Информационные процессы, Том 17, № 4, 2017, стр. 284–293
© *2017 Рыбин, Иванов.*

Оценка экспоненты вероятности ошибки для регулярных, основанных на графах МПП-кодов с конечной длиной[1]

Рыбин П. С., Иванов Ф. И.

Институт проблем передачи информации, Российская академия наук, Москва, Россия
Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

Поступила в редколлегию 15.12.2017

Аннотация—В данной работе исследована экспонента вероятности ошибки для регулярных, основанных на графах кодов с малой плотностью проверок при передаче по двоичному симметричному каналу и декодировании по максимуму правдоподобия. В отличие от большинства работ, где исследование экспонент вероятности ошибки проводилось в предположении, что длина кода стремится к бесконечности (асимптотический анализ), мы рассматриваем случай, когда длина кода конечна. В работе описан метод получения нижней границы на экспоненту вероятности ошибки для регулярного кода с малой плотностью проверок, описанного с точки зрения графа. Кроме того, были исследованы зависимости экспоненты вероятности ошибки от различных параметров кодов с малой плотностью проверок. Также в статье приведены численные результаты, полученные для выведенной нижней границы.

КЛЮЧЕВЫЕ СЛОВА: МПП-код, экспонента вероятности ошибки, конечная длина.

1. ВВЕДЕНИЕ

Коды с малой плотностью проверок (МПП-коды) были предложены Р. Галлагером в работе [1]. В данной работе также была получена нижняя граница на минимальное расстояние таких кодов. В работах [2] и [3] были получены верхняя и нижняя границы экспоненты вероятности ошибки декодирования МПП-кодов для алгоритма декодирования по максимуму правдоподобия. Кроме того, из работы [2] следует, что нижняя граница экспоненты вероятности ошибки МПП-кодов при декодировании по максимуму правдоподобия асимпотически стремится к экспоненте вероятности ошибки декодирования “хороших” линейных кодов, которая была получена в [4].

Следует отметить, что во всех указанных выше работах экспоненты вероятности ошибки были получены для асимптотического случая: т. е. в предположении, что длина кода стремится к бесконечности. Данное предположение позволяет нам использовать некоторые хорошо известные результаты, используемые при исследования асимптотического поведения, для решения поставленной нами задачи анализа МПП-кодов с конечными длинами.

Существует достаточно ограниченное число работ, где рассматривалось неасимптотическое поведение МПП-кодов. В статье [5] исследовалось поведение МПП-кодов конечной длины в области “начала водопада” (таких отношений сигнал-шум, где начинается уменьшение выходной вероятности ошибки на кодовое слово после декодирования). Авторы предложили алгоритм, основанный на методе оценке плотностей (англ. Density Evolution), который предсказывает

[1] Работа выполнена в ИППИ РАН при финансовой поддержке Российского научного фонда, грант № 14-50-00150.

корректирующие характеристики МПП-кодов конечной длины при передаче кодовых слов через различные двоичные симметиричные каналы без памяти. В работах [6], [7], [8] для описания ошибок декодирования были использованы различные комбинаторные методики. Такой подход позволил авторам получить оценки для средней вероятности ошибки на блок и на бит при использовании итеративных алгоритмов декодирования. Некоторые модификации данных методов были получены в работе [9]. Большинство рассмотренных выше работ оперируют с только двоичным стирающим каналом и алгоритмами декодирования с малой сложностью (например, с мажоритарным или алгоритмом “распространения доверия”). Лишь небольшое число работ посвящено каналам с ошибками.

Данная работа преимущественно опирается на работы [10] и [11], где были разработаны методы анализа МПП-кодов, основанных на графах. Также следует отметить, что исследование усреденного спектра МПП-кодов конечных длин представлено в работе [7]. Тем не менее в указанной работе не была рассмотрена экспонента вероятности ошибки. В данной работе мы исследуем корректирующие свойства МПП-кодов фиксированной длины в предположении их декодирования по максимуму правдоподобия. Мы представим метод получения нижней границы экспоненты ошибочного декодирования, используя при этом производящие функции и некоторые другие комбинаторные методы. Кроме того, мы продемонстрируем численные результаты расчетов и проанализируем поведение выведенной границы для различных параметров кодов.

2. КОНСТРУКЦИЯ МПП-КОДОВ

В некоторых случаях удобно представлять МПП-код в виде графа Таннера [12]. Данный граф является двудольным, где $m > 0$ вершин-проверок (проверочных символов соединяются с $n > m$ вершинами-символами (кодовыми битами), при этом n является длиной соответствующего МПП-кода. Хотя в данной работе рассматриваются только регулярные МПП-коды (это значит, что степени всех вершин-проверок и вершин-символов одинаковы и равны l и n_0 соответственно), мы представим более общий случай описания ансамблей МПП-кодов. В этом случае граф Таннера может быть описан при помощи двух векторов, составленных из вероятностей:

$$\tilde{\lambda} = \left(\tilde{\lambda}_2, \ldots, \tilde{\lambda}_l\right),$$

$$\tilde{\rho} = \left(\tilde{\rho}_1, \ldots, \tilde{\rho}_{n_0}\right),$$

где $\tilde{\lambda}_k$ является долей вершин-проверок степени k, а $\tilde{\rho}_t$ — доля вершин-символов степени t. Для удобства дальнейшего изложения мы также введем два многочлена:

$$\tilde{\lambda}\left(x\right) = \sum_{k=2}^{l} \tilde{\lambda}_k x^{k-1},$$

$$\tilde{\rho}\left(x\right) = \sum_{t=2}^{n_0} \tilde{\rho}_t x^{t-1}.$$

Пусть E равно общему числу ребер в графе, n является числом вершин-символов, а m определяет число вершин-проверок. Тогда $n\tilde{\lambda}_k k$ равно числу ребер, выходящих из вершин-символов степени k и $m\tilde{\rho}_t t$ совпадает с числом ребер, выходящих из вершин-проверок степени t. Таким образом,

$$n = \frac{E}{\sum\limits_{k=2}^{l} \tilde{\lambda}_k k} = \frac{E}{1 + \tilde{\lambda}'\left(1\right)},$$

$$m = \frac{E}{\sum\limits_{t=2}^{n_0} \tilde{\rho}_t t} = \frac{E}{1 + \tilde{\rho}'(1)},$$

где $\tilde{\lambda}'(1)$ and $\tilde{\rho}'(1)$ производные функций $\tilde{\lambda}(x)$ и $\tilde{\rho}(x)$ по переменной x, вычисленные в точке $x = 1$.

Каждой вершине-символу степени i назначим i гнезд-символов (мест подсоединения ребер к данной вершине-символу). Аналогично каждой вершине-проверке степени i назначим i гнезд-проверок. Общее число гнезд-символов и общее число гнезд-проверок совпадают между собой и равны общему числу ребер E. Ансамбль двудольных графов определяется выбором случайной перестановки π из множества равновероятных перестановок длины E. Для каждого $1 \leq i \leq E$ мы соединим вершину-символ, соответствующему i-ому гнезду-символу с вершиной-проверкой, соответствующей с π_i-ому гнезду-проверке. Следует обратить внимание, что при таком подходе между парами вершин могут образоваться кратные ребра. Отображение из двудольного графа в проверочную матрицу $\mathbf{H}$ МПП-кода осуществляется следующим образом: элемент $\mathbf{H}_{i,j}$ равен 1 в том случае, если соответствующие i-ая вершина-проверка и j-ая вершина-символ соединены нечетным числом ребер, в противном случае $\mathbf{H}_{i,j} = 0$.

Скорость R каждого кода в построенном нами ансамбле удовлетворяет неравенству: $R \geq R'$, где

$$R' = 1 - \frac{m}{n} = 1 - \frac{\sum\limits_{k=2}^{l} \tilde{\lambda}_k k}{\sum\limits_{t=2}^{n_0} \tilde{\rho}_t t} = 1 - \frac{1 + \tilde{\lambda}(1)}{1 + \tilde{\rho}'(1)}$$

является проектируемой скоростью кода (неравенство может иметь место ввиду того, что среди m проверочных уравнений в матрице $\mathbf{H}$ найдутся линейно-зависимые).

Частным случаем представленного ансамбля нерегулярных кодов является ситуация, когда все вершины-символы имеют одну и ту же степень l, а вершины-проверки - фиксированную степень n_0. В этом случае ансамбль называется регулярным, причем найдется такая константа $m \in \mathbb{N}$, что $nl = mn_0$, $R \geq 1 - l/n_0$, и

$$\tilde{\lambda}(x) = x^{l-1},$$

$$\tilde{\rho}(x) = x^{n_0-1}.$$

В данной работе рассматриваются только регулярные ансамбли МПП-кодов. Граф Таннера регулярного (l, n_0) МПП-кода представлен на рисунке . 1.

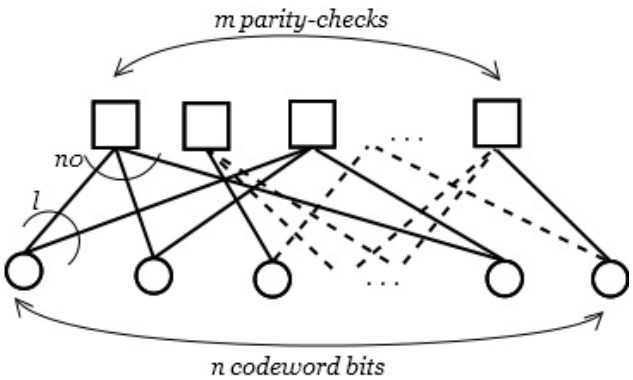

Рис. 1. Граф Таннера регулярного (2,3) МПП-кода

3. ОСНОВНОЙ РЕЗУЛЬТАТ

Для изучения вероятности ошибки декодирования P мы рассмотрим двоичный симметричный канал с вероятностью перехода p. Вероятность ошибки P мы будем оценивать в следующем виде:

$$P \leq \exp\{-nE(\cdot)\},$$

где $E(\cdot)$ — экспонента вероятности ошибки.

В данной работе мы рассмотрим экспоненту $E(\cdot)$ для (l, n_0) регулярного МПП-кода длины n при его декодировании по максимуму правдоподобия. Известно, что данный алгоритм декодирования имеет экспоненциальную сложность.

В работе [10] была доказана следующая теорема:

Теорема. *Среднее число $\overline{N}(W)$ кодовых слов веса W нерегулярного МПП-кода, определенного многочленами $\left(\tilde{\lambda}(x), \tilde{\rho}(x)\right)$, удовлетворяет следующему соотношению:*

$$\overline{N}(W) = \sum_{j=0}^{E} \frac{t(W,j)\, q(j)}{\binom{E}{j}}, \tag{1}$$

где $t(W,j)$ определяется следующим образом:

$$t(W,j) = \left(\prod_k \left(1 + xy^k\right)^{\tilde{\lambda}_k n}\right)[x^W y^j],$$

где запись $f_i = f(x)[x^i]$ означает, что $f(x) = \sum\limits_i f_i x^i$, и $q(j)$ определяется как:

$$q(j) = \left(\prod_k g_0^{\tilde{\rho}_k (1-R)n}(x,k)\right)[x^j],$$

где $g_0 = \frac{1}{2}\left((1+x)^k + (1-x)^k\right)$

Для регулярного (l, n_0) МПП-кода достаточно легко рассчитать точное значение $t(W,j)$. В самом деле, так как $\tilde{\lambda}_k = 1$ когда $k = l$, и $\tilde{\lambda}_k = 0$ иначе, то:

$$t(W,j) = (1 + xy^l)^n [x^W y^j],$$

$$(1 + xy^l)^n [x^W y^j] = \sum_{k=0}^{n} \binom{n}{k} x^k y^{lk} [x^W y^j],$$

$$t(W,j) = \begin{cases} \binom{n}{W}, \ j = lW \\ 0, \ j \neq lW \end{cases}.$$

Также можно упростить выражение для $q(j)$:

$$q(j) = (g_0^m(x, n_0))[x^j],$$

$$q(j) = \left(\sum_{k=0}^{\lfloor \frac{n_0}{2} \rfloor} \binom{n_0}{2k} x^{2k}\right)^m [x^{lW}].$$

Таким образом,

$$\overline{N}\left(W\right)=\frac{\binom{n}{W}q\left(lW\right)}{\binom{E}{lW}}, \tag{2}$$

В связи с тем, что точное вычисление $q\left(lW\right)$ в формуле (2) является достаточно сложной задачей, мы используем теорему, приведенную в [13], для оценки данного коэффициента:

Теорема (Hayman, 1956). *Пусть* $f(x)=\sum\limits_{k}a_kx^k$ *и* r_k *является положительным корнем уравнения* $a(r_k)=k$ *для всех* $k=1,2,\ldots,$ *где* $a(x)=x\frac{f'(x)}{f(x)}$. *Тогда*

$$a_k\approx\frac{f(r_k)}{r_k^k\sqrt{2\pi b(r_k)}}, \tag{3}$$

где $b(x)=xa'(x)$

В нашем случае $f(x)=\left(g_0\left(x,n_0\right)\right)^m$ и $k=lW$. Поэтому

$$a(x)=x\frac{m\left(g_0\left(x,n_0\right)\right)^{m-1}g_0'\left(x,n_0\right)}{\left(g_0\left(x,n_0\right)\right)^m}.$$

Очевидно, что $f(x)\neq 0$ для $x\geq 0$, тогда

$$a(x)=mx\frac{g_0'\left(x,n_0\right)}{g_0\left(x,n_0\right)}.$$

Таким образом,

$$mr_{lW}\frac{g_0'\left(r_{lW},n_0\right)}{g_0\left(r_{lW},n_0\right)}=lW.$$

Окончательно имеем:

$$q\left(lW\right)\leq a_{lW},$$

где a_{lW} получено из (3). Наконец, мы можем оценить среднее число $\overline{N}\left(W\right)$ кодовых слов веса W регулярного (l,n_0) МПП-кода следующим образом:

$$\overline{N}\left(W\right)\leq\frac{a_{lW}\binom{n}{W}}{\binom{E}{lW}} \tag{4}$$

Неравенство (4) позволяет нам оценить не только спектр МПП-кода, но также и его относительное минимальное расстояние $\delta=\frac{d}{n}$. Мы представим полученные результаты оценки в следующей главе данной работы.

Теперь определим усредненную спектральную функцию $\psi\left(z\right)$ для регулярного (l,n_0) МПП-кода следующим образом :

$$\psi\left(z\right)=\sum_{W=d}^{n}\overline{N}\left(W\right)z^W.$$

Применяя (4), получим следующую очевидную оценку для $\psi\left(z\right)$:

$$\psi\left(z\right)\leq n\sum_{W=d}^{n}\overline{N}\left(W\right)z^W\leq n\sum_{W=d}^{n}\frac{a_{lW}\binom{n}{W}}{\binom{E}{lW}}z^W.$$

Оценка спектральной функции будет использована нами в следующей теореме, полученной в работе [14]. Данная теорема позволяет оценить вероятность ошибки декодирования по максимуму правдоподобия для заданного регулярного (l,n_0) МПП-кода.

Теорема. *Вероятность ошибки $P(R,p)$ линейного кода со спектральной функцией $\psi(z)$, скоростью R и длиной n при декодировании по максимуму правдопобия в двоичном симметричном канале с переходной вероятностью p оценивается сверху следующим образом:*

$$P(R,p) \le \exp\left(-nE(R,p)\right),$$

где $E(R,p)$ определяется как:

$$E(R,p) = \max_{s\ge 0, t\le 0, r\le 0} \left\{ \frac{r}{s-r}\ln g(s) - \frac{s}{s-r}\left(\ln g(r) + \frac{1}{n}\ln\psi\left(\frac{g(r,t)}{g(r)}\right)\right)\right\},$$

где $g(s)$ и $g(r,t)$ — производящие функции, которые вычисляются следующим образом:

$$g(s) = (1-p)^{1-s} + p^{1-s},$$

$$g(r,t) = (1-p)^{1-r+t}p^{-t} + (1-p)^{-t}p^{1-r+t}.$$

4. ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

Вначале рассмотрим оценку среднего числа $\overline{N}(W)$ кодовых слов веса W для регулярных МПП-кодов с фиксированными параметрами l и n_0, но различными длинами n. Для более наглядного представления численных результатов мы введем в рассмотрение специальную функцию $\nu(\omega) = \frac{\log \overline{N}(\omega n)}{n}$, где $\omega = \frac{W}{n}$, $0 \le \omega \le 1$

На рис. 2 представлена оценка спектра для регулярных МПП-кодов параметрами $(l, n_0) = (7, 14)$ и различными длинами: $n = 100$, $n = 300$, $n = 1000$ а также с фиксированной скоростью $R = 0.5$. Легко заметить, что $\nu(\omega)$ растет вместе с увеличением длины кода n.

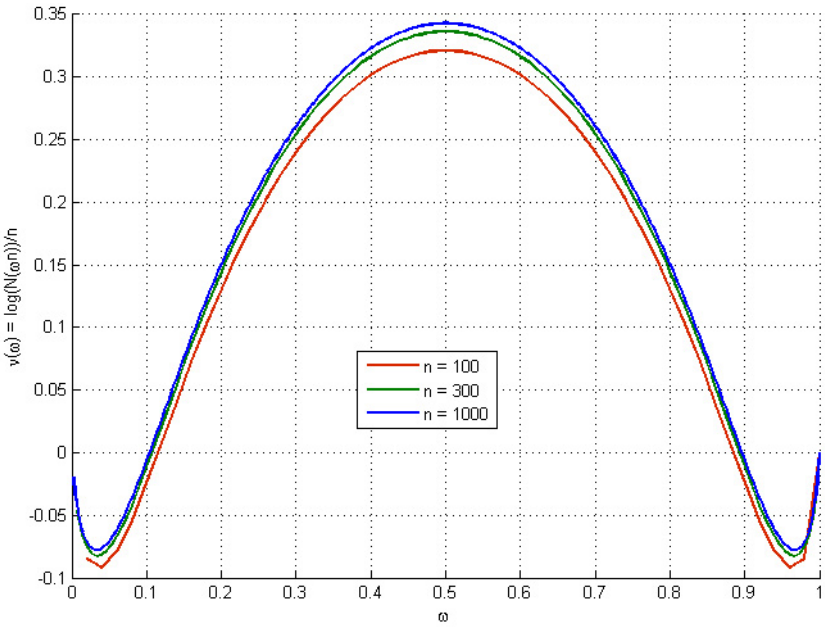

Рис. 2. Сравнение зависимостей $\nu(\omega)$ от n для фиксированных $l = 7$, $n_0 = 14$, $R = 0.5$

На рис. 3 представлена зависимость $\nu(\omega)$ от l для фиксированной длины $n = 300$ и скорости $R = 0.5$.

Легко заметить, что $\nu(\omega)$ увеличивается с уменьшением l (особенно это заметно для больших и малых значений ω). Данный факт означает, что относительное минимальное расстояние

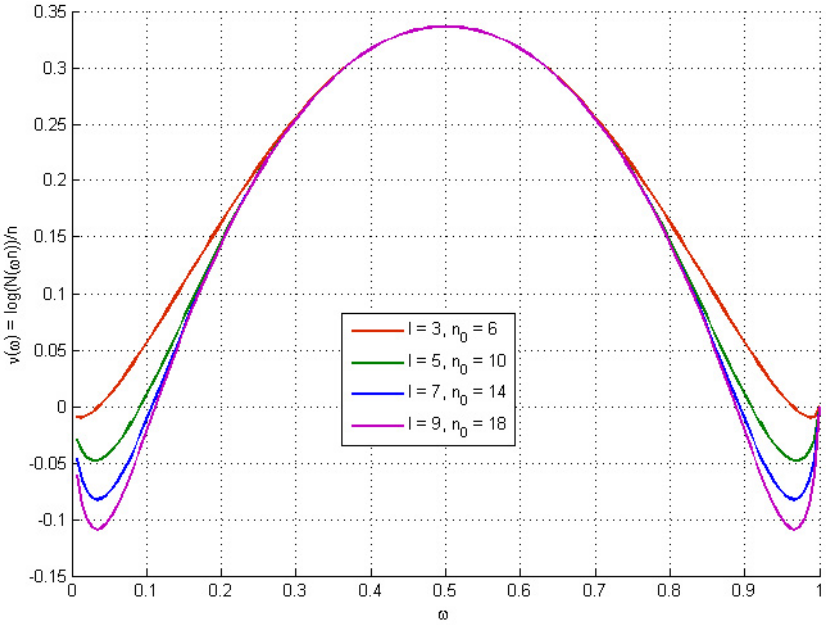

Рис. 3. Сравнение зависимостей $\nu(\omega)$ от l для заданных $n = 300$ и $R = 0.5$

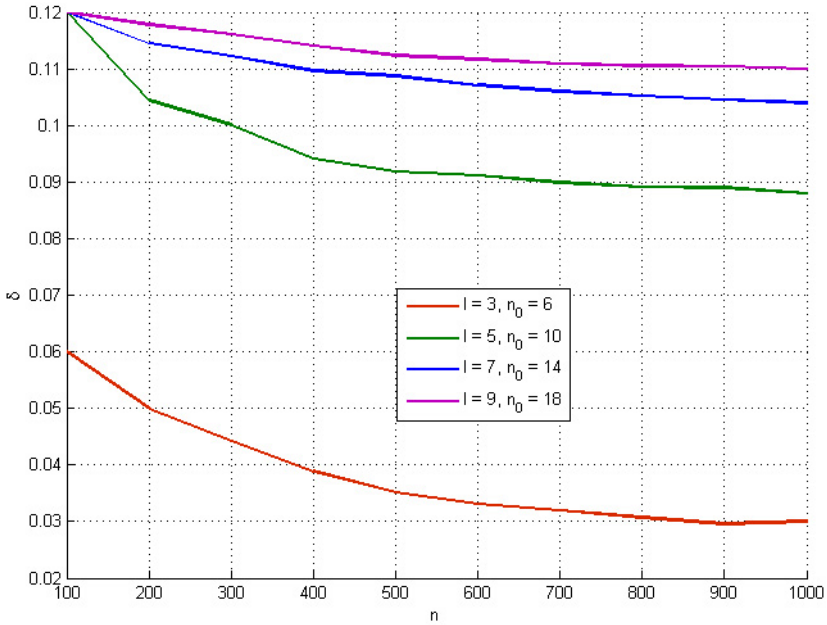

Рис. 4. Сравнение зависимостей δ от l для заданной скорости $R = 0.5$ и различных n

$\delta = \frac{d}{n}$ является функцией от l, которая монотонно возрастает по l (для заданных n и R). Данный факт более заметен на рис. 4.

Численные результаты, представленные на рис. 4, позволяют нам сделать вывод о том, что относительное минимальное расстояние уменьшается с ростом длины кода n. Также можно обратить внимание на то, что МПП-коды с параметрами $n \approx 100$, $R = 0.5$ и $l \in \{5, 7, 9\}$ имеют $\delta \approx 0.12$, в то время как в асимптотическом случае $\delta_{VG}(0.5) \approx 0.11$, где $\delta_{VG}(R)$ - граница Варшамова-Гилберта. Таким образом, при малых длинах данные МПП-коды лежат выше границы Варшамова-Гилберта.

Теперь представим результаты расчетов экспоненты вероятности ошибки $E(R,p)$, полученные согласно теореме 3.

На рис. 5 представлена зависимость экспоненты вероятности ошибки $E(R,p)$ от длины МПП-кода n и переходной вероятности в канале p. Остальные параметры кода фиксированы: $l = 7$, $n_0 = 14$, $R = 0.5$. Можно заменить, что экспонента растет вместе с n и уменьшается вместе с ростом p (когда n фиксированно).

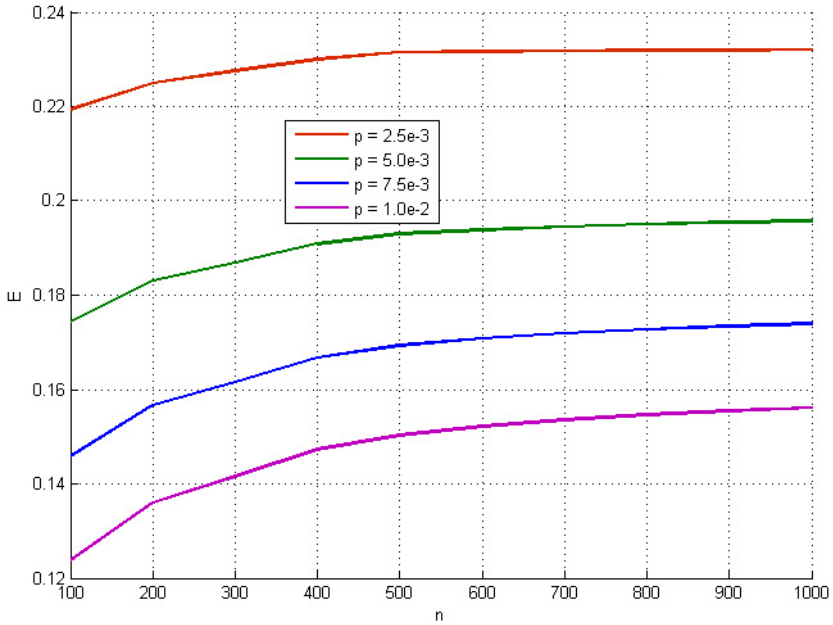

Рис. 5. Сравнение зависимостей $E(R,p)$ от n и p для заданных $R = 0.5$ и $l = 7$

Рис. 6 демонстрирует зависимость $E(R,p)$ от n и l при заданной скорости $R = 0.5$ и вероятности ошибки в канале $p = 2.5 \cdot 10^{-3}$. Можно заметить, что экспонента вероятности ошибки убывает при увеличении n в случае, когда $l = 3$ или $l = 5$, но растет вместе с увеличением n для МПП-кодов с $l = 7$ или $l = 9$. Данный факт объясняется существенным ухудшением относительного минимального расстояния при росте длины для кодов с $l = 3$ и $l = 5$ (см. 4).

В завершение данной главы мы представим результаты сравнения экспоненты вероятности ошибки, полученной в данной работе (сплошные линии на графике), с экспонентой вероятности ошибки, полученной с применением границы Полянского [15] для случая случайный двоичных кодов (пунктирные графики). Данное сравнение приведено на рис. 7. В качестве МПП-кодов были выбраны $(7, 14)$ регулярные коды скорости $R = 0.5$ с длинами от 100 до 1000. Сравнения приведены для различных входных вероятностей ошибки в канале. Легко заметить, что граница Полянского для случайных кодов всегда выше, чем граница, полученная в данной работе. Этот факт говорит о том, что регулярные МПП-коды (особенно при коротких длинах) обладают меньшими потенциальными корректирующими свойствами, чем произвольные двоичные коды.

5. ЗАКЛЮЧЕНИЕ

В работе описан метод оценки спектральной функции двоичного регулярного МПП-кода конечной длины, представленного в виде двудольного графа. Данная оценка была использована для вывода нижней границы экспоненты вероятности ошибочного декодирования по максимуму правдоподобия при передача информации по двоичному симметричному каналу. Чис-

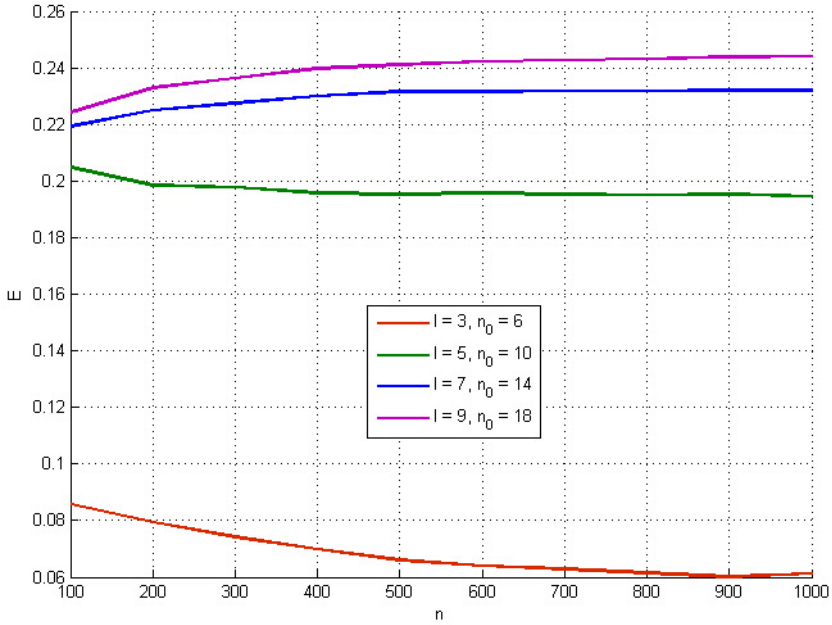

Рис. 6. Сравнение зависимостей $E(R,p)$ от n и l для заданных $R = 0.5$ и $p = 2.5 \cdot 10^{-3}$

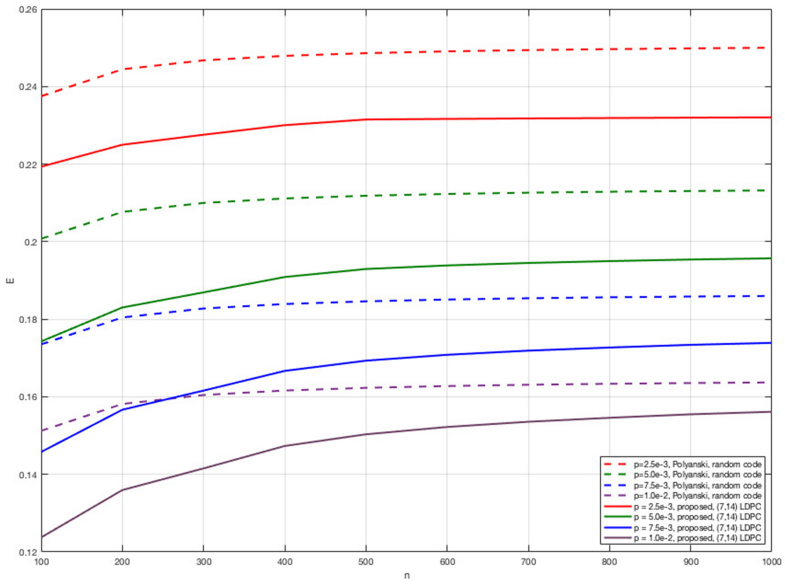

Рис. 7. Сравнение зависимостей $E(R,p)$ от n и p для заданных $R = 0.5$ и $l = 7$ для МПП-кодов и случайных двоичных кодов

ленные результаты, представленные в работе, показывают, что регулярные $(3,6)$ МПП-коды, которые широко используются на практике ввиду их хороших корректирующих характеристик при итеративном декодировании, имеют наихудшую экспоненту при декодировании по максимуму правдоподобия. Более того, эта экспонента убывает с ростом длины кода. Данные факты означают, что при проектировании сигнально-кодовых конструкций, где требуется при достичь очень малых вероятностей ошибки, данный класс кодов следует заменить на некоторый другой, обладающий лучшей экспонентой.

СПИСОК ЛИТЕРАТУРЫ

1. Gallager R. G. *Low-Density Parity-Check Codes.* Cambridge, MA, USA: MIT Press, 1963.

2. Barak O., Burshtein D. Lower Bounds on the Error Rate of LDPC Code Ensembles. *IEEE Trans. Inform. Theory*, 2007, vol. 53, pp. 4225 – 4236.

3. Burshtein D., Barak O. Upper Bounds on the Error Exponents of LDPC Code Ensembles. *2006 IEEE International Symposium on Information Theory (ISIT).* Seattle, USA:2006, pp. 401 – 405.

4. Gallager R. G., *Information Theory and Reliable Communication.* New York, USA: John Wiley and Sons Inc., 1968.

5. Noor-A-Rahim M., Nguyen K. D., Lechner G. Finite length analysis of LDPC codes. *IEEE Wireless Comm. and Networking Conference (WCNC).* Istambul, Turkey: 2014.

6. Proietti C. D., Telatar I. E., Richardson T. J., Urbanke R. L. Finite-length analysis of low-density parity-check codes on the binary erasure channel. *IEEE Trans. Inform. Theory*, 2002, vol. 48, no. 6, pp. 1570–1579.

7. Richardson T. J., Urbanke R. L. Finite-length density evolution and the distribution of the number of iterations on the binary erasure channel. *unpublished manuscript, available at hhttp://lthcwww.epfl.ch/papers/RiU02.psi.*

8. Zhang H., Orlitsky A. Finite-length analysis of LDPC codes with large left degrees. *Proc. International Symposium on Information Theory (ISIT).* Lausanne, Switzerland: 2002.

9. Johnson S. J. A finite-length algorithm for LDPC codes without repeated edges on the binary erasure channel. *IEEE Transactions on Information Theory*, 2009, vol. 55, no. 1, pp. 27–32.

10. Burshtein D., Miller G. Asymptotic enumeration methods for analyzing LDPC codes. *Information Theory, IEEE Transactions on*, 2004, vol. 50, no. 6, pp. 1115–1131.

11. Rybin P. S. On the error-correcting capabilities of low-complexity decoded irregular LDPC codes. *IEEE Int. Symposium on Inform. Theory (ISIT).* Hawaii, USA: 2014, pp. 3165 – 3169.

12. Tanner R. A recursive approach to low complexity codes. *IEEE Trans. Inform. Theory*, 1981, vol. 27, no. 5, pp. 533–547.

13. Wilf H. S. *Generatingfunctionology.* USA: Academic Press Inc., 1992.

14. Blokh E. L., Zyablov V. V., *Linear concatenated codes.* Moscow, U.S.S.R: Nauka, 1982.

15. Polyanskiy Y., Poor H. V., Verdu S. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theory*, 2010, vol. 56, no. 5, pp. 2307–2359.

On the Finite Length Analysis of Regular Graph-Based LDPC Codes Error Exponent

Rybin P. S., Ivanov F. I.

This paper deals with the error exponent of the regular graph-based binary low-density parity-check (LDPC) codes under the maximum likelihood (ML) decoding algorithm in the binary symmetric channel (BSC). Unlike other papers where error exponents are considered for the case when the length of LDPC codes tends to infinity (asymptotic analysis) we considered the finite length case (finite length analysis). In this paper we describe the method of deriving the lower bound on the error exponent for regular graph-based LDPC code with finite length under ML decoding and analyze the dependency of the error exponent on various LDPC code parameters. The numerical results, obtained for the considered lower bound, are represented and analyzed at the end of the paper.

KEYWORDS: LDPC code, error exponent, finite length.