

УДК 621.391 : 519.72

© 2024 г. И.А. Мельников, А.Ю. Угловский, А.А. Крещук,
А.А. Куреев, Е.М. Хоров

ИСПОЛЬЗОВАНИЕ МЕТОДА ИНФОРМАЦИОННОГО СЖАТИЯ ДЛЯ СНИЖЕНИЯ СЛОЖНОСТИ ДЕКОДЕРА МПП-КОДОВ¹

Одним из способов снижения сложности алгоритмов распространения доверия для декодирования кодов с малой плотностью проверок на четность является хранение предварительно вычисленной суммы сообщений в узлах переменных. В свою очередь, объем обрабатываемой информации может быть значительно снижен с помощью метода информационного сжатия (МИС), снижающего разрядность всех обновляемых сообщений. Предлагается алгоритм построения бинарной функции на основе МИС, соответствующей вычитанию. Использование разработанной функции позволяет уменьшить количество хранимых и используемых таблиц поиска для узлов переменных.

Ключевые слова: Алгоритм распространения доверия, итеративное декодирование, коды с малой плотностью проверок на четность (МПП), метод информационного сжатия, последовательное декодирование.

DOI: 10.31857/S0555292324030033, EDN: QUNUCD

§ 1. Введение

Коды с малой плотностью проверок на четность (МПП) являются одними из самых широко используемых помехоустойчивых кодов в системах связи [1]. МПП-декодеры могут использовать такие подходы как алгоритм распространения доверия (РД, англ.: belief propagation) [2] и метод суммы наименьших значений (СН, англ.: min-sum) [3], которые итеративно обновляют логарифм отношения правдоподобия (ЛОП, англ.: log-likelihood ratios).

К сожалению, большим недостатком РД и СН является их высокая вычислительная сложность и большая разрядность ЛОП. Одним из методов снижения вычислительной сложности является хранение суммы нескольких сообщений ЛОП и последующее вычитание каждого сообщения ЛОП. В свою очередь, для снижения объема распространяемой информации можно использовать метод информационного сжатия (МИС, англ.: information bottleneck) [4, 5]. Этот метод использует специальные дискретные функции, определяемые таблицами поиска (англ.: lookup tables), которые заменяют обычные арифметические операции [6], например, сложение. Эти таблицы строятся с помощью метода дискретной эволюции плотности (ДЭП, англ.: discrete density evolution) [7]. МИС приводит к неравномерному квантованию ЛОП, и следовательно, полученные функции не являются коммутативными и ассоциативными. В результате вычитание, используемое для снижения вычислительной сложности, становится едва возможным.

¹ Исследование осуществлено в рамках Госзадания № FFNU-2022-0035 ИППИ РАН.

Настоящая статья посвящена решению данной проблемы и предлагает метод построения дискретной двоичной функции для построения операции вычитания. Численные результаты демонстрируют сокращение вдвое количества операций в МИС-декодере без значительной потери производительности.

Статья устроена следующим образом. В § 2 представлено описание МИС. В § 3 предлагается алгоритм построения дискретной двоичной функции вычитания. Численные результаты и оценка производительности приводятся в § 4. Наконец, § 5 завершает статью.

§ 2. Метод информационного сжатия

Для снижения размерности используемых ЛОП может применяться метод информационного сжатия. Этот метод позволяет найти компромисс между сжатием наблюдаемой случайной величины Y и информацией об истинной случайной величине X , при заданном совместном распределении вероятностей $p(x, y)$ величин X и Y . Такой подход вводит случайную величину T , такую что релевантная информация $I(T; X)$ велика, но информация сжатия $I(T; Y)$ мала. Пусть x , y и t являются элементами из пространств событий случайных величин X , Y и T соответственно, а $p(t|y)$ определяет отображение Y в T , т. е. $p(t|y) = 1$, если t является сжатым представлением y , и $p(t|y) = 0$ в противном случае.

Цель метода информационного сжатия состоит в том, чтобы найти такое распределение $p(t|y)$, которое минимизирует следующий функционал L :

$$L[p(t|y)] = I(T; Y) - \beta I(T; X) \rightarrow \min, \quad (1)$$

где β определяет компромисс между $I(T; X)$ и $I(T; Y)$. Если $\beta = 0$, то сжатие будет максимальным, а $\beta \rightarrow \infty$ указывает на максимальное сохранение релевантной информации [8]. Глобально оптимальное решение (1) можно найти с помощью алгоритма динамического программирования для метода информационного сжатия [9].

В контексте задачи декодирования случайная величина X представляет собой сообщение на входе канала, наблюдаемая случайная величина Y – сообщение на выходе канала, а случайная величина T – сжатое представление Y . При декодировании ЛОП дискретизируются, а стандартные операции суммирования заменяются операциями, описываемыми таблицами поиска, которые возвращают сжатую сумму $t_\ell = y_\ell \oplus t_{\ell-1}$ между ЛОП y_ℓ и частичной сжатой суммы $t_{\ell-1}$ предыдущих $\ell-1$ ЛОП, где $\ell > 1$, $t_1 = y_1$. Другими словами, таблица поиска ℓ принимает на вход y_ℓ и $t_{\ell-1}$ и возвращает t_ℓ .

Таблицы поиска строятся последовательно для каждого ℓ , начиная с $\ell = 2$. На каждом шаге таблица поиска строится путем решения задачи метода информационного сжатия для распределения

$$p(y_\ell, t_{\ell-1} | x_1 = \dots = x_\ell),$$

где x_ℓ равно 0 или 1. Стоит отметить, что $x_1 = \dots = x_\ell$, потому что наблюдаемые величины соответствуют одному и тому же биту входных данных канала.

Используя таблицы поиска, можно представить работу алгоритма распространения доверия следующим образом. Алгоритм распространения доверия присваивает значения ЛОП P_v на выходе канала каждому узлу переменных $v \in V$ и два значения ЛОП Q_{vc} и R_{cv} ребрам (v, c) , где $c \in C$ и $v \in V$. При этом Q_{vc} является сообщением, отправляемым от v к c , а R_{cv} – сообщением, отправляемым в противоположном направлении. Затем алгоритм рекурсивно обновляет Q_{vc} и R_{cv} , учитывая только

соседние узлы:

$$\begin{aligned} Q_{vc} &\leftarrow P_v \oplus \left(\bigoplus_{c' \in N(v) \setminus c} R_{c'v} \right), \\ R_{cv} &\leftarrow -s_{cv} \left(\bigoplus_{v' \in N(c) \setminus v} Q_{v'c} \right), \end{aligned} \quad (2)$$

где $N(k)$ – множество узлов соседей узла k , а

$$s_{cv} = \prod_{v' \in N(c) \setminus v} \text{sign}(Q_{v'c}).$$

§ 3. Вычитание с помощью метода информационного сжатия

Определим функцию $\ominus$, которая вычитает одно значение ЛОП из общей суммы ЛОП, и реализуем ее, используя таблицы поиска. Для этого перепишем первое уравнение (2) следующим образом:

$$\begin{aligned} A_v &\leftarrow P_v \oplus \left(\bigoplus_{c' \in N(v)} R_{c'v} \right), \\ Q_{vc} &\leftarrow A_v \ominus R_{cv}. \end{aligned} \quad (3)$$

Таблица поиска для операции вычитания $\ominus$ наблюдаемой величины Y из частичной суммы строится с помощью метода информационного сжатия для распределения

$$p(t, y_j | x_i = b, i \neq j),$$

где t – сумма ЛОП по множеству узлов N , x_i – i -й бит на входе канала, j – индекс вычитаемого входа, b равно 0 или 1. Пусть $s_j = t \ominus y_j$. Для простоты далее примем $j = 1$. Аналогично $I(T; X)$ в (1) рассмотрим следующую взаимную информацию:

$$I(s_1; x_i = b) = \sum_{(s_1, b)} P(s_1, b | x_i = b) \log \frac{P(s_1, b | x_i = b)}{P(s_1)P(b | x_i = b)}, \quad (4)$$

где $i \neq 1$. Тогда в случае использования вычитания в декодере и с параллельным расписанием (ПрР) имеем

$$P(s_1, b | x_i = b) = \sum_{\left(\bigoplus_{k=1}^n y_k \right) \ominus y_1 = s_1} P(y_1)P(b | x_i = b) \prod_{r=2}^n P(y_r, b | x_r = b). \quad (5)$$

Остаточная информация $I(s_1; x_i = b)$ показывает объем информации о “невывчитенных” битах, оставшийся после вычисления общей суммы и вычитания первого входа преобразования. Она является релевантной информацией, которую необходимо максимизировать.

$I(s_1; x_1 = b)$ – взаимная информация между значениями, которые должны быть независимыми в методе дискретной эволюции плотности. Она соответствует информации о первом вычитаемом бите после его вычитания. Для $i = 1$ имеем

$$I(s_1; x_1 = b) = \sum_{(s_1, b)} P(s_1, b | x_1 = b) \log \frac{P(s_1, b | x_1 = b)}{P(s_1)P(b | x_1 = b)}. \quad (6)$$

В случае ПрР получим

$$P(s_1, b | x_1 = b) = \sum_{\left(\bigoplus_{k=1}^n y_k\right) \ominus y_1 = s_1} P(y_1, b | x_1 = b) \prod_{r=2}^n P(y_r).$$

Для оценки предложенного метода также рассматривается взаимная информация $I(a_1; x_i = b)$, $i \neq 1$, которая соответствует информации $I(s_1; x_i = b)$ в базовом методе, где a_1 – результат суммирования всех чисел, кроме первого:

$$I(a_1; x_i = b) = \sum_{(a_1, b)} P(a_1, b | x_i = b) \log \frac{P(a_1, b | x_i = b)}{P(a_1)P(b | x_i = b)}. \quad (7)$$

В случае ПрР $P(a_1, b | x_i = b)$ может быть выражено следующим образом:

$$P(a_1, b | x_i = b) = \sum_{\bigoplus_{k=2}^n y_k = a_1} P(b | x_i = b) \prod_{r=2}^n P(y_r, b | x_r = b).$$

В лучшем случае значения, получаемые с помощью (3), должны быть равны значениям, получаемым с помощью (2). Однако из-за сжатия эти значения могут отличаться. Близость (4) и (7) показывает, насколько метод вычитания близок к сумме без одного, а близость (6) к нулю показывает, насколько успешно информация о вычтенном бите удалена. Все вышесказанное приводит к задаче минимизации следующего функционала:

$$L[p(t | y)] = I(s_1; t) + \gamma I(s_1; x_1 = b) - \beta I(s_1; x_i = b, i \neq 1) \rightarrow \min, \quad (8)$$

где $I(s_1, t)$ – взаимная информация между общей суммой t и суммой без одного s_1 .

§ 4. Численные результаты

Рассмотрим регулярную протоматрицу проверок на четность $\{3, 6\}$ размера 1536 на 3072. Каждый переменный узел связан с тремя проверочными узлами, и мощность входных распределений равна $|y_i| = 2^{10} - 1$. Для метода информационного сжатия выбрано $\beta = \infty$, чтобы сохранить максимальное количество релевантной информации.

Как показывает рис. 1, построенная функция вычитания дает взаимную информацию (4), очень близкую к (7), что означает сохранение значительного количества информации об оставшихся битах. Незначительное расхождение наблюдается только при высоком отношении сигнал/шум (ОСШ).

Остаточная информация при низких значениях ОСШ достаточно близка к нулю, но с увеличением ОСШ она быстро растет. На самом деле после каждой итерации декодера уровень шума фактически уменьшается, поэтому предлагаемый алгоритм должен хорошо работать в первых итерациях, а на последних итерациях появляется слишком много информации о вычитаемом бите и точность декодера снижается.

Результаты, продемонстрированные на рис. 2, а, показывают, что эффективность коррекции ошибок декодера ПрР с разработанным методом вычитания практически не отличается от базового декодера без использования разработанной оптимизации сложности.

Рассмотрим предлагаемые декодеры для регулярных протоматриц $\{3, 6\}$, $\{4, 8\}$ и $\{5, 10\}$ с учетом 4- и 5-битных сообщений. Разница в разрядности рассматрива-

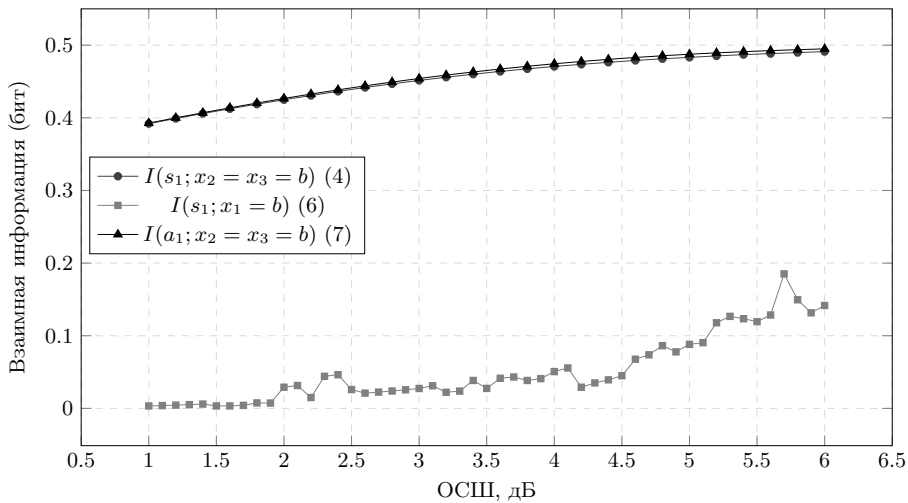


Рис. 1. Анализ взаимной информации

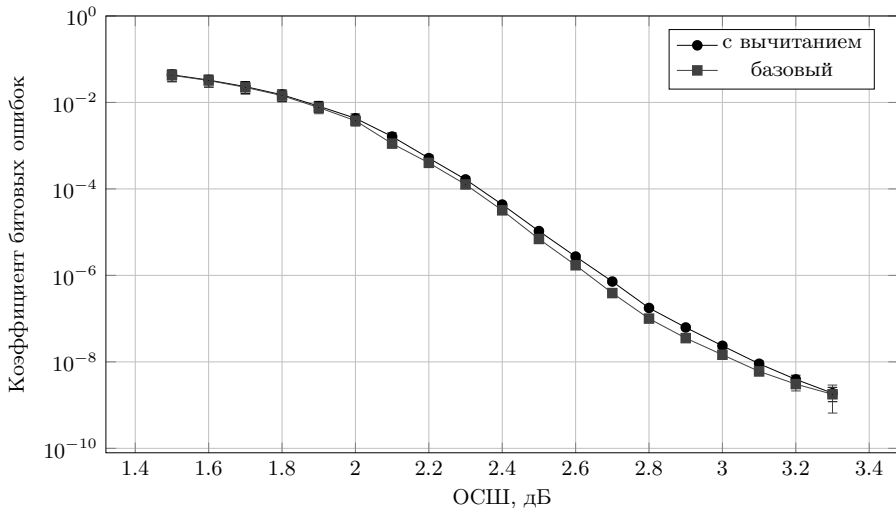
емых декодеров объясняется затратами на память, которые растут с увеличением разрядности сообщений и весов матрицы.

Результаты моделирования для регулярной протоматрицы $\{3, 6\}$ и 5-битных сообщений показаны на рис. 2,б. Предлагаемые и стандартные методы имеют одинаковую производительность исправления ошибок на пороге декодирования, их небольшое различие объясняется чувствительностью метода МИС к точности. Однако, что более важно, предлагаемый метод сопоставим со стандартным и имеет более простую архитектуру. Различие в уровнях ошибок объясняется тем, что “вычитание” с увеличением ОСШ начинает оставлять слишком много остаточной информации о вычитаемом бите, и точность декодера начинает снижаться (см. рис. 1).

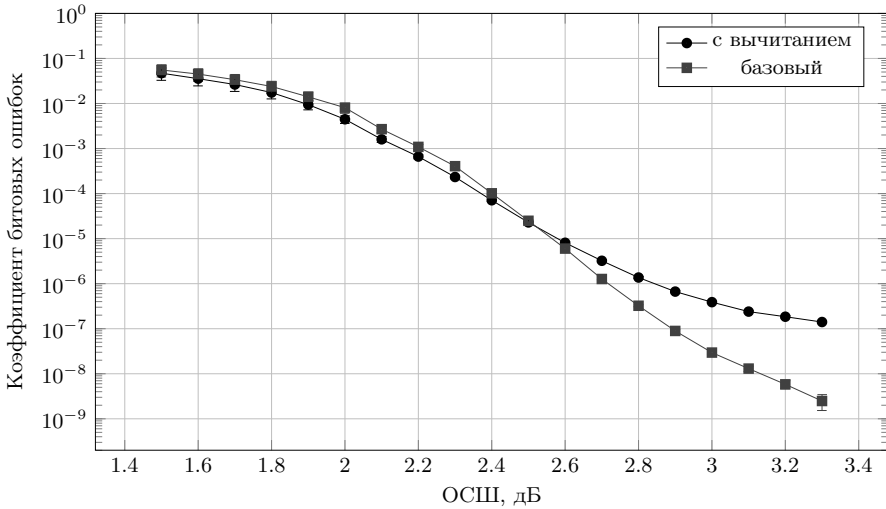
Результаты доказывают эффективность разработанного алгоритма. Однако его использование приводит к экспоненциальной по памяти сложности оптимизации параметров декодера, так как требует хранения $(N + 1)$ -мерных распределений во время работы, где N зависит от весов матрицы проверок на четность. Нахождение алгоритма с меньшей сложностью является задачей дальнейших исследований. Расширенная версия представленных результатов опубликована в работе [10].

§ 5. Заключение

Данное исследование посвящено разработке метода снижения вычислительной и пространственной сложности МПП-декодеров. В статье предложен и оценен метод реализации оптимизации вычитания для уменьшения сложности операций в узлах и количества хранимых таблиц поиска. Предложенный метод вычитания реализован в декодерах с использованием различных типов расписаний: параллельного и последовательного расписания. Это позволило значительно снизить сложность и количество обращений к памяти декодера при потере примерно 0,1 дБ в эффективности коррекции ошибок при использовании последовательного расписания и без потерь при использовании параллельного расписания. Выигрыш в количестве используемых таблиц поиска увеличивается с весом столбца матрицы проверок на четность и для веса 5 составляет 50% для обоих типов расписаний. Однако оптимизация таблицы поиска для вычитания в случае последовательного расписания требует памяти, которая растет экспоненциально с весом столбца матрицы проверки на четность.



а) ПРР, 10 итераций с 5-битными сообщениями с регулярной {3,6} протоматрицей.



б) ПСР, 5 итераций с 5-битными сообщениями с регулярной {3,6} протоматрицей

Рис. 2. Сравнение базового декодера и декодера с предложенной оптимизацией

СПИСОК ЛИТЕРАТУРЫ

1. Угловский А.Ю., Мельников И.А., Алексеев И.А., Куреев А.А. Оценка низкого уровня ошибок с помощью выборки по значимости с равномерным распределением // Пробл. передачи информ. 2023. Т. 59. № 4. С. 3–12. <https://doi.org/10.31857/S0555292323040010>
2. Gallager R. Low-Density Parity-Check Codes // IRE Trans. Inform. Theory. 1962. V. 8. № 1. P. 21–28. <https://doi.org/10.1109/TIT.1962.1057683>
3. Fossorier M.P.C., Mihaljevic M., Imai H. Reduced Complexity Iterative Decoding of Low-Density Parity Check Codes Based on Belief Propagation // IEEE Trans. Commun. 1999. V. 47. № 5. P. 673–680. <https://doi.org/10.1109/26.768759>

4. *Lewandowsky J., Bauch G.* Information-Optimum LDPC Decoders Based on the Information Bottleneck Method // IEEE Access. 2018. V. 6. P. 4054–4071. <https://doi.org/10.1109/ACCESS.2018.2797694>
5. *Lewandowsky J., Bauch G., Stark M.* Information Bottleneck Signal Processing and Learning to Maximize Relevant Information for Communication Receivers // Entropy. 2022. V. 24. № 7. Paper No. 972 (30 pp.). <https://doi.org/10.3390/e24070972>
6. *Stark M., Lewandowsky J., Bauch G.* Information-Bottleneck Decoding of High-Rate Irregular LDPC Codes for Optical Communication Using Message Alignment // Appl. Sci. 2018. V. 8. № 10. Paper No. 1884 (17 pp.). <https://doi.org/10.3390/app8101884>
7. *Kurkoski B.M., Yamaguchi K., Kobayashi K.* Noise Thresholds for Discrete LDPC Decoding Mappings // Proc. 2008 IEEE Global Telecommunications Conf. (GLOBECOM'08). New Orleans, LA, USA. Nov. 30–Dec. 4, 2008. P. 1–5. <https://doi.org/10.1109/GLOCOM.2008.ECP.214>
8. *Фернандес М., Кабатянский Г.А., Круглик С.А., Мяо И.* Коды для точного нахождения носителя разреженного вектора по ошибочным линейным измерениям и их декодирование // Пробл. передачи информ. 2023. Т. 59. № 1. С. 17–24. <https://doi.org/10.31857/S0555292323010023>
9. *He X., Cai K., Song W., Mei Z.* Dynamic Programming for Sequential Deterministic Quantization of Discrete Memoryless Channels // IEEE Trans. Commun. 2021. V. 69. № 6. P. 3638–3651. <https://doi.org/10.1109/TCOMM.2021.3062838>
10. *Melnikov I.A., Uglovskii A.Yu., Kreshchuk A.A., Kureev A.A., Khorov E.M.* Reducing the Complexity of the Layer Scheduled LDPC Decoder Based on the Information Bottleneck Method // Probl. Inf. Transm. 2024. V. 60. № 3. P. 199–208. <https://doi.org/10.1134/S0032946024030049>

Мельников Игнат Алексеевич

Угловский Артем Юрьевич

Крещук Алексей Андреевич

Институт проблем передачи информации

им. А.А. Харкевича РАН, Москва

melnikov@iitp.ru

uglovski@iitp.ru

krsh@iitp.ru

Куреев Алексей Андреевич

Московский физико-технический институт

(национальный исследовательский университет), Москва

Институт проблем передачи информации

им. А.А. Харкевича РАН, Москва

kureev@wireless.iitp.ru

Хоров Евгений Михайлович

Институт проблем передачи информации

им. А.А. Харкевича РАН, Москва

khorov@wireless.iitp.ru

Поступила в редакцию

20.09.2024

После доработки

20.09.2024

Принята к публикации

30.11.2024