

Министерство образования и науки Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ»

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

**«Исследование зависимости вероятности ошибки на блок от спектра
графа Таннера для МППЧ-кодов»**

Автор: Ковшаров Антон Павлович _____

Направление подготовки (специальность): 01.04.02 Прикладная математика и
информатика

Квалификация: Магистр

Руководитель: Буздалов М.В., канд. техн. наук _____

К защите допустить

Зав. кафедрой Васильев В.Н., докт. техн. наук, проф. _____

« ____ » _____ 20 ____ г.

Санкт-Петербург, 2017 г.

Студент Ковшаров А.П. **Группа** М4239 **Кафедра** компьютерных технологий
Факультет информационных технологий и программирования

Направленность (профиль), специализация Технологии проектирования и разработки программного обеспечения

Консультанты:

а) Кудряшов Б.Д., докт. техн. наук, профессор _____

б) Бочарова И.Е., канд. техн. наук, доцент _____

Квалификационная работа выполнена с оценкой _____

Дата защиты _____ «15» июня 2017 г.

Секретарь ГЭК _____

Листов хранения _____

Демонстрационных материалов/Чертежей хранения _____

Министерство образования и науки Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ»

УТВЕРЖДАЮ

Зав. каф. компьютерных технологий
докт. техн. наук, проф.

_____ Васильев В.Н.

« ____ » _____ 20 ____ г.

ЗАДАНИЕ
НА МАГИСТЕРСКУЮ ДИССЕРТАЦИЮ

Студент Ковшаров А.П. **Группа** М4239 **Кафедра** компьютерных технологий
Факультет информационных технологий и программирования **Руководитель** Буздалов
Максим Викторович, канд. техн. наук, научный сотрудник Университета ИТМО

1 Наименование темы: Исследование зависимости вероятности ошибки на блок от спектра графа Таннера для МППЧ-кодов

Направление подготовки (специальность): 01.04.02 Прикладная математика и информатика

Направленность (профиль): Технологии проектирования и разработки программного обеспечения

Квалификация: Магистр

2 Срок сдачи студентом законченной работы: «31» мая 2017 г.

3 Техническое задание и исходные данные к работе.

В рамках работы требуется исследовать зависимость между спектром графа Таннера МППЧ-кода и его эффективностью при декодировании. Для проведения исследования необходимо разработать эффективный алгоритм вычисления спектра, позволяющий провести отбор кодов с хорошим спектром среди широкого спектра сгенерированных кодов. Также необходимо разработать итеративный декодер, позволяющий измерить эффективность кода посредством симуляции передачи кодовых слов через канал с шумом.

4 Содержание магистерской диссертации (перечень подлежащих разработке вопросов)

- а) Обоснование важности установления зависимости между спектром и эффективностью для исследования и отбора МППЧ-кодов;
- б) Разработка и реализация итеративного декодера для быстрой оценки эффективности кода посредством моделирования;
- в) Разработка и реализация алгоритма подсчета спектра графа Таннера;
- г) Описание плана исследования. Порядок отбора кодов для проведения тестирования;
- д) Результаты исследования.

5 Перечень графического материала (с указанием обязательного материала)

Не предусмотрено

6 Исходные материалы и пособия

- а) Б.Д.Кудряшов. Основы теории кодирования;
- б) М.Холл. Комбинаторика;
- в) D.J.C.MacKay. Encyclopedia of Sparse Graph Codes.
<http://www.inference.phy.cam.ac.uk/mackay/codes/data.html>.

7 Календарный план

№№ пп.	Наименование этапов магистерской диссертации	Срок выполнения этапов работы	Отметка о выполнении, подпись руков.
1	Ознакомление с основами теории кодирования	12.2015	
2	Ознакомление с имеющимся набором программ для исследования и отбора МППЧ-кодов	05.2016	
3	Ознакомление с существующими итеративными декодерами	07.2016	
4	Разработка и реализация итеративного декодера заточенного под нужды исследования	09.2016	
5	Ознакомление с существующими подходами подсчета спектра кода	11.2016	
6	Разработка и реализация алгоритма подсчета спектра графа Таннера МППЧ-кода	12.2016	
7	Проведение исследования зависимости эффективности кода от спектра	03.2017	
8	Написание пояснительной записки	05.2017	

8 Дата выдачи задания: «01» сентября 2015 г.

Руководитель _____

Задание принял к исполнению _____ «01» сентября 2015 г.

Министерство образования и науки Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ»

АННОТАЦИЯ
МАГИСТЕРСКОЙ ДИССЕРТАЦИИ

Студент: Ковшаров Антон Павлович

Наименование темы работы: Исследование зависимости вероятности ошибки на блок от спектра графа Таннера для МППЧ-кодов

Наименование организации, где выполнена работа: Университет ИТМО

ХАРАКТЕРИСТИКА МАГИСТЕРСКОЙ ДИССЕРТАЦИИ

1 Цель исследования: Установить существование критерия оценки эффективности МППЧ-кода, основанного на спектре графа Таннера.

2 Задачи, решаемые в работе:

- а) разработка и реализация эффективного алгоритма вычисления спектра графа Таннера;
- б) проведение исследования зависимости эффективности кода от спектра графа Таннера;
- в) формулировка критерия оценки эффективности МППЧ-кода.

3 Число источников, использованных при составлении обзора: _____

4 Полное число источников, использованных в работе: 14

5 В том числе источников по годам

Отечественных			Иностранных		
Последние 5 лет	От 5 до 10 лет	Более 10 лет	Последние 5 лет	От 5 до 10 лет	Более 10 лет

6 Использование информационных ресурсов Internet: _____

7 Использование современных пакетов компьютерных программ и технологий: C++, CUDA C — для создания итеративного декодера. Python и zsh скрипты для автоматизация исследования. Java — алгоритм подсчета спектра. Python, matplotlib, pandas — обработка и визуализация результатов. \LaTeX , Git.

8 Краткая характеристика полученных результатов: В результате была продемонстрирована зависимость между спектром графа Таннера и эффективностью кода. Разработан вычислительно эффективный алгоритм подсчета спектра графа Таннера. Результаты могут быть использованы для ускорения поиска эффективных МППЧ-кодов.

9 Гранты, полученные при выполнении работы: Грантов или других форм государственной поддержки и субсидирования в процессе работы не предусматривалось.

10 Наличие публикаций и выступлений на конференциях по теме работы:

- 1 *Ковшаров А., Кудряшов Б.* Исследование зависимости вероятности ошибки на блок от спектра графа Таннера для МППЧ-кодов // Сборник тезисов докладов конгресса молодых ученых. Электронное издание. — 2017.
- 2 *Ковшаров А., Анохина И.* LDPC-codes frame error rate and Tanner's graph spectrum correlation research // Сборник тезисов докладов конгресса молодых ученых. Электронное издание. — 2017.

Выпускник: Ковшаров А.П. _____

Руководитель: Буздалов М.В. _____

« ____ » _____ 20 ____ г.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	5
1. Общие сведения	6
1.1. Линейные коды	6
1.2. МППЧ-коды	7
1.3. Квазициклические МППЧ-коды	7
2. Декодер	9
3. Алгоритм подсчета спектра графа Таннера	10
3.1. Веса как символьные переменные	15
3.2. Фиксированные веса	15
3.3. Алгоритм	20
3.4. Оценка сложности	20
4. Численный анализ влияния спектров циклов на вероятность ошибки БП-декодирования	22
4.1. Описание ансамблей кодов	22
4.1.1. Ансамбль Галлагера	22
4.1.2. Ансамбль Ричардсона-Урбанке	23
4.1.3. Ансамбль квазициклических кодов	24
4.2. Описание эксперимента	24
ЗАКЛЮЧЕНИЕ	27
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	28

ВВЕДЕНИЕ

Интуиция подсказывает, что большой обхват и малое число коротких циклов положительно сказываются на эффективности итеративного декодирования.

Существуют и другие критерии для поиска хороших МППЧ кодов, например, ACE. В любом случае, все сводится к анализу структуры циклов графа. Поиск кодов, обладающих хорошей структурой – вычислительно сложная задача. Наша задача - упростить вычисления.

Известные подходы

Новый подход имеет простое описание и низкую вычислительную сложность

ГЛАВА 1. ОБЩИЕ СВЕДЕНИЯ

1.1. Линейные коды

Определение 1. Линейным двоичным (n, k) —кодом называется любое k —мерное подпространство пространства всевозможных двоичных векторов длины n .

Определение 2. Отношение $R = k/n$ называется скоростью линейного (n, k) кода.

Определение 3. Порождающей матрицей линейного (n, k) —кода называется матрица размера $k \times n$, строки которой его базисные вектора.

Пример 4.

$$G = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

задает код 000, 101, 100, 001.

Определение 5. Двоичный вектор \mathbf{h} длины n для которого все кодовые слова некоторого (n, k) кода $C = \{\mathbf{c}_i\}, i = 0, \dots, 2^k - 1$ удовлетворяют тождеству

$$(\mathbf{c}_i, \mathbf{h}) = 0, i = 0, \dots, 2^k - 1$$

называется *проверкой* по отношению к коду C .

Заметим, что предыдущее определение проверки эквивалентно

$$G \cdot \mathbf{h}^T = 0$$

так как для выполнения тождества для любого кодового слова достаточно выполнения тождества для базисных векторов.

Определение 6. Пространство проверок называется пространством, ортогональным линейному коду, или *проверочным пространством*.

Теорема 7. Размерность проверочного пространства линейного (n, k) —кода равна $r = n - k$.

Определение 8. Проверочной матрицей линейного (n, k) —кода называется матрица размера $r \times n$, строки которой составляют базис проверочного пространства

Для проверочной и порождающей матрицы выполнено следующее соотношение

$$G \cdot H^T = 0$$

Если же принятая последовательность \mathbf{y} из-за шума в канале перестала быть кодовым словом то соответственно

$$\mathbf{s} = \mathbf{y} \cdot \mathbf{H}^T \neq 0$$

и \mathbf{s} называется *синдромом* вектора \mathbf{y} , неравенство нулевому вектору которого указывает на ошибки в принятой последовательности \mathbf{y} .

1.2. МППЧ-коды

Линейный код может быть задан проверочной матрицей \mathbf{H} .

Галлагер [1] предложил идею выбора матрицы \mathbf{H} разреженной (с малой плотностью) для уменьшения сложности кодирования и декодирования: в матрице должно быть мало единиц, строки и столбцы не должны иметь большое число общих элементов. Он также подкрепил свою идею анализом с использованием метода случайного декодирования.

Для интуитивного понимания почему малое число единиц приводит к более эффективному декодированию следует заметить, что в случае когда строки проверок мало между собой зависят, декодирование может производиться методом проб и ошибок, пытаясь подобрать последовательность символов, исправление которых будет уменьшать вес синдрома, с каждой следующей попыткой.

Используемый алгоритм декодирования, описанный далее, существенно опирается на факт, что влияние конкретного столбца на синдром не сильно зависит от остальных столбцов.

Определение 9. МППЧ-код называется (J, K) *регулярным* если его проверочная матрица \mathbf{H} содержит ровно J единиц в каждом столбце и ровно K единиц в каждой строке. Иначе МППЧ-код называется *нерегулярным*.

1.3. Квазициклические МППЧ-коды

P_M^i — квадратная матрица порядка M , полученная из единичной сдвигом строк вправо i раз. Например:

$$P_3^0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad P_3^1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad P_3^2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Матрицы P_M^i содержат одинаковые элементы на всех диагоналях параллельных главной — такие матрицы называются *циркулянтными матрицами* или *циркулянтами*.

МППЧ-код называется квазициклическим если его проверочная матрица может быть представлена в виде блочной матрицы из блоков циркулянтов единичной матрицы.

Для построения б

ГЛАВА 2. ДЕКОДЕР

Пример ссылок на литературные источники: [2–14].

ГЛАВА 3. АЛГОРИТМ ПОДСЧЕТА СПЕКТРА ГРАФА ТАННЕРА

Эта задача порождена проблемой анализа и оптимизации МППЧ кодов. Пусть B – (двоичная) базовая матрица кода. Для нее определен двудольный граф Таннера $T = \{V, E\}$ с множеством вершин $V = V_s \cup V_c$, где V_s и V_c – множества символьных и проверочных вершин, соответственно. Единицам матрицы B соответствуют ребра графа T .

Пример 10. Рассмотрим базовую матрицу B

$$B(D) = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix}, \quad (1)$$

для задающей квазициклический МППЧ код полиномиальной проверочной матрицы

$$H(D) = \begin{pmatrix} D^{w_{11}} & D^{w_{12}} & 0 & D^{w_{14}} \\ D^{w_{21}} & D^{w_{22}} & D^{w_{23}} & 0 \\ D^{w_{31}} & 0 & D^{w_{33}} & D^{w_{34}} \end{pmatrix} \quad (2)$$

Для удобства переназначим веса переходов

$$H(D) = \begin{pmatrix} D^{w_1} & D^{w_4} & 0 & D^{w_8} \\ D^{w_2} & D^{w_5} & D^{w_6} & 0 \\ D^{w_3} & 0 & D^{w_7} & D^{w_9} \end{pmatrix} \quad (3)$$

Соответствующая матрица инцидентности графа Таннера

$$T(D) = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ D^{w_1} & 0 & 0 & D^{w_4} & 0 & 0 & 0 & D^{w_8} & 0 \\ 0 & D^{w_2} & 0 & 0 & D^{w_5} & D^{w_6} & 0 & 0 & 0 \\ 0 & 0 & D^{w_3} & 0 & 0 & 0 & D^{w_7} & 0 & D^{w_9} \end{pmatrix}$$

Сам граф показан на рис. 2. Кружками и квадратами показаны символьные и проверочные узлы.

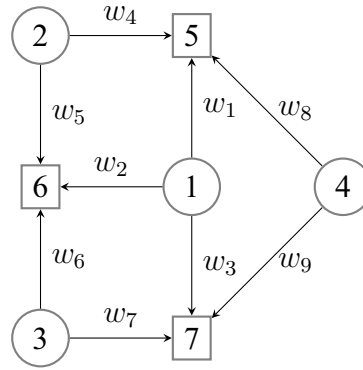


Рисунок 1 – Граф Таннера для кода из примера 10

Задача состоит в подсчете числа циклов заданной длины в расширенном графе с заданной степенью расширения M . Длина цикла равна числу переходов в пути, начинающемся и заканчивающемся в одном и том же узле и таком, что сумма весов переходов по модулю M равна нулю. Веса суммируются с учетом знаков, зависящих от направления перехода (см. рис. 2).

Такой тип циклов часто называют замкнутым обходом, однако при подсчете рассматриваемых объектов необходимо учесть следующие дополнительные ограничения.

- Запрещено двигаться обратно по последнему пройденному ребру. Например, путь $6 \rightarrow 1 \rightarrow 6$ веса 0 запрещен.
- Циклические сдвиги путей и инверсии путей должны учитываться как один цикл. Например, путь $p = 6 \rightarrow 1 \rightarrow 7 \rightarrow 3 \rightarrow 6$ образует цикл при условии $w_2 - w_6 + w_7 - w_3 = 0$. При этом пути $6 \rightarrow 3 \rightarrow 7 \rightarrow 1 \rightarrow 6$ и $1 \rightarrow 7 \rightarrow 3 \rightarrow 6 \rightarrow 1$ тоже циклы, но они уже не вносят вклад в число циклов длины 4, если цикл p учтен.

Хотелось бы применить стандартные методы, используемые при анализе систем на основе конечных автоматов. Данный граф не является конечным автоматом, поскольку перемещение из состояния в состояние зависит от предыдущего состояния. Например, на рис. 2 после состояния 4 возможно только 5, если предыдущим было 7 и, наоборот, только 7, если предыдущим было 5.

Чтобы свести задачу к анализу конечных автоматов, введем новое множество состояний $U = \{e, \xi\}$, где e задает ребро исходного графа, а ξ – направление перехода. Сокращенно будем записывать пары в виде $+e$ и $-e$.

Из графа на рис. 2 получится граф с 18 состояниями $\{\pm 1, \pm 2, \dots, \pm 9\}$. Заметим, однако, что после отрицательного ребра следуют только положитель-

ные и после положительного отрицательные. Это позволит записать матрицу переходов компактно в виде двух матриц, матрицы положительных и матрицы отрицательных переходов. Например, следующими состояниями (ребрами графа Таннера) после положительного перехода $+1$ возможны отрицательные $-4, -8$. После отрицательного перехода -1 возможны положительные $2, 3$.

В нашем примере две матрицы переходов имеют вид

$$A_- = \begin{pmatrix} 0 & 0 & 0 & D^{-w_4} & 0 & 0 & 0 & D^{-w_8} & 0 \\ 0 & 0 & 0 & 0 & D^{-w_5} & D^{-w_6} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & D^{-w_7} & 0 & D^{-w_9} \\ D^{-w_1} & 0 & 0 & 0 & 0 & 0 & 0 & D^{-w_8} & 0 \\ 0 & D^{-w_2} & 0 & 0 & 0 & D^{-w_6} & 0 & 0 & 0 \\ 0 & D^{-w_2} & 0 & 0 & D^{-w_5} & 0 & 0 & 0 & 0 \\ 0 & 0 & D^{-w_3} & 0 & 0 & 0 & 0 & 0 & D^{-w_9} \\ D^{-w_1} & 0 & 0 & D^{-w_4} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & D^{-w_3} & 0 & 0 & 0 & D^{-w_7} & 0 & 0 \end{pmatrix}$$

$$A_+ = \begin{pmatrix} 0 & D^{w_2} & D^{w_3} & 0 & 0 & 0 & 0 & 0 & 0 \\ D^{w_1} & 0 & D^{w_3} & 0 & 0 & 0 & 0 & 0 & 0 \\ D^{w_1} & D^{w_2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & D^{w_5} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & D^{w_4} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & D^{w_7} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & D^{w_6} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & D^{w_9} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & D^{w_8} & 0 \end{pmatrix}$$

Заметим, что все циклы имеют четную длину и состоят из пар переходов (по стрелке, против стрелки). Можно построить матрицу переходов за два

шага, как произведение матриц A_+, A_-

$$A = A_+ A_- = \begin{pmatrix} 0 & 0 & 0 & 0 & \omega_{45} & 0 & 0 & 0 & \omega_{89} \\ 0 & 0 & 0 & \omega_{54} & 0 & 0 & \omega_{67} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \omega_{76} & 0 & \omega_{98} & 0 \\ 0 & \omega_{12} & \omega_{13} & 0 & 0 & 0 & 0 & 0 & \omega_{89} \\ \omega_{21} & 0 & \omega_{23} & 0 & 0 & 0 & \omega_{67} & 0 & 0 \\ \omega_{21} & 0 & \omega_{23} & \omega_{54} & 0 & 0 & 0 & 0 & 0 \\ \omega_{31} & \omega_{32} & 0 & 0 & 0 & 0 & 0 & \omega_{98} & 0 \\ 0 & \omega_{12} & \omega_{13} & 0 & \omega_{45} & 0 & 0 & 0 & 0 \\ \omega_{31} & \omega_{32} & 0 & 0 & 0 & \omega_{76} & 0 & 0 & 0 \end{pmatrix}$$

где $\omega_{ij} = D^{-w_i+w_j}$.

В общем случае матрица переходов за 2 шага будет иметь размер, равный числу единиц в базовой матрице B .

Для подсчета производящей функции числа путей длины $2L$, начинающихся с ребра 1 в положительном направлении нужно начальный вектор $a_0 = (1, 0, \dots, 0)$ умножить на A^L . Для нашего примера при

$$a_2 = a_0 A = \begin{pmatrix} 0 & 0 & 0 & 0 & \omega_{45} & 0 & 0 & 0 & \omega_{89} \end{pmatrix} \quad (4)$$

$$a_4 = a_2 A = \begin{pmatrix} \omega_{4521,8931} & \omega_{8932} & \omega_{4523} & 0 & 0 & \omega_{8976} & \omega_{4567} & 0 & 0 \end{pmatrix} \quad (5)$$

где первая компонента является сокращенной записью полинома

$$\omega_{4521,8931} = D^{-w_4+w_5-w_2+w_1} + D^{-w_8+w_9-w_3+w_1}$$

Кроме того необходимо отметить:

- Соблюдение условия неповторения ребра на стыке цикла также соблюдается.
- Хотя мы пишем $+$, $-$ в выражениях типа $D^{-w_4+w_5-w_2+w_1}$, w_i не коммутируют между собой так как неупорядоченный набор ребер не задает однозначно цикл (11), таким образом до подстановки конкретных значений w_i не могут быть сложены.

Пример 11. Обозначим пути $c_1 = 1 \rightarrow -4 \rightarrow 5 \rightarrow -2$, $c_2 = 1 \rightarrow -4 \rightarrow 6 \rightarrow -3$, $c_3 = 2 \rightarrow -5 \rightarrow 6 \rightarrow -3$. Обратные к этим путям по

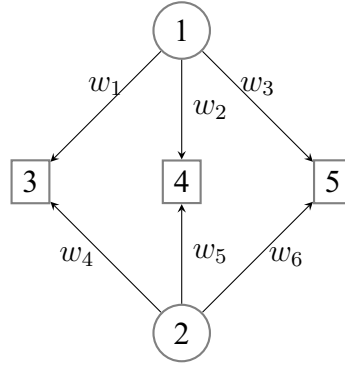


Рисунок 2 – Граф Таннера для кода из примеров 11, 12

направлению соответственно c_{-1}, c_{-2}, c_{-3} . Тогда набор ребер как объединением $\{c_1, c_{-2}, c_{-3}\}$ может обозначать два пути – $c_1 c_{-3} c_{-2}$ и $c_1 c_{-2} c_{-3}$, которые не могут быть получены друг из друга с помощью циклического сдвига и инверсии.

Из (9) видим, что существуют 2 потенциальных цикла длины 4. Цикл образуется в том случае, когда сумма в показателе степени равна нулю. Например, если $-w_4 + w_5 - w_2 + w_1 = 0$, а $-w_8 + w_9 - w_3 + w_1 \neq 0$ то

$$a_{41}(D) = 1 + D^{-w_8+w_9-w_3+w_1},$$

и $a_{41}(0) = 1$. Пусть $a_{2L,i}^j$ – обозначает коэффициент при D^j в i -ой компоненте a_{2L} .

(здесь и далее считаем, что $D^0 = 1$, несмотря на то что D может быть 0. Таким образом свободный член полинома равен количеству циклов веса 0 без учета эквивалентных циклов) В общем случае все циклы длины $2L$, проходящие через ребро $+1$ учтены в $a_{2L,1}(0)$, однако некоторые из них могут быть учтены многократно (12). Кроме того необходимо аналогично учесть циклы, проходящие через ребро $+2, +3$ и так далее. Такое суммарное рассмотрение неизбежно будет учитывать циклы повторно, однако количество повторений не обязательно равно длине цикла и зависит от его состава.

Пример 12. Пути $1 \rightarrow -4 \rightarrow 5 \rightarrow -2 \rightarrow 1 \rightarrow -4 \rightarrow 6 \rightarrow -3$ и $1 \rightarrow -4 \rightarrow 6 \rightarrow -3 \rightarrow 1 \rightarrow -4 \rightarrow 5 \rightarrow -2$, начинающиеся с 1 эквивалентны, так как являются циклическим сдвигом друг друга, но будут учтены дважды.

Дальнейшие рассуждения можно проводить двумя способами:

- Рассматривать w_i в качестве символьных переменных. Можно выписать все возможные комбинации весов, приводящие к потенциальным циклам

длины $2L$ и затем вычислять спектры циклов, подставляя разные наборы разметок в полученные уравнения. Число циклов равно числу нулей в уравнениях.

- При заданных (фиксированных) весах ребер подсчитать спектр длин циклов в заданном диапазоне.

3.1. Веса как символьные переменные

В случае рассмотрения w_i как символьных переменных необходимо избавиться от эквивалентных относительно сдвига и инверсии комбинаций. Для этого каждый путь приводится к минимальному лексикографическому виду после чего убираются дубликаты.

Заметим, что цикл в обратном направлении к рассматриваемому циклу не может быть эквивалентен относительно сдвига исходному. Обозначим \bar{p} путь в обратном направлении к пути $p = e_1, e_2, \dots, e_n$. Предположим p и \bar{p} эквивалентны относительно сдвига – тогда найдется индекс i , такой что $\bar{e}_i, \bar{e}_{i-1}, \dots, \bar{e}_1 = e_1, e_2, \dots, e_i$. Если i нечетно, тогда $\bar{e}_{(i+1)/2} = e_{(i+1)/2}$ – противоречие. Если i четно, то $e_{i/2+1} = \bar{e}_{i/2}$, что противоречит ограничению что путь не может идти обратно по последнему пройденному ребру.

Таким образом каждому циклу соответствует ровно два пути в разных направлениях. Непосредственно инвертируя каждый путь несложно оставить ровно один из каждой пары.

По причине того, что число различных путей растет экспоненциально при построении всех возможных символьных комбинаций, лучше воспользоваться подходом *meet-in-the-middle*. Это позволяет для нахождения всех возможных циклов длины $2L$ рассматривать только пути длины L , в то время как описанный алгоритм рассматривает пути длины $2L$.

3.2. Фиксированные веса

В случае когда веса изначально зафиксированы можно считать, что веса коммутируют, поэтому достаточно считать не более M членов в каждом полиноме при умножении на матрицу A , по одному значению для каждой возможной суммы весов. После чего устранить дубликаты из результата с помощью обращения Мебиуса.

Рассмотрим для цикла p сколько раз он был учтен. Порядком цикла назовем максимальное r , такое что p можно представить как

$$p = \underbrace{ss \dots s}_{r \text{ раз}}$$

Периодом цикла p назовем длину $|s| = \frac{|p|}{r}$. Таким образом цикл p периода l имеет l различных циклических сдвигов – следовательно будет учтен $2l$ раз, с учетом последовательностей в обратном порядке.

Последовательно для каждого i зафиксируем

$$a_0 = (\underbrace{0, 0, \dots, 0}_{i \text{ раз}}, 1, 0, \dots, 0)$$

и вычислим

$$a_{2L} = a_0 A^L$$

проводя все вычисления в кольце многочленов по модулю D^M .

Запомним количество циклов (всех возможных весов) проходящих через $+i$, содержащееся в многочлене $a_{2L,i}$ как $b_{2L,i}$.

Суммируя по всем возможным a_0 получаем

$$b_{2L} = \sum_i b_{2L,i}$$

- многочлен, с коэффициентами при D^w соответственно равными количеству циклов длины $2L$ и веса w , где каждый цикл учтен дважды столько, сколько он имеет различных циклических сдвигов (в обоих направлениях).

Далее временно забудем об ограничении инверсии, так как для устранения путей эквивалентных относительно разворота достаточно разделить результирующий спектр на два.

Обозначим за $g(l)$ – многочлен, коэффициент при D^w которого равен числу циклов длины и периода l веса w .

Определение 13. Введем операцию $T_d(p(D))$ где $p(D)$ – многочлен, а d – натуральное число как:

$$T_d(c_0 + c_1 D + c_2 D^2 + \dots + c_n D^n) = c_0 + c_1 D^d + c_2 D^{2d} + \dots + c_n D^{nd} \pmod{D^M}$$

Тогда при фиксированной длине L имеем равенство:

$$\sum_{d|L} d \cdot T_{\frac{L}{d}}(g(d)) = b_L$$

Равенство справедливо, так как каждый цикл длины L периода d веса W состоит из повторенного $\frac{L}{d}$ раз цикла длины и периода d веса w , такого что $w \cdot \frac{L}{d} = W \pmod{M}$ и однозначно им задается. Каждый такой цикл имеет d различных циклических сдвигов, поэтому учтен в b_L d раз. Таким образом суммирование ведется по всем возможным длинам периодов, после чего благодаря $T_{\frac{L}{d}}$ вес каждого цикла из $g(d)$ домножается на число повторений цикла для достижения длины L и каждый из циклов учитывается с коэффициентов d так как входит в b_L в виде d различных линейных последовательностей ребер.

В равенствах вида

$$f(n) = \sum_{d|n} g(d)$$

$g(d)$ может быть выражено с помощью формулы обращения Мебиуса:

$$g(n) = \sum_{d|n} \mu(d) f(n/d)$$

где

$$n = p_1^{e_1} \cdot p_2^{e_2} \cdot \dots \cdot p_r^{e_r}$$

- разложение n на простые множители

$$\mu(n) = \begin{cases} 1 & \text{при } n = 1, \\ 0 & \text{если } \exists i \quad e_i > 1, \\ (-1)^r & \text{если } e_1 = e_2 = \dots = e_r = 1 \end{cases}$$

В рассматриваемом случае можно ввести и использовать следующее обобщение формулы Мебиуса:

Теорема 14. Если

$$f(n) = \sum_{d|n} T_{\frac{n}{d}}(g(d))$$

где T_k удовлетворяет свойствам:

$$T_k(c \cdot p(D)) = c \cdot T_k(p(D)) \quad (6)$$

$$T_k(p(D) + q(D)) = T_k(p(D)) + T_k(q(D)) \quad (7)$$

$$T_{k_1}(T_{k_2}(p(D))) = T_{k_1 \cdot k_2}(p(D)) \quad (8)$$

$$T_1(p(D)) = p(D) \quad (9)$$

Тогда $g(d)$ может быть выражено:

$$g(n) = \sum_{d|n} \mu(d) T_d(f(n/d))$$

Доказательство.

Лемма 15.

$$\sum_{d|n} \mu(d) = 0 \quad \text{при } n > 1$$

Доказательство.

$$n = p_1^{e_1} \cdot p_2^{e_2} \cdot \dots \cdot p_r^{e_r}$$

$$n^* = p_1 \cdot p_2 \cdot \dots \cdot p_r$$

$$\sum_{d|n} \mu(d) = \sum_{d|n^*} \mu(d)$$

так как любое $\mu(d)$ где $p_i^2 | d$ равно 0.

$$\sum_{d|n^*} \mu(d) = 1 - \binom{r}{1} + \binom{r}{2} - \dots + (-1)^k \cdot \binom{r}{k} + \dots = (1 - 1)^r = 0$$

так как существует $\binom{r}{k}$ делителей n^* состоящих из k простых множителей, каждый из которых внесет вклад $\mu(d) = (-1)^k$.

$$\begin{aligned} \sum_{d|n} \mu(d) \cdot T_d(f(n/d)) &= \\ \sum_{d|n} \mu(d) \cdot T_d\left(\sum_{d'|\frac{n}{d}} T_{\frac{n}{dd'}}(g(d'))\right) &= \\ \sum_{d|n} \mu(d) \cdot \sum_{d'|\frac{n}{d}} T_d(T_{\frac{n}{dd'}}(g(d'))) &= \end{aligned}$$

$$\begin{aligned}
& \sum_{d|n} \mu(d) \cdot \sum_{d'|\frac{n}{d}} T_{\frac{n}{d}}(g(d')) = \\
& \sum_{d|n} \sum_{d'|\frac{n}{d}} \mu(d) \cdot T_{\frac{n}{d}}(g(d')) = \\
& \sum_{d'|n} \sum_{d|\frac{n}{d'}} \mu(d) \cdot T_{\frac{n}{d}}(g(d')) = \\
& \sum_{d'|n} T_{\frac{n}{d'}}(g(d')) \cdot \sum_{d|\frac{n}{d'}} \mu(d) = \\
& T_{\frac{n}{n}}(g(n)) = T_1(g(n)) = g(n)
\end{aligned}$$

Нетрудно заметить что операция $T_d(p(D))$ из определения 13 удовлетворяет ограничениям из теоремы 14. Действительно, первые два свойства следуют из линейности кольца многочленов по модулю. Третье и четвертое свойства очевидно следуют из определения операции 13.

Таким образом, можно воспользоваться теоремой 14 для разрешения выражения 3.2 относительно $g(L)$. В качестве $f(L)$ выступает b_L , а в качестве $g(d)$ необходимо взять $d \cdot g(d)$, внеся множитель d внутрь операции T_d , согласно свойству один:

$$\begin{aligned}
L \cdot g(L) &= \sum_{d|L} \mu(d) T_d(b_{L/d}) \\
g(L) &= \frac{\sum_{d|L} \mu(d) T_d(b_{L/d})}{L}
\end{aligned}$$

Напомним, что $g(L)$ содержит многочлен, коэффициент при D^w которого равен числу циклов длины и периода L веса w . Обозначим результирующую величину количества циклов длины L веса 0 с учетом эквивалентности относительно сдвига за C_L . Для подсчета величины C_L с помощью $g(L)$ необходимо произвести суммирование по всем возможным длинам периодов d в циклах длины L и весам w таким, что период повторенный $\frac{L}{d}$ раз приведет к циклу нулевого веса:

$$C_L = \sum_{d|L} \sum_{\frac{w \cdot L}{d} \equiv 0 \pmod{M}} g^{(w)}(0)$$

где $g^{(w)}$ – w -ая производная g , используемая для получения коэффициента при D^w .

Наконец для устранения дубликатов относительно инверсии достаточно разделить C_L на два.

3.3. Алгоритм

$$a_0^i = (\underbrace{0, 0, \dots, 0}_{i \text{ раз}}, 1, 0, \dots, 0) \quad (10)$$

$$a_{2L}^i = a_0^i \cdot A^L \quad (11)$$

$$b_{2L} = \sum_i a_{2L}^i \quad (12)$$

$$g(L) = \frac{\sum_{d|L} \mu(d) T_d(b_{L/d})}{L} \quad (13)$$

$$C_L = \sum_{d|L} \sum_{\substack{w \cdot L \\ d} = 0 \pmod{M}} g^{(w)}(0) \quad (14)$$

3.4. Оценка сложности

Пусть исходная базовая матрица имела размер $b \times c$, а коэффициент расширения равен M . Для простоты будем рассматривать (J, K) -регулярный МППЧ код с весом столбцов и строк J и K соответственно.

Тогда число ребер в графе Таннера обозначим за $E = b \cdot K = c \cdot J$. Ограничение на максимальную длину в спектре обозначим S .

Таким образом размер матрицы A составляет $E \times E$, каждый из элементов которой представляет собой многочлен степени не больше M . И для получения всех интересующих A^L , посредством перемножения матриц, достаточно затратить $O(S \cdot E^3 \cdot M)$ операций, так как перемножение и сложение многочленов по модулю D^M требует $O(M)$ времени в отличии от $O(1)$ для обычных чисел.

При каждом фиксированном стартовом ребре i и длине циклов $2L$, многочлен $a_{2L}^i = a_0^i \cdot A^L$ 11 может быть получен за время $E^2 \cdot M$, посредством перемножения вектора a_0 длины E на матрицу A^L размера $E \times E$, опять же по причине того что перемножение и сложение многочленов по модулю требует $O(M)$ операций. Таким образом суммарно получение всех необходимых a_{2L} займет время $O(E \cdot S \cdot E^2 \cdot M) = O(S \cdot E^3 \cdot M)$.

Сложение всех необходимых a_{2L} для получения b_{2L} (12) может быть осуществлено за суммарный размер полиномов a , а именно $O(E \cdot S \cdot M)$.

Все необходимые значения функции Мебиуса $\mu(d)$ могут быть подсчитаны за время $O(S \cdot \ln S)$ с помощью решета Эратосфена. При оценке времени суммарно затраченного на подсчет $g(L)$ заметим, что суммирование ведется по всем делителям чисел L от 1 до S . Как известно суммарное количество делителей чисел от 1 до n имеет порядок $O(n \cdot \ln n)$. Действительно, число d является делителем для $\lfloor \frac{n}{d} \rfloor$ чисел: $d, 2d, 3d, \dots, \lfloor \frac{n}{d} \rfloor \cdot d$. Получаем сумму гармонического ряда $\sum_d \lfloor \frac{n}{d} \rfloor = O(n \cdot \ln n)$. Таким образом для подсчета всех $g(L)$ необходимо затратить $O(M \cdot S \cdot \ln S)$, так как операция T_d и сложение многочленов степени M может быть осуществлено за $O(M)$.

Пользуясь оценкой суммы гармонического ряда для 14 с учетом суммирования по весам получаем время необходимое для подсчета C_L – $O(M \cdot S \cdot \ln S)$, так как теперь складываются коэффициенты многочлена, а не целые многочлены.

Итого по всем шагам алгоритма получаем:

$$O(S \cdot E^3 \cdot M) + O(S \cdot E^3 \cdot M) + O(E \cdot S \cdot M) + O(M \cdot S \cdot \ln S) + O(M \cdot S \cdot \ln S) =$$

$$O(S \cdot E^3 \cdot M) + O(M \cdot S \cdot \ln S) = O(M \cdot S \cdot \max(\ln S, E^3))$$

Очевидно для всех разумных входных данных член E^3 мажорирует $\ln S$ таким образом итоговая сложность:

$$O(M \cdot S \cdot E^3)$$

ГЛАВА 4. ЧИСЛЕННЫЙ АНАЛИЗ ВЛИЯНИЯ СПЕКТРОВ ЦИКЛОВ НА ВЕРОЯТНОСТЬ ОШИБКИ БП-ДЕКОДИРОВАНИЯ

Общий подход к генерации случайного регулярного МППЧ-кода состоит из следующих шагов:

- а) Выбор весов столбцов и строк базовой матрицы (J, K) . В исследовании будут рассмотрены $(3, 6)$ и $(4, 8)$ -регулярные МППЧ-коды, как наиболее применяемые на практике.
- б) Выбор размера базовой матрицы (n_b, r_b) . Размер должен удовлетворять соотношению $J \cdot n_b = K \cdot r_b$ (два способа подсчитать число единиц в матрице — по строкам и по столбцам). В исследовании рассмотрен размер $(24, 12)$, который удовлетворяет $(3, 6)$ и $(4, 8)$ с коэффициентам 4 и 3 соответственно.
- в) Выбор базовой матрицы из фиксированного ансамбля кодов. В исследовании рассмотрены ансамбли Галлагера, Ричардсона-Урбанке и Квази-циклических кодов.
- г) Выбор размера результирующего кода после лифтинга. В исследовании рассмотрены размеры 576 и 2304, используемые в стандарте 802.16e WiMAX. Коэффициенты лифтинга $M = 24$ и $M = 96$ соответственно.
- д) Разметка базовой матрицы весами по модулю M . Все разметки приняты равновероятными.

4.1. Описание ансамблей кодов

Генерация случайных матриц заданного размера с фиксированным числом единиц в строках и столбцах для задания регулярного МППЧ-кода может производиться различными способами. При проведении тестирования были рассмотрены следующие ансамбли кодов.

4.1.1. Ансамбль Галлагера

Матрицы в ансамбле Галлагера состоят из полос с фиксированным числом строк в каждой. Каждый столбец полосы содержит ровно одну единицу. Таким образом число полос равно весу столбца.

Например, рассмотрим (3,6)-код, $M = 4$. Такой код состоит из 6 полос, каждая из которых состоит из $M = 4$ строк. Первая строка имеет вид

[illegible]

Остальные $M - 1$ строк этой полосы — сдвиги первой строки на 6 позиций. Таким образом строится первая полоса. Оставшиеся 2 полосы — случайные перестановки первой полосы.

В результате получен (24,12)-код, он же (3,6)-регулярный МППЧ-код.

4.1.2. Ансамбль Ричардсона-Урбанке

В ансамбле Ричардсона-Урбанке все (3,6)-регулярные коды равновероятные.

Рассмотрим способ на примере того же (24,12)-кода. Возьмем последовательность (номера строк единиц):

$$\left(1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 2 \ 2 \ 2 \ 2 \ 2 \ 2 \ 3 \ 3 \ 3 \ 3 \ 3 \ 3 \dots 12 \ 12 \ 12 \ 12 \ 12 \ 12 \right)$$

Возьмем случайную перестановку чисел этой последовательности. Берем первые $J = 3$ числа, скажем 7,3,11. Они указывают номера строк единиц первого столбца. Берем еще 3 числа и находим второй столбец и так далее.

Этот способ не гарантирует регулярность кода, так как на одной позиции может оказаться две единицы и таким образом столбец будет содержать меньше чем J строк с единицами. Это оказалось существенно при проведении экспериментов. Матрицы в которых столбец содержит меньше J строк с единицами более разряженные и соответственно их граф Таннера содержит меньше циклов. Таким образом такие матрицы становятся лучшими согласно спектру, но вполне ожидаемо показывают довольно плохие результаты при моделировании.

Таким образом был рассмотрен немного модифицированный ансамбль Ричардсона-Урбанке, который контролировал регулярность кода, отбрасывая иррегулярные МППЧ-коды. Так как регулярные коды составляют значительную часть кодов из ансамбля Ричардсона-Урбанке, данная модификация незначительно увеличивает время генерации.

4.1.3. Ансамбль квазициклических кодов

4.2. Описание эксперимента

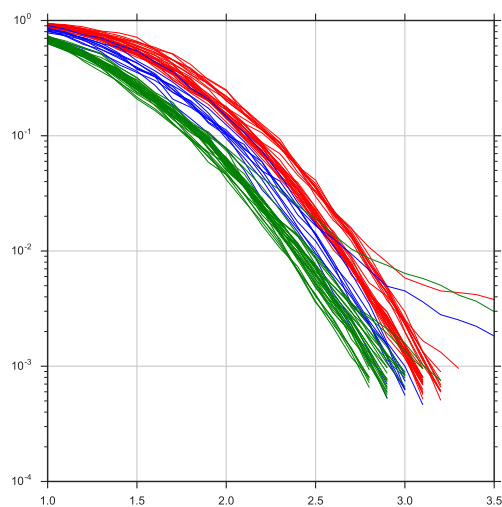


Рисунок 3 – Ричардсон-Урбанке 4x8 576

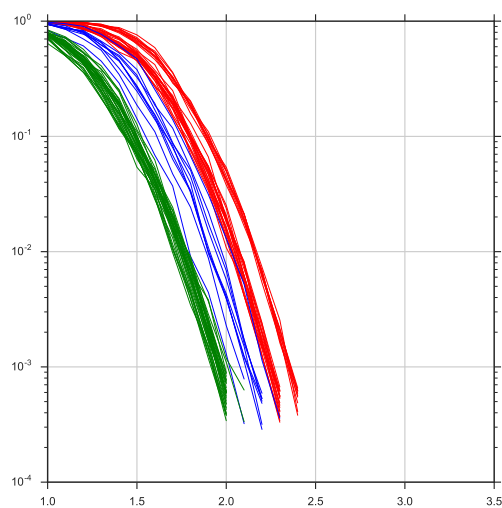


Рисунок 4 – Ричардсон-Урбанке 4x8 2304

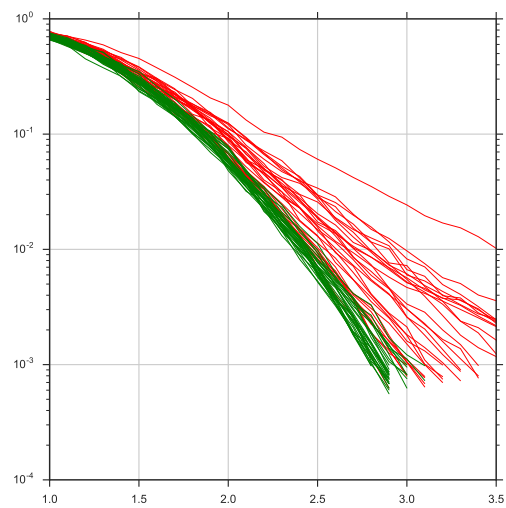


Рисунок 5 – Галлагер 3x6 576

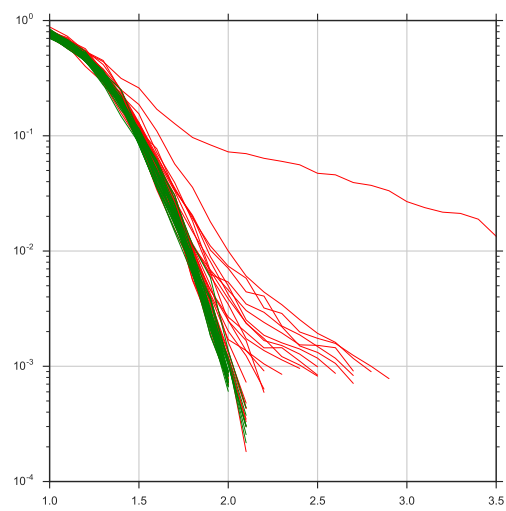


Рисунок 6 – Галлагер 3x6 2304

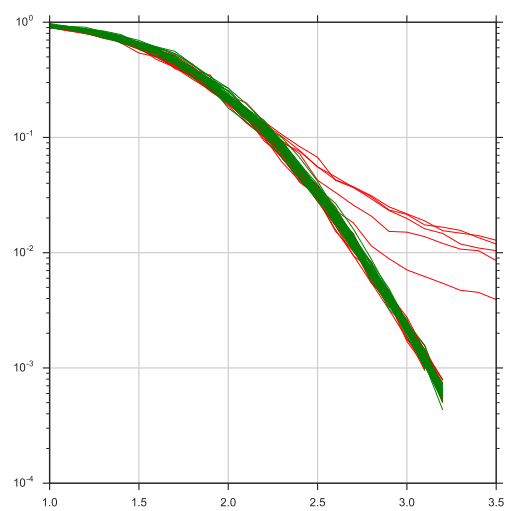


Рисунок 7 – Галлагер 4x8 576

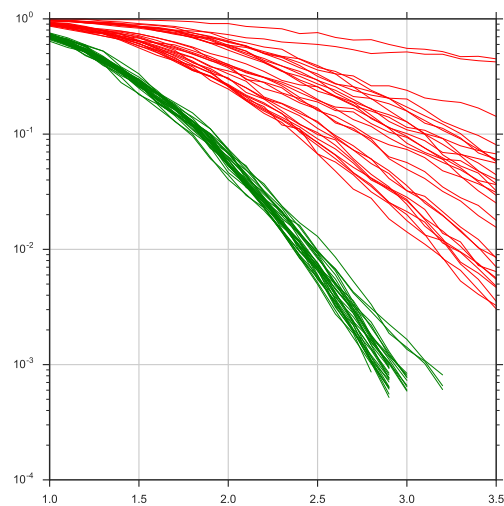


Рисунок 8 – Квазициклические коды 3x6 576

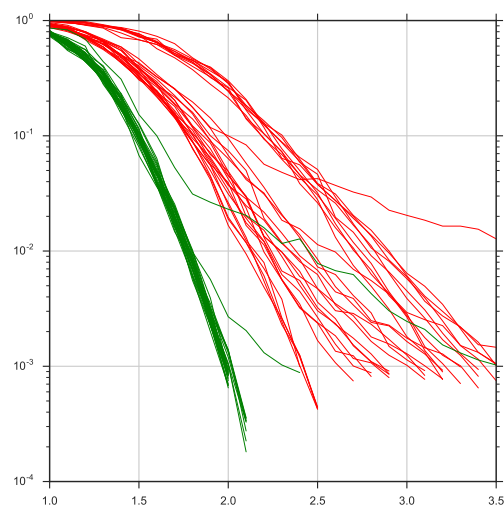


Рисунок 9 – Квазициклические коды 3x6 2304

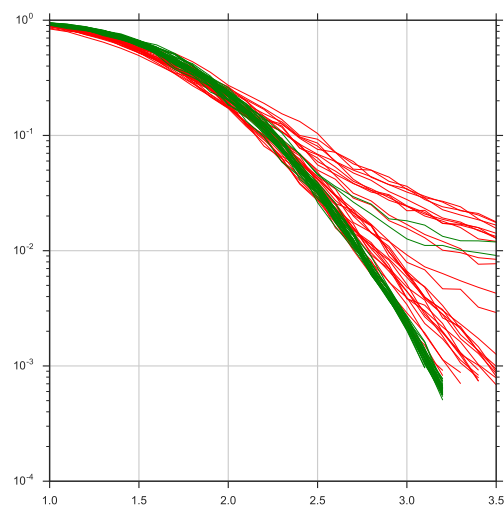


Рисунок 10 – Квазициклические коды 4x8 576

ЗАКЛЮЧЕНИЕ

- а) Получены экспериментальные доказательства прямой зависимости между количеством коротких циклов и вероятностью ошибки на блок для различных ансамблей кодов
- б) Разработан вычислительно эффективный алгоритм подсчета спектра графа Таннера, который может быть использован для ускорения отбора кодов.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 *Gallager R.* Low-Density Parity-Check Codes. — 1963.
- 2 *Холл М.* Комбинаторика. — Издательство "Мир"Москва, 1970.
- 3 *Кудряшов Б.* Основы теории кодирования. — БХВ-Петербург, 2016.
- 4 *Alon N., Yuster R., Zwick U.* Finding and counting given length cycles // *Algorithmica*. — 1997. — P. 209–223.
- 5 *Karimi M., Banihashemi A.* Counting short cycles of quasi cyclic protograph LDPC-codes // *IEEE Communications Letters*. — 2012. — Mar. — Vol. 16, no. 3. — P. 400–403.
- 6 *Karimi M., Banihashemi A.* A message-passing algorithm for counting short cycles in graph. — 2013.
- 7 *Monien B.* How to find long paths efficiently // *Annals of Discrete Mathematics*. — 1985. — No. 25. — P. 239–254.
- 8 *Alon N., Yuster R., Zwick U.* Color-coding. — 1995.
- 9 *Halford T., Chugg K.* An algorithm for counting short cycles in bipartite graphs // *IEEE Transactions on Information Theory*. — 2006. — Jan. — Vol. 52, no. 1.
- 10 *Broulim J., Ayriyan A., Georgiev V.* OpenCL/CUDA algorithms for parallel decoding of any irregular LDPC code using GPU // *Journal of L^AT_EXclass files*. — 2015. — Aug. — Vol. 14, no. 8.
- 11 Stressing the BER simulation of LDPC codes in the error floor region using GPU clusters / G. Falcao [et al.]. — 2013.
- 12 *Harary F., Manvel B.* On the number of cycles in a graph // *Matematicky casopis*. — 1971. — Vol. 21, no. 1. — P. 55–63.
- 13 *Yedidia J., Freeman W., Weiss Y.* Understanding Belief Propagation and its Generalizations. — 2001. — Ноябрь.
- 14 *MacKay D.* Encyclopedia of Sparse Graph Codes. — URL: <http://www.inference.phy.cam.ac.uk/mackay/codes/data.html> (visited on 05/08/2017).