МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ М. В. ЛОМОНОСОВА
МЕХАНИКО–МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

На правах рукописи

**Козачинский Александр Николаевич**

# Сравнение коммуникационной, информационной и вопросной сложности

01.01.06 — математическая логика, алгебра и теория чисел

Диссертация
на соискание учёной степени
кандидата физико-математических наук

Научный руководитель
доктор физико-математических наук
профессор Н. К. Верещагин

Москва — 2019

# Оглавление

# Введение

## Актуальность темы и степень ее разработанности

Диссертация посвящена различным вопросам в коммуникационной сложности. Исследуется как коммуникационная сложность конкретных функций, так и связь коммуникационной сложности с другими сложностными мерами — а именно, с информационной и вопросной сложностью. Коммуникационная сложность была введена в работе Яо [43] в 1979 году. В модели коммуникационной сложности есть два игрока, Алиса и Боб. С каждым из игроков связано конечное множество, элемент которого игрок получает на вход. При этом Алиса не видит вход Боба, а Боб не видит входа Алисы. Игрокам нужно ответить на какой-то вопрос о своих входах, ответ на который зависит как от входа Алисы, так и от входа Боба, что делает невозможным ответить какому-то игроку на вопрос самостоятельно. Например, Алиса и Боб могут хотеть вычислить значение известной игрокам функции $f$, принимающей два аргумента, первый из которых — вход Алисы, а второй — вход Боба.

Для этого у Алисы и Боба есть канал коммуникации, по которому они могут общаться, посылая друг другу битовые сообщения. Алгоритмы их общения в этой науке называются коммуникационными протоколами. Для Алисы и Боба нужно придумать такой протокол, чтобы в конце общения, основываясь на полученной друг от друга информации, они могли ответить на вопрос, например, выдать значение $f$.

Алиса и Боб доверяют друг другу, при этом они стараются минимизировать количество бит, которое они друг другу посылают. Подчеркнем, что Алисе и Бобу даются неограниченные вычислительные ресурсы — можно представлять себе, что передача по каналу намного дороже затрат на локальные вычисления.

В простейшей модели, называемой детерминированной коммуникационной сложностью, Алиса и Боб должны всегда правильно выдавать значение $f$, а минимальное $d$, для которого найдется протокол, вычисляющий $f$, в котором на любой паре входов передается не более $d$ битов, называется коммуникационной сложностью $f$. Для детерминированной коммуникационной сложности $f$ используется обозначение $D(f)$. Рассматриваются и более сложные модели — недетерминированная, вероятностная сложность и т. д. (большинство моделей имеют аналогии в списке сложностных классов классической теории вычислений). Например, часто в этой диссертации мы будем рассматривать вероятностную коммуникационную сложность, в которой предполагается, что Алиса и Боб могут подбрасывать монетки, основывая на результатах бросаний дальнейшие действия. При этом, по аналогии с определением класса BPP, им разрешается на любой паре входов с небольшой вероятностью выдавать неправильный ответ. Можно требовать, чтобы ошибка была «односторонней» — тогда возникает аналогия с классом RP.

При определении различных моделей возникают и другие детали, связанные скорее с коммуникационной спецификой задачи. Например, одна и та же монетка у Алисы и Боба или у каждого своя? Рассматриваются оба варианта, причем первый называется моделью с общими случайными битами, а второй — с частными случайными битами. Взаимоотношение между общими и частными случайными битами также будет рассмотрено в диссертации. Подробный обзор этих двух и многих других моделей коммуникационной сложности можно

найти в классической монографии Кушилевица и Нисана [29]. Коммуникационная сложность показала себя как очень успешный инструмент в получении нижних оценок в других разделах сложности вычислений. Среди такого рода результатов наибольшую известность, вероятно, имеет работа Вигдерсона и Карчмера [28], устанавливающая связь между коммуникационной и схемной сложностью. Помимо этого коммуникационная сложность использовалась для нижних оценок на конечные автоматы, в потоковых алгоритмах, в тестировании свойств, сложности доказательств и т. д. (см. например, монографию Ружгардена [35], целиком посвященную применениям коммуникационной сложности).

Далее мы чуть более подробно изложим те вопросы в коммуникационной сложности, в рамках которых получены основные результаты диссертации.

Примером конкретной функции, коммуникационная сложность которой изучается в диссертации, является функция Gap Hamming Distance. Соответствующую коммуникационную задачу можно сформулировать так. Алиса и Боб получают по $n$-битовой строке. Гарантируется, что расстояние Хемминга между их входами либо не больше $L$, либо не меньше $U$. Алиса и Боб должны выдать 0 в первом случае и 1 во втором. Здесь $L \le U \le n$ — некоторые натуральные числа. Эту задачу мы обозначим через $\mathrm{GHD}(n, L, U)$.

Формально говоря, задача Алисы и Боба состоит в вычислении следующей частично определенной функции:

$$\mathrm{GHD}(n, L, U) : \{0,1\}^n \times \{0,1\}^n \to \{0,1\},$$
$$\mathrm{GHD}(n, L, U)(x, y) = \begin{cases} 0 & \text{если } d(x,y) \le L, \\ 1 & \text{если } d(x,y) \ge U, \\ \text{не определено} & \text{если } L < d(x,y) < U, \end{cases}$$

где через $d$ обозначено расстояние Хемминга.

Эта функция часто использовалась для получения нижних оценок в потоковых алгоритмах [41, 12] и в тестировании свойств [4]. В серии работ [27, 10, 11, 15, 36, 39] была полностью исследована вероятностная коммуникационная сложность этой функции для случая, когда $L$ и $U$ симметричны относительно $n/2$. Разные верхние оценки при других $L$ и $U$ были получены в работах [30, 44, 26]. Эти верхние оценки интересны тем, что не зависят от $n$ (т. е. от размерности входа). Исследовалась также сложность Gap Hamming Distance с односторонней ошибкой [13, 22]. Именно сложности с односторонней ошибкой и будет посвящен наш результат.

Второе направление в коммуникационной сложности, затронутое в диссертации, имеет отношение к теореме Раза — Маккинзи о связи коммуникационной сложности и сложности деревьев разрешения (вопросной сложности) [33]. Начнем с деревьев разрешения. Они вычисляют функции вида $f : \{0,1\}^n \to \{0,1\}$. На входе $x \in \{0,1\}^n$, чтобы вычислить $f(x)$, дерево может запрашивать значения $x$ в различных координатах. Номер очередной запрашиваемой координаты может зависеть от результатов предыдущих запросов. При этом нас не интересует время работы дерева. Единственное, что мы минимизируем, это суммарное число запросов (в худшем случае). Наименьшее $d$, для которого найдется дерево, вычисляющее $f$ и делающее не более $d$ запросов, называется вопросной сложностью $f$ и обозначается ниже через $D^{dt}(f)$. Здесь мы рассматриваем только детерминированные деревья и, соответственно, только детерминированную вопросную сложность. О других видах вопросной сложности можно прочитать в обзоре Бурмана и де Вольфа [14]. Раз и Маккинзи предложили рассматривать композиции функций, чтобы связать коммуникационную и вопросную сложность. Точнее, для функции $f : \{0,1\}^n \to \{0,1\}$, у которой можно измерить вопросную сложность, и для булевой функции $g$, вход которой каким-то образом разбит на два аргумента, благодаря чему определена коммуникационную сложность $g$, можно естественным образом определить

композицию $f \circ g$. А именно, надо рассмотреть $n$ пар входов для $g$, к каждому применить $g$ и к полученному $n$-битовому вектору применить $f$. Из такого определения вытекает неравенство $D(f \circ g) \leq D^{dt}(f) \cdot D(g)$ — дерево разрешения для $f$ легко преобразовать в протокол для $f \circ g$, в котором каждому запросу $f$ соответствует вычисление функции $g$ при помощи $D(g)$ битов.

Раз и Маккинзи исследовали вопрос: можно ли доказать обратное неравенство, т. е. неравенство вида $D(f \circ g) = \Omega(D^{dt}(f) \cdot D(g))$. Надежды на то, что оно выполнено для всех $f$ и $g$, нет — достаточно рассмотреть пример

$$f(x_1, x_2, \ldots, x_n) = x_1 \oplus x_2 \oplus \ldots \oplus x_n, \qquad g(x, y) = x \oplus y.$$

Тем не менее, Раз и Маккинзи доказали, что если взять в качестве $g$ «функцию адресации»:

$$\mathrm{IND}_k : \{0,1\}^k \times \{1, \ldots, k\} \to \{0,1\}, \qquad \mathrm{IND}_k(y, x) = y_x,$$

то неравенство $D(f \circ \mathrm{IND}_k) = \Omega(D^{dt}(f) \cdot D(\mathrm{IND}_k))$ будет выполнено для всех $f$ арности не более $k^{1/20}$ (арностью $f$ мы называем количество переменных $f$, выше это число обозначалось через $n$).

Примечательно в этом результате то, что коммуникационная сложность оценивается снизу через вопросную. Считается общепринятым, что нижние оценки на вопросную сложность зачастую намного легче получать, чем на коммуникационную сложность. При этом теорема Раза — Маккинзи автоматически преобразует нижние оценки на вопросную сложность в нижние оценки на коммуникационную сложность. В литературе такого рода результаты называются *лифтингом* (подъемом нижних оценок из более простой модели в более сложную). В данной области используется еще один специфический термин — функция $g$ называется гаджетом. Это отражает тот факт, что $g$ используется для «переноса» сложности «внешней» функции $f$.

Благодаря методу лифтинга, изобретенному Разом и Маккинзи, было получено много прорывных результатов в разных областях сложности вычислений. Например, сами Раз и Маккинзи при помощи своего метода (а также на основе упомянутой выше работы Вигдерсона и Карчмера) доказали неравенство сложностных классов $\mathrm{mNC}^i$ и $\mathrm{mNC}^{i+1}$ для всех $i \in \mathbb{N}$. В серии работ Ватсона, Гуса и Питасси при помощи лифтинга Раза – Маккинзи и его обобщений были выведены разделения между различными моделями в коммуникационной сложности из разделений соответствующих им моделей в вопросной сложности [24, 25, 23]. Техника лифтинга также использовалась в сложности доказательств и схем [20, 21] и в нижних оценках для структур данных [17].

Помимо этого, работы [42, 16] пытались ослабить ограничения на арность $f$ в теореме Раза — Маккинзи. Напомним, что неравенство $D(f \circ \mathrm{IND}_k) = \Omega(D^{dt}(f) \cdot D(\mathrm{IND}_k))$ было доказано лишь для $f$, арность которых не превосходит некоторой степени от длины входа гаджета $\mathrm{IND}_k$ (причем эта степень даже меньше единицы). В упомянутых работах удалось для других гаджетов получить аналогичный результат для всех $f$, арность которых экспоненциальна по длине входа гаджета. Кроме того, в работе [16] был даже сформулирован признак того, что для гаджета выполнен аналог лифтинга Раза — Маккинзи. Признак заключается в наличии так называемых протыкающих распределений в коммуникационной матрице гаджета. Тем не менее, наличие протыкающих распределений с хорошими параметрами было доказано лишь для нескольких примеров гаджетов. В диссертации будет представлен новый метод получения таких примеров, а также будут исследованы ограничения этой техники.

Еще одна сложностная мера, про связь которой с коммуникационной сложностью будет идти речь в диссертации, называется информационной сложностью. Впервые информационная сложность была введена в работе [1]. В этой области предлагается измерять не только

коммуникационную длину, но и так называемое *информационное разглашение* коммуникационных протоколов. Грубо говоря, информационное разглашение протокола равно количеству информации, которое игроки (Алиса и Боб) узнали из протокола о входе противоположного игрока. Точное определение можно дать при помощи Шенновской теории информации.

В первую очередь здесь надо отметить то, что информационное разглашение не может превосходить коммуникационной длины (один бит, посланный по каналу, приносит максимум один бит информации принимающему и ничего — посылающему). Можно легко придумать пример, когда разглашение будет намного меньше — игроки посылают друг другу случайные биты, не зависящие от входа. Более содержательным примером является следующий протокол. Пусть Алиса и Боб пересылают другу другу биты своих входов по очереди, пока не будет найдено первое различие. Коммуникационная длина такого протокола будет порядка $n$, где $n$ — длина входа. В то же время его информационное разглашение не будет превосходить $O(\log n)$. Дело в том, что игроки, по сути, узнают только номер первого различия, которое может быть задано порядка $\log n$ битами. Основным применением информационной сложности является так называемая проблема прямой суммы в коммуникационной сложности (см. обзор [40]). Ее суть в том, можно ли вычислить с точки зрения коммуникационной сложности $n$ копий функции в среднем быстрее, чем одну копию. Оказывается (впервые это было доказано в [2]), что аналогичное утверждение неверно для информационной сложности. Таким образом подходом к проблеме прямой суммы в коммуникационной сложности может быть более подробное исследование связи коммуникационной сложности с информационной.

Такого рода результаты о связи этих двух сложностных мер известны в литературе как *сжатие* коммуникационных протоколов. Типичный результат о сжатии формулируется так. Показывается, что протокол с маленьким информационным разглашением можно преобразовать в протокол, делающий «то же самое» (например, вычисляющий ту же функцию), с маленькой коммуникационной длиной. В идеале мы хотим, чтобы коммуникационная длина результирующего протокола была ограничена с точностью до мультипликативной константы информационным разглашением исходного протокола. Для произвольных протоколов такого достичь не удается, известны лишь частичные результаты [2, 5]. Но, например, для протоколов, использующих только общие случайные биты, доказаны почти что оптимальные результаты о сжатии [9, 32, 3]. Это обстоятельство мотивирует вопрос о роли частных случайных битов в информационной сложности. Точнее, всегда ли можно так сгенерировать частные случайные биты в общем источнике случайности, чтоб информационная сложность не сильно увеличилась? В литературе был дан положительный ответ на этот вопрос для одноразундовых протоколов, то есть для протоколов, в которых передает всегда только один игрок [9, 6]. В диссертации будет получена некая нетривиальная оценка для произвольных протоколов, которая однако позволяет лишь передоказать, а не усилить предыдущие результаты о сжатии.

Наконец, в диссертации будет рассмотрен еще один вопрос, берущий свое начало из теории информации. А именно, будут исследованы интерактивные аналоги теоремы Вольфа — Слепяна [37]. Так же, как в этой теореме, мы будем предполагать, что имеется посылающий (Алиса) и принимающий (Боб), причем на входе у Алисы случайная величина $X$, а у Боба — случайная величина $Y$. Задача Алисы — передать $X$ Бобу. Случайные величины $X$ и $Y$ совместно распределены, поэтому $Y$ можно воспринимать как частичную информацию об $X$ у Боба (подчеркнем, что $Y$ не виден Алисе!)

Будет несколько отличий от исходной формулировки теоремы Вольфа — Слепяна. Во-первых случайные величины $X$ и $Y$ предполагаются произвольными; в частности не требуется, чтобы они были получены в результате большого числа независимых испытаний. Кроме того, Бобу будет разрешено посылать сообщения Алисе (раньше передавала только Алиса). Этим изменением исходной формулировки объясняется связь с коммуникационной сложностью. Кроме того, нас будет интересовать средняя длина коммуникации, а не длина

в худшем случае.

В подобной постановке за последние годы в литературе было получено несколько верхних оценок [7, 9, 8]. В диссертации будут усилены эти верхние оценки, а также будут получены нижние.

## Цели и задачи исследования

Целью данной работы является получение новых результатов о связи коммуникационной сложности с другими сложностными мерами, а именно с информационной сложностью и с вопросной сложностью, а также получение новых оценок на коммуникационную сложность конкретных функций.

## Объект и предмет исследования

Основным объектом и предметом исследования данной работы является коммуникационная сложность. Исследуются также и другие сложностные меры, такие как информационная сложность и вопросная сложность, их взаимосвязь с коммуникационной сложностью.

## Научная новизна

Полученные в диссертации результаты являются новыми.

## Теоретическая и практическая значимость

Работа имеет теоретический характер. Полученные результаты представляют интерес для специалистов в области коммниукационной сложности, схемной сложности, вопросной сложности, интерактивного кодирования из Московского Государственного Университета имени М. В. Ломоносова, Санкт-Петербургского Государственного Университета, математического института имени В. А. Стеклова Российской Академии наук и других академических институтов и университетов.

## Основные методы исследования

В работе используются комбинаторные и вероятностные методы. Применены различные технические инструменты из теории информации (цепное правило, неравенство Пинскера) и из теории псевдослучайности (хеш-функции, экспандеры).

## Основные положения, выносимые на защиту

На защиту выносятся: обоснование актуальности, научная значимость работы, а также следующие положения, которые подтверждаются результатами исследования, представленными в заключении диссертации.

1. Нижняя оценка $\Omega(L^2/U+1)$ и верхняя оценка $O((L^2/U+1)\log L)$ на сложность функции $\mathrm{GHD}(n, L, U)$ с односторонней ошибкой (когда нельзя ошибаться на входах на расстоянии не больше $L$), а также SMP-протокол, доказывающий верхнюю оценку.

2. Способ, как из экспандеров получать гаджеты с протыкающими распределениями. Новый гаджет с рекордным взаимоотношением между арностью внешней функции и длиной входа гаджета, полученный при помощи этого способа. Невозможность улучшить это взаимоотношение при помощи текущей техники.

3. Способ безошибочно моделировать протоколы с частными случайными битами протоколами с общими случайными битами так, что информационное разглашение моделирующего протокола не превосходит $O(\sqrt{Id})$, где $I$ и $d$ — информационное разглашение и коммуникационная длина исходного протокола.

4. Протокол, решающий задачу Вольфа – Слепяна для пары случайных величин $X, Y$ с вероятностью ошибки $\varepsilon$, со средней длиной не более $(1 + 1/r)H + r + O(\log(1/\varepsilon))$ и со средним количеством раундов не более $2H/r+2$, где $H = H(X|Y)$ — условная энтропия $X$ при известном $Y$, а $r$ — произвольное натуральное число. Пример случайных величин $X, Y$, показывающий, что от члена порядка $O(\log(1/\varepsilon))$ в верхней оценке избавиться нельзя.

# Достоверность

Достоверность результатов исследования подтверждается теоретическими выкладками, а также сравнением с результатами других исследователей.

# Апробация работы

Основные результаты, полученные в диссертации, докладывались на следующих научных семинарах и конференциях:

- Колмогоровский семинар по сложности вычислений и сложности определений, механико-математический факультет МГУ им. М. В. Ломоносова, 2014–2018 гг.
- Научно-исследовательский семинар кафедры математической логики и теории алгоритмов, механико-математический факультет МГУ им. М. В. Ломоносова, 2018 год.
- Конференция “10th International Computer Science Symposium in Russia” в Листвянке в 2015 году.
- Конференция “Проблемы теоретической информатики” в Москве в 2015 году.
- Воркшоп “Algorithms in Communication Complexity, Property Testing and Combinatorics” в Москве в 2016 году.
- Конференция “11th International Computer Science Symposium in Russia” в Санкт-Петербурге в 2016 году.
- Конференция “Проблемы теоретической информатики” в Москве в 2016 году.
- Конференция “43rd International Symposium on Mathematical Foundations of Computer Science” в Ливерпуле в 2018 году.

На конференции “International Computer Science Symposium in Russia” в Санкт-Петербурге в 2016 году работа автора получила награду Yandex Best Student Paper Award.

## Публикации автора

Работа автора, посвященная роли частных случайных битов в информационной сложности, опубликована в сборнике трудов конференции "10th International Computer Science Symposium in Russia" ([46], Серия Lecture Notes in Computer Science, входит в систему Scopus). Работа автора, посвященная интерактивному аналогу теоремы Вольфа — Слепяна , опубликована в специальном выпуске журнала Theory of Computing Systems ([47], входит в систему Web of Science). Работа, посвященная задаче Gap Hamming Distance (совместная с Е. Клениным) и работа о теореме Раза — Маккинзи опубликованы в сборнике трудов конференции "43rd International Symposium on Mathematical Foundations of Computer Science" ([48-49], серия Leibniz International Proceedings in Informatics, входит в систему Scopus).

## Используемые обозначения

- $CC(\tau)$ — коммуникационная длина коммуникационного протокола $\tau$;
- $D(f)$ — детерминированная коммуникационная сложность функции $f$;
- $R(f)$ — вероятностная коммуникационная сложность функции $f$;
- $R^0(f), R^1(f)$ — вероятностная коммуникационная сложность функции $f$ с односторонней ошибкой (верхний индекс $i \in \{0,1\}$ указывает на то, что нельзя ошибаться на входах со значением $f$, равным $i$).
- $R^{\|}(f), R^{0,\|}(f), R^{1,\|}(f)$ — аналогичные обозначения в модели параллельных протоколов (SMP-протоколов);
- $ACC_\nu(\tau)$ — средняя длина коммуникационного протокола $\tau$ по входному распределению $\nu$;
- $IC_\nu(\tau)$ — информационное разглашение коммуникационного протокола $\tau$ по входному распределению $\nu$;
- $D^{dt}(f)$ — вопросная сложность функции $f$.

## Благодарности

Автор глубоко признателен своему научному руководителю Николаю Константиновичу Верещагину за неоценимую помощь при работе над диссертацией и постоянную поддержку. Автор благодарен всем участникам колмогоровского семинара на мехмате МГУ за научное сотрудничество и интерес к работе.

# Глава 1

# Основные понятия и вспомогательные результаты

## 1.1 Технические факты

### 1.1.1 Комбинаторика булева куба.

Расстоянием Хемминга между двумя $n$-битовыми строками $x = x_1 \dots x_n, y = y_1 \dots y_n \in \{0,1\}^n$ назовем следующую величину:

$$d(x,y) = |\{i \in \{1,2,\dots,n\} : x_i \neq y_i|\,.$$

Очевидно, $d(x,y)$ задает метрику на множестве $\{0,1\}^n$.

Шаром Хемминга с центром $x \in \{0,1\}^n$ и радиусом $r \in \{0,1,\dots,n\}$ назовем следующее множество

$$B_r^n(x) = \{y \in \{0,1\}^n : d(x,y) \leq r\}.$$

Размер $B_r^n(x)$ зависит только от $n$ и $r$ и равен $\binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{r}$. Введем обозначение $V_2(n,r) = |B_r^n(x)|$.

*Функцией Шеннона* назовем следующую функцию, определенную на отрезке $[0,1]$:

$$h(x) = x\log_2(1/x) + (1-x)\log_2(1/(1-x)).$$

Мы будем пользоваться следующей верхней оценкой на $V_2(n,r)$ через функцию Шеннона:

**Предложение 1** ([18], лемма 2.4.4)**.** *Предположим $r \leq n/2$. Тогда $V_2(n,r) \leq 2^{h(r/n)n}$.*

*Диаметром* множества $B \subset \{0,1\}^n$ назовем величину:

$$\operatorname{diam}(B) = \max_{x,y \in B} d(x,y).$$

Нам понадобится следующее утверждение о диаметре (известное как теорема Клейтмана):

**Предложение 2** ([18], теорема 2.4.16)**.** *Предположим $r < n/2$. Пусть диаметр множества $B \subset \{0,1\}^n$ не превосходит $2r$. Тогда $|B| \leq V_2(n,r)$.*

### 1.1.2 Теория информации

Всюду в этом разделе мы рассматриваем случайные величины, пробегающие конечное множество значений.

Пусть $X$ является случайной величиной, принимающей значения в множестве $\mathcal{X}$. *Энтропией Шеннона* $X$ и *энтропией различения* $X$ называются следующие две величины:

$$H(X) = \sum_{x \in \mathcal{X}} \Pr[X = x] \log_2(1/\Pr[X = x]),$$

$$H_2(X) = \log_2 \left( \frac{1}{\sum\limits_{x \in \mathcal{X}} \Pr[X = x]^2} \right).$$

Заметим, что энтропия Шеннона случайной величины, принимающей всего два значения с вероятностями $p$ и $1 - p$, равна функции Шеннона от $p$, т.е. $h(p)$.

Простым следствием неравенства Йенсена, примененного к логарифмической функции является то, что энтропия Шеннона не меньше энтропии различения:

**Предложение 3.** *Для любой случайной величины $X$ выполнено неравенство $H(X) \geq H_2(X)$.*

Зачастую мы будем подставлять в энтропию кортежи из случайных величин. При этом входящие в кортеж случайные величины мы будем перечислять просто через запятую, например $H(X_1, X_2)$.

Некоторые простые свойства энтропии Шеннона:

**Предложение 4** ([19])**.** *Пусть $X$ — случайная величина, принимающая $n$ возможных значений. Тогда выполнено неравенство*

$$H(X) \leq \log_2(n).$$

*Равенство $H(X) = \log_2(n)$ выполнено тогда и только тогда, когда распределение $X$ равномерно.*

**Предложение 5** ([19])**.** *Пусть случайные величины $X$, $Y$ и $Z$ принимают значения в одном и том же множестве $\mathcal{V}$, причем для некоторого $\delta \in [0, 1]$ и для любого $v \in \mathcal{V}$ выполнено:*

$$\Pr[X = v] = \delta \Pr[Y = v] + (1 - \delta) \Pr[Z = v].$$

*Тогда $H(X) \geq \delta H(Y) + (1 - \delta) H(Z)$.*

Рассмотрим теперь пару совместно распределенных случайных величин $(X, Y)$. Пусть значения $X$ пробегают множество $\mathcal{X}$, а значения $Y$ – множество $\mathcal{Y}$. Для всякого $y \in Y$ через $X|Y = y$ обозначим случайную величину, принимающую значения в множестве $\mathcal{X}$, распределение которой совпадает с условным распределением $X$ при условии $Y = y$. *Условной энтропией Шеннона* $X$ при известном $Y$ называется следующая величина:

$$H(X|Y) = \sum_{y \in \mathcal{Y}} H(X|Y = y) \Pr[Y = y].$$

Легко получить по определению величины $H(X|Y = y)$ следующее выражение для $H(X|Y)$:

$$H(X|Y) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \Pr[X = x, Y = y] \log_2(1/\Pr[X = x|Y = y]). \tag{1.1}$$

Обычную энтропию Шеннона можно рассматривать как частный случай условной энтропии. А именно, для любой случайной величины $X$ рассмотрим совместно распределенную с ней случайную величину $\Lambda$, принимающую единственное значение. Тогда $H(X) = H(X|\Lambda)$.

Добавления условия не увеличивает энтропию:

**Предложение 6** ([19]). *Для любых двух совместно распределенных случайных величин $X$ и $Y$ выполнено неравенство. $H(X|Y) \leq H(X)$.*

С другой стороны, применение некоторой функции к условию может только увеличить энтропию:

**Предложение 7** ([19]). *Для любых двух совместно распределенных случайных величин $X$ и $Y$ и любой функции $f$ выполнено: $H(X|Y) \leq H(X|f(Y))$.*

Следующее предложение (известное как цепное правило для энтропии) является основным инструментом при работе с энтропией кортежей случайных величин.

**Предложение 8** ( [19]). *Для любого набора совместно распределенных случайных величин $X_1, X_2, \ldots, X_n$ выполнено следующее равенство:*

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i|X_1, \ldots, X_{i-1}).$$

Цепное правило выполнено и при релятивизации:

**Следствие 1.** *Для любого набора совместно распределенных случайных величин $X_1$, $X_2$, …, $X_n$, $Y$ выполнено следующее равенство:*

$$H(X_1, \ldots, X_n|Y) = \sum_{i=1}^{n} H(X_i|X_1, \ldots, X_{i-1}, Y).$$

Действительно, достаточно применить предыдущее предложение ко всем кортежам вида $X_1, \ldots, X_n|Y = y$, где $y$ пробегает возможные значения $Y$.

Пусть теперь $X, Y, Z$ — три совместно распределенных случайных величины. Назовем *взаимной информацией между $X$ и $Y$* и *взаимной информацией между $X$ и $Y$ при условии $Z$* следующие две величины:

$$I(X : Y) = H(X) - H(X|Y), \qquad I(X : Y|Z) = H(X|Z) - H(X|Y, Z).$$

Из предложения 6 вытекает, что взаимная информация неотрицательна:

**Следствие 2.** *Для любых трех совместно распределенных случайных величин $X, Y, Z$ выполнены следующие неравенства:*

$$I(X : Y) \geq 0, \qquad I(X : Y|Z) \geq 0.$$

Взаимная информация симметрична:

**Предложение 9** ([19]). *Для любых трех совместно распределенных случайных величин $X, Y, Z$ выполнены следующие равенства:*

$$I(X : Y) = I(Y : X), \qquad I(X : Y|Z) = I(Y : X|Z).$$

Взаимная информация независимых случайных величин равна нулю:

**Предложение 10** ([19]). *Для любых двух совместно распределенных независимых случайных величин $X, Y$ выполнено $I(X : Y) = 0$.*

Аналог цепного правила выполнен и для взаимной информации:

**Предложение 11** ([19]). *Для любого набора совместно распределенных случайных величин $X_1, \ldots, X_n, Y$ выполнено следующее равенство:*

$$I(X_1, \ldots, X_n : Y) = \sum_{i=1}^{n} I(X_i : Y | X_1, \ldots, X_{i-1}).$$

По тем же соображениям, что и при получения следствия 1, легко вывести и релятивизованное цепное правило для взаимной информации:

**Предложение 12.** *Для любого набора совместно распределенных случайных величин $X_1, \ldots, X_n, Y, Z$ выполнено следующее равенство:*

$$I(X_1, \ldots, X_n : Y | Z) = \sum_{i=1}^{n} I(X_i : Y | X_1, \ldots, X_{i-1}, Z).$$

Пусть теперь даны два распределения вероятностей $P$ и $Q$ на конечном множестве $A$. Назовем *статистическим расстоянием* между $P$ и $Q$ и *расстоянием Кульбака — Лейблера* между $P$ и $Q$ следующие две величины:

$$\delta(P, Q) = \max_{B \subset A} |P\{B\} - Q\{B\}|, \qquad D_{KL}(P||Q) = \sum_{a \in A} P(a) \log_2(P(a)/Q(a)).$$

Расстояние Кульбака — Лейблера всегда неотрицательно:

**Предложение 13** ([19]). *Для любых двух вероятностных распределений $P$, $Q$ на одном и том же конечном множестве выполнено $D_{KL}(P||Q) \geq 0$.*

Статистическое расстояние и расстояние Кульбака — Лейблера связаны следующим неравенством (известным как неравенство Пинскера):

**Предложение 14** ([45]). *Для любых двух вероятностных распределений $P$, $Q$ на одном и том же конечном множестве выполнено следующее неравенство:*

$$\delta(P, Q) \leq \sqrt{\frac{\ln(2) \cdot D_{KL}(P||Q)}{2}}.$$

Пусть $X, Y$ — две случайные величины, пробегающие одно и то же множество значений. Тогда можно говорить о статистическом расстоянии и о расстоянии Кульбака-Лейблера между $X$ и $Y$. А именно, положим $\delta(X, Y) = \delta(P_X, P_Y), D_{KL}(X||Y) = D_{KL}(P_X||P_Y)$, где $P_X$ — распределение $X$, а $P_Y$ — распределение $Y$.

Непосредственной проверкой легко убедиться в следующем взаимоотношении между расстоянием Кульбака — Лейблера и взаимной информацией:

**Предложение 15.** *Для любых двух совместно распределенных случайных величин $X$ и $Y$ выполнено равенство:*

$$I(X : Y) = \sum_{x \in \mathcal{X}} D_{KL}((Y|X = x)||Y) \cdot \Pr[X = x].$$

*Здесь через $\mathcal{X}$ обозначено множество значений $X$.*

Если $x \in \{0,1\}^*$, то через $|x|$ обозначим длину $x$. Два слова $x, y \in \{0,1\}^*$ назовем сравнимыми, если существует $w \in \{0,1\}^*$ такой, что $x = yw$ или $y = xw$. Множество $D \subset \{0,1\}^*$ назовем префиксным, если любые два различных элемента $D$ не сравнимы.

**Предложение 16** ([19]). *Пусть $X$ — случайная величина, принимающая значения в конечном префиксном множестве. Тогда*

$$\mathbb{E}|X| \geq H(X).$$

Наконец нам понадобится следующее неравенство, известное как неравенство Фано:

**Предложение 17** ([19]). *Пусть $X, Y$ — две совместно распределенных случайных величины, принимающих значения в конечных множествах $\mathcal{X}$ и $\mathcal{Y}$. Тогда для любой функции $f : \mathcal{Y} \to \mathcal{X}$ выполнено:*

$$H(X|Y) \leq \Pr[f(Y) \neq X] \cdot \log_2(|\mathcal{X}|) + 1.$$

### 1.1.3 Экспандеры

Мы рассматриваем неориентированные графы без кратных ребер, но в них могут быть петли. Пусть $G$ — граф с множеством вершин $V$ и множеством ребер $E$ (формально говоря, $E$ — это какой-то набор из одноэлементных и двухэлементных подмножеств $V$). Рассмотрим произвольное подмножество $S \subset V$. Через $\Gamma(S)$ обозначим множество вершин $G$, соединенных ребром хотя бы с одной вершиной из $S$:

$$\Gamma(S) = \{u \in V : \text{найдется } v \in S \text{ такое, что } \{u, v\} \in S\}.$$

Если $s \in V$ — вершина $G$, то обозначим $\Gamma(s) = \Gamma(\{s\})$.

Граф называется $d$-регулярным, если степень любой его вершины равна $d$ (петля дает вклад 1 к степени вершины). Через $M_G$ обозначим матрицу смежности $G$. Матрица $M_G$ симметрична, поэтому у нее есть собственный базис. Заметим, что если граф $G$ является $d$-регулярным, то $d$ является собственным значением $M_G$ (этому собственному значению соответствует вектор из одних единиц). Пусть $\gamma$ — число от 0 до 1. Назовем $d$-регулярный граф $G$ с $m$ вершинами *спектральным* $(m, d, \gamma)$*-экспандером*, если кратность собственного значения $d$ у матрицы $M_G$ равна 1, а все остальные собственные значения $M_G$ не превосходят по модулю $\gamma d$.

Нам потребуется следующее свойство спектральных экспандеров:

**Предложение 18** ([38], теорема 4.6). *Предположим граф $G$ с множеством вершин $V$ является спектральным $(m, d, \gamma)$-экспандером, $d > 0$. Тогда для любого $A \subset V$ выполнено следующее неравенство:*

$$\frac{|\Gamma(A)|}{|A|} \geq \frac{1}{\gamma^2 + (1 - \gamma^2)\frac{|A|}{m}}.$$

Нам понадобится следующая явная конструкция спектральных экспандеров. Пусть $q$ — степень простого числа. Конечное поле размера $q$ обозначим через $\mathbb{F}_q$. Графом $AP_q$ назовем следующий граф. Его вершинами будут упорядоченные пары элементов $\mathbb{F}_q$. Две вершины $(x, y), (a, b) \in \mathbb{F}_q^2$ мы соединяем ребром тогда и только тогда, когда $ax = b + y$ в поле $\mathbb{F}_q$.

**Предложение 19** ([34], лемма 5.1). *Граф $AP_q$ является спектральным $(q^2, q, 1/\sqrt{q})$ - экспандером.*

## 1.2 Коммуникационная сложность

В коммуникационной сложности рассматривается следующая ситуация. Есть два конечных множества, $\mathcal{X}$ и $\mathcal{Y}$, и два игрока, Алиса и Боб. Алиса получает на вход $x \in \mathcal{X}$, а Боб — $y \in \mathcal{Y}$. Алиса не знает $y$, а Боб — $x$. Тем не менее, они хотят ответить на некоторый вопрос

об $x$ и $y$. Например, они хотят узнать значение некоторой заранее фиксированной функции $g : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ на паре $(x, y)$. Для этого они *коммуницируют* между собой согласно некоторому алгоритму. Такие алгоритмы мы будем называть *коммуникационными протоколами*. За один такт работы протокола кто-то из игроков передает другому один бит. Распределение ролей (кто принимает, а кто — передает) в каждый момент времени определяется предыдущей коммуникацией. Таким образом невозможна ситуация, когда оба игрока молчат либо оба пытаются что-то передать. Бит, посылаемый передающим, определяется предыдущей коммуникацией, а также входом передающего.

Поскольку коммуникационная сложность является основным объектом изучения диссертации, данный раздел является самым объемным в текущей главе. Более подробное введение можно найти в монографиях [29, 35].

**Глава 5** и **Глава 6**, в которых фигурирует информационная сложность, требуют, к сожалению, аккуратного определения протоколов и введения ряда технических обозначений.

### 1.2.1 Коммуникационный протоколы и их характеристики

Напомним, что длину двоичного слова $a \in \{0, 1\}^*$ мы обозначаем через $|a|$. Для $b \in \{0, 1\}$ через $\overline{b}$ обозначим 0, если $b = 1$, и 1, если $b = 0$.

Под двоичным деревом с корнем мы понимаем конечное множество $D \subset \{0, 1\}^*$ такое, что

- пустое слово (которое мы будем обозначать через $\Lambda$) принадлежит $D$;
- если непустое слово $a_1 \dots a_n$ принадлежит $D$, то $D$ также принадлежат слова $a_1 \dots a_{n-1}$ и $a_1 \dots a_{n-1}\overline{a_n}$.

Мы используем стандартную терминологию для деревьев с корнем. Элементы $D$ мы будем называть вершинами. Глубиной вершины $v \in D$ назовем ее длину как двоичного слова. Далее, *корнем* $D$ называется пустое слово. Элемент $a_1 \dots a_n \in D$ называется *листом* $D$, если $a_1 \dots a_n 0 \notin D$. Элемент $D$, не являющийся листом, мы называем *внутренней вершиной* $D$. Из определения вытекает, что если $a_1 \dots a_n \in D$ — внутренняя вершина, то слова $a_1 \dots a_n 0$ и $a_1 \dots a_n 1$ принадлежат $D$. Эти два слова являются *сыновьями* $a_1 \dots a_n$, а слово $a_1 \dots a_n$ для них, в свою очередь, является *родителем*.

Множество листьев дерева с корнем $D$ мы будем обозначать через $\mathcal{L}(D)$, а множество его внутренних вершин — через $\mathcal{I}(D)$. *Глубиной* $D$ будем называть максимально возможную глубину вершины $D$ (которая, очевидно, достигается на одном из листов).

**Определение 1.** *Пусть $\mathcal{X}, \mathcal{Y}$ — конечные множества. Протоколом с частными случайными битами над $(\mathcal{X}, \mathcal{Y})$ называется пятерка $\langle D, A, B, \phi, \psi \rangle$, где*

- *$D$ — это двоичное дерево с корнем;*
- *$A, B$ — непересекающиеся подмножества $\mathcal{I}(D)$, причем $A \cup B = \mathcal{I}(D)$;*
- *$\phi, \psi$ — это функции следующего вида:*

$$\phi : \mathcal{X} \times A \to [0, 1], \qquad \psi : \mathcal{Y} \times B \to [0, 1].$$

Листья и внутренние вершины $D$ будем также называть листьями и внутренними вершинами протокола.

Коммуникацию в протоколе $\tau = \langle D, A, B, \phi, \psi \rangle$ можно себе представлять следующим образом. Пусть $x \in \mathcal{X}$ — вход Алисы, а $y \in \mathcal{Y}$ — вход Боба. В каждый момент времени они находятся в какой-то вершине $v \in D$ (в начале — в корне $D$). Если $v \in A$, передает Алиса.

Она генерирует случайный бит $b \in \{0,1\}$ т. ч. $\Pr[b=0] = \phi(x,v)$ (каждый новый бит генерируется независимо от предыдущих). Алиса посылает $b$ Бобу и они переходят в вершину $vb$. Если же $v \in B$, передает Боб. Он действует по аналогии, генерируя случайный бит с вероятностью нуля, равной $\psi(y,v)$, и посылая его Алисе. В конце концов $v$ становится равным какому-то листу $D$.

Можно себе представлять, что для генерации $b$ Алиса и Боб подбрасывают монетку (*случайные биты*). При этом бросания Алисы не видны Бобу, а бросания Боба — Алисе. Поэтому случайные биты в этом определении называются *частными*.

Можно интересоваться вероятностью, с которой для данной пары входов $(x,y)$ в протоколе $\tau$ будет достигнута вершина $v$. Мы эту вероятность будем обозначать через $p^\tau(v,x,y)$. Формально определить функцию

$$p^\tau : \{0,1\}^* \times \mathcal{X} \times \mathcal{Y} \to [0,1]$$

можно так: для $v \in D$ индукцией по глубине $v$ положить:

$$\begin{aligned} & p^\tau(\Lambda,x,y) = 1, \\ \forall v \in \mathcal{I}(D), \qquad & p^\tau(v0,x,y) = \begin{cases} p^\tau(v,x,y)\phi(x,v) & v \in A, \\ p^\tau(v,x,y)\psi(y,v) & v \in B. \end{cases} \\ \forall v \in \mathcal{I}(D), \qquad & p^\tau(v1,x,y) = \begin{cases} p^\tau(v,x,y)(1-\phi(x,v)) & v \in A, \\ p^\tau(v,x,y)(1-\psi(y,v)) & v \in B, \end{cases} \end{aligned} \tag{1.2}$$

а для $u \in \{0,1\}^* \setminus D$ положить $p^\tau(u,x,y) = 0$. Легко видеть, что функция $p^\tau(\cdot,x,y)$ задает распределение вероятностей на листьях протокола, т.е.

$$\forall (x,y) \in \mathcal{X} \times \mathcal{Y} \qquad \sum_{v \in \mathcal{L}(D)} p^\tau(v,x,y) = 1. \tag{1.3}$$

Коммуникационная длина $\tau$ — это, грубо говоря, максимально возможное количество переданных бит на в $\tau$. Формально:

**Определение 2.** *Коммуникационной длиной протокола $\tau = \langle D, A, B, \phi, \psi \rangle$ с частными случайными битами называется глубина $D$. Коммуникационную длину $\tau$ обозначим через $CC(\tau)$.*

Нас также будет интересовать, сколько раз в протоколе происходит «переключение» между передачей от Алисы к Бобу и передачей от Боба к Алисе. Точнее, пусть фиксирован протокол с частными случайными битами $\tau = \langle D, A, B, \phi, \psi \rangle$. Рассмотрим произвольный лист $l \in \mathcal{L}(D)$. Коммуникацию, которая привела Алису и Боба в $l$, можно разбить на непрерывные отрезки времени, когда либо передавала только Алиса, либо передавал только Боб. Количество таких отрезков мы обозначим через $r_\tau(l)$. Это число на единицу больше количества «переключений» от Алисы к Бобу и от Боба к Алисе, поэтому формально определить $r_\tau(l)$ можно так:

$$\begin{aligned} r_\tau(l) = 1 + |\{ i \in \{0,1,\ldots,|l|-2\} : {} & l_1 \ldots l_i \in A, l_1 \ldots l_{i+1} \in B \\ & \text{либо } l_1 \ldots l_i \in B, l_1 \ldots l_{i+1} \in A \}|. \end{aligned}$$

**Определение 3.** *Протокол $\tau = \langle D, A, B, \phi, \psi \rangle$ с частными случайными битами называется $k$-раундовым, если для любого $l \in \mathcal{L}(D)$ выполнено $r_\tau(l) \le k$.*

Частным случаем протоколов с частными случайными битами являются *детерминированные* протоколы. В них Алиса и Боб, грубо говоря, не используют случайность.

**Определение 4.** *Коммуникационный протокол с частными случайными битами* $\tau = \langle D, A, B, \phi, \psi \rangle$ *над* $(\mathcal{X}, \mathcal{Y})$ *называется детерминированным, если* $\phi(\mathcal{X}, A), \psi(\mathcal{Y}, B) \subset \{0, 1\}$.

Если $\tau = \langle D, A, B, \phi, \psi \rangle$ — детерминированный протокол над $(\mathcal{X}, \mathcal{Y})$, то однозначно можно определить, в какие вершины $D$ на данной паре входов $(x, y) \in \mathcal{X} \times \mathcal{Y}$ придут Алиса и Боб. Формально, будем говорить, что детерминированный протокол $\tau$ на паре $(x, y)$ *приходит в вершину* $v \in D$, если $p^\tau(v, x, y) = 1$. В частности, для данной пары $(x, y) \in \mathcal{X} \times \mathcal{Y}$ существует и единственен лист $l \in \mathcal{L}(D)$, в который приходит $\tau$ на $(x, y)$.

Основным инструментом для анализа детерминированных коммуникационных протоколов является их «прямоугольное свойство»:

**Предложение 20.** *Пусть* $\tau = \langle D, A, B, \phi, \psi \rangle$ *— детерминированный протокол над* $(\mathcal{X}, \mathcal{Y})$. *Тогда для любого* $v \in D$ *существуют* $\mathcal{X}_0 \subset \mathcal{X}$ *и* $\mathcal{Y}_0 \subset \mathcal{Y}$ *такие, что для всех* $(x, y) \in \mathcal{X} \times \mathcal{Y}$ *пара* $(x, y)$ *приходит в* $v$ *в протоколе* $\tau$ *тогда и только тогда, когда* $(x, y) \in \mathcal{X}_0 \times \mathcal{Y}_0$.

Оно легко выводится из следующего более общего свойства:

**Предложение 21.** *Пусть* $\tau = \langle D, A, B, \phi, \psi \rangle$ *— протокол с частными случайными битами над* $(\mathcal{X}, \mathcal{Y})$, *а* $v$ *— произвольная вершина* $D$. *Тогда ранг матрицы*

$$M_v : \mathcal{X} \times \mathcal{Y} \to [0, 1], \qquad M_v(x, y) = p^\tau(v, x, y),$$

*не превосходит* 1.

*Доказательство.* Это легко доказать индукцией по глубине $v$. Матрица $M_\Lambda$ заполнена единицами, и поэтому ее ранг равен 1. С другой стороны, матрицы $M_{v0}$ и $M_{v1}$ получаются из $M_v$ домножением всех строк (если $v \in A$) или домножением всех столбцов (если $v \in B$) на какие-то числа, отчего ранг не может увеличиться. □

Мы будем рассматривать более широкий класс протоколов, чем протоколы с частными случайными битами. А именно, мы будем разрешать Алисе и Бобу подбрасывать монетку, бросания которой видны обоим игрокам (*общие случайные биты*). Протоколы, использующие как общую, так и частную случайность, мы будем называть просто *вероятностными*.

**Определение 5.** *Вероятностным протоколом над* $(\mathcal{X}, \mathcal{Y})$ *называется случайная величина* $\tau$, *принимающая значения в конечном множестве протоколов с частными случайными битами над* $(\mathcal{X}, \mathcal{Y})$.

Бросания общей монетки, грубо говоря, определяют значение $\tau$. Затем Алиса и Боб используют частные случайные биты для того, чтобы общаться согласно выбранному значению $\tau$.

В этом определении $\tau$ — это случайная величина с конечным числом значений, которую можно задать как функцию на каком-то конечном вероятностном пространстве. Нам зачастую будет удобно иметь это вероятностное пространство в явном виде. Иными словами, вероятностный протокол $\tau$ над $(\mathcal{X}, \mathcal{Y})$ можно задать как отображение, которое каждому элементу $r \in \mathcal{R}$ некоторого конечного вероятностного пространства $(\mathcal{R}, \mu)$ сопоставляет протокол с частными случайными битами $\tau_r$ над $(\mathcal{X}, \mathcal{Y})$ . При этом мы будем называть $(\mathcal{R}, \mu)$ *пространством общих случайных битов* протокола $\tau$.

**Определение 6.** *Коммуникационная сложность вероятностного протокола* $\tau$ *с пространством общих случайных битов* $(\mathcal{R}, \mu)$ *равна*

$$\max_{r \in \mathcal{R}} CC(\tau_r)$$

*и обозначается через* $CC(\tau)$. *Вероятностный протокол* $\tau$ *называется* $k$*-раундовым, если для всех* $r \in \mathcal{R}$ *протоколы* $\tau_r$ *являются* $k$*-раундовыми.*

Вероятностный протокол $\tau$ с пространством общих случайных битов $(\mathcal{R}, \mu)$ называется *протоколом с общими случайными битами*, если для любого $r \in \mathcal{R}$ протокол $\tau_r$ является детерминированным (Алиса и Боб не используют частных случайных битов). Для вероятностного протокола $\tau$ над $(\mathcal{X}, \mathcal{Y})$ с пространством общих случайных битов $(\mathcal{R}, \mu)$ определим функцию:

$$p^\tau : \{0,1\}^* \times \mathcal{X} \times \mathcal{Y} \to [0,1],$$
$$p^\tau(v, x, y) = \mathbb{E}_{r \sim \mu} p^{\tau_r}(v, x, y).$$

Значение $p^\tau(v, x, y)$, как и раньше, соответствует вероятности достижения вершины $v$ в протоколе $\tau$ на входе $(x, y)$.

Наконец, определим так называемые SMP-протоколы (сокращение от Simultaneous Message Passing), которые, формально говоря, не вписываются ни в одно из данных ранее определений. Дело в том, что в SMP-протоколах Алиса и Боб не общаются между собой. Вместо этого они посылают по одному двоичному слову третьему игроку, Чарли. Задача Чарли — ответить на какой-то вопрос о входах Алисы и Боба.

Мы сразу дадим определение вероятностных SMP-протоколов с общими случайными битами (рассматривать SMP-протоколы других видов нам не потребуется).

**Определение 7.** *Пусть $\mathcal{X}, \mathcal{Y}$ — конечные множества. SMP-протоколом $\tau$ над $(\mathcal{X}, \mathcal{Y})$ называется пятерка $\langle (\mathcal{R}, \mu), m_A, m_B, p, q \rangle$, где*

- *$(\mathcal{R}, \mu)$ — конечное вероятностное пространство;*
- *$p, q$ — натуральные числа;*
- *$m_A$, $m_B$ — это функции вида:*

$$m_A : \mathcal{X} \times \mathcal{R} \to \{0,1\}^p, \qquad m_B : \mathcal{Y} \times \mathcal{R} \to \{0,1\}^q.$$

*При этом сумма $p+q$ называется коммуникационной сложностью $\tau$ и обозначается через $CC(\tau)$.*

### 1.2.2 Средняя коммуникационная длина и информационное разглашение

Определим теперь *среднюю* коммуникационную длину протокола. Пусть фиксирован вероятностный протокол $\tau$ над $(\mathcal{X}, \mathcal{Y})$, а также распределение вероятностей $\nu$ на его входах, то есть на $\mathcal{X} \times \mathcal{Y}$. Величина $CC(\tau)$ равна, грубо говоря, количеству переданных между Алисой и Бобом битов *в худшем случае*. Можно также интересоваться средним количеством переданных бит, когда входы для протокола выбираются по распределению $\nu$. Формально эту величину удобно определить сначала для протоколов с частными случайными битами, а уже затем — для произвольных вероятностных.

**Определение 8.** *Назовем средней коммуникационной длиной протокола с частными случайными битами $\tau = \langle D, A, B, \phi, \psi \rangle$ по распределению $\nu$ следующую величину:*

$$\mathbb{E}_{(x,y) \sim \nu} \sum_{l \in \mathcal{L}(D)} p^\tau(l, x, y) \cdot |l|.$$

*Будем обозначать эту величину через $ACC_\nu(\tau)$.*

Напомним, что $p^\tau(l,x,y)$ — это вероятность получения листа $l$ в $\tau$ на входе $(x,y)$, а $|l|$ — его длина. Таким образом, сумма

$$\sum_{l\in\mathcal{L}(D)} p^\tau(l,x,y)\cdot|l|$$

соответствует в определении 8 средней длине протокола (по частным случайным битам) для данной входной пары $(x,y)$, которая потом усредняется по $\nu$.

**Определение 9.** *Назовем средней коммуникационной длиной вероятностного протокола $\tau$ по распределению $\nu$ следующую величину:*

$$\mathbb{E}_{r\sim\mu} ACC_\nu(\tau_r),$$

*где $(\mathcal{R},\mu)$ — пространство общих случайных битов $\tau$. Будем обозначать эту величину через $ACC_\nu(\tau)$.*

Совершенно аналогично определяется среднее количество раундов протокола $\tau$ по данному входному распределению $\nu$ — просто надо всюду выше заменить длину листа $l$ на $r_\tau(l)$ (количество раундов коммуникации на листе $l$).

Также дадим здесь определение информационного разглашения коммуникационного протокола. Впервые один из вариантов этого понятия появился в 2004-м году в работе [1] («внешнее» информационное разглашение). Мы используем другой вариант («внутреннее» информационное разглашение) из работы [2] и называем его просто информационным разглашением.

Неформально говоря, информационное разглашение протокола равно количеству информации, которое игроки передают друг другу в протоколе о своих входах. Формальное определение использует Шенноновскую теорию информации. Как и со средней длиной, мы сначала определяем информационное разглашение для протоколов с частными случайными битами.

**Определение 10.** *Пусть $\mathcal{X},\mathcal{Y}$ — конечные множества, $\nu$ — распределение вероятностей на $\mathcal{X}\times\mathcal{Y}$, а $\tau=\langle D,A,B,\phi,\psi\rangle$ — протокол с частными случайными битами над $(\mathcal{X},\mathcal{Y})$. Информационным разглашением $\tau$ по $\nu$ назовем следующие величину:*

$$IC_\nu(\tau) = I(Y:T|X) + I(X:T|Y),$$

*где $(X,Y)$ — пара случайных величин, распределенная согласно $\nu$, а $T$ — совместно распределенная с $(X,Y)$ случайная величина, принимающая значения в множестве $\mathcal{L}(D)$ такая, что для любой $(x,y)\in\mathcal{X}\times\mathcal{Y}$ и любого $v\in\mathcal{L}(D)$ выполнено следующее:*

$$\Pr[T=v|X=x,Y=y] = p^\tau(v,x,y).$$

Здесь подразумевается, что Алиса получает на вход $X$, а Боб — $Y$. Величина $T$ часто называется *транскриптом* протокола $\tau$ и по сути представляет собой конкатенацию всех битов, посылаемых в $\tau$ (которая задает лист, достигаемый в $\tau$). В этом определении величина $I(Y:T|X)$ соответствует количеству информации, которую Алиса узнает об $Y$ (входе Боба) из $T$. В условии для взаимной информации находится $X$ (вход Алисы), которое Алиса знает заранее. Аналогично величина $I(X:T|Y)$ соответствует количеству информации, которую Боб узнает из протокола об $X$.

**Определение 11.** *Пусть $\mathcal{X},\mathcal{Y}$ — конечные множества, $\nu$ — распределение вероятностей на $\mathcal{X}\times\mathcal{Y}$, а $\tau$ — вероятностный протокол над $(\mathcal{X},\mathcal{Y})$ с пространством общих случайных битов $(\mathcal{R},\mu)$. Информационным разглашением $\tau$ по $\nu$ назовем следующие величину:*

$$IC_\nu(\tau) = \mathbb{E}_{r\sim\mu} IC_\nu(\tau_r).$$

Введенные нами величины удовлетворяют следующему взаимоотношению:

**Предложение 22.** *Для любого вероятностного протокола $\tau$ и любого распределения $\nu$ выполнено:*

$$IC_\nu(\tau) \le ACC_\nu(\tau) \le CC(\tau).$$

Второе неравенство очевидно — среднее не может превосходить максимума. Первое неравенство соответствует интуиции, что информационное разглашение не может превосходить информационной длины. Действительно, каждый бит, пересылаемый в протоколе, приносит максимум один бит информацию принимающему и ничего — посылающему. Формальное доказательство можно найти в [5].

### 1.2.3 Вычисление функций и различные виды коммуникационной сложности

В этом разделе мы определим детерминированную и различные виды вероятностной коммуникационной сложности. Итак, пусть $\mathcal{X}, \mathcal{Y}$ — конечные множества, а $g : \mathcal{X} \times \mathcal{Y} \to \{0, 1, \perp\}$ — некоторая функция. Здесь $\perp$ — произвольный символ, отличный от 0 и 1. Содержательно под $\perp$ понимается «символ неопределенности». А именно, мы будем говорить, что $g$ не определено на $(x, y) \in \mathcal{X} \times \mathcal{Y}$, если $g(x, y) = \perp$.

Мы будем использовать обозначение $\neg g$ для отрицания функции $g$. Формально $\neg g$ это функция вида $\neg g : \mathcal{X} \times \mathcal{Y} \to \{0, 1, \perp\}$, где

$$\neg g(x, y) = \begin{cases} 0 & \text{если } g(x, y) = 1, \\ 1 & \text{если } g(x, y) = 0, \\ \perp & \text{если } g(x, y) = \perp. \end{cases}$$

Задача Алисы и Боба состоит в вычислении $g(x, y)$, где $x$ — вход Алисы, а $y$ — вход Боба, когда $g(x, y)$ определено. Значение $g$ Алиса и Боб будут определять по листу, который они достигают в протоколе. А именно, будем говорить, что детерминированный коммуникационный протокол $\tau = \langle D, A, B, \phi, \psi \rangle$ *вычисляет функцию* $g$, если для некоторой функции $f : \mathcal{L}(D) \to \{0, 1\}$ и для всех $(x, y) \in \mathcal{X} \times \mathcal{Y}$ выполнено следующее:

- если $g(x, y) = 0$ и $l$ — это лист $D$, в который приходит $\tau$ на паре $(x, y)$, то $f(l) = 0$;
- если $g(x, y) = 1$ и $l$ — это лист $D$, в который приходит $\tau$ на паре $(x, y)$, то $f(l) = 1$.

**Определение 12.** *Назовем детерминированной коммуникационной сложностью функции $g : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ минимальное $d \in \mathbb{N}$ такое, что существует детерминированный протокол $\tau$, вычисляющий $g$, для которого выполнено $CC(\tau) = d$. Детерминированная коммуникационная сложность $g$ обозначается через $D(g)$.*

Перейдем к вероятностным протоколам. Для них, естественно, имеет смысл разрешить вычислять функцию $g : \mathcal{X} \times \mathcal{Y} \to \{0, 1, \perp\}$ с небольшой вероятностью ошибки.

Более точно, будем говорить, что вероятностный протокол $\tau$ с пространством общих случайных битов $(\mathcal{R}, \mu)$ *вычисляет функцию $g$ с ошибкой $\varepsilon$*, если для каждого $r \in \mathcal{R}$ можно определить функцию $f_r : \mathcal{L}_r \to \{0, 1\}$, где $\mathcal{L}_r$ — множество листьев протокола $\tau_r$, так, что для всех $(x, y) \in \mathcal{X} \times \mathcal{Y}$ выполнено следующее:

- если $g(x, y) = 0$, то

$$\mathbb{E}_{r \sim \mu} \left[ \sum_{l \in \mathcal{L}_r : f_r(l) = 1} p^{\tau_r}(l, x, y) \right] \le \varepsilon$$

(вероятность на $(x, y)$ прийти в лист, в котором выдается 1, не превосходит $\varepsilon$);

- если $g(x,y)=1$, то

$$\mathbb{E}_{r\sim\mu}\left[\sum_{l\in\mathcal{L}_r:f_r(l)=0} p^{\tau_r}(l,x,y)\right]\leq\varepsilon$$

(вероятность на $(x,y)$ прийти в лист, в котором выдается 0, не превосходит $\varepsilon$).

Также определим вычисление $g$ с односторонней ошибкой. Здесь мы уже будем требовать, чтобы протокол никогда не ошибался на входах, на которых значение $g$ равно нулю. Формально, будем говорить, что вероятностный протокол $\tau$ с пространством общих случайных битов $(\mathcal{R},\mu)$ вычисляет функцию $g$ с *односторонней* ошибкой $\varepsilon$, если для каждого $r\in\mathcal{R}$ можно определить функцию $f_r:\mathcal{L}_r\to\{0,1\}$, где $\mathcal{L}_r$ — множество листьев протокола $\tau_r$, так, что для всех $(x,y)\in\mathcal{X}\times\mathcal{Y}$ выполнено следующее:

- если $g(x,y)=0$, то

$$\mathbb{E}_{r\sim\mu}\left[\sum_{l\in\mathcal{L}_r:f_r(l)=1} p^{\tau_r}(l,x,y)\right]=0$$

(вероятность на $(x,y)$ прийти в лист, в котором выдается 1, равна 0);

- если $g(x,y)=1$, то

$$\mathbb{E}_{r\sim\mu}\left[\sum_{l\in\mathcal{L}_r:f_r(l)=0} p^{\tau_r}(l,x,y)\right]\leq\varepsilon$$

(вероятность на $(x,y)$ прийти в лист, в котором выдается 0, не превосходит $\varepsilon$).

Выбор нуля в качестве значения $g$, на котором нельзя ошибаться, не существеннен. Можно либо определить аналогичное понятие для единицы, либо рассматривать при необходимости $\neg g$.

Определим теперь различные виды вероятностной коммуникационной сложности.

**Определение 13.** *Пусть $g:\mathcal{X}\times\mathcal{Y}\to\{0,1,\perp\}$ — некоторая функция.*

- *Через $R_\varepsilon(g)$ обозначим минимальное $d\in\mathbb{N}$ такое, что найдется протокол $\tau$ с общими случайными битами, вычисляющий $g$ с ошибкой $\varepsilon$, для которого $CC(g)=d$.*
- *Через $R^{pr}_\varepsilon(g)$ обозначим минимальное $d\in\mathbb{N}$ такое, что найдется протокол $\tau$ с частными случайными битами, вычисляющий $g$ с ошибкой $\varepsilon$, для которого $CC(g)=d$.*
- *Через $R^0_\varepsilon(g)$ обозначим минимальное $d\in\mathbb{N}$ такое, что найдется протокол $\tau$ с общими случайными битами, вычисляющий $g$ с односторонней ошибкой $\varepsilon$, для которого $CC(g)=d$.*
- *Через $R^1_\varepsilon(g)$ обозначим минимальное $d\in\mathbb{N}$ такое, что найдется протокол $\tau$ с общими случайными битами, вычисляющий $\neg g$ с односторонней ошибкой $\varepsilon$, для которого $CC(g)=d$.*

Также нужно определить аналогичные величины для SMP-протоколов. Для этого скажем, что SMP-протокол $\tau=\langle(R,\mu),\mu_A,\mu_B,p,q\rangle$ вычисляет функцию $g:\mathcal{X}\times\mathcal{Y}\to\{0,1,\perp\}$ с ошибкой $\varepsilon$, если существует функция $f:\mathcal{R}\times\{0,1\}^p\times\{0,1\}^q\to\{0,1\}$ такая, что для любого $x\in\mathcal{X}\times\mathcal{Y}$ выполнено:

- если $g(x,y)=0$, то

$$\Pr_{r\sim\mu}[f(r,m_A(x,r),m_B(y,r))=1]\leq\varepsilon;$$

- если $g(x,y)=1$, то
$$\Pr_{r\sim\mu}[f(r,m_A(x,r),m_B(y,r))=0]\le\varepsilon.$$

Более того, скажем, что протокол $\tau$ вычисляет $G$ с *односторонней* ошибкой $\varepsilon$, если при $g(x,y)=0$ выполнено более сильное условие:

$$\Pr_{r\sim\mu}[f(r,m_A(x,r),m_B(y,r))=1]=0.$$

Дадим теперь для SMP-протокол определения, аналогичные определению 13.

**Определение 14.** *Пусть $g:\mathcal{X}\times\mathcal{Y}\to\{0,1,\perp\}$ — некоторая функция.*

- *Через $R^{\|}_{\varepsilon}(g)$ обозначим минимальное $d\in\mathbb{N}$ такое, что найдется SMP-протокол $\tau$, вычисляющий $g$ с ошибкой $\varepsilon$, для которого $CC(g)=d$.*
- *Через $R^{0,\|}_{\varepsilon}(g)$ обозначим минимальное $d\in\mathbb{N}$ такое, что найдется SMP-протокол $\tau$, вычисляющий $g$ с односторонней ошибкой $\varepsilon$, для которого $CC(g)=d$.*
- *Через $R^{1,\|}_{\varepsilon}(g)$ обозначим минимальное $d\in\mathbb{N}$ такое, что найдется SMP-протокол $\tau$, вычисляющий $\neg g$ с односторонней ошибкой $\varepsilon$, для которого $CC(g)=d$.*

Как оказывается, все эти величины не сильно зависят от $\varepsilon$. Более точно, при помощи стандартной техники амплификации (см., например, [29], упражнение 3.4) можно получить:

**Предложение 23.** *Для любых фиксированных $\varepsilon_1,\varepsilon_2\in(0,1/2)$ величины $R_{\varepsilon_1}(g)$ и $R_{\varepsilon_2}(g)$ отличаются не более, чем в константу раз (не зависящую от $g$). То же самое верно и для величин $R^{pr},R^0,R^1,R^{\|},R^{0,\|},R^{1,\|}$.*

В связи с этим мы иногда будем опускать индекс $\varepsilon$, имея в виду под $\varepsilon$ фиксированное число из интервала $(0,1/2)$, для определенности $\varepsilon=1/3$:

$$R(g)=R_{1/3}(g),$$

и аналогично для других видов вероятностной сложности.

Укажем очевидные неравенства, связывающие все эти величины. Эти неравенства основаны на следующих неформальных наблюдениях:

- любую функцию вычислить с односторонней ошибкой не легче, чем с двухсторонней;
- любую функцию вычислить SMP-протоколом не легче, чем обычным протоколом;
- любую функцию вычислить с частными случайными битами вычислить не легче, чем с общими.

Формализовать, например, первое наблюдение очень просто — любой протокол, вычисляющий функцию $g$ с односторонней ошибкой $\varepsilon$, вычисляет ее (просто) с ошибкой $\varepsilon$ (второе требование более слабое, чем первое). Так же просто формализуются другие два наблюдения, что дает следующие неравенства:

$$R(g)\le\min\{R^0(g),R^1(g)\},\qquad R^{\|}(g)\le\min\{R^{0,\|}(g),R^{1,\|}(g)\},$$
$$R(g)\le R^{\|}(g),\qquad R^0(g)\le R^{0,\|}(g),\qquad R^1(g)\le R^{1,\|}(g),$$
$$R(g)\le R^{pr}(g).$$

Наконец, определим еще две сложностные меры булевых функций. Для обоих будет предполагаться, что на входах задано некоторое вероятностное распределение. Нам, как и раньше, нужно вычислить некоторую функцию, при этом разрешается с некоторой вероятностью ошибаться. Отличие в том, что вероятность ошибки считается по *распределению на входах*, а не для каждой пары входов в отдельности.

Кроме того, в этой модели нас будет интересовать вычисление не только какой-то одной функции, а также $n$ *копий* одной функции.

Перейдем к точным определениям. Пусть $\nu$ — распределение вероятностей на $\mathcal{X} \times \mathcal{Y}$, где $\mathcal{X}, \mathcal{Y}$ — некоторые конечные множества, и пусть $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ — некоторая функция. Будем говорить, что вероятностный протокол $\tau$ над $(\mathcal{X}, \mathcal{Y})$ с пространством общих случайных битов $(\mathcal{R}, \mu)$ *вычисляет функцию $g$ с ошибкой $\varepsilon$ по распределению $\nu$*, если для любого $r$ можно определить функцию $f_r : \mathcal{L}_r \to \{0,1\}$, где $\mathcal{L}_r$ — множество листьев протокола $\tau_r$, так, что

$$\mathbb{E}_{(x,y)\sim\nu}\mathbb{E}_{r\sim\mu}\left[\sum_{l\in\mathcal{L}_r : f_r(l)\neq g(x,y)} p^{\tau_r}(l,x,y)\right] \leq \varepsilon.$$

**Определение 15.** *Через $D^{\nu}_{\varepsilon}(g)$ обозначим минимальное $d \in \mathbb{N}$ такое, что найдется детерминированный протокол с коммуникационной длиной $d$, вычисляющий $g$ с ошибкой $\varepsilon$ по распределению $\nu$.*

В последнем определении не случайно упоминаются только детерминированные протоколы. Случайные биты никак не могут помочь, потому что их можно зафиксировать так, чтобы ошибка по входному распределению не увеличилась. Иными словами, если использовать также и вероятностные протоколы, то величина $D^{\nu}_{\varepsilon}(g)$ не изменится. Тем не менее, мы дали определение вычисления с ошибкой по данному распределению для произвольных протоколов, поскольку будем использовать это понятие не только для детерминированных протоколов.

Между величинами $R_{\varepsilon}(g)$ и $D^{\nu}_{\varepsilon}(g)$ существует следующее интересное взаимоотношение, называемое принципом Яо:

**Предложение 24** ([29], Теорема 3.20)**.** *Для любой функции $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ выполнено:*

$$R_{\varepsilon}(g) = \max_{\nu} D^{\nu}_{\varepsilon}(g),$$

*где максимум берется по всем вероятностным распределениям $\nu$ на $\mathcal{X} \times \mathcal{Y}$.*

Неравенство $R_{\varepsilon}(g) \geq D^{\nu}_{\varepsilon}(g)$ почти что очевидно, но тем не менее весьма полезно — оно позволяет сводить нижние оценки на $R_{\varepsilon}(g)$ к анализу детерминированных протоколов на некотором «трудном» распределении $\nu$. То, что всегда найдется $\nu$, для которого выполнено равенство $R_{\varepsilon} = D^{\nu}_{\varepsilon}(g)$, означает универсальность метода трудных распределений.

Наконец, перейдем к определению сложности $n$ копий функции. Как и раньше, у нас фиксировано распределение $\nu$ на $\mathcal{X} \times \mathcal{Y}$, где $\mathcal{X}, \mathcal{Y}$ — конечные множества, а также фиксирована какая-то функция $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$. Теперь мы предполагаем, что Алиса получает на вход $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$, Боб получает на вход $y = (y_1, \ldots, y_n) \in \mathcal{Y}^n$, а их задача — вычислить вектор

$$(g(x_1, y_1), \ldots, g(x_n, y_n)).$$

Для этого Алиса и Боб выдают $n$-битовый вектор из нулей и единиц в надежде, что $i$-й бит равен $g(x_i, y_i)$. При этом каждая пара $(x_i, y_i)$ подается на вход с вероятностью $\nu((x_i, y_i))$, для разных $i$ независимо. Иными словами, мы рассматриваем на $\mathcal{X}^n \times \mathcal{Y}^n$ распределение $\nu^n$, являющееся распределением $n$ независимых копий $\nu$ и которое формально определяется следующим образом:

$$\nu^n((x,y)) = \nu((x_1,y_1)) \cdot \ldots \cdot \nu((x_n,y_n)) \qquad \forall (x,y) \in \mathcal{X}^n \times \mathcal{Y}^n.$$

Есть также тонкость в том, как мы будем считать вероятность ошибки. Можно было бы потребовать, чтобы весь вектор

$$(g(x_1, y_1), \ldots, g(x_n, y_n))$$

вычислялся правильно с вероятностью не меньше $1 - \varepsilon$ для какого-то $\varepsilon > 0$. Мы же будем добиваться того, что для каждого $i$ вероятность вычислить $g(x_i, y_i)$ правильно была не меньше $1 - \varepsilon$. Это не то же самое — может так оказаться, что при этом всегда хоть какая-то координата вычисляется неправильно (для разных исходов разная).

Переходя к формальному определению, скажем, что детерминированный протокол $\tau = \langle D, A, B, \phi, \psi \rangle$ над $(\mathcal{X}^n, \mathcal{Y}^n)$ *вычисляет $n$ копий функции $g$ с ошибкой $\varepsilon$ по распределению $\nu$*, если найдется функция $f : \mathcal{L}(D) \to \{0,1\}^n$ такая, что для любого $i \in \{1, \ldots, n\}$ выполнено:

$$\Pr_{(x,y)\sim\nu^n}[f(l_\tau(x,y))_i \neq g(x_i, y_i)] \leq \varepsilon,$$

где $l_\tau(x, y)$ — это лист, в который приходит $\tau$ на входе $(x, y) \in \mathcal{X}^n \times \mathcal{Y}^n$.

**Определение 16.** *Через $D^{\nu,n}_\varepsilon(g)$ обозначим минимальное $d \in \mathbb{N}$ такое, что найдется детерминированный протокол с коммуникационной длиной $d$, вычисляющий $n$ копий $g$ с ошибкой $\varepsilon$ по распределению $\nu$.*

То, как мы подсчитываем ошибку при определении величины $D^{\nu,n}_\varepsilon(g)$, позволяет выписать следующее неравенство:

$$D^{\nu,n}_\varepsilon(g) \leq n \cdot D^\nu_\varepsilon(g).$$

Действительно, для каждого $i \in \{1, \ldots, n\}$ Алиса и Боб могут запустить на $(x_i, y_i)$ протокол, доставляющий величину $D^\nu_\varepsilon(g)$. При этом вероятность вычислить $g(x_i, y_i)$ для каждого $i$ будет не меньше $1 - \varepsilon$, что нам и нужно.

## 1.3 Вопросная сложность

Гораздо более полным введением в вопросную сложность является обзор [14]. Здесь мы дадим лишь самые базовые определения, касающиеся только детерминированной вопросной сложности.

**Определение 17.** *Дерево разрешения $T$ от $n$ переменных задается*

- *двоичным деревом с корнем $D$;*
- *функцией $\phi$, которое каждой внутренней вершине $D$ сопоставляет элемент $\{1, \ldots, n\}$.*
- *функцией $\psi$, которое каждому листу $D$ сопоставляет элемент $\{0, 1\}$.*

*Глубиной дерева разрешения называется глубина $D$. Внутренние вершины (листья) $D$ будем также называть внутренними вершинами (листьями) $T$.*

Дерево разрешения от $n$ переменных производит вычисления следующим образом. Оно получает на вход строку $x = x_1 \ldots x_n$. Двигаясь от корня $D$ к листу, оно запрашивает значения каких-то битов $x$. Если текущая внутренняя вершина $v$ помечена числом $i$ (пометка определяется функцией $\phi$), то дерево переходит в вершину $vx_i$. В листе дерево выдает пометку листа (которая определяется функцией $\psi$). Максимальное количество запросов, которое дерево делает на одном входе, как легко понять, равно глубине дерева.

Более строго, в обозначениях определения 17 скажем, что дерево разрешения $T$ от $n$ переменных вычисляет функцию $f : \{0,1\}^n \to \{0,1\}$, если для любого $x$ в дереве $T$ существует лист $z_1 \ldots z_d$ такой, что

- $\psi(z_1 \dots z_d) = f(x)$;
- для любого $i \in \{0, 1, \dots, d-1\}$ выполнено равенство $z_{i+1} = x_{\phi(z_1 \dots z_i)}$.

**Определение 18.** *Вопросной сложностью функции* $f : \{0,1\}^n \to \{0,1\}$ *назовем минимальное* $d \in \mathbb{N}$ *такое, что найдется дерево разрешения от* $n$ *переменных глубины* $d$*, вычисляющее* $f$*. Обозначим вопросную сложность* $f$ *через* $D^{dt}(f)$*.*

# Глава 2

# Коммуникационная сложность задачи Gap Hamming Distance с односторонней ошибкой

Напомним, что через $d(x, y)$ мы обозначали расстояние Хемминга между двоичными словами одной длины. В данной главе мы изучаем следующую функцию

$$\mathrm{GHD}(n, L, U) : \{0,1\}^n \times \{0,1\}^n \to \{0, 1, \perp\},$$
$$\mathrm{GHD}(n, L, U)(x, y) = \begin{cases} 0 & \text{если } d(x, y) \leq L, \\ 1 & \text{если } d(x, y) \geq U, \\ \perp & \text{если } L < d(x, y) < U \end{cases}$$

Здесь $0 \leq L < U \leq n$ — некоторые целочисленные параметры. GHD является аббревиатурой для Gap Hamming Distance.

Как и просто вероятностная сложность этой функции, так и ее вероятностная сложность с односторонней ошибкой вызывают интерес начиная с 2000-х годов. Нижние оценки для GHD использовались для получении нижних оценок на память потоковых алгоритмов [41, 12])и в тестировании свойств [4].

Обычная вероятностная сложность (с «двухсторонней» ошибкой) функции $\mathrm{GHD}(n, L, U)$ известна с точностью до константного множителя когда $L + U = n$, т. е. когда $L$ и $U$ симметричны относительно $n/2$. А именно, выполнена оценка [15, 36, 39]:

$$R(\mathrm{GHD}(n, (1/2 - \gamma)n, (1/2 + \gamma)n)) = \Theta\left(\min\{n, 1/\gamma^2\}\right).$$

Известны также следующие верхние оценки на $R^{\|}(\mathrm{GHD}(n, L, U))$:

$$R^{\|}(\mathrm{GHD}(n, L, U)) = O(L^2/(U - L)^2), \tag{2.1}$$
$$R^{\|}(\mathrm{GHD}(n, L, U)) = O(L \log L) \tag{2.2}$$

(оценка (2.1) — из работ [30, 44], оценка (2.2) — из работы [26]). Вторая оценка не зависит от $U$ и выполнена, даже когда $U = L + 1$. Первая оценка при $U = L + 1$ хуже второй — она имеет вид $O(L^2)$. С другой стороны, при $U = (1 + \Omega(1))L$ первая оценка уже приобретает вид $O(1)$.

В литературе изучалась также и сложность GHD с односторонней ошибкой. Фольклорным результатом (см., например, [22]) является оценка

$$R^0(\mathrm{GHD}(n, L, U)) = O(L \log n). \tag{2.3}$$

Кроме того, в этом контексте изучалась сходная с $\mathrm{GHD}(n, L, U)$ функция, а именно:

$$\mathrm{EGHD}(n, L, U) : \{0,1\}^n \times \{0,1\}^n \to \{0, 1, \perp\},$$
$$\mathrm{EGHD}(n, L, U)(x, y) = \begin{cases} 0 & \text{если } d(x, y) = L, \\ 1 & \text{если } d(x, y) = U, \\ \perp & \text{иначе} \end{cases}$$

(отличие от GHD в том, что расстояние между $x$ и $y$ либо *в точности* $L$, либо *в точности* $U$). Для этой функции была получена оценка:

$$R^1(\mathrm{EGHD}(n, 0, U)) = \Omega(n), \qquad \text{при четном } U = (1 - \Omega(1))n, \tag{2.4}$$

из чего сразу вытекает (задача GHD сложнее EGHD) аналогичная оценка $R^1(\mathrm{EGHD}(n, 0, U)) = \Omega(n)$ при $U = (1 - \Omega(1))n$ (уже не обязательно четном).

Основным результатом данного раздела является почти точное (с точностью до множителя $O(\log L)$) выяснение вероятностной сложности функции $\mathrm{GHD}(n, L, U)$ с односторонней ошибкой:

**Теорема 1.** *Для любых целых $L, U, n$ таких, что $0 \le L < U \le n$ выполнены следующие оценки:*

- $R^0(\mathrm{GHD}(n, L, U)) = \Omega(L^2/U + 1)$;
- $R^{0,\|}(\mathrm{GHD}(n, L, U)) = O((L^2/U + 1)\log(L + 2))$.

Поскольку $R^0 \le R^{0,\|}$, эта теорема утверждает, что $R^0(\mathrm{GHD}(n, L, U))$ и $R^{0,\|}(\mathrm{GHD}(n, L, U))$ равны $L^2/U + 1$ с точностью до множителя $O(\log L)$.

При $U = L + 1$ наша оценка имеет ту же асимптотику, что и (2.2), в то время как наш протокол, в отличие от протокола из оценки (2.2), имеет одностороннюю ошибку. Мы также усиливаем оценку (2.3), поскольку наша верхняя оценка не зависит от $n$.

Помимо этого, наша нижняя оценка является уточнением оценки (2.4). А именно, легко видеть, что

$$R^1(GHD(n, 0, U)) = R^0(\mathrm{GHD}(n, n - U, n)) = \Omega((n - U)^2/n),$$

откуда при $U = (1 - \Omega(1))n$ получается (2.4) (правда только для GHD, а не для EGHD).

## 2.1 Нижняя оценка

Доказательство сводится к следующей лемме.

**Лемма 1.** *Для всех целых $U, n$ таких, что $0 < U \le n$, выполнена следующая оценка* $R^1(\mathrm{GHD}(n, 0, U)) = \Omega\left(\frac{(n-U)^2}{n} + 1\right)$.

Нужная нам нижняя оценка вытекает из этой леммы благодаря следующей очевидной цепочке неравенств и равенств:

$$\begin{aligned} R^0(\mathrm{GHD}(n, L, U)) &\ge R^0(\mathrm{GHD}(U, L, U)) \\ &= R^1(\mathrm{GHD}(U, 0, U - L)) = \Omega(L^2/U + 1). \end{aligned}$$

Остается доказать лемму.

*Доказательство леммы 1.* По определению найдется протокол $\tau$ с общими случайными битами длины $d = R^1(\mathrm{GHD}(n, 0, U))$, в котором Алиса и Боб получают по $n$-битовой строке и для которого выполнено следующее:

- на парах равных входов (эти пары мы будем в дальнейшем называть диагональными) с вероятностью $2/3$ протокол выдает 0;
- если расстояние между входами не меньше $U$, протокол всегда выдает 1.

В среднем (по выбору общих случайных битов) доля диагональных пар, на который протокол выдает 0, не меньше $2/3$. Значит общие случайные биты можно зафиксировать так, чтобы для полученного *детерминированного* протокола $\pi$ выполнялось следующее:

- доля диагональных пар, на которых протокол выдает 0, не меньше $2/3$;
- на всех парах входах, расстояние между которыми не меньше $U$, протокол выдает 1.

При этом длина $\pi$ не превосходит $d$. Рассмотрим листья, в которых $\pi$ выдает 0. Этих листьев не больше $2^d$. Хотя бы $(2/3)\cdot 2^n$ диагональных пар приходят в эти листья. Значит есть лист $l$, выдающий 0, в который приходят хотя бы $(2/3)\cdot 2^{n-d}$ диагональных пар. Через $A$ обозначим множество всех $x \in \{0,1\}^n$ таких, что $(x,x)$ приходит в $l$. С одной стороны, как мы показали, $|A| \geq (2/3)\cdot 2^{n-d}$. С другой стороны, $\operatorname{diam}(A) \leq U-1$. Действительно, в противном случае найдутся две строки $x, y \in \{0,1\}^n$ т. ч. $d(x,y) \geq U$ и $x, y \in A$ (что означает, что $(x,x)$ и $(y,y)$ приходят в $l$). Но тогда согласно предложению 20 в $l$ приходит и пара $(x,y)$. Это означает, что на $(x,y)$ протокол выдает 0 — противоречие.

Итак, $\operatorname{diam}(A) \leq U-1$. Положим $r = \lfloor U/2 \rfloor$. Без ограничения общности можно считать, что $U < n$ (при $U = n$ нижняя оценка, которую мы хотим доказать, является константой, и потому очевидно выполнена). Значит $2r \leq 2\cdot(U/2) < n$. Одновременно диаметр $A$ не превосходит $2r$ (потому что $U-1 = 2(U/2-1/2) \leq 2\lfloor U/2 \rfloor$). Значит по теореме Клейтмана (предложение 2) размер $A$ не превосходит $V_2(n,r)$. С другой стороны, поскольку $r < n/2$, для величины $V_2(n,r)$ выполнена оценка $V_2(n,r) \leq 2^{h(r/n)n}$ (предложение 1). В итоге мы получаем следующие неравенства:

$$(2/3)\cdot 2^{n-d} \leq |A| \leq 2^{h(r/n)n}. \tag{2.5}$$

Убедимся, что для некоторой константы $c > 0$ и для всех $\alpha \in [0, 1/2]$ выполнено неравенство $h(\alpha) \leq 1 - c(1/2-\alpha)^2$. Действительно, в силу того, что $h(1/2) = 1, h'(1/2) = 0$ и $h''(1/2) < 0$, найдутся две константы $\delta > 0$ и $c' > 0$ такие, что неравенство $h(\alpha) \leq 1 - c'(1/2-\alpha)^2$ выполнено для всех $\alpha \in [1/2-\delta, 1/2]$. С другой стороны, поскольку функция $h$ строго возрастает на $[0,1/2]$, то для всех $\alpha \in [0, 1/2-\delta]$ выполнено следующее:

$$h(\alpha) \leq h(1/2-\delta) = 1 - (1 - h(1/2-\delta)) \leq 1 - 4(1/2-\alpha)^2(1-h(1/2-\delta)).$$

Поэтому достаточно положить $c = \min\{c', 4(1-h(1/2-\delta))\}$. Применим доказанное неравенство при $\alpha = r/n$ (что корректно, поскольку $r < n/2$) к (2.5):

$$-\log_2(3/2) + n - d \leq (1 - c(1/2 - r/n)^2)n.$$

Отсюда вытекает:

$$d \geq c(1/2 - r/n)^2 \cdot n - \log_2(3/2).$$

Поскольку $r \leq U/2 < n/2$, то $U/(2n)$ не дальше от $1/2$, чем $r/n$, поэтому

$$R^1(\mathrm{GHD}(n,0,U)) = d \geq \frac{c(n-U)^2}{4n} - \log_2(3/2).$$

С другой стороны, очевидно выполнено оценка $R^1(\mathrm{GHD}(n,0,U)) \geq 1$. Складывая эти два неравенства с весами $1/100$ и $99/100$ получаем требуемую оценку:

$$R^1(\mathrm{GHD}(n,0,U)) = \Omega\left(\frac{(n-U)^2}{n} + 1\right).$$

□

## 2.2 Верхняя оценка

Итак, наша цель — построить SMP-протокол длины $O((L^2/U+1)\log(L+1)$, который получает на вход пару $n$-битовых строк и

- всегда выдает 0, если расстояние Хемминга между входами не больше $L$;
- с константной положительной вероятностью выдает 1, если расстояние Хемминга между входами не меньше $U$.

Мы не будем конкретизировать вероятность выдачи 1 во втором случае — мы лишь покажем, что она не меньше некоторой абсолютной константы $\delta > 0$. Стандартной техникой амплификации, увеличив длину протокола лишь в константу раз, можно добиться, чтобы эта вероятность была, скажем, не меньше 0.99.

Для начала зафиксируем обозначения для вероятностных распределений, которые мы будем использовать в доказательстве. Через $S_n$ обозначим распределение одномерного симметричного случайного блуждания с $n$ шагами. Более точно, $S_n$ — это распределение суммы $n$ независимых случайных величин, каждая из которых принимает два значения, 1 и $-1$, с одинаковыми вероятностями. Через $B(n,p)$ обозначим биномиальное распределение с параметрами $n$ и $p$, то есть сумму $n$ независимых Бернуллиевских случайных величин, каждая из которых принимает значение 1 с вероятностью $p$. Зафиксируем какое-нибудь $\alpha > 0$ такое, что для всех $m \in \mathbb{N}$ выполнено неравенство $\Pr[S_m \geq \sqrt{m}] \geq \alpha$.

Наш протокол является комбинацией трех различных протоколов. Самый важный из них, который мы будем называть HT-протоколом (поскольку он использует неравенство треугольника), имеет длину $O((L^2/U+1)\log n)$ и решает задачу $\mathrm{GHD}(n,L,U)$ с односторонней ошибкой для всех $L, U$ таких, что отношение $U/L$ не меньше некоторой фиксированной константы.

### 2.2.1 HT-протокол

Обозначим входные строки Алисы и Боба через $x, y \in \{0,1\}^n$ соответственно. Положим $b = \lceil CL^2/U + 1 \rceil$, где $C = 360/\alpha^2$. При любых $L, U$ число $b$ положительно. Игроки используют общий источник случайности, чтобы сгенерировать функцию $\chi : \{1,2,\ldots,n\} \to \{1,2,\ldots,b\}$ (с равномерным распределением по всем таким функциям). Игроки используют $\chi$ для разделения $x$ и $y$ на $b$ блоков:

$$x^1, \ldots, x^b, \qquad y^1, \ldots, y^b.$$

Координата $i$ попадает в блок $\chi(i)$. То есть $x^j$ получается из $x$ выбрасыванием всех координат $i \in \{1,2,\ldots,n\}$ таких, что $\chi(i) \neq j$. Аналогично определяется $y^j$. Заметим, что разбиение на блоки (в какой блок попадает каждая координата) известно всем участниками (как Алисе с Бобом, так и Чарли).

Также в общем источнике случайности независимо генерируется $b$ случайных строк $r^1$, …, $r^b$, строка $r^j$ генерируется с равномерным распределением среди всех строк той же длины, что у $x^j$ и $y^j$. Алиса посылает Чарли расстояния Хемминга от $x^j$ до $r^j$ по всем $j \in \{1,2,\ldots,b\}$. То же самое делает Боб, только он посылает расстояния от $y^j$ до $r^j$. Каждый участник таким образом посылает Чарли по $b$ чисел от 0 до $n$. Значит длину описанного протокола можно оценить сверху величиной $O(b\log n) = O((L^2/U+1)\log n)$. Коммуникация на этом заканчивается, далее Чарли вычисляет ответ. А именно, он вычисляет величину

$$T = \sum_{j=1}^{b} \left| d(x^j, r^j) - d(y^j, r^j) \right|.$$

Если $T$ не превосходит $L$, Чарли выдает 0, иначе 1. Убедимся, что на $(x, y)$ таких, что $d(x, y) \leq L$, Чарли всегда выдает 0. Действительно, при таких $(x, y)$ величина $T$ всегда не превосходит $L$, что легко следует из неравенства треугольника:

$$T = \sum_{j=1}^{b} \left| d(x^j, r^j) - d(y^j, r^j) \right| \leq \sum_{j=1}^{b} d(x^j, y^j) = d(x, y) \leq L.$$

Для того, чтобы оценить вероятность выдачи 1 на парах входов, находящихся на расстоянии хотя бы $U$, докажем следующую лемму:

**Лемма 2.** *Предположим $U \geq 2b$ и $d(x, y) \geq U$, где $x, y$ — какие-то $n$-битовые строки. Тогда на $(x, y)$ HT-протокол выдает 1 с вероятностью хотя бы $\alpha/6$.*

*Доказательство.* Пусть фиксированы $x, y$ т.ч. $d(x, y) \geq U$. Как распределена случайная величина $d(x^j, y^j)$? Каждая координата, в которой $x$ и $y$ различается, независимо от других координат с вероятностью $1/b$ попадает в $j$-й блок. Таким образом, $d(x^j, y^j)$ есть сумма $d(x, y)$ независимых Бернуллиевских случайных величин, каждая из которых принимает значение 1 с вероятностью $1/b$. Иными словами, $d(x^j, y^j) \sim B(d(x, y), 1/b)$. Предположим, разбиение на блоки фиксировано. Как тогда распределена разность $d(x^j, r^j) - d(y^j, r^j)$? Предположим длина $j$-го блока равна $l$. Тогда

$$d(x^j, r^j) - d(y^j, r^j) = \sum_{i=1}^{l} \left( |x_i^j - r_i^j| - |y_i^j - r_i^j| \right).$$

В этой сумме $l$ слагаемых. Эти слагаемые независимы друг от друга (они зависят от разных координат $r^j$). Если $x_i^j = y_i^j$, то соответствующее слагаемое всегда равно 0. Если же $x_i^j \neq y_i^j$, то соответствующее слагаемое принимает два возможных значения, 1 и $-1$, оба с вероятностью $1/2$. Таким образом при фиксированном разбиении на блоки мы имеем $d(x^j, r^j) - d(y^j, r^j) \sim S_{d(x^j, y^j)}$. Далее нам понадобится следующая лемма о биномиальном распределении:

**Лемма 3.** *Если $X \sim B(n, p)$ и $pn \geq 2$, то*

$$\Pr\left[X > \frac{pn}{10}\right] \geq \frac{1}{3}.$$

*Доказательство.* Матожидание и дисперсия $X$ имеют следующий вид:

$$\mathbb{E}X = pn, \qquad \mathbb{D}V = p(1-p)n \leq pn.$$

Поэтому по неравенству Чебышева:

$$\Pr\left[X \leq \frac{pn}{10}\right] \leq \frac{\mathbb{D}V}{\left(pn - \frac{pn}{10}\right)^2} \leq \frac{\frac{100}{81}}{pn} \leq \frac{100}{162} \leq \frac{2}{3}.$$

□

Зафиксируем $j \in \{1, 2, \ldots, b\}$. Поскольку $d(x, y)/b \geq U/b \geq 2$, то из этой леммы вытекает, что с вероятностью хотя бы $1/3$ выполнено неравенство $d(x^j, y^j) \geq d(x, y)/(10b)$. С другой стороны, при любом фиксированном разбиении на блоки с вероятностью хотя бы $\alpha$ выполнено $d(x^j, r^j) - d(y^j, r^j) \geq \sqrt{d(x^j, y^j)}$.

Назовем $j$-й блок хорошим, если $|d(x^j, r^j) - d(y^j, r^j)| \geq \sqrt{d(x, y)/(10b)}$. Из сказанного в предыдущем параграфе следует, что для любого $j \in \{1, 2, \ldots, b\}$ с вероятностью хотя бы $\alpha/3$ блок $j$ является хорошим. Отсюда следует, что с вероятностью хотя бы $\alpha/6$ больше чем $(\alpha/6)b$

блоков хорошие. Действительно, обозначим долю хороших блоков через $\theta$. Матожидание $\theta$, как мы показали, не меньше $\alpha/3$. Поэтому

$$\alpha/3 \le \mathbb{E}\theta \le \Pr[\theta > \alpha/6] \cdot 1 + \Pr[\theta \le \alpha/6] \cdot (\alpha/6) \le \Pr[\theta > \alpha/6] + \alpha/6,$$

откуда $\Pr[\theta > \alpha/6] \ge \alpha/6$. Таким образом с вероятностью $\alpha/6$ величина $T$ больше

$$T > (\alpha/6)b\sqrt{d(x,y)/(10b)} \ge \sqrt{\frac{Ub}{\frac{360}{\alpha^2}}} \ge \sqrt{\frac{U \cdot \frac{CL^2}{U}}{\frac{360}{\alpha^2}}} = L.$$

Значит с вероятностью хотя бы $\alpha/6$ Чарли выдает 1. □

Убедимся, что если $U \ge 2CL + 4$, то $U \ge 2b$. Действительно, в этом случае

$$2b = 2\left\lceil CL^2/U + 1\right\rceil \le \frac{2CL^2}{U} + 4 \le 2CL + 4 \le U.$$

Таким образом, нами доказана оценка $R^{0,\|}(\mathrm{GHD}(n, L, U)) = O((L^2/U + 1)\log n)$ при $U \ge 2CL + 4$. Наша следующая задача состоит в том, чтобы доказать оценку $R^{0,\|}(\mathrm{GHD}(n, L, L + 1)) = O((L + 1)\log n)$. Из этого будет вытекать, что оценка

$$R^{0,\|}(\mathrm{GHD}(n, L, U)) = O((L^2/U + 1)\log n)$$

выполнена при всех $L < U$ (действительно, при $U < 2CL+4$ выполнено $L+1 = O(L^2/U+1)$).

### 2.2.2 Протокол для $\mathrm{GHD}(n, L, L + 1)$

Оценка $R^{0,\|}(\mathrm{GHD}(n, L, L + 1)) = O((L + 1)\log n)$ по сути была доказана в [22] (только в этой работе не было явно сформулировано, что эта оценка выполнена и для односторонней ошибки). Для полноты изложения приведем протокол для этой оценки.

Введем следующие обозначения. Пусть $a, b \in \{0, 1\}^n$, $a = a_1 \dots a_n$, $b = b_1 \dots b_n$. Через $a \oplus b$ обозначим результат побитового сложения по модулю 2 строк $a$ и $b$. Также через $\langle \cdot, \cdot \rangle$ скалярное произведение $a$ и $b$ по модулю 2:

$$\langle a, b\rangle = \left(\sum_{i=1}^{n} a_i b_i\right) \pmod 2.$$

Предположим Алиса и Боб получают на вход $n$-битовые строки $x$ и $y$ соответственно. Используя общие случайные биты, Алиса и Боб генерируют $N$ независимых $n$-битовых строк $R^1, \dots, R^N$, каждую с равномерным распределением (значение $N$ мы укажем явно позднее). Алиса посылает Чарли $\langle x, R^1\rangle, \dots, \langle x, R^N\rangle$. Боб посылает Чарли $\langle y, R^1\rangle, \dots, \langle y, R^N\rangle$. Чарли проверяет, существует ли $n$-битовая строка $h$ с не более чем $L$ единицами такая, что

$$\langle x \oplus h, R^1\rangle = \langle y, R^1\rangle, \dots, \langle x \oplus h, R^N\rangle = \langle y, R^1\rangle. \tag{2.6}$$

В силу линейности $\langle \cdot, \cdot \rangle$ по первому аргументу Чарли действительно в состоянии проверить наличие такого $h$. Если оно найдется, Чарли выдает 0, иначе 1.

Данный протокол никогда не ошибается на $(x, y)$ т. ч. $d(x, y) \le L$. Действительно, в качестве $h$ тогда можно взять $h = x \oplus y$.

С другой стороны, предположим $d(x, y) \ge L + 1$. Зафиксируем какую-нибудь $h \in \{0, 1\}^n$ с не более чем $L$ единицами. Поскольку $x \oplus h \ne y$ (иначе $d(x, y) \le L$), то вероятность (по выбираемой с равномерным распределением $n$-битовой строке $R$) того, что будет выполнено равенство $\langle x \oplus h, R\rangle = \langle y, R\rangle$, равна $1/2$. Значит вероятность того, что $h$ удовлетворит (2.6), равна $2^{-N}$. Суммируя по всем $h$ с не более чем $L$ единицами, получаем оценку $V_2(n, L)2^{-N}$ на вероятность ошибки в случае $d(x, y) \ge L+1$. Напомним, что $V_2(n, L)$ — размер шара Хемминга радиуса $L$. Поскольку $V_2(n, L) \le (n + 1)^L$, то достаточно положить $N = \lceil L \log_2(n + 1)\rceil + 2$, чтобы эта вероятность не превосходила $1/3$. При этом длина протокола будет равна $O(N) = O((L + 1)\log n)$, что и требовалось.

### 2.2.3 Модификация HT-протокола

Итак, нами доказана на данный момент для всех $L, U$ оценка

$$R^{0,\|}(\mathrm{GHD}(n, L, U)) = O((L^2/U + 1)\log n).$$

Далее мы покажем, как уменьшить множитель $\log n$ до $\log L$. Для этого нам понадобится отдельно доказать оценку:

$$R^{0,\|}(\mathrm{GHD}(n, L, (4L + 2 + N_0)^4)) = O(\log(L + 2)). \tag{2.7}$$

Здесь $N_0$ — это константа из следующей технической леммы, которую мы докажем в конце этого подраздела.

**Лемма 4.** *Найдутся положительное целое число $N_0$ и положительное число $c$, для которых выполнено следующее. Предположим $m$ и $N$ — это какие-то положительные целые числа, удовлетворяющие неравенству $N \geq \max\{N_0, m^2\}$. Рассмотрим $N$ независимых Бернуллиевских случайных величин $Z_1, \ldots, Z_N$, каждая из которых принимает значение 1 с вероятностью $1/2$. Тогда для любого $i \in \{0, 1, \ldots, m-1\}$ выполнено неравенство:*

$$\Pr[Z_1 + \ldots + Z_N = i \pmod m] \geq c/m.$$

Отметим, что оценка (2.7) является частными случаем нужной нам верхней оценки из теоремы 1. Теперь опишем протокол, доставляющий (2.7). Этот протокол будет заимствовать идею измерения расстояния до случайной строки и использования неравенства треугольника из HT-протокола.

Предположим, входами Алисы и Боба являются $n$-битовые строки $x$ и $y$ соответственно. При помощи общих случайных битов генерируется строка $R \in \{0,1\}^n$ с равномерным распределением. Алиса вычисляет $a = d(x, R)$, Боб вычисляет $b = d(y, R)$. По неравенству треугольника $|a - b| \leq d(x, y)$. Поэтому если $d(x, y) \leq L$, то $|a - b| \leq L$. С другой стороны, если $d(x, y) \geq (4L + 2 + N_0)^4$, то с вероятностью хотя бы $\alpha$ разница $|a - b|$ не меньше корня из $d(x, y)$, то есть не меньше $(4L + 2 + N_0)^2 > (4L + 2)^2 + N_0^2$ (см. аналогичные рассуждения о случайном блуждании из подраздела о HT-протоколе).

Таким образом оценка (2.7) сводится к следующей задаче. Алиса получает на вход число $a \in \{0, 1, \ldots, n\}$, Боб получает на вход число $b \in \{0, 1, \ldots, n\}$. Гарантируется, что либо $|a - b| \leq L$, либо $|a - b| > (4L + 2)^2 + N_0^2$. В первом случае нужно всегда выдавать 0, во втором — с вероятностью, не меньшей некоторой абсолютной положительной константы, нужно выдавать 1. Остается доказать существование SMP-протокола с общими случайными битами длины $O(\log(L + 2))$, удовлетворяющий этим требованиям.

Этот протокол строится следующим образом. Положим $m = 4L + 2$. При помощи общих случайных битов генерируется $n + 1$ независимая бернуллиевская случайная величина

$$Z_0, Z_1, \ldots, Z_n,$$

каждая из которых принимает значение 1 с вероятностью $1/2$.

Алиса посылает Чарли число $\sum_{i=0}^{a} Z_i \pmod m$. Боб посылает Чарли число $\sum_{i=0}^{b} Z_i \pmod m$. На этом коммуникация заканчивается. Длина протокола составляет $O(\log m) = O(\log(L+2))$ бит, как и требовалось.

Чарли находит какую-нибудь пару целых чисел $(s, t)$, удовлетворяющих следующим трем условиям:

$$s \equiv \sum_{i=0}^{a} Z_i \pmod m \tag{2.8}$$

$$t \equiv \sum_{i=0}^{b} Z_i \pmod{m} \tag{2.9}$$

$$|s-t| = \min\left\{|s'-t'| : s' \equiv \sum_{i=0}^{a} Z_i \pmod{m}, t' \equiv \sum_{i=0}^{b} Z_i \pmod{m}\right\}. \tag{2.10}$$

Затем Чарли проверяет, верно ли, что $|s-t| \le L$. Если верно, он выдает 0, иначе 1.

Проверим, что если $|a-b| \le L$, то Чарли всегда выдает 0. Действительно, пара

$$(\sum_{i=0}^{a} Z_i, \sum_{i=0}^{b} Z_i)$$

удовлетворяет (2.8), (2.9), а значит из (2.10) вытекает:

$$|s-t| \le \left|\sum_{i=0}^{a} Z_i - \sum_{i=0}^{b} Z_i\right| \le |a-b|.$$

Таким образом из $|a-b| \le L$ вытекает $|s-t| \le L$ (это и означает, что Чарли всегда выдает 0).

Теперь предположим, что $|a-b| > (4L+2)^2 + N_0^2$. Пусть $a < b$ (случай $a > b$ рассматривается аналогично). Надо показать, что с вероятностью, не меньшей некоторой положительной абсолютной константы, выполнено $|s-t| > L$. Через $E$ обозначим событие, заключающееся в том, что не нашлось целого $r \in [-L, L]$ такого, что $Z_{a+1} + \ldots + Z_b \equiv r \pmod{m}$. Заметим, что если $E$ выполнено, то $|s-t| > L$. Действительно, для $r = t - s$ благодаря (2.8), (2.9) выполнено равенство $Z_{a+1} + \ldots + Z_b \equiv r \pmod{m}$. Поэтому если $|t-s| \le L$, то $E$ не выполнено.

Для окончания доказательства остается продемонстрировать, что $\Pr[E]$ не меньше некоторой положительной абсолютной константы. Для этого мы воспользуемся леммой 4, примененной к $Z_{a+1}, \ldots, Z_b$. Чтобы ее можно было применить, должно выполняться неравенство $N = b - a \ge \max\{N_0, m^2\}$. Это неравенство вытекает из условия $|b-a| > (4L+2)^2 + N_0^2 = m^2 + N_0^2$. Если $Z_{a+1} + \ldots + Z_b \equiv i \pmod{m}$ и $i$ не сравнимо с $r$ по модулю $m$ ни для какого целого $r \in [-L, L]$, то событие $E$ выполнено. Количество таких $i$ не меньше $m - (2L+1) = 2L+1$. Значит по лемме 4:

$$\Pr[E] \ge (2L+1) \cdot \frac{c}{m} = \frac{c}{2}.$$

*Доказательство леммы 4.* Положим $N_0$ наименьшему натуральному числу такому, что

$$[N/2 - \sqrt{N}, N/2 + \sqrt{N}] \subset [0, N]$$

для всех $N \ge N_0$. Найдется $d > 0$ такое, что для всех целых $N \ge N_0$ и $k \in [N/2 - \sqrt{N}, N/2 + \sqrt{N}]$ выполнено:

$$\Pr[Z_1 + \ldots + Z_N = k] = \binom{N}{k} 2^{-N} \ge \frac{d}{\sqrt{N}}.$$

Существование такого $d$ легко вытекает из формулы Стирлинга, примененной к $\binom{N}{k}$.

Теперь покажем, что для всех $m > 0, N \ge m^2$ и $i \in \{0, 1, \ldots, m-1\}$ количество целых $k \in [N/2 - \sqrt{N}, N/2 + \sqrt{N}]$ таких, что $k \equiv i \pmod{m}$, не меньше $\sqrt{N}/m$. Количество таких $k$ равно количеству $r \in \mathbb{Z}$ таких, что

$$N/2 - \sqrt{N} \le mr + i \le N/2 + \sqrt{N},$$

а это количество не меньше:

$$\left\lfloor \frac{N/2+\sqrt{N}-i}{m} \right\rfloor - \left\lceil \frac{N/2-\sqrt{N}-i}{m} \right\rceil + 1 \geq \frac{2\sqrt{N}}{m} - 1.$$

При условии $N \geq m^2$, последнее выражение не меньше $\sqrt{N}/m$.

Таким образом, при $N \geq \max\{N_0, m^2\}$ для любого $i \in \{0, \ldots, m-1\}$ получаем:

$$\begin{aligned}\Pr[Z_1 + \ldots + Z_N = i \pmod m] &\geq \sum_{\substack{k \in [N/2-\sqrt{N}, N/2+\sqrt{N}] \\ k \equiv i \pmod m}} \Pr[Z_1 + \ldots + Z_N = k] \\ &\geq (\sqrt{N}/m) \cdot \frac{d}{\sqrt{N}} = \frac{d}{m}.\end{aligned}$$

Остается положить $c = d$. □

### 2.2.4 Финальный протокол

В данном подразделе мы наконец опишем протокол, доказывающий оценку

$$R^{0,\|}(\mathrm{GHD}(n, L, U)) = O((L^2/U + 1)\log(L+2)).$$

Предположим, Алиса и Боб получают на вход $n$-битовые строки $x$ и $y$ соответственно.

*Шаг 1.* Сначала игроки на $(x, y)$ запускают модификацию HT-протокола (имеется в виду протокол, описанный в предыдущем подразделе для получения оценки (2.7)).

*Шаг 2.* Положим $w = 2(4L + 2 + N_0)^8$. Алиса и Боб разбивают $x$ и $y$ на $w$ случайных блоков (таким же способом, как и при построении HT-протокола). Обозначим получившиеся блоки через $x^1, \ldots, x^w$ и $y^1, \ldots, y^w$. Пусть $u_i$ будем суммой по модулю 2 всех битов строки $x^i$, а $v^i$ — суммой по модулю 2 всех битов строки $y^i$. Алиса кладет $u = u_1 \ldots u_w$, Боб кладет $v = v_1 \ldots v_n$.

Напомним также, что нами уже доказана оценка

$$R^{0,\|}(\mathrm{GHD}(n, L, U)) = O((L^2/U + 1)\log(n)).$$

Игроки запускают протокол, доказывающий эту оценку, при $n = w$ на паре $w$-битовых строк $(u, v)$.

На первый шаг уходит $O(\log(L+2))$ битов коммуникации, на второй шаг уходит $O((L^2/U+1)\log(w)) = O((L^2/U+1)\log(L+2))$ битов, как раз сколько нужно. Если либо в протоколе из первого шага, либо в протоколе из второго шага выдается 1, то и в качестве окончательного ответа на задачу $\mathrm{GHD}(n, L, U)$ Чарли выдает 1. В противном случае Чарли выдает 0.

Последнее условие означает, что Чарли всегда выдает 0, когда $d(x, y) \leq L$. Действительно, протоколы из первого и второго шага всегда выдают 0, когда у них на входе пара строк, расстояние Хемминга между которыми не превосходит $L$. Правда, протокол из второго шага запускается не на $(x, y)$, а на $(u, v)$. Это не проблема — легко понять, что $d(u, v) \leq d(x, y)$. Действительно, если $u^i \neq v^j$, то и хотя бы в одном месте блок $x^i$ должен отличаться от блока $y^i$. А значит если $d(x, y) \leq L$, то и $d(u, v) \leq L$.

Остается доказать, что если $d(x, y) \geq U$, то с вероятностью не меньше абсолютной положительной константы на одном из шагов выдается 1. Мы рассмотрим два случая.

*Случай 1.* Предположим $d(x, y) \geq (4L+2+N_0)^4$. Тогда построенный в предыдущем подразделе протокол, доказывающий оценку (2.7), на $(x, y)$ с вероятностью не меньше абсолютной положительной константы выдает 1.

*Случай 2.* Предположим $U \leq d(x,y) \leq (4L+2+N_0)^4$. Достаточно доказать, что с вероятностью $1/2$ выполнено $d(u,v) = d(x,y)$. Из этого будет вытекать, что $d(u,v) \geq U$, а значит протокол, запускаемый на втором шаге, выдаст 1 с вероятностью не меньше абсолютной положительной константы .

Рассмотрим любые две координаты, в которых $x, y$ отличаются. Вероятность того, что эти две координаты попадут в один блок, равна $1/w$. Всего координат, в которых $x, y$ отличаются, не больше $(4L+2+N_0)^4$. Значит с вероятностью не меньше

$$1 - ((4L+2+N_0)^4)^2 \cdot (1/w) = 1/2$$

все координаты, в которых $x$ и $y$ различаются, попадут в разные блоки. Но тогда для всех $i \in \{1, 2, \ldots, w\}$ блок $x^i$ будет отличаться от $y^i$ максимум в одной координате. Значит если $u_i \neq v_i$, то $d(x^i, y^i) = 1$, а если $u_i = v_i$, то $d(x^i, y^i) = 0$. Их этого вытекает, что $d(u,v) = d(x,y)$.

# Глава 3

# Связь вопросной и коммуникационной сложности

Эта глава посвящена различным обобщениям теоремы Раза — Маккинзи. Для начала сформулируем эту теорему.

Композицией $f : \{0,1\}^n \to \{0,1\}$ и $g : \mathcal{X} \times \mathcal{Y} \to \{0,1,\perp\}$ назовем функцию $f \circ g$: $\mathcal{X}^n \times \mathcal{Y}^n \to \{0,1,\perp\}$, которая на входе $((x_1,\dots,x_n),(y_1,\dots,y_n))$ равна $f(g(x_1,y_1),\dots,g(x_n,y_n))$, если $g(x_1,y_1),\dots,g(x_n,y_n) \in \{0,1\}$, и $\perp$ иначе. При этом $f$ мы будем называть внешней функцией $f \circ g$, а $g$ — гаджетом $f \circ g$. Число $n$ мы будем называть арностью $f$, а число $k = \max\{\lceil \log_2 |\mathcal{X}| \rceil, \lceil \log_2 |\mathcal{Y}| \rceil\}$ — длиной входа $g$.

Мы будем интересоваться взаимосвязью детерминированной вопросной сложности $f$ и детерминированной коммуникационной сложности $g$ и $f \circ g$. Уточним, что говоря о коммуникационной сложности $f \circ g$, мы имеем в виду, что Алиса получает на вход элемент $\mathcal{X}^n$, а Боб — элемент $\mathcal{Y}^n$.

Очевидно, выполнено следующее неравенство $D(f \circ g) \le D^{dt}(f) \cdot D(g)$. Действительно, дерево разрешения глубины $d$, вычисляющее $f$, можно преобразовать в протокол длины $d \cdot D(g)$, вычисляющий $f \circ g$. А именно, Алиса и Боб симулируют дерево разрешения. Если оно делает запрос к $i$-й переменной, то Алисе и Бобу надо вычислить значение $g(x_i, y_i)$. На это уходит не более $D(g)$ бит.

Для некоторых $f$ и $g$ можно доказать, что и обратное неравенство

$$D(f \circ g) \ge D^{dt}(f) \cdot D(g)$$

выполнено с точностью до мультипликативной константы. А именно, определим функцию

$$\mathrm{IND}_k : \{0,1\}^k \times \{1,2,\dots,k\} \to \{0,1\},$$
$$\mathrm{IND}_k(y_1 \dots y_k, x) = y_x.$$

Известно, что $D(\mathrm{IND}_k) = \Theta(\log k)$. Для этой функции выполнена следующая

**Теорема 2** (Раз — Маккинзи, [33, 24])**.** *Для всех $f : \{0,1\}^n \to \{0,1\}$ арности не более $k^{1/20}$ выполнено:*

$$D(f \circ \mathrm{IND}_k) = \Omega(D^{dt}(f) \cdot \log k) = \Omega(D^{dt}(f) \cdot D(\mathrm{IND}_k)).$$

Нас будет интересовать вопрос, насколько можно улучшить оценку на арность внешней функции в теореме Раза — Маккинзи (в зависимости от длины входа гаджета). При этом можно рассматривать и другие гаджеты. Чтобы поставить этот вопрос совсем строго, назовем функцию $n : \mathbb{N} \to \mathbb{R}$ *достижимой оценкой на арность в теореме Раза — Маккинзи*, если найдется константа $c > 0$ такая, что для бесконечно многих $k$ найдется гаджет $g_k$ с длиной входа $k$ такой, что неравенство

$$D(f \circ g) \ge c \cdot D^{dt}(f) \cdot D(g)$$

выполнено для всех внешних функций $f$ арности не более $n(k)$. Из теоремы 2 вытекает, что функция $k \mapsto k^{1/20}$ является достижимой оценкой на арность в теореме Раза — Маккинзи. Можно ли то же самое доказать для более быстро растущих функций?

Продвижение в этом направлении было достигнуто в работах Чаттоподхая и др. [16] и Ву и др. [42]. А именно, Чаттопадхай и др. обобщили теорему 2 следующим образом. Пусть $M$ — некоторая матрица с множеством строк $\mathcal{X}$ и множеством столбцов $\mathcal{Y}$, элементами которой являются $0, 1, \perp$. Подматрицу $M$ будем называть *$i$-одноцветной* (где $i \in \{0,1\}$), если все элементы этой подматрицы равны $i$. Будем говорить, что распределение вероятностей $\mu$ на подматрицах $M$ является *$(\delta, h)$-протыкающим* для матрицы $M$ (здесь $\delta \in (0,1)$ и $h \in \mathbb{N}$), если для любых $A \subset \mathcal{X}, B \subset \mathcal{Y}$, т. ч. $|A| \geq 2^{-h}|\mathcal{X}|, |B| \geq 2^{-h}|\mathcal{Y}|$, выполнено следующее: если подматрица $R$ выбирается случайно по распределению $\mu$, то $A \times B$ с вероятностью не меньше $1 - \delta$ имеет с $R$ общий элемент.

Для функции $g : \mathcal{X} \times \mathcal{Y} \to \{0, 1, \perp\}$ определим матрицу $M_g$ с множеством строк $\mathcal{X}$, множеством столбцов $\mathcal{Y}$, в которой на пересечении строки $x \in \mathcal{X}$ и столбца $y \in \mathcal{Y}$ стоит $g(x, y)$.

**Теорема 3** (Чаттопадхай и др., [16])**.** *Предположим для гаджета $g\colon \mathcal{X} \times \mathcal{Y} \to \{0, 1, \perp\}$ и для любого $i \in \{0, 1\}$ найдется $\left(\frac{1}{1000}, h\right)$-протыкающее распределение вероятностей $\mu_i$ для $M_g$, сосредоточенное на $i$-одноцветных подматрицах $M_g$. Тогда для любого $\varepsilon \geq 6/h$ и любой функции $f : \{0,1\}^n \to \{0,1\}$ арности не более $2^{(1-\varepsilon)h}$ выполнено неравенство:*

$$D(f \circ g) \geq \frac{\varepsilon h}{4} \cdot D^{dt}(f).$$

В доказательстве некоторым явным (но не очень эффективным) способом произвольный протокол для $f \circ g$ преобразуется в дерево разрешения для $f$ (глубины примерно в $h$ раз меньше).

Эта теорема означает, что для того, чтобы гаджет удовлетворял аналогу теорему Раза — Маккинзи, достаточно наличия у этого гаджета протыкающих распределений. Результаты настоящей главы в основном как раз и посвящены протыкающим распределениям. В начале сформулируем, что было про них известно.

Чаттопадхай и др. для гаджета $\mathrm{IP}_k$, называемого *функцией скалярного произведения* и определяемого следующим образом:

$$\mathrm{IP}_k : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}, \qquad \mathrm{IP}_k(x, y) = \sum_{i=1}^{n} x_i y_i \pmod 2,$$

построили два $\left(\frac{1}{1000}, k/2 - O(1)\right)$-протыкающих распределения, одно сосредоточенное на 0-одноцветных подматрицах $M_{\mathrm{IP}_k}$, другое сосредоточенное на 1-одноцветных подматрицах $M_{\mathrm{IP}_k}$. Благодаря теореме 3 это означает, что для любого $\varepsilon > 0$ и любой функции $f : \{0,1\}^n \to \{0,1\}$ арности не более $2^{(1/2-\varepsilon)k}$ выполнено неравенство:

$$D(f \circ \mathrm{IP}_k) = \Omega(D^{dt}(f) \cdot k) = \Omega(D^{dt}(f) \cdot D(\mathrm{IP}_k)),$$

где константа в $\Omega(\cdot)$ зависит только от $\varepsilon$. При этом длина входа в $\mathrm{IP}_k$ равна $k$. Это означает, что для любого $\varepsilon > 0$ функция $k \mapsto 2^{(1/2-\varepsilon)k}$ является достижимой оценкой на арность в теореме Раза — Маккинзи. Как видим, оценка на арность внешней функции для гаджета $\mathrm{IP}_k$ экспоненциально лучше, чем у гаджета $\mathrm{IND}_k$. То же самое для гаджета $\mathrm{IP}_k$ независимо доказали Ву и др. [42] (с намного меньшей константой перед $k$ в экспоненте). Кроме того, Чаттопадхай и др. построили $\left(\frac{1}{1000}, 0.45k\right)$-протыкающие распределения на одноцветных подматрицах обоих типов для гаджета $g = \mathrm{GHD}(k, k/4, 3k/4)$, из чего получается чуть более слабые, чем для $\mathrm{IP}_k$, оценки на арность внешней функции.

## 3.1 Результаты

Мы начнем с нового метода получения гаджетов с хорошими протыкающими распределениями из спектральных экспандеров. Пусть в неориентированном графе $G$ с множеством вершин $V$ и множеством ребер $E$ у любых двух различных вершин $u$ и $v$ есть не более одного общего соседа (вершины, соединенной и с $u$, и с $v$ ребром в $G$). Зафиксируем также функцию $c : V \to \{0,1\}$. Тогда корректно определена следующая функция

$$g(G,c) : V \times V \to \{0,1,\perp\}$$
$$g(G,c)(u,v) = \begin{cases} 1 & u \neq v \text{ и } \exists w \in \Gamma(u) \cap \Gamma(v) \text{ т. ч. } c(w) = 1, \\ 0 & u \neq v \text{ и } \exists w \in \Gamma(u) \cap \Gamma(v) \text{ т. ч. } c(w) = 0, \\ \perp & u = v \text{ или } \Gamma(u) \cap \Gamma(v) = \varnothing. \end{cases}$$

Иными словами, здесь $c$ воспринимается как раскраска вершин графа в два цвета. Функция $g(G,c)$ определена только на парах различных вершин, имеющих общего соседа (и тогда этот сосед единственный). Значение $g(G,c)$ на такой паре равно цвету общего соседа.

Назовем $c : V \to \{0,1\}$ сбалансированной раскраской $G$, если $|V|/3 \leq |c^{-1}(1)| \leq 2|V|/3$.

Мы показываем, что если $G$ — хороший спектральный экспандер, а $c$ — сбалансированная раскраска $G$, то $g(G,c)$ обладает хорошими протыкающими распределениями над одноцветными подматрицами обоих типов. Более точно, мы докажем следующую теорему.

**Теорема 4.** *Пусть граф $G = (V,E)$ является спектральным $(m,d,\gamma)$-экспандером, в котором у любых двух различных вершин есть не более одного общего соседа, а $c : V \to \{0,1\}$ является сбалансированной раскраской $G$. Пусть также $m \geq 1/\gamma^2$. Тогда для любого $i \in \{0,1\}$ для матрицы $M_{g(G,c)}$ существует $\left(\frac{1}{1000}, \lfloor 2\log_2(1/\gamma)\rfloor - 100\right)$-протыкающее распределение $\mu_i$, сосредоточенное на $i$-одноцветных подматрицах $M_{g(G,c)}$.*

Константа $\frac{1}{1000}$ не имеет особого значения. Легко понять из доказательства, что для любой константы $\delta > 0$ найдется натуральное число $r$ такое, что распределения $\mu_i$ из теоремы 4 будут $(\delta, \lfloor 2\log_2(1/\gamma)\rfloor - r)$-протыкающими. Тем не менее, для применений нам будет достаточно данной выше формулировки теоремы 4.

Более того, мы докажем эффективную версию теоремы 4. А именно, мы покажем, что протыкающие распределения $\mu_0, \mu_1$ из этой теоремы можно задать за полиномиальное от размера матрицы гаджета время.

**Теорема 5** (Усиление теоремы 4)**.** *Существует **полиномиальный** алгоритм со следующим свойством. Пусть граф $G = (V,E)$ является спектральным $(m,d,\gamma)$-экспандером, в котором у любых двух различных вершин есть не более одного общего соседа, а функция $c : V \to \{0,1\}$ является его сбалансированной раскраской. Пусть также $m \geq 1/\gamma^2$. Тогда, получив на вход матрицу смежности $G$, таблицу истинности $c$ и бит $i \in \{0,1\}$, алгоритм выдает последовательность пар*

$$(R_1, q_1), \ldots, (R_l, q_l)$$

*где $R_1, \ldots, R_l$ — различные $i$-одноцветные подматрицы $M_{g(G,c)}$, а $q_1, \ldots, q_l$ — неотрицательные рациональные числа, в сумме дающие единицу. Кроме того, распределение вероятностей $\mu_i$ на множестве $\{R_1, \ldots, R_l\}$, определенное следующим образом:*

$$\mu_i(R_1) = q_1, \ldots, \mu_i(R_l) = q_l,$$

*является $\left(\frac{1}{1000}, \lfloor 2\log_2(1/\gamma)\rfloor - 100\right)$-протыкающим для матрицы $M_{g(G,c)}$.*

Выход алгоритма из теоремы 5 на тройке $(G, c, i)$, если в графе $G$ содержится $m$ вершин, будет иметь полиномиальную по $m$ длину. В частности носители протыкающих распределений, получающихся на выходе алгоритма, будут иметь полиномиальный от $m$ размер. А $m$ — это как раз размер матрицы $M_{g(G,c)}$.

Объясним, для чего мы доказываем эффективную версию. Для этого отметим следующее свойство протыкающих распределений. Пусть $g : \mathcal{X} \times \mathcal{Y} \to \{0, 1, \perp\}$ — некоторый гаджет, а $M_1, \ldots, M_r$ — произвольные $r$ различных подматриц матрицы $M_g$, размер каждой из которых не меньше $2^{-h}|\mathcal{X}| \times 2^{-h}|\mathcal{Y}|$. Пусть мы хотим найти, скажем, 0-одноцветную подматрицу $M_g$, пересекающую как можно больше подматриц из $\{M_1, \ldots, M_r\}$. Если существует $(\delta, h)$-протыкающие распределение $\mu$ для матрицы $M_g$, сосредоточенное на 0-одноцветных подматрицах, то всегда можно найти 0-одноцветную подматрицу, пересекающую не меньше $(1-\delta)r$ элементов $\{M_1, \ldots, M_r\}$. Действительно, если мы сгенерируем подматрицу $R$ по распределению $\mu$, то $R$ в среднем будет пересекать не меньше $(1-\delta)r$ элементов $\{M_1, \ldots, M_r\}$.

Именно это свойство протыкающих распределений использовали Чаттопадхай и др. в доказательстве теоремы 3. Теперь, предположим, что мы хотим находить такую одноцветную подматрицу (пересекающую много элементов $\{M_1, \ldots, M_r\}$) детерминированно. Мы знаем, что она существует, значит, можно перебрать все такие одноцветные подматрицы, но их может быть экспоненциальное от размера матрицы количество. С другой стороны, как мы видели, достаточно перебирать матрицы из носителя протыкающего распределения. Этим и объясняется наш интерес в рассмотрении протыкающих распределений, у которых носитель может быть выписан за полиномиальное от размера матрицы время. Также теорему 5 можно рассматривать как шаг на пути к эффективному варианту теоремы 3 (эффективному в том смысле, что по протоколу для $f \circ g$ *эффективно* выдается дерево разрешения для $f$).

Отметим, что протыкающие распределения, которые Чаттопадхай и др. строят для гаджета $\mathrm{IP}_k$ не обладают этим свойством — их носители имеют размер $2^{\Omega(k^2)}$, в то время как размер матрицы $\mathrm{IP}_k$ есть $2^k \times 2^k$. Например, для 0-одноцветных подматриц Чаттопадхай и др. строят такое протыкающее распределение. Они генерируют случайное подпространство $\mathbb{F}_2^k$ размерности $k/2$ и в качестве 0-одноцветной подматрицы $M_{\mathrm{IP}_k}$ берут произведение этого подпространства на его "ортогональное дополнение". Размер носителя у такого распределения равно количеству $k/2$-мерных подпространств, а их, как известно, $2^{\Omega(k^2)}$.

С другой стороны, для гаджетов $\mathrm{GHD}(k, k/4, 3k/4)$ и $\mathrm{IND}_k$ протыкающие распределения как раз могут быть выписаны за полиномиальное от размера матрицы время (зато, напомним, их параметры хуже, чем у протыкающих распределений для $\mathrm{IP}_k$).

Теорема 5 автоматически гарантирует, что если матрица смежности $G$ и таблица истинности $c$ заданы эффективно, то и протыкающие распределения для $M_{g(G,c)}$ можно задать эффективно. Так и будет в полученном нами применении теоремы 5.

А именно, мы применяем эту теорему к графу $G = AP_q$. Напомним, как определяется этот граф. Его вершинами являются элементы $\mathbb{F}_q \times \mathbb{F}_q$. Вершина $(a, b)$ соединяется с вершиной $(x, y)$ тогда и только тогда, когда $ax = b + y$ в $\mathbb{F}_q$. Согласно предложению 19, граф $AP_q$ является спектральным $(q^2, q, 1/\sqrt{q})$-экспандером. Для того, чтобы применить к нему теоремы 4, 5, еще надо убедиться, что в нем у любых двух различных вершин не более одного общего соседа. Это легко проверить непосредственно, но мы, кроме того, доказываем следующее предложение:

**Предложение 25.** *Рассмотрим произвольный граф $G$, являющийся спектральным $(m, d, \gamma)$-экспандером, причем*

$$2d + 4 > d^2 \left(2\gamma^2 + \frac{4(1-\gamma^2)}{m}\right).$$

*Тогда в $G$ любые две различные вершины имеют не более одного общего соседа.*

Из этого предложения видно, что в любых спектральных $(t^2, t, 1/\sqrt{t})$-экспандерах у любых двух различных вершин имеется не более одного общего соседа, а значит к таким экс-

пандерам теоремы 4, 5 применимы автоматически. Правда, нам, к сожалению, неизвестно других примеров таких экспандеров, помимо $AP_q$.

Тем не менее, теоремы 4, 5 применимы не только к таким экспандерам. Например, Любоцкий, Филлипс и Сарнак [31] для бесконечно многих пар натуральных чисел $(p, q)$ построили явный спектральный

$$(q(q^2-1)/2, p+1, 2\sqrt{p}/(p+1))\text{-экспандер}$$

(обозначим его через $X^{p,q}$), в котором самый короткий цикл имеет длину не менее $2\log_p(q)$ и в котором нет петель, если $p < q^2$. Значит при $p < \sqrt{q}$ у любых двух различных его вершин есть не более одного общего соседа (иначе в нем был бы цикл длины 4). Вернемся к графу $AP_q$. Из теоремы Чаттопадхая и др. (Теорема 3) и доказанной нами теоремы 5 вытекает такое утверждение:

**Следствие 3.** *В графе $AP_q$ у любых двух различных вершин есть не более одного общего соседа. Кроме того, для любой сбалансированной раскраски $c$ графа $AP_q$ и любой функции $f : \{0,1\}^n \to \{0,1\}$ арности не более $2^{\log_2(q)-200}$ выполнено неравенство:*

$$D(f \circ g(AP_q, c)) \geq \frac{\log_2(q/n) - 200}{4} \cdot D^{dt}(f).$$

Подбором подходящей сбалансированной раскраски $c$ можно таким образом пытаться доказывать аналоги теоремы Раза — Маккинзи для интересных гаджетов. Мы рассматриваем такой гаджет:

$$\mathrm{SQR}^q : \mathbb{F}_{q^2} \times \mathbb{F}_{q^2} \to \{0,1\},$$
$$\mathrm{SQR}^q(x,y) = \begin{cases} 1 & \text{если } \exists z \in \mathbb{F}_{q^2} : x - y = z^2, \\ 0 & \text{иначе.} \end{cases}$$

Можно естественным образом отождествить $\mathbb{F}_{q^2}$ с $\mathbb{F}_q^2$, зафиксировав базис вида $(1, w)$ поля $\mathbb{F}_{q^2}$ над полем $\mathbb{F}_q$, где 1 — единица $\mathbb{F}_{q^2}$, а $w$ — элемент $\mathbb{F}_{q^2}$, не лежащий в подполе размера $q$. Тогда можно считать, что у $\mathrm{SQR}^q$ и $g(AP_q, c)$ одно и то же множество входов. Мы доказываем следующее предложение:

**Предложение 26.** *Для всех $q$, начиная с некоторого, выполнено следующее. Если $q$ — степень нечетного простого числа, то за полиномиальное от $q$ время можно выписать таблицу истинности сбалансированной раскраски с графа $AP_q$, для которой выполнено следующее:*

$$\text{если } g(AP_q, c)(u, v) \neq \perp, \text{ то } g(AP_q, c)(u, v) = \mathrm{SQR}^q(u, v).$$

Это предложение означает, что $\mathrm{SQR}^q$ (для $q$, являющихся степенями нечетных простых чисел) наследует протыкающие распределения $g(AP_q, c)$. Таким образом, для $\mathrm{SQR}^q$ выполнен аналог следствия 3. При этом длина входа $\mathrm{SQR}^q$ равна $k = \lceil 2\log_2(q) \rceil$. Кроме того, легко видеть из следствия 3, что для любого $\varepsilon > 0$ и любой функции $f : \{0,1\}^n \to \{0,1\}$ арности не более $2^{(1/2-\varepsilon)k}$ выполнено следующее:

$$D(f \circ \mathrm{SQR}^q) = \Omega_\varepsilon(D^{dt}(f) \cdot k) = \Omega_\varepsilon(D^{dt}(f) \cdot D(\mathrm{SQR}^q)).$$

Ровно такая же оценка, как мы видели, выполнена для $\mathrm{IP}_k$, а для других гаджетов оценки слабее (например, из графов $X^{p,q}$ в лучшем случае может получиться оценка $2^{k/6}$). Кроме того, у $\mathrm{SQR}^q$, в отличии от $\mathrm{IP}_k$, протыкающие распределения могут быть выписаны за полиномиальное от размера матрицы время (в силу того, что матрица смежности $AP_q$ и таблица истинности $c$ из предложения 26 могут быть выписаны за полиномиальное от $q$ время).

Итак, для любого $\varepsilon > 0$ функция $k \mapsto 2^{(1/2-\varepsilon)k}$ является достижимой оценкой на арность в теореме Раза — Маккинзи, причем есть по-крайней мере два гаджета, доказывающих эту оценку. Можно ли ее улучшить? Мы показываем, что с текущей техникой этого сделать нельзя.

Например, эту оценку можно было бы улучшить, получив для некоторого $d > 1/2$ гаджет с длиной входа $k$ и с двумя $\left(\frac{1}{1000}, dk\right)$-протыкающими распределениями, где одно распределение сосредоточено на 0-одноцветных подматрицах, а другое — на 1-одноцветных. Мы показываем, что такого гаджета не существует.

**Предложение 27.** *Для любого $g : \{0,1\}^k \times \{0,1\}^k \to \{0,1,\perp\}$ и любого целого $h \geq 1$ найдется $i \in \{0,1\}$ такое, что выполнено следующее. Пусть $\mu$ — распределение вероятностей на $i$-одноцветных подматрицах $M_g$. Тогда найдутся два подмножества $A \subset \{0,1\}^k, B \subset \{0,1\}^k$ размера хотя бы $2^{k-h}$ такие, что*

$$\Pr_{R\sim\mu}[R \cap A \times B \neq \varnothing] \leq 2^{k-2h+1}.$$

Из этого вытекает, что если у гаджета с длиной входа $k$ есть два $\left(\frac{1}{1000}, h\right)$-протыкающих распределениями, где одно распределение сосредоточено на 0-одноцветных подматрицах, а другое — на 1-одноцветных, то $h \leq k/2 + O(1)$. То же самое верно, если вместо $\frac{1}{1000}$ иметь любую константу, отделенную сверху от 1.

Кроме того, мы доказываем неулучшаемость так называемой *леммы о толщине*. Эта лемма использовалась во всех упомянутых выше работах о связи коммуникационной и вопросной сложности. Именно из этой леммы в доказательствах возникает ограничение на арность внешней функции. Для начала сформулируем эту лемму.

Введем следующие обозначения. Через $[n]$ обозначим множество чисел $\{1, 2, \ldots, n\}$. Пусть $A$ — некоторое конечное множество, а $X$ — подмножество $A^n$. Зафиксируем также некоторое подмножество $S \subset [n]$, элементами которого, если их упорядочить по возрастанию, являются $i_1, i_2, \ldots, i_{|S|}$. Обозначим

$$X_S = \{(x_{i_1}, \ldots, x_{i_{|S|}}) : (x_1, x_2, \ldots, x_n) \in X\} \subset A^{|S|}.$$

Для $X \subset A^n$ и $i \in [n]$ определим двудольный граф $G_i(X)$. Левой долей $G_i(X)$ будет множество $A$, правой — множество $X_{[n]\setminus\{i\}}$. Мы соединяем левую вершину $x \in A$ с правой вершиной $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots x_n) \in X_{[n]\setminus\{i\}}$ тогда и только тогда, когда

$$(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n) \in X.$$

Через $MinDeg_i(X)$ обозначим минимальную степень правой вершины в $G_i(X)$. Определим $AvgDeg_i(X) = |X|/|X_{[n]\setminus\{i\}}|$. Это число равно количеству ребер в $G_i(X)$ поделить на количество правых вершин. Иными словами, $AvgDeg_i(X)$ — это средняя степень правой вершины.

Лемма о толщине утверждает, грубо говоря, что из любого $X$, у которого для всякого $i \in [n]$ средняя степень правой вершины в $G_i(X)$ велика, можно выбрать достаточно большое подмножество $X'$, у которого для любого $i \in [n]$ *минимальная* степень правой вершины в $G_i(X')$ велика. Более точно, выполнена следующая лемма:

**Лемма 5** (о толщине, [33])**.** *Пусть $A$ — конечное множество, $X \subset A^n$ и для любого $i \in [n]$ выполнено $AvgDeg_i(X) \geq d$. Тогда для любого $\delta \in (0,1)$ существует $X' \subset X$ размера хотя бы $(1-\delta)|X|$ такое, что для любого $i \in [n]$ выполнено $MinDeg_i(X') \geq \frac{\delta d}{n}$.*

Изучив доказательство теоремы 3, можно убедиться в следующем: если для некоторой возрастающей функции $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ оценку $MinDeg_i(X') \geq \frac{\delta d}{n}$ в лемме о толщине можно улучшить до $MinDeg_i(X') \geq \frac{\delta d}{\rho(n)}$, то неравенство

$$D(f \circ g) \geq \frac{\varepsilon h}{4} D^{dt}(f)$$

из теоремы 3 было бы выполнено для всех внешних функций арности не более $\rho^{-1}(2^{(1-\varepsilon)h})$. Например, если доказать лемму о толщине с оценкой $MinDeg_i(X') \geq \frac{\delta d}{\sqrt{n}}$ (то есть для $\rho(n) = \sqrt{n}$), то, как показывали бы в этом случае гаджеты $\mathrm{IP}_k, \mathrm{SQR}^q$, функция $k \mapsto 2^{(1-\varepsilon)k}$ была бы достижимой оценкой на арность в теореме Раза – Маккинзи для любого $\varepsilon > 0$.

Тем не менее, доказать это ни для какого $\rho(n) = o(n)$ невозможно. Заметим, что в лемме о толщине не гарантируется даже просто существования непустого $X'$, у которого для всех $i \in [n]$ выполнено $MinDeg_i(X') > \frac{d}{n}$. Мы для любого $\varepsilon > 0$ приводим пример, когда непустого $X'$ с оценкой $MinDeg_i(X') \geq \frac{d}{n} + \varepsilon$ действительно не существует.

**Теорема 6.** *Для $\varepsilon > 0$ и любых любых целых $n \geq 2, s \geq 1$ существует $m \in \mathbb{N}$ и непустое подмножество $X \subset \{0,1,\ldots,m-1\}^n$ такое, что*

- *для всех $i \in [n]$ выполнено $AvgDeg_i(X) \geq s(n-\varepsilon)$.*
- *не существует непустого $X' \subset X$ такого, что для всех $i \in [n]$ выполнено $MinDeg_i(X') \geq s+1$.*

## 3.2 Доказательства

### 3.2.1 Доказательство теоремы 5

Нам понадобится следующее предложение.

**Предложение 28** ([38], Следствие 3.34)**.** *Для любых натуральных $t$ и $n$ существует функция $\psi_{t,n} : \{0,1\}^{tn} \times \{0,1\}^n \to \{0,1\}$ такая, что для любых различных $x_1,\ldots,x_t \in \{0,1\}^n$ и для любых $b_1,\ldots,b_t \in \{0,1\}$ при случайном выборе $s \in \{0,1\}^{tn}$ с равномерным распределением выполнено:*

$$\Pr[\psi_{t,n}(s,x_1) = b_1, \ldots, \psi_{t,n}(s,x_t) = b_t] = 2^{-t}.$$

*Кроме того, существует полиномиальный алгоритм, который, получив на вход $t, n \in \mathbb{N}, s \in \{0,1\}^{tn}$ и $x \in \{0,1\}^n$, выдает $\psi_{t,n}(s,x)$.*

Предположим, мы получаем на вход тройку $(G,c,i)$, где $G = (V,E)$ — граф на $m$ вершинах, в котором у любых двух различных вершин есть не более одного общего соседа, $c$ — функция из $V$ в $\{0,1\}$, а $i \in \{0,1\}$. Определим распределение вероятностей $\mu_i(G,c)$ на $i$-одноцветных подматрицах $M_{g(G,c)}$ следующим образом. Положим $k = \lceil \log_2 m \rceil$. Рассмотрим следующую случайную величину. Сгенерируем $u \in c^{-1}(i)$ и $s \in \{0,1\}^{20k}$ независимо с равномерным распределением. Разобьем $\Gamma(u)$ на два непересекающихся множества $A = A_{s,u}$ и $B = B_{s,u}$ так: поместим в $A_{s,u}$ все $v \in \Gamma(u)$, для которых выполнено $\psi_{20,k}(s,v) = 1$, а в $B_{s,u}$ поместим все $w \in \Gamma(u)$, для которых выполнено $\psi_{20,k}(s,w) = 0$ (здесь $\psi_{20,k} : \{0,1\}^{20k} \times \{0,1\}^k \to \{0,1\}$ — функция[1] , определенная в предложении 28).

Распределение $\mu_i(G,c)$ будет распределением случайной величины $A \times B$. Заметим, что $\mu_i(G,c)$ действительно сосредоточено на $i$-одноцветных подматрицах $M_{g(G,c)}$. Чтобы понять, почему это так, надо проверить, что для любых $v \in A, w \in B$ выполнено $g(G,c)(v,w) = i$. В самом деле, $v$ и $w$ различны (поскольку $A$ и $B$ не пересекаются), и у них есть общий сосед $u$, для которого $c(u) = i$.

На входе $(G,c,i)$ наш алгоритм выдаст последовательность пар

$$(R_1,q_1),\ldots,(R_l,q_l)$$

[1] формально говоря, второй аргумент $\psi_{20,k}(\cdot,\cdot)$ должен быть элементом $\{0,1\}^k$, а мы подставляем туда вершину $G$. Но поскольку $k \geq \log_2(m)$, можно отождествить $j$-ю вершину графа с $j$-м в лексикографическом порядке элементом $\{0,1\}^k$.

такую, что $R_1, \dots, R_l$ — различные подматрицы $M_{g(G,c)}$ и $\{R_1, \dots, R_l\}$ есть носитель $\mu_i(G,c)$, а $q_j = \mu_i(G,c)(R_j)$ для всех $j = 1, \dots, l$. Объясним, как это сделать за полиномиальное от $m$ время. Перебирая все $s \in \{0,1\}^{20k}$ (число которых есть $m^{O(1)}$) и все $u \in c^{-1}(i)$ (число которых не превосходит $m$), мы можем найти список $R_1, \dots, R_l$ всех подматриц $M_{g(G,c)}$, представляющихся в виде $A_{s,u} \times B_{s,u}$. Затем, чтобы найти $\mu_i(G,c)(R_j)$, надо найти число пар $(s,u)$, $s \in \{0,1\}^{20k}$, $u \in c^{-1}(i)$ таких, что $A_{s,u} \times B_{s,u} = R_j$ ( это опять можно сделать перебором), и поделить это число на $2^{20k} \cdot |c^{-1}(i)|$.

Чтобы доказать теорему, нам остается показать, что если в $G$ у любых двух вершин есть не более одного общего соседа, при этом $G$ является спектральным $(m,d,\gamma)$-экспандером и $m \geq 1/\gamma^2$, а $c$ является сбалансированной раскраской $G$, то распределение $\mu_i(G,c)$ будет $\left(\frac{1}{1000}, \lfloor 2\log_2(1/\gamma)\rfloor - 100\right)$-протыкающим для матрицы $M_{g(G,c)}$.

Зафиксируем $P, Q \subset V$ такие, что $|P|, |Q| \geq 2^{-h}|V|$, где

$$h = \lfloor 2\log_2(1/\gamma)\rfloor - 100.$$

Нам надо доказать, что с вероятностью хотя бы 0.999 выполнено $A \times B \cap P \times Q \neq \varnothing$. Мы докажем более сильное утверждение, а именно, что $\Pr[A \cap P \neq \varnothing] \geq 0.9996$ и $\Pr[B \cap Q \neq \varnothing] \geq 0.9996$. Мы докажем только первое из этих двух неравенств (легко проверить, что второе доказывается точно так же). Чтобы доказать первое неравенство, мы сначала докажем, что $\Pr[|\Gamma(u) \cap P| \geq 20] \geq 0.9997$. Покажем, как из этого неравенства вытекает неравенство $\Pr[A \cap P \neq \varnothing] \geq 0.9996$. При любой фиксации $u$ такого, что $|\Gamma(u) \cap P| \geq 20$, вероятность по случайному выбору $s \in \{0,1\}^{20k}$ с равномерным распределением того, что $A_{s,u}$ не пересекается с $P$, не превосходит $2^{-20}$. Действительно, зафиксируем произвольные 20 различных элементов $v_1, \dots, v_{20} \in \Gamma(u) \cap P$. Если $A_{s,u} \cap P$ пусто, это означает, что

$$\psi_{20,k}(s, v_1) = 0, \dots, \psi_{20,k}(s, v_{20}) = 0.$$

Вероятность последнего события по случайному выбору $s \in \{0,1\}^{20k}$ с равномерным распределением равна $2^{-20}$ согласно предложению 28. Поэтому

$$\begin{aligned}\Pr[A \cap P \neq \varnothing] &\geq \Pr[A \cap P \neq \varnothing \,\big|\, |\Gamma(u) \cap P| \geq 20] \cdot \Pr[|\Gamma(u) \cap P| \geq 20] \\ &\geq (1 - 2^{-20}) \cdot 0.9997 \geq 0.9999 \cdot 0.9997 \geq 0.9996.\end{aligned}$$

Итак, для доказательства теоремы остается доказать неравенство

$$\Pr[|\Gamma(u) \cap P| \geq 20] \geq 0.9997.$$

Размер $P$ не меньше $2^{100}\gamma^2 m \geq 2^{100}\lfloor \gamma^2 m \rfloor$. Разделим $P$ произвольным образом на 20 непересекающихся множеств $P_1, \dots, P_{20}$ размера хотя бы $2 \cdot 10^6 \lfloor \gamma^2 m \rfloor$. Из условия $m \geq 1/\gamma^2$ вытекает, что $\gamma^2 m \geq 1$. Это значит, что $\lfloor \gamma^2 m \rfloor \geq \gamma^2 m/2$, из чего вытекает, что $|P_1|, \dots, |P_{20}| \geq 10^6 \gamma^2 m$.

Мы докажем, что для всякого $j \in \{1, 2, \dots, 20\}$ выполнено $\Pr[\Gamma(u) \cap P_j = \varnothing] \leq 15 \cdot 10^{-6}$. Этого нам достаточно, поскольку из того, что $|\Gamma(u) \cap P| < 20$ вытекает, что с одним из множеств $P_1, \dots, P_{20}$ множество $\Gamma(u)$ не пересекается, и поэтому $\Pr[|\Gamma(u) \cap P| < 20] \leq 20 \cdot 15 \cdot 10^{-6} = 0.0003$.

Заметим, что $\Gamma(u)$ не пересекается с $P_j$ тогда и только тогда, когда $u \notin \Gamma(P_j)$ (действительно, и то, и другое означает, что у $u$ нет соседей в $P_j$). Поэтому

$$\Pr[\Gamma(u) \cap P_j = \varnothing] = \frac{|c^{-1}(u) \setminus \Gamma(P_j)|}{|c^{-1}(u)|} \leq \frac{m - |\Gamma(P_j)|}{m/3}. \tag{3.1}$$

В последнем неравенстве мы пользуемся тем, что $c$ является сбалансированной. С другой стороны, согласно предложению 18, выполнено следующее:

$$|\Gamma(P_j)| \geq |P_j| \cdot \frac{1}{\gamma^2 + (1-\gamma^2)\frac{|P_j|}{m}} \geq \frac{|P_j|}{\gamma^2 + \frac{|P_j|}{m}} \geq \frac{|P_j|}{\frac{|P_j|}{10^6 m} + \frac{|P_j|}{m}}$$
$$= \frac{10^6 m}{10^6 + 1} \geq (1 - 10^{-6})m.$$

Здесь в третьем неравенстве мы пользуемся тем, что $|P_j| \geq 10^6 \gamma^2 m$. Подставляя полученную оценку на $|\Gamma(P_j)|$ в (3.1), получаем

$$\Pr[\Gamma(u) \cap P_j = \varnothing] \leq 3 \cdot 10^{-6} < 15 \cdot 10^{-6},$$

что и требовалось.

### 3.2.2 Вывод следствия 3.

Покажем сначала, что графе $AP_q$ у любых двух различных вершин есть не более одного общего соседа. Зафиксируем произвольные $(x, y), (a, b) \in \mathbb{F}_q^2$ такие, что $(x, y) \neq (a, b)$. Если $(u, v)$ является общим соседом $(x, y)$ и $(a, b)$, то $xu = y + v, au = b + v$. Покажем, что существует не более одной пары $(u, v)$, удовлетворяющей двум этим равенствам. В матричном виде эти равенства можно записать так:

$$\begin{pmatrix} x & -1 \\ a & -1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} y \\ b \end{pmatrix}. \tag{3.2}$$

Если $x \neq a$, то определитель матрицы системы (3.2) отличен от нуля, значит у этой системы единственное решение. Если $x = a$, то из существования корня системы (3.2) вытекало бы $y = b$, значит корней нет.

Итак, граф $AP_q$ является спектральным $(q^2, q, 1/\sqrt{q})$-экспандером (см. предложение 19) и в нем у любых двух различных вершин есть не более одного общего соседа. При этом выполнено ограничение $m \geq 1/\gamma^2$ из теоремы 5. Тогда если $c$ — сбалансированная раскраска $AP_q$, то по теореме 5 при

$$h = \lfloor 2 \log_2(\sqrt{q}) \rfloor - 100$$

для матрицы $M_{g(AP_q,c)}$ есть два $\left(\frac{1}{1000}, h\right)$-протыкающих распределения, одно сосредоточено на 0-одноцветных подматрицах, а другое сосредоточено на 1-одноцветных подматрицах. Теперь рассмотрим произвольную функцию $f : \{0,1\}^n \to \{0,1\}$ арности $n \leq 2^{\log_2(q)-200}$. Применяя теорему 3 к $f$ при

$$\varepsilon = 1 - \frac{\log_2(n)}{\lfloor \log_2(q) \rfloor - 100},$$

мы получаем:

$$D(f \circ g(AP_q, c)) \geq \frac{\varepsilon h}{4} \cdot D^{dt}(f) = \frac{\lfloor \log_2 q \rfloor - 100 - \log_2(n)}{4} \cdot D^{dt}(f)$$
$$\geq \frac{\log_2(q/n) - 200}{4} \cdot D^{dt}(f).$$

Остается проверить, что выполнены все ограничения теоремы 3, а именно $\varepsilon \geq 6/h$ и $n \leq 2^{(1-\varepsilon)h}$.

Проверим первое ограничение. Поскольку $n \leq 2^{\log_2(q)-200}$, то

$$h - \log_2(n) = \lfloor \log_2(q) \rfloor - 100 - \log_2(n) \geq \log_2(q) - 101 - \log_2(q) + 200 \geq 6.$$

Поэтому $\varepsilon = 1 - \frac{\log_2(n)}{h} = \frac{h - \log_2(n)}{h} \geq \frac{6}{h}$.

Второе ограничение выполнено благодаря следующему равенству:

$$(1 - \varepsilon)h = \frac{\log_2(n)}{h} \cdot h = \log_2(n).$$

### 3.2.3 Доказательство предложения 25

Будем считать, что вершины $G$ пронумерованы числами от 1 до $m$. Для $A \subset [m] = \{1, 2, \ldots, m\}$ через $\mathbb{I}_A \in \mathbb{R}^m$ обозначим вектор, у которого в координатах из $A$ стоит 1, а в других координатах — 0.

Рассмотрим любые две различные вершины $u$ и $v$ графа $G$. Обозначим $w = \{u, v\}$. Покажем, что

$$\|M\mathbb{I}_w\|^2 \le d^2 \left(2\gamma^2 + \frac{4(1-\gamma^2)}{m}\right), \tag{3.3}$$

где $M$ — матрица смежности $G$.

Заметим, что

$$\mathbb{I}_w = \frac{2}{m} \cdot \mathbb{I}_{[m]} + \left(\mathbb{I}_w - \frac{2}{m} \cdot \mathbb{I}_{[m]}\right),$$

причем вектор $\mathbb{I}_{[m]}$ перпендикулярен вектору $\left(\mathbb{I}_w - \frac{2}{m} \cdot \mathbb{I}_{[m]}\right)$ (поскольку сумма координат последнего вектора равна нулю). Матрица $M$ — симметричная, сумма элементов по каждой строке и по каждому столбцу в ней равна $d$, у матрицы $M$ кратность собственного значения $d$ равна 1 (ему соответствует собственный вектор $\mathbb{I}_{[m]}$), а все остальные собственные значения $M$ не превосходят по модулю $\gamma d$. Из этого вытекают две вещи:

- Подпространство векторов, перпендикулярных $\mathbb{I}_{[m]}$ является инвариантным для $M$. В частности, вектора
$$M\frac{2}{m} \cdot \mathbb{I}_{[m]} \qquad \text{и} \qquad M\left(\mathbb{I}_w - \frac{2}{m} \cdot \mathbb{I}_{[m]}\right)$$
также перпендикулярны.

- Операторная норма $M$, если ее сузить на подпространство векторов, перпендикулярных $\mathbb{I}_{[m]}$, не превосходит $\gamma d$.

Используя эти соображения, легко вывести (3.3):

$$\begin{aligned}
\|M\mathbb{I}_w\|^2 &= \left\|M\frac{2}{m} \cdot \mathbb{I}_{[m]}\right\|^2 + \left\|M\left(\mathbb{I}_w - \frac{2}{m} \cdot \mathbb{I}_{[m]}\right)\right\|^2 \\
&\le \frac{4d^2}{m} + \gamma^2 d^2 \cdot \left\|\left(\mathbb{I}_w - \frac{2}{m} \cdot \mathbb{I}_{[m]}\right)\right\|^2 \\
&= \frac{4d^2}{m} + \gamma^2 d^2 \cdot \left((m-2)\frac{4}{m^2} + 2\left(1 - \frac{2}{m}\right)^2\right),
\end{aligned}$$

а последнее выражение, как легко убедиться, как раз равно правой части (3.3).

Теперь предположим от противного, что $|\Gamma(u) \cap \Gamma(v)| \ge 2$. Противоречие получится, если доказать

$$\|M\mathbb{I}_w\|^2 \ge 2d + 4. \tag{3.4}$$

Действительно, из (3.3), (3.4) тогда вытекало бы, что

$$2d + 4 \le d^2 \left(2\gamma^2 + \frac{4(1-\gamma^2)}{m}\right),$$

что противоречило бы условию предложения.

Докажем (3.4). Заметим, что

$$|\Gamma(w)| = |\Gamma(u) \cup \Gamma(v)| = |\Gamma(u)| + |\Gamma(v)| - |\Gamma(u) \cap \Gamma(v)| \le 2d - 2.$$

Пусть в $M\mathbb{I}_w$ ровно $t$ ненулевых координат. Имеем $t = |\Gamma(w)| \leq 2d-2$. Обозначим значения этих координат через $\xi_1, \ldots, \xi_t$. Поскольку сумма элементов в каждом столбце $M$ равна $d$, это означает, что сумма координат $M\mathbb{I}_w$ равна $2d$ (вектор $M\mathbb{I}_w$ равен сумме двух столбцов $M$), то есть $\xi_1 + \ldots + \xi_t = 2d$. С другой стороны, числа $\xi_1 - 1, \ldots, \xi_t - 1$ являются неотрицательными числами с суммой $2d - t \geq 2$. Значит $(\xi_1 - 1)^2 + \ldots + (\xi_t - 1)^2 \geq 2$ (иначе среди $\xi_1 - 1, \ldots, \xi_t - 1$ могло бы быть максимум одно положительное, и оно бы равнялось 1). Поэтому

$$\begin{aligned}\|M\mathbb{I}_w\|^2 &= \xi_1^2 + \ldots + \xi_t^2 \\ &= (\xi_1 - 1)^2 + \ldots + (\xi_t - 1)^2 + \sum_{i=1}^{t}(2\xi_i - 1) \\ &\geq 2 + 4d - t \geq 2d + 4.\end{aligned}$$

### 3.2.4 Доказательство предложения 26

Элементы поля $\mathbb{F}_{q^2}$, которые можно представить в виде $z^2$ для некоторого $z \in \mathbb{F}_{q^2}$, назовем квадратами. Подполе $\mathbb{F}_{q^2}$ размера $q$ будем просто обозначать через $\mathbb{F}_q$. Напомним, что у нас фиксирован базис $(1, w)$ поля $\mathbb{F}_{q^2}$ над $\mathbb{F}_q$, где 1 — единица поля, а $w \notin \mathbb{F}_q$. Мы отождествляем $\mathbb{F}_q^2$ и $\mathbb{F}_{q^2}$ при помощи следующей биекции:

$$(x, y) \mapsto x + yw.$$

Нам понадобится следующая лемма:

**Лемма 6.** *Пусть $q$ — степень нечетного простого числа, а $\alpha$ — примитивный корень $\mathbb{F}_{q^2}$. Тогда*

- *множество $\{0, \alpha^2, \alpha^4, \ldots, \alpha^{q^2-1}\}$ является множеством квадратов $\mathbb{F}_{q^2}$;*
- *все элементы $\mathbb{F}_q$ являются квадратами в $\mathbb{F}_{q^2}$.*

*Доказательство.* Докажем первый пункт. То, что все элементы указанного множества являются квадратами, очевидно. Докажем, что других квадратов нет. Любой другой элемент $\mathbb{F}_{q^2}$ можно представить как $\alpha^j$, где $j \in \{1, 2, \ldots, q^2 - 1\}$ — нечетное число. Если $\alpha^j$ – квадрат, то $\alpha^j = \alpha^{2i}$ для некоторого целого $i$. Это означает, что $\alpha^{j-2i} = 1$, то есть $q^2 - 1$ делит $j - 2i$. Но $j - 2i$ нечетно, а $q^2 - 1$ — четно.

Докажем второй пункт. Любой элемент $\mathbb{F}_q$ является корнем $x^q - x$ в $\mathbb{F}_{q^2}$. Надо доказать, что все корни этого многочлена являются квадратами. Пусть $\alpha^j$ — корень $x^q - x$. Тогда $\alpha^{qj} = \alpha^j$. Следовательно $\alpha^{(q-1)j} = 1$, а значит $q^2 - 1$ делит $(q-1)j$. Из этого вытекает, что $(q+1)$ делит $j$, то есть $j$ четно. Поэтому $\alpha^j$ — квадрат. □

Определим следующую раскраску $c$ графа $AP_q$:

$$c : \mathbb{F}_q^2 \to \{0, 1\}, \qquad c(a, b) = \begin{cases} 1 & \text{если } 1 + aw \text{ — квадрат,} \\ 0 & \text{иначе.} \end{cases}$$

Очевидно, таблицу истинности $c$ можно вычислить за полиномиальное от $q$ время (проверять квадратность можно перебором всех элементов поля, число которых — $q^2$).

Теперь предположим, что $(u, v), (x, y) \in \mathbb{F}_q^2$ — это две различных вершины графа $AP_q$, для которых $g(AP_q, c)((u, v), (x, y)) \neq \perp$. Это означает, что у $(u, v)$ и $(x, y)$ в графе $AP_q$ есть общий сосед $(a, b)$. Тогда по определению:

$$g(AP_q, c)((u, v), (x, y)) = c(a, b).$$

Нам надо показать, что из этого вытекает.

$$\mathrm{SQR}^q(u + v \cdot w, x + y \cdot w) = c(a, b). \tag{3.5}$$

Действительно, по определению графа $AP_q$ выполнено:

$$au = b + v, \qquad ax = b + y.$$

Поэтому $a(u-x) = v-y$. Поскольку $(u, v)$ и $(x, y)$ различны, это значит, что $u-x \neq 0$. Кроме того:

$$(u + v \cdot w) - (x + y \cdot w) = (u - x)(1 + a \cdot w).$$

Заметим, что $u-x$ — ненулевой элемент $\mathbb{F}_q$, являющийся, согласно второму пункту леммы 6, квадратом в $\mathbb{F}_{q^2}$. Поэтому разница $(u+v\cdot w)$ и $(x+y\cdot w)$ является квадратом тогда и только тогда, когда $1 + a \cdot w$ является квадратом. Отсюда вытекает (3.5).

Чтобы закончить доказательство предложения, нам остается показать, что раскраска $c$ является сбалансированной для всех достаточно больших $q$. Сгенерируем $(a, b, \lambda) \in \mathbb{F}_q \times \mathbb{F}_q \times (\mathbb{F}_q \setminus \{0\})$ случайно с равномерным распределением. Заметим, что

$$\begin{aligned} |c^{-1}(1)| &= \left|\{(a, b) \in \mathbb{F}_q^2 : c(a, b) = 1\}\right| \\ &= q \cdot |\{a \in \mathbb{F}_q : 1 + a \cdot w \text{ — квадрат}\}| \\ &= q^2 \Pr[1 + a \cdot w \text{ — квадрат}]. \end{aligned}$$

Поэтому нам достаточно показать, что $1/3 \leq \Pr[1 + a \cdot w \text{ — квадрат}] \leq 2/3$. Заметим, что согласно пункту 2 леммы 6 выполнено следующее: $1+a\cdot w$ является квадратом тогда и только тогда, когда $\lambda + (\lambda a) \cdot w$ является квадратом. С другой стороны, $\lambda + (\lambda a) \cdot w$ распределено равномерно на множестве $\{i + j \cdot w : i \in \mathbb{F}_q \setminus \{0\}, j \in \mathbb{F}_q\}$. Размер этого множества равен $q^2 - q$. Согласно первому пункту леммы (6), всего в $\mathbb{F}_{q^2}$ есть $(q^2 + 1)/2$ квадратов. Поэтому для всех достаточно больших $q$ количество квадратом в множестве $\{i+j\cdot w : i \in \mathbb{F}_q \setminus \{0\}, j \in \mathbb{F}_q\}$ не меньше $0.4q^2$ и не больше $0.6q^2$. Значит для всех достаточно больших $q$ вероятность $\Pr[\lambda + (\lambda a) \cdot w \text{ — квадрат}]$ принадлежит отрезку $[1/3, 2/3]$. А эта вероятность, как мы уже показали, равна вероятности $\Pr[1 + a \cdot w \text{ — квадрат}]$.

### 3.2.5 Доказательство предложения 27

. Нам понадобится следующая лемма о биномиальных коэффициентах:

**Лемма 7.** *Для любых натуральных $k, m$ таких, что $k \leq m/2$, выполнено следующее: $\binom{m-k}{k}/\binom{m}{k} \geq 1 - \frac{k^2}{m-k}$.*

*Доказательство.* Заметим, что

$$\begin{aligned} \frac{\binom{m-k}{k}}{\binom{m}{k}} &= \frac{m-k}{m} \cdot \ldots \cdot \frac{m-2k+1}{m-k+1} \geq \left(\frac{m-2k}{m-k}\right)^k \\ &= \left(1 - \frac{k}{m-k}\right)^k \geq 1 - \frac{k^2}{m-k}. \end{aligned}$$

Первое неравенство в этой цепочке выполнено потому, что для всех положительных $i$:

$$\begin{aligned} \frac{k}{m-k+i} \leq \frac{k}{m-k} &\implies 1 - \frac{k}{m-k+i} \geq 1 - \frac{k}{m-k} \\ &\implies \frac{m-2k+i}{m-k+i} \geq \frac{m-2k}{m-k}. \end{aligned}$$

Второе неравенство — это неравенство Бернулли. Оно применимо здесь, поскольку $\frac{k}{m-k} \leq 1$ в силу услови $k \leq m/2$. □

Перейдем к доказательству предложения. Обозначим $s = 2^{k-h}$. Предположим сначала, что в матрице $M_g$ найдется 0-одноцветная подматрица $X \times Y$ такая, что $|X| \geq s$ и $|Y| \geq s$. Тогда утверждение предложения выполнено для $i = 1$ и $A = X, B = Y$. Действительно, для любого распределения вероятностей $\mu$ на 1-одноцветных подматрицах $M_g$ выполнено:

$$\Pr_{R\sim\mu}[R \cap X \times Y \neq \varnothing] = 0.$$

Теперь предположим, что для любой 0-одноцветной подматрицы $X \times Y$ матрицы $M_g$ либо $|X|$, либо $|Y|$ меньше $s$. Предложение тогда будет выполнено для $i = 0$. Существование требуемых $A$ и $B$ мы докажем вероятностным методом. А именно, мы покажем, что если выбрать $A, B \subset \{0,1\}^k$ независимо с равномерным распределением среди всех подмножеств $\{0,1\}^k$ размера $s = 2^{k-h}$, то будет выполнено:

$$\mathbb{E}_{A,B} \Pr_{R\sim\mu}[R \cap A \times B \neq \varnothing] \leq 2^{k-2h+1} \tag{3.6}$$

для любого распределения вероятностей $\mu$ на 0-одноцветных подматрицах $M_g$. Если доказать (3.6), это будет означать, что $A$ и $B$ можно зафиксировать так, чтобы они удовлетворяли условию предложения.

Перепишем (3.6) следующим образом:

$$\mathbb{E}_{A,B} \Pr_{R\sim\mu}[R \cap A \times B \neq \varnothing] = \mathbb{E}_{R\sim\mu} \Pr_{A,B}[R \cap A \times B \neq \varnothing].$$

Мы покажем, что для любого фиксированного $R = X \times Y$ из носителя $\mu$ выполнено:

$$\Pr_{A,B}[R \cap A \times B \neq \varnothing] \leq 2^{k-2h+1}.$$

Действительно, $R$ является 0-одноцветной, а значит либо $|X| < s$, либо $|Y| < s$. Рассмотрим первый случай, второй рассматривается точно также. Заметим, что

$$\Pr_{A,B}[R \cap A \times B \neq \varnothing] \leq \Pr_{A}[X \cap A \neq \varnothing] = 1 - \frac{\binom{2^k-|X|}{s}}{\binom{2^k}{s}} \leq 1 - \frac{\binom{2^k-s}{s}}{\binom{2^k}{s}}$$

К последнему выражению можно применить лемму 7 (действительно, $s = 2^{k-h} \leq 2^k/2$ в силу того, что $h \geq 1$) и получить, что $\Pr_{A,B}[R \cap A \times B \neq \varnothing]$ не превосходит $\frac{s^2}{2^k-s}$. Опять воспользовавшись тем, что $s \leq 2^{k-1}$, мы получаем требуемую оценку:

$$\Pr_{A,B}[R \cap A \times B \neq \varnothing] \leq \frac{s^2}{2^{k-1}} = 2^{k-2h+1}.$$

### 3.2.6 Доказательство теоремы 6

Рассмотрим произвольное подмножество $X \subset \{0,1,\ldots,m-1\}^n$ и число $i \in [n]$. Будем говорить, что $x \in X$ является *$i$-уникальным* в $X$, если не найдется $x' \in X$, отличного от $X$, такого, что

$$x_1 = x'_1, \ldots, x_{i-1} = x'_{i-1}, x_{i+1} = x'_{i+1}, \ldots, x_n = x'_n.$$

Назовем множество $X \subset \{0,1,\ldots,m-1\}^n$ *редуцируемым*, если для любого непустого $X' \subset X$ найдется $i \in [n]$ такой, что $MinDeg_i(X') = 1$. Заметим, что множество $X$ является редуцируемым тогда и только тогда, когда для всех непустых $X' \subset X$ найдется $x \in X'$ и $i \in [n]$ такой, что $x$ является $i$-уникальным в $X'$.

Мы сначала докажем теорему 6 для случая $s = 1$:

**Лемма 8.** *Для любого натурального $n \geq 2$ и любого $\varepsilon > 0$ существует $m \in \mathbb{N}$ и редуцируемое подмножество $X \subset \{0, 1, \ldots, m-1\}^n$ такое, что для всех $i \in [n]$ выполнено $AvgDeg_i(X) \geq n - \varepsilon$.*

Как мы отмечали, редуцируемость как раз и означает, что не существует непустого $X' \subset X$ такого, что $MinDeg_i(X') \geq 2$ для всех $i \in [n]$.

*Доказательство леммы 8.* Рассмотрим произвольное $m > 0$. Для каждого $n \geq 2$ зададим множество $X_n \subset \{0, 1, \ldots, m-1\}^n$ по следующему рекурсивному правилу:

$$X_2 = \Big\{(j, j) : j \in \{0, 1, \ldots, m-1\}\Big\} \cup \Big\{(j, j+1) : j \in \{0, 1, \ldots, m-2\}\Big\},$$

$$X_{n+1} = \Big\{(x, j) : x \in X_n, j \in \{0, 1, \ldots, m-1\}\Big\} \cup \Big\{(y, 0) : y \in \{0, 1, \ldots, m-1\}^n \setminus X_n\Big\}.$$

Между размером $X_{n+1}$ и размером $X_n$ по определению имеется следующее взаимоотношение:

$$|X_{n+1}| = m|X_n| + m^n - |X_n| = |X_n|(m-1) + m^n.$$

Покажем, используя это взаимоотношение (индукцией по $n$), что

$$|X_n| \geq n(m-1)^{n-1}. \tag{3.7}$$

Проверим это неравенство для $n = 2$. Имеем: $|X_2| = 2m-1 \geq 2(m-1)$. Теперь, предположим, что мы уже доказали, что $|X_n| \geq n(m-1)^{n-1}$. Тогда

$$|X_{n+1}| \geq n(m-1)^{n-1} \cdot (m-1) + m^n \geq (n+1)(m-1)^n.$$

Из (3.7) вытекает, что для любого целого $n \geq 2$ и $i \in [n]$ выполнено:

$$AvgDeg_i(X_n) = \frac{|X_n|}{|(X_n)_{[n]\setminus\{i\}}|} \geq \frac{n(m-1)^{n-1}}{m^{n-1}} = n \cdot \left(1 - \frac{1}{m}\right)^{n-1}.$$

Последнее выражение стремится к $n$ при $m \to \infty$. Таким образом, чтобы доказать лемму, достаточно доказать, что множества $X_n$ являются редуцируемыми. Мы докажем это индукцией по $n$.

Рассмотрим случай $n = 2$ и возьмем произвольное непустое $X' \subset X_2$. Элементы $\{0, 1, \ldots, m-1\}^2$ упорядочим лексикографически (сначала сравниваем первую координату, потом вторую). Рассмотрим наименьший элемент в $X'$. Если он имеет вид $(j, j)$ для некоторого $j \in \{0, 1, \ldots, m-1\}$, то $(j, j)$ является 1-уникальным в $X'$ (а значит $MinDeg_1(X') = 1$). Действительно, если бы в $X'$ существовал другой элемент со второй координатой $j$, то его первая координата была бы $j-1$, а значит он был бы меньше $(j, j)$ в лексикографическом порядке. Если же минимальный в лексикографическом порядке элемент $X'$ имеет вид $(j, j+1)$ для некоторого $j \in \{0, 1, \ldots, m-2\}$, то он является 2-уникальным в $X'$ (а значит $MinDeg_2(X') = 1$); действительно, иначе нашелся бы другой элемент $X'$ с первой координатой, равной $j$, а это может быть только $(j, j)$, который меньше $(j, j+1)$ в лексикографическом порядке.

Остается показать, что если $X_n$ — редуцируемо, то $X_{n+1}$ — тоже редуцируемо. Рассмотрим произвольное непустое $X' \subset X_{n+1}$. Допустим $X'$ пересекается с $\Big\{(y, 0) : y \in \{0, 1, \ldots, m-1\}^n \setminus X_n\Big\}$. Тогда в $X'$ есть элемент вида $(y, 0)$ для некоторого $y \in \{0, 1, \ldots, m-1\}^n \setminus X_n$. Этот элемент является $(n+1)$-уникальным в $X'$. Действительно, если в $X'$ есть элемент вида

$(y, j)$ для некоторого $j > 0$, то по определению $X_{n+1}$ это означает, что $y \in X_n$, противоречие. Значит в этом случае $MinDeg_{n+1}(X') = 1$.

Теперь предположим, что $X'$ не пересекается с $\Big\{(y, 0) : y \in \{0, 1, \ldots, m-1\}^n \setminus X_n\Big\}$, то есть

$$X' \subset \Big\{(x, j) : x \in X_n, j \in \{0, 1, \ldots, m-1\}\Big\}.$$

Поскольку $X'$ не пусто, то и для некоторого $j \in \{0, 1, \ldots, m-1\}$ множество

$$X'' = \{x \in X_n : (x, j) \in X'\}$$

не пусто. Согласно предположению о редуцируемости $X_n$, найдутся $x \in X''$ и $i \in [n]$ такие, что $x$ является $i$-уникальным в $X''$. Покажем, что отсюда следует, что $(x, j)$ является $i$-уникальным в $X'$ (а значит $MinDeg_i(X') = 1$).

В первую очередь отметим, что по определению $X''$ элемент $(x, j)$ действительно принадлежит $X'$. С другой стороны, предположим в $X'$ есть другой элемент, совпадающий с $(x, j)$ по всем координатам, кроме $i$-й. Тогда он имеет вид $(y, j)$ (последняя, $(n+1)$-ая координата у них обязана совпадать, поскольку $i \neq n+1$) для некоторого $y \in \{0, 1, \ldots, m-1\}^n, y \neq x$. Отсюда следует, что $y \in X''$, но $y$ отличен от $x$ и совпадает с $x$ во всех координатах, кроме $i$-й, что противоречит $i$-уникальности $x$ в $X''$. □

Чтобы доказать теорему 6 для всех $s$, мы введем следующее определение.

**Определение 19.** *Пусть $s, m, n$ — целые положительные числа и $X \subset \{0, 1, \ldots, m-1\}^n$. Определим множество $In(X, s) \subset \{0, 1, \ldots, sm-1\}^n$ следующим образом:*

$$In(X, s) = \Big\{(sx_1 + r_1, \ldots, sx_n + r_n) : \\ (x_1, \ldots, x_n) \in X, \qquad r_1, \ldots, r_n \in \{0, 1, \ldots, s-1\}\Big\}.$$

Легко понять, что между размер $In(X, s)$ в $s^n$ раз больше, чем размер $X$ (для каждого элемента $x \in X$ есть ровно $s^n$ элементов $In(X, s)$, у которых частное от деления любой координаты на $s$ равно соответствующей координате $x$). Аналогичное свойство выполнено и для любой проекции:

$$|(In(X, s))_{[n]\setminus\{i\}}| = s^{n-1} \cdot |X_{[n]/\{i\}}|.$$

Отсюда вытекает:

**Лемма 9.** *Для любого $i \in [n]$ выполнено $AvgDeg_i(In(X, s)) = s \cdot AvgDeg_i(X)$.*

Нам остается доказать следующую лемму:

**Лемма 10.** *Если $X \subset \{0, 1, \ldots, m-1\}^n$ редуцируемо, то для любого непустого $Y \subset In(X, s)$ найдется $i \in [n]$ такой, что $MinDeg_i(Y) \leq s$.*

*Доказательство.* Рассмотрим множество $X'$, состоящее из всех элементов $(x_1, \ldots, x_n) \in X$ таких, что

$$(sx_1 + r_1, \ldots, sx_n + r_n) \in Y$$

для некоторых $r_1, \ldots, r_n \in \{0, 1, \ldots, s-1\}$. Поскольку $Y$ непусто, то и $X'$ непусто. Значит для некоторого $i \in [n]$ в $X'$ есть элемент $x = (x_1, \ldots, x_n)$, являющийся $i$-уникальным в $X'$. По определению, для некоторых чисел $r_1, \ldots, r_n \in \{0, 1, \ldots, s-1\}$ выполнено:

$$y = (y_1, \ldots, y_n) \in Y,$$

где $y_i = sx_i + r_i$. Покажем, что в графе $G_i(Y)$ у правой вершины

$$(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$$

есть не более $s$ соседей слева, из чего вытекает $MinDeg_i(Y) \leq s$. Более точно, мы покажем, что все левые соседи этой вершины принадлежат множеству $\{sx_i, sx_i + 1, \ldots, sx_i + s - 1\}$. Действительно, если это не так, то для для какого-то $q \in \{0, 1, \ldots, m-1\} \setminus \{x_i\}$ и $r \in \{0, 1, \ldots, s-1\}$ выполнено:

$$(y_1, \ldots, y_{i-1}, sq + r, y_{i+1}, \ldots, y_n) \in Y \subset In(X, s).$$

Это значит, что $(x_1, \ldots, x_{i-1}, q, x_{i+1}, \ldots, x_n)$ принадлежит $X'$, а это противоречит $i$ - уникальности $x$ в $X'$. □

Теперь окончательно докажем теорему 6. Нам даются $\varepsilon > 0$ и целые $n \geq 2, s \geq 1$. По лемме 8 найдется $m' > 0$ и редуцируемое множество $Y \subset \{0, 1, \ldots, m'-1\}^n$ такое, что для всех $i \in [n]$ выполнено $AvgDeg_i(Y) \geq n - \varepsilon$. Положим $m = sm'$, $X = In(Y, s) \subset \{0, 1, \ldots, m-1\}^n$. По лемме 9 для любого $i \in [n]$ выполнено:

$$AvgDeg_i(X) = s \cdot AvgDeg_i(Y) \geq s(n - \varepsilon).$$

С другой стороны, поскольку $Y$ редуцируемо, то по лемме 10 не найдется непустого $X' \subset X$ такого, что для любого $i \in [n]$ выполнено $MinDeg_i(X') \geq s + 1$.

# Глава 4

# Частные и общие случайные биты в информационной сложности

В этой главе мы изложим наш результат, касающийся сравнения частных и общих случайных битов в информационной сложности. Но начнем мы с описания предыдущих результатов, тесно связанных с нашим.

Как мы писали во введении, примером применения информационной сложности в коммуникационной сложности является проблема прямой суммы. Напомним, в чем суть этой проблемы — она о взаимосвязи коммуникационной сложности одной копии и $n$ копий функции, то есть, формально говоря, величин $D^{\nu}_{\varepsilon}(g)$ и $D^{\nu,n}_{\varepsilon}(g)$ (см. определения 15 и 16). Вопрос заключается в том, насколько точна тривиальная оценка $D^{\nu,n}_{\varepsilon}(g) \leq n \cdot D^{\nu}_{\varepsilon}(g)$.

Сначала упомянем результат Барака, Бравермана, Чена и Рао о связи информационной сложности и коммуникационной сложности $n$ копий функции.

**Теорема 7** ([2]). *Для любого натурального $n$ выполнено следующее: существует вероятностный протокол $\tau$, вычисляющий функцию $g$ с ошибкой $\varepsilon$ по распределению $\nu$ такой, что*

$$IC_{\nu}(\tau) \leq D^{\nu,n}_{\varepsilon}(g)/n, \qquad CC(\tau) = D^{\nu,n}_{\varepsilon}(g).$$

Если бы оценка $D^{\nu,n}_{\varepsilon}(g)/n$ выполнялась не для информационной, а для коммуникационной сложности $\tau$, проблема прямой суммы была бы решена: было бы доказано равенство $D^{\nu,n}_{\varepsilon}(g) = n \cdot D^{\nu}_{\varepsilon}(g)$.

Еще надо как-то связать информационную и коммуникационную сложность одной копии функции. Мы приведем следующий результат.

**Теорема 8** ([2]). *Для любого вероятностного протокола $\tau$ над $(\mathcal{X}, \mathcal{Y})$, любой функции $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, любого распределения вероятностей $\nu$ на $\mathcal{X} \times \mathcal{Y}$ и любых $\varepsilon, \delta \in (0,1)$ выполнено следующее: если $\tau$ вычисляет функцию $g$ с ошибкой $\varepsilon$ по распределению $\nu$, то найдется детерминированный протокол $\tau'$, вычисляющий функцию $g$ с ошибкой $\varepsilon + \delta$ по распределению $\nu$, такой, что*

$$CC(\tau') = O\Big(\frac{\sqrt{IC_{\nu}(\tau) \cdot CC(\tau)} \cdot \log(CC(\tau)/\delta)}{\delta}\Big).$$

Такого рода результаты называют «сжатием» коммуникационных протоколов: по протоколу с небольшой коммуникационной сложностью надо построить моделирующий его протокол с небольшой коммуникационной сложностью.

Продемонстрируем, как из этих двух теорем вытекает следующий результат, относящийся к проблеме прямой суммы. Этот результат, грубо говоря, утверждает, что коммуникационная сложность $n$ копий функции хотя бы в $\sqrt{n}$ раз больше коммуникационной сложности одной копии. Он был получен в [2].

**Теорема 9.** *Для любой функции $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, любого распределения вероятностей $\nu$ на $\mathcal{X} \times \mathcal{Y}$ и любых $n \in \mathbb{N}, \varepsilon, \delta \in (0,1)$ выполнено следующее неравенство:*

$$D^{\nu,n}_{\varepsilon}(g) \cdot \log_2(D^{\nu,n}_{\varepsilon}(g)/\delta) = \Omega\big(\delta\sqrt{n} \cdot D^{\nu}_{\varepsilon+\delta}(g)\big).$$

*Доказательство.* Возьмем протокол $\tau$ из теоремы 7. Он вычисляет функцию $g$ с ошибкой $\varepsilon$ по $\nu$, при этом:

$$IC_\nu(\tau) \le D^{\nu,n}_{\varepsilon}(g)/n, \qquad CC(\tau) = D^{\nu,n}_{\varepsilon}(g).$$

Применим к $\tau$ теорему 8. Получим детерминированный протокол $\tau'$, вычисляющий $g$ с ошибкой $\varepsilon + \delta$ по $\nu$, для которого выполнено:

$$CC(\tau') = O\left(\frac{\sqrt{\frac{D^{\nu,n}_{\varepsilon}(g)}{n} \cdot D^{\nu,n}_{\varepsilon}(g)} \cdot \log(D^{\nu,n}_{\varepsilon}(g)/\delta)}{\delta}\right).$$

С другой стороны, по определению $D^{\nu}_{\varepsilon+\delta}(g)$ получаем оценку:

$$D^{\nu}_{\varepsilon+\delta}(g) \le CC(\tau'),$$

откуда получаем требуемое утверждение. □

Можно ли вместо $\sqrt{n}$ получить более высокую степень $n$, остается открытым вопросом. Еще один (не сравнимый с теоремой 8) был получен Браверманом:

**Теорема 10** ([5])**.** *Для любого вероятностного протокола $\tau$ над $(\mathcal{X}, \mathcal{Y})$, любой функции $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, любого распределения вероятностей $\nu$ на $\mathcal{X} \times \mathcal{Y}$ и любых $\varepsilon, \delta \in (0,1)$ выполнено следующее: если $\tau$ вычисляет функцию $g$ с ошибкой $\varepsilon$ по распределению $\nu$, то найдется детерминированный протокол $\tau'$, вычисляющий функцию $g$ с ошибкой $\varepsilon + \delta$ по распределению $\nu$, такой, что*

$$CC(\tau') = 2^{O\left(IC_\nu(\tau)/\delta^2\right)}.$$

Здесь оценка на $CC(\tau')$ не зависит от $CC(\tau)$, но зато она экспоненциальна по $IC_\mu(\tau)$. Отсюда тем же способом можно получить следующую оценку для проблемы прямой суммы:

$$D^{\nu,n}_{\varepsilon}(g) = \Omega\Big(\delta^2 n \cdot \log(D^{\nu}_{\varepsilon+\delta}(g))\Big).$$

То есть, грубо говоря, коммуникационная сложность $n$ копий функции $g$ хотя бы в $n$ раз больше, чем *логарифм* сложности одной копии.

## 4.1 Общие и частные случайные биты в сжатии протоколов и формулировка нашего результата

Резюмировать доказательство теоремы 9 можно так: мы берем протокол $\tau$ из теоремы 7 и применяем к нему сжатие из теоремы 8. При этом протокол $\tau$ вероятностный; для доказательства теоремы 7 важно, чтобы $\tau$ использовал как общие, так и частные случайные биты. С другой стороны оказывается, что протоколы, использующие только общие случайные биты, можно «сжимать лучше».

**Теорема 11** ([9, 32]). *Для любого протокола с* ***общими случайными битами*** $\pi$ *над* $(\mathcal{X}, \mathcal{Y})$*, любой функции* $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$*, любого распределения вероятностей* $\nu$ *на* $\mathcal{X} \times \mathcal{Y}$ *и любых* $\varepsilon, \delta \in (0,1)$ *выполнено следующее: если* $\tau$ *вычисляет функцию* $g$ *с ошибкой* $\varepsilon$ *по распределению* $\nu$*, то найдется детерминированный протокол* $\pi'$*, вычисляющий функцию* $g$ *с ошибкой* $\varepsilon + \delta$ *по распределению* $\nu$*, такой, что*

$$CC(\pi') = O\left(\frac{IC_\nu(\pi) \cdot \log(CC(\pi)/\delta)}{\delta}\right).$$

**Теорема 12** ([3]). *Для любого протокола с* ***общими случайными битами*** $\pi$ *над* $(\mathcal{X}, \mathcal{Y})$*, любой функции* $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$*, любого распределения вероятностей* $\nu$ *на* $\mathcal{X} \times \mathcal{Y}$ *и любых* $\varepsilon, \delta \in (0,1)$ *выполнено следующее: если* $\tau$ *вычисляет функцию* $g$ *с ошибкой* $\varepsilon$ *по распределению* $\nu$*, то найдется детерминированный протокол* $\pi'$*, вычисляющий функцию* $g$ *с ошибкой* $\varepsilon + \delta$ *по распределению* $\nu$*, такой, что*

$$CC(\pi') = O\left(\frac{(IC_\nu(\pi))^2 \cdot \log\log(CC(\pi))}{\delta^2}\right).$$

В первом результе оценка линейна по $IC_\nu(\tau)$ и логарифмична по $CC(\tau)$, а во втором — дважды логарифмична по $CC(\tau)$, зато квадратична по $IC_\nu(\tau)$.

Возникает вопрос, можно ли усилить оценку на сжатие произвольных протоколов так: сначала избавиться от частных случайных битов, не слишком увеличив информационную сложность, а потом применить результаты теорем 11, 12.

Несколько работ получили результаты в этом направлении. Для того, чтобы их сформулировать, дадим определение эквивалентного протокола.

**Определение 20.** *Пусть* $\tau$ *и* $\tau'$ *— два вероятностных протокола над* $(\mathcal{X}, \mathcal{Y})$*. Будем называть их эквивалентными, если для любого* $v \in \{0,1\}^*$ *и любой пары* $(x, y) \in \mathcal{X} \times \mathcal{Y}$ *выполнено равенство:*

$$p^\tau(v, x, y) = p^{\tau'}(v, x, y).$$

Браверман и Гарг доказали следующую теорему:

**Теорема 13** ([6]). *Для любого* ***однораундового*** *протокола с частными случайными битами* $\tau$ *над* $(\mathcal{X}, \mathcal{Y})$ *и любого вероятностного распределения* $\nu$ *на* $\mathcal{X} \times \mathcal{Y}$ *существует протокол с общими случайными битами* $\pi$*, эквивалентный* $\tau$*, такой, что*

$$IC_\nu(\pi) \le IC_\nu(\tau) + \log_2(IC_\nu(\tau) + O(1)) + O(1).$$

Ранее также для однораундовых протоколов Броди и др. [9] доказали чуть более слабую оценку.

Можно применять эти результаты пораундово и к многораундовым протоколам. Однако никакой нетривиальной оценки не получается, если количество раундов в протоколе $\tau$ сопоставимо с его коммуникационной длиной — добавочный член $O(1)$ суммируется по всем раундам, в результате получается оценка $IC_\nu(\pi) = O(CC(\tau))$, которую можно было получить, сгенерировав частные случайные биты $\tau$ в общем источнике случайности.

Наш основной результат находится в том же русле, но не зависит от количества раундов $\tau$:

**Теорема 14** (основной результат). *Для любого протокола частными случайными битами* $\tau$ *над* $(\mathcal{X}, \mathcal{Y})$ *и любого вероятностного распределения* $\nu$ *на* $\mathcal{X} \times \mathcal{Y}$ *существует протокол с общими случайными битами* $\pi$*, эквивалентный* $\tau$*, такой, что*

$$IC_\nu(\pi) = O\Big(\sqrt{IC_\nu(\tau) \cdot CC(\tau)}\Big), \qquad CC(\pi) \le CC(\tau).$$

Это результат интересен тем, что из него при помощи теоремы 11 можно вывести новое доказательство теоремы 8 о сжатии произвольных протоколов. Объясним идею, как это сделать. Нам нужно доказать, что вероятностный протокол $\tau$ с информационным разглашением $I$ и коммуникационной длиной $C$ можно сжать до протокола длины $O(\sqrt{IC} \cdot \log C)$ (мы опускаем для простоты множители, связанные с вероятностью ошибки). Для этого мы сначала избавляемся по нашей теореме 14 от частных случайных битов $\tau$, получая протокол $\pi$ с общими случайными битами, эквивалентный $\tau$, информационное разглашение которого не превосходит $O(\sqrt{IC})$ (а коммуникационная длина та же самая). Формально говоря, мы не можем это сделать сразу, поскольку наша теорема 14 сформулирована только для случая, когда $\tau$ — протокол с частными случайными битами. Тем не менее, если $\tau$ — вероятностный, то есть если $\tau$ является случайной величиной, принимающей значения в протоколах с частными случайными битами, то надо просто применить теорему 14 к каждому значению $\tau$. Таким образом мы получим искомый $\pi$.

Но поскольку $\pi$ не использует частных случайных битов, то теперь мы можем сжать $\pi$ согласно теореме 11 до протокола с коммуникационной длиной

$$O\Big(IC_\nu(\pi) \cdot \log(CC(\pi))\Big) = O(\sqrt{IC} \cdot \log C),$$

что и требовалось. Недостающие детали легко восстановить, но это не представляет большого интереса.

Доказательство теоремы 14 будет разбито на два шага. На первом шаге мы докажем лемму, которая, по сути, является частным случаем теоремы 14 для протоколов, в которых *пересылается* 1 *бит*. А именно, в этом частном случае теорема 14 утверждает следующее: для любого 1-битового протокола с частными случайными битами с информационным разглашением $I$ существует эквивалентный ему протокол с общими случайными битами с информационным разглашением $O(\sqrt{I})$ (поскольку речь идет об 1-битовых протоколах, то $I \leq 1$). Из этого уже несложно вывести теорему 14 для общего случая. А именно, применяя лемму побитово к протоколу, в котором пересылается $N$ бит, мы получим протокол с общими случайными битами, эквивалентный исходному, информационная разглашение которого благодаря неравенству Коши — Буняковского не будет превосходить:

$$O(\sqrt{I_1} + \ldots + \sqrt{I_N}) = O(\sqrt{(I_1 + \ldots + I_N) \cdot N},$$

где $I_k$ — это, грубо говоря разглашение $k$-го бита исходного протокола. Таким образом и будет получена оценка, являющаяся средним геометрическим $(I_1 + \ldots + I_N)$ (т. е. разглашения исходного протокола) и $N$ (т. е. коммуникационной длины исходного протокола).

## 4.2 Основная лемма

Здесь и до конца доказательства мы используем обозначение

$$r \in_R W \subset \mathbb{R}^n,$$

под которым имеется в виду, что $r$ выбирается с равномерным распределением из измеримого подмножества $W \subset \mathbb{R}^n$.

**Лемма 11.** *Для некоторого $D > 0$ выполнено следующее. Пусть $(X, B)$ — пара совместно распределенных случайных величин, где $X$ принимает значения в конечном множестве $\mathcal{X}$, а $B$ — в множестве $\{0, 1\}$. Определим*

$$f : \mathcal{X} \times [0, 1] \to \{0, 1\}, \qquad f(x, r) = \begin{cases} 0 & \text{если } r \leq \Pr[B = 0 | X = x] \\ 1 & \text{иначе.} \end{cases}$$

*Тогда $\mathbb{E}_{r \in_R [0,1]} H(F(X, r)) \leq D \cdot \sqrt{I(X : B)}$.*

*Доказательство.* Матожидание $\mathbb{E}_{r\in_R[0,1]}H(F(X,r))$ запишем в виде интеграла:

$$\mathbb{E}_{r\in_R[0,1]}H(FX,r)) = \int_0^1 H(F(X,r))dr,$$

а взаимную информацию распишем через расстояние Кульбака-Лейблера (см. предложение 15):

$$I(X:B) = \sum_{x\in\mathcal{X}} D_{KL}((B|X=x)||B)\cdot \Pr[X=x].$$

По неравенству Пинскера (см. предложение 14) выполнено:

$$I(X:B) \geq \frac{2}{\ln(2)}\cdot V,$$

где

$$V = \sum_{x\in\mathcal{X}} \delta^2((B|X=x),B)\cdot \Pr[X=x]. \tag{4.1}$$

Поэтому над достаточно доказать, что для некоторой абсолютной константы $D>0$ выполнено неравенство:

$$\int_0^1 H(F(X,r))dr \leq D\cdot\sqrt{V}. \tag{4.2}$$

Определим множество:

$$\Omega = \Big\{r\in[0,1] : |r-\Pr[B=0]| > \sqrt{2V}\Big\}.$$

Мы оценим интеграл из (4.2) отдельно по точкам из $\Omega$ и отдельно по точкам из $[0,1]\setminus\Omega$. Во-первых заметим, что

$$\int_{[0,1]\setminus\Omega} H(F(X,r))dr \leq 2\sqrt{2V}, \tag{4.3}$$

поскольку мера $[0,1]\setminus\Omega$ не превосходит $2\sqrt{2V}$, а $H(F(X,r))\leq 1$.

С другой стороны, мы сейчас установим, что

$$r\in\Omega \implies \min\{\Pr[F(X,r)=0],\Pr[F(X,r)=1]\} \leq \frac{V}{(r-\Pr[B=0])^2}. \tag{4.4}$$

Действительно, пусть, скажем, $r<\Pr[B=0]-\sqrt{2V}$. Зафиксируем произвольный $x\in\mathcal{X}$ такой, что $F(x,r)=1$. Из этого по определению $F$ вытекает, что $\Pr[B=0|X=x]<r$. С другой стороны, поскольку $\Pr[B=0]>r+\sqrt{2V}$, отсюда вытекает, что

$$\delta((B|X=x),B) = \Big|\Pr[B=0|X=x]-\Pr[B=0]\Big| \geq \Big|r-\Pr[B=0]\Big|.$$

Поэтому по неравенству Маркова, примененному к (4.1), мы получаем:

$$\begin{aligned}\Pr[F(X,r)=1] &= \sum_{x\in\mathcal{X}:\ F(x,r)=1} \Pr[X=x]\\ &\leq \sum_{x\in\mathcal{X}:\ \delta((B|X=x),B)\geq\left|r-\Pr[B=0]\right|} \Pr[X=x]\\ &\leq \frac{V}{(r-\Pr[B=0])^2}.\end{aligned}$$

Совершенно аналогично можно установить, что всех $r \in \Omega$ таких, что $r > \Pr[B = 0] + \sqrt{2V}$, выполнено $\Pr[F(X, r) = 0] \leq \frac{V}{(r - \Pr[B=0])^2}$. Таким образом, (4.4) доказано.

Для окончания доказательства леммы нам понадобится следующее техническое предложение:

**Предложение 29.** *Пусть* $h(\alpha) = \alpha \log_2 \left(\frac{1}{\alpha}\right) + (1 - \alpha) \log_2 \left(\frac{1}{1-\alpha}\right)$ *— функция Шеннона. Тогда при* $\alpha \in [0, 1/2]$ *выполнено:* $h(\alpha) \leq 2\alpha \log_2 \left(\frac{1}{\alpha}\right)$.

*Доказательство.* Достаточно доказать, что при $\alpha \in [0, 1/2]$ выполнено:

$$(1 - \alpha) \log_2 \left(\frac{1}{1 - \alpha}\right) \leq \alpha \log_2 \left(\frac{1}{\alpha}\right).$$

Рассмотрим функцию

$$f(\alpha) = \alpha \log_2 \left(\frac{1}{\alpha}\right) - (1 - \alpha) \log_2 \left(\frac{1}{1 - \alpha}\right).$$

Производная этой функции равна:

$$f'(\alpha) = \frac{1}{\ln(2)} \left(\ln \left(\frac{1}{\alpha(1 - \alpha)}\right) - 2\right).$$

Корни производной — это корни уравнения $\alpha(1 - \alpha) = 1/e^2$. Поскольку $1/e^2 < 1/4$, это уравнение имеет два различных корня $\alpha_0$ и $\alpha_1$, где $\alpha_0 \in (0, 1/2)$, $\alpha_1 = 1 - \alpha_0$. Легко также видеть, что при $\alpha < \alpha_0$ производная $f'(\alpha)$ положительна, а при $\alpha \in (\alpha_0, \alpha_1)$ — отрицательна. Поэтому $f(\alpha)$ возрастает на $[0, \alpha_0]$ и убывает на $[\alpha_0, 1/2]$. Поскольку $f(0) = 0, f(1/2) = 0$, это означает, что $f(\alpha) \geq 0$ для всех $\alpha \in [0, 1/2]$, что и требовалось. □

Чтобы оценить интеграл

$$\int_{\Omega} H(F(X, r))dr,$$

мы для каждого $r \in \Omega$ применим оценку:

$$\begin{aligned} H(F(X, r)) &= h\left(\min\{\Pr[F(X, r) = 0], \Pr[F(X, r) = 1]\}\right) \\ &\leq h\left(\frac{V}{(r - \Pr[B = 0])^2}\right) \\ &\leq 2\frac{V}{(r - \Pr[B = 0])^2} \log_2 \left(\frac{(r - \Pr[B = 0])^2}{V}\right). \end{aligned}$$

Здесь в первом неравенстве мы воспользовались (4.4), а также тем обстоятельством, что функция Шеннона возрастает на $[0, 1/2]$, причем

$$\frac{V}{(r - \Pr[B = 0])^2} \leq 1/2$$

для $r \in \Omega$. Во втором неравенстве мы воспользовались предложением, а также опять тем,

что $\frac{V}{(r-\Pr[B=0])^2} \le 1/2$ для $r \in \Omega$. Отсюда получаем следующую оценку:

$$\begin{aligned}\int_\Omega H(F(X,r)) &\le 2\int_\Omega \frac{V}{(r-\Pr[B=0])^2}\log_2\left(\frac{(r-\Pr[B=0])^2}{V}\right)dr \\ &= 2\sqrt{V}\int_\Omega \frac{V}{(r-\Pr[B=0])^2}\log_2\left(\frac{(r-\Pr[B=0])^2}{V}\right)d\left(\frac{r-\Pr[B=0]}{\sqrt{V}}\right) \\ &\le 2\sqrt{V}\int_{|y|\ge\sqrt{2}} \frac{\log_2(y^2)}{y^2}dy,\end{aligned}$$

а поскольку интеграл $\int_{|y|\ge\sqrt{2}} \frac{\log_2(y^2)}{y^2}dy$ сходится, последнее выражение не превосходит $D\sqrt{V}$ для некоторой абсолютной константы $D > 0$. Отсюда и из (4.3) вытекает (4.2), а вместе с ним и лемма. □

## 4.3 Завершение доказательства

Итак, нам дан протокол $\tau = (D, A, B, \phi, \psi)$ с частными случайными битами над $(\mathcal{X}, \mathcal{Y})$, а также распределение $\nu$ на $\mathcal{X}\times\mathcal{Y}$. Мы покажем, что найдется протокол $\pi$ с общими случайными битами, эквивалентный $\tau$, такой, что

$$IC_\nu(\pi) \le 2D\cdot\sqrt{IC_\nu(\tau)\cdot CC(\tau)}. \tag{4.5}$$

Здесь и всюду ниже $D$ — это константа из леммы 11. Протокол $\pi$ мы зададим как случайную величину, принимающую значения в конечном множестве детерминированных протоколов. Сначала опищем множество этих детерминированных протоколов. Пусть $N = CC(\tau)$. Для $r = (r_0, \dots, r_{N-1}) \in [0,1]^N$ через $\pi_r$ обозначим следующий детерминированный протокол. Пусть у Алисы на входе $x \in \mathcal{X}$, а у Боба на входе $y \in \mathcal{Y}$. Алиса и Боб заводят переменную $v$, пробегающую множество вершин $D$. В начале кладется $v = \Lambda$. Пока $v \notin \mathcal{L}(D)$, делается следующее. В случае $v \in A$ передает Алиса. Если $r_{|v|} \le \phi(x, v)$, она посылает 0, иначе 1. Аналогично в случае $v \in B$ передает Боб, и он посылает 0, если $r_{|v|} \le \psi(y, v)$, и 1 иначе. В момент, когда $v \in \mathcal{L}(D)$, протокол $\pi_r$ завершается. К этому моменту передано $|v|$ битов. Поэтому совершенно очевидно, что коммуникационная длина $\pi_r$ не превосходит $N = CC(\tau)$.

Мы зададим случайную величину $\pi$ как функцию, которая сопоставляет случайно выбранному $r \in [0,1]^N$ с равномерным распределением протокол $\pi_r$. Хоть и вероятностное пространство, на котором определено $\pi$, бесконечно, количество различных протоколов вида $\pi_r$ конечно, поскольку конечно количество различных детерминированных протоколов коммуникационной длины не больше $N$. Т. е. $\pi$ действительно является случайной величиной, принимающей конечное множество значений — просто мы задали ее распределение, используя бесконечное вероятностное пространство.

Покажем, что протокол $\pi$ эквивалентен $\tau$. Зафиксируем $r \in [0,1]^N$, а также $v = v_1 \dots v_k \in D$ и $(x, y) \in \mathcal{X}\times\mathcal{Y}$. Определим для всех $i \in \{1, 2, \dots, k\}$ функцию $J_i : [0,1] \to \{0,1\}$ такую, что $J_i(x)$ равно единице тогда и только тогда, когда выполнено одно из следующих условий:

- $v_1 \dots v_{i-1} \in A, v_i = 0$ и $x \le \phi(x, v_1 \dots v_{i-1})$;
- $v_1 \dots v_{i-1} \in A, v_i = 1$ и $x > \phi(x, v_1 \dots v_{i-1})$;
- $v_1 \dots v_{i-1} \in B, v_i = 0$ и $x \le \psi(y, v_1 \dots v_{i-1})$;

- $v_1 \dots v_{i-1} \in B, v_i = 0$ и $x > \psi(y, v_1 \dots v_{i-1})$.

По определению $\pi_r$ этот протокол приходит в $v$ на $(x, y)$ тогда и только тогда, когда $J_1(r_0) \cdot \dots \cdot J_k(r_{k-1}) = 1$. Поэтому:

$$p^\pi(v, x, y) = \mathbb{E}_{r \in_R [0,1]^N} J_1(r_0) \cdot \dots \cdot J_k(r_{k-1}) = \prod_{i=1}^{k} \mathbb{E}_{r_{i-1} \in_R [0,1]} J_i(r_{i-1}). \tag{4.6}$$

С другой стороны, для любого $i \in \{1, \dots, k\}$ матожидание $\mathbb{E}_{r_{i-1} \in_R [0,1]} J_i(r_{i-1})$ легко явно выписать:

$$\mathbb{E}_{r_{i-1} \in_R [0,1]} J_i(r_{i-1}) = \begin{cases} \phi(x, v_1 \dots v_{i-1}) & \text{если } v_1 \dots v_{i-1} \in A, v_i = 0, \\ 1 - \phi(x, v_1 \dots v_{i-1}) & \text{если } v_1 \dots v_{i-1} \in A, v_i = 1, \\ \psi(y, v_1 \dots v_{i-1}) & \text{если } v_1 \dots v_{i-1} \in B, v_i = 0, \\ 1 - \psi(y, v_1 \dots v_{i-1}) & \text{если } v_1 \dots v_{i-1} \in B, v_i = 1. \end{cases}$$

Поэтому согласно формальному определению функции $p^\tau(v, x, y)$ (см (1.2)) выполнено следующее:

$$\mathbb{E}_{r_{i-1} \in_R [0,1]} J_i(r_{i-1}) = p^\tau(v_1 \dots v_i, x, y) / p^\tau(v_1 \dots v_{i-1}, x, y).$$

Отсюда из (4.6) вытекает:

$$p^\pi(v, x, y) = \frac{p^\tau(v_1, x, y)}{p^\tau(\Lambda, x, y)} \cdot \frac{p^\tau(v_1 v_2, x, y)}{p^\tau(v_1, x, y)} \cdot \dots \cdot \frac{p^\tau(v_1 \dots v_k, x, y)}{p^\tau(v_1 \dots v_{k-1}, x, y)} = p^\tau(v, x, y)$$

поскольку $p^\tau(\Lambda, x, y) = 1$, то есть $\pi$ эквивалентен $\tau$.

Остается оценить информационное разглашение $\pi$. Рассмотрим сначала $IC_\nu(\tau)$. Пусть пара случайных величин $(X, Y)$ распределена согласно $\nu$. Величина $IC_\nu(\mu)$ по определению равна:

$$IC_\nu(\tau) = I(Y : T|X) + I(X : T|Y),$$

где $T$ — совместно распределенная с $(X, Y)$ случайная величина, принимающая значение в множестве $\mathcal{L}(D)$, такая, что для всех $v \in \mathcal{L}(D), (x, y) \in \mathcal{X} \times \mathcal{Y}$ выполнено:

$$\Pr[T = v | X = x, Y = y] = p^\tau(v, x, y).$$

Мы распишем $IC_\nu(\tau)$ с помощью цепного правила (см. предложение 11). Для этого нам будет удобно ввести случайную величину

$$\Pi = T \underbrace{00 \dots 0}_{N - |T|},$$

т. е. мы приписываем к $T$ столько нулей, чтобы получилось слово длины $N$ (иногда ничего не надо приписывать). Очевидно, что $\Pi$ является функцией от $T$, но верно и обратное, поскольку множество значений $T$ является префиксным множеством. Поэтому согласно предложению 7 имеем $H(Y|T, X) = H(Y|\Pi, X), H(X|T, Y) = H(X|\Pi, Y)$, откуда

$$I(Y : T|X) = I(Y : \Pi|X), \qquad I(X : T|Y) = I(X : \Pi|Y).$$

Таким образом, выполнено:

$$\begin{aligned} IC_\nu(\tau) &= I(Y : \Pi|X) + I(X : \Pi|Y) \\ &= \sum_{i=1}^{N} I(Y : \Pi_i | X \Pi_{<i}) + \sum_{i=1}^{N} I(X : \Pi_i | Y \Pi_{<i}), \end{aligned} \tag{4.7}$$

где $\Pi_{<i} = \Pi_1 \dots \Pi_{i-1}$. Здесь мы используем цепное правило (предложение 11).

Теперь рассмотрим величину $IC_\nu(\pi)$. По определению она равна:

$$IC_\nu(\pi) = \mathbb{E}_{r\in_R[0,1]^N} IC_\nu(\pi_r).$$

Теперь, чтобы расписать $IC_\nu(\pi_r)$, введем случайную величину $T^r$, являющуюся функцией $(X,Y)$ и равную листу, в который протокол $\pi_r$ приходит на $(X,Y)$. Тогда по определению:

$$IC_\nu(\pi_r) = I(Y : T^r|X) + I(X : T^r|Y) = H(T^r|X) + H(T^r|Y).$$

Здесь мы использовали определение взаимной информации и тот факт, что $H(T^r|X,Y) = 0$ (это выполнено, поскольку $T^r$ является функцией от $(X,Y)$). Аналогично тому, что у нас было раньше, определим:

$$\Pi^r = T^r \underbrace{00\dots0}_{N-|T^r|},$$

и, используя цепное правило для условной энтропии (предложение 8), напишем:

$$\begin{aligned} IC_\nu(\pi) &= \mathbb{E}_{r\in_R[0,1]^N} IC_\nu(\pi_r) \\ &= \mathbb{E}_{r\in_R[0,1]^N} \left( \sum_{i=1}^N H(\Pi_i^r|X\Pi_{<i}^r) + \sum_{i=1}^N H(\Pi_i^r|Y\Pi_{<i}^r) \right) \\ &= \sum_{i=1}^N \mathbb{E}_{r\in_R[0,1]^N} H(\Pi_i^r|X\Pi_{<i}^r) + \sum_{i=1}^N \mathbb{E}_{r\in_R[0,1]^N} H(\Pi_i^r|Y\Pi_{<i}^r). \end{aligned} \tag{4.8}$$

где мы использовали обозначение $\Pi_{<i}^r = \Pi_1^r \dots \Pi_{i-1}^r$.

Мы докажем, используя лемму 11, что для любого $i \in \{1,2,\dots,N\}$ выполнено следующее:

$$\mathbb{E}_{r\in_R[0,1]^N} H(\Pi_i^r|X\Pi_{<i}^r) \le D\sqrt{I(Y : \Pi_i|X\Pi_{<i})}, \tag{4.9}$$

$$\mathbb{E}_{r\in_R[0,1]^N} H(\Pi_i^r|Y\Pi_{<i}^r) \le D\sqrt{I(X : \Pi_i|Y\Pi_{<i})}, \tag{4.10}$$

откуда благодаря (4.7),(4.8) через неравенство Коши-Буняковского легко вытекает (4.5). Мы докажем только (4.10), (4.9) доказывается совершенно аналогично.

Заметим, что при фиксированных $X = x, Y = y$ величина $\Pi_{<i}^r$ зависит только от $r_0,\dots,r_{i-2}$, а величина $\Pi_i^r$ — только от $r_0,\dots,r_{i-1}$. Поэтому для $z \in [0,1]^{i-1}, t \in [0,1]$ мы введем функции:

$$\Pi_{<i}^z : \mathcal{X}\times\mathcal{Y} \to \{0,1\}^{i-1}, \qquad \Pi_{<i}^z(x,y) = \Pi_{<i}^r(x,y),$$

$$\Pi_i^{z,t} : \mathcal{X}\times\mathcal{Y} \to \{0,1\}, \qquad \Pi_i^{z,t}(x,y) = \Pi_i^r(x,y),$$

где $r = (z,t,\underbrace{0,\dots,0}_{N-i})$.

В этих обозначениях (4.10) перепишется так:

$$\int_0^1 dt \int_{[0,1]^{i-1}} H(\Pi_i^{z,t}|Y\Pi_{<i}^z)dz \le D\sqrt{I(X : \Pi_i|Y\Pi_{<i})}. \tag{4.11}$$

Для каждого $y \in \mathcal{Y}, v \in \{0,1\}^{i-1}$ введем функцию $f_{y,v} : \mathcal{X}\times[0,1] \to \{0,1\}$. Если $v \in A$, положим:

$$f_{y,v}(x,t) = \begin{cases} 0 & \text{если } t \le \phi(x,v) \\ 1 & \text{иначе.} \end{cases}$$

Если $v \in B$, положим

$$f_{y,v}(x,t) = \begin{cases} 0 & \text{если } t \le \psi(y,v) \\ 1 & \text{иначе.} \end{cases}$$

Если же $v \notin \mathcal{I}(D)$, положим $f_{y,v}(x,t) = 0$. Заметим, что по определению протокола $\pi$ выполнено следующее: для любых $y \in \mathcal{Y}, v \in \{0,1\}^{i-1}$ и для любых $z \in [0,1]^{i-1}, t \in [0,1]$ при условии $Y = y, \Pi^z_{<i} = v$ случайная величина $\Pi^{z,t}_i$ равна 0 тогда и только тогда, когда $f_{y,v}(X,t) = 0$. В частности отсюда следует:

$$\Pr[\Pi^{z,t}_i = 0 | Y = y, \Pi^z_{<i} = v] = \Pr[f_{y,v}(X,t) = 0 | Y = y, \Pi^z_{<i} = v]. \tag{4.12}$$

Мы в начале покажем, что для любого $t \in [0,1]$ выполнено следующее:

$$\int\limits_{[0,1]^{i-1}} H(\Pi^{z,t}_i | Y \Pi^z_{<i}) dz \le H(f_{Y,\Pi_{<i}}(X,t) | Y \Pi_{<i}). \tag{4.13}$$

Распишем левую часть (4.13) по определению условной энтропии:

$$\begin{aligned} &\int\limits_{[0,1]^{i-1}} H(\Pi^{z,t}_i | Y \Pi^z_{<i}) dz \\ &= \int\limits_{[0,1]^{i-1}} dz \sum_{y,v} \Pr[Y = y, \Pi^z_{<i} = v] H(\Pi^{z,t}_i | Y = y, \Pi^z_{<i} = v) \\ &= \sum_{y,v} \Pr[Y = y, \Pi_{<i} = v] \int\limits_{[0,1]^{i-1}} \frac{\Pr[Y = y, \Pi^z_{<i} = v]}{\Pr[Y = y, \Pi_{<i} = v]} H(\Pi^{z,t}_i | Y = y, \Pi^z_{<i} = v) dz, \end{aligned} \tag{4.14}$$

где суммирование ведется по всем $y \in \mathcal{Y}, v \in \{0,1\}^{i-1}$. Мы оценим интеграл

$$\int\limits_{[0,1]^{i-1}} \frac{\Pr[Y = y, \Pi^z_{<i} = v]}{\Pr[Y = y, \Pi_{<i} = v]} H(\Pi^{z,t}_i | Y = y, \Pi^z_{<i} = v) dz$$

сверху выражением $H(f_{y,v}(X,t) | Y = y, \Pi_{<i} = v)$. Для этого мы воспользуемся выпуклостью энтропии вверх (см. предложение 5). А именно, из выпуклости энтропии вытекает следующее. Пусть $\Omega$ — измеримое подмножество $\mathbb{R}^k$, а $w(z)$ — неотрицательная функция, интеграл которой по $\Omega$ равен единице. Пусть для каждого $z \in \Omega$ задано распределение вероятностей $\mu_z$ на множестве $\{0,1\}$. Зададим распределение $\mu$ как усреднение $\mu_z$ с весом $w(z)$:

$$\mu(0) = \int\limits_\Omega w(z) \mu_z(0) dz, \qquad \mu(1) = \int\limits_\Omega w(z) \mu_z(1) dz.$$

Тогда $\int\limits_\Omega w(z) H(\mu_z) \le H(\mu)$.

В нашем случае $\Omega = [0,1]^{i-1}$, «весовая» функция $w(z)$ имеет вид

$$w(z) = \frac{\Pr[Y = y, \Pi^z_{<i} = v]}{\Pr[Y = y, \Pi_{<i} = v]},$$

а распределение $\mu_z$ есть условное распределение $\Pi^{z,t}_i | Y = y, \Pi_{<i} = v$.

Проверим, что интеграл $w(z)$ по $\Omega$ действительно равен 1. Это следует из уже доказанной эквивалентности протоколов $\pi$ и $\tau$:

$$\int\limits_{[0,1]^{i-1}} \frac{\Pr[Y=y,\Pi^z_{<i}=v]}{\Pr[Y=y,\Pi_{<i}=v]}dz = \frac{\int\limits_{[0,1]^{i-1}} \Pr[Y=y,\Pi^z_{<i}=v]dz}{\Pr[Y=y,\Pi_{<i}=v]} = \frac{\Pr[Y=y,\Pi_{<i}=v]}{\Pr[Y=y,\Pi_{<i}=v]} = 1. \tag{4.15}$$

Теперь убедимся, что усреднение $\mu_z$ с весом $w(z)$ будет условным распределением $f_{y,v}(X,t)|Y=y,\Pi_{<i}=v$:

$$\begin{aligned}
&\int\limits_{[0,1]^{i-1}} \frac{\Pr[Y=y,\Pi^z_{<i}=v]}{\Pr[Y=y,\Pi_{<i}=v]} \cdot \Pr[\Pi^{z,t}_i=0|Y=y,\Pi^z_{<i}=v]dz \\
&= \int\limits_{[0,1]^{i-1}} \frac{\Pr[Y=y,\Pi^z_{<i}=v]}{\Pr[Y=y,\Pi_{<i}=v]} \cdot \Pr[f_{y,v}(X,t)=0|Y=y,\Pi^z_{<i}=v]dz \\
&= \frac{\int\limits_{[0,1]^{i-1}} \Pr[f_{y,v}(X,t)=0,Y=y,\Pi^z_{<i}=v]dz}{\Pr[Y=y,\Pi_{<i}=v]} \\
&= \frac{\Pr[f_{y,v}(X,t)=0,Y=y,\Pi_{<i}=v]}{\Pr[Y=y,\Pi_{<i}=v]} \\
&= \Pr[f_{y,v}(X,t)=0|Y=y,\Pi_{<i}=v].
\end{aligned} \tag{4.16}$$

Здесь во второй строчке мы воспользовались (4.12), а в четвертой — опять тем, что $\pi$ и $\tau$ эквивалентны.

Поэтому из (4.15), (4.16) мы получаем:

$$\begin{aligned}
&\int\limits_{[0,1]^{i-1}} \frac{\Pr[Y=y,\Pi^z_{<i}=v]}{\Pr[Y=y,\Pi_{<i}=v]} H(\Pi^{z,t}_i|Y=y,\Pi^z_{<i}=v)dz \\
&\le H(f_{y,v}(X,t)|Y=y,\Pi_{<i}=v),
\end{aligned}$$

и наконец, из (4.14) получаем (4.13)

Учитывая (4.11), для доказательства (4.10) нам остается показать, что

$$\int\limits_0^1 H(f_{Y,\Pi_{<i}}(X,t)|Y,\Pi_{<i})dt \le D\sqrt{I(X:\Pi_i|Y,\Pi_{<i})}.$$

В силу выпуклости корня вверх достаточно доказать это неравенство при любых фиксированных $Y=y,\Pi_{<i}=v$:

$$\int\limits_0^1 H(f_{y,v}(X,t)|Y=y,\Pi_{<i}=v)dt \le D\sqrt{I(X:\Pi_i|Y=y,\Pi_{<i}=v)}.$$

Действительно, если $v \notin A$, то при условии $Y=y,\Pi_{<i}=v$ случайная величина $f_{y,v}(X,t)$ является константой, поэтому левая часть равна нулю и неравенство выполнено по очевидным соображениям. Если же $v \in A$, то мы воспользуемся нашей основной леммой 11, примененной к паре случайных величин $(X,\Pi_i)|Y=y,\Pi_{<i}=v$. Нам остается только проверить, что

$$\phi(x,v) = \Pr[\Pi_i=0|X=x,Y=y,\Pi_{<i}=v].$$

Действительно, по определению случайной величины $\Pi$ имеем:

$$\begin{aligned}\Pr[\Pi_i = 0 | X = x, Y = y, \Pi_{<i} = v] &= \frac{\Pr[\Pi_i = 0, X = x, Y = y, \Pi_{<i} = v]}{\Pr[X = x, Y = y, \Pi_{<i} = v]} \\ &= \frac{\Pr[\Pi_{<i+1} = v0, X = x, Y = y]}{\Pr[X = x, Y = y, \Pi_{<i} = v]} \\ &= \frac{\Pr[\Pi_{<i+1} = v0 | X = x, Y = y]}{\Pr[\Pi_{<i} = v | X = x, y = y]} \\ &= \frac{p^\tau(x, y, v0)}{p^\tau(x, y, v)} = \phi(x, v).\end{aligned}$$

# Глава 5

# Интерактивные аналоги теоремы Вольфа — Слепяна

Эта глава посвящена верхним и нижним оценкам для интерактивного аналога теоремы Вольфа — Слепяна.

Для начала напомним формулировку теоремы Вольфа — Слепяна ([37]). Рассмотрим пару случайных величин $X$ и $Y$, имеющих вид

$$X = X_1 \dots X_n, \qquad Y = Y_1 \dots Y_n.$$

где пары случайных величин $(X_1, Y_1), \dots, (X_n, Y_n)$ независимы и одинаково распределены. Множество значений $X_i$ обозначим через $\Sigma$, множество значений $Y_i$ — через $\Gamma$. Предположим, Алиса получает на вход $X$, а Боб — $Y$. Задача Алисы — передать $X$ Бобу, быть может, с не большой вероятностью ошибки. Теорема Вольфа — Слепяна утверждает, грубо говоря, что Алисе для этого достаточно передать одно сообщение фиксированной битовой длины, а именно $\approx H(X_1|Y_1)n$. Более строго эту теорему можно сформулировать так: для любого $\varepsilon > 0$ существуют функции $E : \Sigma^n \to \{0,1\}^k$ и $D : \{0,1\}^k \times \Gamma^n \to \Sigma^n$, где $k = H(X_1|Y_1)n + O_\varepsilon(\sqrt{n})$, такие, что:

$$\Pr[D(E(X), Y) \neq X] \leq \varepsilon.$$

Алиса, получив на вход $X$, посылает Бобу сообщение $E(X)$. Боб декодирует $X$, используя функцию $D$ — он применяет $D$ к $E(X)$ и $Y$.

Задача, изучаемая в этой главе, также предполагает, что на входе у Алисы есть значение некоторой случайной величины $X$, а у Боба — значение некоторой другой случайной величины $Y$, совместно распределенной с $X$, при этом Алиса хочет передать $X$ Бобу. Отличаться от теоремы Вольфа — Слепяна наша задача будет в следующих отношениях.

- Мы считаем случайные величины $X$ и $Y$ произвольными. В частности, мы не предполагаем, что они получены серией большого числа независимых испытаний.
- Мы разрешаем Бобу посылать сообщения Алисе.
- Мы интересуемся средней длиной коммуникации, а не длиной коммуникации в худшем случае.

Перейдем к формальным определениям. Пусть фиксированы совместно распределенные случайные величины $X$ и $Y$, где $X$ принимает значения в конечном множестве $\mathcal{X}$, а $Y$ — в конечном множестве $\mathcal{Y}$. Скажем, что детерминированный протокол $\tau = \langle D, A, B, \phi, \psi \rangle$ над $(\mathcal{X}, \mathcal{Y})$ *решает задачу Вольфа — Слепяна для* $(X, Y)$ *с ошибкой* $\varepsilon$, если существует функция $\theta : \mathcal{L}(D) \times \mathcal{Y} \to \mathcal{X}$, такая, что:

$$\Pr[\theta(\tau(X, Y), Y) \neq X] \leq \varepsilon.$$

Напомним, что $\mathcal{L}(D)$ — это множество листьев $D$, а через $\tau(x, y)$ обозначен элемент $\mathcal{L}(D)$, в который приходит протокол $\tau$ на входе $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Иными словами, процесс передачи $X$ от Алисы к Бобу происходит следующим образом. Сначала Алиса и Боб общаются согласно протоколу $\tau$ на входе $(X, Y)$. Затем Боб выдает свою догадку о том, чему равно $X$. Он это делает на основе своего входа $Y$, а также на основе своей коммуникации с Алисой (то есть на основе $\tau(X, Y)$). Формально, Боб выдает $\theta(\tau(X, Y), Y)$.

Мы будем также рассматривать протоколы с общими случайными битами для этой задачи. Будем говорить, что протокол с общими случайными битами $\tau$ над $(\mathcal{X}, \mathcal{Y})$ с пространством общих случайных битов $(\mathcal{R}, \mu)$ *решает задачу Вольфа — Слепяна для* $(X, Y)$ *с ошибкой* $\varepsilon$, если для любого $r \in \mathcal{R}$ существует $\varepsilon_r \in [0, 1]$ такое, что:

- для любого $r \in \mathcal{R}$ детерминированный протокол $\tau_r$ решает задачу Вольфа — Слепяна для $(X, Y)$ с ошибкой $\varepsilon_r$;
- $\mathbb{E}_{r \sim \mu} \varepsilon_r \leq \varepsilon$.

## 5.1 Верхние и нижние оценки для интерактивного аналога теоремы Вольфа — Слепяна: формулировки результатов

Мы получаем следующую верхнюю оценку.

**Теорема 15.** *Пусть $(X, Y)$ — пара совместно распределенных случайных величин, где $X$ принимает значения в конечном множестве $\mathcal{X}$, а $Y$ — в конечном множестве $\mathcal{Y}$. Обозначим через $\nu$ распределение $(X, Y)$. Тогда для любого $\varepsilon > 0$ и любого целого положительного $r$ существует протокол $\tau$ с общими случайными битами, решающий задачу Вольфа — Слепяна для $(X, Y)$ с ошибкой $\varepsilon$ такой, что*

- *средняя длина $\tau$ по распределению $\nu$ не превосходит*

$$H(X|Y) + \frac{H(X|Y)}{r} + r + \log_2\left(\frac{1}{\varepsilon}\right) + 2;$$

- *среднее число раундов $\tau$ по распределению $\nu$ не превосходит*

$$\frac{2H(X|Y)}{r} + 2.$$

Если забыть о среднем числе раундов $\tau$ и минимизировать лишь среднюю длину $\tau$, то надо положить $r = \lceil\sqrt{H(X|Y)}\rceil$. Мы получим протокол с общими случайными битами со средней длиной $H(X|Y) + 2\sqrt{H(X|Y)} + O(\log(1/\varepsilon))$, решающий задачу Вольфа — Слепяна для $(X, Y)$ с ошибкой $\varepsilon$. До этого наилучшая верхняя оценка принадлежала Браверману и Рао ([7]) и имела вид $H(X|Y) + 5\sqrt{H(X|Y)} + O(\log(1/\varepsilon))$.

Как в нашем результате, так и в результате Бравермана и Рао, среднее число раундов при этом равно $O(\sqrt{H(X|Y)})$. Что будет, если ограничить среднее число раундов сильнее? Для любого $\alpha \in [0, \frac{1}{2}]$, положив $r = H^{1-\alpha}$, из теоремы 15 мы получаем протокол со средней длиной $H + O(H^{1-\alpha})$ и средним числом раундов $O(H^{\alpha})$ (здесь $H = H(X|Y)$). Ранее протоколы с такими свойствами были построены лишь для $\alpha = 0$ в [8, 9] (в этом случае среднее число раундов в протоколе константно, а средняя длина равна $O(H)$) и для $\alpha = 1/2$ (этому случаю соответствует упомянутый выше результат Бравермана — Рао).

В упомянутых выше оценках на среднюю длину коммуникации членом первого порядка является условная энтропия $H(X|Y)$. Из теоретико-информационных соображений легко выводится, что меньший член первого порядка получить нельзя. А именно, верна следующая оценка:

**Предложение 30.** *Рассмотрим произвольную пару совместно распределенных случайных величин $(X, Y)$, принимающих конечное множество значений. Пусть $\mathcal{X}$ — множество значений $X$, а $\mathcal{Y}$ — множество значений $Y$. Предположим $\tau$ — протокол с общими случайными битами, решающий задачу Вольфа – Слепяна для $(X, Y)$ с ошибкой $\varepsilon$. Тогда средняя длина $\tau$ по распределению $(X, Y)$ не меньше*

$$H(X|Y) - \varepsilon \log_2(|\mathcal{X}|) - 1.$$

*Доказательство.* Через $\nu$ обозначим распределение $(X, Y)$.

Достаточно доказать предложение для детерминированных протоколов, поскольку нижняя оценка зависит от $\varepsilon$ линейно. Чуть более подробно, пусть протокол с общими случайными битами $\tau$ с пространством общих случайных битов $(\mathcal{R}, \mu)$ решает задачу Вольфа — Слепяна для $(X, Y)$ с ошибкой $\varepsilon$. Тогда для любого $r \in \mathcal{R}$ найдется $\varepsilon_r$ такое, что

- детерминированный протокол $\tau_r$ решает задачу Вольфа — Слепяна для $(X, Y)$ с ошибкой $\varepsilon_r$;
- $\mathbb{E}_{r\sim\mu}\varepsilon_r \leq \varepsilon$.

Если считать доказанным предложение для детерминированных протоколов, то $ACC_\nu(\tau_r) \geq H(X|Y) - \varepsilon_r \log_2(\mathcal{X}) - 1$. Тогда по определению средней длины протокола с общими случайными битами:

$$\begin{aligned} ACC_\nu(\tau) &= \mathbb{E}_{r\sim\mu} ACC_\nu(\tau_r) \geq \mathbb{E}_{r\sim\mu}(H(X|Y) - \varepsilon_r \log_2(|\mathcal{X}|) - 1) \\ &\geq H(X|Y) - \varepsilon \log_2(|\mathcal{X}|) - 1. \end{aligned}$$

Таким образом, можно считать $\tau = \langle D, A, B, \phi, \psi\rangle$ детерминированным. Рассмотрим взаимную информацию $I(X : \tau(X, Y)|Y)$. С одной стороны:

$$I(X : \tau(X, Y)|Y) \leq H(\tau(X, Y)) \leq ACC_\nu(\tau). \tag{5.1}$$

Поясним последнее неравенство. Множество значений $\tau(X, Y)$ есть префиксное множество. Поэтому согласно предложению 16 энтропия $\tau(X, Y)$ не превосходит средней длины $\tau(X, Y)$. А средняя длина $\tau(X, Y)$ как раз равна $ACC_\nu(\tau)$.

С другой стороны, $I(X : \tau(X, Y)|Y) = H(X|Y) - H(X|\tau(X, Y), Y)$. По определению, существует функция $\theta : \mathcal{L}(D) \times \mathcal{Y} \to \mathcal{X}$ такая, что

$$\Pr[\theta(\tau(X, Y), Y) \neq X] \leq \varepsilon.$$

Применяя неравенство Фано (предложение 17) к функции $\theta$, из (5.1) получаем:

$$ACC_\nu(\tau) \geq H(X|Y) - \varepsilon \log_2(|\mathcal{X}|) - 1.$$

□

Получаем при $\varepsilon = \Omega(1/\log|\mathcal{X}|)$ нижнюю оценку $H(X|Y) - O(1)$. С другой стороны, из этого предложения получить нижнюю оценку больше $H(X|Y)$ нельзя (что не удивительно, поскольку предложение выполнено для любых $X, Y$). Интересно получить пример $X, Y$ с нижней оценкой, значительно превышающей $H(X|Y)$. Максимум на что мы можем рассчитывать — это нижняя оценка вида $H(X|Y) + \Omega(\sqrt{H(X|Y)} + \log_2(1/\varepsilon))$. Мы же приводим

пример $X, Y$ с нижней оценкой вида $H(X|Y) + \Omega(\log_2(1/\varepsilon))$ ( значительно более слабой). Если быть точным, нижняя оценка будет выполняться для всех $\varepsilon$ в интервале $[1/n, 1/\log_2(n)]$, где $n$ — количество различных значений, принимаемых $X$.

Наш пример $X, Y$ выглядит следующим образом. Для $n \in \mathbb{N}, \gamma \in (0, 1/2)$ определим пару совместно распределенных случайных величин $X_n^\gamma, Y_n^\gamma$, каждая из которых принимает значения в множестве $\{0, 1, \ldots, n\}$, следующим образом:

$$\Pr[X_n^\gamma = i, Y_n^\gamma = j] = \frac{(1-\gamma)\delta_{ij} + \frac{\gamma}{n}(1-\delta_{ij})}{n+1}, \qquad i, j \in \{0, 1, 2, \ldots, n\}, \tag{5.2}$$

где

$$\delta_{ij} = \begin{cases} 1 & \text{если } i = j, \\ 0 & \text{если } i \neq j. \end{cases}$$

Легко убедиться, что сумма $\Pr[X_n^\gamma = i, Y_n^\gamma = j]$ по всем $i, j \in \{0, 1, \ldots, n+1\}$ равна 1, то есть формула (5.2) действительно задает вероятностное распределение. Кроме того видно, что и $X_n^\gamma$, и $Y_n^\gamma$ распределены равномерно на $\{0, 1, \ldots, n+1\}$. Конечно, при этом $X_n^\gamma$ и $Y_n^\gamma$ не являются независимыми. А именно, при известном $Y_n^\gamma = j$ случайная величина $X_n^\gamma$ принимает значение $j$ с вероятностью $1 - \gamma$, а каждое из оставшихся $n$ значений — с вероятностью $\frac{\gamma}{n}$. Отсюда:

$$H(X_n^\gamma | Y_n^\gamma) = (1-\gamma)\log_2\left(\frac{1}{1-\gamma}\right) + \gamma\log_2\left(\frac{n}{\gamma}\right) = \gamma\log_2(n) + O(1). \tag{5.3}$$

Мы получаем следующую нижнюю оценку на среднюю длину протоколов, решающих задачу Вольфа — Слепяна для $(X_n^\gamma, Y_n^\gamma)$.

**Теорема 16.** *Пусть $\pi$ — протокол с общими случайными битами, решающий задачу Вольфа — Слепяна для $(X_n^\gamma, Y_n^\gamma)$ с ошибкой $\varepsilon$. Тогда для $n \geq 2$ средняя длина $\pi$ по распределению пары $(X_n^\gamma, Y_n^\gamma)$ не меньше:*

$$(1 - \gamma - \gamma/n)\log_2\left(\frac{\gamma}{\varepsilon + \gamma/n}\right) + (\gamma - 2\varepsilon)\log_2(n+1).$$

*При $\frac{1}{n} \leq \varepsilon \leq \frac{1}{\log_2(n)}$ и $\gamma = \Omega(1)$ это выражение приобретает вид:*

$$\gamma\log_2(n) + (1-\gamma)\log_2\left(\frac{1}{\varepsilon}\right) - O(1) = H(X_n^\gamma | Y_n^\gamma) + (1-\gamma)\log_2\left(\frac{1}{\varepsilon}\right) - O(1).$$

Интересно, что при $\varepsilon = 1/n$ нижняя оценка становится равной $\log_2(n) - O(1)$, что лишь на константу битов хуже тривиального протокола, в котором Алиса пересылает Бобу $X_n^\gamma$, используя $\log_2(n+1)$ бит.

## 5.2 Верхняя оценка

Пусть $A$ — конечное множество. Рассмотрим произвольные $\varepsilon > 0$ и $r \in \mathbb{N}$. Мы сейчас определим вспомогательный протокол с общими случайными битами, который мы будем обозначать через $\tau_{r,\varepsilon}^A$. В нем Алиса получает на вход элемент $a \in A$, а Боб — распределение вероятностей $\mu$ на множестве $A$. Алиса и Боб воспринимают общую случайность как достаточно длинную последовательность независимых бернуллиевских случайных величин, каждая из которых принимает значение 1 с вероятностью 1/2. Первые $|A|$ битов этой последовательности естественным образом определяют функцию $h_1 : A \to \{0, 1\}$, вторые $|A|$ битов — функцию $h_2 : A \to \{0, 1\}$, и так далее.

Положим $k = \lceil \log_2(1/\varepsilon) \rceil$. Также для всех $i = 0, 1, \ldots$ положим:

$$S_i = \left\{ x \in A : \mu(x) \in (2^{-i-1}, 2^i] \right\}.$$

Множества $S_i$ известны Бобу. Они, очевидно, не пересекаются и задают разбиение множества $A$:

$$A = \bigcup_{i \geq 0} S_i.$$

В начале Алиса посылает Бобу $h_1(a) \ldots h_k(a)$. Далее протокол работает по этапам, нумерация которых начинается с единицы. В конце Боб выдает элемент $a' \in A$.

На этапе номер $t$ происходит следующее:

1. Алиса посылает Бобу
$$h_{k+r(t-1)+1}(a) \ldots h_{k+rt}(a).$$
2. Для каждого $i \in \{r(t-1), \ldots, rt-1\}$ Боб вычисляет множества $S_i'$, определяемые следующим образом:
$$S_i' = \{x \in S_i : h_1(x) \ldots h_{k+rt}(x) = h_1(a) \ldots h_{k+rt}(a)\}.$$
3. Если для некоторого $i \in \{r(t-1), \ldots, rt-1\}$ множество $S_i'$ непусто, то для наименьшего такого $i$ Боб кладет $a'$ равным какому-нибудь элементу $S_i'$ (не важно какому), посылает 1 Алисе, после чего протокол завершается. Иначе Боб посылает Алисе 0, и они переходят к этапу $t+1$.

Нам потребуется доказать несколько лемм о протоколе $\tau^A_{r,\varepsilon}$. Через $t_{max}$ обозначим такое целое число, что

$$r(t_{max} - 1) \leq \left\lfloor \log_2 \left( \frac{1}{\mu(a)} \right) \right\rfloor \leq r t_{max} - 1. \tag{5.4}$$

**Лемма 12.** *Если в $\tau^A_{r,\varepsilon}$ на входе у Алисы $a$, а у Боба — $\mu$, то протокол длится не более $t_{max}$ этапов, где $t_{max}$ определяется из (5.4).*

*Доказательство.* Положим $i = \left\lfloor \log_2 \left( \frac{1}{\mu(a)} \right) \right\rfloor$. Заметим, что

$$2^{-i-1} < \mu(a) = 2^{-\log_2\left(\frac{1}{\mu(a)}\right)} \leq 2^{-i},$$

поэтому $a \in S_i$. Значит $S_i'$ не пусто. При этом $S_i'$ рассматривается Бобом как раз на этапе $t_{max}$, поэтому больше этапов не может быть. □

**Лемма 13.** *Если в $\tau^A_{r,\varepsilon}$ на входе у Алисы $a$, а у Боба — $\mu$, то*

- *Алиса передает не больше*
$$\log_2 \left( \frac{1}{\mu(a)} \right) + r + \log_2 \left( \frac{1}{\varepsilon} \right) + 1 \text{ битов;}$$
- *Боб передает не больше*
$$\frac{\log_2 \left( \frac{1}{\mu(a)} \right)}{r} + 1 \text{ битов.}$$

*Доказательство.* Поскольку по лемме 12 протокол длится не более $t_{max}$ этапов, то Алиса передает не более $k+rt_{max}$ битов, а Боб — не более $t_{max}$ битов. Из левой части (5.4) вытекает, что:

$$t_{max} \le \frac{\log_2\left(\frac{1}{\mu(a)}\right)}{r} + 1,$$

что сразу дает оценку на количество битов, переданных Бобом. Алиса же передает не более:

$$\begin{aligned} k + rt_{max} &\le k + r\left(\frac{\log_2\left(\frac{1}{\mu(a)}\right)}{r} + 1\right) \\ &= \lceil \log_2(1/\varepsilon) \rceil + \log_2\left(\frac{1}{\mu(a)}\right) + r \\ &= \log_2\left(\frac{1}{\mu(a)}\right) + r + \log_2\left(\frac{1}{\varepsilon}\right) + 1 \end{aligned}$$

битов. □

**Лемма 14.** *Для любого $a \in A$ и любого распределения вероятностей $\mu$ на $A$, когда в $\tau^A_{r,\varepsilon}$ на входе у Алисы $a$, а у Боба — $\mu$, то с вероятностью хотя бы $1-\varepsilon$ (по общим случайным битам) выполнено $a' = a$.*

*Доказательство.* Для всякого натурального $i$ через $t(i)$ обозначим натуральное число, удовлетворяющее:

$$r \cdot (t(i) - 1) \le i \le r \cdot t(i) - 1.$$

Через $E_i$ обозначим событие, что для некоторого $x \in S_i$, отличного от $a$, выполнено:

$$h_1(x) \dots h_{k+r\cdot t(i)}(x) = h_1(a) \dots h_{k+r\cdot t(i)}(a). \tag{5.5}$$

Для каждого фиксированного $x \in S_i \setminus \{a\}$ равенство (5.5) выполняется с вероятностью $2^{-k-r\cdot t(i)}$, поэтому вероятность $E_i$ можно оценить так:

$$\Pr[E_i] \le |S_i| \cdot 2^{-k-r\cdot t(i)} \le |S_i| \cdot 2^{-k-i-1}, \tag{5.6}$$

где последнее неравенство выполнено в силу $i \le r \cdot t(i) - 1$.

С другой стороны, если $a' \ne a$, то по определению протокола $\tau^A_{r,\varepsilon}$ для некоторого $i$ выполнено событие $E_i$, поэтому благодаря (5.6) мы получаем оценку:

$$\begin{aligned} \Pr[a' \ne a] &\le \sum_{i=0}^{\infty} \Pr[E_i] \le \sum_{i=0}^{\infty} |S_i| \cdot 2^{-k-i-1} \\ &= 2^{-\lceil \log_2(1/\varepsilon) \rceil} \sum_{i=0}^{\infty} |S_i| \cdot 2^{-i-1} \le \varepsilon \sum_{i=0}^{\infty} |S_i| \cdot 2^{-i-1}. \end{aligned}$$

Остается показать, что $\sum\limits_{i=0}^{\infty} |S_i| \cdot 2^{-i-1} \le 1$. Заметим, что по определению для любого $x \in S_i$ выполнено $\mu(x) \ge 2^{-i-1}$. Поэтому:

$$\sum_{x \in S_i} \mu(x) \ge |S_i| \cdot 2^{-i-1}.$$

А значит сумма $\sum\limits_{i=0}^{\infty} |S_i| \cdot 2^{-i-1}$ оценивается так:

$$\sum_{i=0}^{\infty} |S_i| \cdot 2^{-i-1} \le \sum_{i=0}^{\infty} \sum_{x \in S_i} \mu(x) = \sum_{x \in A} \mu(x) = 1.$$

Здесь мы пользуемся тем, что множества $S_i, i = 0, 1, \dots$ задают разбиение множества $A$. □

Все готово для того, чтобы доказать теорему 15. Предположим Алиса получает на вход $x \in \mathcal{X}$, а Боб — $y \in \mathcal{Y}$. Они запускают протокол $\tau^{\mathcal{X}}_{r,\varepsilon}$, на входе $(x, \mu)$, где $\mu$ — условное распределение $X|Y = y$. Согласно лемме 14 полученный протокол будет решать задачу Вольфа — Слепяна для $(X, Y)$ с ошибкой $\varepsilon$. С другой стороны, в полученном протоколе на входе $(x, y)$ будет передаваться (согласно лемме 13) не более:

$$\log_2\left(\frac{1}{\mu(x)}\right) + r + \log_2\left(\frac{1}{\varepsilon}\right) + 1 + \frac{\log_2\left(\frac{1}{\mu(x)}\right)}{r} + 1$$
$$= \log_2\left(\frac{1}{\Pr[X = x|Y = y]}\right) \cdot \left(1 + \frac{1}{r}\right) + \log_2\left(\frac{1}{\varepsilon}\right) + 2$$

битов. Среднее этой величины по распределению пары $(X, Y)$ как раз равно (см. (1.1)):

$$H(X|Y) + \frac{H(X|Y)}{r} + \log_2\left(\frac{1}{\varepsilon}\right) + 2.$$

Остается оценить среднее число раундов. Очевидно из определения протокола $\tau^{A}_{r,\varepsilon}$, что оно не превосходит удвоенного среднего числа битов, передаваемых от Бобу к Алисе, то есть согласно лемме 13 не превосходит среднего величины:

$$2 \cdot \left(\frac{\log_2\left(\frac{1}{\Pr[X=x|Y=y]}\right)}{r} + 1\right).$$

Это среднее по распределению $(X, Y)$ как раз равно:

$$\frac{2H(X|Y)}{r} + 2.$$

## 5.3 Нижняя оценка

Перепишем нижнюю оценку следующим образом.

$$(1 - \gamma - \gamma/n)\log_2\left(\frac{\gamma}{\varepsilon + \gamma/n}\right) + (\gamma - 2\varepsilon)\log_2(n + 1)$$
$$= \gamma \log_2(n + 1) + (1 - \gamma)\log_2\left(\frac{1}{\varepsilon}\right)$$
$$- \left((1 - \gamma)\log_2\left(\frac{\varepsilon + \gamma/n}{\gamma\varepsilon}\right) + \frac{\gamma}{n}\log_2\left(\frac{\gamma}{\varepsilon + \gamma/n}\right) + 2\varepsilon \log_2(n + 1)\right)$$

При $\frac{1}{n} \le \varepsilon \le \frac{1}{\log_2(n)}$ и $\gamma = \Omega(1)$ выражение в скобках в последней строчке можно оценить сверху так:

$$(1 - \gamma)\log_2\left(\frac{\varepsilon + \gamma\varepsilon}{\gamma\varepsilon}\right) + \frac{\gamma}{n}\log_2\left(\frac{1}{\varepsilon}\right) + 2\varepsilon \log_2(n + 1)$$
$$\le (1 - \gamma)\log_2(2/\gamma) + \frac{\gamma}{n}\log_2(n) + \frac{2\log_2(n + 1)}{\log_2(n)} = O(1).$$

Иными словами, при $\frac{1}{n} \le \varepsilon \le \frac{1}{\log_2(n)}$ и $\gamma = \Omega(1)$ нижняя оценка теоремы 16 приобретает вид:

$$\gamma \log_2(n) + (1 - \gamma)\log_2\left(\frac{1}{\varepsilon}\right) - O(1).$$

Благодаря (5.3) это выражение равно $H(X_n^\gamma|Y_n^\gamma) + (1-\gamma)\log_2\left(\frac{1}{\varepsilon}\right) - O(1)$.

Перейдем непосредственно к доказательству оценки:

$$ACC_\nu(\pi) \geq (1-\gamma-\gamma/n)\log_2\left(\frac{\gamma}{\varepsilon+\gamma/n}\right) + (\gamma-2\varepsilon)\log_2(n+1), \tag{5.7}$$

где $\nu$ —- распределение пары $(X_n^\gamma, Y_n^\gamma)$, а $\pi$ — протокол, решающий задачу Вольфа — Слепяна для $(X_n^\gamma, Y_n^\gamma)$ с ошибкой $\varepsilon$. Сначала, рассмотрим случай, когда $\pi$ — детерминированный. Через $\pi(x,y)$ для $x,y \in \{0,1,\ldots,n\}$ обозначим лист протокола $\pi$, в который он приходит на входе $(x,y)$. По определению:

$$ACC_\nu(\pi) = \mathbb{E}|\pi(X_n^\gamma, Y_n^\gamma)|.$$

С другой стороны, поскольку множество значений случайной величины $\pi(X_n^\gamma, Y_n^\gamma)$ является конечным префиксным множеством, то согласно предложению 16 выполнена оценка:

$$ACC_\nu(\pi) \geq H(\pi(X_n^\gamma, Y_n^\gamma)).$$

Поэтому нам достаточно доказать (5.7) не для $ACC_\nu(\pi)$, а для $H(\pi(X_n^\gamma, Y_n^\gamma))$.

Пусть в протоколе $\pi$ есть $S$ различных листьев и множество листьев протокола $\pi$ есть множество $\{l_1,\ldots,l_S\}$. Согласно предложению 20 для любого $i \in \{1,\ldots,S\}$ множество пар, приходящих в лист $l_i$, есть «прямоугольник», что означает, что найдутся $A_i, B_i \subset \{0,1,\ldots,n\}$ такие, что:

$$A_i \times B_i = \left\{(x,y) \in \{0,1,\ldots,n\}^2 : \pi(x,y) = l_i\right\}.$$

Согласно определению, величина $\nu(x,y)$ для $x,y \in \{0,1,\ldots,n\}^2$ зависит от того, равны $x$ и $y$ или нет:

$$\nu(x,y) = \begin{cases} \frac{1-\gamma}{n+1} & \text{если } x = y, \\ \frac{\gamma}{n(n+1)} & \text{если } x \neq y. \end{cases}$$

Обозначим через $d_i$ количество пар вида $(x,x)$ («диагональных» пар) в $A_i \times B_i$. Кроме того, определим $|A_i| = w_i, |B_i| = h_i$. Тогда вероятность того, что $\pi(X_n^\gamma, Y_n^\gamma)$ равно $l_i$, выразится так:

$$\Pr[\pi(X_n^\gamma, Y_n^\gamma) = l_i] = \sum_{(x,y)\in A_i\times B_i} \nu(x,y) = \frac{(1-\gamma)d_i}{n+1} + \frac{\gamma(w_ih_i - d_i)}{n(n+1)}. \tag{5.8}$$

Рассмотрим следующие распределения вероятностей на множестве $\{l_1,\ldots,l_S\}$:

$$\xi(l_i) = \frac{d_i}{n+1}, \qquad \eta(l_i) = \frac{w_ih_i}{(n+1)^2},$$

где $i \in \{1,\ldots,S\}$. Во-первых, убедимся, что эти формулы действительно задают распределения вероятностей. Заметим, что $w_ih_i$ — это количество пар, приходящих в $l_i$, а $d_i$ — это сколько из этих пар диагональных. Всего пар $(n+1)^2$, а диагональных пар — $(n+1)$. Поэтому:

$$\sum_{i=1}^S d_i = n+1, \qquad \sum_{i=1}^S w_ih_i = (n+1)^2,$$

что и означает, что $\xi, \eta$ задают вероятностные распределения.

Согласно (5.8) распределение случайной величины $\pi(X_n^\gamma, Y_n^\gamma)$ является «взвешенной суммой» распределений $\xi$ и $\eta$, точнее:

$$\Pr[\pi(X_n^\gamma, Y_n^\gamma) = l_i] = (1-\gamma-\gamma/n)\xi(l_i) + (\gamma+\gamma/n)\eta(l_i).$$

А тогда в силу выпуклости энтропии (см. предложение 5) имеем оценку:

$$\begin{aligned} H(\pi(X_n^\gamma, Y_n^\gamma)) &\geq (1-\gamma-\gamma/n)H(\xi) + (\gamma+\gamma/n)H(\eta) \\ &\geq (1-\gamma-\gamma/n)H(\xi) + \gamma H(\eta). \end{aligned}$$

Мы оценим $H(\xi)$ и $H(\eta)$ по отдельности и докажем:

$$H(\xi) \geq \log_2 \left( \frac{\gamma}{\varepsilon + \gamma/n} \right) \tag{5.9}$$

$$H(\eta) \geq (1 - 2\varepsilon/\gamma) \log_2(n+1). \tag{5.10}$$

Очевидно из (5.9) и (5.10) вытекает (5.7) для детерминированных протоколов.

**Доказательство (5.9).** Поскольку детерминированный протокол $\pi$ решает задачу Вольфа — Слепяна для $(X_n^\gamma, Y_n^\gamma)$ с ошибкой $\varepsilon$, то существует функция $\theta : \{l_1, \ldots, l_S\} \times \mathcal{Y} \to \mathcal{X}$ такая, что

$$\Pr[\theta(\pi(X_n^\gamma, Y_n^\gamma), Y_n^\gamma) \neq X_n^\gamma] \leq \varepsilon.$$

Оценим снизу количество пар $(x, y) \in A_i \times B_i$ таких, что $\theta(\pi(x, y), y) \neq x$. Поскольку для таких $(x, y)$ выполнено $\pi(x, y) = l_i$, то достаточно оценить количество пар $(x, y) \in A_i \times B_i$ таких, что $\theta(l_i, y) \neq x$. Количество пар вида $(x, x)$ из $A_i \times B_i$ равно $d_i$, и оно же равно $|A_i \cap B_i|$. С другой стороны, для каждого $y \in A_i \cap B_i$ есть максимум один $x \in A_i \cap B_i$ такой, что $\theta(l_i, y) = x$. Иными словами, если ограничиться парами из $(A_i \cap B_i) \times (A_i \cap B_i)$, то среди них будет хотя бы $d_i^2 - d_i$ пар $(x, y)$ таких, что $\theta(l_i, y) \neq x$, а все такие пары лежат в $A_i \times B_i$. Поэтому

$$\begin{aligned} \varepsilon \geq \Pr[\theta(\pi(X_n^\gamma, Y_n^\gamma), Y_n^\gamma) \neq X_n^\gamma] &= \sum_{\substack{(x,y)\in\{0,1,\ldots,n\}^2 \\ \theta(\pi(x,y),y)\neq y}} \nu((x,y)) \\ = \sum_{i=1}^{S} \sum_{\substack{(x,y)\in\{0,1,\ldots,n\}^2, \\ \pi(x,y)=l_i, \\ \theta(\pi(x,y),y)\neq y}} \nu((x,y)) &= \sum_{i=1}^{S} \sum_{\substack{(x,y)\in A_i\times B_i, \\ \theta(\pi(x,y),y)\neq y}} \nu((x,y)) \\ \geq \sum_{i=1}^{S} (d_i^2 - d_i) \cdot \frac{\gamma}{n(n+1)} &= -\frac{\gamma}{n(n+1)} \sum_{i=1}^{S} d_i + \gamma \sum_{i=1}^{S} \frac{d_i^2}{n(n+1)} \\ \geq -\frac{\gamma}{n} + \gamma \sum_{i=1}^{S} \left( \frac{d_i}{n+1} \right)^2, & \end{aligned}$$

откуда вытекает:

$$\sum_{i=1}^{S} \left( \frac{d_i}{n+1} \right)^2 \leq \frac{\varepsilon + \gamma/n}{\gamma}.$$

Поскольку энтропия Шеннона не меньше энтропии различения (предложение 3), то мы получаем требуемую оценку:

$$H(\xi) \geq H_2(\xi) = \log_2 \left( \frac{1}{\sum\limits_{i=1}^{S} \left( \frac{d_i}{n+1} \right)^2} \right) \geq \log_2 \left( \frac{\gamma}{\varepsilon + \gamma/n} \right).$$

**Доказательство (5.10).** Количество пар $(x, y) \in A_i \times B_i$, для которых выполнено $\theta(l_i, y) = x$, не превосходит $|B_i| = h_i$ (для каждого $y \in B_i$ есть максимум один соответствующий $x$).

Из этих пар не больше $d_i$ имеют $\nu$-вероятность, равную $\frac{1-\gamma}{n+1}$ (поскольку всего столько пар в $A_i \times B_i$ имеют такую $\nu$-вероятность, а другие пары имеют меньшую $\nu$-вероятность, а именно $\gamma/(n(n+1))$). Поэтому

$$\Pr[\pi(X_n^\gamma, Y_n^\gamma) = l_i, \theta(l_i, Y_n^\gamma) = X_n^\gamma] \le \frac{(1-\gamma)d_i}{n+1} + \frac{\gamma(h_i - d_i)}{n(n+1)}.$$

Отсюда следует, что

$$\begin{aligned} 1 - \varepsilon &\le \Pr[\theta(\pi(X_n^\gamma, Y_n^\gamma), Y_n^\gamma) = X_n^\gamma] \\ &= \sum_{i=1}^{S} \Pr[\pi(X_n^\gamma, Y_n^\gamma) = l_i, \theta(l_i, Y_n^\gamma) = X_n^\gamma] \\ &\le \sum_{i=1}^{S} \left( \frac{(1-\gamma)d_i}{n+1} + \frac{\gamma(h_i - d_i)}{n(n+1)} \right) = 1 - \gamma - \gamma/n + \frac{\gamma}{n(n+1)} \sum_{i=1}^{S} h_i. \end{aligned}$$

Переписывая последнее неравенство, получаем:

$$\begin{aligned} \sum_{i=1}^{S} h_i &\ge \frac{(\gamma + \gamma/n - \varepsilon) \cdot n(n+1)}{\gamma} = \frac{(\gamma(n+1) - \varepsilon n) \cdot (n+1)}{\gamma} \\ &\ge \frac{(\gamma(n+1) - \varepsilon(n+1)) \cdot (n+1)}{\gamma} = (1 - \varepsilon/\gamma)(n+1)^2. \end{aligned}$$

Используя то, что $h_i \le n+1$ для всех $i$, можно оценить $H(\eta)$ снизу так:

$$\begin{aligned} H(\eta) &= \sum_{i=1}^{S} \frac{w_i h_i}{(n+1)^2} \log_2 \left( \frac{(n+1)^2}{w_i h_i} \right) \\ &\ge \sum_{i=1}^{S} \frac{w_i h_i}{(n+1)^2} \log_2 \left( \frac{(n+1)^2}{(n+1) w_i} \right) \\ &= -\log(n+1) + \sum_{i=1}^{S} h_i \frac{w_i}{(n+1)^2} \log_2 \left( \frac{(n+1)^2}{w_i} \right). \end{aligned}$$

Чтобы получить нужную оценку, нам потребуется доказать следующее неравенство:

$$\frac{w_i}{(n+1)^2} \log_2 \left( \frac{(n+1)^2}{w_i} \right) \ge \frac{\log_2((n+1)^2)}{(n+1)^2}. \tag{5.11}$$

Если его доказать, то оценка на $H(\eta)$ примет нужный нам вид:

$$\begin{aligned} H(\eta) &\ge -\log_2(n+1) + \left( \sum_{i=1}^{S} h_i \right) \cdot \frac{\log_2((n+1)^2)}{(n+1)^2} \\ &\ge -\log_2(n+1) + (2 - 2\varepsilon/\gamma) \log_2(n+1) = (1 - 2\varepsilon/\gamma) \log_2(n+1), \end{aligned}$$

где во втором неравенстве мы воспользовались ранее полученной оценкой на $\sum_{i=1}^{S} h_i$.

Еще недоказанное нами неравенство (5.11) вытекает из такого наблюдения:

**Лемма 15.** *Функция $f(x) = x \log_2(1/x)$ возрастает на $(0, 1/e)$.*

*Доказательство.* Достаточно взять производную функции $f$:

$$f'(x) = \left(\frac{-x\ln(x)}{\ln(2)}\right)' = \frac{-1-\ln(x)}{\ln(2)} = \frac{-\ln(ex)}{\ln(2)}.$$

□

Действительно, (5.11) эквивалентно следующему неравенству:

$$f\left(\frac{1}{(n+1)^2}\right) \le f\left(\frac{w_i}{(n+1)^2}\right),$$

что верно благодаря лемме, поскольку:

$$\frac{1}{(n+1)^2} \le \frac{w_i}{(n+1)^2} \le \frac{1}{n+1} \le 1/e,$$

где в последнем неравенстве мы воспользовались ограничением $n \ge 2$.

Итак, теорема доказана для детерминированных протоколов. Теперь обобщим ее на протоколы с общими случайными битами. Пусть $\pi$ — протокол с общими случайными битами с пространством общих случайных битов $(\mathcal{R}, \mu)$, решающий задачу Вольфа — Слепяна для $(X_n^\gamma, Y_n^\gamma)$ с ошибкой $\varepsilon$. По определению это означает, что для любого $r \in \mathcal{R}$ существует $\varepsilon_r \in [0,1]$ такое, что детерминированный протокол $\pi_r$ решает задачу Вольфа — Слепяна для $(X_n^\gamma, Y_n^\gamma)$ с ошибкой $\varepsilon_r$ и:

$$\mathbb{E}_{r\sim\mu}\varepsilon_r \le \varepsilon.$$

С другой стороны, по определению средней длины протокола с общими случайными битами:

$$ACC_\nu(\pi) = \mathbb{E}_{r\sim\mu} ACC_\nu(\pi_r).$$

Воспользовавшись уже доказанной оценкой на $ACC_\nu(\pi_r)$ (благо протоколы $\pi_r$ детерминированные), а также выпуклостью логарифма, получаем:

$$\begin{aligned} ACC_\nu(\pi) &\ge \mathbb{E}_{r\sim\mu}\left[(1-\gamma-\gamma/n)\log_2\left(\frac{\gamma}{\varepsilon_r+\gamma/n}\right) + (\gamma-2\varepsilon_r)\log_2(n+1)\right] \\ &\ge (1-\gamma-\gamma/n)\log_2\left(\frac{\gamma}{\varepsilon+\gamma/n}\right) + (\gamma-2\varepsilon)\log_2(n+1), \end{aligned}$$

что и требовалось.

# Заключение

Работа содержит четыре основных результата, один совместный и три полученных без соавторов.

Первый результат (полученный совместно с Е. Клениным) относится к коммуникационной сложности задачи Gap Hamming Distance с односторонней ошибкой. А именно, доказывается нижняя оценка $\Omega(L^2/U+1)$ и верхняя оценка $O((L^2/U+1)\log L)$ на сложность функции $\mathrm{GHD}(n, L, U)$ с односторонней ошибкой (когда нельзя ошибаться на входах на расстоянии не больше $L$). Таким образом указанная величина вычислена с точностью до множителя порядка $O(\log L)$. Показывается, что верхняя оценка выполнена также для SMP-протоколов (в которых Алиса и Боб не общаются между собой, а посылают по сообщению третьему игроку, Чарли, который, не видя входов Алисы и Боба, вычисляет ответ).

Второй результат связан с теоремой Раза — Маккинзи. Во-первых, предложен способ, как из экспандеров получать гаджеты с хорошими протыкающими распределениями. Из экспандера, основанного на аффинной плоскости над конечным полем, получено новое семейство гаджетов, для которого верен аналог теоремы Раза — Маккинзи. При этом взаимоотношение между арностью внешней функции и длиной входа гаджета у этого семейства такое же, как у гаджета из работы [16] (у других гаджетов, изучавшихся в литературе, оно хуже). Кроме того, получен ряд результатов, показывающих, что улучшить это взаимоотношение с текущей техникой нельзя. А именно, показано, что не существует гаджетов с лучшими, чем у нашего нового семейства гаджетов, протыкающими распределениями. Также доказана неулучшаемость леммы о толщине (Thickness Lemma) — технического утверждения из доказательства теоремы Раза — Маккинзи. Наконец, мы показываем, что наша конструкция протыкающих распределений явная в том смысле, что носитель протыкающего распределения может быть выписан за полиномиальное от размера матрицы гаджета время (для упомянутого выше гаджета из работы [16] это неизвестно). Интерес к явным протыкающим распределениям мотивируется как шаг к эффективному варианту теоремы Раза — Маккинзи.

Третий результат исследует роль частных случайных битов в информационной сложности. Доказывается, что любой протокол с частными случайными битами можно безошибочно моделировать протоколом, не использующим частных случайных битов, так, что информационное разглашение моделирующего протокола не превосходит $O(\sqrt{Id})$, где $I$ и $d$ — информационное разглашение и коммуникационная длина исходного протокола. Под безошибочным моделированием мы понимаем следующее: на любой паре входов коммуникация у исходного и у моделирующего протокола имеет одно и то же распределение. Из этого результата можно вывести новое доказательство сжатия из работы [2].

Четвертый результат относится к интерактивному аналогу теоремы Вольфа — Слепяна. Показано, что для любого натурального $r$ и любого $\varepsilon$ задачу Вольфа — Слепяна для пары случайных величин $X, Y$ можно решить с вероятностью ошибки $\varepsilon$ при помощи протокола со средней длиной не более $(1+1/r)H+r+O(\log(1/\varepsilon))$ и со средним количеством раундов не более $2H/r+2$, где $H = H(X|Y)$ — условная энтропия $X$ при известном $Y$.. Отсюда для любого $\alpha \in [1/2, 1]$, подобрав нужное $r$, легко получить протокол со средней длиной $H + O(H^\alpha)$ и со средним числом раундов $O(H^{1-\alpha})$. Ранее такая оценка была известна лишь для $\alpha = 1/2$ [7] и для $\alpha = 1$ [9, 8]. Кроме того, если минимизировать лишь среднюю длину и забыть о числе раундов, то, положив, $r = \sqrt{H}$, мы получаем оценку $H + 2\sqrt{H} + O(\log(1/\varepsilon))$, что немного

улучшает оценку $H + 5\sqrt{H} + O(\log(1/\varepsilon))$ из работы [7] . Кроме того, в этой главе мы исследуем вопрос о возможности дальнейших улучшений этих оценок. Мы показываем, грубо говоря, что в общем случае от члена порядка $O(\log(1/\varepsilon))$ избавиться нельзя. Более точно, для любого натурального $n$ мы приводим пример случайных величин $X$ и $Y$, принимающих $n$ возможных значений, для которых выполнена следующая нижняя оценка. Любой протокол, решающий для них задачу Вольфа — Слепяна с ошибкой $\varepsilon$ при $\log_2 n \leq 1/\varepsilon \leq n$, передает в среднем не меньше $H(X|Y) + \Omega(\log(1/\varepsilon))$ бит. Константа в $\Omega(\cdot)$ может быть сделана сколь угодно близкой к единице.

# Литература


[1] Bar-Yossef, Z., Jayram, T. S., Kumar, R., and Sivakumar, D. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences 68*, 4 (2004), 702–732.

[2] Barak, B., Braverman, M., Chen, X., and Rao, A. How to compress interactive communication. *SIAM Journal on Computing 42*, 3 (2013), 1327–1363.

[3] Bauer, B., Moran, S., and Yehudayoff, A. Internal compression of protocols to entropy. In *LIPIcs-Leibniz International Proceedings in Informatics* (2015), vol. 40, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

[4] Blais, E., Brody, J., and Matulef, K. Property testing lower bounds via communication complexity. *Computational Complexity 21*, 2 (2012), 311–358.

[5] Braverman, M. Interactive information complexity. *SIAM Journal on Computing 44*, 6 (2015), 1698–1739.

[6] Braverman, M., and Garg, A. Public vs private coin in bounded-round information. In *International Colloquium on Automata, Languages, and Programming* (2014), Springer, pp. 502–513.

[7] Braverman, M., and Rao, A. Information equals amortized communication. *IEEE Transactions on Information Theory 60*, 10 (2014), 6058–6069.

[8] Braverman, M., Rao, A., Weinstein, O., and Yehudayoff, A. Direct product via round-preserving compression. In *International Colloquium on Automata, Languages, and Programming* (2013), Springer, pp. 232–243.

[9] Brody, J., Buhrman, H., Koucký, M., Loff, B., Speelman, F., and Vereshchagin, N. Towards a reverse newman's theorem in interactive information complexity. *Algorithmica 76*, 3 (2016), 749–781.

[10] Brody, J., and Chakrabarti, A. A multi-round communication lower bound for gap hamming and some consequences. In *Computational Complexity, 2009. CCC'09. 24th Annual IEEE Conference on* (2009), IEEE, pp. 358–368.

[11] Brody, J., Chakrabarti, A., Regev, O., Vidick, T., and De Wolf, R. Better gap-hamming lower bounds via better round elimination. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, 2010, pp. 476–489.

[12] Brody, J., and Woodruff, D. P. Streaming algorithms with one-sided estimation. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, 2011, pp. 436–447.

[13] Buhrman, H., Cleve, R., and Wigderson, A. Quantum vs. classical communication and computation. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing* (1998), ACM, pp. 63–68.

[14] Buhrman, H., and De Wolf, R. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science 288*, 1 (2002), 21–43.

[15] Chakrabarti, A., and Regev, O. An optimal lower bound on the communication complexity of gap-hamming-distance. *SIAM Journal on Computing 41*, 5 (2012), 1299–1317.

[16] Chattopadhyay, A., Kouckỳ, M., Loff, B., and Mukhopadhyay, S. Simulation theorems via pseudorandom properties. *arXiv preprint arXiv:1704.06807* (2017).

[17] Chattopadhyay, A., Kouckỳ, M., Loff, B., and Mukhopadhyay, S. Simulation beats richness: new data-structure lower bounds. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing* (2018), ACM, pp. 1013–1020.

[18] Cohen, G., Honkala, I., Litsyn, S., and Lobstein, A. *Covering codes*, vol. 54. Elsevier, 1997.

[19] Cover, T. M., and Thomas, J. A. *Elements of information theory*. John Wiley & Sons, 2012.

[20] de Rezende, S. F., Nordström, J., and Vinyals, M. How limited interaction hinders real communication (and what it means for proof and circuit complexity). In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)* (2016), IEEE, pp. 295–304.

[21] Garg, A., Göös, M., Kamath, P., and Sokolov, D. Monotone circuit lower bounds from resolution. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing* (2018), ACM, pp. 902–911.

[22] Gavinsky, D., Kempe, J., and de Wolf, R. Quantum communication cannot simulate a public coin. *arXiv preprint quant-ph/0411051* (2004).

[23] Göös, M., Kamath, P., Pitassi, T., and Watson, T. Query-to-communication lifting for p np. In *Proceedings of the 32nd Computational Complexity Conference* (2017), Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, p. 12.

[24] Göös, M., Pitassi, T., and Watson, T. Deterministic communication vs. partition number. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on* (2015), IEEE, pp. 1077–1088.

[25] Göös, M., Pitassi, T., and Watson, T. Query-to-communication lifting for bpp. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)* (2017), IEEE, pp. 132–143.

[26] Huang, W., Shi, Y., Zhang, S., and Zhu, Y. The communication complexity of the hamming distance problem. *Information Processing Letters 99*, 4 (2006), 149–153.

[27] Jayram, T. S., Kumar, R., and Sivakumar, D. The one-way communication complexity of hamming distance. *Theory of Computing 4*, 1 (2008), 129–135.

[28] Karchmer, M., and Wigderson, A. Monotone circuits for connectivity require super-logarithmic depth. *SIAM Journal on Discrete Mathematics 3*, 2 (1990), 255–265.

[29] Kushilevitz, E., and Nisan, N. *Communication Complexity*. Cambridge University Press, 2006.

[30] Kushilevitz, E., Ostrovsky, R., and Rabani, Y. Efficient search for approximate nearest neighbor in high dimensional spaces. *SIAM Journal on Computing 30*, 2 (2000), 457–474.

[31] Lubotzky, A., Phillips, R., and Sarnak, P. Ramanujan graphs. *Combinatorica 8*, 3 (1988), 261–277.

[32] Pankratov, D. Direct sum questions in classical communication complexity. *Master's thesis, University of Chicago* (2012).

[33] Raz, R., and McKenzie, P. Separation of the monotone nc hierarchy. In *Foundations of Computer Science, 1997. Proceedings., 38th Annual Symposium on* (1997), IEEE, pp. 234–243.

[34] Reingold, O., Vadhan, S., and Wigderson, A. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Annals of mathematics* (2002), 157–187.

[35] Roughgarden, T. Communication complexity (for algorithm designers). *Foundations and Trends® in Theoretical Computer Science 11*, 3–4 (2016), 217–404.

[36] Sherstov, A. A. The communication complexity of gap hamming distance. *Theory of Computing 8*, 1 (2012), 197–208.

[37] Slepian, D., and Wolf, J. Noiseless coding of correlated information sources. *IEEE Transactions on information Theory 19*, 4 (1973), 471–480.

[38] Vadhan, S. P. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science 7*, 1–3 (2012), 1–336.

[39] Vidick, T. A concentration inequality for the overlap of a vector on a large set. *Chicago Journal of Theoretical Computer Science 1* (2012), 1–12.

[40] Weinstein, O. Information complexity and the quest for interactive compression. *ACM SIGACT News 46*, 2 (2015), 41–64.

[41] Woodruff, D. Optimal space lower bounds for all frequency moments. In *Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms* (2004), Society for Industrial and Applied Mathematics, pp. 167–175.

[42] Wu, X., Yao, P., and Yuen, H. S. Raz-mckenzie simulation with the inner product gadget. In *Electronic Colloquium on Computational Complexity (ECCC)* (2017), vol. 24.

[43] Yao, A. C.-C. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the eleventh annual ACM symposium on Theory of computing* (1979), ACM, pp. 209–213.

[44] Yao, A. C.-C. On the power of quantum fingerprinting. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing* (2003), ACM, pp. 77–81.

[45] Yeung, R. W. *Information theory and network coding*. Springer Science & Business Media, 2008.

# Список публикаций автора по теме диссертации

## Статьи в рецензируемых научных изданиях, рекомендуемых для защиты в диссертационном совете МГУ по специальности

[46] Kozachinskiy, A. Making randomness public in unbounded-round information complexity // Proceedings of the 10th International Computer Science Symposium in Russia. – 2015. – Lecture Note in Computer Science, Springer, V. 9139. – P. 296-309.
DOI: 10.1007/978-3-319-20297-6_19. Импакт-фактор: Scopus — 1,174.

[47] Kozachinskiy, A. On Slepian-Wolf theorem with interaction // Theory of Computing Systems. – 2018. – V. 62, №3. – P. 583-599. DOI: 10.1007/s00224-016-9741-x. Импакт-фактор: Scopus — 0,950; Web of Science — 0.604.

[48] Klenin, E., and Kozachinskiy, A. One-sided error communication complexity of Gap Hamming Distance // Proceedings of the 43rd International Symposium on Mathematical Foundations of Computer Science. – 2018. – Leibniz International Proceedings in Informatics, V. 117. – P. 7:1-7:15. DOI: 10.4230/LIPIcs.MFCS.2018.7. Импакт-фактор: Scopus — 0,940.

*Автору принадлежит доказательство следствия 4 (нижняя оценка). В доказательстве теоремы 2 (верхняя оценка) Е. Кленину принадлежит конструкция HT-протокола, а автору принадлежит анализ HT-протокола (Лемма 15), а также следующие за HT-протоколом шаги доказальства.*

[49] Kozachinskiy, A. From expanders to hitting distributions and simulation theorems // Proceedings of the 43rd International Symposium on Mathematical Foundations of Computer Science. – 2018. – Leibniz International Proceedings in Informatics, V. 117. – P. 4:1-4:15.
DOI:10.4230/LIPIcs.MFCS.2018.4. Импакт-фактор: Scopus — 0,940.