ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ УНИТАРНОЕ
ПРЕДПРИЯТИЕ «НИИ «КВАНТ»

На правах рукописи
УДК 510.5

Сергеев Игорь Сергеевич

НЕКОТОРЫЕ ВОПРОСЫ СИНТЕЗА ПАРАЛЛЕЛЬНЫХ СХЕМ

01.01.06 — математическая логика, алгебра и теория чисел

ДИССЕРТАЦИЯ

на соискание учёной степени
доктора физико-математических наук

Научный консультант:
доктор физико-математических наук,
профессор С. Б. Гашков

МОСКВА — 2021

Содержание

Посвящается памяти
моего школьного учителя математики,
замечательного педагога и наставника
Михаила Борисовича Анохина (1948–2019).

1 Общее введение

Актуальность и разработанность темы. Задачи синтеза схем, удовлетворяющих критериям эффективности, учитывающим глубину, в теории булевых функций стали рассматриваться ненамного позже, чем задачи оптимизации числа элементов схем (сложности). Еще в 1956 г. О. Б. Лупанов [41] поставил и решил задачу оптимального синтеза вентильных схем глубины 2 — за несколько лет до получения своих основных результатов об асимптотически оптимальном синтезе в различных моделях вычислений.

Также, к 1960-м годам вопросы синтеза быстрых схем для арифметики стали активно изучаться, исходя из потребностей электроники. Сама теория быстрых вычислений, как принято считать, ведет отсчет с опубликованной по инициативе А. Н. Колмогорова работы [19] о быстром умножении чисел, в которой наряду с рекордным по сложности методом А. А. Карацубы представлен метод параллельного умножения Ю. П. Офмана.

Ряд фундаментальных асимптотических проблем теории синтеза параллельных схем был решен О. Б. Лупановым: в частности, определена асимптотика функции Шеннона глубины схем над произвольным булевым базисом, при этом решена задача об одновременной минимизации глубины и сложности [44], получена асимптотика функции Шеннона сложности формул ограниченной глубины (альтернирования) [42, 46] и решена аналогичная задача для схем [47]. Вопросы сложности вентильных схем ограниченной глубины всесторонне изучены Э. И. Нечипоруком, см. [57]. Исследование асимптотических вопросов было продолжено в Московском университете учениками О. Б. Лупанова. Так, С. Б. Гашков установил величину функции Шеннона глубины булевых функций в стандартном базисе с точностью до аддитивной постоянной [5] и аналогичный результат получил для многочленов (включая важный случай функций k-значной

логики) [7]. С. А. Ложкиным получена асимптотика функции Шеннона глубины для полных базисов с нулевыми весами элементов [34] и серия асимптотических оценок высокой точности в различных моделях параллельных вычислений [35, 36, 37, 38, 40, 39], в частности, наиболее точные оценки функции Шеннона глубины схем для ряда базисов, включая стандартный и монотонный. А. Е. Андреев (ученик В. Б. Кудрявцева), расширяя результат Нечипорука, установил асимптотику сложности реализации классов недоопределенных матриц вентильными схемами глубины 2 [3]. А. Б. Угольников получил описание порядков функции Шеннона глубины для всех конечных систем булевых функций [76]. О. М. Касим-Заде установил значение функции Шеннона глубины произвольного бесконечного полного булева базиса с точностью до аддитивной постоянной [22, 23]. Его ученик А. В. Кочергин получил асимптотику функции Шеннона глубины для любого полного базиса k-значной логики [29].

С 1980-х гг. активно развивается направление получения нижних оценок сложности индивидуальных функций при ограничении на глубину вычисления. Число работ в этой области измеряется сотнями. Дело в том, что ограничение на глубину позволяет доказывать в традиционных моделях вычислений (схемы, формулы над полными базисами) сверхполиномиальные и даже экспоненциальные нижние оценки. Первый результат такого рода был получен еще одним учеником Лупанова Г. А. Ткачевым [73]. Далее, в работах М. Фёрста, Дж. Сакса, М. Сипсера [131], Э. Яо [213] и Й. Хостада [140] были заложены основы теории. Фундаментальный вклад в теорию нижних оценок сложности схем ограниченной глубины внесли работы А. А. Разборова, также представителя школы Московского университета (схемы в базисе $\{\vee, \wedge, \oplus\}$ [63], схемы из пороговых элементов [189], арифметические схемы [135]).

Развитие методов синтеза параллельных схем для конкретных функций или классов функций, а также исследование пределов возможностей этих методов стимулируется приложениями, прежде всего, микроэлектроникой. Широко известные результаты в области синтеза параллельных схем принадлежат еще одному математику из школы Лупанова В. М. Храпченко: конструкция асимптотически минимального по глубине сумматора [79],

эффективные параллельные схемы для оператора умножения и симметрических функций [83], метод параллельного перестроения булевых формул [94, 84]. К числу важнейших результатов в указанной области следует отнести метод синтеза параллельных префиксных схем Р. Ладнера и М. Фишера [156], параллельные схемы сортировки М. Айтаи, Я. Комлоша и Э. Семереди [96], параллельные схемы для деления, возведения в степень и других целочисленных операций П. Бима, С. Кука и Дж. Гувера [100]. Для построения быстрых методов умножения в различных алгебраических структурах сегодня широко применяются идеально распараллеливаемые алгоритмы быстрого преобразования Фурье (БПФ), идея которого восходит к работам И. Гуда [133], Дж. Кули и Дж. Тьюки [118].

За прошедшие полвека существенно изменился понятийный аппарат теории: скажем, сегодня исследования глубины функций часто выполняются в терминах коммуникационной сложности подходящим образом определенных протоколов. Понятие коммуникационной сложности было введено Э. Яо [212]. Ее связь с параллельными мерами сложности (в том числе, с глубиной схем) была прояснена в работах А. Хайнала, В. Маасса, Г. Турана [138] и М. Карчмера, А. Вигдерсона [148]. Эта связь привела к появлению специальных методов нижних оценок глубины конкретных функций. Прежде такие оценки, как правило, извлекались лишь в качестве следствий из известных нижних оценок сложности формул или схем. Наиболее яркие результаты получены А. Вигдерсоном с М. Карчмером [148] (нижняя оценка монотонной глубины функции проводимости) и Р. Разом [188] (нижние оценки монотонной глубины пермамента и кликовых функций).

Взаимоотношения между понятиями глубины и сложности варьируются в зависимости от вычислительной модели. Сложность функции в модели деревьев решений (решающих диаграмм) определяется глубиной реализующего ее дерева. В терминах построения деревьев решений формулируются многие задачи теории алгоритмов, например, задачи сортировки или поиска за минимальное число сравнений.

Цель настоящей работы — развитие методов синтеза параллельных схем; решение ряда задач, относящихся к классическим направлениям теории синтеза параллельных схем. Основные задачи: получение соотношений

между глубиной и сложностью булевых формул, синтез экономных формул для симметрических булевых функций, асимптотически оптимальный синтез вентильных схем ограниченной глубины, синтез минимальных параллельных префиксных схем, асимптотически оптимальный синтез схем и формул ограниченной глубины из многовходовых элементов, асимптотическая сложность сортировки. Постановки перечисленных задач восходят к 1950–70-м гг.

Объектом исследований служат схемы из функциональных элементов или формулы над булевыми базисами, а также деревья решений. Предмет исследования — эффективность параллельного вычисления функций.

Перейдем к краткой характеристике **содержания работы**. Подробные исторические справки и точные формулировки понятий по каждой задаче помещены во вводные разделы соответствующих глав.

В **главе 2** исследуются вопросы о соотношении между глубиной и сложностью булевых формул, о построении эффективных по глубине или сложности формул для симметрических булевых функций и о нижних оценках сложности формул в базисе k-местных функций.

В 1960-х гг. В. М. Храпченко [94] доказал, что функционально полные конечные булевы базисы являются равномерными, что значит: глубина и логарифм сложности формул для любой функции над такими базисами совпадают по порядку. Позднее А. Б. Угольников [75] и М. Рагаз [187] установили аналогичный факт для произвольных конечных булевых систем. Так возникла задача исследования констант равномерности — пределов отношений глубины и логарифма сложности функций в заданных базисах.

Начиная с 1970-х гг. было получено множество оценок констант равномерности для различных базисов. Основной интерес представляли арифметические базисы из операций сложения и умножения, а также деления, и булев аналог — монотонный базис из операций конъюнкции и дизъюнкции — в связи с практической задачей о параллельном перестроении арифметических выражений. Рекордные верхние оценки для этих базисов получены в [165, 83, 153]. Нетривиальные нижние оценки констант равномерности были известны только для монотонных арифметических базисов (из работ [196, 120]) до тех пор, пока автору [225] не удалось получить анало-

гичный результат для булева монотонного базиса, см. раздел 2.2. Задача оставалась нерешенной на протяжении 50 лет.

Разделы 2.3–2.5 посвящены симметрическим функциям. Симметрические функции — это функции, значения которых не изменяются при перестановке аргументов. В булевом случае значение симметрической функции определяется арифметической суммой переменных.

Вопросы эффективной реализации симметрических функций в различных вычислительных моделях всегда находились в фокусе внимания теории сложности. Для приложений, практических и теоретических, представляют интерес методы синтеза как конкретных симметрических функций, так и подклассов (пороговые, периодические функции), реже симметрических функций вообще. Одно из самых известных приложений — параллельные схемы умножения чисел, основанные на эффективном вычислении арифметической суммы битов. Известные конструкции параллельных схем для деления и некоторых других арифметических операций с числами или элементами конечных полей (см., например, [100, 69, 11]) также опираются на быстрые схемы для симметрических функций.

В настоящей работе мы ограничиваем рассмотрение вычислительной моделью формул над полными конечными базисами, интересуясь прежде всего бинарными базисами. Именно оптимизация глубины формулы представляет интерес в задаче о быстром умножении чисел.

Первые результаты о сложности, а затем и о глубине формул для симметрических булевых функций были получены В. М. Храпченко в начале 1970-х гг.: как верхние оценки [81, 83], так и нижние [80] (все — для стандартного базиса). Позже к ним были добавлены аналогичные результаты для полного бинарного базиса (нижняя оценка — в [129]). Серия уточнений верхних оценок продолжалась до начала 1990-х гг., когда усилиями М. Патерсона, Н. Пиппенджера, У. Цвика [169, 171] были определены принципы оптимального конструирования формул из заданных подформул-компрессоров, и тем самым обобщены ранее известные методы.

Автором показано [219, 220, 221], что для реализации симметрических функций можно эффективно использовать популярные (в теории быстрых алгоритмов) идеи использования нескольких взаимно простых модулей и

приближенных вычислений. Новый метод приводит к примерно на 10–20% лучшим верхним оценкам глубины и логарифма сложности формул, чем известные ранее (однако рекордные полученные оценки неконструктивны).

В разделе 2.5 отдельно рассмотрен вопрос о реализации периодических симметрических функций с малыми простыми числами в качестве периодов (MOD-функций). Эта задача связана, в том числе, и с упомянутым модулярным методом синтеза симметрических функций, использующим недвоичные системы счисления. Нетривиальные верхние оценки глубины и сложности формул для ряда периодических симметрических функций получены в работах [157, 113] до 1990 г. Автором в [222] получены новые оценки — часть из них при помощи предложенного общего приема сведения к задаче об оптимальном покрытии матриц прямоугольниками.

В разделе 2.6 изложен метод получения нижних оценок сложности формул в базисе U_k из работы автора [229]. Базис U_k является максимальным базисом k-местных функций, в котором сложность линейной булевой функции нелинейна. Предлагаемый метод служит расширением известного метода Храпченко [80] нижних оценок сложности формул в базисе B_0 (или U_2). С помощью этого метода автором усилены известные из работ [62, 89, 114] нижние оценки сложности линейной функций при всех $k \geq 3$. В определенном, хотя и довольно слабом смысле, полученные оценки являются точными.

Глава 3 посвящена линейным схемам и объединяет результаты из двух направлений. Линейная схема, в одной из интерпретаций — это схема из многовходовых элементов сложения в некоторой коммутативной полугруппе $(G, +)$. Схема реализует некоторое линейное преобразование, но принято говорить, что схема реализует саму матрицу данного преобразования. Классическая вентильная схема [41] — это линейная схема над булевой полугруппой $(\mathbb{B}, \vee)$, где $\mathbb{B} = \{0, 1\}$. Под сложностью линейной схемы понимается число ребер в ней.

В разделе 3.2 приводится решение задачи асимптотически оптимального синтеза линейных схем глубины 3 для класса $\mathbb{B}^{m\times n}$ булевых матриц размера $m \times n$. Оставим в стороне случай очень узких матриц, когда $m = O(\log n)$. Выше уже упоминалась работа О. Б. Лупанова [41] 1956 г.,

в которой получена асимптотика сложности класса $\mathbb{B}^{m\times n}$ при реализации вентильными схемами в случае достаточно узких матриц, $\log m = o(\log n)$. Причем использовались схемы глубины 2. Это один из первых результатов теории асимптотически оптимального синтеза. Чуть позже Э. И. Нечипорук [53] при помощи конструкции глубины 3 получил асимптотику сложности при определенных соотношениях между m и n, включая важный случай $m = n$. В конце 1970-х гг. Н. Пиппенджер [176] решил задачу при любых m и n, но использовал схемы растущей глубины. Отталкиваясь от конструкций Нечипорука и Пиппенджера, автор [223] показал, что асимптотика сложности на самом деле достигается на схемах глубины 3.

Раздел 3.3 посвящен вопросу об экстремальных отношениях сложности булевой матрицы при реализации линейными схемами разных видов. Рассматриваются схемы над $(\mathbb{B}, \vee)$ (вентильные схемы, OR-схемы), схемы над $(\mathbb{B}, \oplus)$ (вентильные схемы по модулю 2, XOR-схемы), схемы над $(\mathbb{Z}, +)$ (аддитивные схемы, SUM-схемы).

Задача об экстремальных отношениях ведет отсчет с работы Б. С. Митягина и Б. Н. Садовского [50], в которой поставлен и почти решен вопрос о максимуме отношения OR-сложности и XOR-сложности в классе булевых (n, n)-матриц без прямоугольников (сплошь единичных подматриц) размера 2×2. Однако активные исследования в области начались почти полвека спустя в результате совместных работ автора с С. Б. Гашковым [215] и М. И. Гринчуком [216]. Практически окончательное решение получила проблема [50], и заодно аналогичные проблемы для класса циркулянтных матриц и всех (n, n)-матриц. Построены как конструктивные, так и неконструктивные примеры.

Затем, в течение нескольких лет последовала серия работ [149, 230, 174, 107, 128] нескольких коллективов авторов, в которых помимо отношений сложности типа OR/XOR рассмотрены отношение типа SUM/OR, отношение OR-сложности матриц и их дополнений, в том числе, с ограничением на глубину схем. Промежуточные итоги развития теории подведены в соответствующей главе совместной с С. Юкной работы автора [230].

Автору принадлежит пример последовательности матриц с растущим отношением XOR- и OR-сложности в глубине 2 [230] и примеры матриц

с близким к максимальному возможному отношением OR-сложности к сложности дополнительной матрицы, в том числе, в ограниченной глубине [230, 238].

Часть перечисленных результатов доказывается с использованием конструкций экстремальных редких множеств — множеств, свободных от сумм $A + B$ с достаточно большими значениями $|A|$ и $|B|$. Автором в [218] доказан результат, расширяющий область применения известных конструкций редких множеств, в частности, [152].

Главу закрывает раздел 3.4, в котором показано, как простой факт из теории линейных схем позволяет для сложности реализации монотонной симметрической функции с порогом 2 монотонными схемами вывести оценки высокой точности [226]. Тем самым решается проблема, поставленная еще в 1970-х гг. Л. Адлеманом и П. Блониарцем [103].

В **главе 4** изучается задача минимизации сложности параллельных префиксных схем. Префиксная схема реализует систему префиксных сумм x_1, $x_1 \circ x_2$, ..., $x_1 \circ x_2 \circ \ldots \circ x_m$ в полугруппе $(G, \circ)$, используя только элементы $\circ$. В общем случае к бинарной операции $\circ$ не предъявляется никаких требований кроме ассоциативности.

Задача оптимизации сложности префиксной схемы становится нетривиальной при ограничении на глубину. Параллельные префиксные схемы используются в ряде приложений. В частности, на базе префиксных схем строятся эффективные схемы сумматоров чисел (префиксные сумматоры). Достоинством префиксных сумматоров является возможность балансирования различных характеристик схемы (сложности, глубины, ветвления элементов) за счет подбора конструкции опорной префиксной схемы.

Различные конструкции параллельных префиксных схем предлагались с конца 1950-х гг. Наиболее яркий результат получили Р. Ладнер и М. Фишер [156] около 1980 г., построив схемы минимально возможной глубины $\lceil \log_2 m \rceil$ и линейной сложности, около $4m$ (работа [156] относится к числу самых упоминаемых в теории синтеза). Чуть позже Ф. Фич [127] уточнила эту верхнюю оценку и доказала нетривиальную нижнюю оценку для случая $m = 2^n$.

Спустя 30 лет автору удалось установить точное значение сложности

минимальной префиксной схемы $m = 2^n$ входов и глубины n (оно приблизительно равно $3.5m$) [217, 235]. Заодно было установлено, что сложность схемы может быть понижена при дополнительных предположениях относительно операции $\circ$, в частности, для операции сложения по модулю 2 — эффект, видимо, ранее не замеченный.

В **главе 5** изучаются асимптотические вопросы синтеза схем и формул, использующих многовходовые функциональные элементы конъюнкции и дизъюнкции и либо элементы отрицания, либо отрицания переменных в качестве входов (в последнем случае вычислительные модели названы AC-схемами и AC-формулами).

Указанные модели исследуются, в первую очередь, в направлении получения нижних оценок сложности индивидуальных функций при ограничении на глубину. Модель AC-схем используется при определении классов сложности AC^k. Значительное число результатов получено для расширений рассматриваемого базиса линейными, симметрическими, пороговыми или MOD-функциями.

По не вполне ясным причинам поведение функции Шеннона сложности в указанных моделях изучалось мало. Лишь работа В. Данцика [121] специально посвящена оценкам для функции Шеннона сложности схем с отрицаниями, еще несколько оценок вытекают из результатов Э. И. Нечипорука [52] о синтезе схем и формул в базисах с нулевыми весами элементов. Асимптотически точные результаты, по всей видимости, не были сформулированы до работы автора [224] (хотя, например, для менее естественной модели схем из пороговых элементов очень нетривиальный асимптотически точный результат [45] был получен О. Б. Лупановым в 1970-е гг.).

В [224] установлена асимптотика функции Шеннона сложности AC-схем, причем она достигается на схемах глубины 3. (Схемы или формулы глубины 2 редуцируются до ДНФ или КНФ — этот случай не требует отдельного изучения.) Кроме того, асимптотически точные результаты получены для схем глубины 3 и формул глубины 4 с отрицаниями. Центральным местом в доказательстве верхних оценок является построение покрытий булева куба множествами специального вида, названными псевдосферическими.

Глава 6 посвящена классической задаче минимизации числа сравнений, необходимых для сортировки n элементов линейно упорядоченного множества в наихудшем случае. Из теоретико-информационных соображений это число не может быть меньше $\log_2 n! \sim n \log_2 n$. Указанная оценка достигается асимптотически на нескольких простых алгоритмах, известных с 1950-х гг. Лучший из них, алгоритм Форда—Джонсона [130] (вариант метода бинарных вставок) использует не более $\log_2 n! + cn$ сравнений, где $c < 0.12$. В нескольких разработанных позднее модификациях этого метода константа c в верхней оценке последовательно понижалась и по итогам работы [159] имела величину в районе 0.07. Для средней сложности сортировки по всем перестановкам входного набора известные к 2020 г. верхние оценки также имели вид $\log_2 n! + \Theta(n)$.

Автором в [228] получен следующий принципиальный результат: сложность сортировки составляет $\log_2(n!) + o(n)$ сравнений в наихудшем случае (и тем более в среднем). Другими словами, сортировка может быть реализована деревом решений (сравнений), почти идеально сбалансированным по глубине. Оценка опирается на эффективную процедуру вставки большого числа элементов в упорядоченный набор, организованную подобно системе массового обслуживания.

Итак, к **основным результатам** настоящей работы относятся:

— нижняя оценка константы равномерности монотонного булева базиса;

— методы синтеза эффективных по глубине или сложности формул для симметрических булевых функций, основанные на применении модулярной арифметики и приближенного суммирования;

— метод синтеза эффективных по глубине или сложности формул для элементарных периодических симметрических функций (MOD-функций), основанный на сведении к задаче о покрытии;

— метод доказательства нижних оценок сложности функций при реализации формулами в k-арном базисе, основанный на применении специальных мер сложности двудольных графов;

— решение задачи об асимптотически оптимальном синтезе вентильных схем глубины 3 в общем случае (когда соотношение между числом входов и числом выходов схемы не слишком мало или велико);

— результат о возможности эффективного погружения многомерных редких множеств в пространства меньшей размерности;

— пример последовательности булевых матриц с растущим отношением XOR- и OR-сложности в глубине 2; метод построения примеров последовательности булевых матриц с почти экстремальным отношением OR-сложности к сложности дополнительных матриц;

— точное значение сложности минимальной универсальной префиксной схемы глубины n на 2^n входах; верхние оценки сложности префиксных схем при различных ограничениях на глубину, и отдельно для префиксных XOR-схем;

— метод оптимального синтеза схем и формул глубины 3 из многовходовых элементов (типа конъюнкций, дизъюнкций) и отрицаний, основанный на построении специальных покрытий булева куба;

— решение задачи о сортировке за минимальное число сравнений в асимптотическом смысле с высокой точностью.

Теоретическая и практическая значимость. Работа носит теоретический характер. Ее результаты могут найти применение прежде всего в теоретических исследованиях закономерностей синтеза параллельных схем. Часть результатов имеет прикладной потенциал в схемотехнике или практике быстрых вычислений. Это относится, в первую очередь, к конструкциям параллельных умножителей, параллельных префиксных схем и отчасти к методу быстрой сортировки.

Методы исследования. В работе используются методы теории сложности вычислений и дискретной математики, в том числе, теории асимптотически оптимального синтеза и комбинаторики, а также элементарные методы алгебры и теории чисел.

Публикации. Перечень публикаций, в которых изложены основные результаты, приведен в конце работы: в рецензируемых изданиях опубликовано 16 работ.

В совместной работе [215] автору принадлежат примеры из §§2–3, отвечающие на основной вопрос, и их анализ. В работе [216] автору принадлежит технический результат §3, приводящий к уточнению оценок [12], а также следствия 3 и 4. В работе [218] автору принадлежат лемма 13, тео-

рема 4 и следствия из них (следствие 1, теорема 5 в общей формулировке). В работе [230] автору принадлежат результат §5.5 и пример в конце §5.6.

Апробация результатов. Результаты диссертации неоднократно докладывались на научных семинарах «Синтез и сложность управляющих систем» и «Математические вопросы кибернетики» в МГУ (2009–2012 гг.), на международных семинарах серии «Дискретная математика и ее приложения» (МГУ, 2010, 2019 гг.), на молодежных научных школах по дискретной математике и ее приложениям в ИПМ им. Келдыша (2013, 2015 гг.), на международной конференции «Проблемы теоретической кибернетики» (Нижний Новгород 2011 г.), на международной конференции «Современные проблемы анализа и преподавания математики» в МГУ в 2010 г., на семинаре «Теоретическая кибернетика» в ИПМ им. Келдыша в 2019 г. и на семинаре «Дискретная математика и математическая кибернетика» в МГУ в 2020 г.

2 Сложность и глубина формул. Формулы для симметрических булевых функций

В этом разделе представлен результат о соотношении между глубиной и сложностью формул в монотонном булевом базисе $B_M = \{\vee, \wedge\}$. Далее рассматриваются сложность и глубина реализации симметрических булевых функций формулами над базисом B_2 всех двуместных булевых функций и над стандартным базисом $B_0 = \{\wedge, \vee, {}^-\}$. Функция называется *симметрической*, если ее значения сохраняются при любых перестановках значений аргументов; в булевом случае это определение эквивалентно тому, что значения функции зависят только от арифметической суммы аргументов. Также предлагается метод нижней оценки сложности (глубины) реализации булевых функций формулами в k-арном базисе U_k, состоящем из всех k-местных функций, монотонно невозрастающих или монотонно неубывающих по каждой переменной.

Для сравнения порядков роста мы используем обозначения: $f = \omega(g)$ равносильно $g = o(f)$; $f = \Omega(g)$ равносильно $g = O(f)$ и может быть записано как $g \preceq f$ или $f \succeq g$; $f \asymp g$ означает $f = \Theta(g)$. Обозначения $f \sim g$, $f \gtrsim g$, $f \lesssim g$ используются соответственно для асимптотического равенства и неравенств.

2.1 Основные понятия и известные факты

Формулы

Напомним, что множество формул над базисом B, сложность формулы, глубина формулы и функция, реализуемая формулой, определяются индуктивно следующим образом: 0) константы базиса являются формулами сложности и глубины 0; 1) символы переменных являются формулами сложности 1, глубины 0 и реализуют соответствующие тождественные функции; 2) выражение $G(F_1, ..., F_k)$, где G — символ, обозначающий отличную от константы k-местную функцию $g \in B$, а F_i — формула сложности L_i и глубины D_i, реализующая функцию f_i, является формулой сложности $L_1 + \ldots + L_k$, глубины $\max\{D_1, ..., D_k\} + 1$ и реализует функцию

$g(f_1, \ldots, f_k)$.[1] (Неформально, сложность формулы — это число символов переменных в ней.)

Если базис B состоит из не более чем двуместных булевых функций, то используют компактное правило записи формулы: $(F_1 \circ F_2)$, где $\circ$ обозначает двуместную операцию базиса B, или $\overline{F}_1$ в случае одноместной операции отрицания. При этом скобки опускают в тех случаях, когда приоритет операций определен или не важен.

Сложность $L_B(f)$ (глубина $D_B(f)$) реализации булевой функции f формулами над базисом B определяется как минимум сложности (глубины) формул, реализующих f. Сложность $L_B(K)$ (глубина $D_B(K)$) класса функций K определяется как $\max_{f \in K} L_B(f)$ (соответственно $\max_{f \in K} D_B(f)$). Формула, реализующая булев оператор, определяется как совокупность формул, реализующих отдельные функции — компоненты оператора. Сложность булевого оператора определяется как сумма сложностей его компонент, а глубина — как максимум глубины компонент. Более подробно введенные понятия обсуждаются в [49, 59, 88, 95, 124, 145] (там же см. понятие схемы из функциональных элементов, которое встречается ниже, но несущественно для изложения основных результатов).

Известно (и просто проверяется), что любой полный бинарный базис относится к одному из двух типов: внутри одного типа базисов сложность любой функции одинакова с точностью до постоянного множителя. Принадлежность к одному из двух типов определяется наличием в базисе линейной функции. Сложности формул в базисах разных типов соотносятся как $L_{B_2}(f) \leq L_{B_0}(f) \preceq (L_{B_2}(f))^{\log_3 10}$. Верхняя оценка установлена в [179].

Аналогичное утверждение в отношении глубины гласило бы, что глубина функции при реализации формулами в базисах одного из указанных типов отличается не более чем на константу. В действительности, разница может быть более существенной: на нее влияет свойство базиса содержать или не содержать одновременно как обобщенную конъюнкцию $x^\alpha y^\beta$, так и обобщенную дизъюнкцию $x^\gamma \vee y^\delta$, где $\alpha, \beta, \gamma, \delta \in \{0, 1\}$. В работе [79]

[1] Сложность и глубина формул могут определяться иначе: например, сложность — как число (сумма весов) базисных функций, используемых при построении формулы, в глубине — не учитываться одноместные функции. Выводы настоящего раздела сохраняют силу при указанных вариациях определений.

В. М. Храпченко построил примеры функций и пар базисов, для которых достигается двукратное отношение глубин. Общая классификация бинарных базисов с точки зрения глубины изучалась в работе [161]. В частности, в ней установлено наличие не менее 4-х неэквивалентных типов базисов.

Сложность и глубина формул

Тривиальным образом в любом базисе, состоящем из не более чем k-местных функций, выполняется $D_B(f) \geq \log_k L_B(f)$. Базисы, для которых также выполнено $D_B(f) = O(\log L_B(f))$, называются равномерными. В равномерных базисах косвенно формульная сложность служит мерой параллельности функций.

Около 1968 г. свойство равномерности было установлено В. М. Храпченко для любого функционально полного конечного булева базиса [94]. Чуть позже Р. Брентом с соавторами в работах [109, 110] была доказана равномерность основных арифметических базисов $\{+, *\}$ и $\{+, *, /\}$. Наконец, к 1987 г. А. Б. Угольников [75] и М. Рагаз [187] независимо доказали равномерность произвольной (не обязательно полной) конечной булевой системы. Также они построили примеры неравномерных конечных систем в алгебре трехзначной логики. Равномерность полного конечного базиса функций k-значной логики устанавливается так же, как и в булевом случае. Исследование равномерности неполных конечных систем в k-значных логиках при $k \geq 3$ продолжено, например, в работах [66, 72] (подробная библиография приводится там же).

Следуя [85], равномерность базиса B можно охарактеризовать величиной (равной константе или ∞)

$$c_B = \overline{\lim_{N\to\infty}} \max_{L_B(f)=N} \frac{D_B(f)}{\log_2 N}.$$

Определение означает, что для любой выражаемой в базисе B функции f выполнено $D_B(f) \leq (c_B + o(1)) \log L_B(f)$, а также существует бесконечная последовательность функций f_k, для которой $D_B(f_k) \geq (c_B - o(1)) \log L_B(f_k)$. Здесь и далее по тексту логарифмы без указания основания полагаются двоичными.

Оценкам констант равномерности было посвящено множество работ,

преимущественно в 1970-х гг. Особый интерес представляют базисы арифметического типа, т. е. состоящие из операций сложения и умножения, и, возможно, вычитания или деления, в некотором полукольце, в связи с практической задачей о параллельном преобразовании арифметических выражений. В основном рассматривается базис $B_A = \{+, *\}$ и реже $B_{AD} = \{+, *, /\}$. Булевы аналоги арифметического базиса: монотонный базис $B_M = \{\vee, \wedge\}$, дополняемый до стандартного базиса $B_0 = \{\wedge, \vee, \bar{}\}$, и базис $\{\oplus, \wedge\}$, дополняемый до базиса Жегалкина $\{\oplus, \wedge, 1\}$. Тривиально выполняются соотношения

$$c_{B_0} \leq c_{B_M} \leq c_{B_A}, \qquad c_{\{\oplus,\wedge\}} \leq c_{B_A}.$$

Рекордные на сегодняшний день верхние оценки для указанных базисов $c_{B_M} < 1.73$ и $c_{B_A} \leq 2$ получены соответственно В. М. Храпченко [84] и С. Р. Косараю [153]. Нетривиальные нижние оценки известны только для общего арифметического базиса: соотношения $c_{B_A} > 1.16$ и $c_{B_A} \geq 1.5$ были получены последовательно в работах [196] и [120]. В булевом случае подобные нижние оценки известны только для некоторых неарифметических базисов: первая из них была установлена еще в 1967 г. В. М. Храпченко [79] для базиса $B_S = \{\backslash\}$, состоящего из единственной функции «штрих Шеффера». Отметим, что существуют и идеально равномерные базисы: так, для $B = \{\oplus\}$ тривиально выполняется $c_B = 1$.

Известные автору результаты для булевых и общих арифметических базисов сведены в таблице 1. Символ $\Rightarrow$ обозначает функцию импликации. Дизъюнкции и конъюнкции k входов обозначены соответственно как $\bigvee_k$ и $\bigwedge_k$. Через m_3 обозначена функция голосования трех переменных. B_2 обозначает базис всех двуместных булевых функций.

В базисе B_M нетривиальное соотношение между глубиной и сложностью, хотя и не приводящее к оценке $c_{B_M} > 1$, получила Б. Комменц-Вальтер [116] для функций вида

$$f_n = x_n \vee y_n(x_{n-1} \vee y_{n-1}(\ldots(x_1 \vee y_1)\ldots)).$$

Результат можно записать как $D_{B_M}(f_n) \geq \log n + \log\log n - O(1)$. Впоследствии М. И. Гринчук [14] доказал аналогичную верхнюю оценку. Два

	базис B	оценка c_B	год
Храпченко [79]	B_S	$c_B \geq 2$	1967
Храпченко [94]	ф.п. кон. булев базис	$c_B < \infty$	1968
Spira [202]	B_0	$c_B < 3.42$	1971
Brent, Kuck, Maruyama [109]	B_A	$c_B < 2.47$	1973
Brent [110]	B_{AD}	$c_B \leq 4$	1974
Preparata, Muller [180]	B_A	$c_B < 2.16$	1975
Barak, Shamir [99]	B_M	$c_B \leq 2$	1975
Muller, Preparata [165]	B_A B_{AD}	$c_B < 2.09$ $c_B < 2.89$	1975
Preparata, Muller [181]	B_M	$c_B < 1.82$	1976
McColl [161, 162]	$\{\Rightarrow, 0\},\ \{\Rightarrow, \setminus\}$ B_S B_2	$c_B > 1.44$ $c_B < 2.89$ $c_B < 2.47$	1977
Preparata, Muller, Barak [182]	$\{\bigvee_3, \bigwedge_3\}$ $\{\bigvee_4, \bigwedge_4\}$ $\{\bigvee_5, \bigwedge_5\}$	$c_B < 1.38$ $c_B < 1.18$ $c_B \leq 1$	1977
Храпченко [84]	B_M	$c_B < 1.73$	1978
Shamir, Snir [196]	B_A	$c_B > 1.16$	1980
Храпченко [85]	$\{m_3, {}^-, 0, 1\}$	$1 \leq c_B < 1.45$	1981
Kosaraju [153]	B_A	$c_B \leq 2$	1986
Угольников [75], Ragaz [187]	кон. булева система	$c_B < \infty$	1987
Coppersmith, Schieber [120]	B_A	$c_B \geq 1.5$	1992
Сергеев [225]	B_M	$c_B > 1.06$	2019

Таблица 1: Перечень известных оценок для констант равномерности

результата в совокупности устанавливают соотношение

$$D_{B_M}(f_n) = \log L_{B_M}(f_n) + \log\log L_{B_M}(f_n) \pm O(1).$$

В работе [117] метод [116] был распространен на полный базис B_0 и получена оценка

$$D_{B_0}(f_n) \geq \log L_{B_0}(f_n) + (1 - o(1)) \log\log\log L_{B_0}(f_n).$$

Автор в [225] путем оценки глубины функций достаточно естественно определяемых последовательностей установил первую нетривиальную

нижнюю оценку константы равномерности для монотонного булева базиса: $c_{B_M} > 1.06$. Этот результат изложен в разделе 2.2.

Сложность и глубина симметрических функций

Обозначим через S_n класс всех симметрических булевых функций n переменных. Пусть T_n^k обозначает пороговую симметрическую функцию n переменных с порогом k — по определению[2], $T_n^k(x_1, \ldots, x_n) = (x_1 + \ldots + x_n \geq k)$. Функция $T_n^{n/2}$ также называется функцией голосования[3].

Известные верхние оценки сложности и глубины реализации симметрических функций как формулами, так и схемами из функциональных элементов над полными базисами, связаны с эффективной реализацией булевого (n, m)-оператора $C_n(x_1, \ldots, x_n) = (C_{n,m-1}, ..., C_{n,0})$ подсчета числа единиц в булевом наборе $(x_1, \ldots, x_n)$, где $m = \lceil \log_2(n+1) \rceil$. Сведение к вычислению C_n используется при минимизации глубины и сложности формул для умножения двоичных чисел.

Предварительное представление о сравнительной сложности реализации функции $T_n^{n/2}$, оператора C_n и класса функций S_n дают следующие известные или легко выводимые оценки. (Параллельно получаемые соотношения для глубины опустим, во избежание дублирования.)

Теорема 2.1. *Для любого полного конечного базиса B справедливы соотношения*

$$L_B(C_n) \preceq \log n \cdot L_B(S_n), \qquad L_B(T_n^{n/2}) \preceq L_B(S_n),$$
$$L_B(T_n^{n/2}) \preceq L_B(C_{2n}), \qquad L_B(S_n) \preceq \frac{n}{\log n \log\log n} \cdot L_B(C_n),$$
$$L_B(C_n) \preceq \sum_{k=0}^{n} L_B(T_n^k), \qquad L_B(S_n) \preceq \sum_{k=0}^{n} L_B(T_n^k), \qquad L_B(T_n^k) \preceq L_B(T_{2n}^n).$$

Доказательство. Первые три неравенства очевидны: компоненты оператора C_n и функция голосования являются симметрическими функциями;

[2]Здесь и в аналогичных случаях ниже значение функции полагается равным 1, если выражение в правой части истинно; 0, иначе.

[3]Часто функция голосования определяется как $T_n^{(n+1)/2}$. Для дальнейшего рассмотрения это расхождение в определениях непринципиально.

функция $T_n^{n/2}$ является подфункцией функции $C_{2n,m}$, $m = \lfloor \log_2 n \rfloor$. Следующие три соотношения легко устанавливаются в случае $B = B_0$. Первое из них вытекает из того, что произвольная симметрическая функция представляется как функция компонент оператора C_n — эту функцию можно реализовать асимптотически оптимальным методом О. Б. Лупанова, см. [49, §14], [124, §4.1], учитывая то, что более половины компонент C_n имеют сложность не выше $L_B(C_n)/\log n$ по порядку. Пятое и шестое неравенства вытекают из простых формул, выражающих компоненты оператора C_n и функции из S_n через пороговые симметрические функции T_n^k. Наконец, любая из функций T_n^k является подфункцией функции T_{2n}^n (седьмое неравенство).

Для перехода к произвольному базису B используется соотношение

$$L_B(f) \le L + O(2^D), \tag{2.1}$$

где L и D — сложность и глубина некоторой формулы F, реализующей f над базисом B_0. Это соотношение следует и того, что любая из функций $\overline{x}$, $x \vee y$, xy выразима формулой в базисе B, в которой переменные x и y не повторяются. Это вытекает из свойства полных базисов (наличие немонотонной и нелинейной функций), см. [88, §8.3], [95, §I.1.6]. Остается проверить, что при подстановке в формулу F вместо функций базиса B_0 упомянутых выше формул над B, реализующих их, сложность формулы F возрастает не более чем на $O(2^D)$ (рост проистекает за счет несущественных переменных, присутствующих в замещающих формулах). □

При специальном выборе базиса, например, $B = S_l$, легко получить верхние оценки для $L_B(C_n), L_B(T_n^{n/2})$ вида $n^{1+\varepsilon(l)}$, где $\varepsilon(l) \to 0$ (см. также [82]). При достаточно больших l эти оценки очевидно близки к точным: известные нижние оценки имеют вид $L_B(T_n^{n/2}) \succeq n \log n$ [90] (см. также [124, §4.2.2]).

Методы синтеза, о которых далее пойдет речь, могут применяться в любом полном базисе. Исходя из того, что наибольший интерес представляют бинарные базисы, для демонстрации результатов мы выбираем по одному достаточно выразительному бинарному базису из разных сложностных классов (см. выше): базис B_2 всех двуместных булевых функций и стан-

дартный базис $B_0 = \{\wedge, \vee, {}^-\}$.

В случае $B \in \{B_0, B_2\}$ оценки теоремы 2.56 характеризуют соотношения между величинами $L_B(C_n)$, $L_B(T_n^{n/2})$ и $L_B(S_n)$ с удовлетворительной точностью ввиду большого расхождения между известными нижними и верхними оценками для каждой из этих величин.

А именно, наилучшие известные нижние оценки сложности для функции $T_n^{n/2}$ (и, как следствие, для C_n и S_n) имеют вид $L_{B_0}(T_n^{n/2}) \succeq n^2$ [80] (см. также [59, §8.5], [124, §4.5.1], [145, §6.8]) и $L_{B_2}(T_n^{n/2}) \succeq n \log n$ [129] (см. также [59, §8.7], [124, §4.2.3]).

Для глубины иных нижних оценок, чем те, которые тривиально вытекают из перечисленных оценок сложности, по-видимому, не известно.

Начиная с работы [81], для вывода верхних оценок сложности (и, аналогично, глубины, см., например, [168]) симметрических функций вместо грубого соотношения теоремы 2.56 используется следующее утверждение и следствие из него.

Утверждение 2.1. *Пусть K — класс булевых функций n переменных, булев (m,n)-оператор $\xi(x_1, \ldots, x_n) = (\xi_1, \ldots, \xi_m)$ таков, что любая функция $f \in K$ представляется в виде $f = \varphi(\xi_1(x_1, \ldots, x_n), \ldots, \xi_m(x_1, \ldots, x_n))$. Тогда для любого полного конечного базиса B справедливо*

$$L_B(K) \preceq \sum_{i=1}^{m} 2^i L_B(\xi_i) + 2^{2m},$$

а в случае $B \in \{B_0,\, B_2\}$ выполнено

$$D_B(K) \lesssim \max_{1 \le i \le m} \{D_B(\xi_i) + i\}.$$

Доказательство. Над базисом B_0 функция φ как функция компонент оператора ξ реализуется методом разложения по переменным (метод каскадов), см., например, [49, §5], [59, §4.3], [81], [95, §V.2.3]. Для произвольного базиса используем соотношение (2.1).

Для доказательства оценки глубины можно применить стандартную формулу разложения по нескольким переменным. Пусть $B = B_0$. Выбе-

рем параметр $k \approx \sqrt{m}$ и воспользуемся представлением

$$\varphi(y_1, \ldots, y_m) = \bigvee_{\alpha=(\alpha_1,\ldots,\alpha_k)\in\{0,1\}^k} y_1^{\alpha_1} \cdot \ldots \cdot y_k^{\alpha_k} \cdot \varphi_\alpha(y_{k+1}, \ldots, y_m).$$

К подфункциям φ_α также применим разложение по k переменным и т.д. Поскольку элементарные конъюнкции реализуются с глубиной $\lceil \log k \rceil$, глубина формулы относительно входа y_{ks+t}, $1 \le t \le k$, не превосходит

$$s(k+1) + \log k + 1 = ks + t + O(\sqrt{m}). \qquad \square$$

Следствие 2.1. *Пусть $m = \lceil \log_2(n+1) \rceil$. Для любого полного конечного базиса B справедливо*

$$L_B(S_n) \preceq \sum_{i=0}^{m-1} 2^{m-i} L_B(C_{n,i}),$$

а в случае $B \in \{B_0,\, B_2\}$ выполнено

$$D_B(S_n) \le \max_{0\le i<m} \{D_B(C_{n,i}) + m - i\}.$$

Аналогичный метод позволяет получать более точные, чем следуют из теоремы 2.56, оценки сложности пороговых симметрических функций, опираясь на следующее утверждение.

Утверждение 2.2. *Пусть $m = \lceil \log_2 k \rceil$. Для любого полного конечного базиса B справедливо*

$$L_B(T_n^k) \preceq \sum_{i=0}^{m-1} L_B(C_{n,i}) + L_B\left(T_n^{2^m}\right).$$

Доказательство. Функция T_n^k реализуется как дизъюнкция функции $T_n^{2^m}$ и функции сравнения m-разрядного числа $[C_{n,m-1}, ..., C_{n,0}]$ с числом k. Последняя реализуется формулой линейной сложности и глубины $O(\log m)$, см., например, [88, §10.1], [124, §2.5]. $\square$

Сведение к реализации функции $T_n^{2^m}$ объясняется тем, что обычно эту функцию легко вычислить попутно с вычислением C_n. Методы [81, 168] позволяют одновременно получать оценки $L_B(S_n) \preceq n^{1+\alpha}$ и $L_B(T_n^k) \preceq kn^\alpha$

(при небольших значениях k из метода [168] извлекаются даже лучшие оценки сложности T_n^k).

Одним из наиболее популярных приложений оператора C_n является реализация умножения и многократного сложения чисел, хотя обычно речь идет о реализации схемами или программами, а не формулами. Обозначим через M_n булев $(2n, 2n)$-оператор умножения двоичных n-разрядных чисел, а через $\Sigma_{n,k}$ булев $(kn, k+m)$-оператор сложения n штук k-разрядных чисел, $m = \lceil \log_2 n \rceil$.

Утверждение 2.3. *Пусть* $m = \lceil \log_2 n \rceil$. *Для любого полного конечного базиса* B *справедливо*

$$L_B(\Sigma_{n,k}) \preceq (k+m) \log^{O(1)} n \cdot L_B(C_n), \qquad L_B(M_n) \preceq L_B(\Sigma_{n,2n}).$$

Для $B \in \{B_0, B_2\}$ *справедливо*

$$D_B(\Sigma_{n,k}) \lesssim D_B(C_n) + \log_2(k + \log n), \qquad D_B(M_n) \lesssim D_B(\Sigma_{n,2n}).$$

Доказательство. Первое соотношение вытекает из неравенств

$$L_B(\Sigma'_{n,h}) \le L_B(C_n) \cdot L_B(\Sigma'_{m,h}), \qquad L_B(C_n) \preceq n^{O(1)}, \qquad L_B(\Sigma'_{n,h}) \preceq n^{O(1)}.$$

где оператор $\Sigma'_{n,h}$ вычисляет первые h разрядов суммы n чисел. Первое из этих неравенств является интерпретацией «школьного» метода сложения чисел, второе доказано в [81], третье является их следствием. Второе соотношение утверждения вытекает из школьного метода умножения.

В оценке глубины используется результат Храпченко [79], позволяющий складывать k-разрядные числа с глубиной $(1 + o(1)) \log_2 k$. □

Отметим, что основанный на реализации оператора C_n метод получения оценки глубины умножения (о нем подробнее пойдет речь ниже) подразумевает построение схем достаточно большой, квадратичной, сложности. Поэтому немалый интерес представляют параллельные модификации быстрых методов умножения чисел, скажем, методов Карацубы [19], Тоома [74] и Шёнхаге—Штрассена [193], не приводящие к существенному росту сложности. Имеется, по меньшей мере, два подхода к параллельному перестроению числовых алгоритмов — оба связаны с применением специального кодирования чисел, в котором аддитивные операции выполняются

с глубиной $O(1)$. Первый способ использует запись в знаковой четверичной системе счисления и известен с 1960-х гг. [98], см. также [209]. Во втором способе, предложенном А. В. Чашкиным [87, 88], число u кодируется парой чисел (u_1, u_2), разность которых доставляет истинное значение $u = u_1 - u_2$. Аккуратные оценки глубины параллельных модификаций методов Карацубы и Шёнхаге—Штрассена также приведены в [231].

Набор пороговых функций $T_n = (T_n^1, \ldots, T_n^n)$ составляет оператор сортировки n булевых переменных. Чаще встречается задача сортировки многоразрядных чисел. Э. Ш. Коспанов [28] указал экономный способ сведения сортировки n штук m-разрядных чисел к вычислению оператора T_n, получив верхнюю оценку глубины $D_{B_0}(T_n) + (1 + o(1)) \log_2 n + \log_2 m$.

Синтез эффективных схем

Рассмотрим вопрос о реализации оператора C_n. Почти все известные эффективные по сложности и глубине формулы и схемы для C_n строятся из компрессоров[4] [5]. Двоичный (k, l)-компрессор ширины 1 — это формула (или схема), реализующая булев оператор $(x_1, \ldots, x_k) \to (y_1, \ldots, y_l)$, определяемый условием $\sum 2^{a_i} x_i = \sum 2^{b_j} y_j$, где $k > l$, $a_i, b_j \in \mathbb{Z}$. (k, l)-компрессор произвольной ширины строится из параллельных копий компрессоров ширины 1 и позволяет сводить сложение k чисел к сложению l чисел[6]. Действительно, k чисел $x^i = [x^i_{m-1}, ..., x^i_0]$, $1 \le i \le k$, параллельными преобразованиями $(x^1_{j+a_1}, \ldots, x^k_{j+a_k}) \to (y^1_{j+b_1}, \ldots, y^l_{j+b_l})$, $j \in \mathbb{Z}$, переводятся в l чисел $y^i = [y^i_{m+h}, ..., y^i_0]$, $1 \le i \le l$, где $h < \log_2 k$, с сохранением суммы (все не определенные разряды в приведенных формулах полагаются равными нулю).

Вывод верхних оценок формульной или схемной сложности или глубины оператора C_n обычно состоит из двух этапов: 1) конструирование подходящего (k, l)-компрессора (при малых k и l), 2) синтез формулы (схемы) из компрессоров и вспомогательных формул (схем). Второй этап в случае

[4] Западный термин для компрессора — CSA (carry save adder)

[5] Впрочем, для построения коротких формул в базисе B_0 пригодна простая конструкция, описанная в [49, §22], [95, §V.2.8]. С ее помощью несложно вывести оценки $L_{B_0}(C_{2^m,i}) \le \left(C^i_{m+i-1} - C^{i-2}_{m+i-1}\right) \cdot 8^m$ и, как следствие, $L_{B_0}(C_n) \preceq n^5$, $L_{B_0}(S_n) \preceq n^{5.17}$.

[6] Поэтому дальнейшее рассмотрение можно ограничить компрессорами ширины 1.

оценки схемной сложности не представляет затруднений, в случае оценки сложности и глубины формул он в значительной степени унифицирован в работах [168, 169, 171]. Поэтому в любом случае задача фактически сводится к построению элементарных компрессоров.

Простейшим компрессором является $(3, 2)$-компрессор (для него принято обозначение FA_3 (full adder)). Он реализует сумму трех битов по правилу $x_1 + x_2 + x_3 = 2y_2 + y_1$. Имеются упоминания об этом компрессоре (и о компрессорах вообще) в контексте быстрого вычисления сумм, относящиеся к 1960 г., например [70], но возможно есть и более ранние. Впоследствии этот компрессор и принцип его применения переоткрывался в работах многих авторов (перечень этих работ приводится, например, в [169]).

Рекордные оценки сложности и глубины (как для формул, так и для схем) получаются при помощи сложно устроенных компрессоров. В частности, используется идея использования дополнительных способов кодировки двоичных наборов [203, 171] в конструкциях компрессоров. Примеры эффективных компрессоров с несколькими типами кодировки входов и выходов построены в [203] (сложность схем, вычисляющих сумму по модулю 4, над базисом B_2), [171] (глубина формул над базисом B_2), [122] (сложность схем над базисом B_2), [234] (сложность формул над базисами B_0 и B_2).

Первые аккуратные оценки сложности и глубины оператора C_n, а также класса симметрических функций были получены В. М. Храпченко [81, 83] (для базиса B_0). Серия работ разных авторов 1970-х гг. была посвящена уточнениям оценок в базисе B_2. Общий подход к доказательству верхних оценок сложности и глубины был предложен в [169], а именно, был установлен оптимальный способ размещения компрессоров, составляющих формулы. Известные верхние оценки для оператора C_n и класса симметрических функций n переменных представлены в табл. 2 в хронологическом порядке (неконструктивные оценки отмечены символом *).

Последние результаты получены автором в [219, 220, 221]. Оценки [219, 220] основаны на простом применении модулярной арифметики[7]. Вместо прямого подсчета арифметической суммы σ булевых переменных $x_1, \ldots, x_n$

[7] Отметим, что модулярная арифметика уже нашла применение при реализации симметрических функций контактными схемами, см., например, [43].

	L_{B_0}	L_{B_2}	D_{B_0}	D_{B_2}
C_n	$n^{4.62}$ [81] $n^{4.60}$ [168] $n^{4.57}$ [171] $n^{4.54}$ [234] $n^{4.47}$ [220] $n^{3.91}$ [221]*	$n^{3.55}$ [175] $n^{3.42}$ [167] $n^{3.32}$ [173] $n^{3.16}$ [168] $n^{3.13}$ [171] $n^{3.06}$ [234] $n^{3.03}$ [220] $n^{2.84}$ [221]*	$5.12 \log n$ [83] $5.07 \log n$ [168] $4.95 \log n$ [170] $4.94 \log n$ [136] $4.87 \log n$ [219] $4.14 \log n$ [221]*	$4 \log n$, фольклор $3.71 \log n$ [168] $3.57 \log n$ [170] $3.48 \log n$ [171] $3.44 \log n$ [136] $3.34 \log n$ [219] $3.02 \log n$ [221]*
S_n	$n^{4.93}$ [81] $n^{4.85}$ [168] $n^{4.82}$ [234] $n^{4.48}$ [220] $n^{4.01}$ [221]*	$n^{3.55}$ [175] $n^{3.42}$ [167] $n^{3.37}$ [173] $n^{3.30}$ [168] $n^{3.23}$ [234] $n^{3.04}$ [220] $n^{2.95}$ [221]*	$4.88 \log n$ [219] $4.24 \log n$ [221]*	$3.81 \log n$ [168] $3.34 \log n$ [219] $3.10 \log n$ [221]*

Таблица 2: Верхние оценки для оператора C_n и класса S_n

вычисляются числа $(\sigma \bmod 2^k)$ и $(\sigma \bmod 3^l)$, где $2^k \cdot 3^l > n$, причем для вычисления $(\sigma \bmod 3^l)$ используется троичная система счисления. Искомое число σ может быть затем определено при помощи Китайской теоремы об остатках. Эффективность способа обусловлена тем, что самые сложные для метода компрессоров разряды суммы σ — старшие — вычисляются почти «бесплатно» из младших средствами модулярной арифметики.

В работе [221] предлагается еще один прием, упрощающий формулы. Он состоит в дополнительном приближенном вычислении суммы σ. Отрезок $[0, n]$ накрывается интервалами длины T. Значение σ определяется по номеру интервала и остаткам $\sigma \bmod 2^k$, $\sigma \bmod 3^l$ при условии $2^k \cdot 3^l \geq T$. Размер формул сокращается из-за того, что процесс вычисления суммы компрессорами останавливается раньше (когда найденных разрядов достаточно для восстановления числа по модулю T). Для вычисления номера интервала применяется вероятностная процедура, предложенная Л. Вэльянтом [207], поэтому полученные оценки оказываются неконструктивными.

Отметим, что приближенное суммирование используется в широком спектре задач теории сложности. Например, идея локализации значений суммы в интервалах играет ключевую роль в [199] при вычислении пороговых монотонных функций контактными схемами.

MOD-функции

Интересный для исследования подкласс симметрических функций составляют элементарные периодические функции, которые мы для краткости будем называть MOD-функциями.

Обозначим через MOD_n^m булев (n, m)-оператор сложения n одноразрядных чисел по модулю m. Компоненты оператора — MOD-функции — определяются как

$$\mathrm{MOD}_n^{m,r}(x) = \left(\sum_{i=1}^{n} x_i \equiv r \bmod m \right), \tag{2.2}$$

$r \in \mathbb{Z}_m$, где $x = (x_1, \ldots, x_n)$.

Универсальный способ вычисления MOD-функции, как и произвольной симметрической булевой функции, состоит в сведении к реализации оператора C_n. Но для конкретных MOD-функций известны лучшие оценки. Основной интерес представляет случай простого модуля m. Сложность MOD-функций с составным модулем m определяется сложностью MOD-функций для отдельных взаимно простых множителей числа m (Китайская теорема об остатках). Сложность MOD-функций с примарными модулями p^k определяется сложностью MOD-функций с модулем p, а именно

$$L_B\left(\mathrm{MOD}_n^{p^k}\right) \preceq n^{o(1)} L_B(\mathrm{MOD}_n^p), \quad D_B\left(\mathrm{MOD}_n^{p^k}\right) \leq D_B(\mathrm{MOD}_n^p) + o(\log n)$$

при постоянном k, см. [129]. Таким образом, достаточно ограничить рассмотрение простыми m.

Известные нижние оценки для $m \neq 2^k$ далеко отстоят от верхних. А именно, $L_{B_0}(\mathrm{MOD}_n^m) \succeq n^2$ (следствие из метода [80]) и $L_{B_2}(\mathrm{MOD}_n^m) \succeq n \log n$ при $m > 2$ [129]. Для оценки глубины предлагается пользоваться стандартным для бинарных базисов соотношением $D_B(f) \geq \log L_B(f)$.

Нетривиальные верхние оценки были получены для сложности формул в базисе B_2 при $m = 3, 5, 7$ [157] и для глубины в базисе B_0 при $m =$

$3, 5, 11$ [113][8].

Автором в [222] предложен новый метод синтеза эффективных формул для MOD-функций, сводящийся к построению покрытий малого ранга для циклических (циркулянтных) булевых матриц. Наилучшие результаты сведены в табл. 3 (неконструктивная оценка отмечена символом *).

m	L_{B_0}	L_{B_2}	D_{B_0}	D_{B_2}
3	$n^{2.59}$ [43]	n^2 [129, 157]	$2.80 \log n$ [222]	$2 \log n$ [161]
5	$n^{3.22}$ [222]	$n^{2.84}$ [221]* n^3 [157]	$3.35 \log n$ [222]	$3 \log n$, следует из [157]
7	$n^{3.63}$ [222]	$n^{2.59}$ [157]	$3.87 \log n$ [222]	$2.93 \log n$ [222]

Таблица 3: Наилучшие известные верхние оценки для операторов MOD_n^m

Отметим, что для сложности реализации $\mathrm{MOD}_n^{m,r}$-функций схемами из функциональных элементов нетривиальные верхние оценки известны только в случае $m \leq 4$, см. [151].

Ниже в разделах 2.3–2.5 излагаются конструктивные методы синтеза формул для симметрических функций [219, 220], неконструктивные оценки сложности и глубины симметрических функций [221] и методы синтеза формул для MOD-функций [222].

Формулы в k-арных базисах

С позиции получения нижних оценок сложности (глубины) формул симметрических функций в базисе B_0 самым сильным инструментом пока остается метод Храпченко [80]. Он позволяет получать квадратичные нижние оценки сложности для функций, принимающих разные значения в средних слоях булева куба. (В более выразительном базисе B_2 рекордные нижние оценки имеют величину $\Omega(n \log n)$ [129].)

Оценка Храпченко связана с понятием *чувствительности* булевой функции f, которая определяется как

$$s(f) = \max_{N \subset f^{-1}(0),\, P \subset f^{-1}(1)} \frac{|R(N,P)|^2}{|N| \cdot |P|}, \tag{2.3}$$

[8]Последняя оценка (для $m = 11$) уже уступает общей оценке глубины симметрических функций.

где $R(N,P)$ — множество пар соседних наборов из N и P, т. е. отличающихся в одной координате. Храпченко [80] доказал соотношение $L_{B_0}(f) \geq s(f)$ для стандартного базиса $B_0 = \{\vee, \wedge, {}^{-}\}$.

Максимальную чувствительность $s(f) = n^2$ среди функций n переменных имеет линейная функция $f = l_n(x_1, \ldots, x_n) = x_1 \oplus \ldots \oplus x_n$. Другие популярные примеры — функция голосования $m_n(x_1, \ldots, x_n) = (\sum x_i \geq n/2)$, для которой $s(m_n) \sim n^2/4$, монотонные пороговые функции вообще, определитель матрицы над $GF(2)$, см. [80].

В ряде работ изучалась сложность вычисления линейной функции и функции голосования в полных k-арных базисах. В основном нижние оценки получались методом сжатия формул при случайных подстановках, восходящем к работе Б. А. Субботовской [71]. Получаемые оценки имеют вид $L_B(l_n) = \Omega\left(n^{\Gamma_B}\right)$, где Γ_B — экспонента сжатия базиса B (подробнее об экспонентах сжатия см., например, в [145]).

Сама Субботовская (Мучник) [51] получила оценки $\Gamma_B \geq 1 + \frac{1}{3k-4}$ для некоторых k-арных базисов B. Среди k-арных базисов особый интерес представляет максимальный базис U_k, в котором сложность линейной функции нелинейна. Легко проверить, что базис U_k содержит все функции, по каждой переменной монотонно невозрастающие или монотонно неубывающие, т.е. ровно те функции, из которых нельзя получить линейную функцию двух переменных путем подстановки констант, инверсий и отождествлений переменных. Тривиально выполняется $U_k \subset U_{k+1}$ при любом k. В силу определения базис U_k строго эквивалентен[9] базису из всех k-местных монотонных функций и отрицания.

Н. А. Перязев [62] доказал, что $\Gamma_{U_k} \geq 1 + \frac{1}{3k-4}$. Потом этот результат был независимо получен Д. Ю. Черухиным [89] (см. также [92]), а также Х. Чоклер и У. Цвиком [114]. Для базиса U_3 авторы [114] доказали более сильную оценку $\Gamma_{U_3} \geq 4/3$. Последний результат приводит к наилучшей известной оценке сложности линейной функции в тернарном базисе $L_{U_3}(l_n) = \Omega\left(n^{4/3}\right)$. В работе [91] Черухин выполнил классификацию k-арных базисов относительно сложности линейной функции: в лю-

[9]Базисы B_1 и B_2 называются строго эквивалентными, если $L_{B_1}(f) = L_{B_2}(f)$ для любой представимой в них функции f.

бом базисе B выполнено либо $L_B(l_n) = \Theta(n)$, либо $L_B(l_n) = \Theta(n^2)$, либо $n^\beta \preceq L_B(l_n) \preceq n^\gamma$ при некоторых $1 < \beta < \gamma < 2$. Здесь $L_{U_k}(l_n) = \Theta(n^\beta)$.

Верхняя оценка сложности линейной функции $L_{U_3}(l_n) \preceq n^{1.74}$ получена в работе [114] и также фактически содержится в [91]. Для функции голосования К. Уено [205] получил верхнюю оценку $L_{U_3}(m_n) \preceq n^{3.8}$. Отметим, что эта оценка может быть существенно понижена применением методов из [220, 221].

Для полноты картины перечислим известные оценки сложности функции голосования в монотонных базисах. Обозначим через M_k базис из всех k-местных монотонных функций. В случае $k = 3$ имеют место соотношения $n^{\log_2 3} \preceq L_{M_3}(m_n) \preceq n^{4.3}$. Нижняя оценка получена М. М. Рохлиной в [64], а верхняя — А. Гуптой и С. Махаджаном в [137]. В работе [137] установлена и общая оценка вида $L_{M_k}(m_n) \preceq n^{3+c/\ln k}$.

Автором в [229] предложен обобщающий [80] метод оценки сложности формулы через чувствительность функции. Для булева базиса B определим *экспоненту Храпченко* χ_B как максимальное число χ, такое, что для любой функции f, выразимой в базисе B, выполнено[10]

$$L_B(f) \geq s^\chi(f). \tag{2.4}$$

Из определения немедленно следуют оценки $L_B(l_n) = \Omega\left(n^{2\chi_B}\right)$, $L_B(m_n) = \Omega\left(n^{2\chi_B}\right)$.

Всегда выполняется $\chi_B \geq 1/2$, поскольку для любой функции f, существенно зависящей от n переменных, $L_B(f) \geq n$ и $s(f) \leq n^2$. Для полных бинарных булевых базисов значения экспонент Храпченко доставляют известные классические результаты: $\chi_B = 1$ для $B \subset U_2$ и $\chi_B = 1/2$ для прочих базисов. Для любого полного k-арного базиса B, не содержащегося в U_k, тривиально выполнено $\chi_B = 1/2$ (так как сложность линейной функции в базисе B линейна).

В работе [229] показано, что при любом $k \geq 2$ справедливы соотношения

$$\frac{1}{2} + \frac{1}{10\ln k} \leq \chi_{U_k} \leq \frac{1}{2} + \frac{1}{2\log(k/2)}.$$

[10] Максимум существует, поскольку множество подходящих чисел χ ограничено и замкнуто.

Нижняя оценка получается как следствие из более общего результата о специальных мерах сложности двудольных графов. Для тернарного базиса установлены более точные оценки $0.769 \leq \chi_{U_3} \leq \log_4 3 \approx 0.792$. Как следствие, при любом $k \geq 2$ имеет место

$$L_{U_k}(l_n), L_{U_k}(m_n) \in \Omega\left(n^{1+1/(5\ln k)}\right),$$

(указанная оценка сильнее ранее известных при $k \geq 4$), а при $k = 3$ справедливо

$$L_{U_3}(l_n), L_{U_3}(m_n) \in \Omega\left(n^{1.538}\right).$$

Простые верхние оценки показывают, что на самом деле $L_{U_k}(l_n) = O\left(n^{1+c/\ln k}\right)$ при любом k и подходящей константе c. Эти результаты изложены в разделе 2.6.

2.2 Глубина и сложность монотонных формул

Как и в работах [196, 120], нижняя оценка глубины доказывается для бесповторных функций, т. е. выражаемых формулами, в которых переменные не повторяются. (На самом деле, при выводе как нижних, так и верхних оценок констант c_B достаточно ограничиться бесповторными функциями.) Стратегия рассуждения состоит в рассмотрении двух главных подформул минимальной по глубине формулы для данной функции. Доказывается, что либо одна из подформул содержит бесповторную функцию, «похожую» на исходную (т. е. при переходе на меньшую глубину функция не сильно упрощается), либо суммарное число существенных переменных, от которых зависят две подфункции, реализуемые главными подформулами, заметно возрастает. Это позволяет оценить сложность формул при различных ограничениях на глубину и воспользоваться тем фактом, что сложность формулы глубины d не может быть больше, чем 2^d.

Метод работает и в монотонном арифметическом базисе B_A, позволяя получать потенциально более высокие нижние оценки, чем в булевом случае, однако уступающие уже известным оценкам для этого базиса.

2.2.1 Расщепление монотонных функций

Примем соглашение о том, что запись $X = \alpha$, где $\alpha \in \{0, 1\}$, означает, что вместо всех переменных группы X подставлена константа α.

Напомним, что моном $x^S = \prod_{i \in S} x_i$ называется *импликантой* монотонной булевой функции f, если $f \geq x^S$. Если дополнительно при любом $i \in S$ свойство $f \geq x^{S \setminus \{i\}}$ не выполнено, то x^S называется *простой импликантой* функции f.

В основе стратегии доказательства лежит

Лемма 2.1. *Пусть f, g — отличные от констант монотонные булевы функции непересекающихся групп переменных X, Y. Тогда, если*

$$f(X) \vee g(Y) = p(X, Y) \cdot q(X, Y),$$

где p, q — монотонные функции, то либо $p(X, 0) = f(X)$, либо $q(0, Y) = g(Y)$; симметричным образом, либо $q(X, 0) = f(X)$, либо $p(0, Y) = g(Y)$.

Доказательство. Пусть p и q отличны от констант (иначе с очевидностью утверждение леммы выполнено). Запишем $p = p_X \vee p_Y \vee p_{XY}$ и $q = q_X \vee q_Y \vee q_{XY}$, где $p_X = p(X, 0)$, $q_X = q(X, 0)$, $p_Y = p(0, Y)$, $q_Y = q(0, Y)$, а функции p_{XY} и q_{XY} объединяют простые импликанты функций p и q, в которые входят переменные как из X, так и из Y.

По построению, $p_X q_X = f$ и $p_Y q_Y = g$. В частности, $p_X \geq f$ и $q_Y \geq g$. Не ограничивая общности, предположим, что $p_X \neq f$, т. е. p_X содержит некоторую импликанту I, которой нет у функции f. Если при этом q_Y содержит импликанту J, которой нет у g, то произведение pq содержит импликанту IJ, которой нет у $f(X) \vee g(Y)$, что невозможно. Поэтому $q_Y \leq g$. Учитывая, что одновременно выполнено $q_Y \geq g$, получаем $q_Y = g$. Рассуждая симметрично, устанавливаем, что либо $q_X = f$, либо $p_Y = g$. □

В силу принципа двойственности (см., например, [95]) аналогичное утверждение справедливо для представлений типа $f(X)g(Y) = p \vee q$.

Содержательно, лемма устанавливает, что либо функции p и q имеют общие подфункции (т. е. суммарное число существенных переменных у p и q больше, чем у исходной функции $f \vee g$), либо одна из функций p и q содержит f и g в качестве подфункций.

2.2.2 Нижняя оценка

Пусть $t \geq 2$ — параметр. Последовательность обобщенных чисел Фибоначчи $\Phi_k^{(t)}$ зададим соотношением $\Phi_n^{(t)} = \Phi_{n-1}^{(t)} + \Phi_{n-t}^{(t)}$ и начальными условиями $\Phi_{1-t}^{(t)} = \ldots = \Phi_0^{(t)} = 1$.

Построим специальную последовательность функций $f_k^\vee, f_k^\wedge$. Функции с индексом k зависят от $\Phi_k^{(t)}$ переменных. При $k \leq 0$ положим $f_k^\vee(x) = f_k^\wedge(x) = x$, а далее определим рекурсивно как

$$f_k^\vee(X,Y) = f_{k-1}^\wedge(X) \vee f_{k-t}^\wedge(Y), \qquad f_k^\wedge(X,Y) = f_{k-1}^\vee(X) \cdot f_{k-t}^\vee(Y), \quad (2.5)$$

где X, Y — непересекающиеся группы из соответственно $\Phi_{k-1}^{(t)}$ и $\Phi_{k-t}^{(t)}$ переменных. По построению, функции $f_k^\vee$ и $f_k^\wedge$ двойственны[11].

Предлагаемый способ построения «труднораспараллеливаемых» функций (2.5) созвучен примеру из [120]. Чередование операций применяется, чтобы в основной формуле не возникали выражения типа конъюнкции или дизъюнкции большого числа элементов — такие выражения допускали бы простую балансировку по глубине. Важно, чтобы сложность слагаемых в правой части формул (2.5) существенно различалась — тогда вычисления прямо по этим формулам становятся невыгодными с точки зрения глубины. И напротив, можно ожидать, что параллельная реализация функций $f_k^\vee, f_k^\wedge$ должна идти вразрез с их естественной структурой, приводя к избыточной сложности. С другой стороны, разница сложности слагаемых в (2.5) не должна быть чрезмерно большой, так как выражения вида $fg \vee h$ можно вычислять по формуле $(f \vee h)(g \vee h)$, достаточно экономной в случае, когда h — простая функция. Предполагаем, что выполнение последних двух условий может быть обеспечено грамотным выбором параметра t.

Далее, для удобства рассуждений будем работать с расширенным константами базисом $B'_M = B_M \cup \{0, 1\}$. Легко проверить, что константы бесполезны для реализации отличных от констант функций: их можно удалить из любой формулы, не увеличивая сложность и глубину. Поэтому и константа равномерности одна и та же для базисов B_M и B'_M.

[11] Напомним, что функции $f(X)$ и $g(X)$ называются двойственными, если $f(X) = \overline{g(\overline{X})}$, где $\overline{X}$ — вектор из отрицаний переменных X.

Из определения следует, что функции $f_k^\vee, f_k^\wedge$ выражаются бесповторными формулами от $\Phi_k^{(t)}$ своих переменных. Эти формулы (при $k \geq 0$) имеют глубину $k \sim \log_\phi 2 \cdot \log \Phi_k^{(t)}$, где ϕ — единственный положительный корень многочлена $x^t - x^{t-1} - 1$.

Через $\nu_d^0(f)$ и $\nu_d^1(f)$ обозначим минимум сложности реализации функции f формулами над B'_M глубины не более d с внешней операцией дизъюнкции и, соответственно, конъюнкции. Тогда $\nu_d(f) = \min\{\nu_d^0(f), \nu_d^1(f)\}$ означает просто сложность реализации функции f формулами глубины не выше d. Если реализация с глубиной d невозможна, то формально положим $\nu_d(f) = \infty$. В силу двойственности,

$$\nu_d^0(f_k^\vee) = \nu_d^1(f_k^\wedge), \qquad \nu_d^1(f_k^\vee) = \nu_d^0(f_k^\wedge). \tag{2.6}$$

Для компактности обозначений положим $f_k = f_k^\vee$.

Наша дальнейшая цель — получение нижней оценки для $\nu_d(f_k)$. Тогда из неравенства $\nu_d(f_k) > 2^d$ извлекается оценка глубины $D_{B_M}(f_k) > d$.

Следующий факт относительно сложности формул тривиален.

Утверждение 2.4. *Пусть $g_1, \ldots, g_s$ — подфункции функции f, попарно не имеющие общих переменных. Тогда $\nu_d^\alpha(f) \geq \nu_d^\alpha(g_1) + \ldots + \nu_d^\alpha(g_s)$, где $\alpha \in \{0, 1\}$.*

Доказательство. Рассмотрим формулу F, на которой достигается оценка $\nu_d^\alpha(f)$. Подставляя константы в эту формулу, получим формулу F_i для каждой подфункции g_i. По условию, любая переменная x функции f может встречаться не более чем в одной формуле F_i. При этом по построению число вхождений x в формулы F и F_i одинаково. □

Далее это утверждение мы будем комбинировать с оценкой сложности формулы через сумму сложностей подформул.

Лемма 2.2. *Справедливы соотношения:*

$$\nu_d^0(f_k) \geq \nu_d^1(f_{k-1}) + \nu_d^1(f_{k-t}). \tag{2.7}$$

$$\begin{aligned}\nu_d^1(f_k) \geq \min\{\, & 2\nu_{d-1}(f_{k-1}) + \nu_d^0(f_{k-t}), \\ & 2\nu_{d-1}(f_{k-t}) + \nu_d^0(f_{k-1}),\ \ \nu_{d-1}(f_{k-1}) + \nu_{d-1}(f_{k-t})\,\}.\end{aligned} \tag{2.8}$$

Доказательство. Соотношение (2.7) выполнено в силу определения функций f_k, утверждения 2.4 и свойства (2.6). Второе соотношение устанавливается при помощи леммы 2.1. Рассмотрим формулу минимальной сложности с внешним элементом конъюнкции, реализующую функцию $f_k = f^{\wedge}_{k-1}(X) \vee f^{\wedge}_{k-t}(Y)$ при ограничении d на глубину. Пусть p, q — функции, реализуемые на входах финального элемента.

Согласно лемме 2.1, выполнено одно из трех. Если $p(X,0) = q(X,0) = f^{\wedge}_{k-1}(X)$, то с учетом утверждения 2.4 и (2.6) величина $\nu^1_d(f_k)$ оценивается снизу первым выражением под знаком минимума в (2.8). Иначе, если $p(0,Y) = q(0,Y) = f^{\wedge}_{k-t}(Y)$, то величина $\nu^1_d(f_k)$ оценивается вторым выражением в (2.8). В последнем случае, скажем, для функции p выполнено $p(X,0) = f^{\wedge}_{k-1}(X)$ и $p(0,Y) = f^{\wedge}_{k-t}(Y)$. Так получаем третье выражение в (2.8). □

Теорема 2.2. *Справедливо соотношение:* $c_{B_M} > 1.06$.

Доказательство. Первая часть доказательства основана на стандартном приеме: для оценки величины $\nu_d(f_k)$ (по индукции при помощи леммы 2.2) подберем подходящее простое выражение, в нашем случае зависящее от d, k, параметра t и вспомогательных параметров $a > 1, c > 0$ и γ.

Обозначим $r_{d,k} = ca^{k-\gamma d}\phi^k$. Покажем, что при надлежащем выборе параметров t, a, c, γ справедливо

$$\nu^1_d(f_k) \geq r_{d,k}, \qquad \nu_d(f_k),\ \nu^0_d(f_k) \geq r_{d,k-1} + r_{d,k-t}. \tag{2.9}$$

Более точно, мы определим условия на параметры, при которых (2.9) заведомо выполнено.

По определению величин $r_{d,k}$ и в силу $(a\phi)^t > (a\phi)^{t-1} + 1$ справедливо

$$r_{d,k} \geq r_{d,k-1} + r_{d,k-t}. \tag{2.10}$$

Значит, для $\nu^1_d(f_k)$ используется более высокая оценка, чем для $\nu^0_d(f_k)$. Поэтому мы оцениваем $\nu_d(f_k)$ так же, как $\nu^0_d(f_k)$.

При заданных t, γ и a константа c определяется так, чтобы заявленные оценки выполнялись при $k \leq t$ и допустимых d (поскольку не имеет смысла рассматривать значения $d \geq k$, то оставшихся пар k, d — конечное число).

Очевидно, $c > 0$, так как все величины $\nu_d(f_k)$ положительны, а их число конечно.

Далее, считая, что t, γ и a выбраны подходящим образом, применяем индукцию по $d + k$. Индуктивный переход состоит в доказательстве того, что оценки (2.9) не противоречат неравенствам (2.7) и (2.8).

Соотношение (2.7) удовлетворено автоматически видом выбранной оценки для $\nu_d^0(f_k)$. Рассмотрим (2.8). Первое выражение под знаком минимума в (2.8) всегда больше третьего. Действительно,

$$2\nu_{d-1}(f_{k-1}) + \nu_d^0(f_{k-t}) > 2\nu_{d-1}(f_{k-1}) \geq \nu_{d-1}(f_{k-1}) + \nu_{d-1}(f_{k-t}).$$

Тогда (2.9) будет согласовано с (2.8), если $r_{d,k}$ не превосходит оценки величины двух последних выражений под знаком минимума, т. е. справедливо

$$r_{d,k} \leq r_{d,k-2} + r_{d,k-t-1} + 2r_{d-1,k-t-1} + 2r_{d-1,k-2t},$$
$$r_{d,k} \leq r_{d-1,k-2} + 2r_{d-1,k-t-1} + r_{d-1,k-2t},$$

или, если записать иначе,

$$(a\phi)^{2t} \leq (a\phi)^{2t-2} + (a\phi)^{t-1} + 2a^{\gamma}((a\phi)^{t-1} + 1), \tag{2.11}$$
$$a^{-\gamma}(a\phi)^{2t} \leq (a\phi)^{2t-2} + 2(a\phi)^{t-1} + 1. \tag{2.12}$$

Тем самым доказано, что соотношения (2.9) имеют место, если выполнены неравенства (2.11) и (2.12).

Теперь несложно получить оценку глубины функции f_k. Неравенство $\nu_d(f_k) > 2^d$ вытекает из $r_{d,k} \geq c'2^d$ при некоторой константе c'. Последнее имеет место при $d = k \cdot \frac{\log(a\phi)}{1+\gamma \log a} - O(1)$. Как следствие,

$$D_{B_M}(f_k) \geq \frac{\log(a\phi) \cdot k}{1 + \gamma \log a} - O(1) = \frac{1 + \log_\phi a}{1 + \gamma \log a} \log \Phi_k^{(t)} - O(1). \tag{2.13}$$

С целью максимизации оценки (2.13) выберем $t = 250$, при этом $\phi \approx 1.016596$. Тогда условия (2.11) и (2.12) выполнены, например, при $a = 1.00134$ и $\gamma = 8.93$. При подстановке выбранных значений параметров получаем $D_{B_M}(f_k) > 1.063 \log \Phi_k^{(t)} - O(1)$. □

2.3 Конструктивные верхние оценки для симметрических функций. Модулярный метод

2.3.1 Описание метода

Обозначим через $C_n^{(3)}(x_1,\ldots,x_n) = \left(C_{n,m-1}^{(3)},\ldots,C_{n,0}^{(3)}\right)$, $m = \lceil \log_3(n+1) \rceil$, булев $(n,2m)$-оператор вычисления арифметической суммы булевых переменных $x_1,\ldots,x_n$ в троичной системе счисления. Компонента $C_{n,i}^{(3)}$ является 2-битным кодом соответствующей цифры из троичной записи числа.

Лемма 2.3. *Пусть $2^k \cdot 3^l > n$. Для любого полного конечного базиса B справедливо*

$$L_B(C_n) \preceq 2^{O\left(\log^2\log n\right)} \left(\sum_{i=0}^{k-1} L_B(C_{n,i}) + \sum_{i=0}^{l-1} L_B\left(C_{n,i}^{(3)}\right) \right),$$
$$D_B(C_n) \le D_B\left(C_{n,k-1},\ldots,C_{n,0},C_{n,l-1}^{(3)},\ldots,C_{n,0}^{(3)}\right) + O(\log^2\log n).$$

Доказательство. Для вычисления C_n перепишем число $\left[C_{n,l-1}^{(3)},\ldots,C_{n,0}^{(3)}\right]$ в двоичной системе счисления и выполним восстановление результата из остатков по модулям 2^k и 3^l. Перезапись $O(\log n)$-разрядного числа из одной системы счисления в другую может быть реализована методом «деления пополам» А. Шёнхаге, см. [8, Гл. 14]. Следующий шаг реализуется при помощи сложения, вычитания, умножения и деления с остатком $O(\log n)$-разрядных чисел согласно тождеству

$$(\sigma \bmod 2^k\cdot 3^l) = (\sigma \bmod 2^k) + 2^k\left(\tau\left((\sigma \bmod 3^l) - (\sigma \bmod 2^k)\right) \bmod 3^l\right),$$

где константа τ определяется сравнением $\tau 2^k \equiv 1 \mod 3^l$. Все перечисленные арифметические операции реализуются с глубиной $O\left(\log^2\log n\right)$,[12] поэтому увеличивают сложность формул, реализующих входы, в $2^{O\left(\log^2\log n\right)}$ раз. □

[12] Глубина преобразования N-разрядных чисел из одной системы счисления в другую по порядку не более чем в $\log N$ раз превосходит глубину умножения. Деление с остатком сводится к двум умножениям и нескольким сложениям (см., например, [8, Гл. 16]), поэтому имеет глубину по порядку не большую, чем глубина умножения и сложения. Сложение и вычитание имеют глубину $O(\log N)$, см., например, [88, §10.1], [124, §2.5]. Умножение N-разрядных чисел выполняется с глубиной $O(\log N)$, например, методом компрессоров, см. также [88, §10.3].

Произвольную симметрическую функцию можно представить как функцию разрядов числа $\sigma_2 = \sigma \bmod 2^k$ и кода числа $\sigma_3 = \sigma \bmod 3^l$ и далее реализовать методом каскадов. Для эффективной реализации требуется предварительно выполнить специальную перекодировку числа σ_3. Троичная кодировка не подходит, потому что является избыточной. Переход к двоичной (как в лемме 2.3) устраняет избыточность, но приводит к нежелательному выравниванию сложности, и также глубины разрядов кода. Однако можно указать «компромиссную» кодировку, асимптотически оптимальную по длине и сохраняющую неравномерность сложностей и глубин разрядов, присущую исходному троичному представлению числа σ_3.[3] Для этого число следует разбить на достаточно длинные блоки и перекодировать их в двоичное представление.

Пусть $l = \lambda b$, где $\lambda, b \in \mathbb{N}$, и $b = \Theta(\log n / \log\log n)$. Перепишем число $\sigma_3 = \sigma \bmod 3^l$ в системе счисления с основанием 3^b, где цифры представляются двоичными числами. Обозначим компоненты оператора, вычисляющего σ_3 в указанной системе счисления, через $C^{(3^b)}_{n,i}$, $i = 0, \ldots, \lambda\mu - 1$, где $\mu = \lceil b \log_2 3 \rceil$. При этом, как следует из алгоритма смены системы счисления, см. выше в доказательстве леммы 2.3,

$$L_B\left(C^{(3^b)}_{n,j\mu}\right), \ldots, L_B\left(C^{(3^b)}_{n,j\mu+\mu-1}\right) \in 2^{O\left(\log^2 b\right)} \cdot \max_{0 \le i < b} L_B\left(C^{(3)}_{n,jb+i}\right),$$
$$D_B\left(C^{(3^b)}_{n,j\mu}\right), \ldots, D_B\left(C^{(3^b)}_{n,j\mu+\mu-1}\right) \le \max_{0 \le i < b} D_B\left(C^{(3)}_{n,jb+i}\right) + O(\log^2 b),$$

где B — произвольный полный конечный базис. Код для числа σ теперь будет состоять из $k + \lambda\mu = (1 + O(1)/b)\log n + O(1) = \log n + O(\log\log n)$ бит.

Обозначим через $R_n(x_1, ..., x_n) = (R_{n,n-1}, \ldots, R_{n,0})$ оператор упорядочения набора из n чисел: $R_{n,n-1} \ge \ldots \ge R_{n,0}$.

Лемма 2.4. *Пусть* $2^k \cdot 3^{\lambda b} > n$, $\mu = \lceil b \log_2 3 \rceil$, $s = k + \mu\lambda$, $B \in \{B_0, B_2\}$.

[3] Неравномерность имеет место, если σ_3 вычислено методом компрессоров, см. ниже.

Положим

$$(L_{s-1},\dots,L_0) = R_s\left(L_B(C_{n,k-1}),\dots,L_B(C_{n,0}),L_B\left(C^{(3^b)}_{n,\mu\lambda-1}\right),\dots,L_B\left(C^{(3^b)}_{n,0}\right)\right),$$

$$(D_{s-1},\dots,D_0) = R_s\left(D_B(C_{n,k-1}),\dots,D_B(C_{n,0}),D_B\left(C^{(3^b)}_{n,\mu\lambda-1}\right),\dots,D_B\left(C^{(3^b)}_{n,0}\right)\right).$$

Тогда

$$L_B(S_n) \preceq \sum_{i=0}^{s-1} 2^{s-i} L_i, \qquad D_B(S_n) \le \max_{0\le i<s}\{D_i + s - i\} + O\left(\sqrt{\log n}\right).$$

Доказательство. Симметрическая функция реализуется методом каскадов, исходя из описанного выше кода, как в доказательстве утверждения 2.1. □

Лемма 2.4 подходит для компьютерных расчетов, однако не очень удобна для аналитических. Приведем ее ослабленный вариант, приводящий к явной оценке сложности формул.

Лемма 2.5. *Пусть* $l = \lceil \alpha \log_3 n \rceil$, $k = \lceil (1-\alpha)\log_2 n \rceil$, $B \in \{B_0, B_2\}$. *Пусть выполнены соотношения*

$$L_B(C_{n,i}) \preceq 2^{\tau_1 \cdot i} n^{\omega_1}, \qquad L_B\left(C^{(3)}_{n,i}\right) \preceq 2^{\tau_2 \cdot i} n^{\omega_2},$$

а также $\tau_1(\eta-\alpha) > \eta$ *и* $\tau_2 \alpha \log_3 2 > \eta$ *при некотором* η, *где* $\alpha \le \eta \le 1$. *Тогда*

$$L_B(S_n) \preceq n^{\max\{\tau_1(1-\alpha)+\omega_1,\, \tau_2\alpha\log_3 2+\omega_2\}+o(1)}.$$

Доказательство. Представляем симметрическую функцию так же, как в лемме 2.4. Далее используем формулу разложения симметрической функции по переменным, в качестве которых подставляем последовательно чередуемые блоки из компонент оператора C_n и компонент оператора $C_n^{(3^b)}$ в порядке убывания номера разряда, где длины блоков соотносятся как $(\eta-\alpha)/\alpha$, до исчерпания кода числа σ_3 (код имеет длину $(\alpha+o(1))\log_2 n$). Оставшиеся $(1-\eta-o(1))\log_2 n$ младших разрядов оператора C_n подставляются вместо самых внутренних переменных в разложении.

Длина блока выбирается медленно растущей функцией от n. Неравенства $\tau_1(\eta-\alpha) \geq \eta$ и $\tau_2 \alpha \log_3 2 \geq \eta$ означают, что вклад компонент соответственно из двоичной и из троичной частей кода в сложность формулы при движении от старших разрядов к младшим убывает: сложность компоненты кода убывает быстрее, чем растет число ее вхождений в формулу. Поэтому сложность построенной формулы с точностью до множителя величины $n^{o(1)}$ определяется сложностью вычисления старших компонент кода: $C_{n,(1-\alpha)\log_2 n}$ и $C^{(3)}_{n,\alpha\log_3 n}$. □

Эффективный метод, позволяющий вычислять различные разряды суммы с различной формульной сложностью, предложен в [168]. Рассмотрим некоторый компрессор над базисом B. Через $x_{s,i}$ и $y_{s,i}$ обозначим соответственно входы и выходы, относящиеся к s-му разряду, $s \geq 0$. Через $\Phi(x)$ обозначим размер формулы, реализующей бит x (позволим $\Phi(x)$ принимать произвольное положительное вещественное значение). При любом s положим

$$A_s(p) = \sum_i \Phi(x_{s,i})^p - \sum_i \Phi(y_{s,i})^p, \tag{2.14}$$

где суммы по пустому множеству индексов считаются равными нулю. Из [168] извлекается

Лемма 2.6. *Пусть при некоторых $p \geq 1$, $\Phi(x_{s,i}) > 0$ и $\nu \geq 1$ выполнено*

$$\sum_s A_s(p)\nu^{-s} > 0. \tag{2.15}$$

Тогда $L_B(C_{n,l}) \preceq \left(\nu^l n\right)^{1/p+o(1)}$ (или $L_B\left(C^{(3)}_{n,l}\right) \preceq \left(\nu^l n\right)^{1/p+o(1)}$ в троичном случае).

Для полноты изложения доказательство леммы приведено далее в §2.3.5.

Аналогично получается результат для глубины. Пусть входы некоторого компрессора над базисом B, относящиеся к разряду s, имеют глубины $d^x_{s,i}$, а выходы, относящиеся к тому же разряду — $d^y_{s,i}$. Пусть r — максимальный номер разряда, к которому относятся выходы компрессора. Положим

$$a_s(\lambda) = \sum_i \lambda^{d^x_{s,i}} - \sum_i \lambda^{d^y_{s,i}},$$

где сумма по пустому множеству индексов считается равной нулю. Многочлен $a(\lambda; x) = \sum_{s=0}^{r} a_s(\lambda) x^{r-s}$ назовем *характеристическим многочленом* компрессора. Из [168] извлекается

Лемма 2.7. *Пусть при некоторых* $\lambda > 1$ *и* $\nu \geq 1$ *выполнено* $a(\lambda; \nu) > 0$. *Положим* $\delta_\mu = \log_\lambda 2 \cdot (\mu \log_2 \nu + 1)$ (*в троичном случае* $\delta_\mu^{(3)} = \log_\lambda 2 \cdot (\mu \log_3 \nu + 1)$). *Тогда* $D_B(C_{n,\mu\log_2 n}) \lesssim \delta_\mu \log_2 n$ (*в троичном случае* $D_B\left(C^{(3)}_{n,\mu\log_3 n}\right) \lesssim \delta_\mu^{(3)} \log_2 n$).

Доказательство этой леммы также приведено в §2.3.5.

Для демонстрации предложенной техники требуется указать эффективные по сложности или глубине компрессоры, двоичные и троичные. Для получения оценок мы воспользуемся известными конструкциями двоичных компрессоров [83, 171, 234] и простыми троичными компрессорами [219, 220].

2.3.2 Троичные компрессоры

В этом параграфе мы опишем реализации троичного $(4, 2)$-компрессора (т. е. простейшего полного компрессора). Компрессор вычисляет (троичную) сумму $[U_1, U_0]$ четырех чисел $X_1, X_2, X_3, X_4 \in \{0, 1, 2\}$, т. е. $3U_1 + U_0 = X_1 + X_2 + X_3 + X_4$.

Конструкция для оценки сложности в базисе B_0

При вычислениях в стандартном базисе используем монотонную кодировку. Троичная цифра $X \in \{0, 1, 2\}$ кодируется упорядоченной парой двоичных битов $(x^\wedge, x^\vee)$, $x^\wedge \leq x^\vee$, арифметическая сумма которых равна X (т. е. цифры 0, 1, 2 имеют соответственно коды 00, 01, 11). Обозначим коды входов X_k через $(x_k^\wedge, x_k^\vee)$, а коды выходов U_j — через $(u_j^\wedge, u_j^\vee)$.

Предварительно заметим, что биты $u_j^\wedge$ и $u_j^\vee$ выражаются через входы двойственными формулами: формула, реализующая $u_j^\wedge$, превращается в формулу для $u_j^\vee$ при замене $x_k^\wedge \leftrightarrow x_k^\vee$, $\vee \leftrightarrow \wedge$. Это можно проверить, если 1) рассмотреть x_k°, u_j°, $\circ \in \{\vee, \wedge\}$ как функции трехзначной логики от переменных X_i; 2) в таблице значений функций x_k°, u_j° заменить 1 на 2, вместо функций базиса B_0 подставить в формулу аналогичные функции, опреде-

ленные на наборах из нулей и двоек и принимающие значения из $\{0,2\}$; 3) применить троичный вариант принципа двойственности [95, §I.1.3], считая константы k и $2-k$ двойственными, $k\in\{0,1,2\}$.

Таким образом, достаточно построить формулы для $u_0^\wedge$ и $u_1^\vee$. Построение выполним стандартным способом выражения через пороговые функции. Обозначим $\chi_1^t=(X_1+X_2\ge t)$ и $\chi_2^t=(X_3+X_4\ge t)$ — это пороговые функции троичных переменных, заданных двоичными кодами. Тогда

$$\begin{gathered}u_1^\vee=\chi_1^3\vee\chi_1^2\chi_2^1\vee\chi_1^1\chi_2^2\vee\chi_2^3,\\ u_0^\wedge=(\chi_1^1\overline{\chi_1^2}\vee\chi_1^4)(\chi_2^1\overline{\chi_2^2}\vee\chi_2^4)\vee\chi_1^2\overline{\chi_1^3}(\chi_2^3\overline{\chi_2^4}\vee\overline{\chi_2^1})\vee\chi_2^2\overline{\chi_2^3}(\chi_1^3\overline{\chi_1^4}\vee\overline{\chi_1^1}).\end{gathered}\tag{2.16}$$

Функции χ_i^1 и χ_i^2 реализуются формулами

$$\chi_i^1=x_{2i-1}^\vee\vee x_{2i}^\vee,\qquad \chi_i^2=x_{2i-1}^\vee x_{2i}^\vee\vee\left(x_{2i-1}^\wedge\vee x_{2i}^\wedge\right),\tag{2.17}$$

а функции χ_i^4 и χ_i^3 — двойственными к ним в указанном выше смысле.

Если обозначить через $\Phi(X_i)$ и $\Phi(U_j)$ сложность формул, реализующих любой из битов кода X_i и U_j соответственно, то из (2.16) и (2.17) получаем

$$\begin{gathered}\Phi(U_0)\le 12(\Phi(X_1)+\Phi(X_2)+\Phi(X_3)+\Phi(X_4)),\\ \Phi(U_1)\le 5(\Phi(X_1)+\Phi(X_2)+\Phi(X_3)+\Phi(X_4)).\end{gathered}\tag{2.18}$$

Конструкция для оценки глубины в базисе B_0

Опишем эффективную по глубине модификацию компрессора с характеристическим многочленом $(2+2\lambda-\lambda^7)x-\lambda^5$. Считая, что входы X_1,X_2 имеют глубину 0, а входы X_3,X_4 — глубину 1, для вычисления кода числа U_i используем формулы

$$\begin{aligned}u_0^\wedge=\left(\left(\overline{\chi_1^2}\vee\chi_1^4\right)\left(\chi_1^1\chi_2^1\right)\right)\left(\overline{\chi_2^2}\vee\chi_2^4\right)\vee\\ \vee\left(\chi_1^2\overline{\chi_1^3}\,\overline{\chi_2^4}\right)\left(\chi_2^3\vee\overline{\chi_2^1}\right)\vee\left(\chi_2^2\overline{\chi_2^3}\right)\left(\left(\chi_1^3\vee\overline{\chi_1^1}\right)\overline{\chi_1^4}\right),\end{aligned}$$

$$u_1^\vee=\left(\left(\chi_1^3\vee x_3^\vee x_4^\wedge\right)\vee\chi_1^2\chi_2^1\right)\vee\left(\chi_1^1\vee x_3^\wedge x_4^\vee\right)\chi_2^2.$$

Функции χ_i^1 и χ_i^2 реализуются формулами (2.17), а функции χ_i^4 и χ_i^3 — двойственными формулами. Формулы для $u_0^\vee$ и $u_1^\wedge$ строятся двойственным образом по отношению к формулам для $u_0^\wedge$ и $u_1^\vee$.

Конструкция в базисе B_2

Для вычислений в базисе B_2 будем кодировать троичные цифры X тройками битов x_i^0, x_i^1, x_i^2, где $x_i^k = (X_i = k)$. Любые два из трех битов составляют неизбыточную кодировку. Аналогично обозначим биты кода U.

Из соображений двойственности (см. выше) достаточно реализовать только биты u_i^1 и u_i^2. При построении двойственных формул над B_2 дополнительно применяются правила $x_i^k \leftrightarrow x_i^{2-k}$ и $\oplus \leftrightarrow \sim$, где $\sim$ — операция эквивалентности.

Для упрощения восприятия и проверки формул определим троичные числа $[E_1, E_0] = X_1 + X_2$ и $[H_1, H_0] = X_3 + X_4$, и введем естественные обозначения e_i^k, h_i^k для битов кода. Тогда

$$\begin{aligned} u_0^k &= \left(e_0^k \sim h_0^0\right)\left(e_0^{k+1} \sim h_0^2\right), \\ u_1^1 &= u_0^1 \oplus \left(x_1^1 \oplus x_2^1 \oplus x_3^1 \oplus x_4^1\right), \\ u_1^2 &= \overline{e_1^0 \vee h_1^0} \vee \left(e_0^2\left(x_3^2 x_4^2\right) \vee h_0^2\left(x_1^2 x_2^2\right)\right). \end{aligned} \tag{2.19}$$

Вспомогательные функции реализуются формулами:

$$e_0^k = \left(x_1^k \sim x_2^0\right)\left(x_1^{k+1} \sim x_2^2\right), \qquad e_1^0 = x_1^2\overline{x_2^0} \vee x_1^1 x_2^2, \tag{2.20}$$

аналогично вычисляются h_i^k. Отметим, что формулы для u_0^k, e_0^k получаются из стандартной реализации оператора сложения по модулю 3, см. (2.35) ниже.

Обозначая через $\Phi(X_i)$ и $\Phi(U_j)$ сложность формул, реализующих любой из битов кода X_i и U_j соответственно, из (2.19) и (2.20) получаем

$$\begin{aligned} \Phi(U_0) &\le 4(\Phi(X_1) + \Phi(X_2) + \Phi(X_3) + \Phi(X_4)), \\ \Phi(U_1) &\le 5(\Phi(X_1) + \Phi(X_2) + \Phi(X_3) + \Phi(X_4)). \end{aligned} \tag{2.21}$$

Характеристический многочлен компрессора равен $(4 - \lambda^4)x - \lambda^5$ (все входы расположены на одной глубине).

2.3.3 Двоичные компрессоры

Оценки сложности выводятся при помощи (сложносоставных) компрессоров из [234].

Конструкция для оценки сложности в базисе B_0

В реализации компрессора над базисом B_0 некоторые пары битов v_1, v_2 кодируются тройками $\widehat{v} = (v^\wedge, v^\vee, v^\oplus)$, где $v^\wedge = v_1 v_2$, $v^\vee = v_1 \vee v_2$, $v^\oplus = v^\vee \overline{v^\wedge}$, а некоторые тройки битов v_1, v_2, v_3 кодируются четверками $\widetilde{v} = (v', v'', v''', v^\oplus)$:

$$\begin{gathered} v' = T_3^1(v_1, v_2, v_3) = v_1 \vee v_2 \vee v_3, \\ v'' = T_3^2(v_1, v_2, v_3) = v_1(v_2 \vee v_3) \vee v_2 v_3, \\ v''' = T_3^3(v_1, v_2, v_3) = v_1 v_2 v_3, \\ v^\oplus = v_1 \oplus v_2 \oplus v_3 = v_1(v_2 v_3 \vee \overline{v_2}\,\overline{v_3}) \vee \overline{v_1}(v_2 \overline{v_3} \vee \overline{v_2} v_3). \end{gathered} \tag{2.22}$$

Пара битов $[v'', v^\oplus]$ является двоичной записью суммы $v_1 + v_2 + v_3$.

На рис. 1 изображен $(17, 6)$-компрессор, функционирующий по правилу

$$x_1 + \ldots + x_{17} = y_1 + y_2 + y_3 + 2y_4 + 4q_2^\oplus + 8q_2''.$$

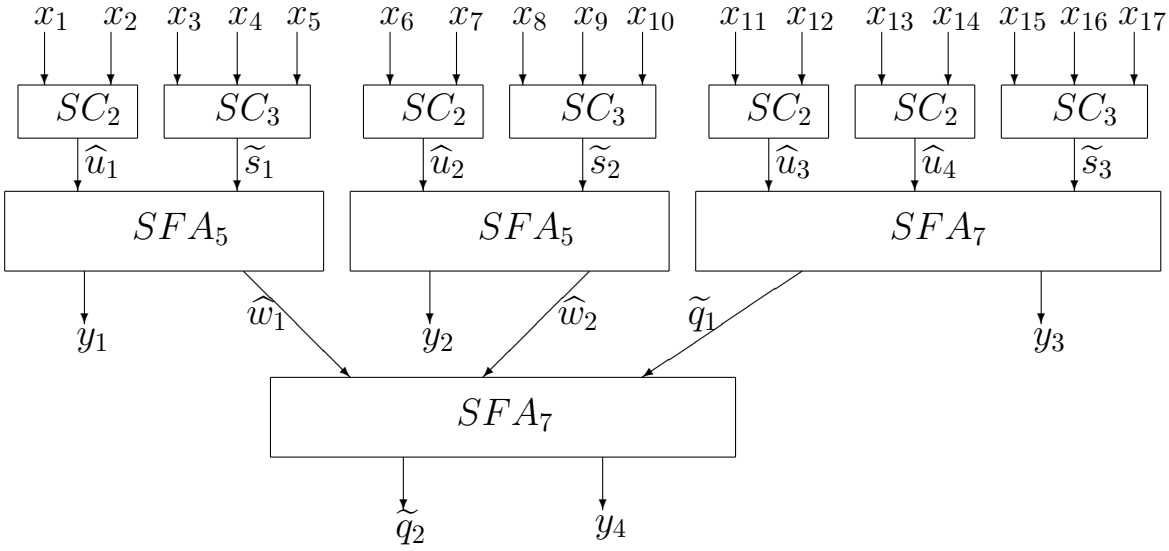


Рис. 1: Двоичный компрессор для базиса B_0

Блоки SC_2 и SC_3 выполняют перекодирование пар и троек входов соответственно в тройки типа $\widehat{v}$ и четверки типа $\widetilde{v}$ согласно приведенным выше формулам. Компрессор SFA_5 выполняет преобразование $(\widehat{u}, \widetilde{s}) \to (\widehat{w}, y)$ по правилу $u^\wedge + u^\vee + s' + s'' + s''' = y + 2(w^\wedge + w^\vee)$ согласно формулам

$$y = u^\oplus \overline{s^\oplus} \vee \overline{u^\oplus} s^\oplus, \qquad w^\vee = u^\wedge \vee s'' \vee u^\vee s', \qquad w^\wedge = u^\wedge s'' \vee u^\vee s'''. \tag{2.23}$$

Компрессор SFA_7 выполняет преобразование $(\widehat{u},\widehat{w},\widetilde{s}) \to (\widehat{q},y)$ по правилу

$$u^{\wedge}+u^{\vee}+w^{\wedge}+w^{\vee}+s'+s''+s'''=y+2(q'+q''+q''')$$

согласно формулам

$$\begin{gathered}
y = s^{\oplus}\left(u^{\oplus}w^{\oplus} \vee \overline{u^{\oplus}}\,\overline{w^{\oplus}}\right) \vee \overline{s^{\oplus}}\left(u^{\oplus}\overline{w^{\oplus}} \vee \overline{u^{\oplus}}w^{\oplus}\right), \\
q' = s'T_4^1(\widehat{u},\widehat{w}) \vee T_4^2(\widehat{u},\widehat{w}) \vee s'', \\
q'' = s'''T_4^1(\widehat{u},\widehat{w}) \vee s''T_4^2(\widehat{u},\widehat{w}) \vee s'T_4^3(\widehat{u},\widehat{w}) \vee T_4^4(\widehat{u},\widehat{w}), \\
q''' = s'''T_4^3(\widehat{u},\widehat{w}) \vee s''T_4^4(\widehat{u},\widehat{w}), \\
q^{\oplus} = \overline{s_2'}\,T_4^2(\widehat{u},\widehat{w})\overline{T_4^4(\widehat{u},\widehat{w})} \vee s_2'\overline{s_2''}\,T_4^1(\widehat{u},\widehat{w})\overline{T_4^3(\widehat{u},\widehat{w})} \;\vee \\
\vee\; s_2''\overline{s_2'''}\left(T_4^4(\widehat{u},\widehat{w}) \vee \overline{T_4^2(\widehat{u},\widehat{w})}\right) \vee s_2'''\left(T_4^3(\widehat{u},\widehat{w}) \vee \overline{T_4^1(\widehat{u},\widehat{w})}\right),
\end{gathered} \tag{2.24}$$

где

$$\begin{gathered}
T_4^1(\widehat{u},\widehat{w}) = u^{\vee} \vee w^{\vee}, \quad T_4^2(\widehat{u},\widehat{w}) = u^{\vee}w^{\vee} \vee u^{\wedge} \vee w^{\wedge}, \\
T_4^3(\widehat{u},\widehat{w}) = u^{\wedge}w^{\vee} \vee u^{\vee}w^{\wedge}, \quad T_4^4(\widehat{u},\widehat{w}) = u^{\wedge}w^{\wedge}.
\end{gathered}$$

Обозначая через $\Phi(x)$ сложность формулы, реализующей бит x, и полагая для удобства

$$\begin{gathered}
\Phi_1 = \Phi(x_1) = \Phi(x_2) = \Phi(x_3) = \Phi(x_6) = \Phi(x_7) = \Phi(x_8), \\
\Phi_2 = \Phi(x_4) = \Phi(x_5) = \Phi(x_9) = \Phi(x_{10}), \\
\Phi_3 = \Phi(x_{11}) = \Phi(x_{12}) = \Phi(x_{13}) = \Phi(x_{14}), \\
\Phi_4 = \Phi(x_{15}), \quad \Phi_5 = \Phi(x_{16}) = \Phi(x_{17}),
\end{gathered}$$

из (2.22), (2.23), (2.24) получаем соотношение:

$$\begin{pmatrix} \Phi(y_1) \\ \Phi(y_2) \\ \Phi(y_3) \\ \Phi(y_4) \\ \Phi(q_2^{\oplus}) \\ \Phi(q_2'') \end{pmatrix} \le \begin{pmatrix} 12 & 16 & 0 & 0 & 0 \\ 12 & 16 & 0 & 0 & 0 \\ 0 & 0 & 32 & 4 & 16 \\ 96 & 96 & 96 & 12 & 32 \\ 144 & 144 & 96 & 14 & 40 \\ 72 & 72 & 48 & 7 & 20 \end{pmatrix} \begin{pmatrix} \Phi_1 \\ \Phi_2 \\ \Phi_3 \\ \Phi_4 \\ \Phi_5 \end{pmatrix}. \tag{2.25}$$

Конструкция для оценки сложности в базисе B_2

В конструкции компрессора над базисом B_2 некоторые пары битов v_1, v_2 кодируются как[13] $\breve{v} = (v^0,\, v^{\oplus})$, где $v^0 = v_1$, $v^{\oplus} = v_1 \oplus v_2$. Мы используем $(15, 6)$-компрессор, изображенный на рис. 6, который получается обобщением конструкции из [171]. Он функционирует по правилу

$$16x_1+8(x_2+x_3+x_4)+4(x_5+x_6+x_7)+2(x_8+x_9+x_{10})+x_{11}+\ldots+x_{15} = \sum_{i=1}^{6} 2^{i-1} y_i.$$

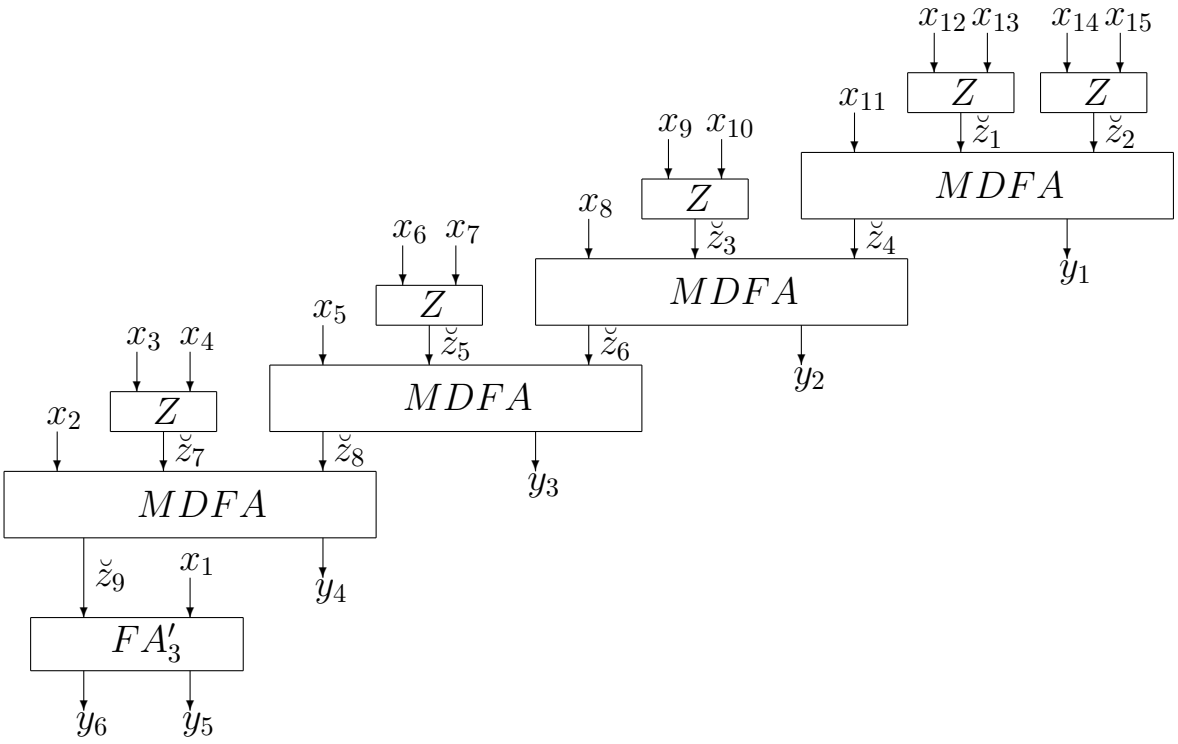


Рис. 2: Двоичный компрессор для базиса B_2

Блок Z выполняет кодирование $(v_1,\, v_2) \to \breve{v}$. Блок MDFA является $(5, 3)$-компрессором из работы [122], выполняющим преобразование $(x,\, \breve{u}_1,\, \breve{u}_2) \to (y,\, \breve{z})$ согласно условию

$$x + u_1^0 + (u_1^0 \oplus u_1^{\oplus}) + u_2^0 + (u_2^0 \oplus u_2^{\oplus}) = 2(z^0 + (z^0 \oplus z^{\oplus})) + y$$

[13] Такой способ кодирования применялся в [203].

по формулам

$$\begin{gathered} y = x \oplus u_1^{\oplus} \oplus u_2^{\oplus}, \quad z^0 = (x \oplus u_1^0)\, u_1^{\oplus} \oplus u_1^0, \\ z^{\oplus} = ((x \oplus u_1^0) \vee u_1^{\oplus}) \oplus (x \oplus u_1^{\oplus} \oplus u_2^0)\, \overline{u_2^{\oplus}}. \end{gathered} \tag{2.26}$$

Компрессор FA_3' совпадает с FA_3 с точностью до кодировки пары входов. Он выполняет преобразование $(x_1,\, \breve{z}_9) \to (y_5, y_6)$ по формулам

$$y_5 = x_1 \oplus z_9^{\oplus}, \qquad y_6 = (x_1 \oplus z_9^0)\, z_9^{\oplus} \oplus z_9^0. \tag{2.27}$$

Обозначая $\vec{\Phi} = (\Phi(x_1), \ldots, \Phi(x_{15}))^T$, из (2.26) и (2.27) получаем

$$\begin{pmatrix} \Phi(y_1) \\ \Phi(y_2) \\ \Phi(y_3) \\ \Phi(y_4) \\ \Phi(y_5) \\ \Phi(y_6) \end{pmatrix} \le \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 3 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 3 & 3 & 3 & 6 & 1 & 2 \\ 0 & 1 & 1 & 1 & 2 & 2 & 3 & 3 & 3 & 6 & 3 & 3 & 6 & 1 & 2 \\ 1 & 2 & 2 & 3 & 3 & 3 & 6 & 3 & 3 & 6 & 3 & 3 & 6 & 1 & 2 \\ 1 & 4 & 4 & 9 & 3 & 3 & 6 & 3 & 3 & 6 & 3 & 3 & 6 & 1 & 2 \end{pmatrix} \cdot \vec{\Phi}. \tag{2.28}$$

Конструкция для оценки глубины в базисе B_0

Для построения схемы над базисом B_0 используем модифицированный $(7,3)$-компрессор [81], вычисляющий арифметическую сумму $[y_2, y_1, y_0]$ семи булевых переменных $x_1, \ldots, x_7$. Модификация состоит в повышении глубины одного из входов[14]. Характеристический многочлен компрессора имеет вид $(6+\lambda^2-\lambda^6)x^2-\lambda^7 x-\lambda^6$, компоненты y_0 и y_1 вычисляются как в [81] по формулам

$$y_0 = \varphi(\varphi(\varphi(x_1,x_2),\varphi(x_3,x_4)),\varphi(\varphi(x_5,x_6),x_7)), \qquad \varphi(x,y) = x\overline{y} \vee \overline{x}y,$$

$$y_1 = \left(\left(\overline{\chi_1^1} \vee \chi_1^4\right)\chi_2^2 \vee \left(\chi_1^1\overline{\chi_1^2}\right)\left(\chi_2^1\overline{\chi_2^3}\right)\right) \vee \left(\left(\chi_1^2\overline{\chi_1^3}\right)\overline{\chi_2^2} \vee \chi_1^3\overline{\chi_1^4}\left(\overline{\chi_2^1} \vee \chi_2^3\right)\right),$$

а y_2 реализуется формулой

$$y_2 = \chi_1^2\left(\chi_2^2 \vee \chi_1^4\right) \vee \left(\chi_1^1\chi_2^3 \vee \chi_1^3\chi_2^1\right),$$

где $\chi_1^k = T_4^k(x_1,x_2,x_3,x_4)$, $\chi_2^k = T_3^k(x_5,x_6,x_7)$, глубина входа x_7 считается равной 2, глубина остальных входов — 0. Пороговые функции χ_i^k, $k \le 2$,

[14] Вероятно, этот же компрессор используется в [136].

реализуются стандартным образом:

$$\chi_1^1 = x_1 \vee x_2 \vee x_3 \vee x_4, \qquad \chi_1^2 = (x_1 \vee x_2)(x_3 \vee x_4) \vee (x_1x_2 \vee x_3x_4),$$
$$\chi_2^1 = (x_5 \vee x_6) \vee x_7, \qquad \chi_2^2 = (x_5 \vee x_6)x_7 \vee x_5x_6,$$

недостающие функции χ_i^k, $k > 2$, — двойственным образом, функция $\chi_2^2 \vee \chi_1^4$ — как $x_7(x_5 \vee x_6) \vee (x_5x_6 \vee x_1x_2x_3x_4)$.

Конструкция для оценки глубины в базисе B_2

Для построения схемы над базисом B_2 воспользуемся двоичным $(6,3)$-компрессором [171]. Компрессор выполняет преобразование $(x_1, \ldots, x_6) \to (y_2, y_1, y_0)$ с условием $4y_2 + 2y_1 + y_0 = 2x_6 + x_5 + \ldots + x_1$ и характеризуется многочленом $(4 + \lambda^2 - \lambda^3)x^2 + (\lambda^4 - \lambda^5)x - \lambda^6$. Компоненты компрессора вычисляются по формулам

$$z_1 = x_1 \oplus x_2 \oplus x_3 \oplus x_4, \qquad z_2 = (x_1 \oplus x_2)(x_3 \oplus x_4) \oplus (x_1x_2 \oplus x_3x_4),$$
$$z_4 = x_1x_2x_3x_4, \qquad w = x_5z_1 \oplus z_2,$$
$$y_0 = x_5 \oplus z_1, \qquad y_1 = x_6 \oplus w, \qquad y_2 = x_6w \oplus ((x_5z_1)z_2 \oplus z_4),$$

где глубина входов x_6 и x_5 полагается равной 4 и 2 соответственно, глубина остальных входов — 0.

2.3.4 Оценки

Проиллюстрируем применение метода §2.3.1 на материале описанных выше конструкций компрессоров.

Теорема 2.3. *Справедливы соотношения:*

$$L_{B_0}(C_n) \preceq n^{4.47}, \quad L_{B_0}(S_n) \preceq n^{4.48}, \quad L_{B_2}(C_n) \preceq n^{3.03}, \quad L_{B_2}(S_n) \preceq n^{3.04},$$
$$D_{B_0}(C_n) \lesssim 4.87 \log_2 n, \qquad D_{B_0}(S_n) \lesssim 4.88 \log_2 n,$$
$$D_{B_2}(C_n) \lesssim 3.34 \log_2 n, \qquad D_{B_2}(S_n) \lesssim 3.34 \log_2 n.$$

Доказательство. Начнем с оценок глубины. Ограничимся случаем базиса B_2; оценки в базисе B_0 получаются аналогично.

Применяя лемму 2.7 к конструкции двоичного компрессора с выбором параметра $\lambda = 1.409$ (при этом $x_\lambda \approx 1.8397$), получаем оценку

$$D_{B_2}(C_{n,i}) \lesssim 1.778i + 2.0215 \log_2 n. \tag{2.29}$$

Для троичного компрессора при выборе $\lambda = 1.3342$ (при этом $x_\lambda \approx 5.0858$) получаем оценку

$$D_{B_2}\left(C^{(3)}_{n,i}\right) \lesssim 5.6409i + 2.404 \log_2 n. \tag{2.30}$$

При помощи леммы 2.3 с выбором $k \approx 0.739 \log_2 n$ и $l \approx 0.261 \log_3 n$ устанавливаем заявленную оценку для $D_{B_2}(C_n)$. Доказательство оценки глубины для класса симметрических функций иллюстрирует рис. 3.

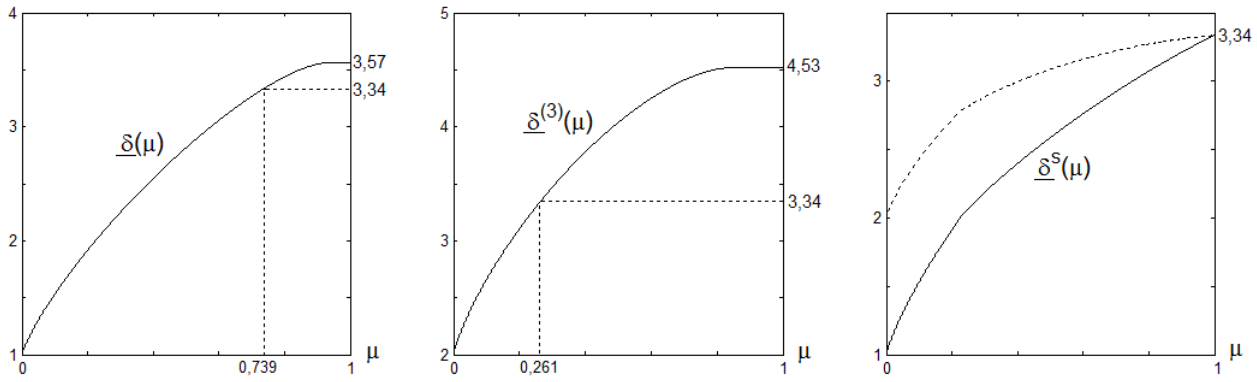


Рис. 3: Графики функций $\underline{\delta}(\mu)$, $\underline{\delta}^{(3)}(\mu)$, $\underline{\delta}^s(\mu)$.

Слева на рис. 3 изображен график функции $\underline{\delta}(\mu) = \min_\lambda \delta(\lambda, \mu)$, которая определяется из леммы 2.7 для двоичного $(6, 3)$-компрессора. Напомним, что величина $(\underline{\delta}(\mu) + o(1)) \log_2 n$ служит верхней оценкой глубины разряда $C_{n,\mu \log_2 n}$. Второй график изображает аналогично определяемую функцию $\underline{\delta}^{(3)}(\mu)$ для троичного компрессора[15]. (Отметим, что линейные оценки (2.29) и (2.30) изображались бы касательными к графикам функций $\underline{\delta}(\mu)$ и $\underline{\delta}^{(3)}(\mu)$ в точках с абсциссами 0.739 и 0.261 соответственно.)

Третий график изображает функцию $\underline{\delta}^s(\mu) = D_{\mu \log_2 n}/\log_2 n$ из леммы 2.4: величина $(\underline{\delta}^s(\mu) + o(1)) \log_2 n$ служит верхней оценкой глубины для

[15]Графики построены при помощи ЭВМ.

$\mu \log_2 n$ битов кода входного набора, используемого в лемме 2.4 при реализации симметрической функции. График для $\underline{\delta}^s$ получается сортировкой участков графиков $\underline{\delta}([0, 0.739])$ и $\underline{\delta}^{(3)}([0, 0.261])$. Пунктиром показан график функции $\underline{\delta}^s(\mu) + 1 - \mu$, максимум которой определяет глубину реализации симметрических функций методом каскадов (см. лемму 2.4). Так получается заявленная оценка для $D_{B_2}(S_n)$.

Приведенная методика расчета оценки глубины класса S_n существенно использует ЭВМ, поэтому не является вполне строгим обоснованием. Впрочем, требуемые оценки (во всяком случае, в том приближенном виде, в котором они указаны) могут быть выведены аналитически. Для этого мы каждую из функций $\underline{\delta}(\mu)$ и $\underline{\delta}^{(3)}(\mu)$ мажорируем кусочно-линейной функцией, составленной из участков линейных приближений вида (2.29) и (2.30) (одной пары приближений (2.29), (2.30) недостаточно, но достаточно, например, трех таких пар). После этого подходящее кусочно-линейное приближение для $\underline{\delta}^s$ и, как следствие, для $\underline{\delta}^s(\mu)+1-\mu$, может быть построено явно. Соответствующую выкладку мы здесь не приводим.

Докажем оценки сложности для базиса B_0. Из (2.18) и леммы 2.6 при выборе $\Phi(x_1) = \ldots = \Phi(x_4)$, $p = 0.2812$ и $\nu = 2.255$ следует $L_{B_0}\left(C^{(3)}_{n,l}\right) \preceq 2^{4.172 \cdot l} n^{3.5562}$.

Из (2.25) и леммы 2.6 при выборе $\Phi_1 = 0.45$, $\Phi_2 = 0.21$, $\Phi_3 = 0.4$, $\Phi_4 = 1$, $\Phi_5 = 0.5$, $p = 0.271$ и $\nu = 1.2511$ следует $L_{B_0}(C_{n,l}) \preceq 2^{1.193 \cdot l} n^{3.6901}$.

Из двух приведенных оценок и леммы 2.3 при выборе $k = \lceil 0.6531 \log_2 n \rceil$ и $l = \lceil 0.3469 \log_3 n \rceil$ вытекает $L_{B_0}(C_n) \preceq n^{4.47}$.

Для реализации симметрических функций используем соотношения

$$L_{B_0}\left(C^{(3)}_{n,l}\right) \preceq 2^{4.776 \cdot l} n^{3.4341}, \qquad L_{B_0}(C_{n,l}) \preceq 2^{1.5703 \cdot l} n^{3.4543}. \tag{2.31}$$

Первое вытекает из (2.18) и леммы 2.6 при подстановке $\Phi(x_1) = \ldots = \Phi(x_4)$, $p = 0.2912$ и $\nu = 2.622$. Второе вытекает из (2.25) и леммы 2.6 при подстановке $\Phi_1 = 0.45$, $\Phi_2 = 0.2$, $\Phi_3 = 0.4$, $\Phi_4 = 1$, $\Phi_5 = 0.5$, $p = 0.2895$ и $\nu = 1.3704$.

В отличие от разобранного выше примера с оценкой глубины, для вывода требуемой оценки сложности нам достаточно одного линейного приближения (2.31). В этой ситуации можно применить лемму 2.5; значения

параметров определим как $\alpha = 0.3469$ и $\eta = 1$.

Перейдем к базису B_2. Из (2.21) и леммы 2.6 при выборе $\Phi(x_1) = \ldots = \Phi(x_4)$, $p = 0.4325$ и $\nu = 5.3513$ следует $L_{B_2}\left(C_{n,l}^{(3)}\right) \preceq 2^{5.5951 \cdot l} n^{2.3122}$.

Из (2.28) и леммы 2.6 при выборе $\Phi(x_1) = 10$, $\Phi(x_2) = \Phi(x_3) = 1.27$, $\Phi(x_4) = 0.55$, $\Phi(x_5) = \Phi(x_6) = 0.53$, $\Phi(x_7) = 0.26$, $\Phi(x_8) = \Phi(x_9) = 0.19$, $\Phi(x_{10}) = 0.09$, $\Phi(x_{11}) = \Phi(x_{12}) = 0.066$, $\Phi(x_{13}) = 0.033$, $\Phi(x_{14}) = 0.15$, $\Phi(x_{15}) = 0.08$, $p = 0.444$ и $\nu = 1.3479$ следует $L_{B_2}(C_{n,l}) \preceq 2^{0.9701 \cdot l} n^{2.2523}$.

Из указанных двух соотношений и леммы 2.3 при выборе $k = \lceil 0.7978 \log_2 n \rceil$ и $l = \lceil 0.2022 \log_3 n \rceil$ вытекает $L_{B_2}(C_n) \preceq n^{3.03}$.

Для реализации симметрических функций воспользуемся соотношениями

$$L_{B_2}\left(C_{n,l}^{(3)}\right) \preceq 2^{6.3776 \cdot l} n^{2.22718}, \qquad L_{B_2}(C_{n,l}) \preceq 2^{1.33293 \cdot l} n^{1.9763}.$$

Первое вытекает из (2.21) и леммы 2.6 при подстановке $\Phi(x_1) = \ldots = \Phi(x_4)$, $p = 0.449$ и $\nu = 7.278$. Второе вытекает из (2.28) и леммы 2.6 при подстановке $\Phi(x_1) = 10$, $\Phi(x_2) = \Phi(x_3) = 1.27$, $\Phi(x_4) = \Phi(x_5) = \Phi(x_6) = 0.53$, $\Phi(x_7) = 0.26$, $\Phi(x_8) = \Phi(x_9) = 0.19$, $\Phi(x_{10}) = 0.09$, $\Phi(x_{11}) = \Phi(x_{12}) = 0.068$, $\Phi(x_{13}) = 0.033$, $\Phi(x_{14}) = 0.16$, $\Phi(x_{15}) = 0.08$, $p = 0.506$ и $\nu = 1.596$. Применяем лемму 2.5 с параметрами $\alpha = 0.202$ и $\eta = 0.8128$. □

Хотя приведенные численные оценки служат иллюстрацией эффективности модулярного метода, они далеко не окончательные, и могут быть улучшены путем конструирования более удачных двоичных или троичных компрессоров или же путем привлечения схем сложения в других системах счисления (пятиричной, семиричной и т.д.). Вопрос о перспективности применения других систем счисления косвенно затрагивается в разделе 2.5.

На качественном уровне, применение троичного компрессора позволило несколько уточнить оценки для оператора C_n, а оценки для класса симметрических функций улучшить более существенно: причина в спрямлении графика сложности (глубины) разрядов кода (сравните графики $\underline{\delta}(\mu)$ и $\underline{\delta}^s(\mu)$ на рис. 3). Заметим, что отдельно метод [168], применяемый вместе с рассмотренными двоичными компрессорами, позволил бы указать оценки $D_{B_2}(S_n) \lesssim 3.66 \log_2 n$, $D_{B_0}(S_n) \lesssim 5.16 \log_2 n$.

2.3.5 Приложение. Доказательство технических лемм

Доказательство леммы 2.6. По построению, функция размера $\Phi(y_{s,j})$ формулы выхода компрессора является линейной комбинацией размеров формул входов с неотрицательными целочисленными коэффициентами. Поэтому неравенство (2.15) остается в силе при пропорциональном изменении всех $\Phi(x_{s,i})$. Без ограничения общности можем считать, что $\min\{\Phi(x_{s,i})\} = 1$.

В силу непрерывной зависимости $\Phi(y_{s,j})$ от $\Phi(x_{s,i})$ найдется такое $\delta > 0$, что неравенства (2.15) остаются справедливыми при подстановке в (2.14) параметров $\Phi'_{s,i} \in [\Phi(x_{s,i})-\delta,\ \Phi(x_{s,i})]$ и $\Psi'_{s,i} \in [\Phi(y_{s,i}),\ \Phi(y_{s,i})+\delta]$ вместо соответствующих $\Phi(x_{s,i})$ и $\Phi(y_{s,i})$. Тогда найдется (достаточно малое) $\lambda > 1$, для которого существуют $d^x_{s,i}, d^y_{s,i} \in \mathbb{Z}$ такие, что $\lambda^{d^x_{s,i}/p} \in [\Phi(x_{s,i})-\delta,\ \Phi(x_{s,i})]$ и $\lambda^{d^y_{s,i}/p} \in [\Phi(y_{s,i}),\ \Phi(y_{s,i})+\delta]$ для всех s, i. Назовем число $d^x_{s,i}$ (соответственно $d^y_{s,i}$) уровнем входа $x_{s,i}$ (выхода $y_{s,i}$). Можно считать, что $\min\{d^x_{s,i}\} = 0$. Обозначим $d = \max\{d^y_{s,i}\}$. Также обозначим через s' максимальный номер разряда, к которому относится выход компрессора.

Формулу, реализующую оператор C_n (аналогично $C^{(3)}_n$), построим по следующему шаблону. Формула состоит из компрессоров, расположенных на различных уровнях и относящихся к различным разрядам: входами компрессоров могут быть входы формулы, выходы компрессоров, расположенных на уровнях с меньшими номерами, а также тождественно нулевые формулы. Далее ограничимся рассмотрением двоичного случая.

Пусть компрессор разряда l и уровня k принимает входы разряда $s+l$ на уровнях $d^x_{s,i}+k$ и производит выходы разряда $s+l$ на уровнях $d^y_{s,i}+k$. Рассмотрим формулу, в которой при любом l, $-s' \le l \le \log_2 n+1$, на уровне k, $0 \le k \le \log_\lambda(\nu^l n)$, расположено $\lfloor c\nu^l n\lambda^{-k} \rfloor$ компрессоров разряда l, где c — некоторая константа, которая будет определена позднее. Ненулевые входы формулы располагаются в нулевом разряде на уровнях d и выше.

Оценим число входов и выходов формулы, относящихся к фиксированному разряду l, в зависимости от уровня k. По построению, все выходы формулы на уровнях меньше d являются нулевыми. Нулевыми являются и все входы в отрицательных разрядах. Суммарное число входов (все они

нулевые) на тех же уровнях есть $O(n)$. Если $d \le k \le \log_\lambda(\nu^l n)$, то разность между числом входов и числом выходов разряда $l \ge 0$ и уровня k компрессоров внутри формулы есть

$$\sum_{s,i} \left\lfloor c\nu^{l-s} n \lambda^{d^x_{s,i}-k} \right\rfloor - \sum_{s,i} \left\lfloor c\nu^{l-s} n \lambda^{d^y_{s,i}-k} \right\rfloor = \\ = c\nu^l n \lambda^{-k} \sum_s A_s(p) \nu^{-s} \pm O(1) = \Theta(\nu^l n \lambda^{-k}) \pm O(1),$$

следовательно, формула на уровне k имеет не более $O(1)$ выходов разряда l (т. е. таких выходов компрессоров, которые не присоединяются ко входам других компрессоров). На уровнях выше $\log_\lambda(\nu^l n)$ формула суммарно принимает и производит $O(1)$ входов и выходов разряда l. Таким образом, формула в каждом разряде производит $O(\log n)$ выходов. Подходящий выбор константы c обеспечивает не менее n входов в нулевом разряде.

Согласно выбору λ, размер формул на уровне k оценивается сверху как $\lambda^{k/p}$. Поэтому формулы, реализующие выходы, относящиеся к разряду l, имеют сложность не выше $\lambda^{(\log_\lambda(\nu^l n)+O(1))/p} = O((\nu^l n)^{1/p})$.

Заключительное сложение $O(\log n)$ чисел (составленных из выходов формулы в разрядах не старше l-го) можно реализовать произвольной формулой полиномиальной сложности, поэтому окончательно сложность вычисления l-го разряда оператора C_n оценивается как $O\left((\nu^l n)^{1/p} \log^{O(1)} n\right)$. □

Доказательство леммы 2.7. Схема доказательства в целом такая же, как у предыдущей леммы, при этом рассуждение проще (ограничимся случаем двоичных компрессоров).

Пусть $d^x_{s,i}$ и $d^y_{s,i}$ обозначают соответственно глубину i-го входа и i-го выхода компрессора, относящегося к разряду s. Пусть $\min\{d^x_{s,i}\} = 0$, положим $d = \max\{d^y_{s,i}\}$. Обозначим через s' максимальный номер разряда, к которому относится выход компрессора.

Дальше действуем как в доказательстве леммы 2.6. Пусть компрессор разряда l и уровня k принимает входы разряда $s + l$ на глубинах $d^x_{s,i} + k$ и производит выходы разряда $s + l$ на глубинах $d^y_{s,i} + k$. Рассмотрим формулу, в которой при любом l, $-s' \le l \le \log_2 n + 1$, на глубине k, $0 \le k \le \log_\lambda(c\nu^l n)$, расположено $\left\lfloor c\nu^l n \lambda^{-k} \right\rfloor$ компрессоров разряда l. Будем

считать, что ненулевые входы формулы располагаются в нулевом разряде на глубине d.

Оценка разности между числом входов и числом выходов произвольного разряда $l \geq 0$ и глубины k, $d \leq k \leq \log_\lambda(c\nu^l n)$, компрессоров внутри формулы принимает вид

$$\sum_{s,i} \left\lfloor c\nu^{l-s} n \lambda^{d^x_{s,i}-k} \right\rfloor - \sum_{s,i} \left\lfloor c\nu^{l-s} n \lambda^{d^y_{s,i}-k} \right\rfloor = c\nu^l n \lambda^{-k} a(\lambda;\nu) \pm O(1).$$

Повторяя рассуждение из доказательства леммы 2.6, получаем, что в построенной формуле все выходы, относящиеся к разрядам не выше l-го, вычисляются на глубине $\log_\lambda(\nu^l n) + O(1)$ — эти выходы могут быть сгруппированы в $O(\log n)$ слагаемых чисел. Окончательно, младшие $l+1$ разрядов суммы вычисляются с дополнительной глубиной $O(\log\log n + \log l)$. $\square$

2.4 Неконструктивные верхние оценки для симметрических функций. Метод приближений

2.4.1 Формулы для приближенного суммирования

Обозначим через $\Psi_n^{k,t}$ частично определенную[16] пороговую функцию n переменных с порогом k и интервалом неопределенности радиуса t:

$$\Psi_n^{k,t}(x_1,\ldots,x_n) = 1 \iff \sum x_i \geq k+t,$$
$$\Psi_n^{k,t}(x_1,\ldots,x_n) = 0 \iff \sum x_i \leq k-t.$$

Как следует из [207], такие функции реализуются сравнительно простыми монотонными формулами. Обозначим $\alpha = \frac{3-\sqrt{5}}{2}$.

Лемма 2.8 (Вэльянт [207]). *Пусть заданы частично определенная булева функция f от n переменных и вероятностное распределение Φ на множестве формул этих же переменных в базисе $B_M = \{\vee, \wedge\}$, такое, что для любого набора $x \in f^{-1}(0)$ выполнено неравенство*

$$\mathbf{P}(F(x) = 1 \mid F \in \Phi) \leq \alpha - \varepsilon,$$

[16] Под частично определенной булевой функцией здесь и далее понимается произвольная булева функция, имеющая предписанные значения на области определения.

а для любого набора $x \in f^{-1}(1)$ *выполнено неравенство*

$$\mathbf{P}(F(x) = 0 \mid F \in \Phi) \leq 1 - \alpha - \varepsilon$$

при некотором $\varepsilon > 0$. *Тогда* $D_{B_M}(f) \leq 2\left(\log_{4\alpha}(1/\varepsilon) + \log_2 n\right) + O(1)$.

Зададим вероятностное распределение Φ примерно так же, как в [207]:

$$\mathbf{P}(F \equiv x_i) = \frac{1}{3n},\ 1 \leq i \leq n,\ \ \mathbf{P}(F \equiv 1) = \alpha - \frac{k}{3n},\ \ \mathbf{P}(F \equiv 0) = \frac{2}{3} - \alpha + \frac{k}{3n}.$$

Легко проверяются неравенства

$$\mathbf{P}\left(F(x) = 1 \,\middle|\, \sum x_i \leq k - t\right) \leq \alpha - \frac{t}{3n},$$
$$\mathbf{P}\left(F(x) = 0 \,\middle|\, \sum x_i \geq k + t\right) \leq 1 - \alpha - \frac{t}{3n}.$$

Таким образом, получаем

Следствие 2.2. *Справедливо неравенство*

$$D_{B_M}\left(\Psi_n^{k,t}\right) \leq 2\left(\log_{6-2\sqrt{5}}(n/t) + \log_2 n\right) + O(1).$$

Пусть $n = (2^r - 1)(2t-1) - 1$. Разобьем отрезок $[0, n]$ на $2^r - 1$ интервалов $I_j = [(j-1)(2t-1),\ j(2t-1) - 1]$. Пусть $X_j = \Psi_n^{j(2t-1)-t,t}(x)$ — пороговая функция с интервалом неопределенности I_j. Положим $X_{2^r} = 0$.

Номер интервала $J = (J_{r-1}, \ldots, J_0)$, в который попадает арифметическая сумма булевых переменных $\sigma = x_1 + \ldots + x_n$, можно вычислить из X_j с точностью до ± 1, например, следующим образом.

Величина $g_j = X_{j-1} \cdot \overline{X_{j+1}}$ имеет смысл признака попадания суммы в окрестность интервала I_j. Действительно, если $\sigma \in I_j$, то $g_j = 1$; если $g_j = 1$, то $\sigma \in I_{j-1} \cup I_j \cup I_{j+1}$. Положим

$$Z_i = \bigvee_{k=1}^{2^{r-i-1}} X_{2^{i+1}k - 2^i} \cdot \overline{X_{2^{i+1}k}}, \qquad G_i = \bigvee_{k=1}^{2^{r-i}-1} g_{2^i k}, \qquad i = 2, \ldots, r-1.$$

Функция Z_i имеет смысл предположительного значения разряда J_i; ее область неопределенности — интервалы $I_{2^i k}$. Функция G_i имеет смысл попадания σ в окрестность области неопределенности функции Z_i. Определим

$$J_{r-1} = G_{r-1} \vee Z_{r-1}, \quad J_i = \overline{G_{i+1}}(G_i \vee Z_i),\ \ i = 2, \ldots, r-2, \quad J_1 = \overline{G_2},\ \ J_0 = 0,$$

что означает: при попадании в область неопределенности номер заканчивается на $10\ldots0$. Теперь несложно проверяется

Лемма 2.9. *Оператор J удовлетворяет условию $\sigma \in I_{J-1} \cup I_J \cup I_{J+1}$. Кроме того, для глубины компонент оператора J как функций от переменных X_j справедлива оценка $D_{B_0}(J_i) \le r - i + O(1)$.*

Доказательство. Если для истинного номера J^* интервала, которому принадлежит σ, выполнено сравнение $J^* \equiv 2 \mod 4$, то этот номер точно вычисляется функциями J_i (поскольку $G_i = 0$ при всех i). Если $J^* \equiv 0 \mod 4$, то номер также вычисляется точно: интервал I_{J^*} фиксируется условием $G_i = 1$, $G_{i+1} = 0$, где число i таково, что $J^* \equiv 2^i \mod 2^{i+1}$. Если J^* нечетно, то J есть номер одного из соседних интервалов, какого именно — зависит от того, выполнялось ли условие $G_i = 1$ при некотором i. Оценка глубины непосредственно вытекает из вида применяемых формул. □

Для оценки сложности построенных формул используем тривиально выполняемое в бинарных базисах B соотношение $L_B(f) \le 2^{D_B(f)}$.

2.4.2 Формулы для симметрических функций

Через $J_n^t(x_1, \ldots, x_n) = \left(J_{n,r-1}^t, \ldots, J_{n,0}^t\right)$, $r = \left\lceil \log_2 \left(\frac{n+1}{2t-1} + 1\right)\right\rceil$, обозначим построенный выше (n, r)-оператор вычисления номера интервала, в котором находится арифметическая сумма σ переменных. Следующая лемма расширяет лемму 2.3.

Лемма 2.10. *Пусть $2^k \cdot 3^l \ge 6t - 3$. Тогда для любого полного конечного базиса B справедливы оценки*

$$L_B(C_n) \le 2^{O\left(\log^2 \log n\right)} \cdot L_B\left(J_n^t, C_{n,k-1}, \ldots, C_{n,0}, C_{n,l-1}^{(3)}, \ldots, C_{n,0}^{(3)}\right),$$
$$D_B(C_n) \le D_B\left(J_n^t, C_{n,k-1}, \ldots, C_{n,0}, C_{n,l-1}^{(3)}, \ldots, C_{n,0}^{(3)}\right) + O\left(\log^2 \log n\right).$$

Доказательство. При помощи Китайской теоремы об остатках находится остаток $\sigma \bmod 2^k \cdot 3^l$, затем находится единственное число в интервале $I_{J-1} \cup I_J \cup I_{J+1}$, где $J = J_n^t$, имеющее такой остаток. Вычисление состоит из нескольких простых арифметических действий с $\log n$-разрядными числами, см. доказательство леммы 2.3. □

Напомним, что при реализации симметрических функций в качестве кода аргумента используется номер интервала, разряды числа $\sigma_2 = \sigma \bmod 2^k$ и разряды числа $\sigma_3 = \sigma \bmod 3^l$, предварительно переписанного в системе счисления с основанием 3^b, где $b = \Theta(\log n / \log\log n)$, и представлением цифр двоичными числами, см. §2.3.1. Обозначаем компоненты оператора, вычисляющего σ в указанной системе счисления, через $C_{n,i}^{(3^b)}$.

Напомним, что через $R_n(x_1, ..., x_n) = (R_{n,n-1}, \ldots, R_{n,0})$ обозначается оператор упорядочения набора из n чисел: $R_{n,n-1} \geq \ldots \geq R_{n,0}$. Следующая лемма обобщает лемму 2.4.

Лемма 2.11. *Пусть* $2^k \cdot 3^{\lambda b} \geq 6t - 3$, $\nu = \lceil b \log_2 3 \rceil$, $r = \left\lceil \log_2 \left(\frac{n+1}{2t-1} + 1 \right) \right\rceil$, $s = k + \nu\lambda + r$. *Пусть* $B \in \{B_0, B_2\}$. *Положим*

$$(L_{s-1}, \ldots, L_0) = R_s \Big(L_B(J^t_{n,r-1}), \ldots, L_B(J^t_{n,0}), \\ L_B(C_{n,k-1}), \ldots, L_B(C_{n,0}), L_B\left(C^{(3^b)}_{n,\nu\lambda-1}\right), \ldots, L_B\left(C^{(3^b)}_{n,0}\right)\Big),$$

$$(D_{s-1}, \ldots, D_0) = R_s \Big(D_B(J^t_{n,r-1}), \ldots, D_B(J^t_{n,0}), \\ D_B(C_{n,k-1}), \ldots, D_B(C_{n,0}), D_B\left(C^{(3^b)}_{n,\nu\lambda-1}\right), \ldots, D_B\left(C^{(3^b)}_{n,0}\right)\Big).$$

Тогда

$$L_B(S_n) \leq 2^{O\left(\log^2 \log n\right)} \left(\sum_{i=0}^{s-1} 2^{s-i} L_i \right),$$

$$D_B(S_n) \leq \max_{0 \leq i < s} \{D_i + s - i\} + O\left(\sqrt{\log n}\right).$$

Доказательство. Симметрическая функция реализуется методом каскадов, исходя из описанного выше кода, как в доказательстве утверждения 2.1. □

Применяем леммы 2.9–2.11, следствие 2.2 и формулы из §2.3.2, §2.3.3 для реализации младших разрядов операторов C_n, $C_n^{(3)}$. Аналогично теореме 2.3 доказывается

Теорема 2.4. *Имеют место оценки:*

$$L_{B_0}(C_n) \preceq n^{3.91}, \quad L_{B_0}(S_n) \preceq n^{4.01}, \quad L_{B_2}(C_n) \preceq n^{2.84}, \quad L_{B_2}(S_n) \preceq n^{2.95},$$
$$D_{B_0}(C_n) \lesssim 4.14 \log_2 n, \qquad D_{B_2}(C_n) \lesssim 3.02 \log_2 n,$$
$$D_{B_0}(S_n) \lesssim 4.24 \log_2 n, \qquad D_{B_2}(S_n) \lesssim 3.1 \log_2 n.$$

Доказательство. Рассуждение проводится как в доказательстве теоремы 2.3, только код для симметрических функций составляется из трех частей. Ограничимся пояснением к оценкам глубины функций в базисе B_2.

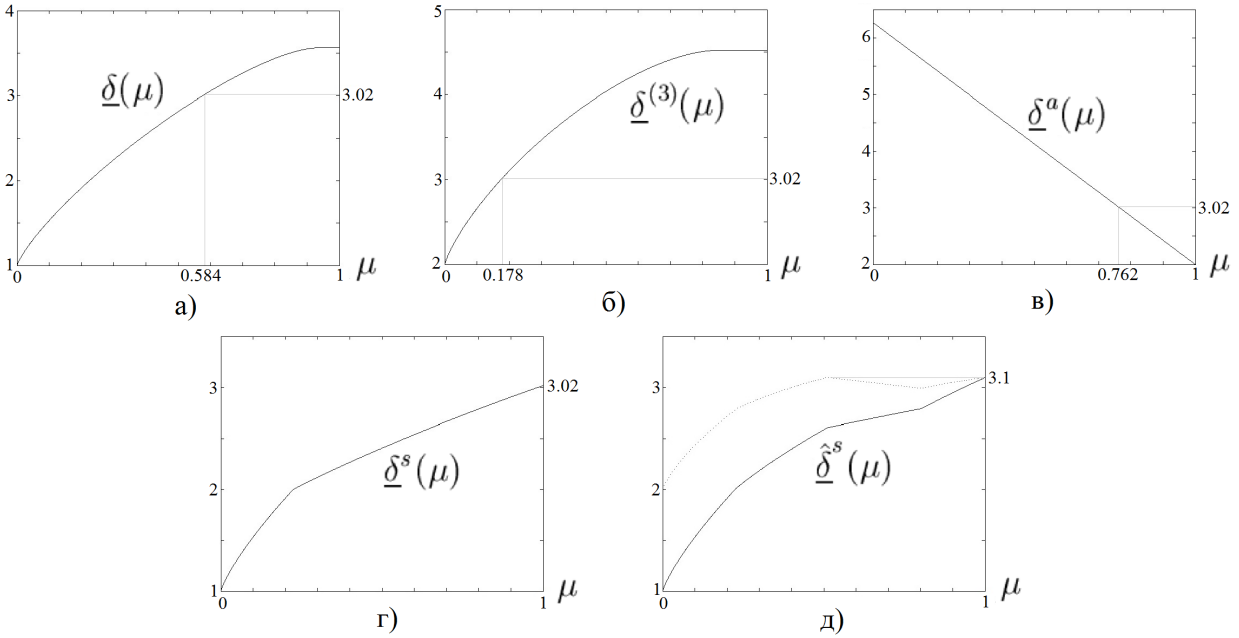


Рис. 4: Графики функций $\underline{\delta}(\mu)$, $\underline{\delta}^{(3)}(\mu)$, $\underline{\delta}^{a}(\mu)$, $\underline{\delta}^{s}(\mu)$, $\hat{\underline{\delta}}^{s}(\mu)$.

На рис. 4а мы повторяем график функции $\underline{\delta}(\mu)$, такой, что величина $(\underline{\delta}(\mu) + o(1)) \log_2 n$ служит верхней оценкой глубины разряда $C_{n,\mu \log_2 n}$. На рис. 4б изображен график аналогично определяемой функции $\underline{\delta}^{(3)}(\mu)$. На рис. 4в изображен график функции

$$\underline{\delta}^{a}(\mu) = 2 + (1 - \mu)\left(1 + \log_{6-2\sqrt{5}} 2\right),$$

такой, что арифметическая сумма σ переменных $x_1, \ldots, x_n$ может быть вычислена с точностью $\pm n^{\mu + o(1)}$ с глубиной $\underline{\delta}^{a}(\mu) \log_2 n$ методом леммы 2.9 и следствия 2.2.

График рис. 4г изображает функцию $\underline{\delta}^s(\mu)$ глубины (деленной на $\log n$) компонент кода, используемого в лемме 2.10 для вычисления оператора C_n (при изображении графика троичные цифры считаются отдельными компонентами, поэтому длина кода $\log_2 n + O(1)$). График получается упорядочением объединения участков графиков $\underline{\delta}([0,\, 0.584])$, $\underline{\delta}^{(3)}([0,\, 0.178])$ и $\underline{\delta}^a([0.762,\, 1])$.

График рис. 4д изображает функцию $\hat{\underline{\delta}}^s(\mu) = D_{\mu \log_2 n}/\log_2 n$ из леммы 2.11: величина $(\hat{\underline{\delta}}^s(\mu)+o(1))\log_2 n$ служит верхней оценкой глубины для $\mu \log_2 n$ битов кода входного набора, используемого в лемме 2.11 при реализации симметрической функции. График для $\underline{\delta}^s$ получается упорядочением участков графиков $\underline{\delta}([0,\, 0.608])$, $\underline{\delta}^{(3)}([0,\, 0.197])$ и $\underline{\delta}^a([0.805,\, 1])$. Пунктиром показан график функции $\hat{\underline{\delta}}^s(\mu)+1-\mu$, максимум которой определяет глубину реализации симметрических функций методом леммы 2.11. □

Пользуясь утверждением 2.3, получаем

Следствие 2.3. *Справедливы соотношения:*

$$L_{B_0}(M_n) \preceq n^{4.91}, \qquad L_{B_2}(M_n) \preceq n^{3.84},$$
$$D_{B_0}(M_n) \lesssim 5.14 \log_2 n, \qquad D_{B_2}(M_n) \lesssim 4.02 \log_2 n.$$

2.5 Синтез формул для MOD-функций

В свете материала предыдущих двух разделов, операторы вычисления арифметической суммы булевых переменных по модулям p^k могут быть интегрированы в общую схему реализации симметрических функций. Изучение функций сложения по малым простым модулям p (т. е. вычисления младших p-ичных разрядов суммы) дает предварительный ответ на вопрос, какой эффект можно ожидать от такой интеграции. Результаты настоящего раздела (см. табл. 3) указывают, что оценки табл. 2 (глубины и сложности вычисления симметрических функций) могут быть несколько улучшены за счет применения пятиричных компрессоров в базисе B_0 и, возможно, за счет применения семиричных компрессоров в базисе B_2.

Изложение следует работе [222] и построено следующим образом: в §§2.5.1–2.5.3 представлен обзор известных конструкций; общий способ по-

строения формул меньшей сложности и глубины в базисе B_0 приводится в §2.5.4; оценка глубины оператора MOD_n^3 в стандартном базисе доказывается в §2.5.5; в §2.5.6 доказывается оценка глубины оператора MOD_n^7 в базисе B_2. Новые оценки суммирует

Теорема 2.5. *Справедливы соотношения:*

$$L_{B_0}(\mathrm{MOD}_n^5) \preceq n^{3.22}, \qquad L_{B_0}(\mathrm{MOD}_n^7) \preceq n^{3.63},$$
$$D_{B_0}(\mathrm{MOD}_n^3) \lesssim 2.8\log_2 n, \qquad D_{B_0}(\mathrm{MOD}_n^5) \lesssim 3.35\log_2 n,$$
$$D_{B_0}(\mathrm{MOD}_n^7) \lesssim 3.87\log_2 n, \qquad D_{B_2}(\mathrm{MOD}_n^7) \lesssim 2.93\log_2 n.$$

2.5.1 Простые формулы в базисе B_0

Разобьем переменные на 2 группы: $x = (x^1, x^2)$, $|x^i| = n_i$ (через $|x|$ обозначается длина булева набора). Справедливы элементарные формулы [43]:

$$\mathrm{MOD}_{n_1+n_2}^{m,r}(x) = \bigvee_{k=0}^{m-1} \mathrm{MOD}_{n_1}^{m,k}(x^1) \cdot \mathrm{MOD}_{n_2}^{m,r-k}(x^2), \tag{2.32}$$

$$\mathrm{MOD}_{n_1+n_2}^{m,r}(x) = \bigwedge_{k=0}^{m-1} \left(\mathrm{MOD}_{n_1}^{m,k}(x^1) \vee \overline{\mathrm{MOD}_{n_2}^{m,r-k}(x^2)}\right). \tag{2.33}$$

Напомним, что у функций f и $\overline{f}$ совпадают и глубина, и сложность при реализации формулами в любом из базисов B_0 и B_2, см. [49, 145].

При помощи любого из тождеств (2.32), (2.33) методом деления пополам можно получить верхнюю оценку сложности формул

$$L_{B_0}(\mathrm{MOD}_n^m) \preceq n^{1+\log m}, \tag{2.34}$$

указанную еще в работе [43]. В частности, при малых m имеем

$$L_{B_0}(\mathrm{MOD}_n^3) \preceq n^{2.59}, \qquad L_{B_0}(\mathrm{MOD}_n^5) \preceq n^{3.33}, \qquad L_{B_0}(\mathrm{MOD}_n^7) \preceq n^{3.81}.$$

При построении эффективных по глубине формул разложения (2.32) и (2.33) выгодно использовать поочередно, а множество переменных разбивать на части разной длины. По сути в этом состоит метод [113], в оригинальной работе изложенный в терминах коммуникационной сложности.

Например, при $m = 3$ разложения (2.32) и (2.33) приводят к конъюнкциям или дизъюнкциям формул глубины k и $k-1$. Обозначим через N_k максимальное n, такое, что функции $\mathrm{MOD}_n^{3,r}$ представимы как конъюнкциями, так и дизъюнкциями формул глубины k и $k-1$. Тогда имеем

$$N_{k+2} \geq N_k + N_{k-2},$$

откуда следует

$$D_{B_0}(\mathrm{MOD}_n^3) \leq 2\log_\varphi n + O(1) < 2.89\log n + O(1),$$

где $\varphi = \frac{1+\sqrt{5}}{2}$. Аналогично выводится оценка

$$D_{B_0}(\mathrm{MOD}_n^5) \leq \log_\alpha n + O(1) < 3.48\log n + O(1),$$

где $\alpha^4 = \alpha + 1$.

При $m = 7$ указанный прием не позволяет улучшить тривиально вытекающую из (2.32) или (2.33) оценку

$$D_{B_0}(\mathrm{MOD}_n^7) \leq 4\log n + O(1).$$

2.5.2 Простые формулы в базисе B_2

В базисе B_2 справедливы чуть более короткие формулы [161]:

$$\mathrm{MOD}_{n_1+n_2}^{m,r}(x) = \bigwedge_{k=1}^{m-1} \left(\mathrm{MOD}_{n_1}^{m,k}(x^1) \sim \mathrm{MOD}_{n_2}^{m,r-k}(x^2)\right), \qquad (2.35)$$

где «$\sim$» означает булеву операцию эквивалентности, см. также [124, §4.4].

Эти формулы приводят к оценке

$$L_{B_2}(\mathrm{MOD}_n^m) \preceq n^{1+\log(m-1)}.$$

При малых m получаем оценки сложности

$$L_{B_2}(\mathrm{MOD}_n^3) \preceq n^2, \qquad L_{B_2}(\mathrm{MOD}_n^5) \preceq n^3$$

и одновременно оценки глубины

$$D_{B_2}(\mathrm{MOD}_n^3) \leq 2\log n + O(1), \qquad D_{B_2}(\mathrm{MOD}_n^5) \leq 3\log n + O(1).$$

2.5.3 Алгебраический метод

Общая идея состоит в том, чтобы найти подходящее представление

$$(\mathbb{Z}_m, +) \cong (G, *),$$

где групповая операция $*$ в G выполняется просто.

В работе [157] предложено использовать мультипликативные группы двоичных конечных полей. В случае $m = 7$ применяется мультипликативная группа поля $GF(2^3)$, элементы которого представляются двоичными матрицами размера 3×3. Групповая операция является обычным умножением матриц над $GF(2)$.

Тогда оператор $\mathrm{MOD}_n^7(x)$ характеризуется матрицей $H_n(x)$, элементы которой вычисляются рекурсивно по формулам

$$H_{n_1+n_2}[i,j](x) = \bigoplus_{k=0}^{2} H_{n_1}[i,k](x^1) \cdot H_{n_2}[k,j](x^2). \tag{2.36}$$

Получаем оценки

$$L_{B_2}(\mathrm{MOD}_n^7) \preceq n^{\log 6} < n^{2.59}, \qquad D_{B_2}(\mathrm{MOD}_n^7) \le 3 \log n + O(1),$$

так как переход между различными представлениями группы $\mathbb{Z}_7$ выполняется с глубиной $O(1)$.

В случае $m = 3$ и $m = 5$ в методе [157] используются поля $GF(2^2)$ и $GF(2^4)$ соответственно; это приводит к таким же оценкам, как в §2.5.2.

2.5.4 Формулы в расширенной кодировке

Пусть $S \subset \mathbb{Z}_m$. Определим функции

$$\mathrm{MOD}_n^{m,S}(x) = \left(\sum_{i=1}^{n} x_i \bmod m \in S \right),$$

обобщающие (2.2). Набор всех функций $\mathrm{MOD}_n^{m,S}$, $0 < |S| < m$, задает значение суммы переменных по модулю m, используя максимальную однозначную кодировку элементов множества $\mathbb{Z}_m$. Любая другая двоичная однозначная кодировка содержится в максимальной.

Укажем простой способ уточнить оценки сложности (2.34). Будем искать формулы вида

$$\mathrm{MOD}_n^{m,S}(x) = \bigvee_k \mathrm{MOD}_{n_1}^{m,A_k}(x^1) \cdot \mathrm{MOD}_{n_2}^{m,B_k}(x^2). \tag{2.37}$$

Их можно интерпретировать в терминах покрытий множеств или матриц. Каждому множеству S сопоставим $(m, m)-$матрицу I_m^S, строки и столбцы которой занумерованы числами из $\mathbb{Z}_m$, а элементы определяются как $I_m^S[i,j] = (i+j \in S)$. Тогда формула (2.37) отвечает покрытию матрицы I_m^S сплошь единичными подматрицами (прямоугольниками) на пересечении строк A_k и столбцов B_k.

Из (2.32) вытекает тождество [43]

$$\mathrm{MOD}_{n_1+n_2}^{m,S}(x) = \bigvee_{k=0}^{m-1} \mathrm{MOD}_{n_1}^{m,k}(x^1) \cdot \mathrm{MOD}_{n_2}^{m,S-k}(x^2), \tag{2.38}$$

отвечающее покрытию матрицы отдельными строками ($S-k$ обозначает множество $\{r-k \mid r \in S\}$). Ранг тривиального покрытия (число покрывающих матриц) равен m. Нас интересуют матрицы I_m^S с рангом минимального покрытия (или OR-рангом) $\mathrm{rk}_\vee(I_m^S) < m$. Через $\overline{S}$ будем обозначать $\mathbb{Z}_m \setminus S$.

Известный пример матрицы малого ранга получается при $S = \overline{\{r\}}$ (матрица I_m^S имеет одну нулевую циклическую диагональ и единицы — в остальных позициях), см., например, [145]. Легко проверить, что если $m \le \binom{t}{\lfloor t/2 \rfloor}$, то $\mathrm{rk}_\vee(I_m^S) \le t$. Соответствующее покрытие для матрицы с нулевой главной диагональю получается так: занумеруем строки матрицы подмножествами множества $\{1, \ldots, t\}$ мощности $\lfloor t/2 \rfloor$, а столбцы — дополняющими их подмножествами мощности $\lceil t/2 \rceil$. $i-$ю матрицу из покрытия образуют строки и столбцы, номерам которых принадлежит i.

В случае $m = 3$ при любом $S \ne \emptyset, \mathbb{Z}_m$ матрицы I_m^S имеют полный ранг, зато при $m = 5$ матрицы $I_m^{\overline{\{r\}}}$ имеют ранг 4 (причем покрытие образуют прямоугольники со сторонами 2 или 3), см. рис. 5.

Как следствие, функция $\mathrm{MOD}_n^{5,S}$ при $|S| = 4$ реализуется формулой (2.37) с 4-мя слагаемыми, а при $|S| = 1$ — двойственной формулой (конъюнкция дизъюнкций) такого же размера. Для остальных функций выбираем реализацию (2.38), поскольку $\mathrm{rk}_\vee(I_5^S) = 5$ при $2 \le |S| \le 3$.

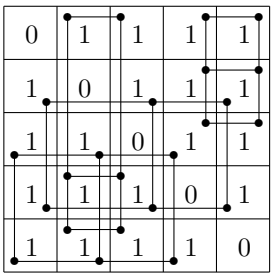


Рис. 5: Покрытие матрицы $\overline{I_5}$

Оценим сложность формул, к которым приводит указанная стратегия. Обозначим через X_n сложность формул для $\mathrm{MOD}_n^{5,S}$, $|S| = 1$ или 4, а через Y_n — сложность формул для $\mathrm{MOD}_n^{5,S}$, $|S| = 2$ или 3. Пусть $Z_n = \max\{X_n, Y_n/a\}$, где параметр a будет выбран позднее. Из (2.37) и (2.38) получаем

$$Z_{2n} \le \max\{8Y_n, 5(X_n + Y_n)/a\}.$$

С целью минимизировать отношение Z_{2n}/Z_n выбираем $a = \frac{5+\sqrt{185}}{16}$, в результате получаем $Z_{2n} \le 8aZ_n$. Следовательно,

$$L_{B_0}(\mathrm{MOD}_n^5) \preceq n^{3+\log a} < n^{3.22}.$$

Незначительное улучшение оценки возможно при дополнительном подборе оптимального отношения n_1/n_2.

Аналогичным образом, матрица $I_7^{\overline{\{r\}}}$ может быть покрыта 5-ю прямоугольниками, при этом один из них имеет размер 1×6. Обозначая, как и выше, через X_n сложность формул для $\mathrm{MOD}_n^{7,S}$, $|S| = 1$, через Y_n — сложность формул для $\mathrm{MOD}_n^{7,S}$, $2 \le |S| \le 3$, и вводя обозначение $Z_n = \max\{X_n, Y_n/a\}$, получаем соотношение

$$Z_{2n} \le \max\{2X_n + 8Y_n, 7(X_n + Y_n)/a\}.$$

При выборе $a = \frac{5+\sqrt{249}}{16}$ имеем $Z_{2n} \le (8a + 2)Z_n$ и

$$L_{B_0}(\mathrm{MOD}_n^7) \preceq n^{\log(8a+2)} < n^{3.64}.$$

Оценку можно уточнить до $O(n^{3.63})$, если применять разбиение на неравные группы переменных: $4n_1 \approx 5n_2$. Иначе, это уточнение можно получить, используя соотношение $rk_\vee(I_7^S) = 6$, $|S| = 5$.

Построенные формулы позволяют улучшить и оценки глубины из [113].

Обозначим через N_k максимальное n, такое, что функции $\mathrm{MOD}_n^{5,S}$ представимы дизъюнкциями и конъюнкциями двух формул глубины k и $k-2$. Докажем соотношение

$$N_{k+10} \geq N_{k+6} + 2N_{k+1} + 2N_k.$$

При помощи формул (2.38) и их отрицаний можно реализовать все функции $\mathrm{MOD}^{5,S}_{N_{k+1}+N_k}$ дизъюнкциями или конъюнкциями формул глубины $k+4$ и $k+2$, а при $|S| = 1$ — формулой глубины $k+4$, используя минимальное покрытие матрицы $I_5^{\overline{\{r\}}}$. Далее, применяя тождества (2.32), (2.33) с $n_1 = n_2 = N_{k+1} + N_k$, строим формулы для функций $\mathrm{MOD}^{5,r}_{2N_{k+1}+2N_k}$, которые являются дизъюнкциями или конъюнкциями формул глубины $k+7$ и $k+5$. Окончательно, применяем формулы (2.38) и их отрицания с параметрами $n_1 = 2N_{k+1} + 2N_k$ и $n_2 = N_{k+6}$, чтобы реализовать функции $\mathrm{MOD}^{5,S}_{N_{k+6}+2N_{k+1}+2N_k}$ дизъюнкциями и конъюнкциями формул глубины $k+10$ и $k+8$. Таким образом,

$$D_{B_0}(\mathrm{MOD}_n^5) \leq \log_\alpha n + O(1) < 3.35 \log n + O(1),$$

где $\alpha^{10} = \alpha^6 + 2\alpha + 2$.

Аналогично поступаем в случае $m = 7$. Обозначим через N_k максимальное n, такое, что функции $\mathrm{MOD}_n^{7,S}$ представимы формулами глубины k. Докажем соотношение

$$N_{k+12} \geq N_{k+8} + 2N_{k+1} + 2N_k.$$

Используя минимальное покрытие матрицы $I_7^{\overline{\{r\}}}$, реализуем функции $\mathrm{MOD}^{7,r}_{2N_k}$ и $\mathrm{MOD}^{7,r}_{2N_{k+1}}$, конъюнкциями формул глубины $k+3$ и $k+1$, и конъюнкциями формул глубины $k+4$ и $k+2$ соответственно. Далее, при помощи формулы (2.32) с глубиной $k+8$ вычисляются функции $\mathrm{MOD}^{7,r}_{2N_{k+1}+2N_k}$. Теперь все функции $\mathrm{MOD}^{7,S}_{N_{k+8}+2N_{k+1}+2N_k}$ могут быть вычислены с глубиной $k+12$ формулами (2.38) с параметрами $n_1 = 2N_{k+1} + 2N_k$ и $n_2 = N_{k+8}$. Как следствие,

$$D_{B_0}(\mathrm{MOD}_n^7) \leq \log_\alpha n + O(1) < 3.87 \log n + O(1),$$

где $\alpha^{12} = \alpha^8 + 2\alpha + 2$.

2.5.5 Оценка глубины оператора MOD_n^3 в базисе B_0

Рассмотренный выше способ не приводит к более эффективным формулам для функций $\mathrm{MOD}_n^{3,r}$. Оценку глубины оператора MOD_n^3 все же можно уточнить, используя специальные формулы, отвечающие разбиению множества переменных на 3 группы.

Разделим набор n переменных на три группы: $x = (x^1, x^2, x^3)$, $|x^i| = n_i$. Введем сокращенное обозначение $F_i^r = \mathrm{MOD}_{n_i}^{3,r}(x^i)$. Справедлива формула

$$\mathrm{MOD}_n^{3,r}(x) = (F_1^0 \vee F_2^0 \vee F_3^r)(F_1^1 \vee F_2^1 \vee F_3^{r+1})(F_1^2 \vee F_2^2 \vee F_3^{r+2}) \vee \\ \vee (F_1^0 \vee F_2^2 \vee F_3^{r+1})(F_1^1 \vee F_2^0 \vee F_3^{r+2})(F_1^2 \vee F_2^1 \vee F_3^r), \quad (2.39)$$

которую можно также записать в виде

$$\mathrm{MOD}_n^{3,r}(x) = (F_1^0 \vee F_2^0 \vee F_3^r)(F_1^1 \vee F_2^1 \vee F_3^{r+1})(F_1^2 \vee F_2^2 \vee F_3^{r+2}) \vee \\ \vee F_1^0 \cdot F_2^0 \cdot F_3^r \vee F_1^1 \cdot F_2^1 \cdot F_3^{r+1} \vee F_1^2 \cdot F_2^2 \cdot F_3^{r+2}. \quad (2.40)$$

Пусть N_k — максимальное n, такое, что компоненты оператора MOD_n^3 представимы формулами глубины k. Докажем неравенство

$$N_{k+11} \geq 2N_{k+4} + 2N_{k+3} + N_{k+2} + 4N_k.$$

При помощи (2.32) и (2.33) реализуем функции $\mathrm{MOD}_{2N_k}^3$ дизъюнкциями или конъюнкциями формул глубины $k+2$ и $k+1$. Используя (2.40) с параметрами $n_1 = n_2 = 2N_k$ и $n_3 = N_{k+2}$, получим формулу для $\mathrm{MOD}_{N_{k+2}+4N_k}^3$, которая является дизъюнкцией формул глубины $k+6$, $k+5$ и $k+4$. Посредством (2.32) реализуем функции $\mathrm{MOD}_{2N_{k+3}}^3$ и $\mathrm{MOD}_{2N_{k+4}}^3$ дизъюнкциями формул глубины $k+5$ и $k+4$, и дизъюнкциями формул глубины $k+6$ и $k+5$ соответственно. Окончательно, строим формулу глубины $k+11$ для $\mathrm{MOD}_{2N_{k+4}+2N_{k+3}+N_{k+2}+4N_k}^3$, применяя (2.39) с параметрами $n_1 = N_{k+2} + 4N_k$, $n_2 = 2N_{k+4}$, $n_3 = 2N_{k+3}$. Таким образом,

$$D_{B_0}(\mathrm{MOD}_n^3) \leq \log_\alpha n + O(1) < 2.8 \log n + O(1),$$

где $\alpha^{11} = 2\alpha^4 + 2\alpha^3 + \alpha^2 + 4$.

2.5.6 Оценка глубины оператора MOD_n^7 в базисе B_2

Как и в предыдущем параграфе, рассмотрим разбиение набора переменных на 3 группы. Исходные формулы (2.36) конкретизируются следующим образом. Положим $T = \{0, 1, 2, 5\} \subset \mathbb{Z}_7$. Функции $\mathrm{MOD}_n^{7,T+r}$ вычисляют сумму переменных в специальном представлении множества $\mathbb{Z}_7$. Для $x = (x^1, x^2)$, $|x^i| = n_i$, выполнено

$$\mathrm{MOD}_{n_1+n_2}^{7,T+r}(x) = \bigoplus_{k\in\{0,1,3\}} \mathrm{MOD}_{n_1}^{7,T+k-3r}(x^1) \cdot \mathrm{MOD}_{n_2}^{7,T+k-3r}(x^2). \tag{2.41}$$

Теперь пусть $x = (x^1, x^2, x^3)$, $|x^i| = n_i$, $|x| = n$. Введем сокращенное обозначение $F_i^r = \mathrm{MOD}_{n_i}^{7,T+r}(x^i)$. Тогда из (2.41) следует

$$\mathrm{MOD}_n^{7,T+r}(x) = \bigoplus_{j=0}^{6} \mathrm{MOD}_{n_3}^{7,r+j}(x^3) \cdot \mathrm{MOD}_{n_1+n_2}^{7,T-j}(x^1, x^2) =$$

$$= \bigoplus_{j=0}^{6} \mathrm{MOD}_{n_3}^{7,r+j}(x^3) \bigoplus_{k\in\{0,1,3\}} F_1^{k+3j} \cdot F_2^{k+3j},$$

откуда приведением подобных слагаемых с учетом очевидного тождества

$$\mathrm{MOD}_n^{m,S}(x) = \bigoplus_{r\in S} \mathrm{MOD}_n^{m,r}(x)$$

получаем

$$\mathrm{MOD}_n^{7,T+r}(x) = \bigoplus_{k=0}^{6} F_1^k \cdot F_2^k \cdot \overline{F_3^{3+r-2k}}. \tag{2.42}$$

Непосредственно из (2.42) вытекает соотношение

$$N_{k+5} \geq N_{k+1} + 2N_k,$$

где N_k — максимальное n, такое, что функции $\mathrm{MOD}_n^{7,T+r}$ имеют глубину k. Следовательно,

$$D_{B_2}(\mathrm{MOD}_n^7) \leq \log_\alpha n + O(1) < 2.93 \log n + O(1),$$

где $\alpha^5 = \alpha + 2$.

2.6 Сложность формул в k-арных базисах

В данном разделе изложены результаты работы [229] о нижних оценках сложности формул в базисе U_k.

2.6.1 Экспонента Храпченко и меры сложности двудольных графов

Для произвольного двудольного графа $G = (A, B, E)$ на множествах вершин A, B и с множеством ребер E определим величину $s(G)$ по аналогии с (2.3) как

$$s(G) = \max_{X \subset A,\, Y \subset B} \frac{|E \cap (X \times Y)|^2}{|X| \cdot |Y|}. \tag{2.43}$$

Эту величину можно интерпретировать как произведение средних степеней вершин в долях X и Y или, в определенном смысле, как максимальную плотность графа G. Из определения следует, что если $G' \subset G$, то $s(G') \leq s(G)$.

Любой булевой функции f от n переменных можно поставить в соответствие двудольный подграф G_f булева куба $\{0,\, 1\}^n$ с долями $f^{-1}(0)$ и $f^{-1}(1)$, и множеством ребер, соединяющих пары вершин из $R(f^{-1}(0), f^{-1}(1))$. Тогда $s(f) = s(G_f)$ ввиду (2.3) и (2.43). Таким образом, величина (2.43) служит расширением понятия чувствительности функции на любые двудольные графы.

Множество двудольных графов $G_i = (A_i, B_i, E_i)$ назовем *покрытием* двудольного графа $G = (A, B, E)$, если для всех i выполняется $A_i \subset A$, $B_i \subset B$, при этом $E \subset \bigcup E_i$. Покрытие назовем *монотонным*, если для любого множества индексов I одно из двух множеств $A \setminus \bigcup_{i \in I} A_i$ и $B \setminus \bigcup_{i \notin I} B_i$ пусто.

При любом $k \geq 2$ определим *экспоненту сложности* двудольных графов χ_k (соответственно *экспоненту монотонной сложности* χ_k^*) как максимальное число χ, такое, что для любого двудольного графа G и для любого его покрытия (соответственно монотонного покрытия) $G_1, \ldots, G_k$ выполнено условие субаддитивности

$$s^\chi(G_1) + \ldots + s^\chi(G_k) \geq s^\chi(G).$$

Так определенные функционалы s^{χ_k} и $s^{\chi_k^*}$ являются субаддитивными мерами сложности двудольных графов (но по отношению к различным операциям). Подробнее о применении субаддитивных мер в теории сложности см. в [145, Ch. 6].

Тривиально выполняется $\chi_2 \geq \chi_3 \geq \ldots$ и $\chi_2^* \geq \chi_3^* \geq \ldots$ Более содержательным является следующий результат.

Утверждение 2.5. *При любом $k \geq 2$ справедливо $\chi_k \leq \chi_k^* \leq \chi_{U_k}$.*

Доказательство. Первое неравенство тривиально, поскольку монотонное покрытие тоже является покрытием. Второе неравенство докажем в более общей форме.

Утверждение 2.6. *Пусть $T \subset \{0,1\}^n \times \{0,1\}^n$. Для множеств $A, B \subset \{0,1\}^n$ определим $R_T(A,B) = (A \times B) \cap T$. Для произвольной булевой функции $f(x_1, \ldots, x_n)$ положим*

$$s_T(f) = \max_{N \subset f^{-1}(0),\, P \subset f^{-1}(1)} \frac{|R_T(N,P)|^2}{|N| \cdot |P|}.$$

Тогда

$$L_{U_k}(f) \geq \left(\frac{s_T(f)}{s}\right)^{\chi_k^*}, \qquad s = \max_{1 \leq i \leq n} \max\{s_T(x_i),\, s_T(\overline{x}_i)\}. \tag{2.44}$$

Доказательство. Напомним, что базис U_k строго эквивалентен базису из всех монотонных k-местных функций и отрицания. Стандартным образом все отрицания в формуле над этим базисом могут быть опущены на место переменных (правила де Моргана).

Рассуждение проведем индукцией по сложности функций. Для константы $\sigma \in \{0,1\}$ справедливо $s_T(\sigma) = 0$. Для любой переменной x и ее отрицания выполняется $s_T(x), s_T(\overline{x}) \leq s$ (база индукции).

Предположим, что (2.44) справедливо для всех функций сложности меньше L. Рассмотрим произвольную функцию f сложности L. Пусть в минимальной неизбыточной формуле она вычисляется как $f = \varphi(f_1, \ldots, f_k)$, где φ — монотонная функция. Тогда $L_{U_k}(f_i) < L$ при всех i. Определим $G = (A, B, E)$, где $A \subset f^{-1}(0)$, $B \subset f^{-1}(1)$, $E = R_T(A,B)$ — множества, на

которых достигается $s_T(f)$. Положим $G_i = (A_i, B_i, E_i)$, где $A_i = f_i^{-1}(0) \cap A$, $B_i = f_i^{-1}(1) \cap B$, $E_i = (A_i \times B_i) \cap E$.

По определению, $s_T(f) = s(G)$ и $s_T(f_i) \geq s(G_i)$ в силу $A_i \subset f_i^{-1}(0)$ и $B_i \subset f_i^{-1}(1)$. Покажем, что графы G_i образуют монотонное покрытие графа G.

Условия $A_i \subset A$ и $B_i \subset B$ выполнены по построению. Если $f(\alpha) = 0$ и $f(\beta) = 1$, то найдется функция f_i, для которой также $f_i(\alpha) = 0$ и $f_i(\beta) = 1$. В противном случае из $f_i(\alpha) \geq f_i(\beta)$ для всех i следовало бы $f(\alpha) \geq f(\beta)$ из-за монотонности φ. Таким образом, если $\rho \in E$, то $\rho \in \bigcup E_i$, иначе говоря, $E \subset \bigcup E_i$.

Предположим, что нарушено условие монотонности. Это значит, что для некоторого множества индексов I найдутся наборы $\alpha \in A \setminus \bigcup_{i \in I} A_i$ и $\beta \in B \setminus \bigcup_{i \notin I} B_i$. Иначе говоря, $f(\alpha) = 0$ и $f_i(\alpha) = 1$ при всех $i \in I$, а также $f(\beta) = 1$ и $f_i(\beta) = 0$ при всех $i \notin I$. С другой стороны, ввиду $(f_1(\alpha), \ldots, f_k(\alpha)) \geq (f_1(\beta), \ldots, f_k(\beta))$ и монотонности функции φ должно выполняться $f(\alpha) \geq f(\beta)$. Противоречие.

Поэтому при $\chi = \chi_k^*$ в силу предположения индукции и определения χ_k^* имеет место

$$L_{U_k}(f) = L_{U_k}(f_1) + \ldots + L_{U_k}(f_k) \geq \frac{s_T^\chi(f_1) + \ldots + s_T^\chi(f_k)}{s^\chi} \geq$$
$$\geq \frac{s^\chi(G_1) + \ldots + s^\chi(G_k)}{s^\chi} \geq \left(\frac{s(G)}{s}\right)^\chi = \left(\frac{s_T(f)}{s}\right)^\chi. \quad \square$$

При выборе $T = R(\{0,1\}^n, \{0,1\}^n)$ соотношение (2.44) превращается в $L_{U_k}(f) \geq s^{\chi_k^*}(f)$ ввиду $s_T(f) = s(f)$ и $s = s(x_i) = s(\overline{x_i}) = 1$. Утверждение 2.5 доказано. $\square$

Утверждение 2.5 позволяет оценки на χ_{U_k} извлекать в качестве следствий из оценок на χ_k и χ_k^*. Последние величины могут быть проще для анализа, так как связаны с менее специфичными по сравнению с булевыми функциями объектами — двудольными графами.

Для разминки рассмотрим случай $k = 2$ и выведем несколько известных результатов в новой интерпретации. Следующее утверждение по сути повторяет теорему 8.1 из [209], где результат Храпченко доказывается на

языке формальных мер сложности — такой способ доказательства предложен независимо А. Е. Андреевым и М. Патерсоном.

Утверждение 2.7. $\chi_2^* = 1$.

Доказательство. Докажем $\chi_2^* \geq 1$, т. е. что для произвольного двудольного графа $G = (A, B, E)$ и его монотонного покрытия $G_1 = (A_1, B_1, E_1)$, $G_2 = (A_2, B_2, E_2)$ выполняется $s(G_1) + s(G_2) \geq s(G)$. Не ограничивая общности[17], полагаем $s(G) = \frac{|E|^2}{|A| \cdot |B|}$.

По условию монотонности либо $A \setminus A_2 = \emptyset$, либо $B \setminus B_1 = \emptyset$. И наоборот, либо $A \setminus A_1 = \emptyset$, либо $B \setminus B_2 = \emptyset$.

Если $A \setminus A_2 = B \setminus B_2 = \emptyset$, то $A_2 = A$ и $B_2 = B$, значит $G_2 = G$, и доказывать нечего.

Если $A \setminus A_2 = A \setminus A_1 = \emptyset$, то $A_1 = A_2 = A$. Получаем $s(G_1) \geq \frac{|E_1|^2}{|A| \cdot |B_1|}$ и, поскольку $|E \setminus E_1|$ ребер соединяют A с $B \setminus B_1 \subset B_2$, то $s(G_2) \geq \frac{(|E| - |E_1|)^2}{|A| \cdot (|B| - |B_1|)}$. Применяя простое неравенство $\frac{x_1^2}{y_1} + \frac{x_2^2}{y_2} \geq \frac{(x_1 + x_2)^2}{y_1 + y_2}$, справедливое при $y_1, y_2 > 0$, выводим

$$s(G_1) + s(G_2) \geq \frac{|E_1|^2}{|A| \cdot |B_1|} + \frac{(|E| - |E_1|)^2}{|A| \cdot (|B| - |B_1|)} \geq \frac{|E|^2}{|A| \cdot |B|} = s(G).$$

Оставшиеся два случая симметричны рассмотренным.

Верхняя оценка $\chi_2^* \leq 1$ тривиальна. Для графа G, состоящего из двух ребер с общей вершиной, и (монотонного) покрытия его графами G_1, G_2, состоящими из одного ребра каждый, выполнено $s(G) = 2$ и $s(G_1) = s(G_2) = 1$. Следовательно, $2 \geq 2^{\chi_2^*}$. □

Как следствие из утверждений 2.5 и 2.7 получаем результат Храпченко $\chi_{U_2} \geq 1$ [80]. Легко вывести и обобщение, принадлежащее К. Л. Рычкову [65]:

$$L_{U_2}(f) \geq s_d(f) = \max_{N \subset f^{-1}(0),\, P \subset f^{-1}(1)} \frac{|R_d(N, P)|^2}{\left(C_{n-1}^{d-1} + \ldots + C_{n-1}^1 + 1\right) |N| \cdot |P|}, \quad (2.45)$$

где $R_d(N, P)$ — множество пар наборов из N и P с расстоянием не более d, а n — число аргументов функции f. Для этого применяем утверждение 2.6

[17] Иначе рассмотрим подграф $G' \subset G$, на котором достигается $s(G)$, и его покрытие графами $G_1 \cap G'$ и $G_2 \cap G'$. Из $s(G_1 \cap G') + s(G_2 \cap G') \geq s(G')$ тогда следует $s(G_1) + s(G_2) \geq s(G)$.

с выбором $R_T = R_d$. Остается заметить, что $s_d(x) = s_d(\overline{x}) = C_{n-1}^{d-1} + \ldots + C_{n-1}^1 + 1$. Оценка (2.45) позволяет получать квадратичные нижние оценки сложности для некоторых функций, связанных с плотно упакованными кодами, см. [65].

Любопытно, что $\chi_2 < \chi_2^*$. Докажем простую оценку.

Утверждение 2.8. $\chi_2 < 0.95$.

Доказательство. Рассмотрим пример графа G и его покрытия графами G_1, G_2 на рис. 6.

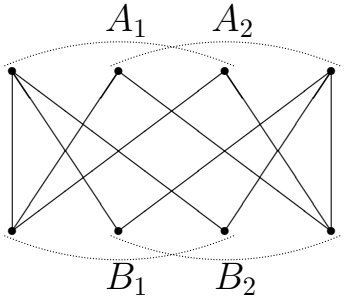


Рис. 6: Пример покрытия двудольного графа

Легко проверить, что $s(G) \geq \frac{10^2}{4^2} = \frac{25}{4}$, при этом $s(G_1) = s(G_2) = 3$ (последние величины достигаются на подграфах с долями из одной и трех вершин). Неравенство $2 \cdot 3^\chi \geq (25/4)^\chi$ выполняется только при $\chi \leq \log_{25/12} 2 < 0.95$. □

2.6.2 Оценки для экспонент сложности в общем случае

Теорема 2.6. *При любом* $k \geq 2$ *выполнено* $\chi_{U_k} \leq \log_{\lceil k/2 \rceil (\lfloor k/2 \rfloor + 1)} k$.

Доказательство. Рассмотрим функцию голосования m_k. Выберем в качестве множеств N и P соседние слои булева куба, на которых функция меняет значение: это соответственно наборы веса $p-1$ и p, где $p = \lceil k/2 \rceil$. Очевидно $L_{U_k}(m_k) = k$. С другой стороны

$$s(m_k) \geq \frac{|R(N,P)|}{|P|} \cdot \frac{|R(N,P)|}{|N|} =$$

$$= p(k-p+1) = \lceil k/2 \rceil (\lfloor k/2 \rfloor + 1) = (L_{U_k}(m_k))^{1/\chi},$$

где $\chi = \log_{\lceil k/2 \rceil (\lfloor k/2 \rfloor + 1)} k$. Таким образом, при бóльших значениях χ условие (3.1) не выполняется. $\square$

Заметим, что оценку теоремы 2.6 можно записать более грубо как

$$\chi_{U_k} \le \frac{\log_2 k}{\log_2(\lceil k/2 \rceil (\lfloor k/2 \rfloor + 1))} < \frac{\log_2 k}{2\log_2(k/2)} = \frac{1}{2} + \frac{1}{2\log_2(k/2)}$$

(при $k = 2$ неравенство выполняется в смысле $1 < \infty$).

Теорема 2.7. *При любом $k \ge 2$ справедливо $\chi_k \ge 1/2 + 1/(10 \ln k)$.*

Доказательство. Пусть графы $G_i = (A_i, B_i, E_i)$ составляют покрытие графа $G = (A, B, E)$. Требуется доказать, что при $\chi = 1/2 + 1/(10 \ln k)$ выполнено

$$s^\chi(G_1) + \ldots + s^\chi(G_k) \ge s^\chi(G),$$

где k — число графов в покрытии. Обозначим $a_i = |A_i|/|A|$, $b_i = |B_i|/|B|$, $e_i = |E_i|/|E|$. Как и выше в утверждении 2.7, можно считать, что $s(G) = \frac{|E|^2}{|A| \cdot |B|}$.

Рассуждение проведем индукцией по k. В вырожденном случае $k = 1$ примем для удобства $\chi_1 = 1$. Докажем переход от $k-1$ к $k \ge 2$. Рассмотрим два случая.

I. Пусть один из графов покрытия, скажем, G_1, содержит значительную часть обеих долей графа G. Более конкретно, предположим, что $a_1 + b_1 \ge 1.63$.

Заметим, что графы $G' = (A, B \setminus B_1, E \setminus E_1)$ и $G'' = (A \setminus A_1, B, E \setminus E_1)$ имеют покрытия $G'_2, \ldots, G'_k$ и соответственно $G''_2, \ldots, G''_k$, где $G'_i = (A_i, B_i \setminus B_1, E_i \setminus E_1)$, $G''_i = (A_i \setminus A_1, B_i, E_i \setminus E_1)$, при этом $G'_i, G''_i \subset G_i$. Вместе графы G' и G'' содержат все ребра графа G, не связывающие A_1 и B_1, т. е. суммарно не менее $(1-e_1)|E|$ ребер. Используя простое неравенство

$$\max\left\{\frac{x_1}{y_1}, \frac{x_2}{y_2}\right\} \ge \frac{x_1 + x_2}{y_1 + y_2}, \tag{2.46}$$

справедливое при $y_1, y_2 > 0$, получаем

$$\max\{s(G'), s(G'')\} = \max\{\sqrt{s(G')}, \sqrt{s(G'')}\}^2 \ge$$
$$\ge \frac{(1-e_1)^2|E|^2}{(\sqrt{1-a_1} + \sqrt{1-b_1})^2 |A| \cdot |B|}. \tag{2.47}$$

Согласно индуктивному предположению при $\chi = 1/2 + 1/(10\ln(k-1))$ (при $\chi = 1$ в случае $k = 2$) выполнено

$$s^\chi(G_2) + \ldots + s^\chi(G_k) \geq$$
$$\geq \max\{s^\chi(G'_2) + \ldots + s^\chi(G'_k),\, s^\chi(G''_2) + \ldots + s^\chi(G''_k)\} \geq$$
$$\geq \max\{s^\chi(G'),\, s^\chi(G'')\} \geq \left(\frac{(1-e_1)^2|E|^2}{(\sqrt{1-a_1}+\sqrt{1-b_1})^2|A|\cdot|B|}\right)^\chi. \quad (2.48)$$

В силу неравенства Минковского

$$\sum x_i^p \geq \Big(\sum x_i\Big)^p, \quad p \leq 1, \quad x_i \geq 0,$$

(см. [77, гл. II]), соотношение (2.48) выполнено и при всех меньших значениях χ.

Как следствие, при $\chi = 1/2 + 1/(10\ln k)$ имеем

$$s^\chi(G_1) + \ldots + s^\chi(G_k) \geq s^\chi(G_1) + \max\{s^\chi(G'),\, s^\chi(G'')\} \geq$$
$$\left(\left(\frac{e_1^2}{a_1 b_1}\right)^\chi + \left(\frac{(1-e_1)^2}{(\sqrt{1-a_1}+\sqrt{1-b_1})^2}\right)^\chi\right)\left(\frac{|E|^2}{|A|\cdot|B|}\right)^\chi \geq$$
$$\left[\left(\frac{e_1}{t}\right)^{2\chi} + \left(\frac{1-e_1}{2\sqrt{1-t}}\right)^{2\chi}\right]\left(\frac{|E|^2}{|A|\cdot|B|}\right)^\chi, \quad (2.49)$$

где $t = (a_1 + b_1)/2$ (последний переход использует неравенства между средними геометрическим, арифметическим и квадратичным: $\sqrt{ab} \leq (a+b)/2 \leq \sqrt{(a^2+b^2)/2}$).

Применяя к выражению в квадратных скобках неравенство Гёльдера

$$\sum x_i y_i \leq \Big(\sum x_i^p\Big)^{1/p}\Big(\sum y_i^q\Big)^{1/q}, \quad \frac{1}{p} + \frac{1}{q} = 1, \quad x_i, y_i \geq 0$$

(подробнее см. в [77, гл. II]) с параметрами $x_1 = e_1/t$, $x_2 = (1-e_1)/(2\sqrt{1-t}\,)$, $y_1 = t$, $y_2 = 2\sqrt{1-t}$, $p = 2\chi$, $q = 2\chi/(2\chi-1)$, получаем

$$\left(\frac{e_1}{t}\right)^{2\chi} + \left(\frac{1-e_1}{2\sqrt{1-t}}\right)^{2\chi} \geq \frac{(e_1 + (1-e_1))^{2\chi}}{\left(t^{\frac{2\chi}{2\chi-1}} + (2\sqrt{1-t}\,)^{\frac{2\chi}{2\chi-1}}\right)^{2\chi-1}} =$$
$$\left(t^{\frac{2\chi}{2\chi-1}} + (4(1-t))^{\frac{\chi}{2\chi-1}}\right)^{1-2\chi}. \quad (2.50)$$

Ввиду $\chi < 1$ функция $g(t) = t^{\frac{2\chi}{2\chi-1}} + (4(1-t))^{\frac{\chi}{2\chi-1}}$ является возрастающей в точке 1. Поэтому она не превосходит единицы на отрезке $[t_\chi, 1]$, где t_χ — ближайший к единице корень уравнения $g(t) = 1$. Если $\chi \le 2/3$ (а это так при $k \ge 2$), то на отрезке $[3/4, 1]$ выполнено $g(t) \le t^4 + 16(1-t)^2$. Следовательно, $t_\chi \le t_{2/3} \approx 0.812 < 0.815$.

Поскольку по условию $t = (a_1 + b_1)/2 \ge 0.815$, то (2.49) с учетом (2.50) продолжается как

$$s^\chi(G_1) + \ldots + s^\chi(G_k) \ge (g(t))^{1-2\chi} \left(\frac{|E|^2}{|A| \cdot |B|}\right)^\chi \ge \left(\frac{|E|^2}{|A| \cdot |B|}\right)^\chi = s^\chi(G).$$

II. В оставшемся случае для любого i выполняется $a_i b_i \le (a_i + b_i)^2/4 \le 0.815^2$. Тогда, опять используя неравенство Гёльдера (с параметрами $x_i = e_i$, $y_i = 1$, $p = 2\chi$, $q = 2\chi/(2\chi-1)$) и подставляя $\chi = 1/2 + 1/(10\ln k)$, выводим

$$s^\chi(G_1) + \ldots + s^\chi(G_k) \ge \left(\left(\frac{e_1}{0.815}\right)^{2\chi} + \ldots + \left(\frac{e_k}{0.815}\right)^{2\chi}\right) s^\chi(G) \ge$$
$$\frac{(e_1 + \ldots + e_k)^{2\chi}}{k^{2\chi-1} \cdot 0.815^{2\chi}} \cdot s^\chi(G) \ge \frac{s^\chi(G)}{e^{1/5} \cdot 0.815} > s^\chi(G). \quad \square$$

Объединяя утверждение 2.5, теоремы 2.6 и 2.7, получаем

$$\frac{1}{2} + \frac{1}{10\ln k} \le \chi_k \le \chi_k^* \le \chi_{U_k} \le \frac{1}{2} + \frac{1}{2\log_2(k/2)}.$$

2.6.3 Уточнение нижней оценки экспоненты Храпченко при $k = 3$

Далее мы попробуем аккуратнее оценить величину χ_{U_3}. Практически нет сомнений в том, что оценка теоремы 2.6 в случае $k = 3$ точна. Но на пути доказательства нижней оценки мы вынуждены прибегнуть к нескольким упрощениям, каждое из которых отдаляет нас от истинного значения экспоненты.

Теорема 2.8. *Справедливо* $\chi_3^* \ge 0.769$.

Доказательство. Пусть графы $G_i = (A_i, B_i, E_i)$, $i = 1, 2, 3$, образуют монотонное покрытие графа $G = (A, B, E)$. При этом $s(G) = \frac{|E|^2}{|A| \cdot |B|}$ (тогда, в

частности, $A = \cup A_i$ и $B = \cup B_i$). Докажем, что при $\chi = 0.769$ выполнено

$$s^\chi(G_1) + s^\chi(G_2) + s^\chi(G_3) \geq s^\chi(G). \tag{2.51}$$

I. Сначала разберем простой случай, когда одна из долей графа G принадлежит одному из графов покрытия. Например, пусть $B_3 = B$. Рассмотрим граф $G' = (A \setminus A_3,\, B,\, E \setminus E_3)$. Легко видеть, что графы G' и G_3 образуют монотонное покрытие графа G. Поэтому $s(G') + s(G_3) \geq s(G)$ согласно утверждению 2.7.

Если также выполнено $B \setminus (B_1 \cup B_2) \neq \emptyset$, то $A_3 = A$ ввиду монотонности исходного покрытия, следовательно, $G_3 = G$, и доказывать нечего. Поэтому можно полагать $B = B_1 \cup B_2$. Пусть

$$G'_1 = (A_1 \cap (A \setminus A_3),\, B_1,\, E_1 \cap (E \setminus E_3)) \subset G_1,$$
$$G'_2 = (A_2 \cap (A \setminus A_3),\, B_2,\, E_2 \cap (E \setminus E_3)) \subset G_2.$$

Проверим, что графы G'_1 и G'_2 образуют монотонное покрытие графа G'. В проверке нуждается только условие монотонности. Пусть оно нарушено, тогда, скажем, выполнено $(A \setminus A_3) \setminus (A_2 \cap (A \setminus A_3)) = A \setminus (A_2 \cup A_3) \neq \emptyset$ и $B \setminus B_1 \neq \emptyset$. Получается противоречие с монотонностью покрытия графа G. Поэтому заключаем $s(G'_1) + s(G'_2) \geq s(G')$ по утверждению 2.7. Окончательно имеем

$$s(G_1) + s(G_2) + s(G_3) \geq s(G'_1) + s(G'_2) + s(G_3) \geq s(G') + s(G_3) \geq s(G).$$

Следовательно, при любом $\chi \leq 1$ соотношение (2.51) выполнено.

II. Перейдем к основному случаю, в котором ни одна из долей графа G не лежит в графе из покрытия. В этом случае любая вершина графа G принадлежит хотя бы двум графам из G_1, G_2, G_3. В противном случае, если бы выполнялось, скажем, $A \setminus (A_1 \cup A_2) \neq \emptyset$, то в силу монотонности покрытия имело место $B = B_3$. На языке формул рассматриваемая ситуация соответствует вычислению функции f как $f = m_3(f_1, f_2, f_3)$.

План доказательства — принципиально такой же как в теореме 2.7. В конечном счете рассуждение сводится к рассмотрению двух подслучаев, определяемых наличием в покрытии графа G_i, накрывающего значительную часть графа G (пп. IV и V–VII). Разбору подслучаев предшествует

подготовительная работа, состоящая в учете свойства монотонности покрытия.

Разложим множество A произвольным образом на три непересекающихся подмножества N_1, N_2, N_3 с условиями

$$N_1 \subset A_2 \cap A_3, \quad N_2 \subset A_1 \cap A_3, \quad N_3 \subset A_1 \cap A_2.$$

Аналогично поступим с множеством B:

$$P_1 \subset B_2 \cap B_3, \quad P_2 \subset B_1 \cap B_3, \quad P_3 \subset B_1 \cap B_2.$$

По условию основного случая все множества N_i и P_i непусты. Обозначим $e_{ij} = |E \cap (N_i \times P_j)|/|E|$, $a_i = |N_i|/|A|$, $b_i = |P_i|/|B|$.

С целью компактности записи введем обозначение

$$s_*(G') = s(G') \cdot \frac{|A| \cdot |B|}{|E|^2}$$

для произвольного графа $G' \subset G$. Тогда $s_*(G) = 1$.

В силу $(N_i \cup N_j) \subset A_k$ и $(P_i \cup P_j) \subset B_k$ при $\{i,\, j,\, k\} = \{1,\, 2,\, 3\}$ выполнено

$$s_*(G_k) \geq \max \left\{ \frac{(e_{ii} + e_{ij} + e_{ji} + e_{jj})^2}{(a_i + a_j) \cdot (b_i + b_j)},\, \frac{e_{ij}^2}{a_i \cdot b_j},\, \frac{e_{ji}^2}{a_j \cdot b_i} \right\}. \tag{2.52}$$

Конечно, также верны соотношения вида $s_*(G_k) \geq (e_{ii} + e_{ij})^2/(a_i(b_i + b_j))$, но мы исключаем их из дальнейшего анализа с целью его упрощения, тем самым уходя с пути доказательства оценки $\chi_3^* \geq \log_4 3$.

Заметим, что графы $G'_k = (N_k, B, E \cap (N_k \times B))$ и $G''_k = (A, P_k, E \cap (A \times P_k))$ обладают монотонными покрытиями G'_{ik}, G'_{jk} и соответственно G''_{ik}, G''_{jk}, где $G'_{ik} = (N_k,\, B_i,\, E_i \cap (N_k \times B))$ и $G''_{ik} = (A_i,\, P_k,\, E_i \cap (A \times P_k))$, при этом $G'_{ik}, G''_{ik} \subset G_i$. Поэтому согласно утверждению 2.7

$$s_*(G_i) + s_*(G_j) \geq \max\{s_*(G'_k),\, s_*(G''_k)\} \geq \max \left\{ \frac{(e_{k1} + e_{k2} + e_{k3})^2}{a_k},\, \frac{(e_{1k} + e_{2k} + e_{3k})^2}{b_k} \right\}. \tag{2.53}$$

Положим $e^*_{ij} = e_{ij} + e_{ji}$ при $i \neq j$ и обозначим $\alpha_{ij} = \sqrt{(a_i + a_j)(b_i + b_j)}$, $\beta_{ij} = \sqrt{a_i b_j} + \sqrt{a_j b_i}$. Применяя неравенство (2.46), оценки (2.52) и (2.53) перепишем как

$$s_*(G_k) \geq \max\{H_k,\, Q_k\}, \qquad H_k = \frac{(e_{ii} + e^*_{ij} + e_{jj})^2}{\alpha^2_{ij}}, \quad Q_k = \frac{(e^*_{ij})^2}{\beta^2_{ij}}, \tag{2.54}$$

$$s_*(G_i) + s_*(G_j) \geq F_{ij} = \frac{(2e_{kk} + e^*_{ik} + e^*_{jk})^2}{\gamma^2_k}, \tag{2.55}$$

где $\gamma_k = \sqrt{a_k} + \sqrt{b_k}$. Напомним, что при $\chi \leq 1$ из (2.55) вытекает $s^\chi_*(G_i) + s^\chi_*(G_j) \geq F^\chi_{ij}$.

III. Итак, далее мы доказываем, что из (2.54) и (2.55) при подходящем выборе χ вытекает

$$1 = s^\chi_*(G) \leq s^\chi_*(G_1) + s^\chi_*(G_2) + s^\chi_*(G_3). \tag{2.56}$$

Теперь задача сведена к доказательству числовых неравенств.

При фиксации значений всех сумм $a_i + b_i$ знаменатели в оценках (2.54) и (2.55) принимают максимальные значения при $a_i = b_i$ для всех i:

$$\alpha_{ij} = \sqrt{(a_i + a_j)(b_i + b_j)} \leq \frac{a_i + b_i}{2} + \frac{a_j + b_j}{2},$$
$$\beta_{ij} = \sqrt{a_i b_j} + \sqrt{a_j b_i} \leq \sqrt{(a_i + b_i)(a_j + b_j)},$$
$$\gamma_k = \sqrt{a_k} + \sqrt{b_k} \leq \sqrt{2(a_k + b_k)}$$

(проверяется возведением в квадрат и применением простых неравенств о средних). Поэтому далее будем полагать, что $a_i = b_i$ при всех i. Следовательно, $\alpha_{ij} = a_i + a_j = 1 - a_k$ (где $k \neq i, j$), $\beta_{ij} = 2\sqrt{a_i a_j}$ и $\gamma_k = 2\sqrt{a_k}$.

Кроме того, можно считать, что $H_k \leq Q_k$ при всех k. Действительно, если, например, $H_1 > Q_1$, то при некотором $\varepsilon > 0$ замена $e^*_{23} := e^*_{23} + 2\varepsilon$, $e_{22} := e_{22} - \varepsilon$, $e_{33} := e_{33} - \varepsilon$ приводит к равенству $H_1 = Q_1$. При этом указанная замена сохраняет сумму $\sum e_{ij}$ и не увеличивает оценки (2.54) и (2.55) для величин $s_*(G_i)$, поскольку все оценки для H_i, Q_i, F_{ij}, за исключением оценки Q_1, не увеличиваются.

Из $H_k \leq Q_k$ следует

$$\frac{\alpha_{ij}}{\beta_{ij}} \cdot e^*_{ij} \geq e_{ii} + e_{jj} + e^*_{ij} \tag{2.57}$$

при всех $i \neq j$ и далее при обозначениях $x_{ij} = e^*_{ij}/\beta_{ij}$

$$(\alpha_{12} + \beta_{12})x_{12} + (\alpha_{13} + \beta_{13})x_{13} + (\alpha_{23} + \beta_{23})x_{23} \geq 2\sum e_{ij} = 2. \qquad (2.58)$$

IV. Теперь воспроизведем фрагмент доказательства теоремы 2.7. Из (2.57) следует

$$F_{ij} = \frac{(1 + e_{kk} - e_{ii} - e_{jj} - e^*_{ij})^2}{\gamma_k^2} \geq \frac{\left(1 - \frac{\alpha_{ij}}{\beta_{ij}} \cdot e^*_{ij}\right)^2}{\gamma_k^2} = \frac{(1 - \alpha_{ij}x_{ij})^2}{\gamma_k^2}.$$

Этот переход еще немного понижает потенциально возможную оценку для χ, но упрощает дальнейшие рассуждения. Теперь при любой перестановке $\{i, j, k\} = \{1, 2, 3\}$ выполнено

$$s^\chi_*(G_1) + s^\chi_*(G_2) + s^\chi_*(G_3) \geq Q^\chi_k + F^\chi_{ij} \geq x^{2\chi}_{ij} + \left(\frac{1 - \alpha_{ij}x_{ij}}{\gamma_k}\right)^{2\chi}. \qquad (2.59)$$

Применяя неравенство Гёльдера к правой части (2.59), получаем

$$x^{2\chi}_{ij} + \left(\frac{1 - \alpha_{ij}x_{ij}}{\gamma_k}\right)^{2\chi} \geq \frac{(\alpha_{ij}x_{ij} + (1 - \alpha_{ij}x_{ij}))^{2\chi}}{\left(\alpha_{ij}^{\frac{2\chi}{2\chi-1}} + \gamma_k^{\frac{2\chi}{2\chi-1}}\right)^{2\chi-1}} = \left(\alpha_{ij}^{\frac{2\chi}{2\chi-1}} + \gamma_k^{\frac{2\chi}{2\chi-1}}\right)^{1-2\chi}.$$

Таким образом, при условии

$$\alpha_{ij}^{\frac{2\chi}{2\chi-1}} + \gamma_k^{\frac{2\chi}{2\chi-1}} \leq 1 \qquad (2.60)$$

правая часть (2.59) не меньше $s^\chi_*(G) = 1$ при любом возможном $x_{ij} \geq 0$.

Напомним, что $\alpha_{ij} = 1 - a_k$ и $\gamma_k = 2\sqrt{a_k}$. Ввиду $\chi < 1$ функция $g(t) = (1 - t)^{\frac{2\chi}{2\chi-1}} + (2\sqrt{t})^{\frac{2\chi}{2\chi-1}}$ является убывающей в нуле. Поэтому она не превосходит единицы на отрезке $[0, t_\chi]$, где t_χ — минимальный положительный корень уравнения $g(t) = 1$. Тогда при $a_k \leq t_\chi$ справедливо (2.60) и как следствие выполняется (2.56).

V. Нам остается рассмотреть случай, когда при всех i выполнено

$$a_i \geq t_\chi. \qquad (2.61)$$

В силу (2.54) имеет место соотношение

$$s^\chi_*(G_1) + s^\chi_*(G_2) + s^\chi_*(G_3) \geq Q^\chi_1 + Q^\chi_2 + Q^\chi_3 = x^{2\chi}_{12} + x^{2\chi}_{13} + x^{2\chi}_{23}.$$

Оценим значение χ, при котором неравенство

$$x_{12}^{2\chi} + x_{13}^{2\chi} + x_{23}^{2\chi} \geq 1 \tag{2.62}$$

справедливо для любых x_{ij} и значений параметров a_i при ограничениях (2.58) и (2.61). Достаточно рассмотреть случай, когда в (2.58) выполняется равенство; с обозначением $c_{ij} = (\alpha_{ij} + \beta_{ij})/2$ условие (2.58) принимает вид

$$c_{12}x_{12} + c_{13}x_{13} + c_{23}x_{23} = 1.$$

Применяя неравенство Гёльдера, получаем

$$x_{12}^{2\chi} + x_{13}^{2\chi} + x_{23}^{2\chi} \geq \frac{(c_{12}x_{12} + c_{13}x_{13} + c_{23}x_{23})^{2\chi}}{\left(c_{12}^{\frac{2\chi}{2\chi-1}} + c_{13}^{\frac{2\chi}{2\chi-1}} + c_{23}^{\frac{2\chi}{2\chi-1}}\right)^{2\chi-1}} = \left(c_{12}^{\frac{2\chi}{2\chi-1}} + c_{13}^{\frac{2\chi}{2\chi-1}} + c_{23}^{\frac{2\chi}{2\chi-1}}\right)^{1-2\chi}.$$

Теперь достаточно при ограничениях (2.61) доказать соотношение

$$c_{12}^{\frac{2\chi}{2\chi-1}} + c_{13}^{\frac{2\chi}{2\chi-1}} + c_{23}^{\frac{2\chi}{2\chi-1}} \leq 1. \tag{2.63}$$

VI. Заметим, что $c_{ij} = (\sqrt{a_i} + \sqrt{a_j})^2/2$. Положим

$$g(x,y) = \left(\sqrt{x} + \sqrt{y}\right)^\theta + \left(\sqrt{x} + \sqrt{z}\right)^\theta + \left(\sqrt{y} + \sqrt{z}\right)^\theta, \quad z = 1 - x - y,$$

где $\theta = \frac{4\chi}{2\chi-1}$. Обозначим $D = \{(x,\, y) \mid x, y \geq t_\chi,\ x + y \leq 1 - t_\chi\}$. Для вывода (2.63) достаточно убедиться в том, что

$$\max_{(x,\,y)\in D} g(x,y) \leq 2^{\theta/2}. \tag{2.64}$$

Выберем $\chi = 0.769$, тогда $\theta \approx 5.717$ и $t_\chi \approx 0.09314$. Неравенство (2.64) легко проверить, применяя ограниченный компьютерный перебор значений функции g в узлах решетки с подходящим малым шагом, с учетом ограниченности частных производных. Однако приведем аналитическое решение.

Максимум функции $g(x,y) \in C^1(D)$ достигается либо на границе области D, либо в некоторой точке локального экстремума, где $g'_x = g'_y = 0$. Сначала проверим второй вариант.

Обозначим $A_{xy} = \left(\sqrt{x} + \sqrt{y}\right)^{\theta-1}$. Тогда условие $g'_x(x,y) = g'_y(x,y) = 0$ записывается как

$$\frac{A_{xy} + A_{xz}}{\sqrt{x}} = \frac{A_{xz} + A_{yz}}{\sqrt{z}} = \frac{A_{xy} + A_{yz}}{\sqrt{y}}. \tag{2.65}$$

Отсюда, в частности, следует $(\sqrt{x} - \sqrt{y})(A_{xz} + A_{yz}) = \sqrt{z}(A_{xz} - A_{yz})$ и далее в силу симметрии

$$\frac{A_{xy}}{\sqrt{x} + \sqrt{y} - \sqrt{z}} = \frac{A_{xz}}{\sqrt{x} + \sqrt{z} - \sqrt{y}} = \frac{A_{yz}}{\sqrt{y} + \sqrt{z} - \sqrt{x}} \tag{2.66}$$

(ни один из знаменателей не обращается в ноль, поскольку A_{xy}, A_{xz}, A_{yz} не равны нулю).

Мы не будем искать сами точки экстремума, т. е. решение уравнений (2.66). Вместо этого заметим, что в точке экстремума (такие точки в области D есть, например, $x = y = 1/3$) функция g принимает значение

$$\begin{aligned}(\sqrt{x} + \sqrt{y})A_{xy} + (\sqrt{x} + \sqrt{z})A_{xz} + (\sqrt{y} + \sqrt{z})A_{yz} = \\ \left(\sqrt{x} + \sqrt{y} + \frac{(\sqrt{x} + \sqrt{z})(\sqrt{x} + \sqrt{z} - \sqrt{y}) + (\sqrt{y} + \sqrt{z})(\sqrt{y} + \sqrt{z} - \sqrt{x})}{\sqrt{x} + \sqrt{y} - \sqrt{z}}\right) \cdot \\ \cdot A_{xy} = \frac{2A_{xy}}{\sqrt{x} + \sqrt{y} - \sqrt{z}} = \frac{2\left(\sqrt{x} + \sqrt{y}\right)^{\theta-1}}{\sqrt{x} + \sqrt{y} - \sqrt{z}}.\end{aligned} \tag{2.67}$$

Покажем, что выражение в правой части (2.67) не превосходит $2^{\theta/2}$ при любых $(x,\, y) \in D$. Не ограничивая общности, полагаем $x \geq z \geq y$.

Если сумма $\sigma = \sqrt{x} + \sqrt{y}$ фиксирована, то значение (2.67) возрастает с ростом z, т. е. при убывании суммы $x + y$. При $z \neq x$ и $z \neq y$ значение (2.67) можно увеличить, сближая x с y и сохраняя σ. Поэтому мы ищем максимум величины (2.67) при условиях $x = z \geq y$ или $x \geq z = y$. В обоих случаях правая часть (2.67) представляется в виде $\xi(u) = \frac{2\left(\sqrt{u} + \sqrt{(1-u)/2}\right)^{\theta-1}}{\sqrt{u}}$, где $u = x$ или $u = y$, а интервал для поиска максимума можно определить как $u \in [t_\chi,\, 1]$ в силу $t_\chi \leq y \leq x < 1$.

Анализируя производную, устанавливаем, что на рассматриваемом отрезке функция $\xi(u)$ принимает максимальное значение в точке экстремума,

определяемой из уравнения

$$(\theta-1)\left(1-\frac{1}{2}\sqrt{\frac{2u}{1-u}}\right)=1+\sqrt{\frac{1-u}{2u}}.$$

Уравнение разрешается как $u=\frac{1}{1+2\tau^2}$, где $\tau=\left(\theta-2-\sqrt{\theta^2-6\theta+6}\right)/2\approx 0.8118$ (другое решение лежит за пределами отрезка). В указанной точке $\xi(u)\approx 6.92<2^{\theta/2}\approx 7.25$.

VII. Осталось исследовать значения функции $g(x,y)$ на границе области D. Можем считать, что $z=t_\chi$. Покажем, что максимум функции одной переменной

$$g_1(x)=\left(\sqrt{x}+\sqrt{y}\right)^\theta+\left(\sqrt{x}+\sqrt{t_\chi}\right)^\theta+\left(\sqrt{y}+\sqrt{t_\chi}\right)^\theta,\quad y=1-t_\chi-x,$$

достигается при $x=y=(1-t_\chi)/2$. Полагая $x\le y$, ограничим рассмотрение отрезком $x\in[t_\chi,\,(1-t_\chi)/2]$. Теперь достаточно доказать, что на указанном отрезке выполнено $g_1'(x)\ge 0$.

Производная функции g_1 имеет вид

$$g_1'(x)=\frac{\theta\cdot(\varphi(y)-\varphi(x))}{2\sqrt{xy}},\quad \varphi(x)=\sqrt{x}\left(\left(\sqrt{x}+\sqrt{y}\right)^{\theta-1}+\left(\sqrt{y}+\sqrt{t_\chi}\right)^{\theta-1}\right).$$

Нам достаточно показать, что функция $\varphi(x)$ возрастает на отрезке $[t_\chi,\,(1-t_\chi)/2]$, т.е. $\varphi'(x)\ge 0$. Действительно, в силу $t_\chi\le x\le y$ и $\theta>4$

$$\begin{gathered}2\varphi'(x)=\frac{1}{\sqrt{x}}\left(\sqrt{x}+\sqrt{y}\right)^{\theta-1}+(\theta-1)\sqrt{x}\left(\frac{1}{\sqrt{x}}-\frac{1}{\sqrt{y}}\right)\left(\sqrt{x}+\sqrt{y}\right)^{\theta-2}+\\ \frac{1}{\sqrt{x}}\left(\sqrt{y}+\sqrt{t_\chi}\right)^{\theta-1}-(\theta-1)\frac{\sqrt{x}}{\sqrt{y}}\left(\sqrt{y}+\sqrt{t_\chi}\right)^{\theta-2}\ge\\ 4(x+y)\left(\sqrt{x}+\sqrt{y}\right)^{\theta-4}+\\ +\left(y+3\sqrt{t_\chi y}-(\theta-1)\left(\sqrt{xy}+t_\chi+2\sqrt{t_\chi x}\right)\right)\left(\sqrt{y}+\sqrt{t_\chi}\right)^{\theta-4}\ge\\ \left(4x+5y-(\theta-1)\sqrt{xy}-(\theta-1)t_\chi-(2\theta-5)\sqrt{t_\chi x}\,\right)\left(\sqrt{y}+\sqrt{t_\chi}\right)^{\theta-4}>0,\end{gathered}$$

поскольку

$$4x + 5y - (\theta - 1)\sqrt{xy} - (\theta - 1)t_\chi - (2\theta - 5)\sqrt{t_\chi x} \geq$$
$$(10 - \theta) \cdot \frac{x + y}{2} - (\theta - 1)t_\chi - (2\theta - 5)\sqrt{t_\chi (x + y)/2} =$$
$$(10 - \theta)(1 - t_\chi)/2 - (\theta - 1)t_\chi - (2\theta - 5)\sqrt{t_\chi (1 - t_\chi)/2} \approx 0.18.$$

Окончательно получаем

$$\max_{(x,\, y) \in D} g(x, y) = g_1((1 - t_\chi)/2) \approx 7.252 < 2^{\theta/2} \approx 7.254.$$

Таким образом, при выбранном значении χ установлено (2.64), а вместе с тем (2.56), откуда следует (2.51). □

2.6.4 Верхние оценки сложности

В заключение мы покажем, что полученные нижние оценки сложности до определенной степени точны применительно к линейной функции. Для функции голосования и симметрических функций вообще, как и в известных результатах для бинарных базисов, зазор между нижними и верхними оценками существенно шире.

Теорема 2.9. (*i*) *При всех* $k \geq 2$ *справедливо*

$$L_{U_{2k}}(l_n) \leq 2n^{1+1/\log k}, \qquad D_{U_{2k}}(l_n) \leq \lceil \log_k n \rceil.$$

(*ii*) *При любом* $k \geq 3$ *выполнено*

$$L_{U_{2k}}(m_n) = O\left(n^{1+c_1 \ln\ln k/\ln k} \cdot (\ln n)^{c_2 \ln\ln k}\right),$$

где c_1, c_2 *— некоторые константы.*

Доказательство. Сначала заметим, что любая функция f от k переменных реализуется в базисе U_{2k} формулой, в которую каждая переменная входит не более двух раз, в частности, имеет место

$$L_{U_{2k}}(f) \leq 2k. \tag{2.68}$$

Для этого рассмотрим произвольную дизъюнктивную нормальную форму (ДНФ) функции $f(x_1, \ldots, x_k)$ и заменим в ней вхождения отрицаний переменных x_i новыми переменными y_i. Получим формулу для некоторой монотонной функции $\varphi(x_1, \ldots, x_k, y_1, \ldots, y_k) \in U_{2k}$. Из построения очевидно, что формула

$$\varphi(x_1, \ldots, x_k, \overline{x_1}, \ldots, \overline{x_k}) \tag{2.69}$$

сложности $2k$ реализует f.

Докажем п. (i). Из (2.68) следует, что при $n \leq k$ верно $L_{U_{2k}}(l_n) \leq 2n$. Теперь несложно вывести по индукции оценку $L_{U_{2k}}(l_n) \leq n \cdot 2^{\lceil \log_k n \rceil}$. Используя формулу типа (2.69) для функции l_k, при $n = n_1 + \ldots + n_k$ получаем рекуррентное соотношение

$$L_{U_{2k}}(l_n) \leq 2(L_{U_{2k}}(l_{n_1}) + \ldots + L_{U_{2k}}(l_{n_k})).$$

Для обоснования индуктивного перехода остается заметить, что произвольное число $n \in (k^{d-1}, k^d]$ представляется суммой k чисел $n_i \leq k^{d-1}$. Глубина построенной формулы равна $\lceil \log_k n \rceil$.

Перейдем к доказательству п. (ii). Вычислим арифметическую сумму булевых переменных $x_1, \ldots, x_n$. Для этого воспользуемся простейшим вариантом метода компрессоров. В качестве компрессора выбираем полный сумматор k битов. Каждый из $r = \lceil \log k \rceil$ разрядов суммы k битов имеет сложность не выше $2k$ согласно (2.68). Посредством параллельных копий компрессора k многоразрядных чисел можно преобразовать в r чисел с сохранением суммы. При этом максимальная сложность разряда слагаемых возрастает не более чем в $2k$ раз. При помощи дерева компрессоров сложение $n \leq (k/r)^d$ чисел (в нашем случае роль чисел играют булевы переменные x_i) сводится к сложению r чисел со сложностью каждого разряда не выше $(2k)^d$.

Далее заметим, что любой разряд суммы двух p-разрядных чисел имеет линейную сложность, не выше cp, даже над базисом U_2. Тогда посредством дерева из $r - 1$ сумматоров сумма r штук p-разрядных чисел вычисляется с увеличением максимальной сложности разряда в $(cp)^{\lceil \log r \rceil}$ раз.

Наконец, функция сравнения p-разрядного числа с фиксированным порогом выражается бесповторной формулой в базисе $\{\vee, \wedge\}$. Объединяя все

перечисленные оценки, для сложности функции голосования (и любой другой пороговой симметрической функции) n переменных получаем верхнюю оценку (по условию задачи $p = \lceil \log_2 n \rceil$ и $d = \lceil \log_{k/r} n \rceil$):

$$L_{U_{2k}}(m_n) \leq p \cdot (2k)^d (cp)^{\lceil \log r \rceil} \leq n^{\log_{k/r}(2k)} (\ln n)^{O(\ln r)}. \qquad \square$$

Как следствие из теорем 2.7 и 2.9, при $k \geq 4$ для глубины линейной функции получаем двусторонние оценки:

$$\log_k n + \frac{\ln n}{5 \ln^2 k} \leq D_{U_k}(l_n) \leq \lceil \log_{\lfloor k/2 \rfloor} n \rceil \leq \log_k n + \frac{2 \ln n}{\ln^2 k} + 1.$$

3 Линейные схемы ограниченной глубины

3.1 Введение

Пусть $(S, +)$ — коммутативная полугруппа, т. е. множество S замкнуто относительно бинарной ассоциативной и коммутативной операции $+$. Рассматривается задача вычисления системы сумм

$$y_i = \sum_{j \in T_i} x_j, \qquad i = 1, \ldots, m$$

с использованием только операции сложения в полугруппе. Это вычисление отвечает реализации линейного оператора $y = Ax$ с булевой матрицей A.

Естественная вычислительная модель, возникающая при решении данной задачи — *линейные схемы* над $(S, +)$. Линейная схема определяется как ориентированный ациклический граф с n вершинами-входами $x_1, \ldots, x_n$, не имеющими входящих ребер, и m вершинами-выходами $y_1, \ldots, y_m$, не имеющими исходящих ребер. В каждой вершине, кроме входов, вычисляется сумма над $(S, +)$ по всем входящим ребрам. Сложность схемы — общее число ребер в ней, а глубина — число ребер в самом длинном ориентированном пути.

Мы сосредоточим внимание на трех основных полугруппах: OR-полугруппе $(\{0,\, 1\}, \vee)$, XOR-группе $(\{0,\, 1\}, \oplus)$ и SUM-полугруппе $(\mathbb{N}, +)$.

При вычислениях в OR-полугруппе существенную роль играет правило сокращения $x + x = x$, при вычислениях в XOR-группе — правило сокращения $x + x = 0$. При вычислениях в SUM-полугруппе указанные правила не действуют.

Линейные схемы над OR-полугруппой (OR-схемы) представляют собой простейшую монотонную модель вычислений, XOR-схемы — простейшую групповую модель вычислений, SUM-схемы — универсальны в том смысле, что схема, вычисляющая оператор с матрицей A над SUM-полугруппой, вычисляет оператор с этой же матрицей над произвольной коммутативной полугруппой $(S, +)$.

Понятие OR-схемы совпадает с понятием вентильной схемы, введенной О. Б. Лупановым в работе [41] (именно с этой работы начинается изуче-

ние вентильных и, более широко, линейных схем). XOR-схемы в литературе часто называются линейными булевыми схемами. Понятие SUM-схемы совпадает с понятием вентильной схемы в современном смысле [32] и соответствует понятию (векторной) аддитивной цепочки (см., например, [25, §4.6.3]).

Важно отметить, что вычисление линейного оператора Ax с матрицей $A = (a_{i,j})$ линейной схемой является «кодированием» матрицы путями в ориентированном ациклическом графе. А именно, если $p_{i,j}$ — число путей между входом x_j и выходом y_i в схеме, то схема вычисляет матрицу A в следующем смысле:

– SUM-схема: $a_{i,j} = p_{i,j}$;

– OR-схема: $a_{i,j} = (p_{i,j} > 0)$;

– XOR-схема: $a_{i,j} = (p_{i,j} \bmod 2)$.

Сложность (число ребер) минимальной L-схемы, $\mathsf{L} \in \{\mathsf{SUM}, \mathsf{OR}, \mathsf{XOR}\}$, для матрицы A будем обозначать через $\mathsf{L}(A)$; сложность минимальной схемы глубины d — через $\mathsf{L}_d(A)$.

Можно выделить следующие основные направления исследования сложности линейных булевых операторов:

– Получение асимптотически точных оценок сложности для классов булевых матриц. К этому направлению относятся самые первые результаты, полученные О.Б. Лупановым и Э.И. Нечипоруком в 1950–60 гг.

– Получение нижних оценок сложности для индивидуальных конкретно (эффективно) заданных матриц. Для монотонных линейных схем известны почти квадратичные оценки. Проблема нелинейных нижних оценок XOR-сложности является одной из основных открытых проблем в теории сложности.

– Изучение сложности часто встречающихся матриц (полные треугольные матрицы, матрицы Сильвестра, матрицы Серпинского и др.).

– Выяснение вопроса, насколько сильно могут отличаться SUM-, OR- и XOR-сложность одной и той же матрицы. Это направление активно развивается в последние годы.

– Сравнительное исследование вычислительных возможностей линейных схем (прежде всего, XOR-схем) и схем более общего вида, в вершинах

которых допускается вычисление произвольных булевых функций.

Краткие обзоры результатов в обсуждаемой области (с акцентом на асимптотических оценках) выполнены О. Б. Лупановым [48] и В. В. Кочергиным [32]. Подробный обзор опубликован совместно С. Юкной и автором в [230].

В данной главе представлено решение проблемы асимптотически оптимального синтеза линейных схем ограниченной глубины для класса булевых матриц вместе с некоторыми обобщениями [223] (раздел 3.2), а также изложен ряд результатов о расхождениях между различными мерами сложности матриц [215, 216, 230, 238] (раздел 3.3).

В заключительном разделе 3.4 применение аппарата линейных схем демонстрируется в доказательстве нижних оценок монотонной сложности симметрической пороговой функции T_n^2 [226].

3.2 Асимптотические оценки сложности для классов булевых и целочисленных матриц при ограничении глубины

В материале настоящего раздела L обозначает сложность любого из типов линейных схем: SUM, OR, XOR, если речь идет о булевых матрицах. Оценки, относящиеся к небулевым матрицам, подразумевают использование SUM-схем.

Обозначим $E_q = \{0, 1, \ldots, q-1\}$. Пусть $\mathsf{L}_d(q, m, n)$ означает функцию Шеннона сложности реализации класса матриц размера $m \times n$ (m строк и n столбцов) с элементами из E_q с глубиной d; $\mathsf{L}(q, m, n)$ — обозначение для функции Шеннона без ограничения на глубину.

В силу симметрии, $\mathsf{L}_d(q, m, n) = \mathsf{L}_d(q, n, m)$ (схема для произвольной матрицы A превращается в схему для транспонированной матрицы A^T изменением ориентации ребер). Поэтому при изложении следующих известных результатов полагаем $m \leq n$.

Для компактности формулировок, в рамках этого раздела определим $\log x = \max\{\log_2 x, 1\}$.

В работе [41] О. Б. Лупанов получил первые результаты о сложности

класса булевых матриц. Он доказал, что при $m = \omega(\log n)$

$$\mathsf{L}_2(2, m, n) \sim \frac{mn}{\log n},$$

а при дополнительном ограничении $\log m = o(\log n)$

$$\mathsf{L}(2, m, n) \sim \mathsf{L}_2(2, m, n).$$

В случае полиномиально эквивалентных размеров $\log m \asymp \log n$ асимптотики $\mathsf{L}(2, m, n)$ и $\mathsf{L}_2(2, m, n)$ расходятся. Мощностное рассуждение (см., например, [57]) показывает, что при $m = \Omega\left(\log^2 n\right)$ справедлива нижняя оценка

$$\mathsf{L}(2, m, n) \geq \frac{mn}{\log(mn)} \left(1 + \Theta\left(\frac{\log\log n}{\log n}\right)\right). \tag{3.1}$$

Э. И. Нечипорук [53, 57] установил, что в ряде случаев асимптотика достигается на схемах глубины 3. А именно, если $\log m \sim c_{p,r} \log n$, где $c_{p,r} = \frac{p}{p(r-1)+r}$, $p, r \in \mathbb{N}$, то

$$\mathsf{L}(2, m, n) \sim \mathsf{L}_3(2, m, n) \sim \frac{mn}{\log(mn)}. \tag{3.2}$$

Для классов узких матриц $m \asymp \log n$ асимптотика в глубине 2 почти для всех соотношений между m и n получена В. А. Орловым [60]:

$$\mathsf{L}_2(2, m, n) \sim \lfloor \alpha + 1 \rfloor n, \qquad \frac{m}{\log n} \to \alpha \notin \mathbb{N}.$$

Вопрос об асимптотике $\mathsf{L}(2, m, n)$ остается открытым. А. В. Чашкин [86] показал, что при переходе от OR-схем к схемам в расширенном базисе $\{\vee, \wedge\}$, ожидаемая асимптотика $mn/\log n$ получается.

Н. Пиппенджер [176] провел доказательство в общем случае, получив при $H = mn \log q \to \infty$ универсальную верхнюю оценку

$$\mathsf{L}(q, m, n) \leq 3m \log_3(q-1) + (1 + \tau(H)) \frac{H}{\log H} + O(n), \tag{3.3}$$

где $\tau(H) \asymp \sqrt{\frac{\log\log H}{\log H}}$, асимптотически точную в случае $\log n = o(m \log q)$, как показывает обобщающая (3.1) нижняя оценка, приведенная в [176]:

$$\mathsf{L}(q, m, n) \geq 3m \log_3(q-1) + \left(1 - \Theta\left(\frac{\log\log H}{\log H}\right)\right) \frac{H}{\log H}. \tag{3.4}$$

В частности, этот результат Пиппенджера закрывает вопрос об асимптотике сложности класса булевых матриц размера $m \times n$, $m = \omega(\log n)$. Однако, формально схемы Пиппенджера имеют растущую глубину, поэтому в ряде работ (например, [32, 230]) поставлен вопрос о достижимости асимптотики сложности класса булевых матриц схемами ограниченной глубины.

Автором в работе [223] получен утвердительный ответ на этот вопрос, о чем речь пойдет ниже[18]. Будет показано, что при $m = \Omega\left(\log^{3/2} n\right)$ оценка

$$\mathsf{L}_d(q, m, n) \leq (1 + \tau(n))\frac{mn}{\log_q(mn)} \tag{3.5}$$

в случае $q = n^{o(1)}$ выполняется при $\tau(n) = o(1)$ и $d = 3$; в случае $q = \log^{O(1)} n$ — при $\tau(n) \asymp \sqrt{\frac{\log\log n}{\log n}}$ и $d = 3$, а в случае $q = o(n/\log^2 n)$ — при $\tau(n) \asymp \sqrt{\frac{\log\log n}{\log n}}$ и $d = 4$.

Более того, при дополнительных ограничениях $m = \Omega\left(\log^2 n\right)$ и $m \in n^{\mu}\log^{\pm O(1)} n$ для некоторой постоянной $\mu \in \mathbb{Q}$ оценка (3.5) доказывается с остаточным членом «правильного» порядка $\tau(n) \asymp \frac{\log\log n}{\log n}$ при $q = \log^{O(1)} n$ и $d = 3$, а также при $q = o(n/\log^2 n)$ и $d = 4$.

В заключение, мы приводим универсальную оценку в форме

$$\mathsf{L}(q, m, n) \leq 3m\log_3(q-1) + (1 + \tau(H))\frac{H}{\log H} + n$$

при $H = mn\log q \to \infty$ и $\tau(H) \asymp \sqrt{\frac{\log\log H}{\log H}}$. Эта оценка при $q = n^{O(1)}$ достигается на схемах глубины $O(1)$, а при $n = \log^{O(1)} q$ остаточный член в ней можно уточнить до $\tau(H) \asymp \frac{\log\log H}{\log H}$.

Принципиальная возможность получения перечисленных результатов заложена в конструкции Пиппенджера [176], и требуется лишь взглянуть на нее под нужным углом зрения. По существу, оптимальная схема глубины 3 получается из схемы [176], если, оставив нетронутым средний слой, на котором сосредоточена основная сложность схемы, остальные слои выше и, соответственно, ниже него «склеить». Прием склеивания слоев также применял Нечипорук при выводе оценки (3.2).

[18] Есть сведения о том, что сам Э. И. Нечипорук получил указанный результат около 1970 г., но не успел опубликовать.

Отметим, что задаче о сложности реализации матрицы вентильными схемами подобна задача о сложности реализации матрицы (векторными) *аддитивными цепочками*. В частности, методы оптимального синтеза вентильных схем, как правило, могут быть перестроены в методы оптимального синтеза аддитивных цепочек, и наоборот. Так, оценки, аналогичные (3.3) и (3.4), для аддитивных цепочек доказаны Пиппенджером в [178], а уточнение верхней оценки получено С. Б. Гашковым и В. В. Кочергиным в [9]. Впоследствии В. В. Кочергин в серии работ получил ряд обобщений, в частности, для класса матриц с индивидуальными ограничениями на размер каждого коэффициента, см., например, [30, 31].

Понятия глубины для вентильных схем и для аддитивных цепочек не вполне аналогичны (глубина аддитивной цепочки соответствует глубине вентильной схемы с ограничением 2 на число входящих в каждую вершину ребер), поэтому прямой переносимости результатов, связанных с глубиной, из одной модели в другую, как правило, нет. Решение задачи одновременной минимизации глубины и сложности в асимптотическом смысле для обычной аддитивной цепочки приведено в [10].

Материал излагается в следующем порядке. В §3.2.1 мы приводим вспомогательный результат о свойствах приближений специального вида, возникающих в конструкциях схем типа [176]. В §3.2.2 доказывается основной результат о синтезе схем глубины 3. В §3.2.3 доказывается расширение для глубины 4. В §3.2.4 устанавливаются оценки для класса матриц с ограничением общего вида на размер коэффициентов.

3.2.1 Приближение

В конструкции [176] используются приближения действительных чисел из отрезка [1/2, 1] дробями вида

$$P(r_1,\dots,r_k)=\frac{r_1\cdot\ldots\cdot r_k-r_2\cdot\ldots\cdot r_k+\ldots+(-1)^{k-1}r_k+(-1)^k}{r_1\cdot\ldots\cdot r_k}=$$
$$1-\frac{1}{r_1}\left(1-\frac{1}{r_2}\left(1-\frac{1}{r_3}\left(\ldots\left(1-\frac{1}{r_k}\right)\ldots\right)\right)\right), \quad (3.6)$$

где $r_i\in\mathbb{N}$. Положим формально $P()=1$.

Лемма 3.1. *Пусть* $\alpha \in [1/2,\, 1]$.

(*i*) *При любом* $\delta \in (0,\, 1/2]$ *существуют четное число* $k \geq 0$ *и натуральные числа* $r_1, \ldots, r_k$ *такие, что*

$$0 \leq \varepsilon = P(r_1, \ldots, r_k) - \alpha \leq \delta, \quad R = r_1 \cdot \ldots \cdot r_k \leq \frac{2}{\delta}, \quad k \leq \log_2(2/\delta).$$

(*ii*) *Если* $\alpha = \frac{u}{v}$, *где* $u, v \in \mathbb{N}$, *то при некотором четном* k *имеет место представление*

$$\alpha = P(r_1, \ldots, r_k), \qquad R = r_1 \cdot \ldots \cdot r_k \leq v^v.$$

Доказательство. Будем строить приближение (3.6) для числа α градиентным алгоритмом, как описано ниже.

Введем вспомогательные величины h_i, связанные с представлением (3.6) и определяемые из соотношений

$$h_0 = \alpha, \qquad h_{i-1} = 1 - \frac{1}{r_i} h_i, \quad i > 0. \tag{3.7}$$

Градиентный метод заключается в том, что мы на каждом шаге выбираем максимальное r_i такое, что $h_i \in [0,\, 1]$. Если $h_i = 1$, то процесс заканчивается в силу $\alpha = P(r_1, \ldots, r_i)$.

Несложно проверить, что выбор всегда возможен. Справедливы формулы

$$r_i = \left\lfloor \frac{1}{1 - h_{i-1}} \right\rfloor, \qquad h_i = r_i(1 - h_{i-1}), \tag{3.8}$$

из которых следует, что h_i имеет вид $\lfloor x \rfloor / x$ при $x \geq 1$, поэтому $h_i \in [0,\, 1]$. Кроме того, последовательность $\{h_i\}$ — возрастающая, в силу

$$1 - h_i = 1 - \frac{\lfloor x \rfloor}{x} < \frac{1}{x} = 1 - h_{i-1}.$$

Как следствие, если $h_0 \geq 1/2$, то $h_i > 1/2$ и $r_i \geq 2$ при всех $i > 0$.

Положим $R_0 = 1$ и обозначим $R_i = r_1 \cdot \ldots \cdot r_i$. Вследствие (3.7) и (3.8),

$$P(r_1, \ldots, r_i) - \alpha = (-1)^i \cdot \frac{1 - h_i}{R_i} = (-1)^i \cdot \frac{h_{i+1}}{R_{i+1}}. \tag{3.9}$$

Если $1 - \alpha \leq \delta$, то условия леммы выполнены при $k = 0$. Иначе, определим t из условий $R_t < 1/\delta$ и либо $1/\delta \leq R_{t+1}$, либо $h_t = 1$. Теперь неравенство $0 \leq (-1)^t (P(r_1, \ldots, r_t) - \alpha) \leq \delta$ следует из (3.9).

В случае четного t положим $k = t$. В случае нечетного t положим $k = t + 1$ и $r'_{t+1} = \lceil 1/(\delta R_t) \rceil$. При этом $1/\delta \le R'_{t+1} = R_t r'_{t+1} < 2/\delta$ и $r'_{t+1} \ge 2$. Если $h_t \ne 1$, то $r'_{t+1} \le r_{t+1}$, а точность приближения гарантируется оценкой

$$P(r_1, \ldots, r_t, r'_{t+1}) - \alpha = \frac{1 - \frac{r'_{t+1}}{r_{t+1}} \cdot h_{t+1}}{R'_{t+1}} \in \left[0,\ \frac{1}{R'_{t+1}}\right] \subset [0,\ \delta].$$

Если $h_t = 1$, то

$$P(r_1, \ldots, r_t, r'_{t+1}) - \alpha = P(r_1, \ldots, r_t, r'_{t+1}) - P(r_1, \ldots, r_t) = \frac{1}{R'_{t+1}} \le \delta.$$

Неравенство $k \le \log_2(2/\delta)$ тривиально в силу $r_i \ge 2$ для всех i. Утверждение (i) доказано.

Если α — рациональное число со знаменателем дроби v, то последовательность $\{h_i\}$ состоит из рациональных чисел со знаменателями дробей v. Действительно, обозначая $h_i = u_i/v$, с учетом $u_0 = u \in \mathbb{N}$, по формуле (3.8) получаем $u_i = r_i(v - u_{i-1}) \in \mathbb{N}$ при всех i. Кроме того, из $u_i \le v$ следует $r_i \le v$.

Ввиду $h_i > h_{i-1}$, все u_i различны, поэтому последовательность $\{h_i\}$ содержит не более $v - 1$ членов.

Вместо представления $\alpha = P(r_1, \ldots, r_k)$ с нечетным k можно выбрать представление $\alpha = P(r_1, \ldots, r_{k-1}, r_k - 1, r_k)$. Равенство справедливо в силу тождества $\frac{1}{r} = \frac{1}{r-1}\left(1 - \frac{1}{r}\right)$.

Оценка на R в (ii) следует из $k \le v - 1$ и $r_i \le v$. □

Более аккуратно свойства приближения (3.6) исследуются в [176].

3.2.2 Схемы глубины 3

Как правило, схемы, рассматриваемые далее, имеют слоистую структуру. Это означает, что длина всех путей в схеме одинакова, а множество ребер естественным образом распадается на слои: слой образуют ребра, расположенные на одинаковом расстоянии от входов (или выходов).

В следующих двух леммах описываются простые приемы синтеза схем, комбинация которых приводит к основному результату. Первый прием известен как метод разрезания на полосы Лупанова [41]. Второй является частью метода Нечипорука [57] синтеза схем глубины 3.

Лемма 3.2. *Пусть $s \in \mathbb{N}$. Произвольная (m, n)-матрица над E_q может быть реализована вентильной схемой глубины 2 с $q^s n/s$ внутренними вершинами, $q^{s+1}n$ вентилями на первом слое и $m(n/s+1)$ вентилями на втором слое*[19].

Доказательство. Разобьем матрицу на вертикальные полосы ширины s. На первом уровне реализуем всевозможные суммы в полосах (напомним, что вентильную схему можно интерпретировать как схему из элементов сложения): всего для каждой полосы q^s сумм, для неполной полосы ширины $s' < s$ их менее $(s'/s)q^s$. Для реализации одной суммы требуется не более sq вентилей. На втором слое суммы в строках складываются из сумм в полосах. □

Далее, как обычно, под *весом* булевой матрицы понимается число единиц в ней.

Лемма 3.3. *Пусть $p, r \in \mathbb{N}$. Произвольная булева (m, n)-матрица веса V может быть реализована схемой глубины 2 с np^{r-1} внутренними вершинами, rnp^{r-1} вентилями на первом слое и $V/r + m(n/p+1)$ вентилями на втором слое.*

Доказательство. Разобьем матрицу на вертикальные полосы ширины p. На первом уровне реализуем всевозможные суммы не более чем r переменных из каждой полосы. Для одной полосы таких сумм не более p^r, для неполной полосы ширины $p' < p$ их менее $(p'/p)p^r$.

Каждую строку матрицы внутри каждой полосы представим в виде суммы подстрок веса r и, при необходимости, одной подстроки меньшего веса. На втором слое схемы выполняется сложение подстрок. □

Помимо схем из двух указанных лемм, мы будем применять двойственные к ним схемы. Это значит, что схема для матрицы A получена из схемы, построенной одним из способов лемм 3.2 и 3.3, для транспонированной матрицы A^T с последующим обращением ориентации вентилей. Чтобы сформулировать двойственное утверждение, нужно в исходной формулировке каждой из лемм поменять ролями m и n.

[19] Здесь и далее в верхних оценках мы опускаем округления.

Из леммы 3.2 вытекает решение асимптотической проблемы при малых m и q.

Следствие 3.1. *Пусть* $qm = n^{o(1)}$. *Тогда*

$$L_2(q,m,n) \le (1+\tau(q,m,n))\frac{mn}{\log_q(mn)} + n, \qquad \tau(q,m,n) \asymp \frac{\log(qm\log n)}{\log n}.$$

Доказательство. Оценка достигается на схеме из леммы 3.2 (транспонированная версия) с параметром $s = \lfloor \log_q n - 2\log_q \log n - 1 \rfloor$. □

В доказательстве следующей теоремы фактически строится упрощенная схема Пиппенджера [176].

Теорема 3.1. (*i*) *Пусть* $m \le n$, $m = \Omega\left(\log^{3/2} n\right)$ *и* $q = n^{o(1)}$. *Тогда*

$$L_3(q,m,n) \le (1+\tau(q,n))\frac{mn}{\log_q(mn)}, \qquad \tau(q,n) \asymp \sqrt{\frac{\log(q\log n)}{\log n}}.$$

(*ii*) *Пусть дополнительно выполнено* $q = \log^{O(1)} n$, $m = \Omega\left(\log^2 n\right)$ *и* $m \in n^{\mu}\log^{\pm O(1)} n$ *для некоторой постоянной* $\mu \in \mathbb{Q}$. *Тогда*

$$L_3(q,m,n) \le (1+\tau(n))\frac{mn}{\log_q(mn)}, \qquad \tau(n) \asymp \frac{\log\log n}{\log n}.$$

Доказательство. (*i*) Пусть $m = n^{\alpha}$. Если $\frac{(3/2)\log\log n - O(1)}{\log n} \le \alpha \le \sqrt{\frac{\log(q\log n)}{\log n}}$, то нужную оценку обеспечивают схемы глубины 2 из следствия 3.1. Поэтому далее полагаем, что $\alpha \ge \sqrt{\frac{\log(q\log n)}{\log n}}$.

Положим $\delta = \frac{1}{2}\sqrt{\frac{\log(q\log n)}{\log n}}$ и при помощи леммы 3.1 построим приближение

$$\frac{1}{1+\alpha} = P(r_t,\ldots,r_1) - \varepsilon, \quad 0 \le \varepsilon \le \delta, \qquad R = r_1 \cdot \ldots \cdot r_t \le 2/\delta,$$

где $t = 2k$. Параметры r_i умышленно пронумерованы в обратном порядке по отношению к (3.6).

Можно считать, что $k > 0$, поскольку в силу выбора δ случай $k = 0$ соответствует малым значениям параметра α: $\alpha \lesssim \frac{1}{2}\sqrt{\frac{\log(q\log n)}{\log n}}$.

Далее мы опишем итерационную процедуру построения многослойной схемы, вычисляющей произвольную (m, n)-матрицу A. Предварительно

дадим неформальное общее описание этой процедуры. На первом шаге методом леммы 3.2 для матрицы строится схема глубины 2. При этом один слой схемы — условно говоря, простой (с относительно небольшим числом ребер), второй — сложный. Пусть A_0 — булева матрица, вычисляемая сложным слоем. На следующем шаге процедуры слой, реализующий матрицу A_0, заменяется двухслойной схемой, построенной методом леммы 3.3. В этой схеме также один из слоев простой, а другой — сложный. Булева матрица, реализуемая сложным слоем, обозначается через A_1. На очередном шаге вместо этого слоя подставляется двухслойная схема, построенная методом леммы 3.3, и так далее. При реализации матриц A_i используются поочередно прямая и двойственная версия схем леммы 3.3. В итоге мы получим схему из $2k+1$ слоев, среди которых только один внутренний слой сложный — он будет определять сложность всей схемы.

Перейдем к формальному описанию процедуры, характеризующейся параметрами β, $a_1, \ldots, a_{2k}$, значения которых будут определены позднее.

1) Пусть β — некоторый параметр, при котором $\beta \log_q n$ — целое число. Для реализации матрицы A применим транспонированное преобразование леммы 3.2 с разрезанием на горизонтальные полосы высоты $s = \beta \log_q n$. Второй слой схемы заполняется пучками из не более чем qs ребер, а на первом слое реализуется некоторая булева матрица A_0 из $m_0 \leq mn^{\beta}/(\beta \log_q n)$ строк, $n_0 = n$ столбцов, веса $v_0 \leq n(m/(\beta \log_q n) + 1)$.

Размеры возникающих далее в процессе трансформации схемы булевых матриц A_i обозначим через $m_i \times n_i$, а вес — через v_i.

2) Нечетная итерация. Разбиваем матрицу A_{2i} на вертикальные полосы ширины $n^{a_{2i+1}}$. Реализуем при помощи леммы 3.3 с параметром $r = r_{2i+1}$ матрицу из каждой полосы. Первый слой схемы составляют пучки из не более чем r_{2i+1} вентилей. На втором слое реализуется некоторая матрица A_{2i+1} с размерными параметрами $m_{2i+1} = m_{2i}$, $n_{2i+1} \leq n_{2i} n^{a_{2i+1}(r_{2i+1}-1)}$ и весом $v_{2i+1} \leq v_{2i}/r_{2i+1} + m_{2i}(n_{2i} n^{-a_{2i+1}} + 1)$. Подставим эту двухслойную реализацию в схему для матрицы A.

3) Четная итерация. Разбиваем матрицу A_{2i-1} на горизонтальные полосы высоты $n^{a_{2i}}$. Реализуем при помощи леммы 3.3 с параметром $r = r_{2i}$ транспонированную матрицу из каждой полосы. Второй слой схемы со-

ставляют пучки из не более чем r_{2i} вентилей. На первом слое реализуется некоторая матрица A_{2i} с размерными параметрами $m_{2i} \leq m_{2i-1} n^{a_{2i}(r_{2i}-1)}$, $n_{2i} = n_{2i-1}$ и весом $v_{2i} \leq v_{2i-1}/r_{2i} + n_{2i-1}(m_{2i-1} n^{-a_{2i}} + 1)$. Подставим описанную конструкцию в схему для матрицы A.

Окончательно получаем схему, состоящую из $2k+2$ слоев, в которой внутренний слой (с номером $k+1$, считая от входов) вычисляет матрицу A_{2k}, размеры и вес которой удовлетворяют оценкам:

$$m_{2k} \leq \frac{m \cdot n^{\beta + a_2(r_2-1) + a_4(r_4-1) + \ldots + a_{2k}(r_{2k}-1)}}{\beta \log_q n}, \tag{3.10}$$

$$n_{2k} \leq n^{1 + a_1(r_1-1) + a_3(r_3-1) + \ldots + a_{2k-1}(r_{2k-1}-1)}, \tag{3.11}$$

$$\begin{aligned} v_{2k} \leq \frac{mn}{R \cdot \beta \log_q n} + \\ \frac{m \cdot n^{1+\beta + a_1(r_1-1) + a_2(r_2-1) + \ldots + a_{2k-1}(r_{2k-1}-1) - a_{2k}}}{\beta \log_q n} + \\ \frac{m \cdot n^{1+\beta + a_1(r_1-1) + a_2(r_2-1) + \ldots + a_{2k-2}(r_{2k-2}-1) - a_{2k-1}}}{r_{2k} \cdot \beta \log_q n} + \ldots \\ \ldots + \frac{m \cdot n^{1+\beta-a_1}}{r_2 \cdot \ldots \cdot r_{2k} \cdot \beta \log_q n} + \\ n_{2k-1} + \frac{m_{2k-2}}{r_{2k}} + \ldots + \frac{m_0}{r_2 \cdot \ldots \cdot r_{2k}} + \frac{n}{r_1 \cdot \ldots \cdot r_{2k}}. \end{aligned} \tag{3.12}$$

Положим $\beta = (1+\alpha)/R - \gamma_0$, $a_1 = \beta + \gamma$ при некоторых положительных параметрах $\gamma_0, \gamma \in o(\delta)$ с условием n^{a_1} — целое число. Пусть далее $a_i = a_{i-1} r_{i-1}$ для всех $i > 1$.

При таком выборе для любого $i \leq 2k-1$ имеем

$$\begin{aligned} \beta + a_1(r_1-1) + a_2(r_2-1) + \ldots + a_i(r_i-1) = \\ = \beta + (a_2 - a_1) + (a_3 - a_2) + \ldots + (a_{i+1} - a_i) = a_{i+1} - \gamma. \end{aligned}$$

Поэтому числители в средней группе слагаемых оценки (3.12) равны $mn^{1-\gamma}$. Сумма этих членов в таком случае не превосходит

$$\frac{tmn^{1-\gamma}}{\beta \log_q n}. \tag{3.13}$$

Сумму членов в последней строке (3.12) грубо оценим сверху как

$$t(n_{2k} + m_{2k}). \tag{3.14}$$

В результате получаем оценку

$$v_t \leq \frac{mn}{(1 + \alpha - \gamma_0 R)\log_q n} + \frac{tmn^{1-\gamma}}{\beta \log_q n} + t(n_{2k} + m_{2k}). \tag{3.15}$$

Схема глубины 3 для исходной матрицы A получается следующим образом. На среднем слое реализуется матрица A_{2k}. Остальные слои построенной схемы вокруг слоя, реализующего матрицу A_{2k}, склеиваются. Первый слой полученной таким образом схемы составляют n_{2k} пучков из не более чем $r_1 r_3 \cdot \ldots \cdot r_{2k-1}$ вентилей, а третий слой — m_{2k} пучков из не более чем $r_2 r_4 \cdot \ldots \cdot r_{2k} \cdot q\beta \log_q n$ вентилей. Суммарное число вентилей в этих двух слоях можно оценить как

$$Rq\beta \log_q n \cdot (n_{2k} + m_{2k}). \tag{3.16}$$

Используя формулы $a_i = r_{i-1} \cdot \ldots \cdot r_1 a_1$ и $\beta = a_1 - \gamma$, показатель степени у n в (3.10) переписывается как

$$\begin{aligned}
&\beta + a_2(r_2 - 1) + a_4(r_4 - 1) + \ldots + a_{2k}(r_{2k} - 1) = \\
&\quad a_1 R \cdot P(r_{2k}, \ldots, r_1) - \gamma = (1 + \alpha - (\gamma_0 - \gamma)R)\left(\frac{1}{1+\alpha} + \varepsilon\right) - \gamma \\
&\qquad = 1 + \varepsilon(1+\alpha) - (\gamma_0 - \gamma)R\left(\frac{1}{1+\alpha} + \varepsilon\right) - \gamma.
\end{aligned}$$

Аналогично, сумма в показателе степени в (3.11) преобразуется как

$$\begin{aligned}
&a_1(r_1 - 1) + a_3(r_3 - 1) + \ldots + a_{2k-1}(r_{2k-1} - 1) = \\
&\quad a_1 R(1 - P(r_{2k}, \ldots, r_1)) = (1 + \alpha - (\gamma_0 - \gamma)R)\left(\frac{\alpha}{1+\alpha} - \varepsilon\right) \\
&\qquad = \alpha - \varepsilon(1+\alpha) - (\gamma_0 - \gamma)R\left(\frac{\alpha}{1+\alpha} - \varepsilon\right).
\end{aligned}$$

Теперь неравенства (3.10) и (3.11) можно переписать как

$$m_{2k} \leq \frac{mn^{1+\varepsilon(1+\alpha)-(\gamma_0-\gamma)R\left(\frac{1}{1+\alpha}+\varepsilon\right)-\gamma}}{\beta \log_q n}, \tag{3.17}$$

$$n_{2k} \leq mn^{1-\varepsilon(1+\alpha)-(\gamma_0-\gamma)R\left(\frac{\alpha}{1+\alpha}-\varepsilon\right)}. \tag{3.18}$$

Осталось указать выбор параметров. Выберем $\gamma \asymp \frac{1}{R}\sqrt{\frac{\log(q\log n)}{\log n}}$ и $\gamma_0 \sim 2\gamma$ так, чтобы показатели степеней у n в оценках (3.13), (3.17), (3.18) не превосходили $1 - \log\left(q\log^3 n\right)/\log n$, и при этом $\beta\log_q n$ и n^{a_1} были целыми числами. (Благодаря выбору множителя $\frac{1}{2}$ в определении величины δ, разность $\frac{\alpha}{1+\alpha} - \varepsilon$ в показателе степени в (3.18) неотрицательна и по порядку равна α.) Требование $\beta\log_q n \in \mathbb{Z}$ выполнимо, поскольку $\gamma\log_q n = \Omega(1)$.

Тогда вклад слагаемых (3.13), (3.14) и (3.16) в оценку сложности является величиной $O(mn/\log^2 n)$. Теперь сложность схемы оценивается как сумма (3.15) и (3.16):

$$\frac{mn}{(1+\alpha-\gamma_0 R)\log_q n} + O\left(\frac{mn}{\log^2 n}\right) = \frac{mn}{(1+\alpha)\log_q n}\left(1 + O\left(\sqrt{\frac{\log(q\log n)}{\log n}}\right)\right).$$

Докажем (ii). Если $\mu = 0$, то достаточно воспользоваться следствием 3.1. Иначе, пусть $m = n^\alpha$. Как следует из п. (ii) леммы 2.56, существует приближение

$$\frac{1}{1+\alpha} = P(r_1, \ldots, r_k) - \varepsilon, \quad |\varepsilon| = O\left(\frac{\log\log n}{\log n}\right), \quad r_1 \cdot \ldots \cdot r_k = O(1).$$

Если $\varepsilon > 0$, то просто применяем конструкцию из доказательства п. (i). При этом можно положить $\delta, \gamma \asymp \frac{\log\log n}{\log n}$.

Случай $\varepsilon < 0$ сводится к случаю $\varepsilon > 0$. Для этого матрица разбивается на вертикальные полосы ширины $\hat{n} = \frac{n}{\log^c n}$, где константа c выбирается из условия $m \geq \hat{n}^\mu$. Подматрицы в полосах реализуются независимо. Затем схемы для подматриц объединяются путем отождествления выходов. $\square$

3.2.3 Схемы глубины 4

Ценой роста глубины, можно ослабить зависимость сложности схемы теоремы 3.1 от q. В схеме теоремы 3.1 увеличению q препятствуют два обстоятельства: множитель q в оценке (3.16) сложности граничного слоя и требование целочисленности параметра ширины полосы $s = \beta\log_q n$. Оба препятствия обусловлены конструкцией леммы 3.2.

Первое препятствие (лишний множитель) можно устранить, добавив в схему еще один слой. Второе препятствие устраняется, если разрешить

разбиение на полосы дробной ширины. Оба приема совмещены в следующей лемме, идея которой просматривается в методе [176], и которую можно использовать в качестве альтернативы лемме 3.2.

Лемма 3.4. *Пусть $\sigma > 0$. Произвольная (m, n)-матрица над E_q может быть реализована вентильной схемой глубины 3 с $2nq/\log q$ вентилями на первом слое, $n(\log q + 1)$ вершинами на глубине 1, $12q^\sigma(n/\sigma + 1)(\sigma + 3)\log q$ вентилями на среднем слое, $4q^\sigma(n/\sigma + 1)$ вершинами на глубине 2 и $m(n/\sigma + 1)$ вентилями на третьем слое.*

Доказательство. Положим *емкость* ячейки матрицы над E_q — число возможных значений, которые она может заключать — равной q. Допустим возможность разбиения некоторых ячеек на части, которые будем называть неполными ячейками. Неполная ячейка может заключать числа из $aE_b = \{0, a, 2a, \ldots, (b-1)a\}$ и имеет, по определению, емкость b. Ячейка для чисел из aE_b (в случае $a = 1$ и $b = q$ это целая ячейка, иначе — неполная) может быть разбита на две части по следующему правилу. Пусть $b_1 b_2 \geq b$ и $b_1, b_2 < b$. Разрешим одной части ячейки заключать числа из множества aE_{b_1}, а другой — числа из $ab_1E_{b_2}$. Произвольное число из aE_b можно представить суммой двух чисел из aE_{b_1} и $ab_1E_{b_2}$, и разместить слагаемые в подходящих частях ячейки.

Емкость множества ячеек, включая, возможно, неполные, естественно определить как произведение емкостей составляющих (емкость целой ячейки может быть меньше емкости множества ее частей).

Разобьем множество ячеек строки матрицы на подмножества с емкостью каждого подмножества в интервале $[2q^\sigma, 4q^\sigma]$ (включая, возможно, одно подмножество меньшей емкости). Для этого мы совершаем проход вдоль строки, добавляя в текущее подмножество емкости C очередную ячейку емкости b, если $Cb \leq 2q^\sigma$, иначе — часть этой ячейки с емкостью $b_1 = \lceil 2q^\sigma/C \rceil$. Вторая часть ячейки приобретает емкость $\lceil b/b_1 \rceil$ и становится очередной для следующего подмножества.

Указанное разбиение делит матрицу на вертикальные полосы. Строка в полосе может содержать не более $\sigma + 1$ целых ячеек и не более двух неполных. Число полос можно оценить, исходя из того, что выделение од-

ного подмножества емкости не менее $2q^\sigma$ уменьшает емкость оставшейся части множества минимум в q^σ раз. Отталкиваясь от емкости исходного множества q^n, получаем, что матрица содержит не более $n/\sigma + 1$ полос.

Теперь опишем схему. На первом слое схемы для каждого входа x вычисляется вектор кратностей: $\left(1,\, 2,\, 4,\, \ldots,\, 2^{\lfloor \log(q/\log q)\rfloor}\right) x$. Один такой вектор реализуется при помощи не более чем $2q/\log q$ вентилей. На втором слое вычисляются всевозможные суммы в полосах: число сумм в одной полосе не превосходит $4q^\sigma$ по построению, для вычисления одной суммы используется не более $\sigma + 3$ ячеек, для вычисления значения в ячейке из кратностей, вычисленных на первом уровне схемы, достаточно $3\log q$ вентилей. Последняя оценка опирается на возможность разложения произвольного числа $a \in E_q$ в сумму $a = a_1 l + a_2$, $l = 2^{\lfloor \log(q/\log q)\rfloor}$, при этом $a_1 \leq 2\log q$, $a_2 < l$. На заключительном слое, как и в лемме 3.2, суммы в строках складываются из сумм в полосах. □

Доказанная лемма позволяет расширить результат следствия 3.1 и дополнить результат теоремы 3.1 о возможностях асимптотически оптимального синтеза в глубине 3.

Следствие 3.2. *Пусть* $q = o\left(n/\log^2 n\right)$, $m = n^{o(1)}$. *Тогда*

$$L_3(q,m,n) \leq (1+\tau(m,n))\frac{mn}{\log_q(mn)} + n, \qquad \tau(m,n) \asymp \frac{\log(m\log n)}{\log n}.$$

Доказательство. Оценка достигается на схеме из леммы 3.4 (транспонированная версия) с параметром $\sigma = \log_q(n/\log^2 n)$. □

Теорема 3.2. (*i*) *Пусть* $m \leq n$, $m = \Omega\left(\log^{3/2} n\right)$ *и* $q = o\left(n/\log^2 n\right)$. *Тогда*

$$L_4(q,m,n) \leq (1+\tau(n))\frac{mn}{\log_q(mn)}, \qquad \tau(n) \asymp \sqrt{\frac{\log\log n}{\log n}}.$$

(*ii*) *Пусть дополнительно* $m = \Omega\left(\log^2 n\right)$ *и* $m \in n^\mu \log^{\pm O(1)} n$ *для некоторой постоянной* $\mu \in \mathbb{Q}$. *Тогда*

$$L_4(q,m,n) \leq (1+\tau(n))\frac{mn}{\log_q(mn)}, \qquad \tau(n) \asymp \frac{\log\log n}{\log n}.$$

Доказательство. (*i*) Положим $\delta = \frac{1}{2}\sqrt{\frac{\log\log n}{\log n}}$. Величины $\alpha, \varepsilon, k, t, R$ определим как в теореме 3.1.

При $\alpha \le 2\delta$ схемы глубины 3 из следствия 3.2 обеспечивают нужный результат. При бóльших α повторяем построение теоремы 3.1, только используем в качестве отправной точки схему леммы 3.4 вместо схемы леммы 3.2 (т. е. модифицируется шаг п. 1 в доказательстве теоремы 3.1).

1) Пусть β — некоторый параметр. Для реализации (m, n)-матрицы A применим транспонированное преобразование леммы 3.4 с параметром $\sigma = \beta \log_q n$. В полученной схеме третий слой содержит не более $2mq$ вентилей. Второй слой схемы образуют пучки из не более чем $(\beta \log_q n + 3)(\log q + 1)$ ребер, а на первом слое реализуется некоторая булева матрица A_0 из $m_0 \le 4n^{\beta}(m + \beta \log_q n)/(\beta \log_q n)$ строк, $n_0 = n$ столбцов, веса $v_0 \le n(m/(\beta \log_q n) + 1)$.

Далее действуем как описано в пп. 2) и 3) теоремы 3.1. В итоге приходим к схеме из $2k+3$ слоев, в которой средний слой (с номером $k+1$) вычисляет матрицу A_{2k}, для размеров и веса которой справедливы оценки:

$$m_{2k} \le \frac{4(m + \beta \log_q n) n^{\beta + a_2(r_2-1) + a_4(r_4-1) + \ldots + a_{2k}(r_{2k}-1)}}{\beta \log_q n}, \tag{3.19}$$

$$n_{2k} \le n^{1 + a_1(r_1-1) + a_3(r_3-1) + \ldots + a_{2k-1}(r_{2k-1}-1)}, \tag{3.20}$$

$$\begin{aligned} v_{2k} \le \frac{mn}{R \cdot \beta \log_q n} + \\ \frac{4(m + \beta \log_q n) n^{1 + \beta + a_1(r_1-1) + a_2(r_2-1) + \ldots + a_{2k-1}(r_{2k-1}-1) - a_{2k}}}{\beta \log_q n} + \\ \frac{4(m + \beta \log_q n) n^{1 + \beta + a_1(r_1-1) + a_2(r_2-1) + \ldots + a_{2k-2}(r_{2k-2}-1) - a_{2k-1}}}{r_{2k} \cdot \beta \log_q n} + \ldots \\ \ldots + \frac{4(m + \beta \log_q n) n^{1 + \beta - a_1}}{r_2 \cdot \ldots \cdot r_{2k} \cdot \beta \log_q n} + \\ n_{2k-1} + \frac{m_{2k-2}}{r_{2k}} + \ldots + \frac{m_0}{r_2 \cdot \ldots \cdot r_{2k}} + \frac{n}{r_1 \cdot \ldots \cdot r_{2k}}. \end{aligned}$$

Положим $\beta = (1 + \alpha)/R - \gamma_0$, $a_1 = \beta + \gamma$ и $a_i = a_{i-1} r_{i-1}$ для всех $i > 1$. В результате (см. доказательство теоремы 3.1), для веса матрицы

A_{2k} выводится оценка

$$v_t \le \frac{mn}{(1+\alpha-\gamma_0 R)\log_q n} + \frac{4t(m+\beta\log_q n)n^{1-\gamma}}{\beta\log_q n} + t(n_{2k}+m_{2k}). \qquad (3.21)$$

Схема глубины 4 для исходной матрицы A строится следующим образом. Сохраняем внешний слой из ребер, соединенных с выходами, построенный на первом шаге, а также средний слой, где реализуется матрица A_{2k}. Слои в каждой из оставшихся двух связных частей схемы совмещаются. На первом слое полученной таким образом схемы расположены n_{2k} пучков из не более чем $r_1 r_3 \cdot \ldots \cdot r_{2k-1}$ вентилей, на третьем слое — m_{2k} пучков из не более чем $r_2 r_4 \cdot \ldots \cdot r_{2k} \cdot 3(\beta\log_q n + 3)\log q$ вентилей. Тогда суммарное число вентилей на первом, третьем и четвертом слоях оценивается как

$$3R(\beta\log_q n + 3)\log q \cdot (n_{2k}+m_{2k}) + 2mq/\log q. \qquad (3.22)$$

Так же, как в теореме 3.1, оценки (3.19), (3.20) переписываются в виде

$$m_{2k} \le \frac{4(m+\beta\log_q n)n^{1+\varepsilon(1+\alpha)-(\gamma_0-\gamma)R\left(\frac{1}{1+\alpha}+\varepsilon\right)-\gamma}}{\beta\log_q n}, \qquad (3.23)$$

$$n_{2k} \le mn^{1-\varepsilon(1+\alpha)-(\gamma_0-\gamma)R\left(\frac{\alpha}{1+\alpha}-\varepsilon\right)}. \qquad (3.24)$$

Выберем $\gamma \sim \frac{c}{R}\sqrt{\frac{\log\log n}{\log n}}$ и $\gamma_0 \sim 2\gamma$ так, чтобы показатели степеней у n в среднем слагаемом оценки (3.13) и оценках (3.23), (3.24) не превосходили $1 - 3\log\log n/\log n$, и при этом выполнялось условие $n^{a_1} \in \mathbb{Z}$.

Теперь сложность схемы оценивается как сумма (3.21) и (3.22):

$$\frac{mn}{(1+\alpha-\gamma_0 R)\log_q n} + O\left(\frac{mn}{\log^2 n}\right) = \frac{mn}{(1+\alpha)\log_q n}\left(1 + O\left(\sqrt{\frac{\log\log n}{\log n}}\right)\right).$$

Доказательство п. (ii) не отличается от доказательства п. (ii) теоремы 3.1. □

3.2.4 Вычисление матриц с быстро растущими коэффициентами

В этом параграфе рассматривается класс матриц с произвольным ограничением q на размер коэффициентов. Следующая лемма описывает строительный блок для оптимальных схем.

Лемма 3.5. *Пусть* $v = (3^{u_1}, \ldots, 3^{u_s})$, $u_i \in \mathbb{N} \cup \{0\}$, $u = \max_{i=1,\ldots,s}\{u_i\}$. *Тогда при любом* d, $1 \le d \le u$, *справедливо*

$$L_d(v) \le \min\{u,\, 2d-1\} \cdot 3^{u/d} + (s-1)3^{\lfloor (u-1)/d \rfloor}.$$

Доказательство. Удобнее рассмотреть реализацию транспонированного вектора v^T. На рис. 7 изображена базовая схема, вычисляющая вектор $w = (1, 3, \ldots, 3^u)^T$. Она имеет сложность $4u$ и глубину u. Нужная схема для v^T в случае $d = u$ получится, если из висячих ребер оставить только те, которые ведут к выходам кратности 3^{u_i}.

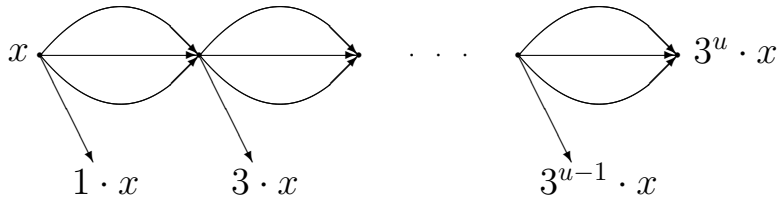


Рис. 7: Схема для вектора w

При произвольном d запишем $u = pd + r$, $0 \le r < d$. Обобщая схему рис. 7, максимальную кратность 3^u вычислим в цепочке из d слоев: из них r слоев содержат группы из 3^{p+1} параллельно соединенных вентилей, остальные $d - r$ слоев — группы из 3^p штук. Для вычисления каждого из недостающих выходов достаточно добавить группу из не более чем 3^p вентилей, а если $r = 0$, то даже не более чем из 3^{p-1} вентилей.

Обозначим $x = r/d$. Число вентилей в цепочке, вычисляющей 3^u, оценивается как

$$(d-r)3^p + r3^{p+1} = (d+2r)3^p = d(1+2x)3^p \le d(1+x)3^{p+x} = d(1+x)3^{u/d},$$

в силу $1+2x \le (1+x)(1+\ln 3 \cdot x) \le (1+x)3^x$ для $x \ge 0$. Осталось заметить, что $r \le \min\{u-d,\, d-1\}$. □

Замечание. Если к выходу схемы ведет единственное ребро, и этот выход соединяется со входом другой схемы, то ребро можно удалить, отождествив его концы, не нарушая функционирования объединенной схемы. Висячие ребра, ведущие к выходам, как в схеме рис. 7, нужны только затем, чтобы

удовлетворить требованию определения схемы, запрещающей пути между входами или выходами. Соответственно, когда выходы становятся внутренними вершинами, необходимость в этих ребрах отпадает. По построению, в схеме из доказательства леммы по меньшей мере $s-1-(u-d)$ выходов являются концами висячих ребер.

Теорема 3.3. *Пусть $m \le n$, $H = mn\log q \to \infty$.*

(i) Пусть $1 \le d \le \lfloor \log_3(q-1) \rfloor$. Определим $s_d = \lfloor \log_3(q-1) \rfloor - d$ при $d \ge \log_3(q-1)/2$ и $s_d = d-1$, иначе. Тогда

$$L_{d+3}(q,m,n) \le (1+\tau(H))\frac{H}{\log H} + n + mq^{1/d}(\lfloor \log_3(q-1) \rfloor + s_d), \quad (3.25)$$

где $\tau(H) \asymp \sqrt{\frac{\log\log H}{\log H}}$.

(ii) Пусть дополнительно выполнено условие $n = \log^{O(1)} q$. Тогда справедливо

$$L_d(q,m,n) \le 3m\log_3(q-1) + (1+\tau(H))\frac{H}{\log H}, \quad (3.26)$$

где $d = \lfloor \log_3 q + \sqrt{\log_3 q} \rfloor + 3$ и $\tau(H) \asymp \frac{\log\log H}{\log H}$.

Доказательство. Случай $q \le n/\log^3 n$ охватывают следствие 3.2, теоремы 3.1 и 3.2 (последнее слагаемое в оценке (3.25) в этом случае несущественно). Рассмотрим основной случай $q \ge n/\log^3 n$.

Базовый метод. Пусть $3^{k(t-1)} \le q-1 < 3^{kt}$ при некоторых $k,t \in \mathbb{N}$. Используя запись в троичной системе счисления, каждый из элементов $a_{i,j}$ матрицы A представим в виде

$$a_{i,j} = bD_{i,j}c^T, \quad b = \left(1,\, 3^k,\, 3^{2k}, \ldots,\, 3^{(t-1)k}\right), \quad c = (1,\, 3,\, 3^2, \ldots,\, 3^{k-1}),$$

где $D_{i,j}$ — матрица размера $t \times k$ над E_3.

Тогда справедливо

$$A = (a_{i,j}) = BDC, \qquad B = \begin{pmatrix} b & 0 & \cdots & 0 \\ 0 & b & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & b \end{pmatrix},$$

$$D = \begin{pmatrix} D_{1,1} & \cdots & D_{1,n} \\ \cdots & \cdots & \cdots \\ D_{m,1} & \cdots & D_{m,n} \end{pmatrix}, \qquad C = \begin{pmatrix} c^T & 0 & \cdots & 0 \\ 0 & c^T & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & c^T \end{pmatrix}.$$

Схему для матрицы A можно построить, соединяя последовательно схемы для матриц C, D, B.

Если вектора-подматрицы матриц B и C реализуются схемами леммы 3.5, то висячие ребра обращены внутрь схемы, и их можно удалить, согласно замечанию к лемме.

Для доказательства п. (i) выберем $k = 1$. Тогда

$$L_{d+3}(A) \le L'_d(B) + L_3(D) \le mL'_d(b) + L_3(3, mt, n),$$

где функционал L'_d отличается от L_d отсутствием учета висячих ребер, ведущих к выходам. Согласно лемме 3.5 с учетом замечания к ней,

$$L'_d(b) \le (\min\{t-1, 2d-1\} + t - 1 - d)3^{(t-1)/d} \le (\lfloor \log_3(q-1) \rfloor + s_d)q^{1/d}.$$

Величину $L_3(3, mt, n)$ оценим при помощи следствия 3.1 или теоремы 3.1 в зависимости от соотношения между mt и n как

$$L_3(3, mt, n) \le (1 + \tau(H))\frac{H}{\log H} + n.$$

Здесь существенно, что $t = \lceil \log_3 q \rceil = (1 + O(1/\log H)) \log_3 q$ в силу ограничения снизу на q.

Для доказательства п. (ii) достаточно выбрать $k \le t$ так, чтобы выполнялось условие $mt \in (nk)^\mu \log^{\pm O(1)}(nk)$ для некоторой постоянной $\mu \in \mathbb{Q}$. Случай $m = n = 1$ тривиален. Иначе, пусть $n \ge 2$, $m = n^\alpha$ и $\log_3 q = n^\beta$. Выберем $\mu \in [\alpha, \min\{1, \alpha+\beta/2\}] \cap \mathbb{Q}$ и положим $k = \left\lfloor n^{\frac{\mu(1+\beta)-\alpha}{1+\mu}} \right\rfloor$, $t = \left\lceil \frac{\log_3 q}{k} \right\rceil$.

Реализуем матрицы B и C схемами леммы 3.5 (максимальной) суммарной глубины $kt - 1 \leq \log_3 q + k - 1$, а матрицу D реализуем схемой глубины 3 как в п. (i). По построению, $kt \leq \log_3 q + k = \left(1 + O\left(1/\sqrt{H}\right)\right) \log_3 q$. $\square$

В частном случае $d = \lfloor \log_3(q-1) \rfloor$ оценка п. (i) теоремы является незначительным уточнением оценки (3.3).

Следствие 3.3. *При условиях $m \leq n$, $H = mn \log q \to \infty$ справедливо*

$$L(q, m, n) \leq 3m \log_3(q-1) + (1 + \tau(H)) \frac{H}{\log H} + n, \qquad (3.27)$$

где $\tau(H) \asymp \sqrt{\frac{\log\log H}{\log H}}$, а при дополнительном ограничении $q = n^{O(1)}$ оценка (3.27) *достигается на схемах глубины $O(1)$.*

Доказательство. Вторая часть утверждения получается при надлежащем выборе $d \geq \log_q n + 1$. $\square$

В п. (ii) мы ограничились указанием только достаточно общего случая (очень большое q), когда оценка (3.26) может быть доказана с улучшенным остаточным членом. Способ доказательства и результат теоремы 3.1 позволяют получить «правильный» порядок величины $\tau(H)$ и в некоторых других ситуациях.

В целом, оценка п. (ii) служит обобщением своего частного случая $m = n = 1$, для которого результат может быть получен адаптацией метода А. Брауэра [108] построения аддитивных цепочек к модели вентильных схем. По существу, применяемая в доказательстве теоремы 3.3 декомпозиция $A = BDC$ является центральным местом в методе Брауэра.

3.3 Экстремальные расхождения между линейными мерами сложности булевых матриц

Асимптотические результаты свидетельствуют о том, что для почти всех $n \times n$ матриц SUM-, OR- и XOR-сложность примерно совпадают. Ситуация, однако, может измениться, если мы рассмотрим конкретную «нетипичную» матрицу. Более формально, можно поставить вопрос: насколько велико может может быть отношение $\mathsf{L}(A)/\mathsf{L}'(A)$ для $n \times n$ матрицы A, где

$\mathsf{L}, \mathsf{L}' \in \{\mathsf{SUM}_d, \mathsf{OR}_d, \mathsf{XOR}_d\}$. Предварительно заметим, что согласно асимптотическим результатам такое отношение не может по порядку превосходить $n/\log n$.

По-видимому, впервые подобный вопрос (немного в других терминах) был поставлен в работе [50] при $\mathsf{L} = \mathsf{OR}$ и $\mathsf{L}' = \mathsf{XOR}$ для класса булевых матриц без прямоугольников (т. е. без сплошь единичных подматриц размера 2×2). Для конкретной $n \times n$ матрицы A (матрицы Зингера, см. ниже) там была получена оценка $\mathsf{OR}(A)/\mathsf{XOR}(A) = n^{1/2-o(1)}$, но доказательство оказалось некорректным, и вопрос не был закрыт еще около полувека.

Основные результаты об экстремальных расхождениях между различными мерами сложности получены уже в 21-м веке. Активная разработка темы открылась работой С. Б. Гашкова и автора [215], в которой результат [50] получил корректное доказательство. Кроме того, там было доказано отношение $n^{1-o(1)}$ для конкретно заданной булевой $n \times n$ матрицы (норм-матрицы из работы [152]). Из другой совместной работы автора с М. И. Гринчуком [216] вытекала неконструктивная оценка $n/\log^{6+o(1)} n$ для класса булевых $n \times n$ матриц.

В итоге, работы [215, 216] стимулировали исследования указанного круга вопросов. Почти одновременно в работах [128, 107] 2013 г. разными способами была получена почти окончательная оценка $n/\log^2 n$ порядка OR/XOR отношения, и одновременно С. Юкна заметил, что эта же оценка вытекает из его более раннего результата [143]. Причем последний результат позволяет доказать отношение $n/\log^2 n$ в любой глубине, начиная с двух (полное доказательство приведено в [230]).

Кроме того, в работе [128] впервые получена нетривиальная оценка для SUM/OR отношения. Она имела величину порядка $\sqrt{n}/\log n$ и доказывалась неконструктивно. Однако, результат допускает и конструктивный аналог с величиной отношения $n^{1/2-o(1)}$, см. [230]. Метод [128] не позволяет получать растущее отношение SUM- и OR-сложностей в глубине 2. Пример $n \times n$ матрицы A с отношением $\mathsf{SUM}_2(A)/\mathsf{OR}_2(A) \asymp \log n$ построил Т. Пинто [174] в том же 2013 г.

Чуть раньше, Н. Кац, отвечая на вопрос из [145], построил метод, позволяющий неконструктивно доказывать высокие оценки отношений типа

$\mathsf{OR}_d(\overline{A})/\mathsf{OR}_d(A)$. Сами оценки приведены в [230]. Позднее автором предложен метод получения высоких конструктивных оценок [238].

В упоминавшейся выше работе С. Юкны и автора [230] предпринято систематическое изложение теории сложности линейных схем, охватывающее бóльшую часть известных результатов, в частности, подведены промежуточные итоги исследований экстремальных расхождений. Также там представлено несколько новых результатов, в частности, построен пример $n \times n$ матрицы A, для которой XOR_2-сложность существенно больше $\mathsf{OR}_2(A)$-сложности.

В табл. 4 приведены сведения о порядках величины неконструктивно и конструктивно доказанных расхождений между мерами сложности линейных схем (схемы с ограничением на глубину представлены схемами глубины 2)[20]. Тривиальные верхние оценки во всех случаях имеют величину порядка $n/\ln n$. Но для отношения $\mathsf{SUM}_2(A)/\mathsf{SUM}(A)$ несложно получить оценку $O(\sqrt{n})$.

	нижняя оценка	эффективная нижняя оценка
$\mathsf{OR}(A)/\mathsf{XOR}(A)$	$n/\ln^2 n$ [230, 128, 107]	$n2^{-\Theta\left(\sqrt{\ln n \ln\ln n}\right)}$ [215]
$\mathsf{OR}_2(A)/\mathsf{XOR}_2(A)$	$n/\ln^2 n$ [230]	$\sqrt{n}/\ln n$, фольклор
$\mathsf{SUM}(A)/\mathsf{OR}(A)$	$\sqrt{n}/\ln n$ [128]	$\sqrt{n}2^{-\Theta\left(\sqrt{\ln n \ln\ln n}\right)}$ [230]
$\mathsf{SUM}_2(A)/\mathsf{OR}_2(A)$	$\ln n$ [174]	
$\mathsf{XOR}_2(A)/\mathsf{OR}_2(A)$	$\ln\ln\ln n$ [230]	
$\mathsf{OR}(\overline{A})/\mathsf{OR}(A)$	$n/\ln^3 n$ [230]	$n2^{-\Theta\left(\sqrt{\ln n \ln\ln n}\right)}$ [238]
$\mathsf{OR}_2(\overline{A})/\mathsf{OR}_2(A)$	$n/\ln^3 n$ [230]	$\sqrt{n}2^{-\Theta\left(\ln^{2/3} n\right)}$ [230]
$\mathsf{OR}_2(A)/\mathsf{OR}(A)$	$\sqrt{n/\ln n}$ [230]	
$\mathsf{SUM}(\overline{A})/\mathsf{SUM}(A)$	$n^{1/4-o(1)}$ [147]	
$\mathsf{SUM}_2(A)/\mathsf{SUM}(A)$	$\sqrt{n/\ln n}$ [230]	
$\mathsf{XOR}_2(A)/\mathsf{XOR}(A)$	$n^{(3-\omega)/2}/\ln n$ [146]	$\ln^2 n/\ln^2\ln n$ [132]

Таблица 4: Экстремальные отношения типа $\mathsf{L}(A)/\mathsf{L}'(A)$

[20]Примечание к таблице: ω — экспонента сложности матричного умножения.

Решение задач об экстремальных отношениях, естественно, опирается на методы получения высоких нижних оценок сложности матриц. Сегодня можно выделить около десятка методов, оперирующих теми или иными характеристиками, либо структурными особенностями матриц, см. [230]. Для рассматриваемого круга задач особую роль играет метод оценки OR-сложности матрицы через показатель «редкости» — максимальный размер прямоугольников (сплошь единичных подматриц) в матрице. Этот метод был предложен Э. И. Нечипоруком [54, 57] в 1960-х гг. И пока это единственный известный метод, позволяющий получать близкие к квадратичным (т. е. максимально возможным) оценки сложности.

Для применения метода Нечипорука требуется построение редких матриц большого веса. Легко проверить, что большинство $n \times n$ матриц являются такими: имеют вес порядка n^2 и не содержат прямоугольников размера $2\log n \times 2\log n$ (см., например, [145, 230]). Гораздо труднее построить эффективный пример. Первый пример матрицы, имеющей OR-сложность $n^{2-o(1)}$, построил А. Е. Андреев [2]. Наилучшие известные оценки достигаются на «норм-матрицах» из работ Я. Коллара, Л. Роньяи, Т. Сабо и Н. Алона [152, 97]. Автором в работе [218] предложен метод, позволяющий расширять круг примеров, в частности, построить циркулянтный (циклический) аналог норм-матриц. Новые примеры оказываются полезными при исследовании задач об экстремальных отношениях.

Далее, в §3.3.1 вводятся понятия редкого множества и редкой матрицы, и изложен технический результат [218] о погружении многомерного редкого множества в пространство меньшей размерности. В §3.3.2 перечислены некоторые методы получения нижних оценок сложности матриц, которые используются в выводе последующих результатов. В §3.3.3 приводятся результаты об экстремальных отношениях типа OR/XOR в некоторых классах матриц, в той или иной степени использующих описанный в §3.3.1 метод. В §3.3.4 строится пример матрицы с растущим отношением $\mathsf{XOR}_2/\mathsf{OR}_2$. В §3.3.5 приводится эффективный метод построения матриц, OR-сложность которых существенно отличается от сложности дополнительных матриц.

3.3.1 Редкие множества. Погружение многомерного редкого множества в пространство меньшей размерности

Редкие множества и редкие матрицы

Подмножество H полугруппы $(G, +)$ назовем (k, l)-*редким*, если оно не содержит подмножеств вида $A \oplus B = \{a + b \mid a \in A,\ b \in B\}$, где $|A| = k$ и $|B| = l$. Кроме того, (k, k)-редкое подмножество будем называть сокращенно k-редким[21].

В абелевой группе $(G, +)$ без элементов порядка 2 понятие 2-редкого множества совпадает с понятием *множества Сидона*. Напомним, что подмножество H является множеством Сидона тогда и только тогда, когда для любых элементов $a, b, c, d \in H$ справедливо

$$a + b = c + d \ \Rightarrow\ ((a = c)\ \&\ (b = d))\ \vee\ ((a = d)\ \&\ (b = c)).$$

Покажем эквивалентность определений. Пусть $H \subset G$ не является 2-редким, т. е. для некоторых элементов $a \neq b$, $c \neq d$ выполнено $a + c, a + d, b + c, b + d \in H$. Но тогда $0 \neq (b + c) - (a + c) = (b + d) - (a + d)$, при этом $b + c \neq b + d$, значит, H не является множеством Сидона.

Пусть теперь H не является множеством Сидона, т. е. для некоторых элементов $a, b, c, d \in H$ выполняется $a + b = c + d$, при этом $a \neq c$, $a \neq d$. Но тогда $\{a, d\} \oplus \{c - a, 0\} = \{a, b, c, d\} \subset H$, т. е. H не является 2-редким.

К понятию редкого множества близко более часто употребимое понятие *редкой матрицы* (англ. — free matrix). Булева матрица называется (k, l)-редкой, если она не содержит сплошь единичных подматриц размера $k \times l$.

Очевидно, (k, l)-редкому множеству $M \subset \mathbb{Z}_m^n$ соответствует булева (k, l)-редкая и одновременно (l, k)-редкая симметричная относительно главной диагонали булева матрица $(\mu_{\alpha,\beta})$ размера $m^n \times m^n$ (строки и столбцы нумеруются элементами $\mathbb{Z}_m^n$), которая определяется условием

$$\mu_{\alpha,\beta} = 1 \Longleftrightarrow \alpha + \beta \in M.$$

Такая матрица оказывается в определенном смысле циклической, а при $n = 1$ — циркулянтной (состоящей из единичных циклических диагоналей).

[21] Термин «редкий» предложен в работе [50].

С редкими матрицами связана известная проблема К. Заранкевича. Она ставит вопрос: насколько большим может быть вес (k,l)-редкой булевой матрицы размера $m \times n$. К настоящему времени в случае $k,l \geq 4$ неизвестен даже порядок величины веса экстремальных матриц.

Конструкции редких матриц и связанных с ними редких множеств используются в доказательствах нижних оценок сложности во многих задачах дискретной математики. Несколько таких задач будут рассмотрены ниже.

Обзор конструкций редких множеств

Выше отмечалось, что в абелевой группе 2-редкое подмножество — это то же самое, что множество Сидона.

Обзор известных конструкций множеств Сидона, а также оценок мощности таких множеств приводится в [166] и в работах из библиографического указателя к [166]. Для нас важно отметить следующие факты: мощность множества Сидона в E_n и в группе $\mathbb{Z}_n$ асимптотически не превосходит $\sqrt{n}$. Эта оценка достигается в первом случае при всех (достаточно больших) n, а во втором — при некоторых n. Перечислим наиболее известные примеры.

а) Множество Дж. Зингера [198] мощности $q+1$ в $\mathbb{Z}_{q^2+q+1}$, где q — степень простого числа, определяется как $\{0\} \cup \{s_i \mid \theta^{s_i}/(\theta+\alpha_i) \in GF(q), GF(q) = \{\alpha_1, \ldots, \alpha_q\}\}$, где θ — примитивный элемент в поле $GF(q^3)$. Более подробно см. в [78].

б) Множество Р. Боуза [106] мощности q в $\mathbb{Z}_{q^2-1}$, где q — степень простого числа, определяется как $\{s_i \mid \theta^{s_i} = \theta + \alpha_i, GF(q) = \{\alpha_1, \ldots, \alpha_q\}\}$, где θ — примитивный элемент в поле $GF(q^2)$.

в) Множество В. Е. Алексеева [1] мощности $p-1$ в $\mathbb{Z}_{p(p-1)}$, где p — простое число, определяется как $\{s_i \mid s_i \equiv i \mod (p-1),\ \ s_i \equiv \zeta^i \mod p\}$, где ζ — порождающий элемент мультипликативной группы поля $GF(p)$.

Любой из перечисленных примеров трансформируется в множество Сидона в E_n мощности $(1-o(1))\sqrt{n}$. Это устанавливается, во-первых, тем, что редкое подмножество в $\mathbb{Z}_Q$ тем более является редким в E_Q (в смысле естественного взаимно однозначного гомоморфизма из E_Q в $\mathbb{Z}_Q$). Во-вторых, если $Q_R(m)$ — ближайшее снизу к m число вида $R(p)$, где R — (фиксиро-

ванный) многочлен, а p — простое число, то, как известно, $Q_R(m) \sim m$ при $m \to \infty$ (это следует из результатов о плотности распределения простых чисел, см., например, [190]).

Несколько известных примеров многомерных редких множеств приводятся в следующей теореме из [218]. Пусть $N = \mathbb{N} \cup \{0\}$. Обозначим через ψ_n следующее (достаточно естественное) отображение из $GF(2^n)$ в N^n (через $GF(q)$ обозначается поле Галуа порядка q). Пусть используется представление поля $GF(2^n)$ как векторного пространства размерности n над $GF(2)$. Тогда ψ_n переводит вектора над $GF(2)$ в вектора над N, при этом единичные компоненты — в единицы, нулевые компоненты — в нули.

Теорема 3.4.

(i.a) В группе $(GF(q)^2, +)$, *где* q *— нечетно, парабола* $\{(x, x^2) \mid x \in GF(q)\}$ *является 2-редким подмножеством мощности* q.

(i.b) В полугруппе $(N^{2n}, +)$ *«кубическая парабола»* $\{(\psi_n(x), \psi_n(x^3)) \mid x \in GF(2^n)\}$ *является 2-редким подмножеством мощности* 2^n.

(ii.a) В группе $(GF(q)^3, +)$, *где* q *— нечетно, сфера* $\{(x, y, z) \mid x^2 + y^2 + z^2 = \gamma\}$, *где* $-\gamma$ *— квадратичный невычет в* $GF(q)$, *является 3-редким подмножеством мощности* $q^2 - q$.

(ii.b) В группе $(GF(q)^3, +)$, *где* $q = 2^{2k+1}$, *поверхность* $\{(x, y, z) \mid x^3 + y^7 + z^{15} = 1\}$ *является* $(47, 315)$*-редким подмножеством мощности* q^2.

(iii) В группе $(GF(q^t), +)$ *множество* $\left\{ x \left| x^{\frac{q^t-1}{q-1}} = 1 \right. \right\}$ *элементов единичной нормы является* $(t, t! + 1)$*-редким подмножеством мощности* $(q^t - 1)/(q - 1)$.

Пример из п. $(i.a)$, вероятно, является фольклорным.

Пример из п. $(i.b)$ предложен Б. Линдстрёмом в [158].

Утверждение п. $(ii.a)$ фактически доказано У. Брауном в [112] для случая простого числа q. Более точно, он доказал, что пересечение любых трех различных сфер $S(a, b, c) = \{(x, y, z) \mid (x - a)^2 + (y - b)^2 + (z - c)^2 = \gamma\}$ состоит не более чем из двух точек. Рассуждение проходит и в общем случае.

Пример из п. $(ii.b)$ предложен С. Б. Гашковым в [6].

Утверждение п. (iii) непосредственно вытекает из результата работы

Я. Коллара, Л. Роньяи, Т. Сабо [152], в которой показано, что любые t различных множеств $N(a) = \left\{ x \,\middle|\, (x-a)^{\frac{q^t-1}{q-1}} = 1 \right\}$ пересекаются не более, чем по $t!$ точкам; наоборот, любые $t!+1$ таких множеств имеют не более $t-1$ общих точек.

Теорема 3.4 доставляет примеры редких множеств большой мощности в многомерных векторных пространствах. Излагаемый далее способ [218] позволяет такие множества трансформировать в числовые («одномерные») редкие множества.

Трансформация редких множеств

Лемма 3.6. *Обозначим* $N_{q,t} = \left\{ \sum_{i=0}^{t-1} a_i (2q-1)^i \,\middle|\, a_i \in E_q \right\}$. *Пусть* $A, B \subset N$ *и* $A \oplus B \subset N_{q,t}$. *Тогда существуют взаимно однозначные отображения* $\xi_A : A \to N_{q,t}$ *и* $\xi_B : B \to N_{q,t}$ *такие, что для любых* $a \in A$, $b \in B$ *выполняется* $a + b = \xi_A(a) + \xi_B(b)$.

Содержательно лемма означает, что если в записи элементов суммы Минковского $A \oplus B$ двух числовых множеств A, B в системе счисления с основанием $2q-1$ присутствуют только «цифры» от 0 до $q-1$, то элементы самих множеств можно видоизменить так, что их запись в указанной системе счисления также будет содержать только цифры от 0 до $q-1$, а попарные суммы при этом не изменятся.

Доказательство. Доказательство проведем индукцией по t. Если $t = 1$, то неизбежно $A, B \subset E_q$, поэтому в качестве ξ_A и ξ_B можно взять тождественные отображения. Доказательству индуктивного перехода предпошлем вспомогательное рассуждение.

В кольце вычетов $\mathbb{Z}_m = \{\overline{0}, \ldots, \overline{m-1}\}$ введем понятие *отрезка* $[\overline{a}, \overline{b}]$, определяемое при $a \le b$ как $\{\overline{a}, \overline{a+1}, \ldots, \overline{b}\}$, а при $a > b$ как $\{\overline{a}, \ldots, \overline{m-1}, \overline{0}, \ldots, \overline{b}\}$. *Длину* отрезка определим как $|[\overline{a}, \overline{b}]| - 1$, т. е. в первом случае как $b-a$, а во втором — как $b+m-a$. Для любого подмножества $M \subset \mathbb{Z}_m$ найдется отрезок минимальной длины (возможно, не один), содержащий M. Ясно, что оба конца такого отрезка должны принадлежать M. Длину минимального отрезка, содержащего M, назовем *диаметром* M и обозначим $d(M)$. Заметим, что если $d(M) \le (m-1)/2$, то минимальный

отрезок для M определен однозначно и он содержится в любом содержащем M отрезке длины не более $(m-1)/2$. Легко показать, что

$$\max\{d(M_1),\, d(M_2)\} \le d(M_1 \oplus M_2) \le \min\{d(M_1) + d(M_2),\, m-1\}.$$

Докажем следующее утверждение. Если $d(M_1 \oplus M_2) \le (m-1)/2$, то $d(M_1 \oplus M_2) = d(M_1) + d(M_2)$. При этом если $[\overline{a},\, \overline{b}]$ и $[\overline{c},\, \overline{d}]$ — отрезки минимальной длины для множеств M_1 и M_2 соответственно, то для множества $M_1 \oplus M_2$ минимальным является отрезок

$$\rho = [\overline{(a+c) \bmod m},\, \overline{(b+d) \bmod m}]$$

(напомним, при рассматриваемых ограничениях все минимальные отрезки определены однозначно).

Случай, когда одно из множеств имеет нулевой диаметр, является тривиальным, поэтому без ограничения общности будем считать, что $0 < d(M_1) \le d(M_2)$. В силу того, что $d(M_2) \le (m-1)/2$, отрезок ρ имеет длину $d(M_1) + d(M_2)$ и содержит множество $M_1 \oplus M_2$. Предположим, что этот отрезок не является минимальным для $M_1 \oplus M_2$. Поскольку все вычеты

$$\overline{(a+c) \bmod m},\ \overline{(b+c) \bmod m},\ \overline{(a+d) \bmod m},\ \overline{(b+d) \bmod m}$$

принадлежат множеству $M_1 \oplus M_2$, то любой отрезок, содержащий данное множество, содержит при $d(M_2) > d(M_1)$ один из отрезков

$$\rho = [\overline{(a+c) \bmod m},\, \overline{(b+d) \bmod m}], \quad [\overline{(b+c) \bmod m},\, \overline{(a+c) \bmod m}],$$
$$[\overline{(a+d) \bmod m},\, \overline{(b+c) \bmod m}], \quad [\overline{(b+d) \bmod m},\, \overline{(a+d) \bmod m}],$$

а при $d(M_1) = d(M_2)$ — один из перечисленных отрезков, не считая третьего.

Второй и четвертый отрезок имеют длину $m - d(M_1) > (m-1)/2$, а третий отрезок при $d(M_2) > d(M_1)$ имеет длину $m - (d(M_2) - d(M_1)) > (m-1)/2$. Следовательно, минимальный отрезок для $M_1 \oplus M_2$ должен содержать отрезок ρ, что противоречит предположению.

Перейдем к доказательству индуктивного перехода леммы. Пусть $t \ge 2$. Предположим, что утверждение леммы справедливо при подстановке $t-1$ вместо t. Определим множество «младших разрядов» $A_0 =$

$\left\{\overline{a \bmod (2q-1)}\middle| \, a \in A\right\}$ и аналогично — множество B_0. По условию леммы в кольце вычетов $\mathbb{Z}_{2q-1}$ справедливо вложение $A_0 \oplus B_0 \subset [\overline{0}, \overline{q-1}]$. Пусть $\rho_A = [\overline{l_A}, \overline{r_A}]$ и $\rho_B = [\overline{l_B}, \overline{r_B}]$ — минимальные отрезки для A_0 и B_0 соответственно. Тогда по доказанному выше отрезок

$$\rho = [\overline{(l_A + l_B) \bmod (2q-1)}, \overline{(r_A + r_B) \bmod (2q-1)}]$$

является минимальным для $A_0 \oplus B_0$. Ясно, что $\rho \subset [\overline{0}, \overline{q-1}]$ (поскольку любой содержащий некоторое подмножество $\mathbb{Z}_m$ отрезок длины не более $(m-1)/2$ содержит также минимальный отрезок для этого подмножества).

Теперь заметим, что если $l_A = 0$ или $l_B = 0$, то $r_A, r_B \leq q-1$. Действительно, пусть, например, $l_A = 0$. Тогда $l_B \leq q-1$ и $r_A \leq q-1$. Если предположить, что $r_B \geq q$, то в случае $l_B \leq r_A$ получаем

$$d(A_0 \oplus B_0) = (r_A - l_A) + (r_B - l_B) \geq r_B \geq q$$

(противоречие с условием $d(A_0 \oplus B_0) \leq q-1$), а в случае $l_B \geq r_A$ получаем

$$q \leq r_A + r_B = (l_A + l_B) + d(A_0 \oplus B_0) \leq l_B + q - 1 \leq 2q - 2$$

(противоречие с условием $\rho \subset [\overline{0}, \overline{q-1}]$). Поэтому $r_B \leq q-1$.

Далее заметим, что если $\overline{0} \in (\rho_A \cap \rho_B)$, то $\overline{0} \in \rho$, следовательно $l_A = l_B = 0$. Поэтому если $l_A \neq 0$ и $l_B \neq 0$, то $\overline{0} \notin (\rho_A \cap \rho_B)$, другими словами, либо $0 < l_A \leq r_A$, либо $0 < l_B \leq r_B$.

Следовательно, в любом случае выполнено либо $l_A \leq r_A$, либо $l_B \leq r_B$. Без ограничения общности предположим, что выполнено первое неравенство — оно означает, что $\forall_{a \in A}(a \geq l_A)$. Определим отображения ξ'_A и ξ'_B следующим образом: $\xi'_A(a) = a - l_A$, $\xi'_B(b) = b + l_A$.

Ясно, что $\xi'_A(a) + \xi'_B(b) = a + b$. В силу $l_{\xi'_A(A)} = 0$ имеем также $0 \leq l_{\xi'_B(B)}, r_{\xi'_A(A)}, r_{\xi'_B(B)} \leq q-1$ (где $l_{\xi'_A(A)}$ и $r_{\xi'_A(A)}$ определяются по аналогии с l_A и r_A). Как следствие, имеет место $\xi'_A(A), \xi'_B(B) \subset (2q-1)N \oplus E_q$, что значит: элемент a любого из множеств $\xi'_A(A), \xi'_B(B)$ можно записать в виде $a = (2q-1)a' + a''$, где $a'' \in E_q$, $a' \in N$. Определим новые множества

$$A' = \{a' \,|\, a = (2q-1)a' + a'' \in \xi'_A(A),\, a'' \in E_q\},$$
$$B' = \{b' \,|\, b = (2q-1)b' + b'' \in \xi'_B(B),\, b'' \in E_q\}.$$

Для любых $a' \in A'$ и $b' \in B'$ выполнено $a' + b' = \lfloor (a+b)/(2q-1) \rfloor$ при некоторых $a \in \xi'_A(A)$ и $b \in \xi'_B(B)$, потому что $0 \le a'' + b'' \le 2q-2$, где $a'' = a \bmod (2q-1)$ и $b'' = b \bmod (2q-1)$, и переносов из младшего разряда (в системе счисления с основанием $2q-1$) не возникает. Поэтому $A' \oplus B' \subset N_{q,t-1}$. Следовательно, по индуктивному предположению существуют взаимно однозначные отображения $\xi''_A : A' \to N_{q,t-1}$ и $\xi''_B : B' \to N_{q,t-1}$, сохраняющие суммы.

Окончательно положим $\xi_A(a) = (2q-1)\xi''_A(a') + a''$, где $\xi'_A(a) = (2q-1)a' + a''$, $a'' \in E_q$, и аналогично определим $\xi_B(b)$. По построению, отображение ξ_A (аналогично ξ_B) состоит в прибавлении (вычитании) некоторой целой константы, поэтому, в частности, является взаимно однозначным. $\square$

Замечание. Утверждение леммы останется справедливым, если сумму по степеням $2q-1$ в определении множества $N_{q,t}$ заменить суммой по степеням произвольного числа $Q \ge 2q-1$.

Рассмотрим отображение $\psi_{q,s,t}$ из E_q^{st} в $E^s_{(2q-1)^t}$, переводящее вектор $(a_0, \ldots, a_{st-1})$ в вектор

$$\left(\sum_{i=0}^{t-1} a_i (2q-1)^i, \ \sum_{i=0}^{t-1} a_{t+i} (2q-1)^i, \ldots, \sum_{i=0}^{t-1} a_{(s-1)t+i} (2q-1)^i \right). \qquad (3.28)$$

Теорема 3.5. *Если подмножество $M \subset E_q^{st}$ полугруппы $(N^{st}, +)$ является (k,l)-редким, то подмножество $\psi_{q,s,t}(M)$ полугруппы $(N^s, +)$ также является (k,l)-редким.*

Доказательство. Через $pr_i(A)$ будем обозначать множество i-х компонент множества $A \subset N^s$, а именно $pr_i(A) = \{a_i \mid (a_0, \ldots, a_{s-1}) \in A\}$. Оператор pr_i обладает очевидным свойством: $pr_i(A \oplus B) = pr_i(A) \oplus pr_i(B)$.

Проверим, что отображение $\psi_{q,s,t}$ сохраняет редкость множества. Действительно, согласно лемме 3.6, если для некоторых множеств $A, B \subset N^s$ верно, что множество $A \oplus B$ содержится в $\psi_{q,s,t}(M)$, то для каждого $i = 1, \ldots, s$ найдется пара взаимно однозначных отображений $\xi_{i,A} : pr_i(A) \to N_{q,t}$ и $\xi_{i,B} : pr_i(B) \to N_{q,t}$, обладающих свойством $\xi_{i,A}(a) + \xi_{i,B}(b) = a + b$ (где $N_{q,t}$ — из леммы 3.6). Следовательно, вектор-отображения $\xi_A = (\xi_{0,A}, \ldots, \xi_{s-1,A})$ и $\xi_B = (\xi_{0,B}, \ldots, \xi_{s-1,B})$ взаимно однозначно отображают

соответственно множество A и множество B в $N^s_{q,t} = \psi_{q,s,t}(E^{st}_q)$ и также обладают свойством $\xi_A(a) + \xi_B(b) = a + b$.

Теперь, поскольку отображение $\psi^{-1}_{q,s,t}$ сохраняет суммы и является взаимно однозначным, получаем, что $\psi^{-1}_{q,s,t}(\xi_A(A)) \oplus \psi^{-1}_{q,s,t}(\xi_B(B)) \subset M$ и при этом $|\psi^{-1}_{q,s,t}(\xi_A(A))| = |\xi_A(A)| = |A|$, $|\psi^{-1}_{q,t}(\xi_B(B))| = |\xi_B(B)| = |B|$. □

Следствия

Теорема 3.5 позволяет преобразовать редкие множества [152] (см. п. (*iii*) теоремы 3.4) в k-редкие подмножества мощности $n^{1-o(1)}$ в циклической группе $\mathbb{Z}_n$ при медленно растущем k и, как следствие, предъявить k-редкие циркулянтные матрицы веса $n^{2-o(1)}$. Более точно, из п. (*iii*) теоремы 3.4 (с выбором простого числа в качестве q) и теоремы 3.5 вытекает

Следствие 3.4. *При любом n можно эффективно*[22] *указать (k,l)-редкую циркулянтную матрицу порядка n и веса αn^2, где $k = O\left(\sqrt{\frac{\log n}{\log\log n}}\right)$, $l, \alpha^{-1} \in 2^{O(\sqrt{\log n \log\log n})}$.*

Помимо этого, указанный пример редкого множества дает наиболее сильную форму основного результата [218]: пример многочлена степени n от одной переменной с коэффициентами 0, 1, имеющего аддитивную сложность $n^{1-o(1)}$ и мультипликативную сложность $n^{1/2-o(1)}$ при реализации схемами в монотонном арифметическом базисе $\{\times, +\}$.

Отметим, что примеры редких матриц из работ [2, 97] не сводятся к конструкциям редких множеств.

3.3.2 Известные методы получения нижних оценок сложности

Метод Нечипорука

Метод получения нижних оценок, эксплуатирующий свойство редкости матрицы, был предложен Э. И. Нечипоруком в 1960-х гг.

Теорема 3.6 (Нечипорук [54, 57]). *Для $(k+1, l+1)$-редкой матрицы A*

$$\mathsf{OR}(A) \geq \frac{|A|}{k \cdot l}, \qquad \mathsf{OR}_2(A) \geq \frac{|A|}{\max\{k, l\}}.$$

[22] Матрица считается эффективной, если ее элементы выражаются булевой функцией от (двоичной записи) своих координат, реализуемой схемой полиномиальной сложности.

В частности, для 2-редкой матрицы A теорема дает точное значение сложности $\mathsf{OR}(A) = \mathsf{OR}_1(A) = |A|$.

Некоторое время этот результат оставался незамеченным[23], а затем был независимо и почти одновременно переоткрыт К. Мельхорном [163], Н. Пиппенджером [177] и И. Вегенером [208].

Первый пример матрицы с почти квадратичной нижней оценкой сложности построил А. Е. Андреев [2]. На сегодняшний день, рекордная оценка для конкретно заданных матриц имеет вид $n^2 \cdot 2^{-\Theta(\sqrt{\log n \log\log n})}$: она достигается на норм-матрицах из работы [152].

Обе оценки теоремы являются асимптотически неулучшаемыми [230], в том числе при растущих k и l. В качестве примера достаточно рассмотреть матрицу[24] $A = M \otimes J$, где M — 2-редкая матрица размера $m \times m$ большого веса, а J — сплошь единичная матрица размера $k \times k$ (по построению, матрица $M \otimes J$ является $(k+1)$-редкой):

$$\mathsf{OR}(A) = |A|/k^2 + O(mk), \qquad \mathsf{OR}_2(A) = |A|/k + O(mk).$$

Оценки через устойчивость

Устойчивость матрицы A над полем $\mathbb{F}$ определяется как функция $R_A(r)$, равная минимальному числу элементов матрицы, которое необходимо изменить для того, чтобы уменьшить ранг матрицы до $\leq r$. В интересующем нас случае поля $\mathbb{F} = GF(2)$ определение можно записать как

$$R_A(r) = \min\{|B| : \operatorname{rk} A \oplus B \leq r\}.$$

Несколько способов вывода нижних оценок XOR-сложности через оценку устойчивости предложено в работах Л. Вэльянта [206] (в которой и введено понятие устойчивости) и П. Пудлака [183]. Применение этих методов осложняется отсутствием достаточно высоких нижних оценок устойчивости. Нетривиальные оценки сложности в глубине 2 позволяет получать следующий способ.

[23] Кроме тривиального частного случая $k = l = 1$, который использовался позднее в работах Митягина и Садовского [50] и Нечипорука [58].

[24] Через $M \otimes J$ обозначается кронекерово произведение матриц M и J.

Теорема 3.7 (Пудлак [183]). *Пусть A — матрица размера $n \times n$, $f : \mathbb{N} \to \mathbb{R}_+$ — неубывающая функция, $1 \le a \le b \le n$. Если*

$$R_A(r) \ge \frac{f(n)^2}{r}$$

выполнено для всех $a \le r \le b$, то

$$\mathsf{XOR}_2(A) \ge 2f(n) \ln \frac{b}{a}.$$

Доказательство приводится также в [230].

Как пример приложения метода, можно привести полную треугольную матрицу $T = T_n$. В работе П. Пудлака и З. Ваврина [186] найдено точное значение устойчивости $R_T(r) \ge (1 - o(1))n^2/4r$. Как следствие, из теоремы 3.7 вытекает точная по порядку оценка $\mathsf{XOR}_2(T) \ge (1 - o(1))n \ln n$. Рекордные оценки сложности $\mathsf{XOR}_2(A) \succeq n \log^{3/2} n$ этим методом получены для «хороших» кодовых матриц, см. [184]. Однако в работе [132] для этого же класса матриц другим способом доказана неулучшаемая по порядку оценка $\mathsf{XOR}_2(A) = \Omega(n(\ln n / \ln \ln n)^2)$.

3.3.3 Оценки OR/XOR отношений в некоторых классах матриц

Матрицы без прямоугольников

Матрицей без прямоугольников называется 2-редкая матрица. Согласно теореме 3.6, OR-сложность матрицы без прямоугольников совпадает с ее весом. В работе Б. С. Митягина и Б. Н. Садовского [50] фактически поставлен вопрос, может ли такая матрица вычисляться существенно более простыми XOR-схемами. Таким образом, максимизация отношения OR/XOR состоит в построении примера легко вычисляемой XOR-схемами матрицы возможно большего веса.

В работе [50] была рассмотрена матрица Зингера: циркулянтная матрица, порождаемая 2-редким множеством Зингера [198]. Известно, что матрица, построенная на основе множества Зингера, фактически совпадает с матрицей, строки которой представляют из себя характеристические функции прямых конечной проективной плоскости порядка q. Тот факт, что матрица не содержит прямоугольников, вытекает из того, что любые две

прямые проективной плоскости пересекаются ровно в одной точке. Матрица Зингера, как легко проверяется, имеет экстремально большой вес среди всех (n, n)-матриц без прямоугольников, асимптотически $n^{3/2}$.

Хотя в [50] было высказано верное соображение о возможности простой реализации циркулянтных матриц при помощи быстрых алгоритмов умножения, доказательство основного результата оказалось неверным, как заметил Э. И. Нечипорук в своей рецензии [56] (подробнее см. в [215]). В итоге, результат [50] оставался неподтвержденным, пока автором совместно с С. Б. Гашковым не было опубликовано корректное доказательство в [215]. Оно очень простое.

Как известно (см., например, [101]), линейное преобразование с циркулянтной (n, n)-матрицей A есть циклическая свертка с вектором, образованным коэффициентами первой строки матрицы. Поэтому для XOR-сложности матрицы справедлива оценка

$$\mathsf{XOR}(A) = O(M(n)), \tag{3.29}$$

где $M(n)$ — число битовых операций сложения в билинейном алгоритме умножения многочленов над $GF(2)$ (билинейный алгоритм при фиксации одного из сомножителей становится линейным преобразованием коэффициентов второго сомножителя). Рекордная на сегодняшний день оценка сложности умножения двоичных многочленов составляет $O\left(n \log n \cdot 2^{O(\log^* n)}\right)$ [139]. Таким образом, справедливо

Следствие 3.5 ([215]). *Максимум отношения* $\lambda(n) = \mathsf{OR}(A)/\mathsf{XOR}(A)$ *по всем булевым* (n, n)*-матрицам без прямоугольников заключен в пределах:*

$$\sqrt{n}/\log^{1+o(1)} n \preceq \lambda(n) \preceq \sqrt{n},$$

причем нижняя оценка доказывается эффективно.

В работе [215] приведен еще один пример, на котором достигается такая же по порядку нижняя оценка — матрица инциденций точек и прямых конечной проективной плоскости над $GF(p)$, предложенная Т. Ковари, В. Шош и П. Тураном в работе [154]. Эта матрица также использовалась в работе Нечипорука [58]. Она является подматрицей матрицы Зингера, и

асимптотически настолько же плотная, как последняя. В работе [215] показано, что эта матрица является матрицей дискретного преобразования Фурье порядка p над кольцом $GF(2)[x]/(x^p+1)$ и, как следствие, допускает вычисление XOR-схемами сложности $O(M(n))$.

Отметим, что для случая реализации схемами глубины 3 из двухвходовых элементов примеры 2-редких матриц с существенным расхождением OR- и XOR-сложности построены С. Б. Гашковым в курсовой работе 1973 г. и К. А. Зыковым [17] (при такой постановке отношение сложностей не может превосходить константу).

Циркулянтные матрицы

Оценка (3.29) указывает широкий класс матриц, просто реализуемых XOR-схемами — циркулянтные матрицы. Естественно рассмотреть вопрос, насколько сложным этот класс является для OR-схем.

Этот вопрос изучал М. И. Гринчук в работе [12]. С использованием вероятностного метода он доказал существование k-редких циркулянтных $n \times n$ матриц веса $\Omega\left(k^{-4}n^{2-\sqrt{3/k}}\right)$ и получил следствия для OR- и OR_2-сложности. Наибольшие оценки получаются при выборе $k \asymp \log^2 n$: соответственно $n^2 \log^{-12} n$ и $n^2 \log^{-10} n$ по порядку. Более слабые, но тоже почти квадратичные оценки для этой задачи были позднее получены в [204].

Автором было предложено уточнение результата [12], опубликованное в совместной работе [216]. Уточнение состояло в применении более аккуратной оценки мощности суммы двух множеств в евклидовом пространстве в доказательстве [12]. В результате было установлено, что существуют (k,l)-редкие циркулянтные матрицы порядка n с весом $\Omega\left(\frac{k+l}{k^2l^2}\cdot n^{2-\frac{k+l+2}{kl}}\right)$. Для сравнения, классический результат Эрдёша—Спенсера [93] дает лишь чуть-чуть лучшую оценку $\Omega_{k,l}\left(n^{2-\frac{k+l-2}{kl-1}}\right)$ в классе всех (k,l)-редких матриц.

При выборе $k = l \asymp \log n$ согласно теореме 3.6 устанавливаем существование циркулянтных матриц OR_2-сложности $\Omega(n^2 \log^{-4} n)$ и OR-сложности $\Omega(n^2 \log^{-5} n)$.

В работе [216] также было замечено, что оценка OR-сложности класса циркулянтных матриц влечет оценку монотонной сложности функции булевой свертки, а если более точно, то оценку числа дизъюнкторов в вы-

числяющей свертку монотонной схеме.

Булева свертка порядка N — это функция

$$U_N(x_0,\dots,x_{N-1},y_0,\dots,y_{N-1}) = (u_0,\dots,u_{2N-2}), \quad u_k = \bigvee_{i+j=k} x_i y_j.$$

Циклическая булева свертка порядка N определяется как

$$Z_N(x_0,\dots,x_{N-1},y_0,\dots,y_{N-1}) = (z_0,\dots,z_{N-1}), \quad z_k = \bigvee_{i+j\equiv k \bmod N} x_i y_j.$$

Очевидная верхняя оценка сложности свертки над монотонным базисом $B_M = \{\vee, \wedge\}$ квадратична. Наилучшие нижние оценки числа конъюнкторов принадлежат Н. Блюму: $\Omega(N^{4/3})$, полученная в [104], и $\Omega(N^{3/2})$, анонсированная в [105].

Если обозначить через $V(f)$ минимальное число дизъюнкторов в схеме, реализующей функцию f над базисом B_M, то прямо из определений функций свертки легко вывести соотношения

$$V(Z_N) \le V(U_N) + N - 1, \qquad V(U_N) \le V(Z_{2N-1}).$$

Ранее известная наилучшая нижняя оценка числа дизъюнкторов в схеме для булевой свертки была установлена Ю. Вайсом в работе [211] и имела вид $V(Z_N) = \Omega(N^{3/2})$.

На самом деле, почти квадратичная нижняя оценка вытекает уже непосредственно из работы [12]. Циклическую булеву свертку (с точностью до перестановки компонент) можно интерпретировать как систему булевых сумм переменных $x_0,\dots,x_{N-1}$ с переменной циркулянтной матрицей, определяемой строкой $y_{N-1},\dots,y_0$. Так как сложность схемы (в данном случае, в смысле меры $V(f)$) не возрастает при подстановке констант вместо некоторых входов, то заключаем, что сложность циклической свертки порядка N не меньше, чем сложность системы булевых сумм с произвольной циркулянтной матрицей порядка N. Используя приведенные выше оценки из [216], получаем

Следствие 3.6. $V(U_N)$, $V(Z_N) = \Omega\left(N^2 \log^{-6} N\right)$.

Указанное следствие пока доставляет наилучшую нижнюю оценку и для общей монотонной сложности оператора булевой свертки.

Оценки в ограниченной глубине

В работе [236] автором доказан аналог (3.29) для схем ограниченной глубины. Этот результат позволяет построить примеры с растущим отношением $\mathsf{OR}_d/\mathsf{XOR}_d$ и в классе циркулянтных матриц, и в классе матриц без прямоугольников.

Теорема 3.8. *При $k \in \mathbb{N}$ произвольную булеву циркулянтную $n \times n$ матрицу Z можно реализовать* XOR*-схемой:*

а) глубины $2k-1$ и сложности не более $f(2k-1)n^{1+1/k}$;

б) глубины $2k$ и сложности не более $f(2k)n\left(\frac{n}{\log n}\right)^{1/k}$.

Доказательство. Рассуждение проведем по индукции. При $k = 1$ предъявляются тривиальная схема глубины 1 и сложности $O(n^2)$, а также схема глубины 2 и сложности $O(n^2/\log n)$, которая строится методом О. Б. Лупанова [41].

Докажем индуктивный переход от $k-1$ к k. Для доказательства используется метод А. Л. Тоома умножения многочленов [74] вместе с приемом А. Шёнхаге [191], позволяющим распространить метод на двоичные многочлены. Умножение многочленов с глубиной d сводится к нескольким параллельным умножениям глубины $d-2$.

Разобьем вектор переменных на q частей длины n/q. Эти части интерпретируем как вектора коэффициентов многочленов из кольца

$$R = GF(2)[y]/(y^{2\cdot 3^s} + y^{3^s} + 1)$$

при младших степенях y. Параметр s выбирается из условия $3^s \geq n/q$.

Сведем умножение двоичных многочленов степени $n-1$ к умножению многочленов степени $n/q-1$ над R. Последнее умножение выполним при помощи дискретного преобразования Фурье (ДПФ) порядка $3^m \geq 2q$ с примитивным корнем $\zeta = y^{s+1-m} \in R$.

Опишем схему. На входе многочлен $B(x) = \sum B_i x^i \in R[x]$ степени $q-1$. Постоянный сомножитель обозначим через $A(x) = \sum A_i x^i$. На выходе — произведение $C(x) = A(x)B(x) = \sum C_i x^i$.

1. Вычисляем $B(\zeta^0), \ldots, B(\zeta^{3^m-1})$.

2. Вычисляем $C(\zeta^i) = A(\zeta^i)B(\zeta^i)$ для всех $i = 0, \ldots, 3^m - 1$.

3. Вычисляем коэффициенты многочлена $C(x)$.

Этапы 1 и 3 реализуем схемами глубины 1, а этап 2 — схемой глубины $d-2$. Оценим сложность схемы, обозначим ее $M(d, n)$.

1. Умножение на степень y в кольце R выполняется с линейной сложностью, поэтому с линейной сложностью вычисляется значение многочлена $F(x)$ в точке y^p. Поэтому сложность первой схемы оценим как $O(3^m 3^s q)$.

2. Каждое из умножений на шаге 2 есть умножение двоичных многочленов степени $2 \cdot 3^s - 1$ с последующим приведением по модулю. Умножение выполним методом из индуктивного предположения глубины $d-2$ и сложности $M(d-2, 2 \cdot 3^s)$. Приведение многочлена (в данном случае, степени $4 \cdot 3^s - 1$) по модулю $y^{2 \cdot 3^s} + y^{3^s} + 1$ достигается дублированием некоторых его коэффициентов (выходов предшествующей схемы) и отождествлением некоторых коэффициентов, так как каждый коэффициент приводимого многочлена используется не более чем дважды. Следовательно, приведение по модулю можно реализовать без увеличения глубины и с не более чем двукратным увеличением сложности. Общую сложность второго этапа теперь можно оценить как $2 \cdot 3^m M(d-2, 2 \cdot 3^s)$.

3. Согласно основному свойству ДПФ искомые коэффициенты многочлена $C(x)$ находятся как $C_i = C^*(\zeta^{-i})$, где многочлен $C^*(x)$ имеет коэффициенты $C(\zeta^i)$. Поэтому сложность этапа 3 можно оценить так же, как и сложность этапа 1, величиной $O(3^m 3^s q)$.

Чтобы результат умножения многочленов над R преобразовать (обратно) в результат умножения двоичных многочленов, надо выполнить подстановку $x = y^{2 \cdot 3^s}$ и привести подобные. Это преобразование реализуется отождествлением выходов и не влияет на глубину и сложность схемы.

Оценки теоремы получаются при следующем выборе параметров: $q = n^{1/k}$ при $d = 2k-1$ и $q = (n/\log n)^{1/k}$ при $d = 2k$; $3^s = \Theta(n/q)$, $3^m = \Theta(q)$.

По построению, $f(k) = 2^{O(k)}$. □

Следствие 3.7. *Для $n \times n$ матрицы Зингера S_n справедливо*

$$\mathsf{OR}_4(S_n)/\mathsf{XOR}_4(S_n) \succeq \sqrt{\log n}, \qquad \mathsf{OR}_{2t-1}(S_n)/\mathsf{XOR}_{2t-1}(S_n) \succeq n^{1/2-1/t}$$

при $t \geq 3$.

По-видимому, 4 — это минимальная глубина, в которой получено расхождение между OR и XOR-сложностью матриц без прямоугольников.

Оценки в классе всех матриц

Опираясь на комбинаторный результат [185], С. Юкна показал, что специальные подматрицы матриц Сильвестра позволяют достичь практически экстремального соотношения между OR- и XOR-сложностью (доказательство фактически содержится в [143], а в явном виде — в [230]).

Напомним, что $n \times n$ *матрицу Сильвестра* H_n при $n = 2^k$ можно определить как матрицу скалярных произведений над $GF(2)$: $H_n[u, v] = \langle u, v \rangle$, где строки и столбцы занумерованы булевыми векторами длины k.

Пусть $n = 2^r$ и $S \subset \{0, 1\}^{2r}$, $|S| = 2^r$. Определим $n \times n$ матрицу H_S как подматрицу матрицы H_{n^2}, образованную строками и столбцами, отмеченными векторами из S.

Результат [143] состоит в том, что почти для всех подмножеств S матрица H_S является $2r$-рамсеевой, т. е. не содержит ни сплошь нулевых, ни сплошь единичных подматриц размера $2r \times 2r$. Как следует из леммы 3.6, для почти любой матрицы H_S выполнено

$$\mathsf{OR}_2(H_S) \succeq \frac{n^2}{\log n}, \qquad \mathsf{OR}(H_S) \succeq \frac{n^2}{\log^2 n}.$$

С другой стороны, для любой матрицы H_S имеют место простые верхние оценки $\mathsf{XOR}_2(H_S) \preceq n \log n$ и $\mathsf{XOR}_3(H_S) \asymp n$, см., например, [230]. Поэтому почти для всех матриц H_S справедливы соотношения

$$\mathsf{OR}_2(H_S)/\mathsf{XOR}_2(H_S) \succeq \frac{n}{\log^2 n}, \qquad \mathsf{OR}(H_S)/\mathsf{XOR}_3(H_S) \succeq \frac{n}{\log^2 n}.$$

Альтернативные доказательства достижимости этих отношений (на других примерах матриц) получены в работах [107, 128].

Про экстремальные отношения для конкретно заданных матриц известно следующее. Для само́й матрицы Сильвестра $H = H_n$ выполняется соотношение

$$\mathsf{OR}_2(H)/\mathsf{XOR}_2(H) \asymp \frac{\sqrt{n}}{\log n}.$$

В глубине 2 этот пример является пока рекордным. Применяя к матрице A_n из следствия 3.4 соотношение (3.29), теорему 3.8 и лемму 3.6, получаем

Следствие 3.8. *Для некоторой явно заданной $n \times n$ матрицы A_n выполнено*

$$\mathsf{OR}(A_n)/\mathsf{XOR}(A_n) \succeq \frac{n}{\Delta}, \qquad \mathsf{OR}(A_n)/\mathsf{XOR}_{2t-1}(A_n) \succeq \frac{n^{1-1/t}}{\Delta}$$

при любом $t \in \mathbb{N}$, где $\Delta = 2^{\Theta(\sqrt{\log n \log\log n})}$.

Первое из неравенств следствия получено в [215] непосредственно для норм-матрицы из [152] с использованием вместо перестройки в циркулянтную матрицу быстрого алгоритма многомерной свертки для доказательства верхней оценки XOR-сложности.

3.3.4 Пример последовательности матриц с растущим XOR/OR отношением в глубине 2

Проблема построения матриц A с растущим отношением $\mathsf{XOR}(A)/\mathsf{OR}(A)$ близка к проблеме доказательства нелинейных нижних оценок XOR-сложности и является по этой причине сложной. К настоящему времени не доказано даже существование таких матриц. Имеющийся запас методов позволил автору в работе [230] построить пример $n \times n$ матрицы K_n, для которой $\mathsf{XOR}_2(K_n)/\mathsf{OR}_2(K_n) \to \infty$. Материал настоящего параграфа содержится в [230].

При $n = 2^r$ определим $n \times n$ *матрицу пересечений* D_n. Занумеруем строки и столбцы матрицы всевозможными подмножествами множества E_r. Положим $D_n[u, v] = (|u \cap v| > 0)$. Важнейшее свойство матрицы D_n — разложимость над OR-полугруппой в произведение $U \cdot U^T$, где U — $n \times r$ матрица, образованная всевозможными булевыми строками длины r.

Определим подматрицу $D_{n,k}$ матрицы пересечений D_n, образованную столбцами и строками, индексируемыми непустыми множествами мощности $\leq k$. Матрица $D_{n,k}$ имеет размер $N(n,k) \times N(n,k)$, где

$$N(n,k) = \sum_{i=1}^{k} \binom{n}{i}.$$

Рассмотрим кронекерово произведение матриц вида $A = D_{n,k} \otimes D_{m,p}$. Строки и столбцы матрицы A нумеруются парами (S, T) подмножеств $S \subset X$ и $T \subset Y$ размера $1 \le |S| \le k$, $1 \le |T| \le p$, где X, Y — непересекающиеся множества мощности n и m соответственно. По определению, $A[(S,T),(S',T')] = 1$ тогда и только тогда, когда $S \cap S' \neq \emptyset$ и $T \cap T' \neq \emptyset$.

Что существенно, такие матрицы имеют полный ранг над $GF(2)$. Это есть обобщение того факта, что сами матрицы $D_{n,k}$ имеют полный ранг, см., например, [145, лемма 4.11].

Лемма 3.7. *Матрица $A = D_{n,k} \otimes D_{m,p}$ имеет полный ранг над $GF(2)$.*

Доказательство. Воспользуемся стратегией доказательства невырожденности матрицы $D_{n,k}$ из работы А. А. Разборова [63]. Матрица A имеет размер $NM \times NM$, где $N = N(n,k)$, $M = N(m,p)$. Покажем, что строки матрицы линейно независимы над $GF(2)$, т. е. для любого ненулевого вектора

$$\lambda = (\lambda_{I_1,J_1}, \lambda_{I_1,J_2}, \ldots, \lambda_{I_N,J_M}) \in GF(2)^{NM},$$

индексируемого парами подмножеств $I \subset E_n$, $J \subset E_m$ размера $1 \le |I| \le k$, $1 \le |J| \le p$, выполнено $\lambda^T A \neq 0$; множества I, J состоят из индексов элементов множеств X и Y.

Рассмотрим булеву функцию

$$f(x_0, \ldots, x_{n-1}, y_0, \ldots, y_{m-1}) = \bigoplus_{1 \le |I| \le k,\ 1 \le |J| \le p} \lambda_{I,J} \left(\bigvee_{i \in I} x_i \right) \cdot \left(\bigvee_{j \in J} y_j \right).$$

Поскольку хотя бы один из коэффициентов $\lambda_{I,J}$ не равен нулю, найдется пара (I_0, J_0), такая что $\lambda_{I_0,J_0} \neq 0$ и $\lambda_{I,J} = 0$ для всех прочих пар (I, J), удовлетворяющих условию $I_0 \subset I$, $J_0 \subset J$. Иначе говоря, пара (I_0, J_0) максимальна. Без ограничения общности можно считать, что $I_0 = \{0, \ldots, t\}$, $J_0 = \{0, \ldots, u\}$. В функцию f подставим $x_i = 0$ для всех $i \notin I_0$ и $y_j = 0$ для всех $j \notin J_0$.

Полученная подфункция зависит от переменных $x_0, \ldots, x_t, y_0, \ldots, y_u$. При этом, ее многочлен Жегалкина содержит моном $x_0 \cdot \ldots \cdot x_t \cdot y_0 \cdot \ldots \cdot y_u$ в силу максимальности пары (I_0, J_0) и того факта, что многочлен дизъюнкции имеет максимальную степень. Поэтому подфункция принимает зна-

чение 1 на некотором наборе $(a_0, \ldots, a_t, b_0, \ldots, b_u)$ значений переменных. Следовательно, сама функция f принимает значение 1 на наборе

$$c = (a_0, \ldots, a_t, 0, \ldots, 0, b_0, \ldots, b_u, 0, \ldots, 0).$$

Таким образом,

$$1 = f(c) = \bigoplus_{I,J} \lambda_{I,J} \Big(\bigvee_{i \in I} c_i \Big) \cdot \Big(\bigvee_{j \in J} c_{j+n} \Big).$$

Положим $I' = \{i : a_i = 1\}$, $J' = \{j : b_j = 1\}$. Тогда $|I'| \le k$, $|J'| \le p$. Далее, условие $(\bigvee_{i \in I} c_i)(\bigvee_{j \in J} c_{j+n}) = 1$ эквивалентно $I \cap I' \ne \emptyset$, $J \cap J' \ne \emptyset$, а это равносильно $A[(I, J), (I', J')] = 1$. Окончательно получаем

$$\bigoplus_{I,J} \lambda_{I,J} \cdot A[(I, J), (I', J')] = 1,$$

что означает: координата (I', J') вектора $\lambda^T A$ — ненулевая. □

Матрица K_n определяется для $n = N(k^2, k) \cdot N(p^2, p)$ как $D_{k^2,k} \otimes D_{p^2,p}$ с условием $k^2 \asymp N(p^2, p)/p$. Таким образом,

$$k \asymp \frac{\log n}{\log \log n}, \qquad p \asymp \frac{\log \log n}{\log \log \log n}.$$

Теорема 3.9.

$$\mathsf{OR}_2(K_n) \preceq n \cdot \frac{\log n}{\log \log n}, \qquad \mathsf{XOR}_2(K_n) \succeq n \cdot \frac{\log n \cdot \log \log \log n}{\log \log n}.$$

Доказательство. Верхняя оценка OR_2-сложности следует из представления матрицы K_n в виде произведения $B \cdot B^T$ над булевым полукольцом, где B — $n \times (k^2 + p^2)$ матрица, определяемая условием $B[(S, T), q] = (q \in S \cup T)$ (столбцы матрицы B индексируются элементами из $X \cup Y$). В силу $|S| \le k$, $|T| \le p$ каждая строка B содержит не более $k + p$ единиц, следовательно вес матрицы B не превосходит $(k + p)n \asymp n \log n / \log \log n$.

Нижняя оценка XOR_2-сложности доказывается при помощи теоремы 3.7. Напомним, что эта теорема гарантирует оценку $\mathsf{XOR}_2(A) \ge 2f(n) \ln(b/a)$ при условии $R_A(r) \ge f^2(n)/r$ для всех r, $a \le r \le b$.

Мы докажем, что при некоторых постоянных $c_1, c_2 > 0$ неравенство

$$R_{K_n}(c_1 r) \geq \frac{c_2 n^2}{r} \tag{3.30}$$

выполняется для чисел $r = r(t,u) = N(k^2,t)\cdot N(p^2,u)$, $1 \leq t \leq k$, $1 \leq u \leq p$.

Однако, условия теоремы 3.7 требуют оценок устойчивости не только для отдельных значений $r = c_1 r(t,u)$, но и для всех промежуточных значений r в интервале между $c_1 r(1,1)$ и $c_1 r(k,p)$. (Именно требование заполнения этих интервалов вынуждает рассматривать кронекеровы произведения вместо, скажем, непосредственно матриц $D_{n,k}$.) Для оценки устойчивости с промежуточными значениями параметра мы используем простые соотношения

$$\frac{r(t,u+1)}{r(t,u)} = \frac{N(p^2,u+1)}{N(p^2,u)} \leq 1 + \binom{p^2}{u+1} \Big/ \binom{p^2}{u} \leq p^2,$$
$$\frac{r(t+1,1)}{r(t,p)} = \frac{N(k^2,t+1)\cdot N(p^2,1)}{N(k^2,t)\cdot N(p^2,p)} \leq \frac{k^2 p^2}{N(p^2,p)} \asymp p,$$

справедливые согласно выбору $k^2 \asymp N(p^2,p)/p$. Указанные соотношения гарантируют, что при любом r в интервале между $c_1 r(1,1)$ и $c_1 r(k,p)$ найдутся t, u, такие что $r \leq c_1 r(t,u) \leq p^2 r$. Как следствие,

$$R_{K_n}(r) \geq R_{K_n}(c_1 r(t,u)) \geq \frac{c_2 n^2}{r(t,u)} \geq \frac{c_2 (n/p)^2}{r}.$$

Применяя теорему 3.7 с параметрами $f(n) = \sqrt{c_2}n/p$, $a = c_1 r(1,1)$, $b = c_1 r(k,p)$, получаем

$$\mathsf{XOR}_2(K_n) \succeq (n/p)\log n \asymp n \cdot \frac{\log n \cdot \log\log\log n}{\log\log n}.$$

Осталось доказать (3.30). Нам пригодятся простые неравенства

$$\binom{k^2}{t} \leq N(k^2,t) \leq \left(1 + t\cdot\frac{t}{k^2-t}\right)\binom{k^2}{t} \leq 6\cdot\binom{k^2}{t},$$

справедливые при $t \leq 2k$ и $k \geq 10$, а также соотношения

$$\binom{k^2}{t} \geq \binom{k^2-2k}{t} \geq e^{-32}\binom{k^2}{t},$$

справедливые при $t \le 2k$ и $k \ge 4$. Таким образом,

$$N(k^2, t) \ge N(k^2 - 2k, t) \ge e^{-32} N(k^2, t).$$

Выберем $1 \le t \le k$, $1 \le u \le p$ и отметим минимальное множество коэффициентов матрицы K_n, инвертирование которых позволяет уменьшить ранг матрицы до $c_1 r(t, u)$.

Определим (t, u)-*подматрицу* матрицы K_n, зафиксировав пару непересекающихся подмножеств $I, I' \subset X$ мощности $k - t$ и пару непересекающихся подмножеств $J, J' \subset Y$ мощности $p - u$. Подматрица образуется пересечением всех строк (S, T) и столбцов (S', T'), таких что $I \subset S$, $I' \subset S'$, $J \subset T$, $J' \subset T'$ (все включения строгие). Рассмотрим подматрицу B (t, u)-подматрицы A, для индексов строк и столбцов которой выполняются дополнительные условия $I \cap S' = I' \cap S = J \cap T' = J' \cap T = \emptyset$.

По построению, матрица B с точностью до перестановки строк и столбцов совпадает с матрицей $D_{k^2-2(k-t),t} \otimes D_{p^2-2(p-u),u}$. Это так, поскольку $B[(S, S'), (T, T')] = 1$ равносильно $(S \setminus I) \cap (S' \setminus I') \ne \emptyset$ и $(T \setminus J) \cap (T' \setminus J') \ne \emptyset$. Согласно лемме 3.7, такая матрица имеет полный ранг

$$\operatorname{rk} B = N(k^2 - 2(k-t), t) \cdot N(p^2 - 2(p-u), u).$$

Как следствие, $\operatorname{rk} B \ge c_3 r(t, u)$ с некоторой константой $c_3 > 0$. Положим $c_1 = c_3/2$. Поскольку ранг B не может быть уменьшен вдвое изменением менее чем $(\operatorname{rk} B)/2$ коэффициентов B, в подматрице B (и следовательно, в (t, u)-подматрице A) не менее $c_1 r(t, u)$ коэффициентов отмечены.

Теперь оценим общее число отмеченных коэффициентов в K_n. В любой (t, u)-подматрице A их по меньшей мере $c_1 r(t, u)$. С другой стороны, любой коэффициент K_n относится, самое большее, к $\binom{k}{t}^2 \cdot \binom{p}{u}^2$ таким подматрицам. Действительно, рассмотрим произвольный коэффициент матрицы K_n с номером $[(S, T), (S', T')]$. Определяемая парами (I, J) и (I', J') (t, u)-подматрица A включает его только при $I \subset S$, $I' \subset S'$, $J \subset T$, $J' \subset T'$. Поскольку $|I| = |I'| = k - t$ и $|J| = |J'| = p - u$, есть $\binom{|S|}{k-t} \le \binom{k}{t}$ возможностей для выбора I или I', а также $\binom{|T|}{p-u} \le \binom{p}{u}$ возможностей для выбора J или J'.

Так как имеется $\binom{k^2}{k-t} \cdot \binom{k^2-(k-t)}{k-t}$ пар дизъюнктных множеств (I, J) и $\binom{p^2}{p-u} \cdot \binom{p^2-(p-u)}{p-u}$ пар дизъюнктных множеств (I', J'), умноженное на $r(t,u)$ число отмеченных коэффициентов матрицы K_n оценивается снизу как

$$c_1 r^2(t,u) \frac{\binom{k^2}{k-t}\binom{k^2-(k-t)}{k-t}\binom{p^2}{p-u}\binom{p^2-(p-u)}{p-u}}{\binom{k}{t}^2\binom{p}{u}^2} \geq$$

$$\geq c_4 \left[\frac{\binom{k^2}{t}\binom{k^2}{k-t}\binom{p^2}{u}\binom{p^2}{p-u}}{\binom{k}{t}\binom{p}{u}}\right]^2 \geq c_4 \left[\frac{(k^2-k)^k (p^2-p)^p}{k!p!}\right]^2 \geq c_5 n^2$$

при некоторых положительных постоянных c_4, c_5. Как следствие, неравенство $R_{K_n}(c_1 r) \geq c_2 n^2/r$ выполняется при $c_2 = c_5$ для всех $r = r(t,u)$. Соотношение (3.30), а вместе с ним и теорема, доказаны. □

Обобщение матрицы K до многократного кронекерова произведения K' позволяет достичь отношения

$$\mathsf{XOR}_2(K')/\mathsf{OR}_2(K') \succeq (\log\log n)^{1-o(1)}.$$

3.3.5 Отношение OR-сложностей матрицы и ее дополнения

Поставим вопрос: может ли сложность матрицы существенно отличаться от сложности дополнительной матрицы. В случае XOR-схем разница, разумеется, не может превышать $2n$. В случае OR-схем отношение $\mathsf{OR}(\overline{A})/\mathsf{OR}(A)$ доказуемо может быть близко к предельной величине $n/\log n$.

Метод построения экстремальных в смысле отношения $\mathsf{OR}(\overline{A})/\mathsf{OR}(A)$ матриц предложен Н. Кацем [149] для решения близкой задачи. В работе [230] аналогичным методом получен следующий результат.

Теорема 3.10. *Существует $n \times n$ матрица A, такая, что матрица $\overline{A}$ является $(\ln n)$-редкой, имеет вес $|\overline{A}| \asymp n^2$, но при этом*

$$\mathsf{OR}_2(A) \preceq n \log^2 n, \qquad \mathsf{OR}_3(A) \preceq n \log n.$$

Как следствие (из этой теоремы и теоремы 3.6), устанавливается существование матрицы A, для которой справедливо:

$$\mathsf{OR}_2(\overline{A})/\mathsf{OR}_2(A),\ \mathsf{OR}(\overline{A})/\mathsf{OR}_3(A) \succeq \frac{n}{\log^3 n}.$$

Оригинальное рассуждение [149] доказывает существование матрицы A, $\mathrm{rk}_\vee(A) \preceq \log n$, дополнительная матрица к которой является 2-редкой и при этом достаточно плотной: $|\overline{A}| \succeq n^{1.1}$. Можно построить также матрицу A сложности $\mathsf{OR}_2(A) \preceq n \log^2 n$, для которой дополнительная матрица является 2-редкой и имеет вес $|\overline{A}| \succeq n^{5/4}$ [230].

В основе доказательства теоремы 3.10 лежит вероятностная конструкция. Дерандомизация рассуждения позволила автору явно построить примеры дополнительных друг к другу матриц с большим отношением сложностей в глубине 2 [230].

Теорема 3.11. *Можно эффективно указать $n \times n$ матрицу A, для которой справедливо соотношение*

$$\mathsf{OR}_2(\overline{A})/\mathsf{OR}_2(A) \succeq \sqrt{n} 2^{-\Theta(\log^{2/3} n)}.$$

Доказательство. Пусть $n = \binom{p}{k}$, $p = 2^r$ и r — четно. Рассмотрим схему глубины 3 с n входами и выходами, занумерованными различными k-элементными подмножествами множества E_p. Схема также имеет по p вершин, занумерованных числами из E_p, на каждом из двух средних уровней. Соединим каждый вход и каждый выход $a \in \binom{E_p}{k}$ схемы со всеми k вершинами $b \in a$ соседнего уровня. Пусть на среднем слое схемы реализуется матрица Сильвестра H_p. Через A обозначим $n \times n$ матрицу, вычисляемую описанной схемой.

По построению, $\mathrm{rk}_\vee(A) \le p$, поэтому $\mathsf{OR}_2(A) \preceq pn$. С другой стороны, число нулей в матрице A совпадает с числом сплошь нулевых $k \times k$ подматриц H_p. Последнее число можно оценить снизу как произведение числа $n = \binom{p}{k}$ способов выбрать k строк $\{a_1, \ldots, a_k\}$ и числа $\binom{p2^{-k}}{k}$ способов выбрать k штук из по меньшей мере $2^{r-k} = p2^{-k}$ решений x системы линейных уравнений $\langle a_1, x\rangle = \ldots = \langle a_k, x\rangle = 0$ над $GF(2)$. Следовательно, $|\overline{A}| \ge n\binom{p2^{-k}}{k}$.

Покажем, что матрица $\overline{A}$ — K-редкая при $K = \binom{\sqrt{p}+1}{k}$. Заметим, что для любого множества S из $m \ge k$ вершин первого (или второго) из средних уровней схемы (соответствующих строкам/столбцам H_p) ровно $\binom{m}{k}$ входов (выходов) схемы соединяются исключительно с вершинами из S — это в

точности те входы или выходы, метки которых содержатся в S. Напомним, что матрица Сильвестра H_p является $(\sqrt{p}+1)$-рамсеевой, см., например, [145]. Если бы матрица $\overline{A}$ не была K-редкой, то матрица H_p содержала бы сплошь нулевую $m \times m$ подматрицу, $m = \sqrt{p}+1$, а это невозможно.

Согласно теореме 3.6,

$$\mathsf{OR}_2(\overline{A}) \geq |\overline{A}|/K \geq n\binom{p2^{-k}}{k}\Big/\binom{\sqrt{p}+1}{k}.$$

Окончательно, выберем $k \asymp \log^{1/3} n$, что значит $p = 2^{\Theta(\log^{2/3} n)}$. □

Эффективные примеры матриц с высоким отношением $\mathsf{OR}(\overline{A})/\mathsf{OR}(A)$ построены автором в [238]. Назовем *k-прямоугольником* сплошь единичную матрицу размера $k \times k$.

Теорема 3.12.

(i) Для некоторой конкретно заданной булевой $n \times n$ матрицы C выполнено $\mathsf{OR}(\overline{C})/\mathsf{OR}(C) = n \cdot 2^{-O(\sqrt{\ln n \ln\ln n})}$.

(ii) Для некоторой конкретно заданной булевой $n \times n$ матрицы C выполнено: $\mathsf{OR}(C) = O(n)$, *матрица $\overline{C}$ является 2-редкой и* $|\overline{C}| = \Omega(n^{4/3})$.

(Напомним, что 2-редкая матрица не может иметь вес выше $n^{3/2}+n$.)

Доказательство теоремы опирается на простую комбинаторную лемму.

Лемма 3.8. *Пусть $n \times n$ матрица A имеет вес $|A| \geq 2n^{3/2}$. Тогда она содержит $\Omega\left((|A|/n)^4\right)$ 2-прямоугольников.*

Доказательство. Скажем, что строка матрицы покрывает пару столбцов u, если в позициях на пересечении строки и этих столбцов стоят единицы. Если обозначить через a_i число единиц в i-й строке матрицы A, то общее число покрытий строками пар столбцов можно оценить как

$$\sigma = \sum_{i=1}^{n}\binom{a_i}{2} = \frac{1}{2}\sum_{i=1}^{n} a_i^2 - \frac{|A|}{2} \geq \frac{\left(\sum_{i=1}^{n} a_i\right)^2}{2n} - \frac{|A|}{2} = \frac{|A|^2}{2n} - \frac{|A|}{2} \geq \frac{|A|^2}{4n}.$$

Обозначим через b_u число строк, покрывающих пару столбцов u. Тогда

$\sum_u b_u = \sigma$. Число 2-прямоугольников в матрице A при этом равно

$$\sum_u \binom{b_u}{2} = \frac{1}{2}\sum_u b_u^2 - \frac{\sigma}{2} \geq \frac{\left(\sum_u b_u\right)^2}{n(n-1)} - \frac{\sigma}{2} = \\ = \frac{\sigma^2}{n(n-1)} - \frac{\sigma}{2} \geq \frac{\sigma^2}{2n^2} = \Omega\left((|A|/n)^4\right). \quad \square$$

Пусть $n = \binom{m}{2}$. По булевой $m \times m$ матрице A построим $n \times n$ матрицу B следующим образом. Занумеруем строки и столбцы матрицы B 2-элементными подмножествами множества E_m. Положим $B[a,b] = 1$ в том и только том случае, когда на пересечении строк a и столбцов b в матрице A расположен 2-прямоугольник.

Лемма 3.9. *Если матрица A является k-редкой, то матрица B является K-редкой, $K = \binom{k-1}{2} + 1$.*

Доказательство. Если матрица B содержит K-прямоугольник на пересечении строк $s_1, \ldots, s_K$ и столбцов $t_1, \ldots, t_K$, то матрица A содержит прямоугольник на пересечении строк $\cup s_i$ и столбцов $\cup t_i$. При этом $|\cup s_i|, |\cup t_i| \geq k$, что противоречит k-редкости матрицы A. $\square$

Лемма 3.10. *Если матрица A является k-редкой и $|A| \geq 2m^{3/2}$, то*

$$\mathsf{OR}(B) = \Omega\left(\left(\frac{|A|}{kn}\right)^4\right),$$

при этом $\mathsf{OR}_3(\overline{B}) = O(n)$.

Доказательство. Согласно лемме 3.8, $|B| = \Omega((|A|/n)^4)$, а из леммы 3.9 следует, что B является K-редкой. Поэтому по теореме 3.6

$$\mathsf{OR}(B) \geq \frac{|B|}{K^2} = \Omega\left(\left(\frac{|A|}{kn}\right)^4\right).$$

Покажем, что матрицу $\overline{B}$ можно реализовать схемой глубины 3 и линейной сложности. На втором и третьем уровнях схемы разместим по m вершин и занумеруем их числами из E_m. Вход $a = \{i, j\}$ соединим с вершинами i, j второго уровня. Аналогично поступим с выходами и вершинами

третьего уровня. Вершины второго и третьего уровня соединим ребрами согласно матрице $\overline{A}$.

По построению, схема имеет $O(m^2)$ ребер. То, что схема реализует матрицу $\overline{B}$, вытекает из того, что путь, соединяющий вход a и выход b содержится в схеме в том и только том случае, когда подматрица матрицы $\overline{A}$, образованная строками b и столбцами a, не является сплошь нулевой. $\square$

Для вывода п. (i) теоремы в качестве матрицы A возьмем норм-матрицу из работы [152], которая при подходящем выборе параметров является Δ-редкой и имеет вес m^2/Δ, где $\Delta = 2^{O(\sqrt{\log m \log\log m})}$. Положим $C = \overline{B}$. Для доказательства п. (ii) выберем в качестве матрицы A 3-редкую матрицу Брауна [112] веса $\Theta(m^{5/3})$. Положим $C = \overline{B}$. Теорема 3.12 доказана.

Вопрос о получении нетривиальных оценок для соотношения $\mathsf{SUM}(\overline{A})/\mathsf{SUM}(A)$ остается открытым. Недавний результат [155] показывает, что сложность матриц с $O(n)$ нулями линейна, следовательно матрицы веса $O(n)$ не могут служить примерами.

3.4 Нижние оценки монотонной сложности функции T_n^2

В этом разделе изучаются монотонные схемы, т. е. схемы из функциональных элементов [49] над базисом $B_M = \{\vee, \wedge\}$, реализующие пороговую функцию T_n^2. Задачу получения нижних оценок мы сводим к анализу сложности систем булевых сумм при реализации схемами над базисом $\{\vee\}$. Такие схемы соответствуют линейным OR-схемам. Функционал сложности схем над базисом B будем обозначать через C_B.

Напомним историю вопроса. Нижняя оценка $C_{B_M}(T_n^2) \geq C_{B_2}(T_n^k) \geq 2n-3$, $2 \leq k \leq n-1$, доказана Б. М. Клоссом [24] в 1960-х гг. — в то время она была рекордной нижней оценкой сложности реализации индивидуальных булевых функций схемами в бинарном базисе, см. также [124, 209, 13]. В монотонном базисе оценка уточняется до $C_{B_M}(T_n^2) \geq 2n+\log_2 n-4$, исходя из того, что число дизъюнкторов в любой монотонной схеме, вычисляющей T_n^2, не меньше $2n-4$, а число конъюнкторов — не меньше $\log_2 n$. Эти оценки достигаются (оба результата, видимо, принадлежат Ф. Яо,

см. [103, 209, 124]), но, как показал П. Блониарц [103], не на одной схеме во всяком случае при $n = 2^k \geq 4$.

В 1970-х гг. Л. Адлеман (см. [103, 124, 209, 13]) получил верхнюю оценку $C_{B_M}(T_n^2) \leq 2n + O(\sqrt{n})$. При этом, согласно [103], ему удалось доказать и нижнюю оценку $2n + \Omega(\sqrt{n})$ при следующих ограничениях на структуру схем: (i) схемы содержат минимально возможное число $\lceil \log_2 n \rceil$ конъюнкторов, (ii) схемы синхронные (т. е. все пути от входов к выходам имеют одинаковую длину).

Автором [226] получено простое доказательство оценки $C_{B_M}(T_n^2) = 2n + \Theta(\sqrt{n})$, а в классе схем глубины альтернирования 3 (иначе говоря, $\bigvee\bigwedge\bigvee$-схем), т. е. схем с одним слоем конъюнкторов, установлена более точная оценка $2n + 2\sqrt{n} + O(n^{1/4}) - O(1)$. Эти результаты излагаются далее.

3.4.1 Общая нижняя оценка

Напомним, что при помощи схем над базисом $\{\vee\}$ можно вычислять булевы операторы Ax, линейные над полугруппой $(\{0, 1\}, \vee)$ (системы булевых сумм). Сложность оператора с булевой матрицей A при реализации такими схемами обозначим через $C^+(A)$ (и в пределах настоящего раздела будем называть ее просто сложностью матрицы A). Такие схемы подробно рассматриваются в [124, 209].

Принцип транспонирования для схем из функциональных элементов формулируется несколько иначе, чем для линейных схем, см. [50] или [9]. Если матрица A размера $m \times n$ не имеет нулевых строк и столбцов, то

$$C^+(A^T) = C^+(A) + m - n. \tag{3.31}$$

В основе предлагаемого доказательства нижней оценки сложности функции T_n^2 лежит

Лемма 3.11. *Пусть A — булева матрица размера $n \times m$, $m < n$, не имеющая нулевых столбцов, и строки которой попарно несравнимы. Тогда*

$$C^+(A) \geq n - m + \sqrt{2(n-m)}.$$

Доказательство. По условию, матрица A не имеет нулевых строк. Из условия несравнимости строк следует, что столбцы, которым принадлежат

единицы строк веса 1, тоже имеют вес 1. Пусть матрица содержит $q \leq m$ строк веса 1. Удалим из матрицы эти строки вместе с соответствующими столбцами и рассмотрим оставшуюся матрицу A' размера $(n-q) \times (m-q)$. Очевидно $C^+(A') = C^+(A)$.

В матрице A' нет строк веса 0 и 1. Поэтому каждая ее строка σ_i (точнее, булева сумма переменных в строке) вычисляется на выходе некоторого дизъюнктора схемы как $\sigma_i = b_{i,0} \vee b_{i,1}$. При этом никакой выход схемы не может использоваться для вычисления другого выхода (несравнимость строк). Обозначим $B = \{b_{i,k}\} = X \cup Y$, где X — множество строк из B веса 1, Y — остальные элементы B.

Для мощности множества B при условии $|B \vee B| \geq n-q$ можно указать тривиальную оценку $|B| \geq \sqrt{2(n-q)}$.

Мощность множества X оценим как $|X| \leq m - q$. Булевы суммы из множества Y вычисляются на внутренних дизъюнкторах схемы, которых, таким образом не меньше $|Y| \geq \sqrt{2(n-q)} - m + q$. Складывая эту оценку с числом выходных дизъюнкторов $n - q$, получаем утверждение леммы. □

Легко видеть, что множество импликант функции T_n^k — всевозможные мономы n переменных степени не меньше k. Мономы степени k составляют множество простых импликант[25].

Теорема 3.13. *Для $n \in \mathbb{N}$ выполнено $C_{B_M}(T_n^2) > 2n + \sqrt{n/8 - 1/4} - 4$.*

Доказательство. Пусть схема S, реализующая функцию $T_n^2(x_1, \ldots, x_n)$, содержит L конъюнкторов, вычисляющих функции $u_i = g_{i,0} \cdot g_{i,1}$, $i = 1, \ldots, L$ (здесь $g_{i,\alpha}$ — функции на входах конъюнкторов). Рассмотрим основной случай, когда $n \geq 3$ и $2L < n$.

Подсхема, состоящая из всех дизъюнкторов схемы S, вычисляет функции $g_{i,\alpha}$ как булевы суммы переменных $x_1, \ldots, x_n$ и функций $u_1, \ldots, u_L$. Подсхема также вычисляет еще одну сумму, соответствующую выходу схемы S, если он является дизъюнктором, но эту сумму мы не будем брать в расчет[26].

[25] Определения импликанты и простой импликанты см. в §2.2.1.

[26] В вырожденном случае сумма $g_{i,\alpha}$ может совпадать с переменной или вычисляться на выходе конъюнктора. Тогда мы все равно формально считаем, что для ее вычисления используется пустое множество дизъюнкторов.

Таким образом, частью схемы S является линейная схема, реализующая булев оператор с матрицей A_0 размера $2L \times (n+L)$. Рассмотрим подматрицу A_1 размера $2L \times n$ матрицы A_0, оставив столбцы, соответствующие только переменным x_i. Пусть матрица A_1 содержит l нулевых строк, что значит: l сумм $g_{i,\alpha}$ имеют только $u_1, \ldots, u_L$ в качестве слагаемых. Удалим из A_1 нулевые строки и получим матрицу A размера $(2L-l) \times n$. Очевидно, $C^+(A) = C^+(A_1) \leq C^+(A_0)$. Для оценки снизу величины $C^+(A)$ рассмотрим матрицу A^T. Ее строки соответствуют переменным $x_1, \ldots, x_n$, а столбцы — булевым суммам $g_{i,\alpha}$, использующим эти переменные. По построению, матрица не содержит нулевых столбцов.

Заметим, что строки матрицы A^T попарно несравнимы. Предположим противное. Пусть, скажем, для первых двух строк σ_1 и σ_2 выполнено $\sigma_1 \geq \sigma_2$. Это означает, что в любую сумму $g_{i,\alpha}$ переменная x_2 входит только вместе с x_1. Как следствие, никакая из функций u_i не содержит простую импликанту x_1x_2. А значит, функция, вычисляемая на выходе схемы S (она является булевой суммой каких-то из $u_1, \ldots, u_L$), также не содержит простую импликанту x_1x_2, поэтому отлична от T_n^2. Противоречие.

Теперь при помощи леммы 3.11 и соотношения (3.31) мы можем оценить сложность матрицы A как

$$C^+(A) \geq 2(n-2L+l) + \sqrt{2(n-2L+l)}.$$

Тогда сложность схемы S оценивается снизу как

$$C^+(A_0) + L \geq C^+(A) + L \geq$$
$$\geq 2n - 3L + 2l + \sqrt{2(n-2L+l)} \geq 2n - 3L + \sqrt{2(n-2L)}.$$

Поэтому, используя универсальную нижнюю оценку $2n-4$ для числа дизъюнкторов в любой схеме, вычисляющей T_n^2, получаем

$$C_{B_M}(T_n^2) \geq \min_L \max \left\{ 2n + L - 4,\ 2n - 3L + \sqrt{2(n-2L)} \right\}.$$

Легко проверить, что минимум в правой части неравенства достигается при $L = \sqrt{n/8 - 15/64} + 7/8$, откуда следует утверждение теоремы (при $2L < n$). В случае $2L \geq n$ просто воспользуемся оценкой $C_{B_M}(T_n^2) \geq 2n + L - 4$. $\square$

3.4.2 Нижняя оценка в классе схем глубины 3

Более аккуратные рассуждения позволяют уточнить оценку теоремы 3.13 до $C_{B_M}(T_n^2) \geq 2n + \sqrt{2n/3} - O(1)$. Но и такая оценка оставляет зазор величины $\Theta(\sqrt{n})$ до наилучшей известной верхней оценки $C_{B_M}(T_n^2) \leq 2n + 2\sqrt{n} + O(\sqrt[4]{n})$ (см. [103]). Однако при дополнительном естественном ограничении на вид схем этот зазор можно существенно сократить. Известные конструкции схем, доставляющие верхнюю оценку, содержат один слой конъюнкторов, т. е. схемы имеют глубину (альтернирования) 3, см. [103, 209, 124, 13].

Далее мы покажем, что в классе схем глубины 3 (иначе говоря, $\bigvee\bigwedge\bigvee$-схем) указанная верхняя оценка не может быть существенно улучшена, установив нижнюю оценку $C^{(3)}_{B_M}(n) \geq 2n+2\sqrt{n}-O(1)$. Через $C^{(3)}$ обозначен функционал сложности схем глубины 3.

Для полноты изложения напомним вывод верхней оценки. Пусть множество из n переменных занумеровано двумя индексами $x_{i,j}$, $i = 1, \ldots, p$, $j = 1, \ldots, s$, где $ps \geq n$. Обозначим $u_i = \bigvee x_{i,j}$ и $v_j = \bigvee x_{i,j}$. Тогда

$$T_n^2(\{x_{i,j}\}) = T_p^2(u_1, \ldots, u_p) \vee T_s^2(v_1, \ldots, v_s),$$

поэтому

$$C^{(3)}_{B_M}(T_n^2) \leq 2n - p - s + 1 + C^{(3)}_{B_M}(T_p^2) + C^{(3)}_{B_M}(T_s^2),$$

поскольку все суммы u_i и v_j вычисляются со сложностью $2n - p - s$. Достаточно выбрать $p = s \approx \sqrt{n}$.

Доказываемая далее нижняя оценка опирается на известную теорему В. Мантеля [160]: если граф на N вершинах не содержит треугольников (полных графов на трех вершинах), то в нем не более $N^2/4$ ребер. Несколько доказательств теоремы приведены в [144].

Лемма 3.12. *Пусть A — булева матрица размера $n \times 2m$, $n > 2m$, не имеющая нулевых столбцов, со следующими свойствами: при любом $i = 1, \ldots, m$ столбцы с номерами $2i - 1$ и $2i$ (нумерация с 1) не имеют двух единиц в одной и той же строке; булева сумма любых двух строк содержит две единицы в некоторой паре столбцов с номерами $2i - 1$ и*

$2i$. *Тогда*

$$C^+(A) \geq n + 2\sqrt{n - 2m + 3} - 5.$$

Доказательство. Легко проверить, что при $n \geq 3$ матрица содержит не более одной строки веса 1. Остальные не менее чем $n-1$ строк вычисляются на выходах элементов схемы. Как в доказательстве леммы 3.11, обозначим через $B = \{b_{i,\alpha} \mid i = 1, \ldots, n,\ \alpha \in \{0, 1\}\}$ множество строк, суммы которых дают все строки σ_i матрицы, $\sigma_i = b_{i,0} \vee b_{i,1}$, $|\sigma_i| > 1$.

Рассмотрим граф G, вершины которого без повторений помечены элементами множества B, а ребро между вершинами b и b' имеется в том и только том случае, когда $b \vee b'$ является строкой матрицы A.

Граф не содержит треугольников. Действительно, если вершины b_1, b_2, b_3 образуют треугольник, то матрица A содержит строки $b_1 \vee b_2$, $b_1 \vee b_3$ и $b_2 \vee b_3$. По условию леммы, сумма (любых двух из) этих строк $b_1 \vee b_2 \vee b_3$ имеет единицы в некоторых столбцах $2i$ и $2i + 1$, но тогда обе единицы присутствуют в какой-то из сумм $b_j \vee b_k$, $1 \leq j < k \leq 3$, значит, эта сумма не может быть строкой матрицы. Противоречие.

Обозначив через X множество строк из B веса 1, заметим, что никакая вершина графа не может иметь более двух общих ребер с вершинами из X. Действительно, если $e_1, e_2, e_3 \in X$, то $\sigma \vee e_1$, $\sigma \vee e_2$, $\sigma \vee e_3$ не могут быть тремя строками матрицы A (поскольку оба условия леммы не выполнимы для строк e_1, e_2, e_3).

Оценим мощность множества $Y = B \setminus X$. Граф G, по определению, содержит не менее $n - 1$ ребер. При этом подграф на вершинах из Y содержит не более $|Y|^2/4$ ребер (по теореме Мантеля), подграф на вершинах из X — не более $|X|$ ребер, множества X и Y соединяются не более чем $2|Y|$ ребрами (из-за ограничения на число общих ребер с вершинами из X). Поэтому $n \geq |Y|^2/4 + 2|Y| + |X| + 1$, откуда следует

$$|Y| \geq 2\sqrt{n - |X| + 3} - 4 \geq 2\sqrt{n - 2m + 3} - 4.$$

Число внутренних дизъюнкторов в схеме, вычисляющей матрицу A, не меньше чем $|Y|$, т. к. все строки из Y имеют вес не менее 2. □

Теорема 3.14. *Для $n \in \mathbb{N}$ выполнено $C^{(3)}_{B_M}(T^2_n) \geq 2n + 2\sqrt{n + 27} - 14$.*

Доказательство. Неизбыточная $\bigvee\bigwedge\bigvee$-схема S с L конъюнкторами имеет строго следующую структуру: дизъюнкторы первого слоя вычисляют $2L$ булевых сумм переменных, подаваемых на входы конъюнкторов; на втором слое расположены параллельно все L конъюнкторов; на третьем слое — дерево из $L-1$ дизъюнкторов, в котором суммируются результаты умножений.

Остается только оценить сложность матрицы A, вычисляемой дизъюнкторами первого слоя. Пусть входам каждого конъюнктора соответствуют соседние строки матрицы. Тогда при $n > 2L$ матрица A^T размера $n \times 2L$ удовлетворяет условиям леммы 3.12: первое условие запрещает функциям, вычисляемым конъюнкторами, иметь импликанты длины 1; второе условие обеспечивает вычисление всевозможных импликант степени 2.

Следовательно, из леммы 3.12 с учетом (3.31) выводим

$$C^{(3)}_{B_M}(T^2_n) \geq C^+(A^T) + n - 1 \geq 2n + 2\sqrt{n - 4L + 3} - 6$$

и далее

$$C^{(3)}_{B_M}(T^2_n) \geq \min_L \max \left\{ 2n + L - 4,\ 2n + 2\sqrt{n - 4L + 3} - 6 \right\}.$$

Минимум достигается при $L = 2\sqrt{n + 27} - 10$, откуда вытекает требуемое. При $n \leq 2L$, как легко убедиться, уже оценка $2n + L - 4$ выше доказываемой. $\square$

4 Параллельные префиксные схемы

4.1 Введение

Пусть $\circ$ — бинарная ассоциативная операция на некотором множестве $\mathbf{G}$. Множество функций

$$x_1 \circ \ldots \circ x_i, \qquad 1 \le i \le m, \tag{4.1}$$

называется системой *префиксов* (или *префиксных сумм*) упорядоченного набора переменных $x_1, \ldots, x_m$, принимающих значения в $\mathbf{G}$. Схемы из функциональных элементов [49] над базисом $\{\circ\}$, реализующие систему (1), называют *префиксными схемами*.

Префиксную схему можно также определить как линейную схему с ограничением 2 на число входящих в вершину ребер. При этом в функционале сложности удобнее подсчитывать число вершин схемы (не считая входов), а не ребер. Также следует иметь в виду, что операция $\circ$ может быть некоммутативной, поэтому фиксируется порядок ребер, входящих в каждую (функциональную) вершину.

Число m (входов схемы) далее будем называть *порядком*[27] схемы.

Префиксные схемы применяются во многих теоретических и прикладных задачах синтеза, например, в задаче реализации сумматора двоичных чисел, а также в задачах сортировки, решения линейных рекуррентных последовательностей. Об этих и некоторых других приложениях префиксных схем рассказывается в [102].

Сложность системы (4.1) очевидно равна $m - 1$, однако глубина минимальной схемы тоже равна $m - 1$. Еще в 1960-е годы в связи с потребностями ряда приложений возникла необходимость строить *параллельные* префиксные схемы, т. е. схемы глубины $O(\log m)$. Вопрос: какова может быть сложность таких схем.

Далее, если особо не оговаривается иное, мы будем рассматривать *универсальные* префиксные схемы, т. е. подходящие для вычислений в произвольном множестве с операцией $(\mathbf{G}, \circ)$ — это существенно для обоснования нижних оценок сложности.

[27] В некоторых работах число входов называется *шириной* схемы.

Пусть $L(m)$ означает сложность минимальной (универсальной) префиксной схемы порядка m и глубины $\lceil \log_2 m \rceil$.

Несколько простых конструкций префиксных схем, приводящих к оценке $L(m) = O(m \log m)$, было предложено в 1950–70 гг. (см., например, [200, 150]). В 1978 г. Р. Ладнер и М. Фишер [156] получили линейную верхнюю оценку сложности

$$L(m) \leq (4 - o(1))m.$$

В случае $m = 2^n$ была указана более точная оценка

$$L(2^n) \leq 4 \cdot 2^n - \Phi_{n+5} + 1 = 4 \cdot 2^n - O(\varphi^n),$$

где Φ_k — k-е число Фибоначчи[28], $\varphi = (1 + \sqrt{5})/2$.

Чуть позже Ф. Фич [126, 127][29] доказала следующие нижнюю и верхнюю оценки сложности:

$$\left(3\tfrac{1}{3} - o(1)\right) 2^n \leq L(2^n) \leq \left(3\tfrac{421}{792} - o(1)\right) 2^n.$$

Основной результат, излагаемый далее, состоит в доказательстве соотношений

$$L(2^n) = 3.5 \cdot 2^n - (8.5 + 3.5(n \bmod 2))2^{\lfloor n/2 \rfloor} + n + 5, \qquad L(m) \leq (3.5 - o(1))m,$$

и получен автором в [217, 235].

Незначительное ослабление условия минимальности глубины префиксной схемы позволяет существенно улучшить ее сложность, а также некоторые другие представляющие прикладной интерес характеристики, например, ветвление выходов элементов. Сложность, впрочем, может быть улучшена только до известного предела, определяемого соотношением

$$L + D \geq 2m - 2, \tag{4.2}$$

[28] Число Фибоначчи Φ_k является ближайшим к $\varphi^k/\sqrt{5}$ целым числом.

[29] Автор не имел возможности ознакомиться с работой [126], поэтому результаты Фич излагаются по [127].

справедливым для произвольной префиксной схемы порядка m, сложности L и глубины D[30]. Схемы, на которых достигается оценка (4.2), называются *оптимальными*. Синтез оптимальных префиксных схем с различными дополнительными характеристиками является популярным направлением в схемотехнике.

Несложно построить оптимальную схему глубины $2\lceil\log_2 m\rceil - 2$ (см., например, [61]). Минимально возможная глубина для оптимальных схем составляет $\log_\varphi m - O(1)$. Такие схемы построены в работе [214]. Более точно, в [214] показано, что оптимальные схемы порядка m и глубины d существуют для $m \le \Phi_{d+3} - 1$ и только для таких m.

Обозначим через $L(m,k)$ минимальную сложность префиксной схемы порядка m и глубины $\lceil\log_2 m\rceil + k$. В силу вышесказанного содержательным является случай $k \le \log_\varphi m - \log_2 m - O(1)$. В [156, 127] строились схемы, дополнительно удовлетворяющие соглашению реализации максимального префикса $x_1 \circ \ldots \circ x_m$ с минимально возможной глубиной $\lceil\log_2 m\rceil$. Сложность префиксных схем из этого класса обозначим через $L'(m,k)$. Очевидно, $L(m,k) \le L'(m,k)$.

В работе [156] доказаны оценки

$$L'(m,k) < (2 + 2^{1-k})m - 2, \qquad L'(2^n,k) \le (2 + 2^{1-k})2^n - \Phi_{n+5-k} - k + 1,$$

а в работе [127]:

$$\left(2 + \tfrac{1}{3}\cdot 2^{1-k} - o(1)\right)2^n \le L'(2^n,k) \le \left(2 + \tfrac{421}{792}\cdot 2^{1-k} - o(1)\right)2^n - k.$$

Указанные оценки уточнены автором в [217, 235]. Справедливо

$$L'(2^n,k) = (2 + 2^{-k})2^n - (5 + 2((n-k) \bmod 2))2^{\lfloor (n-k)/2\rfloor} - k + 2,$$
$$L'(m,k) \le (2 + 2^{-k} - o(1))m.$$

Вторая оценка выполнена при $1 \le k \le \lceil\log_2 m\rceil - 2$ и $m \to \infty$.

Сложность префискной схемы может быть также уменьшена при ослаблении требования универсальности. В качестве примера рассмотрим

[30] Это соотношение, по-видимому, впервые установлено Фич [127] в содержательном частном случае. Общая формулировка дана М. Шниром в [201], но при этом сопровождается излишне громоздким доказательством.

префиксные схемы «по модулю 2»: схемы над базисом $\{\oplus\}$, где $\oplus$ — ассоциативная бинарная операция, для которой справедливо тождество $x \oplus y \oplus y = x$ (такой операцией, в частности, является операция сложения по модулю 2)[31]. Фактически в этом случае задача состоит в построении экономных XOR-схем для полной треугольной матрицы. Для сложности вычислений по модулю 2 введем обозначения $L^{\oplus}(m)$ и $L'^{\oplus}(m,k)$, аналогичные введенным выше в случае универсальных схем.

В [217, 235] показано, что

$$L^{\oplus}(m) \le \left(3\tfrac{3}{11} - o(1)\right) m, \qquad L'^{\oplus}(m,k) \le \left(2 + \tfrac{3}{11} \cdot 4^{1-k} - o(1)\right) m,$$

где $1 \le k \le \lceil (\log_2 m)/2 \rceil - 1$ и $m \to \infty$.

Изложение следует работе [235]. В §4.2 вводятся понятия, предназначенные для описания структуры префиксной схемы. В §4.3 и §4.4 доказываются соответственно нижняя и верхняя оценки для $L(2^n)$. В §4.5 извлекаются следствия в отношении сложности префиксных схем различной глубины. В §4.6 устанавливается верхняя оценка для $L^{\oplus}(2^n)$ и аналогичные следствия. В §4.7 собраны некоторые сведения и замечания о параллельных префиксных схемах с ограничением на ветвление выходов элементов.

4.2 Предварительные понятия

Введем некоторые понятия (в основном, заимствованные из [127]), полезные для анализа структуры префиксной схемы.

Прежде заметим, что при вычислении префиксных сумм универсальными схемами (или формулами) с выходами схемы могут соединяться ориентированными путями только элементы, на выходах которых реализуются функции вида $x_i \circ x_{i+1} \circ \ldots \circ x_j$.

Действительно, если это не так, то найдется формула, реализующая некоторую префиксную сумму из (4.1), в которой есть пара переменных x_j, x_k, где $j < k$, такая, что вхождение символа x_k в формуле предшествует вхождению символа x_j. Покажем, что такая формула не может вычислять

[31] Кстати, префиксная схема над $(GF(2), \oplus)$ выполняет преобразование из кодировки Грея в обычную двоичную кодировку числа, подробнее см. в [27].

префиксную сумму в некоммутативной группе $(\mathbf{G}, \circ)$ с элементами бесконечного порядка (например, в группе симметрий окружности). Положим $x_j = a$, $x_k = b$ и $x_i = e$ при $i \neq j, k$, где $a, b, e \in \mathbf{G}$ и e — единица группы. Префиксная сумма на этом наборе аргументов принимает значение $a \circ b$, а формула — либо $b \circ a$, либо $a^{l_1} \circ b^{l_2} \circ a^{l_3} \circ \ldots$, где все $l_i \geq 0$ и $\sum_i l_i \geq 3$. В первом случае выберем a и b так, что $a \circ b \neq b \circ a$. Во втором случае одно из чисел $l_a = \sum l_{2i-1}$, $l_b = \sum l_{2i}$ больше 1. Считая, без ограничения общности, что $l_a > 1$, выберем в качестве a элемент бесконечного порядка и положим $b = e$. Тогда в силу $a^{l_a} \neq a$ значения префиксной суммы и формулы на указанном наборе будут различны.

Поэтому далее мы можем ограничить рассмотрение такими схемами, у которых на выходах всех элементов реализуются функции вида $x_i \circ x_{i+1} \circ \ldots \circ x_j$.

Если на выходе элемента v реализуется функция $x_i \circ \ldots \circ x_j$, то поставим ему в соответствие *метку* $\lambda(v) = [i; j]$. Каждому входу схемы x_i припишем метку $[i; i]$. Также введем обозначения для левого и правого концов метки: $l(v) = i$ и $r(v) = j$ соответственно, и обозначим $w(v) = r(v) - l(v) + 1$ — число слагаемых вычисляемой на выходе элемента v суммы, которое будем называть *шириной* элемента v.

Пусть в элемент v ведут ребра из элементов v' и v''. Тогда для одного из этих элементов (допустим, что для v') выполнено $l(v') = l(v)$, а для другого — $r(v'') = r(v)$. В таком случае элемент v' будем называть *левым входом*, а v'' — *правым входом* элемента v. Очевидно, $w(v) = w(v') + w(v'')$. Глубину, на которой расположен элемент v, обозначим через $d(v)$.

Классифицируем элементы префиксной схемы S. Подсхему, вычисляющую максимальную префиксную сумму $x_1 \circ \ldots \circ x_m$, назовем *каркасной*. Она является деревом, содержащим $m - 1$ функциональных элементов, из которых не более $D + 1$ являются выходами схемы S, где D — глубина рассматриваемой подсхемы. Элементы этой подсхемы, не совпадающие с выходами схемы S, будем называть *каркасными*. Те элементы схемы S, которые не являются ни выходами, ни каркасными, будем называть *избыточными*.

Заметим, что схемы без избыточных элементов, в которых глубина реа-

лизации максимального префикса совпадает с глубиной самой схемы, в точности образуют множество оптимальных схем, определенных во введении (это рассуждение легко превратить в доказательство соотношения (4.2) — при этом (4.2) остается верным, если вместо D подставить глубину реализации максимального префикса).

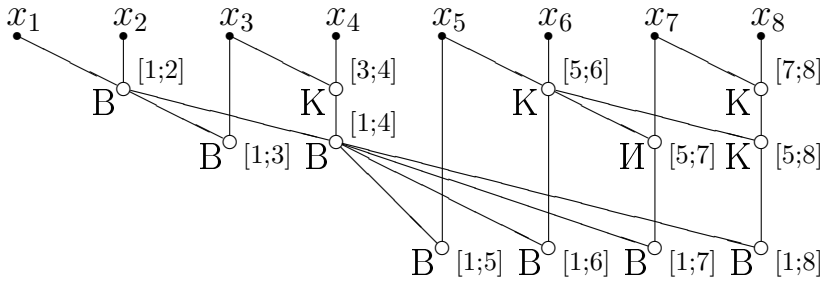


Рис. 8: Префиксная схема порядка 8

На рис. 8 изображена префиксная схема порядка 8 (ориентация ребер сверху вниз). Элементы-выходы, каркасные элементы и избыточный элемент обозначены символами «В», «К», «И» соответственно.

Каркасная подсхема префискной схемы порядка $m = 2^n$ и глубины n определяется однозначно. Элементы каркасной подсхемы имеют метки $[i2^k + 1; (i + 1)2^k]$, где $i = 0, \ldots, 2^{n-k} - 1$ и $k = 1, \ldots, n$.

4.3 Нижняя оценка

Первоначальное представление о структуре минимальной префиксной схемы дает следующая лемма. Обозначим через $\mathbf{S}_r(m, d)$ множество префиксных схем порядка m и глубины не выше d, у которых любой выход, зависящий от более чем r входов схемы, расположен на большей глубине, чем выход, в котором реализуется префикс $x_1 \circ \ldots \circ x_r$.

Лемма 4.1 (Фич [127]). *Пусть на схеме S достигается минимум сложности схем из $\mathbf{S}_r(m, d)$. Тогда выход u^* схемы S, реализующий функцию $x_1 \circ \ldots \circ x_r$, соединен ориентированными путями с каждым из выходов, зависящих более чем от r входов схемы.*

Доказательство. Если для некоторой схемы последнее условие не выполняется, то в ней имеются элементы, не являющиеся выходами и зависящие от обоих входов x_r и x_{r+1}. Обозначим множество таких элементов через M. Преобразуем схему следующим образом. Элемент $v \in M$, которому не предшествуют другие элементы из M, удалим из схемы, а начало ребер, исходящих из v, перенесем в правый вход v' элемента v. Заметим, что $l(v') = r + 1$, поскольку никакой из входов v не принадлежал множеству M. Будем повторять указанную процедуру до тех пор, пока в схеме не останется элементов из M.

Теперь если некоторый элемент u исходной схемы имел вход с меткой $[i; j]$, $1 < i \leq r < j$, то в новой схеме элемент u (при условии, что он не был удален) будет иметь вход с меткой $[r + 1; j]$.

Окончательно, ребра, соединявшие выходы u'' и u', для которых $r(u'') < r = r(u^*) < r(u')$, заменим ребрами, ведущими из u^* в u'. Легко проверить, что новая схема также принадлежит множеству $\mathbf{S}_r(m, d)$ и имеет меньшую сложность. Несколько подробнее см. в [127]. □

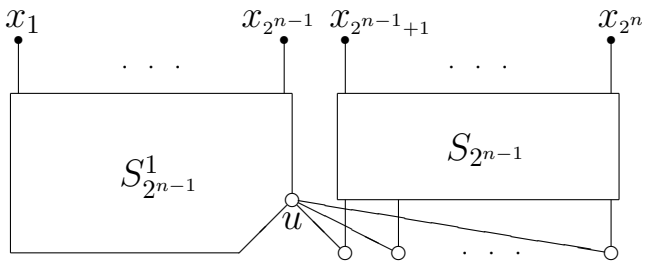


Рис. 9: Структура схемы S_{2^n}

Очевидно, любая минимальная схема S_{2^n} порядка 2^n и глубины n принадлежит множеству $\bigcap_{i=0}^{n-1} \mathbf{S}_{2^i}(2^n, n)$. В частности, $S_{2^n} \in \mathbf{S}_{2^{n-1}}(2^n, n)$. Тогда она содержит в качестве подсхемы (минимальную) схему $S_{2^{n-1}}$, реализующую систему префиксов 2^{n-1} переменных $x_{2^{n-1}+1}, \ldots, x_{2^n}$. При этом у любого выхода u' схемы S_{2^n}, для которого $w(u') > 2^{n-1}$, левым входом является выход u, реализующий сумму $x_1 \circ \ldots \circ x_{2^{n-1}}$, а правым входом — элемент с меткой $[2^{n-1} + 1; r(u')]$, который является выходом подсхемы $S_{2^{n-1}}$ (см. рис. 9).

Далее, как и в [127], будем оценивать снизу количество избыточных

элементов в схеме S_{2^n}.

Доказательство следующей леммы основано на простом наблюдении: если $d(v) = h$, то $w(v) \leq 2^h$. Наоборот, если $w(v) > 2^{h-1}$, то $d(v) \geq h$.

Лемма 4.2. *Пусть элемент v схемы S расположен на глубине h, и в схеме нет каркасных элементов с правым концом метки $r(v)$, расположенных на глубине большей чем k, где $k < h$. Пусть $w(v) > 2^h - 2^{h-s} + 2^k$. Тогда в схеме содержится s избыточных элементов $v_1, \ldots, v_s$ таких, что $r(v_i) = r(v)$, $w(v_i) \geq \big(w(v) \bmod 2^{h-i}\big) > 2^{h-i} - 2^{h-s} + 2^k$. При этом v_1 является входом для v, при любом i элемент v_i является входом для v_{i-1}, и $d(v_s) \leq h - s$.*

Доказательство. Лемма доказывается индукцией по s. При $s = 0$ доказывать нечего.

Пусть $s \geq 1$, и предположим, что лемма доказана для всех меньших значений s. Пусть элементы v' и v_1 являются соответственно левым и правым входами элемента v. Из соотношений $w(v) = w(v') + w(v_1)$ и $w(v') \leq 2^{h-1}$ следует

$$w(v_1) \geq w(v) - 2^{h-1} = \big(w(v) \bmod 2^{h-1}\big) > 2^{h-1} - 2^{h-s} + 2^k.$$

Поскольку $w(v_1) > 2^k$, то элемент v_1 не может быть каркасным, следовательно является избыточным. Очевидно, $d(v_1) \leq h - 1$. Кроме того, если $s > 1$, то для элемента v_1 выполнены условия леммы (где h надо заменить на $h-1$, а s — на $s-1$). Тогда, по индуктивному предположению, найдутся еще $s-1$ избыточных элементов $v_2, \ldots, v_s$, удовлетворяющих соотношению на $w(v_i)$:

$$w(v_i) \geq \Big(w(v_1) \bmod 2^{(h-1)-(i-1)}\Big) \geq \big(w(v) \bmod 2^{h-i}\big) > 2^{h-i} - 2^{h-s} + 2^k.$$

Каждый из элементов v_i является входом для v_{i-1}. При этом $d(v_s) \leq d(v_1) - (s-1) \leq h - s$. □

В частности, доказанная лемма обосновывает наличие избыточного элемента в схеме из рис. 8 при условии, что выход v с меткой $[1;7]$ должен быть расположен на глубине 3.

Далее под S подразумевается произвольная префиксная схема порядка 2^n и глубины n. Ключевой в доказательстве нижней оценки сложности является следующая лемма.

Лемма 4.3 (основная). *Пусть $k, N, R \in \mathbb{N}$, $N < 2^k$ и $R < 2^{n-2k-1}$, R не является степенью двойки. Тогда в схеме S содержится не менее N избыточных элементов с правыми концами меток из интервала*

$$J_{N,R,k} = [N2^{n-k-1} + R2^k,\ N2^{n-k-1} + (R+1)2^k - 1]. \tag{4.3}$$

При выводе нижней оценки в [127] фактически используется частный случай этого утверждения, в котором $N = 1$.

Для доказательства леммы 4.3 нам потребуется ввести некоторые дополнительные понятия. Но прежде поясним смысл выбора параметров.

Лемма 4.4. *В условиях леммы 4.3 для любого каркасного элемента e схемы S с правым концом метки $r(e) \in J_{N,R,k}$ выполнено $w(e) < R2^k$.*

Доказательство. Заметим, что $w(e) \leq 2^\nu$, где 2^ν — максимальная степень двойки, которая делит $r(e)$ (см. §4.2). Из $0 < r(e) - N2^{n-k-1} < 2^{n-k-1}$ следует, что $2^\nu \mid \left(r(e) - N2^{n-k-1}\right)$. Если R — не степень двойки, тогда и $r(e) - N2^{n-k-1} = R2^k + R_0$, где $0 \leq R_0 < 2^k$, не является степенью двойки. Значит, $2^\nu \leq (R2^k + R_0)/3 < R2^k$. Окончательно имеем $w(e) \leq 2^\nu < R2^k$. □

Запишем N в виде $(k+1)$-разрядного двоичного числа:

$$N = \underbrace{0\ldots0}_{p_1}\,\underbrace{1\ldots1}_{s_1}\,\underbrace{0\ldots0}_{p_2}\,\ldots\ldots\,\underbrace{1\ldots1}_{s_q}\,\underbrace{0\ldots0}_{p_{q+1}}. \tag{4.4}$$

Число N представляется разбитым на блоки из последовательно идущих нулей и единиц: p_1 нулей в старших разрядах, затем блок из s_1 единиц и т.д. По построению, $p_{q+1} \geq 0$, прочие параметры p_i и s_i положительны.

Следующая лемма описывает ситуацию, в которой мы будем применять лемму 4.2. Предварительно введем обозначения $P_t = \sum_{i=1}^{t-1}(p_i + s_i) + p_t$ и

$$N_t = N - \left(N \bmod 2^{k+1-P_t}\right) = N - \left(2^{k+1-P_t} - 2^{k+1-P_t-s_t}\right),$$

где $t = 1, \ldots, q+1$ (здесь и везде далее сумма по пустому множеству индексов полагается равной нулю).

Лемма 4.5. *Пусть $t \le q$ и элемент v схемы S с правым концом метки $r(v) \in J_{N,R,k}$ расположен на глубине $n - P_t$, при этом $l(v) \le N_t 2^{n-k-1} + 1$. Тогда в схеме имеются s_t отличных от v избыточных элементов $v_1, \ldots, v_{s_t}$ таких, что $r(v_i) = r(v)$, $d(v) > d(v_1) > \ldots > d(v_{s_t})$ и $l(v_{s_t}) \le N_{t+1} 2^{n-k-1} + 1$.*

Доказательство. Действительно,

$$\begin{aligned} w(v) \ge r(v) - N_t 2^{n-k-1} \ge (N - N_t) 2^{n-k-1} + R2^k = \\ = \left(N \bmod 2^{k+1-P_t}\right) 2^{n-k-1} + R2^k = \\ = \left(2^{k+1-P_t} - 2^{k+1-P_t-s_t}\right) 2^{n-k-1} + R2^k > 2^{n-P_t} - 2^{n-P_t-s_t} + 2^{k_0}, \end{aligned}$$

где k_0 — максимальная глубина каркасного элемента с правым концом метки из $J_{N,R,k}$; последний переход справедлив в силу леммы 4.4. В этом случае лемма 4.2 гарантирует существование искомых s_t избыточных элементов и выполнение для имеющего наименьшую глубину среди них элемента v_{s_t} соотношений $d(v_{s_t}) \le n - P_t - s_t$ и

$$\begin{aligned} w(v_{s_t}) \ge \left(w(v) \bmod 2^{n-P_t-s_t}\right) \ge \left((r(v) - N_t 2^{n-k-1}) \bmod 2^{n-P_t-s_t}\right) = \\ = r(v) - N2^{n-k-1} + \left((N - N_t)2^{n-k-1} \bmod 2^{n-P_t-s_t}\right) = \\ = r(v) - N2^{n-k-1} + \left(\left(N \bmod 2^{k+1-P_t}\right) 2^{n-k-1} \bmod 2^{n-P_t-s_t}\right) = \\ = r(v) - N2^{n-k-1} + \left(N \bmod 2^{k+1-P_t-s_t}\right) 2^{n-k-1} = \\ = r(v) - N2^{n-k-1} + \left(N \bmod 2^{k+1-P_{t+1}}\right) 2^{n-k-1} = r(v) - N_{t+1} 2^{n-k-1}. \end{aligned}$$

Следовательно, $l(v_{s_t}) \le N_{t+1} 2^{n-k-1} + 1$. □

4.3.1 Граф связей

Для схемы S и множества $J_{N,R,k}$ введем понятие *графа связей* $G_{N,R,k}(S)$.

Ориентированный граф $G_{N,R,k}(S)$ строится поэтапно следующим образом. Первоначально граф состоит из 2^k изолированных вершин z_i, где $i = 0, \ldots, 2^k - 1$. Каждой вершине z_i соответствует элемент-выход $v^{0,i}$ схемы S с меткой $[1; N2^{n-k-1} + R2^k + i]$. Все вершины графа $G_{N,R,k}(S)$ имеют тип 0.

Далее последовательно выполним шаги $0, \ldots, q$, где q определяется из (4.4). Шаг с номером t состоит в следующем.

Для каждой из вершин z_i типа t в порядке уменьшения индекса i выполним следующую процедуру.

1) Пусть $i \neq 0$, правый вход элемента $v^{t,i}$ имеет левый конец метки $N2^{n-k-1} + R2^k + i' + 1$ и является либо каркасным элементом, либо входом схемы. Обозначим левый вход элемента $v^{t,i}$ через $v_0^{t,i'}$. Если элемент $v^{t,i'}$ уже определен, то переобозначим через $v^{t,i'}$ тот из элементов $v^{t,i'}$, $v_0^{t,i'}$, который имеет меньшую глубину (при равенстве глубин элемент назначается произвольно). Иначе, положим $v^{t,i'} = v_0^{t,i'}$. Поставим элемент $v^{t,i'}$ в соответствие вершине $z_{i'}$. Если из вершины z_i уже выходило ребро, то удалим его. Соединим вершины z_i и $z_{i'}$ ребром, ориентированным в направлении $z_{i'}$, и назначим вершине $z_{i'}$ тип t.

2) (Это условие не проверяется на последнем шаге q.) Если $d(v^{t,i}) = n - P_{t+1}$, то изменим тип вершины z_i на $t+1$. Поставим ей в соответствие элемент $v^{t+1,i}$ схемы S, который в лемме 4.2 при условии $v = v^{t,i}$ и $s = s_{t+1}$ соответствует элементу v_s.

3) В противном случае ничего не делаем.

Заметим, что условия 1) и 2) не могут выполняться одновременно, т. к. в случае $d(v^{t,i}) = n - P_{t+1}$ правый вход элемента $v^{t,i}$ является избыточным элементом.

По построению, в графе связей все компоненты связности являются корневыми деревьями. *Корневое дерево*, в данном случае, — это дерево с ориентацией ребер, в котором есть вершина, не имеющая исходящих ребер (корень), и любая из вершин соединена с корнем ориентированным путем. При этом тип вершин графа связей не убывает при движении по ребрам.

Структура графа $G_{N,R,k}(S)$ позволяет оценить снизу количество избыточных элементов в схеме S с правыми концами меток из $J_{N,R,k}$. Сформулируем несколько предварительных наблюдений.

Лемма 4.6. *Справедливы следующие утверждения:*

а) Все определенные в результате построения графа связей элементы $v^{t,i}$ различны.

б) При $t > 0$ элементы $v^{t,i}$ являются избыточными.

в) $d(v^{t,i}) \le n - P_{t+1} + p_{t+1}$. *Если* $t < q$, *то* $d(v^{t,i}) \ge n - P_{t+1}$.

г) Для любого $t < q$ *в графе связей не может содержаться ориентированная цепочка, соединяющая* $p_{t+1} + 1$ *вершин типа* t.

д) Если при $i > 0$ *вершина* z_i *типа* t *является корневой, то правый вход элемента* $v^{t,i}$ *является избыточным.*

Доказательство. Справедливость п. *а)* устанавливается тем, что элементы, различающиеся в первом индексе, имеют разную глубину (это следует из п. *в)*, а различающиеся во втором — разные правые концы меток.

Утверждение п. *б)* вытекает из того, что при $t > 0$ элемент $v^{t,i}$ не является выходом, кроме того, он не является каркасным согласно лемме 4.4. Действительно, $l(v^{t,i}) \le N2^{n-k-1} + 1$ по лемме 4.5, следовательно $w(v^{t,i}) \ge R2^k$.

Соотношения п. *в)* при $t = 0$ выполнены, поскольку элементы $v^{0,i}$ являются выходами схемы S. Для произвольного элемента $v^{t,i}$ верхнюю границу глубины $n - P_t - s_t = n - P_{t+1} + p_{t+1}$ доставляет условие 1) или 2), которым данный элемент определяется. Нижнюю границу дает оценка ширины элемента $v^{t,i}$. Из леммы 4.5 следует $l(v^{t,i}) \le N_{t+1}2^{n-k-1} + 1$, откуда при $t < q$ вытекает

$$w(v^{t,i}) \ge 2^{n-P_{t+1}} - 2^{n-P_{t+1}-s_{t+1}} + R2^k > 2^{n-P_{t+1}-1}.$$

Утверждение п. *г)* является простым следствием п. *в)* и условий 1), 2).

Утверждение п. *д)* следует из того, что условие 1) для вершины z_i не выполняется на шаге t. □

Перейдем к оценке числа избыточных элементов в рассматриваемом фрагменте схемы S. Исходя из структуры графа связей, включающей соответствие типов и вершин, оценим число избыточных элементов схемы с правым концом метки $N2^{n-k-1} + R2^k + i$, где $i = 0, \ldots, 2^k - 1$.

Пусть вершина z_i имеет тип t. Пусть в нее входит ребро, исходящее из вершины типа $t' > 0$. Тогда после построения графа связей определен избыточный элемент $v^{t',i}$. Если же вершине z_i предшествует цепочка из $p_{t'}$ вершин типа $t'-1$, то также определен элемент $v^{t'-1,i}$, причем $d(v^{t'-1,i}) = n - P_{t'}$ согласно п. *в)* леммы 4.6, следовательно, согласно условию 2), найдутся избыточные элементы $v^{t'-1,i}$, $v^{t',i}$, а также еще $s_{t'} - 1$ расположенных (с точки

зрения глубины) между ними избыточных элементов, никакой из которых, как следует, например, из п. *в*) леммы 4.6, не совпадает с каким-либо элементом $v^{t'',i}$. Если вершина z_i, $i > 0$, является корневой, то можно указать еще один избыточный элемент — правый вход элемента $v^{t,i}$.

Поскольку вершина z_0 всегда является корневой, количество учтенных таким образом избыточных элементов можно записать в виде

$$I_{N,R,k}(S) = \mu(G_{N,R,k}(S)) - 1 + \sum_{z \in G_{N,R,k}(S)} \sum_{t=1}^{q} c(z,t), \tag{4.5}$$

где $\mu(G)$ обозначает число компонент связности графа G, а функция $c(z,t)$ определяется при $t \geq 1$ как

$$c(z,t) = \begin{cases} s_t, & \text{вершине } z \text{ предшествует цепочка из } p_t \text{ вершин типа } t-1; \\ 1, & \text{иначе, если вершина } z \text{ имеет тип } t \text{ или в вершину } z \\ & \text{входит ребро, исходящее из вершины типа } t; \\ 0, & \text{иначе.} \end{cases}$$

Величина $I_{N,R,k}(S)$ обладает тем недостатком, что структура графа связей недостаточна для ее определения, а требуется также информация о типах вершин. Наша следующая цель заключается в том, чтобы перейти от $I_{N,R,k}(S)$ к более универсальной характеристике, которая бы доставляла оценку числа избыточных элементов, опираясь только на топологию графа связей.

4.3.2 Стоимость графа

Обозначим через Δ множество графов, все компоненты связности которых являются корневыми деревьями (с ориентацией ребер в направлении корня). Под глубиной вершины z графа $G \in \Delta$ будем понимать максимальную длину (количество ребер) ориентированного пути, ведущего от листа графа к данной вершине, обозначим ее через $d_G(z)$. Под листом, как обычно, понимается вершина, в которую не входят ребра — глубина листа полагается равной нулю. Для обозначения того, что граф G' является подграфом графа G, будем использовать запись $G' \subset G$.

Для графа $G \in \Delta$ введем понятие $(p_1, s_1, \ldots, p_q, s_q)$-*стоимости*, где $p_i, s_i \in \mathbb{N}$ для всех $i = 1, \ldots, q$. Предварительно для краткости обозначений определим целочисленные интервалы:

$$M_t(p_1, \ldots, p_q) = \left[\sum_{i=1}^{t} p_i,\ \sum_{i=1}^{t+1} p_i - 1 \right],\ 0 \le t < q,$$
$$M_q(p_1, \ldots, p_q) = \left[\sum_{i=1}^{q} p_i,\ +\infty \right).$$

Определим функцию $(p_1, s_1, \ldots, p_q, s_q)$-стоимости вершины z графа G как

$$C^G_{p_1,s_1,\ldots,p_q,s_q}(z) = \sum_{t=1}^{q} c^G_{p_1,s_1,\ldots,p_q,s_q}(z,t),$$

где

$$c^G_{p_1,s_1,\ldots,p_q,s_q}(z,t) = \begin{cases} s_t, & \text{в вершину } z \text{ ведет ребро из вершины } z', \text{ такой,} \\ & \text{что } d_G(z') = \sum_{i=1}^{t} p_i - 1; \\ 1, & \text{иначе, если в вершину } z \text{ ведет ребро из} \\ & \text{вершины } z', \text{ такой, что } d_G(z') \in M_t(p_1, \ldots, p_q); \\ 0, & \text{иначе.} \end{cases}$$

Тогда $(p_1, s_1, \ldots, p_q, s_q)$-стоимость графа G определяется как

$$C_{p_1,s_1,\ldots,p_q,s_q}(G) = \mu(G) - 1 + \sum_{z \in G} C^G_{p_1,s_1,\ldots,p_q,s_q}(z).$$

Формально говоря, стоимость определена и при $q = 0$ как $C^G(z) = 0$ и $C(G) = \mu(G) - 1$.

Прежде чем разъяснить связь стоимости со введенной раньше характеристикой $I_{N,R,k}(S)$, докажем простое соотношение между типом и глубиной вершины графа связей.

Лемма 4.7. *Пусть* $d_{G_{N,R,k}(S)}(z) \ge \sum_{i=1}^{t} p_i$, *где параметры* N, R, k *определены условиями леммы* 4.3, *а* p_i *определяются из* (4.4). *Тогда тип вершины* z *не меньше, чем* t.

Доказательство. Рассмотрим ориентированную цепочку, определяющую глубину вершины z: по условию, она соединяет не менее $\sum_{i=1}^{t} p_i+1$ вершин, в том числе, согласно п. *в*) леммы 4.6, не более p_{i+1} вершин каждого типа i. Отсюда следует, что тип вершины z не меньше, чем t. □

Лемма 4.8. *Пусть параметры N, R, k определены условиями леммы* 4.3, *а p_i, s_i определяются из* (4.4). *Пусть любая вершина графа $G_{N,R,k}(S)$ имеет минимальный тип, удовлетворяющий лемме* 4.7. *Тогда $I_{N,R,k}(S) = C_{p_1,s_1,\ldots,p_q,s_q}(G_{N,R,k}(S))$.*

Доказательство. Утверждение леммы следует из совпадения значений $c(z,t)$ в формуле (4.5) и $c^G_{p_1,s_1,\ldots,p_q,s_q}(z,t)$. □

Из доказанной леммы следует, что стоимость графа связей $G_{N,R,k}(S)$ совпадает с $I_{N,R,k}(S)$ в случае, когда его вершины имеют минимально возможный (с точки зрения леммы 4.7) тип. В общем случае величина $C_{p_1,s_1,\ldots,p_q,s_q}(G_{N,R,k}(S))$ не может служить нижней оценкой для $I_{N,R,k}(S)$. Однако справедливо более слабое утверждение, позволяющее все же использовать стоимость для оценки снизу величины $I_{N,R,k}(S)$.

Естественным образом распространим понятие и обозначение стоимости графа на непустое множество графов $\Gamma \subset \Delta$:

$$C_{p_1,s_1,\ldots,p_q,s_q}(\Gamma) = \min_{G\in\Gamma} C_{p_1,s_1,\ldots,p_q,s_q}(G).$$

Для ориентированного графа T введем обозначение

$$\delta(T) = \{G \mid G \subset T, G \text{ содержит все вершины } T\} \cap \Delta.$$

Справедлива

Лемма 4.9. *Пусть параметры N, R, k определены условиями леммы* 4.3, *а p_i, s_i определяются из* (4.4). *Тогда*

$$I_{N,R,k}(S) \geq C_{p_1,s_1,\ldots,p_q,s_q}(\delta(G_{N,R,k}(S))).$$

Доказательство. Отталкиваясь от графа связей $G_{N,R,k}(S)$, опишем процедуру, в процессе которой удаляются некоторые ребра, приводящую в результате к графу, стоимость которого не превосходит $I_{N,R,k}(S)$.

Граф в процессе преобразований будем обозначать через G'. Подразумевая наличие соответствия типов и вершин, определим для графа G' величины $c'(z,t)$ и $I(G')$, аналогичные величинам $c(z,t)$ и $I_{N,R,k}(S)$, определенным для графа связей. Перед началом преобразований $G' = G_{N,R,k}(S)$ и $I(G') = I_{N,R,k}(S)$.

Далее последовательно для каждой из вершин z_i в порядке уменьшения индекса i выполним следующее действие. Если тип t вершины z_i в графе G' не является минимально возможным (исходя из соотношения леммы 4.7, применяемой к графу G' вместо $G_{N,R,k}(S)$), то изменим тип вершины на минимальный и удалим выходящее из нее ребро, если такое имеется.

Заметим, что перед этой операцией коэффициент $c'(z_i,t)$ положителен, а после нее он обнуляется. Остальные коэффициенты $c'(z,t')$, $z \in G'$, $t' = 1,\ldots,q$, не увеличиваются (измениться, и только в меньшую сторону, могут лишь коэффициент $c'(z',t)$ для вершины z', в которую ведет ребро из z_i, и коэффициент $c'(z'',t+1)$ для некоторой вершины z'' в цепочке, исходящей из z_i). Значит, на этом шаге величина $I(G')$ не увеличивается за счет того, что $c'(z_i,t)$ строго уменьшается, а число $\mu(G')$ компонент связности увеличивается не более чем на 1.

Порядок перебора вершин обеспечивает то, что все вершины с большими чем i индексами имеют минимальный тип. В результате получается граф $G \in \delta(G_{N,R,k}(S))$, для которого выполнены условия леммы 4.8 (с точностью до переобозначений). Следовательно, $I(G) = C_{p_1,s_1,\ldots,p_q,s_q}(G)$, откуда в силу $I(G) \le I_{N,R,k}(S)$ вытекает утверждение леммы. □

Доказанная лемма позволяет ограничиться изучением универсальной, т. е. применимой к любому графу из Δ, и не зависящей от исходной схемы S характеристики стоимости.

Стратегия дальнейших рассуждений состоит в том, чтобы показать, что наименьшую стоимость среди допустимых графов имеют графы определенного вида, а именно подграфы гиперпары. Вычисление минимального значения стоимости для таких графов приведет к нижней оценке стоимости графа связей $G_{N,R,k}(S)$ и, как следствие, к сформулированной в лемме 4.3 нижней оценке числа избыточных элементов в рассматриваемом фрагменте схемы S.

4.3.3 Множество допустимых графов. Гиперпары

Определим множество допустимых графов, т. е. множество, в котором найдется граф, изоморфный любому возможному графу связей. Для этого определим граф T_k с вершинами $z_0, \ldots, z_{2^k-1}$ таким образом, что ребро, соединяющее вершины z_i и $z_{i'}$, и ориентированное в направлении $z_{i'}$, содержится в графе тогда и только тогда, когда $i - i' = 2^t$ и $2^t \mid i$ для некоторого $t \geq 0$.

Лемма 4.10. $G_{N,R,k}(S) \in \delta(T_k)$.

Доказательство. По определению, граф T_k включает ребро, ведущее из z_i в $z_{i'}$, в том и только том случае, когда элемент схемы S с меткой $[N2^{n-k-1} + R2^k + i' + 1; N2^{n-k-1} + R2^k + i]$ является либо каркасным, либо входом схемы. Следовательно, граф T_k содержит все ребра, которые могут присутствовать в графе $G_{N,R,k}(S)$ (в данном случае, независимо от N и R). Значит, $G_{N,R,k}(S) \subset T_k$. □

Лемма показывает, что в качестве множества допустимых графов может быть выбрано множество $\delta(T_k)$. Из нее и леммы 4.9 немедленно вытекает

Следствие 4.1. *Пусть параметры N, R, k определены условиями леммы* 4.3*, а p_i, s_i определяются из* (4.4)*. Тогда*

$$I_{N,R,k}(S) \geq C_{p_1,s_1,\ldots,p_q,s_q}(\delta(T_k)).$$

Для удобства дальнейших рассуждений дадим еще один, рекурсивный способ построения графа T_k, попутно определяя множество *отмеченных* вершин, *правильных* и *неправильных* ребер. Граф T_0 состоит из единственной вершины, она же является отмеченной. Граф T_k получается из двух графов T_{k-1} следующим образом. Добавляются ребра, ведущие из корневой вершины первого из графов ко всем отмеченным вершинам второго графа. (Под корневой вершиной графа T_k понимается вершина, из которой не выходят ребра.) Отмеченными вершинами графа T_k считаются отмеченные вершины первого графа, а также корневая вершина второго графа, она

же корневая в графе T_k. Множество правильных ребер графа T_k образуют правильные ребра обоих графов T_{k-1} и ребро, соединяющее корневые вершины этих графов. Остальные ребра считаются неправильными.

Простейшие графы из семейства $\{T_k\}$ изображены на рис. 10: ориентация ребер слева направо, для вертикальных ребер — снизу вверх, выделены отмеченные вершины и правильные ребра.

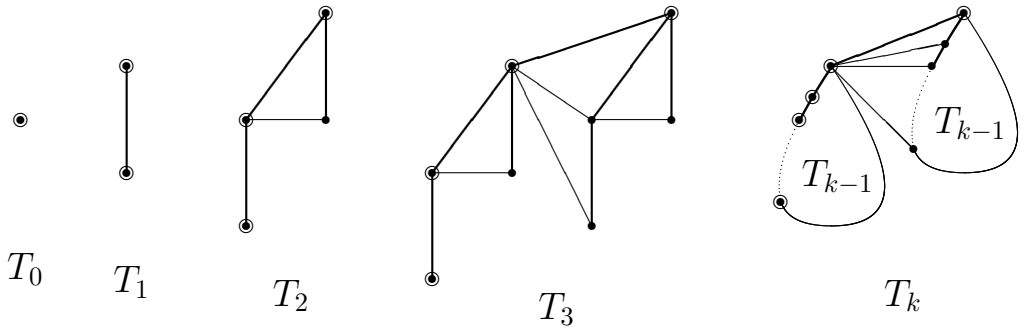


Рис. 10: Графы T_k

Соответствие между двумя определениями можно усмотреть путем ввода нумерации вершин во втором определении. Пусть вершина графа T_0 не имеет номера. При построении графа T_k из двух графов T_{k-1} к номерам вершин первого графа будем приписывать слева единицу, а к номерам вершин второго — ноль. Тогда вершине z_i из первого определения будет соответствовать вершина с номером i в двоичной записи.

Несложно проверить, что в каждую вершину графа T_k входит не более одного неправильного ребра, более точно, в отмеченные вершины входят только правильные ребра, а в остальные — помимо правильных ровно по одному неправильному.

Определим еще одно важное семейство графов — гиперпары[32].

Гиперпара H_k — это корневое дерево, которое определяется рекурсивно следующим образом. Гиперпара H_0 состоит из единственной вершины, она же является корнем. Далее при $k > 0$ гиперпара H_k образуется из двух гиперпар H_{k-1} путем соединения их корневых вершин ребром. Простейшие гиперпары изображены на рис. 11 (ориентация ребер снизу вверх).

[32]Понятие гиперпары позаимствовано из работы [192].

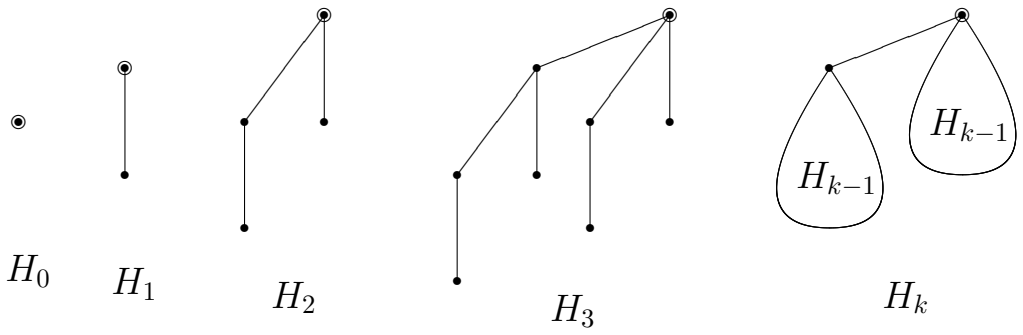


Рис. 11: Гиперпары

Легко видеть, что гиперпара H_k получается из графа T_k удалением всех неправильных ребер.

Пусть в графе $H \in \Delta$ одна из корневых вершин отмечена как *главная корневая вершина*. Назовем *композицией* графов $T \in \Delta$ и H граф $T \circ H$, получаемый построением графа T на главных корневых вершинах графов, изоморфных графу H. Заметим, что $T \circ H \in \Delta$.

В дальнейшем, если выбор главной корневой вершины графа H понятен из контекста, то эта вершина будет именоваться просто корневой вершиной графа H. В частности, главной в корневом дереве является единственная корневая вершина. Если граф H' выбирается из множества $\delta(H)$, $H \in \Delta$, то главной корневой вершиной графа H' считаем главную корневую вершину графа H.

Отметим следующие простые свойства гиперпар.

1) Композиция двух гиперпар является гиперпарой: гиперпара $H_{k_1+k_2}$ изоморфна графу $H_{k_1} \circ H_{k_2}$.

2) Для любой вершины z гиперпары H_k максимальный связный подграф с корнем в данной вершине является гиперпарой H_d, где $d = d_{H_k}(z)$.

3) Глубина корневой вершины гиперпары H_k равна k, в нее входят k ребер.

В графе T_k (равно как и в H_k) любая вершина является корневой для некоторого подграфа T_i (или H_i). Определим *порядок* вершины z в графе T_k (H_k) как наибольший из индексов i, таких, что T_k (H_k) содержит подграф, изоморфный графу T_i (H_i), с корневой вершиной z (далее по тексту в подобных случаях вместо «граф, изоморфный T_i» будет употребляться

просто «граф T_i»). Как видно из определения графа T_k, альтернативным образом можно определить порядок вершины как число входящих в нее правильных ребер. Как было замечено выше, порядок вершины гиперпары равен $d_{H_k}(z)$.

По построению, из вершины порядка i в графе T_k правильное ребро ведет к вершине большего порядка, а неправильные ребра (в случае $i > 0$) — к вершинам порядка $0, \ldots, i-1$, причем эти вершины соединяются цепочкой правильных ребер.

Ниже, говоря об изоморфизме на множестве графов $\{T_k\}$ и их подграфов, всякий раз мы будем иметь в виду не только топологическое совпадение, но также совпадение порядков соответствующих вершин, исключая корневые. В частности, нам удобно, чтобы выражение «подграф T_i с корневой вершиной v в графе T_k» однозначно указывало на граф из описанной выше рекурсивной процедуры построения T_k (только у него порядки внутренних вершин такие, как в определении графа T_i).

Далее мы перейдем к выяснению значений $C_{p_1,s_1,\ldots,p_q,s_q}(\delta(H_k))$ и затем $C_{p_1,s_1,\ldots,p_q,s_q}(\delta(T_k))$. В следующем разделе доказываются технические леммы 4.12, 4.13, 4.14, предоставляющие инструментарий для всех дальнейших построений.

4.3.4 Стоимость множества подграфов композиции корневых деревьев

Будем говорить, что ребро ρ' графа $G \in \Delta$ *зависит* от ребра ρ, если при удалении ребра ρ глубина вершины, из которой выходит ребро ρ', уменьшается.

Лемма 4.11. *Справедливы следующие утверждения:*

а) Множество зависящих от ρ ребер графа $G \in \Delta$ образует ориентированную цепочку (возможно, пустую), исходящую из вершины, в которую ведет ребро ρ.

Пусть эта цепочка состоит из j ребер $\rho_1, \ldots, \rho_j$, соединяющих последовательно вершины $z_0, \ldots, z_j$. Пусть z — вершина, из которой исходит ребро ρ, а G' — граф, получаемый из G удалением ребра ρ.

б) Для $j > 0$ при $i < j$ справедливо $d_G(z_i) = d_G(z) + i + 1$.

в) При любом $i = 1, \ldots, j$ множество зависящих от ρ_i ребер графа G совпадает с $\rho_{i+1}, \ldots, \rho_j$.

г) Если $z' \neq z_i$ для любого $i = 0, \ldots, j$, то $C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z') = C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z')$.

Доказательство. Утверждение п. *а)* следует из простого анализа множества вершин, глубина которых может измениться при удалении ребра ρ.

Утверждение п. *б)* следует из того, что в графе G (единственная) длиннейшая ориентированная цепочка от листа к вершине z_i проходит через z.

Для проверки п. *в)* обозначим через G'' граф, получаемый удалением из G ребра ρ_i. Для любого $i' \geq i$ справедливо $d_{G''}(z_{i'}) \leq d_{G'}(z_{i'})$, поскольку множество листьев, с которыми соединена ориентированными путями вершина $z_{i'}$ в графе G'', содержится в аналогично определяемом множестве графа G'. Следовательно, при $i \leq i' < j$ имеем $d_{G''}(z_{i'}) < d_G(z_{i'})$.

Для вершины z_j при этом есть две возможности. Либо она является корневой, либо $d_{G'}(z_j) = d_G(z_j)$. В первом случае доказывать нечего, во втором заметим, что в графе G найдется ориентированная цепочка длины $d_G(z_j)$, ведущая от листа к вершине z_j, не содержащая ребро ρ_j и, как следствие, никакое из ребер ρ_i. Поэтому $d_{G''}(z_j) = d_G(z_j)$.

Утверждение п. *г)* следует из того, что глубина вершин, из которых ведут ребра в вершину z', не изменяется при удалении ребра ρ. □

Лемма 4.12. *Произвольное ребро ρ можно удалить из графа $G \in \Delta$, возможно, удалив при этом также несколько зависящих от него ребер, так, что стоимость графа увеличится не более чем на* 1.

Доказательство. Воспользуемся индукцией по числу j зависящих от ρ ребер.

При $j = 0$ стоимость вершин графа при удалении ребра не увеличивается, а увеличивается на 1 только число компонент связности, следовательно утверждение леммы в этом случае справедливо.

Докажем индуктивный переход от $j - 1$ к j. Воспользуемся обозначениями леммы 4.11.

Ясно, что $C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_0) \leq C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_0)$. Рассмотрим два случая.

а) Пусть $C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_0) < C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_0)$. Тогда удалим из G ребро ρ_1, а также, возможно, еще несколько ребер из $\{\rho_i\}$, увеличив стоимость не более, чем на 1 (это можно сделать по индуктивному предположению — от ребра ρ_1 в графе G зависят $j-1$ ребер). Потом удалим ребро ρ — при этом стоимость не увеличится, т. к. увеличение числа компонент связности графа компенсируется уменьшением стоимости вершины z_0.

б) Пусть $C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_0) = C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_0)$. Пусть $d_G(z) \in M_t(p_1,\ldots,p_q)$. Всегда верно $c^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_0,t') = c^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_0,t')$ при $t' < t$ и $t' > t+1$, т. к. наличие ребра ρ несущественно для указанных коэффициентов.

Если предположить $d_G(z) = \sum_{i=1}^{t+1} p_i - 1$, то, учитывая $d_{G'}(z_0) < d_G(z_0)$, имеем $c^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_0,t) \le c^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_0,t)$ и $c^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_0,t+1) = 0 < s_{t+1} = c^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_0,t+1)$. Это противоречит равенству стоимости вершины z_0 в графах G и G'. Значит, $d_G(z) \ne \sum_{i=1}^{t+1} p_i - 1$ и, как следствие, $c^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_0,t+1) = c^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_0,t+1) = 0$. Тогда $c^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_0,t) = c^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_0,t) \ne 0$. Поэтому заключаем, что $d_{G'}(z_0) \in M_t(p_1,\ldots,p_q)$.

В силу соотношений

$$d_{G'}(z_0) + i - 1 \le d_{G'}(z_{i-1}) < d_G(z_{i-1}) = d_G(z_0) + i - 1$$

равенства $c^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_i,t) = c^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_i,t)$ и

$$c^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_i,t+1) = c^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_i,t+1) = 0,$$

а вместе с ними $C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_i) = C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_i)$ сохраняются при всех $i \le j'$, где либо $j' = j$, либо $d_G(z_{j'}) = \sum_{i=1}^{t+1} p_i - 1$.

б.1) В случае $j' = j$ при удалении ребра ρ стоимость всех вершин графа не изменяется, а стоимость самого графа увеличивается на 1 за счет увеличения числа компонент связности, поэтому утверждение леммы выполнено в этом случае.

б.2) Иначе, если $d_G(z_{j'}) = \sum_{i=1}^{t+1} p_i - 1$, то

$$c^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_{j'+1},t+1) \le c^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_{j'+1},t+1) = s_{t+1}.$$

При этом равенство возможно лишь в случае $j'+1 = j$. Но тогда $C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_j) = C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_j)$, и можно применить рассуждение п. б.1).

Если $j' + 1 < j$, то $c^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_{j'+1}, t+1) = 0$, откуда вытекает $C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_{j'+1}) < C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_{j'+1})$. Аналогично, как и в случае а), пользуясь индуктивным предположением, удалим из G ребро $\rho_{j'+2}$, а также, возможно, еще несколько ребер из $\{\rho_i\}$ с увеличением стоимости не более, чем на 1. Потом удалим ребро ρ, в результате чего стоимость не увеличится, благодаря уменьшению стоимости вершины $z_{j'+1}$. Случай б) полностью рассмотрен. □

Лемма 4.13. *Пусть в графе $G \in \Delta$ подграф H с корневой вершиной z таков, что в графе G нет ребер, соединяющих вершины из $H \setminus \{z\}$ и $G \setminus H$. Пусть $H' \in \Delta$, и граф G' получается из G заменой подграфа H графом H' (при которой корневая вершина графа H' совмещается с вершиной z). Тогда, если выполнены условия $d_H(z) \geq d_{H'}(z)$ и $d_G(z)$, $d_{G'}(z) \in M_t(p_1, \ldots, p_q)$ для некоторого t, а также*

$$\left(C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z) - C^{H}_{p_1,s_1,\ldots,p_q,s_q}(z)\right) + C_{p_1,s_1,\ldots,p_q,s_q}(H) = \left(C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z) - C^{H'}_{p_1,s_1,\ldots,p_q,s_q}(z)\right) + C_{p_1,s_1,\ldots,p_q,s_q}(H'),$$

то существует граф G'', получаемый из графа G' удалением некоторых ребер в ориентированной цепочке, исходящей из вершины z, такой, что

$$C_{p_1,s_1,\ldots,p_q,s_q}(G'') \leq C_{p_1,s_1,\ldots,p_q,s_q}(G).$$

Доказательство. Условие расположения подграфа H внутри графа G (подграф H соединяется с остальным графом только через вершину z) означает, что изменение стоимости графа G, возникающее при замене H на H', складывается из разности стоимостей этих подграфов, изменения стоимости вершины z (при этом надо иметь в виду, что частично ее стоимость учитывается в стоимости подграфов), а также из изменения стоимости вершин в цепочке, исходящей из z.

Нетрудно видеть, что изменение стоимости подграфа вместе со стоимостью вершины z выражается величиной $A_2 - A_1$, где

$$A_1 = \left(C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z) - C^{H}_{p_1,s_1,\ldots,p_q,s_q}(z)\right) + C_{p_1,s_1,\ldots,p_q,s_q}(H)$$

(стоимость соответствующего фрагмента до замены), а

$$A_2 = \left(C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z) - C^{H'}_{p_1,s_1,\ldots,p_q,s_q}(z)\right) + C_{p_1,s_1,\ldots,p_q,s_q}(H')$$

(стоимость после замены). По условию леммы $A_1 = A_2$. Следовательно, разница в стоимости графов G и G' определяется стоимостью вершин в цепочке, исходящей из z.

Если из вершины z не выходит ребер, то $C_{p_1,s_1,\ldots,p_q,s_q}(G') = C_{p_1,s_1,\ldots,p_q,s_q}(G)$, поэтому положим $G'' = G'$. В противном случае рассмотрим последовательность вершин $z = z_0, z_1, z_2, \ldots$ в ориентированной цепочке, начинающейся в вершине z.

Если $d_G(z) = d_{G'}(z)$, то стоимость вершин цепочки не изменится, поэтому также положим $G'' = G'$. Иначе заметим, что в силу $d_H(z) \geq d_{H'}(z)$ для любого j справедливо $d_G(z_j) \geq d_{G'}(z_j)$.

Если $d_G(z_j) \in M_t(p_1, \ldots, p_q)$, где $j \geq 1$, то $C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_j) = C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_j)$, т. к. $d_G(z_{j-1})$, $d_{G'}(z_{j-1}) \in M_t(p_1, \ldots, p_q)$, что следует из условия $d_G(z)$, $d_{G'}(z) \in M_t(p_1, \ldots, p_q)$. Поэтому если условие $d_G(z_j) \in M_t(p_1, \ldots, p_q)$ выполнено для всех вершин цепочки, то замена H на H' не изменит стоимости графа. Тогда положим $G'' = G'$.

Иначе, пусть z_k — первая в цепочке вершина, для которой $d_G(z_k) \geq \sum_{i=1}^{t+1} p_i$. Изменение величины $C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_k)$ в результате замены H на H' может произойти в единственном случае $d_G(z_{k-1}) = \sum_{i=1}^{t+1} p_i - 1 > d_{G'}(z_{k-1})$ и только в сторону уменьшения (за счет изменения коэффициента $c^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_k, t+1)$).

Если $d_{G'}(z_k) = d_G(z_k)$, то $C_{p_1,s_1,\ldots,p_q,s_q}(G') \leq C_{p_1,s_1,\ldots,p_q,s_q}(G)$, поэтому положим $G'' = G'$.

В противном случае, т. е. если $d_{G'}(z_k) < d_G(z_k)$, заметим, что $d_G(z_k) = \sum_{i=1}^{t+1} p_i$, следовательно

$$c^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_k, t+1) = s_{t+1} > 0 = c^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_k, t+1),$$

и значит $C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z_k) > C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z_k)$. В этом случае удалим ребро, выходящее из вершины z_k графа G, если оно есть, методом леммы 4.12. Полученный при замене H на H' граф назовем G''. Возможное увеличение стоимости графа на 1 при удалении ребра компенсируется последующим уменьшением стоимости вершины z_k при замене, а стоимость других

вершин цепочки не изменяется. Поэтому действительно, $C_{p_1,s_1,\ldots,p_q,s_q}(G'') \leq C_{p_1,s_1,\ldots,p_q,s_q}(G)$. $\square$

В доказанной лемме ограничения $d_G(z)$, $d_{G'}(z) \in M_t(p_1,\ldots,p_q)$ и $A_1 = A_2$ (фактически, $A_1 \geq A_2$) можно снять при дополнительных предположениях относительно графа H'.

Лемма 4.14. *Пусть в графе $G \in \Delta$ подграф H с корневой вершиной z таков, что в графе G нет ребер, соединяющих вершины из $H\backslash\{z\}$ и $G\backslash H$. Пусть $H' \in \Delta$, и граф G' получается из G заменой подграфа H графом H' (при которой корневая вершина z' графа H' совмещается с вершиной z). Пусть $d_H(z) \in M_t(p_1,\ldots,p_q)$, $d_{H'}(z') \in M_{t'}(p_1,\ldots,p_q)$ и $C^{H'}_{p_1,s_1,\ldots,p_q,s_q}(z') = \sum_{i=1}^{t'} s_i$.*

Тогда, если выполнено условие

$$\begin{cases} C_{p_1,s_1,\ldots,p_q,s_q}(H') < C_{p_1,s_1,\ldots,p_q,s_q}(H), & d_{H'}(z') > d_H(z) \\ C_{p_1,s_1,\ldots,p_q,s_q}(H') + \sum_{t'+1}^{t} s_i \leq C_{p_1,s_1,\ldots,p_q,s_q}(H), & d_{H'}(z') \leq d_H(z) \end{cases},$$

то существует граф G'', получаемый из графа G' удалением некоторых ребер в ориентированной цепочке, исходящей из вершины z, для которого выполнено

$$C_{p_1,s_1,\ldots,p_q,s_q}(G'') \leq C_{p_1,s_1,\ldots,p_q,s_q}(G).$$

При этом, граф G'' может быть выбран так, что если z — корневая вершина в графе G и $d_G(z) < d_{G''}(z)$, то

$$C_{p_1,s_1,\ldots,p_q,s_q}(G'') < C_{p_1,s_1,\ldots,p_q,s_q}(G).$$

Доказательство. а) Рассмотрим случай $d_{H'}(z') > d_H(z)$. Тогда в графе G удалим ребро, исходящее из вершины z, если оно есть, способом леммы 4.12, затем заменим подграф H графом H'. Полученный граф выберем в качестве графа G''. Покажем, что выбор корректен.

Рассмотрим изменение стоимости вершины z. По условию леммы для любого $i \leq t'$ выполнено $c^{H'}_{p_1,s_1,\ldots,p_q,s_q}(z',i) = s_i$. Следовательно, при любом i справедливо $c^{H'}_{p_1,s_1,\ldots,p_q,s_q}(z',i) \geq c^{H}_{p_1,s_1,\ldots,p_q,s_q}(z,i)$. Отсюда вытекает

$$C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z) - C^{H'}_{p_1,s_1,\ldots,p_q,s_q}(z') \leq C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z) - C^{H}_{p_1,s_1,\ldots,p_q,s_q}(z).$$

Таким образом, стоимость графа увеличивается не более чем на 1 при удалении ребра и затем уменьшается при подстановке H' вместо H на величину $A_1 - A_2$, где

$$A_1 = \left(C^G_{p_1,s_1,\ldots,p_q,s_q}(z) - C^H_{p_1,s_1,\ldots,p_q,s_q}(z)\right) + C_{p_1,s_1,\ldots,p_q,s_q}(H),$$
$$A_2 = \left(C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z) - C^{H'}_{p_1,s_1,\ldots,p_q,s_q}(z')\right) + C_{p_1,s_1,\ldots,p_q,s_q}(H').$$

Проведенные выше рассуждения и условие леммы показывают, что $A_1 > A_2$. Значит, в результате преобразования стоимость не увеличивается.

Если вершина z является корневой в графе G, то при указанном преобразовании стоимость строго уменьшается, поскольку не требуется удаление исходящего из z ребра.

б) Пусть $d_{H'}(z') \le d_H(z)$ (это значит, что $t' \le t$). Оценим разницу в стоимости фрагментов графов G и G' с корневой вершиной z:

$$\begin{aligned} A_1 = \left(C^G_{p_1,s_1,\ldots,p_q,s_q}(z) - C^H_{p_1,s_1,\ldots,p_q,s_q}(z)\right) + C_{p_1,s_1,\ldots,p_q,s_q}(H) \ge \\ \ge \sum_{i=t+1}^{q} c^G_{p_1,s_1,\ldots,p_q,s_q}(z,i) + C_{p_1,s_1,\ldots,p_q,s_q}(H') + \sum_{i=t'+1}^{t} s_i \ge \\ \ge \sum_{i=t'+1}^{q} c^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z,i) + C_{p_1,s_1,\ldots,p_q,s_q}(H') = \\ = \left(C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z) - C^{H'}_{p_1,s_1,\ldots,p_q,s_q}(z')\right) + C_{p_1,s_1,\ldots,p_q,s_q}(H') = A_2. \end{aligned}$$

Если $A_1 > A_2$, то поступим так же, как в случае а).

Иначе заметим, что равенство $A_1 = A_2$ возможно, как показывает проведенная выше выкладка, только при выполнении условий

$$\begin{aligned} c^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z,i) &= c^G_{p_1,s_1,\ldots,p_q,s_q}(z,i), && i > t, \\ c^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z,i) &= c^G_{p_1,s_1,\ldots,p_q,s_q}(z,i) = s_i, && t'+1 \le i \le t. \end{aligned}$$

А это значит, что $d_G(z)$, $d_{G'}(z) \in M_{t''}(p_1,\ldots,p_q)$ при некотором t''. Таким образом, выполнены условия леммы 4.13. Поэтому, при необходимости удалив из графа G' некоторые ребра в цепочке, исходящей из вершины z, можно получить граф G'', имеющий стоимость, не большую, чем граф G. $\square$

Далее число вершин в графе G будем обозначать через $|G|$.

Лемма 4.15. *Пусть $T, H \in \Delta$ и для главной корневой вершины z' графа H выполнено $d_H(z') = d + \sum_{i=1}^{t-1} p_i \in M_{t-1}(p_1, \ldots, p_q)$. Тогда*

$$C_{p_1,s_1,\ldots,p_q,s_q}(T \circ H) = C_{p_t-d,\, s_t,\ldots,p_q,s_q}(T) + |T| \cdot C_{p_1,s_1,\ldots,p_q,s_q}(H).$$

Доказательство. В соответствии с определением композиции рассмотрим граф $T \circ H$ как внешний граф T, построенный на корневых вершинах внутренних графов, изоморфных графу H. Обозначим через Z_T множество вершин внешнего подграфа, а через Z_H — множество остальных вершин графа $T \circ H$.

Пусть $z \in Z_T$. Легко проверяются следующие соотношения:

$$\mu(T \circ H) = \mu(T) + |T|(\mu(H) - 1),$$
$$C^{T\circ H}_{p_1,s_1,\ldots,p_q,s_q}(z) = C^{T}_{p_t-d,\, s_t,\ldots,p_q,s_q}(z) + C^{H}_{p_1,s_1,\ldots,p_q,s_q}(z').$$

Как следствие, получаем

$$\begin{aligned}
&C_{p_1,s_1,\ldots,p_q,s_q}(T \circ H) = \\
&\quad = \mu(T \circ H) - 1 + \sum_{z \in Z_T} C^{T\circ H}_{p_1,s_1,\ldots,p_q,s_q}(z) + \sum_{z \in Z_H} C^{T\circ H}_{p_1,s_1,\ldots,p_q,s_q}(z) = \\
&\quad = \mu(T) - 1 + |T|(\mu(H) - 1) + \\
&+ \sum_{z \in Z_T} \left(C^{T}_{p_t-d,\, s_t,\ldots,p_q,s_q}(z) + C^{H}_{p_1,s_1,\ldots,p_q,s_q}(z) \right) + \sum_{z \in Z_H} C^{H}_{p_1,s_1,\ldots,p_q,s_q}(z) = \\
&\quad = C_{p_t-d,\, s_t,\ldots,p_q,s_q}(T) + |T| \cdot C_{p_1,s_1,\ldots,p_q,s_q}(H). \quad \square
\end{aligned}$$

Следующая лемма является ключевой для установления стоимости множества подграфов гиперпары и формулируется в достаточно общем виде с целью охватить различные случаи, возникающие при анализе стоимости указанных множеств. Перед тем как ее сформулировать, введем некоторые обозначения, мотивируемые леммой 4.14.

Пусть T — корневое дерево с корневой вершиной z. Положим

$$\theta_{p_1,s_1,\ldots,p_q,s_q}(T) = \min\{d_G(z) \mid G \in \delta(T),\, C_{p_1,s_1,\ldots,p_q,s_q}(G) = C_{p_1,s_1,\ldots,p_q,s_q}(\delta(T))\},$$
$$\delta^t_{p_1,\ldots,p_q}(T) = \delta(T) \cap \{G \,|\, d_G(z) \in M_t(p_1, \ldots, p_q)\},$$
$$\Omega^t_{p_1,s_1,\ldots,p_q,s_q}(T) = \begin{cases} \min\limits_{G \in \delta^t_{p_1,\ldots,p_q}(T)} C_{p_1,s_1,\ldots,p_q,s_q}(G), & \delta^t_{p_1,\ldots,p_q}(T) \neq \emptyset \\ +\infty, & \delta^t_{p_1,\ldots,p_q}(T) = \emptyset \end{cases}.$$

Лемма 4.16. *Пусть T, H — корневые деревья,*

$$\theta_{p_1,s_1,\ldots,p_q,s_q}(H) = d + \sum_{i=1}^{t'} p_i \in M_{t'}(p_1,\ldots,p_q)$$

и для любого $t \geq t'$

$$C_{p_1,s_1,\ldots,p_q,s_q}(\delta(H)) + \sum_{i=t'+1}^{t} s_i \leq \Omega^t_{p_1,s_1,\ldots,p_q,s_q}(H). \tag{4.6}$$

Пусть для главной корневой вершины z' некоторого графа H^0 минимальной стоимости из $\delta(H)$ выполняется $d_{H^0}(z') = \theta_{p_1,s_1,\ldots,p_q,s_q}(H)$ и $C^{H^0}_{p_1,s_1,\ldots,p_q,s_q}(z') = \sum_{i=1}^{t'} s_i$. Тогда минимум $(p_1,s_1,\ldots,p_q,s_q)$-стоимости на множестве $\delta(T \circ H)$ и минимум глубины корневой вершины среди минимальных по стоимости графов достигается на графе из $\delta(T) \circ H^0$. Справедливы соотношения:

$$C_{p_1,s_1,\ldots,p_q,s_q}(\delta(T \circ H)) = C_{p_{t'+1}-d,\, s_{t'+1},\ldots,p_q,s_q}(\delta(T)) + |T| \cdot C_{p_1,s_1,\ldots,p_q,s_q}(\delta(H)),$$

$$\theta_{p_1,s_1,\ldots,p_q,s_q}(T \circ H) = \theta_{p_{t'+1}-d,\, s_{t'+1},\ldots,p_q,s_q}(T) + \theta_{p_1,s_1,\ldots,p_q,s_q}(H).$$

Доказательство. В соответствии с определением будем рассматривать граф $T \circ H$ как внешний граф T, построенный на корневых вершинах внутренних графов, изоморфных графу H. Покажем, что существует оптимальный по стоимости подграф графа $T \circ H$, содержащий целиком оптимальные подграфы внутренних графов, и минимум глубины корневой вершины достигается именно на таком графе. Для этого достаточно показать, что в произвольном графе $G \in \delta(T \circ H)$ можно заменить фрагмент, образованный пересечением с любым внутренним графом H, графом H^0, при необходимости произведя удаление некоторых ребер во внешнем подграфе, и при этом стоимость всего графа не возрастет.

Обозначим через H_z фрагмент графа G, образованный пересечением с некоторым внутренним графом с корневой вершиной z. Для подграфа H_z графа G и графа G', получаемого из G заменой H_z на H^0, выполнены условия леммы 4.14. Действительно, если $d_{H_z}(z) < \theta_{p_1,s_1,\ldots,p_q,s_q}(H) = d_{H^0}(z')$, то

H_z не является графом минимальной стоимости на множестве $\delta(H)$, поэтому $C_{p_1,s_1,\ldots,p_q,s_q}(H_z) > C_{p_1,s_1,\ldots,p_q,s_q}(H^0)$. Иначе, т. е. если $d_{H^0}(z') \le d_{H_z}(z) \in M_t(p_1,\ldots,p_q)$, согласно (4.6) выполнено

$$C_{p_1,s_1,\ldots,p_q,s_q}(H_z) \ge C_{p_1,s_1,\ldots,p_q,s_q}(H^0) + \sum_{i=t'+1}^{t} s_i.$$

Выполним замену H_z на H^0 методом леммы 4.14, которая гарантирует, что удаляемые при замене ребра не принадлежам другим внутренним подграфам графа G.

Так поступим со всеми внутренними подграфами графа G. Согласно лемме 4.14 увеличение глубины корневой вершины графа G возможно только в случае уменьшения стоимости, что означает неоптимальность (исходного) графа G.

Теперь можно ограничить рассмотрение графами $G \in \delta(T) \circ H^0$. Проведенное выше рассуждение показывает, что такие графы доставляют как минимум стоимости, так и минимум глубины корневой вершины среди минимальных по стоимости графов.

Тем самым, стоимость множества $\delta(T \circ H)$ совпадает со стоимостью множества $\delta(T) \circ H^0$, а стоимость последнего определяется из леммы 4.15. Утверждение леммы в отношении величины $\theta_{p_1,s_1,\ldots,p_q,s_q}(T \circ H)$ следует немедленно. □

4.3.5 Стоимость множества подграфов гиперпары

В настоящем разделе устанавливаются простые соотношения для минимальной стоимости подграфов гиперпар различного размера.

Лемма 4.17. *Для любых $l \le p_1 - 1$ и $t > 0$ справедливо:*

$$C_{p_1,s_1,\ldots,p_q,s_q}(\delta(H_l)) = 0, \quad \theta_{p_1,s_1,\ldots,p_q,s_q}(H_l) = l, \quad \Omega^t_{p_1,s_1,\ldots,p_q,s_q}(H_l) = +\infty.$$

При этом для любого $k \ge l$ выполняются соотношения:

$$C_{p_1,s_1,\ldots,p_q,s_q}(\delta(H_k)) = C_{1,s_1,\ldots,p_q,s_q}(\delta(H_{k-l})),$$
$$\theta_{p_1,s_1,\ldots,p_q,s_q}(H_k) = \theta_{1,s_1,\ldots,p_q,s_q}(H_{k-l}) + l.$$

Доказательство. Несложно проверить первое утверждение: очевидно, что граф H_l является единственным графом минимальной стоимости в $\delta(H_l)$. Второе утверждение получается применением леммы 4.16 к графу $H_k = H_{k-l} \circ H_l$. □

Лемма 4.18. *Пусть связный подграф гиперпары H_k содержит $l > 0$ вершин, к которым ведут ребра из листьев подграфа. Тогда он содержит всего не более $l(k-1)+2$ вершин.*

Доказательство. Пусть $l = 1$ и t — порядок вершины, соединенной с листом. Тогда рассматриваемый подграф может содержать только ребра, входящие в данную вершину (их не более t) и ориентированную цепочку, выходящую из этой вершины (ее длина не превосходит $k-t$). Следовательно, в нем не более $k+1$ вершин.

Далее заметим, что если связный подграф гиперпары H_k не содержит ребра, соединяющего корневую вершину гиперпары с вершиной порядка $k-1$, то он является подграфом одной из гиперпар H_{k-1} (см. рис. 11).

Докажем индуктивный переход от $l-1$ к l. Граф, имеющий l вершин, соединенных с листьями, можно представить как объединение графа с $l-1$ такими вершинами и графом с одной такой вершиной. Например, в качестве последнего можно взять произвольную вершину z глубины 1 с пучком входящих в нее ребер вместе с цепочкой ребер (возможно, пустой), выходящей из z и содержащей вершины, не соединенные с листьями.

При этом можно считать, что один из графов является подграфом гиперпары H_{k-1}. По индуктивному предположению, в первом графе не более $(l-1)(k-1)+2$ вершин (или $(l-1)(k-2)+2$ в случае подграфа H_{k-1}), а при объединении со вторым графом в силу того, что в результате получается связный граф, добавляется еще не более $k-1$ (или соответственно k) вершин. Тогда число вершин в объединении не превосходит

$$\max\{(l-1)(k-1)+2+(k-1),\ (l-1)(k-2)+2+k\} \le l(k-1)+2. \qquad \square$$

Лемма 4.19. *Пусть $k \le s_1$. Тогда*

$$C_{1,s_1,\ldots,p_q,s_q}(\delta(H_k)) = 2^k - 1, \quad \theta_{1,s_1,\ldots,p_q,s_q}(H_k) = 0.$$

Кроме того,

$$C_{1,s_1,\ldots,p_q,s_q}(\delta(H_{s_1+1})) = 2^{s_1+1} - 2, \quad \theta_{1,s_1,\ldots,p_q,s_q}(H_{s_1+1}) = 1$$

и для любого $t \geq 1$*:*

$$\Omega^t_{1,s_1,\ldots,p_q,s_q}(H_k) \geq 2^k - 1 + (s_1 - k) + \sum_{i=2}^{t} s_i,$$

$$\Omega^t_{1,s_1,\ldots,p_q,s_q}(H_{s_1+1}) \geq 2^{s_1+1} - 2 + \sum_{i=2}^{t} s_i.$$

Доказательство. Пусть граф $G \in \delta(H_k)$ содержит связные компоненты $G_1, \ldots, G_j$, отличные от изолированных вершин. Случай $j = 0$ тривиален, поэтому рассмотрим случай $j \geq 1$. Без ограничения общности, будем считать, что $G_i \in \delta(H_{k-1})$ при $i > 1$. Пусть граф G_i содержит l_i вершин, соединенных с листьями. Тогда по лемме 4.18 в нем всего не более $l_i(k-2)+2$ вершин, если $i > 1$, и не более $l_1(k-1)+2$ вершин в случае $i = 1$.

Пользуясь оценками

$$\mu(G) = |G| - \sum_{i=1}^{j}(|G_i| - 1), \qquad C_{1,s_1,\ldots,p_q,s_q}(G_i) \geq l_i s_1,$$

стоимость графа теперь можно оценить как

$$\begin{aligned} C_{1,s_1,\ldots,p_q,s_q}(G) = \mu(G) - 1 + \sum_{i=1}^{j} C_{1,s_1,\ldots,p_q,s_q}(G_i) \geq \\ \geq |G| - 1 + \sum_{i=1}^{j}(l_i s_1 - |G_i| + 1) \geq \\ \geq 2^k - 1 + l_1(s_1 - k + 1) - 1 + \sum_{i=2}^{j}(l_i(s_1 - k + 2) - 1). \quad (4.7) \end{aligned}$$

При $k \leq s_1$ имеем $C_{1,s_1,\ldots,p_q,s_q}(G) \geq 2^k - 1$. Оценка стоимости $2^k - 1$ достигается на графе, состоящем из всех изолированных вершин, откуда следуют соотношения первой части леммы. Если глубина корневой вершины графа G не меньше 1, то, подставляя $l_1 \geq 1$ в (4.7), получаем $\Omega^1_{1,s_1,\ldots,p_q,s_q}(H_k) \geq 2^k - 1 + (s_1 - k)$.

При $k = s_1+1$ приходим к оценке $C_{1,s_1,\ldots,p_q,s_q}(G) \geq 2^k-2$, причем равенство возможно только при $|G_1| = l_1(k-1)+2$, значит, корневая вершина гиперпары H_k не может быть изолированной в графе G. Указанная оценка достигается на графе, состоящем из пучка ребер, входящих в корневую вершину гиперпары, и изолированных остальных вершин, откуда следуют соотношения $C_{1,s_1,\ldots,p_q,s_q}(\delta(H_{s_1+1})) = 2^{s_1+1}-2$ и $\theta_{1,s_1,\ldots,p_q,s_q}(H_{s_1+1}) = 1$.

Докажем последние соотношения леммы при $t \geq 2$. В графе $G \in \delta^t_{1,p_2,\ldots,p_q}(H_k)$ найдутся вершины $z_2,\ldots,z_t$, такие, что $d_G(z_j) = 1+\sum_{i=2}^{j} p_i$. Учитывая в оценке стоимости графа слагаемые $c^G_{1,s_1,\ldots,p_q,s_q}(z_j,j) = s_j$, получаем требуемые соотношения. □

Лемма 4.20. *Пусть $p_1 \leq k < p_1+s_1$. Тогда*

$$C_{p_1,s_1,\ldots,p_q,s_q}(\delta(H_k)) = 2^{k-p_1+1}-1, \qquad \theta_{p_1,s_1,\ldots,p_q,s_q}(H_k) = p_1-1.$$

Пусть $k \geq p_1+s_1$. Тогда

$$C_{p_1,s_1,\ldots,p_q,s_q}(\delta(H_k)) = 2^{k-p_1-s_1+1}(2^{s_1}-1) + C_{p_2,s_2,\ldots,p_q,s_q}(\delta(H_{k-p_1-s_1})),$$
$$\theta_{p_1,s_1,\ldots,p_q,s_q}(H_k) = p_1 + \theta_{p_2,s_2,\ldots,p_q,s_q}(H_{k-p_1-s_1}).$$

Доказательство. Рассмотрим гиперпару H_k в первом случае как композицию $H_{k-p_1+1} \circ H_{p_1-1}$, а во втором — как композицию $H_{k-p_1-s_1} \circ H_{s_1+1} \circ H_{p_1-1}$. Затем применим лемму 4.16, используя полученные в леммах 4.17 и 4.19 соотношения. □

Подведем итог серие лемм.

Лемма 4.21. *Пусть $k = \sum_{i=1}^{t}(p_i+s_i)+k'$ при условии: $k' < (p_{t+1}+s_{t+1})$ или $t = q$. Тогда*

$$C_{p_1,s_1,\ldots,p_q,s_q}(\delta(H_k)) = \sum_{t'=1}^{t} 2^{k'+1+\sum\limits_{i=t'+1}^{t}(p_i+s_i)}(2^{s_{t'}}-1) + \\ + \begin{cases} 0, & k' < p_{t+1} \\ 2^{k'-p_{t+1}+1}-1, & k' \geq p_{t+1} \end{cases},$$

$$\theta_{p_1,s_1,\ldots,p_q,s_q}(H_k) = \sum_{i=1}^{t} p_i + \begin{cases} k', & k' < p_{t+1} \\ p_{t+1}-1, & k' \geq p_{t+1} \end{cases}.$$

При этом существует граф H_k^0 минимальной стоимости из $\delta(H_k)$, для корневой вершины z которого выполнено

$$C^{H_k^0}_{p_1,s_1,\ldots,p_q,s_q}(z) = \sum_{i=1}^{t} s_i, \qquad d_{H_k^0}(z) = \theta_{p_1,s_1,\ldots,p_q,s_q}(H_k).$$

Доказательство. Применяем, покуда возможно, лемму 4.20. Граф H_k^0 имеет вид (многократной) композиции графов минимальной стоимости, построенных в леммах 4.17 и 4.19. □

Несложно заметить, что доказанное соотношение для стоимости можно переписать в более компактном виде. Справедливо

Следствие 4.2. *Определим $(k+1)$-разрядное число N как*

$$N = \underbrace{0\ldots0}_{p_1}\,\underbrace{1\ldots1}_{s_1}\,\underbrace{0\ldots0}_{p_2}\ldots$$

Тогда

$$C_{p_1,s_1,\ldots,p_q,s_q}(\delta(H_k)) = N.$$

4.3.6 Оптимальность гиперпары

Далее нам понадобится следующее расширение леммы 4.14.

Лемма 4.22. *Пусть в графе $G \in \Delta$ подграф $W = H \circ F$ с корневой вершиной z таков, что в графе G нет ребер, соединяющих вершины из $W \setminus \{z\}$ и $G \setminus W$. Пусть $H' \in \Delta$, $|H'| = |H|$, $W' = H' \circ F$ и граф G' получается из G заменой подграфа W графом W' (при которой корневая вершина z' графа W' совмещается с вершиной z). Пусть для главной корневой вершины z_F графа F выполнено $d_F(z_F) = d + \sum_{i=1}^{t-1} p_i \in M_{t-1}(p_1,\ldots,p_q)$ и $C^F_{p_1,s_1,\ldots,p_q,s_q}(z_F) = \sum_{i=1}^{t-1} s_i$.*

Положим $(p'_1, s'_1, \ldots, p'_{q'}, s'_{q'}) = (p_t - d, s_t, \ldots, p_q, s_q)$. Пусть для главных корневых вершин z_H и z'_H графов H и H' соответственно выполнено $d_H(z_H) \in M_\tau(p'_1,\ldots,p'_{q'})$, $d_{H'}(z'_H) \in M_{\tau'}(p'_1,\ldots,p'_{q'})$ и $C^{H'}_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(z'_H) = \sum_{i=1}^{\tau'} s'_i$.

Тогда, если выполнено условие

$$\begin{cases} C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H') < C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H), & d_{H'}(z'_H) > d_H(z_H) \\ C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H') + \sum_{\tau'+1}^{\tau} s'_i \le C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H), & d_{H'}(z'_H) \le d_H(z_H) \end{cases},$$

то существует граф G'', получаемый из графа G' удалением некоторых ребер в ориентированной цепочке, исходящей из вершины z, для которого

$$C_{p_1,s_1,\ldots,p_q,s_q}(G'') \le C_{p_1,s_1,\ldots,p_q,s_q}(G).$$

Доказательство. Сведем доказываемую лемму к лемме 4.14. Для этого рассмотрим изменение стоимости $A_2 - A_1$ вершины z вместе с подграфом W внутри графа G при замене W на W':

$$A_1 = \left(C^G_{p_1,s_1,\ldots,p_q,s_q}(z) - C^W_{p_1,s_1,\ldots,p_q,s_q}(z)\right) + C_{p_1,s_1,\ldots,p_q,s_q}(W),$$
$$A_2 = \left(C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z) - C^{W'}_{p_1,s_1,\ldots,p_q,s_q}(z')\right) + C_{p_1,s_1,\ldots,p_q,s_q}(W').$$

По условиям леммы при всех $i < t$

$$c^G_{p_1,s_1,\ldots,p_q,s_q}(z,i) = c^W_{p_1,s_1,\ldots,p_q,s_q}(z,i) = c^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z,i) = c^{W'}_{p_1,s_1,\ldots,p_q,s_q}(z',i) = s_i.$$

Кроме того, из леммы 4.15 следует, что

$$C_{p_1,s_1,\ldots,p_q,s_q}(W) - C_{p_1,s_1,\ldots,p_q,s_q}(W') = C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H) - C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H').$$

Поэтому $A_1 - A_2 = A'_1 - A'_2$, где

$$A'_1 = \left(\sum_{i \ge t} c^G_{p_1,s_1,\ldots,p_q,s_q}(z,i) - C^H_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(z_H)\right) + C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H),$$
$$A'_2 = \left(\sum_{i \ge t} c^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z,i) - C^{H'}_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(z'_H)\right) + C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H').$$

Основываясь на полученном соотношении, доказательство леммы можно завершить повторением рассуждения из доказательства леммы 4.14 (доказательство леммы 4.14 в рассматриваемой ситуации проходит с точностью до некоторых переобозначений). □

Лемма 4.23. $C_{p_1,s_1,\ldots,p_q,s_q}(\delta(T_k)) = C_{p_1,s_1,\ldots,p_q,s_q}(\delta(H_k))$.

Доказательство. Покажем, что по произвольному графу $G \in \delta(T_k)$ можно построить граф $G' \in \delta(H_k)$, имеющий не большую стоимость. Напомним, что условие $G \in \delta(T_k)\backslash\delta(H_k)$ означает, что граф G содержит неправильные ребра, т. е. ребра, ведущие из вершин с бóльшим порядком в вершины с меньшим порядком.

Для характеризации «неправильности» графа $G \subset T_k$, т. е. того, что $G \not\subset H_k$, удобно ввести численную величину $e(G)$, определяемую следующим образом. Пусть ρ — ребро графа G, ведущее из вершины v в вершину v'. Положим $e(\rho) = 0$, если ρ — правильное, и $e(\rho)$ равно разности порядков вершин v и v' в графе T_k, если ρ — неправильное. Далее положим $e(G) = \sum_{\rho\in G} e(\rho)$. Таким образом, $e(G) = 0$, если $G \subset H_k$, и $e(G) > 0$ — иначе.

Поэтому для доказательства утверждения леммы достаточно указать такое преобразование графа $G \in \delta(T_k)\backslash\delta(H_k)$, которое уменьшает величину $e(G)$ и не увеличивает стоимость графа.

Стратегия доказательства состоит в том, чтобы в произвольном графе $G \in \delta(T_k) \setminus \delta(H_k)$ выбрать подходящее неправильное ребро ρ и затем построить не увеличивающее стоимость преобразование графа, при котором ребро ρ либо удаляется из графа, либо перенаправляется в вершину с бóльшим порядком, при этом новых неправильных ребер не возникает.

Пусть вершина z' имеет наименьший порядок j' среди тех вершин графа G, из которых выходят неправильные ребра. Обозначим через ρ (неправильное) ребро, выходящее из нее, через z — вершину, в которую ведет это ребро, через j — порядок вершины z. Обозначим через z^* вершину, в которую ведет правильное ребро (его обозначим через ρ_1) из z в графе T_k. Через ρ^* обозначим ребро графа T_k, соединяющее вершины z' и z^* (существование ребра ρ^* гарантируется определением графа T_k).

Через H_z и H_{z^*} обозначим пересечение графа G с подграфами T_j графа T_k на корневых вершинах z и z^* соответственно. Все эти подграфы определены однозначно, см. выше в §4.3.3. Введенные обозначения и фрагмент графа G, к которому они относятся, иллюстрирует рис. 12.

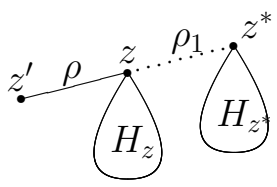


Рис. 12: Фрагмент графа G

Заметим, что изображенный на рис. 12 фрагмент графа G не содержит, за исключением ρ, неправильных ребер, поскольку все вершины в нем, кроме z' и, возможно, z^*, имеют порядок меньше j'. В частности, $H_z, H_{z^*} \in \delta(H_j)$. Таким образом, рассматриваемый фрагмент графа G может быть связан с остальной частью графа только через вершины z' и z^*.

Пусть $j = \sum_{i=1}^{r-1}(p_i + s_i) + a$, где $0 \le a < p_r + s_r$. Представим граф H_j в виде композиции $H_l \circ H_{j_{2r-1}} \circ \ldots \circ H_{j_2} \circ H_{j_1}$, где при $i < r$ положим $j_{2i-1} = p_i - 1$ и $j_{2i} = s_i + 1$; $j_{2r-1} = \min\{p_r - 1, a\}$, $l = a - j_{2r-1}$. По построению, $0 \le l \le s_r$.

Используя лемму 4.22, покажем, что подграфы H_z и H_{z^*} можно заменить графами из $\delta(H_l) \circ H^0_{j-l}$, где H^0_k — оптимальный по стоимости граф с минимальной глубиной корневой вершины из леммы 4.21, не увеличивая стоимости графа G. Ограничимся рассмотрением подграфа H_z (случай подграфа H_{z^*} ничем не отличается).

Пусть первоначально $F = H^0_0$ и $(p'_1, s'_1, \ldots, p'_{q'}, s'_{q'}) = (p_1, s_1, \ldots, p_q, s_q)$.

Представим H_j в виде композиции $T \circ U$ внешнего графа $T = H_l \circ H_{j_{2r-1}} \circ \ldots \circ H_{j_2}$ и внутреннего графа $U = H_{j_1}$. Отметим, что $H_z = H_z \circ F$. В графе H_z рассмотрим произвольный подграф $H \circ F$ с корневой вершиной z^0 (принадлежащей внешнему графу), образованный пересечением H_z с соответствующим внутренним подграфом графа H_j. Согласно лемме 4.22 подграф H можно заменить графом $H' = H^0_{j_1}$, при необходимости удалив некоторые ребра, зависящие от z^0, не увеличивая стоимости графа G. Условия для применения леммы 4.22 обеспечиваются леммой 4.17 (на самом деле, на этом первом шаге вместо леммы 4.22 можно использовать лемму 4.14). Применяя указанное преобразование ко всем внутренним подграфам графа H_z, в итоге получаем на месте H_z граф вида $T' \circ H' \circ F$,

$T' \in \delta(T)$.

Выполним переобозначения:

$$H_z := T' \circ H' \circ F, \quad F := H' \circ F, \quad p'_1 := p'_1 - j_1.$$

Если $r > 1$, продолжим. В этом случае $p'_1 = 1$, $j_2 = s'_1 + 1$.

Теперь запишем $H_j = T \circ U$, где $T = H_l \circ H_{j_{2r-1}} \circ \ldots \circ H_{j_3}$ и $U = H_{j_2} \circ H_{j_1}$. В графе H_z рассмотрим произвольный внутренний подграф $H \circ F \in \delta(U)$ с корневой вершиной z^0, образованный пересечением H_z с соответствующим внутренним подграфом графа H_j. Методом леммы 4.22 подграф $H \circ F$ заменим графом $H' \circ F$, где $H' = H^0_{j_2}$. Условия для применения леммы 4.22 обеспечивает лемма 4.19. Так поступим со всеми внутренними подграфами графа H_z, в итоге получим вместо H_z граф вида $T' \circ H' \circ F$, $T' \in \delta(T)$.

Выполним переобозначения:

$$H_z := T' \circ H' \circ F, \quad F := H' \circ F, \quad (p'_1, s'_1, \ldots, p'_{q'}, s'_{q'}) := (p'_2, s'_2, \ldots, p'_{q'}, s'_{q'}).$$

Будем продолжать в том же духе, пока на месте H_z не получим граф из $\delta(H_l) \circ H^0_{j-l}$.

Проведенное рассуждение позволяет ограничиться рассмотрением ситуации $H_z = H \circ H^0_{j-l}$, $H^*_z = H^* \circ H^0_{j-l}$, где $H, H^* \in \delta(H_l)$. Положим

$$(p'_1, s'_1, \ldots, p'_{q'}, s'_{q'}) = (p_r - j_{2r-1}, s_r, \ldots, p_q, s_q).$$

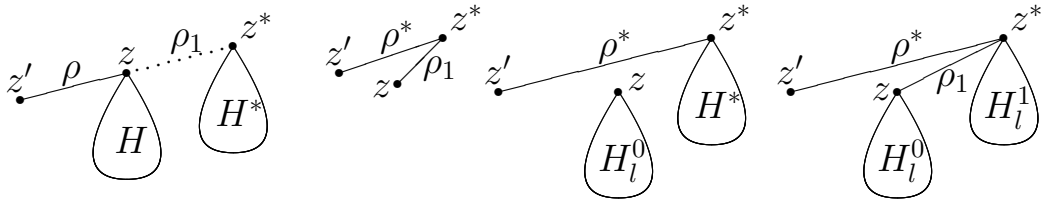
Напомним, что $l \le s'_1$. При этом либо $p'_1 > 0$ и $l = 0$, либо $p'_1 = 1$.

Далее рассмотрим несколько возможных случаев.

а) Пусть $d_G(z') < d_{H^0_{j-l}}(z^0) = \theta_{p_1,s_1,\ldots,p_q,s_q}(H_{j-l}) = \sum_{i=1}^{r-1} p_i + j_{2r-1}$, где z^0 — корневая вершина графа H^0_{j-l}. Поскольку $C^{H^0_{j-l}}_{p_1,s_1,\ldots,p_q,s_q}(z^0) = \sum_{i=1}^{r-1} s_i$, то в графе G ребро ρ можно заменить ребром ρ^*: при этом стоимость графа не изменяется, поскольку сохраняется число компонент связности, глубина всех вершин и, как следствие, стоимость всех отличных от z и z^* вершин, а стоимость вершин z и z^* не зависит от наличия ребер ρ и ρ^*.

Далее считаем, что $d_G(z') \ge \theta_{p_1,s_1,\ldots,p_q,s_q}(H_{j-l})$. Ниже будут рассматриваться преобразования фрагмента графа G, изображенного на рис. 12, в которые будут вовлечены ребра ρ, ρ_1, ρ^* и внешние подграфы H и H^* графов H_z и H^*_z. Дополнительно могут удаляться некоторые ребра в цепочке,

исходящей из z^*, методами лемм 4.12 и 4.13. При этом изменение стоимости графа, происходящее за пределами рассматриваемого фрагмента, может быть оценено леммами 4.12 и 4.13, а изменение стоимости внутри фрагмента фактически определяется изменением $(p'_1, s'_1, \ldots, p'_{q'}, s'_{q'})$-стоимости графа W, изображенного на рис. 13а, в котором глубина вершины z' формально полагается равной $d = d_G(z') - \theta_{p_1,s_1,\ldots,p_q,s_q}(H_{j-l})$ (более формально — ниже).



а) W б) W^+ $[l = 0]$ в) W^- г) $W^\times$

Рис. 13: Рассматриваемые подграфы

Положим

$$A = C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H) + \left(C^W_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(z) - C^H_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(z)\right). \qquad (4.8)$$

б) Пусть $\rho_1 \notin G$ и $A \geq C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(\delta(H_l))+1$. Тогда удалим ребро ρ и заменим подграф H графом H^0_l. При этом увеличение числа компонент связности при удалении ребра компенсируется уменьшением стоимости подграфа.

в) Иначе, если $\rho_1 \in G$ и $A \geq C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(\delta(H_l)) + 2$, то удалим ребро ρ_1 методом леммы 4.12, затем удалим ребро ρ и выполним замену подграфа H на граф H^0_l.

Далее считаем, что либо $\rho_1 \notin G$ и $A = C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(\delta(H_l))$, либо $\rho_1 \in G$ и $A \leq C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(\delta(H_l)) + 1$.

Ниже мы рассмотрим несколько вариантов преобразования графа G в граф G', осуществляемого подстановкой вместо подграфа W одного из графов, изображенных на рис. 13б–г. При такой подстановке изменение стоимости графа G складывается из четырех слагаемых:

$$C_{p_1,s_1,\ldots,p_q,s_q}(G') - C_{p_1,s_1,\ldots,p_q,s_q}(G) = \sigma_1 + \sigma_2 + \sigma_3 + \sigma_4,$$

где σ_1 — изменение стоимости вершины z вместе с подграфом H, σ_2 — изменение стоимости вершины z^* вместе с подграфом H^*, σ_3 — изменение числа компонент связности при возможном добавлении или удалении ребра ρ_1, σ_4 — изменение стоимости вершин в цепочке, исходящей из z^*.

Обозначая через H'_z и H'_{z^*} графы, в которые превращаются соответственно подграфы H_z и H_{z^*} графа G при рассматриваемой замене, приведем формулы для σ_1 и σ_2:

$$\sigma_1 = \left(C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z) - C^{H'_z}_{p_1,s_1,\ldots,p_q,s_q}(z)\right) + C_{p_1,s_1,\ldots,p_q,s_q}(H'_z) - \\ - \left(C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z) - C^{H_z}_{p_1,s_1,\ldots,p_q,s_q}(z)\right) - C_{p_1,s_1,\ldots,p_q,s_q}(H_z) = \\ = C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H^0_l) - A,$$

$$\sigma_2 = \left(C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z^*) - C^{H'_{z^*}}_{p_1,s_1,\ldots,p_q,s_q}(z^*)\right) + C_{p_1,s_1,\ldots,p_q,s_q}(H'_{z^*}) - \\ - \left(C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z^*) - C^{H_{z^*}}_{p_1,s_1,\ldots,p_q,s_q}(z^*)\right) - C_{p_1,s_1,\ldots,p_q,s_q}(H_{z^*}).$$

По условию на A имеем $\sigma_1 \in \{0, -1\}$. Очевидно, $\sigma_3 \in \{0, \pm 1\}$. Также заметим, что $\sigma_4 = 0$, если из z^* в графе G не выходит ребер.

г) Пусть $p'_1 > 1$ и $l = 0$. В этом случае графы H и H^* состоят из единственной вершины.

Рассмотрим подстановку W^+ вместо W, превращающую граф G в граф G', см. рис. 13б.

г.1) Пусть $\rho_1 \notin G$. Заметим, что в этом случае $C^{W}_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(z) = 0$. Удалим ребро, выходящее из вершины z^*, если оно имеется, методом леммы 4.12, и заменим подграф W графом W^+. Стоимость при этом не увеличится, т. к. возможное увеличение стоимости графа при удалении ребра компенсируется уменьшением стоимости при замене W на W^+: $\sigma_1 = \sigma_2 = 0$ (т. к. стоимость вершин z и z^* не изменяется) и $\sigma_3 = -1$.

г.2) Пусть $\rho_1 \in G$. В этом случае $C^{W}_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(z) \leq 1$ и $\sigma_3 = 0$. Справедливо $\sigma_1 + \sigma_2 \leq 0$, т. к.

$$C^{W^+}_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(z) = 0, \quad C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z^*) \leq C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z^*) + C^{W}_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(z).$$

г.2.1) Если $\sigma_1 + \sigma_2 < 0$, то поступим так же, как в п. г.1).

г.2.2) Иначе (т. е. если $\sigma_1 + \sigma_2 = 0$) справедливо равенство

$$C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z^*) = C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z^*) + C^{W}_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(z). \tag{4.9}$$

Можно проверить, что это возможно только в случае $d_G(z^*), d_{G'}(z^*) \in M_t(p_1, s_1, \ldots, p_q, s_q)$ при некотором t. Действительно, предположим противное. Т.к. по построению $d_G(z^*) - 1 \leq d_{G'}(z^*) \leq d_G(z^*)$, то $d_G(z^*) = d_{G'}(z^*) + 1 = \sum_{i=1}^{t} p_i$ при некотором t. Но тогда

$$C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z^*) \leq C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z^*) - s_t + C^{W}_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(z),$$

что приводит к противоречию.

Рассмотрим произвольный подграф U графа G с корневой вершиной z^*, содержащий фрагмент, изображенный на рис. 12, и связанный с остальной частью графа G только через вершину z^*. Лемма 4.13 позволяет заменить его графом U', отличающимся от U подстановкой W^+ вместо W, не увеличивая стоимость графа G. Условие леммы $d_U(z^*) \geq d_{U'}(z^*)$ выполнено в силу $\rho_1 \in G$. Последнее условие леммы 4.13 выполнено в силу того, что стоимость графов U и U' различается только стоимостью вершин z и z', т. е. записывая формально,

$$\begin{aligned} &C_{p_1,s_1,\ldots,p_q,s_q}(U) - C_{p_1,s_1,\ldots,p_q,s_q}(U') = \\ &\quad = C^{U}_{p_1,s_1,\ldots,p_q,s_q}(z) - C^{U'}_{p_1,s_1,\ldots,p_q,s_q}(z) + C^{U}_{p_1,s_1,\ldots,p_q,s_q}(z^*) - C^{U'}_{p_1,s_1,\ldots,p_q,s_q}(z^*), \end{aligned}$$

и в силу вытекающего из (4.9) соотношения

$$C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z^*) + C^{U}_{p_1,s_1,\ldots,p_q,s_q}(z) = C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z^*) + C^{U'}_{p_1,s_1,\ldots,p_q,s_q}(z).$$

д) Иначе, пусть $p'_1 = 1$.

Согласно лемме 4.19 из неравенства $A \leq C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(\delta(H_l)) + 1$ вытекает, что либо $l \geq s'_1 - 1$, либо $A = C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(\delta(H_l)) + 1$ и $d_H(z) = 0$. Вторая возможность означает, что $d, d_W(z) \in M_\tau(p'_1, \ldots, p'_{q'})$ при некотором τ.

Действительно, если $l \leq s'_1 - 2$, то $d_H(z) = 0$ — иначе $C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H) \geq C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(\delta(H_l)) + 2$. Тогда $C^{W}_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(z) - C^{H}_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(z) \geq 1$, при этом равенство возможно лишь в случае $d, d_W(z) \in M_\tau(p'_1, \ldots, p'_{q'})$ при некотором τ. Одновременно это означает, что $A = C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(\delta(H_l)) + 1$.

д.1) Пусть $A = C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(\delta(H_l)) + 1$ и $d_H(z) = 0$. Напомним, что в этом случае $\rho_1 \in G$.

Рассмотрим граф G', получаемый из G подстановкой W^- вместо W, см. рис. 13в. Имеем $\sigma_1 = -1$, $\sigma_3 = 1$. Из отмеченного выше следует, что $d_G(z'), d_G(z) \in M_t(p_1, \ldots, p_q)$ при некотором t. Тогда σ_2 выражается формулой

$$\sigma_2 = C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z^*) - C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z^*) = -c^{G}_{p_1,s_1,\ldots,p_q,s_q}(z^*, t+1) \le 0.$$

д.1.1) Если $\sigma_2 < 0$, то удалим ребро, выходящее из вершины z^*, если оно имеется, методом леммы 4.12, и выполним подстановку W^- вместо W.

д.1.2) Если $\sigma_2 = 0$, тогда $d_G(z^*), d_{G'}(z^*) \in M_{t'}(p_1, \ldots, p_q)$ при некотором t'. Выполним подстановку W^- вместо W в графе G методом леммы 4.13 так же, как в п. г.2.2).

д.2) Пусть $l \ge s'_1 - 1$.

Рассмотрим граф G', получаемый из G подстановкой $W^\times$ вместо W (см. рис. 13г: граф H^1_l состоит из пучка ребер, входящих в корневую вершину, и изолированных остальных вершин).

В этом случае $\sigma_3 \in \{0, -1\}$. Для оценки σ_2 укажем несколько соотношений. По построению,

$$C_{p_1,s_1,\ldots,p_q,s_q}(H'_{z^*}) - C_{p_1,s_1,\ldots,p_q,s_q}(H_{z^*}) = C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H^1_l) - C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H^*)$$

и, как легко проверить,

$$C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H^1_l) = C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H^0_l) + (s'_1 - l).$$

Пусть $d_{H^*}(z^*) \in M_t(p'_1, \ldots, p'_{q'})$. Согласно лемме 4.19 в случае $t \ge 1$:

$$C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H^*) \ge C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H^1_l) + \sum_{i=2}^{t} s'_i, \tag{4.10}$$

а в случае $t = 0$ (т. е. $d_{H^*}(z^*) = 0$):

$$C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H^*) \ge C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(H^1_l) - (s'_1 - l). \tag{4.11}$$

Лемма 4.19 позволяет сделать следующие выводы. Если $\sigma_1 = 0$, т. е. $A = C_{p'_1,s'_1,\ldots,p'_{q'},s'_{q'}}(\delta(H_l))$, то $l = s'_1$ и $d_W(z) \in M_1(p'_1, \ldots, p'_q)$. Если $\sigma_1 = -1$,

то, считая ввиду п. д.1), что $d_H(z) > 0$, есть три возможных ситуации: либо $d_W(z) \in M_1(p'_1, \ldots, p'_q)$, либо $l = s'_1$ и $d, d_W(z) \in M_\tau(p'_1, \ldots, p'_{q'})$ при некотором τ, либо $s'_2 = 1$, $l = s'_1$ и $d_W(z) \in M_2(p'_1, \ldots, p'_q)$. Сравнивая компоненты стоимости вершины z^* в графах G и G', получаем справедливую во всех четырех случаях оценку

$$C^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z^*) \le \sum_{i=1}^{r-1} s_i + s'_1 + \sum_{i=2}^{t} s'_i + \sum_{i \ge r+t} c^{G}_{p_1,s_1,\ldots,p_q,s_q}(z^*, i) \le$$

$$\le \sum_{i=1}^{r-1} s_i + C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z^*) - C^{H_{z^*}}_{p_1,s_1,\ldots,p_q,s_q}(z^*) + s'_1 + \sum_{i=2}^{t} s'_i =$$

$$= C^{G}_{p_1,s_1,\ldots,p_q,s_q}(z^*) - C^{H_{z^*}}_{p_1,s_1,\ldots,p_q,s_q}(z^*) + C^{H'_{z^*}}_{p_1,s_1,\ldots,p_q,s_q}(z^*) + \sum_{i=2}^{t} s'_i. \quad (4.12)$$

С учетом соотношений (4.10), (4.11) из (4.12) следует: $\sigma_2 \le s'_1 - l$ в первых трех случаях и $\sigma_2 \le 0$ в четвертом случае (когда $\sigma_1 = -1$, $s'_2 = 1$, $l = s'_1$ и $d_W(z) \in M_2(p'_1, \ldots, p'_q)$).

Следовательно, $\sigma_1 + \sigma_2 + \sigma_3 \le 0$, причем равенство возможно только в случае $\sigma_2 = s'_1 - l = -\sigma_1$ и $\sigma_3 = 0$, что означает: $\rho_1 \in G$ и либо $\sigma_1 = 0$, либо $\sigma_1 = -1$ и $\tau = 1$.

д.2.1) Пусть $\sigma_1 + \sigma_2 + \sigma_3 < 0$. Тогда удалим ребро, выходящее из вершины z^*, если оно имеется, методом леммы 4.12, и выполним подстановку $W^\times$ вместо W.

д.2.2) Пусть $\sigma_1 + \sigma_2 + \sigma_3 = 0$, т. е. $\sigma_2 = s'_1 - l = -\sigma_1$ и $\sigma_3 = 0$.

Равенство $\sigma_2 = s'_1 - l$ означает, что неравенство (4.12) обращается в равенство, а это, в свою очередь, влечет $d_{G'}(z^*) \ge \sum_{i=1}^{r+t-1} p_i$ и $c^{G}_{p_1,s_1,\ldots,p_q,s_q}(z^*, i) = c^{G'}_{p_1,s_1,\ldots,p_q,s_q}(z^*, i)$ при любом $i \ge r + t$. Поэтому $d_G(z^*)$, $d_{G'}(z^*) \in M_{t'}(p_1, \ldots, p_q)$ при некотором t'.

В этом случае выполним подстановку $W^\times$ вместо W в графе G методом леммы 4.13. Лемма 4.13 применяется в том же стиле, что и в п. г.2.2). □

Доказанная лемма вместе со следствиями 4.1 и 4.2 влечет справедливость основной леммы 4.3. Тем самым лемма 4.3 доказана.

4.3.7 Собственно нижняя оценка

Теорема 4.1. *Справедлива нижняя оценка сложности*

$$L(2^n) \geq 3.5 \cdot 2^n - (8.5 + 3.5(n \bmod 2))2^{\lfloor n/2 \rfloor} + n + 5.$$

Доказательство. Из леммы 4.1 следует, что минимальная схема S_{2^n} имеет структуру, показанную на рис. 9. Она включает аналогичную подсхему $S_{2^{n-1}}$, реализующую с минимальной глубиной $n-1$ префиксные суммы переменных $x_{2^{n-1}+1}, \ldots, x_{2^n}$, подсхему $S^1_{2^{n-1}}$, реализующую с глубиной n префиксные суммы переменных $x_1, \ldots, x_{2^{n-1}}$, а также 2^{n-1} элементов-выходов. Имеем,

$$L(2^n) \geq L(2^{n-1}) + L\left(S^1_{2^{n-1}}\right) + 2^{n-1}.$$

Подсхема $S^1_{2^{n-1}}$ содержит $2^{n-1}-1$ элементов-выходов и $2^{n-1}-n$ каркасных элементов. Остается оценить количество избыточных элементов в этой подсхеме. Используя лемму 4.3, покажем, что оно не меньше, чем мощность множества

$$\{(N, R, k) \mid N < 2^k,\ N \text{ нечетно},\ R < 2^{n-2k-1},\ R \text{ — не степень двойки}\} \cap \mathbb{N}^3. \tag{4.13}$$

Будем применять лемму 4.3, последовательно увеличивая параметр k. Пусть $k = 1$, тогда при подходящем выборе параметров $N = 1$ (других вариантов в этом случае нет) и R лемма гарантирует наличие одного избыточного элемента с правым концом метки в соответствующем интервале из двух значений.

Далее, каждая тройка $(N = 2N' + n_0, R, k > 1)$, где $n_0 \in \{0, 1\}$, удовлетворяющая условиям леммы 4.3, вбирает информацию о существовании $2N'$ избыточных элементов с правыми концами меток из интервала $J_{N,R,k}$ (см. (4.3)), которые учитывались тройками с меньшими значениями первого параметра, и в случае $n_0 = 1$ добавляет информацию о существовании еще одного элемента с правым концом метки из рассматриваемого интервала. Действительно, тройками $(N', R' = n_0 2^{n-2k-2} + 2R, k-1)$ и $(N', R'+1, k-1)$ учитывалось по N' избыточных элементов в каждом из интервалов $J_{N',R',k-1}$ и $J_{N',R'+1,k-1}$, в объединении дающих рассматриваемый интервал.

Следовательно, можно установить взаимно однозначное соответствие между тройками (N, R, k), в которых числа N нечетны, и некоторым подмножеством множества избыточных элементов схемы $S^1_{2^{n-1}}$. Тем самым, задача сведена к вычислению мощности множества (4.13).

Легко видеть, что для заданного k существуют 2^{k-1} способов выбора нечетного числа N и независимо $2^{n-2k-1} - n + 2k$ способов выбора параметра R, где полагается, что $k < \lfloor n/2 \rfloor$. Следовательно, число избыточных элементов в подсхеме $S^1_{2^{n-1}}$ можно оценить снизу как

$$\sum_{k=1}^{\lfloor n/2 \rfloor - 1} 2^{k-1} \left(2^{n-2k-1} - n + 2k\right) =$$

$$= \sum_{k=1}^{\lfloor n/2 \rfloor - 1} 2^{n-k-2} - n \sum_{k=1}^{\lfloor n/2 \rfloor - 1} 2^{k-1} + \sum_{k=1}^{\lfloor n/2 \rfloor - 1} k2^k =$$

$$= \left(2^{n-2} - 2^{n-\lfloor n/2 \rfloor - 1}\right) - n\left(2^{\lfloor n/2 \rfloor - 1} - 1\right) + (\lfloor n/2 \rfloor - 2)2^{\lfloor n/2 \rfloor} + 2 =$$

$$= 2^{n-2} - (2.5 + (n \bmod 2))2^{\lfloor n/2 \rfloor} + n + 2.$$

Получаем рекуррентное соотношение для $L(2^n)$ в виде

$$L(2^n) \geq L(2^{n-1}) + 3.5 \cdot 2^{n-1} - (2.5 + (n \bmod 2))2^{\lfloor n/2 \rfloor} + 1. \qquad (4.14)$$

Очевидно, $L(1) = 0$, поэтому для $n = 0$ утверждение теоремы верно. Остается проверить индуктивный переход.

$$L(2^n) \geq 3.5 \cdot 2^{n-1} - (12 - 3.5(n \bmod 2))2^{\lceil n/2 \rceil - 1} + n + 4 +$$

$$+ 3.5 \cdot 2^{n-1} - (2.5 + (n \bmod 2))2^{\lfloor n/2 \rfloor} + 1 =$$

$$= 3.5 \cdot 2^n - (12 - 3.5(n \bmod 2))2^{\lceil n/2 \rceil - 1} - (2.5 + (n \bmod 2))2^{\lfloor n/2 \rfloor} + n + 5.$$

Легко проверяется тождество

$$(12 - 3.5(n \bmod 2))2^{\lceil n/2 \rceil - 1} = (6 + 2.5(n \bmod 2))2^{\lfloor n/2 \rfloor},$$

которым устанавливается справедливость индуктивного перехода и, как следствие, утверждения теоремы. □

4.4 Верхняя оценка

Покажем, что оценка теоремы 4.1 является точной. Для этого предъявим оптимальный способ построения схем $S^1_{2^k}$, который фактически является модификацией метода [127].

Через Q_{2^k} обозначим минимальную (т. е. не содержащую избыточных элементов) префиксную схему порядка 2^k, реализующую максимальный префикс $x_1 \circ \ldots \circ x_{2^k}$ на глубине k и имеющую сложность $2^{k+1} - k - 2$ и глубину $2k - 2$. Такую схему несложно построить, см. [156, 127].

Для $i = 1, \ldots, \lceil n/2 \rceil$ положим $l_i = 2^n - 2^{n+1-i}$ и $l_{\lceil n/2 \rceil + 1} = 2^n$. Также для $i = 1, \ldots, \lceil n/2 \rceil - 1$ положим $m_i = n - 2i$ и $m_{\lceil n/2 \rceil} = 1 - (n \bmod 2)$. Определим семейство схем $S^1_{2^n}$ так, как показано на рис. 14–15.

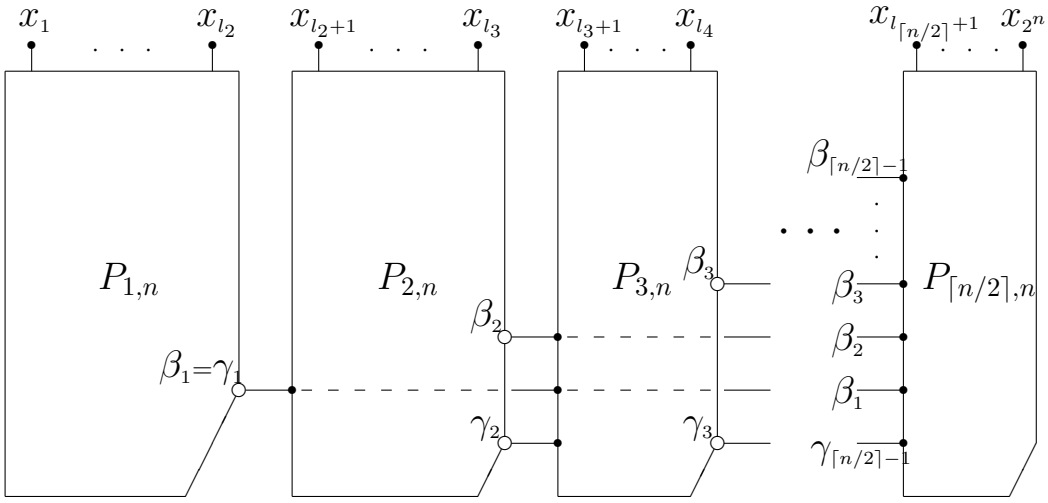


Рис. 14: Структура схемы $S^1_{2^n}$

Сделаем необходимые пояснения к рисункам. Схема $S^1_{2^n}$ (рис. 14) состоит из $\lceil n/2 \rceil$ подсхем, обозначенных $P_{i,n}$. В подсхеме $P_{i,n}$ расположены выходы схемы $S^1_{2^n}$ с правыми концами меток в интервале от $l_i + 1$ до l_{i+1}. В каждой такой подсхеме вычисляется также функция $\beta_i = x_{l_i+1} \circ \ldots \circ x_{l_{i+1}}$, которая используется подсхемами $P_{j,n}$, где $j > i$.

Если $i > 1$, то входами схемы $P_{i,n}$ помимо входов переменных также являются функции $\beta_1, \ldots, \beta_{i-1}$ и $\gamma_{i-1} = x_1 \circ \ldots \circ x_{l_i}$.

В подсхеме $P_{i,n}$ входы переменных разбиваются на группы: в первых двух группах по 2^{i-1} входов, в остальных — по 2^i (см. рис. 15). В каждой

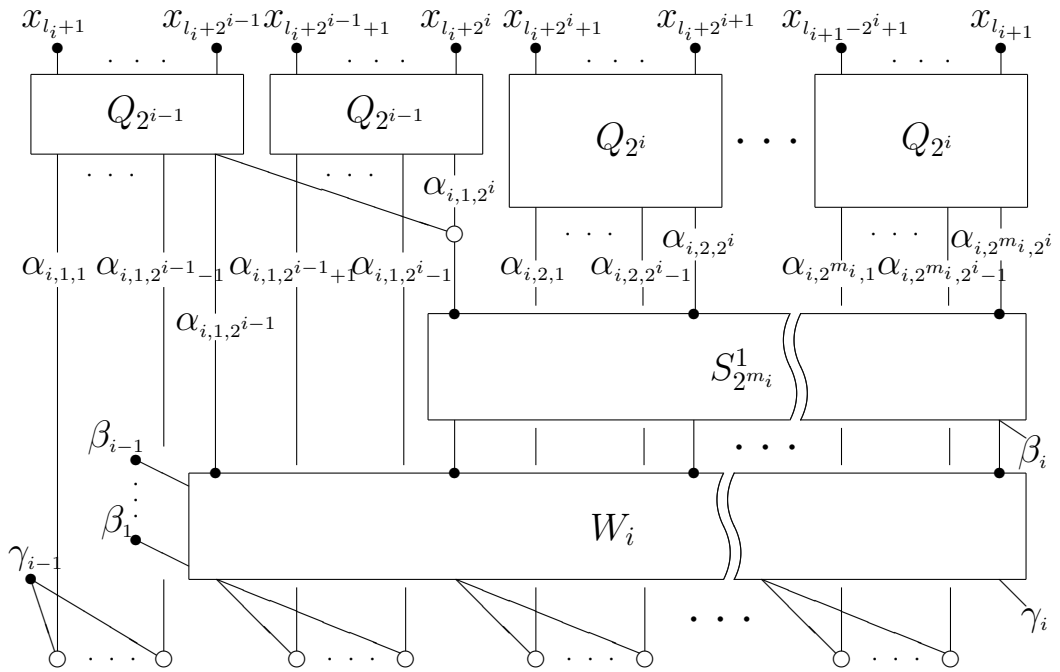


Рис. 15: Структура схемы $P_{i,n}$

из групп вычисляются префиксные суммы при помощи схем $Q_{2^{i-1}}$ и Q_{2^i}. Эти суммы обозначены через $\alpha_{i,j,k}$, где

$$\alpha_{i,j,k} = \begin{cases} x_{l_i+2^{i-1}+1} \circ \ldots \circ x_{l_i+k}, & j = 1,\ k > 2^{i-1} \\ x_{l_i+(j-1)2^i+1} \circ \ldots \circ x_{l_i+(j-1)2^i+k}, & \text{иначе} \end{cases}.$$

Любая из функций $\alpha_{i,j,k}$ вычисляется на глубине не выше $2i-2 \le n$.

Выходы элементов, реализующих функции $\alpha_{i,1,2^{i-1}} \circ \alpha_{i,1,2^i}$ и $\alpha_{i,j,2^i}$, где $j > 1$ (все элементы расположены на глубине i), подаются на входы подсхемы $S^1_{2^{m_i}}$. Соответствующий максимальному префиксу выход этой подсхемы реализует функцию β_i. Таким образом, при любом $i < \lceil n/2 \rceil$ функция β_i вычисляется на глубине, не превосходящей $i + m_i = n - i$, а функция $\beta_{\lceil n/2 \rceil}$ вычисляется на глубине $\lceil n/2 \rceil + m_{\lceil n/2 \rceil} = n - (\lceil n/2 \rceil - 1)$.

Выходы элемента, реализующего функцию $\alpha_{i,0} = \alpha_{i,1,2^{i-1}}$, и выходы $\alpha_{i,1}, \ldots, \alpha_{i,2^{m_i}}$ подсхемы $S^1_{2^{m_i}}$ подаются на входы подсхемы W_i, которая вычисляет функции $\alpha_{i,k} \circ \beta_1 \circ \ldots \circ \beta_{i-1}$, в том числе γ_i (при $k = 2^{m_i}$) на глубине не выше n. Ясно, что на выходах подсхемы W_i реализуются функции $x_1 \circ \ldots \circ x_k$, где $k \in \{l_i + 2^{i-1}\} \cup \{l_i + j2^i \mid j = 1, \ldots, 2^{m_i}\}$.

Следовательно, глубина любой из подсхем $P_{i,n}$ не превосходит $n+1$.

Лемма 4.24. *Сложность схемы* $S^1_{2^n}$ *равна*

$$L(S^1_{2^n}) = 5 \cdot 2^{n-1} - (3.5 - (n \bmod 2))2^{\lceil n/2 \rceil} + 1.$$

Доказательство. Вычислим сложность схемы $P_{i,n}$:

$$\begin{aligned} L(P_{i,n}) &= L(S^1_{2^{m_i}}) + 2L(Q_{2^{i-1}}) + (2^{m_i} - 1)L(Q_{2^i}) + 1 + \\ &\quad + (2^{m_i} + 1)(i-1) + 2(2^{i-1} - 1) + (2^{m_i} - 1)(2^i - 1) = \\ &= L(S^1_{2^{m_i}}) + 2(2^i - i - 1) + (2^{m_i} - 1)(2^{i+1} - i - 2) + \\ &\quad + (2^{m_i} + 1)(i-1) + 2^{m_i}(2^i - 1) = \\ &= L(S^1_{2^{m_i}}) + 2^{m_i}(3 \cdot 2^i - 4) - 1. \end{aligned}$$

Используя соотношения $L(S^1_1) = 0$ и $L(S^1_2) = 1$ в качестве базы индукции, докажем индуктивный переход.

$$\begin{aligned} L(S^1_{2^n}) &= \sum_{i=1}^{\lceil n/2 \rceil} L(P_{i,n}) = \sum_{i=1}^{\lceil n/2 \rceil} \left(L(S^1_{2^{m_i}}) + 2^{m_i}(3 \cdot 2^i - 4) - 1 \right) = \\ &= \sum_{i=1}^{\lceil n/2 \rceil} \left(5 \cdot 2^{m_i - 1} - (3.5 - (m_i \bmod 2))2^{\lceil m_i/2 \rceil} + 2^{m_i}(3 \cdot 2^i - 4) \right) = \\ &= \sum_{i=1}^{\lceil n/2 \rceil} \left(3(2^{m_i + i} - 2^{m_i - 1}) - (3.5 - (m_i \bmod 2))2^{\lceil m_i/2 \rceil} \right). \end{aligned}$$

Подставляя $m_i = n - 2i$, отдельно вычислим сумму первых $\lceil n/2 \rceil - 1$ слагаемых:

$$\begin{aligned} \Sigma_1 &= \sum_{i=1}^{\lceil n/2 \rceil - 1} \left(3(2^{m_i + i} - 2^{m_i - 1}) - (3.5 - (m_i \bmod 2))2^{\lceil m_i/2 \rceil} \right) = \\ &= \sum_{i=1}^{\lceil n/2 \rceil - 1} 3(2^{n-i} - 2^{n-2i-1}) - \sum_{i=1}^{\lceil n/2 \rceil - 1} (3.5 - (n \bmod 2))2^{\lceil n/2 \rceil - i} = \\ &= 3(2^n - 2^{\lfloor n/2 \rfloor + 1}) - \left(2^{n-1} - 2^{1 - (n \bmod 2)} \right) - (3.5 - (n \bmod 2))(2^{\lceil n/2 \rceil} - 2) = \\ &= 5 \cdot 2^{n-1} - (3.5 - (n \bmod 2))2^{\lceil n/2 \rceil} - 3 \cdot 2^{\lfloor n/2 \rfloor + 1} + 9 - 3(n \bmod 2). \end{aligned}$$

Используя $m_{\lceil n/2 \rceil} = 1 - (n \bmod 2)$, последнее слагаемое исходной суммы вычисляется как

$$\Sigma_2 = 3\left(2^{\lceil n/2 \rceil + 1 - (n \bmod 2)} - 2^{-(n \bmod 2)}\right) - (2.5 + (n \bmod 2))2^{1-(n \bmod 2)} =$$
$$= 3 \cdot 2^{\lfloor n/2 \rfloor + 1} - (4 + (n \bmod 2))2^{1-(n \bmod 2)} = 3 \cdot 2^{\lfloor n/2 \rfloor + 1} - 8 + 3(n \bmod 2).$$

Складывая Σ_1 и Σ_2, получаем требуемое соотношение. □

Теорема 4.2. *Справедлива верхняя оценка сложности*

$$L(2^n) \leq 3.5 \cdot 2^n - (8.5 + 3.5(n \bmod 2))2^{\lfloor n/2 \rfloor} + n + 5.$$

Доказательство. Воспользуемся следствием из леммы 4.1 (см. рис. 9):

$$L(2^n) \leq L(2^{n-1}) + L(S^1_{2^{n-1}}) + 2^{n-1}.$$

Используя построенные выше схемы $S^1_{2^k}$, для которых справедлива оценка леммы 4.24, получаем рекуррентное соотношение

$$L(2^n) \leq L(2^{n-1}) + 3.5 \cdot 2^{n-1} - (2.5 + (n \bmod 2))2^{\lfloor n/2 \rfloor} + 1,$$

которое разрешается как (4.14) с точностью до знака неравенства. □

Вместе теоремы 4.1 и 4.2 определяют точное значение сложности минимальной префиксной схемы порядка 2^n и глубины n. Схемы Ладнера—Фишера [156] оказываются неминимальными, начиная с $n = 6$. Недавно последовательность префиксных схем порядка 2^n, глубины n и сложности $L(2^n)$ (по крайней мере, при $n \leq 25$) была обнаружена М. Ширан [197] компьютерным счетом.

Выделяя из построенной схемы подсхему, зависящую от m первых переменных, где $2^{n-1} < m \leq 2^n$, получаем

Следствие 4.3. *Для любого m справедливо $L(m) \leq (3.5 - o(1))m$.*

4.5 Реализация с почти минимальной глубиной

Теорема 4.3. *При $1 \leq k \leq n-2$ имеет место соотношение*

$$L'(2^n, k) = \left(2 + 2^{-k}\right)2^n - (5 + 2((n-k) \bmod 2))2^{\lfloor (n-k)/2 \rfloor} - k + 2.$$

Доказательство. Для доказательства нижней оценки, как следует из леммы 4.1, достаточно рассмотреть подсхему схемы $S_{2^{n+k}}$, зависящую от первых 2^n переменных.

Верхнюю оценку в случае $k = 1$ доставляет схема $S^1_{2^n}$, построенная в §4.4. При помощи этой схемы методом [156] строятся минимальные схемы $S^k_{2^n}$ для остальных значений k (см. рис. 16). □

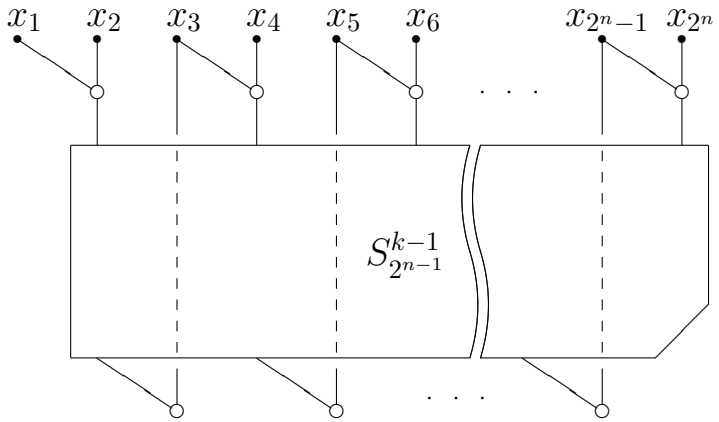


Рис. 16: Структура схемы $S^k_{2^n}$

Для произвольного m верхняя оценка получается так же, как и в следствии 4.3.

Следствие 4.4. *Для любых m, k, где $1 \le k \le \lceil \log_2 m \rceil - 2$, при $m \to \infty$ справедливо*

$$L'(m,k) \le (2 + 2^{-k} - o(1))m.$$

4.6 Префиксные XOR-схемы

В этом параграфе будет показано, что при использовании некоторых дополнительных свойств операции $\circ$ префиксные суммы могут быть вычислены параллельной схемой меньшей сложности, нежели в общем случае. В качестве примера рассмотрим ассоциативную операцию $\oplus$ с аксиомой $x \oplus y \oplus y = x$. Если $\oplus$ является групповой операцией, то ее можно интерпретировать как операцию сложения по модулю 2.

Преимуществом операции $\oplus$ является возможность вычисления функции $x_i \oplus \ldots \oplus x_j$ как

$$x_i \oplus \ldots \oplus x_j = (x_i \oplus \ldots \oplus x_{j+k}) \oplus (x_{j+1} \oplus \ldots \oplus x_{j+k}).$$

Префиксную схему Λ_{2^n} порядка 2^n и глубины n над базисом $\{\oplus\}$ построим согласно рис. 9, используя вместо схемы $S^1_{2^n}$ схему $\Lambda^1_{2^n}$. Конструкция схемы $\Lambda^1_{2^n}$ в основном аналогична конструкции схемы $S^1_{2^n}$, описанной в §4.4. Отличия следующие.

Вместо схем Q_{2^k} используем схемы Ψ_{2^k}. Схема Ψ_{2^k} имеет сложность $2^{k+1} - 2k - 1$ и глубину $k + \lceil k/2 \rceil - 1$ и вычисляет суммы $x_1 \oplus \ldots \oplus x_i$ и $x_{2^{k-1}+i} \oplus \ldots \oplus x_{2^k}$, где $i = 1, \ldots, 2^{k-1}$, а также $x_1 \oplus \ldots \oplus x_{2^k}$, причем последнюю сумму — на глубине k. Способ построения такой схемы показан на рис. 17.

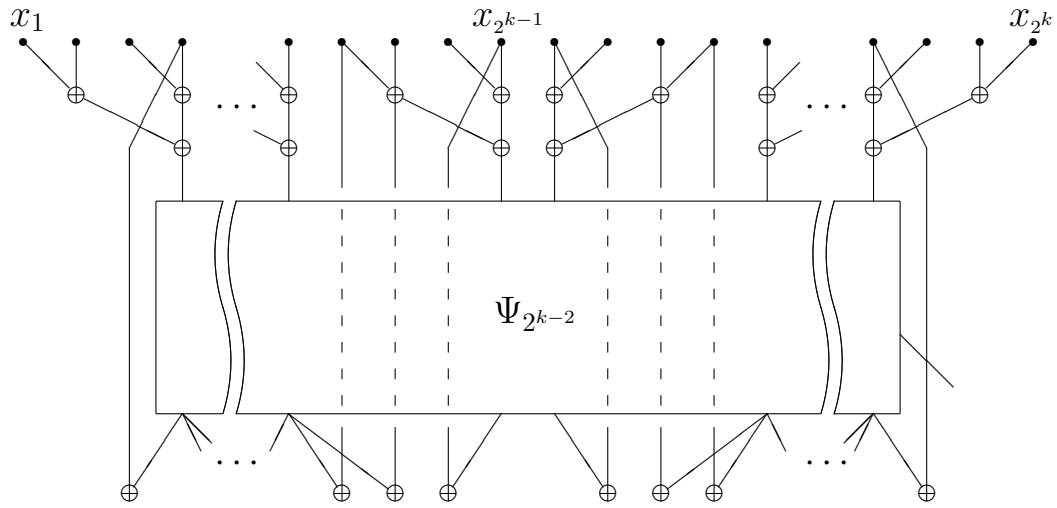


Рис. 17: Структура схемы Ψ_{2^k}

Для $i = 1, \ldots, n - \lceil n/3 \rceil$ обозначим $\lambda_i = 2^n - 2^{n+1-i}$ и положим $\lambda_{n-\lceil n/3 \rceil+1} = 2^n$. Схема $\Lambda^1_{2^n}$ состоит из подсхем $\Pi_{i,n}$, где $i = 1, \ldots, n - \lceil n/3 \rceil$, которые устроены и соединены так же, как схемы $P_{i,n}$ на рис. 14.

В подсхеме $\Pi_{i,n}$ расположены выходы схемы $\Lambda^1_{2^n}$ с правыми концами меток в интервале от λ_i+1 до λ_{i+1}. Кроме того, в подсхеме $\Pi_{i,n}$ вычисляется функция $\beta_i = x_{\lambda_i+1} \oplus \ldots \oplus x_{\lambda_{i+1}}$. Если $i > 1$, то входами схемы $\Pi_{i,n}$ помимо

входов переменных также являются функции $\beta_1, \ldots, \beta_{i-1}$ и $\gamma_{i-1} = x_1 \oplus \ldots \oplus x_{\lambda_i}$.

Структура подсхемы $\Pi_{i,n}$ с незначительными отличиями повторяет структуру подсхемы $P_{i,n}$ (см. рис. 15). Входы переменных подсхемы разбиваются на группы: в первых двух группах по $2^{\lceil i/2 \rceil}$ входов, в остальных — по $2^{\lceil i/2 \rceil + 1}$. К каждой из групп присоединена соответствующая схема из семейства Ψ_{2^k}. Выходы этих подсхем, реализующие максимальные префиксы, присоединены ко входам подсхемы $\Lambda^1_{2^{\mu_i}}$, где $\mu_i = n - i - \lceil i/2 \rceil - 1$ для $i < n - \lceil n/3 \rceil$, и $\mu_{n - \lceil n/3 \rceil} = 2\lceil n/3 \rceil + \lfloor n/3 \rfloor - n$.

Последним отличием схемы $\Pi_{i,n}$ от схемы $P_{i,n}$ является ее заключительный уровень, на котором вычисляются выходные функции $x_1 \oplus \ldots \oplus x_j$ (он соответствует пучкам элементов, расположенных снизу на рис. 15). Если переменная x_j подается на вход подсхемы Ψ_{2^k}, то функция $x_1 \oplus \ldots \oplus x_j$ вычисляется как (обозначим $r = j \bmod 2^k$)

$$\begin{cases} (x_1 \oplus \ldots \oplus x_{j-r}) \oplus (x_{j-r+1} \oplus \ldots \oplus x_j), & r \le 2^{k-1} \\ (x_1 \oplus \ldots \oplus x_{j+2^k-r}) \oplus (x_{j+1} \oplus \ldots \oplus x_{j+2^k-r}), & r > 2^{k-1} \end{cases},$$

где второе «слагаемое» реализуется на выходе подсхемы Ψ_{2^k}.

Лемма 4.25. *Сложность схемы $\Lambda^1_{2^n}$ равна*

$$L(\Lambda^1_{2^n}) = 2\tfrac{3}{11} \cdot 2^n - \sigma_n,$$

где σ_n определяется из рекуррентного соотношения

$$\sigma_n = 2\sigma_{n-3} + \sigma_{n-4} + 1$$

с начальными условиями

$$\sigma_0 = \tfrac{25}{11}, \quad \sigma_1 = \tfrac{39}{11}, \quad \sigma_2 = \tfrac{56}{11}, \quad \sigma_3 = \tfrac{79}{11}.$$

Замечание. Разумеется, σ_n можно задать явно аналитической формулой, но она будет слишком громоздкой. Для этого следует положить $\sigma_n = \chi_n - 0.5$, а χ_n определить из рекуррентного соотношения $\chi_n = 2\chi_{n-3} + \chi_{n-4}$ с соответствующими начальными условиями. Решение этого соотношения

может быть получено в виде линейной комбинации степеней корней многочлена $x^4 - 2x - 1$. В частности, справедливо $\sigma_n \sim c\chi^n$, где $\chi = 1.3953\ldots$ — максимальный по абсолютной величине корень, а $c = 2.86\ldots$

Доказательство. Вычислим сложность схемы $\Pi_{i,n}$ для $i > 1$:

$$\begin{aligned}
L(\Pi_{i,n}) &= L(\Lambda^1_{2^{\mu_i}}) + 2L(\Psi_{2^{\lceil i/2\rceil}}) + (2^{\mu_i} - 1)L(\Psi_{\lceil i/2\rceil+1}) + 1 + \\
&\quad + (2^{\mu_i} + 1)(i-1) + 2\left(2^{\lceil i/2\rceil} - 1\right) + (2^{\mu_i} - 1)\left(2^{\lceil i/2\rceil+1} - 1\right) = \\
&= L(\Lambda^1_{2^{\mu_i}}) + 2\left(2^{\lceil i/2\rceil+1} - 2\lceil i/2\rceil - 1\right) + (2^{\mu_i} - 1)\left(2^{\lceil i/2\rceil+2} - 2\lceil i/2\rceil - 3\right) + \\
&\quad + (2^{\mu_i} + 1)(i-1) + 2^{\mu_i}\left(2^{\lceil i/2\rceil+1} - 1\right) = \\
&\quad = L(\Lambda^1_{2^{\mu_i}}) + 2^{\mu_i}\left(3 \cdot 2^{\lceil i/2\rceil+1} - (i \bmod 2) - 5\right) - (i \bmod 2).
\end{aligned}$$

Сложность схемы $\Pi_{1,n}$ выражается на 1 меньшим значением, что объясняется отсутствием элемента $\gamma_0 \oplus x_1$ (см. рис. 15).

Несложно проверить, что $L(\Lambda^1_{2^0}) = 0$, $L(\Lambda^1_{2^1}) = 1$, $L(\Lambda^1_{2^2}) = 4$, $L(\Lambda^1_{2^3}) = 11$. Используя эти соотношения в качестве базы индукции, докажем индуктивный переход.

С целью удобства записи введем обозначение $\omega_3(n)$ для неопределенной функции, зависящей только от $n \bmod 3$. В соответствии с определением для любого целого k справедливо $\omega_3(n) = \omega_3(n + 3k)$.

$$\begin{aligned}
L(\Lambda^1_{2^n}) &= \sum_{i=1}^{n-\lceil \frac{n}{3}\rceil} L(\Pi_{i,n}) = \\
&= \sum_{i=1}^{n-\lceil \frac{n}{3}\rceil} \left(L(\Lambda^1_{2^{\mu_i}}) + 2^{\mu_i}\left(3 \cdot 2^{\lceil i/2\rceil+1} - (i \bmod 2) - 5\right) - (i \bmod 2)\right) - 1 = \\
&= \sum_{i=1}^{n-\lceil \frac{n}{3}\rceil} \left(2^{\mu_i}\left(3 \cdot 2^{\lceil i/2\rceil+1} - (i \bmod 2) - 2\tfrac{8}{11}\right) - \sigma_{\mu_i} - (i \bmod 2)\right) - 1.
\end{aligned}$$

Подставляя $\mu_i = n - i - \lceil i/2\rceil - 1$, отдельно вычислим сумму всех слагаемых, кроме последнего:

$$\Sigma_1 = \Sigma_1^1 - \Sigma_1^2 - \Sigma_1^3 - \Sigma_1^4 - \Sigma_1^5 - 1,$$

где

$$\Sigma_1^1 = \sum_{i=1}^{n-\lceil \frac{n}{3} \rceil - 1} 3 \cdot 2^{\mu_i + \lceil i/2 \rceil + 1} = 3 \sum_{i=1}^{n-\lceil \frac{n}{3} \rceil - 1} 2^{n-i} = 3(2^n - 2^{\lceil n/3 \rceil + 1}),$$

$$\Sigma_1^2 = \sum_{i=1}^{n-\lceil \frac{n}{3} \rceil - 1} (i \bmod 2) 2^{\mu_i} = \sum_{i=1}^{n-\lceil \frac{n}{3} \rceil - 1} (i \bmod 2) 2^{n-i-\lceil i/2 \rceil - 1} =$$

$$= \sum_{j=1}^{\lfloor n/3 \rfloor} 2^{n-3j} = \tfrac{1}{7} \cdot 2^n + \omega_3(n),$$

$$\Sigma_1^3 = \sum_{i=1}^{n-\lceil \frac{n}{3} \rceil - 1} 2\tfrac{8}{11} \cdot 2^{\mu_i} = 2\tfrac{8}{11} \sum_{i=1}^{n-\lceil \frac{n}{3} \rceil - 1} 2^{n-i-\lceil i/2 \rceil - 1} =$$

$$= 2\tfrac{8}{11} \left(\sum_{j=1}^{\lfloor n/3 \rfloor} 2^{n-3j} + \sum_{j=1}^{n-\lceil \frac{n}{3} \rceil - \lfloor \frac{n}{3} \rfloor - 1} 2^{n-3j-1} \right) =$$

$$= 2\tfrac{8}{11} \left(\tfrac{1}{7} \cdot 2^n + \tfrac{1}{7} \cdot 2^{n-1} + \omega_3(n) \right) = \tfrac{45}{77} \cdot 2^n + \omega_3(n),$$

$$\Sigma_1^4 = \sum_{i=1}^{n-\lceil \frac{n}{3} \rceil - 1} \sigma_{\mu_i} = \sum_{i=1}^{n-\lceil \frac{n}{3} \rceil - 1} \sigma_{n-i-\lceil i/2 \rceil - 1}, \quad \Sigma_1^5 = \sum_{i=1}^{n-\lceil \frac{n}{3} \rceil - 1} (i \bmod 2) = \lfloor n/3 \rfloor.$$

Учитывая $\mu_{n-\lceil n/3 \rceil} = 2\lceil n/3 \rceil + \lfloor n/3 \rfloor - n$, для последнего слагаемого исходной суммы имеем:

$$\Sigma_2 = 3 \cdot 2^{\mu_{n-\lceil n/3 \rceil} + \lceil (n - \lceil n/3 \rceil)/2 \rceil + 1} + \omega_3(n) = 3 \cdot 2^{\lceil n/3 \rceil + 1} + \omega_3(n).$$

Окончательно получаем

$$L(\Lambda_{2^n}^1) = \Sigma_1 + \Sigma_2 = 2\tfrac{3}{11} \cdot 2^n - n/3 - \sum_{i=1}^{n-\lceil \frac{n}{3} \rceil - 1} \sigma_{n-i-\lceil i/2 \rceil - 1} + \psi(n \bmod 3),$$

откуда следует

$$\sigma_n = \sum_{i=1}^{n-\left\lceil \frac{n}{3} \right\rceil - 1} \sigma_{n-i-\lceil i/2 \rceil - 1} + n/3 - \psi(n \bmod 3) =$$

$$= \sigma_{n-3} + \sigma_{n-4} + 1 + \sum_{i=1}^{(n-3)-\left\lceil \frac{n-3}{3} \right\rceil - 1} \sigma_{(n-3)-i-\lceil i/2 \rceil - 1} + \frac{n-3}{3} - \psi((n-3) \bmod 3) =$$

$$= \sigma_{n-3} + \sigma_{n-4} + 1 - L(\Lambda^1_{2^{n-3}}) + 2\tfrac{3}{11} \cdot 2^{n-3} = 2\sigma_{n-3} + \sigma_{n-4} + 1. \quad \square$$

Из леммы следует

Теорема 4.4. *Справедлива верхняя оценка сложности*

$$L^{\oplus}(2^n) \le 3\tfrac{3}{11} \cdot 2^n - \tau_n,$$

где

$$\tau_n = \frac{\sigma_{n+3} + \sigma_{n+2} + \sigma_{n+1} - \sigma_n - n - 7}{2}.$$

Замечание. С учетом замечания к лемме 4.25 можно заключить, что $\tau_n \sim c'\chi^n$, где $c' = 7.235\ldots$

Аналогично следствию 4.3 доказывается

Следствие 4.5. *Для любого m справедливо $L^{\oplus}(m) \le \left(3\frac{3}{11} - o(1)\right) m$.*

В случае $k \ge \lceil (\log_2 m)/2 \rceil - 1$ несложно показать, что $L'^{\oplus}(m,k) = 2m - \lceil \log_2 m \rceil - 2$. Нижняя оценка вытекает из упомянутого в §4.2 соотношения между сложностью и глубиной реализации максимального префикса в схеме [127, 201]. Для доказательства верхней оценки строится схема, в которой входы разбиты на группы по 4; вычисляются суммы всех входов в каждой из групп; они подаются на входы аналогичной схемы для $\lceil m/4 \rceil$ переменных; недостающие префиксы вычисляются с дополнительной глубиной 1 относительно выходов данной подсхемы. Также группируя входы по 4 и в остальном используя метод [156], с учетом результата леммы 4.25 получаем

Следствие 4.6. *Для любых m, k, где $1 \le k \le \lceil (\log_2 m)/2 \rceil - 1$, при $m \to \infty$ справедливо*

$$L'^{\oplus}(m,k) \le \left(2 + \tfrac{3}{11} \cdot 4^{1-k} - o(1)\right) m.$$

Представляет интерес вопрос, может ли верхняя оценка теоремы 4.2 быть улучшена при других предположениях относительно операции $\circ$. Особенный интерес представляют коммутативные операции, идемпотентные операции ($x \circ x = x$) или, если более узко, булевы операции конъюнкции и дизъюнкции.

4.7 Замечания о префиксных схемах с ограниченным ветвлением элементов

Технически обусловленной является задача синтеза префиксных схем с ограниченным ветвлением элементов. (Задается ограничение q на степень ветвления входов и элементов схемы, при этом степень ветвления выходов не должна превышать $q - 1$.) При построении таких схем помимо функциональных элементов $\circ$ используются также тождественные элементы $\triangledown$, которые используются в качестве элементов ветвления. В случае $q = 2$ параллельные префиксные схемы без элементов ветвления строить, вообще говоря, невозможно. Для обозначения сложности префиксных схем с ограничением q на степень ветвления элементов будем использовать введенные ранее обозначения сложности с нижним индексом q.

Сложность параллельных префиксных схем при различных значениях q исследовалась Ф. Фич [127]. Так, для $q \geq 3$ ей были построены схемы минимальной глубины и линейной сложности $O(m)$. Также было доказано, что в случае $q = 2$ функция сложности имеет нелинейный порядок роста, $L_2(m) = \Theta(m \log m)$. В частности, для $m = 2^n$ были получены соотношения

$$(n + 1 - o(1))2^{n-1} \leq L_2(2^n) \leq (3n - 3.5 - o(1))2^{n-1}.$$

Следует заметить, что лучшая верхняя оценка получается более ранним методом Коге—Стоуна 1973 г. [150]. Схема Коге—Стоуна имеет сложность $(n-0.5)2^n$ (в том числе $(n-1)2^n+1$ функциональных элементов $\circ$). Схема порядка 8 изображена на рис. 18.

Отталкиваясь от схемы Коге—Стоуна, методом [156] (как на рис. 16) можно построить префиксную схему порядка 2^n, глубины $n+k$ и сложности

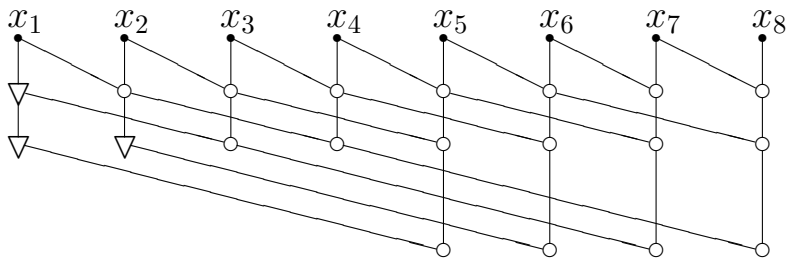


Рис. 18: Схема Коге—Стоуна

$(n-k-3)2^{n-k}+5\cdot 2^{n-1}-k$. Таким образом, имеют место соотношения

$$L_2(2^n)\le (n-0.5)2^n, \qquad L_2'(2^n,k)\le (n-k-3)2^{n-k}+5\cdot 2^{n-1}-k.$$

Вторую оценку можно слегка уточнить, используя модификацию схемы Коге—Стоуна, изображенную на рис. 19 (она имеет сложность также $(n-0.5)2^n$, однако достраивается до схемы с нулевой степенью ветвления выходов меньшей сложности, а именно, $(n-0.25)2^n$). Справедливо

$$L_2'(2^n,k)\le (n-k-3.25)2^{n-k}+5\cdot 2^{n-1}-k. \tag{4.15}$$

(Детали доказательства восстанавливаются несложно.)

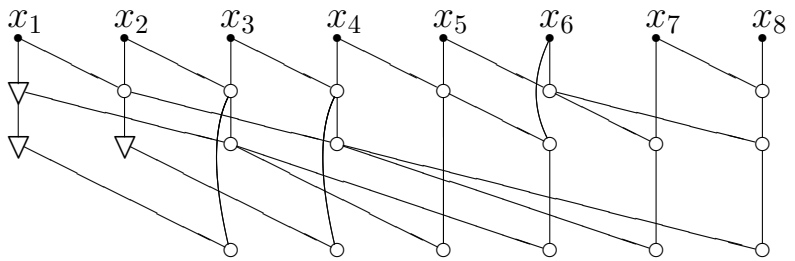


Рис. 19: Модифицированная схема Коге—Стоуна

Оценка (4.15) при малых k примерно в 2 раза отличается от нижней оценки $(n-2)2^{n-k-1}+2^n-O(n(n+k)2^{-k})$ из [127], а при $k\asymp n$ асимптотически в 1.25 раза превосходит стандартную нижнюю оценку 2^{n+1}. В то же время несколько усовершенствованная конструкция из [201] позволяет получить оценку $L_2'(2^n,n-1)\lesssim 2.3\cdot 2^n$, на ее основе может быть построена схема сложности $(2+o(1))2^n$ и глубины $(2+o(1))n$. В целом, асимптотическое поведение функций $L_2(2^n)$ и $L_2'(2^n,k)$ остается пока невыясненным.

5 Схемы ограниченной глубины из многовходовых элементов

5.1 Введение

Основной объект, изучаемый в настоящей главе — схемы из функциональных элементов конъюнкции и дизъюнкции с неограниченным числом входов. Входами схем являются булевы переменные и их отрицания. Для краткости далее будем называть такие схемы *AC-схемами*, мотивируя название тем, что эти схемы используются в определении классов сложности AC^k.[33] *Сложность* схемы определяется как число функциональных элементов в ней, *глубина* — как максимальное число элементов в ориентированном пути от входа к выходу схемы. Через $C_d(n)$ и $C(n)$ обозначим сложность реализации класса булевых функций n переменных (т. е. функцию Шеннона) AC-схемами глубины d и, соответственно, неограниченной глубины.

Дополнительно рассмотрим схемы над бесконечным базисом $U = \{x_1 \vee \ldots \vee x_k,\ x_1 \cdot \ldots \cdot x_k \mid k \in \mathbb{N}\} \cup \{^-\}$ (входами схем являются, как обычно, только булевы переменные). Для функций Шеннона сложности таких схем введем обозначения $C_d^U(n)$ и $C^U(n)$ в зависимости от вида ограничения на глубину. Условимся, что при подсчете глубины пропускаются элементы отрицания.

Для функций Шеннона сложности схем над базисом U, в которых не учитываются элементы отрицания, введем обозначения $C_d^*(n)$ и $C^*(n)$. По сути, такие схемы можно эквивалентным образом определить как схемы над бесконечным базисом $U^\infty = \{(x_1^{\alpha_1} \cdot \ldots \cdot x_k^{\alpha_k})^\beta \mid k \in \mathbb{N}; \alpha_i, \beta \in \{0, 1\}\}$ из обобщенных конъюнкций.

Две последних модели введены для прояснения роли элементов отрицания в оптимальных методах синтеза. Из определений следуют соотношения $C^*(n) \le C^U(n)$, $C^*(n) \le C(n)$, $C^U(n) \le C(n) + n$ (такие же соотношения имеют место при ограничении на глубину).

Под *формулами* будем понимать схемы, в которых выходы функцио-

[33] AC — сокращение от alternating circuits. Обзор некоторых результатов, связанных с классами AC^k, имеется в [145, 210].

нальных элементов не имеют ветвлений. Для функций Шеннона сложности AC-формул введем обозначения $L_d(n)$ и $L(n)$. Обозначим через $L_d^U(n)$ и $L^U(n)$ функции Шеннона сложности формул над базисом U. Понятие формулы с «бесплатными» отрицаниями по существу совпадает с понятием AC-формулы, поскольку любое отрицание может быть «опущено» на уровень входов переменных. Вследствие определения, $L(n) \leq L^U(n)$ и $L_d(n) \leq L_d^U(n)$. Принципиальным преимуществом AC-формул является отсутствие ограничений на ветвление отрицаний переменных.

Задача синтеза с глубиной 2 сводится к построению ДНФ или КНФ. Рассмотрев линейную функцию, можно убедиться в справедливости оценок

$$L_2(n) \sim C_2(n) \sim C_2^U(n) \sim C_2^*(n) \sim 2^{n-1}, \qquad L_2^U(n) \sim n \cdot 2^{n-2}.$$

Вопрос о поведении функций Шеннона сложности для рассматриваемых моделей вычислений становится нетривиальным в глубине 3 и выше.

В работе [42] О. Б. Лупанов показал, что асимптотика функции Шеннона сложности формул достигается на формулах глубины 3 и равна $2^n/\log_2 n$, если под сложностью понимать число вхождений символов переменных в формулу. В работе [47] он установил, что асимптотика функции Шеннона сложности схем из двухвходовых элементов тоже достигается на схемах глубины 3 и составляет $2^n/n$. Под глубиной здесь подразумевается глубина альтернирования — увеличенное на единицу число чередований элементов дизъюнкции и конъюнкции от входов к выходам схемы. Число двухвходовых элементов приблизительно соответствует числу ребер в эквивалентной схеме из многовходовых элементов. Влияние глубины альтернирования на сложность формул изучено С. А. Ложкиным и В. А. Коноводовым в работе [40], где для соответствующих функций Шеннона установлены асимптотические оценки высокой степени точности.

Конкретные оценки сложности функций Шеннона, относящиеся к введенным выше моделям, получены в работе [121]. Их можно записать как

$$1.914 \cdot 2^{n/2} \lesssim C^U(n) \lesssim C_3(n) \lesssim 2.122 \cdot 2^{n/2}, \qquad C_3(2m) \lesssim 2 \cdot 2^m. \qquad (5.1)$$

Кроме того, оценки $C^*(n) \sim \sqrt{2} \cdot 2^{n/2}$ и $L(n) \lesssim 2^n/n$ вытекают из результатов Э. И. Нечипорука [52]. Отметим, что О. М. Касим-Заде [21] получил

универсальную оценку $O(2^{n/2})$ функции Шеннона сложности схем над произвольным бесконечным базисом. Таким образом, рассматриваемые здесь базисы относятся к наиболее слабым по выразительным возможностям. Добавим, что в работах Э. И. Нечипорука [54] и О. Б. Лупанова [45] установлена асимптотика $2 \cdot \sqrt{2^n/n}$ функции Шеннона сложности схем над объемлющим базисом из всех пороговых функций[34].

Автором в работе [224] перечисленные оценки расширены следующим образом:

$$C(n) \sim C_3(n) \sim C_3^U(n) \sim C_3^*(n) \sim 2 \cdot 2^{n/2},$$
$$C^*(n) \sim \sqrt{2} \cdot 2^{n/2}, \qquad 1.944 \cdot 2^{n/2} \lesssim C^U(n),$$
$$0.63 \cdot 2^n/n \lesssim L(n) \lesssim L_3(n) \lesssim 2^n/n,$$
$$L^U(n) \sim L_4^U(n) \sim 2^n/n, \qquad L_3^U(n) \lesssim 2 \cdot 2^n/n.$$

Таким образом, вопрос о сложности AC-схем полностью решен в асимптотическом смысле. Естественно предположить, что на самом деле должно выполняться соотношение $C^U(n) \sim 2 \cdot 2^{n/2}$, но пока удается подойти к этой оценке только сверху.

Содержательно, основным результатом работы [224] являются верхние оценки сложности в глубине 3. В общем, соответствующие схемы и формулы строятся по принципам, известным из классических работ О. Б. Лупанова и Э. И. Нечипорука (например, [42, 47, 49, 52]). С технической стороны, для синтеза указанных схем и формул потребовались специальные покрытия булева куба $\mathbb{B}^n$ асимптотически оптимального размера, обобщающие покрытие сферами. Необходимые покрытия в [224] получены единым способом, который, в частности, приводит к альтернативному доказательству результата А. Е. Липатовой [33] об оптимальном покрытии куба псевдосферами.

Модель схем с бесплатными отрицаниями выделяется тем, что асимптотика сложности не достигается на схемах ограниченной глубины. Можно проверить, что при любом постоянном $d \geq 2$ справедлива оценка $C_{d+1}^*(n) \gtrsim \sqrt{2d/(d-1)} \cdot 2^{n/2}$. Таким образом, оценка $C^*(n) \sim \sqrt{2} \cdot 2^{n/2}$

[34] При любом k имеется $2^{\Theta(k^2)}$ различных пороговых функций k переменных (см., например, [16]), среди которых порядка 3^k конъюнкций и дизъюнкций.

достижима только на схемах растущей глубины. Соответствующий метод получается переложением метода Э. И. Нечипорука [52]. Отправной точкой для метода служит идея многоярусного представления функций, но в полученном в [52] представлении многоярусность выражена не вполне явно. Мы предложим альтернативный способ доказательства, в том числе, явную форму для многоярусного представления функций. Помимо точной оценки в неограниченной глубине, из метода извлекаются верхние оценки вида $C^*_{d+O(1)}(n) \lesssim \sqrt{2d/(d-1)} \cdot 2^{n/2}$.

Отметим, что постановочно к рассматриваемой группе задач примыкают вопросы асимптотического синтеза схем и формул из многовходовых элементов конъюнкции и суммы по модулю 2. Эта область практически совсем не изучена. Прямо переносятся разве что простые методы синтеза формул (глубины 4) над базисом U, позволяя получить асимптотику функции Шеннона сложности $2^n/n$. Нетривиальная верхняя оценка сложности схем (также глубины 4) вытекает из конструкции С. Н. Селезневой [195] ($2.5 \cdot 2^{n/2}$ при нижней оценке $2 \cdot 2^{n/2}$). Кроме того, в ее работах [67, 68] установлен порядок функции Шеннона сложности $\bigoplus\bigwedge\bigoplus$-схем (представлений суммами мультиаффинных функций) $\Theta(2^n/n^2)$.

Изложение в основном следует работе [224]. В §5.2 доказываются нижние оценки функций Шеннона для рассматриваемых моделей. В §5.3 приводятся простые методы синтеза схем оптимальной или почти оптимальной сложности с ослабленным ограничением на глубину. В §5.4 строятся необходимые разбиения булева куба. В §5.5 с помощью этих разбиений получены методы синтеза схем и формул глубины 3. В §5.6 строится многоярусное представление функций и метод синтеза схем над базисом U^∞.

5.2 Нижние оценки сложности

Нижние оценки функционалов сложности для всех рассматриваемых моделей получаются более-менее стандартными мощностными рассуждениями.

Лемма 5.1. $L^U(n) \gtrsim 2^n/n, \quad L(n) \gtrsim \log_3 2 \cdot 2^n/n > 0.63 \cdot 2^n/n.$

Доказательство. Оценим сверху число формул $N^U(s)$ (число AC-формул $N(s)$) n переменных сложности не более s. Для каждого элемента форму-

лы есть не более 2^n (соответственно, 3^n) способов определить набор присоединенных к нему входов переменных (или переменных и их отрицаний). Введем произвольный порядок на множестве таких наборов и упорядочим элементы формулы согласно выбранному порядку. Тогда число способов соединить элементы формулы с литералами есть число сочетаний с повторениями $\binom{2^n+s-1}{s} \le (2^n+s)^s/s!$ (соответственно, $\binom{3^n+s-1}{s} \le (3^n+s)^s/s!$). Кроме того, для каждого элемента не более чем тремя способами можно определить его тип, не более чем s способами можно определить номер элемента, в который ведет исходящее из данного элемента ребро. Еще есть s способов выбрать выход формулы. Таким образом, получаем

$$N^U(s) \le s\cdot(3s(2^n+s))^s/s! \le (c\cdot 2^n)^s, \quad N(s) \le s\cdot(3s(3^n+s))^s/s! \le (c\cdot 3^n)^s$$

для некоторой константы c в предположении $s = o(2^n)$. Требуемые оценки следуют из условия $N^U(s) \ge 2^{2^n}$ (или $N(s) \ge 2^{2^n}$). □

Не исключено, что уточнение результата для AC-формул возможно применением более аккуратной оценки числа функций, обладающих ДНФ (или КНФ) из s компонент (в доказательстве леммы неявно используется грубая оценка $3^{ns}/s!$).

Лемма 5.2. $C(n) \gtrsim 2\cdot 2^{n/2}$.

Доказательство. Оценим число схем $N(a,b)$ от n переменных с a элементами дизъюнкции и b элементами конъюнкции. Пусть на схеме задана естественная нумерация элементов. Есть не более 2^{a+b} способов выбрать типы элементов. Можем считать, что элементы одного типа не соединяются ребрами. Тогда есть всего 2^{ab} возможностей соединить или не соединить пары элементов разных типов (ориентация ребер устанавливается автоматически, согласно нумерации вершин). Также, для каждой вершины имеется не более 3^n возможностей соединения со входами схемы. Получаем оценку $N(a,b) \le 2^{ab}(2\cdot 3^n)^{a+b}$.

Тогда число схем сложности s оценивается как

$$N(s) \le s\cdot \max_a N(a, s-a) \le s\cdot 2^{s^2/4}(2\cdot 3^n)^s,$$

откуда с учетом требования $N(s) \ge 2^{2^n}$ следует утверждение леммы. □

Лемма 5.3. *При любом постоянном $d \geq 2$ справедливо*

$$C^*_{d+1}(n) \gtrsim \sqrt{2d/(d-1)} \cdot 2^{n/2}.$$

Доказательство. Если отрицания бесплатны, то можем считать, что схема состоит из пар: (элемент типа $\vee$ или $\wedge$, его отрицание). Как и выше, можно считать, что в схеме нет ребер, соединяющих элементы одного типа. Дополнительно можно считать, что элементы типа «отрицание конъюнктора» не могут соединяться с дизъюнкторами, и наоборот: отрицания дизъюнкторов не могут соединяться с конъюнкторами. Как следствие, есть только одно «легальное» ребро, которым можно соединить два парных элемента.

Оценим число $N(s_1, \ldots, s_d)$ легальных схем n переменных глубины $d+1$ и сложности $s = s_1 + \ldots + s_d + 1$, в которых на каждом i-м слое расположено s_i (парных) элементов, $i = 1, \ldots, d$. Имеется 2^s способов выбрать типы элементов, $2^{s-1+\sum_{i<j} s_i s_j}$ способов соединить их и 3^{ns} способов присоединить к ним входы переменных. В итоге имеем $N(s_1, \ldots, s_d) \leq 2^{\sum_{i<j} s_i s_j} (2 \cdot 3^n)^s$. Поскольку

$$\sum_{i<j} s_i s_j = \frac{\left(\sum s_i\right)^2 - \sum s_i^2}{2} \leq \frac{\left(\sum s_i\right)^2 - \left(\sum s_i\right)^2 / d}{2} = \frac{d-1}{2d}(s-1)^2,$$

то для числа легальных схем глубины $d+1$ и сложности не выше s справедлива оценка

$$N(s) \leq s^d \cdot \max_{s_1 + \ldots + s_d = s-1} N(s_1, \ldots, s_d) \leq s^d \cdot 2^{\frac{d-1}{2d} s^2} (2 \cdot 3^n)^s.$$

Утверждение леммы следует из необходимого условия $N(s) \geq 2^{2^n}$. □

Из леммы, в частности, вытекает неравенство $C^*(n) \gtrsim \sqrt{2} \cdot 2^{n/2}$.

В заключение укажем на возможность уточнения нижней оценки величины $C^U(n)$. Сначала напомним схему доказательства [121] соотношения $C^U(n) \gtrsim 1.914 \cdot 2^{n/2}$. Удобно разбить множество элементов схемы на 4 типа: конъюнкторы, выходы которых не подаются на вход элементов отрицания; дизъюнкторы, выходы которых не подаются на вход элементов отрицания; конъюнкторы вместе с их отрицаниями; дизъюнкторы вместе с их отрицаниями. Последние два типа — парные, дают удвоенный вклад в сложность схемы.

Далее оценивается число схем с заданным числом элементов каждого типа: a', a, b', b соответственно парных и непарных дизъюнкторов, парных и непарных конъюнкторов. Предварительно фиксируется естественная нумерация элементов схемы. Для асимптотики оценки сложности существенно только число способов соединить между собой функциональные элементы.

Множество подсчитываемых схем можно ограничить. Как обычно, запрещается соединять индивидуальные элементы одного типа. Дополнительно запрещается оба выхода парного элемента присоединять ко входу одного и того же элемента.

Тогда для любых двух элементов схемы есть от одного до трех способов установить соединение между ними, в зависимости от типов: нет ребра, ребро выходит из элемента типа $\vee$ или $\wedge$, для парных элементов дополнительно возможно ребро, исходящее из элемента отрицания.

Число возможностей может зависеть от порядка элементов в схеме: так, парный дизъюнктор может быть присоединен к непарному конъюнктору тремя способами, а наоборот — только двумя. В методе [121] информация о порядке элементов в схеме не извлекается, принимаются во внимание только типы элементов, и оценка исходит из наихудшего возможного порядка их взаимного расположения.

Окончательно, верхняя оценка для числа вариантов соединений элементов схемы записывается как

$$\nu(a, a', b, b') = 3^{a'b+a'b'+ab'} 2^{a'(a'-1)/2+b'(b'-1)/2+ab+aa'+bb'}. \qquad (5.2)$$

Максимум величины $\nu(a, a', b, b')$ при фиксированной сложности схемы $s = 2(a' + b') + a + b$ и определяет нижнюю оценку функции Шеннона.

Лемма 5.4. $C^U(n) \gtrsim 1.944 \cdot 2^{n/2}$.

Доказательство. Оценка (5.2) предполагает наихудший случай взаимного расположения элементов, когда парные элементы предшествуют непарным. По существу, это значит, что в нижней части схемы все элементы снабжены отрицаниями. Однако, к этой части схемы можно применить правила упрощения из доказательства леммы 5.3, и сократить число возможностей.

Попробуем сыграть на этом обстоятельстве. Добавим правило из леммы 5.3: если все входы элемента типа $\vee$ или $\wedge$ относятся к парным элементам, то отрицание этого элемента не может присоединяться к элементу противоположного типа (соответственно, $\wedge$ или $\vee$).

В естественной нумерации элементов схемы (здесь считаем парные элементы за один) выделим зону из первых t элементов (t — некоторый параметр). Пусть в этой зоне расположены a'_0, a_0, b'_0, b_0 соответственно парных и непарных дизъюнкторов, парных и непарных конъюнкторов; еще a'_1, a_1, b'_1, b_1 парных и непарных дизъюнкторов, парных и непарных конъюнкторов расположены в оставшейся части схемы.

Оценим число функций $N(a'_0, a_0, b'_0, b_0, a'_1, a_1, b'_1, b_1)$, которые вычисляют такие схемы, двумя способами.

Первый способ точно такой же, как в методе [121], только дополнительно учитывает, что элементы выделенной зоны заведомо предшествуют элементам вне зоны. Получаем оценку

$$\nu_1(a_0, a'_0, b_0, b'_0, a_1, a'_1, b_1, b'_1) = \nu(a_0, a'_0, b_0, b'_0)\nu(a_1, a'_1, b_1, b'_1)\cdot \\ \cdot 3^{a'_0(b_1+b'_1)+b'_0(a_1+a'_1)} 2^{(a_0+b'_0)(b_1+b'_1)+(b_0+a'_0)(a_1+a'_1)}.$$

Во втором способе добавим отрицания ко всем элементам в выделенной зоне, которые их не имели, и применим к полученным схемам правила упрощения. Заметим, что множество функций, реализуемых схемами, при этом не изменилось. В силу правила упрощения, любой элемент выделенной зоны соединяется с любым другим элементом схемы не более чем двумя способами. Поэтому число реструктурированных схем оценивается как

$$\nu_2(a_0, a'_0, b_0, b'_0, a_1, a'_1, b_1, b'_1) = 2^{t(t-1)/2+t(a_1+a'_1+b_1+b'_1)}\nu(a_1, a'_1, b_1, b'_1).$$

Теперь можно использовать оценку

$$N(a'_0, \ldots, b_1) \le \min\{\nu_1(a_0, \ldots, b'_1),\, \nu_2(a_0, \ldots, b'_1)\} \cdot h, \qquad (5.3)$$

где добавочный множитель h имеет величину $2^{o(s^2)}$ относительно сложности схемы $s = 2(a'_0 + a'_1 + b'_0 + b'_1) + a_0 + a_1 + b_0 + b_1$ (см. доказательства предыдущих лемм).

Остается подобрать значение параметра t, при котором максимум оценки (5.3) по всем допустимым наборам $a'_0, \ldots, b_1$ минимален. Вычислительный эксперимент показывает, что $t \approx 0.079s$, при этом $N(s) \prec 2^{0.2646 \cdot s^2}$. $\square$

Резерв для уточнения нижней оценки $C^U(n)$ еще остается. В подсчитываемых схемах можно запретить треугольники, образуемые элементами определенных типов. Например, выход элемента отрицания не должен присоединяться к дизъюнктору (конъюнктору) напрямую и одновременно транзитом через конъюнктор (дизъюнктор) или его отрицание. Учет подобных запретов позволяет еще чуть-чуть улучшить оценку, но не обещает достижения окончательного результата.

5.3 Простые методы синтеза

В этом параграфе приводятся простые способы доказательства оценок $L(n), L^U(n) \lesssim 2^n/n$, $C(n) \lesssim 2 \cdot 2^{n/2}$, но не оптимальные с точки зрения глубины используемых конструкций формул и схем.

Для набора переменных $X = (x_1, \ldots, x_n)$ и булева вектора $\sigma = (\sigma_1, \ldots, \sigma_n)$ обозначим через X^σ элементарную конъюнкцию (э.к.) $x_1^{\sigma_1} \cdot \ldots \cdot x_n^{\sigma_n}$, через X_σ — элементарную дизъюнкцию (э.д.) $x_1^{1-\sigma_1} \vee \ldots \vee x_n^{1-\sigma_n}$, через $\dot{X}^\sigma$ — моном $\bigwedge_{i:\,\sigma_i=1} x_i$, а через $\dot{X}_\sigma$ — дизъюнкцию $\bigvee_{i:\,\sigma_i=0} x_i$. Через $|X|$ будем обозначать число переменных в наборе X.

В доказательстве оценок сложности формул мы воспользуемся результатом О. Б. Лупанова [41] об оптимальных покрытиях булевых матриц прямоугольниками. Напомним, что прямоугольником называется сплошь единичная подматрица матрицы. Пусть $P(R)$ означает множество строк, $Q(R)$ — множество столбцов, $|R| = |P(R)| + |Q(R)|$ — вес прямоугольника R. Следующее утверждение является переформулировкой леммы 3.2.

Утверждение 5.1. *Пусть $s \in \mathbb{N}$. Произвольная булева матрица размера $p \times q$ может быть покрыта не более чем $2^s q/s$ прямоугольниками R_i суммарным весом $\sum |P(R_i)| \leq p(q/s + 1)$ и $\sum |Q(R_i)| \leq 2^{s+1} q$.*

Теорема 5.1. $L_4(n) \lesssim 2^n/n$.

Доказательство. Разобьем множество из n переменных на 2 группы X

и Y. Произвольную булеву функцию можно представить в виде

$$f(X,Y) = \bigvee_{\sigma\in\mathbb{B}^{|X|},\ \tau\in\mathbb{B}^{|Y|}} f_{\sigma,\tau} X^\sigma Y^\tau, \tag{5.4}$$

коэффициенты $f_{\sigma,\tau}$ образуют булеву матрицу размера $2^{|X|}\times 2^{|Y|}$.

Выберем $|Y| \in \omega(\log n)\cap n^{o(1)}$. Применяя утверждение 3.2 с выбором параметра $s \approx |X| - 2\log_2 n$, находим покрытие $R_1,\ldots,R_t$ матрицы $(f_{\sigma,\tau})$, где $t = O(2^n/n^3)$. Тогда (5.4) можно переписать в виде

$$f(X,Y) = \bigvee_{i=1}^{t} \left(\bigvee_{\sigma\in P(R_i)} X^\sigma\right)\cdot\left(\bigvee_{\tau\in Q(R_i)} Y^\tau\right). \tag{5.5}$$

Оценим сложность формулы (5.5). На первом уровне формулы вычисляются необходимые э.к. X^σ, Y^τ — их количество совпадает с весом покрытия, $\sum_i |R_i| = (1+o(1))2^n/n$. На втором уровне — дизъюнкции по столбцам и строкам прямоугольников, всего $2t$ штук. На третьем уровне формулы — t конъюнкторов. □

Теорема 5.2. $L_4^U(n) \lesssim 2^n/n$.

Доказательство. Сначала опишем простую конструкцию глубины 6, которая фактически содержится в работе Э. И. Нечипорука [52]. Разобьем множество переменных на две группы X, Y и сгруппируем наборы переменных X по слоям булева куба (k-й слой куба $\mathbb{B}^m$ будем обозначать через $\mathbb{B}_k^m$). Тогда вместо (5.4) имеем

$$f(X,Y) = \bigvee_{k=0}^{|X|} \bigvee_{\sigma\in\mathbb{B}_k^{|X|},\ \tau\in\mathbb{B}^{|Y|}} f_{k,\sigma,\tau} X^\sigma Y^\tau, \tag{5.6}$$

где при фиксированном k коэффициенты $f_{k,\sigma,\tau}$ составляют булеву матрицу размера $\binom{|X|}{k}\times 2^{|Y|}$.

Для каждой матрицы $(f_{k,\sigma,\tau})$, применяя утверждение 5.1 с некоторым параметром s, подберем покрытие $R_{k,1},\ldots,R_{k,t_k}$.

Через $E_k(X)$ обозначим характеристическую функцию k-го слоя булева куба $\mathbb{B}^{|X|}$. Заметим, что $E_k(X)\cdot\dot{X}^\sigma = X^\sigma$ для любого набора $\sigma\in\mathbb{B}_k^{|X|}$. Тогда

(5.6) можно преобразовать как

$$f(X,Y) = \bigvee_{k=0}^{|X|} E_k(X) \cdot \bigvee_{i=1}^{t_k} \left(\bigvee_{\sigma \in P(R_{k,i})} \dot{X}^\sigma \right) \cdot \left(\bigvee_{\tau \in Q(R_{k,i})} Y^\tau \right). \qquad (5.7)$$

Выберем $|Y| \asymp \log^2 n$ и $s = |X| - |Y|$. Тогда $t_k = O\left(2^{|X|}\right)$, $\sum_{k,i} |Q(R_{K,i})| = O\left(|X| \cdot 2^{|X|}\right)$ и $\sum_{k,i} |P(R_{k,i})| \sim 2^n/n$.

Пусть функции $E_k(X)$ вычисляются посредством ДНФ. Тогда формула (5.7) имеет глубину 6. Сложность формулы складывается из следующих компонент: $O\left(|X| \cdot 2^{|X|}\right)$ элементов для реализации ДНФ всех функций E_k, $O(|Y| \cdot \sum |Q(R_{k,i})|)$ элементов для вычисления э.к. Y^τ, $\sum |P(R_{k,i})|$ конъюнкторов, вычисляющих мономы $\dot{X}^\sigma$, и $O\left(|X| \cdot 2^{|X|}\right)$ прочих элементов. Асимптотика сложности определяется величиной $\sum |P(R_{k,i})| \sim 2^n/n$.

Чтобы доказать оценку с глубиной 4, модифицируем разбиение наборов переменных X. Пусть множество переменных X состоит из r групп $X^{(i)}$ одинаковой мощности m. Для любого вектора $K = (k_1, \ldots, k_r) \in [0, m]^r$ обозначим $\mathbb{B}_K^{mr} = \mathbb{B}_{k_1}^m \times \ldots \times \mathbb{B}_{k_r}^m$ и $E_K(X) = \bigwedge_{i=1}^r E_{k_i}\left(X^{(i)}\right)$. Сгруппируем наборы переменных X по множествам $\mathbb{B}_K^{mr}$. В силу $E_K(X) \cdot \dot{X}^\sigma = X^\sigma$ при любом $\sigma \in \mathbb{B}_K^{mr}$, имеет место представление

$$f(X,Y) = \bigvee_{K \in [0,m]^r} \bigvee_{\sigma \in \mathbb{B}_K^{mr},\, \tau \in \mathbb{B}^{|Y|}} f_{K,\sigma,\tau} X^\sigma Y^\tau = \\ = \bigvee_{K \in [0,m]^r} \bigvee_{\sigma \in \mathbb{B}_K^{mr},\, \tau \in \mathbb{B}^{|Y|}} E_K(X) f_{K,\sigma,\tau} \dot{X}^\sigma Y^\tau,$$

Применяя утверждение 5.1 с подходящим параметром s, при каждом K строим покрытие $R_{K,1}, \ldots, R_{K,t_K}$ матрицы $(f_{K,\sigma,\tau})$ (матрица имеет размер $|\mathbb{B}_K^{mr}| \times 2^{|Y|}$). В итоге получаем формулу

$$f(X,Y) = \bigvee_{K \in [0,m]^r} \bigvee_{i=1}^{t_K} E_K(X) \left(\bigvee_{\sigma \in P(R_{K,i})} \dot{X}^\sigma \right) \cdot \left(\bigvee_{\tau \in Q(R_{K,i})} Y^\tau \right). \qquad (5.8)$$

Выберем $|Y| \asymp \log^2 n$, $r \sim \log_2 n$ и $s \approx |X| - 2n/r$. Тогда, согласно оценкам утверждения 3.2, $t_K = O(2^{n-2n/r})$ и $\sum_{K,i} |P(R_{K,i})| \sim 2^n/n$.

Пусть функции $E_k(X^{(i)})$ вычисляются при помощи ДНФ. Тогда формула (5.8) имеет глубину 4. Общий вклад ДНФ функций $E_k\left(X^{(i)}\right)$ в сложность формулы можно оценить как $O(rmn^r 2^m 2^{n-2n/r}) = o(2^n/n)$. Асимптотика сложности формулы определяется $\sum_{K,i} |P(R_{K,i})|$ конъюнкторами, вычисляющими мономы $\dot{X}^\sigma$. Сложность реализации всех э.к. Y^τ можно грубо оценить как $O\left(n^r 2^{n-2n/r}|Y| \cdot 2^{|Y|}\right) = o(2^n/n)$. На глубине 2 и выше располагаются еще $O(rn^r 2^{n-2n/r}) = o(2^n/n)$ элементов. □

Простая верхняя оценка функционала $C^U(n)$ приведена в [121]. Фактически там установлены оценки $C_3(n) \lesssim 2 \cdot 2^{n/2}$ при четном n и $C_3(n) \lesssim (3/\sqrt{2}) \cdot 2^{n/2}$ при нечетном n.

Напомним сам метод. Разобьем множество переменных на две группы X и Y. Булеву функцию f разложим на подфункции по переменным X, а эти подфункции представим в виде КНФ:

$$f(X,Y) = \bigvee_{\sigma\in\mathbb{B}^{|X|}} X^\sigma \cdot f_\sigma(Y) = \bigvee_{\sigma\in\mathbb{B}^{|X|}} X^\sigma \cdot \bigwedge_{\tau\in f_\sigma^{-1}(0)} Y_\tau. \tag{5.9}$$

На первом уровне схемы вычисляются всевозможные э.д. Y_τ, $2^{|Y|}$ штук. На втором уровне — $2^{|X|}$ конъюнкций $X^\sigma \cdot \bigwedge Y_\tau$. Дизъюнктор-выход схемы имеет глубину 3. Оценки получаются при выборе $|X| = \lfloor n/2 \rfloor$.

Используя прием Э. И. Нечипорука разложения булева куба по двум ортогональным системам [52][35], можно выровнять оценки сложности для четного и нечетного случаев, сохраняя простоту доказательства ценой повышения глубины.

Предварительно напомним полезный и просто проверяемый факт о двойственности дизъюнктивных и конъюнктивных форм.

Утверждение 5.2. *Пусть $g_1(X), \ldots, g_t(X)$ — система попарно ортогональных функций, т. е. $g_i(X)g_j(X) = 0$ при $i \neq j$. Обозначим $g_k^*(X) = \bigvee_{i\neq k} g_i(X)$. Тогда при любых $A_i \in \mathbb{B}$ справедливо тождество*

$$\bigvee_{i=1}^{t} A_i \cdot g_i(X) = \bigwedge_{i=1}^{t} (A_i \vee g_i^*(X)) \cdot \bigvee_{i=1}^{t} g_i(X). \tag{5.10}$$

[35] Наглядно этот прием демонстрируется при доказательстве теоремы 4 [52].

Отметим, что в случае полной системы функций $\{g_i(X)\}$ тождество (5.10) упрощается ввиду $\bigvee g_i(X) = 1$.

Теорема 5.3. $C_4(n) \lesssim 2 \cdot 2^{n/2}$.

Доказательство. Множество переменных разобьем на 3 группы: X, Y и Z с условием $|X| = |Y|$.

На множестве всех э.к. Z^σ введем нумерацию с двумя индексами: $Z^{i,j}$, $0 \le i, j < 2^{|Z|/2} + 1$.

Положим $h_i(Z) = \bigvee_j Z^{i,j}$ и $g_j(Z) = \bigvee_i Z^{i,j}$. Введенное множество функций позволяет выразить любую э.к. как $Z^{i,j} = h_i(Z)g_j(Z)$ и удовлетворяет условиям ортогональности $h_i(Z)h_{i'}(Z) = 0$ при $i \ne i'$, $g_j(Z)g_{j'}(Z) = 0$ при $j \ne j'$.

Запишем

$$f(X,Y,Z) = \bigvee_{\sigma,\tau \in \mathbb{B}^{|X|}} X^\sigma Y^\tau f_{\sigma,\tau}(Z) = \bigvee_{\sigma,i} X^\sigma h_i(Z) \cdot \bigvee_{\tau,j:\ Z^{i,j} \le f_{\sigma,\tau}(Z)} Y^\tau g_j(Z). \tag{5.11}$$

Теперь внутренние дизъюнкторы в формуле (5.11) заменим конъюнкторами, используя утверждение 5.2. Положим $g_j^* = \bigvee_{j' \ne j} g_{j'}$. Применяя утверждение 5.2 к ортогональной системе $Y^\tau g_j(Z)$, получаем тождество

$$\bigvee_{(\tau,j) \in T} Y^\tau g_j(Z) = \bigwedge_{(\tau,j) \notin T} (Y_\tau \vee g_j^*(Z)).$$

Окончательно, имеем представление

$$f(X,Y,Z) = \bigvee_{\sigma,i} X^\sigma h_i(Z) \cdot \bigwedge_{\tau,j:\ Z^{i,j} \cdot f_{\sigma,\tau}(Z) = 0} (Y_\tau \vee g_j^*(Z)). \tag{5.12}$$

Пусть функции $g_j^*(Z)$ и $h_i(Z)$ реализуются в виде ДНФ. Оценим сложность схемы, построенной по формуле (5.12).

На глубине 1 вычисляются всевозможные э.к. Z^σ, $2^{|Z|}$ штук. На глубине 2 расположены дизъюнкторы, вычисляющие все функции $h_i(Z)$, не более $2^{|Z|/2} + 1$ штук, и всевозможные функции $Y_\tau \vee g_j^*(Z)$, не более $2^{|Y|}(2^{|Z|/2} + 1)$ штук. На глубине 3 расположены конъюнкторы, вычисляющие все произведения под знаком дизъюнкции в (5.12), не более $2^{|X|}(2^{|Z|/2} + 1)$ штук.

Таким образом, с учетом $|X| = |Y|$, сложность схемы оценивается как $2 \cdot 2^{n/2} + O(2^{|X|} + 2^{|Z|})$. Достаточно выбрать $|X| \sim |Z|$. $\square$

5.4 Специальные разбиения булева куба

Эффективные методы синтеза схем и формул глубины 3 опираются на специальные разбиения (или покрытия) булева куба. В этом параграфе изложен способ построения разбиений определенного типа, обобщающих классическое разбиение куба на единичные сферы.

Введем понятие псевдосферы, совпадающее с понятием S-множества из работы [33]. *Псевдосферой* называется множество двоичных наборов одинаковой длины, таких, что каждый набор имеет координату, значение которой отличает его от всех других наборов множества.

Простым примером псевдосферы в $\mathbb{B}^n$, причем экстремальной мощности n, является сфера единичного радиуса с центром $(\alpha_1, \ldots, \alpha_n) \in \mathbb{B}^n$:

$$\{(\alpha_1, \ldots, \alpha_{i-1}, \overline{\alpha}_i, \alpha_{i+1}, \ldots, \alpha_n) \mid 1 \leq i \leq n\}.$$

Пусть $\{a_1, \ldots, a_r\} \subset \mathbb{B}^n$ — псевдосфера, $\varphi(X) = \bigvee_{i=1}^{r} X^{a_i}$ — ее характеристическая функция на множестве переменных $x_1, \ldots, x_n$. Тогда любую э.к. X^{a_i} можно вычислить как $\varphi(X) \cdot x_{\pi(i)}^{\beta(i)}$, где $\pi(i)$ — номер, а $\beta(i)$ — значение координаты, которая отличает набор a_i.

При синтезе AC-формул нам потребуется экономное разбиение булева куба $\mathbb{B}^n$ на псевдосферы. Как известно, если $n = 2^k$, то куб можно просто разбить на $2^n/n$ сфер радиуса 1, см., например, [49]. Как следует из результата Г. А. Кабатянского и В. И. Панченко [18], при произвольном n куб также можно покрыть $(1 + o(1))2^n/n$ сферами[36]. Кроме того, асимптотически оптимальное разбиение на $(1 + o(1))2^n/n$ псевдосфер построено А. Е. Липатовой [33]. Ниже мы предложим альтернативный и, пожалуй, более простой вывод этой оценки, причем с чуть лучшим остаточным членом в $o(1)$.

[36] В [18] показано, что куб $\mathbb{B}^n$ можно покрыть $(1 + o(1))2^n/n$ единичными шарами. Как легко проверить, если шары с множеством центров A образуют покрытие $\mathbb{B}^n$, то сферы с множеством центров $A \times \mathbb{B}$ образуют покрытие $\mathbb{B}^{n+1}$.

При синтезе формул над U мы ограничены в использовании отрицаний. В этом случае могут быть полезны разбиения, аналогичные разбиениям на полусферы. Напомним, что верхняя (нижняя) полусфера — это наборы единичной сферы, расположенные слоем выше (ниже) центра сферы. Из результата [119] следует существование разбиения булева куба размерности n на $\Theta(2^n/n)$ верхних или нижних полусфер. Говоря о покрытии куба полусферами, мы имеем в виду покрытие куба без одного набора: сплошь нулевого или сплошь единичного.

По аналогии, назовем *положительной* (*отрицательной*) *псевдополусферой* такую псевдосферу, в которой каждый набор имеет координату со значением 1 (соответственно, 0), отличающую его от остальных наборов псевдосферы. Если $\varphi(X)$ — характеристическая функция псевдополусферы $\{a_1, \ldots, a_r\}$, то любая э.к. X^{a_i} может быть вычислена как $\varphi(X) \cdot x^{\beta}_{\pi(i)}$, где $\pi(i)$ — номер координаты, которая отличает набор a_i, а $\beta = 0$ или $\beta = 1$ в зависимости от того, отрицательная псевдополусфера или положительная.

Ниже мы построим разбиение куба (без одного набора) на $(1 + o(1))2^{n+1}/n$ положительных или отрицательных псевдополусфер. Эта оценка асимптотически точная, так как средний размер псевдополусферы в разбиении куба $\mathbb{B}^n$ не может превышать $n/2 + o(1)$.

Для доказательства оценки сложности AC-схем глубины 3 нам понадобится специальное разбиение булева куба $\mathbb{B}^n$, совмещающее достоинства двух упоминавшихся ранее разбиений: возможность выделения набора в подмножестве домножением на одну конъюнкцию, как в случае разбиения на псевдосферы, и неизбыточность разбиения, как в случае разложения по двум ортогональным системам (свойство $|\{h_i\}| \cdot |\{g_j\}| \sim 2^n$ в обозначениях теоремы 5.3).

Более конкретно, нас интересует возможность разбиения булева куба или основной его части на $(1 + o(1))2^n/q$ множеств мощности не более q, таких, что любой набор σ произвольного множества из разбиения обладает уникальным в рамках своего множества поднабором σ', при этом в каждом случае такой выделяющий поднабор может быть выбран из некоторого общего для всех множеств разбиения множества мощности $(1 + o(1))q$.

Три перечисленных вида разбиений будут далее получены как частные

варианты более общей конструкции.

Рассмотрим множество A, в котором выбрано некоторое непустое подмножество B; элементы подмножества B далее будем называть *отметками*. Для вектора $b = (b_1, \ldots, b_n) \in (B \cup \{*\})^n$ определим множество векторов $A\wr_B^b \subset A^n$:

$$A\wr_B^b = \{(a_1, \ldots, a_n) \in A^n \mid a_i = b_i, \text{ если } b_i \in B; a_i \notin B, \text{ если } b_i = *\}.$$

Иначе говоря, на позициях отметок вектора b векторы из $A\wr_B^b$ имеют такие же отметки, а на позициях звездочек — произвольные элементы из $A \setminus B$. Через $\nu(b)$ далее будем обозначать вес вектора b — число отметок в нем.

Множество $S \subset A^n$ назовем (A, B)-*псевдосферическим*, если каждый вектор из S отличается от остальных значением отметки (т. е. элементом из B) в некоторой позиции. Пусть можно установить соответствие между элементами множества и выделяющими их позициями, при котором каждой позиции соответствует не более одного элемента (далее мы будем рассматривать только такие множества). Тогда назовем *типом* множества S вектор из $(B \cup \{*\})^n$ веса $|S|$, который для любого элемента S содержит выделяющую его отметку в соответствующей позиции. (Тип может быть определен, вообще говоря, неоднозначно.)

Пусть $A\wr_B^{n,[h_1,h_2]}$ означает множество всех векторов из A^n, имеющих не менее h_1 и не более h_2 координат из подмножества B (отметок).

Обозначим

$$E(h,q,t,s) = \sum_{k=\max\{0,\,h-q\}}^{t} \binom{t}{k}\binom{q-1}{h-1-k} s^k (s-1)^{h-1-k}, \tag{5.13}$$

$$E([h_1,\, h_2],q,t,s) = \sum_{h=h_1}^{h_2} E(h,q,t,s).$$

Величины $E(h,q,t,s)$ определены при $s \geq 2$ и $h \leq q+t$, а также при $s = 1$ и $h \leq t+1$ (в последнем случае сумма содержит только одно ненулевое слагаемое $\binom{t}{h-1}$).

Лемма 5.5. *Пусть* $B \subset A$, $r \geq 0$, $1 \leq h_1 \leq h_2$ *и либо* $|B| = 1$ *и* $h_2 \leq t+1$, *либо* $|B| \geq 2$ *и* $h_2 \leq q+t$. *Пусть* $\left|A\wr_B^{q+t,[h_1,h_2]} \times A^r\right| = M$. *Тогда существует*

разбиение множества $A\wr_B^{q+t,[h_1,\,h_2]} \times A^r$ *на*

$$M/q + E([h_1,\,h_2], q, t, |B|) \cdot \binom{q+t}{q} \cdot |B|^q$$

(A, B)*-псевдосферических подмножеств мощности не более* q*.*

Доказательство. Будем строить разбиение на псевдосферические множества, типы которых содержат отметки только в первых $q + t$ позициях, т. е. не используя последние r координат для выделения элементов подмножеств.

Опишем процедуру построения разбиения. Выделим равное число заготовок для подмножеств каждого из возможных $\binom{q+t}{q} \cdot |B|^q$ типов веса q. В заготовке подмножества на заданное место мы можем поместить произвольный вектор, имеющий нужную отметку в соответствующей позиции и не имеющий совпадений с другими $q-1$ отметками в типе заготовки (на оставшиеся $t + r$ позиций не накладывается ограничений).

Для любого вектора из $A\wr_B^b \times A^r \subset A\wr_B^{q+t,[h_1,\,h_2]} \times A^r$ имеется определенное число способов выбрать тип заготовки и место в ней. Число способов определяется только весом $\nu(b)$ вектора b, обозначим это число через $c_{\nu(b)}$. Мощность множества $A\wr_B^b \times A^r$ также определяется только весом вектора b, обозначим ее через $s_{\nu(b)}$. Разобьем множество $A\wr_B^b \times A^r$ произвольно на $c_{\nu(b)}$ частей примерно одинакового размера, не превосходящего $\lceil s_{\nu(b)}/c_{\nu(b)} \rceil$. Каждую часть отнесем к определенному месту в заготовке определенного типа.

Тем самым на определенное место в заготовке определенного типа оказываются распределены векторы из $E([h_1,\,h_2], q, t, |B|)$ множеств $A\wr_B^b \times A^r$. Действительно, выбрать подходящий вектор b можно, сначала выбрав вес h вектора, затем $h-1$ позицию с отметками в нем (одна позиция предопределена местом в заготовке), для $h-1-k$ позиций на пересечении с позициями отметок типа заготовки мы имеем $|B|-1$ вариант для выбора отметки, для k прочих позиций — $|B|$ вариантов, см. (5.13).

Таким образом, на фиксированное место в заготовке определенного ти-

па распределено не более

$$\sum_{h=h_1}^{h_2} E(h,q,t,|B|)\cdot\lceil s_h/c_h\rceil \le \sum_{h=h_1}^{h_2} E(h,q,t,|B|)\cdot s_h/c_h + E([h_1,\,h_2],q,t,|B|)$$

векторов. Значение суммы в правой части — одно и то же для любого места и типа заготовки, и соответствует ситуации, когда мы могли бы делить каждое множество $A\wr_B^b\times A^r$ строго на равные части вообще говоря не целого размера $s_{\nu(b)}/c_{\nu(b)}$. Поэтому оно равно $M/\left(q\binom{q+t}{t}\cdot|B|^q\right)$.

Таким образом, для каждого из $\binom{q+t}{q}\cdot|B|^q$ типов нам достаточно иметь $M/\left(q\binom{q+t}{t}\cdot|B|^q\right)+E([h_1,\,h_2],q,t,|B|)$ заготовок. □

Используя лемму, мы можем получать подходящие разбиения куба $\mathbb{B}^n$, выбирая ограничение q на размер подмножеств из широкого диапазона значений.

Следствие 5.1. *Куб $\mathbb{B}^n$ может быть покрыт $(1+O(\log n/n))\cdot 2^n/n$ псевдосферами.*

Доказательство. Пусть $n = q + r$, где $r = \lceil 2\log_2 n\rceil$. Заметим, что $\mathbb{B}\wr_{\mathbb{B}}^{q,[q,\,q]} = \mathbb{B}^q$. Тогда лемма 5.5 предоставляет разбиение куба $\mathbb{B}^n = \mathbb{B}\wr_{\mathbb{B}}^{q,[q,\,q]}\times\mathbb{B}^r$ на $2^n/q + E([q,\,q],q,0,2)\cdot 2^q$ штук $(\mathbb{B},\mathbb{B})$-псевдосферических множеств, т. е. полусфер, мощности q. Поскольку $E([q,\,q],q,0,2)=1$, число множеств в разбиении не превышает $2^n/q + 2^q = (1+O(r/n))\cdot 2^n/n$. □

Следствие 5.2. *Пусть $B=\{0\}$ или $B=\{1\}$. Тогда куб (с выколотой вершиной) $\mathbb{B}^n\setminus B^n$ может быть покрыт $(1+o(1))2^{n+1}/n$ псевдополусферами.*

Доказательство. Пусть $n = q + t + r$, где $t = q + \Delta$, $\Delta \asymp n^{2/3}$, $r \asymp n^{3/4}$. Таким образом, $q, t \sim n/2$.

Как известно, центральные k слоев n-мерного булева куба содержат почти все наборы при условии $k = \omega(\sqrt{n})$. Более точно, доля наборов в центральных слоях оценивается снизу как $1 - c^{k^2/n}$ для некоторой константы $c < 1$, см., например, [88].

В частности, $\left|\mathbb{B}\wr_B^{q+t,[q,\,t]}\right| = (1-o(1/n))\cdot|\mathbb{B}^{q+t}|$.

При помощи леммы 5.5 строится покрытие множества $\mathbb{B}\wr_B^{q+t,[q,\,t]} \times \mathbb{B}^r$ $(\mathbb{B}, B)$-псевдосферическими подмножествами (т. е. псевдополусферами) мощности не выше q. Покрытие состоит из $(1+o(1))2^n/q + E([q,\,t],q,t,1)\cdot\binom{q+t}{q}$ подмножеств. Поскольку

$$q\cdot E([q,\,t],q,t,1)\cdot\binom{q+t}{q} = q\cdot\sum_{h=q-1}^{t-1}\binom{t}{h}\binom{q+t}{q} \le q(\Delta{+}1)t^{\Delta+1}2^{q+t} = o(2^n),$$

покрытие содержит асимптотически $2^n/q \sim 2^{n+1}/n$ множеств.

Из наборов, не попавших в покрытие куба $\mathbb{B}^n\setminus B^n$, можно сформировать псевдополусферы мощности 1, их $o(2^n/n)$ штук. $\square$

Следствие 5.3. *Пусть $B \subset A = \mathbb{B}^p$, $|B| = 1$, $q = p^2(2^{p-2}-1)$, $t = p^2$. Тогда множество из $(1-o(1/q))2^{p(q+t)}$ элементов куба A^{q+t} может быть покрыто $(1+o(1))2^{p(q+t)}/q$ (A,B)-псевдосферическими подмножествами.*

Доказательство. Число векторов в A^{q+t} с h отметками (элементами из B) равно $\binom{q+t}{h}(2^p-1)^{q+t-h}$ и составляет долю

$$\binom{q+t}{h}(2^p-1)^{-h}\left(1-\frac{1}{2^p}\right)^{q+t} \le \binom{q+t}{h}(2^p-1)^{-h}e^{-p^2/4} \qquad (5.14)$$

от общего числа векторов в A^{q+t} (мы воспользовались известным неравенством $(1-1/x)^x \le 1/e$ при $x \ge 1$).

При $h = 0$ и при $h \ge t$ величина (5.14) оценивается сверху как $e^{-p^2/4}$ в силу

$$\binom{q+t}{h}(2^p-1)^{-h} \le \left(\frac{e(q+t)}{h(2^p-1)}\right)^h \le \left(\frac{e\cdot 2^{p-2}}{2^p-1}\right)^h < 1,$$

где в первом переходе используется известное неравенство $\binom{n}{k} \le (en/k)^k$.

Таким образом, не менее $1 - qe^{-p^2/4} = 1 - o(1/q)$ от общего числа векторов A^{q+t} содержится в $A\wr_B^{q+t,[1,\,t]}$.

Остается применить лемму 5.5 с выбором параметров $r = 0$, $h_1 = 1$, $h_2 = t$. Асимптотическая оценка $2^{p(q+t)}/q$ размера покрытия справедлива в

силу

$$q \cdot E([1,\, t], q, t, 1) \cdot \binom{q+t}{q} \cdot |B|^q = q \cdot \sum_{h=0}^{t-1} \binom{t}{h} \binom{q+t}{t} <$$

$$< 2^q \cdot 2^t \cdot 2^{q+t} = o\left(2^{p(q+t)}\right). \quad \square$$

5.5 Синтез с глубиной 3

Напомним простой факт.

Утверждение 5.3. *Булева функция n переменных, принимающая значение 1 ровно на r наборах, имеет КНФ из rn множителей.*

Доказательство. Если обозначить достаточный для вычисления функции размер КНФ через $l(n, r)$, то разложением по переменной получаем соотношение $l(n, r) \leq \max_k\{l(n-1, k) + l(n-1, r-k)\}$, из которого и из начального условия $l(n, 1) = n - 1$ следует нужная оценка. $\square$

Этой простой оценки нам будет достаточно, но отметим, что вдвое лучшая оценка известна из [15].

Излагаемый далее способ построения формулы по сути является упрощенной применительно к рассматриваемой ситуации версией метода О. Б. Лупанова [42].

Теорема 5.4. $L_3(n) \lesssim 2^n/n$.

Доказательство. Разобьем множество переменных на 2 группы X, Y. Пусть $S_1, \ldots, S_v$ — гарантируемое следствием 5.1, результатом [33] или следствием из [18] покрытие куба $\mathbb{B}^{|X|}$ псевдосферами, $v \sim 2^{|X|}/|X|$. Пусть $\varphi_i(X)$ — характеристическая функция множества S_i. Обозначим через $x_{\pi(\sigma)}^{\beta(\sigma)}$ множитель, выделяющий э.к. X^σ из своей псевдосферы. Рассмотрим пред-

ставление

$$f(X,Y) = \bigvee_{i=1}^{v} \bigvee_{\sigma\in S_i,\,\tau\in\mathbb{B}^{|Y|}} f_{\sigma,\tau} X^\sigma Y^\tau =$$

$$= \bigvee_{i=1}^{v} \varphi_i(X) \cdot \bigvee_{\sigma\in S_i,\,\tau\in\mathbb{B}^{|Y|}} f_{\sigma,\tau}\, x_{\pi(\sigma)}^{\beta(\sigma)} \cdot Y^\tau =$$

$$= \bigvee_{i=1}^{v} \varphi_i(X) \cdot \bigvee_{\tau\in\mathbb{B}^{|Y|}} D_{i,\tau}(X) \cdot Y^\tau, \qquad D_{i,\tau} = \bigvee_{\sigma\in S_i} f_{\sigma,\tau} \cdot x_{\pi(\sigma)}^{\beta(\sigma)}. \quad (5.15)$$

Функции $D_{i,\tau}$ являются дизъюнкциями переменных X. Теперь внутренние дизъюнкции в формуле (5.15) можно преобразовать в конъюнкции согласно тождеству

$$\bigvee_{\tau\in\mathbb{B}^{|Y|}} A_\tau \cdot Y^\tau = \bigwedge_{\tau\in\mathbb{B}^{|Y|}} (A_\tau \vee Y_\tau). \quad (5.16)$$

Тождество является следствием утверждения 5.2 для (ортогональной) системы э.к. $g_\tau(Y) = Y^\tau$.

В итоге, приходим к формуле

$$f(X,Y) = \bigvee_{i=1}^{v} \varphi_i(X) \cdot \bigwedge_{\tau\in\mathbb{B}^{|Y|}} (D_{i,\tau}(X) \vee Y_\tau). \quad (5.17)$$

На первом уровне формула (5.17) содержит $v2^{|Y|}$ дизъюнкторов, вычисляющих всевозможные $D_{i,\tau} \vee Y_\tau$, и еще согласно утверждению 5.3 не более $v|X|^2$ дизъюнкторов для множителей КНФ-представлений функций φ_i. На втором уровне расположены v конъюнкторов. Нужная оценка получается, например, при выборе $|X| \approx n - 3\log_2 n$. □

Используя вместо псевдосфер псевдополусферы в качестве множеств разбиения булева куба, получим аналогичный результат для сложности формул над базисом U.

Теорема 5.5. $L_3^U(n) \lesssim 2 \cdot 2^n/n$.

Доказательство. Разобьем множество переменных на 3 группы X, Y, Z с условием $|Y| = |Z|$. Пусть $S_1, \ldots, S_v$ — гарантируемое следствием 5.2 разбиение куба $\mathbb{B}^{|X|}$ на положительные псевдополусферы и отдельный нулевой вектор, $v \sim 2^{|X|+1}/|X|$.

Обозначим через $\varphi_i(X)$ характеристические функции множеств S_i. Пусть $g_{i,k}(X)$ — множитель, выделяющий k-й набор множества S_i — он либо совпадает с некоторой переменной из X, либо тождественно равен 1 в случае $|S_i| = 1$.

Пусть, как в доказательстве теоремы 5.2, $E_k(Y)$ означает характеристическую функцию k-го слоя $\mathbb{B}_k^{|Y|}$ булева куба $\mathbb{B}^{|Y|}$. Рассмотрим представление

$$f(X,Y,Z) = \bigvee_{i=1}^{v} \bigvee_{j=0}^{|Y|} \bigvee_{k=0}^{|Z|} \varphi_i(X) E_j(Y) E_k(Z) \cdot \bigvee_{\sigma \in \mathbb{B}_j^{|Y|},\, \tau \in \mathbb{B}_k^{|Z|}} \dot{Y}^\sigma \dot{Z}^\tau D_{i,\sigma,\tau}(X),$$

$$D_{i,\sigma,\tau}(X) = \bigvee_{k=1}^{|S_i|} f_{i,\sigma,\tau,k} \cdot g_{i,k}(X). \quad (5.18)$$

Согласно утверждению 5.2, применяемому при фиксированных j, k к ортогональной системе функций $E_j(Y)E_k(Z)\dot{Y}^\sigma \dot{Z}^\tau$, где $\sigma \in \mathbb{B}_j^{|Y|}$, $\tau \in \mathbb{B}_k^{|Z|}$, с учетом тождества

$$E_j(Y) \cdot \bigvee_{\rho \in \mathbb{B}_j^{|Y|} \setminus \{\sigma\}} \dot{Y}^\rho = E_j(Y) \cdot \dot{Y}_\sigma$$

справедливо

$$E_j(Y)E_k(Z) \cdot \bigvee_{\sigma \in \mathbb{B}_j^{|Y|},\, \tau \in \mathbb{B}_k^{|Z|}} A_{\sigma,\tau} \dot{Y}^\sigma \dot{Z}^\tau =$$

$$= E_j(Y)E_k(Z) \cdot \bigwedge_{\sigma \in \mathbb{B}_j^{|Y|},\, \tau \in \mathbb{B}_k^{|Z|}} (A_{\sigma,\tau} \vee \dot{Y}_\sigma \vee \dot{Z}_\tau).$$

Преобразуя (5.18) при помощи указанного тождества, приходим к формуле

$$f(X,Y,Z) = \bigvee_{i=1}^{v} \bigvee_{j=0}^{|Y|} \bigvee_{k=0}^{|Z|} \varphi_i(X) E_j(Y) E_k(Z) \cdot \bigwedge_{\sigma \in \mathbb{B}_j^{|Y|},\, \tau \in \mathbb{B}_k^{|Z|}} (D_{i,\sigma,\tau}(X) \vee \dot{Y}_\sigma \vee \dot{Z}_\tau). \quad (5.19)$$

По построению, внутренние дизъюнкции содержат только переменные без отрицаний.

Характеристические функции $\varphi_i(X)$, $E_j(Y)$, $E_k(Z)$ реализуем при помощи КНФ. Согласно утверждению 5.3, КНФ функции $\varphi_i(X)$ содержит не более $|X|^2$ множителей. Число множителей в КНФ для $E_k(Y)$ оценим тривиально как $2^{|Y|}$.

В формуле (5.19) на глубине 1 расположены $v2^{2|Y|} \sim 2^{n+1}/|X|$ элементов, вычисляющих внутренние дизъюнкции, а также $O\left(v|Y|^2\left(|X|^2 + 2^{|Y|}\right)\right)$ элементов дизъюнкции и отрицания для множителей КНФ-представления характеристических функций. На глубине 2 расположены $v(|Y|+1)^2$ конъюнкторов.

Отметим, что разбиение множества переменных на несколько групп применяется лишь для того, чтобы подавить сложность характеристических функций. Для достижения требуемой оценки следует положить, например, $|Y| \sim 2\log_2 n$. □

Наконец, используя специальное разбиение булева куба на псевдосферические множества, удается установить окончательный результат об асимптотике функции Шеннона сложности AC-схем.

Теорема 5.6. $C_3(n) \lesssim 2 \cdot 2^{n/2}$.

Доказательство. Достаточно ограничиться случаем нечетного n. Выберем параметр $q = o(n/\log n)$ вида $p^2(2^{p-2}-1)$ с условием $q \sim \sqrt{2}\cdot 2^s$, $s \in \mathbb{N}$. Пусть $k = p^3 2^{p-2}$. Множество переменных разобьем на 3 группы: X, Y, Z из $m = \frac{n+1}{2} + s - k$, $r = \frac{n-1}{2} - s$ и k штук соответственно.

Произвольно выберем множество $B \subset \mathbb{B}^p$, $|B| = 1$. Пусть $S_1, \ldots, S_u$ — разбиение почти всего булева куба $\mathbb{B}^k$ на $(\mathbb{B}^p, B)$-псевдосферические множества, гарантируемое следствием 5.3, $u \sim 2^k/q$. Не покрытые разбиением $u' = o(2^k/q)$ векторов из $\mathbb{B}^k$ определим как множества $S_{u+1}, \ldots, S_{u+u'}$ мощности 1.

Обозначим через $h_i(Z)$ характеристические функции множеств S_i, $1 \le i \le u+u'$, а через $g_j(Z)$ — конъюнкции, выделяющие отметки (поднаборы из B) в векторах Z, $1 \le j \le q + p^2$. Пусть $b_i = (b_{i,1}, \ldots, b_{i,q+p^2})$ означает тип множества S_i (при неоднозначности определения выбор значения типа

произволен). Тогда справедлива декомпозиция, аналогичная (5.11):

$$f(X,Y,Z)=\bigvee_{\sigma\in\mathbb{B}^m}\bigvee_{i=1}^{u}X^\sigma h_i(Z)\cdot\bigvee_{\tau\in\mathbb{B}^r}\bigvee_{j\colon b_{i,j}\in B}f_{\sigma,\tau,i,j}\cdot Y^\tau g_j(Z)\vee \bigvee_{\sigma\in\mathbb{B}^m}\bigvee_{i=u+1}^{u+u'}X^\sigma h_i(Z)\cdot\bigvee_{\tau\in\mathbb{B}^r}f_{\sigma,\tau,i}\cdot Y^\tau. \quad (5.20)$$

При фиксированном i для таких j, что $b_{i,j}\in B$, определим $g^*_{i,j}=\bigvee_{j'\neq j\colon b_{i,j'}\in B}g_{j'}$. Заметим, что $h_i(Z)g^*_{i,j}(Z)=h_i(Z)\cdot\overline{g_j(Z)}$. Применяя утверждение 5.2 (при фиксированном i) к ортогональной системе $h_i(Z)Y^\tau g_j(Z)$, $b_{i,j}\in B$, получаем тождество

$$h_i(Z)\cdot\bigvee_{(\tau,j)\in T}Y^\tau g_j(Z)=h_i(Z)\cdot\bigwedge_{(\tau,j)\notin T}(Y_\tau\vee g^*_{i,j}(Z))=h_i(Z)\cdot\bigwedge_{(\tau,j)\notin T}(Y_\tau\vee\overline{g_j(Z)}),$$

которое позволяет переписать (5.20) в виде

$$f(X,Y,Z)=\bigvee_{\sigma\in\mathbb{B}^m}\bigvee_{i=1}^{u}X^\sigma h_i(Z)\cdot\bigwedge_{\tau\in\mathbb{B}^r,\,j\colon b_{i,j}\in B,\,f_{\sigma,\tau,i,j}=0}\left(Y_\tau\vee\overline{g_j(Z)}\right)\vee \bigvee_{\sigma\in\mathbb{B}^m}\bigvee_{i=u+1}^{u+u'}X^\sigma h_i(Z)\cdot\bigwedge_{\tau\in\mathbb{B}^r\colon f_{\sigma,\tau,i}=0}Y_\tau. \quad (5.21)$$

Оценим сложность схемы, построенной по формуле (5.21). На первом уровне схемы вычисляются всевозможные дизъюнкции Y_τ и $Y_\tau\vee\overline{g_j(Z)}$, всего $2^r(q+p^2+1)\sim 2^{n/2}$ штук, а также всевозможные э.д. переменных Z для КНФ-представлений функций $h_i(Z)$, всего $2^k=2^{o(n)}$ штук. На втором уровне схемы располагаются $(u+u')2^m\sim 2^{k+m}/q\sim 2^{n/2}$ конъюнкторов. $\square$

5.6 Синтез схем над базисом U^∞

5.6.1 Специальные системы функций

Определим два типа систем функций $\{g^r_i(X)\}$, в которых индексация указывает соответствие с вершинами многодольного графа: верхний индекс означает номер доли (яруса), нижний — порядковый номер вершины в данной доле.

M-система ранга (t, p) определяется как множество булевых функций $\{g_i^r(X)\}$, $r = 1, \ldots, t$, $i = 1, \ldots, p$, удовлетворяющих условиям:

(i) любые две различные функции g_i^r, g_j^s имеют не более одного общего единичного набора. Иначе говоря, произведение любых двух функций — либо 0, либо некоторая э.к. X^σ;

(ii) ортогональность внутри яруса: $g_i^r g_j^r = 0$ при любых r и $i \neq j$.

M-систему $\{g_i^r(X)\}$ назовем *полной*, если дополнительно выполняется условие: любая э.к. X^σ может быть вычислена единственным образом как произведение некоторых двух функций системы, $X^\sigma = g_i^r g_j^s$.

M-систему $\{h_i^r(X)\}$ назовем *равномерно полной*, если она удовлетворяет условию: для каждой пары ярусов r и s любая э.к. X^σ может быть вычислена единственным образом как произведение некоторых двух связанных с этими ярусами функций, $X^\sigma = h_i^r h_j^s$.

Пусть $|X| = n$. Чтобы предъявить полную M-систему ранга (t, p), где $p^2 t(t-1)/2 \geq 2^n$, достаточно взять полный t-дольный граф с p вершинами в каждой доле и затем всем э.к. X^σ поставить в соответствие инъективно ребра этого графа. Для произвольной i-й вершины доли r определим функцию g_i^r как дизъюнкцию всех э.к., соответствующих ребрам, инцидентным данной вершине. (Этот способ предложен в [52].)

Покажем как построить равномерно полную M-систему ранга (t, p) в случае простого числа $p \geq 2^{n/2}$ и $t \leq p + 1$.

Рассмотрим аффинную плоскость над $GF(p)$. Она содержит p^2 точек, и в ней можно провести $p(p+1)$ различных прямых. Каждая прямая содержит p точек. Любые две непараллельные прямые пересекаются ровно в одной точке. Имеется $p + 1$ семейство параллельных прямых, каждое семейство содержит p прямых и покрывает всю плоскость. Подробнее о геометрии аффинной плоскости см., например, в [20].

Теперь возьмем t-дольный граф с p вершинами в каждой доле и поставим в соответствие различным долям различные семейства параллельных прямых, а различным вершинам внутри доли — различные прямые из семейства. Всем э.к. X^σ поставим в соответствие инъективно точки аффинной плоскости. Для произвольной i-й вершины доли r определим функцию h_i^r как дизъюнкцию всех э.к., соответствующих точкам плоско-

сти, принадлежащим прямой, связанной с данной вершиной.

Прямое произведение M-системы функций $\{g_i^r\}$ ранга (t,p) и M-системы функций $\{h_i^r\}$ ранга (t,q) определим как систему функций $\{e_{i,j}^r = g_i^r h_j^r\}$ ранга (t,pq). Легко доказывается

Лемма 5.6. *Пусть X, Y — непересекающиеся множества переменных. Тогда прямое произведение полной (равномерно полной) M-системы функций переменных X и равномерно полной M-системы функций переменных Y является полной (равномерно полной) M-системой функций переменных X, Y.*

5.6.2 Многоярусное представление

Пусть запись $g \subset f$ означает, что множество единиц функции g содержится во множестве единиц функции f. Также для удобства введем обозначение $\varepsilon(f)$, означающее произвольную функцию, множество единиц которой содержится во множестве единиц f, т. е. $\varepsilon(f) \subset f$.

Пусть задана полная M-система функций $\{g_i^r(X)\}$ ранга (t,p). Используя ее, построим специальное многоярусное представление произвольной функции $f(X)$. Для этого определим характеристические функции яруса

$$G^r = \bigvee_{i=1}^{p} g_i^r$$

и вспомогательные функции

$$f_k^1 = g_k^1,$$

$$f_k^2 = g_k^2 \cdot \bigwedge_{j:\, g_k^2 g_j^1 \not\subset f} \overline{f_j^1},$$

$$\dots$$

$$f_k^t = g_k^t \cdot \bigwedge_{j:\, g_k^t g_j^{t-1} \not\subset f} \overline{f_j^{t-1}} \cdot \ldots \cdot \bigwedge_{j:\, g_k^t g_j^1 \not\subset f} \overline{f_j^1}.$$

Лемма 5.7. *Справедливо представление*

$$f(X) = \bigvee_{i=1}^{p} f_i^t \ \vee\ \overline{G^t} \cdot \bigvee_{i=1}^{p} f_i^{t-1} \ \vee\ \ \dots\ \ \vee\ \overline{G^t} \cdots \overline{G^3} \cdot \bigvee_{i=1}^{p} f_i^2. \qquad (5.22)$$

Доказательство. Начнем с очевидной декомпозиции

$$f = G^t \cdot f \ \vee \ \overline{G^t} \cdot G^{t-1} \cdot f \ \vee \quad \ldots \quad \vee \ \overline{G^t} \cdots \overline{G^3} \cdot G^2 \cdot f. \tag{5.23}$$

Проверим, что имеет место равенство между соответствующими слагаемыми формул (5.22) и (5.23). Для этого достаточно показать, что

$$\overline{G^t} \cdots \overline{G^{r+1}} \cdot f_k^r = \overline{G^t} \cdots \overline{G^{r+1}} \cdot g_k^r \cdot f \tag{5.24}$$

при любых k и $r = 2, \ldots, t$.

Поскольку $f_k^r \subset g_k^r \subset G^r$, выполнено $\overline{G^r} \subset \bigwedge_{j \in T} \overline{f_j^r}$ при любом множестве индексов T. Следовательно, $f_k^r \supset g_k^r \cdot \overline{G^{r-1}} \cdots \overline{G^1}$, откуда получаем $\overline{f_k^r} \subset \overline{g_k^r} \vee G^1 \vee \ldots \vee G^{r-1}$. С учетом $\overline{g_k^r} \subset \overline{f_k^r}$ можно записать

$$\overline{f_k^r} = \overline{g_k^r} \vee \varepsilon(G^1 \vee \ldots \vee G^{r-1}) = \bigvee_{i \neq k} g_i^r \ \vee \ \overline{G^r} \vee \varepsilon(G^1 \vee \ldots \vee G^{r-1}).$$

Как следствие,

$$\bigwedge_{j \notin T} \overline{f_j^r} = \bigvee_{j \in T} g_j^r \ \vee \overline{G^r} \ \vee \ \varepsilon(G^1 \vee \ldots \vee G^{r-1}). \tag{5.25}$$

Используя 5.25, получаем формулу

$$\begin{aligned} f_k^r = g_k^r \cdot \bigwedge_{j:\, g_k^r g_j^{r-1} \not\subset f} \overline{f_j^{r-1}} \cdot \ldots \cdot \bigwedge_{j:\, g_k^r g_j^1 \not\subset f} \overline{f_j^1} = \\ = g_k^r \cdot \left(\bigvee_{j:\, g_k^r g_j^{r-1} \subset f} g_j^{r-1} \ \vee \overline{G^{r-1}} \ \vee \ \varepsilon(G^1 \vee \ldots \vee G^{r-2}) \right) \cdot \ldots \\ \ldots \cdot \left(\bigvee_{j:\, g_k^r g_j^2 \subset f} g_j^2 \ \vee \overline{G^2} \ \vee \ \varepsilon(G^1) \right) \cdot \left(\bigvee_{j:\, g_k^r g_j^1 \subset f} g_j^1 \ \vee \overline{G^1} \right). \end{aligned} \tag{5.26}$$

Перейдем непосредственно к проверке (5.24). Любой единичный набор σ функции $\overline{G^t} \cdots \overline{G^{r+1}} \cdot g_k^r \cdot f$ определяет э.к. X^σ, соответствующую ребру, соединяющему k-ю вершину r-го яруса с некоторой l-й вершиной яруса $s < r$, т. е. $X^\sigma = g_k^r g_l^s$. Из (5.26) видно, что $X^\sigma \subset f_k^r$ — на наборе σ обращается в единицу слагаемое

$$g_k^r \cdot \overline{G^{r-1}} \cdots \overline{G^{s+1}} \left(\bigvee_{j:\, g_k^r g_j^s \subset f} g_j^s \right) \overline{G^{s-1}} \cdots \overline{G^1},$$

возникающее при раскрытии скобок в (5.26).

Осталось проверить, что у функции $\overline{G^t} \cdots \overline{G^{r+1}} \cdot f_k^r$ нет других единиц. Предположим противное: эта функция обращается в 1 на наборе σ, при этом $f(\sigma) = 0$. Тогда неизбежно $X^\sigma = g_k^r g_l^s$ при некоторых l и $s < r$. Однако множитель

$$\alpha = g_k^r \cdot \left(\bigvee_{j:\, g_k^r g_j^s \subset f} g_j^s \ \vee \overline{G^s} \ \vee \ \varepsilon(G^1 \vee \ldots \vee G^{s-1}) \right)$$

в произведении (5.26), а значит и само произведение f_k^r, на наборе σ обращается в 0, потому что $\alpha g_l^s = 0$. Приходим к противоречию с предположением. Тем самым доказано (5.24), значит, формулы (5.22) и (5.23) тождественны. □

5.6.3 Верхняя оценка

Теорема 5.7. $C^*(n) \lesssim \sqrt{2} \cdot 2^{n/2}$.

Доказательство. Вычислим функцию, следуя представлению (5.22). Сначала укажем подходящую полную M-систему функций $\{g_i^r\}$. Пусть t — параметр.

Разобьем множество из n переменных на три группы X, Y, Z примерно одинаковой мощности, $|X| \sim |Y| = |Z| \sim n/3$. Руководствуясь леммой 5.6, построим прямое произведение $\{g_k^r(X,Y,Z)\}$ трех систем: полной M-системы $\{u_k^r(X)\}$ ранга (t,p) и двух равномерно полных M-систем $\{v_k^r(Y)\}$, $\{v_k^r(Z)\}$ ранга (t,q). Как указано выше, такие системы существуют при $p \sim \sqrt{2^{|X|+1}/t(t-1)}$ и $q \sim 2^{|Y|/2}$, поскольку согласно закону распределения простых чисел для любого положительного числа x найдется простое число $q = x + o(x)$ при $x \to \infty$. Таким образом, M-система $\{g_k^r(X,Y,Z)\}$ имеет ранг (t, pq^2).

Схема состоит из следующих элементов. Вычисляются всевозможные э.к. X^σ, Y^τ, Z^τ при помощи $2^{|X|} + 2^{|Y|+1}$ конъюнкторов. Следуя ДНФ-представлению, вычислим все функции $u_k^r(X)$, $v_k^r(Y)$, $v_k^r(Z)$, затратив $t(p+2q)$ дизъюнкторов. Поскольку любая функция g_k^r является произведением некоторых трех функций из множеств $\{u_k^r(X)\}$, $\{v_k^r(Y)\}$ и $\{v_k^r(Z)\}$,

каждая функция f_k^r вычисляется при помощи одного конъюнктора, на входы которого подаются, в том числе, отрицания ранее вычисленных функций f_i^s. Всего для вычисления функций f_k^r используется tpq^2 конъюнкторов.

Характеристическая функция G^r вычисляется схемой из трех дизъюнкторов и одного конъюнктора (она является конъюнкцией характеристических функций систем в прямом произведении), т. е. $4t$ элементов достаточно для вычисления всех функций G^r. Завершая вычисления по формуле (5.22), добавим в схему еще $t-2$ дизъюнкторов, $t-2$ конъюнкторов и один внешний дизъюнктор.

Если t полагается слабо растущим относительно n, то асимптотика сложности определяется слагаемым $tpq^2 \sim \sqrt{2} \cdot 2^{n/2}$. □

Повторяя доказательство теоремы 5.7 с выбором постоянного $t = d$, можно показать что оценка леммы 5.3 достигается на схемах глубины $d + O(1)$.

6 Сложность сортировки

6.1 Введение

В этой главе рассматривается классическая задача сортировки набора из n элементов линейно упорядоченного множества при помощи попарных сравнений. Подробное введение в проблематику см. в [4, гл. 3], [26, §5.3], [164, Ch. 2].

Алгоритм сортировки можно изобразить в виде бинарного корневого дерева с ориентацией ребер в направлении от корня. Такое дерево обычно называется деревом сравнений, а также деревом решений или решающей диаграммой. Внутренняя вершина дерева соответствует операции сравнения некоторых двух элементов. В зависимости от результата сравнения алгоритм осуществляет переход по одному из ребер к следующей вершине. Концевая вершина (лист) дерева соответствует полученному в результате сравнений упорядочению входного набора. *Сложностью* алгоритма называется глубина дерева — максимальное расстояние в ребрах между корнем и листом.

Естественно ограничить рассмотрение алгоритмами, не выполняющими избыточных сравнений (т.е. сравнений, не добавляющих новой информации о порядке элементов). В соответствующих таким алгоритмам деревьях каждая возможная перестановка связана ровно с одним листом. В частности, деревья для сортировки n-элементного набора имеют ровно $n!$ листьев.

Пусть $S(n)$ означает минимальную сложность алгоритма сортировки n-элементного набора. Поскольку глубина дерева с m листьями не меньше[37] $\log m$, имеет место простая нижняя оценка

$$S(n) \geq \log(n!) = n \log n - \log e \cdot n + O(\log n), \tag{6.1}$$

$\log e \approx 1.443$. Более того, поскольку даже средняя глубина дерева по всем его листьям не превосходит $\log m$, то оценка (6.1) действительна и для сложности сортировки в среднем по всем $n!$ возможным перестановкам входного набора, см. также [4, 26, 164]. Подобные нижние оценки называются теоретико-информационными.

[37] Как и выше, у двоичных логарифмов основание не указывается.

С точки зрения верхней оценки, в целом лучшим среди известных остается метод Форда—Джонсона [130] (метод бинарных вставок), предложенный более 60 лет назад. Он приводит к соотношению

$$S(n) \le \log(n!) + cn + O(\log n), \tag{6.2}$$

где константа c в зависимости от n варьируется от $\log(3e/8) \approx 0.028$ (в благоприятном случае $n \sim 2^k/3$) до $\log(3/(4\ln 2)) \approx 0.114$ (в неблагоприятном случае $n \sim \ln 2 \cdot 2^k/3$). Метод бинарных вставок доказуемо оптимален при $n \le 15$ и при некоторых бóльших n, см. [172]. Усовершенствование метода, предложенное в [194] около 1980 г., позволяет уточнить константу в неблагоприятном случае до $c \approx 0.105$. Достаточно большую работу по сглаживанию неравномерности оценок метода бинарных вставок проделали авторы [159]. По результатам этой работы можно указать константу для неблагоприятного случая где-то в пределах $0.06 < c < 0.07$, точнее сказать трудно[38]. Для сложности сортировки в среднем оценки сближены гораздо сильнее. Константа в верхней границе сложности недавно была понижена до $c \approx 0.032$ при любом n, см. [142, 125].

Метод Форда—Джонсона основан на процедурах вставки элементов в линейно упорядоченное множество. Вставка каждого элемента выполняется отдельно. Эффективность метода обеспечивается путем подбора мощности целевого множества близкой к степени двойки.

С другой стороны, известно, что совместная вставка даже двух элементов выполняется быстрее, чем раздельная, если мощность n целевого множества находится в пределах $2^k < n < \frac{17}{14} \cdot 2^k - 1$ [134, 141], см. также [26]. Еще выгоднее может быть группировка элементов по 4 или 5 с последующей их сортировкой до вставки, см. [194]. Высокая эффективность метода [142] (с точки зрения средней сложности) также достигается за счет того, что часть вставок выполняется совместно для пар элементов.

В развитие этой идеи автором в [228] предложен алгоритм групповой вставки большого числа элементов, организованный в некотором смысле как система массового обслуживания. Этот подход перекликается с концепцией массового производства (частично упорядоченных наборов), которой

[38] Наиболее сильная форма результата в [159] представлена графически.

следуют лучшие известные алгоритмы выбора элемента заданного порядка [192, 123]. Предлагаемый метод позволяет приблизить сложность вставки, приходящуюся на один элемент, к теоретико-информационной нижней границе, т.е. $\log n + o(1)$ при любом n. Как следствие, на методе сортировки при помощи групповых вставок достигается оценка

$$S(n) \leq \log(n!) + o(n).$$

Разумеется, такая же оценка справедлива и для сложности сортировки в среднем.

Изложение следует работе [228]. В §6.2 дается краткая справка о методе Форда—Джонсона. В §6.3 приводятся некоторые элементарные соображения, лежащие в основе предлагаемого метода. Обобщенная концепция метода групповой вставки изложена в §6.4. Центральная часть метода — стратегия выбора элементов для сравнений — описана в §6.5. В §6.6 выводятся основные результаты о сложности групповой вставки и сортировки.

6.2 Метод бинарных вставок

Операцию сравнения элементов e_1 и e_2 будем обозначать через $e_1?e_2$, а отношение порядка — неравенствами $e_1 < e_2$ или $e_1 > e_2$. Если элементы e_1 и e_2 совпадают, то операция сравнения все равно возвращает либо $e_1 < e_2$, либо $e_1 > e_2$. Без ограничения общности можно считать, что все элементы входного набора различны. Для краткости линейно упорядоченный набор элементов далее будем называть *цепочкой*. Под *длиной* цепочки будем понимать число элементов в ней. Номер элемента цепочки при нумерации с 1 в порядке возрастания будем называть *рангом*.

Напомним суть метода бинарных вставок [130], см. также [26]. Она заключается в следующем: все элементы разбиваются на пары, внутри пар выполняются сравнения, бóльшие элементы пар сортируются. В результате получается частичный порядок, диаграмма[39] которого изображена на рис. 20: элементы в парах обозначены через α_i, β_i, где $\beta_i > \alpha_i$, нумера-

[39] Диаграмма Хассе: ребра соединяют пары элементов с известным порядком, подробнее см., например, в [26].

ция — в порядке возрастания β_i. В нечетном случае $n = 2k + 1$ один из элементов не имеет пары, он обозначен через α_{k+1}.

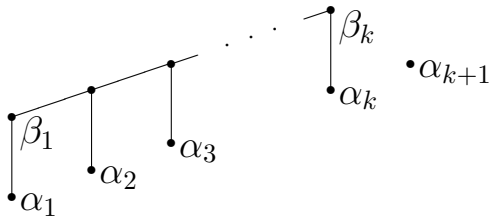


Рис. 20: Частичный порядок в методе бинарных вставок

По построению, элементы $\alpha_1, \beta_1, \beta_2, \ldots, \beta_k$ образуют цепочку. Далее производится последовательная вставка оставшихся элементов α_i в эту цепочку. Первым вставляется элемент α_3 при помощи двух сравнений, затем элемент α_2 — также за два сравнения; далее вставляются все элементы, для которых достаточно трех сравнений (это α_5 и α_4 в таком порядке) и т.д. Всякий раз вставка элемента выполняется в цепочку длины $2^j - 1$, где $j = 2, 3, \ldots$, разве что в финальной серии вставок используется цепочка неполной длины.

Можно проверить, что j сравнений достаточно для вставки каждого из элементов α_i с номерами $u_{j-1} < i \le u_j$, где $u_j = \frac{2^{j+1}+(-1)^j}{3}$. Таким образом, сложность $f(n)$ сортировки n-элементного набора методом Форда—Джонсона при $\lfloor 2^{k+1}/3 \rfloor \le n < \lfloor 2^{k+2}/3 \rfloor$ описывается выражением

$$f(n) = kn - \left\lfloor 2^{k+2}/3 \right\rfloor + \lfloor k/2 \rfloor + 1.$$

Эту формулу удобно переписать в асимптотическом виде. Пусть $n = \tau \cdot 2^{k+1}/3$, где $\tau \in [1,\, 2)$. Тогда

$$f(n) = n \log n - (2/\tau + \log(2\tau/3))\, n + \log n/2 + O(1). \qquad (6.3)$$

Коэффициент в скобках при линейном члене принимает максимальное значение $\log(8/3) \approx 1.415$ при $\tau = 1$, а минимальное значение $\log(4e \ln 2/3) \approx 1.329$ — при $\tau = 2 \ln 2$. Более подробный анализ метода см. в [130, 26].

Введем стандартное обозначение $M(m, n)$ для сложности слияния двух цепочек длины m и n. Одним из результатов работы Шульте-Мёнтинга [194] является оценка $M(5, n) \le 5k - 8$ при $n \le \frac{319}{448} \cdot 2^k - O(1)$. Она

демонстрирует возможность вставки пятерки элементов в цепочку длины n за $5k - 8 + S(5) = 5k - 1$ сравнений, т.е. со средней сложностью $k - 1/5$ на один элемент. Поэтому в общем случае часть вставок финальной серии метода Форда—Джонсона, а именно вставку элементов с номерами после u_{k-1}, выгодно произвести в цепочку меньшей длины со средней сложностью $k - 1/5$ вместо k. Это наблюдение приводит к уточнению оценки (6.3) для всех n, за исключением очень близко расположенных к точкам последовательности $\{2^i/3\}$. Максимальное значение константы в оценке сложности (6.2) при этом понижается до $c \approx 0.105$. Аналогичный результат получен в работе [159] весьма сложным способом. Там вместо вставок в финальной стадии алгоритма используются слияния длинных цепочек методом Кристена [115]. Но при этом оценка для неблагоприятного случая улучшена значительно.

6.3 Предварительные сведения

Для удобства изложения введем еще несколько понятий. В задачах слияния или вставки более длинную цепочку будем называть *главной*. Множество элементов главной цепочки, расположенных между некоторыми элементами α и β, назовем *интервалом* и обозначим через (α, β): если Z — главная цепочка, то $(\alpha, \beta) = \{z \in Z \mid \alpha < z < \beta\}$. В качестве концов интервала также допускаются условные элементы $\pm\infty$: по определению, $-\infty < \alpha < +\infty$ выполнено для всех α. В отличие от длины цепочки, *длину* интервала определим как увеличенное на единицу число элементов в нем, т.е. как число возможностей для вставки нового элемента. Для длины будем использовать обозначение $|(\alpha, \beta)|$.

Обозначим через $Q(a)$ максимальное число n, при котором вставка элемента в интервал длины n выполняется со сложностью не более a или, иначе говоря, $M(1, n-1) \le a$. Тривиально,

$$Q(a) = 2^{\lfloor \log a \rfloor}. \tag{6.4}$$

Аналогично, через $P(a)$ обозначим максимальное n, такое, что $M(2, n-1) \le 2a$: в аргументе функции P записываем число сравнений, приходяще-

еся на один элемент вставляемой пары. Известно [134, 141], что

$$P(a) = \begin{cases} \left\lfloor \frac{17}{14} \cdot 2^a \right\rfloor, & a \in \mathbb{N} \\ \left\lfloor \frac{12}{7} \cdot 2^{a-1/2} \right\rfloor, & (a - 1/2) \in \mathbb{N} \end{cases}. \tag{6.5}$$

Пусть $M(m \times k, n)$ означает сложность слияния m цепочек длины k с цепочкой длины n. Введем обозначения $Q_m(a)$ и $P_m(a)$ для максимального числа n, при котором $M(m \times 1, n-1) \leq am$ и соответственно $M(m \times 2, n-1) \leq 2am$.

По определению, $Q_1(a) = Q(a)$ и $P_1(a) = P(a)$. Также очевидно, что $Q_m(a) \geq Q(a) - m + 1$ и $P_m(a) \geq P(a) - 2(m-1)$.

Первое следствие, оправдывающее введение функции Q_m, можно извлечь из (6.5). Поскольку $Q_2(a) \geq P(a - 1/2)$,

$$Q_2(a) \geq \begin{cases} 2^a - 1, & a \in \mathbb{N} \\ \left\lfloor \frac{17}{14} \cdot 2^{a-1/2} \right\rfloor, & (a - 1/2) \in \mathbb{N} \end{cases}. \tag{6.6}$$

Таким образом, $Q_2(a)$ может быть существенно больше, чем $Q(a)$.

При $m = 1$ осмысленны только полуцелые аргументы у функций $P_m(a)$ и $Q_{2m}(a)$. Скажем, $P_1(r - 1/4) = P_1(r - 1/2) \approx \frac{6}{7} \cdot 2^r$, $r \in \mathbb{N}$. Однако с ростом m открываются новые возможности. Для разминки и иллюстрации основной идеи предлагаемого далее метода выведем оценку

$$P_m\left(r - \frac{1}{4} + \frac{1}{4m}\right) \geq \frac{51}{56} \cdot 2^r - 4m. \tag{6.7}$$

В главной цепочке длины $\left\lceil \frac{51}{56} \cdot 2^r \right\rceil - 4m$ выберем элемент α ранга $\left\lceil \frac{17}{56} \cdot 2^r \right\rceil - 2m + 1$. Тогда интервал $(-\infty, \alpha)$ имеет длину не более $Q_2(r-3/2) - 2(m-1)$ согласно (6.6), а интервал $(\alpha, +\infty)$ — не более $P(r-1) - 2(m-1)$ согласно (6.5).

Произвольная пара $\beta_0 < \beta_1$ обрабатывается следующим образом. Сравниваем β_0 с α. Если $\beta_0 > \alpha$, то пара вставляется в интервал $(\alpha, +\infty)$ за $2(r-1)$ сравнений методом [134, 141]. Иначе, пара разбивается: элемент β_1 вставляется в главную цепочку бинарным методом за r сравнений, а элемент β_0 помещается во временное хранилище — *контейнер*. Если контейнер

был пуст, то переходим к обработке следующей пары. Но если в контейнере оказалось 2 элемента, вставляем их в интервал $(-\infty, \alpha)$ за $2r-3$ сравнений.

По мере выполнения вставок главная цепочка и ее подынтервалы удлиняются. Выбор длин интервалов с запасом $2(m-1)$ обеспечивает возможность вставки в них элементов всех m обрабатываемых пар за планируемое число сравнений: соответственно $2(r-3/2)$ для двух элементов в интервал $(-\infty, \alpha)$ и $2(r-1)$ — для пар в интервал $(\alpha, +\infty)$.

После того, как обработка всех пар завершена, в контейнере еще может оставаться один элемент. Тривиальным способом он вставляется в цепочку за $r-1$ сравнений. Тогда общее число сравнений, выполняемых алгоритмом, в худшем случае оценивается как $m + m(r + r - 3/2) + 1/2 = 2m(r - 1/4) + 1/2$. Соотношение (6.7) доказано.

6.4 Общий метод

В этом параграфе мы опишем общий вид предлагаемого метода групповой вставки вместе со средствами его анализа, попутно вводя необходимые понятия.

Операцию сравнения в любом методе вставки или слияния можно рассматривать как шаг разбиения главной цепочки на все меньшие интервалы, в которых локализуются вставляемые элементы или группы элементов. Это дает возможность рекурсивного построения алгоритма: задача слияния с длинной цепочкой при помощи нескольких сравнений сводится к слияниям с более короткими цепочками.

Опишем основные принципы метода групповой вставки упорядоченных пар элементов в главную цепочку. В главной цепочке выделяются интервалы, которые называются *финишными*. Вставка элемента или пары элементов в финишный интервал выполняется подходящим алгоритмом из доставляющих оценки (6.4), (6.5) и (6.6). Для некоторых финишных интервалов предусмотрены контейнеры. Как в простом примере выше, любой контейнер имеет емкость 2 элемента. Когда в контейнере оказывается два элемента, то выполняется их совместная вставка в соответствующий интервал.

Пары обрабатываются по очереди. Для результата обработки пары могут быть следующие возможности: пара попадает в финишный интервал и применяется метод [134, 141] для окончательной вставки ее в главную цепочку, либо пара в какой-то момент разбивается и элементы обрабатываются по отдельности.

Судьба одиночных элементов может быть следующей: либо элемент попадает в финишный интервал и вставляется в главную цепочку бинарным методом, либо элемент помещается в контейнер. Если в контейнере нет других элементов, то работа продолжается с новой парой. Иначе, выполняется обработка двух элементов из контейнера.

После того как все пары подверглись обработке, любой из вставляемых элементов либо включен в главную цепочку, либо находится один в некотором контейнере. Алгоритм завершается тем, что каждый из оставшихся в контейнерах элементов вставляется в свой интервал тривиальным бинарным методом.

Удельной сложностью обработки одной пары назовем число сравнений, приходящихся на элементы пары, в наихудшем случае при условии, что ни один из элементов пары не остался в контейнере (для этого можно предположить, что все контейнеры не пусты). В ходе подсчета сложности одно сравнение, выполняемое при совместной обработке пары элементов, засчитывается как полсравнения каждому элементу. Удельная сложность вставки пары в примере из предыдущего параграфа равна $2(k - 1/4)$.

Итоговую сложность алгоритма групповой вставки можно оценить сверху как

$$\rho m + C \log(n + 2m) \tag{6.8}$$

где ρ — удельная сложность, m — число вставляемых объектов, n — длина исходной цепочки, а C — общее число контейнеров. Избыточное число сравнений при вставке последнего элемента в контейнере здесь оценивается грубо как $\log(n + 2m)$; более аккуратное рассуждение привело бы к оценке $O(1)$; в реальных алгоритмах эта величина не превосходит 1, как в примере из §6.3.

Теперь введем комплексное понятие стратегии сравнений. Пусть задан некоторый числовой набор $y_1, \ldots, y_d \in \mathbb{R}$ с условием $0 \leq y_1 < y_2 < \ldots <$

$y_d < 1$. Вводятся величины $p_{r,i}$, характеризующие длину интервала $J_{r,i}$, вставка пары в который возможна с удельной сложностью $2(r+y_i)$. Наряду с величинами $p_{r,j}$ будем также использовать удельные величины $x_{r,j} = p_{r,j} \cdot 2^{-r}$.

При любом $i = 1, \ldots, d$ и $r \geq r_0$ задается сокращенное дерево сравнений $T_{r,i}$ для вставки пары в интервал $J_{r,i}$. Под сокращенным деревом мы понимаем поддерево обычного дерева сравнений. Концевые вершины $T_{r,i}$ соответствуют вызову процедуры вставки в интервалы меньшей длины, назовем эти интервалы *терминальными*. Разбиение пары допускается только в концевых вершинах дерева. Так определенные сокращенные деревья сравнений описывают сведение задачи вставки в длинный интервал к задачам меньшей размерности.

Концевой вершине дополнительно ставятся в соответствие ограничения на длины терминальных интервалов, обеспечивающие выполнение оценки удельной сложности вставки пары в интервал $J_{r,i}$. Если при концевой вершине на расстоянии s от корня выполняется вставка пары в терминальный интервал (α', α''), то ограничение имеет вид $|(\alpha', \alpha'')| \leq p_{l,j}$, где $l + y_j + s/2 \leq r + y_i$. Для одиночного элемента β, вставляемого в (α', α''), ограничение может выглядеть либо как $|(\alpha', \alpha'')| \leq p_{l,j}$, либо как $|(\alpha', \alpha'')| \leq 2^l$. В первом случае вклад этого элемента в удельную сложность оценивается как $l + y_j + s/2 + 1/2$, а во втором — как $l + s/2$. Сумма оценок для двух элементов пары не должна превосходить $2(r + y_i)$.

Будем считать, что ранги элементов α', α'' также выражаются линейными комбинациями величин $p_{l,j}$. Некоторые ограничения могут при этом тривиально выполняться (например, в случае тождественности правой и левой частей неравенств). Множество оставшихся определим как *систему ограничений* стратегии.

Итак, под *стратегией сравнений* будем понимать семейство деревьев (фактически алгоритмов) $T_{r,i}$ и систему ограничений, обеспечивающую удельную сложность $2(r + y_i)$ вставки пар в интервалы $J_{r,i}$. *Глубиной* стратегии назовем максимальную глубину деревьев $T_{r,i}$.

Если минимальная удельная сложность вставки пары по всем терминальным интервалам в алгоритме с деревом сравнений $T_{r,i}$ равна $2(r' + y_j)$,

то положим $w_{r,i} = r - r' + 1$.

Мы ограничим рассмотрение *однородными* стратегиями — такими, в которых очередной элемент интервала $J_{r,i}$, выбираемый для сравнения, делит интервал $J_{r,i}$ в отношении, не зависящем r (с точностью до целочисленного округления). Тогда при любом i все деревья $T_{r,i}$ подобны некоторому дереву T_i. Можно считать, что внутренней вершине T_i приписана пара (y, Δ), $y \in \{0, 1\}$, $\Delta \in \mathbb{R}$, кодирующая сравнение $\beta_y ? \alpha$, где (β_0, β_1) — вставляемая пара, а α — элемент ранга $\Delta \cdot |J_{r,i}|$ в интервале $J_{r,i}$. В однородной стратегии $w_{r,i} = w_i$ при всех r. Величину $\max_i w_i$ назовем *шириной* однородной стратегии. Ширина стратегии характеризует максимальную разницу сложности вставки между исходным интервалом и его терминальными подынтервалами. Ниже под стратегией всюду понимается однородная стратегия.

Далее в выкладках позволим длинам интервалов и рангам элементов цепочки принимать нецелые значения, под которыми будут пониматься округления до ближайшего целого в ту или иную сторону. В случае утверждения о сложности вставки в интервал нецелой длины x будем требовать выполнение этого утверждения для интервала длины $\lceil x \rceil$.

Наша ближайшая цель — построить переход от стратегии к конкретному алгоритму. В предположении существования пределов $z_i = \lim_{r\to\infty} x_{r,i}$ ограничения стратегии можно переписать в терминах предельных величин z_i и перейти к задаче линейного программирования. Ниже следует описание и обоснование корректности такого перехода.

Пусть на множестве последовательностей d-мерных вещественных векторов $x_0, x_1, \ldots \in \mathbb{R}^d$, $x_j = (x_{j,1}, x_{j,2}, \ldots, x_{j,d})$, определена система σ нормированных линейных неравенств ширины w вида

$$\sum_{i=1}^{d}\sum_{j=0}^{w-1} a_{i,j,k} \cdot x_{t+j,i} \le b_k, \quad \text{где} \quad \sum_{i=1}^{d}\sum_{j=0}^{w-1} |a_{i,j,k}| = 1, \qquad k = 1, \ldots, K \tag{6.9}$$

при всех $t \ge 0$. Здесь $a_{i,j,k}, b_k$ — вещественные коэффициенты. При помощи формальной подстановки $x_{j,i} = z_i$ для всех i, j в (6.9) получим *приведенную систему*

$$\sum_{i=1}^{d} a_{i,k} \cdot z_i \le b_k, \qquad a_{i,k} = \sum_{j=0}^{w-1} a_{i,j,k}, \qquad k = 1, \ldots, K, \tag{6.10}$$

относительно переменных z_i.

Система (6.10) определяет симплекс $T(\sigma)$ в пространстве $\mathbb{R}^d$. Для произвольного параметра $\varepsilon \geq 0$ определим вложенный симплекс $T^\varepsilon(\sigma) \subset T(\sigma)$ системой неравенств

$$\sum_{i=1}^{d} a_{i,k} \cdot z_i \leq b_k - \varepsilon, \qquad k = 1, \ldots, K. \tag{6.11}$$

Будем говорить, что подпоследовательность $x_l, x_{l+1}, \ldots, x_{l'}$ удовлетворяет (6.9), если ограничения (6.9) выполнены для всех $t = l, \ldots, l' + w - 1$ при подходящем доопределении последовательности векторами $x_{l'+1}, \ldots, x_{l'+w-1}$.

Далее через $||\cdot||$ обозначается l_∞-норма в векторном пространстве (максимум модуля координат вектора), а через $\langle \cdot, \cdot \rangle$ — евклидово скалярное произведение векторов. Нам понадобится простой факт о скорости приближения к заранее известному решению задачи (6.9).

Лемма 6.1. *Пусть $u \in T^\varepsilon(\sigma)$, $v \in T(\sigma)$ и $0 < \varepsilon < ||v - u||$. Тогда на отрезке $[u, v]$ найдется удовлетворяющая системе неравенств σ (6.9) ширины w последовательность $\{x_i\}$ с началом $x_0 = x_1 = \ldots = x_{w-1} = u$, сходящаяся к точке v со скоростью*

$$||v - x_i|| \leq (1 - \varepsilon\Delta)^{i/w-1} \cdot ||v - u||, \qquad \Delta = ||v - u||^{-1}. \tag{6.12}$$

Доказательство. Определим $\varepsilon_i = \varepsilon(1 - \varepsilon\Delta)^i$. При $i \geq 1$ и $j = 0, \ldots, w - 1$ положим

$$x_{iw+j} = x_{(i-1)w+j} + \varepsilon_{i-1}\Delta \cdot (v - u). \tag{6.13}$$

По построению,

$$v - x_{iw+j} = \left(1 - (\varepsilon_0 + \ldots + \varepsilon_{i-1})\Delta\right)(v - u) = (1 - \varepsilon\Delta)^i \cdot (v - u). \tag{6.14}$$

Таким образом, оценка (6.12) выполняется. Осталось проверить, что построенная последовательность удовлетворяет системе σ.

В силу $x_{lw} = x_{lw+1} = \ldots = x_{lw+w-1} \in [u, v] \subset T(\sigma)$ неравенства (6.9) выполнены при всех $t = lw$, где $l \in \mathbb{N}$.

Покажем, что $x_{lw} \in T^{\varepsilon_l}(\sigma)$. При $l = 0$ это гарантировано условием леммы. Обозначим $a_k = (a_{1,k}, \ldots, a_{d,k})$. Тогда при $l > 1$ и при любом $k = 1, \ldots, K$ согласно (6.14) имеет место

$$\langle a_k,\, x_{lw}\rangle = \langle a_k,\, (1-\varepsilon_l/\varepsilon)v + (\varepsilon_l/\varepsilon)u\rangle \le (1-\varepsilon_l/\varepsilon)b_k + (\varepsilon_l/\varepsilon)(b_k-\varepsilon) = b_k - \varepsilon_l$$

в силу линейности скалярного произведения.

Пусть $lw < t < (l+1)w$. В этом случае в левых частях неравенств (6.9) используются только координаты векторов x_{lw} и $x_{(l+1)w}$. Тогда при помощи (6.13) с учетом $x_{lw} \in T^{\varepsilon_l}(\sigma)$ при любом $k = 1, \ldots, K$ получаем

$$\sum_{i=1}^{d}\sum_{j=0}^{w-1} a_{i,j,k}\cdot x_{t+j,i} \le \sum_{i=1}^{d}\sum_{j=0}^{w-1} a_{i,j,k}\cdot x_{lw,i} + \varepsilon_l \cdot \sum_{i=1}^{d}\sum_{j=0}^{w-1} |a_{i,j,k}| \le (b_k - \varepsilon_l) + \varepsilon_l.$$

Следовательно, построенная последовательность удовлетворяет системе неравенств σ. □

Сформулируем основной результат параграфа. Обозначим

$$\pi(a) = \lim_{r\to\infty} P(r+a)\cdot 2^{-r}. \tag{6.15}$$

В частности, $\pi(a) = \frac{17}{14}$ при $0 \le a < \frac{1}{2}$ и $\pi(a) = \frac{12}{7}$ при $\frac{1}{2} \le a < 1$ согласно (6.5).

Лемма 6.2. *Пусть для некоторого набора $y_1, \ldots, y_d \in \mathbb{R}$, где $0 \le y_1 < y_2 < \ldots < y_d < 1$, имеется стратегия сравнений глубины h и ширины w, задающая систему ограничений σ (6.9) на $x_{r,i}$, где $x_{r,i}\cdot 2^r$ — нижняя оценка длины интервала, вставка упорядоченной пары в который выполняется с удельной сложностью $2(r + y_i)$. Пусть $v = (v_1, \ldots, v_d) \in T(\sigma)$, $u = (u_1, \ldots, u_d) \in T^{\varepsilon}(\sigma)$, где $0 < \varepsilon < ||v - u||$, причем $u_i < v_i$ и $u_i \le \pi(y_i)$ для всех $i = 1, \ldots, d$. Обозначим $u_{min} = \min\{u_i\}$. Тогда для любого $r \in \mathbb{N}$ и любых $m \in \mathbb{N}$, $\varphi > 0$, удовлетворяющих условию*

$$m \le \min\{u_{min}, \varepsilon/2\}\cdot 2^{r-\varphi-1} \tag{6.16}$$

выполняется

$$P_m\left(r + y_i + \frac{(r+2)\varphi\cdot 2^{\varphi}}{m}\right) \ge v_i\cdot 2^r - \frac{R}{\varepsilon}\cdot m\cdot 2^{\varphi+2} - R\cdot 2^r\cdot e^{-\frac{\varepsilon}{2R}\left(\frac{\varphi}{w(d+1)(h+1)}-2\right)}, \tag{6.17}$$

где $i = 1, \ldots, d$ и $R = ||v - u||$.

Доказательство. Зададимся некоторым значением r и допустимой парой параметров m и φ. Положим $L = \left\lfloor \frac{\varphi}{(h+1)(d+1)} \right\rfloor$ и $q = r - L$. Из (6.16) и (6.15) вытекает $m \le \frac{6}{7} \cdot 2^{r-\varphi}$, откуда $r > \varphi \ge L$. Следовательно, $q > 0$.

Пусть $\delta \in (0,\, \varepsilon)$ — параметр, который будет выбран позже. На отрезке $[u, v]$ выберем точку $v' = (1-\delta/\varepsilon)v + (\delta/\varepsilon)u$. То, что $v' \in T^{\delta}(\sigma)$, проверяется рассуждением из леммы 6.1: при любом $k = 1, \ldots, K$

$$\langle a_k,\, v' \rangle = \langle a_k,\, (1-\delta/\varepsilon)v + (\delta/\varepsilon)u \rangle \le (1-\delta/\varepsilon)b_k + (\delta/\varepsilon)(b_k - \varepsilon) = b_k - \delta.$$

По определению, ширина системы ограничений не превосходит ширину соответствующей ей стратегии; можно считать, что обе величины равны w. Рассмотрим σ' — систему, получающуюся из σ вычитанием δ из правых частей неравенств (6.9). Тогда имеем $v' \in T(\sigma')$ и $u \in T^{\varepsilon-\delta}(\sigma')$. Применяя лемму 6.1 к системе σ', построим последовательность $\{x_l = (x_{l,1}, \ldots, x_{l,d})\}$, сходящуюся к v', с началом $x_q = \ldots = x_{q+w-1} = u$.

Обозначим $X_{l,i} = x_{q+l,i} \cdot 2^{q+l}$. По условию, заданная стратегия сравнений обеспечивает удельную сложность $2(q + l + y_i)$ вставки пары в интервал длины $X_{l,i}$. Покажем индукцией по $dl+i$, что для вставки m пар в интервал длины

$$X'_{l,i} = X_{l,i} - 2^{(dl+i)h} \cdot 2m \tag{6.18}$$

можно построить алгоритм той же удельной сложности $2(q+l+y_i)$ с $c_{dl+i} \le (dl+i) \cdot 2^{(dl+i)(h+1)}$ контейнерами. (Напомним, что под интервалом длины $X'_{l,i}$ понимается интервал длины $\lceil X'_{l,i} \rceil$.)

Во-первых, убедимся, что $X'_{l,i} > 0$. Действительно, согласно (6.16)

$$X_{l,i} \ge u_i \cdot 2^{q+l} \ge m \cdot 2^{\varphi+1-r+q+l} \ge 2m \cdot 2^{(h+1)(d+1)L+l-L} > 2m \cdot 2^{(d+1)Lh}.$$

Теперь проверим, что запас по длине и число контейнеров достаточны для вставки, затем обеспечим выполнение ограничений (6.9) для подпоследовательности векторов x'_{q+l}, $0 \le l \le L$, с компонентами $x'_{q+l,i} = X'_{l,i} \cdot 2^{-(q+l)}$.

При $0 \le l \le w-1$ (база индукции) требуемое гарантируется условиями леммы на u_i: длины интервалов находятся в области применения оценок (6.5). В этом случае контейнеры не нужны. Для вставки m пар требуется запас длины $2m-2$ по отношению к (6.5) с возможными дополнительными

поправками: 1 на округление длины интервала до $\lceil X'_{0,i} \rceil$ и 1 на ошибку (6.15) при предельном переходе по отношению к (6.5). Поэтому взятый в (6.18) запас минимум в $2m$ достаточен.

При $l \geq w$, отталкиваясь от разбиения интервала длины $X_{l,i}$ на подынтервалы в соответствии со стратегией, построим разбиение интервала длины $X'_{l,i}$. По существу, задача состоит в том, чтобы распределить запас длины $2^{(dl+i)h} \cdot 2m$ между подынтервалами. Будем передвигаться по дереву сравнений от корня к произвольному листу. Пусть в очередной вершине выполняется сравнение с элементом[40] α главной цепочки, и (α', α'') — минимальный интервал, включающий α, границы которого либо являются элементами предшествующих сравнений, либо границами исходного интервала. Тогда запас длины интервала (α', α'') поделим поровну между интервалами (α', α) и (α, α''). В итоге, поскольку дерево сравнений имеет глубину h, любой из подынтервалов, образованный точками сравнений, получит запас длины не менее $2^{(dl+i-1)h} \cdot 2m$ — по индуктивному предположению, достаточный запас для вставки пары с удельной сложностью $2(q + l' + y_j)$ или одиночного элемента — с удельной сложностью[41] $q + l' + y_j + 1/2$, где $l' < l$ или $j < i$. Разумеется, запас выбирается в пределах длины подынтервала, чтобы она оставалась положительной. Также заметим, что терминальные интервалы в алгоритме имеют удельную сложность вставки пары не менее $q + l - w \geq q$, поэтому рекурсивный вызов алгоритма (вставки в терминальный интервал) возможен. Формально не исключается случай, когда стратегия предполагает вставку одиночного элемента во внутренний подынтервал с удельной сложностью не менее $q + l + y_i + 1/2$. В этой ситуации просто вызывается алгоритм вставки в исходный интервал стартовой длины $X'_{l,i}$.

С использованием разбиения интервала длины $X'_{l,i}$ разбиение интервала длины $\lceil X'_{l,i} \rceil$ строится простым округлением рангов всех внутренних элементов разбиения до ближайших целых чисел сверху. При этом длины любых подынтервалов исходного разбиения, в том числе терминальных,

[40] Поскольку стратегия оперирует с нецелыми длинами интервалов, под элементом в этом абзаце понимается условная точка разбиения интервала.

[41] Под удельной сложностью вставки одиночного элемента понимается приходящаяся на него удельная сложность вставки в рамках пары.

изменяются менее чем на 1, т.е. превращаются в округления до ближайших целых сверху или снизу. Таким образом, учет округлений не приводит к изменению оценки удельной сложности алгоритма.

Оценим достаточное число контейнеров. Дерево сравнений имеет не более 2^h листьев, каждому соответствует до двух терминальных интервалов, вставка в которые по предположению индукции выполняется алгоритмами, использующими не более $2c_{dl+i-1}$ контейнеров. Еще не более $2^{h+1}+1$ контейнеров может понадобиться для обслуживания непосредственно терминальных интервалов и исходного интервала. Получаем

$$c_{dl+i} \le 2^{h+1}\left((dl+i-1)\cdot 2^{(dl+i-1)(h+1)}\right) + 2^{h+1} + 1 \le (dl+i)\cdot 2^{(dl+i)(h+1)}.$$

Теперь проверим выполнение ограничений σ для $x'_{l,i}$. Формально положим $x'_{j,i} = x_{j,i}$ при $j > r$. Определим $\delta = 2^{\varphi-r+1}\cdot m$. По условию (6.16) имеет место $\delta < \varepsilon/2$. Кроме того, в силу (6.18) и опредедений параметров φ и δ выполнено $x'_{j,i} - x_{j,i} \in (-\delta, 0]$ для всех $j \ge q$. Значит, при всех $t \ge q$ и любом $k = 1, \ldots, K$ получаем

$$\sum_{i=1}^{d}\sum_{j=0}^{w-1} a_{i,j,k}\cdot x'_{t+j,i} \le \sum_{i=1}^{d}\sum_{j=0}^{w-1} a_{i,j,k}\cdot x_{t+j,i} + \delta\cdot\sum_{i=1}^{d}\sum_{j=0}^{w-1}|a_{i,j,k}| \le (b_k-\delta)+\delta.$$

Таким образом, последовательность $\{x'_l = (x'_{l,1}, \ldots, x'_{l,d})\}$ удовлетворяет системе ограничений σ при $t \ge q$. Тем самым доказан индуктивный переход: значит, вставка в интервал длины $X'_{L,i}$ действительно выполняется с удельной сложностью $2(r+y_i)$ сравнений.

Остается проверить оценку (6.17). Выбором параметра L обеспечено

$$c_{dl+i} \le (d+1)L\cdot 2^{(d+1)(h+1)L} \le \varphi\cdot 2^{\varphi}. \tag{6.19}$$

Из леммы 6.1 следует

$$\begin{aligned}
|x'_{r,i} - v_i| &\le |x'_{r,i} - x_{r,i}| + |x_{r,i} - v'_i| + |v'_i - v_i| \le \\
&\le \delta + (1-(\varepsilon-\delta)/R)^{L/w-1}\cdot R + (\delta/\varepsilon)\cdot|v_i - u_i| \le \\
&\le \delta + e^{-(\varepsilon-\delta)(L/w-1)/R}\cdot R + (\delta/\varepsilon)\cdot R \le \\
&\le \delta + e^{-\frac{\varepsilon}{2R}\left(\frac{\varphi}{w(d+1)(h+1)}-2\right)}\cdot R + (\delta/\varepsilon)\cdot R < \\
&< \frac{2R}{\varepsilon}\cdot m\cdot 2^{\varphi-r+1} + e^{-\frac{\varepsilon}{2R}\left(\frac{\varphi}{w(d+1)(h+1)}-2\right)}\cdot R,
\end{aligned} \tag{6.20}$$

где также использовалось известное неравенство $(1-x)^{1/x} \le e^{-1}$ при $x \in (0,\,1)$. Теперь (6.17) получается подстановкой (6.19) и (6.20) в оценку (6.8), в которой можно положить $n + 2m \le X_{L,i} = x_{r,i} \cdot 2^r < 2^{r+2}$, поскольку из теоретико-информационных соображений $x_{r,i} \le 2^{y_i+1/2} < 4$. □

Оценки в обеих леммах весьма грубы: точность приносится в жертву простоте или общности изложения. В частности, существенно меньшее число контейнеров требуется конкретной стратегии, к которой будет применяться лемма 6.2.

6.5 Универсальная стратегия

В этом параграфе строится стратегия сравнений для доказательства асимптотической оценки высокой точности. Хорошо известно (и вполне очевидно), что для построения оптимального алгоритма сортировки некоторого частично упорядоченного множества желательно, чтобы при каждом сравнении множество возможных доупорядочиваний разбивалось на две примерно равные части, в зависимости от результата сравнения. Такой ход мысли ведет к градиентному методу: сначала выполняем некоторое число сравнений, руководствуясь делением множества исходов как можно точнее пополам; затем полученный частичный порядок обрабатываем каким-то простым способом.

Теоретико-информационная нижняя оценка сложности вставки пары в интервал длины $n-1$ составляет $\log(n(n-1)/2) < 2(\log n - 1/2)$. Представим идеальный случай, предположив, что с удельной сложностью $2a$ пару можно вставить в интервал длины $Y(a) = (1-\lambda) \cdot 2^{a+1/2}$ при любом достаточно большом a, где $\lambda \ge 0$ — не зависящий от a параметр, который будет определен позднее.

Рассмотрим одну из градиентных стратегий для алгоритма вставки пары $\beta_0 < \beta_1$ в интервал длины $Y(a)$. Стратегия характеризуется параметром $s \in \mathbb{N}$. Сразу заметим, что при построении стратегии длины интервалов считаются действительными числами: необходимость округления до целых учитывается при переходе от стратегии к алгоритму леммой 6.2.

Сначала отдельно рассмотрим задачу выбора очередного элемента для

сравнения. Пусть в результате серии сравнений частичный порядок таков, что элемент β_1 принадлежит интервалу I_1, а β_0 принадлежит интервалу $I_0 \cup I_1$, где $I_0 \cap I_1 = \emptyset$ и $|I_0| = b \cdot |I_1|$. Тогда наилучший выбор для сравнения с β_0 — это элемент[42] α, имеющий ранг $(b/2+1/4) \cdot |I_1|$, считая от левого края интервала I_0. (Если $b > 1/2$, то $\alpha \in I_0$.) А наилучший выбор для сравнения с β_1 — это элемент α' ранга $\sqrt{b^2+b+1/2} \cdot |I_1|$, см. рис. 21. Вообще, если бы мы хотели разбить множество исходов в отношении $x : (1-x)$, то для сравнения с β_0 и β_1 нужно было бы выбрать соответственно элементы ранга $x(b+1/2) \cdot |I_1|$ или ранга $\sqrt{b^2+x(2b+1)} \cdot |I_1|$.

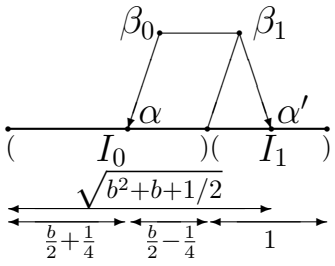


Рис. 21: Оптимальные варианты для очередного сравнения

Исходное разбиение. Не ограничивая общности, считаем, что первое сравнение имеет вид $\beta_1 ? \alpha_1$, тогда в качестве α_1 следует взять элемент ранга $Y(a-1/2) = (1-\lambda) \cdot 2^a$. Если $\beta_1 < \alpha_1$, то вызывается алгоритм вставки пары в интервал $(-\infty, \alpha_1)$ длины $Y(a-1/2)$. Иначе, выполняем сравнение $\beta_0 ? \alpha_2$, где α_2 — элемент оптимального ранга $(1-\lambda) \cdot \frac{\sqrt{2}+1}{4} \cdot 2^a$. Причина, по которой второе сравнение выполняется с элементом β_0, будет разъяснена чуть позже.

Если $\beta_0 < \alpha_2$, то элемент β_0 вставляется в интервал $(-\infty, \alpha_2)$ с удельной сложностью $a - 2 + \log(\sqrt{2}+1)$, а элемент β_1 вставляется в интервал $(\alpha_1, +\infty)$ длины $(1-\lambda)(\sqrt{2}-1) \cdot 2^a$ с удельной сложностью $a + \log(\sqrt{2}-1)$. Следовательно, в этом случае вставка пары выполняется за $2a$ удельных сравнений.

Если же $\beta_0 > \alpha_2$, то $\beta_1 \in I_1$ и $\beta_0 \in I_0 \cup I_1$, где $I_1 = (\alpha_1, +\infty)$ и $I_0 = (\alpha_2, \alpha_1)$. Пусть $b = |I_0|/|I_1| = \frac{2\sqrt{2}+1}{4}$. При помощи s сравнений, разбивая каждый раз множество исходов пополам, мы локализуем β_1 в одном

[42] Здесь и ниже под элементом обычно понимается точка разбиения интервала.

из 2^s подынтервалов интервала I_1. В соответствии с приведенными выше формулами границами этих подынтервалов являются элементы α'_k ранга $b_k \cdot |I_1|$, $b_k = \sqrt{b^2 + k(2b+1)\cdot 2^{-s}}$, $k = 0, \ldots, 2^s - 1$, и условная точка $\alpha'_{2^s} = +\infty$. Обозначим подынтервал $(\alpha'_{k-1}, \alpha'_k)$ через J_k. Общая схема разбиения показана на рис. 22.

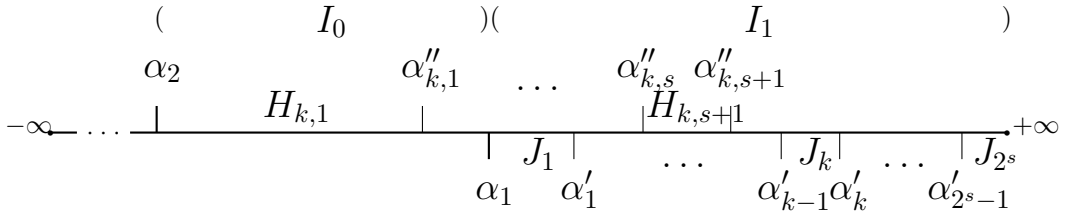


Рис. 22: Универсальное разбиение на интервалы

Пусть элемент β_1 определен в интервал J_k. Следующий этап — вставка элемента β_0. По очереди сравниваем его с элементами $\alpha''_{k,1}, \ldots, \alpha''_{k,s+1}$, определяемыми следующим образом. Положим $\alpha''_{k,0} = \alpha_2$. Обозначим через $H_{k,j}$ интервал $(\alpha''_{k,j-1}, \alpha''_{k,j})$. Длина интервала $H_{k,j}$ выбирается равной

$$|H_{k,j}| = (1-\lambda)\cdot 2^{2a-s-2-j-\log(|J_k|/(1-\lambda))}. \tag{6.21}$$

Если $\beta_0 > \alpha''_{k,1}$, то выполняем сравнение $\beta_0 ? \alpha''_{k,2}$; если $\beta_0 > \alpha''_{k,2}$, то сравниваем β_0 с $\alpha''_{k,3}$ и т.д. Как только $\beta_0 < \alpha''_{k,j}$ элементы β_0 и β_1 вставляются независимо в интервалы $H_{k,j}$ и J_k с суммарной удельной сложностью $2a - s - 2 - j$. В таком случае общая удельная сложность вставки пары составит $2a$.

Наконец, если $\beta_0 > \alpha''_{k,s+1}$, то оба элемента β_0, β_1 принадлежат интервалу $(\alpha''_{k,s+1}, \alpha'_k)$. Хотелось бы вставить пару в этот интервал ценой $2a-2s-3$ сравнений. Это было бы возможно в случае, когда, скажем, $\alpha''_{k,s+1} = \alpha_{k-1}$. Но равенство, вообще говоря, не выполняется, и интервал оказывается длиннее требуемого. Оценим превышение длины. Интервал J_k имеет длину

$$|J_k| = (b_k - b_{k-1})\cdot|I_1| = \frac{2b+1}{2^s(b_k+b_{k-1})}\cdot|I_1| = (1-\lambda)\cdot\frac{\sqrt{2}+1}{2^{s+1}(b_k+b_{k-1})}\cdot 2^a. \tag{6.22}$$

Тогда из (6.21) следует, что

$$|H_{k,j}| = (1-\lambda)\cdot 2^{a-1-j-\log(\sqrt{2}+1)+\log(b_k+b_{k-1})} = \frac{b_k+b_{k-1}}{2^{j+1}}\cdot|I_1|. \tag{6.23}$$

Значит, сумма длин интервалов $H_{k,j}$ при фиксированном k равна

$$\sum_{j=1}^{s+1} |H_{k,j}| = (1 - 2^{-s-1}) \cdot \frac{b_k + b_{k-1}}{2} \cdot |I_1|. \tag{6.24}$$

Получаем

$$|(\alpha''_{k,s+1}, \alpha'_k)| = b_k \cdot |I_1| - \sum_{j=1}^{s+1} |H_{k,j}| = \left(b_k - (1 - 2^{-s-1})(b_k + b_{k-1})/2\right) |I_1| =$$
$$= \left(\frac{b_k - b_{k-1}}{2} + \frac{b_k + b_{k-1}}{2^{s+2}}\right) |I_1| < \left(\frac{2b+1}{2^{s+1} \cdot 2b} + \frac{b+1}{2^{s+1}}\right) |I_1| < \frac{7}{2^{s+2}} \cdot |I_1|.$$

Следовательно,

$$|(\alpha''_{k,s+1}, \alpha'_k)| - Y(a - s - 3/2) < \frac{7}{2^{s+2}} \cdot |I_1| - (1 - \lambda) \cdot 2^{a-s-1} =$$
$$= \frac{7 - 2(\sqrt{2}+1)}{2^{s+2}} \cdot |I_1| < \frac{\ln 2}{2^s} \cdot |I_1|. \tag{6.25}$$

Полученная оценка понадобится чуть позже. Теперь мы модифицируем описанную выше стратегию сравнений, позволив некоторое отступление от идеала, т.е. деления множества исходов ровно пополам. А именно, мы изменим распределение длин подынтервалов J_k внутри интервала I_1, уменьшив (оценочную) сложность вставки элементов в каждый из них, тем самым увеличив допустимую сложность вставки в интервалы $H_{k,j}$ и удлинив их. Для этого у нас имеется резерв — использовать для сложности вставки элемента простую оценку $Q(a)$ вместо $Y(a - 1/2)$, которая лучше, когда значение a близко снизу к целому числу[43]. Длина интервала I_1 не будет изменена.

Заметим, что длины интервалов J_k с ростом k изменяются в пределах от $\frac{2b+1}{2b} \cdot 2^{-s} \cdot |I_1|$ до $\frac{2b+1}{2b+2} \cdot 2^{-s} \cdot |I_1|$, см. (6.22). В силу $b < 1$ выполнено $\left[\frac{2}{3}, \frac{4}{3}\right] \subset \left[1 - \frac{1}{2b+1}, 1 + \frac{1}{2b+1}\right]$, поэтому при любом (достаточно большом) a часть интервалов J_k имеет длину, близкую к степени двойки. Выполненное вторым сравнение $\beta_0 ? \alpha_2$ преследовало именно цель уменьшения показателя b до величины меньше 1.

[43] Едва ли можно построить эффективный алгоритм вставки, не прибегая к оценкам типа $Q(a)$, что видно на примерах алгоритмов [134, 141] и простого алгоритма из §6.3.

Коррекция разбиения. На этом шаге мы вносим поправки в длины интервалов J_k и $H_{k,j}$, корректируя стратегию, и одновременно дискретизируем задачу (новые интервалы будут обозначаться как J_k^* и $H_{k,j}^*$). Введем параметр дискретизации $d \in 2\mathbb{N}$ и положим $y_i = i/d$, $i = 0, \ldots, d-1$. По-прежнему мы ищем оценку длины интервала, достаточной для вставки пары с удельной сложностью $2(r+y_i)$, в виде $Y(r+y_i) = (1-\lambda)\cdot 2^{r+y_i+1/2}$.

Задавшись некоторым $a = r + y_i$, $r \in \mathbb{N}$, опишем модификацию исходного разбиения. Четность параметра d позволяет выбрать элемент α_1 для первого сравнения так же, как и в идеальной схеме. Но сложность вставки элемента в интервал $I_1 = (\alpha_1, +\infty)$ теперь приходится оценить как $a + \lceil d\cdot\log(\sqrt{2}-1)\rceil/d < a + \log(\sqrt{2}-1) + 1/d$. Как следствие, вместо α_2 приходится выбирать элемент α_2^* ранга не выше $(1-\lambda)\cdot 2^{a-2-\lceil d\cdot\log(\sqrt{2}-1)\rceil/d}$. Руководствуясь соображением последующего применения леммы 6.2, создадим дополнительный запас сложности $1/d$ на втором сравнении и выберем в качестве α_2^* элемент чуть меньшего ранга

$$(1-\lambda)\cdot 2^{a-2-\lceil d\cdot\log(\sqrt{2}-1)\rceil/d-1/d} > (1-\lambda)\cdot\frac{\sqrt{2}+1}{4}\cdot 2^{a-2/d}. \tag{6.26}$$

Тогда, используя неравенство $e^x \geq 1+x$, получаем

$$|I_0^*| = |(\alpha_2^*, \alpha_1)| \leq (1-\lambda)\cdot 2^a\left(1 - \frac{\sqrt{2}+1}{4}\cdot 2^{-2/d}\right) \leq$$
$$\leq (1-\lambda)\cdot 2^a\left(\frac{3-\sqrt{2}}{4} + \frac{\sqrt{2}+1}{2d}\cdot\ln 2\right).$$

Следовательно,

$$0 \leq |I_0^*| - |I_0| \leq \frac{(3+2\sqrt{2})}{2d}\cdot\ln 2\cdot|I_1|. \tag{6.27}$$

При d достаточно большом, $|I_0^*| < |I_1|$.

Выполним перераспределение интервалов J_k внутри I_1. Введем параметр $\mu > 0$. Потребуем, чтобы удельная сложность вставки элемента в J_k^* была меньше сложности вставки в J_k в идеальной стратегии хотя бы на $\left(2 - \frac{k}{2^s}\right)\mu$ сравнений, $k = 1, \ldots, 2^s$. Напомним, что в идеальной стратегии

сложность оценивается как $\log(|J_k|/(1-\lambda))$. Прибегать к неравномерности сужения вынуждает эффект смещения правых границ интервалов J_k. При равномерном сужении в $1+\mu$ раз ввиду $|I_0| < |I_1|$ компенсация этого смещения, вообще говоря, была бы затруднена при малых k: сумма длин интервалов $H_{k,j}$ выросла бы примерно на $\mu \cdot |I_0|$, тогда как точка α'_{k+1} (правая граница интервала J_k) могла бы сместиться вправо приблизительно на $\mu \cdot |I_1|$. Поэтому для интервалов с большими номерами выбирается меньший коэффициент сужения.

Сужение. Мы собираемся сузить почти все интервалы J_k, но это сужение будет компенсировано расширением небольшого числа интервалов, длины которых близки к степеням двойки, с опорой на простую оценку (6.4). Сначала оценим сверху величину θ_p общего сужения интервалов с номерами от p до 2^s. Перед этим заметим, что для величины b_k выполняются простые соотношения (проверяются возведением в квадрат)

$$b + \frac{k}{2^s} \cdot b \le b_k \le b + \frac{k}{2^s} \cdot \frac{2b+1}{2b^2}. \tag{6.28}$$

С учетом дискретизации, длину суженного интервала J_k^* можно выбрать в пределах

$$2^{-\left(2-\frac{k}{2^s}\right)\mu - \frac{1}{d}} \cdot |J_k| \le |J_k^*| \le 2^{-\left(2-\frac{k}{2^s}\right)\mu} \cdot |J_k|. \tag{6.29}$$

Исходя из пессимистического предположения, что все интервалы сужают-

ся, снова используя неравенство $e^x \geq 1+x$, а также (6.22) и (6.28), получаем

$$\theta_p \leq \sum_{k=p}^{2^s}(|J_k| - |J_k^*|) \leq \sum_{k=p}^{2^s}|J_k|\left(1 - 2^{-\left(2-\frac{k}{2^s}\right)\mu-\frac{1}{d}}\right) \leq$$

$$\leq \sum_{k=p}^{2^s}|J_k|\left(\left(2-k2^{-s}\right)\mu + 1/d\right)\ln 2 =$$

$$= \left(2\mu + \frac{1}{d}\right)(b+1-b_{p-1})\ln 2\cdot|I_1| - \mu 2^{-s}\ln 2\cdot\sum_{k=p}^{2^s}k|J_k| =$$

$$= \left(2\mu + \frac{1}{d}\right)(b+1-b_{p-1})\ln 2\cdot|I_1| - \mu 2^{-2s}(2b+1)\ln 2\cdot\sum_{k=p}^{2^s}\frac{k}{b_k+b_{k-1}}|I_1| \leq$$

$$\leq \left(2\mu\left(1 - \frac{p-1}{2^s}\cdot b\right) - \mu 2^{-2s}\cdot\frac{2b+1}{2(b+1)}\cdot\frac{2^{2s}-p^2}{2} + \frac{1}{d}\right)\ln 2\cdot|I_1| =$$

$$= \left(\left(2 - 2b\cdot\frac{p-1}{2^s} - \frac{2b+1}{4(b+1)}\cdot\left(1-\frac{p^2}{2^{2s}}\right)\right)\mu + \frac{1}{d}\right)\ln 2\cdot|I_1|.$$

В частности, для суммарного сужения всех интервалов J_k имеем оценку

$$\theta_1 \leq (2\mu + 1/d)\ln 2\cdot|I_1|, \tag{6.30}$$

а при $p \geq 2$ — оценку

$$\theta_p \leq \left(\left(2 - \frac{2b+1}{4(b+1)}\right)\mu + \frac{1}{d}\right)\ln 2\cdot|I_1|, \tag{6.31}$$

поскольку $p^2/2^s \leq p \leq 2(p-1)$ и $\frac{2b+1}{4(b+1)} < b = \frac{2\sqrt{2}+1}{4}$.

Расширение. Теперь оценим снизу величину возможной компенсации. Если

$$(1-\lambda)\cdot 2^{l+2\mu} \leq |J_k| < 2^l, \quad l \in \mathbb{N},$$

то сложность вставки элемента пары в интервал J_k в идеальной стратегии пока оценивалась как минимум в $l + 2\mu$. Тем не менее, длину интервала можно увеличить до 2^l, при этом уменьшив оценку сложности по меньшей мере на 2μ, что и требуется.

Сначала проверим, что если λ и μ не слишком велики, то при некотором $l \in \mathbb{N}$ справедливо

$$|J_{2^s}| \leq (1-\lambda)\cdot 2^{l+2\mu} < 2^l \leq |J_1|. \tag{6.32}$$

При $\mu \ll \lambda$ для этого достаточно выполнения условия $|J_1|/|J_{2^s}| \geq \frac{2}{1-\lambda}$. Используя (6.22) и (6.28), выводим

$$\frac{|J_1|}{|J_{2^s}|} = \frac{b_{2^s} + b_{2^s-1}}{b_1 + b_0} \geq \frac{2b + 2 - \frac{b}{2^s}}{2b + \frac{2b+1}{2b^2 \cdot 2^s}} \geq$$
$$\geq \frac{b+1}{b}\left(1 - \frac{b}{2(b+1)\cdot 2^s}\right)\left(1 - \frac{2b+1}{4b^3 \cdot 2^s}\right) > \frac{b+1}{b}\left(1 - 2^{1-s}\right),$$

где также использовались простые неравенства $\frac{1}{1+x} \geq 1-x$ и $(1-x)(1-y) \geq 1-x-y$, справедливые при $x, y \geq 0$. Подстановкой численных значений легко проверяется, что полученная оценка превосходит $\frac{2}{1-\lambda}$, скажем, при любых $s \geq 10$ и $\lambda \leq 1/50$.

Итак, некоторый отрезок $g = \left[(1-\lambda)\cdot 2^{l+2\mu},\, 2^l\right]$ лежит внутри отрезка $[|J_{2^s}|,\, |J_1|]$. Оценим возможное увеличение длины только для тех интервалов J_k, длины которых оказались в g. Напомним, что длины интервалов монотонно убывают с ростом k, см. (6.22). Длину отрезка g оценим снизу при помощи (6.22) и справедливого при $0 \leq x \leq 1$ неравенства $2^x \leq 1 + x$ как

$$\left(1 - (1-\lambda)2^{2\mu}\right)\cdot 2^l \geq (1 - (1-\lambda)(1+2\mu))\cdot |J_{2^s}| \geq$$
$$\geq L = (\lambda - 2\mu)\,\frac{2b+1}{2(b+1)\cdot 2^s}\cdot |I_1|. \quad (6.33)$$

Теперь оценим сверху разность длин соседних интервалов:

$$\frac{2^s}{(2b+1)\cdot |I_1|}\cdot(|J_k| - |J_{k+1}|) = \frac{1}{b_k + b_{k-1}} - \frac{1}{b_{k+1} + b_k} \leq$$
$$\leq \frac{b_{k+1} - b_{k-1}}{4b^2} = \frac{2(2b+1)}{4b^2(b_{k+1} + b_{k-1})\cdot 2^s} \leq \frac{2b+1}{4b^3 \cdot 2^s}.$$

Следовательно, при любом k

$$|J_k| - |J_{k+1}| \leq \Delta = \frac{(2b+1)^2}{4b^3 \cdot 2^{2s}}\cdot |I_1|. \quad (6.34)$$

Легко проверить, что если множество точек делит отрезок длины L на подотрезки с длинами не более Δ, то сумма расстояний от этих точек до границы отрезка не меньше $\frac{L(L-\Delta)}{2\Delta}$. Подставляя в эту формулу значения L и Δ, приходим к оценке

$$\frac{L(L-\Delta)}{2\Delta} \geq \left((\lambda-2\mu)^2 \cdot \frac{b^3}{2(b+1)^2} - (\lambda-2\mu)\cdot\frac{2b+1}{4(b+1)\cdot 2^s}\right)|I_1| =$$
$$= (\lambda-2\mu)\left(\lambda-2\mu-\frac{(2b+1)(b+1)}{2b^3\cdot 2^s}\right)\frac{b^3}{2(b+1)^2}\cdot|I_1| >$$
$$> (\lambda-2\mu)\left(\lambda-2\mu-2^{2-s}\right)\frac{b^3}{2(b+1)^2}\cdot|I_1|.$$

Так мы оценили снизу максимум суммарного возможного расширения интервалов J_k с длинами из отрезка g (до длины 2^l).

Компенсация сужения остальных интервалов J_k осуществима при $\frac{L(L-\Delta)}{2\Delta} \geq \theta_1$. Используя (6.30), достаточное для этого условие можно теперь записать как

$$(\lambda-2\mu)\left(\lambda-2\mu-2^{2-s}\right)\frac{b^3}{2(b+1)^2} \geq \left(2+\frac{1}{d}\right)\ln 2\cdot\mu. \tag{6.35}$$

Пока ограничимся тривиальным замечанием: при достаточно большом s можно задать параметры в виде[44] $d \asymp 2^s$, $\lambda \asymp 2^{-s/2}$, $\mu \asymp 2^{-s}$ так, что (6.35) будет удовлетворено.

Сведение. Пришло время проверить, позволяют ли предпринятые меры компенсировать дефицит длины интервалов $H_{k,j}$, отраженный в соотношении (6.25). Обратим внимание на изменение условий, в которых выполнялась оценка (6.25). Для величины $Y(a-s-3/2)$ используется прежняя оценка, но при этом произошло удлинение интервала I_0, см. (6.27), а правая граница интервала J_k могла сместиться правее на величину, оцененную сверху как θ_{k+1}, см. (6.31).

Таким образом, мы требуем, чтобы при любом k прирост суммарной длины интервалов $H_{k,j}$ составил не менее

$$|I_0^*| - |I_0| + \theta_{k+1} + \frac{\ln 2}{2^s}\cdot|I_1| + \delta\cdot|I_1|, \tag{6.36}$$

где $\delta\cdot|I_1|$ — дополнительный запас, который потребуется позже для применения леммы 6.2.

[44] Символ $\asymp$ означает равенство порядков роста.

По построению, $|H^*_{k,j}| \geq |H_{k,j}| \cdot 2^{\left(2-\frac{k}{2^s}\right)\mu-\frac{1}{d}}$ с учетом поправки на дискретизацию. Поэтому, подставляя (6.24) и используя (6.28), получаем

$$\sum_{j=1}^{s+1}(|H^*_{k,j}| - |H_{k,j}|) \geq \left(2^{\left(2-\frac{k}{2^s}\right)\mu-\frac{1}{d}} - 1\right) \cdot \sum_{j=1}^{s+1} |H_{k,j}| \geq$$

$$\geq \left(\left(2-\frac{k}{2^s}\right)\mu - \frac{1}{d}\right) \ln 2 \cdot \left(1 - \frac{1}{2^{s+1}}\right) \cdot \frac{b_k + b_{k-1}}{2} \cdot |I_1| \geq$$

$$\geq \left(2-\frac{k}{2^s}\right)\left(1-\frac{1}{2^{s+1}}\right)\left(1+\frac{k-\frac{1}{2}}{2^s}\right)\mu b \cdot \ln 2 \cdot |I_1| - \frac{b+1}{d} \cdot \ln 2 \cdot |I_1| \geq$$

$$\geq \left(2 + \frac{k-2}{2^s} - \frac{k^2}{2^{2s}}\right)\mu b \cdot \ln 2 \cdot |I_1| - \frac{b+1}{d} \cdot \ln 2 \cdot |I_1| \geq$$

$$\geq \left(\left(2-2^{1-s}\right)\mu b - \frac{b+1}{d}\right) \cdot \ln 2 \cdot |I_1|.$$

Используя эту оценку, а также (6.27) и (6.31), достаточное для выполнения (6.36) условие можно записать как

$$\left(2-2^{1-s}\right)\mu b - \frac{b+1}{d} \geq \frac{(3+2\sqrt{2})}{2d} + \left(2 - \frac{2b+1}{4(b+1)}\right)\mu + \frac{1}{d} + \frac{1}{2^s} + \delta \cdot \log e. \tag{6.37}$$

Положим $d = 2^s$ и $\delta = 2^{-s}$. Тогда, чтобы удовлетворить (6.37), можно выбрать $\mu = 30 \cdot 2^{-s}$ (принимаем, что $s \geq 10$). Условие (6.35) при этом будет выполнено, например, если $\lambda = 20 \cdot 2^{-s/2} + 64 \cdot 2^{-s}$. Обеспечивающее (6.32) частное условие $\lambda \leq 1/50$ выполнено при любом $s \geq 20$.

Описание стратегии. Фиксируя выбранные значения параметров d, δ, μ, λ, получаем стратегию вставки с системой ограничений, которую обозначим через $\sigma[s]$. Стратегия заключается в том, что элементы главной цепочки, участвующие в сравнениях, разбивают цепочку на интервалы, вставка в которые выполняется с такой сложностью, которая получается в вышеприведенном расчете. Опишем систему ограничений стратегии.

Пусть $p(a)$ является оценкой длины интервала с удельной сложностью вставки пары $2a$ сравнений, а $p_c(J)$ — оценкой длины интервала с удельной сложностью вставки элемента на c выше, что и в некоторый интервал J из построенного выше разбиения. Для вставки в некоторые интервалы J мы выбираем оценки типа $Q(a)$, а для остальных вставок — оценки типа $p(a)$

при подходящих a.

Первое сравнение, в силу выбора элемента α_1, не порождает ограничений. Для второго сравнения вводится неравенство

$$p(a) \le p(a - 1/2) + p_{1/d}(I_1). \tag{6.38}$$

Здесь мы используем запас сложности, обеспеченный выбором элемента α_2^*, см. (6.26), и можем позволить вставку в интервал I_1 с чуть большей удельной сложностью. Это понадобится только для формального удовлетворения условиям применения леммы 6.2.

Следующие s сравнений, локализующие β_1 в одном из интервалов J_k^*, относятся к внутренним вершинам дерева сравнений. Для того, чтобы обеспечить в дальнейшем вставку элемента β_1 в любой из этих интервалов с заданной сложностью, вводится ограничение

$$p(a) \le p(a - 1/2) + \sum_{k=1}^{2^s} p_0(J_k^*). \tag{6.39}$$

Ограничение всего одно, так как длины интервалов J_k^*, $k > 1$, мы можем задать свободно, выбирая подходящим образом пограничные элементы $\alpha_k'^*$, тогда только длина интервала J_1^* будет определяться длиной остальных интервалов.

Длины интервалов $H_{k,j}^*$ тоже устанавливаются свободно (выбором элементов для сравнений), и лишь вставки после $2s+3$ сравнений порождают ограничения вида

$$p(a) \le p_0(I_{-1}) + \sum_{j=1}^{s+1} p_0(H_{k,j}^*) + p(a - s - 3/2) + \sum_{i=k+1}^{2^s} p_0(J_i^*), \tag{6.40}$$

где $I_{-1} = (-\infty, \alpha_2^*)$.

Таким образом, система ограничений $\sigma[s]$ складывается из (6.38), (6.39) и (6.40) при всевозможных $a = r + i/d$. Для перехода к форме записи из определения стратегии следует в указанных неравенствах заменить выражения $p(x)$ и $p_c(J)$ величинами $p_{r,i}$ или числовыми значениями в случаях, когда применяется оценка $Q(a)$.

Глубина стратегии с системой ограничений $\sigma[s]$ равна $2s+3$. Проверим, что ширина (стратегии и, следовательно, системы $\sigma[s]$) не превосходит $s+4$. Действительно,

$$\frac{|I_1|}{p(a)} = \sqrt{2} - 1 > \frac{1}{2^2}, \qquad \frac{|I_{-1}|}{p(a)} \geq \frac{\sqrt{2}+1}{4 \cdot 2^{2/d}} > \frac{1}{2}$$

согласно (6.26) (напомним, что $p(a) = Y(a)$). Далее, в силу (6.29) и (6.22)

$$\frac{|J_k^*|}{p(a)} \geq \frac{|J_k|}{p(a) \cdot 2^{2\mu+1/d}} \geq \frac{\sqrt{2}+1}{(b+1) \cdot 2^{s+2+2\mu+1/d}} > \frac{1}{2^{s+2}}.$$

Наконец, из (6.23) следует

$$\frac{|H_{k,s+1}^*|}{p(a)} \geq \frac{|H_{k,s+1}|}{p(a)} \geq \frac{b}{2^{s+1}} \cdot \frac{|I_1|}{p(a)} > \frac{1}{2^{s+3}}.$$

Таким образом, если $p(a) = p_{r',i'}$, то при перезаписи (6.38)–(6.40) в терминах величин $p_{r,i}$, среди этих величин будут встречаться только такие, для которых $r \geq r' - s - 3$.

Лемма 6.3. *Пусть* $s \geq 20$, $d = 2^s$, $\mu = 30 \cdot 2^{-s}$, $\lambda = 20 \cdot 2^{-s/2} + 64 \cdot 2^{-s}$ *и* $\varepsilon = \frac{1}{25 \cdot 2^s}$. *Пусть также* $y_i = i/d$ *и* $v_i = (1-\lambda) \cdot 2^{y_i + 1/2}$. *Тогда*

(*i*) $v = (v_0, \ldots, v_{d-1}) \in T(\sigma[s])$;

(*ii*) $u = v/2 \in T^{\varepsilon}(\sigma[s])$.

Доказательство. Первая часть леммы уже доказана, так как сама стратегия подогнана под заданное решение. Остается проверить (*ii*).

Напомним, что при переходе к канонической системе (6.9) мы переписываем неравенства в терминах переменных $x_{r,i} = p_{r,i} \cdot 2^{-r}$ и выполняем нормировку, а для перехода к приведенной системе (6.10) производится подстановка $x_{r,i} = z_i$.

Пусть $p(a) = p_{r',i'}$. Выполним предварительную нормировку неравенств (6.38)–(6.40) домножением на $2^{-r'}$ и перепишем их в терминах величин $x_{r,i}$. Полученные неравенства обозначим через (6.38')–(6.40'). Оценим сумму абсолютных величин коэффициентов при $x_{r,i}$ в каждом из этих неравенств для определения итоговых коэффициентов нормировки.

Вклад от слагаемого $p(a)$ в каждую сумму коэффициентов равен 1. Тогда вклад любого из слагаемых $p(a - 1/2)$, $p_{1/d}(I_1)$, $p_0(I_{-1})$ и $p_0(H_{k,1})$ не

выше 1, вклад $p(a-s-3/2)$ не превосходит 2^{-s-1}, при любом j вклад слагаемого $p_0(H_{k,j+1})$ вдвое меньше, чем у $p_0(H_{k,j})$. Поскольку $|J_k^*| \le |J_1| \le 2^{a-s}$ согласно (6.22), то вклад каждого из слагаемых $p_0(J_k^*)$ не превосходит 2^{-s}. Эти оценки означают, что суммы коэффициентов при $x_{r,i}$ в неравенствах (6.38'), (6.39') и (6.40') не превосходят соответственно 3, 3 и 5.

Теперь оценим запас, с которым вектор u удовлетворяет приведенным неравенствам (6.38')–(6.40'). По построению, для основного решения v приведенное неравенство (6.38') выполнялось бы и при замене $p_{1/d}(I_1)$ на $p_0(I_1)$ в (6.38). Таким образом, правая часть (6.38) при $p_{r,i} = v_i \cdot 2^r$ превосходит левую на величину не менее

$$p_{1/d}(I_1) - p_0(I_1) \ge |I_1|(2^{1/d}-1) \ge \frac{\ln 2}{d} \cdot |I_1| = (1-\lambda)(\sqrt{2}-1)\ln 2 \cdot 2^{a-s} > 2^{a-s-2}.$$

После нормировки (делением не более чем на $3 \cdot 2^a$) и подстановки u вместо v величина запаса в приведенном неравенстве (6.38') сокращается не более чем в $6 \cdot 2^a$ раз и остается не меньшей чем $\frac{1}{24 \cdot 2^s} > \varepsilon$.

Оценим свободный член, возникающий в (6.39') из-за того, что для сложности вставки в некоторые интервалы J_k^* используется оценка $Q(a)$. Число таких интервалов не меньше $\lfloor L/\Delta \rfloor$, см. выше, а применяемая оценка имеет величину $2^l \ge |J_{2^s}| \ge 2^{a-s-2}$ согласно (6.32) и (6.22). Поэтому, используя значения L и Δ из (6.33), (6.34), заключаем, что свободный член в правой части (6.39') не меньше

$$\frac{L-\Delta}{\Delta \cdot 2^{s+2}} = (\lambda - 2\mu) \cdot \frac{b^3}{2(b+1)(2b+1)} - \frac{1}{2^{s+2}} > \frac{1}{2^{s/2}}.$$

При подстановке u вместо v в приведенное неравенство (6.39') все члены сокращаются вдвое, тогда и свободный член можно уменьшить наполовину, т.е. с учетом нормировки — на $\frac{1}{6 \cdot 2^{s/2}} > \varepsilon$.

Каждое из неравенств (6.40) при $p_{r,i} = v_i \cdot 2^r$ выполняется с заложенным в сумму длин интервалов $H_{k,j}^*$ запасом

$$\delta \cdot |I_1| = (1-\lambda)\delta(\sqrt{2}-1) \cdot 2^a > 2^{a-s+1}/5,$$

см. (6.36). При нормировке (делением не более чем на $5 \cdot 2^a$) и подстановке u вместо v запас сокращается не более чем в $10 \cdot 2^a$ раз, до величины не менее $\varepsilon = \frac{1}{25 \cdot 2^s}$.

Следовательно, при уменьшении свободных членов в приведенных неравенствах (6.38')–(6.40') на ε вектор u остается решением системы. $\square$

6.6 Сортировка

Теорема 6.1. *При любом $a \geq 2$ и $2^{a/4} \leq m \leq 2^{a/2}$ выполнено*

$$P_m(a) \geq 2^{a+1/2} - O\left(a^{-\gamma} \cdot 2^a\right), \tag{6.41}$$

где γ мало, например, $\gamma = \frac{1}{5}$.

Доказательство. Воспользуемся результатом леммы 6.3 и применим лемму 6.2. Значения всех параметров заимствуются из леммы 6.3. Имеем $v \in T(\sigma[s])$ и $u \in T^{\varepsilon}(\sigma[s])$ для системы ограничений $\sigma[s]$ глубины $h = 2s+3$ и ширины $w \leq s+4$. При этом справедливо $u_i < 2^{y_i - 1/2} < \pi(y_i)$, см. (6.5). Кроме того, $\epsilon < 1 - \lambda < R = ||v - u|| < \sqrt{2}$. При $a < 100$ неравенство (6.41) заведомо выполнено при достаточно больших значениях константы под знаком «O». Поэтому считаем, что $a \geq 100$. Положим $\varphi = \log\left(\frac{m\varepsilon}{4a^2}\right)$ и $s = \lceil 2\gamma \log a \rceil$. Если $\gamma \leq 1/4$, то при этом $m\varepsilon \geq 2^{a/4}/(50 \cdot a^{2\gamma}) \geq 4a^2$, следовательно, $\varphi > 0$. Условие (6.16) леммы 6.2 выполнено автоматически.

Пусть $r + y_i \leq a - \frac{(a+2)\varphi \cdot 2^{\varphi}}{m} \leq r + y_i + \frac{1}{d}$, где $r \in \mathbb{N}$. С помощью леммы 6.2 получаем оценку

$$\begin{aligned}
P_m(a) &\geq P_m\left(r + y_i + \frac{(r+2)\varphi \cdot 2^{\varphi}}{m}\right) \geq \\
&\geq v_i \cdot 2^r - Rm \cdot 2^{\varphi+2}/\varepsilon - R \cdot 2^r \cdot e^{-\frac{\varepsilon}{2R}\left(\frac{\varphi}{w(d+1)(h+1)} - 2\right)} \geq \\
&\geq \left(1 - O\left(2^{-s/2}\right)\right) \cdot 2^{a + \frac{1}{2} - \frac{(a+2)\varphi \cdot 2^{\varphi}}{m} - \frac{1}{d}} - O\left(m^2/a^2\right) - O\left(2^a \cdot e^{-\Theta\left(a \cdot s^{-2} \cdot 2^{-2s}\right)}\right) = \\
&= \left(1 - O\left(2^{-s/2}\right)\right) \cdot 2^{a+1/2-O(2^{-s})} - O(2^a/a^2) - 2^{a - O\left(a \cdot s^{-2} \cdot 2^{-2s}\right)} = \\
&= 2^{a+1/2} \cdot \left(1 - O\left(a^{-\gamma}\right)\right),
\end{aligned}$$

если, скажем, $\gamma \leq 1/5$. $\square$

Следствие 6.1. *При любом $a \geq 2$ и $2^{a/4} \leq m \leq 2^{a/2}$ выполнено*

$$Q_{2m}(a) \geq 2^a - O\left(a^{-1/5} \cdot 2^a\right).$$

Реализованное доказательство главного технического результата (теорема 6.1) довольно громоздко. Поэтому для большей ясности дадим неформальное резюме схемы рассуждения.

1. Построим некоторую идеальную серию сравнений, всякий раз выполняя деление множества исходов пополам. К сожалению, она не дает сведения к подобной задаче меньшей размерности.

2. Отступление от идеала (поправочный коэффициент λ) позволяет разбалансировать длины интервалов, используя оценку (6.4) для сложности вставки, близкой к целому числу, и получить сходящийся рекурсивный алгоритм.

3. Дискретизация (параметр d) вводится, чтобы перейти от континуального к конечному семейству алгоритмов. В приведенном доказательстве шаги 2 и 3 совмещены. В результате мы имеем алгоритм с хорошей оценкой сложности, но который рекурсивно сводится к себе самому.

4. Обеспечим прогрессивный рост качества оценки сложности с увеличением размерности задачи, стартуя из точки, в которой оценка тривиально выполняется (лемма 6.2).

С технической стороны, доказательство состоит в контролируемом учете поправок из трех источников: неравномерности разбиения, дискретизации и малой размерности.

Отметим, что даже упрощенный вариант стратегии §6.5 с выбором $s = 1$ и глубиной 6, только не универсальный, а с подбором оптимальных разбиений при каждом i, позволил бы улучшить результаты метода бинарных вставок при всех достаточно больших n. Для получения более точной оценки, однако, приходится выбирать параметр s растущим. Главный содержательный результат формулирует следующая

Теорема 6.2. *При любых n выполняется*

$$S(n) = \log(n!) + O\left(n \log^{-1/5} n\right).$$

Доказательство. Начинаем как в методе бинарных вставок. Разобьем набор из n элементов на пары, упорядочим их и выполним сортировку старших элементов пар (они образуют главную цепочку). На это требуется $n/2 + S(n/2)$ сравнений. Пусть младшие элементы пар обозначаются через

α_i с нумерацией согласно возрастанию ранга старших элементов в главной цепочке, см. фиг. 1.

Вставим первые $n_0 = n/\log n$ из младших элементов в главную цепочку методом бинарных вставок, потратив на это $n + O(n_0)$ сравнений. Оставшиеся элементы разобьем на группы по $m \approx \sqrt{n}/\log n$ штук: группа с номером k содержит элементы $\alpha_{n_0+(k-1)m+1}, \ldots, \alpha_{n_0+km}$. Группы вставляются в главную цепочку по очереди в порядке возрастания номеров методом следствия 6.1. Последняя группа может содержать менее m элементов — их вставляем тривиальным бинарным методом.

Группа с номером t должна быть вставлена в интервал длины $L_t = 2(n_0 + tm) - m$. Согласно следствию 6.1 при некотором $\rho_t = \log L_t + O(\log^{-1/5} n)$ выполнено $Q_m(\rho_t) \geq L_t$. Поэтому сложность вставки t-й группы не превосходит $\rho_t \cdot m$.

Представим идеализированный случай, когда элемент α_k мог бы быть вставлен в главную цепочку за $\log(2k-1)$ сравнений. Тогда, если вставлять элементы в порядке возрастания номеров, было бы затрачено $\log \prod_{k=1}^{\lceil n/2 \rceil}(2k-1) = \log(n!) - \log((n/2)!) - n/2 + O(\log n)$ сравнений.

В групповом методе мы тратим на вставку элемента α_k не более $\log(2k+m) + O(\log^{-1/5} n) = \log(2k-1) + O(\log^{-1/5} n)$ сравнений при $n_0 < k < n/2 - m$ и не более $\log(2k-1) + O(1)$ сравнений при $k \leq n_0$ и $k \geq n/2 - m$. Общее превышение сложности вставки по отношению к идеализированной ситуации тогда можно оценить как

$$O(n_0 + m) + O\left((n/2 - n_0)\log^{-1/5} n\right).$$

Получаем рекуррентное соотношение

$$S(n) \leq n/2 + S(n/2) + \log(n!) - \log((n/2)!) - n/2 + O\left(n \log^{-1/5} n\right),$$

или, если выразить через величины $s(n) = S(n) - \log(n!)$,

$$s(n) \leq s(n/2) + O\left(n \log^{-1/5} n\right).$$

Отсюда следует $s(n) = O\left(n \log^{-1/5} n\right)$. □

7 Заключение

В работе изложены следующие основные результаты автора.

1) Доказана первая нетривиальная нижняя оценка константы равномерности монотонного булева базиса $B_M = \{\vee, \wedge\}$, а именно, для подходящей последовательности функций f_n показано, что $D_{B_M}(f_n) \gtrsim 1.06 \log_2 L_{B_M}(f_n)$.

2) Предложен новый метод реализации симметрических булевых функций над полными базисами, основанный на идеях суммирования по нескольким взаимно простым модулям и приближенного вычисления суммы. Метод позволил уточнить ранее известные оценки глубины и логарифма сложности формул примерно на 10–20%. Как следствие, в частности, для глубины оператора M_n умножения n-разрядных чисел получена неконструктивная оценка $D_{B_2}(M_n) \lesssim 4.02 \log_2 n$ и конструктивная оценка $4.34 \log_2 n$.

3) Предложен новый метод реализации булевыми формулами элементарных периодических симметрических функций (MOD-функций) с малыми периодами, основанный на сведении к задаче о покрытии булевых матриц прямоугольниками. Для таких функций получены новые верхние оценки глубины и сложности.

4) Предложен новый метод нижней оценки сложности булевых функций при реализации формулами в k-арном базисе U_k. В частности, для линейной функции l_n метод позволяет получить оценку сложности, в определенном смысле близкую к окончательной, а именно $L_{U_k}(l_n) = n^{1+\Theta_k(1/\ln k)}$.

5) Доказано, что асимптотика функции Шеннона сложности вентильных (m, n)-схем в общем случае (когда $m = \omega(\log n)$ и $n = \omega(\log m)$) достигается на схемах глубины 3. Этот результат окончательный.

6) Предложен метод трансформации (k, l)-редких (т. е. свободных от сумм $A + B$, $|A| = k$, $|B| = l$) множеств в многомерных пространствах в (k, l)-редкие множества в пространствах меньшей размерности, в частности, одномерных. Метод позволяет строить практически экстремальные эффективные примеры в некоторых задачах получения нижних оценок. Например, многочлены степени n одной переменной, имеющие близкую к

максимально возможной схемную сложность $n^{1-o(1)}$ в монотонном арифметическом базисе $\{+, \times\}$, или k-редкие циркулянтные (n, n)-матрицы веса $n^{2-o(1)}$ при слабо растущем k.

7) Построен первый пример последовательности булевых (n, n)-матриц с растущим отношением XOR- и OR-сложности в глубине 2. Также построены эффективные примеры последовательностей булевых матриц с почти экстремальным отношением $n^{1-o(1)}$ OR-сложности к сложности дополнительных матриц, или с отношением $n^{1/2-o(1)}$ в глубине 2.

8) Установлено точное значение сложности минимальной универсальной (т. е. подходящей для вычислений в любой полугруппе) префиксной схемы глубины n на 2^n входах. Оно имеет величину $(3.5 - o(1))2^n$. Показано, что можно строить параллельные префиксные XOR-схемы с меньшей сложностью, не выше $(36/11 - o(1))2^n$. Также получены верхние оценки сложности префиксных схем при различных ограничениях на глубину.

9) Получены оценки сложности реализации булевых функций n переменных схемами и формулами, использующими многовходовые функциональные элементы конъюнкции и дизъюнкции и либо элементы отрицания, либо отрицания переменных в качестве входов. В ряде случаев эти оценки оказываются асимптотически точными. В частности, для сложности схем со входами переменными и их отрицаниями (AC-схем) доказана асимптотика функции Шеннона $2 \cdot 2^{n/2}$, которая достигается на схемах глубины 3. Предложенный метод синтеза опирается на представляющий самостоятельный интерес результат о существовании специальных покрытий булева куба.

10) Число сравнений, необходимых для сортировки набора из n элементов линейно упорядоченного множества в наихудшем случае, установлено с точностью до величины $o(n)$. Оно равно $\log_2(n!) + o(n)$.

В развитие результатов диссертации могут быть поставлены следующие задачи (перечислим несколько из них):

— получить нетривиальные нижние оценки констант равномерности достаточно выразительных полных базисов, в первую очередь, стандартного базиса B_0;

— построить экономные схемы приближенного суммирования n битов над полными базисами (B_0 или B_2). Особенный интерес представляют конструктивные методы синтеза;

— продолжить исследование пределов величины отношений различных мер сложности линейных схем, в частности, получить оценки для XOR/OR и $\mathsf{OR}_2/\mathsf{OR}$ отношений;

— выяснить, какие алгебраические свойства групповой бинарной операции допускают, а какие не допускают существование параллельных префиксных схем сложности, меньшей чем у универсальных схем. В первую очередь, вопрос относится к свойству коммутативности;

— получить оценки, по возможности, асимптотически точные, для функций Шеннона сложности схем и формул ограниченной глубины из многовходовых элементов конъюнкции и сложения по модулю 2.

Список литературы

[1] Алексеев В. Е. Две конструкции разностных множеств. Проблемы кибернетики. Вып. 38. М.: Наука, 1981, 259–262.

[2] Андреев А. Е. Об одном семействе булевых матриц. Вестник Московского Университета. Серия 1. Математика. Механика. 1986. №2, 97–100.

[3] Андреев А. Е. О сложности реализации вентильными схемами недоопределенных матриц. Математические заметки. 1987. **41**(1), 77–86.

[4] Ахо А., Хопкрофт Дж., Ульман Дж. Проектирование и анализ вычислительных алгоритмов. М.: Мир, 1979.

[5] Гашков С. Б. О глубине булевых функций. Проблемы кибернетики. Вып. 34. М.: Наука, 1978, 265–268.

[6] Гашков С. Б. Об одном методе получения нижних оценок сложности монотонных вычислений многочленов. Вестник Московского университета. Серия 1. Математика. Механика. 1987. №5, 7–13.

[7] Гашков С. Б. О параллельном вычислении некоторых классов многочленов с растущим числом переменных. Вестник Московского университета. Серия 1. Математика. Механика. 1990. №2, 88–92.

[8] Гашков С. Б. Занимательная компьютерная арифметика. Быстрые алгоритмы операций с числами и многочленами. М.: Либроком, 2012.

[9] Гашков С. Б., Кочергин В. В. Об аддитивных цепочках векторов, вентильных схемах и сложности вычисления степеней. Методы дискретного анализа в теории графов и сложности. Вып. 52. Новосибирск: ИМ СО РАН, 1992, 22–40.

[10] Гашков С. Б., Сергеев И. С. О применении метода аддитивных цепочек к инвертированию в конечных полях. Дискретная математика. 2006. **18**(4), 56–72.

[11] Гашков С. Б., Сергеев И. С. О построении схем логарифмической глубины для инвертирования в конечных полях. Дискретная математика. 2008. **20**(4), 8–28.

[12] Гринчук М. И. О сложности реализации циклических булевых матриц вентильными схемами. Известия вузов. Математика. 1988. **7**, 39–44.

[13] Гринчук М. И. О монотонной сложности пороговых функций. Методы дискретного анализа в теории графов и сложности. Вып. 52. Новосибирск: ИМ СО РАН, 1992, 41–48.

[14] Гринчук М. И. Уточнение верхней оценки глубины сумматора и компаратора. Дискретный анализ и исследование операций. Серия 1. 2008. **15**(2), 12–22.

[15] Журавлев Ю. И., Коган А. Ю. Реализация булевых функций с малым числом нулей дизъюнктивными нормальными формами и смежные задачи. Доклады АН СССР. 1985. **285**(4), 795–799.

[16] Зуев Ю. А. Пороговые функции и пороговые представления булевых функций. Математические вопросы кибернетики. Вып. 5. М.: Физматлит, 1994, 5–61.

[17] Зыков К. А. О сложности реализации линейных булевых преобразований схемами глубины три. Вестник Московского университета. Серия 1. Математика. Механика. 1998. №2, 68–70.

[18] Кабатянский Г. А., Панченко В. И. Упаковки и покрытия пространства Хэмминга шарами единичного радиуса. Проблемы передачи информации. 1988. **24**(4), 3–16.

[19] Карацуба А. А., Офман Ю. П. Умножение многозначных чисел на автоматах. Доклады АН СССР. 1962. **145**(2), 293–294.

[20] Картеси Ф. Введение в конечные геометрии. М.: Мир, 1980.

[21] Касим-Заде О. М. Общая верхняя оценка сложности схем в произвольном бесконечном полном базисе. Вестник Московского университета. Серия 1. Математика. Механика. 1997. №4, 59–61.

[22] Касим-Заде О. М. О глубине булевых функций над произвольным бесконечным базисом. Дискретный анализ и исследование операций. Серия 1. 2007. **14**(1), 45–69.

[23] Касим-Заде О. М. О глубине булевых функций при реализации схемами над произвольным бесконечным базисом. Вестник Московского университета. Серия 1. Математика. Механика. 2012. №6, 55–57.

[24] Клосс Б. М., Малышев В. А. Оценки сложности некоторых классов функций. Вестник Московского университета. Серия 1. Математика. Механика. 1965. №4, 44–51.

[25] Кнут Д. Искусство программирования. Т. 2. Получисленные алгоритмы. М.: Вильямс, 2004.

[26] Кнут Д. Искусство программирования. Т. 3. Сортировка и поиск. М.: Вильямс, 2007.

[27] Кнут Д. Искусство программирования. Т. 4, вып. 2. Генерация всех кортежей и перестановок. М.: Вильямс, 2008.

[28] Коспанов Э. Ш. Схемная реализация задачи сортировки. Сибирский журнал исследования операций. 1994. **1**(1), 13–19.

[29] Кочергин А. В. О глубине функций многозначной логики. Дисс. на соискание уч. степ. канд. физ.-мат. наук. М.: МГУ, 2013.

[30] Кочергин В. В. О сложности вычисления систем одночленов с ограничениями на степени переменных. Дискретная математика. 1998. **10**(3), 27–34.

[31] Кочергин В. В. О сложности аддитивных вычислений. Дисс. на соискание уч. степ. докт. физ.-мат. наук. М.: МГУ, 2008.

[32] Кочергин В. В. Теория вентильных схем (современное состояние). Сборник лекций «Дискретная математика и ее приложения». Часть VII. М.: Изд-во ИПМ РАН, 2013, 23–40.

[33] Липатова А. Е. Об одном покрытии множества двоичных наборов и реализации конъюнкций контактными схемами. Математические вопросы кибернетики. Вып. 2. М.: Наука, 1989, 161–173.

[34] Ложкин С. А. Асимптотическое поведение функций Шеннона для задержек схем из функциональных элементов. Математические заметки. 1976. **19**(6), 939–951.

[35] Ложкин С. А. О связи между глубиной и сложностью эквивалентных формул и о глубине монотонных функций алгебры логики. Проблемы кибернетики. Вып. 38. М.: Наука, 1981, 269–271.

[36] Ложкин С. А. О глубине функций алгебры логики в некоторых базисах. Annales Univ. Budapest. Sec. Computatorica. 1983. IV, 113–125.

[37] Ложкин С. А. О глубине функций алгебры логики в произвольном полном базисе. Вестник Московского университета. Серия 1. Математика. Механика. 1996. №2, 80–82.

[38] Ложкин С. А. О синтезе формул, сложность и глубина которых не превосходят асимптотически наилучших оценок высокой степени точности. Вестник Московского университета. Серия 1. Математика. Механика. 2007. №3, 19–25.

[39] Ложкин С. А., Данилов Б. Р. О задержке схем в модели, учитывающей значения на входах функциональных элементов. Вестник Московского университета. Серия 15. Вычислительная математика и кибернетика. 2013. №4, 25–33.

[40] Ложкин С. А., Коноводов В. А. О синтезе и сложности формул с ограниченной глубиной альтернирования. Вестник Московского университета. Серия 15. Вычислительная математика и кибернетика. 2012. №2, 28–36.

[41] Лупанов О. Б. О вентильных и контактно-вентильных схемах. Доклады АН СССР. 1956. **111**(6), 1171–1174.

[42] Лупанов О. Б. О реализации функций алгебры логики формулами из конечных классов (формулами ограниченной глубины) в базисе $\&, \vee, \bar{}$. Проблемы кибернетики. Вып. 6. М.: Физматлит, 1961, 5–14.

[43] Лупанов О. Б. К вопросу о реализации симметрических функций алгебры логики контактными схемами. Проблемы кибернетики. Вып. 15. М.: Наука, 1965, 85–99.

[44] Лупанов О. Б. О схемах из функциональных элементов с задержками. Проблемы кибернетики. Вып. 23. М.: Наука, 1970, 43–81.

[45] Лупанов О. Б. О синтезе схем из пороговых элементов. Проблемы кибернетики. Вып. 26. М.: Наука, 1973, 109–140.

[46] Лупанов О. Б. О сложности универсальной параллельно-последовательной сети глубины 3. Труды МИ АН СССР. 1973. Т. 143. М.: Наука, 1973, 127–131.

[47] Лупанов О. Б. О реализации функций алгебры логики схемами из функциональных элементов «ограниченной глубины» в базисе $\&, \vee, \bar{}$. Сб. работ по математической кибернетике. Т. 2. М.: Изд-во ВЦ АН СССР, 1977, 3–8.

[48] Лупанов О. Б. О вентильных схемах. Acta Cybernetica. 1980. **4**(4), 311–315.

[49] Лупанов О. Б. Асимптотические оценки сложности управляющих систем. М.: Изд-во МГУ, 1984.

[50] Митягин Б. С., Садовский Б. Н. О линейных булевских операторах. Доклады АН СССР. 1965. **165**(4), 773–776.

[51] Мучник Б. А. Оценка сложности реализации линейной функции формулами в некоторых базисах. Кибернетика. 1970. **4**, 29–38.

[52] Нечипорук Э. И. О сложности схем в некоторых базисах, содержащих нетривиальные элементы с нулевыми весами. Проблемы кибернетики. Вып. 8. М.: Физматлит, 1962, 123–160.

[53] Нечипорук Э. И. О вентильных схемах. Доклады АН СССР. 1963. **148**(1), 50–53.

[54] Нечипорук Э. И. О самокорректирующихся вентильных схемах. Доклады АН СССР. 1964. **156**(5), 1045–1048.

[55] Нечипорук Э. И. О синтезе схем из пороговых элементов. Проблемы кибернетики. Вып. 11. М.: Наука, 1964, 49–62.

[56] Нечипорук Э. И. Реферат 1.В.206. Реферативный журнал. Математика. 1967, №1.

[57] Нечипорук Э. И. О топологических принципах самокорректирования. Проблемы кибернетики. Вып. 21. М.: Наука, 1969, 5–102.

[58] Нечипорук Э. И. Об одной булевской матрице. Проблемы кибернетики. Вып. 21. М.: Наука, 1969, 237–240.

[59] Нигматуллин Р. Г. Сложность булевых функций. М.: Наука, 1991.

[60] Орлов В. А. Реализация «узких» матриц вентильными схемами. Проблемы кибернетики. Вып. 22. М.: Наука, 1970, 45–52.

[61] Офман Ю. П. Алгоритмическая сложность дискретных функций. Доклады АН СССР. 1962. **145**(1), 48–51.

[62] Перязев Н. А. Сложность представлений булевых функций формулами в немонолинейных базисах. «Дискретная математика и информатика». Вып. 2. Иркутск: Изд-во Иркут. ун-та, 1995.

[63] Разборов А. А. Нижние оценки размера схем ограниченной глубины в полном базисе, содержащем функцию логического сложения. Математические заметки. 1987. **41**(4), 598–607.

[64] Рохлина М. М. О схемах, повышающих надежность. Проблемы кибернетики. Вып. 23. М.: Наука, 1970, 295–301.

[65] Рычков К. Л. Модификация метода В.М. Храпченко и применение ее к оценкам сложности П-схем для кодовых функций. Методы дискретного анализа в теории графов и схем. Вып. 42. Новосибирск: ИМ СО АН СССР, 1985, 91–98.

[66] Сафин Р. Ф. О соотношении между глубиной и сложностью в предполных классах k-значной логики. Математические вопросы кибернетики. Вып. 13. М.: Физматлит, 2004, 223–278.

[67] Селезнева С. Н. О длине булевых функций в классе полиномиальных форм с аффинными множителями в слагаемых. Вестник Московского университета. Серия 15. Вычислительная математика и кибернетика. 2014. №2, 34–38.

[68] Селезнева С. Н. Порядок длины функций алгебры логики в классе псевдополиномиальных форм. Вестник Московского университета. Серия 15. Вычислительная математика и кибернетика. 2016. №3, 27–31.

[69] Сергеев И. С. О схемах логарифмической глубины для инвертирования в конечных полях характеристики два. Математические вопросы кибернетики. Вып. 15. М.: Физматлит, 2006, 35–64.

[70] Столяров Г. К. Способ параллельного умножения в цифровых вычислительных машинах и устройство для осуществления способа. Авт. свид-во кл. 42 т 14, №126668 (1960).

[71] Субботовская Б. А. О реализации линейных функций формулами в базисе $\vee, \&, {}^{-}$. Доклады АН СССР. 1961. **136**(3), 553–555.

[72] Тарасов П. Б. Об условиях равномерности систем функций многозначной логики. Дисс. на соискание уч. степ. канд. физ.-мат. наук. М.: МГУ, 2016.

[73] Ткачев Г. А. О сложности реализации одной последовательности булевых функций схемами из функциональных элементов и π-схемами при дополнительных ограничениях на структуру схем. Комбинаторно-алгебраические методы в прикладной математике. Горький: изд-во Горьк. ун-та, 1980, 161–207.

[74] Тоом А. Л. О сложности схемы из функциональных элементов, реализующей умножение целых чисел. Доклады АН СССР. 1963. **150**(3), 496–498.

[75] Угольников А. Б. О глубине и полиномиальной эквивалентности формул для замкнутых классов двузначной логики. Математические заметки. 1987. **42**(4), 603–612.

[76] Угольников А. Б. О глубине формул в неполных базисах. Математические вопросы кибернетики. Вып. 1. М.: Наука, 1988, 242–245.

[77] Харди Г., Литлвуд Дж., Полиа Г. Неравенства. М.: ИЛ, 1948.

[78] Холл М. Комбинаторика. М.: Мир, 1970.

[79] Храпченко В. М. Об асимптотической оценке времени сложения параллельного сумматора. Проблемы кибернетики. Вып. 19. М.: Наука, 1967, 107–120.

[80] Храпченко В. М. Об одном методе получения нижних оценок сложности π-схем. Математические заметки. 1971. **10**(1), 83–92.

[81] Храпченко В. М. О сложности реализации симметрических функций формулами. Математические заметки. 1972. **11**(1), 109–120.

[82] Храпченко В. М. О сложности реализации симметрических функций алгебры логики формулами в конечных базисах. Проблемы кибернетики. Вып. 31. М.: Наука, 1976, 231–234.

[83] Храпченко В. М. Некоторые оценки для времени умножения. Проблемы кибернетики. Вып. 33. М.: Наука, 1978, 221–227.

[84] Храпченко В. М. О соотношении между сложностью и глубиной формул. Методы дискретного анализа в синтезе управляющих систем. Вып. 32. Новосибирск: ИМ СО АН СССР, 1978, 76–94.

[85] Храпченко В. М. О соотношении между сложностью и глубиной формул в базисе, содержащем медиану. Методы дискретного анализа в изучении булевых функций и графов. Вып. 37. Новосибирск, ИМ СО АН СССР, 1981, 77–84.

[86] Чашкин А. В. О сложности узких систем булевых функций. Дискретная математика. 1999. **11**(3), 149–159.

[87] Чашкин А. В. Быстрое умножение и сложение целых чисел. «Дискретная математика и ее приложения». Ч. II. М.: изд-во ЦПИ при мех.-мат. ф-те МГУ, 2001, 91–110.

[88] Чашкин А. В. Дискретная математика. М.: Академия, 2012.

[89] Черухин Д. Ю. О сложности реализации линейной функции формулами в конечных булевых базисах. Дискретная математика. 2000. **12**(1), 135–144.

[90] Черухин Д. Ю. Нижние оценки формульной сложности симметрических булевых функций. Дискретный анализ и исследование операций. Серия 1. 2000. **7**(3), 86–98.

[91] Черухин Д. Ю. О реализации линейной функции формулами в различных базисах. Вестник Московского университета. Серия 1. Математика. Механика. 2001. №6, 15–19.

[92] Черухин Д. Ю. О схемах из функциональных элементов конечной глубины ветвления. Дискретная математика. 2006. **18**(4), 73–83.

[93] Эрдеш П., Спенсер Дж. Вероятностные методы в комбинаторике. М.: Мир, 1976.

[94] Яблонский С. В., Козырев В. П. Математические вопросы кибернетики. «Информационные материалы» Научного совета по комплексной проблеме «Кибернетика» АН СССР. Вып. 19а. М., 1968, 3–15.

[95] Яблонский С. В. Введение в дискретную математику. М.: Наука, 1986.

[96] Ajtai M., Komlós J., Szemerédi E. Sorting in $c \log n$ parallel steps. Combinatorica. 1983. **3**(1), 1–19.

[97] Alon N., Rónyai L., Szabó T. Norm-graphs: variations and applications. J. Comb. Theory. Ser. B. 1999. **76**(2), 280–290.

[98] Avizienis A. Signed-digit number representations for fast parallel arithmetic. IRE Trans. on Electr. Computers. 1961. **EC10**, 389–400.

[99] Barak A., Shamir E. On the parallel evaluation of Boolean expressions. SIAM J. Comput. 1976. **5**(4), 678–681.

[100] Beame P. W., Cook S. A., Hoover H. J. Log depth circuits for division and related problems. SIAM J. Comput. 1986. **15**(4), 994–1003. [Рус. перевод: Бим П., Кук С., Гувер Г. Схемы логарифмической глубины для деления и связанных с ним проблем. Кибернетический сборник. Вып. 28. М.: Мир, 1991, 134–150.]

[101] Bini D., Pan V. Polynomial and matrix computations. Vol. 1. Boston: Birkhäuser, 1994.

[102] Blelloch G. E. Prefix sums and their applications. Synthesis of parallel algorithms. San Francisco: Morgan Kaufmann, 1993, 35–60.

[103] Bloniarz P. A. The complexity of monotone Boolean functions and an algorithm for finding shortest paths in a graph. Ph.D. thesis. Tech. Report No. 238, Lab. for Computer Science, MIT, 1979.

[104] Blum N. An $\Omega(n^{4/3})$ lower bound on the monotone network complexity of the n^{th} degree convolution. Theor. Comp. Sci. 1985. **36**, 59–69.

[105] Blum N. On negations in Boolean networks. Lecture Notes Comput. Sci. 2009. **5760**, 18–29.

[106] Bose R. C. An affine analogue of Singer's theorem. J. Indian Math. Soc. (N.S.) 1942, **6**, 1–15.

[107] Boyar J., Find M. Cancellation-free circuits in unbounded and bounded depth. Theoret. Comput. Sci. 2015. **590**, 17–26.

[108] Brauer A. On addition chains. Bull. AMS. 1939. **45**, 736–739.

[109] Brent R. P., Kuck D. J., Maruyama K. The parallel evaluation of arithmetic expressions without division. IEEE Trans. Comp. 1973. **C-22**, 532–534.

[110] Brent R. P. The parallel evaluation of general arithmetic expressions in logarithmic time. J. ACM. 1974. **21**(2), 201–206.

[111] Brent R. Multiple-precision zero-finding methods and the complexity of elementary function evaluation. Analytic computational complexity. NY: Academic Press, 1975, 151–176.

[112] Brown W. G. On graphs that do not contain a Thomsen graph. Canad. Math. Bull. 1966. **9**, 281–285. [Рус. перевод: Браун У. Г. Графы, не содержащие графа Томсена. Кибернетический сборник. Вып. 18. М.: Мир, 1981, 34–38.]

[113] Chin A. On the depth complexity of the counting functions. Inf. Proc. Letters. 1990. **35**, 325–328.

[114] Chockler H., Zwick U. Which bases admit non-trivial shrinkage of formulae? Comput. Complexity. 2001. **10**, 28–40.

[115] Christen C. Improving the bounds for optimal merging. Proc. 19th IEEE Conf. on Found. of Comput. Sci. (Ann Arbor, USA, 1978). NY: IEEE, 1978, 259–266.

[116] Commentz-Walter B. Size-depth tradeoff in monotone Boolean formulae. Acta Inf. 1979. **12**, 227–243.

[117] Commentz-Walter B., Sattler J. Size-depth tradeoff in non-monotone Boolean formulae. Acta Inf. 1980. **14**, 257–269.

[118] Cooley J. W., Tukey J. W. An algorithm for the machine calculation of complex Fourier series. Math. Comp. 1965. **19**(90), 297–301.

[119] Cooper J. N., Ellis R. B., Kahng A. B. Asymmetric binary covering codes. J. Comb. Theory, Ser. A. 2002. **100**, 232–249.

[120] Coppersmith D., Schieber B. Lower bounds on the depth of monotone arithmetic computations. Proc. 33rd Symp. on Found. of Comput. Sci. (Pittsburgh, 1992). Washington: IEEE CS, 1992, 288–295.

[121] Dančík V. Complexity of Boolean functions with unbounded fan-in gates. Inf. Proc. Letters. 1996. **57**, 31–34.

[122] Demenkov E., Kojevnikov A., Kulikov A., Yaroslavtsev G. New upper bounds on the Boolean circuit complexity of symmetric functions. Inf. Proc. Letters. 2010. **110**(7), 264–267.

[123] Dor D., Zwick U. Selecting the median. SIAM J. Comput. 1999. **28**(5), 1722–1758.

[124] Dunne P. E. The complexity of Boolean networks. San Diego: Academic Press, 1988.

[125] Edelkamp S., Weiß A., Wild S. QuickXsort: a fast sorting scheme in theory and practice. Algorithmica. 2020. **82**, 509–588.

[126] Fich F. E. Two problems in concrete complexity: cycle detection and parallel prefix computation. IBM research report RJ 3651, 1982. (Ph.D. thesis. Univ. of California, Berkeley, 1982.)

[127] Fich F. E. New bounds for parallel prefix circuits. Proc. 15th Symp. on Theory of Comput. (Boston, 1983). NY: ACM, 1983, 100–109.

[128] Find M., Göös M., Järvisalo M., Kaski P., Koivisto M., Korhonen J. H. Separating OR, SUM, and XOR circuits. J. Computer System Sci. 2016. **82**(5), 793–801.

[129] Fischer M. J., Meyer A. R., Paterson M. S. $\Omega(n \log n)$ lower bounds on length of Boolean formulas. SIAM J. Comput. 1982. **11**(3), 416–427.

[130] Ford L. R., Johnson S. M. A tournament problem. Amer. Math. Monthly. 1959. **66**(5), 387–389.

[131] Furst M., Saxe J., Sipser M. Parity, circuits, and the polynomial time hierarchy. Math. Syst. Theory. 1984. **17**, 13–27.

[132] Gál A., Hansen K. A., Koucký M., Pudlák P., Viola E. Tight bounds on computing error-correcting codes by bounded-depth circuits with arbitrary gates. Proc. 44th Symp. on Theory of Comput. (New York, 2012). NY: ACM, 2012, 479–494.

[133] Good I. J. The interaction algorithm and practical Fourier analysis. J. R. Statist. Soc. B. 1958. **20**(2), 361–372; 1960. **22**(2), 372–375.

[134] Graham R. L. On sorting by comparisons. Computers in Number Theory. London: Academic Press, 1971, 263–269.

[135] Grigoriev D., Razborov A. Exponential lower bounds for depth 3 arithmetic circuits in algebras of functions over finite fields. Appl. Algebra Eng. Comm. Comput. 2000. **10**(6), 465–487.

[136] Grove E. Proofs with potential. Ph.D. thesis. Univ. of California, Berkeley, 1993.

[137] Gupta A., Mahajan S. Using amplification to compute majority with small majority gates. Comput. Complexity. 1996. **6**, 46–63.

[138] Hajnal A., Maass W., Turán G. On the communication complexity of graph properties. Proc. 20th Symp. on Theory of Comput. (Chicago, 1988). NY: ACM, 1988, 186–191.

[139] Harvey D., van der Hoeven J., Lecerf G. Faster polynomial multiplication over finite fields. J. ACM. 2017. **63**(6), Article 52.

[140] Håstad J. Computational limitations of small-depth circuits. MIT Press, 1986.

[141] Hwang F. K., Lin S. Optimal merging of 2 elements with n elements. Acta Inf. 1971. **1**, 145–158.

[142] Iwama K., Teruyama J. Improved average complexity for comparison-based sorting. Theor. Comput. Sci. 2020. **807**, 201–219.

[143] Jukna S. Disproving the single level conjecture. SIAM J. Comput. 2006. **36**(1), 83–98.

[144] Jukna S. Extremal combinatorics: with applications in computer science. Berlin, Heidelberg: Springer–Verlag, 2011.

[145] Jukna S. Boolean function complexity. Berlin, Heidelberg: Springer–Verlag, 2012.

[146] Jukna S., Sergeev I. Solution of problem 7.7. Comments to [230]. http://lovelace.thi.informatik.uni-frankfurt.de/~jukna/Knizka/problem-7.7.html (2017).

[147] Jukna S., Sergeev I. SUM-complexity gaps for matrices and their complements. Comments to [230]. http://lovelace.thi.informatik.uni-frankfurt.de/~jukna/Knizka/sum-gaps.html (2021).

[148] Karchmer M., Wigderson A. Monotone circuits for connectivity require super-logarithmic depth. SIAM J. Discrete Math. 1990. **3**(2), 255–265.

[149] Katz N. H. On the CNF-complexity of bipartite graphs containing no squares. Lithuanian Math. Journal. 2012. **52**(4), 385–389.

[150] Kogge P. M., Stone H. S. A parallel algorithm for the efficient solution of a general class of recurrence equations. IEEE Trans. on Comp. 1973. **22**(8), 786–793.

[151] Kojevnikov A., Kulikov A. S., Yaroslavtsev G. Finding efficient circuits using SAT-solvers. Proc. 12th Intern. Conf. on Theory and Appl. of Satisf. Testing (Swansea, Wales, 2009). Lecture Notes on Comput. Sci. 2009. **5584**, 139–157. [На рус. яз.: Кожевников А. А., Куликов А. С., Ярославцев Г. Н. Схемная сложность MOD-функций. Препринт ПОМИ. 2008. №18.]

[152] Kóllar J., Rónyai L., Szabó T. Norm-graphs and bipartite Turán numbers. Combinatorica. 1996. **16**(3), 399–406.

[153] Kosaraju S. R. Parallel evaluation of division-free arithmetic equations. Proc. 18th Symp. on Theory of Comput. (Berkeley, 1986). NY: ACM, 1986, 231–239.

[154] Kövari T., Sós V. T., Turán P. On a problem of K. Zarankiewicz. Coll. Math. 1954. **3**, 50–57.

[155] Kulikov A., Mikhailin I., Mokhov A., Podolskii V. Complexity of linear operators. Electr. Colloq. on Comput. Complexity. 2019. TR19–002.

[156] Ladner R. E., Fischer M. J. Parallel prefix computation. J. ACM. 1980. **27**(4), 831–838.

[157] van Leijenhorst D. C. A note on the formula size of the "mod k" functions. Inf. Proc. Letters. 1987. **24**, 223–224.

[158] Lindström B. Determination of two vectors from the sum. J. Comb. Theory. 1969. **6**, 402–407.

[159] Manacher G. K., Bui T. D., Mai T. Optimum combinations of sorting and merging. J. ACM. 1989. **36**(2), 290–334.

[160] Mantel W. Problem 28. Wiskundige Opgaven. 1907. **10**, 60–61.

[161] McColl W. F. Some results on circuit depth. Theory of computation. Report No. 18. Coventry, Univ. of Warwick, 1977.

[162] McColl W. F. The circuit depth of symmetric Boolean functions. J. of Comput. and System Sci. 1978. **17**, 108–115.

[163] Mehlhorn K. Some remarks on boolean sums. Acta Inf. 1979. **12**, 371–375. [Рус. перевод: Мельхорн К. Некоторые замечания, касающиеся булевых сумм. Кибернетический сборник. Вып. 18. М.: Мир, 1981, 39–45.]

[164] Mehlhorn K. Data structures and algorithms. Vol. 1. Sorting and searching. Berlin, NY: Springer, 1984.

[165] Muller D. E., Preparata F. P. Restructuring of arithmetic expressions for parallel evaluation. J. ACM. 1976. **23**(3), 534–543. [Рус. перевод: Мюллер Д. Е., Препарата Ф. П. Перестроение арифметических выражений для параллельного вычисления. Кибернетический сборник. Вып. 16. М.: Мир, 1979, 5–22.]

[166] O'Bryant K. A complete annotated bibliography of work related to Sidon sequences. Elect. J. Combinatorics. 2004. Dynamic survey 11.

[167] Paterson M. S. New bounds on formula size. Proc. 3rd GI-Conf. on Theory of Comput. Sci. (Darmstadt, 1977). Lecture Notes on Comput. Sci. 1977. **48**, 17–26.

[168] Paterson M. S., Pippenger N., Zwick U. Faster circuits and shorter formulae for multiple addition, multiplication and symmetric Boolean functions. Proc. 31st Symp. on Found. of Comput. Sci. (St. Louis, 1990). Washington: IEEE, 1990, 642–650.

[169] Paterson M. S., Pippenger N., Zwick U. Optimal carry save networks. LMS Lecture Notes Series. Boolean function complexity. Vol. 169. Cambridge University Press, 1992, 174–201.

[170] Paterson M., Zwick U. Shallow multiplication circuits. Proc. 10th IEEE Symp. on Comp. Arithm. (Grenoble, 1991). IEEE, 1991, 28–34.

[171] Paterson M., Zwick U. Shallow circuits and concise formulae for multiple addition and multiplication. Comput. Complexity. 1993. **3**, 262–291.

[172] Peczarski M. The Ford-Johnson algorithm still unbeaten for less than 47 elements. Inf. Process. Lett. 2007. **101**(3), 126–128.

[173] Peterson G. L. An upper bound on the size of formulae for symmetric Boolean function. Tech. Report 78–03–01. Univ. Washington, 1978.

[174] Pinto T. Biclique covers and partitions. Electr. J. Combinatorics. 2014. **21**(1), 1–19.

[175] Pippenger N. Short formulae for symmetric functions. IBM report RC–5143. Yorktown heights, NY, 1974.

[176] Pippenger N. The minimum number of edges in graphs with prescribed paths. Math. System Theory. 1979. **12**, 325–346.

[177] Pippenger N. On another Boolean matrix. Theor. Comput. Sci. 1980. **11**, 49–56.

[178] Pippenger N. On the evaluation of powers and monomials. SIAM J. Comput. 1980. **9**(2), 230–250.

[179] Pratt V. R. The effect of basis on size of Boolean expressions. Proc. 16th Symp. on Found. of Comput. Sci. (New York, 1975). NY: IEEE, 1975, 119–121. [Рус. перевод: Пратт В. Р. Влияние базиса на сложность булевых формул. Кибернетический сборник. Вып. 17. М.: Мир, 1980, 114–123.]

[180] Preparata F. P., Muller D. E. The time required to evaluate division-free arithmetic expressions. Inf. Proc. Letters. 1975. **3**(5), 144–146.

[181] Preparata F. P., Muller D. E. Efficient parallel evaluation of Boolean expressions. IEEE Trans. Comp. 1976. **C-25**(5), 548–549.

[182] Preparata F. P., Muller D. E., Barak A. B. Reduction of depth of Boolean networks with a fan-in constraint. IEEE Trans. Comp. 1977. **C-26**(5), 474–479.

[183] Pudlák P. Communication in bounded depth circuits. Combinatorica. 1994. **14**(2), 203–216.

[184] Pudlák P., Rödl V. Some combinatorial-algebraic problems from complexity theory. Discrete Math. 1994. **136**(1–3), 253–279.

[185] Pudlák P., Rödl V. Pseudorandom sets and explicit constructions of Ramsey graphs. Quad. Mat. 2004. **13**, 327–346.

[186] Pudlák P., Vavřín Z. Computation of rigidity of order n^2/r for one simple matrix. Comm. Math. Univ. Carol. 1991. **32**(2), 213–218.

[187] Ragaz M. Parallelizable algebras. Arch. math. Logic. 1986/87. **26**, 77–99.

[188] Raz R., Wigderson A. Monotone circuits for matching require linear depth. J. ACM. 1992. **39**(3), 736–744.

[189] Razborov A., Wigderson A. $n^{\Omega(\log n)}$ lower bounds on the size of depth-3 threshold circuits with AND gates at the bottom. Inform. Proc. Letters. 1993. **45**, 303–307.

[190] Rosser J., Schoenfeld L. Approximate formulas for some functions of prime numbers. Ill. J. Math. 1962. **6**, 64–94.

[191] Schönhage A. Schnelle Multiplikation von Polynomen über Körpern der Charakteristik 2. Acta Inf. 1977. **7**, 395–398.

[192] Schönhage A., Paterson M., Pippenger N. Finding the median. J. Comp. Sys. Sci. 1976. **13**, 184–199.

[193] Schönhage A., Strassen V. Schnelle Multiplikation großer Zahlen. Computing. 1971, **7**(3–4), 271–282. [Рус. перевод: Шёнхаге А., Штрассен В. Быстрое умножение больших чисел. Кибернетический сборник. Вып. 10. М.: Мир, 1973, 87–98.]

[194] Schulte Mönting J. Merging of 4 or 5 elements with n elements. Theor. Comput. Sci. 1981. **14**, 19–37.

[195] Selezneva S. N. On the multiplicative complexity of Boolean functions. Fundamenta Informaticae. 2016. **145**, 399–404.

[196] Shamir E., Snir M. On the depth complexity of formulas. Math. Syst. Theory. 1979/80. **13**(4), 301–322. [Рус. перевод: Шамир Э., Шнир М. О глубине формул. Кибернетический сборник. Вып. 19. М.: Мир, 1983, 71–96.]

[197] Sheeran M. Functional and dynamic programming in the design of parallel prefix networks. J. Funct. Programming. 2010. **21**(1), 59–114.

[198] Singer J. A theorem in finite projective geometry and some applications to number theory. Trans. Amer. Math. Soc. 1938. **43**, 377–385.

[199] Sinha R. K., Thathachar J. S. Efficient oblivious branching programs for threshold and Mod functions. J. Comput. and System Sci. 1997. **55**(3), 373–384.

[200] Sklansky J. Conditional-sum addition logic. IRE Trans. Electr. Comput. 1960. **EC-9**, 226–231.

[201] Snir M. Depth-size trade-offs for parallel prefix computation. J. Algorithms. 1986. **4**, 185–201.

[202] Spira P. M. On time-hardware complexity tradeoffs for Boolean functions. Proc. 4th Hawaii Symp. System Sci. N. Hollywood: Western Periodical Company, 1971, 525–527.

[203] Stockmeyer L. J. On the combinational complexity of certain symmetric Boolean functions. Math. Syst. Theory. 1977. **10**, 323–336. [Рус. перевод: Стокмейер Л. Дж. О комбинационной сложности некоторых симметрических булевых функций. Кибернетический сборник. Вып. 16. М.: Мир, 1979, 45–61.]

[204] Tsukiji T. On a small class of Boolean sums. Theor. Comput. Sci. 1996. **163**, 283–289.

[205] Ueno K. Formula complexity of ternary majorities. Proc. Int. Computing and Combinatorics Conf. (Sydney, Australia, 2012). Lecture Notes in Comput. Sci. 2012. **7434**, 433–444.

[206] Valiant L. G. Graph-theoretic methods in low-level complexity. Lecture Notes on Comput. Sci. 1977. **53**, 162–176.

[207] Valiant L. G. Short monotone formulae for the majority function. J. Algorithms. 1984. **5**, 363–366. [Рус. перевод: Вэльянт Л. Простые монотонные формулы для функции голосования. Кибернетический сборник. Вып. 24. М.: Мир, 1987, 97–100.]

[208] Wegener I. A new lower bound on the monotone network complexity of Boolean sums. Acta Inf. 1980. **15**, 147–152.

[209] Wegener I. The complexity of Boolean functions. Stuttgart: Wiley–Teubner, 1987.

[210] Wegener I. Complexity theory. Berlin, Heidelberg: Springer–Verlag, 2005.

[211] Weiß J. An $n^{3/2}$ lower bound on the monotone network complexity of the Boolean convolution. Inf. and Control. 1983. **59**, 184–188.

[212] Yao A. C. Some complexity questions related to distributed computing. Proc. 11th Symp. on Theory of Computing (Atlanta, 1979). NY: ACM, 1979, 209–213.

[213] Yao A. C. Separating the polynomial time hierarchy by oracles. Proc. 26th Symp. on Found. of Comput. Sci. (Portland, 1985). Washington: IEEE, 1985, 1–10.

[214] Zhu H., Cheng C.-K., Graham R. On the construction of zero-deficiency parallel prefix circuits with minimum depth. ACM Trans. on Design Autom. of Electr. Syst. 2006. **11**(2), 387–409.

Работы автора по теме диссертации

Статьи в рецензируемых научных изданиях, рекомендованных для защиты в диссертационных советах МГУ

[215] Гашков С. Б., Сергеев И. С. О сложности линейных булевых операторов с редкими матрицами. Дискретный анализ и исследование операций. 2010. **17**(3), 3–18.

[216] Гринчук М. И., Сергеев И. С. Редкие циркулянтные матрицы и нижние оценки сложности некоторых булевых операторов. Дискретный анализ и исследование операций. 2011. **18**(5), 38–53.

[217] Сергеев И. С. О минимальных параллельных префиксных схемах. Вестник Московского университета. Серия 1. Математика. Механика. 2011. №5, 48–51.

[218] Гашков С. Б., Сергеев И. С. Об одном методе получения нижних оценок сложности монотонных арифметических схем, вычисляющих действительные многочлены. Математический сборник. 2012. **203**(10), 33–70.

[219] Сергеев И. С. Верхние оценки глубины симметрических булевых функций. Вестник Московского университета. Серия 15. Вычислительная математика и кибернетика. 2013. №4, 39–44.

[220] Сергеев И. С. Верхние оценки сложности формул для симметрических булевых функций. Известия высших учебных заведений. Математика. 2014. №5, 38–52.

[221] Сергеев И. С. О сложности и глубине формул для симметрических булевых функций. Вестник Московского университета. Серия 1. Математика. Механика. 2016. №3, 53–57.

[222] Сергеев И. С. Верхние оценки сложности и глубины формул для MOD-функций. Дискретная математика. 2016. **28**(2), 108–116.

[223] Сергеев И. С. Вентильные схемы ограниченной глубины. Дискретный анализ и исследование операций. 2018. **25**(1), 120–141.

[224] Сергеев И. С. О сложности схем и формул ограниченной глубины над базисом из многовходовых элементов. Дискретная математика. 2018. **30**(2), 120–137.

[225] Сергеев И. С. О соотношении между глубиной и сложностью монотонных булевых формул. Дискретный анализ и исследование операций. 2019. **26**(4), 108–120.

[226] Сергеев И. С. О сложности монотонных схем для пороговых симметрических булевых функций. Дискретная математика. 2020. **32**(1), 81–109.

[227] Сергеев И. С. Многоярусное представление и сложность схем из многовходовых элементов. Вестник Московского университета. Серия 1. Математика. Механика. 2020. №3, 42–46.

[228] Сергеев И. С. О верхней границе сложности сортировки. Журнал вычислительной математики и математической физики. 2021. **61**(2), 345–362.

[229] Сергеев И. С. Формульная сложность линейной функции в k-арном базисе. Математические заметки. 2021. **109**(3), 419–435.

Статьи в зарубежных рецензируемых изданиях

[230] Jukna S., Sergeev I. Complexity of linear boolean operators. Foundations and Trends in Theoretical Computer Science. 2013. **9**(1), 1–123.

Прочие работы

[231] Сергеев И. С. О глубине схем для многократного сложения и умножения чисел. Материалы VI молодежной научной школы по дискретной математике и ее приложениям (Москва, 2007 г.). Ч. II. М.: Изд-во ИПМ РАН, 2007, 40–45.

[232] Гашков С. Б., Сергеев И. С. О сложности булевых линейных операторов с редкими матрицами. Материалы X Международного семинара «Дискретная математика и ее приложения» (Москва, 2010 г.). М.: Изд-во мех.-мат. фак-та МГУ, 2010, 100–102.

[233] Сергеев И. С. Некоторые оценки сложности параллельных префиксных схем. Материалы X Международного семинара «Дискретная математика и ее приложения» (Москва, 2010 г.). М.: Изд-во мех.-мат. фак-та МГУ, 2010, 136–139.

[234] Sergeev I. S. Upper bounds for the formula size of the majority function. 2012. arXiv:1208.3874. [Там же рус. перевод: Сергеев И. С. Верхние оценки сложности формул для функции голосования.]

[235] Sergeev I. S. On the complexity of parallel prefix circuits. Electronic Colloquium on Computational Complexity. 2013. TR13–041. [Рус. перевод: http://istina.msu.ru/publications/article/3490078]

[236] Sergeev I. S. Implementation of linear maps with circulant matrices via modulo 2 rectifier circuits of bounded depth. 2013. arXiv:1305.4389. [Там

же рус. перевод: Сергеев И. С. Реализация линейных преобразований с циркулянтными матрицами вентильными схемами по модулю 2 ограниченной глубины.]

[237] Сергеев И. С. Верхние оценки сложности и глубины симметрических булевых функций. Материалы IX молодежной научной школы по дискретной математике и ее приложениям (Москва, 2013 г.). М.: Изд-во ИПМ РАН, 2013, 100–103.

[238] Сергеев И. С. О сравнительной сложности реализации булевой матрицы и ее дополнения вентильными схемами. Материалы XVII Международной конференции «Проблемы теоретической кибернетики» (Казань, 2014 г.). Казань, Отечество, 2014, 262–264. [Eng. transl.: Sergeev I. S. On relative OR-complexity of Boolean matrices and their complements. 2014. arXiv:1407.4626.]

[239] Сергеев И. С. О сложности и глубине формул для MOD-функций. Материалы X молодежной научной школы по дискретной математике и ее приложениям (Москва, 2015 г.). М.: Изд-во ИПМ РАН, 2015, 61–65.

[240] Сергеев И. С. Сложность схем и формул ограниченной глубины над базисом из многовходовых элементов. Труды X Международной конференции «Дискретные модели в теории управляющих систем» (Москва, 2018 г.). М.: Макс-пресс, 2018, 245–247.

[241] Сергеев И. С. О сложности монотонных схем для симметрических пороговых функций. Материалы XIII Международного семинара «Дискретная математика и ее приложения» (Москва, 2019 г.). М.: Изд-во мех.-мат. фак-та МГУ, 2019, 140–142.

[242] Сергеев И. С. Об асимптотической сложности сортировки. Материалы заочного семинара XIX Международной конференции «Проблемы теоретической кибернетики». Казань, 2020, 119–122.

[243] Sergeev I. S. On the asymptotic complexity of sorting. Electronic Colloquium on Computational Complexity. 2020. TR20–096.