Границы минимального кодового расстояния для недвоичных кодов на двудольных графах

А.А. Фролов, В.В. Зяблов
Институт проблем передачи информации им. А.А. Харкевича РАН
{alexey.frolov, zyablov}@iitp.ru

Аннотация

Исследуется минимальное кодовое расстояние кодов на двудольных графах (ДГ-кодов) над полем $GF(q)$. Получена новая верхняя граница минимального кодового расстояния для ДГ-кодов. Показано, что эта граница лежит ниже границы Варшамова–Гилберта (ВГ) при $q \geqslant 32$. Поскольку коды на базе двудольных графов-расширителей (ДГР-коды) являются частным случаем ДГ-кодов, а полученная граница справедлива для любого ДГ-кода, то она также справедлива и для ДГР-кодов. Таким образом, недвоичные ($q \geqslant 32$) ДГ-коды хуже лучших из известных линейных кодов. Этот результат является ключевым результатом работы. Также получены нижняя граница минимального кодового расстояния для ДГ-кодов с кодом-компонентом Рида–Соломона и нижняя граница кодового расстояния для кодов с малой плотностью проверок (МПП-кодов) с кодом-компонентом Рида–Соломона. Нижняя граница для МПП-кодов близка к границе ВГ и лежит выше верхней границы минимального кодового расстояния для ДГ-кодов.

1 Введение

В этой работе рассматриваются недвоичные коды на двудольных графах. Идея кодов на графах была предложена Таннером в [1]. Позже Сипсер и Спилман в [2] использовали графы-расширители, чтобы получить асимптотически хорошие коды с простым декодированием (под “асимптотически хорошими” мы понимаем коды, чьи скорость и относительное кодовое расстояние одновременно отстоят от нуля). Они назвали эти коды кодами на базе графов-расширителей[1]. Наряду со случайными графами-расширителями в работе [2] использовались и явные конструкции графов с хорошим коэффициентом расширения [3, 4], называемые графами Рамануджана. В работах [5,6] рассматривается специальный случай конструкции Сипсера-Спилмана, в которой граф Рамануджана является двудольным. Эта конструкция является частным случаем ДГ-кодов. В этой работе мы исследуем минимальное кодовое расстояние ДГ-кодов.

[1] в англоязычной литературе используется термин “expander codes”

Нижние оценки кодового расстояния для двоичных ДГ-кодов получены в [7, 8]. Кроме того можно воспользоваться результатами работ [9, 10], где получены нижние оценки для обобщенных двоичных МПП-кодов. Все эти результаты можно легко обобщить на случай недвоичных ДГ-кодов. Однако нам не удалось найти ни одной работы, где была бы получена верхняя оценка кодового расстояния для ДГ-кодов. В работе [11] получены верхние оценки кодового расстояния для двоичных МПП-кодов. Однако авторам удалось показать, что построенная ими верхняя оценка лежит ниже границы ВГ только при очень высоких скоростях ($R > 0.975$), на оставшейся части интервала эта оценка лишь улучшает общие верхние оценки, справедливые для всех линейных кодов. Для построения верхней оценки мы воспользуемся одним из методов, описанных в работе [11].

В этой работе получены верхняя и нижняя оценка кодового расстояния для ДГ-кодов над полем $GF(q)$. Показано, что при любом $q \geqslant 32$ имеется расширяющийся с ростом q интервал (не на высоких скоростях), где полученная верхняя оценка лежит ниже границы Варшамова–Гильберта, и тем самым доказан фундаментальный результат, заключающийся в том, что недвоичные ДГ-коды хуже лучших из известных линейных кодов. При этом показано, что нижняя граница кодового расстояния для МПП-кодов с кодом-компонентом Рида–Соломона очень близка к границе ВГ и лежит выше верхней границы для ДГ-кодов.

Казалось бы, в этой работе получен результат, противоречащий результату работы [12], где приводится конструкция кодов (которые также названы ДГР-кодами), минимальное кодовое расстояние которых с ростом q приближается к границе Синглтона, однако конструкции кодов все-таки различны, а кажущееся противоречие – результат некоторой путаницы в терминологии.

В §2 мы опишем структуру ДГ-кодов. В §3 будет предложена и доказана новая верхняя оценка минимального кодового расстояния для ДГ-кодов. В §4 получены нижние оценки. В §5 приведены полученные численные результаты.

2 Структура кода

Пусть $G = (V_1 : V_2, E)$ – это ненаправленный связный двудольный граф с множеством вершин $V = V_1 \bigcup V_2$ ($V_1 \bigcap V_2 = \emptyset$) и множеством ребер E. Пусть $\deg(u_i) = \Delta_1 \, \forall u_i \in V_1$, $\deg(v_j) = \Delta_2 \, \forall v_j \in V_2$, $|E| = n$, тогда $|V_1| = b_1$, $|V_2| = b_2$, где $b_1 = \frac{n}{\Delta_1}$, $b_2 = \frac{n}{\Delta_2}$.

Пусть $\mathbb{F}_q$ – это поле Галуа с q элементами. Поставим в соответствие каждой вершине $u_i \in V_1$, $i = \overline{1, b_1}$ линейный $(\Delta_1, R_1\Delta_1)$ код $C_i^{(1)}$ над полем $\mathbb{F}_q$; каждой вершине $v_j \in V_2$, $j = \overline{1, b_2}$ – линейный $(\Delta_2, R_2\Delta_2)$ код $C_j^{(2)}$ над тем же полем. Далее коды $C_j^{(i)}$ будем называть компонентными кодами.

Для каждой вершины $u \in V$ обозначим через $E(u)$ множество ребер, ицидентных ей. Мы предполагаем, что ребра из E некоторым образом занумерованы и этот порядок фиксирован. Для любого $u \in V$ порядок на $E(u)$ индуцируется порядком на E. Пусть $\mathbf{z} = (z_e)_{e \in E}$, тогда через $(\mathbf{z})_{E(u)}$ обозначим подблок $\mathbf{z}$, в который входят элементы с индексами из $E(u)$.

Теперь мы готовы дать определение ДГ-кода:

Определение 1 . *Код C является ДГ-кодом, если*

$$C = \left\{ \mathbf{c} \in \mathbb{F}_q^{|E|} : \left((\mathbf{c})_{E(u_i)} \in C_i^{(1)} \ \forall u_i \in V_1 \right) \wedge \left((\mathbf{c})_{E(v_j)} \in C_j^{(2)} \ \forall v_j \in V_2 \right) \right\}$$

ДГ-код C является линейным кодом, следовательно, его можно задать проверочной матрицей. Пусть $\mathbf{H}_i^{(1)}$ – это проверочная матрица кода-компонента $C_i^{(1)}$, $\mathbf{H}_j^{(2)}$ – проверочная матрица кода-компонента $C_j^{(2)}$, тогда проверочная матрица $\mathbf{H}$ кода C выглядит следующим образом:

$$\mathbf{H} = \begin{pmatrix} \pi_1 \left(diag \left(\mathbf{H}_1^{(1)}, \mathbf{H}_2^{(1)}, \ldots, \mathbf{H}_{b_1}^{(1)} \right) \right) \\ \pi_2 \left(diag \left(\mathbf{H}_1^{(2)}, \mathbf{H}_2^{(2)}, \ldots, \mathbf{H}_{b_2}^{(2)} \right) \right) \end{pmatrix}, \tag{1}$$

где

$$diag \left(\mathbf{H}_1^{(i)}, \mathbf{H}_2^{(i)}, \ldots, \mathbf{H}_{b_i}^{(i)} \right) = \begin{pmatrix} \mathbf{H}_1^{(i)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2^{(i)} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H}_{b_i}^{(i)} \end{pmatrix}_{(1-R_i)n \times n},$$

π_i – перестановка столбцов матрицы $diag \left(\mathbf{H}_1^{(i)}, \mathbf{H}_2^{(i)}, \ldots, \mathbf{H}_{b_i}^{(i)} \right)$ однозначно определяемая графом G и фиксированным порядком ребер на E.

Замечание 1 . *Размер* $\mathbf{H}$ – $((1 - R_1) + (1 - R_2))\, n \times n$.

Теперь определим параметры полученного кода. Длина n кода C равна $|E|$, скорость кода C удовлетворяет неравенству

$$R \geqslant R_1 + R_2 - 1 \tag{2}$$

Равенство достигается в случае полного ранга матрицы $\mathbf{H}$.

3 Верхняя граница минимального кодового расстояния для ДГ-кодов

Пусть C' – это ДГ-код. Пусть без ограничения общности $R_1 \leqslant R_2$. Проверочная матрица (1) кода C' может быть преобразована к виду:

$$\mathbf{H} = \begin{pmatrix} diag \left(\mathbf{H}_1^{(1)}, \mathbf{H}_2^{(1)}, \ldots, \mathbf{H}_{b_1}^{(1)} \right) \\ \pi_1^{-1} \pi_2 \left(diag \left(\mathbf{H}_1^{(2)}, \mathbf{H}_2^{(2)}, \ldots, \mathbf{H}_{b_2}^{(2)} \right) \right) \end{pmatrix}.$$

Пусть код C соответствует проверочной матрице $\mathbf{H}$. Коды C и C' эквивалентны, следовательно, они имеют одинаковые минимальные кодовые расстояния. Теперь пользуясь структурой первого слоя докажем следующую теорему:

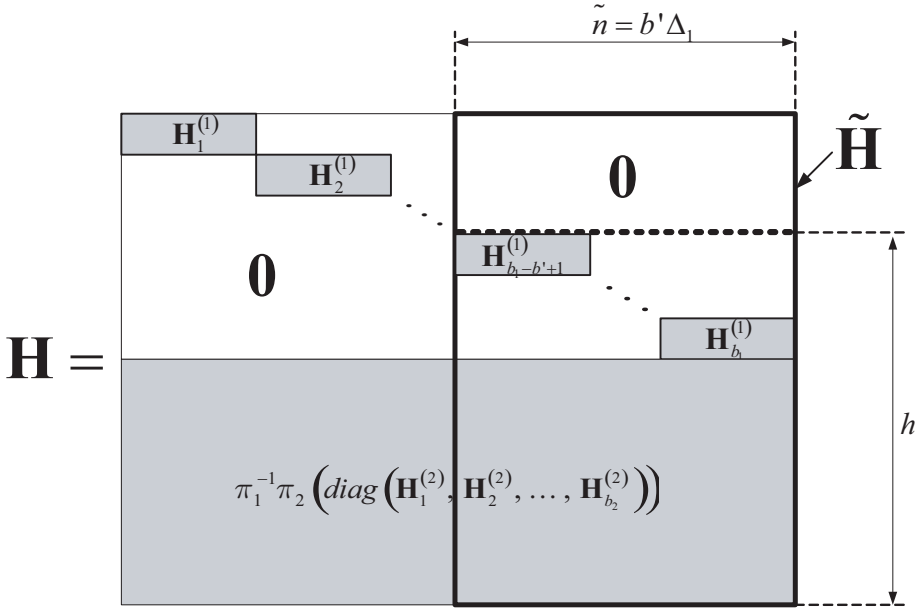

Рис. 1: Проверочная матрица кода $\widetilde{C}$

Теорема 1 . *Пусть C – это ДГ-код, тогда*

$$d(C) \leqslant \min_{b_1 \geqslant b' \geqslant \frac{(R_1-R)}{R_1} b_1 + \frac{1}{R_1\Delta_1}} \left\{ \frac{q^{\widetilde{k}-1}(q-1)}{q^{\widetilde{k}}-1} b'\Delta_1 \right\},$$

где $\widetilde{k} = b'R_1\Delta_1 - (R_1 - R)\, n$, $b' \in \mathbb{N}$.

Доказательство. Рассмотрим код $\widetilde{C}$ длины $\widetilde{n} = b'\Delta_1$, $b' \in \mathbb{N}$. Проверочная матрица $\widetilde{\mathbf{H}}$ этого кода показанана на Рис. 1. Этот код соответствует подкоду C'' кода C. Действительно, нужно лишь добавить префикс из $n - \widetilde{n}$ нулей к слову $\widetilde{\mathbf{c}}$ кода $\widetilde{C}$, чтобы получить слово $\mathbf{c}''$ кода C'', т.е.

$$\mathbf{c}'' = (\mathbf{0}\, \widetilde{\mathbf{c}})\,.$$

Таким образом,

$$d(C) \leqslant d(C'') = d\left(\widetilde{C}\right).$$

Ясно, что избыточность кода $\widetilde{C}$ удовлетворяет следующему условию (через $\widetilde{k}$ обозначим размерность кода $\widetilde{C}$):

$$\widetilde{n} - \widetilde{k} \leqslant (R_1 - R)\, n + b'\,(1 - R_1)\,\Delta_1,$$

следовательно,

$$\widetilde{k} \geqslant b'R_1\Delta_1 - (R_1 - R)\, n.$$

Для того чтобы выполнялось условие $\widetilde{k} \geqslant 1$ достаточно выполнения условия:

$$b'R_1\Delta_1 - (R_1 - R)\, n \geqslant 1.$$

Отсюда получим условие на b':

$$b' \geqslant \frac{R_1 - R}{R_1} b_1 + \frac{1}{R_1 \Delta_1}.$$

Применив границу Плоткина [13] получим необходимый результат. ▲

Замечание 2 . *Вместо границы Плоткина можно применить любую из известных границ для q-ичных линейных кодов (например, границу Бассалыго-Элайеса [14] или границу Мак-Элиса–Родемича–Рамсея–Велча [15]), но для наших целей достаточно уже и границы Плоткина.*

В следующей теореме будет приведена асимптотическая форма верхней границы минимального кодового расстояния:

Теорема 2 . *Пусть* $\{C_i\}_{i=1}^{\infty}$ *– это последовательность ДГ-кодов со скоростями* $R(C_i) = R$ *и длинами* $n(C_i) = i \times HOK(\Delta_1, \Delta_2)$, *тогда*

$$\delta = \lim_{i \to \infty} \left(\frac{d(C_i)}{n(C_i)} \right) \leqslant \frac{q-1}{q} \left(\frac{1-R}{1+R} \right)$$

Доказательство. Выберем

$$b' = \left\lceil b_1 \left(\frac{R_1 - R}{R_1} \right) \right\rceil + f(n),$$

где $f(n) \to \infty$ при $n \to \infty$ и $f(n) = o(n)$, тогда

$$d(C) \leqslant \frac{q^{R_1 \Delta_1 f(n) - 1}(q-1)}{q^{R_1 \Delta_1 f(n)} - 1} \left(\left(\frac{R_1 - R}{R_1} \right) n + (f(n) + 1) \Delta_1 \right)$$

Разделив обе части неравенства на n и перейдя к пределу при $i \to \infty$, получим

$$\delta \leqslant \frac{q-1}{q} \left(\frac{R_1 - R}{R_1} \right). \tag{3}$$

Воспользовавшись соотношениями $R_1 \leqslant R_2$ и (2), получим, что

$$R_1 \leqslant \frac{R_1 + R_2}{2} \leqslant \frac{1+R}{2}. \tag{4}$$

Подставив (4) в (3) получим необходимый результат

$$\delta \leqslant \frac{q-1}{q} \left(\frac{1-R}{1+R} \right).$$

▲

Замечание 3 . *Отметим, что худший случай достигается при* $R_1 = R_2$. *Это следует из неравенства (4). В случае* $R_1 < R_2$ *верхняя граница пройдет еще ниже.*

4 Нижние границы минимального кодового расстояния

В этом разделе мы получим нижние границы минимального кодового расстояния для двух ансамблей кодов. Результаты этого раздела являются

модификацированными версиями результатов работ [7–10, 16]. Так как отличия (хоть и небольшие) все-таки присутствуют мы решили полностью привести все необходимые доказательства.

Отметим, что используемый в данной работе метод требует знания полных распределений весов кодовых слов компонентных кодов. Одновременно с этим, статьи [7] (для двоичных кодов на двудольных графах) и [17] (для двоичных обобщенных МПП-кодов или, как их еще называют, кодов на гиперграфах) предлагают другой способ оценки минимального кодового расстояния, для которого нужно лишь кодовое расстояние компонентных кодов. Результаты, получаемые при таком подходе, слабее, так как используется меньше информации о компонентных кодах, но в некоторых ситуациях это может быть полезно.

Прежде чем перейти к конкретным ансамблям кодов, введем необходимые обозначение и докажем утверждения, справедливые для любого ансамбля кодов.

Пусть $\mathscr{E}$ – это ансамбль кодов, имеющих длину n. Через $\overline{A(W)}$ обозначим среднее по кодам число кодовых слов веса W, т.е.

$$\overline{A(W)} = \frac{1}{|\mathscr{E}|} \sum_{i=1}^{|\mathscr{E}|} A_i(W),$$

где $A_i(W)$ – это число слов веса W в коде $C_i \in \mathscr{E}$.

Теорема 3 (Галлагер, [16]). *Если выполняется условие*

$$\sum_{W=1}^{d} \overline{A(W)} < 1,$$

то в ансамбле $\mathscr{E}$ существует код C, такой что $d(C) > d$.

Доказательство. Заметим, что $\sum\limits_{W=1}^{d} \overline{A(W)} < 1 \Rightarrow \sum\limits_{W=1}^{d} \sum\limits_{i=1}^{|\mathscr{E}|} A_i(W) < |\mathscr{E}|$, а это значит, что суммарное число кодовых слов веса меньше либо равного W в ансамбле $\mathscr{E}$ меньше числа кодов, следовательно, найдется код $C \in \mathscr{E}$, ни одного из этих слов не содержащий:

$$d(C) > d$$

▲

Замечание 4 . *Отметим, что если $\sum\limits_{W=1}^{d} \overline{A(W)} = \beta < 1$, то в ансамбле $\mathscr{E}$ найдется по крайней мере $(1-\beta)|\mathscr{E}|$ кодов C_i, таких что $d(C_i) > d$.*

Замечание 5 . *Отметим, что*

$$\overline{A(W)} = \frac{1}{|\mathscr{E}|} \sum_{i=1}^{|\mathscr{E}|} A_i(W) = \frac{1}{|\mathscr{E}|} \sum_{j=1}^{|V_W|} N\left(\mathscr{E}, \mathbf{v}_j^{(W)}\right), \tag{5}$$

где $V_W = \left\{\mathbf{v}^{(\mathbf{W})} \in \mathbb{F}_q^n : \left|\mathbf{v}^{(W)}\right| = W\right\}$, $N(\mathscr{E}, \mathbf{v})$ – это число кодов из ансамбля $\mathscr{E}$, содержащих кодовое слово $\mathbf{v}$.

Теперь перейдем к конкретным ансамблям.

4.1 Ансамбль ДГ-кодов

Рассмотрим блочную диагональную матрицу

$$\mathbf{H}_b = \begin{pmatrix} \mathbf{H}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_0 & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H}_0 \end{pmatrix}_{(1-R_0)n \times n},$$

на главной диагонали которой находятся b проверочных матриц $\mathbf{H}_0$ линейного $(\Delta_0, R_0\Delta_0)$ кода над полем $\mathbb{F}_q$. Размер $\mathbf{H}_b - (1-R_0)\, n \times n$, где $n = \Delta_0 b$. Через $\varphi(\mathbf{H}_b)$ обозначим матрицу, полученную из матрицы $\mathbf{H}_b$ произвольной перестановкой столбцов и умножением их на произвольные ненулевые элементы поля $\mathbb{F}_q$. Тогда матрица

$$\mathbf{H} = \begin{pmatrix} \varphi_1(\mathbf{H}_b) \\ \varphi_2(\mathbf{H}_b) \end{pmatrix}_{2(1-R_0)n \times n},$$

составленная из двух таких матриц как слоев, является разреженной проверочной матрицей кода из ансамбля $\mathscr{E}_1(\Delta_0, b)$.

Определим ансамбль $\mathscr{E}_1(\Delta_0, b)$ так:

Определение 2 . *Элементы ансамбля $\mathscr{E}_1(\Delta_0, b)$ получаются путем независимого выбора перестановок π_i и ненулевых констант $c_{i,j}, i = 1, 2; j = 1, 2, \ldots, n$, на которые умножаются столбцы полученных в результате перестановок матриц слоев.*

Замечание 6 . *Отметим, что в отличие от определения ансамбля для двоичных кодов здесь добавляется умножение на константы, не равные нулю.*

Замечание 7 . *Каждый код из ансамбля $\mathscr{E}_1(\Delta_0, b)$ является ДГ-кодом ($\Delta_1 = \Delta_2 = \Delta_0$, $R_1 = R_2 = R_0$), поэтому для каждого из них справедлива верхняя граница минимального кодового расстояния из §3, $|\mathscr{E}_1(\Delta_0, b)| = (n!(q-1)^n)^2$.*

Лемма 1 . *Среднее по кодам число кодовых слов веса W в ансамбле $\mathscr{E}_1(\Delta_0, b)$*

$$\overline{A(W)} = \frac{(A_1(W))^2}{(q-1)^W \begin{pmatrix} n \\ W \end{pmatrix}},$$

где $A_1(W)$ – число кодовых слов веса W в первом слое.

Доказательство.

Рассмотрим фиксированный вектор $\mathbf{v}^{(W)}$ длины n, $\left\|\mathbf{v}^{(W)}\right\| = W$. В соответствии с уравнением (5) нужно вычислить $N\left(\mathscr{E}_1(\Delta_0, b), \mathbf{v}^{(W)}\right)$. Рассмотрим отдельно ансамбли первых (L_1) и вторых (L_2) слоев. Если мы знаем число первых слоев, для которых комбинация $\mathbf{v}^{(W)}$ является кодовой ($N\left(L_1, \mathbf{v}^{(W)}\right)$) и число вторых слоев, для которых комбинация $\mathbf{v}^{(W)}$ является кодовой $N\left(L_2, \mathbf{v}^{(W)}\right)$, то

$$N\left(\mathscr{E}_1(\Delta_0, b), \mathbf{v}^{(W)}\right) = N\left(L_1, \mathbf{v}^{(W)}\right) N\left(L_2, \mathbf{v}^{(W)}\right),$$

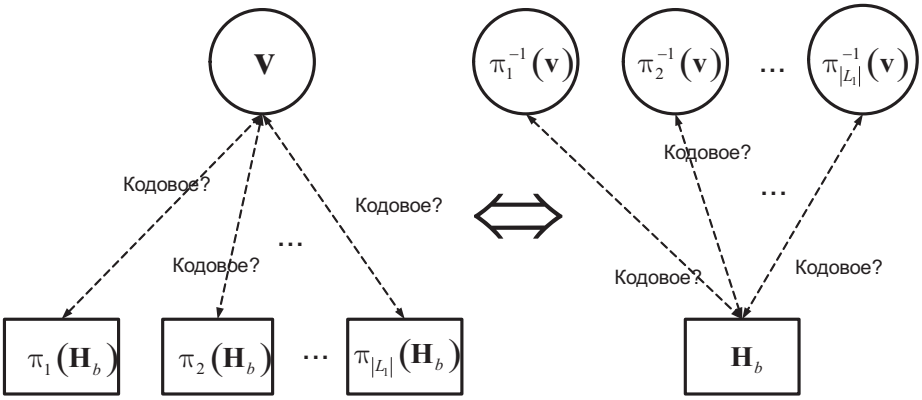

Рис. 2: Подсчет $N\left(L_1, \mathbf{v}^{(W)}\right)$

это следует из того, что перестановки и ненулевые множители для слоев выбираются произвольно и независимо. По этой же причине $L_1 = L_2$, следовательно,

$$N\left(\mathscr{E}_1\left(\Delta_0, b\right), \mathbf{v}^{(W)}\right) = \left(N\left(L_1, \mathbf{v}^{(W)}\right)\right)^2.$$

Для того чтобы вычислить $N\left(L_1, \mathbf{v}^{(W)}\right)$ зафиксируем один конкретный слой, а перестановки и умножения на константы будем производить с элементами вектора $\mathbf{v}^{(W)}$, а не со столбцами проверочной матрицы. Так как все слои являются эквивалентными, пусть это будет слой с тождественным отображением $\varphi = id$. Все сказанное проиллюстрировано на Рис. 2.

В соответствии со свойствами отображений φ_i среди векторов $\left\{{\varphi_i}^{-1}\left(\mathbf{v}^{(W)}\right)\right\}_{i=1}^{|L_1|}$ есть все возможные вектора, причем каждый из них повторяется K раз, где

$$K = W!\,(n-W)!(q-1)^{n-W}.$$

Таким образом, мы приходим к следующему соотношению:

$$N\left(L_1, \mathbf{v}^{(W)}\right) = A_1\left(W\right)K = A_1\left(W\right)W!\,(n-W)!(q-1)^{n-W}.$$

И окончательно,

$$N\left(\mathscr{E}_1\left(\Delta_0, b\right), \mathbf{v}^{(W)}\right) = \left(A_1\left(W\right)W!\,(n-W)!(q-1)^{n-W}\right)^2.$$

Как показывает предыдущее рассуждение $N\left(\mathscr{E}_1\left(\Delta_0, b\right), \mathbf{v}^{(W)}\right)$ одинаково для всех векторов веса W, поэтому в соответствии с (5) получим:

$$\begin{aligned}\overline{A\left(W\right)} &= (q-1)^W \binom{n}{W} \frac{\left(A_1\left(W\right)W!\,(n-W)!(q-1)^{n-W}\right)^2}{\left((q-1)^n n!\right)^2} \\ &= \frac{\left(A_1\left(W\right)\right)^2}{(q-1)^W \binom{n}{W}},\end{aligned}$$

что и завершает доказательство леммы.

▲

В следующей лемме мы получим оценку $\overline{A(W)}$.

Лемма 2 . *Среднее по кодам число кодовых слов веса W в ансамбле $\mathscr{E}_1(\Delta_0, b)$*

$$\overline{A(W)} \leqslant q^{-nF_1(\delta,\Delta_0)},$$

где

$$F_1(\delta,\Delta_0) = h_q(\delta) + \delta\log_q(q-1) + 2\max_{s>0}\left(\delta\log_q(s) - \frac{1}{\Delta_0}\log_q(g_0(s,\Delta_0))\right),$$

$\delta = \frac{W}{n}$, $h_q(\delta) = -\delta\log_q(\delta) - (1-\delta)\log_q(1-\delta)$ – *функция q-ичной энтропии, а* $g_0(s,\Delta_0)$ – *это производящая функция весов кодовых слов кода-компонента.*

Доказательство. Заметим, что в каждом слое множества позиций, занятых кодовыми символами компонентных кодов, не пересекаются. В то же время все позиции покрыты, следовательно, производящая функция слоя $G(s)$ может быть представлена так:

$$G(s) = g_0{}^{\frac{n}{\Delta_0}}(s,\Delta_0),$$

тогда

$$A_1(W) = \left[s^W\right]\left(g_0{}^{\frac{n}{\Delta_0}}(s,\Delta_0)\right)$$

Воспользовавшись очевидной оценкой,

$$A_1(W) \leqslant \min_{s>0}\left(\frac{g_0{}^{\frac{n}{\Delta_0}}(s,\Delta_0)}{s^W}\right),$$

неравенством $\binom{n}{W} \leqslant q^{nh_q(\delta)}$ и леммой 1 получим необходимый результат.

▲

Теорема 4 . *Если существует хотя бы один положительный корень (относительно переменной δ) уравнения*

$$F_1(\delta,\Delta_0) = 0 \tag{6}$$

тогда в ансамбле $\mathscr{E}_1(\Delta_0, b)$ существуют коды $\{C_i\}_{i=1}^{N_1(b)}$ $\left(\lim\limits_{b\to\infty}\frac{N_1(b)}{|\mathscr{E}_1(\Delta_0,b)|} = 1\right)$, такие что $d(C_i) \geqslant (\delta_1 - \varepsilon)n$, где ε – сколь угодно малое положительное число; δ_1 – положительный корень уравнения (6).

Доказательство. В соответствии с леммой 4

$$\lim_{n\to\infty}\left(\sum_{W=1}^{\lfloor(\delta_1-\varepsilon)n\rfloor}\overline{A(W)}\right) = 0.$$

Из замечания 4 следует, что

$$|\mathscr{E}_1(\Delta_0,b)|\left(1 - \sum_{W=1}^{\lfloor(\delta_1-\varepsilon)n\rfloor}\overline{A(W)}\right) \leqslant N_1(b) \leqslant |\mathscr{E}_1(\Delta_0,b)|,$$

что и заканчивает доказательство теоремы. ▲

4.2 Ансамбль МПП-кодов

Рассмотрим матрицу

$$\mathbf{H} = \begin{pmatrix} \varphi_1(\mathbf{H}_b) \\ \varphi_2(\mathbf{H}_b) \\ \vdots \\ \varphi_\ell(\mathbf{H}_b) \end{pmatrix}_{\ell(1-R_0)n \times n}$$

состоящую из ℓ слоев (обозначение $\varphi(\mathbf{H}_b)$ было введено в §4.1). Эта разреженная матрица является проверочной матрицей кода из ансамбля $\mathscr{E}_2(\Delta_0, b)$.

Определение 3 . *Элементы ансамбля $\mathscr{E}_2(\Delta_0, b)$ получаются путем независимого выбора перестановок π_i и ненулевых констант $c_{i,j}, i = 1, 2, \ldots, \ell$; $j = 1, 2, \ldots, n$, на которые умножаются столбцы полученных в результате перестановок матриц слоев.*

Замечание 8 . *Здесь в отличие от ансамбля $\mathscr{E}_1(\Delta_0, b)$ проверочная матрица состоит не из двух, а из ℓ слоев.*

Замечание 9 . *Эти коды не являются ДГ-кодами и, следовательно, верхняя оценка для них не справедлива, мы приводим этот ансамбль для сравнения.*

Все доказательства в этом разделе аналогичны доказательствам из §4.1, мы приведем только основной результат.

Теорема 5 . *Если существует хотя бы один положительный корень (относительно переменной δ) уравнения*

$$F_2(\delta, \Delta_0) = 0 \tag{7}$$

тогда в ансамбле $\mathscr{E}_2(\Delta_0, b)$ существуют коды $\{C_i\}_{i=1}^{N_2(b)}$ $\left(\lim\limits_{b\to\infty} \frac{N_2(b)}{|\mathscr{E}_2(\Delta_0,b)|} = 1\right)$, такие что $d(C_i) \geqslant (\delta_2 - \varepsilon) n$, где ε – сколь угодно малая положительная величина; δ_2 – положительный корень уравнения (7),

$$F_2(\delta, \Delta_0) = (\ell - 1)\left(h_q(\delta) + \delta \log_q(q-1)\right) + \ell \max_{s>0}\left(\delta \log_q(s) - \frac{1}{\Delta_0}\log_q\left(g_0(s, \Delta_0)\right)\right).$$

5 Численные результаты

В качестве компонентного кода будем использовать укороченный код Рида–Соломона длины $\Delta_0 \leqslant q$. Опишем, как его построить. Пусть α – это примитивный элемент поля $\mathbb{F}_q$. Рассмотрим проверочную матрицу удлиненного $(q, q - d_0 + 1)$ кода Рида–Соломона

$$\mathbf{H} = \begin{pmatrix} 1 & \alpha & \alpha^2 & \ldots & \alpha^{(q-2)} & 1 \\ 1 & \alpha^2 & \alpha^4 & \ldots & \alpha^{2(q-2)} & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \alpha^{d_0-1} & \alpha^{2(d_0-1)} & \ldots & \alpha^{(d_0-1)(q-2)} & 0 \end{pmatrix},$$

где d_0 – это минимальное кодовое расстояние. Тогда $(\Delta_0, \Delta_0 - d_0 + 1)$ код Рида–Соломона получается укорочением на $q - \Delta_0$ информационных символов, т.е. проверочная матрица этого кода состоит из Δ_0 столбцов матрицы $\mathbf{H}$. Построенный код также является кодом с максимальным достижимым расстоянием (МДР-кодом), т.е. $d_0 = (1 - R_0)\,\Delta_0 + 1$.

Известно, что число кодовых слов веса W в МДР-коде можно оценить следующим образом:

$$a\,(W) \leqslant \left(\begin{array}{c} \Delta_0 \\ W \end{array} \right) (q-1)^{W-d_0+1}$$

Таким образом, при $s > 0$ для производящей функции весов кодовых слов компонентного кода справедлива следующая оценка:

$$g_0\,(s, \Delta_0) \leqslant 1 + \sum_{i=d_0}^{\Delta_0} \left(\left(\begin{array}{c} \Delta_0 \\ i \end{array} \right) (q-1)^{i-d_0+1} s^i \right).$$

Для ансамбля ДГ-кодов при заданных R и Δ_0 получим

$$d_0 = (1 - R_0)\,\Delta_0 + 1 = \frac{1-R}{2}\Delta_0 + 1.$$

Для ансамбля МПП-кодов зафиксируем $d_0 = 2$, таким образом при заданных R и Δ_0 получим

$$\ell = (1 - R)\,\Delta_0.$$

В обоих случаях выберем Δ_0 таким образом, чтобы максимизировать относительное кодовое расстояние.

На Рис. 3, Рис. 4, Рис. 5 приведены полученные результаты для $q = 64$, $q = 256$, $q = 1024$ соответственно. Численные данные даются в Табл. 1, Табл. 2 и Табл. 3. На каждом из рисунков показаны четыре зависимости:

- Зависимость $\delta_{VG}\,(R)$ – граница Варшамова–Гилберта.
- Зависимость $\delta_{upper}\,(R)$ – верхняя граница минимального кодового расстояния для ДГ-кодов.
- Зависимость $\delta_{lower}\,(R)$ – нижняя граница минимального кодового расстояния для ДГ-кодов с компонентным кодом Рида–Соломона (Δ_0 подбирается таким образом, чтобы получить максимальное значение; Получившиеся значения Δ_0 приведены в таблицах).
- Зависимость $\delta_{LDPC}\,(R)$ – нижняя граница минимального кодового расстояния для МПП-кодов с компонентным кодом Рида–Соломона (Δ_0 подбирается таким образом, чтобы получить максимальное значение; Получившиеся значения Δ_0 и ℓ приведены в таблицах).

В случае $q = 64$ верхняя оценка минимального кодового расстояния для ДГ-кодов лежит ниже границы ВГ при $R \in (0,25; 0,89)$. Этот интервал расширяется с увеличением q, для $q = 256$ интервал такой – $(0,10; 0,97)$, а для $q = 1024$ – $(0,05; 0,99)$.

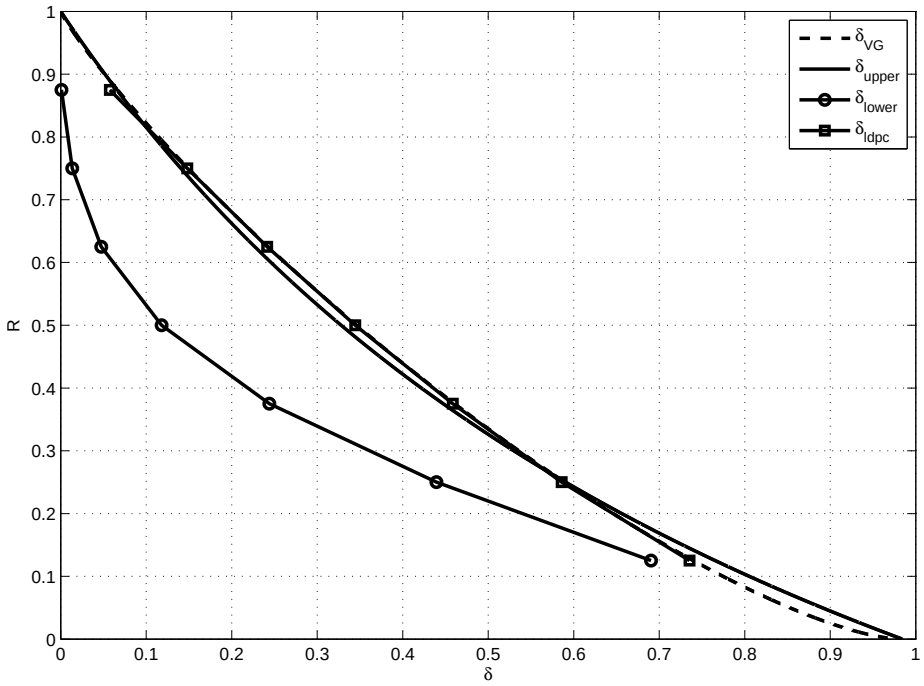

Рис. 3: Результаты для $q = 64$

Таблица 1: Результаты для $q = 64$

R	$\delta_{VG}(R)$	$\delta_{upper}(R)$	$\delta_{lower}(R);\Delta_0$	$\delta_{LDPC}(R);\Delta_0;\ell$
1/8	0,7400	0,7656	0,6905; 64	0,7355; 16; 14
1/4	0,5894	0,5906	0,4395; 64	0,5860; 12; 9
3/8	0,4608	**0,4474**	0,2440; 64	0,4585; 24; 15
1/2	0,3462	**0,3281**	0,1180; 64	0,3445; 28; 14
5/8	0,2427	**0,2272**	0,0475; 64	0,2415; 40; 15
3/4	0,1492	**0,1406**	0,0135; 64	0,1480; 52; 13
7/8	0,0665	**0,0656**	0,0010; 64	0,0575; 64; 8

6 Заключение

При $q \geqslant 32$ существует интервал, в котором полученная верхняя оценка лежит ниже границы Варшамова–Гилберта, следовательно, недвоичные ДГ-коды хуже, чем лучшие из известных недвоичных кодов. Полученная нижняя граница для ансамбля ДГ-кодов с компонентным кодом Рида–Соломона сильно хуже верхней. Полученная нижняя граница для ансамбля МПП-кодов с компонентным кодом Рида–Соломона очень близка к границе Варшамова–Гилберта и лежит выше верхней границы для ДГ-кодов.

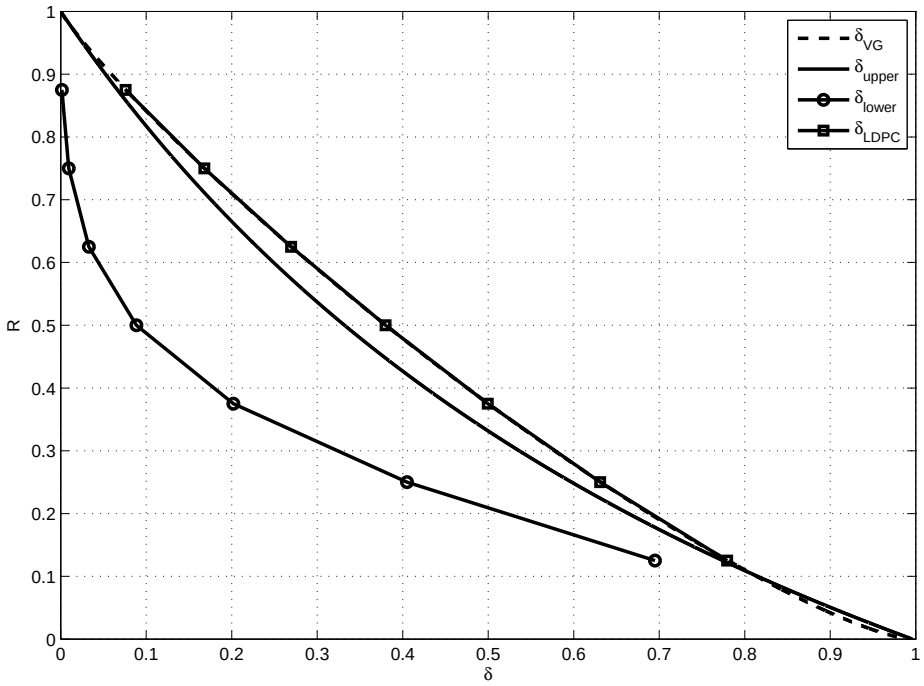

Рис. 4: Результаты для $q = 256$

Таблица 2: Результаты для $q = 256$

R	$\delta_{VG}(R)$	$\delta_{upper}(R)$	$\delta_{lower}(R);\Delta_0$	$\delta_{LDPC}(R);\Delta_0;\ell$
1/8	0,7807	**0,7747**	0,6950; 240	0,7795; 8; 7
1/4	0,6318	**0,5977**	0,4050; 248	0,6310; 16; 12
3/8	0,5004	**0,4528**	0,2020; 240	0,4995; 16; 10
1/2	0,3805	**0,3320**	0,0885; 256	0,3800; 28; 14
5/8	0,2700	**0,2299**	0,0330; 224	0,2695; 40; 15
3/4	0,1684	**0,1423**	0,0095; 192	0,1680; 60; 15
7/8	0,0764	**0,0664**	0,0015; 240	0,0760; 136; 17

ПРИЛОЖЕНИЕ

В этот приложении мы исследуем свойства функции

$$G(\delta) = F_2(\delta, \Delta_0) = (\ell - 1)\left(h_q(\delta) + \delta \log_q(q-1)\right) + \ell \max_{s>0}\left(\delta \log_q(s) - \frac{1}{\Delta_0}\log_q\left(g_0(s, \Delta_0)\right)\right).$$

Пусть функция $g_0(s, \Delta_0) = 1 + \sum_{i=d_0}^{\Delta_0} a(i)s^i$ удовлетворяет следующим

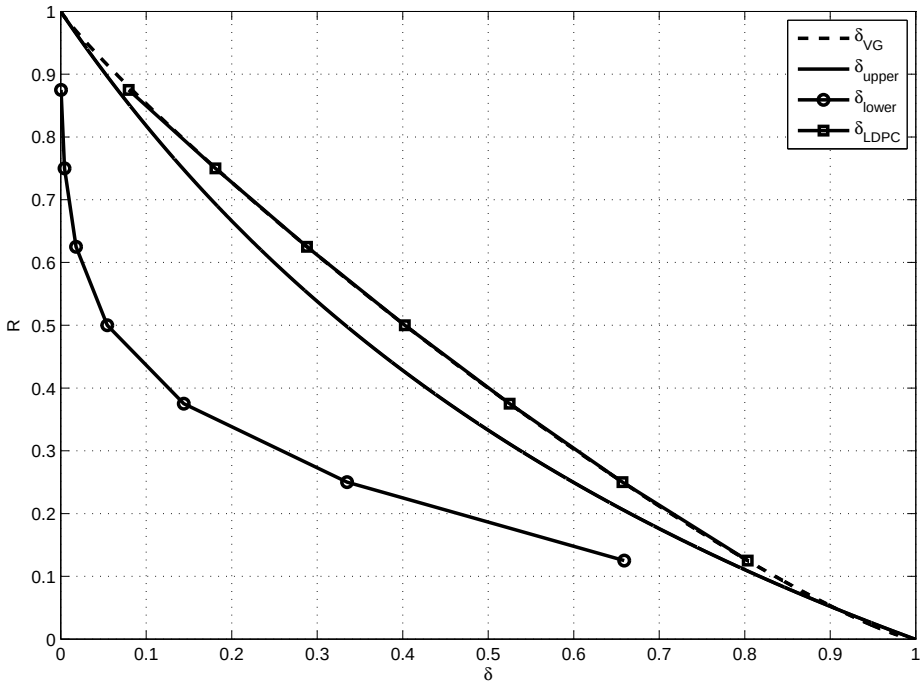

Рис. 5: Результаты для $q = 1024$

Таблица 3: Результаты для $q = 1024$

R	$\delta_{VG}(R)$	$\delta_{upper}(R)$	$\delta_{lower}(R);\Delta_0$	$\delta_{LDPC}(R);\Delta_0;\ell$
1/8	0,8036	**0,7770**	0,6590; 224	0,8035; 16; 14
1/4	0,6573	**0,5994**	0,3350; 248	0,6570; 16; 12
3/8	0,5252	**0,4541**	0,1440; 320	0,5250; 24; 15
1/2	0,4028	**0,3330**	0,0545; 332	0,4025; 28; 14
5/8	0,2884	**0,2305**	0,0180; 352	0,2880; 40; 15
3/4	0,1817	**0,1427**	0,0045; 224	0,1810; 60; 15
7/8	0,0835	**0,0666**	0,0005; 128	0,0795; 96; 12

ограничениям:

$$\begin{cases} a(i) \geqslant 0 \, \forall i \in [d_0, \Delta_0] \\ a(d_0) > 0 \end{cases} \tag{8}$$

Вычислим производную выражения, которое нужно максимизировать, получим:

$$\delta = \frac{1}{\Delta_0} \frac{s g_0'(s, \Delta_0)}{g_0(s, \Delta_0)}.$$

Таким образом, $s = f^{-1}(\delta)$, где $f(s) = \frac{1}{\Delta_0} \frac{s g_0'(s,\Delta_0)}{g_0(s,\Delta_0)}$.

В следующей лемме мы покажем, что функция $f^{-1}(\delta)$ непрерывна на полуинтервале $[0, +\infty)$.

Лемма 3 . *Функция* $f^{-1}(\delta)$, *где* $f(s) = \frac{1}{\Delta_0}\frac{sg_0'(s,\Delta_0)}{g_0(s,\Delta_0)}$, *непрерывна на полуинтервале* $[0, +\infty)$.

Доказательство. Для доказательства этого факта достаточно показать, что $f(s)$ строго возрастает на полуинтервале $[0, +\infty)$. Вычислим производную $f(s)$:

$$f'(s) = \frac{1}{\Delta_0}\frac{\left((sg_0')' g_0 - s(g_0')^2\right)}{g_0^2}$$

Теперь докажем, что $(sg_0')' g_0 - s(g_0')^2 > 0$. Преобразуем это выражение к виду:

$$s(sg_0')' g_0 - (sg_0')^2.$$

Теперь

$$\left[\sum_{i=d_0}^{\Delta_0}(i)^2 a(i) s^i\right]\left[1 + \sum_{i=d_0}^{\Delta_0} a(i) s^i\right] - \left[\sum_{i=d_0}^{\Delta_0} i a(i) s^i\right]^2.$$

Рассмотрим последовательности $\mathbf{x} = \left\{i\sqrt{a(i)s^i}\right\}_{i=d_0}^{\Delta_0}$ и $\mathbf{y} = \left\{\sqrt{a(i)s^i}\right\}_{i=d_0}^{\Delta_0}$ (мы можем это сделать в соответствии с условиями (8) и тем фактом, что $s > 0$). Определим скалярное произведение $(\mathbf{x}, \mathbf{y})$ так:

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=d_0}^{\Delta_0} x_i y_i,$$

тогда справедливо неравенство Коши-Буняковского:

$$\left[\sum_{i=d_0}^{\Delta_0}(x_i)^2\right]\left[\sum_{i=d_0}^{\Delta_0}(y_i)^2\right] \geqslant \left[\sum_{i=d_0}^{\Delta_0} x_i y_i\right]^2.$$

Подставив наши последовательности, получим

$$\left[\sum_{i=d_0}^{\Delta_0}(i)^2 a(i) s^i\right]\left[\sum_{i=d_0}^{\Delta_0} a(i) s^i\right] - \left[\sum_{i=d_0}^{\Delta_0} i a(i) s^i\right]^2 \geqslant 0,$$

следовательно, в связи с условиями (8)

$$\left[\sum_{i=d_0}^{\Delta_0}(i)^2 a(i) s^i\right]\left[1 + \sum_{i=d_0}^{\Delta_0} a(i) s^i\right] - \left[\sum_{i=d_0}^{\Delta_0} i a(i) s^i\right]^2 > 0,$$

что и завершает доказательство леммы. ▲

Следствие 1 . *Исследуемая функция*

$$G(\delta) = (\ell - 1)\left(h_q(\delta) + \delta\log_q(q-1)\right) + \ell\left(\delta\log_q\left(f^{-1}(\delta)\right) - \frac{1}{\Delta_0}\log_q\left(g_0\left(f^{-1}(\delta), \Delta_0\right)\right)\right),$$

непрерывна на $(0, +\infty)$ *ввиду того, что* $f^{-1}(\delta)$ *непрерывна и* $f^{-1}(\delta) > 0$ *при* $\delta > 0$.

Теперь мы готовы доказать лемму:

Лемма 4 . *Если существует хотя бы один положительный корень (относительно переменной δ) уравнения*

$$G(\delta) = 0 \tag{9}$$

то

$$\lim_{n\to\infty}\left(\sum_{W=1}^{\lfloor(\delta_0-\varepsilon)n\rfloor} q^{-nG(\delta)}\right) = 0,$$

где δ_0 – это наименьший положительный корень уравнения (9), ε – сколь угодно малая положительная величина.

Доказательство.

Выберем сколь угодно малую величину ε и рассмотрим следующие два предела:

$$\lim_{n\to\infty}\left(\sum_{W=1}^{\lfloor\varepsilon n\rfloor} q^{-nG(\delta)}\right) \tag{10}$$

и

$$\lim_{n\to\infty}\left(\sum_{\lceil W=\varepsilon n\rceil}^{\lfloor(\delta_0-\varepsilon)n\rfloor} q^{-nG(\delta)}\right). \tag{11}$$

Сначала рассмотрим предел (11):

Функция $G(\delta)$ непрерывна на отрезке $[\varepsilon, \delta_0-\varepsilon]$, следовательно, существует $\min\limits_{\delta\in[\varepsilon,\delta_0-\varepsilon]} G(\delta) = G_0$. $G_0 > 0$, так как $G(\delta) > 0$ при $\delta \in [\varepsilon, \delta_0-\varepsilon]$.

Теперь

$$\begin{aligned} 0 \leqslant \lim_{n\to\infty}\sum_{W=\lceil\varepsilon n\rceil}^{\lfloor(\delta_0-\varepsilon)n\rfloor} q^{-nG(\delta)} &\leqslant \lim_{n\to\infty}\sum_{W=\lceil\varepsilon n\rceil}^{\lfloor(\delta_0-\varepsilon)n\rfloor} q^{-nG_0} \\ &= \lim_{n\to\infty}\left(\left(\lfloor(\delta_0-\varepsilon)n\rfloor - \lceil\varepsilon n\rceil + 1\right) q^{-nG_0}\right) = 0 \end{aligned}$$

Таким образом,

$$\lim_{n\to\infty}\left(\sum_{\lceil W=\varepsilon n\rceil}^{\lfloor(\delta_0-\varepsilon)n\rfloor} q^{-nG(\delta)}\right) = 0.$$

Теперь вернемся к пределу (10)

Введем функцию $G_*(\delta)$:

$$\begin{aligned} G_*(\delta) = (\ell-1)\left(h_q(\delta) + \delta\log_q(q-1)\right) \\ + \ \ell\left(\delta\log_q\left(\delta^{\frac{1}{d_0}}\right) - \frac{1}{\Delta_0}\log_q\left(g_0\left(\delta^{\frac{1}{d_0}}, \Delta_0\right)\right)\right). \end{aligned}$$

Она получается из функции $G(\delta)$ заменой s на $\delta^{\frac{1}{d_0}}$. Очевидно, что

$$G(\delta) \geqslant G_*(\delta),$$

следовательно,

$$\lim_{n\to\infty} \sum_{W=1}^{\lfloor \varepsilon n \rfloor} q^{-nG\left(\frac{W}{n}\right)} \leqslant \lim_{n\to\infty} \sum_{W=1}^{\lfloor \varepsilon n \rfloor} q^{-nG_*\left(\frac{W}{n}\right)}$$

В соответствии с условиями (8) функция $g_0(s, n_0)$ имеет следующий вид:

$$g_0(s, n_0) = 1 + a(d_0)s^{d_0} + \ldots.$$

Заметим, что

$$\log_q \left(g_0\left(\delta^{\frac{1}{d_0}}, \Delta_0\right)\right) \leqslant \frac{1}{\ln q}\left(g_0\left(\delta^{\frac{1}{d_0}}, \Delta_0\right) - 1\right).$$

После преобразований получим следующую оценку для функции $G(\delta)$

$$G(\delta) \geqslant G_*(\delta) = -\left(\ell - 1 - \frac{\ell}{d_0}\right)\delta \log_q \delta + \mathcal{O}(\delta)$$

Пусть $\ell > \frac{d_0}{d_0 - 1}, d_0 \geqslant 2$, тогда

$$G(\delta) \geqslant -c_1 \delta \log_q \delta + c_2 \delta + o(\delta),\ c_1 > 0,$$

и

$$\begin{aligned}
\lim_{n\to\infty} \sum_{W=1}^{\lfloor \varepsilon n \rfloor} q^{-nG\left(\frac{W}{n}\right)} &\leqslant \lim_{n\to\infty} \sum_{W=1}^{\lfloor \varepsilon n \rfloor} q^{nc_1 \frac{W}{n} \log_q \frac{W}{n} - nc_2 \frac{W}{n}} \\
&= \lim_{n\to\infty} \sum_{W=1}^{\lfloor \varepsilon n \rfloor} \left(\frac{W}{n}\right)^{c_1 W} q^{-c_2 W} \\
&\leqslant \lim_{n\to\infty} \sum_{W=1}^{\lfloor \varepsilon n \rfloor} \left((\varepsilon)^{c_1} q^{-c_2}\right)^W \\
&= \frac{(\varepsilon)^{c_1} q^{-c_2}}{1 - (\varepsilon)^{c_1} q^{-c_2}} = 0
\end{aligned}$$

Заметим, что знак c_2 не важен, так как мы всегда можем сделать получившееся значение сколь угодно малым, правильно подобрав ε.

Так как оба предела существуют и конечны, то

$$\lim_{n\to\infty} \sum_{W=1}^{\lfloor (\delta_0 - \varepsilon) n \rfloor} q^{-nG\left(\frac{W}{n}\right)} = 0$$

▲

Список литературы

[1] *Tanner R.* A Recursive Approach to Low Complexity Codes // IEEE Trans. Inform. Theory. 1981. V. 27. № 5. P. 533 – 547.

[2] *Sipser M., Spielman D.* Expander Codes // IEEE Trans. Inform. Theory. 1996. V. 42. № 6. P. 1710 – 1722.

[3] *Lubotzky A., Phillips R., Sarnak P.* Ramanujan Graphs // Combinatorica. 1988. V. 8. P. 261 – 277.

[4] *Маргулис Г. А.* Явные конструкции расширителей // Пробл. передачи информ. 1973. Т. 9. № 4. С. 71 – 80.

[5] *Zemor G.* On Expander Codes // IEEE Trans. Inform. Theory. 2001. V. 47. № 2. P. 835 – 837.

[6] *Skachek V., Roth R.* Generalized Minimum Distance Iterative Decoding of Expander Codes // Proceedings of Information Theory Workshop. 2003. P. 245 – 248.

[7] *Barg A., Zemor G.* Distance Properties of Expander Codes // IEEE Trans. Inform. Theory. 2006. V. 52. № 1. P. 78 – 90.

[8] *Skachek, V.* Minimum distance bounds for expander codes // Information Theory and Applications Workshop. Jan. 27 2008-Feb. 1 2008. P. 366–370.

[9] *Boutros J., Pothier O., Zemor G.* Generalized Low Density (Tanner) Codes // in Proc. IEEE Int. Conf. Communications. 1999. V. 1. Vancouver, BC, Canada. P. 441—445.

[10] *Lentmaier M., Zigangirov K. Sh.* On Generalized Low-Density Parity-Check Codes Based on Hamming Component Codes // IEEE Commun. Lett. 1999. V. 3. № 8. P. 248—260.

[11] *Ben-Haim Y., Litsyn S.* Upper Bounds on the Rate of LDPC Codes as a Function of Minimum Distance // IEEE Trans. Inform. Theory. 2006. V. 52. № 5. P. 2092 – 2100.

[12] *Roth R., Skachek V.* Improved Nearly-MDS Expander Codes // IEEE Trans. Inform. Theory. 2006. V. 52. № 8. P. 3650 – 3661.

[13] *Питерсон У., Уэлдон Э.* Коды, исправляющие ошибки. М.: Мир, 1976.

[14] *Бассалыго Л. А.* Новые верхние границы для кодов, исправляющих ошибки // Пробл. передачи информ. 1965. Т. 1. № 4. С. 41—44.

[15] *McEliece, R., Rodemich, E., Rumsey, H., Welch, L.* New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities // IEEE Trans. Inform. Theory. 1977. V. 23. № 2. P. 157–166.

[16] *Галлагер Р. Дж.* Коды с малой плотностью проверок на четность. М.: Мир, 1966.

[17] *Barg A., Mazumdar A., Zemor G.* Weight distribution and decoding of codes on hypergraphs // Advances in Mathematics of Communications. 2008. V. 2. № 4. PP. 433–450.