сигнатурные коды для суммирующего канала с множественным доступом и шумом, мультимедийные коды цифровых отпечатков пальцев.

**DOI:** 10.31857/S0555292323010023, **EDN:** JDHVDR

## § 1. Введение

В данной статье мы продолжаем (см. [1–6]) исследование задачи нахождения носителя неизвестного разреженного вектора $\boldsymbol{x} \in \mathbb{R}^n$ с помощью $m$ линейных измерений, подверженных влиянию шума. В литературе, посвященной теории сжатых измерений, неоднократно указывалось, что для нахождения хорошей аппроксимации неизвестного $K$-разреженного вектора $\boldsymbol{x}$ достаточно найти его носитель $\operatorname{supp}(\boldsymbol{x}) := \{i : \ x_i \neq 0\}$, а затем применить, например, метод наименьших квадратов. Напомним, что вектор $\boldsymbol{x}$ называется $K$-разреженным, если $|\operatorname{supp}(\boldsymbol{x})| \leqslant K$.

С другой стороны, нахождение хорошего приближения $\widehat{\boldsymbol{x}}$ к вектору $\boldsymbol{x}$ не гарантирует, что носители векторов $\boldsymbol{x}$ и $\widehat{\boldsymbol{x}}$ совпадают, поэтому задача нахождения носителя разреженного вектора по линейным измерениям представляет и самостоятельный интерес. Кроме того, как неоднократно отмечалось, эта задача равносильна нескольким другим известным задачам. Среди них задача построения сигнатурных кодов для суммирующего канала множественного доступа, задача поиска фальшивых монет на точных весах, задача построения мультимедийных кодов, устойчивых к атакам коалиций (подробнее см. обзор [4]). Наконец, самая новая постановка задачи, весьма близкая к задаче нахождения носителя неизвестного вектора, – это

[1] Исследование выполнено при финансовой поддержке гранта правительства Испании TCORISEBLOCK (номер гранта PID2019-110224RB-I00, проект MINECO).

[2] Исследование выполнено при финансовой поддержке Российского научного фонда в рамках гранта РНФ 22-41-02028.

[3] Исследование выполнено при финансовой поддержке Министерства образования Сингапура (Фонд академических исследований 2-го уровня, номера грантов MOE2019-T2-2-083 и MOE-T2EP20121-0007).

[4] Исследование выполнено при финансовой поддержке совместной японско-российского научной программы Японского общества содействия развитию науки (JSPS) и Российского фонда фундаментальных исследований в рамках проекта JPJSBP 120204802.

коды для передачи информации по каналу множественного доступа, когда нужно правильно восстановить только переданную информацию, но не отправителей информации [7].

Напомним кратко постановку задачи сжатых измерений [8, 9]. Требуется найти неизвестный $K$-разреженный вектор $\boldsymbol{x} \in \mathbb{R}^n$ по $m$ линейным измерениям, искаженным вектором шума $\boldsymbol{w} \in \mathbb{R}^m$, что формально можно описать как нахождение $K$-разреженного решения $\boldsymbol{x}$ следующего уравнения:

$$\boldsymbol{s} = H\boldsymbol{x}^T + \boldsymbol{w}, \tag{1}$$

где $H$ – $(m \times n)$-матрица линейных измерений. Нас же будет интересовать только носитель вектора-решения. Этой задаче также посвящено много публикаций (см. работы [10–12] и библиографию в них). В части работ рассматривается вероятностная модель шума (см. [10]), но тогда точное нахождение носителя невозможно. В данной статье мы исследуем задачу точного нахождения носителя неизвестного вектора в предположении, что величина шума $\boldsymbol{w}$ ограничена сверху некоторой известной величиной $t_p$ в метрике $\ell_p$.

Сформулируем основные результаты статьи. Явно строится семейство кодов, позволяющих точно находить носитель $K$-разреженного $n$-мерного вектора $\boldsymbol{x}$, *у которого модули всех ненулевых координат примерно равны*, с помощью $m$ линейных измерений, т.е. находить $\operatorname{supp}(\boldsymbol{x})$ из уравнения (1), где $m = O(K \log n)$ при фиксированном $K$ и $n \to \infty$, а шум имеет порядок $t_p = O(m^{1/p})$. Для этого семейства кодов предлагается алгоритм декодирования с полиномиальной по $Km$ сложностью, позволяющий реализовать половину соответствующей корректирующей способности. Параметры построенных кодов и их алгоритмов декодирования асимптотически оптимальны.

## § 2. Коды, позволяющие точно находить носитель разреженного “почти” двоичного вектора

Начнем со случая, когда априори известно, что все ненулевые координаты вектора $\boldsymbol{x}$ равны (это так, например, для модели 1 из [10]). Будем для простоты предполагать, что вектор $\boldsymbol{x}$ двоичный, т.е. все $x_i \in \{0, 1\}$. Условие, что матрица $H$ позволяет точно найти носитель двоичного вектора $\boldsymbol{x}$ (т.е. и сам вектор) из уравнения (1) при дополнительном ограничении, что $\|\boldsymbol{w}\|_p < t_p$, равносильно тому, что для любых двух различных подмножеств $A, B \subset [n]$, таких что $|A|, |B| \leqslant K$, справедливо неравенство

$$\left\| \sum_{j \in A} \boldsymbol{h}_j - \sum_{j \in B} \boldsymbol{h}_j \right\|_p \geqslant 2t_p. \tag{2}$$

Далее вместо матрицы $H$ мы будем рассматривать код $\mathcal{H}$, состоящий из столбцов этой матрицы.

Определение 1. Двоичный код $\mathcal{H} = \{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_n\} \subset \{0, 1\}^m$ называется $(K, t_p)_2$-*кодом*, если для любых двух различных кодовых подмножеств $F, G \subset \mathcal{H}$, таких что $|F|, |G| \leqslant K$, выполнено

$$\left\| \sum_{h \in F} \boldsymbol{h} - \sum_{h \in G} \boldsymbol{h} \right\|_p \geqslant 2t_p. \tag{3}$$

Отметим, что это определение уже появлялось ранее в [2] в рамках задачи о построении мультимедийного кода, устойчивого к атаке усреднения коалициями мощности не более $K$ и к ошибкам ограниченной евклидовой длины. На самом деле,

задача о мультимедийных кодах, устойчивых к общей линейной атаке $K$-коалиций, почти эквивалентна задаче о нахождении носителя у *произвольного* $K$-разреженного вектора, а задача о мультимедийных кодах, устойчивых к атаке усреднения, близка к рассматриваемой задаче о нахождении *двоичного* вектора. Отметим только, что в постановке задачи о мультимедийных кодах накладывается дополнительное ограничение, что координаты вектора $\boldsymbol{x}$ – это вероятности, т.е. все $x_i \geqslant 0$ и $\sum\limits_i x_i = 1$ (подробнее о сравнении этих задач см. [2, 4]). Более того, приводимая ниже конструкция $(K, t_p)_2$-кодов – это по существу конструкция из [2], у которой расширена область применения (не только двоичные векторы) и, главное, предложен алгоритм декодирования. Мы приведем эту конструкцию для полноты изложения, а также потому, что она осталась незамеченной в теории сжатых измерений.

Рассмотрим два двоичных линейных кода $C$ и $V$ длины $n$ и $m$ соответственно, при этом коды не только исправляют $K$ и $T$ ошибок соответственно, но и известны эффективные алгоритмы их исправления. Пусть $H'$ – какая-то проверочная $(r \times n)$-матрица кода $C$, позволяющая эффективно исправить $K$ ошибок, т.е. найти единственное двоичное решение $\boldsymbol{z}$ веса Хэмминга $\mathrm{wt}(\boldsymbol{z}) \leqslant K$ соответствующего синдромного уравнения $H'\boldsymbol{z}^T = \boldsymbol{s}$. Код $C$ будем называть *внутренним* кодом (как мы сейчас увидим, у предлагаемой конструкции есть определенное сходство с каскадной конструкцией).

Код $V$, называемый *внешним*, – это двоичный линейный код длины $m$ с $r$ информационными символами, и пусть $\theta\colon \mathbb{F}_2^r \to V$ – его систематическое кодирование. Мы предполагаем, что код $V$ исправляет $T$ ошибок с помощью алгоритма декодирования $\psi\colon \mathbb{F}_2^m \to V$.

Код $\mathcal{H}(C, V) = \{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_n\}$ строится следующим образом. Все $n$ столбцов матрицы $H'$ кодируются с помощью отображения $\theta$, превращаясь в $n$ двоичных векторов $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_n$ длины $m$, где $\boldsymbol{h}_j = \theta(\boldsymbol{h}'_j)$.

Лемма 1. *Код $\mathcal{H}(C, V)$ является $(K, t_p)_2$-кодом с $2t_p = (2T+1)^{1/p}$.*

Доказательство. Рассмотрим два различных подмножества $A, B \subset [n]$, таких что $|A|, |B| \leqslant K$, соответствующие им векторы

$$\boldsymbol{h}_A := \sum_{j\in A} \boldsymbol{h}_j = \sum_{j\in A} \theta(\boldsymbol{h}'_j), \quad \boldsymbol{h}_B := \sum_{j\in B} \boldsymbol{h}_j = \sum_{j\in B} \theta(\boldsymbol{h}'_j)$$

и эти же векторы, взятые по модулю 2, а именно $\boldsymbol{h}_A \bmod 2$ и $\boldsymbol{h}_B \bmod 2$. Тогда в силу линейности отображения кодирования $\theta$ имеем

$$\boldsymbol{h}_A \bmod 2 = \sum_{j\in A} \theta(\boldsymbol{h}'_j) \bmod 2 = \theta\Big(\sum_{j\in A} \oplus \boldsymbol{h}'_j\Big).$$

Аналогично

$$\boldsymbol{h}_B \bmod 2 = \theta\Big(\sum_{j\in B} \oplus \boldsymbol{h}'_j\Big).$$

Двоичные векторы $\sum\limits_{j\in A} \oplus \boldsymbol{h}'_j$ и $\sum\limits_{j\in B} \oplus \boldsymbol{h}'_j$ различны, так как это две различные суммы из $K$ или менее столбцов проверочной матрицы кода, исправляющего $K$ ошибок. Поэтому $\boldsymbol{h}_A \bmod 2$ и $\boldsymbol{h}_B \bmod 2$ – это два различных слова кода $V$, и тем самым, они различаются как минимум в $2T+1$ координатах. Тем более, целочисленные векторы $\boldsymbol{h}_A$ и $\boldsymbol{h}_B$ различаются как минимум в $2T+1$ координатах, и следовательно,

$$\|\boldsymbol{h}_A - \boldsymbol{h}_B\|_p^p \geqslant 2T+1. \quad \blacktriangle$$

Вернемся к общей задаче нахождения носителя *произвольного* $K$-разреженного вектора $\boldsymbol{x}$ в присутствии шума $\boldsymbol{w}$ ограниченной $\ell_p$-длины. Заметим, что если не наложить дополнительного ограничения, что значения ненулевых координат вектора $\boldsymbol{x}$ не могут быть слишком малы, то точное нахождение $\operatorname{supp}(\boldsymbol{x})$ невозможно (см. [3]). Действительно, рассмотрим в качестве примера простейший случай, когда $K = 2$, и два вектора $\boldsymbol{x} = (1, \varepsilon, 0, \ldots, 0)$ и $\boldsymbol{x}' = (1, 0, \varepsilon, 0, \ldots, 0)$. Тогда вектор ошибки $\boldsymbol{w} = (0, -\varepsilon, \varepsilon, 0, \ldots, 0)$ переводит вектор $\boldsymbol{x}$ в вектор $\boldsymbol{x}'$, но при этом длина вектора ошибки мала.

Теорема 1. *Если неизвестный $K$-разреженный вектор $\boldsymbol{x}$ "почти" двоичный, т.е. все его ненулевые координаты лежат в диапазоне $[1-\tau, 1+\tau]$, $0 < \tau$, то код $\mathcal{H}(C, V)$ позволяет точно найти носитель вектора, если длина шума $\boldsymbol{w}$ в метрике $\ell_p$ меньше $\Delta/2$, где $\Delta = (2T+1)^{1/p} - 2K\tau m^{1/p}$.*

Доказательство. Нужно доказать, что для любых двух "почти" двоичных $K$-разреженных векторов $\boldsymbol{x}, \boldsymbol{y}$ с различными носителями $A, B \subset [n]$, такими что $|A|, |B| \leqslant K$, справедливо неравенство

$$\left\| \sum_{j \in A} x_j \boldsymbol{h}_j - \sum_{j \in B} y_j \boldsymbol{h}_j \right\|_p \geqslant \Delta. \tag{4}$$

Представим $x_j = 1 + \alpha_j$, $y_j = 1 + \beta_j$, где $|\alpha_j|, |\beta_j| \leqslant \tau$. Тогда

$$\left\| \sum_{j \in A} x_j \boldsymbol{h}_j - \sum_{j \in B} y_j \boldsymbol{h}_j \right\|_p \geqslant \left\| \sum_{j \in A} \boldsymbol{h}_j - \sum_{j \in B} \boldsymbol{h}_j \right\|_p - \left\| \sum_{j \in A} \alpha_j \boldsymbol{h}_j - \sum_{j \in B} \beta_j \boldsymbol{h}_j \right\|_p \geqslant$$
$$\geqslant (2T+1)^{1/p} - (|A| + |B|)\tau \max \|\boldsymbol{h}_j\|_p \geqslant \Delta. \quad \blacktriangle$$

## § 3. Коды, позволяющие находить носитель разреженного вектора, у которого модули всех ненулевых координат примерно равны

То, что коды конструкции, предложенной в [2] и описанной выше, позволяют находить носитель не только двоичного, но и почти двоичного разреженного вектора, было отмечено в [5], но без алгоритма декодирования. Отметим, что эта конструкция кодов была впервые предложена Т. Эриксоном и В.И. Левенштейном в [13] как конструкция сигнатурных кодов для канала множественного доступа, суммирующего по модулю 2, и после этого она переоткрывалась для разных других задач, например, для задачи построения кодов, исправляющих ошибки в канале и синдроме [14].

Предложенная конструкция очевидно ограничена применимостью только к векторам, у которых все координаты одного знака. Сейчас мы опишем более общую конструкцию, позволяющую находить разреженный вектор (а следовательно, и его носитель), все ненулевые координаты которого равны $-1$ или $+1$. Матрица измерений $H$ будет троичной, т.е. $h_{i,j} \in \{-1, 0, +1\}$. То, что матрица $H$ позволяет однозначно найти троичный вектор из уравнения (1) при дополнительном ограничении, что $\|\boldsymbol{w}\|_p < t_p$, означает, что для любых двух различных $K$-разреженных векторов $\boldsymbol{a}, \boldsymbol{b} \in \{-1, 0, +1\}^m$ справедливо неравенство

$$\left\| H\boldsymbol{a}^T - H\boldsymbol{b}^T \right\|_p \geqslant 2t_p. \tag{5}$$

Как и выше, вместо матрицы $H$ мы будем рассматривать *троичный* код $\mathcal{H}$, состоящий из столбцов этой матрицы.

Определение 2. Троичный код $\mathcal{H} = \{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_n\} \subset \{-1, 0, +1\}^m$ называется $(K, t_p)_3$-*кодом*, если для любых двух различных $K$-разреженных векторов $\boldsymbol{a}, \boldsymbol{b} \in$

$\in \{-1,0,+1\}^m$ выполнено

$$\left\| \sum_{j=1}^{n} a_j \boldsymbol{h}_j - \sum_{j=1}^{n} b_j \boldsymbol{h}_j \right\|_p \geqslant 2t_p. \tag{6}$$

Построение троичных $(K, t_p)_3$-кодов дословно повторяет построение их двоичных аналогов. А именно, выбираются два троичных линейных кода $C$ и $V$ длины $n$ и $m$ с известными эффективными алгоритмами исправления $K$ и $T$ ошибок соответственно. Пусть $H'$ – какая-то проверочная $(r \times n)$-матрица кода $C$, позволяющая эффективно исправить $K$ ошибок, т.е. найти единственное троичное решение $\boldsymbol{z}$ веса Хэмминга $\mathrm{wt}(\boldsymbol{z}) \leqslant K$ соответствующего синдромного уравнения $H'\boldsymbol{z}^T = \boldsymbol{s}$ (в поле $\mathbb{F}_3$). Код $V$ представляет собой троичный линейный код длины $m$ с $r$ информационными символами, и пусть $\theta\colon \mathbb{F}_3^r \to V$ – его систематическое кодирование. Мы предполагаем, что код $V$ исправляет $T$ ошибок с помощью алгоритма декодирования $\psi\colon \mathbb{F}_3^m \to V$. Код $C$ будем называть *внутренним* кодом, а код $V$ – *внешним*.

Код $\mathcal{H}(C,V) = \{\boldsymbol{h}_1, \ldots, \boldsymbol{h}_n\}$ состоит из $n$ троичных векторов длины $m$, полученных применением кодирования $\theta$ к столбцам матрицы $H'$, т.е. $\boldsymbol{h}_j = \theta(\boldsymbol{h}'_j)$.

Лемма 2. *Код $\mathcal{H}(C,V)$ является $(K, t_p)_3$-кодом с $2t_p = (2T+1)^{1/p}$.*

Доказательство. Рассмотрим два произвольных различных $K$-разреженных троичных вектора $\boldsymbol{a}, \boldsymbol{b} \in \{-1,0,+1\}^m$, соответствующие им векторы

$$\boldsymbol{h}_a := \sum_{j=1}^{n} a_j \boldsymbol{h}_j = \theta\left( \sum_{j=1}^{n} a_j \boldsymbol{h}'_j \right), \quad \boldsymbol{h}_b := \sum_{j=1}^{n} b_j \boldsymbol{h}_j = \theta\left( \sum_{j=1}^{n} b_j \boldsymbol{h}'_j \right)$$

и эти же векторы, взятые по модулю 3, а именно $\boldsymbol{h}_a \bmod 3$ и $\boldsymbol{h}_b \bmod 3$. Тогда в силу линейности отображения кодирования $\theta$ имеем

$$\boldsymbol{h}_a \bmod 3 = \theta\left( \sum_{j=1}^{n} a_j \boldsymbol{h}'_j \right) \bmod 3 = \theta\left( \sum_{j=1}^{n} a_j \boldsymbol{h}'_j \bmod 3 \right).$$

Аналогично

$$\boldsymbol{h}_b \bmod 2 = \theta\left( \sum_{j=1}^{n} b_j \boldsymbol{h}'_j \bmod 3 \right).$$

Троичные векторы $\sum\limits_{j=1}^{n} a_j \boldsymbol{h}'_j \bmod 3$ и $\sum\limits_{j=1}^{n} b_j \boldsymbol{h}'_j \bmod 3$ различны, так как это две различные линейные комбинации над полем $\mathbb{F}_3$ из $K$ или менее столбцов проверочной матрицы кода, исправляющего $K$ ошибок. Поэтому $\boldsymbol{h}_a \bmod 3$ и $\boldsymbol{h}_b \bmod 3$ – это два различных слова кода $V$, и тем самым, они различаются как минимум в $2T+1$ координатах. Тем более, целочисленные векторы $\boldsymbol{h}_a$ и $\boldsymbol{h}_b$ различаются как минимум в $2T+1$ координатах, и следовательно, $\|\boldsymbol{h}_a - \boldsymbol{h}_b\|_p^p \geqslant 2T+1$. ▲

Обобщим теперь теорему 1, рассмотрев вместо "почти" двоичных разреженных векторов разреженные векторы, у которых модули всех ненулевых координат почти равны.

Теорема 2. *Если для всех ненулевых координат $K$-разреженного вектора $\boldsymbol{x}$ справедливо $|x_j - a_j| < \tau$, $0 < \tau$, $a_j \in \{-1,+1\}$, то троичный код $\mathcal{H}(C,V)$ позволяет точно найти носитель вектора $\boldsymbol{x}$, если длина шума $\boldsymbol{w}$ в метрике $\ell_p$ меньше $\Delta/2$, где $\Delta = (2T+1)^{1/p} - 2K\tau m^{1/p}$.*

Доказательство. Нужно доказать, что для любых двух $K$-разреженных векторов $\boldsymbol{x}, \boldsymbol{y}$, чьи ненулевые координаты удовлетворяют неравенствам $|x_j - a_j| < \tau$, $|y_j - b_j| < \tau$, где все $a_j, b_j \in \{-1, +1\}$, с различными носителями $A, B \subset [n]$ справедливо неравенство

$$\left\| \sum_{j \in A} x_j \boldsymbol{h}_j - \sum_{j \in B} y_j \boldsymbol{h}_j \right\|_p \geqslant \Delta. \tag{7}$$

Представим $x_j = a_j + \alpha_j$, $y_j = b_j + \beta_j$, где $|\alpha_j|, |\beta_j| < \tau$. Тогда

$$\left\| \sum_{j \in A} x_j \boldsymbol{h}_j - \sum_{j \in B} y_j \boldsymbol{h}_j \right\|_p \geqslant \left\| \sum_{j \in A} a_j \boldsymbol{h}_j - \sum_{j \in B} b_j \boldsymbol{h}_j \right\|_p - \left\| \sum_{j \in A} \alpha_j \boldsymbol{h}_j - \sum_{j \in B} \beta_j \boldsymbol{h}_j \right\|_p \geqslant$$
$$\geqslant (2T+1)^{1/p} - (|A| + |B|)\tau \max \|\boldsymbol{h}_j\|_p \geqslant \Delta. \quad \blacktriangle$$

## § 4. Декодирование

Начнем с алгоритма нахождения $K$-разреженного двоичного вектора $\boldsymbol{x}$ при условии, что норма шума удовлетворяет условию $\|\boldsymbol{w}\|_1 < 0{,}5(T+1)$. В качестве предварительного (нулевого) шага декодирования выполняется аналог жесткого приема применительно к "принятому" вектору-синдрому $\boldsymbol{s}$ из уравнения (1), т.е. каждая координата $s_i$ заменяется на ближайшее целое число $\widehat{s}_i$. Очевидно, $\widehat{s}_i \neq (H\boldsymbol{x}^T)_i$ только в том случае, если

$$\left| s_i - (H\boldsymbol{x}^T)_i \right| = |w_i| \geqslant 1/2.$$

Следовательно, расстояние Хэмминга удовлетворяет неравенству

$$d_H(\widehat{\boldsymbol{s}}, H\boldsymbol{x}^T) \leqslant 2\|\boldsymbol{w}\|_1.$$

Тем более,

$$d_H(\widehat{\boldsymbol{s}} \bmod 2, H\boldsymbol{x}^T \bmod 2) \leqslant 2\|\boldsymbol{w}\|_1,$$

и если $2\|\boldsymbol{w}\|_1 < T+1$, то число "ошибок" в векторе $\widehat{\boldsymbol{s}} \bmod 2$, т.е. отличий от вектора $H\boldsymbol{x}^T \bmod 2$, не более $T$, и алгоритм декодирования $\psi$, примененный к вектору $\widehat{\boldsymbol{s}} \bmod 2$, выдаст $H\boldsymbol{x}^T \bmod 2$. В свою очередь, первые $r$ символов $H\boldsymbol{x}^T \bmod 2$ – это синдром $H'\boldsymbol{x}^T \bmod 2$, поскольку кодирование $\theta$ является систематическим. Тогда искомый вектор $\boldsymbol{x}$ находится алгоритмом декодирования кода $C$.

Рассмотрим теперь нахождение носителя "почти" двоичного $K$-разреженного вектора $\boldsymbol{x}$, ненулевые координаты которого лежат в диапазоне $[1-\tau, 1+\tau]$, где $0 < \tau < 1/2$. Представим $\boldsymbol{x} = \boldsymbol{a} + \boldsymbol{\alpha}$, где $\boldsymbol{a}$ – характеристический вектор множества $\operatorname{supp}(\boldsymbol{x})$, и следовательно, вектор $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ представляет собой $K$-разреженный вектор, в котором по предположению $|\alpha_i| \leqslant \tau$ для всех $i$. Принятый "вектор-синдром" $\boldsymbol{s}$ представим в виде $\boldsymbol{s} = H(\boldsymbol{a} + \boldsymbol{\alpha})^T + \boldsymbol{w} = H\boldsymbol{a}^T + \boldsymbol{w}'$, где $\boldsymbol{w}' = \boldsymbol{w} + H\boldsymbol{\alpha}^T$, и следовательно, $\|\boldsymbol{w}'\|_1 \leqslant \boldsymbol{w} + K\tau m$. Поэтому описанный выше алгоритм позволяет найти двоичный вектор $\boldsymbol{a}$ и, тем самым, носитель $\operatorname{supp}(\boldsymbol{x})$, если для нормы "шума" $\boldsymbol{w}'$ выполнено условие

$$\|\boldsymbol{w}'\|_1 < 0{,}5(T+1).$$

Эти рассуждения почти дословно переносятся на "троичный" случай, т.е. поиск носителя разреженного вектора, у которого модули всех ненулевых координат примерно равны.

## § 5. Выбор кодов и сложность реализации

Оценим сложность предложенной конструкции для двух наиболее популярных асимптотических режимов. Первый – когда $K$ постоянно, а $n \to \infty$, что соответствует практическим случаям, когда неизвестный разреженный вектор действительно разреженный. В качестве внутреннего кода $C$ длины $n$ с проверочной $(r \times n)$-матрицей $H'$ возьмем коды БЧХ или классические неприводимые коды Гоппы, для которых $n = 2^{r/K} - 1$ или $n = 2^{r/K}$ соответственно.

Что касается внешнего кода $V$, то снова берем код Гоппы или код БЧХ, либо, если $m \gg 1$, то код из семейства "хороших" кодов, т.е. кодов с фиксированной кодовой скоростью и отношением кодового расстояния к длине кода, отличным от нуля, плюс дополнительно с полиномиальной сложностью реализации. Например, можно взять семейство кодов-расширителей [15, 16].

Нулевой шаг декодирования – округление координат синдрома до целых значений – имеет линейную по $m$ сложность. Первый шаг декодирования – декодирование вектора $\widehat{\boldsymbol{s}}$ "внешним" кодом $V$ – имеет полиномиальную по $m$ сложность для кодов БЧХ или для кодов-расширителей. Второй этап – синдромное декодирование "внутреннего" кода $C$ – для кодов БЧХ имеет сложность $\mathrm{poly}(Kn)$. Таким образом, общая сложность равна $\mathrm{poly}(Kn)$ для $K$ – константы и $n \to \infty$.

Второй асимптотический режим – когда $K$ растет линейно по $n$. Тогда в качестве внутреннего и внешнего кодов берутся коды-расширители, что дает линейную по $n$ сложность нахождения носителя разреженного $n$-мерного вектора.

Важно отметить, что предлагаемые коды могут быть явно построены, более того, с полиномиальной или даже линейной по $n$ сложностью, но случайное кодирование позволяет увеличить кодовую скорость в $\log K$ раз (см. [17]).

В заключение, следуя [10], где исследовалась аппроксимация носителя разреженного *двоичного* вектора в условиях гауссовского шума, рассмотрим следующий числовой пример: $K = 25$, $n = 2^{18}$. Возьмем в качестве внутреннего кода $C$ неприводимый код Гоппы с $r = 25 \times 18 = 450$ битами проверки на четность, а в качестве внешнего кода $V$ – также код Гоппы длины $m = 1020$, исправляющий $T = 57$ ошибок. Тогда полученная $(1020 \times 2^{18})$-матрица $H$ способна точно найти носитель неизвестного двоичного вектора разреженности $K = 25$ из уравнения (1) при условии, что сумма абсолютных величин координат шума $\boldsymbol{w}$ меньше, чем 57,5. Более того, предложенный алгоритм декодирования найдет носитель при условии, что сумма абсолютных величин координат шума меньше, чем 28,75.

## СПИСОК ЛИТЕРАТУРЫ


1. *Egorova E., Fernandez M., Kabatiansky G., Lee M.H.* Signature Codes for Weighted Noisy Adder Channel, Multimedia Fingerprinting and Compressed Sensing // Des. Codes Cryptogr. 2019. V. 87. № 2–3. P. 455–462. https://doi.org/10.1007/s10623-018-0551-9
2. *Егорова Е.Е., Фернандес М., Кабатянский Г.А., Мяо И.* Существование и конструкции мультимедийных кодов, способных находить полную коалицию при атаке усреднения и шуме // Пробл. передачи информ. 2020. Т. 56. № 4. С. 97–108. https://doi.org/10.31857/S0555292320040087
3. *Fan J., Gu Y., Hachimori M., Miao Y.* Signature Codes for Weighted Binary Adder Channel and Multimedia Fingerprinting // IEEE Trans. Inform. Theory. 2021. V. 67. № 1. P. 200–216. https://doi.org/10.1109/TIT.2020.3033445
4. *Егорова Е.Е., Кабатянский Г.А.* Разделимые коды для защиты мультимедиа от нелегального копирования коалициями // Пробл. передачи информ. 2021. Т. 57. № 2. С. 90–111. https://doi.org/10.31857/S0555292321020066
5. *Джанабекова А., Кабатянский Г.А., Камель И., Рабие Т.Ф.* Неперекрывающиеся выпуклые многогранники с вершинами из булева куба и другие задачи теории кодирова-

ния // Пробл. передачи информ. 2022. Т. 58. № 4. С. 50–61. https://www.mathnet.ru/rus/ppi2383

6. *Fernandez M., Kabatiansky G., Miao Y.* A Novel Support Recovery Algorithms and Its Applications to Multiple-Access Channels // Proc. 2022 IEEE Int. Multi-Conf. on Engineering, Computer and Information Sciences (SIBIRCON). Yekaterinburg, Russian Federation. Nov. 11–13, 2022. P. 170–173. https://doi.org/10.1109/SIBIRCON56155.2022.10017094
7. *Polyanskiy Y.* A Perspective on Massive Random-Access // Proc. 2017 IEEE Int. Symp. on Information Theory (ISIT'2017). Aachen, Germany. June 25–30, 2017. P. 2523–2527. https://doi.org/10.1109/ISIT.2017.8006984
8. *Donoho D.L.* Compressed Sensing // IEEE Trans. Inform. Theory. 2006. V. 52. № 4. P. 1289–1306. https://doi.org/10.1109/TIT.2006.871582
9. *Candès E.J., Tao T.* Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies? // IEEE Trans. Inform. Theory. 2006. V. 52. № 12. P. 5406–5425. https://doi.org/10.1109/TIT.2006.885507
10. *Gkagkos M., Pradhan A.K., Amalladinne V., Narayanan K., Chamberland J-F., Georghiades C.N.* Approximate Support Recovery Using Codes for Unsourced Multiple Access // Proc. 2021 IEEE Int. Symp. on Information Theory (ISIT'2021). Melbourne, Australia. July 12–20, 2021. P. 2948–2953. https://doi.org/10.1109/ISIT45174.2021.9517995
11. *Wen J., Zhou Z., Wang J., Tang X., Mo Q.* A Sharp Condition for Exact Support Recovery with Orthogonal Matching Pursuit // IEEE Trans. Signal Process. 2017. V. 65. № 6. P. 1370–1382. https://doi.org/10.1109/TSP.2016.2634550
12. *Mehrabi M., Tchamkerten A.* Error-Correction for Sparse Support Recovery Algorithms // Proc. 2021 IEEE Int. Symp. on Information Theory (ISIT'2021). Melbourne, Australia. July 12–20, 2021. P. 1754–1759. https://doi.org/10.1109/ISIT45174.2021.9518027
13. *Ericson T., Levenshtein V.I.* Superimposed Codes in the Hamming Space // IEEE Trans. Inform. Theory. 1994. V. 40. № 6. P. 1882–1893. https://doi.org/10.1109/18.340463
14. *Влэдуц С.Г., Кабатянский Г.А., Ломаков В.В.* Об исправлении ошибок при искажениях в канале и синдроме // Пробл. передачи информ. 2015. Т. 51. № 2. С. 50–56. http://mi.mathnet.ru/ppi2169
15. *Sipser M., Spielman D.A.* Expander Codes // IEEE Trans. Inform. Theory. 1996. V. 42. № 6. Part 1. P. 1710–1722. https://doi.org/10.1109/18.556667
16. *Spielman D.* Linear-Time Encodable and Decodable Error-Correcting Codes // IEEE Trans. Inform. Theory. 1996. V. 42. № 6. P. 1723–1731. https://doi.org/10.1109/18.556668
17. *Vorobyev I.* Complete Traceability Multimedia Fingerprinting Codes Resistant to Averaging Attack and Adversarial Noise with Optimal Rate // Des. Codes Cryptogr. 2023. V. 4. № 4. P. 1183–1191. https://doi.org/10.1007/s10623-022-01144-x

*Фернандес Марсель*
Политехнический университет Каталонии,
Барселона, Испания
marcelf@entel.upc.edu
*Кабатянский Григорий Анатольевич*
Сколковский институт науки и технологий (Сколтех), Москва
g.kabatyansky@skoltech.ru
*Круглик Станислав Александрович*
Наньянский технологический университет, Сингапур
stanislav.kruglik@ntu.edu.sg
*Мяо Ин*
Университет Цукубы, Цукуба, префектура Ибараки, Япония
miao@sk.tsukuba.ac.jp

Поступила в редакцию
30.12.2022
После доработки
21.02.2023
Принята к публикации
21.02.2023

**ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ**

**Том 59** **2023** **Вып. 1**

УДК 621.391 : 519.725

© 2023 г. **П.В. Трифонов**

# ПОСТРОЕНИЕ И ДЕКОДИРОВАНИЕ ПОЛЯРНЫХ КОДОВ С БОЛЬШИМИ ЯДРАМИ: ОБЗОР[1]

Представлены методы построения и декодирования полярных кодов с большими ядрами. Важнейшей проблемой при реализации алгоритма последовательного исключения декодирования полярных кодов и его обобщений является обработка ядра, т.е. быстрое вычисление логарифмических отношений правдоподобия для входных символов ядра. Представлены оконный и рекурсивный решетчатый методы обработки больших ядер. Рассмотрены методы оценки надежности битовых подканалов и получения кодов с улучшенными свойствами расстояния.

*Ключевые слова*: полярные коды, большие ядра.

**DOI:** 10.31857/S0555292323010035, **EDN:** RMFRTN

## § 1. Введение

Полярные коды, предложенные Э. Ариканом [1], уже вошли в стандарт 5G. Однако их применение там ограничено каналом управления, где передаются относительно короткие блоки данных. Для больших блоков данных коды с малой плотностью проверок на четность (LDPC) оказываются более предпочтительны. Причиной этого является плохое масштабирование полярных кодов Арикана и их аналогов. Конструкцию полярных кодов можно обобщить, заменив матрицу Арикана размера $2 \times 2$ на бóльшую [2]. Можно показать, что такие коды, известные как полярные коды с большими ядрами, достигают асимптотически оптимальной экспоненты масштабирования [3–5]. Такие коды могут быть продекодированы с помощью алгоритма последовательного исключения (ПИ) и его обобщений со сложностью $O(n \log n)$. Однако при непосредственной реализации константа, скрытая в асимптотической записи, растет экспоненциально с размером ядра $\ell$. Поэтому до недавнего времени такие коды считались непрактичными. Построение полярных кодов с большими ядрами может быть выполнено методами, основанными на тех же принципах, что и в случае полярных кодов Арикана.

В данном обзоре представлен обзор методов построения и декодирования полярных кодов с большими ядрами.

## § 2. Полярные коды

**2.1. Поляризация канала.** Рассмотрим вначале случай двоичных кодов. Полярный код [2] – множество векторов $c_0^{n-1} = u_0^{n-1} K^{\otimes m}$, где $K$ – обратимая $(\ell \times \ell)$-матрица, называемая ядром поляризации, $n = \ell^m$, $u_i = 0$ для $i \in \mathcal{F}$, $\mathcal{F} \subset [n]$ называется множеством замораживания, $u_a^b = (u_a, u_{a+1}, \ldots, u_b)$ и $[n] = \{0, \ldots, n-1\}$. Это

[1] Работа выполнена при финансовой поддержке Российского научного фонда в рамках гранта РНФ № 22-11-00208.

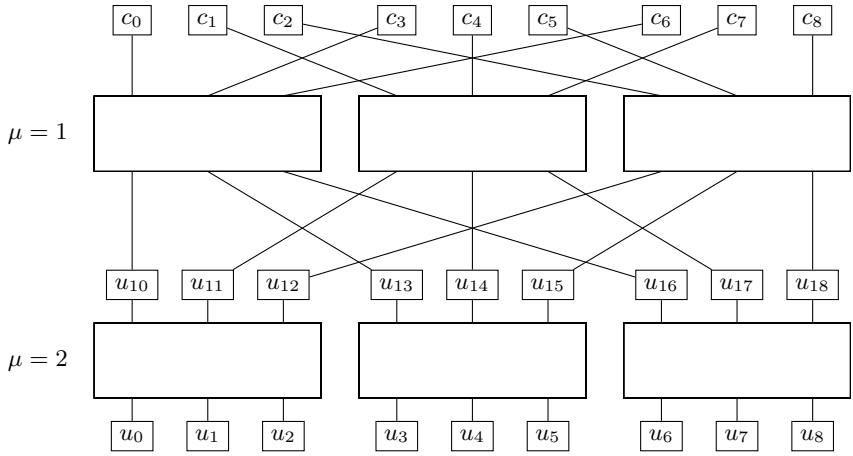


Рис. 1. Схема кодирования для полярного кода длины 9 на основе ядра размерности 3

определение может быть обобщено на случай кодов со смешанными ядрами [6, 7], кодовые слова которых равны $c_0^{n-1} = u_0^{n-1}(K_{\ell_0} \otimes K_{\ell_2} \otimes \ldots \otimes K_{\ell_{m-1}})$, где $K_{\ell_i}$ – ядро размерности $\ell_i$. Если не указано иное, мы рассматриваем случай, когда все ядра одинаковы.

На рис. 1 приведен пример схемы кодирования для полярного кода с ядром размерности $\ell = 3$ и $m = 2$ слоями поляризующего преобразования.

Можно показать, что матрица $K^{\otimes m}$ вместе с симметричным каналом без памяти с двоичным входом и функцией переходных вероятностей $W(r \,|\, c)$ задают синтетические битовые подканалы с функциями переходных вероятностей

$$W_m^{(i)}\big(r_0^{n-1}, u_0^{i-1} \,|\, u_i\big) = \frac{1}{2^{n-1}} \sum_{u_{i+1}^{n-1}} \prod_{j=0}^{n-1} W\Big(r_j \,|\, \big(u_0^{n-1} K^{\otimes m}\big)_j\Big), \quad 0 \leqslant i < \ell^m,$$

где $r_j$ – выходные символы канала. Если $K$ нельзя преобразовать перестановками столбцов в верхнетреугольную матрицу, то пропускные способности этих битовых подканалов сходятся к 0 или 1, а доля подканалов с пропускной способностью, близкой к 1, сходится к $I(W)$ – симметричной пропускной способности исходного канала $W$ (см. [2]). Множество замораживания $\mathcal{F}$ обычно выбирается как набор индексов $i$, соответствующих подканалам с малой пропускной способностью $W_m^{(i)}$. Удобно определить вспомогательные вероятности

$$W_m^{(i)}\big(u_0^i \,|\, r_0^{n-1}\big) = \frac{W_m^{(i)}\big(r_0^{n-1}, u_0^{i-1} \,|\, u_i\big)}{2W_m\big(r_0^{n-1}\big)}.$$

Заметим, что иногда полярные коды определяются [1] с матрицей перестановки $P$ в схеме кодирования, т.е. $c_0^{n-1} = u_0^{n-1} P K^{\otimes m}$. Такая перестановка нужна только для удобства обозначений и может быть исключена без изменения свойств полученного кода. Следует лишь учесть ее наличие или отсутствие в декодере.

**2.2. Алгоритм последовательного исключения декодирования полярных кодов.** Рассмотрим передачу кодового слова $c_0^{n-1}$ полярного кода по симметричному каналу без памяти. Декодирование полярных кодов может быть реализовано алгоритмом

последовательного исключения (ПИ), который принимает решения

$$\widehat{u}_i = \begin{cases} 0, & i \in \mathcal{F}, \\ \arg\max\limits_{u_i \in \mathbb{F}_2} W_m^{(i)}\big(\widehat{u}_0^{i-1} \bullet u_i \,|\, r_0^{n-1}\big), & i \notin \mathcal{F}, \end{cases}$$

где $\bullet$ обозначает оператор конкатенации,

$$W_m^{(i)}\big(u_0^i \,|\, r_0^{n-1}\big) = A \sum_{u_{i+1}^{n-1}} \prod_{j=0}^{n-1} W_0^{(0)}\Big(\big(u_0^{n-1} K^{\otimes \ell}\big)_i \,|\, r_i\Big),$$

$$W(c\,|\,r) = \frac{W(r\,|\,c)}{2W(r)},$$

а $A$ – нормирующий множитель, не зависящий от $u_i$. Из рекурсивной структуры произведения Кронекера следует, что эти вероятности могут быть вычислены как

$$W_\mu^{(\ell i+s)}\big(u_0^{li+s} \,|\, r_0^{\ell^\mu - 1}\big) = A' \sum_{u_{\ell i+s+1}^{\ell i+\ell-1}} \prod_{j=0}^{\ell-1} W_{\mu-1}^{(i)}\Big(\big(u_{\ell t}^{\ell(t+1)-1} K\big)_j,\ t \in [i+1] \,|\, r_{j,\ell}^{\ell^\mu - 1}\Big), \quad (1)$$

где $A'$ – еще один нормирующий множитель, $r_{j,\ell}^{n-1} = (r_j, r_{j+\ell}, \ldots, r_{j+n-\ell})$, $0 \leqslant s < \ell$, $0 \leqslant i < \ell^{\mu-1}$, $1 \leqslant \mu \leqslant m$. Эта операция известна как обработка ядра или маргинализация ядра [8, 9]. Непосредственная реализация этой операции имеет сложность $O(\ell 2^\ell)$. Более эффективные методы обсуждаются в § 4.

Известно, что алгоритм ПИ является крайне субоптимальным. Для случая кодов с ядром Арикана было предложено множество его усовершенствований [10–13], среди которых наиболее известным является списочный алгоритм последовательного исключения Таля – Варди. Большинство этих методов допускает обобщение на случай неарикановских ядер. Для этого необходимо внести следующие изменения:

- Реализовать соответствующий алгоритм обработки ядра;
- Если алгоритм обработки ядра требует сохранения некоторой информации о состоянии на различных фазах $s$, обеспечить доступ к этой информации при рассмотрении декодером различных путей $u_0^{\ell i+s}$.

**2.3. Параметры ядра.** Корректирующая способность полярных кодов конечной длины в значительной степени зависит от вероятности битовых ошибок в подканалах, используемых для передачи незамороженных символов. Она может быть оценена сверху параметром Бхаттачарьи $Z(W_m^{(i)})$ этих подканалов. Следующий параметр [2] показывает, насколько хорошими могут быть подканалы, полученные из матрицы поляризующего преобразования $K^{\otimes m}$.

О п р е д е л е н и е 1. Матрица $K$ имеет *скорость поляризации* $E(K)$, если для любого канала с двоичным входом $W$, $0 \leqslant I(W) < 1$, выполнено следующее:

- Для любого $\beta < E(K)$ справедливо $\liminf\limits_{m\to\infty} \mathbf{P}\big\{Z(W_m^{(i)}) \leqslant 2^{-\ell^{m\beta}}\big\} = I(W)$;
- Для любого $\beta > E(K)$ справедливо $\liminf\limits_{m\to\infty} \mathbf{P}\big\{Z(W_m^{(i)}) \geqslant 2^{-\ell^{m\beta}}\big\} = 1$.

Здесь вероятность следует интерпретировать как долю подканалов $W_m^{(i)}$, удовлетворяющих указанным свойствам.

Из этого определения следует, что для любого $R$, $0 < R < I(W)$, для достаточно больших $m$ вероятность ошибки декодирования методом последовательного исключения для полярного $(n = \ell^m, Rn)$-кода ограничена сверху величиной $2^{-n^\beta}$, $\beta < E(K)$.

Определение 2 (см. [2]). *Частичные расстояния* $\mathcal{D}_i$, $0 \leqslant i < \ell$, $(\ell \times \ell)$-матрицы $K = \begin{pmatrix} K[0] \\ \dots \\ K[\ell-1] \end{pmatrix}$ определяются как

$$\mathcal{D}_i = d_H(K[i], \mathcal{C}^{(i+1)}), \quad 0 \leqslant i < \ell - 1, \tag{2}$$

$$\mathcal{D}_{\ell-1} = \mathrm{wt}(K[\ell-1]), \quad 0 \leqslant i < \ell - 1, \tag{3}$$

где $\mathcal{C}^{(i)} = \langle K[i], \dots, K[\ell-1] \rangle$ – линейный $(\ell, \ell - i)$-код, порожденный строками ядра $i, \dots, \ell - 1$, а $d_H(a, C)$ – минимальное расстояние Хэмминга между вектором $a$ и кодовыми словами кода $C$.

Вектор $\mathcal{D}_0^{\ell-1}$ называется профилем частичных расстояний. Скорость поляризации ядра $K$ может быть вычислена как $E(K) = \frac{1}{\ell} \sum\limits_{i=0}^{\ell-1} \log_\ell \mathcal{D}_i$. Можно легко проверить, что для случая ядра Арикана $K_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ частичными расстояниями являются $\mathcal{D}_0 = 1$, $\mathcal{D}_1 = 2$, так что скорость его поляризации $E(K_2) = 0{,}5$. Можно показать [2], что существуют $(\ell \times \ell)$-ядра $K_\ell$ с $E(K_\ell) \to 1$.

Экспонента масштабирования $\mu$ для семейства кодов со скоростью $R$ показывает длину $n = O\Big(\frac{1}{(I(W) - R)^\mu}\Big)$, необходимую для достижения некоторой фиксированной целевой вероятности ошибки на кодовое слово на канале $W$ с симметричной пропускной способностью $I(W)$. Для случайных кодов можно показать [14], что $\mu = 2$. Для произвольных семейств кодов неизвестно, существует ли такая $\mu$. Можно предположить, что для полярных кодов для любого $\varepsilon$ существует $\mu(W, K_\ell)$ и

$$f = \lim_{m \to \infty} \beta_m \ell^{\frac{m}{\mu(W, K_\ell)} - m}, \quad 0 < f < \infty,$$

где $\beta_m$ – число подканалов $W_m^{(i)}$, таких что $\varepsilon \leqslant Z(W_m^{(i)}) \leqslant 1 - \varepsilon$ (см. [15,16]). Если это предположение верно, то можно показать, что полярные коды Арикана на двоичном стирающем канале имеют экспоненту масштабирования $\mu(BEC, K_2) = 3{,}627$. Точное значение экспоненты масштабирования для полярных кодов на других каналах неизвестно. Однако можно показать [15], что для любого симметричного канала без памяти с двоичным входом она удовлетворяет условию $3{,}579 \leqslant \mu(W, K_2) \leqslant 4{,}714$. Можно показать [3], что существуют $(\ell \times \ell)$-ядра $K_\ell$ с $\lim\limits_{\ell \to \infty} \mu(BEC, K_\ell) = 2$. Более того [17], существует ядро размера 64 с экспонентой масштабирования $\mu \approx 2{,}87$, что уже лучше, чем $\mu_{\text{SC-LDPC}} = 3$ – эвристическая оценка экспоненты масштабирования пространственно связанных LDPC-кодов [18]. В работе [5] было показано, что комбинируя несколько тщательно построенных локальных ядер, можно получить экспоненту масштабирования $\mu = 2 + \alpha$ для любого $\alpha > 0$ и любого симметричного канала без памяти с двоичным входом. Более того, для любого дискретного канала без памяти с пропускной способностью $I(W)$, любых $\pi, \rho > 0$, таких что $\pi + 2\rho < 1$, существуют полярноподобные коды длины $n$ с вероятностью ошибки на кодовое слово $e^{-n^\pi}$, скоростью $R = I(W) - n^{-\rho}$ и сложностью декодирования $O(n \log n)$ (см. [4]).

**2.4. Недвоичные полярные коды.** Полярные коды могут быть построены для каналов с $q$-ичными входными алфавитами, $q > 2$. Однако если поляризующее преобразование построено недостаточно тщательно, может возникнуть многоуровневая поляризация, т.е. пропускные способности подканалов могут сходиться к значениям, отличным от 0 или 1, $q$-ичных символов за одно использование канала, или поляризация может вообще не наблюдаться [19, 20]. В работе [21] было показано,

что обратимая $(\ell \times \ell)$-матрица $K$ над $GF(p^m)$ является поляризующей, если для любого $\overline{K} = VKP$ расширение поля $GF(p)$, порожденное присоединением всех элементов $\overline{K}$ к $GF(p)$, равно $GF(p^m)$, где $V$ – верхнетреугольная матрица, $P$ – матрица перестановок, а $\overline{K}$ – нижнетреугольная матрица.

## § 3. Конструкции поляризующих ядер

Задача нахождения поляризующего ядра состоит из двух подзадач:

1. Нахождение допустимого профиля частичных расстояний (ПЧР) $\mathcal{D}_0^{\ell-1}$;
2. Нахождение конкретного ядра с заданным ПЧР.

Считается, что ПЧР допустим, если существует ядро с таким ПЧР. Для получения необходимых условий допустимости ПЧР можно использовать классические верхние границы минимального расстояния линейных блоковых кодов [2, 22]. К сожалению, эти ограничения применимы только для случая монотонных ПЧР, т.е. $\mathcal{D}_0 \leqslant \mathcal{D}_1 \leqslant \ldots \leqslant \mathcal{D}_{\ell-1}$. В [23, 24] было показано, что ядра с немонотонным ПЧР допускают гораздо более простую обработку по сравнению с ядрами с монотонным ПЧР, имеющими ту же скорость поляризации. Поиск условий допустимости для немонотонных ПЧР остается открытой проблемой.

Даже для фиксированного ПЧР нахождение удовлетворяющего ему ядра в общем случае остается трудной задачей [25]. Из-за большого пространства поиска необходимо использовать некоторые методы для быстрого определения неподходящих и эквивалентных кандидатов. В частности, в [24] было предложено использовать таблицы весов лидеров смежных классов и инварианты эквивалентности кодов для сокращения пространства поиска и получения ядер с почти оптимальной скоростью поляризации.

Простым способом получения большого ядра является использование некоторого семейства вложенных алгебраических кодов [2,22]. Вложенные алгебраические коды были использованы в [17] совместно с компьютерным поиском для получения ядра размера 64 с экспонентой масштабирования $\mu \approx 2{,}87$. Имея некоторое поляризующее ядро, можно попытаться переставить его столбцы, чтобы уменьшить сложность различных алгоритмов обработки ядра [26, 27]. Имея хорошее поляризующее ядро, можно укоротить его, чтобы получить семейство меньших ядер, допускающих унифицированную реализацию обработчика [28].

## § 4. Обработка ядра

Важнейшей проблемой при реализации алгоритма ПИ является эффективное вычисление вероятностей (1). В [29] было предложено аппроксимировать эти значения[2] как

$$W_m^{(i)}\big(u_0^i \,|\, r_0^{n-1}\big) \approx \mathcal{W}_m^{(i)}\big(u_0^i \,|\, r_0^{n-1}\big) = \max_{u_{i+1}^{n-1}} \prod_{j=0}^{n-1} W_0^{(0)}\Big(\big(u_0^{n-1}K^{\otimes \ell}\big)_j \,|\, r_j\Big),$$

так что

$$\mathcal{W}_m^{(\ell i+s)}\big(u_0^{\ell i+s} \,|\, r_0^{n-1}\big) = \max_{u_{s+1}^{\ell-1}} \prod_{j=0}^{\ell-1} \mathcal{W}_{m-1}^{(i)}\Big(\big(u_{\ell t}^{\ell(t+1)-1}K\big)_j,\ t \in [i+1] \,|\, r_{j,\ell}^{n-1}\Big), \tag{4}$$

и $\mathcal{W}_0^{(0)}(c\,|\,r) = W(c\,|\,r)$. Отметим, что $\mathcal{W}_m^{(i)}(u_0^i \,|\, r_0^{n-1})$ – вероятность наиболее вероятного продолжения вектора $u_0^i$, не учитывающая никаких ограничений замораживания на символы $u_j$, $i < j < n$. В [29] было показано, что эта аппроксимация

[2] Здесь мы опускаем коэффициенты нормировки, так как они не влияют на декодирование.

обеспечивает существенное снижение средней сложности последовательного декодирования с незначительной потерей корректирующей способности.

Определим логарифмические отношения правдоподобия (ЛОПП)

$$S_m^{(i)}(u_0^{i-1}, r_0^{n-1}) = \ln \frac{\mathcal{W}_m^{(i)}\big(u_0^{i-1} \bullet 0 \,|\, r_0^{n-1}\big)}{\mathcal{W}_m^{(i)}\big(u_0^{i-1} \bullet 1 \,|\, r_0^{n-1}\big)}.$$

Далее для простоты ограничимся рассмотрением случая $m = 1$. Предполагая, что все $u_j$, $i < j < \ell$, принимают значения 0 и 1 с вероятностью $1/2$, можно заметить, что вычисление $\mathcal{W}_1^{(i)}(u_0^i \,|\, r_0^{\ell-1})$ эквивалентно декодированию по максимуму правдоподобия $r_0^{\ell-1}$ в смежном классе линейного кода, порожденного строками $i+1, \ldots, \ell-1$ ядра $K$, где представитель смежного класса задается линейной комбинацией с коэффициентами $u_0, \ldots, u_i$ верхних $i+1$ строк $K$. Пусть $\widehat{c}_0^{\ell-1}$ – жесткие решения, соответствующие $r_0^{\ell-1}$.

Видно, что

$$\begin{aligned}
S_1^{(i)}(u_0^{i-1}, r_0^{\ell-1}) &= \ln \frac{\max\limits_{u_{i+1}^{\ell-1}} \prod\limits_{j=0}^{\ell-1} W\Big(\big(\big(u_0^{i-1}, 0, u_{i+1}^{\ell-1}\big)K\big)_j \,|\, r_j\Big)}{\max\limits_{u_{i+1}^{\ell-1}} \prod\limits_{j=0}^{\ell-1} W\Big(\big(\big(u_0^{i-1}, 1, u_{i+1}^{\ell-1}\big)K\big)_j \,|\, r_j\Big)} = \\
&= \min_{u_{i+1}^{\ell-1}} \left( \sum_{j=0}^{\ell-1} \Big( \log W(\widehat{c}_j \,|\, r_j) - \log W\Big(\big(\big(u_0^{i-1}, 1, u_{i+1}^{\ell-1}\big)K\big)_j \,|\, r_j\Big)\Big)\right) - \\
&- \min_{u_{i+1}^{\ell-1}} \left( \sum_{j=0}^{\ell-1} \Big( \log W(\widehat{c}_j \,|\, r_j) - \log W\Big(\big(\big(u_0^{i-1}, 0, u_{i+1}^{\ell-1}\big)K\big)_j \,|\, r_j\Big)\Big)\right) = \\
&= \min_{u_{i+1}^{\ell-1}} M\Big(\big(u_0^{i-1}, 1, u_{i+1}^{\ell-1}\big)K, S_0^{\ell-1}\Big) - \min_{u_{i+1}^{\ell-1}} M\Big(\big(u_0^{i-1}, 0, u_{i+1}^{\ell-1}\big)K, S_0^{\ell-1}\Big) = \\
&= \frac{1}{2}\left( \max_{u_{i+1}^{\ell-1}} T\Big(\big(u_0^{i-1}, 0, u_{i+1}^{\ell-1}\big)K, S_0^{\ell-1}\Big) - \max_{u_{i+1}^{\ell-1}} T\Big(\big(u_0^{i-1}, 1, u_{i+1}^{\ell-1}\big)K, S_0^{\ell-1}\Big)\right), \qquad (5)
\end{aligned}$$

где $M(c_0^{\ell-1}, S_0^{\ell-1}) = \sum\limits_{j:\,(-1)^{c_j} S_j < 0} |S_j|$ – корреляционная невязка вектора $c_0^{\ell-1}$ относительно вектора ЛОПП $S_0^{\ell-1}$, $T(c_0^{\ell-1}, S_0^{\ell-1}) = \sum\limits_{j=0}^{\ell-1} (-1)^{c_j} S_j$ – корреляционная функция, а $S_j = \ln \dfrac{W(0 \,|\, r_j)}{W(1 \,|\, r_j)}$ – входные ЛОПП.

В симметричном канале имеем

$$M\big(\big(u_0^i, u_{i+1}^{\ell-1}\big)K, S_0^{\ell-1}\big) = M\big(u_{i+1}^{\ell-1} K[i+1 : \ell-1],\, \overline{S}_0^{\ell-1}\big),$$

где $\overline{S}_i = (-1)^{f_i} S_i$, $0 \leqslant i < \ell$, $f_i = u_0^i K[0:i]$, а $K[a:b]$ обозначает подматрицу матрицы $K$, состоящую из строк $a, a+1, \ldots, b$. Это позволяет переформулировать задачу вычисления выражения (5) следующим образом. Пусть $\overline{K}^{(i)}$ – матрица, полученная добавлением к $K[i : \ell-1]$ столбца $(1, 0, \ldots, 0)^T$. Пространство строк $\overline{K}^{(i)}$ называется $i$-м расширенным ядерным кодом $\overline{\mathcal{C}}^{(i)}$. Этот код имеет длину $\ell+1$ и размерность $\ell - i$. Предполагая, что $\overline{S}_\ell = 0$, получаем, что (5) может быть вычислено путем нахождения двух наиболее вероятных кодовых слов $i$-го расширенного ядерного кода, соответствующих $\overline{S}_0^\ell$, имеющих 0 и 1 в последнем символе. Это может быть реа-

лизовано с помощью алгоритма декодирования по максимуму правдоподобия для расширенного ядерного кода $\overline{C}^{(i)}$.

Примером такого алгоритма является алгоритм Витерби.

**4.1. Применение рекурсивных решеток для обработки больших поляризующих ядер.** Укажем следующие два подхода.

**Рекурсивное декодирование по максимуму правдоподобия.** Решетки позволяют эффективно реализовать декодирование линейных блоковых кодов по максимуму правдоподобия. Однако широко известный алгоритм Витерби не является оптимальным с точки зрения сложности. Существенное снижение сложности может быть получено при использовании рекурсивного алгоритма декодирования по максимуму правдоподобия [30]. Идея этого алгоритма заключается в рекурсивном разбиении полученного зашумленного вектора на несколько секций $[x, y)$, определении для каждой секции нескольких наиболее вероятных векторов $c_x^{y-1} \in \mathbb{F}_2^{y-x}$, соответствующих полученным значениям $r_x^{y-1}$, и их объединении для получения наиболее вероятных векторов для более длинных секций.

Пусть дан линейный блоковый код $C$, а также его подкод $C_{h,h'}$, такой что все его кодовые слова имеют ненулевые символы только в позициях $h \leqslant i < h'$. Пусть $p_{h,h'}(C)$ – линейный код, полученный путем выкалывания из кодовых слов $C$ всех символов, кроме тех, которые находятся в позициях $h \leqslant i < h'$. Далее определим $s_{h,h'}(C) = p_{h,h'}(C_{h,h'})$, т.е. код, полученный из $C$ путем его укорочения на все символы, кроме тех, которые имеют индексы $h \leqslant i < h'$. Коды $s_{h,h'}(C)$ и $p_{h,h'}(C)$ называются *секционными кодами*. Рассмотрим минимальную решетку кода $C$ и ее секции, соответствующие символам от $x$ до $y$. Можно показать [30], что пути между двумя соседними состояниями в этой секции соответствуют смежным классам в $p_{x,y}(C)/s_{x,y}(C)$. Эти смежные классы могут появляться в решетке несколько раз. Следовательно, можно упростить декодирование по максимуму правдоподобия (МП), предварительно вычислив метрики этих путей.

Более конкретно, для каждого смежного класса $D \in p_{x,y}(C)/s_{x,y}(C)$ необходимо определить наиболее вероятный элемент $\ell(D)$, т.е. элемент с минимальной корреляционной невязкой

$$M(D) = M\big(\ell(D), S_x^{y-1}\big),$$

где $S_i$, $x \leqslant i < u$, – ЛОПП, соответствующие декодируемому вектору. Пусть *таблица составных метрик* (*ТСМ*) $T_{x,y}$ – массив, содержащий значения $T_{x,y}[v].\ell = \ell(D)$ и $T_{x,y}[v].m = M(D)$, где $v$ – номер смежного класса $D$. В случае обычного декодирования $(n, k)$-кода $p_{0,n}(C)/s_{0,n}(C)$ содержит единственный элемент, поэтому соответствующая ТСМ имеет одну запись $T_{0,n}[0]$, что дает решение задачи МП-декодирования.

Непосредственный подход к построению ТСМ для некоторого кода $C$ заключается в переборе всех кодовых слов из $p_{x,y}(C)$, и нахождении наиболее вероятного из них для каждого смежного класса в $p_{x,y}(C)/s_{x,y}(C)$. Мы предполагаем, что этот метод используется для $y - x < 2$. Однако в [30] для случая $y - x \geqslant 2$ был предложен более эффективный подход. Пусть $z$ таково, что $x < z < y$. Предположим, что порождающая матрица $p_{x,y}(C)$ представлена в виде

$$G_{x,y}^{(p)} = \left(\begin{array}{cc} G_{x,z}^{(s)} & 0 \\ 0 & G_{z,y}^{(s)} \\ G_{x,y}^{(00)} & G_{x,y}^{(01)} \\ \hline G_{x,y}^{(10)} & G_{x,y}^{(11)} \end{array}\right), \tag{6}$$

где $G^{(s)}_{x,y} = \begin{pmatrix} G^{(s)}_{x,z} & 0 \\ 0 & G^{(s)}_{z,y} \\ G^{(00)}_{x,y} & G^{(01)}_{x,y} \end{pmatrix}$ – порождающая матрица $s_{x,y}(C)$, а $G^{(00)}_{x,y}$ и $G^{(01)}_{x,y}$ – некоторые матрицы размера $k'_{x,y} \times (z-x)$ и $k''_{x,y} \times (y-z)$ соответственно, где $k'_{x,y} = k'_{x,y}(C)$ и $k''_{x,y} = k''_{x,y}(C)$ – некоторые целые числа, зависящие от кода. Эти числа и матрицы могут быть получены из минимальной спэновой (minimum-span) формы порождающей матрицы кода $C$.

Существует взаимно-однозначное соответствие между векторами вида $vG'_{x,y}$, где $G'_{x,y} = \begin{pmatrix} G^{(10)}_{x,y} & G^{(11)}_{x,y} \end{pmatrix}$ – матрица размера $k'_{x,y} \times (y-x)$, и смежными классами $D \in \in p_{x,y}(C)/s_{x,y}(C)$. В дальнейшем $v$ используется в качестве номера смежного класса.

Видно, что

$$T_{x,y}[v].m = \min_{c_x^{y-1} \in D} M(c_x^{y-1}, r_x^{y-1}) = \min_{w \in \mathbb{F}_2^{k''_{x,y}}} (T_{x,z}[a].m + T_{z,y}[b].m), \quad v \in \mathbb{F}_2^{k'_{x,y}}, \quad (7)$$

где $a$ и $b$ – индексы смежных классов $D' \in p_{x,z}(C)/s_{x,z}(C)$ и $D'' \in p_{z,y}(C)/s_{z,y}(C)$ соответственно, таких что $\begin{pmatrix} w & v \end{pmatrix} \begin{pmatrix} G^{(00)}_{x,y} \\ G^{(10)}_{x,y} \end{pmatrix} \in D'$ и $\begin{pmatrix} w & v \end{pmatrix} \begin{pmatrix} G^{(01)}_{x,y} \\ G^{(11)}_{x,y} \end{pmatrix} \in D''$. Такие векторы $a, b$ могут быть определены из системы уравнений

$$\begin{pmatrix} a' & a \end{pmatrix} \begin{pmatrix} G^{(s)}_{x,z} \\ G'_{x,z} \end{pmatrix} = \begin{pmatrix} w & v \end{pmatrix} \begin{pmatrix} G^{(00)}_{x,y} \\ G^{(10)}_{x,y} \end{pmatrix},$$
$$\begin{pmatrix} b' & b \end{pmatrix} \begin{pmatrix} G^{(s)}_{z,y} \\ G'_{z,y} \end{pmatrix} = \begin{pmatrix} w & v \end{pmatrix} \begin{pmatrix} G^{(01)}_{x,y} \\ G^{(11)}_{x,y} \end{pmatrix},$$

где $a', b'$ – векторы, необходимые для обеспечения совместности этой системы, но не используемые далее. Решение этой системы имеет вид $a = \begin{pmatrix} w & v \end{pmatrix} \widehat{G}_{x,y}$, $b = \begin{pmatrix} w & v \end{pmatrix} \widetilde{G}_{x,y}$ для некоторых матриц $\widehat{G}_{x,y}$ и $\widetilde{G}_{x,y}$. Соответствующий наиболее вероятный представитель смежного класса равен $T_{x,y}[v].\ell = T_{x,z}[\widehat{a}].\ell \bullet T_{x,z}[\widehat{b}].\ell$, где $\widehat{a}, \widehat{b}$ – значения $a$ и $b$, дающие минимум в (7).

Сложность вычислений по такой схеме составляет $O\bigl(2^{k'_{x,y}+k''_{x,y}}\bigr)$. Она может быть дополнительно снижена за счет использования приемов, предложенных в [30]. Общая сложность декодирования сильно зависит от используемого метода секционирования, т.е. правила выбора точки разбиения $z$ для различных $x, y$. Этот подход, известный как рекурсивное декодирование по максимуму правдоподобия (РДМП), является более эффективным по сравнению с алгоритмом Витерби [30].

**Обработка ядра.** Значения (5) могут быть найдены путем применения алгоритма РДМП к смежным классам расширенных ядерных кодов. Пусть $\overline{\mathcal{C}}(u_0^{i-1}) = w + \overline{\mathcal{C}}^{(i)}$ – смежный класс $\overline{\mathcal{C}}^{(i)}$, задаваемый ранее принятыми решениями $u_0^{i-1}$, где $w = (u_0^{i-1} K[0 : i-1], 0)$. Для любой секции $[x, y)$ смежные классы, связанные с состояниями в рекурсивной решетке для $\overline{\mathcal{C}}(u_0^{i-1})$, получаются из смежных классов для $\overline{\mathcal{C}}^{(i)}$ как $D(u_0^{i-1}) = \{c + w_x^{y-1} \,|\, c \in D\}$, $D \in p_{x,y}(\overline{\mathcal{C}}^{(i)})/s_{x,y}(\overline{\mathcal{C}}^{(i)})$.

Поскольку представители смежных классов $\ell(D)$ не нужны в контексте задачи обработки ядра, в дальнейшем предполагается, что записи ТСМ содержат только значения $m$. Поэтому мы используем запись $T_{x,y}[v]$ вместо $T_{x,y}[v].m$. Также видно, что $p_{0,\ell}(\overline{\mathcal{C}}^{(i)})/s_{0,\ell}(\overline{\mathcal{C}}^{(i)})$ содержит два смежных класса (класса эквивалентности),

которые соответствуют $u_i = 0$ и $u_i = 1$. Следовательно,

$$S_1^{(i)}(u_0^{i-1}, r_0^{\ell-1}) = T_{0,\ell}[1] - T_{0,\ell}[0], \tag{8}$$

где $T_{0,\ell}$ – ТСМ, построенная для $\overline{\mathcal{C}}(u_0^{i-1})$ и заданного зашумленного вектора $r_0^{\ell-1}$.

Более того, мы предлагаем повторно использовать ТСМ или их части, полученные на последовательных фазах $i$. Для этого нам необходимо определить, как ТСМ эволюционируют с $i$, найти способ обработки предыдущих решений $u_0^{i-1}$, разработать эффективные алгоритмы построения ТСМ для коротких секций и найти оптимальную стратегию секционирования.

Предположим, что для всех фаз $i$ используется одно и то же секционирование. Очевидно, что $p_{x,y}(\overline{\mathcal{C}}^{(i+1)}) \subset p_{x,y}(\overline{\mathcal{C}}^{(i)})$ и $s_{x,y}(\overline{\mathcal{C}}^{(i+1)}) \subset s_{x,y}(\overline{\mathcal{C}}^{(i)})$, $i \in [\ell-1]$, для любых $x, y$, таких что $0 \leqslant x < y \leqslant \ell$.

Лемма 1. *Если* $p_{x,y}(\overline{\mathcal{C}}^{(i+1)}) = p_{x,y}(\overline{\mathcal{C}}^{(i)})$ *и* $s_{x,y}(\overline{\mathcal{C}}^{(i+1)}) = s_{x,y}(\overline{\mathcal{C}}^{(i)})$, *то для любого* $u_i \in \mathbb{F}_2$ *таблица составных метрик* $T_{x,y}$, *построенная для* $\overline{\mathcal{C}}(u_0^{i-1})$, *идентична ТСМ* $T'_{x,y}$, *построенной для* $\overline{\mathcal{C}}(u_0^{i})$ *для того же принятого вектора* $r_0^{\ell-1}$.

Из приведенной выше леммы следует, что не нужно пересчитывать ТСМ для тех секций, где секционные коды не меняются от фазы $i$ к фазе $i+1$.

В дальнейшем мы предполагаем, что

$$s_{x,z}(\mathcal{C}^{(i+1)}) = s_{x,z}(\mathcal{C}^{(i)}) \quad \text{и} \quad s_{z,y}(\mathcal{C}^{(i+1)}) = s_{z,y}(\mathcal{C}^{(i)}). \tag{9}$$

Будем также временно считать, что $u_0^{i-1} = 0$.

Даже если секционные коды меняются, все равно можно переиспользовать некоторые результаты, полученные на предыдущих фазах. Пусть $k'_{i,x,y} = k'_{x,y}(\mathcal{C}^{(i)})$ и $k''_{i,x,y} = k''_{x,y}(\mathcal{C}^{(i)})$. Во-первых, заметим, что если $k''_{i+1,x,y} = k''_{i,x,y}$, но $k'_{i+1,x,y} < k'_{i,x,y}$, то $p_{x,y}(\mathcal{C}^{(i+1)})/s_{x,y}(\mathcal{C}^{(i+1)}) \subset p_{x,y}(\mathcal{C}^{(i)})/s_{x,y}(\mathcal{C}^{(i)})$, так что соответствующая ТСМ на фазе $i+1$ может быть получена как подвектор ТСМ на фазе $i$.

Во-вторых, минимизацию можно производить рекурсивно и сохранять все промежуточные результаты. Этот подход был первоначально представлен в [23]. Более конкретно, перепишем (7) как

$$T_{x,y}[v] = \min_{w \in \mathbb{F}_2^{k''_{i,x,y}}} (T_{x,z}[a] + T_{z,y}[b]) = \min_{w_{k''_{i,x,y}-1}} \dots \min_{w_1} \min_{w_0} (T_{x,z}[a] + T_{z,y}[b]). \tag{10}$$

Вместо того чтобы хранить в ТСМ конечные результаты минимизации в (10), можно хранить промежуточные результаты минимизации для всех $w$. Эти значения могут быть представлены в виде двоичного дерева для каждого $v \in \mathbb{F}_2^{k'_{x,y}}$, так что путь от корня в этом дереве задается значениями $w_{k''_{i,x,y}-1}, w_{k''_{i,x,y}-2}, \dots, w_0$. Под лесом минимизации будем понимать множество таких деревьев, полученных на некоторой фазе для данной секции. Поддеревья внутри леса могут быть проиндексированы переменными $w, v$.

Описанное выше дерево минимизации, построенное на некоторой фазе $i_0$, может быть использовано для получения ТСМ для всех $i \geqslant i_0$, где выполняется (9). Пусть $i_1 > i_0$ – наименьшее целое число, при котором это не выполняется.

Лемма 2. *Пусть* $G''_{i,x,y} = \begin{pmatrix} G^{(00)}_{x,y} & G^{(01)}_{x,y} \end{pmatrix}$ *и* $G'_{i,x,y} = \begin{pmatrix} G^{(10)}_{x,y} & G^{(11)}_{x,y} \end{pmatrix}$ – *матрицы, полученные из* (6) *для кода* $\mathcal{C}^{(i)}$. *Если все матрицы* $G''_{i,x,y}$ *являются вложенными, так что* $G''_{i+1,x,y}$ *занимает верхние строки* $G''_{i,x,y}$ *для любого* $i$, *такого что* $i_0 \leqslant i < i_1$, *то лес минимизации для фазы* $i$, $i_0 \leqslant i \leqslant i_1$, *можно получить, взяв поддеревья, задаваемые векторами* $w$ *и* $v$, *деревьев в лесу, построенном на фазе* $i_0$,

*где*

$$\begin{pmatrix} w & v \end{pmatrix} \begin{pmatrix} G''_{i_0,x,y} \\ G''_{i_0,x,y} \end{pmatrix} = \begin{pmatrix} \overline{w} & \overline{v} \end{pmatrix} \begin{pmatrix} G''_{i,x,y} \\ G''_{i,x,y} \end{pmatrix}, \tag{11}$$

*а* $\overline{w} \in \mathbb{F}_2^{k''_i,x,y}, \overline{v} \in \mathbb{F}_2^{k''_i,x,y}$ *обозначают индексы поддеревьев в лесу на фазе* $i$.

Матрицу $G''_{i,x,y}$ всегда можно привести к форме, требуемой леммой 2, с помощью элементарных операций над строками.

Декодер ПИ должен учитывать на фазе $i$ значения $u_0^{i-1}$. Это сводится к декодированию в смежных классах секционных кодов. Соответствующий представитель смежного класса для секции $[x, y)$ может быть вычислен как линейная комбинация подвекторов ядра $\begin{pmatrix} K_{j,x} & K_{j,x+1} & \dots & K_{j,y-1} \end{pmatrix}$, $0 \leqslant j < i$.

Рассмотрим сначала случай секции $[x, y)$, где лес минимизации (10) строится заново. Если существует решение

$$\begin{pmatrix} f_{j,x,y} & h_{j,x,y} \end{pmatrix} \begin{pmatrix} G^{(s)}_{i,x,y} \\ G'_{i,x,y} \end{pmatrix} = \begin{pmatrix} K_{j,x} & K_{j,x+1} & \dots & K_{j,y-1} \end{pmatrix}, \tag{12}$$

то тогда ТСМ для секции $[x, y)$ содержит в позиции $h_{j,x,y}$ запись, соответствующую требуемому представителю смежного класса. В этом случае полагаем $h_{j,x',y'} = 0$ для всех $x < x' < y' < y$. В противном случае будем считать, что $h_{j,x,y} = 0$. Имея вектор ранее принятых решений $u_0^{i-1}$, соответствующее смещение позиции на секции $[x, y)$ может быть вычислено как $h_{x,y} = \sum\limits_{j=0}^{i-1} u_j h_{j,x,y}$, так что (10) преобразуется в

$$T_{x,y}[v] = \min_w \left( T_{x,z}[a + h_{x,z}] + T_{z,y}[b + h_{z,y}] \right). \tag{13}$$

Значения $h_{x,y}$ аналогичны частичным суммам, которые возникают в декодере ПИ полярных кодов Арикана.

Для тех секций, где ТСМ получается путем взятия поддеревьев (10), необходимо проверить, существует ли решение

$$\omega_{j,x,y} \begin{pmatrix} G^{(s)}_{i_0,x,y} \\ G'_{i_0,x,y} \end{pmatrix} = \begin{pmatrix} K_{j,x} & K_{j,x+1} & \dots & K_{j,y-1} \end{pmatrix}. \tag{14}$$

Если это уравнение не имеет решения для некоторого $j < i$, то соответствующие представители смежных классов уже учтены на меньших секциях при построении ТСМ для секции $[x, y)$ на фазе $i_0$, и можно считать, что $\omega_{j,x,y} = 0$. В противном случае, соответствующая ТСМ может быть получена из леса минимизации, построенного на фазе $i_0 < i$, как записи с номерами

$$\begin{pmatrix} w & v \end{pmatrix} = \begin{pmatrix} \overline{w} & \overline{v} \end{pmatrix} M_{i,x,y} + \sum_{j=0}^{i-1} u_i \omega_{j,x,y}.$$

Видно, что результат (8) не меняется, если из всех записей ТСМ на любом участке вычесть одно и то же значение. Это позволяет построить некоторые ТСМ с меньшей сложностью по сравнению с (13). Можно выделить следующие особые случаи:

1. Рассмотрим случай $k'_{i,x,y} = k''_{i,x,y} = 1$, $\widehat{G}_{i,x,y} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\widetilde{G}_{i,x,y} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Это соответствует построению ТСМ с записями

$$T = [\min(w_{00} + w_{01}, w_{10} + w_{11}), \min(w_{10} + w_{01}, w_{00} + w_{11})],$$

где

$$w_{ij} = \begin{cases} 0, & i = \widehat{c}_j, \\ |S_j|, & i \neq \widehat{c}_j, \end{cases}$$

а $\widehat{c}_j$ – жесткое решение, соответствующее ЛОПП $S_j$. Здесь $S_j$ может быть либо канальным ЛОПП, либо величиной $T'[1] - T'[0]$, где $T'$ – ТСМ из двух элементов, соответствующая левой или правой подсекции секции $[x, y)$.

Вычтем из обоих элементов $T$ значение $w_{10} + w_{11}$. Видно, что в результате получается $\widetilde{T} = [\min(-S_0 - S_1, 0), \min(-S_1, -S_0)]$. Далее вычтем из обоих элементов $\widetilde{T}$ значение $\min(-S_0 - S_1, 0)$. В результате получим

$$\widehat{T} = [0, \min(-S_1, -S_0) - \min(-S_0 - S_1, 0)] = [0, \operatorname{sgn}(S_0) \operatorname{sgn}(S_1) \min(|S_0|, |S_1|)].$$

Этот прием позволяет построить ТСМ для секции, используя только одну операцию сравнения, вместо двух сравнений и четырех сложений для непосредственной реализации. Отметим, что $\widehat{T}[1]$ может быть также использован как $S_j$ в аналогичной схеме в большей секции.

2. $k'_{i,x,y} = 1$, $k''_{i,x,y} = 0$, $\widehat{G}_{i,x,y} = \widetilde{G}_{i,x,y} = (1)$, что соответствует построению ТСМ с записями $T_{x,y} = [T_{x,z}[h_{x,z}] + T_{z,y}[h_{z,y}], T_{x,z}[h_{x,z} + 1] + T_{z,y}[h_{z,y} + 1]]$. Это можно преобразовать к

$$\widehat{T}_{x,y} = [0, (-1)^{h_{x,z}}(T_{x,z}[1] - T_{x,z}[0]) + (-1)^{h_{z,y}}(T_{z,y}[1] - T_{z,y}[0])].$$

Если можно гарантировать, что $T_{x,z}[0] = T_{z,y}[0] = 0$, то $\widehat{T}_{x,y}$ может быть построен с помощью всего одного сложения. Если $z - x = 1$ или $y - z = 1$, будем считать, что $T_{x,z} = [0, S_x]$ или $T_{z,y} = [0, S_z]$ соответственно, где $S_j$ – соответствующий ЛОПП.

3. $k'_{i,x,y} = 0$, $k''_{i,x,y} = 2$, $\widehat{G}_{i,x,y} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\widetilde{G}_{i,x,y} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Пусть $a = T_{x,z}[1] - T_{x,z}[0]$, $b = T_{z,y}[1] - T_{z,y}[0]$. Эти значения можно получить без вычислений, если выполняется $T_{x,z}[0] = T_{z,y}[0] = 0$. Пусть $\widehat{a}$ и $\widehat{b}$ – жесткие решения, соответствующие $a$ и $b$. Тогда получаем

$$T_{x,y}\big[\widehat{a}, \widehat{b}\big] = 0, \qquad T_{x,y}\big[1 \oplus \widehat{a}, 1 \oplus \widehat{b}\big] = |a| + |b|,$$
$$T_{x,y}\big[1 \oplus \widehat{a}, \widehat{b}\big] = |a|, \qquad T_{x,y}\big[\widehat{a}, 1 \oplus \widehat{b}\big] = |b|.$$

Это можно вычислить всего за одну операцию.

Можно легко проверить, что в случае $K = BK_2^{\otimes m}$, где $B$ – матрица перестановки обращения битов, и равномерного секционирования (т.е. для всех $x, y$ используется $z = (x + y)/2$) ТСМ для всех секций могут быть построены с помощью описанных выше приемов 1 и 2. Более того, применение этих приемов в процессе работы алгоритма ПИ в точности повторяет min-sum версию алгоритма ПИ для полярных кодов Арикана [31].

Общая сложность обработки ядра предложенным алгоритмом задается величиной

$$\boldsymbol{C} = \sum_{i=0}^{\ell-1} (\delta_i + c_{i,0,\ell}),$$

где $\delta_i$ – сложность вычисления конечного ЛОПП из полученной ТСМ, а $c_{i,x,y}$ – сложность построения ТСМ для участка $[x, y)$ на фазе $i$. В общем случае первая операция сводится к вычислению (8), т.е. требует одного вычитания. Однако если

на участке $[0, \ell)$ возникают описанные выше особые случаи 1 или 2, то $\delta_i = 0$. Более того,

$$c_{i,x,y} = \begin{cases} m_{ixy}, & \text{если ТСМ для подсекций переиспользуются,} \\ m_{ixy} + c_{i,x,z} + c_{i,z,y} & \text{в противном случае,} \end{cases}$$

где

$$m_{ixy} = \begin{cases} 0, & \text{если переиспользование леса возможно,} \\ 1, & \text{если встречается особый случай,} \\ 2^{k'_{ixy}+k''_{ixy}} + 2^{k'_{ixy}}(2^{k''_{ixy}} - 1) & \text{в противном случае.} \end{cases}$$

Первый и второй члены в последнем выражении представляют собой количество сложений и сравнений соответственно. Для каждой секции $[x, y)$ необходимо найти позицию разбиения $z$, которая минимизирует общую сложность. Это можно сделать заранее с помощью алгоритма оптимизации, приведенного в [30]. Отметим, что секционирование должно быть оптимизировано совместно для всех фаз $i$. Этот алгоритм использует метод динамического программирования для нахождения оптимального $z$ и соответствующей сложности $c_{x,y} = \sum_{i=0}^{\ell-1} c_{i,x,y}$ для каждой допустимой комбинации $x$ и $y$. Результаты, полученные для коротких секций, сохраняются и повторно используются для более длинных секций. Такая оптимизация может значительно снизить общую сложность обработки. Было показано, что данный подход обеспечивает существенное снижение сложности обработки по сравнению с реализацией на основе алгоритма Витерби [32].

Следует признать, что описанный выше подход не требует решения каких-либо систем уравнений во время декодирования. Он сводится к сложению и сравнению элементов некоторых массивов, где индексы аргументов этих операций получаются как XOR некоторых предварительно вычисленных значений и смещений, задаваемых частичными суммами решений алгоритма ПИ.

**4.2. Оконный метод обработки.** Ядро Арикана допускает исключительно простую обработку. Действительно, (1) и (4) в случае ядра Арикана включают не более двух членов. Можно попытаться использовать эти простые выражения для обработки ядер большей размерности с лучшими поляризующими свойствами. Рассмотрим для простоты случай ядра $K$ размерности $\ell = 2^\mu$. Поскольку матрица $K$ обратима, можно выразить выходной вектор ядра как $c_0^{\ell-1} = u_0^{\ell-1}K = f_0^{\ell-1}K_2^{\otimes\mu}$, где $u_0^{\ell-1} = f_0^{\ell-1}T$, а $T = K_2^{\otimes\mu}K^{-1}$ – матрица перехода. Действительно, векторы $f_0^{q-1}$ и $u_0^{q-1}$ удовлетворяют системе уравнений

$$\underbrace{(S \quad I)}_{\Theta'}(u_{\ell-1} \quad \dots \quad u_1 \quad u_0 \quad f_0 \quad f_1 \quad \dots \quad f_{\ell-1})^T = 0,$$

где $(\ell \times \ell)$-матрица $S$ получается транспонированием $T^{-1}$ и записью столбцов в полученной матрице в обратном порядке. Применяя элементарные операции со строками, матрицу $\Theta'$ можно преобразовать к минимальной спэновой форме $\Theta$, где $i$-я строка начинается в $i$-м столбце и заканчивается в столбце $z_i$, где все $z_i$ различны, и $\Theta_{i,z_i} = 1$. Это позволяет получить ограничения динамического замораживания в виде

$$f_{\omega_i} = \sum_{s=0}^{i} u_s \Theta_{\ell-1-i,\ell-1-s} + \sum_{t=0}^{\omega_i-1} f_t \Theta_{\ell-1-i,\ell+t}, \tag{15}$$

где $\omega_i = z_{\ell-1-i} - \ell$. Следовательно,

$$W_1^{(i)}(u_0^i \,|\, r_0^{\ell-1}) = \sum_{f_0^{h_i} \in \mathcal{Z}(u_0^i)} \widetilde{W}_\mu^{(h_i)}(f_0^{h_i} \,|\, r_0^{\ell-1}), \tag{16}$$

где $\widetilde{W}_\mu^{(i)}$ – вероятности путей (1), полученные для случая ядра Арикана $K_2$, суммирование производится по множеству $\mathcal{Z}(u_0^i)$ векторов $f_0^{h_i}$, удовлетворяющих (15), и

$$h_i = \max_{0 \leqslant i' \leqslant i} \omega_{i'}.$$

Вычисление (16) сводится к перебору возможных значений символов $u_s$, $s \in \in [h_i + 1] \setminus \{\omega_t \,|\, 0 \leqslant t \leqslant i\}$. Число таких символов равно $\delta_i = h_i - i$. Это число называется размером окна. Такой перебор может быть реализован с помощью алгоритма списочного декодирования Таля – Варди с размером списка $2^{\max_i \delta_i}$, так что операции удаления пути не выполняются.

Аналогично (4) можно заменить суммирование в (16) на максимизацию и преобразовать алгоритм в область ЛОПП. Это приводит к следующим выражениям для ЛОПП, задаваемых матрицей Арикана:

$$\widetilde{S}_\mu^{(2i)}\big(f_0^{2i-1}, r_0^{2^\mu-1}\big) = \operatorname{sgn}(a)\operatorname{sgn}(b)\min(|a|,|b|), \tag{17}$$

$$\widetilde{S}_\mu^{(2i+1)}\big(f_0^{2i}, r_0^{2^\mu-1}\big) = (-1)^{u_{2i}} a + b, \tag{18}$$

где $a = \widetilde{S}_{\mu-1}^{(i)}(f_{0,2}^{2i-1} + f_{1,2}^{2i-1}, r_{0,2}^{2^\mu-1})$, $b = \widetilde{S}_{\mu-1}^{(i)}(f_{1,2}^{2i-1}, r_{1,2}^{2^\mu-1})$, а $\widetilde{S}_0^{(0)}(r_i) = S_i$ – входные ЛОПП. Вес пути, рассматриваемого min-sum версией декодера Таля – Варди, определяется [33] как

$$R\big(f_0^i, r_0^{\ell-1}\big) = R\big(f_0^{-1}, r_0^{\ell-1}\big) + \tau\Big(\widetilde{S}_\mu^{(i)}\big(f_0^{i-1}, r_0^{\ell-1}\big), f_i\Big), \tag{19}$$

где $R(r_0^{\ell-1}) = 0$ и

$$\tau(S, c) = \begin{cases} 0, & (-1)^c S \geqslant 0, \\ |S|, & (-1)^c S < 0, \end{cases}$$

является штрафной функцией. Можно показать [34], что

$$R\big(f_0^i, r_0^{\ell-1}\big) = \rho - \ln \max_{f_{i+1}^{\ell-1}} \widetilde{W}_\mu^{(\ell-1)}\big(f_0^{\ell-1} \,|\, r_0^{\ell-1}\big),$$

где $\rho$ не зависит от $f_0^i$. Кроме того, $R(f_0^{\ell-1}, r_0^{\ell-1}) = M(f_0^{\ell-1} K_2^{\otimes\mu}, S_0^{\ell-1})$. Это позволяет [35] вычислить ЛОПП для входных символов ядра как

$$S_1^{(i)}\big(u_0^{i-1}, r_0^{\ell-1}\big) = \min_{f_0^{h_i} \in \mathcal{Z}(u_0^{i-1}, 1)} R\big(f_0^{h_i}, r_0^{\ell-1}\big) - \min_{f_0^{h_i} \in \mathcal{Z}(u_0^{i-1}, 0)} R\big(f_0^{h_i}, r_0^{\ell-1}\big). \tag{20}$$

Сложность этого вычисления растет экспоненциально по $2^{\delta_i}$. Однако возможны некоторые упрощения [35]:

1. Для некоторых $i$ могут существовать более эффективные методы оценки величин $R(f_0^{h_i}, r_0^{\ell-1})$, $f_0^{h_i} \in \mathcal{Z}(u_0^{i-1}, b)$, $b \in \mathbb{F}_2$, по сравнению с (17)–(19). Например, быстрое преобразование Адамара может быть использовано, если незамороженные символы $s \in [h_i + 1] \setminus \{\omega_t \,|\, 0 \leqslant t \leqslant i\}$ задают код Рида – Маллера первого порядка или схожий с ним код.
2. Число различных значений ЛОПП $\widetilde{S}_\nu^{(j)}$, $0 < \nu < \mu$, может быть намного меньше, чем $2^{\delta_i}$ – число путей, рассматриваемых декодером Таля – Варди. Причина

этого в том, что векторы частичных сумм $f_{0,2}^{2i-1} + f_{1,2}^{2i-1}$ и $f_{1,2}^{2i-1}$ на различных уровнях $\mu$ могут образовывать некоторое подпространство соответствующего линейного пространства. Эти значения могут быть предварительно вычислены и использованы для различных путей в декодере Таля – Варди.

3. Минимизация в (20) может быть выполнена с сохранением промежуточных результатов в виде дерева, которые могут быть повторно использованы на последующих фазах.

Можно целенаправленно строить ядра с заданной скоростью поляризации и достаточно малым размером окна [36]. Основной недостаток оконного алгоритма обработки заключается в том, что пока нет его обобщения на случай ядер размерности, отличной от $2^\mu$, а его сложность может быть довольно высокой, если ядро не построено должным образом.

**4.3. Приближенные методы обработка ядра.** В [29] было предложено использовать так называемый алгоритм декодирования box-and-match [37] для поиска кодовых слов кода $\overline{\mathcal{C}}^{(i)}$, минимизирующих корреляционную невязку в (5). Сложность этого метода может быть уменьшена за счет повторного использования результатов метода Гаусса на разных фазах. Более того, сортировка входных ЛОПП может быть выполнена однократно в каждом обработчике ядра.

В качестве альтернативы можно использовать описанный выше оконный алгоритм обработки с ограниченным размером списка, т.е. оставить только несколько векторов $f_0^i$ с наибольшей оценкой $R(f_0^i, r_0^{\ell-1})$ (см. [8,38]).

## § 5. Построение кодов

**5.1. Оценка надежности битовых подканалов.** Классические полярные коды предполагают, что множество замораживания $\mathcal{F}$ содержит индексы ненадежных битовых подканалов. Существует несколько подходов к оценке надежности битовых подканалов в случае ядра Арикана, в том числе моделирование методом Монте-Карло, рекурсивные выражения для двоичного стирающего канала [1], гауссовская аппроксимация [39], аппроксимация стохастически улучшенными/ухудшенными каналами [40], эволюция плотностей [41, 42]. Некоторые из этих методов допускают относительно простое обобщение на случай кодов с большими ядрами.

**Метод Монте-Карло.** Наиболее простым способом оценки вероятности ошибки в битовых подканалах является использование идеального декодера ПИ. Более конкретно, произвольным образом[3] генерируется вектор $u_0^{n-1}$, вычисляется вектор $c_0^{n-1} = u_0^{n-1} K^{\otimes m}$, который передается по каналу, вычисляются $W_m^{(i)}(u_0^i \,|\, r_0^{n-1})$, $u_i \in \{0,1\}$, $0 \leqslant i < n$, или $S_m^{(i)}(u_0^{i-1}, r_0^{n-1})$, и принимается решение $\widehat{u}_i$ на основе полученных значений. Если $\widehat{u}_i \neq u_i$, то значение $i$-го счетчика ошибок увеличивается. Необходимо отметить, что оценки $\widehat{u}_i$ не используются на последующих фазах. Указанные действия выполняются достаточно большое число раз, после чего вычисляется частота ошибок в битовых подканалах. Недостатком этого метода является его высокая вычислительная сложность.

**Построение кодов для двоичного стирающего канала.** Если кодовые слова полярного кода передаются по двоичному стирающему каналу, то все $W_m^{(i)}$ также являются двоичными стирающими каналами. Это позволяет анализировать их поведение, изучая свойства только одного ядра.

Конфигурации стирания могут быть связаны с двоичными векторами $e$, такими, что $e_i = 1$, если стирается $i$-й символ. Конфигурация стирания $e \in \mathbb{F}_2^\ell$ на фазе $i$ неисправима, если существуют векторы $u_{i+1}^{\ell-1}, v_{i+1}^{\ell-1}$, такие что для всех $u_0^{i-1}$ и для всех $j$,

[3] Для симметричных каналов достаточно рассмотреть $u_0^{n-1} = 0$.

таких что $e_j = 0$, выполняется $((u_0^{i-1}, 0, u_{i+1}^{\ell-1})K)_j = ((u_0^{i-1}, 1, v_{i+1}^{\ell-1})K)_j$, т.е. выходные символы ядра невозможно различить на нестертых позициях. Пусть $E_{i,w}$ – число неисправимых конфигураций стирания с весом $w$ на фазе $i$. Тогда вероятность стирания (т.е. параметр Бхаттачарьи) в $W_1^{(i)}$ может быть вычислена как

$$f_i(z) = \sum_{w=0}^{\ell} E_{i,w} z^w (1-z)^{\ell-w},$$

где $z$ – вероятность стирания в исходном канале. Набор функций $f_0(z), \ldots, f_{\ell-1}(z)$ называется поляризационным поведением ядра. Оно может быть вычислено из решеток кодов $C^{(i)}$, $0 \leqslant i < \ell$ (см. [17, 43]).

Получив поляризационное поведение ядра, можно вычислить параметры Бхаттачарьи подканалов, задаваемых $m$-слойным поляризационным преобразованием, как

$$Z_{m,\ell i+s} = f_s(Z_{m-1}, i), \quad 0 \leqslant i < \ell^{m-1}, \quad 0 \leqslant s < \ell,$$

где $Z_{0,0}$ – параметр Бхаттачарьи исходного двоичного стирающего канала. Поляризационное поведение также может быть использовано для вычисления экспоненты масштабирования ядра $K$ (см. [16]).

Пример 1. Рассмотрим случай ядра Арикана $K_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. Для фазы 0 имеются следующие неисправимые конфигурации стирания: (10), (01), (11). Для фазы 1 существует единственная неисправимая конфигурация стирания – (11). Следовательно,

$$Z_{m,2i} = f_0(Z_{m-1,i}) = 2Z_{m-1,i}(1 - Z_{m-1,i}) + Z_{m-1,i}^2 = 2Z_{m-1,i} - Z_{m-1,i}^2, \tag{21}$$

$$Z_{m,2i+1} = f_1(Z_{m-1,i}) = Z_{m-1,i}^2. \tag{22}$$

Этот подход может быть легко обобщен на случай недвоичных полярных кодов.

**5.2. Гауссовская аппроксимация.** Симметричная пропускная способность $i$-го подканала, задаваемого ядром $K$, может быть вычислена [44] как

$$I_1^{(i)} = 1 - \int_{-\infty}^{\infty} p_i(\xi \,|\, 0) \log_2(1 + e^{-\xi}) d\xi, \tag{23}$$

где $p_i(\xi \,|\, 0)$ – функция плотности вероятности отношения правдоподобия $L_1^{(i)} = \ln \dfrac{W_1^{(i)}(R_0^{\ell-1}, 0 \,|\, 0)}{W_1^{(i)}(R_0^{\ell-1}, 0 \,|\, 1)}$ при условии того, что передаются нулевые символы, а $R_0^{\ell-1}$ – случайные величины, соответствующие выходу канала. Такой интеграл может быть вычислен методом Монте-Карло, т.е. путем подачи случайных векторов $R_0^{\ell-1}$ на блок вычисления ЛОПП и усреднения соответствующих значений.

Оказывается, что подстановка вместо $L_1^{(i)}$ приблизительных ЛОПП, задаваемых выражением (5), приводит к недостоверным результатам (например, $I_1^{(i)} < 0$). Вместо этого можно аппроксимировать (23) как

$$I_1^{(i)} \approx 1 - \int_{-\infty}^{\infty} f_i(\psi \,|\, 0) \log_2 \left(1 + \frac{\mathbf{P}\{u_i = 1 \,|\, \psi\}}{\mathbf{P}\{u_i = 0 \,|\, \psi\}}\right) d\psi =$$

$$= 1 - \int_{-\infty}^{\infty} f_i(\psi \,|\, 0) \log_2 \left(1 + \frac{f_i(-\psi \,|\, 0)}{f_i(\psi \,|\, 0)}\right) d\psi, \tag{24}$$

где $f_i(\psi\,|\,0)$ – функция плотности вероятности $S_1^{(i)}(\mathbf{0}\,|\,R_0^{\ell-1})$, а $\mathbf{P}\{u_i = c\,|\,\psi\}$ – вероятность события $u_i = c$ при условии $S_1^{(i)}(\mathbf{0}\,|\,R_0^{\ell-1}) = \psi$. Следовательно, можно оценить $I_1^{(i)}$ путем построения гистограммы для $f_i(\psi\,|\,0)$ на основе результатов работы алгоритма, вычисляющего (5).

Для построения полярного кода с некоторым ядром $K$ можно предположить, что все подканалы $W_\lambda^{(i)}$, $0 \leqslant \lambda \leqslant m$, $0 \leqslant i < \ell^\lambda$, являются гауссовскими, так что они могут быть полностью охарактеризованы соответствующей взаимной информацией $I_\lambda^{(i)}$. В [45] было предложено построить таблицы $I_1^{(i)}(C)$ значений $I_1^{(i)}$ для некоторого конечного набора параметров АБГШ-канала $W$ с двоичным входом, где $C$ – пропускная способность канала $W$. Эти таблицы могут быть использованы для интерполяции значений $I_1^{(i)}(C)$ для любого $C \in [0,1]$.

Для построения полярного $(\ell^m, k)$-кода для АБГШ-канала с пропускной способностью $C$ можно рекурсивно вычислить пропускные способности битовых подканалов

$$I_m^{(\ell i+j)}(C) \approx I_1^{(j)}(I_{m-1}^{(i)}(C)), \quad 0 \leqslant j < \ell, \quad m > 1, \tag{25}$$

где $I_0^{(0)}(C) = C$, и объявить замороженными символы $u_i$, соответствующие подканалам с наименьшим $I_m^{(i)}(C)$. Этот подход может быть обобщен на случай кодов с недвоичными ядрами [46].

**5.3. Коды с улучшенными дистантными свойствами.** Полярные коды с большими ядрами имеют достаточно малое минимальное расстояние. Вследствие этого их корректирующая способность даже при декодировании списочным алгоритмом ПИ оказывается недостаточной. Она может быть улучшена путем использования полярных подкодов, полярных кодов с CRC [10] или других каскадных конструкций.

В большинстве случаев наилучшая корректирующая способность достигается при использовании полярных подкодов. Их построение опирается на концепцию динамических замороженных символов. Вместо того чтобы задавать $u_{j_i} = 0$, $j_i \in \mathcal{F}$, $0 \leqslant i < n-k$, можно потребовать, чтобы замороженные символы удовлетворяли условиям

$$u_{j_i} = \sum_{s<j_i} V_{i,s} u_j, \quad j_i \in \mathcal{F}, \tag{26}$$

где $V$ – матрица ограничений размера $(n-k) \times n$, такая что последние ненулевые элементы ее строк находятся в различных столбцах $j_i \in \mathcal{F}$. Символы $u_{j_i}$ с хотя бы одним ненулевым членом в правой части (26) называются динамически замороженными (DFS), а символы с $V_{i,s} = 0$, $s < j_i$, – статически замороженными. Такие коды называются полярными подкодами [47]. Их декодирование может быть реализовано простым обобщением списочного алгоритма ПИ.

Один из способов построения полярного $(n = \ell^m, k, d)$-подкода состоит в том, чтобы взять базовый $(n, k' > k, d)$-код с проверочной матрицей $H$, построить для него матрицу ограничений $V' = QH(K^{\otimes m})^T$, где обратимая матрица $Q$ выбирается так, чтобы последние ненулевые элементы строк $V'$ располагались в различных столбцах $j_0, \ldots, j_{n-k'-1}$, и получить $V = \begin{pmatrix} V' \\ V'' \end{pmatrix}$. Здесь матрица $V''$ содержит $k'-k$ строк веса 1, ненулевые элементы которых расположены в столбцах $j_i$, соответствующих наименее надежным битовым подканалам $W_m^{(j_i)}$, где $j_i \notin \{j_0, \ldots, j_{n-k'-1}\}$, $n-k' \leqslant i < n-k$. В качестве базовых кодов в сочетании с расширенными ядрами БЧХ могут быть выбраны, например, расширенные коды БЧХ [47].

Лучшую корректирующую способность можно получить с помощью рандомизированной конструкции, предложенной в [48]. Эта конструкция позволяет также комбинировать различные ядра в поляризующем преобразовании. Эта конструкция опирается на следующее утверждение.

Теорема 1. *Рассмотрим полярный $(n, k)$-код $\mathcal{C}$, заданный поляризующим преобразованием $A = K_{\ell_0} \otimes \ldots \otimes K_{\ell_{m-1}}$ и множеством замораживания $\mathcal{F}$, где $K_{\ell_i}$ – ядро размерности $\ell_i$, частичные расстояния $\mathcal{D}_{i,j}$ которого удовлетворяют условиям*

$$\mathcal{D}_{i,j} = \mathrm{wt}(K_{\ell_i}[j]), \quad 0 \leqslant j < \ell_i, \quad 0 \leqslant i < m, \tag{27}$$

$n = \prod_{i=0}^{m-1} \ell_i$, *а $K_{\ell_i}[j]$ – $j$-я строка матрицы $K_{\ell_i}$. Тогда*:

1. *Минимальное расстояние полярного кода равно $d = \min\limits_{s \notin \mathcal{F}} \mathrm{wt}(A[s])$, где $A[s]$ – $s$-я строка матрицы $A$, $0 \leqslant s < n$;*
2. *Любое кодовое слово $c_0^{n-1} = u_0^{n-1} A \in \mathcal{C}$ веса $d$, где $d$ – минимальное расстояние кода $\mathcal{C}$, имеет $u_s = 1$ для некоторого $s$, такого что $\mathrm{wt}(A_s) = d$.*

Чтобы получить $(n, k)$-код с хорошей корректирующей способностью при декодировании списочным алгоритмом ПИ, необходимо исключить из полярного кода ненулевые кодовые слова с малым весом (НКСМВ). Это можно сделать, введя ограничения динамического замораживания (26). Чтобы получить $(n, k, d)$-код $\mathcal{C}$, можно построить сначала полярный $(n, k+f_A, \overline{d})$-код $\overline{\mathcal{C}}$ (родительский код) с матрицей ограничений $\overline{V}$, и построить матрицу ограничений для кода $\mathcal{C}$ как $V = \begin{pmatrix} \overline{V} \\ V^{(A)} \end{pmatrix}$. Здесь $V^{(A)}$ – $(f_A \times n)$-матрица, которая определяет такие ограничения динамического замораживания, что большинство НКСМВ родительского кода не удовлетворяют им, т.е. $\mathcal{C}$ не содержит этих кодовых слов. Определим $\overline{\mathcal{F}}$ как множество индексов замороженных символов, заданное матрицей $\overline{V}$.

Чтобы уменьшить вероятность того, что правильный путь будет исключен из рассмотрения списочным декодером ПИ на ранних фазах, эти ограничения замораживания должны быть наложены на символы $u_{s_i}$ с наименьшими возможными индексами $s_i$ таким образом, чтобы кодовые слова с малым весом были исключены. Из теоремы 1 следует, что это можно сделать, обобщив конструкцию [49], т.е. выбрав в качестве $s_i$ $f_A$ максимальных индексов, таких, что $\mathrm{wt}(A_{s_i}) = \overline{d}$, где $\overline{d}$ – минимальное расстояние родительского полярного кода $\overline{\mathcal{C}}$, и выбрав $V_{i,j}^{(A)}$, $0 \leqslant j < s_i$, $0 \leqslant i < f_A$, в качестве независимых случайных двоичных величин. Кроме того, положим $V_{i,s_i}^{(A)} = 1$ и $V_{i,j}^{(A)} = 0$, $j > s_i$. Действительно, такой выбор $s_i$ гарантирует, что для любого НКСМВ $c_0^{n-1} = u_0^{n-1} A \in \overline{\mathcal{C}}$ можно найти такие значения $V_{i,j}^{(A)}$, что $u_0^{n-1}$ не удовлетворяет (26), т.е. $c_0^{n-1} \notin \mathcal{C}$. Для малых значений $f_A$ может оказаться невозможным исключить таким образом все НКСМВ, но результаты моделирования показывают, что даже при случайном выборе $V^{(A)}$ полученные коды обеспечивают достаточно хорошую производительность по сравнению с полярными кодами с CRC и LDPC-кодами. Формируемые таким образом ограничения динамического замораживания называются ограничениями типа A.

Можно дополнительно усовершенствовать конструкцию полярных подкодов путем уменьшения вероятности того, что списочный декодер ПИ исключит правильный путь на ранних фазах.

Вероятность $\mathcal{E}(L)$ события, соответствующего тому, что вес (19) правильного пути станет больше, чем вес $L$ неправильных путей на некоторой промежуточной фазе списочного алгоритма декодирования, зависит от того, насколько быстро веса неправильных путей расходятся с весом правильного пути. Поэтому можно вве-

сти дополнительный набор динамически замороженных символов, отображаемых на относительно надежные битовые подканалы, так, чтобы для неправильного пути оценки этих символов, получаемые из ЛОПП $S_m^{(i)}$, с высокой вероятностью отклонялись от значений, использованных кодером, что приводило бы к уменьшению веса соответствующего пути.

Пусть $C_i$ – пропускная способность (или другая мера надежности) битового подканала $W_m^{(i)}$, и рассмотрим последовательность $r_i : C_{r_0} \leqslant C_{r_1} \leqslant \ldots \leqslant C_{r_{n-1}}$. В [49] было предложено наложить нетривиальные ограничения динамического замораживания на $f_B$ символов $u_{r_i}$, передаваемых по подканалам с индексами $r_{n-k-f_A-1}, \ldots, r_{n-k-f_A-f_B}$, т.е. по наиболее надежным битовым подканалам, используемых для передачи статически замороженных символов в классическом полярном коде. Здесь $f_B$ – параметр конструкции кода. Такие ограничения динамического замораживания называются ограничениями типа B.

Вместо случайного выбора коэффициентов $V_{i,s}$ в (26) можно пытаться явно найти такие значения, которые минимизируют число НКСМВ в полученном коде [48, 50].

Как правило, полярные (под)коды с большими ядрами требуют меньшего размера списка для достижения той же корректирующей способности, что и коды, основанные на ядре Арикана. Если использовать эффективные алгоритмы обработки ядра вместе с тщательно подобранными поляризующими ядрами, то количество арифметических операций, выполняемых списочным алгоритмом ПИ для случая кодов на основе больших ядер, может быть меньше по сравнению с кодами на основе ядра Арикана при одинаковой корректирующей способности [35, 51].

Другой способ улучшить корректирующую способность полярных кодов с большими ядрами при декодировании списочным алгоритмом ПИ заключается в использовании модифицированной процедуры выбора номеров замороженных символов, как описано в [7].

## § 6. Заключение

Известно, что полярные коды с большими ядрами обеспечивают асимптотически оптимальную экспоненту масштабирования и скорость поляризации. В данной статье был представлен обзор методов построения и декодирования таких кодов. Многие из методов построения кодов, разработанных для случая кодов с ядром Арикана, относительно просто обобщаются на случай кодов с большими ядрами.

Сложность непосредственной реализации декодера таких кодов крайне высока. Однако для тщательно построенных поляризующих ядер могут быть построены алгоритмы обработки с низкой сложностью. Это позволяет полярным (под)кодам, основанным на больших ядрах, достичь такой же корректирующей способности при декодировании списочным алгоритмом последовательного исключения, как и кодам, основанным на ядре Арикана, с меньшей сложностью декодирования. В зависимости от структуры ядра, наименьшую сложность обработки могут обеспечивать оконный алгоритм или алгоритм на основе рекурсивных решеток. Это позволяет предположить, что более эффективные методы обработки ядра могут быть получены путем комбинирования этих подходов.

## СПИСОК ЛИТЕРАТУРЫ


1. *Arıkan E.* Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels // IEEE Trans. Inform. Theory. 2009. V. 55. № 7. P. 3051–3073. https://doi.org/10.1109/TIT.2009.2021379
2. *Korada S.B., Şaşoğlu E., Urbanke R.* Polar Codes: Characterization of Exponent, Bounds, and Constructions // IEEE Trans. Inform. Theory. 2010. V. 56. № 12. P. 6253–6264. https://doi.org/10.1109/TIT.2010.2080990

3. *Fazeli A., Hassani H., Mondelli M., Vardy A.* Binary Linear Codes with Optimal Scaling: Polar Codes with Large Kernels // IEEE Trans. Inform. Theory. 2021. V. 67. № 9. P. 5693–5710. https://doi.org/10.1109/TIT.2020.3038806
4. *Wang H.-P., Duursma I.M.* Polar Codes' Simplicity, Random Codes' Durability // IEEE Trans. Inform. Theory. 2021. V. 67. № 3. P. 1478–1508. https://doi.org/10.1109/TIT.2020.3041570
5. *Guruswami V., Riazanov A., Ye M.* Arıkan Meets Shannon: Polar Codes with Near-Optimal Convergence to Channel Capacity // IEEE Trans. Inform. Theory. 2022. V. 68. № 5. P. 2877–2919. https://doi.org/10.1109/TIT.2022.3146786
6. *Presman N., Shapira O., Litsyn S.* Mixed-Kernels Constructions of Polar Codes // IEEE J. Select. Areas Commun. 2016. V. 34. № 2. P. 239–253. https://doi.org/10.1109/JSAC.2015.2504278
7. *Bioglio V., Gabry F., Land I., Belfiore J.-C.* Multi-Kernel Polar Codes: Concept and Design Principles // IEEE Trans. Commun. 2020. V. 68. № 9. P. 5350–5362. https://doi.org/10.1109/TCOMM.2020.3006212
8. *Trifonov P.* Binary Successive Cancellation Decoding of Polar Codes with Reed–Solomon Kernel // Proc. 2014 IEEE Int. Symp. on Information Theory (ISIT'2014). Honolulu, HI, USA. June 29 – July 4, 2014. P. 2972–2976. https://doi.org/10.1109/ISIT.2014.6875379
9. *Bioglio V., Land I.* On the Marginalization of Polarizing Kernels // Proc. 2018 IEEE 10th Int. Symp. on Turbo Codes & Iterative Information Processing (ISTC'2018). Hong Kong, China. Dec. 3–7, 2018. P. 1–5. https://doi.org/10.1109/ISTC.2018.8625378
10. *Tal I., Vardy A.* List Decoding of Polar Codes // IEEE Trans. Inform. Theory. 2015. V. 61. № 5. P. 2213–2226. https://doi.org/10.1109/TIT.2015.2410251
11. *Miloslavskaya V., Trifonov P.* Sequential Decoding of Polar Codes // IEEE Commun. Lett. 2014. V. 18. № 7. P. 1127–1130. https://doi.org/10.1109/LCOMM.2014.2323237
12. *Chandesris L., Savin V., Declercq D.* Dynamic-SCFlip Decoding of Polar Codes // IEEE Trans. Commun. 2018. V. 66. № 6. P. 2333–2345. https://doi.org/10.1109/TCOMM.2018.2793887
13. *Trifonov P.* A Score Function for Sequential Decoding of Polar Codes // Proc. 2018 IEEE Int. Symp. on Information Theory (ISIT'2018). Vail, CO, USA. June 17–22, 2018. P. 1470–1474. https://doi.org/10.1109/ISIT.2018.8437559
14. *Polyanskiy Y., Poor H.V., Verdú S.* Channel Coding Rate in the Finite Blocklength Regime // IEEE Trans. Inform. Theory. 2010. V. 56. № 5. P. 2307–2359. https://doi.org/10.1109/TIT.2010.2043769
15. *Mondelli M., Hassani S.H., Urbanke R.L.* Unified Scaling of Polar Codes: Error Exponent, Scaling Exponent, Moderate Deviations, and Error Floors // IEEE Trans. Inform. Theory. 2016. V. 62. № 12. P. 6698–6712. https://doi.org/10.1109/TIT.2016.2616117
16. *Fazeli A., Vardy A.* On the Scaling Exponent of Binary Polarization Kernels // Proc. 52nd Annu. Allerton Conf. on Communication, Control, and Computing (Allerton'2014). Monticello, IL, USA. Sept. 30 – Oct. 3, 2014. P. 797–804. https://doi.org/10.1109/ALLERTON.2014.7028536
17. *Yao H., Fazeli A., Vardy A.* Explicit Polar Codes with Small Scaling Exponent // Proc. 2019 IEEE Int. Symp. on Information Theory (ISIT'2019). Paris, France. July 7–12, 2019. P. 1757–1761. https://doi.org/10.1109/ISIT.2019.8849741
18. *Mondelli M., Hassani S.H., Urbanke R.L.* How to Achieve the Capacity of Asymmetric Channels // IEEE Trans. Inform. Theory. 2018. V. 64. № 5. P. 3371–3393. https://doi.org/10.1109/TIT.2018.2789885
19. *Park W., Barg A.* Polar Codes for $q$-ary Channels, $q = 2^r$ // IEEE Trans. Inform. Theory. 2013. V. 59. № 2. P. 955–969. https://doi.org/10.1109/TIT.2012.2219035
20. *Şaşoğlu E., Telatar E., Arıkan E.* Polarization for Arbitrary Discrete Memoryless Channels // Proc. IEEE 2009 Information Theory Workshop (ITW'2009). Taormina, Italy. Oct. 11–16, 2009. P. 144–148. https://doi.org/10.1109/ITW.2009.5351487
21. *Mori R., Tanaka T.* Source and Channel Polarization over Finite Fields and Reed–Solomon Matrices // IEEE Trans. Inform. Theory. 2014. V. 60. № 5. P. 2720–2736. https://doi.org/10.1109/TIT.2014.2312181

22. *Presman N., Shapira O., Litsyn S., Etzion T., Vardy A.* Binary Polarization Kernels from Code Decompositions // IEEE Trans. Inform. Theory. 2015. V. 61. № 5. P. 2227–2239. https://doi.org/10.1109/TIT.2015.2409257
23. *Trofimiuk G., Trifonov P.* Efficient Decoding of Polar Codes with Some 16 × 16 Kernels // Proc. IEEE 2018 Information Theory Workshop (ITW'2018). Guangzhou, China. Nov. 25–29, 2018. P. 11–15. https://doi.org/10.1109/ITW.2018.8613307
24. *Trofimiuk G.* A Search Method for Large Polarization Kernels // Proc. 2021 IEEE Int. Symp. on Information Theory (ISIT'2021). Melbourne, Australia. July 12–20, 2021. P. 2084–2089. https://doi.org/10.1109/ISIT45174.2021.9517729
25. *Lin H.-P., Lin S., Abdel-Ghaffar K.A.S.* Linear and Nonlinear Binary Kernels of Polar Codes of Small Dimensions with Maximum Exponents // IEEE Trans. Inform. Theory. 2015. V. 61. № 10. P. 5253–5270. https://doi.org/10.1109/TIT.2015.2469298
26. *Moskovskaya E., Trifonov P.* Design of BCH Polarization Kernels with Reduced Processing Complexity // IEEE Commun. Lett. 2020. V. 24. № 7. P. 1383–1386. https://doi.org/10.1109/LCOMM.2020.2984382
27. *Abbasi F., Viterbo E.* Large Kernel Polar Codes with Efficient Window Decoding // IEEE Trans. Veh. Technol. 2020. V. 69. № 11. P. 14031–14036. https://doi.org/10.1109/TVT.2020.3029305
28. *Trofimiuk G.* Shortened Polarization Kernels // Proc. 2021 IEEE Globecom Workshops (GC Wkshps). Madrid, Spain. Dec. 7–11, 2021. P. 1–6. https://doi.org/10.1109/GCWkshps52748.2021.9681982
29. *Miloslavskaya V., Trifonov P.* Sequential Decoding of Polar Codes with Arbitrary Binary Kernel // Proc. IEEE 2014 Information Theory Workshop (ITW'2014). Hobart, TAS, Australia. Nov. 2–5, 2014. P. 376–380. https://doi.org/10.1109/ITW.2014.6970857
30. *Fujiwara T., Yamamoto H., Kasami T., Lin S.* A Trellis-Based Recursive Maximum-Likelihood Decoding Algorithm for Binary Linear Block Codes // IEEE Trans. Inform. Theory. 1998. V. 44. № 2. P. 714–729. https://doi.org/10.1109/18.661515
31. *Trifonov P.* Trellis-Based Decoding Techniques for Polar Codes with Large Kernels // Proc. IEEE 2019 Information Theory Workshop (ITW'2019). Visby, Sweden. Aug. 25–28, 2019. P. 249–253. https://doi.org/10.1109/ITW44776.2019.8989386
32. *Trifonov P., Karakchieva L.* Recursive Processing Algorithm for Low Complexity Decoding of Polar Codes with Large Kernels // IEEE Trans. Commun. Early access June 2023, https://doi.org/10.1109/TCOMM.2023.3285773
33. *Balatsoukas-Stimming A., Bastani Parizi M., Burg A.* LLR-Based Successive Cancellation List Decoding of Polar Codes // IEEE Trans. Signal Process. 2015. V. 63. № 19. P. 5165–5179. https://doi.org/10.1109/TSP.2015.2439211
34. *Trofimiuk G., Iakuba N., Rets S., Ivanov K., Trifonov P.* Fast Block Sequential Decoding of Polar Codes // IEEE Trans. Veh. Technol. 2020. V. 69. № 10. P. 10988–10999. https://doi.org/10.1109/TVT.2020.3006369
35. *Trofimiuk G., Trifonov P.* Window Processing of Binary Polarization Kernels // IEEE Trans. Commun. 2021. V. 69. № 7. P. 4294–4305. https://doi.org/10.1109/TCOMM.2021.3072730
36. *Trofimiuk G., Trifonov P.* Construction of Binary Polarization Kernels for Low Complexity Window Processing // Proc. IEEE 2019 Information Theory Workshop (ITW'2019). Visby, Sweden. Aug. 25–28, 2019. P. 115–119. https://doi.org/10.1109/ITW44776.2019.8989344
37. *Valembois A., Fossorier M.* Box and Match Techniques Applied to Soft-Decision Decoding // IEEE Trans. Inform. Theory. 2004. V. 50. № 5. P. 796–810. https://doi.org/10.1109/TIT.2004.826644
38. *Gupta B., Yao H., Fazeli A., Vardy A.* Polar List Decoding for Large Polarization Kernels // Proc. 2021 IEEE Globecom Workshops (GC Wkshps). Madrid, Spain. Dec. 7–11, 2021. P. 1–6. https://doi.org/10.1109/GCWkshps52748.2021.9681935
39. *Trifonov P.* Efficient Design and Decoding of Polar Codes // IEEE Trans. Commun. 2012. V. 60. № 11. P. 3221–3227. https://doi.org/10.1109/TCOMM.2012.081512.110872
40. *Tal I., Vardy A.* How to Construct Polar Codes // IEEE Trans. Inform. Theory. 2013. V. 59. № 10. P. 6562–6582. https://doi.org/10.1109/TIT.2013.2272694

41. *Mori R., Tanaka T.* Performance of Polar Codes with the Construction Using Density Evolution // IEEE Commun. Lett. 2009. V. 13. № 7. P. 519–521. https://doi.org/10.1109/LCOMM.2009.090428
42. *Kern D., Vorköper S., Kühn V.* A New Code Construction for Polar Codes Using Min-Sum Density // Proc. 2014 8th Int. Symp. on Turbo Codes and Iterative Information Processing (ISTC'2014). Bremen, Germany. Aug. 18–22, 2014. P. 228–232. https://doi.org/10.1109/ISTC.2014.6955119
43. *Miloslavskaya V., Trifonov P.* Design of Binary Polar Codes with Arbitrary Kernels // Proc. 2012 IEEE Information Theory Workshop (ITW'2012). Lausanne, Switzerland. Sept. 3–7, 2012. P. 119–123. https://doi.org/10.1109/ITW.2012.6404639
44. *Richardson T., Urbanke R.* Modern Coding Theory. Cambridge, UK: Cambridge Univ. Press, 2008.
45. *Trifonov P.* On Construction of Polar Subcodes with Large Kernels // Proc. 2019 IEEE Int. Symp. on Information Theory (ISIT'2019). Paris, France. July 7–12, 2019. P. 1932–1936. https://doi.org/10.1109/ISIT.2019.8849672
46. *Karakchieva L., Trifonov P.* An Approximate Method for Construction of Polar Codes with Kernels over $\mathbb{F}_2^t$ // IEEE Commun. Lett. 2020. V. 24. № 9. P. 1857–1860. https://doi.org/10.1109/LCOMM.2020.2995257
47. *Trifonov P., Miloslavskaya V.* Polar Subcodes // IEEE J. Select. Areas Commun. 2016. V. 34. № 2. P. 254–266. https://doi.org/10.1109/JSAC.2015.2504269
48. *Trifonov P.* Randomized Polar Subcodes with Optimized Error Coefficient // IEEE Trans. Commun. 2020. V. 68. № 11. P. 6714–6722. https://doi.org/10.1109/TCOMM.2020.3018781
49. *Trifonov P., Trofimiuk G.* A Randomized Construction of Polar Subcodes // Proc. 2017 IEEE Int. Symp. on Information Theory (ISIT'2017). Aachen, Germany. June 25–30, 2017. P. 1863–1867. https://doi.org/10.1109/ISIT.2017.8006852
50. *Miloslavskaya V., Vucetic B., Li Y., Park G., Park O.-S.* Recursive Design of Precoded Polar Codes For SCL Decoding // IEEE Trans. Commun. 2021. V. 69. № 12. P. 7945–7959. https://doi.org/10.1109/TCOMM.2021.3111625
51. *Trifonov P.* Recursive Trellis Processing of Large Polarization Kernels // Proc. 2021 IEEE Int. Symp. on Information Theory (ISIT'2021). Melbourne, Australia. July 12–20, 2021. P. 2090–2095. https://doi.org/10.1109/ISIT45174.2021.9517783

*Трифонов Петр Владимирович*
Университет ИТМО, Санкт-Петербург
pvtrifonov@itmo.ru

Поступила в редакцию
29.12.2022
После доработки
22.02.2023
Принята к публикации
22.02.2023

**ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ**

**Том 59** **2023** **Вып. 1**

УДК 621.391 : 519.23

© 2023 г. **Г.К. Голубев**

# ПЕРЕПАРАМЕТРИЗОВАННЫЕ ТЕСТЫ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ ДЛЯ ОБНАРУЖЕНИЯ РАЗРЕЖЕННЫХ ВЕКТОРОВ

Рассматривается задача обнаружения разреженного вектора большой размерности на фоне белого гауссовского шума. Предполагается, что неизвестный вектор может иметь только $p$ ненулевых компонент, положение и величина которых неизвестны, а их число, с одной стороны, велико, но с другой – мало по сравнению с его размерностью. Тест максимального правдоподобия (МП) в этой задаче имеет простой вид и, естественно, зависит от $p$. В статье изучаются статистические свойства перепараметризованных тестов МП, т.е. тестов, построенных на основе предположения, что число ненулевых компонент вектора равно $q$ $(q > p)$, в ситуации, когда на самом деле вектор имеет всего лишь $p$ ненулевых компонент. Показывается, что в некоторых случаях перепараметризованные тесты могут быть лучше стандартных тестов МП.

*Ключевые слова*: разреженный вектор, белый гауссовский шум, тест максимального правдоподобия.

**DOI:** 10.31857/S0555292323010047, **EDN:** RMHNYH

## § 1. Введение

Реальный интерес к перепараметризованным статистическим моделям и методам возник сравнительно недавно и связан с задачами глубокого обучения. Оказалось, что во многих случаях обучение перепараметризованной нейронной сети с помощью методов типа градиентного спуска дает хорошие результаты по сравнению с обучением неперепараметризованной сети [1]. Полностью математически удовлетворительного объяснения этого факта, по-видимому, до сих пор не найдено, что связано с очень высокой нелинейностью рассматриваемой задачи [2]. Поэтому внимание привлекли простые статистические модели, в которых наблюдаются похожие эффекты. Одной из таких моделей является модель линейной регрессии со случайными гауссовскими регрессорами, анализ которой в перепараметризованном режиме можно найти, например, в [3]. В обзорной статье [4] приводятся примеры других близких моделей, связанных с обработкой сигналов.

В этой статье будет показано, что в задаче обнаружения разреженных векторов с большим числом ненулевых компонент с помощью метода МП наблюдаются в некотором смысле похожие эффекты, а именно оказывается, что перепараметризованные тесты МП в некоторых случаях могут быть лучше стандартных.

Далее будет рассматриваться задача обнаружения разреженного вектора $\theta^n = (\theta_1, \dots \theta_n)^\top \in \mathbb{R}^n$ по наблюдениям

$$Y_k = \theta_k + \xi_k, \quad k = 1, \dots, n, \tag{1}$$

где $\xi_k$ – независимые стандартные гауссовские случайные величины.

Цель состоит в том, чтобы на основе наблюдений $Y^n = (Y_1, \ldots, Y_n)^\top$ из (1) решить, какая из двух гипотез

$$\mathcal{H}_0 : \theta^n = 0 \quad \text{или} \quad \mathcal{H}_1 : \theta^n \neq 0$$

является справедливой.

Ключевым предположением в этой статье является гипотеза о том, что $\theta^n$ – *разреженный* вектор. Существует очень много математических определений разреженности. Для наглядности в этой статье мы ограничимся самым простым. Пусть $I_p^n = \{i_1, \ldots, i_p\}$ – мультииндекс (носитель $\theta^n$), т.е. множество, состоящее из $p$ различных положительных целых чисел, выбранных из множества $\{1, 2, \ldots, n\}$. Тогда $\theta^n$ называется разреженным, если существует носитель $I_p^n$, такой что

$$\theta_k = 0, \quad k \notin I_p^n.$$

Множество всех разреженных векторов в $\mathbb{R}^n$ будем далее обозначать через $\Theta_p^n$. Заметим, что его можно определить эквивалентным образом как

$$\Theta_p^n = \big\{\theta^n \in \mathbb{R}^n : \theta_{(k)}^2 = 0,\ k > p\big\};$$

здесь и далее $\{x_{(1)}, \ldots, x_{(n)}\}$ означает невозрастающую перестановку элементов множества $\{x_1, \ldots, x_n\}$.

Понятие разреженности становится статистически интересным и значимым, когда носитель $I_p^n$ неизвестен, размерность $n$ велика и при этом $p \ll n$. Поэтому в этой статье задача проверки гипотез будет рассматриваться в асимптотической постановке, предполагая, что $n \to \infty$. При этом $p = p(n)$ может быть любой известной функцией $n$, такой что $p(n) \in [p_\circ(n), p^\circ(n)]$, где $p_\circ(n)$ и $p^\circ(n)$ таковы, что

$$\lim_{n\to\infty} p_\circ(n) = \infty \quad \text{и} \quad \lim_{n\to\infty} \frac{p^\circ(n)\log^{1+\varepsilon}(n)}{n} = 0; \tag{2}$$

здесь и далее $\varepsilon$ обозначает любое строго положительное число. Не оговаривая этого особо, будем предполагать, что условия (2) выполнены.

Отметим, что задача проверки гипотез о разреженных векторах в ситуации, когда разреженность фиксирована, например, $p = 1$, а $n \to \infty$, существенно отличается от рассматриваемой далее. Хорошо известно, что при фиксированной разреженности предельные распределения статистик тестов МП и байесовских тестов при $\mathcal{H}_0$ не являются гауссовскими. Для байесовских тестов эти распределения тесно связаны с устойчивыми распределениями. Этот замечательный факт, по-видимому, впервые был установлен в [5] (см. также [6, 7] для задач обнаружения в белом гауссовском шуме). Также хорошо известно, что если $p \to \infty$, то предельные распределения статистик байесовских тестов и тестов МП будут гауссовскими при нулевой гипотезе (см., например, [7]). Но при этом, в отличие от классических (т.е. имеющих низкую размерность) задач проверки гипотез, гауссовость не влечет асимптотической эквивалентности байесовских тестов и тестов МП.

Тест МП для проверки простой гипотезы $\mathcal{H}_0$ против альтернативы $\mathcal{H}_1$ при условии, что $\theta^n \in \Theta_p^n$, легко построить. Он, естественно, основан на максимуме отношения правдоподобия

$$L(Y^n; \Theta_p^n) = \max_{\theta^n \in \Theta_p^n} \prod_{k=1}^n \exp\left\{-\frac{(Y_k - \theta_k)^2}{2} + \frac{Y_k^2}{2}\right\} =$$

$$= \max_{I_p^n} \exp\left\{\frac{1}{2}\sum_{k\in I_p^n} Y_k^2\right\} = \exp\left\{\frac{1}{2}\sum_{k=1}^p Y_{(k)}^2\right\}.$$

Поэтому в качестве статистики этого теста будем использовать

$$M_p(Y^n) \stackrel{\text{def}}{=} \frac{1}{2}\sum_{k=1}^{p} Y_{(k)}^2.$$

Таким образом, тест МП отвергает гипотезу $\mathcal{H}_0$, если

$$M_p(Y^n) \geqslant t_{M_p}(\alpha),$$

где критический уровень $t_{M_p}(\alpha)$ выбирается так, чтобы вероятность ошибки первого рода была равна $\alpha$, т.е. он определяется как корень уравнения

$$\mathbf{P}_{\mathcal{H}_0}\big\{M_p(Y^n) \leqslant t_{M_p}(\alpha)\big\} = \mathbf{P}\big\{M_p(\xi^n) \leqslant t_{M_p}(\alpha)\big\} = \alpha;$$

здесь и далее $\xi^n = (\xi_1, \ldots, \xi_n)^\top$ обозначает вектор из независимых стандартных гауссовских случайных величин.

Далее, не оговаривая этого особо, будем считать, что вероятность ошибки первого рода $\alpha$ остается фиксированной при $n \to \infty$.

Мощность (качество) теста МП, как, впрочем, и любого другого теста, принято измерять вероятностью ошибки второго рода

$$\beta_{M_p}(\theta^n) = \mathbf{P}_{\theta^n}\big\{M_p(Y^n) \leqslant t_{M_p}(\alpha)\big\},$$

где $\mathbf{P}_{\theta^n}$ – вероятностное распределение, порожденное наблюдениями $Y^n$ при альтернативе $\mathcal{H}_1$.

Естественно, нам хотелось бы найти тест, который минимизировал бы ошибку второго рода равномерно при всех $\theta^n$. Однако хорошо известно, что за исключением случая двух простых гипотез эта задача, как правило, не имеет простого решения. Оптимальный тест можно найти, например, для средней (байесовский подход) или максимальной (минимаксный подход) вероятности ошибки второго рода. При этих подходах, очевидно, нужна априорная информация о $\theta^n$. При байесовском подходе она определяется априорным вероятностным распределением на $\theta^n$, а при минимаксном – предположением, что $\theta^n \in \Theta$, где $\Theta$ – известное подмножество в $\mathbb{R}^n$. Очевидно, что оптимальные тесты в этом случае будут зависеть от априорной информации.

Если $\theta^n$ имеет небольшую размерность, то априорная информация не играет принципиальной роли, поскольку в этом случае практически всю информацию о неизвестном векторе можно извлечь из наблюдений. При больших размерностях $\theta^n$ из наблюдений мы получаем недостаточно информации и должны компенсировать ее недостаток априорной информацией, и поэтому тесты будут существенно от нее зависеть.

Чтобы устранить этот принципиальный недостаток байесовского и минимаксного подходов, в статистике используется множественная проверка гипотез. Грубо говоря, этот метод подразумевает, что тест строится не с помощью одной статистики, а с помощью некоторого семейства статистик. При этом предполагается, что это семейство не является очень богатым. В рассматриваемой здесь задаче роль такого семейства может играть, например, $\{M_p(Y^n),\ p = 1, \ldots, n\}$. Простейшим хорошо известным, но в то же время достаточно наивным методом множественной проверки гипотез является метод Бонферрони [8]. Обзор некоторых более современных эвристических подходов, включая популярный на практике метод False Discovery Rate [9] и его модификаций, содержится в [10]. Что касается математических результатов множественной проверки гипотез о разреженных векторах, то некоторые из них можно найти, например, в [11].

Отметим, что для обнаружения разреженных векторов из $\Theta_p^n$ наряду со стандартным тестом МП можно использовать и любой *перепараметризованный* тест со

статистикой $M_q(Y^n)$ при $q > p$. На первый взгляд кажется, что эта идея противоречит здравому смыслу, но мы увидим, что перепараметризованные тесты могут иметь меньшую вероятность ошибки второго рода, чем классический тест МП. В этом можно убедиться, анализируя вероятность ошибки второго рода при всех $\theta^n \in \Theta_p^n$. Этот анализ, по сути дела, и представляет основное содержание настоящей статьи.

## § 2. Предельное распределение статистики теста МП при $\mathcal{H}_0$

Для вычисления критических значений тестов МП нам потребуется следующий результат о распределении суммы квадратов порядковых статистик независимых стандартных гауссовских случайных величин.

Пусть

$$G(x) = \sqrt{\frac{2}{\pi}} \int\limits_{\sqrt{2x}}^{\infty} e^{-u^2/2}\, du, \tag{3}$$

и обозначим через $G^{-1}(x)$ обратную функцию к $G(x)$.

Теорема 1. *При любых $p = p(n) \in [p_\circ(n), p^\circ(n)]$ (см. (2)) имеем*

$$\frac{M_p(\xi^n) - \mu(p;n)}{\sqrt{2p}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \quad n \to \infty, \tag{4}$$

*где*

$$\begin{aligned}
\mu(p;n) &= \sum_{k=1}^{p} G^{-1}\left(\frac{n+1}{k}\right) = \\
&= p\left\{\frac{n+1}{\sqrt{\pi}p} G^{-1}\left(\frac{p}{n+1}\right) \exp\left[-G^{-1}\left(\frac{p}{n+1}\right)\right] + \frac{1}{2}\right\} + O(1) = \\
&= p\left\{\log\left(\frac{n+1}{\sqrt{\pi}p}\right) - \frac{1}{2}\log\left[1 + \log\left(\frac{n+1}{\sqrt{\pi}p}\right)\right] + 1 + o(1)\right\} + O(1).
\end{aligned} \tag{5}$$

Доказательство этой теоремы, как и других вспомогательных результатов, приведено в § 4.

Из (4) и (5) вытекает, что при больших $p$ критическое значение для теста МП можно вычислить как

$$t_{M_p}(\alpha) = p\left\{\frac{n+1}{\sqrt{\pi}p} G^{-1}\left(\frac{p}{n+1}\right) \exp\left[-G^{-1}\left(\frac{p}{n+1}\right)\right] + \frac{1}{2}\right\} + \sqrt{2p}\, t(\alpha), \tag{6}$$

где $t(\alpha)$ – $\alpha$-квантиль стандартного гауссовского распределения, т.е. корень уравнения

$$\frac{1}{\sqrt{2\pi}} \int\limits_{t(\alpha)}^{\infty} e^{-x^2/2}\, dx = \alpha.$$

Что касается вычисления вероятности ошибки второго рода, то теорема 1 может быть в принципе полезна только в случае, когда эта вероятность не является малой.

Для анализа малых вероятностей ошибки будет использоваться следующее неасимптотическое неравенство. В его формулировке и далее неравенство $\xi \overset{\mathcal{D}}{\geqslant} \eta$ озна-

чает, что

$$\mathbf{P}\{\xi \geqslant x\} \leqslant \mathbf{P}\{\eta \geqslant x\} \quad \text{при всех } x \geqslant 0.$$

Лемма 1. *Справедливо неравенство*

$$M_p(\xi^n) \overset{\mathcal{D}}{\geqslant} \mu\big(p; \Sigma_{n+1}\big) + \left\{1 - \frac{1}{2}\left[1 + \log\left(\frac{\Sigma_{n+1}}{\sqrt{\pi}p^{\circ}(n)}\right)\right]^{-1}\right\} \sum_{s=1}^{p} w(s,p)(1-\chi_s),$$

*где*:

- *функция* $\mu(\cdot\,,\cdot)$ *определена в* (5);
- $\chi_s$ – *независимые стандартные экспоненциально распределенные случайные величины*;
- $\Sigma_{n+1} = \sum\limits_{s=1}^{n+1} \chi_s$;
- $$w(s,p) = \sum_{k=s}^{p} \frac{1}{k}. \tag{7}$$

## § 3. Вероятность ошибки второго рода

К сожалению, простых и одновременно точных оценок для вероятности ошибки второго рода, по-видимому, не существует. Поэтому мы приведем лишь границы сверху для этой величины, предполагая, что $\theta^n \in \Theta_p^n$, но при этом используется тест МП со статистикой $M_q(Y^n)$, $q \geqslant p$.

Чтобы несколько упростить дальнейшее изложение, будем считать, что

$$q \leqslant Kp,$$

где $K \geqslant 1$ – некоторая постоянная.

В силу инвариантности статистики $M_q(Y^n)$ относительно перестановок компонент вектора наблюдений $Y^n$, без ограничения общности можно считать, что

$$\begin{aligned} Y_k &= \theta_k + \xi_k, \qquad k = 1, \ldots, p, \\ Y_k &= \xi_k, \qquad\quad\ k = p+1, \ldots, n, \end{aligned}$$

а компоненты вектора $\theta^n$ таковы, что $\theta_1^2 \geqslant \theta_2^2 \geqslant \ldots \geqslant \theta_p^2$.

Пусть $p' \leqslant p$ – некоторое целое число. Обозначим для краткости

$$Y'_k = Y_{p'+k}, \quad \xi'_k = \xi_{p'+k}, \quad k = 1, \ldots, n-p',$$

и

$$\|\theta^n\|_{p'}^2 = \sum_{k=1}^{p'} \theta_{(k)}^2.$$

Тогда очевидно, что

$$M_q(Y^n) \geqslant \frac{1}{2}\sum_{k=1}^{p'} Y_k^2 + \frac{1}{2}\sum_{k=1}^{q-p'} Y_{(k)}'^2. \tag{8}$$

Первое слагаемое в правой части этого неравенства имеет вид

$$\frac{1}{2}\sum_{k=1}^{p'} Y_k^2 = \frac{\|\theta^n\|_{p'}^2 + p'}{2} + \sum_{k=1}^{p'} \xi_i\theta_i + \frac{1}{2}\sum_{k=1}^{p'}(\xi_k^2 - 1). \tag{9}$$

Для оценки второго слагаемого в правой части (8) можно воспользоваться леммой Андерсона [12] (см. также [13, гл. II, § 10]), а именно интуитивно понятным неравенством

$$\frac{1}{2}\sum_{k=1}^{q-p'} Y_{(k)}'^2 \overset{\mathcal{D}}{\geqslant} \frac{1}{2}\sum_{k=1}^{q-p'} \xi_{(k)}'^2.$$

Продолжив правую часть этого неравенства с помощью леммы 1, получаем

$$\frac{1}{2}\sum_{k=1}^{q-p'} Y_{(k)}'^2 \overset{\mathcal{D}}{\geqslant} \mu\big(q-p';n-p'\big) + (1+o(1))\sum_{k=1}^{q-p'} w(s,q-p')(1-\chi_s),$$

где веса $w(\cdot\,,\cdot)$ определены в (7). Поэтому отсюда и из (8), (9) приходим к следующему неравенству:

$$M_q(Y^n) \overset{\mathcal{D}}{\geqslant} \frac{\|\theta^n\|_{p'}^2 + p'}{2} + \mu(q-p';n-p') - \zeta(\theta^n,p',q),$$

где

$$\zeta(\theta^n,p',q) = (1+o(1))\sum_{k=1}^{q-p'} w(s,q-p')(\chi_s - 1) + \sum_{k=1}^{p'} \xi_i\theta_i + \frac{1}{2}\sum_{k=1}^{p'}(1-\xi_k^2).$$

Тогда отсюда и из (6) вытекает следующая граница сверху для вероятности ошибки второго рода:

$$\beta_{M_q}(\theta^n) = \mathbf{P}_{\theta^n}\Big\{M_q(Y^n) \leqslant \mu(q;n) + \sqrt{2q}t(\alpha)\Big\} \leqslant$$
$$\leqslant \mathbf{P}\Big\{\zeta(\theta^n,p',q) \geqslant \frac{\|\theta^n\|_{p'}^2}{2} - \Delta(q,p';n) + \frac{p'}{2} - \sqrt{2q}t(\alpha)\Big\}, \tag{10}$$

где

$$\Delta(p',q;n) = \mu(q;n) - \mu(q-p';n-p').$$

Величина $\Delta(p',q;n)$ играет принципиальную роль в рассматриваемой задаче, и далее потребуется следующая ее аппроксимация, вытекающая из (5) и формулы Тейлора:

$$\Delta(p',q;n) = q\left\{\log\left(\frac{n}{\sqrt{\pi}q}\right) - \frac{1}{2}\log\left[1+\log\left(\frac{n}{\sqrt{\pi}q}\right)\right] + 1 + o(1)\right\} -$$
$$-\,(q-p')\left\{\log\left(\frac{n-p'}{\sqrt{\pi}(q-p')}\right) - \frac{1}{2}\log\left[1+\log\left(\frac{n-p'}{\sqrt{\pi}(q-p')}\right)\right] + 1\right\} =$$

$$= p'\left\{\log\left(\frac{n}{\sqrt{\pi}p'}\right) - \frac{1}{2}\log\left[1+\log\left(\frac{n}{\sqrt{\pi}p'}\right)\right] + 1 + o(1)\right\} =$$

$$= \mu(p';n) - qh\left(\frac{p'}{q}\right) + o(q), \quad n\to\infty, \tag{11}$$

где

$$h(x) = -x\log(x) - (1-x)h(1-x).$$

Также далее будет нужна аппроксимация распределения случайной величины $\zeta(\theta^n, p', q)$. Обозначим для краткости

$$D(x) = D(x; p', q, \theta^n) = \left[\frac{\|\theta^n\|_{p'}^2}{1+x} + \frac{p'}{2} + \sum_{s=1}^{q-p'} w^2(s,p)\right]. \tag{12}$$

Лемма 2. *Для всех $x>0$, таких что*

$$x \leqslant \frac{(1-\varepsilon)D(0)}{2w^2(1,q-p')}, \tag{13}$$

*справедливо неравенство*

$$\mathbf{P}\left\{\zeta(\theta^n,p',q) \geqslant \sqrt{2xD(0)}\right\} \leqslant$$
$$\leqslant \exp\left[-r_\circ(x) + O\left(\frac{x(q-p')}{D(0)w(1,q-p')}\right)\right], \quad n\to\infty, \tag{14}$$

*где*

$$r_\circ(x) = x + \frac{\log[2\pi(1+x)]}{2}.$$

Следующая теорема, представляющая основной результат статьи, является по сути прямым следствием этой леммы и неравенства (10).

Теорема 2. *Предположим, что для $p'\in[p_\circ(n),p]$, $\theta^n\in\Theta_p^n$ и $A>0$ выполнены следующие условия*:

$$\|\theta^n\|_{p'}^2 \geqslant 2p'\log\left(\frac{n+1}{\sqrt{\pi}p'}\right) - p'\log\left[1+\log\left(\frac{n+1}{\sqrt{\pi}p'}\right)\right] + p' - 2qh\left(\frac{p'}{q}\right) +$$
$$+ 4\sqrt{A}\left[p'\log\left(\frac{n+1}{\sqrt{\pi}p'}\right) + A\right]^{1/2} + 4A + \varepsilon q \tag{15}$$

*и*

$$A \leqslant \frac{(1-\varepsilon)\|\theta^n\|_{p'}^2}{2w^2(1,q-p')}.$$

*Тогда при $n\to\infty$*

$$\beta_{M_q}(\theta^n) \leqslant \exp\left[-r_\circ(A) + O\left(\frac{q-p'}{w^3(1,q-p')}\right)\right]. \tag{16}$$

Доказательство. Из леммы 2 и неравенства (10) непосредственно вытекает, что если для некоторого $p' \in [p_\circ(n), p]$

$$D(0;p',q,\theta^n) - 2\sqrt{2AD(0;p',q,\theta^n)} \geqslant$$

$$\geqslant 2\Delta(q,p';n) - \frac{p'}{2} + \sum_{s=1}^{q-p'} w^2(s,p) + 2\sqrt{2qt}(\alpha), \tag{17}$$

то неравенство (16) выполнено.

Условие (15) является упрощенной формой неравенства (17). Чтобы в этом убедиться, заметим, что неравенство

$$y - 2\sqrt{2Ay} \geqslant x, \quad x > 0,$$

эквивалентно неравенству

$$y \geqslant \left[\sqrt{2A} + \sqrt{2A+x}\,\right]^2 = x + 4A + 2\sqrt{2A}\sqrt{2A+x}.$$

Подставив в него

$$x = 2\Delta(q,p';n) - \frac{p'}{2} + \sum_{s=1}^{q-p'} w^2(s,p) + 2\sqrt{2qt}(\alpha),$$

$$y = D(0;p',q,\theta^n) = \|\theta^n\|_{p'}^2 + \frac{p'}{2} + \sum_{s=1}^{q-p'} w^2(s,p)$$

и воспользовавшись (5) и (11), приходим к (15). ▲

*Замечание* 1. Теорема 2 допускает достаточно простую интерпретацию. Она позволяет оценить разность между минимальными энергиями векторов $\theta^n$, обнаруживаемых стандартным и перепараметризованным тестом МП.

Предположим для простоты, что $A = A(n)$ – медленно растущая функция $n$, например, $A(n) = \log[p^\circ(n)]$. В этом случае, если выполнено условие

$$\lim_{n\to\infty} \frac{A(n)\log(n)}{p_\circ(n)} = 0,$$

то последние слагаемые в (15) имеют порядок $o(p)$. Поэтому стандартный тест МП сможет с достаточно малой вероятностью ошибки второго рода обнаружить векторы $\theta^n \in \Theta_p^n$ при отношениях сигнал/шум $\|\theta^n\|^2/p$, больших чем

$$2\log\left(\frac{n+1}{\sqrt{\pi}p}\right) - \log\left[1 + \log\left(\frac{n+1}{\sqrt{\pi}p}\right)\right] + 1 + \varepsilon,$$

а перепараметризованный тест сможет сделать это с теми же вероятностями ошибок, но при меньших отношениях сигнал/шум, а именно

$$2\log\left(\frac{n+1}{\sqrt{\pi}p}\right) - \log\left[1 + \log\left(\frac{n+1}{\sqrt{\pi}p}\right)\right] + 1 - \frac{q}{p}h\left(\frac{p}{q}\right) + \varepsilon.$$

*Замечание* 2. На первый взгляд кажется, что выбирая большое $q$, можно существенно уменьшить вероятность ошибки второго рода. К сожалению, это слишком оптимистичное предположение с практической точки зрения. Дело в том, что, во-первых, улучшение верхних границ не влечет автоматического улучшения действительной вероятности ошибки. Во-вторых, улучшаются, по сути, только члены третьего порядка в экспоненте вероятности ошибки. И наконец, улучшаются асимптотические границы в ситуации, когда скорость сходимости к ним очень медленная.

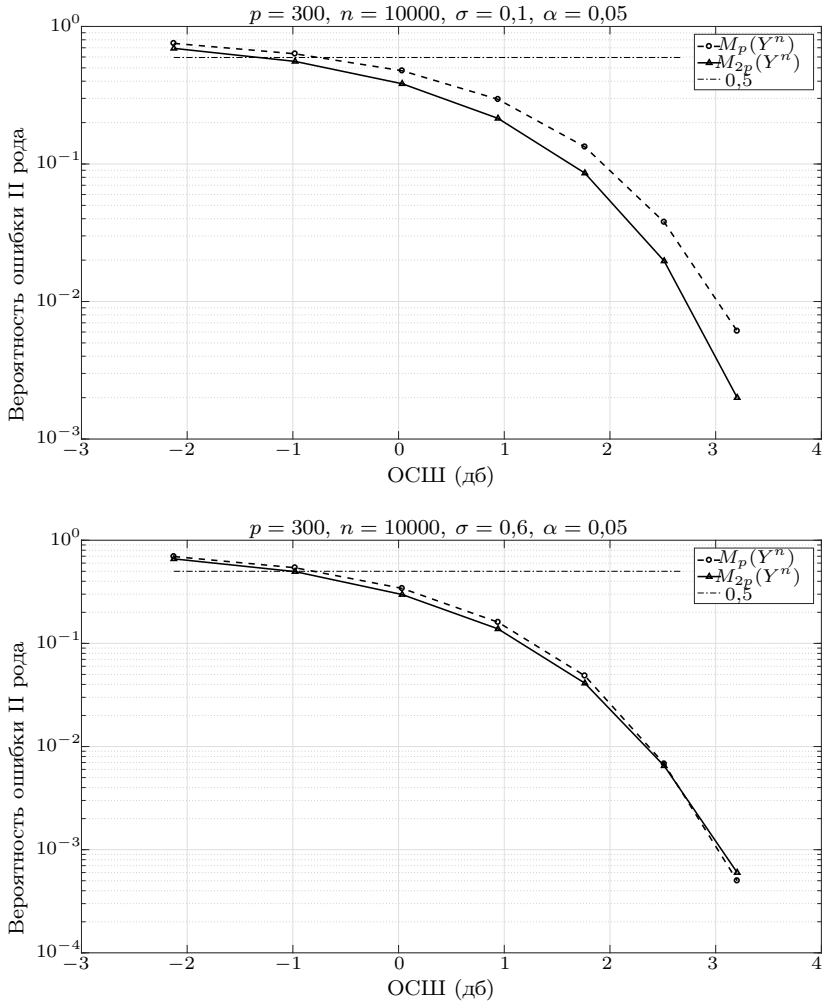


Рис. 1. Ошибки второго рода для стандартного и перепараметризованного теста МП. Графики сверху построены при $\sigma = 0{,}1$, а снизу – при $\sigma = 0{,}6$

К сожалению, медленные скорости сходимости типичны при проверке гипотез о разреженных векторах. На это обращал внимание еще Р.Л. Добрушин в [5].

*Замечание* 3. Наблюдать на практике небольшие преимущества перепараметризованных тестов возможно только лишь при достаточно больших $n$ и $p$. Это иллюстрирует рис. 1, на котором показана зависимость вероятности ошибки второго рода от отношения сигнал/шум при $n = 10000$ и $p = 300$ для стандартного теста МП и перепараметризованного теста с $q = 2p = 600$. При этом в качестве ненулевых $\theta_k$ использовались независимые одинаково распределенные случайные величины, имеющие следующую структуру:

$$\theta_k = A(b_k + \sigma\xi_k),$$

где $A$ и $\sigma$ – постоянные, $b_k$ – независимые случайные величины, принимающие значения $\{+1, -1\}$ с равными вероятностями, $\xi_k$ – независимые $\mathcal{N}(0, 1)$. В этом случае отношение сигнал/шум равно $A^2(1 + \sigma^2)$.

Мы видим, что на втором графике стандартный тест и перепараметризованный оказываются практически эквивалентными. Это объясняется тем, что для "случайных" $\theta_k^2$ неравенство (8) является слишком грубым.

## § 4. Доказательства

Для доказательства теоремы 1 нам потребуется следующий простой факт, который легко проверить с помощью экспоненциального неравенства Чебышева.

Лемма 3. *При* $x \geqslant 0$

$$\mathbf{P}\left\{\frac{1}{k}\sum_{s=1}^{k}\chi_s \geqslant (1+x)\right\} \leqslant \exp\{-k[x-\log(1+x)]\}, \quad x>0,$$

$$\mathbf{P}\left\{\frac{1}{k}\sum_{s=1}^{k}\chi_s \leqslant (1-x)\right\} \leqslant \exp\{k[x+\log(1-x)]\}, \quad x\in(0,1).$$

Доказательство теоремы 1 основано на простых и хорошо известных методах и результатах, таких как теорема Пайка [14].

Введем функцию

$$r(x) = -\log[\sqrt{\pi}G(x)].$$

Тогда очевидно, что

$$G^{-1}(x) = r^{-1}\big[-\log\big(\sqrt{\pi}x\big)\big], \tag{18}$$

где $r^{-1}(x)$ – функция, обратная к $r(x)$.

Интегрируя по частям правую часть в (3), нетрудно проверить, что

$$r(x) = x + \frac{1}{2}\log(1+x) + O\Big(\frac{1}{x^{3/2}}\Big), \tag{19}$$

и отсюда с помощью формулы Тейлора находим

$$r^{-1}(x) = x - \frac{\log(1+x)}{2} + \frac{\log(1+x)}{2(1+x)} + O\Big(\frac{1}{x^{3/2}}\Big). \tag{20}$$

Проверим сначала (5). С помощью (18) и (20) нетрудно убедиться, что

$$\mu(p;n) = \sum_{k=1}^{p} G^{-1}\left(\frac{k}{n+1}\right) = -\int_1^p G^{-1}\left(\frac{x}{n+1}\right)dx + O(1) =$$

$$= -(n+1)\int_{1/(n+1)}^{p/(n+1)} G^{-1}(x)\,dx + O(1) =$$

$$= -(n+1)\int_{G^{-1}[1/(n+1)]}^{G^{-1}[p/(n+1)]} xG'(x)\,dx + O(1) =$$

$$= -\frac{n+1}{\sqrt{\pi}}\int_{G^{-1}[1/(n+1)]}^{G^{-1}[p/(n+1)]} \sqrt{x}e^{-x}\,dx + O(1). \tag{21}$$

Далее, с помощью очевидной замены переменных и интегрирования по частям находим

$$\int\limits_y^\infty \sqrt{x}e^{-x}\,dx = \sqrt{y}e^{-y} + \frac{\sqrt{\pi}}{2}\sqrt{\frac{2}{\pi}}\int\limits_{\sqrt{2y}}^\infty e^{-x^2/2}\,dx = \sqrt{y}e^{-y} + \frac{\sqrt{\pi}}{2}G(y) =$$

$$= \sqrt{\pi}G(y)\biggl(\frac{\sqrt{y}e^{-y}}{\sqrt{\pi}G(y)} + \frac{1}{2}\biggr).$$

Подставив в это равенство (см. (19))

$$e^{-y} = \sqrt{\pi(1+y)}G(y)\biggl[1 + O\biggl(\frac{1}{y^{3/2}}\biggr)\biggr],$$

получаем

$$\frac{1}{\sqrt{\pi}}\int\limits_y^\infty \sqrt{x}e^{-x}\,dx = G(y)\biggl[\sqrt{y(y+1)} + \frac{1}{2} + O\biggl(\frac{1}{\sqrt{y}}\biggr)\biggr] =$$

$$= G(y)\biggl[y + 1 + O\biggl(\frac{1}{\sqrt{y}}\biggr)\biggr].$$

Используя это равенство, (18) и (20), продолжим (21) следующим образом:

$$\mu(p;n) = p\biggl\{\frac{n+1}{\sqrt{\pi}p}G^{-1}\biggl(\frac{p}{n+1}\biggr)\exp\biggl[-G^{-1}\biggl(\frac{p}{n+1}\biggr)\biggr] + \frac{1}{2}\biggr\} -$$

$$- \frac{1}{n+1}\biggl[G^{-1}\biggl(\frac{1}{n+1}\biggr) + 1 + O\biggl(\frac{1}{\sqrt{G^{-1}[1/(n+1)]}}\biggr)\biggr] + O(1) =$$

$$= p\biggl[G^{-1}\biggl(\frac{p}{n+1}\biggr) + 1 + O\biggl(\frac{1}{\sqrt{G^{-1}[p/(n+1)]}}\biggr)\biggr] + O(1) =$$

$$= p\biggl\{\log\biggl(\frac{n+1}{\sqrt{\pi}p}\biggr) - \frac{1}{2}\log\biggl[1 + \log\biggl(\frac{n+1}{\sqrt{\pi}p}\biggr)\biggr] + 1 + o(1)\biggr\} + O(1).$$

Пусть $U_1, \ldots, U_n$ – независимые случайные величины, равномерно распределенные на отрезке $[0,1]$. Тогда (см., например, [14])

$$U_{(k)} \stackrel{\mathcal{D}}{=} 1 - \sum_{s=1}^{k}\chi_s \Big/ \sum_{s=1}^{n+1}\chi_s. \tag{22}$$

Далее воспользуемся следующим очевидным тождеством:

$$\frac{\xi^2_{(k)}}{2} \stackrel{\mathcal{D}}{=} F^{-1}(U_{(k)}), \tag{23}$$

где $F^{-1}(\cdot)$ – функция, обратная к

$$F(x) = \mathbf{P}\Bigl\{\frac{\xi_i^2}{2} \leqslant x\Bigr\} = 1 - G(x).$$

Тогда из (22) и (23) получаем

$$\frac{\xi^2_{(k)}}{2} \stackrel{\mathcal{D}}{=} G^{-1}(1-U_{(k)}) \stackrel{\mathcal{D}}{=} G^{-1}\left[\frac{k}{n+1}\times\left(\frac{1}{k}\sum_{s=1}^{k}\chi_s\right)\Big/\left(\frac{1}{n+1}\sum_{s=1}^{n+1}\chi_s\right)\right]. \tag{24}$$

Для $x^{n+1}\in\mathbb{R}^{n+1}$ обозначим для краткости

$$\mu_k\big(x^{n+1}\big) = \frac{1}{k}\sum_{s=1}^{k} x_s.$$

Тогда из (24) получаем

$$\frac{\xi^2_{(k)}}{2} \stackrel{\mathcal{D}}{=} r^{-1}\left[\log\left(\frac{n+1}{\sqrt{\pi}k}\right) - \log\big[\mu_k\big(\chi^{n+1}\big)\big] + \log\big[\mu_{n+1}\big(\chi^{n+1}\big)\big]\right]. \tag{25}$$

Далее разложим правую часть соотношения (25) по формуле Тейлора в точке $\log[(n+1)/(\sqrt{\pi}k)]$. Это можно сделать, если, например, неравенства

$$\left|\log\big[\mu_k\big(\chi^{n+1}\big)\big]\right| + \left|\log\big[\mu_{n+1}\big(\chi^{n+1}\big)\big]\right| \leqslant \frac{1}{1+\varepsilon}\log\left(\frac{n+1}{\sqrt{\pi}p}\right), \quad k=1,\ldots,p,$$

выполняются с большими вероятностями.

Введем следующее подмножество в $\mathbb{R}^{n+1}$:

$$\Omega^n_{\varepsilon,p} = \bigcap_{k=1}^{n+1}\left\{x\in\mathbb{R}^{n+1} : \left|\log\big[\mu_k\big(x^{n+1}\big)\big]\right| + \left|\log\big[\mu_{n+1}\big(x^{n+1}\big)\big]\right| \leqslant \frac{1}{1+\varepsilon}\log\left(\frac{n+1}{\sqrt{\pi}p}\right)\right\}.$$

Из леммы 3 имеем

$$\mathbf{P}\left\{\left|\log\big[\mu_k\big(\chi^{n+1}\big)\big]\right| + \left|\log\big[\mu_{n+1}\big(\chi^{n+1}\big)\big]\right| \leqslant \frac{1}{1+\varepsilon}\log\left(\frac{n+1}{\sqrt{\pi}p}\right)\right\} \leqslant$$
$$\leqslant \exp\left[-Ck\left(\frac{n}{\sqrt{\pi}p}\right)^{1/(1+\varepsilon)}\right],$$

где здесь и далее $C>0$ – некоторая постоянная, и следовательно,

$$\mathbf{P}\big\{\chi^{n+1}\notin\Omega^n_{\varepsilon,p}\big\} \leqslant \exp\left[-C\left(\frac{n}{\sqrt{\pi}p}\right)^{1/(1+\varepsilon)}\right]. \tag{26}$$

Применяя формулу Тейлора, из (25) получаем, что если $\chi^n\in\Omega^n_{\varepsilon,p}$, то

$$\frac{\xi^2_{(k)}}{2} \stackrel{\mathcal{D}}{=} G^{-1}\left(\frac{n+1}{k}\right) - \big[1+o(1)\big]\log\big[\mu_k\big(\chi^{n+1}\big)\big] + \delta_k(\chi^n). \tag{27}$$

где

$$\delta_k(\chi^n) = [1+o(1)]\log\big[\mu_{n+1}\big(\chi^{n+1}\big)\big] + o(1)\log^2\big[\mu_k\big(\chi^{n+1}\big)\big] +$$
$$+ o(1)\log^2\big[\mu_{n+1}\big(\chi^{n+1}\big)\big].$$

Нетрудно проверить, что

$$\mathbf{E}\,\delta_k^2(\chi^n) \leqslant \frac{C}{n} + \frac{C}{k^2}. \tag{28}$$

Из (27) получаем

$$\sum_{k=1}^{p}\left[\frac{\xi_{(k)}^2}{2}-G^{-1}\left(\frac{n+1}{k}\right)\right]=\sum_{k=1}^{p}\left[\frac{\xi_{(k)}^2}{2}-G^{-1}\left(\frac{n+1}{k}\right)\right]\mathbf{1}\{\chi^{n+1}\in\Omega_{\varepsilon,p}^n\}+$$
$$+\sum_{k=1}^{p}\left[\frac{\xi_{(k)}^2}{2}-G^{-1}\left(\frac{n+1}{k}\right)\right]\mathbf{1}\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^n\}=$$
$$=-(1+o(1))\sum_{k=1}^{p}\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]+\sum_{k=1}^{p}\delta_k(\chi^{n+1})\mathbf{1}\{\chi^{n+1}\in\Omega_{\varepsilon,p}^n\}+$$
$$+(1+o(1))\sum_{k=1}^{p}\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]\mathbf{1}\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^n\}+$$
$$+\sum_{k=1}^{p}\left[\frac{\xi_{(k)}^2}{2}-G^{-1}\left(\frac{n+1}{k}\right)\right]\mathbf{1}\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^n\}. \tag{29}$$

С помощью неравенства Коши – Буняковского и (28) получаем, что

$$\frac{1}{\sqrt{p}}\,\mathbf{E}\sum_{k=1}^{p}\bigl|\delta_k(\chi^{n+1})\bigr|\mathbf{1}\{\chi^{n+1}\in\Omega_{\varepsilon,p}^n\}\leqslant C\left(\sqrt{\frac{p}{n}}+\frac{\log(p)}{\sqrt{p}}\right)\to 0,\quad n\to\infty.$$

Очевидно также, что

$$\frac{1}{\sqrt{p}}\sum_{k=1}^{3}\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]\mathbf{1}\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^n\}\xrightarrow{\mathcal{D}}0,\quad n\to\infty, \tag{30}$$

и в силу (26), леммы 3 и неравенства Коши – Буняковского

$$\frac{1}{\sqrt{p}}\,\mathbf{E}\sum_{k=4}^{p}\Bigl|\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]\Bigr|\mathbf{1}\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^n\}\leqslant$$
$$\leqslant C\sqrt{\mathbf{P}\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^n\}}\,\frac{1}{\sqrt{p}}\sum_{k=4}^{p}\frac{1}{\sqrt{k}}\leqslant C\sqrt{\mathbf{P}\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^n\}}\to 0,\quad n\to\infty. \tag{31}$$

Наконец, для последнего слагаемого в (29), принимая во внимание второе условие в (2), получаем

$$\mathbf{E}\sum_{k=1}^{p}\left[\frac{\xi_{(k)}^2}{2}-G^{-1}\left(\frac{n+1}{k}\right)\right]\mathbf{1}\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^n\}\leqslant$$
$$\leqslant\mathbf{E}\,\mathbf{1}\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^n\}\sum_{k=1}^{n}\xi_{(k)}^2+\mathbf{P}\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^n\}\sum_{k=1}^{n}G^{-1}\left(\frac{n+1}{k}\right)\leqslant$$
$$\leqslant n\sqrt{\mathbf{P}\{\chi^{n+1}\notin\Omega_{\varepsilon,p}^n\}}\leqslant n\exp\left[-C\left(\frac{n}{\sqrt{\pi}p}\right)^{1/(1+\varepsilon)}\right]\to 0,\quad n\to\infty. \tag{32}$$

Таким образом, в силу (29)–(32) осталось проверить, что

$$\frac{1}{\sqrt{2p}}\sum_{k=1}^{p}\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]\xrightarrow{\mathcal{D}}\mathcal{N}(0,1),\quad n\to\infty. \tag{33}$$

Аналогично (30)–(32) получаем, что

$$\frac{1}{\sqrt{2p}}\sum_{k=1}^{p^{1-\varepsilon}}\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]\xrightarrow{\mathcal{D}}0,\quad n\to\infty, \tag{34}$$

и

$$\frac{1}{\sqrt{2p}}\sum_{k=p^{1-\varepsilon}}^{p}\Bigl[\mathbf{E}\log^4\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]\Bigr]^{1/4}\leqslant C. \tag{35}$$

Чтобы доказать (33), воспользуемся формулой Тейлора, точнее, аппроксимацией

$$\log\bigl[\mu_k\bigl(\chi^{n+1}\bigr)\bigr]=\frac{1}{k}\sum_{s=1}^{k}(\chi_s-1)+O(1)\left[\frac{1}{k}\sum_{s=1}^{k}(\chi_s-1)\right]^2,\quad k>p^{1-\varepsilon}, \tag{36}$$

которая будет верна, если, например, неравенство

$$\left|\frac{1}{k}\sum_{s=1}^{k}(\chi_s-1)\right|\leqslant 0{,}5$$

выполняется при всех $k>p^{1-\varepsilon}$.

Из леммы 3 получаем, что

$$\mathbf{P}\left\{\left|\frac{1}{\sqrt{k}}\sum_{s=1}^{k}(\chi_s-1)\right|\geqslant 0{,}5\right\}\leqslant\exp(-Ck),$$

и поэтому

$$\mathbf{P}\left\{\max_{k>p^{1-\varepsilon}}\left|\frac{1}{\sqrt{k}}\sum_{s=1}^{k}(\chi_s-1)\right|\geqslant 0{,}5\right\}\leqslant\exp(-Cp^{1-\varepsilon}).$$

Отсюда и из (34)–(36) получаем

$$\frac{1}{\sqrt{2p}}\sum_{k=1}^{p}\log\left(\frac{1}{k}\sum_{s=1}^{k}\chi_s\right)-\frac{1}{\sqrt{2p}}\sum_{k=1}^{p}\frac{1}{k}\sum_{s=1}^{k}(\chi_s-1)\xrightarrow{\mathcal{D}}0,\quad n\to\infty.$$

Поэтому для завершения доказательства осталось проверить, что

$$\frac{1}{\sqrt{2p}}\sum_{k=1}^{p}\frac{1}{k}\sum_{s=1}^{k}(\chi_s-1)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1),\quad p\to\infty. \tag{37}$$

Очевидно, что

$$\sum_{k=1}^{p}\frac{1}{k}\sum_{s=1}^{k}(\chi_s-1)=\sum_{s=1}^{p}(\chi_s-1)w(s,p), \tag{38}$$

где веса $w(s,p)$ определены в (7). Заметим также, что

$$w(s,p)=\int_s^p\frac{1}{x}\,dx+\int_s^p O\Bigl(\frac{1}{x^2}\Bigr)\,dx=\log\Bigl(\frac{p}{s}\Bigr)+O\Bigl(\frac{1}{s}\Bigr)+O\Bigl(\frac{1}{p}\Bigr).$$

Отсюда, интегрируя по частям, получаем

$$\sum_{s=1}^{p} w^2(s,p) = p\int\limits_{1/p}^{1} \log^2(x)\,dx + O(1) = 2p + O\big(\log^2(p)\big),$$
$$\sum_{s=1}^{p} w^3(s,p) = -p\int\limits_{1/p}^{1} \log^3(x)\,dx + O(1) = 6p + O(\log^3(p)), \tag{39}$$
$$\sum_{s=1}^{p} w^4(s,p) = p\int\limits_{1/p}^{1} \log^4(x)\,dx + O(1) = 24p + O(\log^4(p)).$$

Поэтому с помощью формулы Тейлора находим, что для всех $|\lambda| \leqslant (1-\varepsilon)/\log(p)$

$$\mathbf{E}\exp\left[\lambda\sum_{s=1}^{p}(\chi_s - 1)w(s,p)\right] = \exp\left[-\sum_{s=1}^{p}\log[1-\lambda w(s;p)] - \lambda\sum_{s=1}^{p} w(s,p)\right] =$$
$$= \exp\left[\frac{\lambda^2}{2}\sum_{s=1}^{p} w^2(s,p) + \frac{\lambda^3}{3}\sum_{s=1}^{p} w^3(s,p) + \frac{\lambda^4}{4}\sum_{s=1}^{p} w^4(s,p) + O\Big(\frac{\lambda^5 p}{\varepsilon^5}\Big)\right] =$$
$$= \exp\big\{(1+o(1))p\lambda^2\big\}. \tag{40}$$

Это соотношение очевидным образом доказывает (37) и, следовательно, теорему 1. ▲

Доказательство леммы 1 практически непосредственно вытекает из (25) и выпуклости функции

$$f(x) = r^{-1}\left[\log\left(\frac{1}{\sqrt{\pi}k}\sum_{s=1}^{n+1}\chi_s\right) - \log(1+x)\right]. \quad ▲$$

Доказательство леммы 2 основано на методе Лапласа [15], который используется для вычисления интеграла в правой части следующего тождества:

$$\mathbf{P}\big\{\zeta(\theta^n,p',q) \geqslant y\big\} =$$
$$= \frac{1}{2\pi i}\int\limits_{-\infty}^{\infty} \exp\big\{-ity - \log(t) + \log\big[\mathbf{E}\exp\big(it\zeta(\theta^n,p',q)\big)\big]\big\}\,dt =$$
$$= \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} \exp\Big\{-it - \log(i\sqrt{2\pi}t) + \log\big[\mathbf{E}\exp\big(ity^{-1}\zeta(\theta^n,p',q)\big)\big]\Big\}\,dt. \tag{41}$$

Мы опустим технические детали этого метода, чтобы не загромождать изложение, и сосредоточимся на его принципиальных элементах.

С помощью формулы Тейлора находим (см. (40)), что при $|\lambda| \leqslant (1-\varepsilon)/w(1,q-p')$

$$\log\big\{\mathbf{E}\exp\big[\lambda\zeta(\theta^n,p',q)\big]\big\} = \frac{\lambda p'}{2} - \frac{p'}{2}\log(1+\lambda) + \frac{\lambda^2\|\theta^n\|_{p'}^2}{2(1+\lambda)} -$$
$$- \sum_{s=1}^{q-p'}\big\{\log[1-\lambda w(s,q-p')] + \lambda w(s,q-p')\big\} =$$

$$= \frac{\lambda^2}{2}\left[\frac{\|\theta^n\|^2_{p'}}{1+\lambda} + \frac{p'}{2} + \sum_{s=1}^{q-p'} w^2(s, q-p')\right] + \frac{\lambda^3}{3}\sum_{s=1}^{q-p'} w^3(s, q-p') + \lambda^4 O\left(\frac{q-p'}{\varepsilon^4}\right) =$$

$$= \frac{\lambda^2}{2} D(\lambda) + \lambda^3 M + \lambda^4 O\left(\frac{q-p'}{\varepsilon^4}\right), \tag{42}$$

где функция $D(\cdot)$ определена в (12), а (см. (39))

$$M = \frac{1}{3}\sum_{s=1}^{q-p'} w^3(s, q-p') = 2(q-p')(1+o(1)), \quad q-p' \to \infty.$$

Полагая в (42) $\lambda = it/y$, получаем

$$\log\Big\{\mathbf{E}\exp\big[ity^{-1}\zeta(\theta^n, p', q)\big]\Big\} = -\frac{t^2}{2y^2} D\left(\frac{it}{y}\right) + \frac{it^3}{y^3} M + O\left(\frac{t^4(q-p')}{y^4\varepsilon^4}\right). \tag{43}$$

Метод Лапласа основан на квадратичной аппроксимации (см. (41) и (43)) функции

$$F(t) = -it - \log(i\sqrt{2\pi}t) + \log\big[\mathbf{E}\exp\big(ity^{-1}\zeta(\theta^n, p', q)\big)\big] =$$

$$= -it - \frac{t^2}{2y^2} D\left(\frac{it}{y}\right) + i\frac{t^3}{y^3} M - \log(i\sqrt{2\pi}t) + O\left(\frac{t^4(q-p')}{y^4\varepsilon^4}\right) \tag{44}$$

в окрестности некоторой точки $t_y$, находящейся вблизи точки экстремума $F(\cdot)$. В качестве такой точки можно выбрать, например,

$$t_y = -\frac{iy^2}{D(0)}.$$

Заметим, что поскольку в рассматриваемом случае $y = \sqrt{2xD(0)}$, то

$$t_y = -2ix \quad \text{и} \quad \frac{t_y}{y} = -i\sqrt{\frac{2x}{D(0)}}.$$

Тогда из (13) и (44) получаем

$$\begin{aligned}
&F(t_y) = -x + x\left[\frac{1}{D(0)} D\left(\sqrt{\frac{2x}{D(0)}}\right) - 1\right] - \log\big(2\sqrt{2\pi}x\big) + \\
&+ O\left(\frac{x(q-p')}{D(0)w(1; q-p')}\right), \\
&F'(t_y) = -i + \frac{i}{D(0)}\left[D\left(\sqrt{\frac{2x}{D(0)}}\right) + \frac{1}{2}\sqrt{\frac{2x}{D(0)}} D'\left(\sqrt{\frac{2x}{D(0)}}\right) - D(0)\right] - \frac{1}{x} + \\
&+ O\left(\frac{q-p'}{D(0)\sqrt{w(1; q-p')}}\right), \\
&F''(t_y) = -\frac{1}{2xD(0)}\left[D\left(\sqrt{\frac{2x}{D(0)}}\right) + \frac{3}{2}\sqrt{\frac{2x}{D(0)}} D'\left(\sqrt{\frac{2x}{D(0)}}\right) + \right. \\
&\left. + \frac{2x}{D(0)} D''\left(\sqrt{\frac{2x}{D(0)}}\right)\right] - \frac{1}{x^2} + O\left(\frac{q-p'}{xD(0)\sqrt{w(1; q-p')}}\right).
\end{aligned} \tag{45}$$

Нетрудно проверить, что

$$\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}\exp\left[\frac{F''(t_y)}{2}(t-t_y)^2+F(t_y)(t-t_y)+F(t_y)\right]dt=$$

$$=\exp\left[F(t_y)-\frac{[F'(t_y)]^2}{2F''(t_y)}-\frac{1}{2}\log[-F''(t_y)]\right]. \tag{46}$$

Поэтому заметив, что

$$\frac{1}{D(0)}D\left(\sqrt{\frac{2x}{D(0)}}\right)-1\leqslant 0,$$

из (45) находим

$$F(t_y)-\frac{[F'(t_y)]^2}{2F''(t_y)}-\frac{1}{2}\log[-F''(t_y)]\leqslant -x-\frac{1}{2}\log(x)+O\left(\frac{x(q-p')}{w(1,q-p')}\right).$$

Это неравенство вместе с (41) и (46) доказывает (14). ▲

В заключение автор выражает благодарность анонимному рецензенту за замечания, способствовавшие улучшению статьи.

## СПИСОК ЛИТЕРАТУРЫ

1. *Zhang C., Bengio S., Hardt M., Recht B., Vinyals O.* Understanding Deep Learning (Still) Requires Rethinking Generalization // Commun. ACM. 2021. V. 64. № 3. P. 107–115. https://doi.org/10.1145/3446776
2. *Belkin M.* Fit without Fear: Remarkable Mathematical Phenomena of Deep Learning through the Prism of Interpolation // Acta Numer. 2021. V. 30. P. 203–248. https://doi.org/10.1017/S0962492921000039
3. *Belkin M., Hsu D., Xu J.* Two Models of Double Descent for Weak Features // SIAM J. Math. Data Sci. 2020. V. 2. № 4. P. 1167–1180. https://doi.org/10.1137/20M1336072
4. *Dar Y., Muthukumar V., Baraniuk R.G.* A Farewell to the Bias-Variance Tradeoff? An Overview of the Theory of Overparameterized Machine Learning, https://arxiv.org/abs/2109.02355 [stat.ML], 2021.
5. *Добрушин Р.Л.* Одна статистическая задача теории обнаружения сигнала на фоне шума в многоканальной системе, приводящая к устойчивым законам распределения // Теория вероятн. и ее примен. 1958. Т. 3. № 2. С. 173–185. https://www.mathnet.ru/rus/tvp4928
6. *Бурнашев М.В., Бегматов И.А.* Об одной задаче обнаружения сигнала, приводящей к устойчивым распределениям // Теория вероятн. и ее примен. 1990. Т. 35. № 3. С. 557–560. https://www.mathnet.ru/rus/tvp1261
7. *Ingster Yu.I., Suslina I.A.* Nonparametric Goodness-of-Fit Testing Under Gaussian Models // Lect. Notes Statist. V. 169. New York: Springer-Verlag, 2003. https://doi.org/10.1007/978-0-387-21580-8
8. *Bonferroni C.E.* Teoria statistica delle classi e calcolo delle probabilità // Pubbl. del R. Ist. Super. di Sci. Econ. e Commer. di Firenze. V. 8. Firenze: Seeber, 1936.
9. *Benjamini Y., Hochberg Y.* Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing // J. Roy. Statist. Soc. Ser. B. 1995. V. 57. № 1. P. 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x
10. *Benjamini Y.* Simultaneous and Selective Inference: Current Successes and Future Challenges // Biom. J. 2010. V. 52. № 6. P. 708–721. https://doi.org/10.1002/bimj.200900299

11. *Donoho D., Jin J.* Higher Criticism Thresholding: Optimal Feature Selection When Useful Features are Rare and Weak // Proc. Natl. Acad. Sci. U.S.A. 2008. V. 105. № 39. P. 14790–14795. https://doi.org/10.1073/pnas.0807471105
12. *Anderson T.W.* The Integral of a Symmetric Unimodal Function over a Symmetric Convex Set and Some Probability Inequalities // Proc. Amer. Math. Soc. 1955. V. 6. № 2. P. 170–176. https://doi.org/10.1090/S0002-9939-1955-0069229-1
13. *Ибрагимов И.А., Хасьминский Р.З.* Асимптотическая теория оценивания. М.: Наука, 1979.
14. *Pyke R.* Spacings // J. Roy. Statist. Soc. Ser. B. 1965. V. 27. № 3. P. 395–436; 437–449 (discussion). https://doi.org/10.1111/j.2517-6161.1965.tb00602.x; https://doi.org/10.1111/j.2517-6161.1965.tb00603.x
15. *Федорюк М.В.* Асимптотика: интегралы и ряды. М.: Наука, 1987.

*Голубев Георгий Ксенофонтович*
Институт проблем передачи информации
им. А.А. Харкевича РАН, Москва
golubev.yuri@gmail.com

Поступила в редакцию
16.05.2022
После доработки
06.12.2022
Принята к публикации
03.01.2023

**ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ**

**Том 59** **2023** **Вып. 1**

УДК 621.391 : 519.16

© 2023 г. **М.Н. Вялый**

# О ПРОВЕРКЕ ВЫПОЛНИМОСТИ АЛГЕБРАИЧЕСКИХ ФОРМУЛ НАД ПОЛЕМ ИЗ ДВУХ ЭЛЕМЕНТОВ[1]

Построен вероятностный полиномиальный алгоритм проверки выполнимости алгебраических формул глубины 3 над полем из двух элементов, верхней операцией в которых является сложение. Алгоритм с теми же характеристиками существует для проверки равенства нулю многочлена (задача PIT), задаваемого формулами указанного вида. Однако эти задачи и алгоритмы их решения существенно отличаются. Вероятностный алгоритм для задачи PIT основан на лемме Шварца – Зиппеля, а предложенный в этой статье алгоритм проверки выполнимости основан на лемме Вельянта – Вазирани.

*Ключевые слова*: выполнимость булевых формул, вероятностный алгоритм, алгебраические формулы.

**DOI:** 10.31857/S0555292323010059, **EDN:** RMKBBO

## § 1. Введение

Задача PIT (проверка равенства нулю многочлена, задаваемого алгебраической формулой или схемой) активно изучалась в последние десятилетия. Интерес к этой задаче во многом связан с проблемой дерандомизации. Для задачи PIT существуют вероятностные полиномиальные алгоритмы, но неизвестны детерминированные полиномиальные алгоритмы. Известно, что построение таких алгоритмов приводит к хорошим нижним оценкам в алгебраической сложности [1]. Для многих частных видов формул были найдены полиномиальные детерминированные алгоритмы (см. обзоры [2,3]). Однако уже случай формул глубины 3 оказывается очень трудным. Как было доказано в [4], для любой схемы над полем $\mathbb{C}$ комплексных чисел существует эквивалентная формула глубины 3 субэкспоненциального размера.

С точки зрения теории вычислительной сложности естественно рассмотреть другую задачу: выполнимости алгебраической формулы (или схемы). Над бесконечным или достаточно большим (по сравнению со степенью формулы) конечным полем разницы между этими задачами нет. Однако для малых конечных полей эти задачи различаются, и нет прямой сводимости одной из них к другой.

В этой короткой заметке мы обсуждаем задачу проверки выполнимости алгебраических формул над наименьшим конечным полем $\mathbb{F}_2$ из двух элементов. Эта задача оказывается NP-полной уже для формул глубины 3. Однако, как это принято в исследованиях задачи PIT, если различать формулы по расположению функциональных элементов в слоях, ситуация оказывается более интересной. NP-полнота имеет место для формул вида ΠΣΠ (верхний уровень – умножение). При этом для таких формул задача PIT решается за детерминированное полиномиальное время

[1] Работа выполнена в рамках Программы фундаментальных исследований НИУ ВШЭ, а также частично финансировалась в рамках госзадания 0063-2019-0003.

(см. замечание 1 в § 2). Для второго класса схем глубины 3 – формул вида ΣΠΣ (верхний уровень – сложение) – существование детерминированного полиномиального алгоритма для задачи PIT проблематично в силу упомянутого выше результата о представлении произвольных схем ΣΠΣ-формулами (в [4] используются именно они). Проверка выполнимости ΣΠΣ-формулы над $\mathbb{F}_2$ также выглядит нетривиальной задачей.

Основной результат этой статьи – вероятностный полиномиальный алгоритм проверки выполнимости для ΣΠΣ-формул над $\mathbb{F}_2$. Интересно отметить, что основан этот алгоритм не на лемме Шварца – Зиппеля (см., например, [2]), как вероятностные алгоритмы для задачи PIT, а на лемме Вельянта – Вазирани, изолирующей одну из единиц булевой формулы (см. [5]).

## § 2. Задача выполнимости для алгебраических формул

Здесь мы рассматриваем алгебраические формулы малой глубины над полем $\mathbb{F}_2$ из двух элементов. Входная степень функциональных элементов – сложения и умножения – предполагается неограниченной.

Формулы глубины 1 бывают двух видов: П-формулы, т.е. произведения переменных, и Σ-формулы, которые задают линейные функции.

Среди формул глубины 2 мы выделяем ΣΠ-формулы (суммы произведений переменных) и ΠΣ-формулы (произведения линейных функций).

Аналогично, ΠΣΠ-формулы – это формулы глубины 3, которые являются произведениями ΣΠ-формул, а ΣΠΣ-формулы – это формулы глубины 3, которые являются суммами ΠΣ-формул.

Каждая алгебраическая формула над $\mathbb{F}_2$ задает функцию $\mathbb{F}_2^n \to \mathbb{F}_2$, где $n$ – количество переменных. Если эта функция не равна тождественно нулю, формула называется выполнимой. Задачу проверки выполнимости алгебраической формулы обозначим через AlgSAT. Она лежит в классе NP по очевидным причинам: вычисление значения формулы возможно за полиномиальное время.

Проверка выполнимости Σ- или П-формул тривиальна. Почти столь же проста проверка выполнимости ΣΠ-формулы: нужно сократить одинаковые слагаемые (в поле $\mathbb{F}_2$ выполняется тождество $x + x = 0$), и формула выполнима, если и только если результат сокращения содержит хотя бы один ненулевой моном (многочлен Жегалкина для булевой функции единствен).

Проверка выполнимости ΠΣ-формул также лежит в классе P. Пусть

$$\varphi(x_1, \ldots, x_n) = \prod_{i=1}^{s} \Big( a_{i0} + \sum_{j=1}^{n} a_{ij} x_j \Big). \tag{1}$$

Выполнимость $\varphi$ равносильна существованию решения системы линейных уравнений

$$\sum_{j=1}^{n} a_{ij} x_j = 1 + a_{i0}, \quad 1 \leqslant i \leqslant s, \tag{2}$$

что проверяется стандартными средствами линейной алгебры. Заметим, что множество тех $x \in \mathbb{F}_2^n$, для которых $\varphi(x) = 1$, образует аффинное подпространство координатного пространства $\mathbb{F}_2^n$.

Обозначим через ΠΣΠ-SAT задачу проверки выполнимости ΠΣΠ-формулы. Это уже трудная задача.

Предложение 1. *Задача* ΠΣΠ-SAT NP-*полна.*

Доказательство. Построим полиномиальную сводимость стандартной NP-полной задачи 3SAT (выполнимость 3-КНФ) к ΠΣΠ-SAT. Для этого заметим, что каждый дизъюнкт в КНФ представляется в виде многочлена Жегалкина, и этот многочлен имеет размер $O(1)$, так как в дизъюнкт входит только три литерала. Таким образом, по 3-КНФ получается эквивалентная ΠΣΠ-формула. ▲

*Замечание* 1. Задача PIT для ΠΣΠ-формул проще общего случая, для нее есть детерминированный полиномиальный алгоритм. В кольце многочленов над любым полем нет делителей нуля. Поэтому ΠΣΠ-формула задает нулевой многочлен тогда и только тогда, когда один из множителей верхнего произведения задает нулевой многочлен. Поэтому достаточно в каждом из этих множителей сократить подобные и проверить, остается ли после сокращения хотя бы один ненулевой моном.

Из предложения 1 следует NP-полнота проверки выполнимости любых формул глубины $O(1)$, за исключением случая ΣΠΣ-формул. Именно этот оставшийся случай и является главным предметом изучения в данной статье. Обозначим через ΣΠΣ-SAT задачу проверки выполнимости таких формул. Какова ее алгоритмическая сложность? В следующем параграфе мы приводим вероятностный полиномиальный алгоритм для задачи ΣΠΣ-SAT. Существование такого алгоритма делает крайне сомнительной полноту этой задачи в классе NP.

У задачи ΣΠΣ-SAT есть естественная геометрическая интерпретация. ΣΠΣ-формула обращается в единицу в точности в тех точках, в которых нечетное количество ее ΠΣ слагаемых обращается в единицу. Единицы ΠΣ-формулы – это аффинное подпространство, и любое аффинное подпространство представляется как множество единиц некоторой ΠΣ-формулы. Выбор такой формулы не однозначен. Один из возможных вариантов: если подпространство задано как множество решений системы линейных уравнений (2), то его точки совпадают с множеством единиц формулы (1).

Таким образом, равносильная формулировка задачи ΣΠΣ-SAT состоит в том, что дан набор аффинных подпространств, и нужно узнать, существует ли точка, покрытая нечетным количеством подпространств. Отсюда также следует, что общая задача ΣΠΣ-SAT сводится к случаю ΣΠΣ-формул, в которых каждое слагаемое имеет степень не более $n$ (количество переменных в наших обозначениях).

## § 3. Вероятностный алгоритм для задачи ΣΠΣ-SAT

Напомним, что класс RP (вероятностные вычисления с односторонней ошибкой) состоит из тех языков $L$, для которых существует вероятностный полиномиальный алгоритм $A$, выдающий результаты из множества $\{0,1\}$, такой что

$$x \in L \iff \Pr[A(x)=1] \geqslant \frac{1}{2},$$
$$x \notin L \iff \Pr[A(x)=1] = 0.$$

Теорема 1. *Задача* ΣΠΣ-SAT *принадлежит классу* RP.

Доказательство. Пусть ΣΠΣ-формула имеет вид

$$\varphi(x) = \sum_{i=1}^{s} p_i(x), \quad \text{где } p_i(x) \text{ – ΠΣ-формула.} \tag{3}$$

Обозначим через $M$ матрицу размера $2^n \times s$, матричные элементы которой имеют вид $M_{xi} = p_i(x)$. Условие выполнимости формулы равносильно условию $M\mathbb{1} \neq 0$, так как $\varphi(x) = (M\mathbb{1})_x$. Здесь и далее через $\mathbb{1}$ обозначается вектор-столбец размерности $s$,

все координаты которого равны 1, а через 0 – любой нулевой вектор (размерность которого будет ясна из контекста).

В координатном пространстве векторов-строк $(\mathbb{F}_2^n)^*$ выберем случайный базис $b_1, \ldots, b_n$ по равномерному распределению. Для каждого $0 \leqslant i \leqslant n$ определим вектор-строку $h_i$ размера $2^n$, координаты которого индексированы векторами-столбцами $x \in \mathbb{F}_2^n$. Значения координат задаются следующим правилом: $(h_i)_x = 1$ тогда и только тогда, когда $b_1 x = b_2 x = \ldots = b_i x = 0$, если $i > 0$, и $(h_0)_x = 1$ для всех $x$. Вычислим $h_i M \mathbb{1}$, находя сначала $h_i M$ и умножая затем полученную строку размера $s$ на вектор-столбец $\mathbb{1}$.

Если $h_i M \mathbb{1} = 1$, то выдаем ответ "$\varphi(x)$ выполнима". Если при всех проверках $h_i M \mathbb{1} = 0$, то выдаем ответ "$\varphi(x)$ невыполнима".

Оценим время работы алгоритма. В худшем случае он выполняет $n + 1$ вычислений $h_i M \mathbb{1}$. Вычисление $h_i M$ возможно за полиномиальное время, так как $(h_i M)_k$ равно четности количества точек в аффинном подпространстве

$$L_k = \{x : \, p_k(x) = 1\} \cap \{x : \, b_1 x = b_2 x = \ldots = b_i x = 0\}.$$

Количество точек в подпространстве размерности $d$ равно $2^d$, размерность $L_k$ находится за полиномиальное время стандартными алгоритмами линейной алгебры. Заметим также, что $s$ ограничено длиной входа (описание формулы $\varphi(x)$).

Докажем корректность алгоритма. Если $M\mathbb{1} = 0$, то $hM\mathbb{1} = 0$ для любого вектора-строки $h$. Поэтому вероятность ошибки алгоритма в случае невыполнимой формулы $\varphi(x)$ равна 0.

Предположим, что формула выполнима, т.е. $M\mathbb{1} \neq 0$. Обозначим через $S$ множество тех $x \in \mathbb{F}_2^n$, для которых $(M\mathbb{1})_x = 1$ (другими словами, $\varphi(x) = 1$). Поскольку $\varphi(x)$ выполнима, $S \neq \varnothing$. Лемма Вельянта – Вазирани [5, теорема 2.4] утверждает, что в этом случае с вероятностью не менее $1/2$ для некоторого $0 \leqslant i \leqslant n$ выполняется

$$|S_i| = 1, \quad \text{где} \quad S_i = S \cap \{x : \, b_1 x = b_2 x = \ldots = b_i x = 0\}. \tag{4}$$

Заметим, что $h_i M \mathbb{1}$ равно четности $|S_i|$. Поэтому из (4) следует $h_i M \mathbb{1} = 1$. Значит, алгоритм с вероятностью не менее $1/2$ выдает ответ "$\varphi(x)$ выполнима". Таким образом, ΣΠΣ-SAT $\in$ RP. ▲

## § 4. О дерандомизации основной теоремы

Существует ли детерминированный полиномиальный алгоритм для задачи ΣΠΣ-SAT? Вопрос требует дальнейшего изучения, здесь мы ограничимся несколькими замечаниями.

Возможность дерандомизации леммы Вельянта – Вазирани представляется маловероятной (см. [6, 7]). Прямолинейная дерандомизация алгоритма из теоремы 1 заведомо невозможна по соображениям размерности (линейное отображение пространства размерности $2^n$ в пространство полиномиальной размерности всегда имеет ненулевое ядро). Заметим, что для задачи PIT возможность дерандомизации путем предъявления полиномиального по размеру количества точек, в которых достаточно вычислить значения формулы, считается вполне возможной, и существуют гипотезы о возможном виде таких множеств. Одна из таких гипотез предложена в [1]. Заметим также, что из $\tau$-гипотезы Куарана [8] следует возможность дерандомизации указанного вида для формул над $\mathbb{R}$.

Если $\varphi$ – ΣΠΣ-формула, то и $1 + \varphi$ также ΣΠΣ-формула. Поэтому задача ΣΠΣ-SAT также эквивалентна проверке того, что хотя бы одна точка покрыта четным количеством аффинных подпространств из заданного набора. Заметим, что условие четности существенно слабее других условий на кратность покрытия. В [9] до-

казано, что проверка покрытия $\mathbb{F}_2^n$ заданным набором аффинных подпространств co-NP-полна. С учетом теоремы 1 это означает, что есть значительная разница в трудности проверки существования точки, покрытой четным количеством подпространств (лежит в RP), и существования точки, покрытой нулем подпространств (NP-полна).

Для вероятностного алгоритма из теоремы 1 существенна простота множеств единиц ΠΣ-формул. Действительно, теорема Вельянта – Вазирани [5] предоставляет вероятностную сводимость NP к классу $\oplus$P, полной задачей в которой является $\oplus$SAT: проверка того, что данная булева формула выполняется в нечетном количестве точек. Алгоритм из теоремы 1 неприменим в случае произвольных булевых формул, потому что необходимый для вычисления $h_i M$ подсчет количества единиц булевой формулы является #P-полной задачей.

Однако все эти замечания не исключают, конечно, существования детерминированного полиномиального алгоритма для ΣΠΣ-SAT.

В заключение рассмотрим несколько частных случаев ΣΠΣ-формул, для которых возможна проверка выполнимости за детерминированное полиномиальное время.

Самый простой пример – схемы ограниченной степени. Если все слагаемые в ΣΠΣ-формуле имеют степень $O(1)$, то раскрытием скобок и приведением подобных за полиномиальное время можно преобразовать ΣΠΣ-формулу в равносильную ΣΠ-формулу. Заметим, что в этом случае соответствующие аффинные подпространства имеют ограниченную коразмерность, так как количество уравнений в системе (2) для каждого слагаемого имеет величину $O(1)$.

Построить нетривиальные примеры формул, которые тождественно равны нулю на $\mathbb{F}_2^n$, можно следующим образом. Пусть $u_1, \ldots, u_t$ – набор точек в $\mathbb{F}_2^n$. Возьмем любой граф $G$ на $t$ вершинах, степень каждой из которых четна, и построим набор аффинных подпространств

$$A_{ij} = \{u_i, u_j\}, \quad \{i,j\} \in E(G)$$

(любая пара точек в $\mathbb{F}_2^n$ образует аффинное подпространство). Каждая точка $\mathbb{F}_2^n$ входит в четное (0 или 2) количество таких подпространств. Соответствующие формулы имеют вид

$$\varphi_G(x) = \sum_{\{i,j\}\in E(G)} q_{ij}(x), \quad \text{где} \quad q_{ij} = \prod_{k=1}^{n-1} \ell_k^{\{i,j\}}(x),$$

а множители $\ell_k^{\{i,j\}}(x)$ определяются через векторы $a_1^{\{i,j\}}, \ldots, a_{n-1}^{\{i,j\}}$ базиса ортогонального дополнения к $\mathbb{F}_2(u_j - u_i)$ следующим образом:

$$\ell_k^{\{i,j\}}(x) = 1 - \sum_{m=1}^{n} a_{km}^{\{i,j\}} u_{im} + \sum_{m=1}^{n} a_{km}^{\{i,j\}} x_m$$

(хотя $-1 = +1$ в $\mathbb{F}_2$, знаки расставлены для удобства чтения).

Обобщая этот пример, получаем другое семейство формул, для проверки выполнимости которых существует детерминированный полиномиальный алгоритм. Это такие формулы, в которых размерность аффинного пространства, отвечающего каждому слагаемому формулы, равна $d = O(1)$. Такое подпространство содержит лишь $2^d = O(1)$ точек $\mathbb{F}_2^n$, так что формулы из данного семейства принимают значение 1 лишь на некотором подмножестве объединения всех указанных подпространств. Мощность этого объединения не превосходит $2^d s$, где $s$ – количество слагаемых верхнего уровня (как в (3)). Найти размерность решения системы (2) линейных уравнений и список всех ее решений можно стандартными методами линейной

алгебры, в данном случае это выполнимо за время, полиномиально ограниченное длиной входа (размером записи формулы).

Более сложный алгоритм проверки выполнимости использует формулу для числа единиц суммы по модулю 2 булевых функций, аналогичную формуле включений и исключений. Пусть $f_1, \ldots, f_k$ – функции $\{0,1\}^n \to \{0,1\}$. Сейчас мы рассматриваем 0 и 1 как целые числа. Обозначим

$$f = f_1 \oplus f_2 \oplus \ldots \oplus f_k$$

(складываем значения функций по модулю 2). Тогда

$$(1-2f)(x) = \prod_{i=1}^{k}(1-2f_i(x)),$$

и если обозначить через $N(f)$ количество единиц функции $f$, то получаем

$$N(f) = \sum_{\varnothing \neq S \subset \{0,1\}^k} (-1)^{|S|+1} 2^{|S|-1} N\Big(\prod_{i \in S} f_i(x)\Big) \tag{5}$$

(см. аналогичное вычисление в [10]). Равенство (5) для формулы (3) включает $2^k - 1$ слагаемых. Однако часть этих слагаемых может равняться нулю. Например, пусть никакие $q = O(1)$ слагаемых в (3) не совместны, т.е. соответствующие аффинные подпространства не пересекаются. Тогда количество ненулевых слагаемых в (5) имеет величину $O(n^q)$, и вычисление по этой формуле занимает полиномиальное время (напомним, что подсчет числа точек в аффинном подпространстве возможен за полиномиальное время).

## СПИСОК ЛИТЕРАТУРЫ


1. *Agrawal M.* Proving Lower Bounds via Pseudo-random Generators // FSTTCS 2005: Foundations of Software Technology and Theoretical Computer Science (Proc. 25th Int. Conf. Hyderabad, India. Dec. 15–18, 2005). Lect. Notes Comput. Sci. V. 3821. Berlin: Springer, 2005. P. 92–105. https://doi.org/10.1007/11590156_6
2. *Saxena N.* Progress on Polynomial Identity Testing // Bull. Eur. Assoc. Theor. Comput. Sci. 2009. № 90. P. 49–79.
3. *Saxena N.* Progress on Polynomial Identity Testing-II // Perspectives in Computational Complexity. Progr. Comput. Sci. Appl. Logic. V. 26. Cham: Birkhäuser, 2014. P. 131–146. https://doi.org/10.1007/978-3-319-05446-9_7
4. *Gupta A., Kamath P., Kayal N., Saptharishi R.* Arithmetic Circuits: A Chasm at Depth 3 // SIAM J. Comput. 2016. V. 45. № 3. P. 1064–1079. https://doi.org/10.1137/140957123
5. *Valiant L.G., Vazirani V.V.* NP Is as Easy as Detecting Unique Solutions // Theor. Comput. Sci. 1986. V. 47. № 1. P. 85–93. https://doi.org/10.1016/0304-3975(86)90135-0
6. *Hemaspaandra L.A., Naik A.V., Ogihara M., Selman A.L.* Computing Solutions Uniquely Collapses the Polynomial Hierarchy // SIAM J. Comput. 1996. V. 25. № 4. P. 697–708. https://doi.org/10.1137/S0097539794268315
7. *Dell H., Kabanets, V., van Melkebeek, D., Watanabe O.* Is Valiant–Vazirani's Isolation Probability Improvable? // Comput. Complexity. 2013. V. 22. № 2. P. 345–383. https://doi.org/10.1007/s00037-013-0059-7
8. *Grenet B., Koiran P., Portier N., Strozecki Y.* The Limited Power of Powering: Polynomial Identity Testing and a Depth-Four Lower Bound for the Permanent // Proc. 31st IARCS Annu. Conf. on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2011). Mumbai, India. Dec. 12–14, 2011. Leibniz Int. Proc. Inform. (LIPIcs). V. 13. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany: Dagstuhl Publ., 2011. P. 127–139. https://doi.org/10.4230/LIPIcs.FSTTCS.2011.127

9. *Arvind V., Guruswami V.* CNF Satisfiability in a Subspace and Related Problems // Algorithmica. 2022. V. 84. № 11. P. 3276–3299. https://doi.org/10.1007/s00453-022-00958-4
10. *Леонтьев В.К., Морено О.* О нулях булевых полиномов // Ж. вычисл. матем. и матем. физ. 1998. Т. 38. № 9. С. 1608–1615. https://www.mathnet.ru/rus/zvmmf1832

*Вялый Михаил Николаевич*
Федеральный исследовательский центр
"Информатика и управление" РАН, Москва
Национальный исследовательский университет
"Высшая школа экономики", Москва
Московский физико-технический институт
(государственный университет), Москва
vyalyi@gmail.com

Поступила в редакцию
18.12.2022
После доработки
12.02.2023
Принята к публикации
13.02.2023

**ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ**

**Том 59** **2023** **Вып. 1**

УДК 621.391 : 519.872.6

© 2023 г. **Б.Я. Лихтциндер, А.Ю. Привалов, В.И. Моисеев**

# НЕОРДИНАРНЫЕ ПУАССОНОВСКИЕ МОДЕЛИ ТРАФИКА МУЛЬТИСЕРВИСНЫХ СЕТЕЙ

Появление сетей передачи данных с коммутацией пакетов показало, что пуассоновские модели потоков не являются адекватными, и потребовало разработки новых моделей, основанных на непуассоновских распределениях. Статья посвящена анализу частного случая группового марковского потока – группового неординарного пуассоновского потока событий. В таком потоке выполняются свойство стационарности и отсутствия последействия, но не выполняется свойство ординарности. Рассматривается класс систем массового обслуживания с постоянным временем обслуживания. Приведены результаты аналитических расчетов параметров потока и результаты имитационного моделирования. Показано, что дисперсия очереди зависит от третьего момента размера пачки заявок во входящем групповом пуассоновском потоке.

*Ключевые слова*: системы массового обслуживания, групповой пуассоновский поток, групповой неординарный поток, системы с постоянным временем обслуживания.

**DOI:** 10.31857/S0555292323010060, **EDN:** RMRNDJ

## § 1. Групповой неординарный пуассоновский поток

В попытках адекватного описания поведения трафика в сетях с коммутацией пакетов исследовались модели потоков с распределениями Вейбулла, Эрланга, гамма-распределениями, распределениями Парето и ряд других. Описание сложных коррелированных потоков в современных телекоммуникационных сетях часто производилось с использованием “фрактальных” процессов. Сотни работ посвящены анализу “самоподобного” трафика. Однако из-за сложности моделей этого класса использование их на практике затруднительно. Недостаточная эффективность представления трафика моделями “самоподобных” процессов привела к созданию целого класса моделей потоков, управляемых цепью Маркова. Этапы развития указанных моделей хорошо представлены в обзоре [1].

Особое место среди таких потоков занимают так называемые групповые марковские потоки (BMAP – Batch Markovian Arrival Processes) [2–7].

Одной из разновидностей BMAP-потоков является групповой неординарный пуассоновский поток событий. В таком потоке выполняются свойство стационарности и отсутствия последействия, но не выполняется свойство ординарности. Рассмотрим пуассоновский поток независимых событий с параметром $\lambda$. Каждое событие заключается в одновременном появлении в момент $t_k$ группы (“пачки”) из $\mu_k$ независимых случайно распределенных чисел заявок с распределением

$$\mathbf{P}\,|\mu_k = k| = f_k.$$

Такой поток называют пуассоновским неординарным (групповым) потоком независимых событий [7]. Здесь и далее будем рассматривать класс систем массового обслуживания с постоянным временем обслуживания заявок.

Выделим некоторый интервал времени. Пусть $\tau$ – интервал времени обработки одной заявки. Разделим достаточно большой промежуток времени $T$, в течение которого действует поток указанных событий, на $N_\tau$ последовательных интервалов $\tau$. Пусть $n_i(\tau)$ – число событий, произошедших в течение $i$-го интервала времени.

Поскольку поток событий пуассоновский, вероятности наступления на интервале ровно $n$ событий подчиняются закону Пуассона:

$$\mathbf{P}\,|n_i(\tau) = n| = P_n(\lambda\tau) = \frac{(\lambda\tau)^n}{n!}e^{-\lambda\tau}.$$

Каждому из событий сопутствует появление пачки с распределением вероятностей чисел заявок $f_i$. Введем производящую функцию этого распределения

$$f(z) = \sum_{k=0}^{\infty} f_k z^k.$$

Поскольку все $\mu_k$ взаимно независимы и одинаково распределены, появлению на интервале $\tau$ распределения количества заявок $m(\tau)$ при условии, что на указанном интервале произошло $n$ событий пуассоновского потока, соответствует производящая функция $[f(z)]^n$. Отсюда следует, что производящая функция $G_{m(\tau)}(z)$ числа заявок на интервале $\tau$ определяется соотношением

$$G_{m(\tau)}(z) = \sum_{n=0}^{\infty} P_n(\lambda\tau)[f(z)]^n) = \sum_{n=0}^{\infty} \frac{(\lambda\tau)^n}{n!}e^{-\lambda\tau}[f(z)]^n =$$
$$= e^{-\lambda\tau}\sum_{n=0}^{\infty}\frac{(\lambda\tau f(z))^n}{n!} = e^{\lambda\tau[f(z)-1]}.$$

Для факториальных моментов первых трех порядков, обозначая начальные моменты количества заявок в одной пачке через $\overline{k^r}$, $r = 1,2,3$, а начальные моменты количества заявок на интервале через $\overline{m^r(\tau)}$, $r = 1,2,3$, имеем

$$\frac{\partial G_{m(\tau)}(z)}{\partial z}\bigg|_{z=1} = \overline{m(\tau)} = \lambda\tau f'(z)e^{\lambda\tau[f(z)-1]}\big|_{z=1} = \lambda\tau\overline{k},$$
$$\frac{\partial^2 G_{m(\tau)}(z)}{(\partial z)^2}\bigg|_{z=1} = \overline{m^2(\tau)} - \overline{m(\tau)} = \big(\lambda\tau f''(z) + (\lambda\tau f'(z))^2\big)e^{\lambda\tau[f(z)-1]}\big|_{z=1} =$$
$$= \lambda\tau(\overline{k^2} - \overline{k}) + (\lambda\tau\overline{k})^2,$$
$$\frac{\partial^3 G_{m(\tau)}(z)}{(\partial z)^3}\bigg|_{z=1} = \overline{m^3(\tau)} - 3\overline{m^2(\tau)} + 2\overline{m(\tau)} =$$
$$= \big(\lambda\tau f'''(z) + 3(\lambda\tau)^2 f''(z) + (\lambda\tau f'(z))^3\big)e^{\lambda\tau[f(z)-1]}\big|_{z=1} =$$
$$= \lambda\tau(\overline{k^3} - 3\overline{k^2} + 2\overline{k}) + 3(\lambda\tau)^2\overline{k}(\overline{k^2} - \overline{k}) + (\lambda\tau\overline{k})^3.$$

Отсюда соответствующими подстановками получаем

$$\overline{m(\tau)} = \lambda\tau\overline{k}, \quad \overline{m^2(\tau)} = \lambda\tau\overline{k^2} + (\lambda\tau\overline{k^2}),$$
$$\overline{m^3(\tau)} = \lambda\tau\overline{k^3} + 3(\lambda\tau)^2\overline{k}\overline{k^2} + (\lambda\tau\overline{k})^3.$$

Для центральных моментов $m(\tau)$ отсюда получаем

$$\mu_2(m(\tau)) = D_{m(\tau)} = \overline{m^2(\tau)} - (\overline{m(\tau)})^2 = \lambda\tau\overline{k^2},$$
$$\mu_3(m(\tau)) = \overline{m^3(\tau)} - 3\overline{m^2(\tau)m(\tau)} + 2(\overline{m(\tau)})^3 = \lambda\tau\overline{k^3}.$$

Для дальнейшего нам удобно будет выразить дисперсию и третий центральный момент $m(\tau)$ через коэффициент загрузки, т.е.

$$D_{m(\tau)} = \lambda\tau\overline{k^2} = \lambda\tau(D_k + (\overline{k})^2) = \rho\overline{k}(1 + v_k^2),$$
$$\mu_3(m(\tau)) = \lambda\tau\overline{k^3} = \lambda\tau\overline{k^3} = \rho\frac{\overline{k^3}}{\overline{k}},$$

где $\rho = \lambda\tau\overline{k} = \overline{m(\tau)}$ – общий коэффициент загрузки, а $v_k^2 = \dfrac{D_k}{(\overline{k})^2}$ – квадрат коэффициента вариации чисел заявок в пачках. Полученные соотношения указывают на линейную зависимость дисперсии и третьего центрального момента от коэффициента загрузки $\rho$. В частном случае для пуассоновского потока $v_k^2 = 0$, $\overline{k} = 1$, и в результате $D_{m(\tau)} = \mu_3(m(\tau)) = \rho$.

## § 2. Интервальный метод анализа очередей

Одним из возможных направлений изучения пакетного трафика является разрабатываемый нами интервальный метод, позволяющий заменить анализ интервалов времени между соседними заявками и интервалов времени обработки заявок анализом одной случайной величины – числом заявок, поступающих в течение последовательных интервалов времени обработки каждой из заявок. Нами показано, что дисперсия и корреляционные свойства указанной случайной величины при заданной загрузке полностью характеризуют средний размер очереди в системах массового обслуживания [8]. Для любой одноканальной системы массового обслуживания с неограниченной очередью справедливо рекуррентное соотношение, устанавливающее связь между поступающими и обработанными заявками [9]:

$$q_i(\tau) = q_{i-1}(\tau) + m_i(\tau) - \delta_i(\tau), \tag{1}$$
$$\delta_i(\tau) = \begin{cases} 0, & \text{если } q_{i-1}(\tau) = m_i(\tau) = 0, \\ 1 & \text{в противном случае,} \end{cases}$$

где $m_i(\tau)$ и $q_i(\tau)$ – число заявок, поступивших в течение $i$-го интервала $\tau$ и размер очереди, образовавшейся на указанном интервале, соответственно.

Обратим внимание на некоторые особенности величин $\delta_i(\tau)$:

$$\delta_i^2(\tau) = \delta_i(\tau), \quad \delta_i(\tau)m_i(\tau) = m_i(\tau), \quad \overline{\delta_i(\tau)} = \overline{m(\tau)}, \quad \delta_i(\tau)q_{i-1}(\tau) = q_{i-1}(\tau).$$

Предпоследнее равенство легко получить, найдя математическое ожидание левой и правой частей уравнения (1) в стационарном состоянии системы. Возведем в квадрат левую и правую части (1) и найдем математические ожидания полученных выражений.

После некоторых преобразований получаем

$$\overline{q(\tau)} = \frac{D_m(\tau) + 2\overline{q_{i-1}(\tau)\big[m_i(\tau) - \overline{m(\tau)}\,\big]}}{2[1 - m(\tau)]} - \frac{\overline{m(\tau)}}{2}.$$

Обозначим через

$$\mathrm{Cov}_{q_{i-1}m_i}(\tau) = \overline{\big[q_{i-1}(\tau) - \overline{q(\tau)}\,\big]\big[m_i(\tau) - \overline{m(\tau)}\,\big]} = \overline{q_{i-1}(\tau)\big[m_i(\tau) - \overline{m(\tau)}\,\big]}$$

второй взаимный центральный момент указанных последовательностей, называемый ковариацией. Он определяется как математическое ожидание произведений центрированных значений элементов $q_{i-1}(\tau)$ и $m_i(\tau)$. Учитывая, что $\overline{m(\tau)} = \rho$, окончательно получаем

$$\overline{q(\rho)} = \frac{D_m(\rho) + 2\,\mathrm{Cov}_{q_{i-1}m_i}(\rho)}{2(1-\rho)} - \frac{\rho}{2}. \tag{2}$$

Соотношение (2) носит фундаментальный характер, обобщает формулу Хинчина – Поллачека и справедливо для любых стационарных потоков заявок при постоянном времени обслуживания. Дисперсионная и ковариационная составляющие для потоков трафика определенного типа могут быть получены экспериментально и использованы при инженерных расчетах.

Для рассмотренных нами групповых пуассоновских потоков ковариационная составляющая тождественно равна нулю:

$$\mathrm{Cov}_{q_{i-1}m_i}(\rho) = 0, \quad D_m(\rho) = \rho E_m, \quad \text{где } E_m = \overline{k}(1+v_k^2).$$

При этом формула (2) упрощается:

$$\overline{q(\rho)} = \frac{D_m(\rho)}{2(1-\rho)} - \frac{\rho}{2} = \frac{\rho E_m}{2(1-p)} - \frac{\rho}{2}. \tag{3}$$

Дисперсия $D_m(\rho)$ в формуле линейно зависит от коэффициента загрузки $\rho$, а ее значение пропорционально среднему числу заявок в пачке.

Получим интервальным методом выражения для второго начального момента и дисперсии очереди в одноканальной СМО с групповым пуассоновским потоком на входе. Для нахождения начального момента возведем обе части уравнения (1) в третью степень (для краткости опуская аргумент $\tau$):

$$q_i^3 = q_{i-1}^3 + 3q_{i-1}^2(m_i - \delta_i)^2 + (m_i - \delta_i)^3.$$

После аналогичных предыдущему случаю преобразований с учетом особенностей величин $\delta_i$ получаем

$$q_i^3 = q_{i-1}^3 + 3q_{i-1}^2 m_i - 3q_{i-1}^2 + 3q_{i-1}(m_i^2 - 2m_i + 1) + (m_i^3 - 3m_i^2 + 3m_i - \delta_i).$$

После усреднения и учета стационарного состояния системы получаем

$$3\overline{q_{i-1}^2}(1-\overline{m_i}) = 3\overline{q_{i-1}}(\overline{m_i^2} - 2\overline{m_i} + 1) + (\overline{m_i^3} - 3\overline{m_i^2} + 2\overline{m_i}).$$

И далее, используя дисперсию $m_i$:

$$3\overline{q_{i-1}^2}(1-\overline{m_i}) = 3\overline{q_{i-1}}D_m + 3\overline{q_{i-1}}(\overline{m_i} - 1)^2 + \big(\overline{m_i^3} - 3\overline{m_i^2} + 2\overline{m_i}\big).$$

Далее, для второго начального момента получаем

$$\overline{q_{i-1}^2} = \frac{\overline{q_{i-1}}D_m}{(1-\overline{m_i})} + \overline{q_{i-1}} \cdot (1-\overline{m_i}) + \frac{\overline{m_i^3} - 3\overline{m_i^2} + 2\overline{m_i}}{3(1-\overline{m_i})},$$

$$\overline{q_{i-1}^2} = \overline{q_{i-1}}\left[\frac{D_m}{(1-\overline{m_i})} + (1-\overline{m_i})\right] + \frac{\overline{m_i^3} - 3\overline{m_i^2} + 2\overline{m_i}}{3(1-\overline{m_i})},$$

$$\overline{q_{i-1}^2} = \overline{q_{i-1}}\frac{D_m + (1-\overline{m_i})^2}{(1-\overline{m_i})} + \frac{\overline{m_i^3} - 3\overline{m_i^2} + 2\overline{m_i}}{3(1-\overline{m_i})},$$

$$\overline{q_{i-1}^2} = \frac{\overline{q_{i-1}}}{(1-\overline{m_i})}\big[D_m + (1-\overline{m_i})^2\big] + \frac{\overline{m_i^3} - 3\overline{m_i^2} + 2\overline{m_i}}{3(1-\overline{m_i})}.$$

Если перейти от третьего начального момента $m_i$ к третьему центральному, то после некоторых преобразований получим

$$\overline{q_{i-1}^2} = \frac{\overline{q_{i-1}}}{(1-\overline{m_i})}\big[D_m + (1-\overline{m_i})^2\big] + \frac{\overline{m_i^3} - 3\overline{m_i^2}D_m - 3\overline{m_i^2} + 2\overline{m_i}}{3(1-\overline{m_i})} + \frac{\mu_3}{3(1-\overline{m_i})}.$$

Подставляя сюда выражение (3) для средней очереди пуассоновского потока с групповым прибытием, и учитывая, что $\overline{m(\tau)} = \rho$, получаем

$$\overline{q^2(\rho)} = \frac{D_m(\rho) - \rho(1-\rho)}{2(1-\rho)^2}\big[D_m(\rho) + (1-\rho)^2\big] +$$
$$+ \frac{\rho^3 + 3\rho D_m(\rho) - 3D_m(\rho) - 3\rho^2 + 2\rho}{3(1-\overline{\rho})} + \frac{\mu_3(\rho)}{3(1-\rho)}.$$

Таким образом, второй начальный момент размера очереди в СМО с групповыми пуассоновскими потоками определяется соотношением

$$\overline{q^2(\rho)} = \frac{E_m(\rho) - \rho + \rho^2}{2(1-\rho)^2}\big[E_m(\rho) + 1 - 2\rho + \rho^2\big] +$$
$$+ \frac{\rho^3 + 3E_m\rho^2 - 3E_m\rho - 3\rho^2 + 2\rho}{3(1-\overline{\rho})} + \frac{\mu_3(\rho)}{3(1-\rho)}.$$

В частном случае для простейшего пуассоновского потока имеем $E_m = 1$, $\mu_3(\rho) = \rho$, и выражение упрощается:

$$\overline{q^2(\rho)} = \frac{\rho^2}{2(1-\rho)^2}\Big(1 - \frac{1}{3}\rho + \frac{1}{3}\rho^2\Big).$$

Оно показывает, что второй начальный момент очереди пуассоновского потока определяется исключительно значением коэффициента загрузки.

Для дисперсии размера очереди $\overline{D_q(\rho)} = \overline{q^2(\rho)} - \overline{q(\rho)}^2$ имеем

$$\overline{D_q(\rho)} = \frac{D_m(\rho) - \rho(1-\rho)}{2(1-\rho)^2}\big[D_m(\rho) + (1-\rho)^2\big] +$$
$$+ \frac{\rho^3 + 3\rho D_m(\rho) - 3D_m(\rho) - 3\rho^2 + 2\rho}{3(1-\rho)} + \frac{\mu_3(\rho)}{3(1-\rho)} - \frac{\big[D_m(\rho) - \rho(1-\rho)\big]^2}{4(1-\rho)^2} =$$
$$= \frac{D_m(\rho) - \rho(1-\rho)}{2(1-\rho)^2}\big[D_m(\rho) + (1-\rho)^2 - \frac{D_m(\rho) - \rho(1-\rho)}{2}\big] +$$
$$+ \frac{\rho^3 + 3\rho D_m(\rho) - 3D_m(\rho) - 3\rho^2 + 2\rho}{3(1-\rho)} + \frac{\mu_3(\rho)}{3(1-\rho)} =$$
$$= \frac{D_m(\rho) - \rho(1-\rho)}{4(1-\rho)^2}\big[2D_m(\rho) + 2(1-\rho)^2 - D_m(\rho) + \rho(1-\rho)\big] +$$
$$+ \frac{\rho^3 + 3\rho D_m(\rho) - 3D_m(\rho) - 3\rho^2 + 2\rho}{3(1-\rho)} + \frac{\mu_3(\rho)}{3(1-\rho)} =$$
$$= \frac{D_m(\rho) - \rho(1-\rho)}{4(1-\rho)^2}\big[D_m(\rho) + 2 - 4\rho + 2\rho^2 + \rho - \rho^2\big] +$$
$$+ \frac{\rho^3 + 3\rho D_m(\rho) - 3D_m(\rho) - 3\rho^2 + 2\rho}{3(1-\rho)} + \frac{\mu_3(\rho)}{3(1-\rho)} =$$

$$= \frac{D_m(\rho) - \rho(1-\rho)}{4(1-\rho)^2}\left[D_m(\rho) + 2 - 3\rho + \rho^2\right] +$$
$$+ \frac{\rho^3 + 3\rho D_m(\rho) - 3D_m(\rho) - 3\rho^2 + 2\rho}{3(1-\rho)} + \frac{\mu_3(\rho)}{3(1-\rho)}.$$

В результате

$$\overline{D_q(\rho)} = \frac{D_m(\rho) - \rho(1-\rho)}{4(1-\rho)^2}\left[D_m(\rho) + 2 - 3\rho + \rho^2\right] +$$
$$+ \frac{\rho^3 + 3E_m\rho^2 - 3E_m\rho - 3\rho^2 + 2\rho}{3(1-\rho)} + \frac{\mu_3(\rho)}{3(1-\rho)}.$$

Учитывая, что $D_m(\rho) = \rho\overline{\frac{k^2}{k}}$, а $\mu_3(\rho) = \rho\overline{\frac{k^3}{k}}$, окончательно получаем

$$\overline{D_q(\rho)} = \frac{\rho\overline{\frac{k^2}{k}} - \rho(1-\rho)}{4(1-\rho)^2}\left[\rho\overline{\frac{k^2}{k}} + 2 - 3\rho + \rho^2\right] +$$
$$+ \frac{\rho^3 + 3\rho^2\overline{\frac{k^2}{k}} - 3\rho\overline{\frac{k^2}{k}} - 3\rho^2 + 2\rho}{3(1-\rho)} + \frac{\rho\overline{\frac{k^3}{k}}}{3(1-\rho)}.$$

В частном случае для простейшего пуассоновского потока имеем $E_m = 1$, $D_m(\rho) = \mu_3(\rho) = \rho$, и выражение упрощается:

$$\overline{D_q(\rho)} = \frac{\rho^2}{2(1-\rho)^2}\left(1 - \frac{1}{3}\rho - \frac{1}{6}\rho^2\right).$$

Так же как и второй начальный момент, дисперсия очереди простейшего потока полностью определяется коэффициентом загрузки. Дисперсия очередей групповых пуассоновских потоков зависит от третьего центрального момента, который характеризует симметричность закона распределения размеров пачек заявок. Два различных групповых потока, имеющие одинаковые зависимости средних значений очередей от коэффициента загрузки, имеют различные дисперсии очередей, разность $\overline{\Delta D_q(\rho)}$ которых пропорциональна разности $\Delta\mu_3(\rho)$ их третьих центральных моментов:

$$\overline{\Delta D_q(\rho)} = \frac{\Delta\mu_3(\rho)}{3(1-\rho)}.$$

### § 3. Имитационное моделирование

Для подтверждения полученных зависимостей нами были сгенерированы три групповых пуассоновских потока с различными законами распределения вероятностей чисел заявок в пачках. Все три потока имеют одинаковые выборки по 100 000 заявок. Интервальный метод не учитывает распределение моментов времени поступления заявок внутри каждого интервала времени $\tau$. Имитационное моделирование подтверждает возможность такого допущения.

Поток № 1 имеет постоянные числа заявок в пачках $k = 10$.

Поток № 2 имеет экспоненциальное распределение чисел заявок в пачках со средним значением $\overline{k_e} = 0{,}5k = 5$ и коэффициентом вариации $v^2 = 1$. При моделировании число заявок формируется как округление до ближайшего целого от случайной величины, распределенной по экспоненциальному закону.

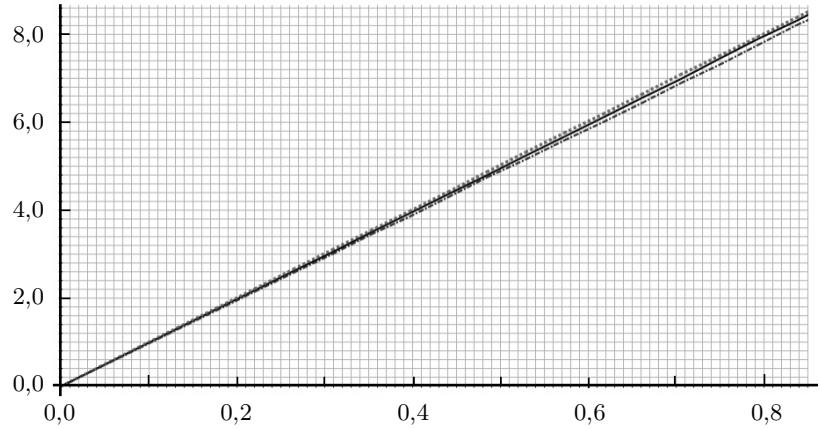


Рис. 1. Зависимости значений дисперсий для трех групповых пуассоновских потоков от коэффициента загрузки (сплошная линия – постоянное число заявок в пачках, точечная линия – пуассоновское распределение чисел заявок в пачках, пунктирная линия – экспоненциальное распределение)

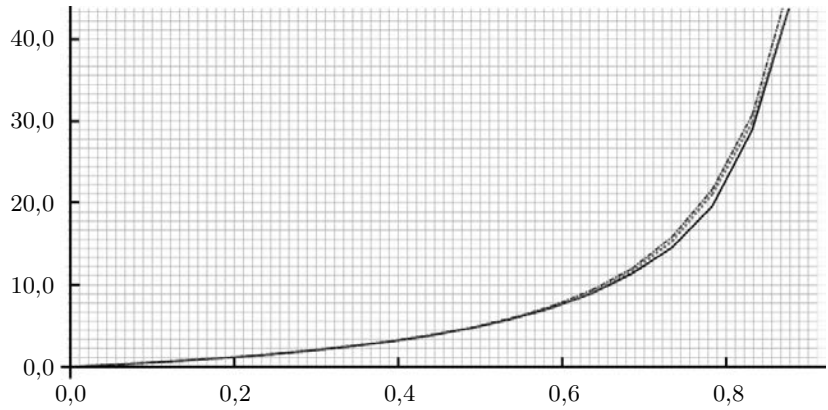


Рис. 2. Зависимости средних значений очередей для трех групповых пуассоновских потоков (сплошная линия – постоянное число заявок в пачках, точечная линия – пуассоновское распределение чисел заявок в пачках, пунктирная линия – экспоненциальное распределение)

Поток № 3 имеет пуассоновское распределение чисел заявок в пачках со средним значением $\overline{k_n} = \dfrac{1}{1{,}1}k = 9{,}1$ и коэффициентом вариации $v^2 = 0{,}1$.

В результате все три потока имеют одинаковые значения дисперсий распределений вероятностей чисел заявок в пачках: $D_m(\rho) = \bar{k}(1+v^2)\rho = k\rho$, что подтверждается совпадением графиков, показанных на рис. 1.

В соответствии с (3) указанные потоки должны иметь одинаковые зависимости средних значений очередей от коэффициента загрузки, что подтверждается совпадением графиков, представленных на рис. 2. Учитывая постоянное время обработки заявки, среднее время задержки в очереди равно произведению среднего размера очереди на указанное время обработки, следовательно, зависимости от коэффициента загрузки будут иметь аналогичный вид.

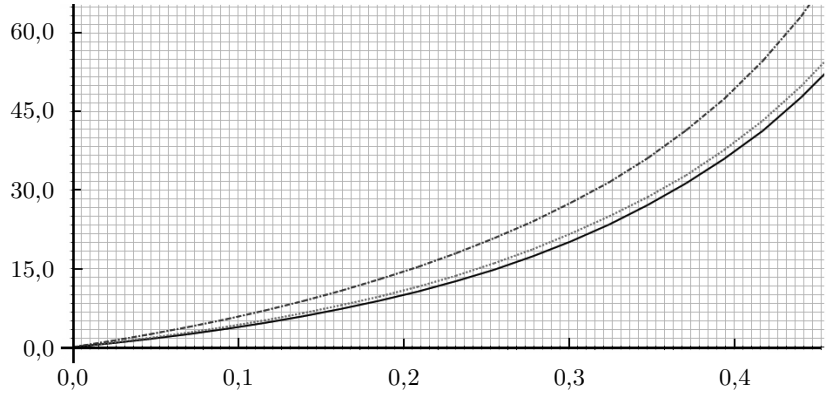


Рис. 3. Зависимости значений дисперсий очередей групповых пуассоновских потоков (сплошная линия – постоянное число заявок в пачках, точечная линия – пуассоновское распределение чисел заявок в пачках, пунктирная линия – экспоненциальное распределение)

Зависимости значений дисперсий очередей, представленные на рис. 3, различны даже при малых загрузках вследствие различия асимметрии распределений вероятностей чисел заявок в пачках рассматриваемых потоков.

## § 4. Заключение

В статье представлено аналитическое исследование статистических характеристик групповых пуассоновских потоков с проверкой результатов методом имитационного моделирования. С помощью интервального метода анализа очередей получены аналитические формулы для средней длины очереди, среднего квадрата очереди и дисперсии очереди в СМО с групповым пуассоновским потоком на входе. Показано, что дисперсия очереди зависит от третьего момента размера пачки заявок во входящем групповом пуассоновском потоке.

## СПИСОК ЛИТЕРАТУРЫ

1. *Вишневский В.М., Дудин А.Н.* Системы массового обслуживания с коррелированными входными потоками и их применение для моделирования телекоммуникационных сетей // Автомат. и телемех. 2017. № 8. С. 3–59. https://www.mathnet.ru/rus/at14562
2. *Neuts M.F.* A Versatile Markovian Point Process // J. Appl. Probab. 1979. V. 16. № 4. P. 764–779. https://doi.org/10.2307/3213143
3. *Wamser F., Gasas P., Seufert M., Moldovan C., Tran-Gia P., Hossfeld T.* Modeling the YouTube Stack: From Packets to Quality of Experience // Comput. Networks. 2016. V. 109. Part 2. P. 211–224. https://doi.org/10.1016/j.comnet.2016.03.020
4. *Appenzeller G., Keslassy I., McKeown N.* Sizing Router Buffers // Proc. 2004 Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM'04). Portland, USA. Aug. 30 – Sept. 3, 2004. P. 281–292. https://doi.org/10.1145/1015467.1015499
5. *Lee J.-B., Kalva H.* The VC-1 and H.264 Video Compression Standards for Broadband Video Services. New York: Springer, 2008. https://doi.org/10.1007/978-0-387-71043-3
6. *Li Z., Zhu X., Gahm J., Pan R., Hu H., Begen A.C., Oran D.* Probe and Adapt: Rate Adaptation for HTTP Video Streaming at Scale // IEEE J. Select. Areas Commun. 2014. V. 32. № 4. P. 719–733. https://doi.org/10.1109/JSAC.2014.140405

7. *Назаров А.А., Лопухова С.В.* Полумарковские процессы и специальные потоки однородных событий. Томск: ИДО ТГУ, 2010. Электронная библиотека (репозиторий) ТГУ: http://vital.lib.tsu.ru/vital/access/manager/Repository/vtls:000405029 (дата обращения 26.10.2022).
8. *Лихтциндер Б.Я.* Трафик мультисервисных сетей доступа (интервальный анализ и проектирование). М.: Науч.-тех. изд-во «Горячая линия – Телеком», 2019.
9. *Клейнрок Л.* Теория массового обслуживания. М.: Машиностроение, 1979.

*Лихтциндер Борис Яковлевич*
Поволжский государственный университет
телекоммуникаций и информатики, Самара
lixt@psuti.ru
*Привалов Александр Юрьевич*
Самарский национальный исследовательский университет
им. академика С.П. Королева
privalov1967@gmail.com
*Моисеев Виктор Игоревич*
Пермский государственный национальный
исследовательский университет
vim@psu.ru

Поступила в редакцию
18.12.2022
После доработки
27.02.2023
Принята к публикации
27.02.2023

Адрес редакции: 127051, Москва, Б. Каретный пер., 19, стр. 1, тел. (495) 650-47-39

Оригинал-макет подготовил *Д.Ю. Ногин*
по контракту с ООО «Объединённая редакция»

**Москва**
**ООО «Объединённая редакция»**